

[19] 中华人民共和国国家知识产权局

[51] Int. Cl⁷

C12N 15/57

C12N 9/52 C12P 21/02



[12] 发明专利说明书

[21] ZL 专利号 98805990.8

[45] 授权公告日 2005 年 2 月 23 日

[11] 授权公告号 CN 1190494C

[22] 申请日 1998.6.4 [21] 申请号 98805990.8

[30] 优先权

[32] 1997.6.10 [33] JP [31] 151969/1997

[86] 国际申请 PCT/JP1998/002465 1998.6.4

[87] 国际公布 WO1998/056926 日 1998.12.17

[85] 进入国家阶段日期 1999.12.9

[71] 专利权人 宝生物工程株式会社

地址 日本大津市

[72] 发明人 高藏晃 森下美绪 下条智子

浅田起代藏 加藤郁之进

审查员 李 珊

[74] 专利代理机构 中国专利代理(香港)有限公司

代理人 卢新华 谭明胜

权利要求书 1 页 说明书 60 页 附图 7 页

[54] 发明名称 表达超热稳定蛋白酶的系统

[57] 摘要

一种具有由序列表 SEQ ID NO: 1 表示的氨基酸序列或经缺失、替换、插入或添加一个至几个氨基酸残基从其产生的序列的超热稳定蛋白酶，一种编码该超热稳定蛋白酶的基因和一种制备该蛋白酶的方法，目的是通过遗传工程技术提供一种有利于工业应用的超热稳定的蛋白酶。

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

1. 一种由如下氨基酸序列组成并具有热稳定蛋白酶活性的蛋白酶，在该氨基酸序列中，一个或多个氨基酸残基从由序列表 SEQ ID NO: 4 表示的氨基酸序列的 C-末端缺失，其中所述蛋白酶含有由 SEQ ID NO: 1 表示的氨基酸序列。

2. 根据权利要求 1 的蛋白酶，它由序列表 SEQ ID NO: 1 表示的氨基酸序列组成。

3. 一种由如下氨基酸序列组成并具有热稳定蛋白酶活性的蛋白酶，在该氨基酸序列中，一至几个氨基酸残基被缺失、替换、插入或添加到由序列表 SEQ ID NO: 1 表示的氨基酸序列中。

4. 一种编码由如下氨基酸序列组成并具有热稳定蛋白酶活性的蛋白的基因，在该氨基酸序列中，从序列表 SEQ ID NO: 4 表示的氨基酸序列的 C-末端缺失一个或多个氨基酸残基，其中所述蛋白酶含有由 SEQ ID NO: 1 表示的氨基酸序列。

5. 根据权利要求 4 的蛋白酶基因，它编码由序列表 SEQ ID NO: 1 表示的氨基酸序列。

6. 根据权利要求 5 的蛋白酶基因，它由序列表 SEQ ID NO: 2 表示的碱基序列组成。

7. 一种编码由如下氨基酸序列组成并具有热稳定蛋白酶活性的蛋白的蛋白酶基因，在该氨基酸序列中，由序列表 SEQ ID NO: 1 表示的氨基酸序列缺失、替换、插入或添加一个至几个氨基酸残基。

8. 一种与根据权利要求 6 的蛋白酶基因在严格条件下杂交且编码具有热稳定蛋白酶活性的蛋白的蛋白酶基因。

25

表达超热稳定蛋白酶的系统

技术领域

- 5 本发明涉及一种在工业应用中作为酶使用的超热稳定蛋白酶，编码该酶的基因以及通过遗传工程技术产生该酶的方法。

技术背景

- 10 蛋白酶是一种切割蛋白中肽键的酶。许多上述酶已在动物、植物和微生物中发现。蛋白酶可作为实验室使用的试剂和作为药物以及在工业领域，如作为添加剂用于去垢剂，用于处理食品及用于利用逆反应进行化学合成。因此，可以说该蛋白酶是工业上极其重要的酶。由于用于工业领域的蛋白酶需要高度的物理和化学稳定性，在其它酶类中热稳定酶是优选使用的。因为芽孢杆菌属的细菌产生的蛋白酶表现为相对高的热稳定性，它们主要作为工业用途的蛋白酶。然而，在寻找更为优异的酶中，想方设法从高温下生长的微生物中获得该酶，例如，从芽孢杆菌属的嗜热细菌或超嗜热菌中获得。

- 例如，熟知的超嗜热菌激烈热球菌产生一种蛋白酶（应用环境微生物学，56：1992 - 1998（1990）；FEMS微生物学通讯，71：17 - 20（1990）；遗传微生物学杂志，137：1193 - 1199（1991））。

此外，据报道嗜热菌热球菌属菌株 KOD1 产生一种硫醇基蛋白酶（半胱氨酸蛋白酶）（应用环境微生物学，60：4559 - 4566（1994））。也已知超嗜热菌热球菌属、葡萄嗜热菌属、栖热拟芽孢杆菌属产生蛋白酶（应用微生物学生物技术，34：715 - 719（1991））。

- 25 如上面描述的来自嗜热菌的蛋白酶具有高度热稳定性。因此，可预计它们可用于代替目前正使用的热稳定蛋白酶或用于还未考虑使用蛋白酶的领域。

- 然而，大部分产生这些酶的微生物仅在高温下生长。例如，激烈热球菌需要在 90 - 100 °C 培养。在上述高温下培养从能量消耗的角度看是不利的。此外，来自超嗜热菌的蛋白酶生产力比常规微生物蛋白酶的生产力低。所以，工业上从超嗜热菌产生蛋白酶的方法存在问题。

另外，用遗传工程技术通过分离目的酶的基因并将其导入到容易培养的宿主微生物中产生该酶的方法，是本领域目前的常规方法。然而，该酶的基因导入到宿主中并不总是如预期的那样有效地表达。据信主要原因是导入基因的 GC 含量或密码子使用与宿主的基因不同。

- 5 因此，需要对导入的每个基因和/或每个宿主优化表达方法，以便达到符合设计用途的酶合适生产力。

本发明的目的

- 10 本发明的目的是提供工业使用有利的来自嗜热菌的蛋白酶，从该嗜热菌中分离编码该蛋白酶的基因和提供利用该基因通过遗传工程技术产生超热稳定性蛋白酶的方法，以便解决上面所述的问题。

本发明概述

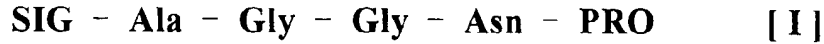
- 15 由嗜热菌产生的蛋白酶中，一些种类根据其氨基酸序列的同源性分类为枯草杆菌蛋白酶型碱性蛋白酶。当编码该蛋白酶的基因被导入到枯草芽孢杆菌中时（枯草芽孢杆菌一般通过遗传工程技术用于生产），该酶的生产力比由枯草芽孢杆菌内源产生的蛋白的生产力更低。

- 20 本发明者经深入的研究并发现，通过把编码来源于枯草杆菌蛋白酶信号肽的基因（信号序列）置于来源于嗜热菌待表达的蛋白酶基因的上游，然后修饰切割位点周围的氨基酸序列，该目的基因将在枯草芽孢杆菌中高效表达。此外，还发现通过把来源于目的嗜热菌的蛋白酶基因中酶活性非必需的部分缺失、可增加所述的酶的表达水平。至此，完成了本发明。

- 25 本发明概括如下。本发明的第一个发明是一种具有由序列 SEQ ID NO：1 给出的氨基酸序列的热稳定蛋白酶和一种具有如下氨基酸序列及热稳定蛋白酶活性的蛋白酶，在该氨基酸序列中，一个或几个氨基酸残基被缺失、替换、插入或添加到序列 SEQ ID NO：1 给出的氨基酸序列中形成。

- 30 本发明的第二个发明是一种编码第一个发明的热稳定蛋白酶的基因和一种能与该基因杂交的热稳定蛋白酶基因。

本发明的第三个发明是通过遗传工程技术将来源于嗜热菌的基因用于生产热稳定蛋白酶，其特征是该基因编码如通式 I 表示的氨基酸序列：



5 其中，SIG 代表来源于枯草杆菌蛋白酶信号肽的氨基酸序列，PRO 代表待表达的蛋白的氨基酸序列。优选地是，SIG 是由序列表 SEQ ID NO: 3 给出的氨基酸序列。优选地是，PRO 是来源于嗜热菌的嗜热稳定蛋白酶的氨基酸序列，更优选地是，来源于激烈热球菌的蛋白酶的氨基酸序列。

10 本发明的第四个发明涉及一种通过遗传工程技术产生蛋白质的方法，其特征在于该方法包括培养已导入第三个发明的基因的芽孢杆菌属细菌，然后从培养物中收集目的蛋白。

本发明的第五个发明是用于通过遗传工程技术生产蛋白的一种质粒，其特征在于第三个发明的基因已插入到质粒中。

15 在氨基酸序列中一个至几个氨基酸残基诸如缺失、替换、插入或添加的突变，可在天然产生的蛋白及由本发明披露的蛋白中产生。上述突变可能由于编码该蛋白的基因的多态性或突变产生，或者由于蛋白的体内修饰或合成后纯化过程中产生。不过，已知上述突变蛋白表现的生理学和生物学活性与未突变的蛋白相等。这可用于如下蛋白中，在该蛋白
20 中上述突变已人为地导入到其氨基酸序列中，在该情况下，可能产生大量的各种突变。例如，众所周知的在人白介素 - 2 (IL - 2) 的氨基酸序列中的半胱氨酸残基被丝氨酸残基替换的多肽仍保持白介素 - 2 的活性 (科学, 224: 1431 (1984))。因此，具有在本发明披露的氨基酸序列中一个或几个氨基酸残基缺失、替换、插入或添加而形成的氨基
25 酸序列及其蛋白酶活性与本发明的蛋白酶相等的蛋白酶也在本发明的范围之内。

至于本文所用的“(与特定的基因)杂交的基因”是一种具有的碱基序列与特定基因的序列相似的基因。有可能具有的碱基序列与特定基因的序列相似的基因编码一种具有氨基酸序列的蛋白，其功能与由特定
30 基因编码的蛋白相似。基因碱基序列的相似性可通过在严格条件下确定基因或其部分互相之间是否形成杂交体(杂交)而检查出来。通过使用

该方法，能获得所编码的蛋白与特定基因编码的蛋白具有相似功能的基因。这就是说具有与本发明的基因的碱基序列相似的基因可通过使用本发明获得的基因或其部分作为探针，按照熟知的方法进行杂交而获得。杂交可按如下方法进行，例如，T. Maniatis 等编著，分子克隆：实验室手册第二版，冷泉港实验室出版，1989，书中所描述的方法。更具体而言，杂交在下列条件下进行。简言之，固定有 DNAs 的杂交膜在含有 0.5% SDS，0.1% 牛血清白蛋白 (BSA)，0.1% 聚乙烯吡咯烷酮，0.1% Ficoll 400，0.01% 变性鲑精 DNA 的 $6 \times \text{SSC}$ ($1 \times \text{SSC}$ 表示为 0.15M NaCl, 0.015M 柠檬酸钠, pH 7.0) 中，和探针在 50 °C 温育 12 - 20 小时。温育后，洗涤杂交膜直到固定的 DNAs 信号与背景能区别为止，开始洗涤在含 0.5% SDS 的 $2 \times \text{SSC}$ 中 37 °C 下进行，然后降低 SSC 浓度至 $0.1 \times$ 和提高温度至 50 °C。

另外，作为杂交的替代方法，可应用基因扩增方法(如，PCR 方法)，该法用本发明获得的基因的部分碱基序列作为引物。因此所获得的基因是否编码具有目的功能的蛋白，能通过利用合适的宿主和合适的表达系统表达该基因，然后检查所得到蛋白的活性的方法确定。

附图简述

图 1 是质粒 pSTC 3 的限制性酶图谱。

图 2 比较了蛋白酶 PFUS、蛋白酶 TCES 和枯草杆菌蛋白酶的氨基酸序列。

图 3 比较了蛋白酶 PFUS、蛋白酶 TCES 和枯草杆菌蛋白酶的氨基酸序列。

图 4 比较了蛋白酶 PFUS、蛋白酶 TCES 和枯草杆菌蛋白酶的氨基酸序列。

图 5 比较了蛋白酶 PFUS、蛋白酶 TCES 和枯草杆菌蛋白酶的氨基酸序列。

图 6 是质粒 pSNP 1 的限制性酶切图谱。

图 7 是质粒 pPS 1 的限制性酶切图谱。

图 8 是质粒 pNAPS 1 的限制性酶切图谱。

发明详述

根据本发明的超热稳定蛋白酶包括来自各种嗜热菌的蛋白酶。例如，WO 95/34645描述的来自激烈热球菌和速生热球菌的蛋白酶。

来自激烈热球菌 DSM 3638 的蛋白酶基因是从该菌株的基因组 DNA 文库中，根据表达热稳定蛋白酶活性而分离得到。含有该基因的质粒被命名为质粒 pTPR 12。用该质粒转化的大肠杆菌 JM 109 被命名并且表示为大肠杆菌 JM 109/pTPR 12，于 1994 年 5 月 24 日（最初保藏日期）根据布达佩斯条约保藏在日本茨城县筑波市东 1 丁目 1 番 3 号，通商产业省工业技术院生命工学和人体技术研究所，登记号 FERM BP - 10 5103。

该蛋白酶下文命名为蛋白酶 PFUL。蛋白酶 PFUL 是一种具有高度热稳定性并且甚至在 95 °C 下也表现蛋白酶活性的蛋白酶。

已经测定了来源于激烈热球菌插入到质粒 pTPR 12 的 DNA 片段的碱基序列。在该 DNA 片段中以 2 个 DraI 位点为边界大约 4.8 kb 的部分碱基序列插入到质粒 pTPR 12 中，在序列表 SEQ ID NO: 5 中显示。此外，根据碱基序列推导的基因产物的氨基酸序列在序列表 SEQ ID NO: 6 中显示。换言之，在序列表 SEQ ID NO: 6 中显示的氨基酸序列是蛋白酶 PFUL 的氨基酸序列。如在序列中显示的那样，蛋白酶 PFUL 由 1398 个氨基酸残基组成，是一种超过 150,000 的高分子量蛋白酶。

将序列表 SEQ ID NO: 6 中显示的蛋白酶 PFUL 的氨基酸序列与熟知的来源于微生物的蛋白酶的氨基酸序列进行比较显示，蛋白酶 PFUL 的第一个半部分的氨基酸序列与以枯草杆菌蛋白酶为代表的系列碱性丝氨酸蛋白酶有同源性（蛋白工程，4: 719 - 737 (1991)），所存在的大约 4 个氨基酸残基的极其高度同源性据信对蛋白酶的催化活性是重要的。

如上所述，发现来源于嗜温菌的蛋白酶中的共同区域在由嗜热菌激烈热球菌产生的蛋白酶 PFUL 的氨基酸序列中是保守的。因此，可预期由不是激烈热球菌的嗜热菌产生的同源蛋白酶也具有该区域。

例如进行 PCR 能筛选出编码超热稳定蛋白酶的基因，用来自各种嗜热菌的染色体 DNA 作为模板，寡核苷酸 PRO - 1F、PRO - 2F、PRO - 2R 和 PRO - 4R 的组合作为引物。根据在蛋白酶 PFUL 基因中

编码显示为与枯草杆菌蛋白酶高度同源的区域碱基序列或蛋白酶 PFUL 的氨基酸序列之间相似的序列,合成这些寡核苷酸。寡核苷酸 PRO - 1F、PRO - 2F、PRO - 2R 和 PRO - 4R 的碱基序列分别在序列表 SEQ ID NOS: 7、8、9 和 10 中显示。

- 5 由于根据本发明的蛋白酶来源于嗜热菌,能用属于激烈热球菌属、热球菌属、葡萄嗜热菌属、热拟芽孢杆菌属等细菌。至于属于热球菌属的细菌,例如能用速生热球菌 DSM 2476。该菌株可从 Deutsche Sammlung Von Mikroorganismen und Zellkulturen GmbH 得到。当用来源于速生热球菌 DSM 2476 的染色体 DNA 作为模板和寡核苷酸 PRO - 10 1F 和 PRO - 2R 或寡核苷酸 PRO - 2F 和 Pro - 4R 的组合作为引物进行 PCR 时,能扩增出特异的 DNA 片段,表明存在蛋白酶基因。此外,通过该 DNA 片段被插入到合适的质粒载体中创造重组质粒,然后用双脱氧法测定该插入 DNA 片段的碱基序列,能推导出由该片段编码的氨基酸序列。结果证明上述的 DNA 片段编码的氨基酸序列与蛋白酶 PFUL 和来自各种微生物的碱性丝氨酸蛋白酶的氨基酸序列同源,以及 PCR 扩
15 DNA 片段是从蛋白酶基因作为模板扩增的。

下一步,用该 PCR - 扩增的 DNA 片段或如上所述的寡核苷酸作为探针,筛选来自嗜热菌的基因文库,能得到编码超热稳定蛋白酶的基因(例如,由速生热球菌产生的编码超热稳定蛋白酶的基因)。

- 20 例如,用 PCR - 扩增的 DNA 片段作为探针,针对文库进行噬菌斑杂交,能获得含有目的基因的噬菌体克隆。通过把 λ GEM - 11 载体 (Promega) 和来自速生热球菌 DSM 2476 用限制性酶 Sau 3AI 部分消化的染色体 DNA 的 DNA 片段连接起来,然后用体外包装方法将它们包装到 λ 噬菌体颗粒中,产生上述的文库。

- 25 通过分析由此获得的一个噬菌体克隆中含有的 DNA 片段,发现蛋白酶基因存在于大约 1.9 kb 的 SacI 片段中。此外,还发现通过测定其碱基序列该片段缺乏蛋白酶基因的 5' 区。用盒和盒引物 (Takara Shuzo 基因技术产物指导, 1994 - 1995, pp. 250 - 251) 用 PCR 能获得 5' 区。因此,能获得包括超热稳定蛋白酶 5' 区的 DNA 片段,在质粒 pTCS 6
30 中缺乏 5' 区。此外,来源于速生热球菌的整个超热稳定蛋白酶基因的碱基序列能从 2 个 DNA 片段的碱基序列确定。

在测定过的碱基序列中找到的可读框的碱基序列在序列表的 SEO ID NO: 11 中显示, 并且从该碱基序列推导的氨基酸序列在序列表的 SEQ ID NO: 12 中显示。因而确定了来源于速生热球菌编码超热稳定蛋白酶基因的碱基序列和该蛋白酶的氨基酸序列。该蛋白酶被命名为蛋白酶 TCES。

可构建表达载体, 在该载体中通过把 2 个 DNA 片段结合在一起重组整个蛋白酶 TCES 基因。然而, 当用大肠杆菌作为宿主时, 目的表达质粒导入其中的转化体没有得到, 可能由于在细胞中来自基因表达的产物的产生对大肠杆菌可能是有害的或致死的。在上述情况下, 例如, 可用枯草芽孢杆菌作为宿主, 使蛋白酶分泌到细胞外, 然后测定其活性。

至于枯草芽孢杆菌菌株, 可用枯草芽孢杆菌 DB 104, 它是在“基因” 83: 215 - 233 (1989) 描述过的熟知的菌株。至于克隆载体, 用质粒 pUB 18 - P43, 它由 Calgary 大学 Sui - Lam Wong 博士馈赠。该质粒含有的卡那霉素抗性基因作为选择标志。

在质粒载体 pUB 18 - P43 中 P43 启动子的下游插入蛋白酶 TCES 基因所形成的重组质粒被命名为质粒 pSTC 3。用该质粒转化的枯草芽孢杆菌 DB 104 被命名并表示为枯草芽孢杆菌 DB 104/pST 3, 于 1995 年 12 月 1 日 (最初的保藏日期) 根据布达佩斯条约保藏在通产省工业技术院生命工学和人体技术研究所, 日本, 茨城县筑波市东 1 丁目 1 番 3 号, 登记号为 FERM BP - 5635。

质粒 pSTC 3 的限制性酶图谱在图 1 显示。在图 1 中, 粗线表示插入到质粒载体 pUB 18 - P43 中的 DNA 片段。

热稳定蛋白酶活性可在枯草芽孢杆菌 DB 104/pSTC 3 的培养物上清液中和细胞抽提液中找到。

从转化体培养物中得到的蛋白酶粗酶制备物的主要特性如下:

(1) 作用:

降解酪蛋白和明胶而产生短链多肽。

水解琥珀酰 - L - 亮氨酸 - L - 亮氨酸 - L - 缬氨酸 - L - 酪氨酸 - 4 - 甲基香豆素 - 7 - 酰胺 (Suc - Leu - Leu - Val - Tyr - MCA) 产生荧光物质 (7 - 氨基 - 4 - 甲基香豆素)。

水解琥珀酰 - L - 丙氨酸 - L - 丙氨酸 - L - 脯氨酸 - L - 苯丙氨酸

酸 - 对 - 硝基酰苯胺 (Suc - Ala - Ala - Pro - Phe - p - NA) 产生黄色物质 (对硝基苯胺) 。

(2) 最适温度:

在 37 - 95 °C 下表现为酶活性, 最适温度为 70 - 80 °C 。

5

(3) 最适 pH:

在 pH 5.5 - 9 下表现为酶活性, 最适 pH 为 pH 7 - 8 。

(4) 热稳定性:

在 80 °C 处理 3 小时后仍保持 90 % 或更高的酶活性。

当把蛋白酶 PFUL、蛋白酶 TCES 和枯草杆菌蛋白酶 (枯草杆菌蛋白酶 BNP' ; 核酸研究, 11: 7911 - 7925 (1983)) 排列在一起, 使同源区互相配对如在图 2 - 5 显示的那样时, 发现蛋白酶 PFUL 的同源区之间和 C - 末端存在的序列, 在蛋白酶 TCES 或枯草杆菌蛋白酶中未找到。根据这些结果, 在激烈热球菌中, 除了蛋白酶 PFUL 外可能存在具有的分子量比蛋白酶 PFUL 更低并与蛋白酶 TCES 或枯草杆菌蛋白酶相似的蛋白酶。

15

因此, 用来自同源区的 DNA 探针进行针对从激烈热球菌制备的染色体 DNA 的 Southern 杂交, 并观察到除蛋白酶 PFUL 基因外的信号, 这表明存在另一种蛋白酶基因。

这种新型的蛋白酶基因通过下列方法分离。

20

例如, 通过合适的限制性酶消化来自激烈热球菌的染色体 DNA, 然后按上面所述的方法对消化过的 DNA 进行 Southern 杂交, 可获得含有编码新蛋白酶基因的 DNA 片段。测定该 DNA 片段的碱基序列, 以确认该碱基序列编码的氨基酸序列与上面提到的蛋白酶同源。如果该 DNA 片段不含有整个目的基因, 剩余的部分可通过逆向 PCR 法或相似的方法进一步获得。

25

例如, 当用限制性酶 SacI 和 SpeI (Takara Shuzo) 消化来自激烈热球菌的染色体 DNA 然后用于 Southern 杂交时, 观察到大小大约为 0.6 kb 的信号。该大小的 DNA 片段被重新得到, 插入到质粒载体 pBluescript SK (-) (Stratagene) 中的 SpeI - SacI 位点之间, 然后用产生的重组质粒转化大肠杆菌 JM 109。用相同的探针, 按上述 Southern 杂交所用的方法进行集落杂交, 从转化体中能获得目的片段插入其中的克隆。从获

30

得的克隆含有的质粒是否拥有编码所述的蛋白酶的序列，能用测定插入到质粒中的 DNA 片段的碱基序列确定。因而确认了在该质粒中存在蛋白酶基因。该质粒被命名为质粒 pSS 3。

5 发现从插入到质粒 pSS 3 中的 DNA 片段的碱基序列推导的氨基酸序列与枯草杆菌蛋白酶、蛋白酶 PFUL、蛋白酶 TCES 等的序列具有同源性。与蛋白酶 PFUL 基因不同的该蛋白酶基因的产物，其部分最近按上述的方法从激烈热球菌中得到，被命名为蛋白酶 PFUS。编码该蛋白酶的 N - 末端和 C - 末端区的区域能用逆向 PCR 法获得。

10 用于逆向 PCR 的引物可根据插入在质粒 pSS 3 中的 DNA 片段的碱基序列制备。用合适的限制性酶消化来自激烈热球菌的染色体 DNA，然后将产生的 DNA 片段进行分子内连接反应。通过用反应混合物作模板和上述提到的引物进行 PCR，能获得在质粒 pSS 3 中含有的该蛋白酶基因片段侧翼区相应的 DNA 片段。由这些区域编码的酶蛋白的氨基酸序列可通过分析所获得的 DNA 片段的碱基序列而推导出来。此外，用来自激烈
15 热球菌的染色体 DNA 为模板可制备能扩增整个蛋白酶 PFUS 基因的引物。可设计引物 NPF - 4 和 NPR - 4。引物 NPF - 4 拥有的碱基序列紧邻于蛋白酶 PFUS 基因的启动密码子上游处，能把 BamHI 5' 位点引入到该序列上。引物 NPR - 4 拥有的序列与蛋白酶 PFUS 基因 3' 部分互补，可把 SphI 5' 位点引入到该序列上。

20 引物 NPK - 4 和 NPR - 4 的碱基序列在序列表的 SEQ ID NOS: 13 和 14 中显示。用来自激烈热球菌的染色体 DNA 为模板，这 2 个引物可用于扩增整个蛋白酶 PFUS 的基因。

与蛋白酶 TCES 一样，蛋白酶 PFUS 能在作为宿主的枯草芽孢杆菌中表达。表达蛋白酶 PFUS 的质粒可按用于蛋白酶 TCES 的表达质粒
25 pSTC 3 构建。具体来说，通过把质粒 pSTC 3 中的蛋白酶 TCES 基因用含有整个蛋白酶 PFUS 基因的 DNA 片段取代，可构建表达蛋白酶 PFUS 的质粒，所述的 DNA 片段用上述的引物由 PCR 扩增得到。因而所构建的表达质粒被命名为质粒 pSNP 1。用该质粒转化的枯草芽孢杆菌 DB 104 被命名为枯草芽孢杆菌 DB 104/pSNP 1，并于 1995 年 12 月 1 日（最初
30 保藏日期）根据布达佩斯条约，保藏在通产省工业技术院生命工学和人体技术研究所，日本，茨城县筑波市东 1 丁目 1 番 3 号，登记号为 FERM

BP - 5634。质粒 pSNP 1 的限制性酶图谱在图 6 显示。

在编码蛋白酶 PFUS 基因中与可读框相应的碱基序列以及从该碱基序列推导的蛋白酶 PFUS 的氨基酸序列分别在序列表的 SEQ ID NOS: 15 和 16 中显示。

- 5 热稳定蛋白酶活性在来自枯草芽孢杆菌 DB 104/pSNP 1 培养物的上清液中和细胞抽提液中发现。即，部分表达的蛋白酶 PFUS 被分泌到培养物上清液中。

从转化体培养物中获得的蛋白酶的主要特性如下：

(1) 作用：

- 10 降解酪蛋白和胶原以产生短链多肽。

水解琥珀酰 - L - 亮氨酸 - L - 亮氨酸 - L - 缬氨酸 - L - 酪氨酸 - 4 - 甲基香豆素 - 7 - 酰胺 (Suc - Leu - Leu - Val - Tyr - MCA) 产生荧光物质 (7 - 氨基 - 4 - 甲基香豆素) 。

- 15 水解琥珀酰 - L - 丙氨酸 - L - 丙氨酸 - L - 脯氨酸 - L - 苯丙氨酸 - 对 - 硝基酰替苯胺 (Suc - Ala - Ala - Pro - Phe - p - NA) 产生黄色物质 (对硝基苯胺) 。

(2) 最适温度：

在 41 - 110 °C 下表现为酶活性，最适温度为 80 - 95 °C。

(3) 最适 pH：

- 20 在 pH 5 - 10 下表现酶活性，最适 pH 为 pH 6 - 8。

(4) 热稳定性：

在 95 °C 处理 8 小时后仍有 90 % 或更高的酶活性。

(5) pH 稳定性：

- 25 在 pH 5 - 11，95 °C 下处理 60 分钟后仍保持 95 % 或更高的酶活性。

(6) 分子量：

在 SDS - PAGE 上出现的分子量大约为 45 kDa。

- 30 使用获得蛋白酶 TCES 基因和蛋白酶 PFUS 基因相似的方法，与蛋白酶 TCES 基因和蛋白酶 PFUS 基因同源的蛋白酶基因能从不是激烈热球菌和速生热球菌的嗜热菌中得到。

可制备大约 1 kb 的 DNA 片段，该片段编码在序列表 SEQ ID NO：

6 中显示的蛋白酶 PFUL 的氨基酸序列从残基 323 位至残基 650 的序列，
用该片段作探针，针对来自海葡萄嗜热菌和解蛋白热拟杆菌 DSM 5265
的染色体 DNAs 进行基因组 Southern 杂交。结果，用 PstI (Takara
Shuzo) 消化的来自海葡萄嗜热菌染色体 DNA 大约 4.8 kb 的位置和用
5 XbaI 消化的来自解蛋白热拟杆菌染色体 DNA 大约 3.5 kb 的位置观察到
信号。

根据这些结果，证明从海葡萄嗜热菌和解蛋白热拟杆菌来的染色体
DNAs 存在编码蛋白酶 PFUL、蛋白酶 PFUS 和蛋白酶 TCES 等基因同
源的序列。用分离和鉴别编码蛋白酶 TCES 和蛋白酶 PFUS 基因相似
10 的方法，检测 DNA 片段，可分离和鉴定在海葡萄嗜热菌和解蛋白热拟杆菌
中编码超热稳定蛋白酶的基因。

一般来说，为了通过遗传工程技术大量制备蛋白，认为应用在宿主
中有效地起作用的启动子而不用编码目的蛋白的基因内部结合的启动子
是有利的。尽管用于构建蛋白酶 TCES 和蛋白酶 PFUS 表达系统的 P43
15 启动子是来自枯草芽孢杆菌的启动子，它不能足够有效地表达这 2 种蛋
白酶。

因此，为了增加表达水平，可利用在枯草芽孢杆菌中高水平表达的
基因，尤其是编码分泌蛋白的基因。可用编码 α -淀粉酶或各种细胞外
蛋白酶的基因。例如，可预期应用枯草杆菌蛋白酶基因的启动子和编码
20 信号肽区可增加蛋白酶 PFUS 的表达水平。

具体而言，通过把整个蛋白酶 PFUS 基因置于编码枯草杆菌蛋白酶
基因信号肽及包括启动子区的区域下游，这样 2 个基因的翻译框能互相
匹配，在枯草杆菌蛋白酶的启动子控制下蛋白酶 PFUS 能作为融合蛋白
表达。

25 例如，编码枯草杆菌蛋白酶 E 的基因能用作枯草杆菌蛋白酶基因，
用在本发明中。可使用把枯草杆菌蛋白酶 E 基因的启动子和编码信号肽
的区域插入到质粒 pKW 2 中，该质粒在“细菌学杂志” 171: 2657 -
2665 (1989) 中描述。包括启动子序列的 5' 上游区的碱基序列在该参
考文献 (同上) 中描述而编码枯草杆菌蛋白酶区的碱基序列在“细菌学
30 杂志” 158: 411 - 418 (1984) 中描述。

根据这些序列，合成引入 EcoRI 位点位于该基因启动子序列上游的

引物 SUB 4 和引入 BamHI 位点位于编码枯草杆菌蛋白酶 E 的信号肽区下游的引物 BmR1。引物 SUB4 和 BmR1 的碱基序列分别在序列表 SEQ ID NOS: 17 和 18 中显示。用质粒 pKWZ 作模板, 经 PCR 用引物 SUB4 和 BmR1 扩增出大约 0.3 kb 的 DNA 片段, 该片段含有枯草杆菌蛋白酶 E 基因的启动子和编码信号肽区。

置于该 DNA 片段下游的蛋白酶 PFUS 基因, 可通过 PCR 法从激烈热球菌的染色体 DNA 中获得。引物 NPF - 4 能用作和该基因 5' 区杂交的引物。根据该基因终止密码子下游的碱基序列设计的引物 NPM - 1 拥有 SphI 位点, 能用作和该基因 3' 区杂交的引物。引物 NPM - 1 的序列在序列表的 SEQ ID NO: 19 中显示。

存在于基因中的一个 BamHI 位点在所述的方法中会产生问题, 在该方法中, BamHI 位点是用于连接蛋白酶 PFUS 基因和 0.3 kb DNA 片段。依据在序列表 SEQ ID NO: 15 中显示的蛋白酶 PFUS 基因的碱基序列, 能制备经 PCR - 诱变法消除 BamHI 位点的引物 mutRR 和 mutFR。引物 mutRR 和 mutFR 的碱基序列分别在序列表 SEQ ID NOS: 20 和 21 中显示。当用这些引物消除 BamHI 位点时, 由该位点编码的氨基酸残基即甘氨酸, 由于引入该位点的碱基替换, 被缬氨酸所替换, 所述的位点在序列表 SEQ ID NO: 16 中显示的蛋白酶 PFUS 氨基酸序列中 560 位。

用这些引物能获得连接有枯草杆菌蛋白酶 E 的启动子和编码信号肽区的蛋白酶 PFUS 基因。具体而言, 用来自激烈热球菌的染色体 DNA 作模板及配对引物 mutRR 和 NPF - 4 或配对引物 mutFR 和 NPM - 1, 进行 2 次 PCRs。此外, 用各自 PCR 扩增的 DNA 片段作模板和引物 NPF - 4 和 NPM - 1 混合形成异源双链, 进行第二轮 PCR。因此, 能扩增出不含有内部 BamHI 位点大约 2.4 kb 的整个蛋白酶 PFUS 基因。

分离用 BamHI 和 SphI 消化的 PCR 扩增的 DNA 片段获得大约 2.4 kb 的一个 DNA 片段, 用该片段取代含有蛋白酶 PFUS 基因的质粒 pSNP 1 中的 BamHI - SphI 片段。由此构建的表达载体被命名为质粒 pPS 1。用该质粒转化的枯草芽孢杆菌 DB 104 命名为枯草芽孢杆菌 DB 104/pPS1。观察含有质粒 pSNP 1 的转化体, 发现在该转化体培养物的上清液和抽提液中相似的蛋白酶活性, 这表明该氨基酸替换不影响酶活性。

质粒 pPS 1 的限制性酶图谱在图 7 中显示。

含有枯草杆菌蛋白酶 E 基因的启动子和编码信号肽区的大约 0.3 kb DNA 片段用 EcoRI 和 BamHI 消化,用于取代在质粒 pPS 1 中含有的 P43 启动子和核糖体结合位点的 EcoRI - BamHI 片段。由此构建的表达质粒被命名为 pNAPS 1。用该质粒转化的枯草芽孢杆菌 DB 104 被命名为枯草芽孢杆菌 DB 104/pNAPS 1。热稳定蛋白酶活性在该转化体培养物的上清液和细胞抽提液中找到,与枯草芽孢杆菌 DB 104/pSNP 1 相比较增加了表达水平。质粒 pNAPS 1 的限制性酶图谱在图 8 中显示。

从该转化体表达的蛋白酶表现的酶学特性与上述的由枯草芽孢杆菌 DB 104/pSNP 1 表达的蛋白酶的那些特性相等。纯化了由该转化体表达的蛋白酶。分析纯化蛋白酶 N - 末端氨基酸序列提出的氨基酸序列在序列表 SEQ ID NO: 22 中显示。该序列与在序列表 SEQ ID NO: 15 中显示的蛋白酶 PFUS 的氨基酸序列从 133 位至 144 位的序列相同,表明成熟的蛋白酶 PFUS 是从该部分开始的多肽所组成的酶。根据这些结果呈现的成熟蛋白酶 PFUS 的氨基酸序列在序列表 SEQ ID NO: 4 中显示。

尽管与枯草芽孢杆菌 DB 104/pSNP 1 (FERM BP - 5634) 产生的蛋白酶量相比较,由枯草芽孢杆菌 DB 104/pNAPS 1 产生的蛋白酶量增加,仍需要更高的生产力。通过修饰枯草芽孢杆菌信号肽和蛋白酶 PFUS 之间由 pNAPS 1 编码的融合肽结合部分,以更有效地移去信号肽,预计能增加该蛋白酶的表达水平。在质粒 pNAPS 1 中,由 3 个氨基酸残基 Ala - Gly - Ser 组成的肽插入在序列表 SEQ ID NO: 3 中显示的枯草杆菌蛋白酶 E 信号肽的 C - 末端氨基酸残基 (Ala) 和蛋白酶 PFUS 的 N - 末端氨基酸残基 (Met) 之间。通过把突变导入到质粒 pNAPS 1 中编码该肽的 DNA 中然后检查已导入突变质粒的转化体的蛋白酶生产性,能获得具有增加蛋白酶表达水平的转化体。

首先,制备突变质粒,其中在质粒 pNAPS 1 中编码融合蛋白,枯草杆菌蛋白酶 E - 蛋白酶 PFUS 的基因中 3 个氨基酸肽中编码 Ser 的部分被修饰,这样该部分的碱基序列编码随机的 2 个氨基酸残基。上述突变质粒可用 PCR 法产生。例如,具有用随机的 6 个碱基替换编码 Ser 密码子 (TCC) - 的序列的引物 SPOFO 和 SPORO(引物 SPOFO 和 SPORO

的碱基序列分别在序列表 SEQ ID NOS: 24 和 25 中显示) 和根据该区周围的碱基序列制备的引物 SUB 3 和 NPR - 10, 用于进行 PCR 以获得一个 DNA 片段, 在与编码 Ser 密码子 (Ser) 相应的部位的定向突变已引入了该 DNA 片段中。由得到的片段取代在质粒 pNAPS 1 中相应的区域能获得含有导入突变的蛋白酶基因的突变质粒。

通过由此获得的突变质粒导入到合适的宿主如枯草芽孢杆菌 DB 104 中, 然后测定由转化体表达的蛋白酶水平, 那么就能获得具有增加表达水平的转化体。由测定所分离的转化体独立培养物中的活性确定蛋白酶的表达水平。另外, 具有增加表达水平的转化体可用含基质的琼脂平板方便地选出。

具体来说, 突变质粒导入其中的转化体生长在含有脱脂乳的琼脂平板上。然后, 平板培养在蛋白酶 PFUS 能表现其活性的温度下, 如在 70 °C 下。在表达蛋白酶的转化体集落周围的脱脂乳被降解而变得清亮。蛋白酶的表达水平能从清亮区的大小估计出来。

与枯草芽孢杆菌 DB 104/pNAPS 1 相比较由此得到的一种表达高水平蛋白酶活性的转化体, 被命名为枯草芽孢杆菌 DB 104/pSPO 124。制备含有该转化体的质粒(该质粒被命名为 pSPO 124)。分析该质粒的碱基序列显示编码 Ser 的部分转变成碱基序列 GGG AAT, 那就是说, 由该质粒编码的是一种 Ser 已转变为 Gly - Asn 的蛋白。

因此, 通过把由 4 个氨基酸残基 Ala - Gly - Gly - Asn 组成的肽置于枯草杆菌蛋白酶信号肽的下游, 将其与目的蛋白的 N - 末端融合, 然后表达融合蛋白, 证明目的蛋白的表达水平能在作为宿主的芽孢杆菌属细胞中增加。除了用于本发明的枯草杆菌蛋白酶 E (来自枯草芽孢杆菌) 外, 来自解淀粉枯草杆菌的枯草杆菌蛋白酶 BPN' (核酸研究, 11: 7911 - 7925 (1983)), 来自地衣枯草杆菌的枯草杆菌蛋白酶 Carlsberg (核酸研究, 13: 8913 - 8926 (1985)) 等都是熟知的由芽孢杆菌属细菌产生的枯草杆菌蛋白酶。来自它们的信号肽能优选地用于本发明, 尽管它们的氨基酸序列互相间稍有不同。在芽孢杆菌属细菌中起作用的各种启动子能用在本发明为控制表达所用的来自枯草杆菌蛋白酶 E 基因的启动子的部位。

关于被表达的蛋白没有限制。只要获得编码蛋白的基因就可应用本

发明通过遗传工程技术高水平地表达该蛋白。根据分类学上与芽孢杆菌属细菌不同的激烈热球菌的蛋白可高水平表达的事实，显然本发明能用于表达不是本发明宿主的生物来源的蛋白。通过遗传工程技术，本发明优选地用于生产蛋白酶 PFUL、蛋白酶 TCES 以及结构上与蛋白酶 PFUS 5 相似，来源于海葡萄嗜热菌和解蛋白热拟杆菌的蛋白酶。

依据和枯草杆菌蛋白酶的同源性，可考虑将蛋白酶 PFUS 表达为具有信号肽和前肽的前体蛋白，然后进行加工以产生成熟的酶。此外，根据成熟蛋白酶 PFUS 酶的 N - 末端氨基酸序列分析的结果，可以设定该成熟酶是一种在序列表 SEQ ID NO: 4 中所示的氨基酸序列组成的酶。10 然而，纯化的成熟蛋白酶 PFUS 的分子量大约为 45 kDa，比根据氨基酸序列计算的分子量更小，表明也进行其 C - 末端肽的加工后，作为前肽表达的蛋白酶 PFUS 转变成成熟的蛋白酶。

如果由加工过程移去的 C - 末端肽对酶活性或酶蛋白折叠成合适的结构是非必需的，预期通过从该基因中缺失编码该部分的区域然后表达15 蛋白酶，也可增加蛋白酶 PFUS 的表达水平。

从枯草芽孢杆菌 DB 104/pNAPS 1 获得的成熟蛋白酶 PFUS 的分子量例如可用质子光谱仪精确地测定。根据测定的分子量和按上述方法确定的成熟蛋白酶 PFUS 的 N - 末端氨基酸序列，发现该蛋白酶是一种与序列表 SEQ ID NO: 15 中所示的氨基酸序列 133 位 Ala 至 552 位 Thr 20 相应的肽。此外，通过把终止密码子导入到在质粒 pNAPS 1 含有的蛋白酶 PFUS 基因中编码 552 位 Thr 部分的附近区域，能构建表达缺乏对其酶活性非必需的多肽的蛋白酶 PFUS 的质粒。具体而言，用引物 NPR 544 和引物 NPFE 81 经 PCR 获得具有所设计的终止密码子引入其中的碱基序列的 DNA 片段，引物 NPR 544 是在质粒 pNAPS 1 中的蛋白酶 PFUS 25 基因从起始密码子至第 544 位氨基酸残基编码密码子的 C - 末端侧 (Ser) 导入一个终止密码子 (TGA) (引物 NPR 544 的碱基序列在序列表的 SEQ ID NO: 28 中显示)，而引物 NPFE 81 拥有在该基因中从 NspV 位点上游区的碱基序列 (引物 NPFE 81 的碱基序列在序列表的 SEQ ID NO: 29 中显示)。然后用该片段取代质粒 pNAPS 1 的相应区30 域能获得含有目的突变导入蛋白酶基因中的突变质粒。该质粒命名为质粒 pNAPSΔC。用该质粒转化的枯草芽孢杆菌 DB 104 被命名为枯草芽孢

杆菌 DB 104 被命名为枯草芽孢杆菌 DB 104/ pNAPSΔC.

该转化体表达蛋白酶活性, 具有的特性与蛋白酶 PFUS 相同, 其表达水平比枯草芽孢杆菌 DB 104/ pNAPS 1 更高。

因此, 可以发现在质粒 pNAPSΔC 中含有的蛋白酶 PFUS 基因拥有表达该酶活性的足够区域。在质粒中存在的编码蛋白酶 PFUS 区的碱基序列在序列表的 SEQ ID NO: 2 中显示。由该碱基序列编码的氨基酸序列在序列表的 SEQ ID NO: 2 中显示。

此外, 通过把在质粒 pNAPSΔC 中相似的突变导入到质粒 pSPO 124 的蛋白酶 PFUS 基因中, 能表达缺乏其 C-末端肽的蛋白酶 PFUS。

具体而言, 通过用 NspV 和 SphI 消化质粒 pNAPSΔC 获得的大约 13 kb 的 DNA 片段和已被 NspV 和 SphI 消化的质粒 pSPO 124 混合并连接, 而构建该目的质粒。该质粒被命名为质粒 pSO124ΔC。用该质粒转化的枯草芽孢杆菌 DB 104 被命名并表示为枯草芽孢杆菌 DB 104/pSO124ΔC, 于 1997 年 5 月 16 日 (最初的保藏日期) 根据布达佩斯条约保藏在通产省工业技术院生命工学工业技术研究所, 日本, 茨城县筑波市东 1 丁目 1 番 3 号, 登记号为 FERM BP-6294。该转化体的蛋白酶表达水平与枯草芽孢杆菌 DB 104/pNAPS 1 相比增加。

由转化体枯草芽孢杆菌 DB 104/pNAPSΔC 和枯草芽孢杆菌 DB 104/ pSO124ΔC 产生的蛋白酶的酶学特性以及物理和化学特性, 似乎与由枯草芽孢杆菌 DB 104/pSNP 1 产生的蛋白酶的那些特性相同。从该 2 个转化体中的培养物中获得的蛋白酶的主要特性如下:

(1) 作用:

降解酪蛋白和胶原产生短链多肽。

水解琥珀酰-L-亮氨酸-L-亮氨酸-L-缬氨酸-L-酪氨酸-4-甲基香豆素-7-酰胺 (Suc-Leu-Leu-Val-Tyr-MCA) 产生荧光物质 (7-氨基-4-甲基香豆素)。

水解琥珀酰-L-丙氨酸-L-丙氨酸-L-脯氨酸-L-苯丙氨酸-对-硝基酰替苯胺 (Suc-Ala-Ala-Pro-Phe-p-NA) 产生黄色物质 (对硝基苯胺)。

(2) 最适温度:

在 40-110℃ 下表现酶活性, 最适温度为 80-95℃。

(3) 最适 pH:

在 pH 5 - 10 下表现酶活性, 最适 pH 为 pH 6 - 8。

(4) 热稳定性:

在 95 °C 下处理 8 小时后保留 90 % 或更高的酶活性。

5 (5) pH 稳定性:

在 95 °C, pH 5 - 11 下处理 60 分钟后保留 95 % 或更高的酶活性。

(6) 分子量:

在 SDS - PAGE 上表现的分子量大约 45 kDa。

因此, 提供了具有高热稳定性的蛋白酶和其基因。此外, 本发明披露了表达蛋白的新系统, 该系统使蛋白酶大量表达。用遗传工程技术将该表达系统用于生产本发明的蛋白酶及各种蛋白。

下列的实施例更详细地说明本发明, 但不能解释为限制本发明的范围。

15 实施例 1

(1) 从激烈热球菌中制备染色体 DNA

按下列条件培养激烈热球菌 DSM 3638。

含有 1% 蛋白胨, 0.5% 酵母抽提物, 1% 可溶性淀粉, 3.5% Jamarine S 固体物 (Jamarine 实验室), 3.5 % Jamarine S 液体物 (Jamarine 实验室), 0.003% $MgSO_4$, 0.001% $NaCl$, 0.0001% $FeSO_4 \cdot 7H_2O$, 0.0001% $CoSO_4$, 0.0001% $CaCl_2 \cdot 7H_2O$, 0.0001% $ZnSO_4$, 0.1 ppm $CuSO_4 \cdot 5H_2O$, 0.1 ppm H_3BO_3 , 0.1 ppm $KAl(SO_4)_2$, 0.1 ppm $Na_2MoO_4 \cdot 2H_2O$, 0.25 ppm $NiCl_2 \cdot H_2O$ 的培养基放在 2L 的培养瓶中, 120 °C 灭菌 20 分钟, 用氮气充泡以移去溶解的氧气, 然后将菌株接种到培养基中, 在 95 °C 下非振荡条件培养 16 小时。培养后, 通过离心收集细胞。

然后得到的细胞悬浮在 4 ml 含有 25 % 蔗糖的 50 mM Tris - HCl (pH 8.0) 中。加 2 ml 的 0.2 M EDTA 和 0.8 ml 溶菌酶 (5 mg/ml) 到悬液中。混合液在 20 °C 下温育 1 小时。然后往混合液中加入 24 ml 的 SET 溶液 (150 mM $NaCl$, 1 mM EDTA, 20 mM Tris - HCl, pH 8.0), 4 ml 的 5 % SDS 和 400 μ l 的蛋白酶 K (10 mg/ml)。继续在 37 °C 下温育 1 小时。反应通过用苯酚 - 氯仿抽提混合物而终止。然后,

进行乙醇沉淀获得大约 3.2 mg 的染色体 DNA。

实施例 2

(1) 为构建质粒 pNSP 1 的引物合成

5 为了合成用于扩增整个蛋白酶 PFUS 基因的引物，含有该整个基因的质粒 pSNP 1 从枯草芽孢杆菌 DB 104/pSNP 1 (FERM BP - 5634) 中分离，然后测定所获得区域的碱基序列。根据该碱基序列，合成引入 BamHI 位点紧邻于蛋白酶 PFUS 基因起始密码子上游的引物 NPF - 4 和能与该基因 3' 区杂交并含有 SphI 识别位点的引物 NPM - 1。引物
10 NPF - 4 和 NPM - 1 的碱基序列分别在序列表 SEQ ID NOS: 13 和 19 中显示。

也合成移去 BamHI 位点的引物 mutRR 和 mutFR，该 BamHI 位点位于蛋白酶 PFUS 基因中自起始密码子下游大约 1.7 kb 处。引物 mutRR 和 mutFR 的碱基序列分别在序列表的 SEQ ID NOS: 20 和 21 中显示。

15 (2) 制备质粒 pPS 1

制备 2 套 LA - PCR 反应混合液，然后进行 94 °C 30 秒 - 55 °C 1 分钟 - 68 °C 3 分钟的 30 个循环反应，每套 LA - PCR 反应混合液含有来自激烈热球菌的染色体 DNA 为模板和引物 NPF - 4 和 mutRR 的组合或引物 mutFR 和 NPM - 1 的组合。用 LA PCR 试剂盒 Ver. 2 (Takara
20 Shuzo) 制备 LA - PCR 反应混合液。等份的反应混合物进行琼脂糖凝胶电泳，分别观察到用引物 NPF - 4 和 mutRR 扩增出一条大约 1.8 kb 的 DNA 片段而用引物 mutFR 和 NPM - 1 扩增出一条大约 0.6 kb 的 DNA 片段。

用 SUPREC - 02 (Takara Shuzo) 从 2 套 PCR 反应混合液中移
25 去引物以制备扩增的 DNA 片段。制备含有这 2 种扩增的 DNA 片段但不含有引物或 LA Taq 的 LA - PCR 反应混合液，94 °C 下热变性 10 分钟，在 30 分钟内冷却至 30 °C，然后在 30 °C 温育 15 分钟以形成异源双链。接着，把 LA Taq (Takara Shuzo) 加到反应混合液中，72 °C 下反应 30 分钟。然后把引物 NPF - 4 和 NPM - 1 加到反应混合液中，然后
30 该反应混合液进行 94 °C 30 分钟 - 55 °C 1 分钟 - 68 °C 3 分钟的 25 个循环反应。在反应混合液中观察到扩增出一条大约 2.4 kb 的 DNA 片段。

大约 2.4 kb 的该 DNA 片段用 BamHI 和 SphI (两者均来自 Takara Shuzo) 消化。该片段和已被 BamHI 和 SphI 消化过移去整个蛋白酶 PFUS 基因的质粒 pSNP 1 混合并连接, 然后导入到枯草芽孢杆菌 DB 104 中。从产生的卡那霉素抗性的转化体中制备质粒, 选出其中仅大约 2.4 kb 片段一个分子插入的质粒, 并且命名为质粒 pPS 1。用该质粒 pPS 1 转化的枯草芽孢杆菌 DB 104 被命名为枯草芽孢杆菌 DB 104/pPS 1。

质粒 pPS 1 的限制性酶图谱在图 7 中显示。

(3) 对枯草杆菌蛋白酶 E 基因的启动子 - 编码信号肽区扩增 DNA 片段

为获得枯草杆菌蛋白酶 E 基因的启动子 - 编码信号肽区而合成引物。首先, 根据在细菌学杂志, 171: 2657 - 2665 (1989) 所描述的枯草杆菌蛋白酶 E 基因启动子区的碱基序列合成引物 SUB 4, 该引物能与该区上游序列杂交并含有 1 个 EcoRI 位点 (引物 SUB 4 的碱基序列在序列表的 SEQ ID NO: 17 中显示)。根据在细胞学杂志, 158: 411 - 418 (1984) 中所描述的枯草杆菌蛋白酶 E 基因的碱基序列, 合成能引入 1 个 BamHI 位点紧邻于编码信号肽区下游的引物 BmR1 (引物 BmR1 的碱基序列在序列表的 SEQ ID NO: 18 中显示)。

制备含有在细菌学杂志, 171: 2657 - 2665 中所描述的枯草杆菌蛋白酶 E 基因的质粒 pKWZ 作模板及引物 SUB 4 和 BmR1 的 PCR 反应混合液, 然后进行 94 °C 30 秒 - 55 °C 1 分钟 - 68 °C 2 分钟的 30 个循环反应。等份的反应混合液进行琼脂糖凝胶电泳, 并观察到扩增出一条大约 0.3 kb 的 DNA 片段。

(4) 构建蛋白酶表达质粒 pNAPS 1

用 EcoRI (Takara Shuzo) 和 BamHI 消化上述大约 0.3 kb 的 DNA 片段, 用实施例 3 中所述的方法已被 EcoRI 和 BamHI 消化的质粒 pPS1 与 DNA 片段混合并连接, 然后导入到枯草芽孢杆菌 DB 104 中。从产生的卡那霉素抗性的转化体中制备质粒, 然后挑选出仅大约 0.3 kb 的片段一个分子插入的质粒, 命名为质粒 pNAPS 1。用该质粒 pNAPS 1 转化的枯草芽孢杆菌 DB 104 被命名为枯草芽孢杆菌 DB 104/pNAPS 1。

质粒 pNAPS 1 的限制性酶图谱在图 8 中显示。

(5) 构建质粒 pSNP 2

为引入 BamHI 位点位于在上面提到的质粒 pNAPS 1 中的枯草杆菌蛋白酶 E 基因的编码信号肽区上游, 合成引物 SUB17R (引物 SUB17R 的碱基序列在序列表的 SEQ ID NO: 23 中显示)。制备含有质粒 pNAPS1 作模板及引物 SUB17R 和 SUB4 的 PCR 反应混合液, 进行 94 °C 30 秒 - 55 °C 1 分钟 - 72 °C 1 分钟的 25 个循环反应。用 EcoRI 和 BamHI 消化扩增的大约 0.21 kb 的 DNA 片段, 以获得一条大约 0.2 kb 的 DNA 片段, 该片段含有枯草杆菌蛋白酶 E 基因的启动子和 SD 序列。该片段和已被 EcoRI 和 BamHI 消化过的质粒 pNAPS1 混合并连接。用该反应混合液转化枯草芽孢杆菌 DB 104。从产生的卡那霉素抗性的转化体中制备质粒, 然后挑选出大约 0.2 kb 的 DNA 片段插入的质粒, 命名为质粒 pSNP 2。

(6) 产生高水平表达蛋白酶的突变质粒

为把位于质粒 pNAPS 1 中枯草杆菌蛋白酶 E 基因的编码信号肽区和蛋白酶 PFUS 基因的起始密码子之间的编码氨基酸残基 Ser 的序列 (碱基序列: TCC) 用编码 2 个随机氨基酸残基的序列替换, 合成引物 SPOFO 和 SPORO (引物 SPOFO 和 SPORO 的碱基序列分别在序列表的 SEQ ID NOS: 24 和 25 中显示)。合成引入的 BamHI 位点紧邻于质粒 pNAPS 1 的枯草杆菌蛋白酶 E 基因中编码信号肽的上游的引物 SUB3 和在蛋白酶 PFUS 编码区中含有 SpeI 位点的引物 NPR - 10 (引物 SUB3 和 NPR - 10 的碱基序列分别在序列表的 SEQ ID NOS: 26 和 27 中显示)。

制备 PCR 反应混合液, 每种混合液含有以质粒 pNAPS 1 作模板和引物 SPOFP 和 NPR - 10 组合或引物 SUB3 和 SPORO 的组合, 将反应混合液进行 94 °C 30 秒 - 50 °C 1 分钟 - 72 °C 1 分钟的 20 个循环反应。在 2 个反应混合液中扩增出的大约 0.13 kb 和大约 0.35 kb 的 DNA 片段混合在一起, 在 94 °C 变性 10 分钟, 逐渐冷却到 37 °C 以形成异源双链。然后用 Taq 聚合酶 (Takara Shuzo) 的方法从异源双链产生双链 DNA。制备含有由上获得的双链 DNA 作模板及引物 SUB3 和 NPR - 10 的 PCR 反应混合液, 进行 94 °C 30 秒 - 50 °C 1 分钟 - 72 °C 1 分钟的 25 个循环反应。用 BamHI 和 SpeI 消化扩增的大约 0.43 kb 的 DNA 片段所获得的

一条 DNA 片段和用 BamHI 和 SpeI 消化过的质粒 pSNP 2 混合并连接。用反应混合液转化枯草芽孢杆菌 DB 104。

将产生的卡那霉素抗性的转化体接种在脱脂乳的平板上（用于高温培养的 LB - 琼脂培养基含有 10 $\mu\text{g/ml}$ 的卡那霉素和 1 % 脱脂乳）以形成集落。接着，平板培养在 70 $^{\circ}\text{C}$ 下，根据在集落周围的脱脂乳的降解程度，检查由各个转化体表达的蛋白酶活性。结果，分离到一个表达特别高活性的克隆，从该克隆制备命名为质粒 pSPO 124 的质粒。用该质粒转化的枯草芽孢杆菌 DB 104 被命名为枯草芽孢杆菌 DB 104/pSPO 124。分析质粒 pSPO 124 的碱基序列，发现在质粒 pNAPS 1 中编码 Ser 的碱基序列被碱基序列 GGG AAT 所替换，那就是说所编码的蛋白中，Ser 被转换为 2 个氨基酸残基 Gly - Asn。另外，证明与蛋白酶 PFUS 基因的 Pro (CCA) 相应的从起始密码子计算的第 25 位密码子转变为编码 Leu 的密码子 (CTA)，这是与上面所述的突变同时发生的。

(7) 构建蛋白酶表达质粒 pNAPS Δ C

1 个终止密码子被导入在质粒 pNAPS 1 中从蛋白酶 PFUS 基因的起始密码子算起的第 544 位氨基酸残基的 C - 末端侧，以构建表达从该点起缺乏下游区的蛋白酶的质粒。合成引物 NPR 544，该引物在该基因中编码第 544 位氨基酸残基的密码子 C - 末端侧导入一个终止密码子(碱基序列: TGA) 和有一个 SphI 位点(引物 NPR 544 的碱基序列在序列表的 SEQ ID NO: 28 中显示)。此外，根据在该基因中从 NspV 位点上游部分的碱基序列，合成引物 NPFE 81 (引物 NPFE 81 的碱基序列在序列表的 SEQ ID NO: 29 中显示)。

制备含有质粒 pNAPS 1 作模板及引物 NPFE 81 和 NPR 544 的 PCR 反应混合液，然后进行 94 $^{\circ}\text{C}$ 30 秒 - 50 $^{\circ}\text{C}$ 1 分钟 - 72 $^{\circ}\text{C}$ 1 分钟的 20 个循环反应。扩增出的大约 0.61 kb 的 DNA 片段用 NspV (Takara Shuzo) 和 SpeI 消化，以获得一条含有终止密码子的大约 0.13 kb 的 DNA 片段。该 DNA 片段和已被限制性酶 NspV 和 SphI 消化过的质粒 pNAPS 1 混合并连接。用反应混合液转化枯草芽孢杆菌 DB 104。从产生的卡那霉素抗性的转化体中制备质粒，选出大约 0.13 kb 的 DNA 片段插入其中的质粒，该质粒被命名为质粒 pNAPS Δ C。用该质粒 pNAPS Δ C 转化的枯草芽孢杆菌 DB 104 被命名为枯草芽孢杆菌 DB 104/pNAPS Δ C。

(8) 构建蛋白酶表达质粒 pSPO124 Δ C

用 NspV 和 SphI 消化质粒 pNAPS Δ C 分离得到一条大约 1.3 kb 的 DNA 片段, 然后该片段与已被 NspV 和 SphI 消化过的质粒 pSPO 124 混合并连接。用反应混合液转化枯草芽孢杆菌 DB 104。从产生的卡那霉素抗性转化体中制备质粒, 挑选出大约 1.3 kb 的 DNA 片段插入的质粒, 并被命名为质粒 pSPO124 Δ C。用该质粒 pSPO124 Δ C 转化的枯草芽孢杆菌 DB 104 被命名为枯草芽孢杆菌 DB 104/pSPO124 Δ C。

实施例 3

10 (1) 用含蛋白酶 PFUS 基因的质粒转化的枯草芽孢杆菌的培养和制备粗酶溶液

用实施例 2 所述的方法将含有蛋白酶 PFUS 基因的质粒 pNAPS 1 导入到枯草芽孢杆菌 DB 104 中出现的枯草芽孢杆菌 DB 104/pNAPS 1, 在 2 ml 含有 10 μ g/ml 卡那霉素的 LB 培养基 (蛋白胨 10 g/L, 酵母抽提物 15 5 g/L, NaCl 5 g/L, pH 7.2) 中, 37 $^{\circ}$ C 下培养 24 小时。离心培养物以得到培养上清液 (制备物 1 - S) 和细胞。

把细胞悬浮在 100 μ l 的 50 mM Tris - HCl, pH 7.5 中, 加 2 mg 溶菌酶 (Sigma) 后 37 $^{\circ}$ C 消化 45 分钟。消化过的样品在 95 $^{\circ}$ C 热处理 10 分钟, 然后通过离心收集上清液以获得无细胞的抽提液 (制备物 1 - L)。

20 相似地, 从含有质粒 pSPO 124 的枯草芽孢杆菌 DB 104/pSPO 124、含有质粒 pNAPS Δ C 的枯草芽孢杆菌 DB 104/pNAPS Δ C 或含有质粒 pSPO124 Δ C 的枯草芽孢杆菌 DB 104/pSPO124 Δ C 中, 获得培养物上清液和无细胞的抽提液。来自枯草芽孢杆菌 DB 104/pSPO 124 的培养物上清液和无细胞抽提液分别称为 124 - S 和 124 - L。来自枯草芽孢杆菌 25 DB 104/pNAPS Δ C 的培养物上清液和无细胞抽提液分别称为 Δ C - S 和 Δ C - L。来自枯草芽孢杆菌 DB 104/pSPO124 Δ C 的培养物上清液和无细胞抽提液分别称为 124 Δ C - S 和 124 Δ C - L。用这些制备物测定蛋白酶活性并测定在每个制备物中含有的蛋白酶浓度。

(2) 比较蛋白酶的生产力

30 用 Suc - Ala - Ala - Pro - Phe - p - NA (Sigma) 作底物光谱法测定酶水解反应中产生的 p - 硝基苯胺的量确定蛋白酶 PFUS 的活

性。简言之，适当稀释待测定其酶活性的酶制备液。50 μ l 在 100 mM 磷酸盐缓冲液 pH 7.0 中的 1 mM Suc - Ala - Ala - Pro - Phe - p - NA 被加到 50 μ l 稀释过的样品溶液中。然后，使反应在 95 $^{\circ}$ C 进行 30 分钟。通过在冰上冷却终止反应后，测定在 405 nm 的吸收值以计算出产生的对硝基苯胺的量。该酶的一个单位被定义为 95 $^{\circ}$ C 每 1 分钟产生 1 μ mol 对硝基苯胺的酶量。假定比活性为 9.5 单位/mg 蛋白酶 PFUS 蛋白，根据所测定的酶活性计算在培养物上清液或细胞中表达的酶蛋白的量。

测定在实施例 3 - (1) 制备的每种酶制备物的蛋白酶活性。根据测定结果计算的每种转化体每 1L 培养物的蛋白酶 PFUS 的生产力在表 1 显示。

在枯草芽孢杆菌 DB 104/pSPO 124 中，在细胞中的蛋白酶 PFUS 的生产力比枯草芽孢杆菌 DB 104/pNAPS 1 增加了 3.6 倍。在枯草芽孢杆菌 DB 104/pNAPS Δ C 中，蛋白酶 PFUS 的生产力分别在培养物上清液中增加 2.4 倍而在细胞中增加 2.2 倍。此外，在枯草芽孢杆菌 DB 104/pSPO124 Δ C 中，蛋白酶 PFUS 的生产力分别在培养物上清液中增加 2 倍而在细胞中增加 2.4 倍。每种细胞的生产力也增加。

在培养物上清液和细胞中产生的蛋白酶 PFUS 的总量，与枯草芽孢杆菌 DB 104/pNAPS 1 相比，分别在枯草芽孢杆菌 DB 104/pSPO 124 中增加 2.1 倍，在枯草芽孢杆菌 DB 104/pNAPS Δ C 中增加 2.1 倍及在枯草芽孢杆菌 DB 104/pSPO124 Δ C 中增加 2.2 倍。

表 1

蛋白酶 PFUS 的生产力 (mg/L 的培养物)

转化体 (质粒)	培养物上清液	细胞	培养物上清液 + 细胞
pNAPS1	15.1	12.5	27.6
pSPO124	13.1	45.4	58.5
pNAPS Δ C	35.5	28.1	63.6
pSPO124 Δ C	30.5	30.1	60.6

实施例 4

(1) 制备纯化的成熟蛋白酶 PFUS 的酶制备物

按实施例 2 中所述的方法将编码本发明的超热稳定蛋白酶的基因导

入枯草芽孢杆菌 DB 104 中出现的 2 种菌株，枯草芽孢杆菌 DB 104/pNAPS 1 和枯草芽孢杆菌 DB 104/pSPO124 Δ C，分别接种到含有 10 μ g/ml 卡那霉素的 5 ml LB 培养基中，37 $^{\circ}$ C 振荡培养 7 小时。5 ml 的培养物接种到 5 L 锥形瓶中 500 ml 含 10 μ g/ml 卡那霉素的 TM 培养基（大豆粉末 5 g/L，聚脲 10 g/L，肉膏 5 g/L，酵母抽提物 2 g/L，葡萄糖 10 g/L，FeSO₄ · 7H₂O 10 mg/L，MnSO₄ · 4H₂O 10 mg/L，ZnSO₄ · 7H₂O 1 mg/L，pH 7.0）中，30 $^{\circ}$ C 振荡培养 3 天。超声处理所产生的培养物，在 95 $^{\circ}$ C 热处理 30 分钟，然后离心收集上清液。加过硫酸铵到上清液中至 25 % 的饱和度，然后由下一步离心获得的上清液上样到用含 25 % 饱和过硫酸铵的 25 mM Tris - HCl 缓冲液平衡过的微量制备甲基 HIC 柱（Bio - Rad）上。用相同的缓冲液洗涤凝胶后，吸附到柱上的蛋白酶 PFUS 用含 40 % 乙醇的 25 mM Tris - HCl 缓冲液（pH 7.6）用分步洗脱法洗脱下来。由此获得的含有蛋白酶 PFUS 的级分用 NAP - 25 柱进行凝胶过滤，该 NAP - 25 柱预先用含 20 % 乙腈的 0.05 % 三氟乙酸平衡，在变性蛋白酶 PFUS 的同时脱盐，然后获得纯化的蛋白酶 PFUS 的制备物。从枯草芽孢杆菌 DB 104/pNAPS 1 和枯草芽孢杆菌 DB 104/pSPO124 Δ C 中得到的制备物分别称为 NAPS - 1 和 SPO - 124 Δ C。

2 种纯化的酶制备物在 0.1 % SDS - 10 % 聚丙烯酰胺凝胶上电泳，然后用考马斯亮蓝 R - 250 染色，结果显示 2 种纯化的酶制备物 NAPS - 1 和 SPO - 124 Δ C 为单一的泳带，估测的分子量大约为 45 kDa。

（2）成熟蛋白酶 PFUS 的 N - 末端氨基酸序列的分析

用 G1000A 蛋白测序仪（Hewlett - Packard），通过自动 Edman 法分析纯化的酶制备物 NAPS - 1 和 SPO - 124 Δ C 的 N - 末端氨基酸序列。2 种纯化的酶制备物的 2 个 N - 末端氨基酸序列在序列表的 SEQ ID NO: 22 中显示。该序列与在序列表的 SEQ ID NO: 15 中所示的蛋白酶 PFUS 氨基酸序列从 133 位至 144 位的序列重合，表明 NAPS - 1 和 SPO - 124 Δ C 二者是从该部分开始的多肽所组成的酶。

（3）成熟蛋白酶 PFUS 的质谱法分析

对纯化的酶制备物 NAPS - 1 和 SPO - 124 Δ C 的质谱法分析用 API 300 四极三质子分光仪（Perkin - Elmer Sciex）进行。根据 NAPS - 1

的估测分子量, 43,744 Da, 说明由枯草芽孢杆菌 DB 104/pNAPS 1 产生的成熟蛋白酶 PFUS 是一种在序列表 SEQ ID NO: 15 中所示的蛋白酶氨基酸序列从 133 位 Ala 至 552 位 Thr 的多肽组成的酶. 此外, 根据 SPO - 124ΔC 的估测分子量 42,906 Da, 说明由枯草芽孢杆菌 DB 104/ 5 pSPO124ΔC 产生的成熟蛋白酶 PFUS 是一种由这样的一种多肽组成的酶, 该多肽为在序列表 SEQ ID NO: 15 中所示的蛋白酶 PFUS 从 133 位 Ala 至 544 位 Ser 的氨基酸序列, 即是序列表 SEQ ID NO: 2 中所示的氨基酸序列。

序列表

申请人名称: TAKARA SHUZO CO., LTD

发明题目: 表达超热稳定蛋白的系统

5 参考/摘要号: 660782

目前申请号:

目前申请日:

在先申请号: 151969/1997

在先申请日: 1997, 6月10日

10 序列数: 29

SEQ ID NO: 1 的资料:

长度: 412

类型: 氨基酸

15 链型: 单链

拓扑结构: 线性

分子类型: 肽

序列描述:

```

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

```

Ala Gly Thr Gly Ala Ala Ser Asn Gly Lys Tyr Lys Gly Met Ala			
	80	85	90
Pro Gly Ala Lys Leu Ala Gly Ile Lys Val Leu Gly Ala Asp Gly			
	95	100	105
Ser Gly Ser Ile Ser Thr Ile Ile Lys Gly Val Glu Trp Ala Val			
	110	115	120
Asp Asn Lys Asp Lys Tyr Gly Ile Lys Val Ile Asn Leu Ser Leu			
	125	130	135
Gly Ser Ser Gln Ser Ser Asp Gly Thr Asp Ala Leu Ser Gln Ala			
	140	145	150
Val Asn Ala Ala Trp Asp Ala Gly Leu Val Val Val Val Ala Ala			
	155	160	165
Gly Asn Ser Gly Pro Asn Lys Tyr Thr Ile Gly Ser Pro Ala Ala			
	170	175	180
Ala Ser Lys Val Ile Thr Val Gly Ala Val Asp Lys Tyr Asp Val			
	185	190	195
Ile Thr Ser Phe Ser Ser Arg Gly Pro Thr Ala Asp Gly Arg Leu			
	200	205	210
Lys Pro Glu Val Val Ala Pro Gly Asn Trp Ile Ile Ala Ala Arg			
	215	220	225
Ala Ser Gly Thr Ser Met Gly Gln Pro Ile Asn Asp Tyr Tyr Thr			
	230	235	240
Ala Ala Pro Gly Thr Ser Met Ala Thr Pro His Val Ala Gly Ile			
	245	250	255
Ala Ala Leu Leu Leu Gln Ala His Pro Ser Trp Thr Pro Asp Lys			
	260	265	270

Val Lys Thr Ala Leu Ile Glu Thr Ala Asp Ile Val Lys Pro Asp
 275 280 285
 Glu Ile Ala Asp Ile Ala Tyr Gly Ala Gly Arg Val Asn Ala Tyr
 5 290 295 300
 Lys Ala Ile Asn Tyr Asp Asn Tyr Ala Lys Leu Val Phe Thr Gly
 305 310 315
 Tyr Val Ala Asn Lys Gly Ser Gln Thr His Gln Phe Val Ile Ser
 10 320 325 330
 Gly Ala Ser Phe Val Thr Ala Thr Leu Tyr Trp Asp Asn Ala Asn
 335 340 345
 Ser Asp Leu Asp Leu Tyr Leu Tyr Asp Pro Asn Gly Asn Gln Val
 350 355 360
 15 Asp Tyr Ser Tyr Thr Ala Tyr Tyr Gly Phe Glu Lys Val Gly Tyr
 365 370 375
 Tyr Asn Pro Thr Asp Gly Thr Trp Thr Ile Lys Val Val Ser Tyr
 380 385 390
 20 Ser Gly Ser Ala Asn Tyr Gln Val Asp Val Val Ser Asp Gly Ser
 395 400 405
 Leu Ser Gln Pro Gly Ser Ser
 410

25 SEQ ID NO: 2 的资料:

长度: 1236

类型: 核酸

链型: 双链

拓扑结构: 线性

30 分子类型: 基因组 DNA

原始来源:**生物: 激烈热球菌****菌株: DSM 3638****序列描述:**

```

GCAGAAATTAG AAGGACTGGA TGAGTCTGCA GCTCAAGTTA TGGCAACTTA CGTTTGGAAC      60
TTGGGATATG ATGGTTCCTGG AATCACAATA GGAATAATTG AACTTGGAAAT TGACGCTTCT      120
CATCCAGATC TCCAAGGAAA AGTAATTGGG TGGGTAGATT TTGTCAATGG TAGGAGTTAT      180
CCATACGATG ACCATGGACA TGGAACTCAT GTAGCTTCAA TAGCAGCTGG TACTGGAGCA      240
GCAAGTAATG GCAAGTACAA GGGAAATGGCT CCAGGAGCTA AGCTGGCGGG AATTAAGGTT      300
CTAGGTGCCG ATGGTTCCTGG AAGCATATCT ACTATAATTA AGGGAGTTGA GTGGGCCGTT      360
GATAACAAAG ATAAGTACGG AATTAAGGTC ATTAATCTTT CTCTTGGTTC AAGCCAGAGC      420
TCAGATGGTA CTGACGCTCT AAGTCAGGCT GTTAATGCAG CGTGGGATGC TGGATTAGTT      480
GTTGTGGTTC CCGCTGGAAA CAGTGGACCT AACAAGTATA CAATCGGTTT TCCAGCAGCT      540
GCAAGCAAAG TTATTACAGT TGGAGCCGTT GACAAGTATG ATGTTATAAC AAGCTTCTCA      600
AGCAGAGGGC CAACTGCAGA CGGCAGGCTT AAGCCTGAGG TTGTTGCTCC AGGAAACTGG      660
ATAATTGCTG CCAGAGCAAG TGGAACTAGC ATGGGTCAAC CAATTAATGA CTATTACACA      720
GCAGCTCCTG GGACATCAAT GGCAACTCCT CACGTAGCTG GTATTGCAGC CCTCTTGCTC      780
CAAGCACACC CGAGCTGGAC TCCAGACAAA GTAAAAACAG CCCTCATAGA AACTGCTGAT      840
ATCGTAAAGC CAGATGAAAT AGCCGATATA GCCTACGGTG CAGGTAGGGT TAATGCATAC      900
AAGGCTATAA ACTACGATAA CTATGCAAAG CTAGTGTTC A CTGGATATGT TGCCAACAAA      960
GGCAGCCAAA CTCACCAGTT CGTTATTAGC GGAGCTTCGT TCGTAACTGC CACATTATAC     1020
TGGGACAATG CCAATAGCGA CCTGATCTT TACCTCTACG ATCCAATGG AAACCAGGTT     1080
GACTACTCTT ACACCGCCTA CTATGGATTC GAAAAGGTTG GTTATTACAA CCCAACTGAT     1140
GGAACATGGA CAATTAAGGT TGTAAGCTAC AGCGGAAGTG CAAACTATCA AGTAGATGTG     1200
GTAAGTGATG GTTCCCTTTC ACAGCCGGA AGTTCA                                     1236

```

SEQ ID NO: 3 的资料:

长度: 29

类型: 氨基酸

链型: 单链

5 拓扑结构: 线性

分子类型: 肽

序列描述:

```

Met Arg Ser Lys Lys Leu Trp Ile Ser Leu Leu Phe Ala Leu Thr
              5             10             15
10 Leu Ile Phe Thr Met Ala Phe Ser Asn Met Ser Ala Gln Ala
              20             25

```

SEQ ID NO: 4 的资料:

长度: 522

15 类型: 氨基酸

链型: 单链

拓扑结构: 线性

分子类型: 肽

特征:

20 其它资料: 在 428 位的 Xaa 是 Gly 或 Val

序列描述:

```

Ala Glu Leu Glu Gly Leu Asp Glu Ser Ala Ala Gln Val Met Ala
              5             10             15
Thr Tyr Val Trp Asn Leu Gly Tyr Asp Gly Ser Gly Ile Thr Ile
              20             25             30
Gly Ile Ile Asp Thr Gly Ile Asp Ala Ser His Pro Asp Leu Gln
              35             40             45
Gly Lys Val Ile Gly Trp Val Asp Phe Val Asn Gly Arg Ser Tyr
              50             55             60

```

Pro Tyr Asp Asp His Gly His Gly Thr His Val Ala Ser Ile Ala
 65 70 75

Ala Gly Thr Gly Ala Ala Ser Asn Gly Lys Tyr Lys Gly Met Ala
 80 85 90

Pro Gly Ala Lys Leu Ala Gly Ile Lys Val Leu Gly Ala Asp Gly
 95 100 105

Ser Gly Ser Ile Ser Thr Ile Ile Lys Gly Val Glu Trp Ala Val
 110 115 120

Asp Asn Lys Asp Lys Tyr Gly Ile Lys Val Ile Asn Leu Ser Leu
 125 130 135

Gly Ser Ser Gln Ser Ser Asp Gly Thr Asp Ala Leu Ser Gln Ala
 140 145 150

Val Asn Ala Ala Trp Asp Ala Gly Leu Val Val Val Val Ala Ala
 155 160 165

Gly Asn Ser Gly Pro Asn Lys Tyr Thr Ile Gly Ser Pro Ala Ala
 170 175 180

Ala Ser Lys Val Ile Thr Val Gly Ala Val Asp Lys Tyr Asp Val
 185 190 195

Ile Thr Ser Phe Ser Ser Arg Gly Pro Thr Ala Asp Gly Arg Leu
 200 205 210

Lys Pro Glu Val Val Ala Pro Gly Asn Trp Ile Ile Ala Ala Arg
 215 220 225

Ala Ser Gly Thr Ser Met Gly Gln Pro Ile Asn Asp Tyr Tyr Thr
 230 235 240

Ala Ala Pro Gly Thr Ser Met Ala Thr Pro His Val Ala Gly Ile
 245 250 255

Ala Ala Leu Leu Leu Gln Ala His Pro Ser Trp Thr Pro Asp Lys
260 265 270

Val Lys Thr Ala Leu Ile Glu Thr Ala Asp Ile Val Lys Pro Asp
275 280 285

Glu Ile Ala Asp Ile Ala Tyr Gly Ala Gly Arg Val Asn Ala Tyr
290 295 300

Lys Ala Ile Asn Tyr Asp Asn Tyr Ala Lys Leu Val Phe Thr Gly
305 310 315

Tyr Val Ala Asn Lys Gly Ser Gln Thr His Gln Phe Val Ile Ser
320 325 330

Gly Ala Ser Phe Val Thr Ala Thr Leu Tyr Trp Asp Asn Ala Asn
335 340 345

Ser Asp Leu Asp Leu Tyr Leu Tyr Asp Pro Asn Gly Asn Gln Val
350 355 360

Asp Tyr Ser Tyr Thr Ala Tyr Tyr Gly Phe Glu Lys Val Gly Tyr
365 370 375

Tyr Asn Pro Thr Asp Gly Thr Trp Thr Ile Lys Val Val Ser Tyr
380 385 390

Ser Gly Ser Ala Asn Tyr Gln Val Asp Val Val Ser Asp Gly Ser
395 400 405

Leu Ser Gln Pro Gly Ser Ser Pro Ser Pro Gln Pro Glu Pro Thr
410 415 420

Val Asp Ala Lys Thr Phe Gln Xaa Ser Asp His Tyr Tyr Tyr Asp
425 430 435

Arg Ser Asp Thr Phe Thr Met Thr Val Asn Ser Gly Ala Thr Lys

	440	445	450
	Ile Thr Gly Asp Leu Val Phe Asp Thr Ser Tyr His Asp Leu Asp		
5	455	460	465
	Leu Tyr Leu Tyr Asp Pro Asn Gln Lys Leu Val Asp Arg Ser Glu		
	470	475	480
	Ser Pro Asn Ser Tyr Glu His Val Glu Tyr Leu Thr Pro Ala Pro		
10	485	490	495
	Gly Thr Trp Tyr Phe Leu Val Tyr Ala Tyr Tyr Thr Tyr Gly Trp		
	500	505	510
	Ala Tyr Tyr Glu Leu Thr Ala Lys Val Tyr Tyr Gly		
15	515	520	

SEQ ID NO: 5 的资料:

长度: 4765

类型: 核酸

链型: 双链

20 拓扑结构: 线性

分子类型: 基因组 DNA

原始来源:

生物: 激烈热球菌

菌株: DSM 3638

25 序列描述:

```

TTTAAATTAT AAGATATAAT CACTCCGAGT GATGAGTAAG ATACATCATT ACAGTCCCAA   60
AATGTTTATA ATTGGAACGC AGTGAATATA CAAAATGAAT ATAACCTCCG AGGTGACTGT   120
AGAATGAATA AGAAGGGACT TACTGTGCTA TTTATAGCGA TAATGCTCCT TTCAGTAGTT   180

```

CCAGTGCCTT TTGTGTCCGC AGAAACACCA CCGGTTAGTT CAGAAAATTC AACAACTTCT 240
 ATACTCCCTA ACCAACAAGT TGTGACAAAA GAAGTTTCAC AAGCGGCGCT TAATGCTATA 300
 ATGAAAGGAC AACCCAACAT GGTTCTTATA ATCAAGACTA AGGAAGGCAA ACTTGAAGAG 360
 GCAAAAACCG AGCTTGAAAA GCTAGGTGCA GAGATTCTTG ACGAAAATAG AGTTCTTAAC 420
 ATGTTGCTAG TTAAGATTAA GCCTGAGAAA GTTAAAGAGC TCAACTATAT CTCATCTCTT 480
 GAAAAACCTT GCCTTAACAG AGAAGTTAAG CTTTCCCCTC CAATTGTCGA AAAGGACGTC 540
 AAGACTAAGG AGCCCTCCCT AGAACCAAAA ATGTATAACA GCACCTGGGT AATTAATGCT 600
 CTCCAGTTCA TCCAGGAATT TGGATATGAT GGTAGTGGTG TTGTTGTTGC AGTACTTGAC 660
 ACGGGAGTTG ATCCGAACCA TCCTTTCTTG AGCATAACTC CAGATGGACG CAGGAAAATT 720
 ATAGAATGGA AGGATTTTAC AGACGAGGGA TTCGTGGATA CATCATTCAG CTTTAGCAAG 780
 GTTGTAATG GGA CTCTTAT AATTAACACA ACATTCCAAG TGGCCTCAGG TCTCACGCTG 840
 AATGAATCGA CAGGACTTAT GGAATACGTT GTTAAGACTG TTTACGTGAG CAATGTGACC 900
 AATGGAAATA TCACTTCTGC TAATGGCATC TATCACTTCG GCCTGCTCCC AGAAAGATAC 960
 TTCGACTTAA ACTTCGATGG TGATCAAGAG GACTTCTATC CTGTCCTTAT AGTTAACTCC 1020
 ACTGGCAATG GTTATGACAT TGCATATGTG GATACTGACC TTGACTACGA CTTACCCGAC 1080
 GAAGTCCAC TTGGCCAGTA CAACGTTACT TATGATGTTG CTGTTTTTAG CTACTACTAC 1140
 GGTCCCTCA ACTACGTGCT TGCAGAAATA GATCCTAACG GAGAATATGC AGTATTTGGG 1200
 TGGGATGGTC ACGGTCACGG AACTCACGTA GCCTGGAAGT TTGCTGGTTA CGACAGCAAC 1260
 AATGATGCTT GGGATTGGCT CAGTATGTAC TCTGGTGAAT GGGAAGTGTT CTCAAGACTC 1320
 TATGGTTGGG ATTATACGAA CGTTACCACA GACACCGTGC AGGGTGTGC TCCAGGTGCC 1380
 CAAATAATGG CAATAAGAGT TCTTAGGAGT GATGGACGGG GTAGCATGTG GGATATTATA 1440
 GAAGGTATGA CATACGCAGC AACCCATGGT GCAGACGTTA TAAGCATGAG TCTCGGTGGA 1500
 AATGCTCCAT ACTTAGATGG TACTGATCCA GAAAGCGTTG CTGTGGATGA GCTTACCGAA 1560
 AAGTACGGTG TTGTATTCTG AATAGCTGCA GGAAATGAAG GTCTTGGCAT TAACATCGTT 1620
 GGAAGTCCTG GTGTTGCAAC AAAGGCAATA ACTGTTGGAG CTGCTGCAGT GCCCATTAAC 1680
 GTTGGAGTTT ATGTTTCCCA AGCACCTGGA TATCCTGATT ACTATGGATT CTATTACTTC 1740

CCCGCCTACA CAAACGTTAG AATAGCATTG TTCTCAAGCA GAGGGCCGAG AATAGATGGT 1800
 GAAATAAAAC CCAATGTAGT GGCTCCAGGT TACGGAATTT ACTCATCCCT GCCGATGTGG 1860
 ATTGGCCGAG CTGACTTCAT GCTTGGAAGT TCGATGGCTA CTCCACATGT CAGCGGTGTC 1920
 GTTGCACTCC TCATAAGCGG GGCAAAGGCC GAGGGAATAT ACTACAATCC AGATATAATT 1980
 AAGAAGGTTT TTGAGAGCGG TGCAACCTGG CTTGAGGGAG ATCCATATAC TGGCCACAAC 2040
 TACTACTGAGC TTGACCAAGG TCATGGTCCT GTTAACGTA CCAAGTCCTG GGAAATCCTT 2100
 AAGGCTATAA ACGGCACCAC TCTCCCAATT GTTGATCACT GGGCAGACAA GTCTTACAGC 2160
 GACTTTGCCG AGTACTTGGG TGTGGACGTT ATAAGAGGTC TCTACGCAAG GAACTCTATA 2220
 CCTGACATTG TCGAGTGGCA CATTAAAGTAC GTAGGGGACA CGGAGTACAG AACTTTTGAG 2280
 ATCTATGCAA CTGAGCCATG GATTAAGCCT TTTGTCAGTG GAAGTGTAAT TCTAGAGAAC 2340
 AATACCGAGT TTGTCCTTAG GGTGAAATAT GATGTAGAGG GTCTTGAGCC AGGTCTCTAT 2400
 GTTGAAGGA TAATCATTGA TGATCCAACA ACGCCAGTTA TTGAAGACGA GATCTTGAAC 2460
 ACAATTGTTA TTCCCGAGAA GTTCACTCCT GAGAACAATT ACACCCTCAC CTGGTATGAT 2520
 ATTAATGGTC CAGAAATGGT GACTCACCAC TTCTTCACTG TGCCTGAGGG AGTGGACGTT 2580
 CTCTACGCGA TGACCACATA CTGGGACTAC GGTCTGTACA GACCAGATGG AATGTTTGTG 2640
 TTCCCATACC AGCTAGATTA TCTTCCCGCT GCAGTCTCAA ATCCAATGCC TGGAAACTGG 2700
 GAGCTAGTAT GGACTGGATT TAACITTTGCA CCCCTCTATG AGTCGGGCTT CCTTGTAAGG 2760
 ATTTACGGAG TAGAGATAAC TCCAAGCGTT TGGTACATTA ACAGGACATA CCTTGACACT 2820
 AACACTGAAT TCTCAATTGA ATTCATATT ACTAACATCT ATGCCCAAT TAATGCAACT 2880
 CTAATCCCCA TTGGCCTTGG AACCTACAAT GCGAGCGTTG AAAGCGTTGG TGATGGAGAG 2940
 TTCTTCATAA AGGGCATTGA AGTTCCTGAA GGCACCGCAG AGTTGAAGAT TAGGATAGGC 3000
 AACCCAAGTG TTCCGAATTC AGATCTAGAC TTGTACCTTT ATGACAGTAA AGGCAATTTA 3060
 GTGGCCTTAG ATGGAAACCC AACAGCAGAA GAAGAGGTTG TAGTTGAGTA TCCTAAGCCT 3120
 GGAGTTTATT CAATAGTAGT ACATGGTTAC AGCGTCAGGG ACGAAAATGG TAATCCAACG 3180
 ACAACCACCT TTGACTTAGT TGTTCAAATG ACCCTTGATA ATGGAAACAT AAAGCTTGAC 3240
 AAAGACTCGA TTATTCTTGG AAGCAATGAA AGCGTAGTTG TAACTGCAAA CATAACAATT 3300

GATAGAGATC ATCCTACAGG AGTATACTCT GGTATCATAG AGATTAGAGA TAATGAGGTC 3360
 TACCAGGATA CAAATACTTC AATTGCGAAA ATACCCATAA CTTTGGTAAT TGACAAGGCG 3420
 GACTTTGCCG TTGGTCTCAC ACCAGCAGAG GGAGTACTTG GAGAGGCTAG AAATTACACT 3480
 CTAATGTAA AGCATGCCCT AACACTAGAG CCTGTGCCAA ATGCTACAGT GATTATAGGA 3540
 AACTACACCT ACCTCACAGA CGAAAACGGT ACAGTGACAT TCACGTATGC TCCAACCTAAG 3600
 TTAGGCAGTG ATGAAATCAC AGTCATAGTT AAGAAAGAGA ACTTCAACAC ATTAGAGAAG 3660
 ACCTTCCAAA TCACAGTATC AGAGCCTGAA ATAACTGAAG AGGACATAAA TGAGCCCAAG 3720
 CTTGCAATGT CATCACCAGA AGCAAATGCT ACCATAGTAT CAGTTGAGAT GGAGAGTGAG 3780
 GGTGGCGTTA AAAAGACAGT GACAGTGGAA ATAACTATAA ACGGAACCGC TAATGAGACT 3840
 GCAACAATAG TGGTTCCTGT TCCTAAGAAG GCCGAAAACA TCGAGGTAAG TGGAGACCAC 3900
 GTAATTTCTT ATAGTATAGA GGAAGGAGAG TACGCCAAGT ACGTTATAAT TACAGTGAAG 3960
 TTTGCATCAC CTGTAACAGT AACTGTTACT TACACTATCT ATGCTGGCCC AAGAGTCTCA 4020
 ATCTTGACAC TTAACCTTCT TGGCTACTCA TGGTACAGAC TATATTCACA GAAGTTTGAC 4080
 GAATTGTACC AAAAGGCCCT TGAATTGGGA GTGGACAACG AGACATTAGC TTTAGCCCTC 4140
 AGCTACCATG AAAAAGCCAA AGAGTACTAC GAAAAGGCC TTAGCCTTAG CGAGGGTAAC 4200
 ATAATCCAAT ACCTTGGAGA CATAAGACTA TTACCTCCAT TAAGACAGGC ATACATCAAT 4260
 GAAATGAAGG CAGTTAAGAT ACTGGAAAAG GCCATAGAAG AATTAGAGGG TGAAGAGTAA 4320
 TCTCCAATTT TTCCACTTT TTCTTTTATA ACATTCCAAG CCTTTTCTTA GCTTCTTCGC 4380
 TCATTCATC AGGAGTCCAT GGAGGATCAA AGGTAAGTTC AACCTCCACA TCTCTTACTC 4440
 CTGGGATTC GAGTACTTTC TCCTCTACAG CTCTAAGAAG CCAGAGAGTT AAAGGACACC 4500
 CAGGAGTGT CATTGTCATC TTTATATATA CCGTTTTGTC AGGATTAATC TTTAGCTCAT 4560
 AAATTAATCC AAGGTTTACA ACATCCATCC CAATTTCTGG GTCGATAACC TCCTTTAGCT 4620
 TTTCCAGAAT CATTCTTCA GTAATTTCAA GGTTCTCATC TTTGGTTTCT CTCACAAACC 4680
 CAATTTCAAC CTGCCTGATA CCTTCTAACT CCCTAAGCTT GTTATATATC TCCAAAAGAG 4740
 TGGCATCATC AATTTTCTCT TTAAA 4765

SEQ ID NO: 6 的资料:

长度: 1398

类型: 氨基酸

链型: 单链

5 拓扑结构: 线性

分子类型: 肽

序列描述:

```

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

```

Tyr Asn Ser Thr Trp Val Ile Asn Ala Leu Gln Phe Ile Gln Glu
 155 160 165

Phe Gly Tyr Asp Gly Ser Gly Val Val Val Ala Val Leu Asp Thr
 170 175 180

Gly Val Asp Pro Asn His Pro Phe Leu Ser Ile Thr Pro Asp Gly
 185 190 195

Arg Arg Lys Ile Ile Glu Trp Lys Asp Phe Thr Asp Glu Gly Phe
 200 205 210

Val Asp Thr Ser Phe Ser Phe Ser Lys Val Val Asn Gly Thr Leu
 215 220 225

Ile Ile Asn Thr Thr Phe Gln Val Ala Ser Gly Leu Thr Leu Asn
 230 235 240

Glu Ser Thr Gly Leu Met Glu Tyr Val Val Lys Thr Val Tyr Val
 245 250 255

Ser Asn Val Thr Ile Gly Asn Ile Thr Ser Ala Asn Gly Ile Tyr
 260 265 270

His Phe Gly Leu Leu Pro Glu Arg Tyr Phe Asp Leu Asn Phe Asp
 275 280 285

Gly Asp Gln Glu Asp Phe Tyr Pro Val Leu Leu Val Asn Ser Thr
 290 295 300

Gly Asn Gly Tyr Asp Ile Ala Tyr Val Asp Thr Asp Leu Asp Tyr
 305 310 315

Asp Phe Thr Asp Glu Val Pro Leu Gly Gln Tyr Asn Val Thr Tyr
 320 325 330

Asp Val Ala Val Phe Ser Tyr Tyr Tyr Gly Pro Leu Asn Tyr Val
 335 340 345

Leu Ala Glu Ile Asp Pro Asn Gly Glu Tyr Ala Val Phe Gly Trp
 350 355 360
 Asp Gly His Gly His Gly Thr His Val Ala Gly Thr Val Ala Gly
 365 370 375
 Tyr Asp Ser Asn Asn Asp Ala Trp Asp Trp Leu Ser Met Tyr Ser
 380 385 390
 Gly Glu Trp Glu Val Phe Ser Arg Leu Tyr Gly Trp Asp Tyr Thr
 395 400 405
 Asn Val Thr Thr Asp Thr Val Gln Gly Val Ala Pro Gly Ala Gln
 410 415 420
 Ile Met Ala Ile Arg Val Leu Arg Ser Asp Gly Arg Gly Ser Met
 425 430 435
 Trp Asp Ile Ile Glu Gly Met Thr Tyr Ala Ala Thr His Gly Ala
 440 445 450
 Asp Val Ile Ser Met Ser Leu Gly Gly Asn Ala Pro Tyr Leu Asp
 455 460 465
 Gly Thr Asp Pro Glu Ser Val Ala Val Asp Glu Leu Thr Glu Lys
 470 475 480
 Tyr Gly Val Val Phe Val Ile Ala Ala Gly Asn Glu Gly Pro Gly
 485 490 495
 Ile Asn Ile Val Gly Ser Pro Gly Val Ala Thr Lys Ala Ile Thr
 500 505 510
 Val Gly Ala Ala Ala Val Pro Ile Asn Val Gly Val Tyr Val Ser
 515 520 525
 Gln Ala Leu Gly Tyr Pro Asp Tyr Tyr Gly Phe Tyr Tyr Phe Pro
 530 535 540

Ala Tyr Thr Asn Val Arg Ile Ala Phe Phe Ser Ser Arg Gly Pro
545 550 555

Arg Ile Asp Gly Glu Ile Lys Pro Asn Val Val Ala Pro Gly Tyr
560 565 570

Gly Ile Tyr Ser Ser Leu Pro Met Trp Ile Gly Gly Ala Asp Phe
575 580 585

Met Ser Gly Thr Ser Met Ala Thr Pro His Val Ser Gly Val Val
590 595 600

Ala Leu Leu Ile Ser Gly Ala Lys Ala Glu Gly Ile Tyr Tyr Asn
605 610 615

Pro Asp Ile Ile Lys Lys Val Leu Glu Ser Gly Ala Thr Trp Leu
620 625 630

Glu Gly Asp Pro Tyr Thr Gly Gln Lys Tyr Thr Glu Leu Asp Gln
635 640 645

Gly His Gly Leu Val Asn Val Thr Lys Ser Trp Glu Ile Leu Lys
650 655 660

Ala Ile Asn Gly Thr Thr Leu Pro Ile Val Asp His Trp Ala Asp
665 670 675

Lys Ser Tyr Ser Asp Phe Ala Glu Tyr Leu Gly Val Asp Val Ile
680 685 690

Arg Gly Leu Tyr Ala Arg Asn Ser Ile Pro Asp Ile Val Glu Trp
695 700 705

His Ile Lys Tyr Val Gly Asp Thr Glu Tyr Arg Thr Phe Glu Ile
710 715 720

Tyr Ala Thr Glu Pro Trp Ile Lys Pro Phe Val Ser Gly Ser Val

725	730	735
Ile Leu Glu Asn Asn Thr Glu Phe Val Leu Arg Val Lys Tyr Asp		
740	745	750
Val Glu Gly Leu Glu Pro Gly Leu Tyr Val Gly Arg Ile Ile Ile		
755	760	765
Asp Asp Pro Thr Thr Pro Val Ile Glu Asp Glu Ile Leu Asn Thr		
770	775	780
Ile Val Ile Pro Glu Lys Phe Thr Pro Glu Asn Asn Tyr Thr Leu		
785	790	795
Thr Trp Tyr Asp Ile Asn Gly Pro Glu Met Val Thr His His Phe		
800	805	810
Phe Thr Val Pro Glu Gly Val Asp Val Leu Tyr Ala Met Thr Thr		
815	820	825
Tyr Trp Asp Tyr Gly Leu Tyr Arg Pro Asp Gly Met Phe Val Phe		
830	835	840
Pro Tyr Gln Leu Asp Tyr Leu Pro Ala Ala Val Ser Asn Pro Met		
845	850	855
Pro Gly Asn Trp Glu Leu Val Trp Thr Gly Phe Asn Phe Ala Pro		
860	865	870
Leu Tyr Glu Ser Gly Phe Leu Val Arg Ile Tyr Gly Val Glu Ile		
875	880	885
Thr Pro Ser Val Trp Tyr Ile Asn Arg Thr Tyr Leu Asp Thr Asn		
890	895	900
Thr Glu Phe Ser Ile Glu Phe Asn Ile Thr Asn Ile Tyr Ala Pro		
905	910	915

Ile Asn Ala Thr Leu Ile Pro Ile Gly Leu Gly Thr Tyr Asn Ala
 920 925 930
 Ser Val Glu Ser Val Gly Asp Gly Glu Phe Phe Ile Lys Gly Ile
 935 940 945
 Glu Val Pro Glu Gly Thr Ala Glu Leu Lys Ile Arg Ile Gly Asn
 950 955 960
 Pro Ser Val Pro Asn Ser Asp Leu Asp Leu Tyr Leu Tyr Asp Ser
 965 970 975
 Lys Gly Asn Leu Val Ala Leu Asp Gly Asn Pro Thr Ala Glu Glu
 980 985 990
 Glu Val Val Val Glu Tyr Pro Lys Pro Gly Val Tyr Ser Ile Val
 995 1000 1005
 Val His Gly Tyr Ser Val Arg Asp Glu Asn Gly Asn Pro Thr Thr
 1010 1015 1020
 Thr Thr Phe Asp Leu Val Val Gln Met Thr Leu Asp Asn Gly Asn
 1025 1030 1035
 Ile Lys Leu Asp Lys Asp Ser Ile Ile Leu Gly Ser Asn Glu Ser
 1040 1045 1050
 Val Val Val Thr Ala Asn Ile Thr Ile Asp Arg Asp His Pro Thr
 1055 1060 1065
 Gly Val Tyr Ser Gly Ile Ile Glu Ile Arg Asp Asn Glu Val Tyr
 1070 1075 1080
 Gln Asp Thr Asn Thr Ser Ile Ala Lys Ile Pro Ile Thr Leu Val
 1085 1090 1095
 Ile Asp Lys Ala Asp Phe Ala Val Gly Leu Thr Pro Ala Glu Gly

1100	1105	1110
Val Leu Gly Glu Ala Arg Asn Tyr Thr Leu Ile Val Lys His Ala		
1115	1120	1125
Leu Thr Leu Glu Pro Val Pro Asn Ala Thr Val Ile Ile Gly Asn		
1130	1135	1140
Tyr Thr Tyr Leu Thr Asp Glu Asn Gly Thr Val Thr Phe Thr Tyr		
1145	1150	1155
Ala Pro Thr Lys Leu Gly Ser Asp Glu Ile Thr Val Ile Val Lys		
1160	1165	1170
Lys Glu Asn Phe Asn Thr Leu Glu Lys Thr Phe Gln Ile Thr Val		
1175	1180	1185
Ser Glu Pro Glu Ile Thr Glu Glu Asp Ile Asn Glu Pro Lys Leu		
1190	1195	1200
Ala Met Ser Ser Pro Glu Ala Asn Ala Thr Ile Val Ser Val Glu		
1205	1210	1215
Met Glu Ser Glu Gly Gly Val Lys Lys Thr Val Thr Val Glu Ile		
1220	1225	1230
Thr Ile Asn Gly Thr Ala Asn Glu Thr Ala Thr Ile Val Val Pro		
1235	1240	1245
Val Pro Lys Lys Ala Glu Asn Ile Glu Val Ser Gly Asp His Val		
1250	1255	1260
Ile Ser Tyr Ser Ile Glu Glu Gly Glu Tyr Ala Lys Tyr Val Ile		
1265	1270	1275
Ile Thr Val Lys Phe Ala Ser Pro Val Thr Val Thr Val Thr Tyr		
1280	1285	1290

Thr Ile Tyr Ala Gly Pro Arg Val Ser Ile Leu Thr Leu Asn Phe
 1295 1300 1305
 Leu Gly Tyr Ser Trp Tyr Arg Leu Tyr Ser Gln Lys Phe Asp Glu
 5 1310 1315 1320
 Leu Tyr Gln Lys Ala Leu Glu Leu Gly Val Asp Asn Glu Thr Leu
 1325 1330 1335
 Ala Leu Ala Leu Ser Tyr His Glu Lys Ala Lys Glu Tyr Tyr Glu
 10 1340 1345 1350
 Lys Ala Leu Glu Leu Ser Glu Gly Asn Ile Ile Gln Tyr Leu Gly
 1355 1360 1365
 Asp Ile Arg Leu Leu Pro Pro Leu Arg Gln Ala Tyr Ile Asn Glu
 1370 1375 1380
 15 Met Lys Ala Val Lys Ile Leu Glu Lys Ala Ile Glu Glu Leu Glu
 1385 1390 1395
 Gly Glu Glu

SEQ ID NO: 7 的资料:

20 长度: 35
 类型: 核酸
 链型: 单链
 拓扑结构: 线性
 分子类型: 其它核酸 (合成的 DNA)
 25 序列描述:

CGWSDRRRTG TTRRHGTHGC DGTDMTYGAC ACBGG

35

SEQ ID NO: 8 的资料:

30 长度: 32
 类型: 核酸
 链型: 单链

拓扑结构: 线性

分子类型: 其它核酸 (合成的 DNA)

序列描述:

KSTCACGGAA CTCACGTDGC BGGHACDGTG GC

32

5

SEQ ID NO: 9 的资料:

长度: 33

类型: 核酸

链型: 单链

10

拓扑结构: 线性

分子类型: 其它核酸 (合成的 DNA)

序列描述:

ASCMGCAACH GTKCCVGCHA CGTGAGTTCC GTG

33

15 SEQ ID NO: 10 的资料:

长度: 34

类型: 核酸

链型: 单链

拓扑结构: 线性

20

分子类型: 其它核酸 (合成的 DNA)

序列描述:

CHCCGSYVAC RTGBGGAGWD GCCATBGAVG TDCC

34

SEQ ID NO: 11 的资料:

25

长度: 1977

类型: 核酸

链型: 双链

拓扑结构: 线性

分子类型: 基因组 DNA

30

原始来源:

生物: 速生热球菌

菌株: DSM 2476**序列描述:**

ATGAAGAGGT TAGGTGCTGT GGTGCTGGCA CTGGTCTCG TGGGTCTTCT GGCCGGAACG 60
 GCCCTTGCGG CACCCGTAAA ACCGGTTGTC AGGAACAACG CGGTTGAGCA GAAGAACTAC 120
 GGACTGCTGA CCCCggGACT GTTCAAGAAA GTCCAGAGGA TGAAGTGGAA CCAGGAAGTG 180
 GACACCGTCA TAATGTTCCG GAGCTACGGA GACAGGGACA GGGCGGTTAA GGTACTGAGG 240
 CTCATGGGCG CCCAGGTCAA GTACTCCTAC AAGATAATCC CTGCTGTGCG GGTAAAATA 300
 AAGGCCAGGG ACCTTCTGCT GATCGCGGGC ATGATAGACA CGGGTACTT CGGTAACACA 360
 AGGGTCTCGG GCATAAAGTT CATAcAGGAG GATTACAAGG TTCAGGTTGA CGACGCCACT 420
 TCCGTCTCCC AGATAGGGGC CGATACCGTC TGGAACTCCC TCGGCTACGA CGGAAGCGGT 480
 GTGGTGGTTG CCATCGTCCA TACGGGTATA GACGCGAACC ACCCCGATCT GAAGGGCAAG 540
 GTCATAGGCT GGTACGACGC CGTCAACGGC AGGTCGACCC CCTACGATGA CCAGGGACAC 600
 GGAACCCACG TTGCGGGTAT CGTTGCCGGA ACCGGCAGCG TTAActCCCA GTACATAGGC 660
 GTCGCCCCCG GCGCGAAGCT CGTCGGCGTC AAGGTTCTCG GTGCCGACGG TTCGGGAAGC 720
 GTCTCCACCA TCATCGCGGG TGTTGACTGG GTCGTCCAGA ACAAGGACAA GTACGGGATA 780
 AGGGTCATCA ACCTCTCCCT CGGCTCCTCC CAGAGCTCCG ACGGAACCGA CTCCTCAGT 840
 CAGGCCGTCA ACAACGCCTG GGACGCCGGT ATAGTAGTCT GCGTCGCCGC CGGCAACAGC 900
 GGGCCGAACA CCTACACCGT CGGCTCACCC GCCGCCGCGA GCAAGGTCAT AACCGTCGGT 960
 GCAGTTGACA GCAACGACAA CATCGCCAGC TTCTCCAGCA GGGGACCGAC CGCGGACGGA 1020
 AGGCTCAAGC CGGAAGTCGT CGCCCCGGC GTTGACATCA TAGCCCCGCG CGCCAGCGGA 1080
 ACCAGCATGG GCACCCCGAT AAACGACTAC TACACCAAGG CCTCTGGAAC CAGCATGGCC 1140
 ACCCCGCACG TTTCGGGCGT TGGCGCGCTC ATCCTCCAGG CCCACCCGAG CTGGACCCCG 1200
 GACAAGGTGA AGACCGCCCT CATCGAGACC GCCGACATAG TCGCCCCCAA GGAGATAGCG 1260
 GACATCGCCT ACGGTGCGGG TAGGGTGAAC GCTTACAAGG CCATCAAGTA CGACGACTAC 1320
 GCCAAGCTCA CCTTCACCGG CTCCGTCCGC GACAAGGGAA GCGCCACCCA CACCTTCGAC 1380
 GTCAGCGGCG CCACCTTCGT GACCGCCACC CTCTACTGGG ACACGGGCTC GAGCGACATC 1440

5 GACCTCTACC TCTACGACCC CAACGGGAAC GAGGTTGACT ACTCCTACAC CGCCTACTAC 1500
 GGCTTCGAGA AGGTCGGCTA CTACAACCCG ACCGCCGGAA CCTGGACGGT CAAGGTCGTC 1560
 AGCTACAAGG GCGCGGCGAA CTACCAGGTC GACGTCGTC GCGACGGGAG CCTCAGCCAG 1620
 5 TCCGGCGGGCG GCAACCCGAA TCCAAACCCC AACCCGAACC CAACCCCGAC CACCGACACC 1680
 CAGACCTTCA CCGGTTCCGT TAACGACTAC TGGGACACCA GCGACACCTT CACCATGAAC 1740
 GTCAACAGCG GTGCCACCAA GATAACCGGT GACCTGACCT TCGATACTTC CTACAACGAC 1800
 CTCGACCTCT ACCTCTACGA CCCCAACGGC AACCTCGMTG ACAGGTCCAC GTCGAGCAAC 1860
 10 AGCTACGAGC ACGTCGAGTA CGCCAACCCC GCCCCGGGAA CCTGGACGTT CCTCGTCTAC 1920
 GCCTACAGCA CCTACGGCTG GCGGACTAC CAGCTCAAGG CCGTCGTCTA CTACGGG 1977

SEQ ID NO: 12 的资料:

15 长度: 659
 类型: 氨基酸
 链型: 单链
 拓扑结构: 线性
 分子类型: 肽
 序列描述:

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

Val Lys Val Leu Arg Leu Met Gly Ala Gln Val Lys Tyr Ser Tyr
 80 85 90
 Lys Ile Ile Pro Ala Val Ala Val Lys Ile Lys Ala Arg Asp Leu
 95 100 105
 Leu Leu Ile Ala Gly Met Ile Asp Thr Gly Tyr Phe Gly Asn Thr
 110 115 120
 Arg Val Ser Gly Ile Lys Phe Ile Gln Glu Asp Tyr Lys Val Gln
 125 130 135
 Val Asp Asp Ala Thr Ser Val Ser Gln Ile Gly Ala Asp Thr Val
 140 145 150
 Trp Asn Ser Leu Gly Tyr Asp Gly Ser Gly Val Val Val Ala Ile
 155 160 165
 Val Asp Thr Gly Ile Asp Ala Asn His Pro Asp Leu Lys Gly Lys
 170 175 180
 Val Ile Gly Trp Tyr Asp Ala Val Asn Gly Arg Ser Thr Pro Tyr
 185 190 195
 Asp Asp Gln Gly His Gly Thr His Val Ala Gly Ile Val Ala Gly
 200 205 210
 Thr Gly Ser Val Asn Ser Gln Tyr Ile Gly Val Ala Pro Gly Ala
 215 220 225
 Lys Leu Val Gly Val Lys Val Leu Gly Ala Asp Gly Ser Gly Ser
 230 235 240
 Val Ser Thr Ile Ile Ala Gly Val Asp Trp Val Val Gln Asn Lys
 245 250 255
 Asp Lys Tyr Gly Ile Arg Val Ile Asn Leu Ser Leu Gly Ser Ser
 260 265 270

Gln Ser Ser Asp Gly Thr Asp Ser Leu Ser Gln Ala Val Asn Asn
 275 280 285

Ala Trp Asp Ala Gly Ile Val Val Cys Val Ala Ala Gly Asn Ser
 290 295 300

Gly Pro Asn Thr Tyr Thr Val Gly Ser Pro Ala Ala Ala Ser Lys
 305 310 315

Val Ile Thr Val Gly Ala Val Asp Ser Asn Asp Asn Ile Ala Ser
 320 325 330

Phe Ser Ser Arg Gly Pro Thr Ala Asp Gly Arg Leu Lys Pro Glu
 335 340 345

Val Val Ala Pro Gly Val Asp Ile Ile Ala Pro Arg Ala Ser Gly
 350 355 360

Thr Ser Met Gly Thr Pro Ile Asn Asp Tyr Tyr Thr Lys Ala Ser
 365 370 375

Gly Thr Ser Met Ala Thr Pro His Val Ser Gly Val Gly Ala Leu
 380 385 390

Ile Leu Gln Ala His Pro Ser Trp Thr Pro Asp Lys Val Lys Thr
 395 400 405

Ala Leu Ile Glu Thr Ala Asp Ile Val Ala Pro Lys Glu Ile Ala
 410 415 420

Asp Ile Ala Tyr Gly Ala Gly Arg Val Asn Val Tyr Lys Ala Ile
 425 430 435

Lys Tyr Asp Asp Tyr Ala Lys Leu Thr Phe Thr Gly Ser Val Ala
 440 445 450

Asp Lys Gly Ser Ala Thr His Thr Phe Asp Val Ser Gly Ala Thr			
	455	460	465
Phe Val Thr Ala Thr Leu Tyr Trp Asp Thr Gly Ser Ser Asp Ile			
	470	475	480
Asp Leu Tyr Leu Tyr Asp Pro Asn Gly Asn Glu Val Asp Tyr Ser			
	485	490	495
Tyr Thr Ala Tyr Tyr Gly Phe Glu Lys Val Gly Tyr Tyr Asn Pro			
	500	505	510
Thr Ala Gly Thr Trp Thr Val Lys Val Val Ser Tyr Lys Gly Ala			
	515	520	525
Ala Asn Tyr Gln Val Asp Val Val Ser Asp Gly Ser Leu Ser Gln			
	530	535	540
Ser Gly Gly Gly Asn Pro Asn Pro Asn Pro Asn Pro Asn Pro Thr			
	545	550	555
Pro Thr Thr Asp Thr Gln Thr Phe Thr Gly Ser Val Asn Asp Tyr			
	560	565	570
Trp Asp Thr Ser Asp Thr Phe Thr Met Asn Val Asn Ser Gly Ala			
	575	580	585
Thr Lys Ile Thr Gly Asp Leu Thr Phe Asp Thr Ser Tyr Asn Asp			
	590	595	600
Leu Asp Leu Tyr Leu Tyr Asp Pro Asn Gly Asn Leu Val Asp Arg			
	605	610	615
Ser Thr Ser Ser Asn Ser Tyr Glu His Val Glu Tyr Ala Asn Pro			
	620	625	630
Ala Pro Gly Thr Trp Thr Phe Leu Val Tyr Ala Tyr Ser Thr Tyr			
	635	640	645

Gly Trp Ala Asp Tyr Gln Leu Lys Ala Val Val Tyr Tyr Gly

650

655

SEQ ID NO: 13 的资料:

- 5 长度: 28
 类型: 核酸
 链型: 单链
 拓扑结构: 线性
 分子类型: 其它核酸 (合成的 DNA)
 10 序列描述:

AGAGGGATCC ATGAAGGGGC TGAAAGCT

28

SEQ ID NO: 14 的资料:

- 15 长度: 28
 类型: 核酸
 链型: 单链
 拓扑结构: 线性
 分子类型: 其它核酸 (合成的 DNA)
 序列描述:

20 AGAGGCATGC GCTCTAGACT CTGGGAGAGT

28

SEQ ID NO: 15 的资料:

- 25 长度: 1962
 类型: 核酸
 链型: 双链
 拓扑结构: 线性
 分子类型: 基因组 DNA
 原始来源:
 生物: 激烈热球菌
 30 菌株: DSM 3638

序列描述:

ATGAAGGGGC TGAAAGCTCT CATATTAGTG ATTTTAGTTC TAGGTTGGT AGTAGGGAGC 60
 GTAGCGGCAG CTCCAGAGAA GAAAGTTGAA CAAGTAAGAA ATGTTGAGAA GAACTATGGT 120
 CTGCTAACGC CAGGACTGTT CAGAAAAATT CAAAAATTGA ATCCTAACGA GGAAATCAGC 180
 ACAGTAATTG TATTTGAAAA CCATAGGGAA AAAGAAATTG CAGTAAGAGT TCTTGAGTTA 240
 ATGGGTGCAA AAGTTAGGTA TGTGTACCAT ATTATACCCG CAATAGCTGC CGATCTTAAG 300
 GTTAGAGACT TACTAGTCAT CTCAGGTTTA ACAGGGGGTA AAGCTAAGCT TTCAGGTGTT 360
 AGGTTTATCC AGGAAGACTA CAAAGTTACA GTTTCAGCAG AATTAGAAGG ACTGGATGAG 420
 TCTGCAGCTC AAGTTATGGC AACTTACGTT TGGAACTTGG GATATGATGG TTCTGGAATC 480
 ACAATAGGAA TAATTGACAC TGGAAATTGAC GCTTCTCATC CAGATCTCCA AGGAAAAGTA 540
 ATTGGGTGGG TAGATTTTGT CAATGGTAGG AGTTATCCAT ACGATGACCA TGGACATGGA 600
 ACTCATGTA CTTCAATAGC AGCTGGTACT GGAGCAGCAA GTAATGGCAA GTACAAGGGA 660
 ATGGCTCCAG GAGCTAAGCT GCGGGGAATT AAGGTTCTAG GTGCCGATGG TTCTGGAAGC 720
 ATATCTACTA TAATTAAGGG AGTTGAGTGG GCCGTTGATA ACAAAGATAA GTACGGAATT 780
 AAGGTCATTA ATCTTTCTCT TGGTTCAAGC CAGAGCTCAG ATGGTACTGA CGCTCTAAGT 840
 CAGGCTGTTA ATGCAGCGTG GGATGCTGGA TTAGTTGTTG TGGTTGCCGC TGGAAACAGT 900
 GGACCTAACA AGTATACAAT CGGTTCTCCA GCAGCTGCAA GCAAAGTTAT TACAGTTGGA 960
 GCCGTTGACA AGTATGATGT TATAACAAGC TTCTCAAGCA GAGGGCCAAC TGCAGACGGC 1020
 AGGCTTAAGC CTGAGGTTGT TGCTCCAGGA AACTGGATAA TTGCTGCCAG AGCAAGTGA 1080
 ACTAGCATGG GTCAACCAAT TAATGACTAT TACACAGCAG CTCCTGGGAC ATCAATGGCA 1140
 ACTCCTCAG TAGCTGGTAT TGCAGCCCTC TTGCTCCAAG CACACCCGAG CTGGACTCCA 1200
 GACAAAGTAA AAACAGCCCT CATAGAACT GCTGATATCG TAAAGCCAGA TGAAATAGCC 1260
 GATATAGCCT ACGGTGCAGG TAGGGTTAAT GCATACAAGG CTATAAACTA CGATAACTAT 1320
 GCAAAGCTAG TGTTCACTGG ATATGTTGCC AACAAAGGCA GCCAACTCA CCAGTTCGTT 1380
 ATTAGCGGAG CTTCGTTGTT AACTGCCACA TTATACTGGG ACAATGCCAA TAGCGACCTT 1440
 GATCTTTACC TCTACGATCC CAATGGAAAC CAGGTTGACT ACTCTTACAC CGCCTACTAT 1500

```

GGATTCGAAA AGGTTGGTTA TTACAACCCA ACTGATGGAA CATGGACAAT TAAGGTTGTA 1560
AGCTACAGCG GAAGTGCAA CTATCAAGTA GATGTGGTAA GTGATGGTTC CCTTTCACAG 1620
CCTGGAAGTT CACCATCTCC ACAACCAGAA CCAACAGTAG ACGCAAAGAC GTTCCAAGGA 1680
5 TCCGATCACT ACTACTATGA CAGGAGCGAC ACCTTTACAA TGACCGTTAA CTCTGGGGCT 1740
ACAAAGATTA CTGGAGACCT AGTGTGTTGAC ACAAGCTACC ATGATCTTGA CCTTACCTC 1800
TACGATCCTA ACCAGAAGCT TGTAGATAGA TCGGAGAGTC CCAACAGCTA CGAACACGTA 1860
GAATACTTAA CCCCCGCCC AGGAACCTGG TACTTCCTAG TATATGCCTA CTACACTTAC 1920
10 GGTGTTGGCTT ACTACGAGCT GACGGCTAAA GTTATTATG GC 1962

```

SEQ ID NO: 16 的资料:

长度: 654

类型: 氨基酸

链型: 单链

15 拓扑结构: 线性

分子类型: 肽

序列描述:

```

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

```

Ile Ile Pro Ala Ile Ala Ala Asp Leu Lys Val Arg Asp Leu Leu			
	95	100	105
Val Ile Ser Gly Leu Thr Gly Gly Lys Ala Lys Leu Ser Gly Val			
	110	115	120
Arg Phe Ile Gln Glu Asp Tyr Lys Val Thr Val Ser Ala Glu Ileu			
	125	130	135
Glu Gly Leu Asp Glu Ser Ala Ala Gln Val Met Ala Thr Tyr Val			
	140	145	150
Trp Asn Leu Gly Tyr Asp Gly Ser Gly Ile Thr Ile Gly Ile Ile			
	155	160	165
Asp Thr Gly Ile Asp Ala Ser His Pro Asp Leu Gln Gly Lys Val			
	170	175	180
Ile Gly Trp Val Asp Phe Val Asn Gly Arg Ser Tyr Pro Tyr Asp			
	185	190	195
Asp His Gly His Gly Thr His Val Ala Ser Ile Ala Ala Gly Thr			
	200	205	210
Gly Ala Ala Ser Asn Gly Lys Tyr Lys Gly Met Ala Pro Gly Ala			
	215	220	225
Lys Leu Ala Gly Ile Lys Val Leu Gly Ala Asp Gly Ser Gly Ser			
	230	235	240
Ile Ser Thr Ile Ile Lys Gly Val Glu Trp Ala Val Asp Asn Lys			
	245	250	255
Asp Lys Tyr Gly Ile Lys Val Ile Asn Leu Ser Leu Gly Ser Ser			
	260	265	270
Gln Ser Ser Asp Gly Thr Asp Ala Leu Ser Gln Ala Val Asn Ala			
	275	280	285
Ala Trp Asp Ala Gly Leu Val Val Val Val Ala Ala Gly Asn Ser			

290	295	300
Gly Pro Asn Lys Tyr Thr Ile Gly Ser Pro Ala Ala Ala Ser Lys		
305	310	315
Val Ile Thr Val Gly Ala Val Asp Lys Tyr Asp Val Ile Thr Ser		
320	325	330
Phe Ser Ser Arg Gly Pro Thr Ala Asp Gly Arg Leu Lys Pro Glu		
335	340	345
Val Val Ala Pro Gly Asn Trp Ile Ile Ala Ala Arg Ala Ser Gly		
350	355	360
Thr Ser Met Gly Gln Pro Ile Asn Asp Tyr Tyr Thr Ala Ala Pro		
365	370	375
Gly Thr Ser Met Ala Thr Pro His Val Ala Gly Ile Ala Ala Leu		
380	385	390
Leu Leu Gln Ala His Pro Ser Trp Thr Pro Asp Lys Val Lys Thr		
395	400	405
Ala Leu Ile Glu Thr Ala Asp Ile Val Lys Pro Asp Glu Ile Ala		
410	415	420
Asp Ile Ala Tyr Gly Ala Gly Arg Val Asn Ala Tyr Lys Ala Ile		
425	430	435
Asn Tyr Asp Asn Tyr Ala Lys Leu Val Phe Thr Gly Tyr Val Ala		
440	445	450
Asn Lys Gly Ser Gln Thr His Gln Phe Val Ile Ser Gly Ala Ser		
455	460	465
Phe Val Thr Ala Thr Leu Tyr Trp Asp Asn Ala Asn Ser Asp Leu		
470	475	480

Asp Leu Tyr Leu Tyr Asp Pro Asn Gly Asn Gln Val Asp Tyr Ser
 485 490 495
 Tyr Thr Ala Tyr Tyr Gly Phe Glu Lys Val Gly Tyr Tyr Asn Pro
 500 505 510
 Thr Asp Gly Thr Trp Thr Ile Lys Val Val Ser Tyr Ser Gly Ser
 515 520 525
 Ala Asn Tyr Gln Val Asp Val Val Ser Asp Gly Ser Leu Ser Gln
 530 535 540
 Pro Gly Ser Ser Pro Ser Pro Gln Pro Glu Pro Thr Val Asp Ala
 545 550 555
 Lys Thr Phe Gln Gly Ser Asp His Tyr Tyr Tyr Asp Arg Ser Asp
 560 565 570
 Thr Phe Thr Met Thr Val Asn Ser Gly Ala Thr Lys Ile Thr Gly
 575 580 585
 Asp Leu Val Phe Asp Thr Ser Tyr His Asp Leu Asp Leu Tyr Leu
 590 595 600
 Tyr Asp Pro Asn Gln Lys Leu Val Asp Arg Ser Glu Ser Pro Asn
 605 610 615
 Ser Tyr Glu His Val Glu Tyr Leu Thr Pro Ala Pro Gly Thr Trp
 620 625 630
 Tyr Phe Leu Val Tyr Ala Tyr Tyr Thr Tyr Gly Trp Ala Tyr Tyr
 635 640 645
 Glu Leu Thr Ala Lys Val Tyr Tyr Gly
 650

SEQ ID NO: 17 的资料:

长度: 25

类型: 核酸

链型: 单链

5 拓扑结构: 线性

分子类型: 其它核酸 (合成的 DNA)

序列描述:

TCTGAATTCG TTCCTTTCG TATGG

25

10 **SEQ ID NO: 18 的资料:**

长度: 20

类型: 核酸

链型: 单链

拓扑结构: 线性

15 分子类型: 其它核酸 (合成的 DNA)

序列描述:

TGTAATGCTG GATCCGCGAG

20

SEQ ID NO: 19 的资料:

20 长度: 25

类型: 核酸

链型: 单链

拓扑结构: 线性

分子类型: 其它核酸 (合成的 DNA)

25 序列描述:

AGAGGCATGC GTATCCATCA GATTTTGGAG

30

SEQ ID NO: 20 的资料:

长度: 20

30 类型: 核酸

链型: 单链

拓扑结构: 线性

分子类型: 其它核酸 (合成的 DNA)

序列描述:

AGTGAACGGA TACTTGGAAC

20

5

SEQ ID NO: 21 的资料:

长度: 20

类型: 核酸

链型: 单链

10

拓扑结构: 线性

分子类型: 其它核酸 (合成的 DNA)

序列描述:

GTTCCAAGTA TCCGTTCACT

20

15 SEQ ID NO: 22 的资料:

长度: 12

类型: 氨基酸

链型: 单链

拓扑结构: 线性

20

分子类型: 肽

序列描述:

Ala Glu Leu Glu Gly Leu Asp Glu Ser Ala Ala Gln

5

10

25 SEQ ID NO: 23 的资料:

长度: 24

类型: 核酸

链型: 单链

拓扑结构: 线性

30

分子类型: 其它核酸 (合成的 DNA)

序列描述:

TCATGGATCC ACCCTCTCCT TTTA

24

SEQ ID NO: 24 的资料:

5

长度: 26

类型: 核酸

链型: 单链

拓扑结构: 线性

分子类型: 其它核酸 (合成的 DNA)

10

序列描述:

GTCCTGCGCAG GCTGCCGGAN NNNNNATGAA GGGGCTGAAA GCTCTC

26

SEQ ID NO: 25 的资料:

长度: 29

15

类型: 核酸

链型: 单链

拓扑结构: 线性

分子类型: 其它核酸 (合成的 DNA)

序列描述:

20

GAGAGCTTTC AGCCCTTCA TNNNNNTTCC GGCAGCCTGC GCAGACATG

29

SEQ ID NO: 26 的资料:

长度: 27

25

类型: 核酸

链型: 单链

拓扑结构: 线性

分子类型: 其它核酸 (合成的 DNA)

序列描述:

30

AGAGGGGGAT CCGTGAGAAG CAAAAA

27

SEQ ID NO: 27 的资料:

长度: 20

类型: 核酸

链型: 单链

5 拓扑结构: 线性

分子类型: 其它核酸 (合成的 DNA)

序列描述:

GATGACTAGT AAGTCTCTAA

20

10 **SEQ ID NO: 28 的资料:**

长度: 20

类型: 核酸

链型: 单链

拓扑结构: 线性

15 分子类型: 其它核酸 (合成的 DNA)

序列描述:

AAGCCTGAGG TTGTTGCTCC

20

SEQ ID NO: 29 的资料:

20 长度: 29

类型: 核酸

链型: 单链

拓扑结构: 线性

分子类型: 其它核酸 (合成的 DNA)

25 序列描述:

GGGCATGCTC ATGAACTTCC AGGCTGTGA

29

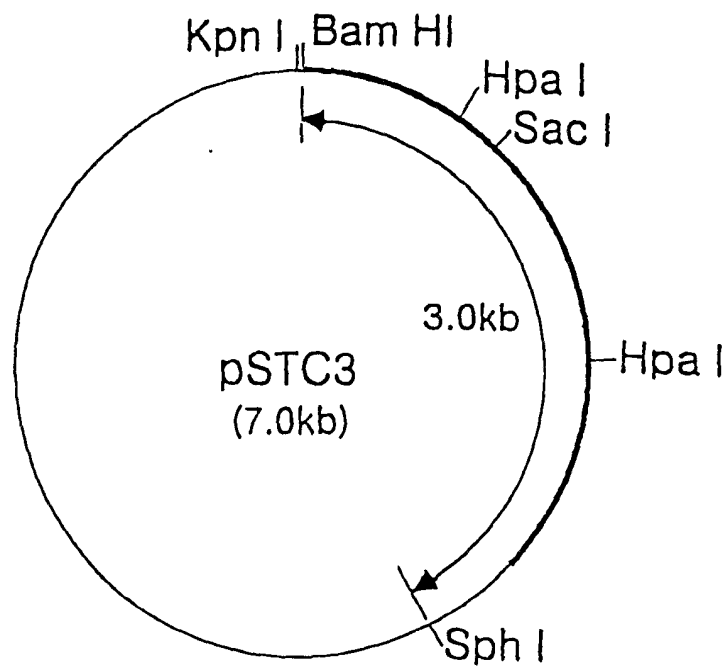


图 1

PFUL	10	20	30	40	50
TCES	MNKKGLTVLF	I A I M L L S V V P	V H F V S A E T P P	V S S E N S T S I	L P N Q Q V V T K E
枯草杆菌蛋白酶		M K R L G A V V	L A L V L V G L L A	G T A L A A P V K P	V V R N N A V Q Q K
					M R G K K V W I S L
PFUL	60	70	80	90	100
TCES	V S Q A A L N A I M	K G Q P N M V L I I	K T K E G K L E E A	K T E L E K L G A E	I L D E N R V L N M
枯草杆菌蛋白酶					L R L M G A Q V K Y
					M S A A K K K D V I
PFUL	110	120	130	140	150
TCES	L L V K I K P E K V	K E L N Y I S S L E	K A W L N R E V K L	S P P I V E K D V K	T K E P S L E P K M
枯草杆菌蛋白酶					Q E D Y K V Q V D D
					D H V A H A Y A Q S
PFUL	160	170	180	190	200
TCES	Y N S T W V I N A L	Q F I Q E F G Y D G	S G V V A V I D T	G V D P N H P F L S	I T P D G R R K I I
枯草杆菌蛋白酶					G K V I G W Y D A V
					V A G G A S M V P S
PFUL	210	220	230	240	250
TCES	E W K D F T D E G F	V D T S F S F S K V	V N G T L I I N T T	F Q V A S G L T L N	E S T G L M E Y V V
枯草杆菌蛋白酶					
PFUL	260	270	280	290	300
TCES	K T V Y V S N V T I	G N I T S A N G I Y	H F G L L P E R Y F	D L N F D G D Q E D	F Y P V L L V N S T
枯草杆菌蛋白酶					

图 2

PFUL	310	320	330	340	350
TCES	---	---	---	---	---
枯草杆菌蛋白酶	GNGYDIAYVD	TDLDYDFTDE	VPLGQYNVTY	DVAVFSYYG	PLNYVLAEID
PFUL	360	370	380	390	400
TCES	---	---	---	---	---
枯草杆菌蛋白酶	PNGEYAVFGW	DGHGHGTHVA	GTVAGYDSNN	DAWDWLSMYS	GEWEVFSRLY
PFUL	410	420	430	440	450
TCES	---	---	---	---	---
枯草杆菌蛋白酶	GWDYTNVTTD	TVQGVAPGAG	IMAIRVLRSD	GRGSMMDIIE	GMTYAATHGA
PFUL	460	470	480	490	500
TCES	---	---	---	---	---
枯草杆菌蛋白酶	---DVISMS	LGGNAPYLDG	TDPEESVAVDE	LTEKYGVVEV	IAAGNEGPGI
PFUL	510	520	530	540	550
TCES	---	---	---	---	---
枯草杆菌蛋白酶	N--IVGSPGV	ATRAITVGA	AVPINVGVVV	SQALGYPDYY	GFYYPAYTN
PFUL	560	570	580	590	600
TCES	---	---	---	---	---
枯草杆菌蛋白酶	VRIAFFSSRG	PRI DGEI KPN	VVAPGYCTYS	SLEPMWIGGAD	F-----MS
PFUL	---	---	---	---	---
TCES	---	---	---	---	---
枯草杆菌蛋白酶	---LAFSSRG	PTA DGRLLKPE	VVAPGVDTIA	PRASGTSMTG	PINDYTKAS
PFUL	---	---	---	---	---
TCES	---	---	---	---	---
枯草杆菌蛋白酶	---RASESSVC	PELD-----	VMAFGVSIQS	TLPFGNKYGA-	-----YN

图 3

PFUL	610	620	630	640	650
TCES	GTSMATPHVS	GVVALISGA	KAEGIYNPD	IISKVLESGA	AWLECDPYTG
枯草杆菌蛋白酶	GTSMATPHVS	GVVALIQAH	PSWTPDKVKT	-----ALLETA	DIVAPKEIAD
	GTSMASPHVA	GAVALISKEH	ENWANTQVRS	-----SLENTT	TKL-GDS---
PFUL	660	670	680	690	700
TCES	QKYTELDQGH	GLVNVTKSWE	ILKAINGTTL	PIVDHWADKS	YSDFAEYLGV
枯草杆菌蛋白酶	-----IAYGA	GRVNVYKALK	YDDYAKLTFT	GSVADKGSAT	HTFDVSGATF
	-----FYXGK	GLINVAQAQ*			
PFUL	710	720	730	740	750
TCES	DVIRGLYARN	SIPDIVEWHI	KYVGDTEYRT	FEIYATEPWI	KPFVSGSVIL
	VTATLYWDTG	SSDIDLLYD	PNGNEVDYSY	TAYYGFKEVG	YYNPTAGTWT
PFUL	760	770	780	790	800
TCES	ENNTEFVLRV	KYDVEGLEPG	LYVGRIIIDD	PPTPVIIDEI	LNTIIVIPEKF
	VKVVSYKGA	NYQVDVSDG	SLSQSGGPN	NPNPNPPTP	TTDTQTFTGS
PFUL	810	820	830	840	850
TCES	TPENNYTLTW	YDINGPEMVT	HHFFTVP	DVLYAMTTYW	DYGLYRDPDM
	VNDYWDTSDT	FTMNVNSGAT	KITGDLTFDT	SYNDL	DPNGNLVDRS
PFUL	860	870	880	890	900
TCES	FVFPYQLDYL	PAAVSNPMPG	NWELVWTGFN	FAPLYESGFL	VRIYGV
	TSSNSYEHVE	YANPAPGTWT	FLVYAYRTYG	WADYQLKAVV	YYG*
PFUL	910	920	930	940	950
	SVWYINRTYL	DTNTEFSIEF	NITNIYAPIN	ATLIPIGLGT	YNASVESVGD

图 4

PFUL	960	970	980	990	1000
	GEFFIKGIEV	PEGTAELKIR	IGNPSV PNSD	LDLYLYDSKG	NLVALDGNPT
PFUL	1010	1020	1030	1040	1050
	AEEEEVVEYP	KPGVYSIVVH	GYSVRDENG	PTTTFDLVV	QMTLDNGNIK
PFUL	1060	1070	1080	1090	1100
	LDKDSIILGS	NESVVVTANI	TIDRDHPTGV	YSGIIEIRD	EVYQDNTSI
PFUL	1110	1120	1130	1140	1150
	AKIPITLVID	KADFAVGLTP	AEGVLGEARN	YTLIVKHALT	LEPVPNATVI
PFUL	1160	1170	1180	1190	1200
	IGNYTYLTDE	NGTVTFYAP	TKLGSDEITV	IVKKENFNITL	EKTFQITVSE
PFUL	1210	1220	1230	1240	1250
	PEITEEDINE	PKLAMSSPEA	NATIVSVEME	SEGGVKKTVT	VEITINGTAN
PFUL	1260	1270	1280	1290	1300
	ETATIVVPVP	KKAENIEVSG	DHVISYSIEE	GEYAKYVIIT	VKFASPVTVT
PFUL	1310	1320	1330	1340	1350
	VTYTIYAGPR	VSILTLNFLG	YSWYRLYSQK	FDELYQKALE	LGVDNETLAL
PFUL	1360	1370	1380	1390	1400
	ALSYHEKAKE	YYEKALELSE	GNIIQYLGDI	RLPLRQAY	INEMKAVKIL
PFUL	1410				
	EKAIEELEGE	E*			

图 5

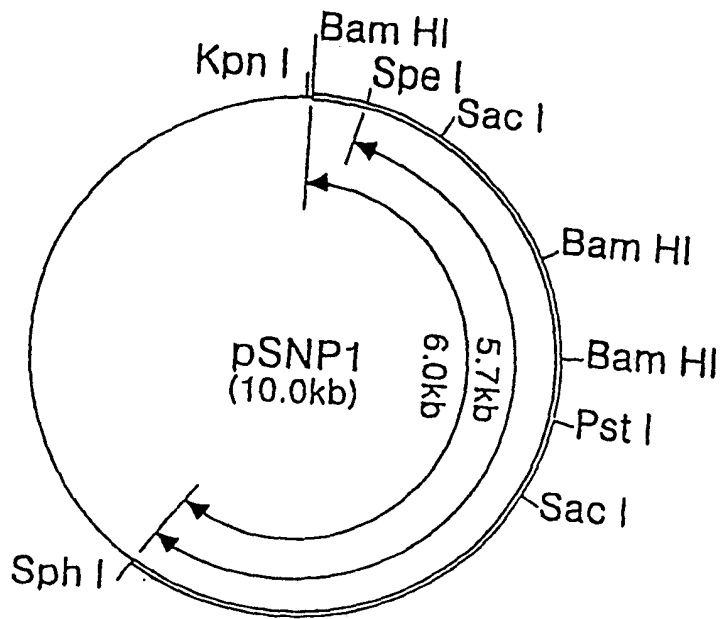


图 6

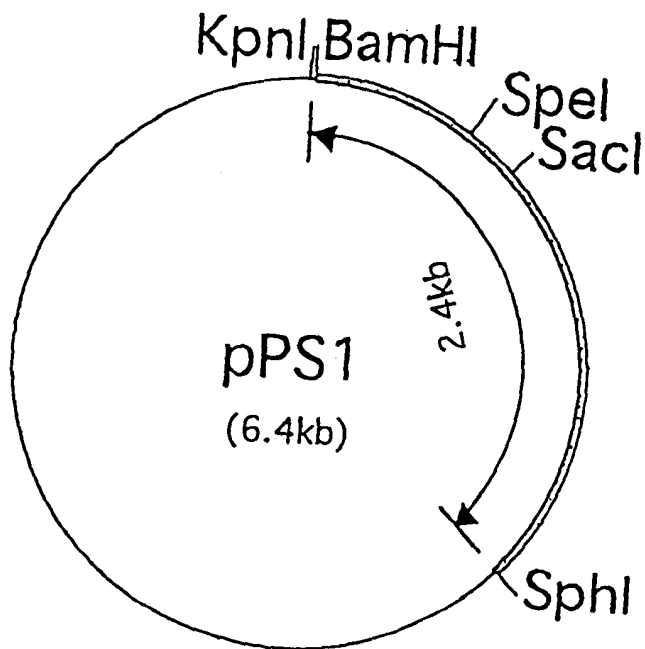


图 7

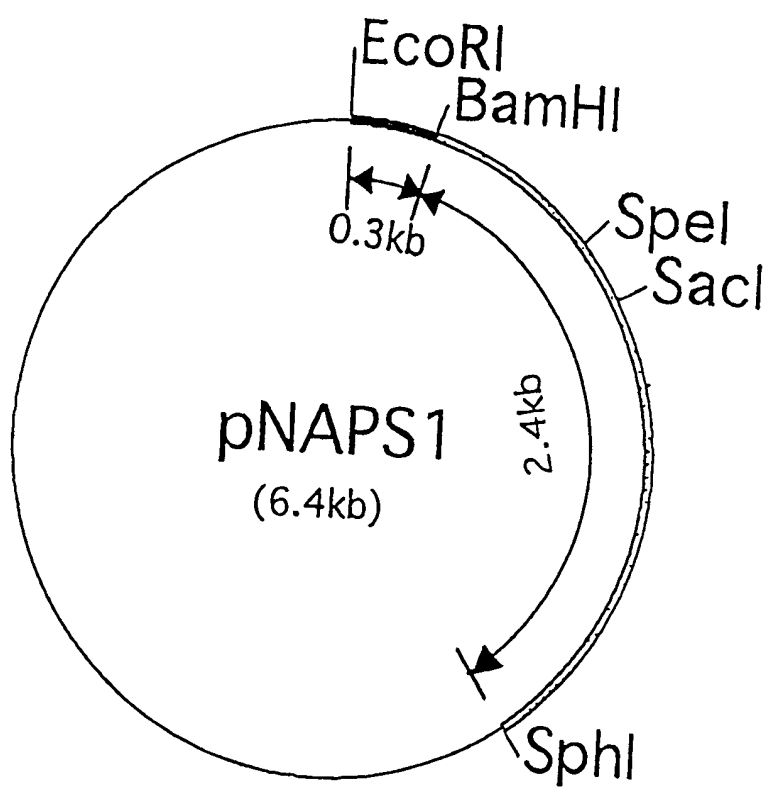


图 8