



# [12] 发明专利申请公开说明书

[21] 申请号 01820348.5

[43] 公开日 2004 年 12 月 8 日

[11] 公开号 CN 1553919A

[22] 申请日 2001.10.10 [21] 申请号 01820348.5

[30] 优先权

[32] 2000.10.10 [33] GB [31] 0024824.8

[86] 国际申请 PCT/GB2001/004509 2001.10.10

[87] 国际公布 WO2002/031157 英 2002.4.18

[85] 进入国家阶段日期 2003.6.10

[71] 申请人 生物科技医药公司

地址 挪威特罗姆瑟

[72] 发明人 英奇·尼尔森 克斯蒂·奥弗博

奥拉夫·莱恩斯

[74] 专利代理机构 北京市柳沈律师事务所

代理人 巫肖南 封新琴

权利要求书 2 页 说明书 24 页 附图 11 页

[54] 发明名称 虾碱性磷酸酶

[57] 摘要

本发明涉及一种核酸分子，该分子包含：SEQ ID Nos. 1 或 20 中所示的核苷酸序列，或与其有至少 55% 的同一性的序列，和/或者能与 SEQ ID Nos. 1 或 20 的互补序列在中等严紧度的条件下杂交的序列，或所述序列的互补序列，其中所述核苷酸序列编码或互补于编码不耐热碱性磷酸酶的序列。本发明还涉及一种重组不耐热碱，其包含：(a) SEQ ID No. 2 所示氨基酸序列的全部或主要部分；或(b) 与 SEQ ID No. 2 所示氨基酸序列有至少 60% 的序列同一性的氨基酸序列的全部或主要部分，及其制造方法。

1 KAYWNRDAQDALKQIGIKLREKQAKNVIFFLGDGMSLSTVTAARIYKGG  
58aaP6:17  
51 LTGKFEREKISWEEFDPAALSKTYNTDRQVDSASAATAYLTGVKTNQGV  
58aaP6:17  
101 IGLDANTVRTNCSYQLDESLFTYSIAHWFQEAQRSTGVVTSRVTHTATPA  
151 GTYAHVADRDWENDSDVVHDREDPEICDDIABQLVFPKPKFKVIMGGG  
58aaP6:26  
201 RRGFFPEEALDIEDGIPGEREDCKHLITDNLDDKASQGATASYVWNRDDL  
823:30  
251 LAVDIRNTDYLMLGLFSYTHLDTVLTRDAEMDPTLPEMTKVAIEMLTKDEN  
301 GFFLLVEGGRIDHMHANQIRQSLAETLDMEEAVSMALSMITDPEETILLV  
351 TADHGHTLITITGYADRNVDILDFAGISLDLDRRYTILDYGSPPGYHTED  
401 GKRYEPTEDLKDINERYASAAPKHSVTHDGTDVGIWVNGPFAHLFTGVY  
823:18  
451 EENYIPHALAYAACVGTGRFTCDK \*

ISSN 1008-4274

1. 一种核酸分子，其包含 SEQ ID No 1 或 20 所示核苷酸序列，或与 SEQ ID No 1 或 20 有至少 55% 的序列同一性的核苷酸序列，和/或者能与 SEQ ID No 1 或 2 的互补序列在中等严紧度的条件下杂交的核苷酸序列，或所述序列的互补序列，其中所述核苷酸序列编码了或互补于编码一种不耐热碱性磷酸酶的序列。
2. 权利要求 1 的核酸分子，其包含与 SEQ ID No 1 或 20 的核苷酸序列有至少 75% 的序列同一性的核苷酸序列。
3. 权利要求 1 或 2 的核酸分子，其包含与 SEQ ID No 1 或 20 的核苷酸序列有至少 85% 的序列同一性的核苷酸序列。
4. 一种分离的的核酸分子，其编码或互补于编码不耐热碱性磷酸酶的序列，所述碱性磷酸酶包含 SEQ ID No. 2 的氨基酸序列或与该氨基酸序列至少 50% 同源的氨基酸序列变体。
5. 权利要求 3 的核酸分子，其中所述碱性磷酸酶具有与 SEQ ID No. 2 至少 70% 相同的氨基酸序列。
6. 权利要求 3 的核酸分子，其中所述碱性磷酸酶具有与 SEQ ID No. 2 至少 80% 相同的氨基酸序列。
7. 前述任一项权利要求的核酸分子，其中所述碱性磷酸酶是北方长额虾碱性磷酸酶。
8. 一种遗传构建体，其包含与启动子序列可操作连接的权利要求 1-7 任一项所述核酸分子。
9. 一种重组表达载体，其包含权利要求 1-7 任一项的核酸分子，并且能在宿主细胞中繁殖。
10. 权利要求 9 的载体，是质粒或病毒载体。
11. 一种细胞，其被权利要求 9-11 任一项所述构建体或载体转化。
12. 权利要求 11 的细胞，是原核细胞。
13. 权利要求 11 的细胞，是真核细胞培养物的一部分。
14. 一种重组细胞或生物，具有含权利要求 1-3 任一项定义的核苷酸序列的核酸分子。
15. 一种重组不耐热碱性磷酸酶，包含：

- a) SEQ ID No. 2 所示氨基酸序列的全部或主要部分; 或
- b) 与 SEQ ID No. 2 有至少 60% 同一性的氨基酸序列的全部或主要部分。

16. 权利要求 15 的重组碱性磷酸酶, 具有 SEQ ID No. 2 所示的氨基酸  
5 序列或与 SEQ ID No. 2 的序列有至少 75% 的序列同一性的氨基酸序列。

17. 权利要求 15 或 16 的重组碱性磷酸酶, 是北方长额虾碱性磷酸酶。

18. 一种能表达不耐热碱性磷酸酶的重组细胞或生物, 所述酶包含 SEQ  
ID No. 2 的氨基酸序列或与 SEQ ID No. 2 有至少 60% 序列同一性的氨基酸  
序列。

10 19. 一种能表达不耐热碱性磷酸酶的重组细胞或生物, 所述酶包含 SEQ  
ID No. 2 的氨基酸序列或与 SEQ ID No. 2 有至少 75% 序列同一性的氨基酸  
序列。

20. 一种制备不耐热碱性磷酸酶的方法, 包括用权利要求 9 的表达载  
体转化宿主细胞, 在培养基中培养该细胞, 分离该细胞所表达的不耐热碱  
15 性磷酸酶。

21. 一种核酸分子, 其包含 SEQ ID No. 15 所示核苷酸序列, 或与 SEQ  
ID No. 15 有至少 55% 的序列同一性的核苷酸序列, 和/或能与 SEQ ID No. 15  
的互补序列在中等严紧度的条件下杂交的核苷酸序列, 或所述序列的互补  
序列, 其中所述核苷酸序列编码了或互补于编码不耐热碱性磷酸酶的序列。

20 22. 一种重组不耐热碱性磷酸酶, 其包含:

- a) SEQ ID No. 16 所示氨基酸序列的全部或主要部分; 或
- b) 与 SEQ ID No. 16 所示氨基酸序列有至少 60% 同一性的氨基酸序列  
的全部或主要部分。

23. 权利要求 22 的重组碱性磷酸酶, 具有 SEQ ID No. 16 所示的氨基  
25 酸序列或与 SEQ ID No. 16 有至少 75% 的序列同一性的氨基酸序列。

24. 一种制备权利要求 的载体的方法。

虾碱性磷酸酶

5 本发明涉及碱性磷酸酶,具体涉及到源自北方长额虾(*Pandalus borealis*)的这种酶和编码该酶的核酸分子。

碱性磷酸酶(E.C. 3.1.3.1),又称为碱性磷酸单酯酶,磷酸单酯酶或甘油磷酸酶,是在碱性条件下有最适宜的酶活性的正磷酸-单酯磷酸水解酶。碱性磷酸酶(ALP)的底物为 DNA, RNA, 和核(糖核)苷与脱氧核(糖核)苷三磷酸。水解产生醇和正磷酸盐。换一角度来说, ALP 使 DNA, RNA, rNTP, dNTP 去磷酸化。不同 ALP 的蛋白质去磷酸化作用也有报道。ALP 在自然界中广泛存在,在从细菌到人类的多种生物中都可以找到。复杂生物一般既含有组织特异性 ALP 又含有非特异性 ALP。这些多肽的大小从 15kDa 到 170kDa 不等。其中一些蛋白结合或“锚定”在细胞膜上。最常见的是,这些酶必需二价的金属阳离子,例如  $Mg^{2+}$ , 或  $Zn^{2+}$ 。

现在在分子生物学中对 ALP 的应用集中在,但并不局限于,三种主要的分析或制备 DNA 的方法: 1)在限制性酶消化后,使载体 DNA 去磷酸化,以便消除克隆载体的自我连接,从而有利于使插入子与载体连接并产生重组构建体, 2)在 PCR 扩增后使 dNTP 去磷酸化,结合应用单链核酸外切酶水解引物使之变成 dNTP,从而省略了 PCR 产物直接进行 DNA 测序前的杂质清除的步骤,或 3)使 DNA 末端去磷酸化,而方便随后用  $[\gamma -^{32}P]NTP$  和 T4 多核苷酸激酶进行  $^{32}P$  标记。所描述的 ALP 酶反应是 DNA 分析过程的中间步骤。本技术领域已知的其它重要应用包括那些把酶活性当作标示的应用,例如在酶联免疫吸附试验(ELISA)中,在基因融合或基因传递系统中,或与寡核苷酸偶联用做杂交探针时。

市售产品中所用的三种主要 ALP 是:

- I) 小牛肠碱性磷酸酶(CIAP)
- II) 产自北极的北方长额虾的虾碱性磷酸酶(SAP)
- III) 细菌碱性磷酸酶(BAP)

30 动物的 CIAP 和 SAP 有着相近的比活性,为 2000-4000 单位/mg 蛋白,而 BAP 蛋白的比活性为 50 单位/mg,因此前两种酶作为高效酶更有吸引力。

如下面所述，SAP 在中等加热条件下失活，所以它优选应用于一些在进行后续步骤之前需要将酶活性除去的应用中。

已有人报道了一种基因工程化温度敏感型 BAP 变体[Shandilya, H. And Chatterjee, D.K., 1995. 一种用于 DNA 去磷酸化的基因工程热敏感型碱性磷酸酶。Focus, 17 (3) : 93-95]，但没有给出其工程化细节。这种由 Life Technologies, Inc.出售的突变酶(TsAP)在只有 EDTA 存在的情况下因加热(65 °C, 15 分钟)失活(95% 或更多)。这种突变体和野生型酶的建议反应温度也是 65 °C。TsAP 的酶活性比野生型 BAP 的酶活性高 40 倍。由于 TsAP 很高的比活性，使得它几乎可以与其它的 ALP 如 CIAP 和 SAP 相当。

10 两种嗜冷微生物的热敏感型 ALP 已经被纯化并定性[de Prada, P., and Brenchley, J.E., 1997, 来自一种嗜冷性节杆菌(*Arthrobacter*)分离株的两种细胞外碱性磷酸酶的纯化与鉴定, Appl. Env. Microbiol.,63(7): 2928-2931]。两种酶的底物特异性和动力学特征不同，它们显示出不同的热敏感性，其中对热最敏感的一种具有与 SAP 相似的热敏感性。没有关于它们的以单位/mg  
15 表示的比活性和一级结构的报道。

一种大西洋鳕的适冷型 ALP 被分离出并定性[Asgeirsson, B., Hartemink, R., and Chlebowski, J.F.,1995, 大西洋鳕(*Gadus morhua*)碱性磷酸酶。能显示其对低温的适应的动力学和结构性质。Comp. Biochem. Physiol., 110B(2) : 315-329]。这种酶显示出与 SAP 相似的热敏感性。没有提供其基因/蛋白的一级结构。

一项关于鲑鱼的 ALP 同工酶的研究[Whitmore, D.H., and Goldberg, E., 1972, Trout intestinal alkaline phosphatase II. The effect of temperatures upon enzymatic activity in vitro and in vivo. J. Exp. Zool., 182:59-68]表明，环境温度会影响同工酶模式，并且一些同工型是热敏感型的。

25 台湾周围温水海域中的虾含有几种 ALP[Lee, A.-C., and Chuang, N.-N., Characterization of different molecular forms of alkaline phosphatase in the hepatopancreas from shrimp *Penaeus monodon* (Crustacea: Decapoda). Comp. Biochem. Physiol., 99B(4): 845-850]，这些酶被定论为热稳定型。没有提供一级结构。

30 在捕虾工业处理废水中发现含有一种来自北极的北方长额虾肝胰脏的碱性磷酸酶[Olsen, R.L., Johansen, A., and Myrnes, B., 1990, Recovery of

enzymes from shrimp waste. *Process Biochem.* 25: 67-68], 后来这种酶被从肝胰脏中纯化出来 [Olsen, R. L., Øverbø, K. and Myrnes, B., 1991, Alkaline phosphatase from hepatopancreas of Shrimp (*Pandalus Borealis*): a dimeric enzyme with catalytically active subunits. *Comp. Biochem. Physiol.* 99B (4): 755-761]。这种纯化蛋白的表观分子量为 65kDa(每个亚单位), 且显示为一种具有催化活性亚单位的二聚体酶, 而其它大多数动物性 ALP 与它相反, 它们必需经过二聚化才能发挥活性。根据该报道, SAP 的等电点为 3.7。虽然 5' 突出末端比平末端或 5' 凹陷末端反应性更强, 这种虾酶可以从限制性内切酶产生的任何 DNA 链末端(5'和 3'突出末端或平末端)有效地去除 5' 末端磷酸。

相对于 CIAP, SAP 活性所需的温度稍低, 但并不认为它是真正的低温激活。在大约 40°C 时酶活性达到最高(CIAP 为 45°C), 在 10°C 或 50°C 时酶活性保留了几乎 40%, 而与之相比的 CIAP 分别保留了 10% 和 90% 的活性。虽然 SAP 活性达到最高时的温度接近 40°C, 但该酶在高于 37°C 的条件下预热 15 分钟就会开始失去活性。在 65°C 的条件下预热后, 酶活性会失去 95% 或更多, 而在 70°C 预热后活性就检测不出来了。与之相比, CIAP 经过相似的热处理还能分别保留 40% 和 20% 的活性。

所以, 相对于它的商业竞争产品, SAP 是热敏感的, 并且是低温激活, 使得它特别适用于那些多步骤的实验室方案, 在这样的方案中一个简单的加热步骤就可以使 SAP 失活, 使它不参与以后的方法步骤。

BIOTEC ASA, Tromsø, Norway, 从捕虾工业废水中生产商用 SAP。在捕虾船上, 刚捕的鲜虾被冻成大块。船靠岸后, 这些虾被再循环的冷水小心地解冻。每 4000 千克虾大约用 1000 升水。在冷冻与解冻的过程中, 虾的肝胰脏破裂, 里面的物质就释放到水中。然后这些废水被回收, 再经过几次层析步骤来纯化 SAP。

生产商们通过一些知名的公司例如 USB, Boehringer-Mannheim 或 Amersham Pharmacia Biotech, 向全球市场供应 SAP。

由于它的高比活性和 DNA 末端的多功能性, 和它的相对热敏感性, SAP 经常被用于在连接反应之前使克隆载体去磷酸化, 在 DNA 测序反应之前对 PCR 扩增产物混合物进行处理, 如美国专利 5,741,676 和 5,756,285 所述。

现在的 SAP 生产受到了捕虾废水质量不稳定的困扰,影响了生产效率。产生质量不稳定的因素有, I)虾体内酶的生产随自然季节而变化,和 II)虾源冷冻之前或期间的操作以及解冻期间或之后对虾或水的操作。

5 还存在对废水的未来可利用性的担忧; I)虾作为一种自然资源,并不能保证一直都能有供应; II)捕虾工业正在寻找一种新的冷冻虾的方法,也就是,单冻法,通过这种方法可检测出废水中酶,例如 SAP 的含量十分小。

除了这些实际的问题,市场还有对重组 SAP 产品的需求,因为重组产品经常被优选应用于分子生物学技术中,特别是当产品纯度成为要注意的问题时,例如,在生产基于 DNA 的治疗药物时或在法医学领域中。SAP 在酶学与物理化学性质上有着很大的优势,所以在应用它的实验室操作中是10 优选的酶。所以,市场会对单一纯化形式的合成或重组 SAP 有可观的需求。然而, SAP 的 DNA 和氨基酸序列此前还没有被阐明。

为了分离基因然后克隆并生产重组 SAP,已有多人尝试获取纯化蛋白的 N 末端序列。但是这些尝试都没有成功。序列分析显示,每一具体 N 末端15 位点有多个(4 个或多个)可替换的氨基酸。所以,即使用上了高度简并寡核苷酸探针,也没有得到可用来获得 SAP 基因的蛋白质序列信息。在表面上看来均相的酶制品中出现非同源 N 末端的原因就只有推测了。可能是因为这些 SAP 同工型是由不同基因产生,经过不同剪接,或经过不同的蛋白质翻译后修饰,再或 SAP 被蛋白酶攻击过,但检测不到分子量的减少。

20 本发明人还发现,北极的北方长额虾肝胰脏碱性磷酸酶与已经过测序的其他物种的碱性磷酸酶之间有限的同一性,妨碍了使用常规方法分离编码北方长额虾碱性磷酸酶的核酸序列。

虽然要得到 SAP 的相关共有氨基酸序列有这些困难,本发明人还是令人惊奇的从北方长额虾的 cDNA 文库中分离出了编码 SAP 的 cDNA,从而25 阐明了 SAP 的编码序列。本发明涉及 SAP cDNA 的制备和它在提供这种酶的另一新来源中的应用。

SEQ ID NO.1 对应于该酶除去 5'端一小部分序列后的 cDNA 全序列。然而,包含了这段序列的核酸分子的表达产物被认为拥有天然酶全部或至少大部分(就是至少 60%, 70%, 80%或通常至少 90%)的活性。这个序列30 末端丢失的几个 N 末端氨基酸不包括底物或辅助因子的结合位点。

所以一般认为这里涉及的 SAP 或热敏感型碱性磷酸酶还包括那些比自然产生的 ALP 小,但是却有相同或基本相同酶活性的分子。更加有利的是,这些本发明的表达产物和其它重组 ALP 酶可以比分离出的自然产生的酶有更强的活性。优选比活性高于 SAP 公布值  $1900 \text{ U mg}^{-1}$ 。测量酶活性所适用方法按 [Olsen et al.] 中所描述。

因此在一方面,本发明提供了一种(分离的)核酸分子,该分子包括(优选组成为)SEQ ID No. 1 所示核苷酸序列,或者与之有至少 55% 的序列同一性的序列,和/或能在中严紧度条件下与 SEQ ID No. 1 的互补序列杂交的序列,或与上述序列的互补序列,其中所述核苷酸序列编码一热敏感型碱性磷酸酶或与该酶编码序列互补序列杂交的序列。

依照本发明的核酸分子因此可以是单链或双链的 DNA, cDNA, 或 RNA。

本发明的核酸分子可以是分离的核酸分子(或换句话说就是从自然界中常见的多种成分中分离出),或可以是重组的或合成的核酸分子。

这里涉及基本上由 SEQ ID No. 1 组成的序列和其变体。因此两个末端也可能有侧翼区。特别是,5'端侧翼区可能有 1-81 或 1-60 个核苷酸,更特别的是 1-30 个核苷酸,这些侧翼区带有用于指导转录/翻译过程中对酶的活性成熟形式进行加工的信号序列。3'侧翼区通常包含 1-60,如 1-30 个额外的核苷酸。

任一给出序列的互补序列将是可以按照碱基配对原则来确定的唯一确定序列。

优选上述核苷酸序列与 SEQ ID No. 1 或其互补序列有至少 65% 的,更优选至少 75%,例如至少 85% 或至少 95% 的序列同一性。同样地,所述核苷酸序列优选在高严紧条件下与 SEQ ID No. 1 或其互补序列杂交。本发明是基于对 SAP 的氨基酸序列和 cDNA 序列的鉴定,但在本领域所共识的是,这些序列的变体,例如等位基因变体和经单个或多个碱基取代、插入或缺失所修饰过的,并且在功能上与新鉴定的序列等价的相关序列是存在并且可以产生的。

本发明一个特别优选的实施例是一种分离的编码真核细胞热敏感型碱性磷酸酶和这些酶的衍生物的核酸分子,优选所述酶来自虾。所述衍生物可以包含保守的氨基酸取代,其中,天然酶的氨基酸可以被具有相似功能

特性的氨基酸取代, 所述特性例如大小, 亲脂性, 极性。氨基酸上与功能相关的基团为本领域技术人员已知。另外可包括不会明显影响酶的催化活性的插入、缺失或取代。一般而言, 酶衍生物与野生型酶相比会有 1-50, 优选 1-30, 例如 1-15 个非保守改变。

5 “功能等价”一词是指核酸分子编码的多肽具有碱性磷酸酶活性, 优选它们的比活性与从北方长额虾中分离出的野生型酶相比, 相同或更高。换句话说, 功能等价的核酸序列编码的多肽可以催化去除由, 例如, 限制性内切酶产生的 DNA 链末端(5'和 3'突出末端或平末端)的末端 5'磷酸基团的反应。

10 编码碱性磷酸酶的核苷酸序列的变异可以发生在北方长额虾物种的不同群体之间, 不同地域的相似群体之间, 也可以发生在相同生物中。这些变异包括在了本发明涉及的范围中。

为了定义严紧度的程度, 中严紧度包括了在 42°C 至 65°C 之间, 0.2 x SSC -2 x SSC 缓冲液, 1% (w/v) SDS 的条件下进行的杂交和/或洗涤, 高严紧度  
15 包括了在至少 55°C 以上, 0.1 x SSC -0.2 x SSC 缓冲液, 0.1% (w/v) SDS 的条件下进行的杂交和/或洗涤。杂交和洗涤的条件是本领域技术人员熟知的。

本发明范围内还包括能在高严紧度条件下与 SEQ ID No. 1 或其互补序列, 再或其部分序列(就是衍生自 SEQ ID No. 1 的杂交探针)杂交并编码碱性磷酸酶或其部分的核酸序列。高严紧度的条件可以容易地根据本领域熟知  
20 的技术来选择, 如 Sambrook et al., 1989, 分子克隆实验手册第二版所述。本发明范围内包括的杂交序列在非严紧条件(室温, 6 x SSC/50% 甲酰胺)下发生结合, 在高严紧度条件(例如, 0.1 x SSC, 68°C)下进行洗涤。其中 SSC= 0.15 M NaCl, 0.015M 柠檬酸钠, pH 7.2。

杂交探针可以是 SEQ ID No. 1(或互补序列)的部分序列, 它的碱基长度  
25 和组成应足以使它能与样品或测试的核酸序列杂交, 从而判断在高严紧度杂交是否发生。所以, 探针长度至少有 15 个碱基, 优选至少 30, 40, 50, 75, 100 或 200 个碱基的长度。有代表性的探针长度从而包括 30-500 个碱基的长度, 例如 30-300, 50-200, 50-150, 75-100 个碱基。

核苷酸序列同一性的判断可以应用 University of Wisconsin 设计的  
30 Genetics Computer Group(GCG)Version 10 软件包的 BestFit 程序。该程序的

使用 Smith 和 Waterman 的局部同源性运算法并设定以下默认值: Gap 生成罚分值=50, Gap 延伸罚分值=3, 平均匹配值=10,000, 平均错配值=-9.000。

产生本文所述具体变体的方法, 例如定点诱变, 随机诱变, 或酶裂解和/或核酸连接的方法, 还有通过, 例如杂交, 确定如此改造过的核酸是否与主题序列有显著同源性的方法, 这些方法都是本领域所熟知的。

换个角度来看, 本发明提供了一种分离的编码了或互补于热敏感型碱性磷酸酶编码序列的核酸分子, 其中所述碱性磷酸酶包含 SEQ ID No. 2 的氨基酸序列或其变体, 该变体的序列至少 50%, 优选至少 60%, 如至少 70% 或 80% 与 SEQ ID No. 2 相同。

10 本发明还提供了一种 cDNA, 它编码包含图 2 中氨基酸序列 (SEQ ID No. 2) 的热敏感型碱性磷酸酶, 或拥有热敏感性和碱性磷酸酶活性的该酶的变体或片段, 优选该 cDNA 衍生自北方长额虾。

本发明还提供了一种 cDNA, 它编码一种热敏感型碱性磷酸酶, 该酶由在 SEQ ID No. 2 序列上缺失, 替换或插入一个或多个氨基酸残基的序列组成, 优选该 cDNA 衍生自北方长额虾。

本发明更进一步提供了一种由本发明里定义过的核酸分子所编码的多肽。根据本发明定义的核酸分子, 使得能够以从前所达不到的数量与纯度来获取重组碱性磷酸酶, 特别是 SAP 和它的变体, 从而能够持续供应这种酶, 导致完全能在任何时间、于可控制的条件下或按指定标准的条件下进行大规模生产。

因此, 本发明内容还扩展到重组多肽, 它包括(或基本组成为)由图 1 所示核酸序列 (SEQ ID No.1) 编码的构成热敏感型碱性磷酸酶的氨基酸序列, 或其功能等价的变体。从另一角度看, 本发明还提供了基本上不含北方长额虾其它成分的重组多肽, 它包括由图 1 所示核苷酸序列 (SEQ ID No.1) 编码的构成热敏感型碱性磷酸酶的氨基酸序列, 或其功能等价的变体。

这里所使用的术语“多肽”包括全长的蛋白质, 和相对短的多肽序列。

上面描述多肽氨基酸序列所用的“功能等价”限定了与上述多肽序列相关或来源于上述多肽序列、但其中的氨基酸序列已经被修饰的、仍保留催化活性的多肽, 其中对多肽的修饰包括一个或多个氨基酸的取代, 添加或缺失, 或化学修饰如糖基化或去糖基化。这样的功能等价的变体可以由自然的生物变异所产生, 也可以用已知的技术制备, 例如功能等价的重组

多肽可以用已知的定点诱变，随机诱变，或酶裂解和/或氨基酸连接技术制备。

5 本发明这一方面的合成多肽可以通过宿主细胞中的表达来制备，该宿主细胞包含重组 DNA 分子，该分子包括与表达调控序列可操纵连接的上文广泛讨论的核苷酸序列，或者宿主细胞包含带有上述重组 DNA 分子的重组 DNA 克隆载体，例如，质粒，粘粒，噬菌体分子或者酵母人工染色体。另外，还可以把本发明的裸 DNA 分子直接注射到宿主细胞中来表达多肽。

10 如上述方法表达出的合成多肽可以是融合多肽，该多肽包含碱性磷酸酶有催化功能的部分，以及由与之融合的重组分子的 DNA 所编码的另一种多肽。例如，可以按照期望产生一种融合蛋白，它包含合成型碱性磷酸酶或另一种本发明多肽，它与诸如  $\beta$ -半乳糖苷酶，磷酸酶，谷胱甘肽-S-转移酶，尿素酶等蛋白融合。绝大多数融合蛋白通过重组基因的表达来产生，在该重组基因中，两种编码序列连接在一起，它们的读码框是同时协调的。还可以按照期望，把本发明的合成肽与具有配体结合功能的蛋白或肽(如抗体)相融合，从而产生例如酶联报告抗体，这也是本技术领域所熟知的。这类编码碱性磷酸酶的核酸分子和它们各自的氨基酸序列的所有这类融合或杂合衍生物都包括在本发明中。这类适当的重组 DNA 和多肽表达的技术在，例如，Sambrook et al., 1989 中描述。另外，本发明多肽还可以用化学方法，例如众所周知的 Merrifield 固相合成方法产生。

20 更具体地，在这一方面本发明提供了一种包含(或基本组成为)如下部分的重组多肽：

(a) SEQ ID No. 2 所示的氨基酸序列的全部或部分；或

(b) 与 SEQ ID No. 2 的全部或部分有至少 60% 序列同一性的氨基酸序列的全部或部分。

25 术语‘重组的’把上述分子与天然多肽区分开，并且强调这样的分子是用‘重组 DNA 技术’(本领域所熟知的名词)产生的。具有 SEQ ID No. 2 氨基酸序列的酶被认为是北方长额虾中天然酶的截短型，因此如果在重组生产中使用相应的核酸，那么生产出的重组酶无论如何都不会与自然产生的非重组分子相同。

30 具体地，上面提到的氨基酸序列可显示出与 SEQ ID No. 2 多肽至少 65%，70%，75%，80%，85%，90%，95% 或 98% 的同一性。或者，上面

提到的氨基酸序列可显示出与 SEQ ID No. 2 多肽至少 70%，75%，80%，85%，90%，95% 或 98% 的相似性。

具体地，上面提到的变异可能发生在 N 末端。本发明的多肽在 N 末端还可包括额外的 4-20 个氨基酸残基，例如 5-10 个。还应注意的是，SEQ ID No. 1 的不完整 cDNA 序列对应于 SEQ ID No. 2 的氨基酸序列，但仅始于第 7 个残基，天冬氨酸。在重组多肽中，氨基酸变异，特别是取代，可能会发生在这前 6 个氨基酸(见图 2 的序列)中。更进一步，图 2 的 N 末端赖氨酸可被精氨酸取代。

可产生对应于 SEQ ID No. 2 氨基酸序列的 cDNA 序列，它由 SEQ ID No. 1 和 5' 末端的衍生核苷酸序列共同组成。这一区域的大部分是由正向引物 SEQ ID No. 7 提供的，该引物是源自天然虾碱性磷酸酶(SAP)的测序片段。由于遗传密码的简并性，不可能根据氨基酸 KAYWNK 确定出比 AArGCnTAyTGGAAYAAr[SEQ ID No. 19]更加准确的序列，其中 r=A 或者 G，n=T,C,A 或者 G，y=C 或者 T。对应 SEQ ID No. 2 的 cDNA 全序列(包括 SEQ ID No. 19 和 SEQ ID No.1)在这里被归为 SEQ ID No. 20。

然而，如在实施例 4 所进行的测序工作已近确立了北方长额虾 (*Pandalus brealis*)碱性磷酸酶的 cDNA 全序列。因此，上面讨论的关于 SEQ ID No. 19 不确定性的问题就被解决了，编码 KAYWNK 的 cDNA 按如下所示：AAAGCATATTGGAACAAA [SEQ ID No. 21]。这个序列可以在 SEQ ID No. 15 中代替 SEQ ID No. 19，连同 SEQ ID No. 17 中编码氨基酸序列 NPITEED 的部分一起产生准确的编码重组酶的 cDNA 序列。该全序列被归于 SEQ ID No. 22。

氨基酸序列的同一性或相似性可以用 University of Wisconsin 的 Genetics Computer Group (GCG) Version 10 软件包的 Bestfit 程序来确定。该程序使用 Smith 和 Waterman 的局部同一性运算法，其默认值为：Gap 生成罚分值=8，Gap 延伸罚分值=2，平均匹配值=2.912，平均错配值=-2.003。

上述 SEQ ID No. 2 氨基酸序列的“部分”或其变体包含至少 20 个连续的氨基酸，优选包含至少 30，40，50，70，100，150，200，300，400 或者 450 个连续的氨基酸。具有超过 200 个，特别是超过 300 个氨基酸的部分被认为是‘显著部分’，这些部分通常带有本发明鉴定出的酶活性位点，优选带有辅助因子的结合位点。

重组多肽优选具有上文定义的功能性活性，如，酶活性。或者它本身不具有功能性活性，但其为整体提供了具有功能性特点的区域，例如，代表酶活性所需要的活性位点或辅助因子结合位点。

正如上面讨论的，SEQ ID No. 1 和 SEQ ID No. 2 不对应天然 SAP 的完整 cDNA 或者氨基酸序列。具有与天然 SAP 相同或基本相同的酶活性，并且保留此酶的热敏感性的截短型多肽构成了本发明的另一方面。插入本文所述截短序列中的完整 cDNA 分子或重组多肽，或者它们的功能等价变体，都是本发明进一步优选的方面。

本发明还有一方面提供了具有与 SAP 相同或基本相同的催化活性和热特性但缺少了它的 N-末端区的多肽。优选这些多肽缺少 N 末端氨基酸以及在天然酶中和与它毗连的 2-50 个氨基酸，特别是 5-30 个，如 5-10 个。‘基本相同的催化活性’是指多肽拥有天然 SAP 的至少 75%，优选至少 85%，更优选至少 90% 的酶活性。除了这些缺失的 N 末端部分，所述多肽优选与 SEQ ID No. 2 所示多肽有至少 60%，65%，70%，75%，80%，85%，90% 或 98% 的同一性。

根据本发明，当碱性磷酸酶经过 65°C 温育 15 分钟(优选 10 分钟)后，其催化活性检测不出时，可认为它是热敏感的。

测定本发明酶的催化活性所适用的方法按照 Olsen et al. [R.L. Olsen, K. Øverbø, and B. Myrnes (1991) 北方长额虾肝胰脏的碱性磷酸酶: 具有催化活性亚单位的二聚体酶(Alkaline phosphatase from the hepatopancreas of shrimp (*Pandalus borealis*): a dimeric enzyme with catalytically active subunits), *Comp. Biochem. Physiol.*, 998: 755-761]。碱性磷酸酶活性通过与 6 mM 对硝基苯磷酸酯在 37°C, pH 10.4 的 0.1 M 甘氨酸/NaOH 缓冲液中温育而测定，缓冲液中还含有 1 mM MgCl<sub>2</sub> 和 1 mM ZnCl<sub>2</sub>，总容积为 0.5ml。15 分钟后，用 0.5 ml 的 2M NaOH 终止反应，所释放出的对硝基苯的量通过测量 405 nm ( $\epsilon_{405}=18.5 \text{ mM}^{-1} \text{ cm}^{-1}$ ) 处的吸光度来判定。一个单位的酶活性力每分钟能产生 1  $\mu\text{mol}$  对硝基苯。SAP 不受锌的不良影响，在最小浓度以上对镁也没有显著量的需求。

本发明的范围内还包括 SEQ ID No. 1 或 SEQ ID No. 2 的变体，该变体编码或者包含保留并展示 SAP 的催化性质，但却是热稳定的蛋白，就是说这些蛋白在 70°C 加热 15 分钟后，不会完全丧失活性。

本发明还提供了制备本发明重组核酸分子的方法，包括把包含本发明核苷酸序列的核酸分子插入到另一核酸分子，例如，载体核酸，如载体 DNA 中。

所以本发明还提供了重组载体，该载体是携有 DNA 片段的载体 DNA，  
5 所述 DNA 片段包含本发明任意 cDNA。

在重组载体中，载体 DNA 优选是一种表达载体，并且能够在如大肠杆菌，酿酒酵母或巴斯德毕赤酵母这样的宿主细胞中繁殖。

本发明的表达载体可包括合适的调控序列，例如翻译调控元件(例如起始密码子和终止密码子，核糖体结合位点)和转录调控元件(如启动子-操纵子区域，末端终止序列)，这些元件与本发明核酸分子在匹配的阅读框中相连接。  
10

本发明载体可以包括本领域已知并在文献中记载的质粒和病毒(包括噬菌体和真核生物的病毒)，并且可以在不同的表达系统中表达，这也是本领域已知并有文献记载的。

现在已知许多技术可以用来把这类载体导入原核或真核细胞进行表达，  
15 或把这类载体导入生殖细胞系或体细胞中用于产生转基因动物。适用的转化或转染技术在文献中有很完善的描述。

包含本文所述本发明核酸分子并使其与启动子序列可操纵地连接的遗传构建体构成了本发明的另一方面。

本发明还包括了被转化或者被转染的原核或真核宿主细胞，或者包含了上面提到过的本发明核酸分子的转基因生物体。为了方便，这里不区分转化和转染。这样的宿主细胞可以是原核细胞，例如大肠杆菌，链霉菌和其他细菌，或真核细胞例如酵母(例如巴斯德毕赤酵母)，或杆状病毒-昆虫细胞系统，被转化的哺乳类细胞和转基因动植物。这里优选原核宿主细胞。  
20

所以本发明还提供了转化体，其是本发明重组载体所转化的细胞。  
25

另一方面，本发明提供了一种重组细胞或生物，它具有含 SEQ ID No. 1 或 20 所示核苷酸序列的核酸分子，或该分子的如本文定义的功能等价变体，并且表达或能够表达热敏感碱性磷酸酶。‘重组的’细胞或生物是说细胞或生物含有并非该生物自然产生的遗传物质(在本例中编码热敏感碱性磷酸酶)，所以北方长额虾被排除在外，因为它虽然表达热敏感碱性磷酸酶，但其基因是自然存在于该物种的。  
30

从另一角度看,那些表达热敏感碱性磷酸酶的细胞或生物可被视为“修饰过的”,即具有自然界中并不存在的基因型和表型的细胞或生物。因此,所述细胞或生物也可被称为“非天然”细胞或生物,因为它包含非天然遗传物质,并且能够产生该物种或细胞类型的天然成员不产生的蛋白(热敏感碱性磷酸酶)。

上述细胞和生物可以被称为转化体,它们或它们的祖先通过导入本文描述的编码热敏感碱性磷酸酶的遗传物质而被转化。

本发明还提供了另一种转化体,该转化体是被重组载体转化的大肠杆菌或酿酒酵母。细菌,真菌和其它培养细胞的适宜表达品系是本领域所熟知的。

本发明还提供了产生热敏感型碱性磷酸酶的方法,包括在培养基中培养本发明转化体,从培养的转化体中分离热敏感型碱性磷酸酶。更具体地,本发明还提供了生产热敏感型碱性磷酸酶的方法,包括用含有如下序列的表达载体转化宿主细胞: SEQ ID No. 1 或 20 所示的核苷酸序列,或与其至少有 55% 同一性的序列,和/或能够在中等严紧度条件下与 SEQ ID No. 1 或 20 的互补序列杂交的序列,或上述序列的互补序列;在培养基中维持宿主细胞,和从所述细胞分离所表达出的热敏感型碱性磷酸酶。培养细胞的方法为本领域所熟知并通常允许细胞繁殖并且同时或随后表达 ALP 基因。基因产物可在细胞内或细胞外积累。回收和提纯由这条途径产生的蛋白的方法是本领域所熟知的。

我们进行了进一步的工作并得到了更加准确的序列信息。修正过的 cDNA 和蛋白质序列相比原来变化很小。修正过的 cDNA 序列如图 4 所示,被命名为 SEQ ID No. 15。除了加入额外的 5' 序列(其等价的氨基酸序列出现在 SEQ ID No. 2 中),新的序列还缺少了 9 个核苷酸,如图 7 的序列对比所示。比较显示序列有 99.4% 相同。

在蛋白质水平上有稍大的改变,因为修正使得蛋白质的一小部分中的阅读框中发生了改变。修正过的氨基酸序列如图 5 所示,并被命名为 SEQ ID No. 16。图 8 中给出了与起始氨基酸序列的对比。该对比显示序列有 96% 相同。

所以,所有以前引用 SEQ ID No. 1 和 2 的地方分别可以用 SEQ ID No. 15 和 16 替代。这些序列包括了那些本发明已经描述过的较佳实施方案。

除此之外, SAP 全长 cDNA 序列已于近期被阐明, 所以其整个分子对应于 SEQ ID No. 17 然后接 SEQ ID No. 15, SAP 全长氨基酸序列(加上信号序列)对应 SEQ ID No. 18 然后接 SEQ ID No. 2 或 16。据信, 在 SEQ ID No. 18 的 47 个氨基酸中, 最后 7 个氨基酸(NPITEED)或者其中的几个, 举例来说, 5 PITEED 或者 TEED 可见于成熟 SAP 蛋白的 N 末端。剩下的氨基酸被认为是信号肽。具有这些序列的分子组成了本发明的另一方面。在本发明的一个较佳实施方案中, 重组的碱性磷酸酶在其 N 末端具有氨基酸序列 NPITEED。

下面将要一些非限制性实施例来进一步描述本发明, 其中:

10 图 1 显示了编码来自北方长额虾(*Pandalus borealis*)的 SAP 的 cDNA(去掉了 5' 末端的一小部分)。终止密码子和包括多腺苷酸尾部的 cDNA 下游序列标有下划线。

图 2 显示了图 1 所示的 SAP cDNA 编码的氨基酸序列(加入了一个额外的 N 末端区域)。与已测序的蛋白片段同源的部分用下划线标出并注明。酶的活性位点序列(根据已知的碱性磷酸酶)用粗体显示; 并且 15

图 3 显示了 SAP 序列与一些从 Swiss-Port 数据库找到的与其高度同源的组织非特异性 ALP 的序列比对, 其中有鼠(登录号 A. n. P09242), 人(A. n. P05186), 小鸡(A. n. Q92058), 和家蚕蛾(*Bombyx mori*)的 ALP(A. n. P29523)。

20 图 4 显示了修正了的编码北方长额 SAP 的 cDNA。多腺苷酸尾部表示为(A)<sub>23</sub>。由于扩增和测序使用了简并 PCR 引物, 5' 序列的 5 个核苷酸位置产生了误差, 所以这几个位置没有特别说明。

图 5 显示了图 4 所示的 SAP cDNA 编码的氨基酸序列。与已测序蛋白片段同源的序列用下划线标出并说明。建议的酶活性位点序列用粗体显示。由于如图 4 讨论的简并 PCR 引物所产生的误差, N 末端的赖氨酸残基(小写 25 字母)可以用精氨酸替换。

图 6 对应上面的图 3, 但是加入了图 5 中的修正的氨基酸序列。

30 图 7 显示图 4 中修正的 cDNA 序列(上列)与图 1 中的起始序列(加入了最初仅在图 2 的等价氨基酸序列中出现的 5' 端)的序列比对。比对是使用 Needle 程序进行的, 应用 Needle-Wunsch global alignment 算法来寻找两序列(包括空隙)的最佳列比 [Needle et al. J. Mol. Biol. (1983) 48, pp443-453]。矩

阵 Blosum 为 62，开放式空隙的罚分为 10.0，空隙延伸值罚分为 0.5。总体同一性百分比 99.4%。

图 8 显示图 5 中修正的氨基酸序列(上列)与图 2 中的起始氨基酸序列的比对。比对是使用上述 Needle 程序进行的，矩阵 Blosum 为 62，开放式空隙的罚分为 10.0，空隙延伸值罚分为 0.5。总体同一性百分比 96.3%。

图 9 给出了北方长额虾碱性磷酸酶的推定的信号序列和 N 末端区域的 cDNA [SEQ ID No. 17]和氨基酸[SEQ ID No. 18]序列。SAP 氨基酸序列后接 KAYWNK...，如图 2 和 4 所描述。

图 10 是一张照片，显示实施例 4 中表达培养物的总细胞蛋白提取物的 SDS-PAGE，其中：泳道 1 和 10：广域标准蛋白，泳道 2 和 9：SAP 1.5  $\mu$ g，泳道 3：阴性对照培养物，未进行诱导，泳道 4：阴性对照培养物，进行了诱导，泳道 5：SAP-培养物 1，未进行诱导，泳道 6：SAP-培养物 1，进行了诱导，泳道 7：SAP-培养物 2，未进行诱导，泳道 8：SAP-培养物 2，进行了诱导。广域标准蛋白中的蛋白质是：肌球蛋白(200 kD)， $\beta$ -半乳糖苷酶(116 kD)，磷酸化酶 b(97.4 kD)，血清白蛋白(66 kD)，卵清蛋白(45 kD)，碳酸酐酶(31 kD)，胰蛋白酶抑制物(21.5 kD)，溶菌酶(14.4 kD)，抑酶肽(6.5 kD)。

图 11 是一张照片，显示实施例 4 中表达培养的总细胞蛋白提取物的免疫印记，其中：1 道和 8 道：SAP 50 ng，2 道：阴性对照培养物，未进行诱导，3 道：阴性对照培养物，进行了诱导，4 道：SAP-培养物 1，未进行诱导，5 道：SAP-培养物 1，进行了诱导，6 道：SAP-培养物 2，未进行诱导，7 道：SAP-培养物 2，进行了诱导。

## 实施例

### 实施例 1- SAP 蛋白的测序和 SAP 基因的分离

#### 25 蛋白分离，片段化和序列分析

如[R. Olsen et al.，]所描述，把虾碱性磷酸酶(SAP)纯化至表现均一的程度。在 10% NuPAGE Bis-Tris 凝胶系统(Novex)中加入含 0.1% SDS 的 50 mM 2-(N-吗啉)乙磺酸电泳缓冲液进行电泳后，用考马斯染色和银染都探测到单个的约 55 kD 的蛋白质条带。

30 冻干 1 mg 蛋白质，把其递交商业序列分析服务公司 (Innovagen, Sweden)。在那里，蛋白质经胰蛋白酶处理成片段，这些片段用反相 HPLC

分离。产生了超过30个片段，选择其中4个作进一步分析。利用质谱分析来分析选定的片段的序列(H23:18, H23:30, 5ReRP6:26, 5ReRP6:17)，分别产生了12, 10, 8和5个氨基酸的序列。

#### 合成来自蛋白质片段序列的寡核苷酸

- 5 根据得到的氨基酸序列预测标准密码子，然后按惯例制成正向的简并寡脱氧核糖核苷酸引物和其反向互补序列(Eugogentec, Belgium)。

#### 虾 cDNA 的合成

- 10 解剖新鲜收获的北方长额虾，将单个的肝胰脏在液氮中储存备用。用 PolyA Tract System 1000 (Promega)从单个肝胰脏中分离出 mRNA。分离的 mRNA 被用于在第二条链合成之前的第一条链 cDNA 的合成，cDNA 的扩增按照 Smart<sup>TM</sup> PCR cDNA 合成试剂盒(Clontech)与 Advantage cDNA 合成试剂盒(Clontech)所给出的说明执行。

#### SAP-基因内部片段的 PCR 扩增和序列分析

- 15 使用小量等份试样的合成 cDNA 为模版，用上述成对的蛋白衍生型寡核苷酸(正向与反向)为引物，进行 PCR。具有标准混合物组成的扩增反应在 Eppendorf 梯度热循环仪中进行。PCR 的产物用琼脂糖凝胶电泳方法检测。寡核苷酸 17F 和 26R 的引物组合得到了一种 600bp 的扩增产物，用 Thermo Sequenase radiolabelled terminator cycle Sequencing Kit (Amersham)并且用同样的 PCR 引物测序。

#### 20 cDNA 末端快速扩增(RACE)

根据 PCR 产物中发现的 DNA 序列，合成出新的特异性正向和反向引物用于 5'和 3'RACE 反应，方法与用 Marathon<sup>TM</sup> cDNA amplification kit (Clontech)从 cDNA 模版进行扩增的描述一样。

- 25 用 SAP 基因特异性正向引物 MalpF 以及试剂盒所含的 API 引物，产生出 1.4 kb 的 3'-RACE 片段，其中所述 API 引物与所连的衔接子具有序列同一性。类似地，用反向基因特异性引物 MalpR 得到了 0.5 kb 的 5'RACE 产物。对 3'RACE 产物的 DNA 序列分析证实，它与 600 bp PCR 产物重叠。把 3'RACE 产物亚克隆到 pCR-Script 载体(Stratagene)中，用来制备 pMalpF 质粒，以便对包含在 600 bp PCR 产物中序列区下游的 cDNA 基因进一步测序。

- 30 用新的基因特异性引物对 600 bp PCR 产物和 pMalpF 插入序列从两个方向进行多次测序。

### 完整 cDNA 基因的高保真 PCR 扩增

用引物 SapF 和 SapR, Pfu 聚合酶(Stratagene), 以 cDNA 为模版进行最终的 PCR 反应, 产生 1.5 kb 的 SAP cDNA 基因扩增产物。对 1.5 kb PCR 产物进行分析, 以确认所得到的序列。结果发现, 该 cDNA 基因包含一开放阅读框和一个 3'序列, 该开放阅读框编码了一个 478 个氨基酸的多肽-见图 1 和 2, 该 3'序列包括公认的转录聚腺苷化信号。

### PCR 所使用的热循环特性

下面的循环特性被用于扩增反应:

-总 cDNA 扩增	1 x: 95°C, 2 分
	16 x: 95°C, 5 秒
	65°C, 5 秒
	68°C, 6 分
-内部 600 bp 片段 PCR	1 x: 94°C, 2 分
	36 x: 94°C, 10 秒
	51°C, 10 秒
	72°C, 1 分
	1 x: 72°C, 5 分
-5'RACE	1 x: 94°C, 30 秒
	5 x: 94°C, 5 秒
	72°C, 4 分
	1 x: 94°C, 30 秒
	5 x: 94°C, 5 秒
	70°C, 4 分
	1 x: 94°C, 30 秒
	23 x: 94°C, 5 秒
	68°C, 4 分
-3'RACE	1 x: 94°C, 30 秒
	36 x: 94°C, 5 秒
	68°C, 4 分
-高保真 PCR	1 x: 94°C, 3 分
	40 x: 94°C, 10 秒

55°C, 15 秒

72°C, 3 分

1 x: 72°C, 5 分

### 结果

SAP 蛋白亚单位最初经估计分子量为 65 kDa。在 USB 的产品表中描述了一种 59 kDa 的 SAP 蛋白，在发明人的 SDS-PAGE 体系中，SAP 蛋白与分子量为 55 kDa 的标准蛋白有相似的迁移模式。这些分歧可以用不同缓冲体系 5 和所用凝胶的质量的多样性来解释。

分离的 cDNA 编码与所分析的 SAP 蛋白片段具有序列同一性的多肽-见图 2。

用 SAP 蛋白序列作为公共数据库中进行同源性搜索的查询序列。分别用 BLAST 和 ClustalW 程序来进行同源性搜索和多序列对比排列。将 SAP 10 序列与它的最高分值同系物进行对比排列，这些同系物是鼠(登录号 P09242)，人(A. n. P05186)，鸡(A. n. Q92058)中找到的组织非特异性 ALP 和家蚕蛾的 ALP(A. n. P29523)。对比排列见图 3。通过公共数据库中的同源性搜索没有找到所得 cDNA 的等价序列。然而，衍生出的蛋白序列与一些已知的 ALP 有相对较高的同源性得分，在氨基酸序列上有 44% 同一性和 60 15 % 的相似性。蛋白同源性在 European Bioinformatics Institute 的互联网服务器上的 SWALL 数据库 (非重复蛋白质序列数据库；Swissport+TrEMBL+TrEMBLNew) 中通过运行 FastA3 程序来进行搜索，其使用默认参数(blosum50 矩阵，gap 开放罚分值=-12，gap 延伸罚分值=-2，KTUP=2)。[W. R. Pearson and D. J. Lipman (1988)，生物序列分析中的改进工 20 具(“Improved tools for biological sequence analysis”), PANS 85: 2444-2448; W. R. Pearson, (1990)，应用 FASTP 和 FASTA 进行的快速灵敏序列比较(“Rapid and sensitive sequence comparison with FASTP and FASTA”), Methods in Enzymology 183: 63-98.]

还有，cDNA 编码的蛋白包含基序 VTDSAASAT，它对应于该序列中 25 由 Prosite PS00 123 限定的 ALP 的活性位点模式。所以，来自虾肝胰脏的 mRNA-转录物的分离 cDNA 代表了北方长额虾(*P. borealis*)的 SAP 基因。

经测序的蛋白质片段 5ReRP6:26 与 cDNA 衍生的序列之间有一个氨基酸不符。它可以解释为 SAP 基因的等位变异或者蛋白质序列分析中的一个错误。

实施例 1 中使用的序列

5 蛋白质片段:

H23:18 DINFRYASAAP(V)K [SEQ ID NO.3]

H23:30 HLITDWLDDK [SEQ ID NO.4]

5ReRP6:26 VIMGGERR [SEQ ID NO.5]

5ReRP6:17 AYWVNK [SEQ ID NO.6]

10 寡核苷酸:

17F: GCITAYTGGAAYAAR [SEQ ID NO.7]

26R: CKYTCICCCICCCATDATIAC [SEQ ID NO.8]

其中 D=A+G+T

I=次黄嘌呤核苷

K=G+T

R=A+G

Y=C+T

MalpF: ATTCGTGGGAAGAATTCGACTTTGC [SEQ ID NO.9]

MalpR: GATCTGCCAGCCTCCTGGAACCA [SEQ ID NO.10]

15 SapF: AARGCNTAYTGGAAYAARGAT [SEQ ID NO.11]

SapR: GAAGGTAATCATCTACATCTCA [SEQ ID NO.12]

20 重复实施例 1, 使序列信息更精确。如图 4 和图 5 所示, 所述 cDNA 基因含有编码 475 个氨基酸的蛋白质的开放阅读框。图 4 的 cDNA 理论上的分子量是 53 kDa, 结果是与纯化型天然 SAP 在 SDS-PAGE 上的分子量估计值 54-55 kDa 相比得到的。根据多序列对比排列结果认为, SAP 在其 N 末端包含 5-10 个额外的氨基酸, 使其分子量接近 54 kDa。正如早先讨论的, 这个 N 末端区域被认为包含七肽 NPITEED 的部分或全部, 见实施例 4。

实施例 2- cDNA 衍生的蛋白质序列与天然 SAP 相比的氨基酸组成

将冻干的 SAP 蛋白质溶解在 6 M 盐酸中，110°C 水解 24 小时。将盐酸蒸发之后，使重新样品悬浮于 0.2 M 柠檬酸钠缓冲液，pH 2.2，然后进行 HPLC 分析，以鉴定水解产物中的氨基酸并测出其摩尔百分比。

注：此系统不能区分天冬氨酸和天冬酰胺或者谷氨酸和谷氨酰胺。所以，天然 SAP 蛋白中这些氨基酸的数据是混合的。此外，也可能会低估半胱氨酸残基的数量，因为蛋白质没有被氧化。

	%	
	cDNA 衍生的蛋白中	天然 SAP 中
丙氨酸	8.3	9.6
天冬氨酸+天冬酰胺	(13.7)	13.6
天冬氨酸	10.0	-
天冬酰胺	3.7	-
谷氨酸+谷氨酰胺	(9.0)	11.7
谷氨酸	6.7	-
谷氨酰胺	2.3	-
半胱氨酸	1.0	0.2
苯丙氨酸	4.1	4.1
甘氨酸	8.4	8.4
组氨酸	3.7	2.8
异亮氨酸	5.2	5.4
赖氨酸	4.6	4.2
亮氨酸	7.3	8.0
甲硫氨酸	1.6	1.9
脯氨酸	2.7	3.1
精氨酸	5.2	5.2
丝氨酸	4.6	4.6
苏氨酸	8.7	8.5
缬氨酸	5.4	4.9
色氨酸	1.6	-
酪氨酸	4.1	3.6

结论：对于大多数氨基酸而言，每种氨基酸在 cDNA 衍生的蛋白质序列中的摩尔比与在纯化的天然蛋白质中的比例几乎相同。明显较小的差异如上讨论。

所以，分离的 cDNA 编码的蛋白质与 SAP 蛋白相比，氨基酸组成非常  
5 近似或完全一致。

对经修正的 cDNA 序列重复了本实施例和分析，其结果如下：

**cDNA 衍生的蛋白质序列与天然 SAP 相比的氨基酸组成。**

	%	
	cDNA 衍生的蛋白中	天然 SAP 中
丙氨酸	8.8	9.6
天冬氨酸+天冬酰胺	-	13.6
天冬氨酸	10.1	-
天冬酰胺	3.4	-
谷氨酸+谷氨酰胺	-	11.7
谷氨酸	7.1	-
谷氨酰胺	2.3	-
半胱氨酸	0.8	0.2
苯丙氨酸	4.0	4.1
甘氨酸	8.0	8.4
组氨酸	3.3	2.8
异亮氨酸	5.4	5.4
赖氨酸	4.6	4.2
亮氨酸	7.6	8.0
甲硫氨酸	2.1	1.9
脯氨酸	2.7	3.1
精氨酸	4.8	5.2
丝氨酸	4.6	4.6
苏氨酸	9.0	8.5
缬氨酸	5.4	4.9

色氨酸	1.5	-
酪氨酸	4.0	3.6

### 实施例 3-SAP 在巴斯德毕赤酵母(*Pichia pastoris*)中的重组表达

编码 SAP 的 cDNA 被用于在适当的微生物中异源表达所述酶，以便通过发酵产生重组酶。有几种表达体系可供选择，其中优选真核表达体系，因为它可表达出相当水平的天然状态的酶。

5 一例这样的表达体系是甲基营养型巴斯德毕赤酵母(Invitrogen)，据报道，其在细胞内表达重组蛋白的水平最高达 12 g/L [Clare, J.J. Rayment, F.B., Ballantine, S.P., Srekrishna, K. and Romans, M.A. (1991): High-level expression of tetanus toxin fragment c in *Pichia pastoris* strains containing multiple tandem integrations of the gene. *Bio/Technology* 9,455-460]，分泌到培养基中的蛋白水平最高达 2.6 g/L [Paifer, E., Margolles, B., Cremata, J., Montesino, R., Herrera, L. and Delgado, JM. (1994): Efficient expression and secretion of recombinant alpha amylase in *Pichia pastoris* using two different signal sequences. *Yeast* 10.1416-1419]。

15 插入的基因受 AOX1(醇氧化酶)启动子的调节，所以表达可被甲醇诱导。

把 SAP cDNA 插入商品载体 pPIC9K 中。重组 SAP 因在载体中与酿酒酵母  $\alpha$  因子的信号肽序列融合而被表达为分泌产物。当整合到 his<sup>-</sup>菌株 (KM71 或者 GS115) 中以后，通过在不含组氨酸的培养基中挑选出 his<sup>+</sup>克隆，20 可选出重组菌株，因为 his 基因型由载体插入子予以弥补。另外，还可以根据对增高浓度的遗传霉素(G418, Invitrogen)的抗性来筛选重组体中的多基因插入子，其中的抗性源自质粒携带的卡那霉素抗性基因。

具体地，PCR 产生的 SAP cDNA 被当作新一轮 PCR 反应的模板，用来产生含有合适的限制性位点的 PCR 产物，所述位点能使所述产物整合到载体 pPIC9K 中，与  $\alpha$  因子信号序列位于相同阅读框内。作为 PCR 反应的正向引物，使用改进的“17F”引物[SEQ ID No. 13]，其中以赖氨酸密码子(AAG)代替 ATG 密码子(Met)，以便重建可被胰蛋白酶酶切的公认的赖氨酸残基。反向引物[SEQ. ID. No. 14]经构建可以在 SAP cDNA 序列的终止密码子的 3' 25 端加入一个 NotI 限制性位点。

产生的 PCR 产物被 NotI 酶切，然后连接到已被 SnaBI 和 NotI 酶切的 pPIC9K 载体中(从而产生质粒 pPIC9K-SAP)。因为该载体经 SnaBI 酶切后 5' 端的连接位点是平末端，而且两末端都在阅读框中，所以不必对已修饰过的 cDNA 的 5' 末端再进行修饰。

- 5 然后用载体转化感受态大肠杆菌菌株如 TOP10F'(Invitrogen)或等价菌株，使其繁殖，用对抗氨苄青霉素的抗性筛选重组体。分离的 pPIC9K-SAP 质粒然后可以用 ScaI 进行限制性酶切，使其线性化，以便产生一种可插入巴斯德毕赤酵母 AOX1 基因中的载体盒。将线性化的 pPIC9K-SAP 构建体转化入巴斯德毕赤酵母细胞(菌株 GS115 或 KM71)之后，根据制造商方案从
- 10 His<sup>+</sup>突变体中筛选出重组体。

- 然后再用制造商推荐的方法在甲醇存在条件下培养细胞，用来检测分离的菌落在培养基中的重组 SAP 的表达。如果需要，还可以从重组菌株中筛选含有多拷贝的所述重组基因的克隆，以获得进一步增加的酶的表达水平。这是通过在不断增加抗生素 G418(Invitrogen)水平的环境中培养 His<sup>+</sup>突
- 15 变体来达到的。生长在高 G418 水平中的克隆很可能含有多拷贝的所述基因。

所有培养，转化，筛选的方法都在巴斯德毕赤酵母体系(Invitrogen, The Netherlands)供应商提供的方案中有详细描述。

- 20 经修饰的 SAP 构建体产生了重组产物，其中添加了来自巴斯德毕赤酵母信号肽酶 KEX2 的剩余信号识别序列的 5 个氨基酸(EAEAY)。该 N 末端序列导致了最终产物被稍微截短，但与其他碱性磷酸酶有相当的共性，仍然拥有巴斯德毕赤酵母信号肽酶的认识序列。

#### 实施例 3 中使用的引物

正向:

- 25 AAGGCITAYTGGAAYAAR [SEQ. ID No. 13]

其中

I=次黄嘌呤核苷

R=A+G

Y=C+T

- 30 反向:

TACAGCGGCCGCCATCTCATTTTTTCG [SEQ. ID No. 14]

## 实施例 4-大肠杆菌 TOP10 中 SAP 的表达

### SAP 信号序列的测定

SAP 的 N 末端信号序列是以 cDNA 为模板用 5'-RACE 来测定的。使用 Oligotex Direct mRNA 小量试剂盒(Qiagen), 按照制造商的方案, 从 100 毫克  
5 克虾肝胰脏中分离 mRNA。虾肝胰脏 cDNA 是用 SMART PCR cDNA 合成试剂盒(Clontech)按照制造商的方案制备的。该 cDNA 是平端结尾, 通过在 50  $\mu$ l cDNA 中加入 40  $\mu$ g 蛋白酶 K, 45°C 保温 45 分钟和 90°C 保温 8 分钟, 然后加入 15 U T4-DNA 聚合酶(Promega), 16°C 保温 30 分钟并于 72°C 保温 10 分钟而纯化。最后, cDNA 在酚/氯仿中提取两次, 再在乙醇中沉淀。再  
10 按照如 Marathon cDNA 扩增试剂盒(Clontech)所述方法, 使该 cDNA 与 RACE-衔接子连接。5'-RACE 反应如下进行: 在终体积 50  $\mu$ l 中, 用 Advantage cDNA 聚合酶混合物(Clontech), AP1-引物(提供自 Marathon cDNA 扩增试剂盒), SAP-特异性引物(5'-GCG TGG TGC ATA TGG TCA ATC CGT CC-3')和 5  $\mu$ l 稀释 50 倍的连接了衔接子的 cDNA 模板进行反应。RACE  
15 PCR-反应在 Biometra TGradient 热循环仪中进行, 初始变性步骤为 94°C30 秒, 接着以 94°C5 秒和 72°C3 分钟循环 5 次, 以 94°C5 秒和 72°C3 分钟循环 5 次, 以 94°C5 秒和 68°C3 分钟循环 30 次。产生的 1,200 bp 的 SAP 片段用琼脂糖凝胶纯化, 用 PCR 再扩增并测序。

### 将 SAP 克隆入 pBAD/gIII A 表达载体

20 为了将 SAP 基因克隆入 pBAD/gIII A 载体, 用分别含有 SacI 和 HindIII 位点(下划线标明)的两种引物对 SAP 基因进行 PCR 扩增。根据得到的全长 SAP 序列(包括主序列和信号序列)设计引物, 用来扩增以 NPITEEDKAYWNK...氨基酸序列为 N 末端的 SAP 基因。另外, 改变了引物 1 中的三个碱基(C 变为 A, A 变为 C, 以及 A 变为 C), 使得三个 N 末端  
25 密码子得到密码优化(在引物序列中用小写字母表示)(引物 1: 5'-ATG GAG CTC AAC CcA ATc ACc GAA GAA GAC AA-3' 引物 2: 5'-ATG AAG CTT TCA TTT TTC GTC ACA GAA AGT G-3')。PCR 反应如下进行: 在终体积为 50  $\mu$ l 中, 含有分别为 10  $\mu$ M 的两种引物, 1.5 U Pfu 聚合酶(Promega)和提供的缓冲液, 0.2 mM dNTP, 并以 10 ng 虾肝胰脏 cDNA 作为模板。PCR-  
30 的初始变性步骤为 94°C30 秒, 接着以 94°C15 秒, 55°C 1 分钟, 72°C 3 分钟循环 30 次。PCR 产物用 Qiaquick PCR 纯化试剂盒(Qiagen)纯化。

PBAD/gIII A 载体和 SAP PCR 产物都分别用 5 U 的 SacI 和 HindIII 限制性内切酶处理。在 10  $\mu$ l 的反应体系中，加入 1 U T4-DNA 连接酶(USB)和分别约 1 mg 的质粒和 SAP PCR 产物，用来把 PCR 产物连接到载体中。反应混合物在 16 $^{\circ}$ C 温育 16 小时。然后向 40  $\mu$ l 电感受态大肠杆菌 TOP10 细胞加入连接混合物，1,800V 条件下电穿孔，以使用连接混合物转化大肠杆菌 TOP10 细胞。电穿孔之后，马上加入 1 毫升 SOC-培养基，在 37 $^{\circ}$ C 温育 1 小时。然后把细胞涂到含有 100 mg/ml 氨苄青霉素的 LB 平板上，以筛选转化的细胞。

### SAP 在大肠杆菌 TOP10 中的表达

10 选择了两个含有重组 pBAD/gIII A 载体(含 SAP 基因)的大肠杆菌 TOP10 菌落进行表达。用含有 pBAD/gIII A 空载体的大肠杆菌 TOP10 作为阴性对照。

在三个含有 50 ml LB 培养基的 Erlenmeyer 烧瓶中接种 2.5 毫升在含有 100  $\mu$ g/ml 氨苄青霉素的 LB 培养基中过夜的培养物。细胞在 30 $^{\circ}$ C 振荡培养，直至细胞浓度达到 OD600  $\approx$  0.5。然后把每份培养物分成两份(2  $\times$  25 毫升)，其中之一加入 L-阿拉伯糖至终浓度 0.05 % (w/v)，以便进行诱导。其后所有的细胞培养物均在 30 $^{\circ}$ C 振荡培养 3 小时，直至回收细胞。

### 表达培养物的分析

每份表达培养物取 200  $\mu$ l 离心，弃去上清。向细胞沉淀物中加入 100  $\mu$ l 的 1X SDS-PAGE 样品缓冲液，100 $^{\circ}$ C 温育 3 分钟，然后 18,000 g 离心 3 分钟。然后对细胞裂解产物直接进行 SDS-PAGE 凝胶电泳分析。一份额外的 SDS-PAGE 凝胶进行电转印，使蛋白转移到硝酸纤维素印迹膜(0.45  $\mu$ m, Bio-Rad)上，然后再用对抗从虾加工废水中纯化的 SAP 的抗体和山羊抗兔 IgG(H+L)(吸收了人类 IgG)辣根过氧化物酶偶联物(Bio-Rad)进行免疫染色。

25 SDS-PAGE，电转印和免疫染色都按照标准方案进行。

如图 10 和 11 所示，以大肠杆菌 TOP10 细胞作为宿主用 pBAD/gIII A 载体表达了重组 SAP。包含重组 pBAD/gIII A-SAP 载体的两种重组大肠杆菌菌株都产生重组 SAP，而阴性对照在 western 印迹中没有给出任何信号(图 11)。



1 KAYWNKDAQDALDKQLGKLRKQAKNVIFFLGDGMSLSTVTAARIYKGG  
    5ReP6:17  
 51 LTGKFEREKISWEEFFAALSKTYNTDKQVTDSAASATAYLTVGVTNQQV  
    活性位置  
 101 IGLDANTVRTNCSYQLDESLFTYSIAHWFQEAAGRSTGVVTSRVTHTATPA  
 151 GTYAHVADRDWENDSDVVHDREDPEICDDIAEQLVFREPGKNFKVIMGGG  
    5ReP6:26  
 201 RRGFFP EEALDIEDGIPGEREDGKHLITDWLDDKASQGATASYVWNRDDL  
    H23:30  
 251 LAVDIRNTDYLMGLFSYTHLDTVLTRDAEMDPTLPEMTKVAIEMLTKDEN  
 301 GFFLLVEGGRIDHMHANQIRQSLAETLDMEESCIHGSLYDRPRGNHFW  
 351 VTADHGHTLTITGYADRNITDILDFAGISDLDDRRRYTILDYGSQPGYHITE  
 401 DGKRYEPTTEEDLKDINFRYASAAPKHSVTHDGTDVGIWVNGPFAHLFTVT  
    H23:18  
 451 GVYEENYIPHALAYAACVGTGRTFCDEK stop - 478

图 2

```

*      20      *      40      *      60      *
ppbt_mouse : -----HISPFVLVAIGTCLTNSFVPEKERDPSYMRQCAQETLKNALKLQKLN---TNVARKV : 54
ppbt_human  : -----HISPFVLVAIGTCLTNSLVPEKEKPKYMRDCAQETLKYALELQKLN---TNVARKV : 54
alpchicken  : -----HKAFLLTLAQLCSASLVPEREKDFEYMRQCAQETLRDALRLQHLN---QNVAKVL : 53
sap         : -----KAYMVKDAGDALDKQLGKLR---EKCAKRV : 28
ppb_bonmo   : HVVSVVAAAAAAGLVRÆDRYHPERLAAGEASAA TRSAAESEA SFVREAEAIERREREGAGAKQAALGHAKV : 74
                W      AQ                                  AKN

      80      *      100     *      120     *      140
ppbt_mouse : IMFLGDGNGVSTVTAARILKQQLHHTGPEETRELDKPFVVAISKTYNTNAQVPPDSAGTATAYLGGVKNREGTV : 128
ppbt_human  : IMFLGDGNGVSTVTAARILKQQLHHTGPEETRELDKPFVVAISKTYNTNAQVPPDSAGTATAYLGGVKNREGTV : 128
alpchicken  : IMFLGDGNGVSTVTAARILKQQLQHRKGEESLLENDKPFVVAISKTYNTNAQVPPDSAGTATAYLGGVKNREGTV : 127
sap         : IMFLGDGNSLSTVTAARIYKGLTG-KFEREKISWEEFDFAISKTYNTDRQVTDASAATAYLITGVVTCGGVI : 101
ppb_bonmo   : VMFLGDGHSVPTLAARLLQRRRGQTGEASLHFECFPTLGLAKTYCVNIAQVPPDSCTATAYLGGVKNREGTV : 148
                FLGDGH T AAR G E F L KTY QV DS ATAYL GVK N G

*      160     *      180     *      200     *      220
ppbt_mouse : GVSAAATERTRNTTQGN--EVTSLRLAKDAKCKSVGIVITITVFNHATPSLAAYAHSAARDQVYSQNEHPPEALS-Q : 199
ppbt_human  : GVSAAATERSRNTTQGN--EVTSLRLAKDAKCKSVGIVITITVFNHATPSLAAYAHSAARDQVYSQNEHPPEALS-Q : 199
alpchicken  : GVSAGVTRDRNTTKGQ--EVTSLRLAKDAKCKAVGIVITITVFNHATPSLAAYAHSAARDQVYSQNEHPPEALS-Q : 198
sap         : CLDANTVBTNGSYQLDESIFTYSIAHDFQEAQRSTGVVISTEVTHATPAGTYAHVAARDQVENSQVVDVREDPE : 175
ppb_bonmo   : SVTAAVPRHDCÆASTDVTKRVCSIAEALADGRDVGIVITITRITHASAGTFKAVMNRNQMENRHDVKQEGHDVN : 222
                G A R C SI W G G VT TR HA P A A R W D

*      240     *      260     *      280     *
ppbt_mouse : GCRDIAYQLHNN--IKDIDVIMGGGRKYHYFKNRTDVEYELDEKARGTRLDGLDLISIKSFRP-RHKHSHYV : 270
ppbt_human  : GCRDIAYQLHNN--IRDIDVIMGGGRKYHYFKNKTQVEYESDEKARGTRLDGLDLVDTKSFRP-RHKHSHFI : 270
alpchicken  : GCRDIARQLVDN--IPDIEVILGGGRKYHYFKNTSVEYFQZERHRGTRLDGKDLVQAGHDTFP-AGKVAKYV : 269
sap         : ICDIARQLVDFREPCKNFVIMGGGRGFFPEEALDIEDGIP----GEREDGKHITDLDNQLASQATASYV : 245
ppb_bonmo   : KCPDIARQLIKMAPGNKFRVIFGGGRREFLPTQVDEEGRG----LRTDGRNLIEEQRDRESQKVSYKYL : 291
                C DIA QL VI GGR P D E R DG L W K V

      300     *      320     *      340     *      360     *
ppbt_mouse : NRTELLALDPSRVVYLLGLFEPGDHQYELNRMNLTDPSSLSENVVALRILTNMLKGFLLVEGGRIDEGHHEGK : 344
ppbt_human  : NRTELLALDPHNVDYLLGLFEPGDHQYELNRMNVTDPSSLSENVVAIQILRNKPKGFLLVEGGRIDEGHHEGK : 344
alpchicken  : HRRELLALNVSRVVYLLGLFEPGDHVEYELDRNETDPSSLSENVVAIRHLQKNPKGFLLVEGGRIDEGHHEGK : 343
sap         : NRDDLAVDIRNTDYLLGLFSEYHLDVTVTRADENDPFLPEHTKVAIENLTKDENGFLVEGGRIDEMHMANQ : 319
ppb_bonmo   : NRDELLKLGSSPPDYLLGLFEGSHLQYELGDESTETLALSLTDVAIRVLSRNERGFLLVEGGRIDEMHMANQ : 365
                R LL D L GLF L P L E VA L GFFL VEGGRIDH HH

      380     *      400     *      420     *      440
ppbt_mouse : AKQALHÆAVEHDOAIGKAGANTSQK-DTLTVVTADESEVFTFGGYTPRGNSIFGLAPHVSDTDKPPFTAILLGN : 417
ppbt_human  : AKQALHÆAVEHDRAIGQAGSLTSE-DTLTVVTADESEVFTFGGYTPRGNSIFGLAPHLSQTDKPPFTAILLGN : 417
alpchicken  : AKQALHÆAVELDRAVGLIGRLTSPR-DTLTVVTADESHVFTFGGYTPRGNPFGNSIFGLAPHQSQVDRKPPFTAILLGN : 416
sap         : IROSLÆTLDHÆESC IHGSLYDRPRGNHFWTADHGHTLITGVADPNTDILDFAGIS-QLDDRRYTILDGGS : 392
ppb_bonmo   : AHLALDÆTIEHRAVKVATDALKED-ESLVVTADETEVMSFNGYSPRGTDVVLGTVRSL-DSNRHPFHVLSHTIN : 437
                L E VTADH H GY R D Y

*      460     *      480     *      500     *      5
ppbt_mouse : GPGYKVVVD-GER-ENVSVDYAHNNYQAQSAVFLRHEHTGCEDEVAVFARQPHAHLLHG--VHØNYIPIHVNYA : 487
ppbt_human  : GPGYKVVG-GER-ENVSVDYAHNNYQAQSAVFLRHEHTGCEDEVAVFSKQPHAHLLHG--VHØNYVPIHVNYA : 487
alpchicken  : GPGYKIVG-GER-ENVSVDFAHANNYQAQAAVFLROETHGCEDEVAVFARQPHAHLLHG--VHØNYIPIHAYA : 486
sap         : GPGYHITEDGR- YEPTEEDLKDINFRYASAAKHSVTEGCTDVGIVVNGPFAHLFTVTGVYENYIPIHAYA : 465
ppb_bonmo   : GPGARIQNGVÆPDVTTDANFGALRWRTHTDVLDESEHTGCEDDVTVFAUGVHHWHFSG--LYØQTHVPIHAYA : 509
                GPG G R P TH G DV G E PH A A

      20      *      540     *
ppbt_mouse : SCICANLDHCÆAVAGSGSAPSPGALLPLAVLSLPTLF- : 524
ppbt_human  : ACICANLCHCÆAPASSAGSLAAGPLLLALALYPLSVLF- : 524
alpchicken  : ACICPNRLHCSSAARPAATA TLLPVLLLLLLLLC---- : 519
sap         : ACVCTGRTFCDEK----- : 478
ppb_bonmo   : ACICPGRHVGSAAATVPTAALLSLLLAFTL RHQCF L : 547
                C G C

```

图 3

AArGCnTAyTGGAAYAArGATGCCCAAGATGCACTGGATAAACAGCTGGGGATAAAACTG 60  
 CGAGAGAAGCAGGCCAAGAACGTGATCTTCTTCCCTAGGAGACGGGATGTCCCTCAGTACA 120  
 GTTACCGCAGCAAGAATCTACAAAGGTGGATTGACCGGGAAATTTGAACGGGAAAAGATT 180  
 TCGTGGGAAGAATTCGACTTTGCTGCTTTGAGTAAGACGTACAATACGGACAAGCAAGTG 240  
 ACAGATAGTGCTGCCAGTGCCACAGCCTACCTCACTGGGGTGAAGACCAACCAGGGAGTC 300  
 ATCGGCCTTGATGCCAATACCGTGAGAACAACTGCTCTTACCAACTTGATGAATCCCTC 360  
 TTCACTTACTCCATCGCCCATTGGTTCCAGGAGGCTGGCAGATCAACAGGTGTCTGTGACA 420  
 TCGACCAGGGTAACTCATGCTACTCCTGCAGGAACTTATGCTCACGTGGCTGACAGAGAT 480  
 TGGGAGAACGACAGTGATGTTGTCCACGACAGAGAAGATCCAGAAATATGTGATGACATA 540  
 GCTGAACAACCTGGTATTCAGAGAACCTGGAAAGAACTTCAAGGTCATCATGGGCGGCGGT 600  
 CGGCGTGGATTCTTTCCCTGAGGAAGCGCTTGATATCGAAGATGGCATTCCAGGGGAAAGG 660  
 GAAGATGGCAAACACCTCATAACCGATTGGTTGGACGATAAGGCATCTCAGGGAGCCACA 720  
 GCATCGTATGTCTGGAACCGAGATGATCTATTGGCTGTTGACATTCGAAACACAGACTAC 780  
 CTTATGGGACTGTTTTTCATACACACATCTAGACACAGTTTTAACCAGAGACGCCGAAATG 840  
 GATCCCACCTTGCCGGAAATGACAAAGGTGCCATCGAAATGCTGACGAAGGATGAAAAT 900  
 GGCTTTTTCTCTTAGTAGAAGGTGGACGGATTGACCATATGCACCACGCCAATCAGATC 960  
 CGTCAGTCGTTAGCGGAGACGCTGGACATGGAGGAAGCTGTATCCATGGCTCTCTCTATG 1020  
 ACAGACCCAGAGGAAACAATCATTCTGGTCCACAGCTGATCACGGTCACACTCTCACCATC 1080  
 ACTGGATACGCAGACAGGAACACCGATATTTTGGATTTGCGCCGCATCAGCGACTTGGAT 1140  
 GATAGGAGGTACACCATCCTGGACTATGGCAGTGGACCTGGATATCATATCACGGAAGAT 1200  
 GGCAAACGTTACGAACCTACAGAGGAAGATTTGAAAGACATCAATTTCCGCTACGCATCG 1260  
 GCAGCGCCCAAGCATTAGTTACCCATGATGGCACAGACGTTGGTATTTGGGTCAACGGA 1320  
 CCTTTCGCTCATCTCTTCACGGGTGTCTATGAAGAGAATTACATTCCCACGCCTTGGCC 1380  
 TACGCAGCCTGCGTTGGTACAGGACGCAC'TTCTGTGACGAAAAATGAGATGTAGATGAT 1440  
 TACCTTCTTAAGTACATTAAAGATTGC - (A)<sub>23</sub> 1490

图 4



```

*          *          *          *          *
20          40          60          80
alp_mouse  : HISFFLVLAIGT-----CLTNSVFPKRRDPSYWRQQAQETLKNALKLQK---LNTNVAKNVIMFLCD : 60
alp_human  : HISFFLVLAIGT-----CLTNSLVPEKREPKYWRDQAQETLKYALELQK---LNTNVAKNVIMFLCD : 60
alp_chicken: -MKAFLLTLLAQ-----LCSASLVPRRRDPPYWRQQAQETLRDALRLQH---LNQNVAKNLLFLCD : 59
SAP_shrimp : -----RAYWRKQADALDRQLCK---LRKQAKNVIFLCD : 34
alp_bommo  : HVVSVVAAAAAAGLVRAEDRYHPERLAACRASAATRSAAESEASTVVRRAQQAIEKRRRERRCAGAKQAACHAKNVIMFLCD : 80
          W AQ          AKN FLCD

*          *          *          *          *
100         120         140         160
alp_mouse  : CHGVSIVTAAIILKQQLHHTGKRETRLENDKPPVVAISKYNTNAQVPPSACTATAYDCVVRANRGTVGVSAATERTRCN : 140
alp_human  : CHGVSIVTAAIILKQQLHNPCKRETRLENDKPPVVAISKYNTKAQVPPSACTATAYDCVVRANRGTVGVSAATERTRCN : 140
alp_chicken: CHGVSIVTAAIILKQQLQHRKCKESLLENDKPPVVAISKYNTNAQVPPSACTATAYDCVVRANRGTVGVSAAGVTRDRCN : 139
SAP_shrimp : CHSLSTVIAAIIYKGLTG-KFERRKLSWRREDFAAISKYNTDKQVTSAASTAYVITETKINQGVICLDANTVTRNGS : 113
alp_bommo  : CHSVPFLAAIITLLQRRRGQTCKKASLHFRQPTLQAKYCVNAQVPPSCTATAYDCVVRANRGTVGVTAAVRHHDCB : 160
          CH T AAR C          E          F          L KTY          QV DS          ATAYL GVK N G G A R C

*          *          *          *          *
180         200         220         240
alp_mouse  : ITTQGNVET--EYLRNARDACKSVGIIHTTRVNHATPSAAYAHSADDDYSDNEMPPALS-QGCKDTAYQLMHNIR--DI : 215
alp_human  : ITTQGNVET--EYLRNARDACKSVGIIHTTRVNHATPSAAYAHSADDDYSDNEMPPALS-QGCKDTAYQLMHNIR--DI : 215
alp_chicken: ITTQGNVET--EYLRNARDACKSVGIIHTTRVNHATPSAAYAHSADDDYSDNEMPPALS-QGCKDTAYQLMHNIR--DI : 214
SAP_shrimp : YQLDESIFTYSTAHNFORARSTGVVTSVTHATPACTYAHVADDDYSDNEMPPALS-QGCKDTAYQLMHNIR--DI : 193
alp_bommo  : ASTDVTRKRVQSLARWALADRVDVITHTTHASACTYAKVANANENDDVVKQEGHDVNRKCPYAHQIKMAPGNKF : 240
          SI W G          C VT TR HA P          A A R W D          C DIA QL

*          *          *          *          *
260         280         300         320
alp_mouse  : DVYHGGCKKHYFKNRTDVEYXLDERKRGTELDGLDLSISIKRSFKP-RRHGSHYVWNTREMLDPSRVYLLGLTEPCD : 294
alp_human  : DVYHGGCKKHYFKNRTDVEYXSDERKRGTELDGLDLDVDTKRSFKP-RRHGSHYVWNTREMLDPSRVYLLGLTEPCD : 294
alp_chicken: DVYHGGCKKHYFKNRTDVEYXSDERKRGTELDGLDLDVQVNDTKP-ACKVAKYVWNTREMLDPSRVYLLGLTEPCD : 293
SAP_shrimp : KYVHGGCKKRCFFPEHALDTEDCIP---GEREDGKHITDLDDEASQCATAASYVWNTREMLDPSRVYLLGLTEPCD : 269
alp_bommo  : KYVHGGCKKRCFFPEHALDTEDCIP---GEREDGKHITDLDDEASQCATAASYVWNTREMLDPSRVYLLGLTEPCD : 315
          VI CCCR P D E          G R DG L W K          W R LL          D L G LF

*          *          *          *          *
340         360         380         400
alp_mouse  : HQYELNRNMLDPSLSSEVVAVALRIITKNLCKEFLVVEGGRIDGCHHECKAKQAALHAIVEMDQAICKAGANTSQDITLTV : 374
alp_human  : HQYELNRNMLDPSLSSEVVAVALRIITKNLCKEFLVVEGGRIDGCHHECKAKQAALHAIVEMDQAICKAGANTSQDITLTV : 374
alp_chicken: HQYELNRNMLDPSLSSEVVAVALRIITKNLCKEFLVVEGGRIDGCHHECKAKQAALHAIVEMDQAICKAGANTSQDITLTV : 373
SAP_shrimp : LDTVLTDRARDPTEPHTKVAIEHTKDNCFYLLVEGGRIDGCHHECKAKQAALHAIVEMDQAICKAGANTSQDITLTV : 349
alp_bommo  : LQYHLEGEKSTREPTLALTDVAIRVUSRWKRCFLVVEGGRIDGCHHECKAKQAALHAIVEMDQAICKAGANTSQDITLTV : 395
          L          P L E VA L          GFLL VEGGRIDH HH          L E          A A

*          *          *          *          *
420         440         460         480
alp_mouse  : VTADHSHVFTFGCYTPRCNSIFGLAPHSVSDTRKPPFTAILYCNCGCYKRVVD-GERENVV-HVDYAHNRYQAQSAVPLRHE : 452
alp_human  : VTADHSHVFTFGCYTPRCNSIFGLAPHSVSDTRKPPFTAILYCNCGCYKRVVC-GERENVV-HVDYAHNRYQAQSAVPLRHE : 452
alp_chicken: VTADHSHVFTFGCYTPRCNPIFGLAPHSVSDTRKPPFTAILYCNCGCYKRVV-GERENVV-HVDYAHNRYQAQSAVPLRHE : 451
SAP_shrimp : VTADHSHVFTFGCYTPRCNPIFGLAPHSVSDTRKPPFTAILYCNCGCYKRVV-GERENVV-HVDYAHNRYQAQSAVPLRHE : 427
alp_bommo  : VTADHSHVFTFGCYTPRCNPIFGLAPHSVSDTRKPPFTAILYCNCGCYKRVV-GERENVV-HVDYAHNRYQAQSAVPLRHE : 474
          VTADH H          GY R          D          Y GPC          G R          P

*          *          *          *          *
500         520         540
alp_mouse  : THGGEDVAVFAKCPMAHLLHGVDKQNYIPHVHAYACIQANLDHCWAGSGSAPSPGALLPLAVLSLPTLFF- : 524
alp_human  : THGGEDVAVFSKCPMAHLLHGVDKQNYIPHVHAYACIQANLDHCWAGSGSAPSPGALLPLAVLSLPTLFF- : 524
alp_chicken: THGGEDVAVFAKCPMAHLLHGVDKQNYIPHVHAYACIQANLDHCWAGSGSAPSPGALLPLAVLSLPTLFF- : 519
SAP_shrimp : THGGEDVAVFAKCPMAHLLHGVDKQNYIPHVHAYACIQANLDHCWAGSGSAPSPGALLPLAVLSLPTLFF- : 475
alp_bommo  : THGGEDVAVFAKCPMAHLLHGVDKQNYIPHVHAYACIQANLDHCWAGSGSAPSPGALLPLAVLSLPTLFF- : 547
          TH G DV          G          G E          PH A A C G          C

```

图 6



```

676   CTCATAACCGATTGGTTGGACGATAAGGCATCTCAGGGAGCCACA 720
      |||
676   CTCATAACCGATTGGTTGGACGATAAGGCATCTCAGGGAGCCACA 720
      |||
721   GCATCGTATGTCTGGAACCGAGATGATCTATTGGCTGTTGACATT 765
      |||
721   GCATCGTATGTCTGGAACCGAGATGATCTATTGGCTGTTGACATT 765
      |||
766   CGAAACACAGACTACCTTATGGGACTGTTTTTCATACACACATCTA 810
      |||
766   CGAAACACAGACTACCTTATGGGACTGTTTTTCATACACACATCTA 810
      |||
811   GACACAGTTTTTAACCAGAGACGCCGAAATGGATCCACCTTGCCG 855
      |||
811   GACACAGTTTTTAACCAGAGACGCCGAAATGGATCCACCTTGCCG 855
      |||
856   GAAATGACAAAGGTCGCCATCGAAATGCTGACGAAGGATGAAAT 900
      |||
856   GAAATGACAAAGGTCGCCATCGAAATGCTGACGAAGGATGAAAT 900
      |||
901   GGCTTTTTCCCTCTAGTAGAAGGTGGACGGATTGACCATATGCAC 945
      |||
901   GGCTTTTTCCCTCTAGTAGAAGGTGGACGGATTGACCATATGCAC 945
      |||
946   CACGCCAATCAGATCCGTCAGTCGTTAGCGGAGACGCTGGACATG 990
      |||
946   CACGCCAATCAGATCCGTCAGTCGTTAGCGGAGACGCTGGACATG 990
      |||
991   GAGG.AAGCTGTATCCATGGCTCTCTCTATGACAGACCCAGAGGA 1034
      |||
991   GAGGAAAGCTGTATCCATGGCTCTCTCTATGACAGACCCAGAGGA 1035
      |||
1035  AACAAATCA.TTCT.GGTCACAGCTGATCACGGTCACACTCTCACC 1077
      |||
1036  AACAAATCATTCTGGGTCACAGCTGATCACGGTCACACTCTCACC 1080
      |||
1078  ATCACTGGATACGCAGACAGGAACACCGATATTTGGATTTGCC 1122
      |||
1081  ATCACTGGATACGCAGACAGGAACACCGATATTTGGATTTGCC 1125
      |||
1123  GGCATCAGCGACTTGGATGATAGGAGGTACACCATCCTGGACTAT 1167
      |||
1126  GGCATCAGCGACTTGGATGATAGGAGGTACACCATCCTGGACTAT 1170
      |||
1168  GGCAGTGGACCTGGATATCATATCACGGAAGATGGCAAACGTTAC 1212
      |||
1171  GGCAGTGGACCTGGATATCATATCACGGAAGATGGCAAACGTTAC 1215
      |||
1213  GAACCTACAGAGGAAGATTTGAAAGACATCAATTTCCGCTACGCA 1257
      |||
1216  GAACCTACAGAGGAAGATTTGAAAGACATCAATTTCCGCTACGCA 1260
      |||
1258  TCGGCAGCGCCCAAGCATTAGTTACCCATGATGGCACAGACGTT 1302
      |||
1261  TCGGCAGCGCCCAAGCATTAGTTACCCATGATGGCACAGACGTT 1305
      |||
1303  GGTATTTGGGTCAACGGACCTTTCGCTCATCTCT.....TCACG 1341
      |||
1306  GGTATTTGGGTCAACGGACCTTTCGCTCATCTCTTCACGGTCACG 1350
      |||
1342  GGTGTCTATGAAGAGAATTACATFCCCCACGCCTTGGCCTACGCA 1386
      |||
1351  GGTGTCTATGAAGAGAATTACATFCCCCACGCCTTGGCCTACGCA 1395
      |||
1387  GCCTGCGTTGGTACAGGACGCCTTCTGTGACGAAAAATGAGAT 1431
      |||
1396  GCCTGCGTTGGTACAGGACGCCTTCTGTGACGAAAAATGAGAT 1440
      |||

```

图 7 (续 1)

```

1432   GTAGATGATTACCTTCTTAAGTACATTAAGATTGCAAAAAAAAA 1476
      |||||||||||||||||||||||||||||||||||||||
1441   GTAGATGATTACCTTCTTAAGTACATTAAGATTGCAAAAAAAAA 1485
1477   AAAAAAAAAAAAAA 1490
      |||||||||||
1486   AAAAAAAAAAAAAA 1499

```

图 7 (续 2)

```

1     KAYWNKDAQDALDKQLGIKLREKQAKNVIFFLGDGMSLSTVTAAR 45
      |||||||||||||||||||||||||||||||||||
1     KAYWNKDAQDALDKQLGIKLREKQAKNVIFFLGDGMSLSTVTAAR 45
46    IYKGGLTGKFEREKISWEEFDFAALSPTYNTDKQVTDASAATAY 90
      |||||||||||||||||||||||||||||||
46    IYKGGLTGKFEREKISWEEFDFAALSPTYNTDKQVTDASAATAY 90
91    LTGVKTNQGVIGLDANTVRTNCSYQLDESLEFTYSIAHWFQEAGRS 135
      |||||||||||||||||||||||||||||||
91    LTGVKTNQGVIGLDANTVRTNCSYQLDESLEFTYSIAHWFQEAGRS 135
136   TGVVTSTRVTHATPAGTYAHVADRDWENDSDVVHDREDPEICDDI 180
      |||||||||||||||||||||||||||
136   TGVVTSTRVTHATPAGTYAHVADRDWENDSDVVHDREDPEICDDI 180
181   AEQLVFREPGKNFVKVIMGGGRRGFFPEEALDIEDGIPGEREDGKH 225
      |||||||||||||||||||||||||||
181   AEQLVFREPGKNFVKVIMGGGRRGFFPEEALDIEDGIPGEREDGKH 225
226   LITDWLDDKASQGATASYVWNRD DLLAVDIRNTDYLMGLFSYTHL 270
      |||||||||||||||||||||||
226   LITDWLDDKASQGATASYVWNRD DLLAVDIRNTDYLMGLFSYTHL 270
271   DTVLTRDAEMDPTLPENTKVAIEMLT KDENGFFLLVEGGRIDMHM 315
      |||||||||||||||||||||||
271   DTVLTRDAEMDPTLPENTKVAIEMLT KDENGFFLLVEGGRIDMHM 315
316   HANQIRQSLAETLDMEEA VSMALSMTDPE.ETIILVTADHGHTLT 359
      |||||||||||: | |||||
316   HANQIRQSLAETLDMEESC IHGSLYDRPRGNNHFWVTADHGHTLT 360
360   ITGYADRNTDILDFAGISDLDDRRYTILDYGS GPGYHITEDGKRY 404
      |||||||||||||||||||||||
361   ITGYADRNTDILDFAGISDLDDRRYTILDYGS GPGYHITEDGKRY 405
405   EPT EEDLKDINFRYASAAPKHSVTHDGT DVG I WNGPFAHLF. .T 447
      ||||||||||||||||||| |
406   EPT EEDLKDINFRYASAAPKHSVTHDGT DVG I WNGPFAHLFTVT 450
448   GVYEENYI PHALAYAACVGTGRTFCDEK 475
      |||||||||||
451   GVYEENYI PHALAYAACVGTGRTFCDEK 478

```

图 8

1            11            21            31            41            51            61  
1 ATGAGGGCTTTCTACTGTCTATTGGGTTACTTTTAACCCCTAAGTCTTCAAGGAGAAGCAACACGTCAAC  
1(+1) M R A F Y C L F G L L L T L S L Q G E A T R Q  
  
71 CAAGTTCCTATGATGGAGATGAGTACCACGAGAGATATCCGGGTCAATCCAACCCATAACAGAAGA  
24(+1) P S S Y D G D E Y H E R Y P G Q S N P I T E E D  
  
141 C  
48(+1)

图 9

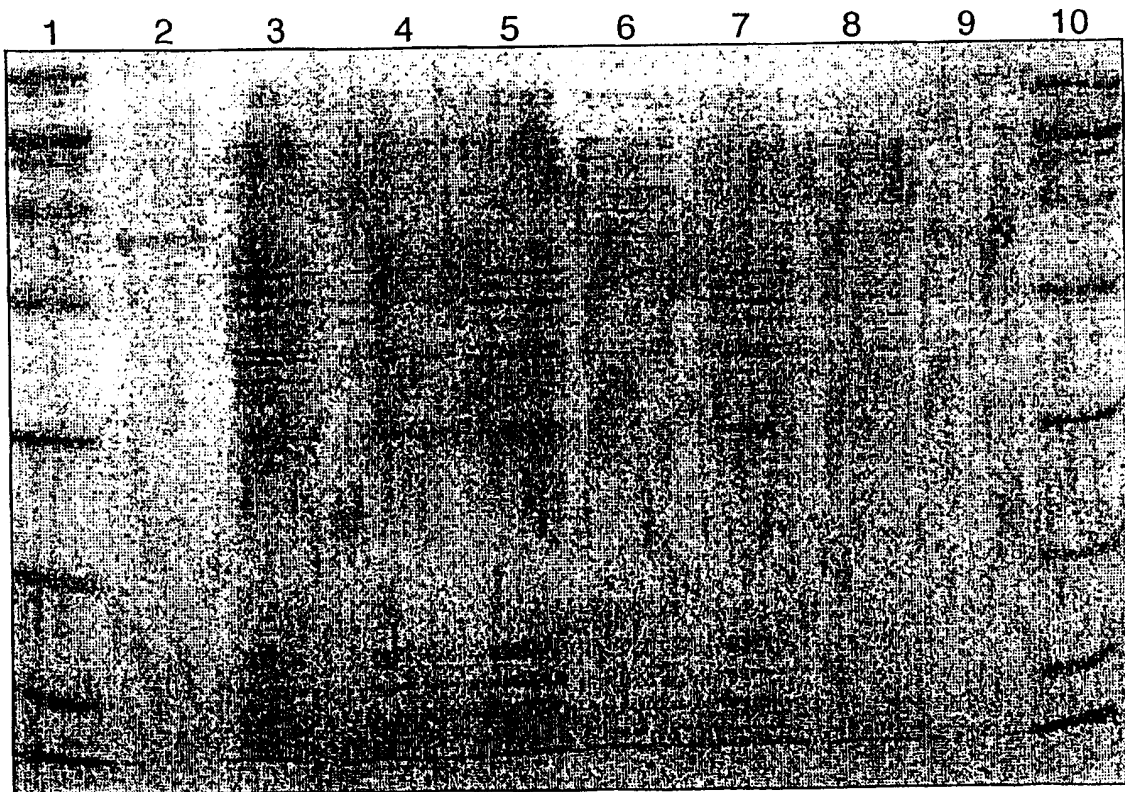


图 10

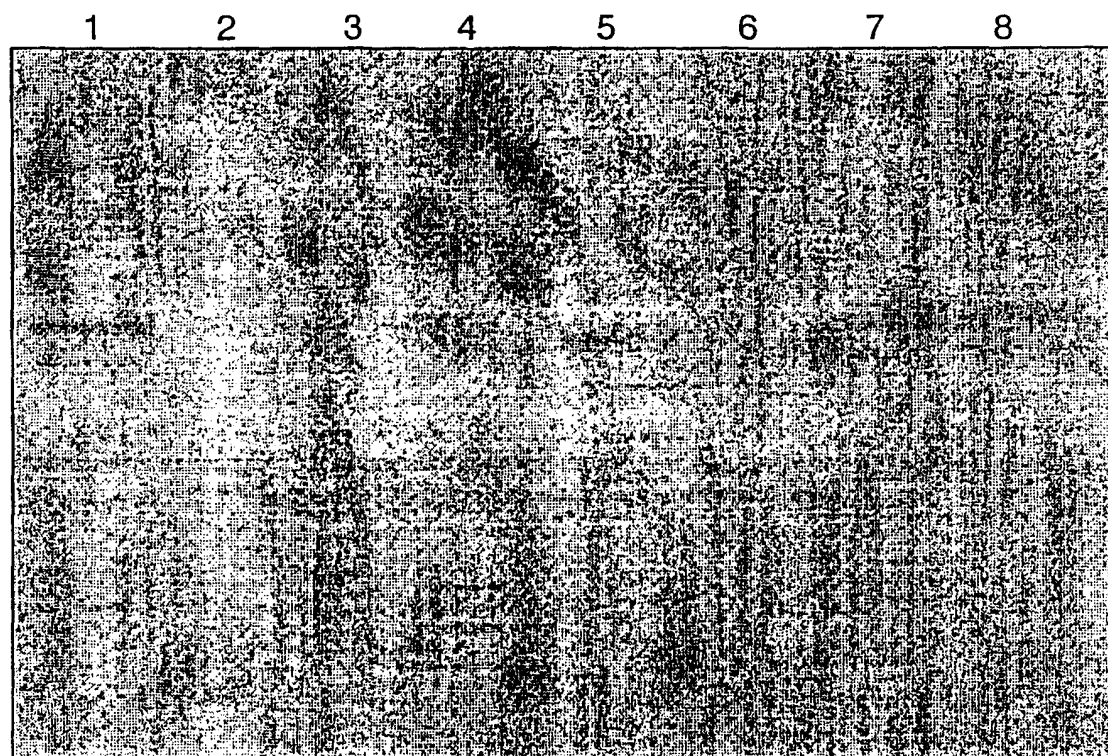


图 11