

[19]中华人民共和国专利局

[51]Int.Cl⁶



[12] 发明专利申请公开说明书

C12N 15/12
C07K 14/47 A61K 38/17
A61K 6/00 G01N 33/68
C12Q 1/68

[21] 申请号 96196921.0

[43]公开日 1998年10月14日

[11] 公开号 CN 1196086A

[22]申请日 96.6.26

[30]优先权

[32]95.7.13 [33]SE[31]9502595-3

[32]95.10.19[33]US[31]60/005,634

[32]95.11.16[33]DK[31]1284/95

[86]国际申请 PCT/IB96/00643 96.6.26

[87]国际公布 WO97/02730 英 97.1.30

[85]进入国家阶段日期 98.3.12

[71]申请人 比奥拉公司

地址 瑞典马尔默

[72]发明人 R·克尔赖 I·斯拉拜

L·哈马尔斯特罗姆 T·乌尔茨

C·D·芳

[74]专利代理机构 中国专利代理(香港)有限公司

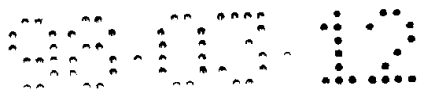
代理人 罗宏 齐曾度

权利要求书 3 页 说明书 36 页 附图页数 7 页

[54]发明名称 釉质基质相关多肽

[57]摘要

本发明涉及编码能调节釉质和细胞表面接触之多肽的新核酸片段。本发明还涉及含本发明核酸片段的表达载体以生产蛋白质，涉及含所述表达载体的生物，生产该多肽的方法，含该多肽的组合物，识别该多肽的抗体或抗体片段，以及治疗各种硬组织疾病或失调症的方法。



权 利 要 求 书

1. 一种至少部分被纯化的核酸片段，它编码能介导釉质与细胞表面的接触的多肽。

2. 一种权利要求 1 的核酸片段，它包括核苷酸序列 SEQ ID NO : 1，其至少 18 个核苷酸的亚序列，或所述核酸序列或亚序列的变异体，该变异体与 SEQ ID NO : 1 或其至少 18 个核苷酸的亚序列至少有 80% 同源性。

3. 一种权利要求 1 的核酸片段，它编码一种多肽，该多肽的氨基酸序列与 SEQ ID NO : 2 所示氨基酸序列至少有 80% 同源性。

4. 一种权利要求 1 的核苷酸序列，它包括核苷酸序列 SEQ ID NO : 3，其至少 18 个核苷酸的亚序列，或所述序列或亚序列的变异体，该变异体与 SEQ ID NO : 3 或其至少 18 个核苷酸的亚序列有至少 80% 同源性。

5. 一种权利要求 1 的核酸片段，它编码一种多肽，该多肽的氨基酸序列与 SEQ ID NO : 4 中所示的氨基酸序列至少有 80% 的同源性。

6. 一种至少部分被纯化的核酸片段，它基本包含 SEQ ID NO : 1 中所示序列。

7. 一种至少部分被纯化的核酸片段，它基本包含 SEQ ID NO : 3 中所示序列。

8. 一种权利要求 1 的核酸片段，它与含核苷酸序列 SEQ ID NO : 1 或其特异部分的核酸片段在严谨杂交条件下杂交。

9. 一种权利要求 1 的核酸片段，它与含核苷酸序列 SEQ ID NO : 3 或其特异部分的核酸片段在严谨杂交条件下杂交。

10. 一种至少部分被纯化的权利要求 1 的核酸片段，它编码含 SEQ ID NO : 2 的氨基酸序列 1 到 407 的多肽。

11. 一种至少部分被纯化的权利要求 1 的核酸片段，它编码含 SEQ ID NO : 4 的氨基酸序列 1 到 302 的多肽。

12. 一种权利要求 1 的核酸片段，它编码含氨基酸序列 SEQ ID NO : 2 或 SEQ ID NO : 4 之一或两者的亚序列的多肽。

13. 一种含权利要求 1-12 之任何的核酸片段的表达系统。

14. 一种可复制表达载体，它载有如权利要求 1-12 之任一中定义的核酸片段并能够介导该核酸片段的表达。



15. 一种携带权利要求 14 的表达系统的生物，如微生物，如细菌，例如 *Escherichia coli*, 酵母，原生动物，或从多细胞生物如真菌，昆虫细胞，植物细胞，哺乳动物得到的细胞或一种细胞系。

16. 一种釉质基质相关多肽，它含至少一个能介导该多肽锚着于细胞粘着分子的序列元件，该序列元件选自四肽 DGEA（天冬氨酸-甘氨酸-谷氨酸-丙氨酸），VTKG（缬氨酸-苏氨酸-赖氨酸-甘氨酸），EKGE（谷氨酸-赖氨酸-甘氨酸-谷氨酸）和 DKGE（天冬氨酸-赖氨酸-甘氨酸-谷氨酸）。

17. 一种权利要求 16 的多肽，它含氨基酸序列 SEQ ID NO：2 或其类似物或其变异体。

18. 一种权利要求 16 的多肽，它含氨基酸序列 SEQ ID NO：4 或其类似物或其变异体。

19. 一种权利要求 16 的具有如此的氨基酸序列的多肽，所述序列中的 20 个氨基酸的连续链段与选自 SEQ ID NO：2 和 SEQ ID NO：4 中所示的氨基酸序列之同样长度的氨基酸链段至少有 80% 程度的同源性。

20. 一种含基本上在 SEQ ID NO：2 中的氨基酸序列 1-407 的多肽。

21. 一种含基本上在 SEQ ID NO：4 中的氨基酸序列 1-324 的多肽。

22. 一种权利要求 16 的多肽，该多肽含氨基酸序列 SEQ ID NO:2 和 /或序列 SEQ ID NO:4 的亚序列。

23. 一种权利要求 16-22 的任何基本纯的多肽。

24. 含权利要求 23 的多肽和非强制性的一种生理可接受性赋形剂的一种组合物。

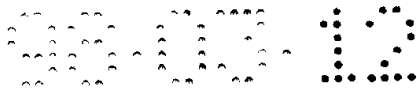
25. 一种产生如权利要求 16 中定义的多肽的方法，包括如下步骤：

(a) 插入权利要求 1-12 之任何中定义的核酸片段到一个表达载体中，

(b) 用步骤(a)中产生的载体转化适当宿主生物，

(c) 在适当的条件下培养步骤(b)产生的宿主生物以表达该多肽，

(d) 收获该多肽，和



(e) 非强制性对该多肽进行翻译后修饰。

26. 一种治疗和/或防止牙周疾病的方法，该方法包括给需要此种治疗的病人施用治疗或预防有效量的权利要求 16 的多肽。

5 27. 一种修复牙中损害的方法，该方法包括非强制性结合适当的填充物材料给需要此种治疗的病人施用有效量的权利要求 16 的多肽。

28. 一种接合两只骨单体的方法，该方法包括给需要此种治疗的病人施用有效量的权利要求 16 的多肽。

10 29. 一种促进或引起硬组织矿化的方法，该硬组织选自骨，釉质，牙质和牙骨质，该方法包括给需要此种治疗的病人施用有效量的权利要求 16 的多肽。

30. 一种有效结合植入物到骨中的方法，该方法包括给需要此种治疗的病人施用有效量的权利要求 16 的多肽。

31. 一种促进植入器具或经皮器具生物相容性的方法，该方法包括用有效量的权利要求 16 的多肽覆盖该植入器具。

15 32. 一种诊断试剂，它包括权利要求 1 的核酸片段或权利要求 16 的多肽。



说明书

釉质基质相关多肽

发明领域

5 本发明涉及编码属于一组被称为 **amelin** 的多肽的新核酸序列，其多肽序列包含与细胞表面识别有关的四肽功能域。 **amelin** 序列的可能用途涉及对硬组织形成失调的诊断，以及在生物物质形成中可作为基质成分或细胞识别标记的 **amelin** 蛋白或其片段的产生。本发明还涉及含
10 本发明核酸片段的表达载体以生产蛋白质，涉及含所述表达载体的生物，生产该多肽的方法，含该多肽的组合物，以及治疗各种硬组织疾病或失调症的方法。

技术背景

在骨，牙质和其他组织中， I 型胶原或类似蛋白质装配成原纤维基质，后者在一些情况下作为结合矿物晶体的支架。邻近的细胞与该基质
15 建立特异性接触，该接触由细胞外蛋白质如胶原和细胞表面受体，如整联蛋白中的功能域的相互作用的介导。已经在数种细胞外蛋白质中鉴定了涉及这些接触的肽功能域（ **Yamada&Kleinman,1992** ）。在釉质中尚未发现相当于骨中胶原纤维的结构网软骨和牙质。而且，还没有在介导其锚着于细胞粘着分子的釉质基质蛋白质中鉴定出序列片段。
20 釉质蛋白质 牙釉蛋白和 **enamelin** 不含有这类蛋白质功能域。新沉积的釉质的矿物成分是总质量的 15% 左右并在以后随蛋白质的降解而增加到 95%（ **Robinson 等， 1988** ）。

在釉质中已鉴定了主要的两组蛋白质：**enamelin** 和牙釉蛋白（ **Termine 等， 1980** ）。在成熟釉质中的蛋白质片段类似于 **enamelin**
25 之一， **tuftelin**，它已被抗体定位于釉质角柱体之间。已确定了对应于 **tuftelin** 的 cDNA 序列，并推测该蛋白质可能在釉质矿化中具有功能（ **Deutsch 等， 1991** ）。其余（迄今所描述的） **enamelin** 对釉质形成的意义可能是有争论的，因为主要蛋白质种类与来自血流的蛋白质是同样的（ **Strawich&Glinmcher,1990**）。是否牙釉蛋白（最常见的蛋白质）
30 为釉质基质提供一种支架仍在讨论之中（ **Simmer 等， 1994** ）。

先前已经汇集了一些从大鼠原位杂交文库随机筛选的部分序列（ **Matsuki 等， 1995** ），其中有些序列显示与本发明的序列有同源性。



没有阅读框被认为来自于这些部分序列。没有说明是否多肽被这些序列编码亦没有给出关于这类多肽之可能功能的假设。

5 已经在猪的未成熟釉质中鉴定了非牙釉蛋白蛋白质（ Uchida 等，1995 ）。一种 15kDa 蛋白质具有与先前已知 **enamelin** 质非同源的 N 末端氨基酸序列（ **VPAFPRQPGTHGVASL-** ）。假设非牙釉蛋白包含 **enamelin** 质的一个新家族但没有提出它们的功能。该蛋白质还没有被完全测序并且它们的基因属于未知。

10 **WO89/08441** 涉及包括用于诱导活矿化组织的各部分之间的结合的一种组合物，其中活性成分来源于牙釉质的一种前体，即所谓的釉质基质。该组合物通过促进矿化组织再生诱导结合。该活性成分是一个蛋白质组分的一部分并以具有高达 **40.000kDa** 分子量为特征但没有鉴定为单一蛋白质。

发明概述

15 虽然通常产生大量矿化的基质蛋白质，它们较差的溶解度妨碍了直接分析。在牙基质中基质蛋白质的生理降解发生在成熟期时的矿物质获得过程中并给分析基质蛋白质造成额外困难。本发明基于这种考虑即由于基质形成细胞合成大量的相应蛋白质，它们应含有高拷贝数的 **mRNA**。因此，对基质形成细胞的主要种类的 **mRNA** 的序列分析可绕过部分问题并帮助研究基质的某些蛋白质成分。

20 这些考虑引入了所采用的导致发现新 **amelin mRNA** 序列的方法，即本发明的基础。简言之，建立了含发育牙齿的 **mRNA** 序列种类的基因文库。从单细菌克隆得到各个序列并用于整个发育中的牙齿组织切片的原位杂交实验。在形成硬组织基质细胞（例如成釉质细胞）中所检测到的序列被确定并被用于查询序列数据库。大多数这样筛选的序列在
25 数据库中都有描述，但在现称为 **amelin** 序列的两个序列没有描述。在釉质基质形成期间新 **mRNA** 的这两个变异体在大鼠成釉质细胞中被高水平表达。所述序列分别含有 **407** 和 **304** 氨基酸残基的开放阅读框。被命名为 **amelin** 的被编码蛋白质富含脯氨酸，亮氨酸和甘氨酸残基并含有与细胞表面相互作用的其他功能域相结合的肽功能域 **Asp-Gly-Glu-Ala**（一种整联蛋白识别序列）。编码 C 末端 **305** 个氨基酸残基的序列，即 **SEQ ID NO 2**：中的氨基酸 **102-407** 和 **SEQ ID NO**： **4** 中的氨基酸 **19-324**，**3'** 非翻译区部分和在非翻译 **5'** 区的微卫星重复序列在
30



两个 mRNA 变异体中是一致的。其余的 5' 区包括长的变异体（在 SEQ ID NO : 1 中核苷酸 12-349）所特有的 338 个核苷酸，共有的 54 个核苷酸和仅在短的变异体中存在的 46 个核苷酸（在 SEQ ID NO : 3 中核苷酸 66-111）。十四个核苷酸有潜力编码两个蛋白质在不同阅读框中的 5 个氨基酸（在 SEQ ID NO : 1 中的 390-403 和 SEQ ID NO : 3 中的 52-65）。较长的变异体的阅读框包括典型 N 末端信号肽的密码子。amelin mRNA 序列的性质表明 amelin 是釉质基质的一种成分并且是迄今涉及成釉质细胞表面与其细胞外基质之间的结合相互作用的唯一蛋白质。

10 预期 amelin 肽或其部分可通过化学或通过使用本文描述的序列信息借助表达载体来翻译合成。进一步预期这些肽可对设计修复牙齿或骨的医疗设备有用。这些肽还可与以提高人工植入材料的生物相容性的目的与这种材料结合使用。人的 amelin mRNA 或基因序列可有助于硬组织形成的遗传病症的诊断。

15 详细描述

为获得可能是难于以直接途径分析的细胞外基质蛋白质的序列资料，以含有基质形成细胞的所有组成部分 mRNA 的 λ 噬菌体建立一个 cDNA 库。以下列途径筛选 amelin RNA 序列：

20 如实施例 4 进行影印嗜斑拓取（replica plaque lift）实验并与分别与 cDNA 和牙釉蛋白以及胶原寡聚物杂交。嗜菌斑呈现与 cDNA 相对强的杂交信号但与寡聚物无信号被进一步分析，设想它们含有常在 cDNA 中被描述但与牙釉蛋白和胶原不同的序列。这些阳性嗜菌斑克隆中的十五个被转变成 Bluescript 质粒。

25 为鉴定在基质形成细胞（即可参与基质产生和生长磨牙的矿化的细胞）中表达的序列，合成 RNA 探针（riboprobe）供原位杂交用。挑选四天龄大鼠，因为涉及釉质基质生成的牙釉蛋白 RNA 的浓度约在此时最高。图 1 显示用 amelin 探针所得结果（见实施例 4 和图 1a），作为与牙釉蛋白 RNA（图 1b）和胶原（图 1c）反应的比较。在分泌期检测含成釉质细胞的内釉质上皮中的 amelin 和牙釉蛋白。胶原探
30 针主要修饰位于间充质髓周边的成牙本质细胞以及牙槽骨的成骨细胞。因此断定 amelin 可对形成釉质基质起作用。对在牙结构中造成显示探针原位杂交信号阳性的十四 cDNA 插入片段进行部分测序。所



述片段用于查询基因库和 EMBL 数据库以对它们作出鉴定。迄今两个新序列没有被描述。

为确定全 *amelin* mRNA 序列, 用从上述初始 *amelin* 序列得到的一个寡核苷酸筛选牙 cDNA 文库并分离到 6 个长度范围在 0.5 和 2kb 之间的另外的插入片段。序列分析表明所有 7 个克隆代表了对应于 3' mRNA 部分的序列。然而, 在两个最长的插入片段中发现了两个不同的 5' 区特称为 *amelin1* 和 *amelin2* (图 2)。为得到全长的序列描述, 从大鼠磨牙建立随机引物库, 并用从所述两个突变体的各自 5' 端(图 2 下划线)得到的两个不同的寡核苷酸筛选。用从 *amelin1* 的 5' 部分得到的 *amelin2* 和 *amelin13* 克隆的 5' 部分杂交分离出 5 个克隆。序列分析证实了前面的结果并扩延了现称为 *amelin1* 和 *amelin2* 的两个变异体的序列并分别在序列表中表示为 SEQ ID NO: 1 和 SEQ ID NO: 3。两个 5' mRNA 序列在最多达 100x (AG) (数据未显示) 的多嘌呤重复处终止。考虑到在 5' 末端的 AG 重复和在 3' 末端的聚腺苷酸尾, 合在一起的序列(图 2)不比由 Northern 印迹(见下)所确定的 mRNA 短。从 polyT primed cDNA 库得到的克隆的序列分析显示出聚腺苷酸添加信号 AATAAA (双下划线) 的一个意外的 3' 变异下游。在某些克隆中观察到如预期的 15 个核苷酸下游的聚腺苷酸尾, 但在其他克隆中观察到其位于最多达 79 核苷酸的较长距离。图 2 中的序列显示最长距离的聚腺苷酸化部位变异体。所有的变异被定位在终止密码子的下游。

两个 cDNA 序列变异体显示了单一的长开放阅读框(图 2)。框内终止密码子出现在聚 (AG) 和开放阅读框之间, 因而看来聚 (AG) 或近侧序列似乎不可能编码蛋白质。*amelin* 的阅读框起始于聚 (AG) 重复序列的 84 个核苷酸下游。前 86 个氨基酸由一个未在 *amelin2* 中显现的序列所编码。*amelin1* 的氨基酸 87 到 99 由 *amelin1* 和 *amelin2* 共有的序列所编码。然而, 该序列不编码 *amelin2* 蛋白质。虽然它包括了 ATG 密码子, 框内终止密码子只容许编码七肽。与七肽终止密码子重叠的下一个 ATG 起始编码 *amelin2* 的最长序列。有趣的是, 它的前十四个核苷酸以不同的框架编码 *amelin1* 和 *amelin2*(图 2 中加阴影部分)。接着的编码 *amelin2* 的 15 个氨基酸的 46 个核苷酸未出现在 *amelin1*RNA 中。在 *amelin2*RNA 中的该“插入片



段”引起两个阅读框的同步化，以至最后的 305 个氨基酸残基对两个蛋白质是共有的。在 *amelin2* 的插入片段中有一个框内 ATG 密码子，它可能是作为另一个翻译的起始。在此情况下，*amelin2* 会少 5 个氨基酸并且没有两个框架编码序列段。最长的可能的开放阅读框含 *amelin1* 的
5 407 个氨基酸残基的密码子和 *amelin2* 的 324 个残基的密码子。

自提出第一次申请以来，已经再次检查了测序结果并做了某些修正。*amelin1* 序列已被修正如下：132 号核苷酸已被从 G 变为 G 没有引起氨基酸变化。191 号核苷酸已被从 G 变为 A 引起 33 位的精氨酸变为 33 位谷氨酰胺。200 号核苷酸已被从 G 变成 C 引起 36 位甘氨酸变为 36
10 丙氨酸。617 号核苷酸已被从 G 变为 C 引起 175 位甘氨酸变为 175 位丙氨酸。809 号核苷酸已被从 G 变为 C 引起 239 位甘氨酸变为 239 位丙氨酸。976 号核苷酸已被从 C 变为 G 引起 295 位脯氨酸变为 295 位丙氨酸。1649 号核苷酸已被从 C 变为 A 未引起氨基酸变化。*amelin2* 序列已被修正如下：326 号核苷酸已被从 G 变为 C 引起 92 位甘氨酸变为 92 位丙氨酸。518 号核苷酸已被从 G 变为 C 引起 156 位甘氨酸变为 156 位丙氨酸。685 号核苷酸已被从 C 变为 G 引起 212 位脯氨酸变为 212 位丙氨酸。1358 号核苷酸从 C 变为 A 没有引起氨基酸变化。

为评估 *amelin* 转录物的大小，进行了从 4 天龄大鼠的磨牙制备的总 RNA 的 Northern 印迹分析（图 3，a 泳道）。DIG 标记的 *amelin* cRNA 探针杂交到—2kb 和—1.9kbRNA 带。如 cDNA 序列分析所确定的，若将 0.2kb 的聚（AG）重复序列和 0.2kb 的聚腺苷酸加入到展示序列中则 *amelin1* 和 *amelin2* mRNAs 是 2.3kb 和 2.0kb 长。这两个测定十分符合说明所述序列包含 *amelin* 的全部或几乎全部 mRNA。为了比较，出示牙釉蛋白的两个主要 mRNAs（1.1kb 和 0.8kb 长）（图 3，
20 b 泳道）。通过液相杂交实验确定了来自磨牙的总 RNA 中的与牙釉蛋白相关的 *amelin* RNA 的大部分（Mathews 等，1989）。若与牙釉蛋白的含量相比，*amelin* RNA 的量占大约 5%。*amelin1* 和 *amelin2* 的序列比较表明两个 RNAs 是同一初级转录物的剪接变体，因为在相匹配的序列部分中未发现改变。

30 在 *amelin1* 和 *amelin2* 中最常见的氨基酸是脯氨酸，甘氨酸和亮氨酸；在任何一个序列中都没有半胱氨酸（参见下表 1）。推定的 *amelin1* 蛋白质的氨基末端具有信号肽的特点：残基 14 到 21 是疏水的并具有亮



氨基酸序列段 (图 2; Leader, 1979)。在 amelin2 序列中没观察到可比较的基元。两个 amelin 都含有肽功能域 DGEA (Asp-Gly-Glu-Ala)(amelin1 中的氨基酸 370-373 和 amelin2 中的氨基酸 287-290) (图 2 框起的部分), 它在早些时候已被鉴定为构成对细胞表面蛋白 a2b1 整联蛋白的 I 型胶原蛋白的识别部位 (Staatz 等, 1991)。此外, 包括了具有 VTKG (Val-Thr-Lys-Gly) 序列的 trombospondin 样细胞粘着功能域 (在 amelin1 中的氨基酸 277-280 和 amelin2 中的氨基酸 194-197) (Yamada&Kleinman 等, 1992)。这两个功能域的存在表明 amelin 是细胞外基质的成分。所预言的 amelin 在水溶液中的低溶解度与这个模型相一致。在 amelin1 中信号序列的存在证实其表现为一种分泌蛋白质。在 amelin2 中没有信号序列不意味着该蛋白质不被分泌。无信号序列的分泌蛋白的先例是鸡卵清蛋白, 其中内在的未被切割的序列提供同样的功能 (在 Leader, 1979 中讨论)。预言在细胞表面相互作用中有意义的两个另外的功能域, EKGE (Glu-Lys-Gly-Glu)(amelin1 中氨基酸 282-285 和 amelin2 中氨基酸 199-202) 和 DKGE (Asp-Lys-Gly-Glu)(amelin1 中氨基酸 298-301 和 amelin 氨基酸 215-218), 簇集在同一区域。如在本段落中所描述四个肽功能域的结合是任何釉质基质蛋白迄今尚未描述的一个特点。

由于所预言的低溶解度, amelin 在 E.coli 细胞中作为一种与硫氧还蛋白在氨基末端的末端的融合蛋白被表达。6 个组氨酸标记被加到羧基末端的末端并在 Ni 柱上纯化蛋白质。洗脱物含一种主要的融合蛋白还有几种在 Western 印迹分析中与抗 amelin 兔血清反应的肽。可通过抗硫氧还蛋白亲和层析进一步纯化该蛋白质。

已产生抗 amelin 蛋白的抗体。用 amelin-硫氧还蛋白融合蛋白免疫兔并通过在偶联到 CNBr-活化的 Sepharose 上的 amelin 融合蛋白亲和层析纯化免疫血清。在硫氧还蛋白偶联的 Sepharose 上可达到进一步的纯化。这些抗体已被用于例如在大鼠牙中 amelin 的免疫组化定位。

还确认了牙提取物中 amelin 的存在。大鼠磨牙在碳酸钠缓冲液 pH10.8, 1 毫摩尔 EDTA+蛋白酶抑制剂中被匀浆。用抗 amelin-硫氧还蛋白免疫血清通过 Western 印迹分析了粗提物的上清液。检测出与两个 amelin 变异体相对应的带。在 SephadexG100 柱上进一步层析粗提物。浓缩对应于 amelin 分子量的组分并进行制备电泳。电泳之后, 现在通



过 N-末端序列分析鉴定电泳带。如果电泳带之一是 **amelin**，测定体内转化起始。

5 通过研究 2，5，10，15，20 和 25 天龄的 Sprague-Darley 大鼠的上颌检测了在牙齿不同发育阶段中 **amelin** 序列的表达。发现 **amelin** 伴随牙釉蛋白出现在原位杂交实验中，即在分泌阶段开始时成釉质细胞伸长期间。在较后的阶段，牙釉蛋白与 **amelin** 的 mRNA 呈现根本不同的杂交型。在成熟阶段牙釉蛋白 mRNA 大量消失仅在成熟了的成釉质细胞的后期有少量存留，这个发现与 Wurtz 等（1995）的发现是一致的。然而用 **amelin** 探针得到的信号在成釉质细胞的成熟阶段没有或只有很少程度的降低。

10 在功能上，两个阶段是不同的因为在成熟期没有额外的釉质基质被沉积。然而，似乎在两个时期都有矿物质沉积，因为新沉积的釉质已经含有矿物质。将这些事件与各自的 mRNA 的出现相关连，**amelin** 有可能参与矿化过程。如上所述 **amelin** mRNA 编码一种含细胞结合功能域的蛋白质，说明它还或选择性地参与成釉质细胞结合到釉质表面。**amelin** 蛋白可起蛋白水解酶的功能。这已通过从丙烯酰胺凝胶切下并电洗脱主要融合蛋白带进行了检测。在室温温孵过夜后，融合蛋白出现 3 条带。4℃ 温孵的对照只产生 1 条带。这说明在较高温度降解发生。要确定是否 **amelin** 事实上起蛋白水解酶的功能需要进一步实验。

20 本发明提供编码对细胞结合功能域有特异结合的蛋白质。该蛋白质是硬组织基质的成分并介导与细胞表面的接触。图 2 描述该蛋白质的编码序列并且该序列从核苷酸的 95 位延伸到 1361 位。细胞结合功能域的新组合占据 969 位到 1259 位核苷酸。各个结合功能域可以现有形式被组合或以不同氨基酸周围的前后关联部分展现或被结合进入非蛋白质性质的聚合物中。所述核苷酸序列和衍生的蛋白质序列均可首先用作按照标准技术的 **amelin** 蛋白质的人工表达工具（Ausubel 等，1994），其次作为肽化学合成的资料。所述序列可被用于建立硬组织形成中失调的鉴别之用的诊断标准，和作为在组织工程中生产生物材料的手段。此外，本发明提供含有位于转录启动子下游的所要求的序列的表达载体，以及生产和分离 **amelin** 的方法，这些方法基于所述表达载体的使用。

30 本发明涉及与含至少一个能介导多肽锚着于细胞粘着分子的序列元件的多肽相关的全部釉质基质。



如下面进一步详细描述，根据术语“釉质基质相关多肽”，就其最广泛的意义来说，意味着一种是釉质基质蛋白质或具有类似性质即能介导釉质和细胞表面之间接触的合成的蛋白质的多肽。

在本说明书和权利要求中，术语“多肽”包括具有至少两个氨基酸残基和最多 10 个氨基酸残基的短肽和寡肽（11-100 个氨基酸残基）以及蛋白质（包括可通过被糖基化，被脂质化，或通过包含辅基而被化学修饰的至少一个短肽，寡肽，或多肽的功能实体）。该多肽定义还包括动物包括人中的肽/蛋白质的天然形式以及以任何形式的表达载体转化任何种类的宿主而得的重组蛋白质或肽，并且包括化学合成的肽。

本发明的已被称做 **amelin** 蛋白质的多肽不同于已知的 **enamelin** 质牙釉蛋白和 **enamalin** 因为它们含有至少一个能介导该多肽锚着于细胞粘着分子的序列元件。特别是，它们含有一个选自包含四肽 **DGEA**（**Asp-Gly-Glu-Ala**），**VTKG**（**Val-Thr-Lys-Gly**），**EKGE**（**Glu-Lys-Gly-Glu**）和 **DKGE**（**Asp-Lys-Gly-Glu**）的序列元件。

本发明的优选实施方案是有氨基酸序列 **SEQ ID NO : 2** 或其类似物或其变异体的多肽以及有氨基酸序列 **SEQ ID NO : 4** 或其类似物或其变异体的多肽，和有氨基酸序列 **SEQ ID NO : 2** 或 **SEQ ID NO : 4** 的亚序列的多肽。

进一步说，本发明涉及编码能介导釉质和细胞表面之间的接触的多肽的核酸片段。通过术语“核酸”意指作为 DNA 或 RNA 存在的高分子量多核苷酸并且可是单链或双链。

虽然编码包含 **SEQ ID NO : 2** 的氨基酸残基 1 到 407 之多肽的核酸片段和编码包含 **SEQ ID NO : 4** 的氨基酸残基 1 到 302 之多肽的核酸片段是优选实施方案，本发明还涉及编码有 **SEQ ID NO : 2** 中所显示的氨基酸序列或其类似物或其变异体之多肽的核酸片段并涉及编码有 **SEQ ID NO : 4** 显示的氨基酸序列或其类似物或其变异体之多肽的核酸序列。

术语“有 **SEQ ID NO : 2**（或 **SEQ ID NO : 4**）显示的氨基酸序列或其类似物或其变异体之多肽”意指有氨基酸序列 **SEQ ID NO : 2**（或 **SEQ ID NO : 4**）之多肽以及有所述序列的类似物或变异体之多肽，它们当本发明的核酸片段在适合的表达系统中被表达时被产生并且



能介导釉质和细胞表面之间的接触，例如通过包含在组织培养中的细胞外基质和基质形成细胞的测试系统所证实的。通过加入多肽片段测试该多肽的浓度依赖型生物活性。通过显微镜观察证实如果该片段能竞争过细胞外基质蛋白质和细胞之间的接触，则细胞就从基质脱离。已知培养的细胞粘着到纤连蛋白,骨桥蛋白,胶原,层粘连蛋白和玻连蛋白。通过蛋白质的 RGD 细胞附着功能域介导细胞结合活性。Amelin 含有择一的细胞结合功能域 DGEA 和 VTKG。例如通过用 amelin, BSA 或纤连蛋白包被细胞培养皿可检测细胞附着。通过检测内源性 N-乙酰-β-D-氨基己糖苷酯酶可对结合 UMR 大鼠骨肉瘤细胞做定量测定。

因此类似物或变异体是并不精确地具有 SEQ ID NO: 2 或 SEQ ID NO: 4 中所显示的氨基酸序列,但仍能如上面定义的介导釉质和细胞表面的接触。例如,一般说,这样的多肽当与实施例中描述的 amelin 蛋白质相比较时将是在氨基酸组成或翻译后修饰例如糖基化或磷酸化方面有一定程度不同的多肽。

因此在本发明上下文中使用术语“类似物”或“变异体”指与从如实施例中描述的 amelin 蛋白质得到的特有氨基酸序列 SEQ ID NO: 2 和 SEQ ID NO: 4 相类似的氨基酸组成或序列的蛋白质或多肽,允许其有少数改变氨基酸序列的变异例如氨基酸的缺失,替换或插入或它们的组合而产生 amelin 蛋白质类似物。这些修饰可产生所述类似物的有趣和有用的新特性。可从动物或人得到类似物多肽或蛋白质或可它们可有部分或完全的合成来源。还可通过使用重组 DNA 技术得到该类似物。

因此本发明的一个重要的实施方案涉及其中至少一个氨基酸残基已被不同的氨基酸残基替换的多肽和/或其中至少一个氨基酸残基已被缺失或被加入以至引起一个含不同于 SEQ ID NO: 2 或 SEQ ID NO: 4 中显示的氨基酸序列或如下面定义的氨基酸序列的亚序列但必须有如上面定义的 amelin 活性的多肽。

本发明的一个有趣的实施方案涉及一种多肽,它是本发明的含 6 到 300 氨基酸的多肽的亚序列,例如至少 10 个氨基酸,至少 30 个氨基酸,如至少 60, 90 或 120 个氨基酸,至少 150 个氨基酸或至少 200 个氨基酸。

本发明的特别重要的实施方案是含 SEQ ID NO: 2 (amelin1)中氨基酸 1-407 的多肽和含 SEQ ID NO: 4 (amelin2)中氨基酸残基 1-



324 的多肽。

已将氨基酸序列 SEQ ID NO : 2 和 SEQ ID NO : 4 与已知氨基酸序列做了比较。与细胞外基质蛋白质（与其同源性最高），即牙釉蛋白和胶原 IV 的同源（同一性）程度是非常的低，分别是 23% 和 26%。

5 同一性散布于整个蛋白质中并且不限于特殊区域。在这方面应该注意 **amelin** 不含重复的三连基元，相反胶原总被重复的三连基元，G-X-Y 所编码。对胶原 IV 和牙釉蛋白的同源性可能由于在两个蛋白质中的高含量的脯氨酸。因此看来 **amelin** 蛋白质与先前已知细胞外蛋白质，特别是釉质基质蛋白质仅有微弱的类似性。

10 本发明的一个重要实施方案涉及有一种氨基酸序列的多肽，其中的连续的一串 20 个氨基酸与选自 SEQ ID NO : 2 或 SEQ ID NO : 4 所显示的氨基酸序列的同样长度的一串氨基酸有至少 80% 程度的同源性。

与 SEQ ID NO : 2 或 SEQ ID NO : 4 中显示的多肽有至少 80% 如至少 85%（例如至少 90%）同源性或同一性的本发明多肽序列构成重要的实施方案。由于 SEQ ID NO : 2 和 SEQ ID NO : 4 显示的序列看上去相当独特，本发明范围还包括对选自 SEQ ID NO : 2 或 SEQ ID NO : 4 中显示的氨基酸序列的类似连续一串 20 个氨基酸的同源程度至少是 25%，如至少 50% 或至少 75% 的多肽。可从其他物种，例如哺乳

20 动物如小鼠，兔，豚鼠，猪，牛或人的类似蛋白质得到这类序列。

通过使用本申请中公开的序列，本领域技术人员将能检测，克隆，测序，生产，和研究 **amelin** 的人译本。由于最方便的可用牙齿材料是被拔掉或切下的牙齿，主要是第三磨牙或赘生牙，一个实际问题是缺少原初材料。这些牙齿的发育通常相当晚并因此参与基质形成的细胞远在

25 第二期之后或不再出现。

另一种选择是，可从合用的组织培养得到原初材料，检测其中提取的 RNA 以测定 **amelin** 信使的存在。在人骨肉瘤细胞（Sao2 细胞）的情况得到 Northern 印迹阳性，尽管阳性 RNA 长度与大鼠 **amelin** mRNA 相比要小得多。

30 因此，建立人骨肉瘤细胞（Saos2 细胞）cDNA 文库以便发现代表一个或多个个人类 **amelin** 或 **amelin** 样结构的 cDNA。以类似方式，从最未发育的牙齿建立 cDNA 库并用 **amelin** 探针或用从 Saos2 库得到的探



针进行筛选。

术语“序列同源性”意指在多肽氨基酸的同一性和位置方面，相匹配的两个或两个以上氨基酸区段中氨基酸序列的同一性。

因此本文使用术语“同源性”以解释一个给定的多肽的氨基酸序列和在 SEQ ID NO : 2 或 SEQ ID NO : 4 中显示的氨基酸序列间同一性的程度。可从一个核苷酸序列如一个 DNA 或 RNA 序列，例如通过如下面定义的杂交得到的或可通过常规氨基酸测序方法得到的这样的序列推断与 SEQ ID NO : 2 或 SEQ ID NO : 4 中显示的氨基酸序列比较的氨基酸序列。优选以成熟多肽的氨基酸序列（即不考虑任何先导序列）确定同源程度。一般说，当比较核苷酸序列时为确定它们的内部同源性只使用编码区。

其中的一个方面是，本发明涉及编码如上定义的本发明多肽的核酸片段。特别是，本发明涉及基本上包含 SEQ ID NO : 1 中显示的序列或基本上包含 SEQ ID NO : 3 中显示的序列的核酸片段。

本发明还涉及与含 SEQ ID NO : 1 中显示的核酸序列或 SEQ ID NO : 3 中显示的核酸序列的核酸片段或所述序列的部分核酸片段杂交的核酸片段，而它们在严谨条件例如 5 毫摩尔单价离子（0.1xSSC），中性 pH 和 65 °C 下是稳定的。

在另一方面，本发明涉及至少 18 个核苷酸的 SEQ ID NO : 1 中显示的核苷酸序列或 SEQ ID NO : 3 中显示的核苷酸序列的类似物或亚序列，它们

1. 与 SEQ ID NO : 1 或 SEQ ID NO : 3 中显示的序列有至少 90% 同源性，和/或

2. 编码一种多肽，其氨基酸序列与 SEQ ID NO : 2 或 SEQ ID NO : 4 中显示的氨基酸序列至少有 80% 同源性。

本发明还涉及编码含氨基酸序列 SEQ ID NO : 2 或 SEQ ID NO : 4 的亚序列之多肽的核酸序列。在本说明书和权利要求中，术语“亚序列”指优选有至少 15 个核苷酸，更优选至少有 18 个核苷酸，并最优选有至少 21 个核苷酸的序列。在本发明的一些实施方案中，本发明核酸片段的亚序列或类似物将包括至少 48 个核苷酸，如至少 75 个核苷酸或至少 99 个核苷酸。“亚序列”应符合至少上述标准 1) 和 2) 中之一或应与含 SEQ ID NO : 1 中显示的核苷酸序列或 SEQ ID NO : 3 显



示的核苷酸序列的核酸片段杂交。

众所周知在 PCR 技术中如本文描述的小片段是有用的。这类片段和亚序列可在其他应用中作为鉴别如实施例 4 中描述的本发明核苷酸序列的 mRNA 片段的探针。

5 关于本发明的核酸片段的术语“类似物”意指编码功能上类似于由 SEQ ID NO: 2 和 SEQ ID NO: 4 所编码的多肽的一种多肽的核酸片段，因为如上述测试所证实的该类似物能介导多肽锚着于细胞粘着分子。

10 众所周知同一氨基酸可被不同密码子编码，密码子的利用特别涉及所论及的生物表达核苷酸序列时的偏爱。因此，本发明核酸片段的一个或一个以上的核苷酸或密码子可被其他当表达时则产生相同于或基本上相同于所谈到的核酸片段编码的多肽的核苷酸或密码子改变。

考虑到少数核苷酸变异对介导上述测试证明的釉质与细胞表面之间的接触没有明显的相反作用，在本上下文中还使用术语“类似物”以指明编码构成 amelin 样多肽的氨基酸序列的核酸片段。

20 术语“明显的相反作用”意指类似物的活性（当按上面描述的而确定时）应至少 10%，更优选至少 20%，甚至更优选至少 25% 如至少 50% 的天然 amelin 附着或脱离活性。可从一个生物体如一个动物或一个人得到同源核酸片段或核苷酸序列，或它们可以是部分或完全合成来源的。还可通过使用重组 DNA 技术得到该类似物。

而且，术语“类似物”和“亚序列”是要考虑到序列中的变异如一个或一个以上核苷酸替换，插入（包括内含子），添加和重排，这些变异对由核酸片段或其序列编码的多肽没有任何基本上的相反作用。

25 术语“替换”意指用一个或一个以上不同的核苷酸替换全长核苷酸序列中的一个或一个以上的核苷酸，“添加”被理解为意指在全长核苷酸序列任何一端加入一个或一个以上的核苷酸，“插入”意指在全长核苷酸序列内部引入一个或一个以上的核苷酸序列，“缺失”意指无论在该序列的任何一端或在序列内的任何适当位点从全长核苷酸序列删除一个或一个以上的核苷酸，和“重排”意指在该核酸内或多肽序列内分
30 别互换两个或两个以上的核苷酸残基。然而，该核酸片段还可通过在将其插入到生物体之前或之后的诱变而被修饰。

根据本发明在本说明书和权利要求中就片段，序列，亚序列和类似



物所使用的术语“片段”，“序列”，“亚序列”和“类似物”当然应被理解为不包含在其天然环境中的现象，而包含例如被分离的，被纯化的，体外或重组形式中的现象。

5 在本发明的一个实施方案中，可通过从细胞或组织提取 RNA 并将其转变成 cDNA 供随后的聚合酶链反应（PCR）中使用来实现对基因突变的检测和对 **amelin**RNA 的定量。可基于本发明的核酸片段如 SEQ ID NO：1 或 SEQ ID NO：3 中显示的核酸片段合成 PCR 引物。用于检测和/或定量的这种方法可被用做诊断以高于或低于正常的量表达 mRNA 之疾病状况的诊断方法。

10 属于本发明范围内的还有包含能检测本发明核酸片段核苷酸探针的诊断试剂以及诊断 **amelin** 表达失调（**deregulated**）的疾病和/或 **amelin** 基因被突变的疾病的方法，包括从被怀疑患有存在高于正常量的 **amelin** 蛋白质或 **amelin** 突变形式的疾病的病人取样进行 PCR 分析，其中样品与按照上面的描述诊断试剂进行接触，使任何核酸片段得到扩增并确定样品中的任何一致的或同源的核酸片段的存在。另一方面，本发
15 明还涉及含根据本发明 **amelin** 多肽的诊断试剂。

用重组 DNA 技术可生产本发明的多肽。本发明的一个重要的实施方案涉及包含本发明的核酸片段的表达系统。特别是，本发明涉及可复制表达载体，它载有根据本发明的核酸片段并能调节该核酸片段的表
20 达。

在本发明范围内有一种携带根据本发明表达载体的生物。本发明这方面可使用的生物包括微生物如芽孢杆菌属，埃希氏杆菌属或沙门氏菌属细菌，酵母如酵母属，毕赤氏酵母属，原生动物或从多细胞生物得到的细胞如真菌，昆虫细胞，植物细胞，哺乳动物细胞或细胞系。若生物
25 体是细菌，该细菌最好是埃希氏菌属，例如 *E.coli*。无论使用何种类型的生物，本发明的核酸片段或直接或用适当的载体被引入到该生物中。另外选择是，通过直接或用一个表达载体引入本发明的核酸片段或类似物或其亚序列可在哺乳动物细胞系中生产多肽。

核酸片段或其类似物或其亚序列还可被克隆到适当的稳定表达载体
30 体中，然后被放入适当的细胞系。然后根据在适合于所使用的载体和细胞系的条件下产量的水平来选择产生需要多肽的细胞。选出的细胞被进一步培养并形成所需要多肽的十分重要和连续不断的来源。被用于生产



本发明多肽的生物也可以是高等生物例如动物。

5 本发明核酸序列的一个特殊类似物的实例是含 SEQ ID NO : 1 或 SEQ ID NO : 3 中显示的 DNA 序列或其部分的 DNA 序列并且该序特别适合在 E.coli 中表达。当与适当的调节序列一起被插入到 E.coli 中时，该 DNA 序列导致基本上具有 SEQ ID NO : 2 或 SEQ ID NO : 4 或其一部分的多肽的表达。因此，该 DNA 序列包含被 E.coli 所识别的特异密码子。

10 在本上下文中，术语“基因”被用于指明参与产生多肽链的核酸序列并且它包括编码区前后的区域（5'上游和3'下游序列）以及间隔序列，位于各编码区段，外显子之间或在5'上游或3'下游区域内的内含子。5'上游区域包括控制基因表达的调节序列，典型的是启动子。3'下游区包括参与基因转录终止的序列和非强制性地包括负责转录物的聚腺苷酸化的序列和3'非翻译区。本发明还涉及含如上述编码本发明多肽核酸片段的表达系统，该系统包含能介导所述核酸片段表达的5'侧翼序列。

15 本发明进一步涉及含编码本发明多肽或如本文定义的融合多肽的核酸序列。在一个特别重要的实施方案中，本发明的核酸片段或类似物或其亚序列或如本文定义的本发明融合核酸片段可被可复制表达载体携带，后者能在宿主生物体或细胞系中复制。

20 该载体可特别是质粒，噬菌体，粘粒，微型染色体或病毒。在本发明的一个有趣的实施方案中载体在被引入进宿主细胞时，被整合到该宿主细胞的基因组中。

25 在本发明的一个特别方面，以产生融合多肽为目的，本发明的核酸可包括编码不同于或相同于本发明多肽的另一核苷酸片段，它在阅读框内被融合于编码 amelin 多肽的 SEQ ID NO : 1 或 SEQ ID NO : 3 或其类似物所显示之序列的核酸片段。在使用 DNA 技术时，融合的核酸序列可被插入到适当的载体或基因组中。另外的选择是，核酸片段之一被插入到已经含有其他核酸片段的载体或基因组中。通过分别插入两个核酸片段并允许表达发生也可产生融合多肽。宿主生物（可属于原核生物来源或真核生物来源）在保证融合序列表达的条件下生长。然后
30 用适当方法纯化融合蛋白并从其融合配偶体分离本发明的多肽。因此，本发明的一个方面涉及产生本发明多肽的方法，包括如下步骤：



- (a) 将本发明的一个核酸片段插入到一个表达载体中，
- (b) 用步骤(a)中产生的载体转化适当的宿主生物，
- (c) 在适当条件下培养步骤(b)中产生的宿主生物以表达多肽，
- (d) 收获多肽，和
- 5 (e) 非强制性地对多肽进行转录后修饰。

如上描述的一种方法也属于本发明范围内，其中通过包含一个或一个以上的步骤的方法如用固定化的 **amelin** 多肽或与所述多肽起反应的抗体进行亲和层析和/或其他层析以及电泳方法分离所产生的多肽。

10 作为热处理，化学处理（ 甲醛，戊二醛等 ）或酶处理（ 肽酶，蛋白水解酶和蛋白修饰的酶 ）的结果，如上述产生的多肽可受到翻译后修饰。与其天然产生环境相比，在生物体中产生时，该多肽可以一种不同的方式被加工。例如，糖基化通常在高等生物如酵母或最好是哺乳动物细胞表达多肽时完成。正常情况下发现糖基化与氨基酸残基天冬酰胺，丝氨酸，苏氨酸或羟脯氨酸有关。这对去除或改变由所谈及的宿主生物引起的加工特点可能是或不是有利的。

15 按照本发明，继多肽在一个生物体或一个细胞系中表达之后，该多肽可被如此使用或可先从生物体或细胞系被纯化。如果多肽被表达成一种被分泌的产物，它可被直接纯化。如果多肽被表达成一种结合物，在纯化前它可能需要宿主的部分或完全破碎。用于多肽纯化的方法的实例是：（ i ）用抗体进行免疫沉淀或亲和层析，（ ii ）用适当的配基进行亲和层析，（ iii ）其他层析方法如凝胶过滤，离子交换或高效液相层析或上述之中的任何衍生方法，（ iv ）电泳方法如聚丙烯酰胺凝胶电泳，变性聚丙烯酰胺凝胶电泳，琼脂糖凝胶电泳和等电聚焦，（ v ）任何其他专门增溶溶解和/或纯化方法。

25 本发明还涉及基本纯的 **amelin** 多肽。在本上下文中，术语“基本纯”被理解为意指所谈及的多肽基本上没有其他成分，例如其他多肽或碳水化合物，它们可形成于多肽生产和/或回收过程中或另外被发现与多肽在一起。蛋白质纯度可以例如通过 SDS 凝胶电泳来评估。

30 本发明的高纯度多肽在该多肽被用于组合物时可能是有利的。还由于其高纯度，可以比常规低纯度多肽低的数量将基本纯的多肽用于大多数目的。

在本发明的一个方面，可从表达本发明多肽的适当细胞系获得纯的



多肽。还可通过熟知的液相或固相肽合成方法利用多肽序列的各个氨基酸的连续偶联制备本发明多肽。另外的选择是，可通过各个氨基酸的偶联形成多肽序列的片段然后再连接这些片段以至形成需要的多肽来合成多肽。因此这些方法构成本发明的另一有趣方面。

5 另一方面，本发明涉及治疗和/或预防牙周疾病的方法，该方法包括根据本发明对需要它的病人施用治疗或预防有效量的多肽。期望本发明多肽将参与牙骨质形成并因此促进牙周韧带的固定。

通过 **amelinRNA** 序列在骨形成细胞中的存在表明 **amelin** 蛋白在人工局部骨形成的过程（**context**）的用途：通过 Northern 杂交在大鼠股骨以及颅盖的骨组织中发现了满足 17 页 1-5 行给出的标准的 **amelinRNA** 的一个大小变异体。用 **amelin** 探针的原位杂交将该 RNA 定位于与生长的骨相关连的成骨细胞。在细胞培养中形成骨的大鼠颅盖细胞在整个骨形成期间也一直表达 **amelinRNA** 的骨-变异体（**C. Brandsten, C. Christersson** 和 **T. Wurtz**, 未发表）

15 在天然和实验骨形成系统中 **amelinRNA** 的存在表明 **amelin** 蛋白在骨形成中的作用。可以想象外部加入的 **amelin** 肽在试管内和医疗应用中都加速或介导骨形成。

而且，本发明涉及修复牙齿中损伤的方法，该方法包括向需要此种治疗的病人施用与充填材料相结合的根据本发明的有效量的多肽。

20 本发明还涉及接合两个骨单体的方法和涉及将植入体有效结合到骨中的方法。与此关联的内容中，如下面详细描述的可将多肽与载体联系起来施用。而且，本发明多肽可被用于促进或引起选自骨，釉质，牙质或牙骨质的硬组织的矿化。

另外，本发明还涉及促进植入器件或经皮器件从例如与在 US 4, 25 578, 079 中描述的方式的生物相容性的方法，该方法包括根据本发明用有效量的多肽覆盖该植入器件，由此例如允许肌肉或韧带与植入物接触。

30 本发明还涉及通过施用本发明的多肽将上皮锚着于硬组织表面的方法，该硬组织表面选自与牙植入物相关连的釉质，牙质或牙骨质。而且，本发明涉及防止与牙的植入关联之上皮的生长，该方法包括给需要此种治疗的病人施用预防有效量的根据本发明的多肽，例如由此防止上皮生长进入牙周韧带。



本发明的一个非常重要的方面涉及含 **amelin** 多肽和生理可接受赋形剂的一种组合物。该组合物可包含本发明的一种纯化了的重组多肽。特别的但非绝对的是，本发明涉及适合于局部施用，例如用于口腔黏膜表面，的组合物。

5 适合于局部施用的本发明组合物可以是擦剂，凝胶，溶液，悬浮液，糊剂，喷雾剂，粉剂，牙膏，和漱口液。

本发明包括通过用牙膏制剂，例如常用类型的市售牙膏，与本发明多肽混合配制的牙膏，它可在常规的基础上被使用以预防例如牙周炎。

10 药膏将常含有磨光剂，表面活性剂，胶凝剂和其他赋形剂如调味剂和着色剂。磨光剂可被选自那些在牙制剂中正在被用于此目的的材料。适合的实例是水不溶性偏磷酸钠或钾，水合或无水磷酸二钙，焦磷酸钙，硅酸锆或其混合物。特别有用的磨光剂是各种形式的硅石。磨光剂通常被细分，具有小于 10 微米的颗粒，例如 2-6 微米。磨光剂可以 10-99% 牙膏重量被使用。一般牙膏制剂将含 20-75% 的磨光剂。

15 适当的表面活性剂通常被包含在牙膏制剂中。表面活性剂一般是水溶性的非皂合成有机去污剂。适合的去污剂是下述构层的水溶性盐：高级脂肪酸单酸甘油酯单硫酸盐（例如氢化椰子脂肪酸单酸甘油酯单硫酸钠）；高级烷基硫酸盐（例如月桂硫酸钠）；烷芳香基磺酸盐（例如十二烷基苯磺酸钠）；和高级烷基磺酰乙酸（例如月桂基磺乙酸钠）。此外，在酰基中可使用含 12-16 碳原子的低级酯族氨基羧酸的饱和高级酯族酰胺并且其中的氨基酸部分得自有 2-6 碳原子低级酯族饱和单氨基羧酸，如甘氨酸，肌氨酸，丙氨酸，3-氨基丙酸和缬氨酸的脂肪酸胺，特别是 N-月桂基，肉豆蔻基和棕榈酰肌氨酸盐化合物。若需要的话也可包括常规的非离子表面活性剂。

25 表面活性物质一般以牙膏制剂重量的约 0.05-10% 的量存在，典型的约 0.5-5%。

典型牙膏流体将主要包括水，甘油，山梨糖醇，丙二醇或其混合物。一个有利的混合物是水和甘油，最好伴有山梨糖醇。可使用胶凝剂如天然或合成树胶或树胶样物质，例如爱尔兰藓或羧甲基纤维素钠。其他可用树胶如西黄蓍胶，聚乙烯吡咯烷酮和淀粉。它们通常以牙膏重量的约 10%，典型的约为牙膏重量的 0.5-5% 被使用。

药膏的 pH 基本上是中性的，如约 6-8 的 pH。若需要的话，可加



人少量调 pH 试剂例如少量酸如柠檬酸或碱性物质。

该牙膏还可含有其他物质如可溶性糖精，调味油（例如绿薄荷油，胡椒薄荷油，冬青油），着色剂或增白剂（例如二氧化钛），防腐剂（例如苯甲酸钠），乳化剂，硅氧烷类，乙醇，薄荷醇和叶绿素化合物（例如叶绿酸钠铜）。

本发明多肽在上述类型或下面讨论的类型的牙膏中的含量通常在以总牙膏组成重量所计算的重量的 1-20% 范围内，如在重量的 5-20% 的范围内，特别是约为重量的 10-20% 如重量的 12-18%。特别指出较后面的范围用于治疗龋炎和牙周炎的牙膏。然而，有兴趣提供有低含量的本发明多肽的牙膏，这类牙膏将适合于防止或预防性目的。对于这类目的，感兴趣的多肽含量范围可以从约为重量的 0.1 到约 5%。

一种特别类型的牙膏是基本上透明的凝胶。这类牙膏可完全不含磨光剂或可含如此细分的磨光剂形式以至凝胶看起来仍基本上清澈。此凝胶牙膏类型可独自被使用也可与含上面讨论的磨光剂的牙膏结合使用。

可以许多不同方式将本发明的一种牙膏制剂的多肽与其他牙齿或口腔制剂混合。通常，将优选形成本发明多肽的悬浮液并将该 amelin 悬浮液与其他制剂结合成膏的形式。另外的选择是，可将 amelin 粉与其他制剂成分混合（先与干制剂成分混合然后与液体制剂成分或半液体制剂成分混合）或将 amelin 本身混合进另外完成的制剂中。一般说，最好将 amelin 粉与磨光剂物质或牙粉一起加入。

在进行 amelin 或其他水不溶性或水微溶性多肽类似物的混合时，最好考虑多肽的物理和化学性质，在牙膏或牙粉或本文讨论的其他制剂方面的考虑将通常是十分简单并将通常在于加入 amelin 多肽到干燥的，溶解的或悬浮形式的制剂或其成分中。

局部施用可以是在或贴近于所谈及的呈现病变的身体局部的一种施用，例如在身体外部如口腔黏膜表面。此应用可以是简单涂抹该组合物，或可借助任何适于促进实现组合物与病理损害之间接触的器具。该组合物可被渗渍或散布于衬垫，膏药，纸片，纱布，海绵材料，cotton wool piece 等。可任意使用一种该组合物注入形式注入进或接近损害处。

根据本发明，局部用组合物可包含 1-80% 重量的活性化合物（基

于该制剂总重量) , 如 0.001-25%w/w 的活性化合物, 例如 0.1-10%, 0.50-5%, 或 2-5%。多于一种的活性成分化合物可被参合到该组合物中; 即含与其他药物化合物合在一起的 amelin 蛋白的组合物也在本发明范围内。按常规该组合物一天使用 1-10 次, 取决于损害的类型, 严重性和部位。

为局部应用, 可根据常规用药实践配制该组合物, 如在口腔局部用药通常使用药物可接受性赋形剂。任何具体组合物制剂中使用的载体性质将取决于想要施用该组合物的方法。可用于组合物的非水载体可包括固体或液体如润滑剂, 溶剂, 湿润剂, 增稠剂和粉剂。预期根据本发明的组合物可仅含有多肽, 非强制性地与水混合, 但该组合物也可含与载体, 稀释剂或黏合剂如纤维素聚合物, 琼脂, 藻酸盐或明胶混合的多肽, 它们对所谈及的目的是可接受的。就牙科应用而言, 载体或稀释剂为牙科学可接受的是便利的。目前优选使用含水溶性聚合物的载体。这类聚合物的非限制性实例是羧基纤维素钠, 微晶纤维素, 羟乙基纤维素, 羟基丙基纤维素, 甲基纤维素, 高分子聚丙烯酸, 藻酸钠, 丙二醇藻酸酯, 山吨树胶, 瓜耳树胶, 洋槐豆树胶, 改性淀粉, 明胶, 果胶或它们的结 10 合物。与活性蛋白质组分混合后, 这些水溶性聚合物可任意被转变成凝胶或膜, 产生由于其便利的物理性质而易于应用的组合物。该组合物可非强制性地含为促进储存稳定性目的的稳定剂或防腐剂。一种适当的赋 15 形剂是藻酸盐, 例如在 EP336967 中描述的。

为局部应用, 该组合物的 pH 可主要在十分宽的范围如 3-9。在本发明的优选实施方案中, 约 4 到 8 的 pH 是优选的。如上述的常规缓冲试剂可被用于获得需要的 pH。

本发明制剂还含有其他添加剂如稳定剂, 防腐剂, 增溶剂, 螯合剂, 25 凝胶形成剂, pH 调节剂, 抗氧化剂等。而且, 可便利的提供改性缓释制剂, 其中活性成分被混合进入聚合物基质, 或毫微颗粒, 或脂质体或微胞, 或吸附在离子交换树脂上, 或由聚合物所携带。

组合物可根据常规用药实践被设计并可以是:

半固体组合物: 凝胶, 膏, 混合物。

30 液体组合物: 溶液, 悬浮液, 渗麸液 (drench), 乳液。

如所指出的, 本发明药物组合物可包含本发明多肽本身或其功能衍生物, 或此类化合物的结合物。适当的功能衍生物实施例包括药用盐,

特别是那些适合于在口腔环境中使用的药用盐。实施例包括氨基功能的药用盐，例如产生阴离子的酸式盐，它们是药用可接受性的，特别是在口腔环境中。实施例包括磷酸盐，硫酸盐，硝酸盐，碘化物，溴化物，氯化物，硼酸盐以及从羧酸衍生的阴离子包括醋酸盐，苯甲酸盐，硬脂酸盐等。氨基功能的其他衍生物包括酰胺，酰亚胺，脲，氨基甲酸酯等。

其他适当的衍生物包括本发明多肽的羧基基团衍生物，包括盐，酯和酰胺。实施例包括带有药物可接受性阳离子的盐，例如锂，钠，钾，镁，钙，锌，铝，正铁，亚铁，铵和低级（ C_{1-6} ）-烷基铵盐。酯包括低级烷基酯。

10 将用数个工作实施例进一步描述本发明，它们不应被认为是对本应用范围的限制。

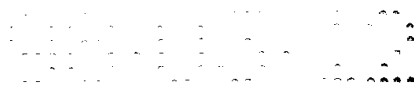
除非另有说明，使用常规方法和试剂盒。按照各个提供者所给的说明书使用试剂盒。本文未描述或提及的方法中的步骤和试剂被解释在：分子生物学现代方法，作者 **F.M. Ausubel, R. Brent, R.E. Kingston, D.D. Moore, J. G. Seidman, J. A. Smith and K. Struhl; John Wiley, New York(1994)**。所有文字引用在此明确引入作为参考。

图解

20 图 1：在生长中的第一磨牙中 RNA 序列的定位。切下四天龄大鼠上颌，将其固定并包埋在石蜡中。用通过 Bluescript 质粒体外转录制备的与 mRNA 互补的 DIG 标记 RNA 序列对整个磨牙远中（distal-mesial）的切片做原位杂交。图 1a：amelin，图 1b：牙釉蛋白，图 1c：I 型胶原。

25 图 2：amelin 1 和 2 序列。数个来自两个变异体的重叠序列被确定并排队。同一序列被相互对应印出，逗号表示缺少各变异体相对应的序列。最长的开放阅读框用单字符密码氨基酸名称简示。具有两个编码框的部分被划阴影（核苷酸 390-403）。下划线是被用于筛选含两个变异体克隆的寡聚物的互补序列（核苷酸 248-272 和 414-430）。方框表示与细胞表面蛋白相互作用的功能域之共有序列。推测的多腺苷酸化信号被双下划线（核苷酸 1892-1897）。

30 图 3：来自大鼠磨牙的 RNA Northern 印迹分析。从四天龄大鼠切下第一磨牙。分离 RNA，在琼脂糖凝胶中每泳道 4 毫克电泳该 RNA 并将其转移到尼龙膜。将各泳道与 amelin（a）和牙釉蛋白（b）DIG



标记的 RNA 探针杂交。kb 长度为已知的 RNA (Gibco BRL)的位置被示明于左边际。

实施例

实施例 1

5 RNA 的分离

用一个市售试剂盒 (Promega Biotech, RNAgents 总 RNA 分离系统) 在 500 升 4 摩尔的异硫氰酸胍盐, 80 毫摩尔 EDTA (Chomczynski&Sacchi,1987)中将从 4 天龄或 7 天龄 Aprague-Dawley 大鼠切下的三只生长中磨牙 (B&K Universal, Sollentuna, Sweden)在
10 玻璃-玻璃匀浆器中匀浆。然后用酚-氯仿提取并两次异丙醇沉淀。溶解 RNA 于 0.2xSET 缓冲液 (0.2%十二烷基硫酸钠, 4 毫摩尔 Tris-Cl pH7.5, 2 毫摩尔 EDTA) 中并通过光密度测量确定其浓度。

实施例 2

cDNA 文库的制备

15 借助于结合在硅酸盐树脂的寡-dT 筛选含聚腺苷酸的 RNA (mRNA) (Quiagen Oligotex mRNA Midi 试剂盒)。从聚腺苷酸末端引发反转录, 并将双链的甲基化的 cDNA 连接到 λ ZAP 载体臂上并包装进入噬菌体颗粒中 (Stratagen ZAP cDNA 克隆试剂盒)。扩增并铺平板后, 通过与总 DIG 标记的 cDNA (见下) 选出含常见表
20 达序列的噬菌体株。从阳性噬菌斑分离噬菌体并借助于 ExAssist 助噬菌体超感染 λ ZAP 感染的 Escherichia coli SOLR 细胞将分离的噬菌体转变成质粒。为获得 5' 端较好的再现, 还建立了 cDNA 文库并在随机位点引发 (Stratagen Random Unidirectional Linker-Primer)。用 Taq 聚合酶, 荧光终止子和半自动序列测试系统 (Applied Biosystems, Taq
25 DyeDeoxy Terminator Cycle Sequencing Kit)的循环测序对在基质形成细胞上产生阳性原位杂交信号的插入片段测序。用 Wisconsin 程序组 (Genetics Computer Group, Inc.)和 DNAid (Fredeic Dardel, fred@botrytis.polytechnique.fr)分析了序列。

实施例 3

30 筛选文库

七天龄大鼠的第一和第二磨牙的 cDNA 文库 (2×10^6 克隆) 噬菌体被铺平板, 并将噬菌体吸附到硝酸纤维素膜 (Schleicher and

Schull)。影印滤膜与 10 毫微克/毫升 cDNA 或胶原-和牙釉蛋白寡核苷酸杂交。杂交在 54 ℃ 进行 15 小时，并洗滤膜和显影 (Boehringer Mannheim, The DIG System)。含牙釉蛋白，胶原或其余常见表达序列的噬菌体被再克隆两次并通过用 ExAssist 助噬菌体 (Stratagen) 的重
5 复感染实现由体内剪切使其转变成 Bluescript 质粒。

实施例 4

杂交实验用探针的制备

用 0.25 mM 核苷酸浓度伴以补充到 0.1 mM 的毛地黄毒苷 (DIG)-dUTP (Boehringer Mannheim) 从有逆转录酶的富含聚腺苷
10 酸的 RNA (Promega Biotech, 逆转录系统) 产生供筛选文库用的 cDNA 探针。

在 DIG 修饰的 UTP (Boehringer Mannheim) 存在下，通过噬菌体 T7 或 T3RNA 聚合酶 (Promega RNA 探针 Gemini II 核心系统， Melton 等， 1994) 的体外转录合成与 mRNA 互补的 RNA 探针。含
15 amelin(1700bp) 的 DNA 模板是 Bluescript 质粒，它是通过体内剪切从 λ 噬菌体得到的。而且，通过限制酶切 Bluescript SK 质粒得到牙釉蛋白 (700bp) 和 I 型胶原 (850bp) 序列。用 [³⁵S] 代替 DIG 标记定量 RNA 测定用的探针。

胶原特异性寡核苷酸有 5 ' -CATGTAGGCAATGCTGTTCTT
20 GCAGTGGTAGGTGATGTTCTGGGAGGC-3 ' 序列 (Yamada 等， 1983) ， 而牙釉蛋白特异性寡核苷酸是 5 ' -
ATCCACTTCTTCCCGCTTGGTCTTGTCTGTCGCTGGCCAAGCT
TC-3 ' (Lau 等， 1992) 。通过 Boehringer 方法中的末端转移酶
反应用 DIG 修饰的 ddUTP 进行 3 ' 标记制备探针。

25 实施例 5

Northern 印迹

为 Northern 印迹分析，在 50% 甲酰胺存在下在使 2 厘米宽度的每
孔中的 15 毫克总 RNA 热变性并在有 2.2 摩尔甲醛， 0.02 摩尔 N-吗啉代
丙烷磺酸， 0.05 摩尔醋酸钠， 1 摩尔 EDTA 的琼脂糖凝胶中电泳
30 (Lehrach 等 1977) 。在 20xSSC (3 摩尔氯化钠， 0.3 摩尔柠檬酸钠) 中 RNA 被过夜转移到尼龙膜中 (Pall Biodyne B Transfer Membrane)。用紫外光使该膜交联并将其切割成条。加入 100 毫微克/

毫升 DIG 标记的 cRNA 探针后，各条在 50%甲酰氨， 5xSSC， 2%封闭试剂（ Boehringer Mannheim）， 0.1%N-月桂酰肌氨酸， 0.02%十二烷基硫酸钠（ SDS ）中 68 ℃ 预杂交 1 小时并随后在同样条件下杂交过夜。然后用 2xSSC， 0.1%SDS 在室温洗膜 5 分钟两次并用 0.1xSSC，
5 0.1%SDS 在 68 ℃ 洗 15 分钟两次。通过磷酸酶偶联的抗 DIG 抗体片段显现 DIG 标记的 RNA 的存在（ Boehringer Mannheim, The DIG System）。

实施例 6

溶液杂交

10 来自切下的磨牙的 RNA 与过量的 ³⁵S-UTP 标记互补 RNA 探针杂交（ Mathews 等， 1989 ）。 40 升 0.6 摩尔氯化钠， 4 摩尔 EDTA， 10 毫摩尔二硫苏糖醇（ DTT ）， 0.1%SDS, 30 毫摩尔 Tris-HCl, pH7.5 和 25%（ v/v ）甲酰氨的反应物含 20,000cpm 探针和不同量的总 RNA。用石蜡油覆盖该混合物，在 70 ℃ 温孵过夜，用 1 毫升 RNA 酶
15 溶液（ 40 克 RNA 酶 A， 2 克 RNA 酶 T1， Boehringer-Mannheim, 100 克鲑鱼色测试（ salmon tests）DNA， Sigma Chemical Co.）稀释并在 37 ℃ 酶解 1 小时。。用 100 升三氯醋酸（ 6 摩尔 ）沉淀 RNA 酶抗性双链 RNA，并在玻璃纤维滤器上收集（ Whatman GF/C）和在 Wallac 1409 流体闪烁计数器中分析之。杂该探针与已知浓度的体外合成 mRNA
20 序列杂交的标准曲线被用于将放射性测试-RNA 中杂交序列的量联系起来。

实施例 7

原位杂交

25 4 ℃ 下用 PBS（ 137 毫摩尔氯化钠， 2.7 毫摩尔氯化钾， 4.3 毫摩尔磷酸氢二钠， 1.4 毫摩尔磷酸二氢钾 ）中的多聚甲醛固定四天龄 Sprague Dawley 大鼠的上颌 24 小时，使其脱水并被包埋在石蜡中。7 微米厚的切片被固定在 vectabond-coated(Vector) 载玻片上。用二甲苯除去石蜡后，在 37 ℃ 用蛋白水解酶 K（ 20 微克/毫升 ）处理样品 30 分钟，用 4%甲酰氨后固定 5 分钟，用三乙醇胺和醋酸酐（ 2.66 毫升
30 三乙醇胺在 200 毫升水中； 0.5 毫升醋酸酐与该载玻片一起被加入其中 ）处理并渗没于 2xSSC 中， 42 ℃ 50%甲酰氨处理 60 分钟。用含 0.5 毫微克/微升 RNA 探针的 20 微升 0.3 摩尔氯化钠， 10 毫摩尔 Tris-HCl



釉蛋白异质性的影响。 *Biochem. Biophys. Res. Commun.* 188, 1253-1260.

-Leader, D. P. (1979). 信号肽氨基酸序列。 *Trends Biochem. Sci.* 4, 205-208.

5 -Leharach, H., Diamond, D., Wozney, J. M. & Boedtker, H. (1977). 通过变性条件下的凝胶电泳确定 RNA 分子量：一种关键的复查方法。 *Biochemistry* 16,4743-4351.

-Mathews, L.S., Enberg, B. & Norstedt, G. (1989). 大鼠生长激素受体基因表达的调节。 *J. Biol. Chem.* 264, 9905-9910.

10 -Matsuki, Y., Nakashima, M., Amizuka, N., Warshawsly, H., Goltzman, D., Yamada, K. M., and Yamada, Y.(1995). 从大鼠剪切物随机选择的 cDNA 克隆中部分序列的复杂情况。 *J. Dent. Res.* 74,307-312.

-Melton, D. A., Krieg, P. A., Rebagliati, M.R., Maniatis, T., Zinn, K. & Green, M. R. (1984). 从含有噬菌体 SP6 启动子质粒体外高效合成生物活性 RNA 和杂交探针 *Nucleic Acids Res.* 12, 7035-7056.

15 -Robinson, C., Kirkham, J. & Hallsworth, A. S. (1988). 发育牛牙齿中蛋白质，矿物质和水的大量分布与浓集。 *Archs. Oral Biol.* 33, 159-162.

-Simmer, J. P., Lau, E. C., Hu, C. C., Aoba, T., Lacey, M., Nelson, D., Zeichner-David, M., Snead, M. L., Slackin, H. C. & Fincham, A. G.(1994). 分离和表征 *Escherichia coli* 中表达的小鼠牙釉蛋白。 *Calcif. Tissue Int.* 54, 312-319.

20 -Ataatz, W. D., Fok, K. F., Zutter, M. M., Adame, S. P., Rodriguez, B. A. & Santoro, S. A.(1991). 胶原中 $\alpha_2(\beta)$ 整联蛋白四肽识别序列的鉴定。 *J. Biol. Chem.* 266, 7363-7367.

25 -Strawich, E. & Glimcher, M. J.(1990). 牙“enamelin”主要被鉴定为血清蛋白质。 *Eur. J. Biochem.* 191,47-56.

-Termine, J. D., Belcourt, A. B., Christner, P. J., Conn, K. M. & Nylen, M. U. (1980).分开提取的小牛牙基质蛋白质的性质。 *J. Biol. Chem.* 255, 9760-9768.

30 -Uchida, T., Fukae, M., Tanabe, T., Uamakoshi, Y., Satoda, T., Murakami, C., Takahashi, O. & Shimizu, M. (1995). 猪的未成熟釉质



中 15kDa 非牙釉蛋白和相关蛋白质的免疫化学及免疫细胞化学研究：提出一组新 enamelin 质“鞘蛋白”。 **Biomed. Res.** 16,131-140.

-Wilkinson, D. L. & Harrison, R. G.(1991). 预断 *Escherichia coli* 中重组蛋白质的溶解度。 **Biochemistry** 9, 443-448.

5 -Watkins, S. (1994). 原位杂交与免疫组化。分子生物学现代方法。
**Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G.,
Smith, J. A. & Struhl, K. John Wiley, New York**

-Yamada, Y. & Kleinman, H. K. (1992). 细胞粘着分子的功能域。
Curr. Opin. Cell Biol. 4, 819-823.

10 -Yamada, Y., Kuhn, K. & deCrombrughe, B. (1983). 编码 C-前肽的一个片段的保守核苷酸序列被发现在不同胶原的相同位置。 **Nucl. Acids Res.** 11, 2733-2744.

-WO 89/08441(Biora AB; 1989 年 9 月 21 日出版)

序列表

15 (1) 一般资料：

(i) 申请人：

(A) 名称：口腔生物学中心

(B) 街： P.O. Box 4064

(C) 城市： Huddinge

20 (E) 国家：瑞典

(F) 邮编 (ZIP) : S-141 04

(ii) 发明题目：新 DNA 和肽序列及相关表达载体

(iii) 序列号： 4

(iv) 计算机可读形式：

25 (A) 介质类型：软盘

(B) 计算机： IBM 兼容 PC

(C) 操作系统： PC-DOS/MS-DOS

(D) 软件：专利发布号 1.0,版本号： 1.30 (EPO)

(2) SEQ ID NO : 1 资料

30 (i) 序列特性：

(A) 长度： 1939 碱基对

(B) 类型：核酸



(C) 链：双链

(D) 拓扑结构：线性

(ii) 分子类型：cDNA

(ix) 特征：

5 (A) 名称/关键字：CDS

(B) 位置：94..1314

(xi) 序列描述：SEQ ID NO : 1 :

```
AGAGAGAGAG CCCAGGAAC AGTCCAGAAA AAAATTAATC TTCTTTTCTT AGAACTGTTT      60
TGATTGGCAT CATCAGGCCT GGGAGCACAG TGA ATG TCA GCA TCT AAG ATT CCA      114
                               Met Ser Ala Ser Lys Ile Pro
                               1                               5
CTT TTC AAA ATG AAG GGC CTG CTC CTG TTC CTG TCC CTA GTG AAA ATG      162
Leu Phe Lys Met Lys Gly Leu Leu Leu Phe Leu Ser Leu Val Lys Met
                               10                               15                               20
AGC CTC GCC GTG CCG GCA TTT CCT CAA CAA CCT GGG GCT CAA GGC ATG      210
Ser Leu Ala Val Pro Ala Phe Pro Gln Gln Pro Gly Ala Gln Gly Met
                               25                               30                               35
```

GCA CCT CCT GGC ATG GCT AGT TTG AGC CTT GAG ACA ATG AGA CAG TTG	258
Ala Pro Pro Gly Met Ala Ser Leu Ser Leu Glu Thr Met Arg Gln Leu	
40 45 50 55	
GGA AGC TTG CAG GGG CTC AAC GCA CTT TCT CAG TAT TCT AGA CTT GGC	306
Gly Ser Leu Gln Gly Leu Asn Ala Leu Ser Gln Tyr Ser Arg Leu Gly	
60 65 70	
TTT GGA AAA GCA CTT AAT AGT TTA TGG TTG CAT GGA CTC CTC CCA CCG	354
Phe Gly Lys Ala Leu Asn Ser Leu Trp Leu His Gly Leu Leu Pro Pro	
75 80 85	
CAT AAT TCT TTC CCA TGG ATA GGA CCA AGG GAA CAT GAA ACC CAA CAG	402
His Asn Ser Phe Pro Trp Ile Gly Pro Arg Glu His Glu Thr Gln Gln	
90 95 100	
CCA TCC TTG CAG CCT CAC CAG CCA GGA CTG AAA CCC TTC CTC CAG CCC	450
Pro Ser Leu Gln Pro His Gln Pro Gly Leu Lys Pro Phe Leu Gln Pro	
105 110 115	
ACT GCT GCA ACC GGT GTC CAG GTC ACA CCC CAG AAG CCA GGG CCT CAT	498
Thr Ala Ala Thr Gly Val Gln Val Thr Pro Gln Lys Pro Gly Pro His	
120 125 130 135	
CCT CCA ATG CAC CCT GGA CAG CTG CCC TTG CAG GAA GGA GAG CTG ATA	546
Pro Pro Met His Pro Gly Gln Leu Pro Leu Gln Glu Gly Glu Leu Ile	
140 145 150	
GCA CCA GAT GAG CCA CAG GTG GCG CCA TCA GAG AAC CCA CCA ACA CCC	594
Ala Pro Asp Glu Pro Gln Val Ala Pro Ser Glu Asn Pro Pro Thr Pro	
155 160 165	
GAG GTA CCA ATA ATG GAT TTT GCC GAT CCA CAA TTC CCA ACA GTG TTC	642
Glu Val Pro Ile Met Asp Phe Ala Asp Pro Gln Phe Pro Thr Val Phe	
170 175 180	
CAG ATC GCC CAT TCG CTG TCT CGG GGA CCA ATG GCA CAC AAC AAA GTA	690
Gln Ile Ala His Ser Leu Ser Arg Gly Pro Met Ala His Asn Lys Val	
185 190 195	
CCC ACT TTT TAC CCA GGA ATG TTT TAC ATG TCT TAT GGA GCA AAC CAA	738
Pro Thr Phe Tyr Pro Gly Met Phe Tyr Met Ser Tyr Gly Ala Asn Gln	
200 205 210 215	
TTG AAT GCT CCT GGC AGA ATC GGC TTC ATG AGT TCA GAA GAA ATG CCT	786
Leu Asn Ala Pro Gly Arg Ile Gly Phe Met Ser Ser Glu Glu Met Pro	
220 225 230	
GGA GAA AGA GGA AGT CCC ATG GCC TAC GGA ACT CTG TTC CCA GGA TAT	834
Gly Glu Arg Gly Ser Pro Met Ala Tyr Gly Thr Leu Phe Pro Gly Tyr	
235 240 245	
GGA GGC TTC AGG CAA ACC CTT AGG GGA CTG AAT CAG AAT TCA CCC AAG	882
Gly Gly Phe Arg Gln Thr Leu Arg Gly Leu Asn Gln Asn Ser Pro Lys	
250 255 260	

GGA GGA GAC TTT ACT GTG GAA GTA GAT TCT CCA GTG TCT GTA ACT AAA	930
Gly Gly Asp Phe Thr Val Glu Val Asp Ser Pro Val Ser Val Thr Lys	
265 270 275	
GGC CCT GAG AAA GGA GAG GGT CCA GAA GGC TCT CCA CTG CAA GAG GCC	978
Gly Pro Glu Lys Gly Glu Gly Pro Glu Gly Ser Pro Leu Gln Glu Ala	
280 285 290 295	
AGC CCA GAC AAG GGC GAA AAC CCG GCT CTC CTT TCA CAG ATT GCC CCC	1026
Ser Pro Asp Lys Gly Glu Asn Pro Ala Leu Leu Ser Gln Ile Ala Pro	
300 305 310	
GGG GCC CAT GCA GGA CTT CTT GCT TTC CCC AAT GAC CAC ATC CCC AAC	1074
Gly Ala His Ala Gly Leu Leu Ala Phe Pro Asn Asp His Ile Pro Asn	
315 320 325	
ATG GCA AGG GGT CCT GCA GGG CAA AGA CTC CTC GGA GTC ACC CCT GCA	1122
Met Ala Arg Gly Pro Ala Gly Gln Arg Leu Leu Gly Val Thr Pro Ala	
330 335 340	
CTC GCA GAC CCA CTG ATC ACC CCT GAA TTA GCA GAA GTT TAT GAA ACC	1170
Ala Ala Asp Pro Leu Ile Thr Pro Glu Leu Ala Glu Val Tyr Glu Thr	
345 350 355	
TAT GGT GCT GAT GTT ACC ACA CCC TTG GGG GAT GGA GAA GCA ACC ATG	1218
Tyr Gly Ala Asp Val Thr Thr Pro Leu Gly Asp Gly Glu Ala Thr Met	
360 365 370 375	
GAT ATC ACC ATG TCC CCA GAC ACT CAG CAG CCA CCG ATG CCT GGA AAC	1266
Asp Ile Thr Met Ser Pro Asp Thr Gln Gln Pro Pro Met Pro Gly Asn	
380 385 390	
AAA GTG CAC CAG CCC CAG GTG CAC AAT GCA TGG CGT TTC CAA GAG CCC	1314
Lys Val His Gln Pro Gln Val His Asn Ala Trp Arg Phe Gln Glu Pro	
395 400 405	
TGACAACCTT GACATAGCAG CTACTTCATG TATGCACAAG CTTTTCAGCT TTGACCCCAT	1374
AGCGTACCTT ATTGCTAAAA CACTTGCTAC CCTTCCACAG CGAAGGTATT AAGAGCACTA	1434
AGCATGTATT AATAAATACA AGTGCCTAGA AATAGTGTAG GTCCCTTCTT GCTTCCATTC	1494
TTATCGAAAT AAAACATATC AACTGTCTCC GTGACTTAGA AATACTATCG ATGATGTCAG	1554
AGCAAGTCTG AGTGTGAGCA CTTGGTGATC TAGCATGTAG CTGTCTTAGG CATCATAAAA	1614
TTCCTCTTAC TACATGACAT TATTATGCCC AGGAAATGTG ACACCGCTTC TTTCTCTACG	1674
CAAAGCACT TAGTTTTAGA ATTCCAAAGT ATTTCAATTA AACCGTATTA AATGGTGATT	1734
GGTGGAGAAT CCTGACTGCT ATTACTGGGT ATCATATATT GGATTTAAAA TTCTTATTTA	1794
TAGAATATTT TATTTAATCT AGGAAAAGAA AAGGCAATTG GCCTGTTTTA AATAAAGAAT	1854
TTTTCTCACT GAAAATGTCA GGAATTGTAT GCTTATTATT TATATGTATT TAAATAGTAA	1914
AGAAAAGCAT ACTCAAAAAA AAAAA	1939



(2) SEQ ID NO : 2 资料:

(i) 序列特性

(A) 长度: 407 氨基酸

(B) 类型: 氨基酸

(D) 拓扑结构: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO : 2 :

Met Ser Ala Ser Lys Ile Pro Leu Phe Lys Met Lys Gly Leu Leu Leu
 1 5 10 15

Phe Leu Ser Leu Val Lys Met Ser Leu Ala Val Pro Ala Phe Pro Gln
 20 25 30

Gln Pro Gly Ala Gln Gly Met Ala Pro Pro Gly Met Ala Ser Leu Ser
 35 40 45

Leu Glu Thr Met Arg Gln Leu Gly Ser Leu Gln Gly Leu Asn Ala Leu
 50 55 60

Ser Gln Tyr Ser Arg Leu Gly Phe Gly Lys Ala Leu Asn Ser Leu Trp
 65 70 75 80

Leu His Gly Leu Leu Pro Pro His Asn Ser Phe Pro Trp Ile Gly Pro
 85 90 95

Arg Glu His Glu Thr Gln Gln Pro Ser Leu Gln Pro His Gln Pro Gly
 100 105 110

Leu Lys Pro Phe Leu Gln Pro Thr Ala Ala Thr Gly Val Gln Val Thr
 115 120 125

Pro Gln Lys Pro Gly Pro His Pro Pro Met His Pro Gly Gln Leu Pro
 130 135 140

Leu Gln Glu Gly Glu Leu Ile Ala Pro Asp Glu Pro Gln Val Ala Pro
 145 150 155 160

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

Pro Gln Phe Pro Thr Val Phe Gln Ile Ala His Ser Leu Ser Arg Gly
 180 185 190

Pro Met Ala His Asn Lys Val Pro Thr Phe Tyr Pro Gly Met Phe Tyr
 195 200 205

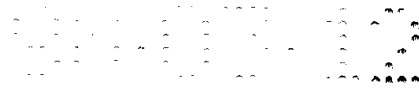
Met Ser Tyr Gly Ala Asn Gln Leu Asn Ala Pro Gly Arg Ile Gly Phe
 210 215 220

Met Ser Ser Glu Glu Met Pro Gly Glu Arg Gly Ser Pro Met Ala Tyr
 225 230 235 240

Gly Thr Leu Phe Pro Gly Tyr Gly Gly Phe Arg Gln Thr Leu Arg Gly
 245 250 255

5

10



(2) SEQ ID NO : 3 资料

(i) 序列特性:

(A) 长度: 1648 碱基对

(B) 类型: 核酸

5

(C) 链: 双链

(D) 拓扑结构: 线性

(ii) 分子类型: cDNA

(ix) 特征:

(A) 名称/关键字: CDS

10

(B) 位置: 52..1023

(xi) 序列描述: SEQ ID NO : 3 :

```
GAGAGAGAGA GCCACCGCAT AATTCTTTCC CATGGATAGG ACCAAGGGAA C ATG AAA      57
                                                    Met Lys
                                                    1
```

```
CCC AAC AGT ATG GAA AAT TCT TTG CCT GTG CAT CCC CCA CCT CTC CCA      105
Pro Asn Ser Met Glu Asn Ser Leu Pro Val His Pro Pro Pro Leu Pro
      5                10                15
```

TCA	CAG	CCA	TCC	TTG	CAG	CCT	CAC	CAG	CCA	GGA	CTG	AAA	CCC	TTC	CTC	153
Ser	Gln	Pro	Ser	Leu	Gln	Pro	His	Gln	Pro	Gly	Leu	Lys	Pro	Phe	Leu	
	20					25					30					
CAG	CCC	ACT	GCT	GCA	ACC	GGT	GTC	CAG	GTC	ACA	CCC	CAG	AAG	CCA	GGG	201
Gln	Pro	Thr	Ala	Ala	Thr	Gly	Val	Gln	Val	Thr	Pro	Gln	Lys	Pro	Gly	
	35				40					45					50	
CCT	CAT	CCT	CCA	ATG	CAC	CCT	GGA	CAG	CTG	CCC	TTG	CAG	GAA	GGA	GAG	249
Pro	His	Pro	Pro	Met	His	Pro	Gly	Gln	Leu	Pro	Leu	Gln	Glu	Gly	Glu	
				55					60					65		
CTG	ATA	GCA	CCA	GAT	GAG	CCA	CAG	GTG	GCG	CCA	TCA	GAG	AAC	CCA	CCA	297
Leu	Ile	Ala	Pro	Asp	Glu	Pro	Gln	Val	Ala	Pro	Ser	Glu	Asn	Pro	Pro	
			70					75					80			
ACA	CCC	GAG	GTA	CCA	ATA	ATG	GAT	TTT	GCC	GAT	CCA	CAA	TTC	CCA	ACA	345
Thr	Pro	Glu	Val	Pro	Ile	Met	Asp	Phe	Ala	Asp	Pro	Gln	Phe	Pro	Thr	
		85					90					95				
GTG	TTC	CAG	ATC	GCC	CAT	TCG	CTG	TCT	CGG	GGA	CCA	ATG	GCA	CAC	AAC	393
Val	Phe	Gln	Ile	Ala	His	Ser	Leu	Ser	Arg	Gly	Pro	Met	Ala	His	Asn	
	100					105					110					
AAA	GTA	CCC	ACT	TTT	TAC	CCA	GGA	ATG	TTT	TAC	ATG	TCT	TAT	GGA	GCA	441
Lys	Val	Pro	Thr	Phe	Tyr	Pro	Gly	Met	Phe	Tyr	Met	Ser	Tyr	Gly	Ala	
	115				120					125				130		
AAC	CAA	TTG	AAT	GCT	CCT	GGC	AGA	ATC	GGC	TTC	ATG	AGT	TCA	GAA	GAA	489
Asn	Gln	Leu	Asn	Ala	Pro	Gly	Arg	Ile	Gly	Phe	Met	Ser	Ser	Glu	Glu	
				135					140					145		
ATG	CCT	GGA	GAA	AGA	GGA	AGT	CCC	ATG	GCC	TAC	GGA	ACT	CTG	TTC	CCA	537
Met	Pro	Gly	Glu	Arg	Gly	Ser	Pro	Met	Ala	Tyr	Gly	Thr	Leu	Phe	Pro	
			150					155					160			
GGA	TAT	GGA	GGC	TTC	AGG	CAA	ACC	CTT	AGG	GGA	CTG	AAT	CAG	AAT	TCA	585
Gly	Tyr	Gly	Gly	Phe	Arg	Gln	Thr	Leu	Arg	Gly	Leu	Asn	Gln	Asn	Ser	
		165					170					175				
CCC	AAG	GGA	GGA	GAC	TTT	ACT	GTG	GAA	GTA	GAT	TCT	CCA	GTG	TCT	GTA	633
Pro	Lys	Gly	Gly	Asp	Phe	Thr	Val	Glu	Val	Asp	Ser	Pro	Val	Ser	Val	
	180					185					190					
ACT	AAA	GGC	CCT	GAG	AAA	GGA	GAG	GGT	CCA	GAA	GGC	TCT	CCA	CTG	CAA	681
Thr	Lys	Gly	Pro	Glu	Lys	Gly	Glu	Gly	Pro	Glu	Gly	Ser	Pro	Leu	Gln	
	195				200					205				210		
GAG	GCC	AGC	CCA	GAC	AAG	GGC	GAA	AAC	CCG	GCT	CTC	CTT	TCA	CAG	ATT	729
Glu	Ala	Ser	Pro	Asp	Lys	Gly	Glu	Asn	Pro	Ala	Leu	Leu	Ser	Gln	Ile	
				215					220					225		
GCC	CCC	GGG	GCC	CAT	GCA	GGA	CTT	CTT	GCT	TTC	CCC	AAT	GAC	CAC	ATC	777
Ala	Pro	Gly	Ala	His	Ala	Gly	Leu	Leu	Ala	Phe	Pro	Asn	Asp	His	Ile	
			230					235					240			



CCC AAC ATG GCA AGG GGT CCT GCA GGG CAA AGA CTC CTC GGA GTC ACC	825
Pro Asn Met Ala Arg Gly Pro Ala Gly Gln Arg Leu Leu Gly Val Thr	
245 250 255	
CCT GCA GCT GCA GAC CCA CTG ATC ACC CCT GAA TTA GCA GAA GTT TAT	873
Pro Ala Ala Ala Asp Pro Leu Ile Thr Pro Glu Leu Ala Glu Val Tyr	
260 265 270	
GAA ACC TAT GGT GCT GAT GTT ACC ACA CCC TTG GGG GAT GGA GAA GCA	921
Glu Thr Tyr Gly Ala Asp Val Thr Thr Pro Leu Gly Asp Gly Glu Ala	
275 280 285 290	
ACC ATG GAT ATC ACC ATG TCC CCA GAC ACT CAG CAG CCA CCG ATG CCT	969
Thr Met Asp Ile Thr Met Ser Pro Asp Thr Gln Gln Pro Pro Met Pro	
295 300 305	
GGA AAC AAA GTG CAC CAG CCC CAG GTG CAC AAT GCA TGG CGT TTC CAA	1017
Gly Asn Lys Val His Gln Pro Gln Val His Asn Ala Trp Arg Phe Gln	
310 315 320	
CAG CCC TGACAACCTT GACATAGCAG CTACTTCATG TATGCACAAG CTTTTCAGCT	1073
Glu Pro	
TTGACCCCAT AGCGTACCTT ATTGCTAAAA CACTTGCTAC CCTTCCACAG CGAAGGTATT	1133
AAGAGCACTA AGCATGTATT AATAAATACA AGTGCCTAGA AATAGTGTAG GTCCCTTCTT	1193
GCTTCCATTC TTATCGAAAT AAAACATATC AACTGTCTCC GTGACTTAGA AATACTATCG	1253
ATGATGTCAG AGCAAGTCTG AGTGTACGCA CTTGGTGATC TAGCATGTAG CTGTCTTAGG	1313
CATCATAAAA TTCCTCTTAC TACATGACAT TATTATGCCC AGGAAATGTG ACACCGCTTC	1373
TTTCTCTACG CAAAAGCACT TAGTTTCAGA ATTCCAAAGT ATTTTATTTA AACCGTATTA	1433
AATGGTGATT GGTGGAGAAT CCTGACTGCT ATTACTGGGT ATCATATATT GGATTTAAAA	1493
TTCTTATTTA TAGAATATTT TATTTAATCT AGGAAAAGAA AAGGCAATTG GCCTGTTTTA	1553
AATAAAGAAT TTTTCTCACT GAAAATGTCA GGAATTGTAT GCTTATTATT TATATGTATT	1613
TAAATAGTAA AGAAAAGCAT ACTCAAAAAA AAAAA	1648



Leu Pro Ser Gln Pro Ser Leu Gln Pro His Gln Pro Gly Leu Lys Pro
20 25 30

Phe Leu Gln Pro Thr Ala Ala Thr Gly Val Gln Val Thr Pro Gln Lys
35 40 45

Pro Gly Pro His Pro Pro Met His Pro Gly Gln Leu Pro Leu Gln Glu
50 55 60

Gly Glu Leu Ile Ala Pro Asp Glu Pro Gln Val Ala Pro Ser Glu Asn
65 70 75 80

Pro Pro Thr Pro Glu Val Pro Ile Met Asp Phe Ala Asp Pro Gln Phe
85 90 95

Pro Thr Val Phe Gln Ile Ala His Ser Leu Ser Arg Gly Pro Met Ala
100 105 110

His Asn Lys Val Pro Thr Phe Tyr Pro Gly Met Phe Tyr Met Ser Tyr
115 120 125

Gly Ala Asn Gln Leu Asn Ala Pro Gly Arg Ile Gly Phe Met Ser Ser
130 135 140

Glu Glu Met Pro Gly Glu Arg Gly Ser Pro Met Ala Tyr Gly Thr Leu
145 150 155 160

Phe Pro Gly Tyr Gly Gly Phe Arg Gln Thr Leu Arg Gly Leu Asn Gln
165 170 175

Asn Ser Pro Lys Gly Gly Asp Phe Thr Val Glu Val Asp Ser Pro Val
180 185 190

Ser Val Thr Lys Gly Pro Glu Lys Gly Glu Gly Pro Glu Gly Ser Pro
195 200 205

Leu Gln Glu Ala Ser Pro Asp Lys Gly Glu Asn Pro Ala Leu Leu Ser
210 215 220

Gln Ile Ala Pro Gly Ala His Ala Gly Leu Leu Ala Phe Pro Asn Asp
225 230 235 240

His Ile Pro Asn Met Ala Arg Gly Pro Ala Gly Gln Arg Leu Leu Gly
245 250 255

Val Thr Pro Ala Ala Ala Asp Pro Leu Ile Thr Pro Glu Leu Ala Glu
260 265 270

Val Tyr Glu Thr Tyr Gly Ala Asp Val Thr Thr Pro Leu Gly Asp Gly
275 280 285

Glu Ala Thr Met Asp Ile Thr Met Ser Pro Asp Thr Gln Gln Pro Pro
290 295 300

Met Pro Gly Asn Lys Val His Gln Pro Gln Val His Asn Ala Trp Arg
305 310 315 320

Phe Gln Glu Pro



说明书附图

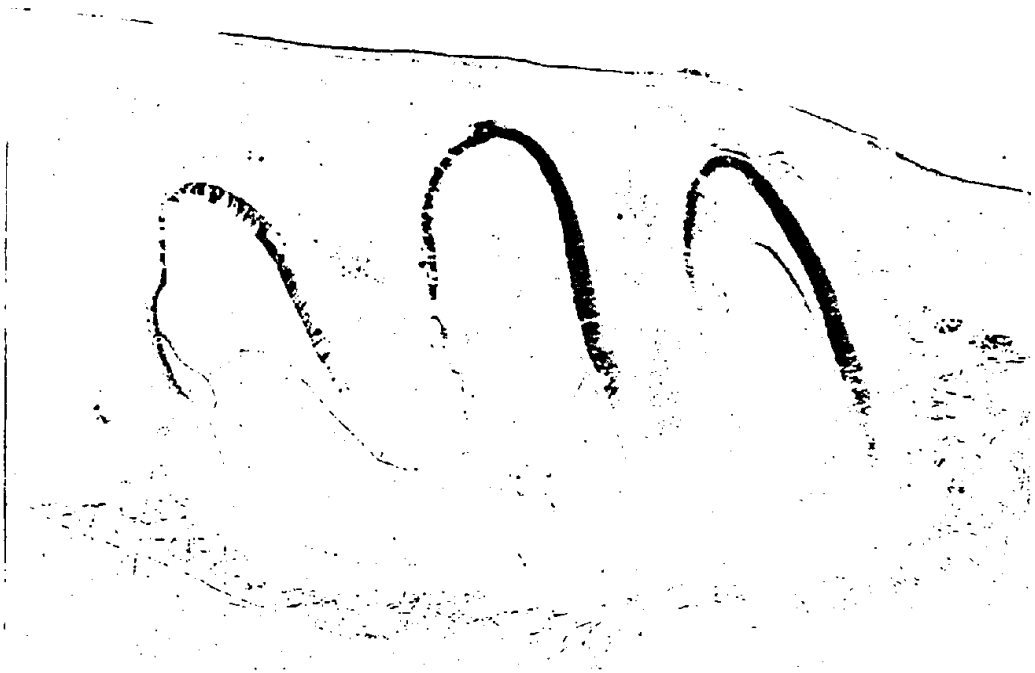


图 1A



图 1 B



图 1 C

1 M S
 amelin1 GAGAGAGAGA GCCCCAGGAA CAGTCCAGAA AAAATTAAT CTCTTTTCT TAGACTGTT TTGATTGGCA TCATCAGGCC TGGGAGCACA GTGAATGTCA 100
 amelin2 GAGAGAGAGA G.....

 amelin1 A S K I P L F K M K G L L L F L S L V K M S L A V P A F P Q Q P G
 amelin2 GCATCTAAGA TTCCACTTTT CAAATGAAG GGCCCTCTCC TGTTCTGTG CCTAGTGAAA ATGAGCCTCG CCGTGCCGGC ATTTCTCTCA CAACCTGGGG 200

 amelin1 A Q G M A P P G M A S L S L E T M R Q L G S L Q G L N A L S Q Y S R
 amelin2 CTCAAGGCAT GGCACCTCCT GGCNTGGCTA GTTTGAGCCT TGAGACAATG AGACAGTGG GAGCTTGGCA GGGGCTCAAC GCACTTTCTC AGTATTCTAG 300

 amelin1 L G F G K A L N S L W L H G L L P P H N S F P W I G P R E H E T Q
 amelin2 ACTTGGCTTT GGAARAGCAC TTANTAGTTT ATGGTTGCAT GGACTCCTCC CACCGCATAA TTCTTTCCCA TGGATAGGAC CAAGGAAAR TGAAGCCCA 400
C CACCGCATAA TTCTTTCCCA TGGATAGGAC CAAGGAAAR TGAAGCCCA
 M K P N

 amelin1 Q CAG.....
 amelin2 CAGTATGGAA ANTCTTTGC CTGTGCATCC CCCACCTCTC CCATCACAGC CNTCCTTCCA GCCTCACCCAG CCAGGACTGA AACCTTCTCT CCAGCCCACT 500
 S M E N S L P V H P P P L P S Q P S L Q P H Q P G L K P F L Q P T

 amelin1 A A T G V Q V T P Q K P G P H P P M H P G Q L P L Q E G E L I A P
 amelin2 GCTGCAACCG GTGTCCAGGT CACACCCAG ANGCCAGGGC CTCATCCTCC ANTGCACCT GGACAGCTGC CTTTGCAGGA AGGAGACTG ATAGCACCAG 600
 GCTGCAACCG GTGTCCAGGT CACACCCAG ANGCCAGGGC CTCATCCTCC ANTGCACCT GGACAGCTGC CTTTGCAGGA AGGAGACTG ATAGCACCAG
 A A T G V Q V T P Q K P G P H P P M H P G Q L P L Q E G E L I A P

 amelin1 D E P Q V A P S E N P P T P E V P I M D F A D P Q F P T V F Q I A H
 amelin2 ATGAGCCACA GTGGGGCCCA TCAGAGAACC CACCAACACC CGAGGTACCA ATATGGATT TTGGCGATCC ACAATTTCCA ACAGTGTTC AGATCGCCCA 700
 ATGAGCCACA GTGGGGCCCA TCAGAGAACC CACCAACACC CGAGGTACCA ATATGGATT TTGGCGATCC ACAATTTCCA ACAGTGTTC AGATCGCCCA
 D E P Q V A P S E N P P T P E V P I M D F A D P Q F P T V F Q I A H



amelin1 S : S R G P M A H N K V P T F Y P G M F Y M S Y G A N Q L N A P A
 TCGCTGCT CGGGACCAA CAAAGTACC ACTTTTACC CAGGATGTT TTACATGCTT TATGGAGCAA ACCAATTGAA TGCTCCTGCC 800
amelin2 S : S R G P M A H N K V P T F Y P G M F Y M S Y G A N Q L N A P A
 TCGCTGCT CGGGACCAA CAAAGTACC ACTTTTACC CAGGATGTT TTACATGCTT TATGGAGCAA ACCAATTGAA TGCTCCTGCC

amelin1 R : G F M S S E E M P G E R G S P M A Y G T L F P G Y G G F R Q T
 AGAATCGCT TCATGAGTTC AGAAGAATG CCTGGAGAAA GAGGAAGTCC CATGGCCTAC GGAATCTGT TCCAGGATA TGGAGGCTC AGGCAACCC 900
amelin2 R : G F M S S E E M P G E R G S P M A Y G T L F P G Y G G F R Q T
 AGAATCGCT TCATGAGTTC AGAAGAATG CCTGGAGAAA GAGGAAGTCC CATGGCCTAC GGAATCTGT TCCAGGATA TGGAGGCTC AGGCAACCC

amelin1 L : R G L N Q N S P K G G D F T V E V D S P V S V T K G P E K G E G P
 TAGGGGCT GAATCAGAAT TCACCCAAGG GAGGAGACTT TACTGTGAA GTAGATTCTC CAGTCTGT AACTAAAGGC CCTGAGAAAAG GAGAGGTTCC 1000
amelin2 L : R G L N Q N S P K G G D F T V E V D S P V S V T K G P E K G E G P
 TAGGGGCT GAATCAGAAT TCACCCAAGG GAGGAGACTT TACTGTGAA GTAGATTCTC CAGTCTGT AACTAAAGGC CCTGAGAAAAG GAGAGGTTCC

amelin1 E S S P L Q E A S P D K G E N P A L L S Q I A P G A H A G L L A F
 AGAAGGCTT CCACTGCAAG AGCCAGGCC AACACAGGCC GAACACCCGG CTCTCCTTC ACAGATTGCC CCGGGGCC ATGCAGACT TCTTGCTTTC 1100
amelin2 E S S P L Q E A S P D K G E N P A L L S Q I A P G A H A G L L A F
 AGAAGGCTT CCACTGCAAG AGCCAGGCC AACACAGGCC GAACACCCGG CTCTCCTTC ACAGATTGCC CCGGGGCC ATGCAGACT TCTTGCTTTC

amelin1 P N D H I P N M A R G P A G Q R L L G V T P A A A D P L I T P E L
 CCAATGACC ACATCCCAA CATGGCAAG GGTCTGCGAG GGCRAAGACT CCTCGGAGT ACCCTGCGAG CTGCAGACC ACTGATCACC CCTGAATTAG 1200
amelin2 P N D H I P N M A R G P A G Q R L L G V T P A A A D P L I T P E L
 CCAATGACC ACATCCCAA CATGGCAAG GGTCTGCGAG GGCRAAGACT CCTCGGAGT ACCCTGCGAG CTGCAGACC ACTGATCACC CCTGAATTAG

amelin1 A E V Y E T Y G A D V T T P L G D G E A T M D I T M S P D T Q Q P P
 CAGAAGTTA TGAACCTAT GTGCTGATG TTACCACACC CTTGGGAT GGAGAGCAA CCATGGATAT CACCATGTCC CCAGACTC AGCAGCCACC 1300
amelin2 A E V Y E T Y G A D V T T P L G D G E A T M D I T M S P D T Q Q P P
 CAGAAGTTA TGAACCTAT GTGCTGATG TTACCACACC CTTGGGAT GGAGAGCAA CCATGGATAT CACCATGTCC CCAGACTC AGCAGCCACC

amelin1 Y P G N K V H Q P Q V H N A W R F Q E P
 GATGCCCTGA AACAAAGTGC ACCAGCCCCA GGTGCACAAT GCATGGCGTT TCCAAGAGCC CTGACAACTT TGACATAGCA GCTACTTCAT GTATGCACAA 1400
amelin2 Y P G N K V H Q P Q V H N A W R F Q E P
 GATGCCCTGA AACAAAGTGC ACCAGCCCCA GGTGCACAAT GCATGGCGTT TCCAAGAGCC CTGACAACTT TGACATAGCA GCTACTTCAT GTATGCACAA



图 2 (续)

起始 - a b

4.40 -

2.40 -

1.40 -

0.24 -



图 3