

[19]中华人民共和国国家知识产权局

[51]Int. Cl⁷

C07K 14/435

[12] 发明专利申请公开说明书

C07K 16/18 C07H 21/00

C12N 15/10 C12N 15/12

C12N 15/63 C12N 15/64

C12P 21/00 G01N 33/68

C12Q 1/68 A61K 38/17

A61K 31/7088 A61K 39/395

[21] 申请号 99124046.4

[11] 公开号 CN 1296958A

[43] 公开日 2001 年 5 月 30 日

[22] 申请日 1999.11.22 [21] 申请号 99124046.4

[71] 申请人 上海博容基因开发有限公司

地址 200092 上海市中山北二路 1111 号 3 号楼 12 层

[72] 发明人 毛裕民 谢毅

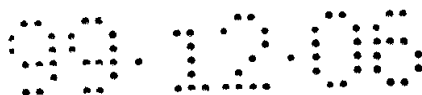
权利要求书 2 页 说明书 29 页 附图页数 2 页

[54] 发明名称 一种新的多肽——人 G 蛋白 44 和编码这种多肽的多核苷酸

[57] 摘要

本发明公开了一种新的多肽——人 G 蛋白 44, 编码此多肽的多核苷酸和经 DNA 重组技术产生这种多肽的方法。本发明还公开了此多肽用于治疗多种疾病的方法, 如恶性肿瘤, 血液病, HIV 感染和免疫性疾病和各类炎症等。本发明还公开了抗此多肽的拮抗剂及其治疗作用。本发明还公开了编码这种新的人 G 蛋白 44 的多核苷酸的用途。

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



权 利 要 求 书

- 1、一种分离的多肽-人G蛋白44,其特征在于它包含有:SEQ ID NO:2所示的氨基酸序列的多肽、或其多肽的活性片段、类似物或衍生物。
- 2、如权利要求1所述的多肽,其特征在于所述多肽、类似物或衍生物的氨基酸序列具有与SEQ ID NO:2所示的氨基酸序列至少95%的相同性。
- 3、如权利要求2所述的多肽,其特征在于它包含具有SEQ ID NO:2所示的氨基酸序列的多肽。
- 4、一种分离的多核苷酸,其特征在于所述多核苷酸包含选自下组中的一种:
 - (a) 编码具有SEQ ID NO:2所示氨基酸序列的多肽或其片段、类似物、衍生物的多核苷酸;
 - (b) 与多核苷酸(a)互补的多核苷酸;或
 - (c) 与(a)或(b)有至少70%相同性的多核苷酸。
- 5、如权利要求4所述的多核苷酸,其特征在于所述多核苷酸包含编码具有SEQ ID NO:2所示氨基酸序列的多核苷酸。
- 6、如权利要求4所述的多核苷酸,其特征在于所述多核苷酸的序列包含有SEQ ID NO:1中105-1295位的序列或SEQ ID NO:1中1-1711位的序列。
- 7、一种含有外源多核苷酸的重组载体,其特征在于它是由权利要求4-6中的任一权利要求所述多核苷酸与质粒、病毒或运载体表达载体构建而成的重组载体。
- 8、一种含有外源多核苷酸的遗传工程化宿主细胞,其特征在于它是选自于下列一种宿主细胞:
 - (a) 用权利要求7所述的重组载体转化或转导的宿主细胞;或
 - (b) 用权利要求4-6中的任一权利要求所述多核苷酸转化或转导的宿主细胞。
- 9、一种具有人G蛋白44活性的多肽的制备方法,其特征在于所述方法包括:
 - (a) 在表达人G蛋白44条件下,培养权利要求8所述的工程化宿主细胞;
 - (b) 从培养物中分离出具有人G蛋白44活性的多肽。
- 10、一种能与多肽结合的抗体,其特征在于所述抗体是能与人G蛋白44特异性结合的抗体。
- 11、一类模拟或调节多肽活性或表达的化合物,其特征在于它们是模拟、促进、拮抗或抑制人G蛋白44的活性的化合物。
- 12、如权利要求11所述的化合物,其特征在于它是SEQ ID NO:1所示的多核苷酸序列或其片段的反义序列。
- 13、一种权利要求11所述化合物的应用,其特征在于所述化合物用于调节人G



蛋白 44 在体内、体外活性的方法。

- 14、一种检测与权利要求 1-3 中的任一权利要求所述多肽相关的疾病或疾病易感性的方法，其特征在于其包括检测所述多肽的表达量，或者检测所述多肽的活性，或者检测多核苷酸中引起所述多肽表达量或活性异常的核苷酸变异。
- 5 15、如权利要求 1-3 中的任一权利要求所述多肽的应用，其特征在于它应用于筛选人 G 蛋白 44 的模拟物、激动剂，拮抗剂或抑制剂；或者用于肽指纹图谱鉴定。
- 16、如权利要求 4-6 中的任一权利要求所述的核酸分子的应用，其特征在于它作为引物用于核酸扩增反应，或者作为探针用于杂交反应，或者用于制造基因芯片或微阵列。
- 10 17、如权利要求 1-6 及 11 中的任一权利要求所述的多肽、多核苷酸或化合物的应用，其特征在于用所述多肽、多核苷酸或其模拟物、激动剂、拮抗剂或抑制剂以安全有效剂量与药学上可接受的载体组成作为诊断或治疗与人 G 蛋白 44 异常相关的疾病的药物组合物。
- 15 18、权利要求 1-6 及 11 中的任一权利要求所述的多肽、多核苷酸或化合物的应用，其特征在于用所述多肽、多核苷酸或化合物制备用于治疗如恶性肿瘤，血液病，HIV 感染和免疫性疾病和各类炎症的药物。



说明书

一种新的多肽——人 G 蛋白 44 和编码这种多肽的多核苷酸

5 本发明属于生物技术领域，具体地说，本发明描述了一种新的多肽——人 G 蛋白 44，以及编码此多肽的多核苷酸序列。本发明还涉及此多核苷酸和多肽的制备方法和应用。

10 多细胞有机体的进化依赖于细胞之间的互相通讯联系。由于细胞间存在着精细的分工，一些细胞群体有赖于其他细胞群体，并要求其他的细胞群体产生应答反应。这种精巧的细胞间通讯网络可以控制细胞的生长、分裂、死亡、分化形成组织以及各种生命过程。

细胞间以 3 种方式进行通讯：（1）细胞通过分泌化学信号进行相互通讯；（2）细胞直接接触，通过与质膜结合的信号分子影响其他细胞；（3）细胞间形成缝隙连接使细胞质相互沟通，通过交换小分子来调节代谢反应。

15 信号传导是将细胞外信号转换为细胞内信使的过程。亲水性信号分子包括神经递质、生长因子、细胞因子、局部化学递质和大多数激素，它们不能穿过质膜，只能与细胞表面的受体结合形成配体受体复合物进行信号转导。根据信号转导的机制和受体蛋白的类型，细胞表面受体可分为 3 种：（1）离子通道耦联的受体；（2）连接酶的表面受体；（3）与 G 蛋白耦联的受体。

20 与 G 蛋白耦联的受体是一条跨越质膜 7 次的多肽链表面受体。配体与受体结合改变受体的构象，使之与质膜胞内侧上的与 GTP 结合的调节蛋白（G 蛋白）的三聚体结合，共同活化质膜上的靶蛋白（酶或离子通道）。可见配体受体复合物是通过 G 蛋白间接作用于效应器的。离子通道通过改变质膜对离子的通透性，酶通过影响胞内底物及其他蛋白来影响细胞的行为。

25 G 蛋白使受体和腺苷酸环化酶偶联起来，使细胞外信号转换为细胞内信号即 cAMP 第二信使，所以，G 蛋白也称偶联蛋白或信号转换蛋白。

GTP 结合蛋白（简称 G 蛋白）在生物体内有着广泛的作用，涉及细胞的增殖、信号转导、蛋白合成及蛋白定位等重要的生物学功能。GTP 结合蛋白又可分为各种超级家族，如：Ras 家族、Rab 家族等，各超级家族又拥有众多的家族成员。
30 蛋白合成延长因子 EF-Tu 是最早被发现的 G 蛋白，随后人们克隆得到了大量 GTP 结合蛋白家族的其它成员。G 蛋白家族的成员均含有保守的 GTP 结合基序，GTP 结合基序在生物体内被作为分子开关，通过其与 GDP、GTP 结合或否来调节蛋白



的表达与不表达。因而，G 蛋白在生物体内有着极为重要的作用，其表达异常将导致组织细胞的异常增殖及蛋白的异常表达，从而引发各种与之相关的疾病，如：各种恶性肿瘤及癌症、各种发育紊乱症、各种免疫系统疾病等。

1997 年，Staoru Senju, Yasuharhu Nishimura 等人从鼠中克隆得到了一种新的编码含有 GTP 结合基序的蛋白 GP-1，同源比较发现，该蛋白与人的 GP-1 蛋白及 AGP-1，CGP-1 蛋白均相似，均为一新的 G 蛋白家族的成员，且含有保守的 GTP 结合基序。

新的 G 蛋白家族的不同的成员具有不同的结构特征，但它们含有高度保守的四个氨基酸序列片段---GTP 结合基序：G1: RVAVVGNVDAGKSTLL; G2: RHKHEIESGRTSSVG; G3: ITFIDLAGHE; G4: FVVVTKID。这四个基序为 G 蛋白与 GTP、GDP 结合行使生物学功能的作用位点。这一新 G 蛋白家族的成员与其它 GTP 结合蛋白相似，在生物体内起着极其重要的作用，调节着体内各种蛋白的正常表达及细胞的增殖。

根据氨基酸同源比较的结果，本发明的多肽被推断鉴定为一种新的人 G 蛋白 44 (HGP44)，其同源蛋白的蛋白号为 Z92773，本发明的多肽被推断与鼠 GP-1、人 GP-1 等相似，同为新 G 蛋白家族的成员，并具有相似的生物学功能。

由于如上所述人 G 蛋白 44 蛋白在机体重要功能中起重要作用，而且相信这些调节过程中涉及大量的蛋白，因而本领域中一直需要鉴定更多参与这些过程的人 G 蛋白 44 蛋白，特别是鉴定这种蛋白的氨基酸序列。新人 G 蛋白 44 蛋白编码基因的分离也为研究确定该蛋白在健康和疾病状态下的作用提供了基础。这种蛋白可能构成开发疾病诊断和/或治疗药的基础，因此分离其编码 DNA 是非常重要的。

本发明的一个目的是提供分离的新的多肽---人 G 蛋白 44 以及其片段、类似物和衍生物。

25 本发明的另一个目的是提供编码该多肽的多核苷酸。

本发明的另一个目的是提供含有编码人 G 蛋白 44 的多核苷酸的重组载体。

本发明的另一个目的是提供含有编码人 G 蛋白 44 的多核苷酸的基因工程化宿主细胞。

本发明的另一个目的是提供生产人 G 蛋白 44 的方法。

30 本发明的另一个目的是提供针对本发明的多肽---人 G 蛋白 44 的抗体。

本发明的另一个目的是提供了针对本发明多肽---人 G 蛋白 44 的模拟化合物、拮抗剂、激动剂、抑制剂。



本发明的另一个目的是提供诊断治疗与人 G 蛋白 44 异常相关的疾病的方法。

本发明涉及一种分离的多肽，该多肽是人源的，它包含：具有 SEQ ID No. 2 氨基酸序列的多肽、或其保守性变体、生物活性片段或衍生物。较佳地，该多肽是具有 SEQ ID NO: 2 氨基酸序列的多肽。

5 本发明还涉及一种分离的多核苷酸，它包含选自下组的一种核苷酸序列或其变体：

(a) 编码具有 SEQ ID No. 2 氨基酸序列的多肽的多核苷酸；

(b) 与多核苷酸 (a) 互补的多核苷酸；

(c) 与 (a) 或 (b) 的多核苷酸序列具有至少 70% 相同性的多核苷酸。

10 更佳地，该多核苷酸的序列是选自下组的一种：(a) 具有 SEQ ID NO: 1 中 102-1295 位的序列；和 (b) 具有 SEQ ID NO: 1 中 1-1711 位的序列。

本发明另外涉及一种含有本发明多核苷酸的载体，特别是表达载体；一种用该载体遗传工程化的宿主细胞，包括转化、转导或转染的宿主细胞；一种包括培养所述宿主细胞和回收表达产物的制备本发明多肽的方法。

15 本发明还涉及一种能与本发明多肽特异性结合的抗体。

本发明还涉及一种筛选的模拟、激活、拮抗或抑制人 G 蛋白 44 蛋白活性的化合物的方法，其包括利用本发明的多肽。本发明还涉及用该方法获得的化合物。

20 本发明还涉及一种体外检测与人 G 蛋白 44 蛋白异常表达相关的疾病或疾病易感性的方法，包括检测生物样品中所述多肽或其编码多核苷酸序列中的突变，或者检测生物样品中本发明多肽的量或生物活性。

本发明也涉及一种药物组合物，它含有本发明多肽或其模拟物、激活剂、拮抗剂或抑制剂以及药学上可接受的载体。

本发明还涉及本发明的多肽和/或多核苷酸在制备用于治疗癌症、发育性疾病或免疫性疾病或其它由于人 G 蛋白 44 表达异常所引起疾病的药物的用途。

25 本发明的其它方面由于本文的技术的公开，对本领域的技术人员而言是显而易见的。

本说明书和权利要求书中使用的下列术语除非特别说明具有如下的含义：

30 “核酸序列”是指寡核苷酸、核苷酸或多核苷酸及其片段或部分，也可以指基因组或合成的 DNA 或 RNA，它们可以是单链或双链的，代表有义链或反义链。类似地，术语“氨基酸序列”是指寡肽、肽、多肽或蛋白质序列及其片段或部分。当本发明中的“氨基酸序列”涉及一种天然存在的蛋白质分子的氨基酸序列时，



这种“多肽”或“蛋白质”不意味着将氨基酸序列限制为与所述蛋白质分子相关的完整的天然氨基酸。

5 蛋白质或多核苷酸“变体”是指一种具有一个或多个氨基酸或核苷酸改变的氨基酸序列或编码它的多核苷酸序列。所述改变可包括氨基酸序列或核苷酸序列中氨基酸或核苷酸的缺失、插入或替换。变体可具有“保守性”改变，其中替换的氨基酸具有与原氨基酸相类似的结构或化学性质，如用亮氨酸替换异亮氨酸。变体也可具有非保守性改变，如用色氨酸替换甘氨酸。

“缺失”是指在氨基酸序列或核苷酸序列中一个或多个氨基酸或核苷酸的缺失。

10 “插入”或“添加”是指在氨基酸序列或核苷酸序列中的改变导致与天然存在的分子相比，一个或多个氨基酸或核苷酸的增加。“替换”是指由不同的氨基酸或核苷酸替换一个或多个氨基酸或核苷酸。

15 “生物活性”是指具有天然分子的结构、调控或生物化学功能的蛋白质。类似地，术语“免疫学活性”是指天然的、重组的或合成蛋白质及其片段在合适的动物或细胞中诱导特定免疫反应以及与特异性抗体结合的能力。

“激动剂”是指当与人G蛋白44结合时，一种可引起该蛋白质改变从而调节该蛋白质活性的分子。激动剂可以包括蛋白质、核酸、碳水化合物或任何其它可结合人G蛋白44的分子。

20 “拮抗剂”或“抑制物”是指当与人G蛋白44结合时，一种可封闭或调节人G蛋白44的生物学活性或免疫学活性的分子。拮抗剂和抑制物可以包括蛋白质、核酸、碳水化合物或任何其它可结合人G蛋白44的分子。

“调节”是指人G蛋白44的功能发生改变，包括蛋白质活性的升高或降低、结合特性的改变及人G蛋白44的任何其它生物学性质、功能或免疫性质的改变。

25 “基本上纯”是指基本上不含天然与其相关的其它蛋白、脂类、糖类或其它物质。本领域的技术人员能用标准的蛋白质纯化技术纯化人G蛋白44。基本上纯的人G蛋白44在非还原性聚丙烯酰胺凝胶上能产生单一的主带。人G蛋白44多肽的纯度可用氨基酸序列分析。

30 “互补的”或“互补”是指在允许的盐浓度和温度条件下通过碱基配对的多核苷酸天然结合。例如，序列“C-T-G-A”可与互补的序列“G-A-C-T”结合。两个单链分子之间的互补可以是部分的或全部的。核酸链之间的互补程度对于核酸链之间杂交的效率及强度有明显影响。

“同源性”是指互补的程度，可以是部分同源或完全同源。“部分同源”是



指一种部分互补的序列，其至少可部分抑制完全互补的序列与靶核酸的杂交。这种杂交的抑制可通过在严格性程度降低的条件下进行杂交（Southern印迹或Northern印迹等）来检测。基本上同源的序列或杂交探针可竞争和抑制完全同源的序列与靶序列在的严格性程度降低的条件下的结合。这并不意味严格性程度降低的条件允许非特异性结合，因为严格性程度降低的条件要求两条序列相互的结合为特异性或选择性相互作用。

“相同性百分率”是指在两种或多种氨基酸或核酸序列比较中序列相同或相似的百分率。可用电子方法测定相同性百分率，如通过MEGALIGN程序（Lasergene software package, DNASTAR, Inc., Madison Wis.）。MEGALIGN程序可根据不同的方法如Cluster法比较两种或多种序列（Higgins, D. G. 和 P.M. Sharp (1988) Gene 73: 237-244）。Cluster法通过检查所有配对之间的距离将各组序列排列成簇。然后将各簇以成对或成组分配。两个氨基酸序列如序列A和序列B之间的相同性百分率通过下式计算：

15

序列A与序列B之间匹配的残基个数

100

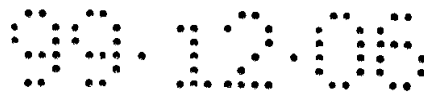
20 序列A的残基数—序列A中间隔残基数—序列B中间隔残基数

也可以通过Cluster法或用本领域周知的方法如Jotun Hein 测定核酸序列之间的相同性百分率（Hein J., (1990) Methods in emzumology 183:625-645）。

“相似性”是指氨基酸序列之间排列对比时相应位置氨基酸残基的相同或保守性取代的程度。用于保守性取代的氨基酸例如，带负电荷的氨基酸可包括天冬氨酸和谷氨酸；带正电荷的氨基酸可包括赖氨酸和精氨酸；具有不带电荷的头部基团有相似亲水性的氨基酸可包括亮氨酸、异亮氨酸和缬氨酸；甘氨酸和丙氨酸；天冬酰胺和谷氨酰胺；丝氨酸和苏氨酸；苯丙氨酸和酪氨酸。

“反义”是指与特定的DNA或RNA序列互补的核苷酸序列。“反义链”是指与“有义链”互补的核酸链。

30 “衍生物”是指HFP或编码其的核酸的化学修饰物。这种化学修饰物可以用烷基、酰基或氨基替换氢原子。核酸衍生物可编码保留天然分子的主要生物学特性的多肽。



“抗体”是指完整的抗体分子及其片段，如 Fa、F(ab')₂ 及 Fv，其能特异性结合人G蛋白44的抗原决定簇。

“人源化抗体”是指非抗原结合区域的氨基酸序列被替换变得与人抗体更为相似，但仍保留原始结合活性的抗体。

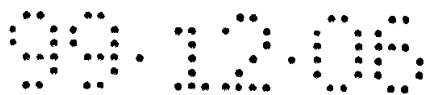
5 “分离的”一词指将物质从它原来的环境（例如，若是自然产生的就指其天然环境）之中移出。比如说，一个自然产生的多核苷酸或多肽存在于活动物中就是没有被分离出来，但同样的多核苷酸或多肽同一些或全部在自然系统中与之共存物质分开就是分离的。这样的多核苷酸可能是某一载体的一部分，也可能这样的多核苷酸或多肽是某一组合物的一部分。既然载体或组合物不是它天然环境的成分，它们仍然是分离的。

15 如本发明所用，“分离的”是指物质从其原始环境中分离出来（如果是天然物质，原始环境即是天然环境）。如活体细胞内的天然状态下的多聚核苷酸和多肽是没有分离纯化的，但同样的多聚核苷酸或多肽如从天然状态中同存在的其他物质中分开，则为分离纯化的。

如本文所用，“分离的人G蛋白44”是指人G蛋白44基本上不含天然与其相关的其它蛋白、脂类、糖类或其它物质。本领域的技术人员能用标准的蛋白质纯化技术纯化人G蛋白44。基本上纯的多肽在非还原聚丙烯酰胺凝胶上能产生单一的主带。人G蛋白44多肽的纯度能用氨基酸序列分析。

20 本发明提供了一种新的多肽——人G蛋白44，其基本上是由SEQ ID NO: 2所示的氨基酸序列组成的。本发明的多肽可以是重组多肽、天然多肽、合成多肽，优选重组多肽。本发明的多肽可以是天然纯化的产物，或是化学合成的产物，或使用重组技术从原核或真核宿主（例如，细菌、酵母、高等植物、昆虫和哺乳动物细胞）中产生。根据重组生产方案所用的宿主，本发明的多肽可以是糖基化的，或可以是非糖基化的。本发明的多肽还可包括或不包括起始的甲硫氨酸残基。

25 本发明还包括人G蛋白44的片段、衍生物和类似物。如本发明所用，术语“片段”、“衍生物”和“类似物”是指基本上保持本发明的人G蛋白44相同的生物学功能或活性的多肽。本发明多肽的片段、衍生物或类似物可以是：（I）这样一种，其中一个或多个氨基酸残基被保守或非保守氨基酸残基（优选的是保守氨基酸残基）取代，并且取代的氨基酸可以是也可以不是由遗传密码子编码的；或者（II）这样一种，其中一个或多个氨基酸残基上的某个基团被其它基团取代包含取代基；或者（III）这样一种，其中成熟多肽与另一种化合物（比如延长



多肽半衰期的化合物，例如聚乙二醇）融合；或者（IV）这样一种，其中附加的氨基酸序列融合进成熟多肽而形成的多肽序列（如前导序列或分泌序列或用来纯化此多肽的序列或蛋白原序列）通过本文的阐述，这样的片段、衍生物和类似物被认为在本领域技术人员的知识范围之内。

5 本发明提供了分离的核酸（多核苷酸），基本由编码具有 SEQ ID NO: 2 氨基酸序列的多肽的多核苷酸组成。本发明的多核苷酸序列包括 SEQ ID NO: 1 的核苷酸序列。本发明的多核苷酸是从人胎脑组织的 cDNA 文库中发现的。它包含的多核苷酸序列全长为 1711 个碱基，其开放读框（105—1295）编码了 396 个氨基酸。根据氨基酸序列同源比较发现，此多肽与鼠的 G 蛋白有 68% 的同源性，可推断出该人 G 蛋白 44 具有鼠的 G 蛋白相似的结构和功能。

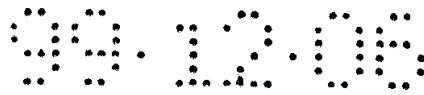
10 本发明的多核苷酸可以是 DNA 形式或是 RNA 形式。DNA 形式包括 cDNA、基因组 DNA 或人工合成的 DNA。DNA 可以是单链的或是双链的。DNA 可以是编码链或非编码链。编码成熟多肽的编码区序列可以与 SEQ ID NO: 1 所示的编码区序列相同或者是简并的变异体。如本发明所用，“简并的变异体”在本发明中是指编码具有 SEQ ID NO: 2 的蛋白质或多肽，但与 SEQ ID NO: 1 所示的编码区序列有差别的核酸序列。

编码 SEQ ID NO: 2 的成熟多肽的多核苷酸包括：只有成熟多肽的编码序列；成熟多肽的编码序列和各种附加编码序列；成熟多肽的编码序列（和任选的附加编码序列）以及非编码序列。

20 术语“编码多肽的多核苷酸”是指包括编码此多肽的多核苷酸和包括附加编码和/或非编码序列的多核苷酸。

25 本发明还涉及上述描述多核苷酸的变异体，其编码与本发明有相同的氨基酸序列的多肽或多肽的片断、类似物和衍生物。此多核苷酸的变异体可以是天然发生的等位变异体或非天然发生的变异体。这些核苷酸变异体包括取代变异体、缺失变异体和插入变异体。如本领域所知的，等位变异体是一个多核苷酸的替换形式，它可能是一个或多个核苷酸的取代、缺失或插入，但不会从实质上改变其编码的多肽的功能。

30 本发明还涉及与以上所描述的序列杂交的多核苷酸（两个序列之间具有至少 50%，优选具有 70% 的相同性）。本发明特别涉及在严格条件下与本发明所述多核苷酸可杂交的多核苷酸。在本发明中，“严格条件”是指：（1）在较低离子强度和较高温度下的杂交和洗脱，如 0.2×SSC，0.1%SDS，60℃；或（2）杂交时加用变性剂，如 50%（v/v）甲酰胺，0.1%小牛血清/0.1%Ficoll，42℃等；或（3）仅在两



条序列之间的相同性至少在 95% 以上, 更好是 97% 以上时才发生杂交。并且, 可杂交的多核苷酸编码的多肽与 SEQ ID NO: 2 所示的成熟多肽有相同的生物学功能和活性。

5 本发明还涉及与以上所描述的序列杂交的核酸片段。如本发明所用, “核酸片段”的长度至少含 10 个核苷酸, 较好是至少 20-30 个核苷酸, 更好是至少 50-60 个核苷酸, 最好是至少 100 个核苷酸以上。核酸片段也可用于核酸的扩增技术(如 PCR)以确定和/或分离编码人 G 蛋白 44 的多核苷酸。

本发明中的多肽和多核苷酸优选以分离的形式提供, 更佳地被纯化至均质。

10 本发明的编码人 G 蛋白 44 的特异的多核苷酸序列能用多种方法获得。例如, 用本领域熟知的杂交技术分离多核苷酸。这些技术包括但不局限于: 1) 用探针与基因组或 cDNA 文库杂交以检出同源的多核苷酸序列, 和 2) 表达文库的抗体筛选以检出具有共同结构特征的克隆的多核苷酸片段。

本发明的 DNA 片段序列也能用下列方法获得: 1) 从基因组 DNA 分离双链 DNA 序列; 2) 化学合成 DNA 序列以获得所述多肽的双链 DNA。

15 上述提到的方法中, 分离基因组 DNA 最不常用。DNA 序列的直接化学合成是经常选用的方法。更经常选用的方法是 cDNA 序列的分离。分离感兴趣的 cDNA 的标准方法是从高表达该基因的供体细胞分离 mRNA 并进行逆转录, 形成质粒或噬菌体 cDNA 文库。提取 mRNA 的方法已有多种成熟的技术, 试剂盒也可从商业途径获得(Qiagene)。而构建 cDNA 文库也是通常的方法(Sambrook, et al., Molecular
20 Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory. New York, 1989)。还可得到商业供应的 cDNA 文库, 如 Clontech 公司的不同 cDNA 文库。当结合使用聚合酶反应技术时, 即使极少的表达产物也能克隆。

可用常规方法从这些 cDNA 文库中筛选本发明的基因。这些方法包括(但不限于): (1) DNA-DNA 或 DNA-RNA 杂交; (2) 标志基因功能的出现或丧失; (3) 测定人
25 G 蛋白 44 的转录本的水平; (4) 通过免疫学技术或测定生物学活性, 来检测基因表达的蛋白产物。上述方法可单用, 也可多种方法联合应用。

在第(1)种方法中, 杂交所用的探针是与本发明的多核苷酸的任何一部分同源, 其长度至少 10 个核苷酸, 较好是至少 30 个核苷酸, 更好是至少 50 个核苷酸, 最好是至少 100 个核苷酸。此外, 探针的长度通常在 2000 个核苷酸之内, 30 较佳的为 1000 个核苷酸之内。此处所用的探针通常是在本发明的基因序列信息的基础上化学合成的 DNA 序列。本发明的基因本身或者片段当然可以用作探针。DNA 探针的标记可用放射性同位素, 荧光素或酶(如碱性磷酸酶)等。



在第(4)种方法中,检测人G蛋白44基因表达的蛋白产物可用免疫学技术如Western印迹法,放射免疫沉淀法,酶联免疫吸附法(ELISA)等。

应用PCR技术扩增DNA/RNA的方法(Saiki, et al. Science 1985; 230: 1350-1354)被优选用于获得本发明的基因。特别是很难从文库中得到全长的cDNA时,可优选使用RACE法(RACE-cDNA末端快速扩增法),用于PCR的引物可根据本文所公开的本发明的多核苷酸序列信息适当地选择,并可用常规方法合成。可用常规方法如通过凝胶电泳分离和纯化扩增的DNA/RNA片段。

如上所述得到的本发明的基因,或者各种DNA片段等的多核苷酸序列可用常规方法如双脱氧链终止法(Sanger et al. PNAS, 1977, 74: 5463-5467)测定。这类多核苷酸序列测定也可用商业测序试剂盒等。为了获得全长的cDNA序列,测序需反复进行。有时需要测定多个克隆的cDNA序列,才能拼接成全长的cDNA序列。

本发明也涉及包含本发明的多核苷酸的载体,以及用本发明的载体或直接用人G蛋白44编码序列经基因工程产生的宿主细胞,以及经重组技术产生本发明所述多肽的方法。

本发明中,编码人G蛋白44的多核苷酸序列可插入到载体中,以构成含有本发明所述多核苷酸的重组载体。术语“载体”指本领域熟知的细菌质粒、噬菌体、酵母质粒、植物细胞病毒、哺乳动物细胞病毒如腺病毒、逆转录病毒或其它载体。在本发明中适用的载体包括但不限于:在细菌中表达的基于T7启动子的表达载体(Rosenberg, et al. Gene, 1987, 56:125);在哺乳动物细胞中表达的pMSXND表达载体(Lee and Nathans, J Bio Chem. 263: 3521, 1988)和在昆虫细胞中表达的来源于杆状病毒的载体。总之,只要能在宿主体内复制和稳定,任何质粒和载体都可以用于构建重组表达载体。表达载体的一个重要特征是通常含有复制起始点、启动子、标记基因和翻译调控元件。

本领域的技术人员熟知的方法能用于构建含编码人G蛋白44的DNA序列和合适的转录/翻译调控元件的表达载体。这些方法包括体外重组DNA技术、DNA合成技术、体内重组技术等(Sambrook, et al. Molecular Cloning, a Laboratory Manual, Cold Spring Harbor Laboratory. New York, 1989)。所述的DNA序列可有效连接到表达载体中的适当启动子上,以指导mRNA合成。这些启动子的代表性例子有:大肠杆菌的lac或trp启动子;λ噬菌体的P_L启动子;真核启动子包括CMV立即早期启动子、HSV胸苷激酶启动子、早期和晚期SV40启动子、反转录病毒的LTRs和其它一些已知的可控制基因在原核细胞或真核细胞或其病毒



中表达的启动子。表达载体还包括翻译起始用的核糖体结合位点和转录终止子等。在载体中插入增强子序列将会使其在高等真核细胞中的转录得到增强。增强子是 DNA 表达的顺式作用因子，通常大约有 10 到 300 个碱基对，作用于启动子以增强基因的转录。可举的例子包括在复制起始点晚期一侧的 100 到 270 个碱基
5 对的 SV40 增强子、在复制起始点晚期一侧的多瘤增强子以及腺病毒增强子等。

此外，表达载体优选地包含一个或多个选择性标记基因，以提供用于选择转化的宿主细胞的表型性状，如真核细胞培养用的二氢叶酸还原酶、新霉素抗性以及绿色荧光蛋白 (GFP)，或用于大肠杆菌的四环素或氨苄青霉素抗性等。

本领域一般技术人员都清楚如何选择适当的载体/转录调控元件（如启动子、增强子等）和选择性标记基因。
10

本发明中，编码人 G 蛋白 44 的多核苷酸或含有该多核苷酸的重组载体可转化或转导入宿主细胞，以构成含有该多核苷酸或重组载体的基因工程化宿主细胞。术语“宿主细胞”指原核细胞，如细菌细胞；或是低等真核细胞，如酵母细胞；或是高等真核细胞，如哺乳动物细胞。代表性例子有：大肠杆菌，链霉菌属；
15 细菌细胞如鼠伤寒沙门氏菌；真菌细胞如酵母；植物细胞；昆虫细胞如果蝇 S2 或 Sf9；动物细胞如 CHO、COS 或 Bowes 黑素瘤细胞等。

用本发明所述的 DNA 序列或含有所述 DNA 序列的重组载体转化宿主细胞可用本领域技术人员熟知的常规技术进行。当宿主为原核生物如大肠杆菌时，能吸收 DNA 的感受态细胞可在指数生长期后收获，用 CaCl_2 法处理，所用的步骤在本领域众所周知。可供选择的是用 MgCl_2 。如果需要，转化也可用电穿孔的方法进行。当宿主是真核生物，可选用如下的 DNA 转染方法：磷酸钙共沉淀法，或者常规机械方法如显微注射、电穿孔、脂质体包装等。
20

通过常规的重组 DNA 技术，利用本发明的多核苷酸序列可用来表达或生产重组的人 G 蛋白 44 (Science, 1984; 224: 1431)。一般来说有以下步骤：

- 25
- (1). 用本发明的编码人 G 蛋白 44 的多核苷酸(或变形体)，或用含有该多核苷酸的重组表达载体转化或转导合适的宿主细胞；
 - (2). 在合适的培养基中培养宿主细胞；
 - (3). 从培养基或细胞中分离、纯化蛋白质。

在步骤 (2) 中，根据所用的宿主细胞，培养中所用的培养基可选自各种常规培养基。在适于宿主细胞生长的条件下进行培养。当宿主细胞生长到适当的细胞密度后，用合适的方法(如温度转换或化学诱导)诱导选择的启动子，将细胞再培养一段时间。
30

在步骤(3)中,重组多肽可包被于细胞内、或在细胞膜上表达、或分泌到细胞外。如果需要,可利用其物理的、化学的和其它特性通过各种分离方法分离和纯化重组的蛋白。这些方法是本领域技术人员所熟知的。这些方法包括但不限于:常规的复性处理、蛋白沉淀剂处理(盐析方法)、离心、渗透破菌、超声波处理、超离心、分子筛层析(凝胶过滤)、吸附层析、离子交换层析、高效液相层析(HPLC)和其它各种液相层析技术及这些方法的结合。

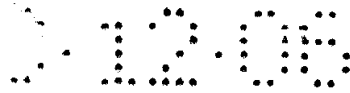
本发明的多肽以及该多肽的拮抗剂、激动剂和抑制剂可直接用于疾病治疗,例如,可治疗恶性肿瘤、肾上腺缺乏症、皮肤病、各类炎症、HIV感染和免疫性疾病等。

本发明的多肽(人G蛋白44)以及该多肽的拮抗剂、激动剂和抑制剂可直接用于疾病治疗,例如,可治疗恶性肿瘤、肾上腺缺乏症、皮肤病、各类炎症、HIV感染和免疫性疾病等。

GTP结合蛋白可分为多个超级家族,具有众多的家族成员。各超级家族的成员均具有保守的GTP结合基序,GTP结合基序具有广泛的生物学功能,在生物体内调节着细胞的增殖、控制着信号的转导及各种蛋白在体内的表达。这些蛋白的表达异常将导致组织细胞的异常增殖及蛋白的异常表达,从而引发各种与之相关的疾病,如:各种恶性肿瘤及癌症、各种发育紊乱症、各种免疫系统疾病等。

具体就本发明的人G蛋白44而言,该蛋白的表达与各种恶性肿瘤及癌症的发生有关;因此本发明的多肽可用于很多疾病的诊断和治疗,如各种与之相关的恶性肿瘤及癌症,这些疾病包括但不限于以下所述,胃癌、肝癌、大肠癌、乳腺癌、肺癌、前列腺癌、宫颈癌、胰腺癌、食道癌、垂体腺瘤、甲状腺良性肿瘤、甲状腺癌、甲状旁腺腺瘤、甲状旁腺癌、肾上腺髓脂肪瘤、嗜铬细胞瘤、胰岛细胞肿瘤、多发性内分泌腺肿瘤、胸腺肿瘤等。

本发明的人G蛋白44也可用于诊断和治疗各种与之相关的发育紊乱症,这些疾病包括但不限于以下所述,脊柱裂、颅脑裂、无脑畸形、脑膨出、孔脑畸形、Down综合症、先天性脑积水、导水管畸形、软骨发育不全性侏儒病、脊柱骨骺发育不良症、假软骨发育不全症、Langer-Giedion综合症、漏斗胸、生殖腺发育不全、先天性肾上腺增生、尿道上裂、隐睾、伴有身材矮小的畸形综合症如Conradi综合症与Danbolt-Closs综合症、先天性青光眼或白内障、先天性晶状体位置异常、先天性小睑裂、视网膜发育异常、先天性视神经萎缩、先天性感觉神经性听觉损失、裂手裂脚症、畸胎、Williams综合症、Alagille综合症、贝



魏二氏综合症等。

本发明的人G蛋白44还可用于诊断和治疗各种与之表达异常相关的免疫系统疾病，这些疾病包括但不限于以下所述，类风湿性关节炎、慢性活动性肝炎、原发性干燥综合症、急性前葡萄膜炎、淋球菌感染后关节炎、强直性脊柱炎、血色素沉着症、免疫复合物型肾小球肾炎、淋球菌感染后心肌炎、系统性红斑狼疮、类风湿性关节炎、硬皮病、多发性肌炎、口眼干燥综合症、结节性多动脉炎、Wegener肉芽肿病、重症肌无力、格林-巴利综合症、自身免疫性溶血性贫血、免疫性血小板减少性紫癜、自身免疫性间质性肾炎、自身免疫性胃炎、胰岛素自身免疫性综合症、自身免疫性甲状腺疾病、自身免疫性心脏病等。

10 本发明也提供了筛选化合物以鉴定提高(激动剂)或阻遏(拮抗剂)人G蛋白44的药剂的方法。激动剂提高人G蛋白44刺激细胞增殖等生物功能，而拮抗剂阻止和治疗与细胞过度增殖有关的紊乱如各种癌症。例如，能在药物的存在下，将哺乳动物细胞或表达人G蛋白44的膜制剂与标记的人G蛋白44一起培养。然后测定药物提高或阻遏此相互作用的能力。

15 人G蛋白44的拮抗剂包括筛选出的抗体、化合物、受体缺失物和类似物等。人G蛋白44的拮抗剂可以与人G蛋白44结合并消除其功能，或是抑制该多肽的产生，或是与该多肽的活性位点结合使该多肽不能发挥生物学功能。

20 在筛选作为拮抗剂的化合物时，可以将人G蛋白44加入生物分析测定中，通过测定化合物对人G蛋白44和其受体之间相互作用的影响来确定化合物是否是拮抗剂。用上述筛选化合物的同样方法，可以筛选出起拮抗剂作用的受体缺失物和类似物。能与人G蛋白44结合的多肽分子可通过筛选由各种可能组合的氨基酸结合于固相物组成的随机多肽库而获得。筛选时，一般应对人G蛋白44分子进行标记。

25 本发明提供了用多肽，及其片段、衍生物、类似物或它们的细胞作为抗原以生产抗体的方法。这些抗体可以是多克隆抗体或单克隆抗体。本发明还提供了针对人G蛋白44抗原决定簇的抗体。这些抗体包括(但不限于)：多克隆抗体、单克隆抗体、嵌合抗体、单链抗体、Fab片段和Fab表达文库产生的片段。

30 多克隆抗体的生产可用人G蛋白44直接注射免疫动物(如家兔，小鼠，大鼠等)的方法得到，多种佐剂可用于增强免疫反应，包括但不限于弗氏佐剂等。制备人G蛋白44的单克隆抗体的技术包括但不限于杂交瘤技术(Kohler and Milstein. Nature, 1975, 256: 495-497)，三瘤技术，人B-细胞杂交瘤技术，EBV-杂交瘤技术等。将人恒定区和非人源的可变区结合的嵌合抗体可用已有的技术生



产 (Morrison et al, PNAS, 1985, 81: 6851)。而已有的生产单链抗体的技术 (U. S. Pat No. 4946778) 也可用于生产抗人 G 蛋白 44 的单链抗体。

抗人 G 蛋白 44 的抗体可用于免疫组织化学技术中, 检测活检标本中的人 G 蛋白 44。

5 与人 G 蛋白 44 结合的单克隆抗体也可用放射性同位素标记, 注入体内可跟踪其位置和分布。这种放射性标记的抗体可作为一种非创伤性诊断方法用于肿瘤细胞的定位和判断是否有转移。

10 抗体还可用于设计针对体内某一特殊部位的免疫毒素。如人 G 蛋白 44 高亲和性的单克隆抗体可与细菌或植物毒素 (如白喉毒素, 蓖麻蛋白, 红豆碱等) 共价结合。一种通常的方法是用巯基交联剂如 SPDP, 攻击抗体的氨基, 通过二硫键的交换, 将毒素结合于抗体上, 这种杂交抗体可用于杀灭人 G 蛋白 44 阳性的细胞。

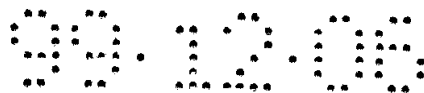
本发明中的抗体可用于治疗或预防与人 G 蛋白 44 相关的疾病。给予适当剂量的抗体可以刺激或阻断人 G 蛋白 44 的产生或活性。

15 本发明还涉及定量和定位检测人 G 蛋白 44 水平的诊断试验方法。这些试验是本领域所熟知的, 且包括 FISH 测定和放射免疫测定。试验中所检测的人 G 蛋白 44 水平, 可以用作解释人 G 蛋白 44 在各种疾病中的重要性和用于诊断人 G 蛋白 44 起作用的疾病。

本发明的多肽还可用作肽谱分析, 例如, 多肽可用物理的、化学或酶进行特异性切割, 并进行一维或二维或三维的凝胶电泳分析, 更好的是进行质谱分析。

20 编码人 G 蛋白 44 的多核苷酸也可用于多种治疗目的。基因治疗技术可用于治疗由于人 G 蛋白 44 的无表达或异常/无活性表达所致的细胞增殖、发育或代谢异常。重组的基因治疗载体 (如病毒载体) 可设计用于表达变异的人 G 蛋白 44, 以抑制内源性的人 G 蛋白 44 活性。例如, 一种变异的人 G 蛋白 44 可以是缩短的、缺失了信号传导功能域的人 G 蛋白 44, 虽可与下游的底物结合, 但缺乏信号传导
25 活性。因此重组的基因治疗载体可用于治疗人 G 蛋白 44 表达或活性异常所致的疾病。来源于病毒的表达载体如逆转录病毒、腺病毒、腺病毒相关病毒、单纯疱疹病毒、细小病毒等可用于将编码人 G 蛋白 44 的多核苷酸转移至细胞内。构建携带编码人 G 蛋白 44 的多核苷酸的重组病毒载体的方法可见于已有文献 (Sambrook, et al.)。另外重组编码人 G 蛋白 44 的多核苷酸可包装到脂质体中转移至细胞内。
30

多核苷酸导入组织或细胞内的方法包括: 将多核苷酸直接注入到体内组织中; 或在体外通过载体 (如病毒、噬菌体或质粒等) 先将多核苷酸导入细胞中, 再



将细胞移植到体内等。

抑制人 G 蛋白 44 mRNA 的寡核苷酸(包括反义 RNA 和 DNA)以及核酶也在本发明的范围之内。核酶是一种能特异性分解特定 RNA 的酶样 RNA 分子,其作用机制是核酶分子与互补的靶 RNA 特异性杂交后进行核酸内切作用。反义的 RNA 和 DNA 及核酶可用已有的任何 RNA 或 DNA 合成技术获得,如固相磷酸酰胺化学合成法合成寡核苷酸的技术已广泛应用。反义 RNA 分子可通过编码该 RNA 的 DNA 序列在体外或体内转录获得。这种 DNA 序列已整合到载体的 RNA 聚合酶启动子的下游。为了增加核酸分子的稳定性,可用多种方法对其进行修饰,如增加两侧的序列长度,核糖核苷之间的连接应用磷酸硫酸酯键或肽键而非磷酸二酯键。

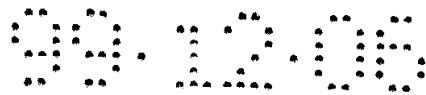
10 编码人 G 蛋白 44 的多核苷酸可用于与人 G 蛋白 44 的相关疾病的诊断。编码人 G 蛋白 44 的多核苷酸可用于检测人 G 蛋白 44 的表达与否或在疾病状态下人 G 蛋白 44 的异常表达。如编码人 G 蛋白 44 的 DNA 序列可用于对活检标本进行杂交以判断人 G 蛋白 44 的表达状况。杂交技术包括 Southern 印迹法, Northern 印迹法、原位杂交等。这些技术方法都是公开的成熟技术,相关的试剂盒都可从商业途径得到。本发明的多核苷酸的一部分或全部可作为探针固定在微阵列 (Microarray) 或 DNA 芯片(又称为“基因芯片”)上,用于分析组织中基因的差异表达分析和基因诊断。用人 G 蛋白 44 特异的引物进行 RNA-聚合酶链反应 (RT-PCR) 体外扩增也可检测人 G 蛋白 44 的转录产物。

20 检测人 G 蛋白 44 基因的突变也可用于诊断人 G 蛋白 44 相关的疾病。人 G 蛋白 44 突变的形式包括与正常野生型人 G 蛋白 44 DNA 序列相比的点突变、易位、缺失、重组和其它任何异常等。可用已有的技术如 Southern 印迹法、DNA 序列分析、PCR 和原位杂交检测突变。另外,突变有可能影响蛋白的表达,因此用 Northern 印迹法、Western 印迹法可间接判断基因有无突变。

25 本发明的序列对染色体鉴定也是有价值的。该序列会特异性地针对某条人染色体具体位置且并可以与其杂交。目前,需要鉴定染色体上的各基因的具体位点。现在,只有很少的基于实际序列数据(重复多态性)的染色体标记物可用于标记染色体位置。根据本发明,为了将这些序列与疾病相关基因相关联,其重要的第一步就是将这些 DNA 序列定位于染色体上。

30 简而言之,根据 cDNA 制备 PCR 引物(优选 15-35bp),可以将序列定位于染色体上。然后,将这些引物用于 PCR 筛选含各条人染色体的体细胞杂合细胞。只有那些含有相应于引物的人基因的杂合细胞会产生扩增的片段。

体细胞杂合细胞的 PCR 定位法,是将 DNA 定位到具体染色体的快捷方法。使



用本发明的寡核苷酸引物，通过类似方法，可利用一组来自特定染色体的片段或大量基因组克隆而实现亚定位。可用于染色体定位的其它类似策略包括原位杂交、用标记的流式分选的染色体预筛选和杂交预选，从而构建染色体特异的cDNA库。

5 将cDNA克隆与中期染色体进行荧光原位杂交(FISH)，可以在一个步骤中精确地进行染色体定位。此技术的综述，参见Verma等，*Human Chromosomes: a Manual of Basic Techniques*, Pergamon Press, New York(1988)。

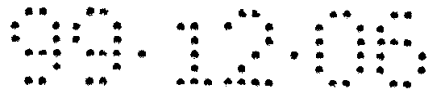
一旦序列被定位到准确的染色体位置，此序列在染色体上的物理位置就可以与基因图数据相关联。这些数据可见于例如，V.Mckusick, *Mendelian*
10 *Inheritance in Man*(可通过与Johns Hopkins University Welch Medical Library联机获得)。然后可通过连锁分析，确定基因与业已定位到染色体区域上的疾病之间的关系。

接着，需要测定患病和未患病个体间的cDNA或基因组序列差异。如果在一些或所有的患病个体中观察到某突变，而该突变在任何正常个体中未观察到，
15 则该突变可能是疾病的病因。比较患病和未患病个体，通常涉及首先寻找染色体中结构的变化，如从染色体水平可见的或用基于cDNA序列的PCR可检测的缺失或易位。根据目前的物理作图和基因定位技术的分辨能力，被精确定位至与疾病有关的染色体区域的cDNA，可以是50至500个潜在致病基因间之一种(假定1兆碱基作图分辨能力和每20kb对应于一个基因)。

20 可以将本发明的多肽、多核苷酸及其模拟物、激动剂、拮抗剂和抑制剂与合适的药物载体组合后使用。这些载体可以是水、葡萄糖、乙醇、盐类、缓冲液、甘油以及它们的组合。组合物包含安全有效量的多肽或拮抗剂以及不影响药物效果的载体和赋形剂。这些组合物可以作为药物用于疾病治疗。

本发明还提供含有一种或多种容器的药盒或试剂盒，容器中装有一种或多种
25 本发明的药用组合物成分。与这些容器一起，可以有由制造、使用或销售药品或生物制品的政府管理机构所给出的指示性提示，该提示反映出生产、使用或销售的政府管理机构许可其在人体上施用。此外，本发明的多肽可以与其它的治疗化合物结合使用。

药物组合物可以以方便的方式给药，如通过局部、静脉内、腹膜内、肌内、
30 皮下、鼻内或皮内的给药途径。人G蛋白44以有效地治疗和/或预防具体的适应症的量来给药。施用于患者的人G蛋白44的量和剂量范围将取决于许多因素，如给药方式、待治疗者的健康条件和诊断医生的判断。



下列附图用于说明本发明的具体实施方案，而不用于限定由权利要求书所界定的本发明范围。

图1是本发明人G蛋白44和鼠的G蛋白的氨基酸序列同源性比较图。上方序列是人G蛋白44，下方序列是鼠的G蛋白。相同氨基酸在两个序列间用单字符氨基酸表示，相似氨基酸用“+”表示。

图2为分离的人G蛋白44的聚丙烯酰胺凝胶电泳图(SDS-PAGE)。44kDa为蛋白质的分子量。箭头所指为分离出的蛋白条带。

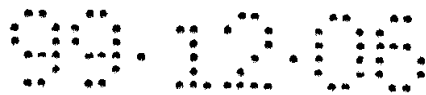
下面结合具体实施例，进一步阐述本发明。应理解，这些实施例仅用于说明本发明而不用于限制本发明的范围。下列实施例中未注明具体条件的实验方法，通常按照常规条件如 Sambrook 等人，分子克隆：实验室手册(New York: Cold Spring Harbor Laboratory Press, 1989)中所述的条件，或按照制造厂商所建议的条件。

实施例1：人G蛋白44的克隆

用异硫氰酸胍/酚/氯仿一步法提取人胎脑总RNA。用Quik mRNA Isolation Kit (Qiegene 公司产品)从总RNA中分离poly(A) mRNA。2ug poly(A) mRNA经逆转录形成cDNA。用Smart cDNA克隆试剂盒(购自Clontech)将cDNA片段定向插入到pBSK(+)载体(Clontech公司产品)的多克隆位点上，转化DH5 α ，细菌形成cDNA文库。用Dye terminate cycle reaction sequencing kit(Perkin-Elmer公司产品)和ABI 377自动测序仪(Perkin-Elmer公司)测定所有克隆的5'和3'末端的序列。将测定的cDNA序列与已有的公共DNA序列数据库(Genebank)进行比较，结果发现其中一个克隆0196E10的cDNA序列为新的DNA。通过合成一系列引物对该克隆所含的插入cDNA片段进行双向测定。结果表明，0196E10克隆所含的全长cDNA为1711bp(如Seq ID NO:1所示)，从第105bp至1295bp有一个1191bp的开放阅读框架(ORF)，编码一个新的蛋白质(如Seq ID NO:2所示)。我们将此克隆命名为pBS-0196E10，编码的蛋白质命名为人G蛋白44。

实施例2：cDNA 克隆的同源检索

将本发明的人G蛋白44的序列及其编码的蛋白序列，用Blast程序(Basiclocal Alignment search tool)[Altschul, SF et al. J.Mol.Biol.1990;215:403-10]，在Genbank、Swissport等数据库进行同源检索。与本发明的人G蛋白44同源性最高的



基因是一种已知的鼠的G蛋白，其编码的蛋白在Genbank的准入号为Z92773。蛋白质同源结果示于图1，两者高度同源，其相同性为68%；相似性为79%。

实施例3：用RT-PCR方法克隆编码人G蛋白44的基因

5 用胎脑细胞总RNA为模板，以oligo-dT为引物进行逆转录反应合成cDNA，用Qiagene的试剂盒纯化后，用下列引物进行PCR扩增：

Primer1: 5'-GGACGGGCCTGCGTTCTCTCCTCC-3' (SEQ ID NO: 3)

Primer2: 5'-TTTTGTCAAATGTTTTATTGAGT-3' (SEQ ID NO: 4)

Primer1为位于SEQ ID NO: 1的5'端的第1bp开始的正向序列；

10 Primer2为SEQ ID NO: 1的中的3'端反向序列。

扩增反应的条件：在50 μ l的反应体积中含有50mmol/L KCl, 10mmol/L Tris-C1, (pH8.5), 1.5mmol/L MgCl₂, 200 μ mol/L dNTP, 10pmol引物, 1U的Taq DNA聚合酶 (Clontech公司产品)。在PE9600型DNA热循环仪 (Perkin-Elmer公司) 上按下列条件反应25个周期：94°C 30sec; 55°C 30sec; 72°C 2min。在RT-PCR时同时设 β -actin为
15 阳性对照和模板空白为阴性对照。扩增产物用QIAGEN公司的试剂盒纯化，用TA克隆试剂盒连接到pCR载体上 (Invitrogen公司产品)。DNA序列分析结果表明PCR产物的DNA序列与SEQ ID NO: 1所示的1-1711bp完全相同。

实施例4：Northern 印迹法分析人G蛋白44基因的表达：

20 用一步法提取总RNA [Anal. Biochem 1987, 162, 156-159]。该法包括酸性硫氰酸胍苯酚-氯仿抽提。即用4M异硫氰酸胍-25mM柠檬酸钠, 0.2M乙酸钠 (pH4.0) 对组织进行匀浆，加入1倍体积的苯酚和1/5体积的氯仿-异戊醇 (49: 1)，混合后离心。吸出水相层，加入异丙醇 (0.8体积) 并将混合物离心得到RNA沉淀。将得到的RNA沉淀用70%乙醇洗涤，干燥并溶于水中。用20 μ g RNA, 在含20mM 3-(N-吗啉代) 丙磺
25 酸 (pH7.0) -5mM乙酸钠-1mM EDTA-2.2M甲醛的1.2%琼脂糖凝胶上进行电泳。然后转移至硝酸纤维素膜上。用 α -³²P dATP通过随机引物法制备³²P-标记的DNA探针。所用的DNA探针为图1所示的PCR扩增的人G蛋白44编码区序列 (105bp至1295bp)。将³²P-标记的探针 (约2 \times 10⁶cpm/ml) 与转移了RNA的硝酸纤维素膜在一溶液中于42°C杂交过夜，该溶液包含50%甲酰胺-25mM KH₂PO₄ (pH7.4) -5 \times SSC-5 \times Denhardt's溶液和200
30 μ g/ml鲑精DNA。杂交之后，将滤膜在1 \times SSC-0.1%SDS中于55°C洗30min。然后，用Phosphor Imager进行分析和定量。

实施例5: 重组人G蛋白44的体外表达、分离和纯化

根据SEQ ID NO: 1和图1所示的编码区序列, 设计出一对特异性扩增引物, 序列如下:

Primer3: 5'-CCCCATATGATGCCCCCTAAAAAGGGAGGTGATGG-3' (Seq ID No: 5)

5 Primer4: 5'-CATGGATCCTTATTTCTTCTTCGGTTGTTGAGG-3' (Seq ID No: 6)

此两段引物的5'端分别含有NdeI和BamHI酶切位点, 其后分别为目的基因5'端和3'端的编码序列, NdeI和BamHI酶切位点相应于表达载体质粒pET-28b(+) (Novagen公司产品, Cat. No. 69865. 3)上的选择性内切酶位点。以含有全长目的基因的pBS-0196E10质粒为模板, 进行PCR反应。PCR反应条件为: 总体积50 μ l中含pBS-0196E10质粒10pg、引物Primer-3和Primer-4分别为10pmol、Advantage polymerase Mix (Clontech公司产品) 1 μ l。循环参数: 94°C 20s, 60°C 30s, 68°C 2 min, 共25个循环。用NcoI和BamHI分别对扩增产物和质粒pET-28(+)进行双酶切, 分别回收大片段, 并用T4连接酶连接。连接产物转化用氯化钙法大肠杆菌DH5 α , 在含卡那霉素(终浓度30 μ g/ml)的LB平板培养过夜后, 用菌落PCR方法筛选阳性克隆, 并进行测序。挑选序列正确的阳性克隆(pET-0196E10)用氯化钙法将重组质粒转化大肠杆菌BL21(DE3)plySs(Novagen公司产品)。在含卡那霉素(终浓度30 μ g/ml)的LB液体培养基中, 宿主菌BL21(pET-0196E10)在37°C培养至对数生长期, 加入IPTG至终浓度1mmol/L, 继续培养5小时。离心收集菌体, 经超声波破菌, 离心收集上清, 用能与6个组氨酸(6His-Tag)结合的亲和层析柱His.Bind Quick Cartridge (Novagen公司产品)进行层析, 得到了纯化的目的蛋白人G蛋白44。经SDS-PAGE电泳, 在44kDa处得到一单一的条带(图2)。将该条带转移至PVDF膜上用Edams水解法进行N-端氨基酸序列分析, 结果N-端15个氨基酸与SEQ ID NO: 2所示的N-端15个氨基酸残基完全相同。

25 实施例6 抗人G蛋白44抗体的产生

用多肽合成仪(PE公司产品)合成下述人G蛋白44特异性的多肽:

30 NH₂-Met-Pro-Pro-Lys-Lys-Gly-Gly-Asp-Gly-Ile-Lys-Pro-Pro-Ile-COOH (SEQ ID NO: 7)。将该多肽分别与血蓝蛋白和牛血清白蛋白耦合形成复合, 方法参见: Avrameas, et al. *Immunochemistry*, 1969; 6: 43。用4mg上述血蓝蛋白多肽复合物加上完全弗氏佐剂免疫家兔, 15天后再用血蓝蛋白多肽复合物加不完全弗氏佐剂加强免疫一次。采用经15 μ g/ml牛血清白蛋白多肽复合物包被的滴定板做ELISA测定兔血清中抗体的滴度。用蛋白A-Sepharose从抗体阳性的家兔血清中分离总IgG。将多肽



结合于溴化氰活化的Sepharose4B柱上，用亲和层析法从总IgG中分离抗多肽抗体。免疫沉淀法证明纯化的抗体可特异性地与人G蛋白44结合。

实施例 7: 本发明的多核苷酸片段用作杂交探针的应用

- 5 从本发明的多核苷酸中挑选出合适的寡核苷酸片段用作杂交探针有多方面的用途，如用该探针可与不同来源的正常组织或病理组织的基因组或 cDNA 文库杂交以鉴定其是否含有本发明的多核苷酸序列和检出同源的多核苷酸序列，进一步还可利用该探针检测本发明的多核苷酸序列或其同源的多核苷酸序列在正常组织或病理组织细胞中的表达是否异常。
- 10 本实施例的目的是从本发明的多核苷酸 SEQ ID NO: 1 中挑选出合适的寡核苷酸片段用作杂交探针，并用滤膜杂交方法鉴定一些组织中是否含有本发明的多核苷酸序列或其同源的多核苷酸序列。滤膜杂交方法包括斑点印迹法、Southern 印迹法、Northern 印迹法和复印方法等，它们都是将待测的多核苷酸样品固定在滤膜上后使用基本相同的步骤杂交。这些相同的步骤是：固定了样品的滤膜首先用不含探针的杂交缓冲液进行预杂交，以使滤膜上样品的非特异性的结合部位被载体和合成的多聚物所饱和。然后预杂交液被含有标记探针的杂交缓冲液替换，并保温使探针与靶核酸杂交。杂交步骤之后，未杂交上的探针被一系列洗膜步骤除掉。本实施例利用较高强度的洗膜条件（如较低盐浓度和较高的温度），以使杂交背景降低且只保留特异性强的信号。本实施例选用的探针包括两类：第一类探针是完全与本发明的多核苷酸 SEQ ID NO: 1 相同或互补的寡核苷酸片段；第二类探针是部分与本发明的多核苷酸 SEQ ID NO: 1 相同或互补的寡核苷酸片段。本实施例选用斑点印迹法将样品固定在滤膜上，在较高强度的洗膜条件下，第一类探针与样品的杂交特异性最强而得以保留。
- 20 一、 探针的选用
- 25 从本发明的多核苷酸 SEQ ID NO: 1 中选择寡核苷酸片段用作杂交探针，应遵循以下原则和需要考虑的几个方面：
- 1, 探针大小优选范围为 18-50 个核苷酸；
 - 2, GC 含量为 30%-70%，超过则非特异性杂交增加；
 - 3, 探针内部应无互补区域；
 - 30 4, 符合以上条件的可作为初选探针，然后进一步作计算机序列分析，包括将该初选探针分别与其来源序列区域（即 SEQ ID NO: 1）和其它已知的基因组序列及其互补区进行同源性比较，若与非靶分子区域的同源性大于 85%或者有超过 15



个连续碱基完全相同，则该初选探针一般就不应该使用；

5, 初选探针是否最终选定为有实际应用价值的探针还应进一步由实验确定。

完成以上各方面的分析后挑选并合成以下二个探针：

5 探针 1 (probe1), 属于第一类探针, 与 SEQ ID NO: 1 的基因片段完全同源或互补 (41nt):

5'-GAGGTGATGGAATTAACCACCCCAATCATTGGAAGATTT-3' (SEQ ID NO: 8)

探针 2 (probe2), 属于第二类探针, 相当于 SEQ ID NO: 1 的基因片段或其互补片段的替换突变序列 (41nt):

5'-GAGGTATCGGAATTAATAATCGCCCAATCATATCGAGATTT-3' (SEQ ID NO: 9)

10 与以下具体实验步骤有关的其它未列出的常用试剂及其配制方法请参考文献:

DNA PROBES G. H. Keller; M. M. Manak; Stockton Press, 1989 (USA) 以及更常用的分子克隆实验手册书籍如《分子克隆实验指南》(1998 年第二版) [美] 萨姆布鲁克等著, 科学出版社。

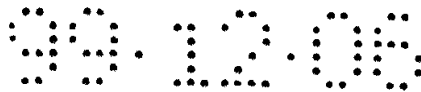
样品制备:

15 1, 从新鲜或冰冻组织中提取 DNA

步骤: 1) 将新鲜或新鲜解冻的正常肝组织放入浸在冰上并盛有磷酸盐缓冲液 (PBS) 的平皿中。用剪刀或手术刀将组织切成小块。操作中应保持组织湿润。2) 以 1000g 离心切碎组织 10 分钟。3) 用冷匀浆缓冲液 (0.25mol/L 蔗糖; 25mmol/L Tris-HCl, pH7.5; 25mmol/L NaCl; 25mmol/L MgCl₂) 悬浮沉淀 (大约 10ml/g)。4) 20 在 4°C 用电动匀浆器以全速匀浆组织悬液, 直至组织被完全破碎。5) 1000g 离心 10 分钟。6) 用重悬细胞沉淀 (每 0.1g 最初组织样品加 1-5ml), 再以 1000g 离心 10 分钟。7) 用裂解缓冲液重悬沉淀 (每 0.1g 最初组织样品加 1ml), 然后接以下的苯酚抽提法。

2, DNA 的苯酚抽提法

25 步骤: 1) 用 1-10ml 冷 PBS 洗细胞, 1000g 离心 10 分钟。2) 用冷细胞裂解液重悬浮沉淀的细胞 (1×10^8 细胞/ml) 最少应用 100ul 裂解缓冲液。3) 加 SDS 至终浓度为 1%, 如果在重悬细胞之前将 SDS 直接加入到细胞沉淀中, 细胞可能会形成大的团块而难以破碎, 并降低的总产率。这一点在抽提 $>10^7$ 细胞时特别严重。4) 加蛋白酶 K 至终浓度 200ug/ml。5) 50°C 保温反应 1 小时或在 37°C 轻轻振摇过夜。6) 30 用等体积苯酚: 氯仿: 异戊醇 (25: 24: 1) 抽提, 在小离心机管中离心 10 分钟。两相应清楚分离, 否则重新进行离心。7) 将水相转移至新管。8) 用等体积氯仿: 异戊醇 (24: 1) 抽提, 离心 10 分钟。9) 将含 DNA 的水相转移至新管。然后进行 DNA



的纯化和乙醇沉淀。

3, DNA 的纯化和乙醇沉淀

步骤: 1) 将 1/10 体积 2mol/L 醋酸钠和 2 倍体积冷 100%乙醇加到 DNA 溶液中, 混匀。在-20°C 放置 1 小时或至过夜。2) 离心 10 分钟。3) 小心吸出或倒出乙醇。

- 5 4) 用 70%冷乙醇 500ul 洗涤沉淀, 离心 5 分钟。5) 小心吸出或倒出乙醇。用 500ul 冷乙醇洗涤沉淀, 离心 5 分钟。6) 小心吸出或倒出乙醇, 然后在吸水纸上倒置使残余乙醇流尽。空气干燥 10-15 分钟, 以使表面乙醇挥发。注意不要使沉淀完全干燥, 否则较难重新溶解。7) 以小体积 TE 或水重悬 DNA 沉淀。低速涡旋振荡或用滴管吹吸, 同时逐渐增加 TE, 混合至 DNA 充分溶解, 每 $1-5 \times 10^6$ 细胞所提取的大约加 1ul。

以下第 8-13 步骤仅用于必须除去污染时, 否则可直接进行第 14 步骤。

- 8) 将 RNA 酶 A 加到 DNA 溶液中, 终浓度为 100ug/ml, 37°C 保温 30 分钟。9) 加入 SDS 和蛋白酶 K, 终浓度分别为 0.5%和 100ug/ml。37°C 保温 30 分钟。10) 用等体积的苯酚: 氯仿: 异戊醇 (25: 24: 1) 抽提反应液, 离心 10 分钟。11) 小心移出水相, 用等体积的氯仿: 异戊醇 (24: 1) 重新抽提, 离心 10 分钟。12) 小心移出水相, 加 1/10 体积 2mol/L 醋酸钠和 2.5 体积冷乙醇, 混匀置-20°C 1 小时。13) 用 70%乙醇及 100%乙醇洗涤沉淀, 空气干燥, 重悬核酸, 过程同第 3-6 步骤。14) 测定 A_{260} 和 A_{280} 以检测 DNA 的纯度及产率。15) 分装后存放于-20°C。

样膜的制备:

- 20 1) 取 4×2 张适当大小的硝酸纤维素膜 (NC 膜), 用铅笔在其上轻轻标出点样位置及样号, 每一探针需两张 NC 膜, 以便在后面的实验步骤中分别用高强度条件和强度条件洗膜。

2) 吸取及对照各 15 微升, 点于样膜上, 在室温中晾干。

- 25 3) 置于浸润有 0.1mol/LNaOH, 1.5mol/LNaCl 的滤纸上 5 分钟 (两次), 晾干置于浸润有 0.5mol/L Tris-HCl (pH7.0), 3mol/LNaCl 的滤纸上 5 分钟 (两次), 晾干。

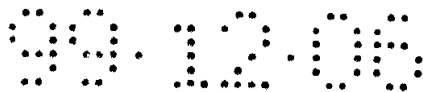
4) 夹于干净滤纸中, 以铝箔包好, 60-80°C 真空干燥 2 小时。

探针的标记

- 30 1) 3μ l Probe (0.10D/10 μ l), 加入 2 μ l Kinase 缓冲液, 8-10 uCi γ - 32 P-dATP+2U Kinase, 以补加至终体积 20 μ l。

2) 37°C 保温 2 小时。

3) 加 1/5 体积的溴酚蓝指示剂 (BPB)。



4) 过 Sephadex G-50 柱。

5) 至有 ^{32}P -Probe 洗出前开始收集第一峰 (可用 Monitor 监测)。

6) 5 滴/管, 收集 10-15 管。

7) 用液体闪烁仪监测同位素量

5 8) 合并第一峰的收集液后即为所需制备的 ^{32}P -Probe (第二峰为游离 γ - ^{32}P -dATP)。

预杂交

将样膜置于塑料袋中, 加入 3-10mg 预杂交液 (10×Denhardt's; 6×SSC, 0.1mg/ml CT DNA (小牛胸腺 DNA。)), 封好袋口后, 68°C 水浴摇 2 小时。

杂交

10 将塑料袋剪去一角, 加入制备好的探针, 封好袋口后, 42°C 水浴摇过夜。

洗膜:

高强度洗膜:

1) 取出已杂交好的样膜。

2) 2×SSC, 0.1%SDS 中, 40°C 洗 15 分钟 (2 次)。

15 3) 0.1×SSC, 0.1%SDS 中, 40°C 洗 15 分钟 (2 次)。

4) 0.1×SSC, 0.1%SDS 中, 55°C 洗 30 分钟 (2 次), 室温晾干。

低强度洗膜:

1) 取出已杂交好的样膜。

20 2) 2×SSC, 0.1%SDS 中, 37°C 洗 15 分钟 (2 次)。

3) 0.1×SSC, 0.1%SDS 中, 37°C 洗 15 分钟 (2 次)。

4) 0.1×SSC, 0.1%SDS 中, 40°C 洗 15 分钟 (2 次), 室温晾干。

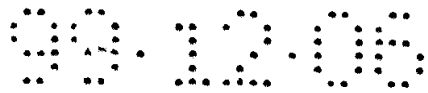
X-光自显影:

25 -70°C, X-光自显影 (压片时间根据杂交斑放射性强弱而定)。

实验结果:

30 采用低强度洗膜条件所进行的杂交实验, 以上四个探针杂交斑放射性强弱没有明显区别; 而采用低强度洗膜条件所进行的杂交实验, 探针 1 的杂交斑放射性强度明显强于其它三个探针杂交斑的放射性强度。因而可用探针 1 定性和定量地分析本发明的多核苷酸在不同组织中的存在和差异表达。

实施例 8 DNA Microarray



基因芯片或基因微矩阵 (DNA Microarray) 是目前许多国家实验室和大制药公司都在着手研制和开发的新技术, 它是指将大量的靶基因片段有序地、高密度地排列在玻璃、硅等载体上, 然后用荧光检测和计算机软件进行数据的比较和分析, 以达到快速、高效、高通量地分析生物信息的目的。本发明的多核苷酸可作为靶 DNA

5

用于基因芯片技术用于高通量研究新基因功能; 寻找和筛选组织特异性新基因特别是肿瘤等疾病相关新基因; 疾病的诊断, 如遗传性疾病。其具体方法步骤在文献中已有多种报道, 如可参阅文献 DeRisi, J. L., Lyer, V. & Brown, P. O. (1997) Science 278, 680-686. 及文献 Helle, R. A., Schema, M., Chai, A., Shalom, D., (1997) PNAS 94: 2150-2155.

10 (一) 点样

各种不同的全长 cDNA 共计 4000 条多核苷酸序列作为靶 DNA, 其中包括本发明的多核苷酸。将它们分别通过 (如实施例_所叙述的方法) PCR 进行扩增, 纯化所得扩增产物后将其浓度调到 500ng/ul 左右, 用 Cartesian 7500 点样仪 (购自美国 Cartesian 公司) 点于玻璃介质上, 点与点之间的距离为 280 μ m。将点样后的玻片

15

1. 潮湿环境中水合 4 小时;
2. 0.2%SDS 洗涤 1 分钟;
3. ddH₂O 洗涤两次, 每次 1 分钟;
4. NaBH₄ 封闭 5 分钟;
5. 95°C 水中 2 分钟;
6. 0.2%SDS 洗涤 1 分钟;
7. ddH₂O 冲洗两次;
8. 凉干, 25°C 储存于暗处备用。

20

25

(二) 探针标记

用一步法分别从正常肝与肝癌中抽提总 mRNA, 并用 Oligotex mRNA Midi Kit (购自 QiaGen 公司) 纯化 mRNA, 通过反转录分别将荧光试剂 Cy3dUTP (5-Amino-propargyl-2'-deoxyuridine 5'-triphosphate coupled to Cy3 fluorescent dye, 购

30

自 Amersham Pharmacia Biotech 公司) 标记正常肝组织的 mRNA, 用荧光试剂 Cy5dUTP (5-Amino-propargyl-2'-deoxyuridine 5'-triphosphate coupled to Cy5 fluorescent dye, 购自 Amersham Pharmacia Biotech 公司) 标记肝癌组织 mRNA, 经

纯化后制备出探针。具体步骤参照及方法 Schena, M., Shalon, D., Heller, R. (1996) Proc. Natl. Acad. Sci. USA. Vol. 93: 10614-10619. Schena, M., Shalon, D., Davis, R. W. (1995) Science. 270. (20): 467-480.

(三) 杂交

- 5 分别将来自以上两种组织的探针与芯片一起在 UniHyb™ Hybridization Solution(购自 TeleChem 公司)杂交液中进行杂交 16 小时, 室温用洗涤液 (1×SSC, 0.2%SDS) 洗涤后用 ScanArray 3000 扫描仪 (购自美国 General Scanning 公司) 进行扫描, 扫描的图象用 Imagen 软件 (美国 Biodiscovery 公司) 进行数据分析处理, 算出每个点的 Cy3/Cy5 比值, 该比值小于 0.5 大于 2 的点被认为是表达有差异的基因。
- 10

实验结果表明, Cy3 signal=9492.81 (取四次实验的平均值), Cy5signal=11478.93 (取四次实验的平均值), Cy3/Cy5=0.8269, 本发明的多核苷酸在以上两种组织中的表达无明显差异。

15

序 列 表

(1) 一般信息:

5

(ii) 发明名称: 人G蛋白44及其编码序列

(iii) 序列数目: 9

(2) SEQ ID NO: 1的信息:

10

(i) 序列特征:

(A) 长度: 1881bp

(B) 类型: 核酸

(C) 链性: 双链

(D) 拓扑结构: 线性

15

(ii) 分子类型: cDNA

(xi) 序列描述: SEQ ID NO: 1:

1 GGACGGGCCTGCGTTCTCTCCTCCTCCTCCCCGCCTCCAGCTGCCGGCAGGACCTTTCT
61 CTCGCTGCCGCTGGGACCCCGTGTCATCGCCCAGGCCGAGCACGATGCCCCCTAAAAAGG
121 GAGGTGATGGAATTAACCACCCCAATCATTGGAAGATTTGGAACCTCACTGAAAATTG
20 181 GTATTGTTGGATTGCCAAATGTTGGGAAATCTACTTTCTTCAATGTGTTAACCAATAGTC
241 AGGCTTCAGCAGAAAACCTCCCGTTCTGCACTATTGATCCTAATGAGAGCAGAGTACCTG
301 TGCCAGATGAAAGGTTTGACTTTCTTTGTCAATACCACAAACCAGCAAGCAAAAATTCCTG
361 CCTTTCTAAATGTGGTGGATATTGCTGGCCTTGTGAAAGGAGCTCACAATGGGCAGGGCC
421 TGGGGAATGCTTTTTTATCTCATATTAGTGCCTGTGATGGCATCTTTCATCTAACACGTG
25 481 CTTTTGAAGATGATGATATCACGCACGTTGAAGGAAGTGTAGATCCTATTCGAGATATAG
541 AAATAATACATGAAGAGCTTCAGCTTAAAGATGAGGAAATGATTGGGCCATTATAGATA
601 AACTAGAAAAGGTGGCTGTGAGAGGAGGAGATAAAAACTAAAACCTGAATATGATATAA
661 TGTGCAAAGTAAAATCCTGGGTTATAGATCAAAAGAAACCTGTTTCGCTTCTATCATGATT
721 GGAATGACAAAGAGATTGAAGTGTGAAATAAACACTTATTTTTGACTTCAAAACCAATGG
30 781 TCTACTTGGTTAATCTTTCTGAAAAAGACTACATTAGAAAGAAAAACAAATGGTTGATAA
841 AAATTAAGAGTGGGTGGACAAGTATGACCCAGGTGCTTTGGTCATTCTTTTAGTGGGG
901 CCTTGAACTCAAGTTGCAAGAATTGAGTGCTGAGGAGAGACAGAAGTATCTGGAAGCGA

961 ACATGACACAAAAGTGCTTTGCCAAAGATCATTAAGGCTGGGTTTGCAGCACTCCAACCTAG
 1021 AATACTTTTTCACTGCAGGCCAGATGAAGTGCCTGCATGGACCATCAGGAAAGGGACTA
 1081 AGGCTCCTCAGGCTGCAGGAAAGATTACACAGATTTTAAAAAGGGATTATTATGGCTG
 1141 AAGTAATGAAATACGAAGATTTTAAAGAGGAAGGTTCTGAAAATGCAGTCAAGGCTGCTG
 5 1201 GAAAGTACAGACAACAAGGCAGAAATTATATTGTTGAAGATGGAGATATTATCTTCTTCA
 1261 AATTTAACACACCTCAACAACCGAAGAAGAAATAAAATTTAGTTATTGCTCAGATAAACA
 1321 TACAACTTCCAAAAGGCATCTGATTTTTAAAAAATTTAAATTTCTGAAAACCAATGCGAC
 1381 AAATAAAGTTGGGGAGATGGGAATCTTTGACAAACAAATTATTTTTATTTGTTTTAAAAAT
 1441 TAAAATACTGTGTACCCCCCCCCCTCCATGAAATGCAGGTTCACTAAATGTGAACAGCTT
 10 1501 TGCTTTTCACGTGATTAAGACCCTACTCCAAATTGTAGAAGCTTTTCAGGAACCATATTA
 1561 CTCTCATGATACTTCATTAATCTCCATCATGTATGCCAAGCCTGACACATTTGACAGTGA
 1621 GGACAATGTGGCTTGCTCCTTTTTGAATCTACAGATAATGCATGTTTTACAGTACTCCAG
 1681 ATGTCTACACTCAATAAAACATTTGACAAAA

15 (3) SEQ ID NO: 2的信息:

(i) 序列特征:

- (A) 长度: 614个氨基酸
- (B) 类型: 氨基酸
- (D) 拓扑结构: 线性

20 (ii) 分子类型: 多肽

(xi) 序列描述: SEQ ID NO: 2:

1	Met	Pro	Pro	Lys	Lys	Gly	Gly	Asp	Gly	Ile	Lys	Pro	Pro	Pro	Ile
16	Ile	Gly	Arg	Phe	Gly	Thr	Ser	Leu	Lys	Ile	Gly	Ile	Val	Gly	Leu
31	Pro	Asn	Val	Gly	Lys	Ser	Thr	Phe	Phe	Asn	Val	Leu	Thr	Asn	Ser
25 46	Gln	Ala	Ser	Ala	Glu	Asn	Phe	Pro	Phe	Cys	Thr	Ile	Asp	Pro	Asn
61	Glu	Ser	Arg	Val	Pro	Val	Pro	Asp	Glu	Arg	Phe	Asp	Phe	Leu	Cys
76	Gln	Tyr	His	Lys	Pro	Ala	Ser	Lys	Ile	Pro	Ala	Phe	Leu	Asn	Val
91	Val	Asp	Ile	Ala	Gly	Leu	Val	Lys	Gly	Ala	His	Asn	Gly	Gln	Gly
106	Leu	Gly	Asn	Ala	Phe	Leu	Ser	His	Ile	Ser	Ala	Cys	Asp	Gly	Ile
30 121	Phe	His	Leu	Thr	Arg	Ala	Phe	Glu	Asp	Asp	Asp	Ile	Thr	His	Val
136	Glu	Gly	Ser	Val	Asp	Pro	Ile	Arg	Asp	Ile	Glu	Ile	Ile	His	Glu
151	Glu	Leu	Gln	Leu	Lys	Asp	Glu	Glu	Met	Ile	Gly	Pro	Ile	Ile	Asp
166	Lys	Leu	Glu	Lys	Val	Ala	Val	Arg	Gly	Gly	Asp	Lys	Lys	Leu	Lys
181	Pro	Glu	Tyr	Asp	Ile	Met	Cys	Lys	Val	Lys	Ser	Trp	Val	Ile	Asp
35 196	Gln	Lys	Lys	Pro	Val	Arg	Phe	Tyr	His	Asp	Trp	Asn	Asp	Lys	Glu
211	Ile	Glu	Val	Leu	Asn	Lys	His	Leu	Phe	Leu	Thr	Ser	Lys	Pro	Met
226	Val	Tyr	Leu	Val	Asn	Leu	Ser	Glu	Lys	Asp	Tyr	Ile	Arg	Lys	Lys



241 Asn Lys Trp Leu Ile Lys Ile Lys Glu Trp Val Asp Lys Tyr Asp
256 Pro Gly Ala Leu Val Ile Pro Phe Ser Gly Ala Leu Glu Leu Lys
271 Leu Gln Glu Leu Ser Ala Glu Glu Arg Gln Lys Tyr Leu Glu Ala
286 Asn Met Thr Gln Ser Ala Leu Pro Lys Ile Ile Lys Ala Gly Phe
5 301 Ala Ala Leu Gln Leu Glu Tyr Phe Phe Thr Ala Gly Pro Asp Glu
316 Val Arg Ala Trp Thr Ile Arg Lys Gly Thr Lys Ala Pro Gln Ala
331 Ala Gly Lys Ile His Thr Asp Phe Glu Lys Gly Phe Ile Met Ala
346 Glu Val Met Lys Tyr Glu Asp Phe Lys Glu Glu Gly Ser Glu Asn
361 Ala Val Lys Ala Ala Gly Lys Tyr Arg Gln Gln Gly Arg Asn Tyr
10 376 Ile Val Glu Asp Gly Asp Ile Ile Phe Phe Lys Phe Asn Thr Pro
391 Gln Gln Pro Lys Lys Lys

(4) SEQ ID NO: 3的信息

(i) 序列特征

- 15 (A) 长度: 24碱基
(B) 类型: 核酸
(C) 链性: 单链
(D) 拓扑结构: 线性

(ii) 分子类型: 寡核苷酸

20 (xi) 序列描述: SEQ ID NO: 3:
GGACGGGCCTGCGTTCTCTCCTCC

24

(5) SEQ ID NO: 4的信息

(i) 序列特征

- 25 (A) 长度: 23碱基
(B) 类型: 核酸
(C) 链性: 单链
(D) 拓扑结构: 线性

(ii) 分子类型: 寡核苷酸

30 (xi) 序列描述: SEQ ID NO: 4:
TTTTGTCAAATGTTTTATTGAGT

23

(6) SEQ ID NO: 5的信息

(i) 序列特征

- 35 (A) 长度: 35碱基



(B) 类型: 核酸

(C) 链性: 单链

(D) 拓扑结构: 线性

(ii) 分子类型: 寡核苷酸

5 (xi) 序列描述: SEQ ID NO : 5:

CCCCATATGATGCCCCCTAAAAAGGGAGGTGATGG

35

(7) SEQ ID NO: 6的信息

(i) 序列特征

10 (A) 长度: 33碱基

(B) 类型: 核酸

(C) 链性: 单链

(D) 拓扑结构: 线性

(ii) 分子类型: 寡核苷酸

15 (xi) 序列描述: SEQ ID NO : 6:

CATGGATCCTTATTTCTTCTTCGGTTGTTGAGG

33

(8) SEQ ID NO: 7的信息:

(i) 序列特征:

20 (A) 长度: 15个氨基酸

(B) 类型: 氨基酸

(D) 拓扑结构: 线性

(ii) 分子类型: 多肽

(xi) 序列描述: SEQ ID NO: 7:

25 Met-Pro-Pro-Lys-Lys-Gly-Gly-Asp-Gly-Ile-Lys-Pro-Pro-Pro-Ile

15

(9) SEQ ID NO: 8的信息

(i) 序列特征

(A) 长度: 41碱基

30 (B) 类型: 核酸

(C) 链性: 单链

(D) 拓扑结构: 线性

(ii) 分子类型: 寡核苷酸

(xi) 序列描述: SEQ ID NO: 4:

GAGGTGATGGAATTAACCACCCCAATCATTGGAAGATT

41

5 (10) SEQ ID NO: 9的信息

(i) 序列特征

(A) 长度: 41碱基

(B) 类型: 核酸

(C) 链性: 单链

10 (D) 拓扑结构: 线性

(ii) 分子类型: 寡核苷酸

(xi) 序列描述: SEQ ID NO: 9:

GAGGTATCGGAATTAATAATCGCCCAATCATATCGAGATT

41



说明书附图

相同性 = 270/396 (68%), 相似性 = 316/396 (79%)

5 Query: 1 MPPKGGDGKPPPIIGRFGTSLKIGIVGLPNVGKSTFFNVLNSQASAENFPFCTIDPN 60
MPPKK KP +IGR GT+LK+GI+GLPNVGKSTFFNVL S+A AENFPFCTIDPN
Sbjct: 1 MPPKKNVVEEKPA-LIGRLGTNLKVGILGLPNVGKSTFFNVLTKSEAQAENFPFCTIDPN 59

10 Query: 61 ESRVVPDERFDFLCQYHKPASKIPAFLNVVDIAGLVKGAHNGQGLGNAFLSHISACDGI 120
ESRV V D+RFD+L ++KPASK+PAFLNV DIAGLVKGA GQGLGNAFLSH+SACD +
Sbjct: 60 ESRVAVQDDRFDLWLVNHYKPAFLNVTDIAGLVKGASEGQGLGNAFLSHVSACDAL 119

15 Query: 121 FHLTRAFEDDDITHVEGSVDPIRDIEIHEELQLKDEEMI-GPIIDKLEKVAVRGGDKKL 179
FHL RAF+DDD+THVEG VDP+RD+EII EL KD + I GP+ DK+EK+ R DK
Sbjct: 120 FHLRAFDDDDVTHVEGEVDPVRDLEIISNELFAKDLQFIDGPL-DKVEKLFTRANDKTK 178

20 Query: 180 KPEYDIMCKVKSVIDQKKPVRFYHDWNDKEIEVLNKHFLFTSKPMVYLVNLSEKDYIRK 239
K EYD + +VK +++KPPVR WN+KEIE+LNKHLFLT+KP+VYLVNLSEKDYIRK
Sbjct: 179 KIEYDTLVRVKK-CLEEKPPVR-QELWNEKEIEILNKHFLFTAKPIVYLVNLSEKDYIRK 236

25 Query: 240 KNKWLKIKWVDKYDPGALVIPFSGALELKLQELSAEERQKYLEANMTQSALPKI KAG 299
KNKWL KIK W+D D GA++IPFSGA ELKL ++ +ERQKYL+ S L KI+ G
Sbjct: 237 KNKWLPKIKAWIDTNDAGAVLIPFSGAFELKLLDMPEDERQKYLKEQGVTSNLDKIVHTG 296

30 Query: 300 FAALQLEYFFTAGPDEVRAWTIRKGTKAPQAAGKIHTDFEKG FIMAEVMKYEDFKKEGSE 359
+ ALQLEYFFT+G DEV+AWTI+ GT AP+AAG+IHTDFEKG FIMAEVMK D E G E
Sbjct: 297 YKALQLEYFFTSGEDEVKAWTIQVGTAPKAAGRIHTDFEKG FIMAEVMKVADLIELGDE 356

Query: 360 NAVKAAGKYRQQGRNYIVEDGDI IFFKFNT PQ--QP KKK 396
KA GK YRQQG+ YIV+DGD+IFFKFN Q KKK
Sbjct: 357 AKCKAGGKYRQQGKTYIVQGDV IFFKFNAGAGLQAKKK 395

图 1

99.12.06

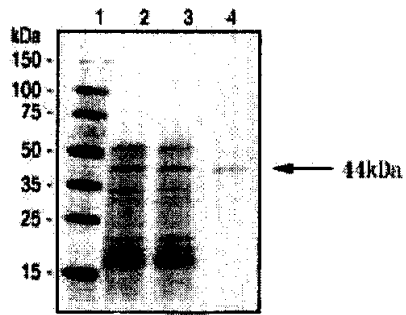


图 2

5