

[19] 中华人民共和国国家知识产权局

[51] Int. Cl⁷

C12P 21/02

C12N 15/12

C12N 15/63 A61K 38/16

[12] 发明专利说明书

[21] ZL 专利号 94117067.5

[45] 授权公告日 2001 年 3 月 7 日

[11] 授权公告号 CN 1062908C

[22] 申请日 1994. 10. 14 [24] 颁证日 2000. 11. 25

[21] 申请号 94117067.5

[30] 优先权

[32] 1993. 10. 14 [33] JP [31] 280505/1993

[73] 专利权人 小野药品工业株式会社

地址 日本大阪府

[72] 发明人 本庶佑 白水伦生 多田秀明

[56] 参考文献

BIOCHEMISTRY VOL 29 1990. 1. 1

PROC NATL ACAD SCI VOL 77 NO12 1980. 1. 1

审查员 孙广秀

[74] 专利代理机构 中国专利代理(香港)有限公司

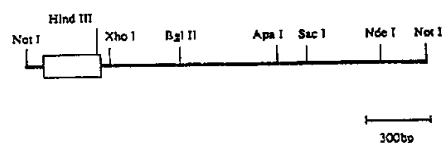
代理人 谭明胜 姜建成

权利要求书 2 页 说明书 35 页 附图页数 3 页

[54] 发明名称 新多肽和编码它们的脱氧核糖核酸

[57] 摘要

在原 B 细胞中生产并分泌本发明多肽,因此可将其用于与生血细胞增殖不足或 不正常、神经增强或抑制、免疫增强和抑制有关的疾病如炎性疾病(类风湿性 关节炎、溃疡性结肠炎等),骨髓移植后的生血干细胞减少,化疗后的白细胞 减少,血小板减少,B 淋巴细胞减少和 T 淋巴细胞减少,贫血,传染性疾病,癌症,白细胞增多,爱滋病,神经变性疾病(阿尔茨海默症,多发性硬化等),预防或治疗神经损伤、骨代谢紊乱(骨质疏松等)或组织修复。制备本发明 多肽时,可用本发明 DNA 作为重要且必需的模板,期望其具有各种用途,或 用于基因疾病诊断和治疗。



Box 是指开放的阅读框

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

权 利 要 求 书

1. 一种含有 SEQ ID No: 1 所示氨基酸序列的基本上纯化的基质衍生的因子 1 α 多肽、所述多肽的具有基质衍生的因子 1 α 生物活性的同系物，或所述多肽的具有基质衍生的因子 1 α 生物活性的片段，或所述片段的具有基质衍生的因子 1 α 生物活性的同系物。

2. 根据权利要求 1 的基质衍生的因子 1 α 多肽，其中所述同系物与 SEQ ID No: 1 所示氨基酸序列有至少 90%的同源性。

3. 根据权利要求 1 的多肽，其中所述片段含有至少 20 个氨基酸。

4. 根据权利要求 1 的多肽，包含有 SEQ ID No: 1 所示的氨基酸序列。

5. 编码权利要求 1 多肽的 DNA。

6. 根据权利要求 5 的 DNA，含有 SEQ ID No: 2 所示的核苷酸序列或与 SEQ ID No: 2 所示的核苷酸序列有至少 90%同源性的片段。

7. 根据权利要求 5 的 DNA，含有 SEQ ID No: 3 所示的核苷酸序列或与 SEQ ID No: 3 所示的核苷酸序列有至少 90%同源性的片段。

8. 一种复制和表达载体，含有权利要求 5 到 7 任一权项所述的 DNA。

9. 用权利要求 8 的复制及表达载体转化或转染的宿主细胞。

10. 一种生产多肽的方法，包括在可有效表达权利要求 1—4 的多肽的条件下，培养权利要求 9 的宿主细胞。

11. 一种抗权利要求 1—4 的多肽的单克隆或多克隆抗体。

12. 一种药物组合物，含有权利要求 1—4 的多肽或权利要求 11 的抗体及药学可接受的稀释剂和/或载体。

13. 一种含有 SEQ ID No: 5 所示氨基酸序列的基本上纯化的基质衍生的因子 1 β 多肽、所述多肽的具有基质衍生的因子 1 β 生物活性的同系物，或所述多肽的具有基质衍生的因子 1 β 生物活性的片段，或所述片段的具有基质衍生的因子 1 β 生物活性的同系物。

14. 根据权利要求 13 的多肽, 其中所述同系物与 SEQ ID No: 1 所示氨基酸序列有至少 90% 的同源性。

15. 根据权利要求 13 的多肽, 其中所述片段含有至少 20 个氨基酸。

16. 根据权利要求 13 的多肽, 含有 SEQ ID No: 5 所示的氨基酸序列。

17. 编码权利要求 13 多肽的 DNA。

18. 根据权利要求 17 的 DNA, 含有 SEQ ID No: 6 所示的核苷酸序列或与 SEQ ID No: 6 所示的核苷酸序列有至少 90% 同源性的片段。

19. 根据权利要求 17 的 DNA, 含有 SEQ ID No: 7 所示的核苷酸序列或与 SEQ ID No: 7 所示的核苷酸序列有至少 90% 同源性的片段。

20. 一种复制及表达载体, 含有权利要求 17 到 19 中任一权项所述的 DNA。

21. 用权利要求 20 的复制及表达载体转化或转染的宿主细胞。

22. 一种生产多肽的方法, 包括在可有效表达权利要求 13-16 的多肽的条件下, 培养权利要求 21 的宿主细胞。

23. 一种抗权利要求 13-16 的多肽的单克隆或多克隆抗体。

24. 一种药物组合物, 含有权利要求 13-16 的多肽或权利要求 23 的抗体以及药理学可接受的稀释剂和/或载体。

25. 一种基本上纯化的成熟的基质衍生的因子 1 α 多肽, 其具有 SEQ ID No: 4 位置 1 的 Lys 至位置 68 的 Lys 的氨基酸序列。

26. 一种基本上纯化的成熟的基质衍生的因子 1 β 多肽, 其具有 SEQ ID No: 8 位置 1 的 Lys 至位置 72 的 Met 的氨基酸序列。

说明书

新多肽和编码它们的脱氧核糖核酸

本发明涉及由人原B细胞系生产的新多肽及编码它们的DNA。

本发明涉及由生血细胞生产的新多肽和编码它们的DNA。已经知道许多种类来自生血细胞的生长和分化因子，如白细胞介素(IL)。

该事实说明，除已发现的已知的分泌性因子外，也可从中分泌具有相似或新功能的因子。

本发明人已将其注意力放在这方面并试图找出由生血细胞生产的新因子(多肽)。本发明人通过使用小鼠SDF-1(基质衍生的因子1；描述于日本专利申请No. 522098)cDNA作为探针的交叉杂交进行筛选，因此得到由人原B细胞生产的人SDF-1(2种， α 和 β)，从而完成本发明。

用计算机对与本发明多肽具有相同或高度同源序列的多肽和编码它们的DNA进行了检索，但未发现这样的多肽和DNA。因此，已证明，本发明多肽和为其编码的DNA是新的。

- 1) 具有SEQ ID No. 1所示氨基酸序列的多肽，
- 2) 编码上述(1)多肽的DNA，
- 3) 具有SEQ ID No. 2所示核苷酸序列的DNA，和
- 4) 具有SEQ ID No. 3所示核苷酸序列的DNA，
- 5) 具有SEQ ID No. 5所示氨基酸序列的多肽，
- 6) 编码上述(5)的多肽的DNA，

- 7) 具有 SEQ ID No. 6 所示核苷酸序列的 DNA, 和
8) 具有 SEQ ID No. 7 所示核苷酸序列的 DNA。

本发明涉及具有 SEQ ID No. 1 或 5 所示氨基酸序列的基本上纯化的形式的多肽, 其同系物或该序列的片段或片段的同系物, 及编码所述多肽的 DNA。更具体地, 本发明涉及具有 SEQ ID No. 2 或 3, 和 6 或 7 所示核苷酸序列的 DNA, 及具有可选择性地与 SEQ ID No. 2 或 3, 和 6 或 7 所示核苷酸序列杂交的片段的 DNA。

通常, 基本上纯化形式的 SEQ ID No. 1 或 5 多肽组成了制剂中的多肽, 该制剂中 90% 以上如 95%, 98% 或 99% 的多肽是 SEQ ID No. 1 或 5。

SEQ ID No. 1 或 5 的多肽同系物与 SEQ ID No. 1 的多肽通常有至少 70%, 优选至少 80 或 90%, 最佳至少为 95% 的同源性, 即至少 20, 优选的至少 30, 例如 40, 60 或 80 个以上邻接氨基酸区域内相同, 所述多肽同系物是指下文中所述本发明多肽。

通常, SEQ ID No. 1 或 5 的片段或其同系物的长度至少为 10, 优选的至少为 15, 如 20, 25, 30, 40, 50 或 60 个氨基酸, 并且包括在本文所用术语“本发明的多肽”中。

能选择性与 SEQ ID No. 2 或 3, 和 6 或 7 的 DNA 杂交的 DNA, 与 SEQ ID No. 2 或 3 的 DNA 有至少 70%, 优选至少 80% 或 90%, 最佳至少为 95% 的同源性, 即其相同性覆盖了至少 20, 优选的至少 30, 如 40, 60 或 100 或更多个邻接核苷酸的区域。所述 DNA 也包括在术语“本发明的 DNA”中。

SEQ ID No. 2 或 3, 6 或 7 DNA 的片段的长度为至少 15, 优选的至少 20, 例如 25, 30 或 40 个核苷酸, 并且也包括在本文所用的术语“本发明的 DNA”中。

本发明进一步的实施方案将提供含有本发明 DNA 的复制和表达载体。所述载体可以是例如质粒, 病毒或噬菌体载体, 它们配有复制原点, 选择性地具有所述 DNA 表达所需的一个启动子, 及任选地含有启动子调节子。所述载体可含有一个或多个选择性标记基因, 例如氨苄青霉素抗性基因。可在体外使用所述载体, 例如生产相应于所述 DNA 的 RNA, 或用于转染或转化宿主细胞。

本发明的另一个实施方案提供用所述载体转化或转染的宿主细胞, 所述载体用于本发明 DNA 的复制和表达, 包括 SEQ ID No. 2 或 3, 和 6 或 7 或其开放阅读框架。可以选择与载体相容的细胞, 可以是如细菌, 酵母, 昆虫或哺乳动物的。

本发明的再一实施方案提供生产多肽的方法, 包括在有效表达本发明多肽的条件下, 培养本发明的宿主细胞。此外, 优选的是, 在表达本发明多肽, 然后由宿主细胞生产的条件下, 完成所述方法。

也可将本发明的 DNA 以反义方向插入上述载体, 以便生产反义 RNA。也可以用合成的方法生产反义 RNA。可以在控制细胞中本发明多肽水平的方法中使用所述的反义 RNA。

本发明也提供抗本发明多肽的单克隆或多克隆抗体。本发明还提供生产抗本发明多肽的单克隆或多克隆抗体的方法。通过用本发明多肽或其片段作为免疫原的常规杂交瘤技术, 制备单克隆抗体。也可以用常规方法制备多克隆抗体, 包括用本发明多肽接种宿主动物, 如大鼠或兔, 然后回收免疫血清。

本发明也提供药物组合物，含有与药理学可接受的稀释剂和/或载体结合在一起的本发明多肽或其抗体。

本发明多肽包括缺失了部分氨基酸序列的多肽（例如，多肽仅有在SEQ ID No. 1或5所示氨基酸序列中，表现生物活性所必需的序列），由其它氨基酸替代了部分氨基酸序列的多肽（如，由具有相似特性的氨基酸替代的那些）和将其它氨基酸加入或插入其部分氨基酸序列后得到的多肽，以及具有SEQ ID No. 1或5所示氨基酸序列的那些。

如熟知的，有一种到6种密码子编码一种氨基酸（例如，已知有一种密码子编码Met，6种密码子编码亮氨酸（Leu））。因此，可以改变DNA的核苷酸序列，以编码具有相同氨基序列的多肽。

在（2）和（6）中说明的本发明的DNA包括一组编码SEQ ID No. 1和5所示多肽的各种核苷酸序列。通过改变核苷酸序列，提高多肽的生产率成为可能。

（3）和（7）中说明的DNA是（2）和（6）中所示DNA的具体实施，并且是天然形式的序列。

（4）和（8）中所示的DNA表示带有非转译区的（3）和（7）中说明的DNA序列。

信号肽是一个疏水区，紧位于转译起始氨基酸Met的下游。据推测，在本发明多肽中的信号肽位于SEQ ID No. 1或5所代表的氨基酸序列中从1位Met到21位Gly的区域。对于生物活性表达所必需的区域相当于缺失了信号肽的SEQ ID No. 1和5的氨基酸序列部分，即成熟蛋白质部分。因此，信号肽决不与活性有关。

可以按下列方法制备含有SEQ ID No. 3或7中所示核

核苷酸序列的DNA，即：

- (I) 从生产本发明多肽的细胞 (如人原 B 细胞系) 中分离 mRNA,
- (II) 从由此得到的 mRNA 制备第一条链 (单链 cDNA), 接着制备第二条链 (双链 cDNA) (cDNA 的合成)。
- (III) 将由此得到的 cDNA 插入到适当的噬菌体载体中,
- (IV) 将重组噬菌体转染至宿主细胞 (构建 cDNA 文库),
- (V) 用小鼠 S D F - 1 cDNA 作为探针, 用噬菌斑杂交筛选 cDNA 文库,
- (VI) 制备含有所得阳性克隆的噬菌体 DNA, 接着将切出的 cDNA 亚克隆至质粒载体中, 然后制备限制酶图谱。
- (VII) 确定用限制酶切出的各片段的核苷酸序列, 接着配成全长序列。

详细地说明, 在用适当的刺激剂 (如 IL - 1 等) 刺激人原 B 细胞系后或不刺激的情况下, 按 Okayama, H 等人的方法 (见 Enzymology, Vol. 154, p3, 1987 中描述的方法) 完成步骤 (I)。

分泌本发明多肽的细胞的实例优选的是人原 B 细胞系 FLEB14。该人细胞系 FLEB14^o 可以由京都大学, 药物学校, 药物化学系第 1 课题组提供。

步骤 (II), (III) 和 (IV) 是制备 cDNA 文库的连续步骤, 主要按 Glubler & Hoffman (Gene, Vol. 25, pp.263, 1983) 的方法, 稍微加以改动, 来完成。

许多已知的质粒载体 (例如 p B 3 2 2, pBluescript 等), 噬菌体载体 (例如 λ gt10, λ DASH II 等) 可作为步骤 (III) 中使用的载体的实例, 噬菌体载体 λ gt10 (43.3kbp, Stratagene) 是优选的。

步骤 (IV) 中所用的宿主细胞优选的是大肠杆菌 NM514

(Stratagene)。

可以按 Molecular Cloning 中描述的方法(由 Sambrook , Fritsh , E. F. 和 Maniatis , T., 编写, 1989 由 Cold Spring Habor Laboratory Press 出版)完成步骤(V)和(VI)。

可以按 Maxam - Gilbert 方法或双脱氧终止方法完成步骤(VII)中的测序。

必须确认得到的cDNA覆盖了全部或几乎全部长度的完整mRNA。可以通过使用cDNA作为探针的 Northern 分析(参见Molecular Cloning)完成这些确认。

如果从杂交带得到的mRNA的大小与cDNA的大小几乎相同, 则认为cDNA几乎是全长的。

一旦确定了SEQ ID No. 2, 3, 6, 7中所示的核苷酸序列, 则可通过化学合成, PCR方法或用本发明的DNA片段作为探针进行杂交, 得到本发明的DNA。此外, 通过将插入有本发明DNA的载体DNA转化到适当的宿主中, 然后培养转化体, 可以得到所需量的本发明DNA。

可经下列方法制备本发明多肽(SEQ ID No. 1或5中所示):

- (1) 从生物体或培养细胞中分离并纯化,
- (2) 化学合成, 或
- (3) 用生物技术,

优选的用(3)中描述的方法。

用生物技术制备多肽时, 表达系统的实例是例如细菌、酵母、昆虫细胞和哺乳动物细胞表达系统。

例如，通过将起始密码子（A T G）加至编码成熟蛋白D N A的5'末端，将由此得到的D N A连至适当启动子（如 trp启动子，lac启动子，IPL启动子，T 7启动子等）的下游，然后，将其插入在大肠杆菌菌株中发挥功能的载体（如pBR322，pUC18，pUC19等）中以制备表达载体，从而完成在大肠杆菌中的表达。

若使用细胞的信号肽（如pelB的信号肽），也可在同质中生产所需的多肽。此外，也很容易生产与其他多肽的融合蛋白。

此外，通过例如，将S E Q I D N o . 3或6中所示的DNA插入适当载体（如反录病毒载体，乳头状瘤病毒载体，牛痘病毒载体，S V 4 0载体，等）中适当启动子（如S V 4 0启动子，L T R启动子，金属硫蛋白启动子等）的下游，以得到表达载体，用由此得到的表达载体转染适当的哺乳动物细胞（如猴C O S - 7细胞，中国仓鼠C H O细胞，小鼠人细胞等），然后在适当的培养基中，培养转化体以便在培养基中得到所需的多肽，从而可完成在哺乳动物细胞中的表达。用常规的生化方法可分离并纯化由此得到的多肽。

在原B细胞中生产并分泌本发明的多肽，这样可将其用于治疗与生血细胞增殖不足或不正常、神经增强或抑制、免疫增强和抑制有关的疾病，例如发炎症性疾病（类风湿性关节炎，溃疡性结肠炎），骨髓移植后的生血干细胞减少症（hematopoietic stemcytopenia），化疗后的白细胞减少，血小板减少，B淋巴细胞减少及T细胞减少，贫血，传染性疾病，癌症，白细胞增多，爱滋病，神经变性性疾病（阿尔茨海默症，多发性硬化等），预防并治疗神经元损伤，预防或治疗骨代谢紊乱（骨质疏松症，等）或组织修复。

在上述活性中，在实验室试验中，可以确定，小鼠S D F - 1 α

刺激小鼠髓样前体细胞系 D A 1 G 的增殖。从而表明人 S D F - 1 α 也有相同的活性。

另外，可以用抗本发明多肽的多克隆或单克隆抗体确定生物体中所述多肽的量，并因此可用于研究所述多肽与疾病间的关系的目的，或用于诊断疾病的目的等。用所述多肽或其片段作为抗原，通过常规方法可以制备其多克隆和单克隆抗体。

在本发明多肽的制备中，可以用本发明 D N A 作为一种重要且必需的模板，期望其拥有各种用途或用于基因疾病的诊断和治疗（基因缺陷性疾病的治疗和通过抑制经反义 D N A (R N A) 表达的多肽进行治疗等）。另外，用本发明的 D N A 作为探针可以分离基因组 D N A。相似地，可以分离人体或其它生物体的与本发明 D N A 有高度同源性的基因。

在原 B 细胞中生产并分泌本发明的多肽，因此，可将其用于与生血细胞增殖不足或不正常；神经元增强或抑制，免疫增强和抑制有关的疾病，例如，发炎性疾病（类风湿性关节炎，溃疡性结肠炎等），骨髓移植后的生血干细胞减少，化疗后的白细胞减少，血小板减少，B 淋巴细胞减少及 T 淋巴细胞减少，贫血，传染性疾病，癌症，白细胞增多，爱滋病，神经变性性疾病（阿尔茨海默症，多发性硬化，等），预防或治疗神经元损伤，预防或治疗骨代谢紊乱（骨质疏松症，等）或组织修复。

通常通过口服或肠胃外施用，优选的经口服、静脉内或心室内，全身地，或部分地施用本发明的多肽。

根据年龄、体重、症状，所需的治疗效果，施用途径和治疗期等确定施用剂量。在成年人中，每人每剂的口服剂量通常在 100 μ g 和

100mg之间，可达每天数次，肠胃外施用为10 μ g和100mg之间，可达每天数次。

如上文提到的，根据各种条件确定使用的剂量。因此，也存在使用低于或高于上述剂量范围的情况。

可以以固体组合物，液体组合物或其它组合物经口服，以注射剂、搽剂或栓剂等经肠胃外施用，来施用本发明的化合物。

用于口服的固体组合物包括压片剂，丸剂，胶囊，可分散粉末、颗粒。胶囊包括软胶囊和硬胶囊。

在所述组合物中，将一种或多种活性化合物与至少一种惰性稀释剂（如乳糖，甘露醇，葡萄糖，羟丙基纤维素，微晶纤维素、淀粉、聚乙烯吡啶烷酮，硅铝酸镁等）混合。组合物也可含有，象通常实践的那样，除惰性稀释剂以外的附加物质：如润滑剂（如硬脂酸镁等），崩解剂（如纤维素甘醇酸钙等），稳定剂（如血清白蛋白，乳糖，等），及溶解助剂（如精氨酸，天冬氨酸等）。

如果需要，可以用一种胃或肠物质胶膜（如糖，明胶，羟丙基纤维素或邻苯二甲酸羟丙基甲基纤维素等）或两种以上胶膜包被片剂或丸剂。另外，包衣可包括在可吸收物质如明胶胶囊内的外壳。

其它口服组合物包括可用已知方法制备并含有一种或多种活性化合物的喷剂组合物。喷剂组合物可含有除惰性稀释剂外的附加物质：如稳定剂（亚硫酸钠等），等渗缓冲剂（氯化钠，柠檬酸钠，柠檬酸等）。为了制备所述的喷剂组合物，如可以用美国专利No. 2,868,691或3,095,355（全文引入本文作为参考）中描述的方法。

注射剂可以含有除惰性稀释剂外的附加成分：如防腐剂，湿润剂，乳化剂，分散剂，稳定剂（如人血清白蛋白，乳糖，等），和助剂，

如溶解助剂（精氨酸，天冬氨酸，等），

例如，通过挡住细菌的滤器过滤，或在组合物中掺入稳定剂或照射将组合物灭菌。如，经冻干也可以将它们加工成无菌固体组合物的形式，在即将使用前，将其溶解于用于注射的无菌水或一些其它无菌稀释剂中。

肠胃外施用的其它组合物包括含有一种或多种活性化合物并可用已知方法制备的外用液体，皮肤搽剂（软膏等），经肛门施用的栓剂和阴道栓剂。

实施例

下列实施例是为了说明而不是限制本发明。

实施例 1：人细胞系FLEB14的Northern分析：

将人原 B 细胞系FLEB14细胞（参见 Katamine S. 等人，Nature 309, 369 (1984)）匀浆。将匀浆物与寡聚-d T 纤维素一起保温。洗涤后，洗脱 poly(A) RNA (Vennstorm, B 等人, Cell, 28, 135 (1982))。在1.0%琼脂糖凝胶中电泳 1 μ g 的多聚(A) RNA，然后吸印到硝酸纤维素膜上。

在 39 °C，50%甲酰胺，5 $\times$ SSC，0.1%SDC，0.1%SDS，5 $\times$ Denhardt's，0.1mg/ml鲑精DNA中，将该膜与³²P-标记的小鼠 SDF-1（在日本专利申请 No. 5-22098中描述为 SEQ ID No. 3；在 SEQ ID No. 9中列出了该序列；由于从小鼠发现了另一种 SDF-1，所以现在称该因子为 SDF-1 α ）cDNA 杂交，在 50 °C，用 0.3M NaCl，30 mM 柠檬酸钠，0.1%SDS 洗涤，然后放射自显影。3.5kb 和 1.9kb mRNA 被杂交。

实施例 2：由人原 B 细胞系 mRNA 制备 cDNA

用常规方法 (参见 Molecular Cloning : Sambrook ,J., Fritsh . E. F., & Maniatis , T, Cold Spring Harbor Laboratory Press (1989)) 从人原 B 细胞系 FLEB14 细胞构建 cDNA 文库。用 Time Saver cDNA 合成盒 (Pharmacia) 合成 cDNA。

用反转录酶和寡聚 - d T 引物, 从 FLEB14 多聚 (A) - R N A (5 μ g) 合成第一条链。然后用 D N A 聚合酶 I , 合成双链 cDNA。

将 cDNA 与 EcoRI - NotI 接头连接 :

AATTCGCGCCGCT

GCGCCGGCGAp

然后磷酸化, 用玻璃粉 (Geneclean II D N A 纯化盒, 可从 Biol01 得到) 从 0.8% 琼脂糖凝胶中回收长于 800bp 的 cDNA。

实施例 3 : 制备 cDNA 文库并交叉杂交

将得到的 cDNA 与含有用磷酸酶处理过的 EcoRI 臂的 λ gt10 噬菌体载体 (可从 Stratagene 得到) 连接。

用体外包装药盒 LAMDA INN (可从 Nihon gene 得到) 的方法完成体外包装。用重组噬菌体转染宿主大肠杆菌 NM514 (可从 Stratagene 得到)。得到含 1×10^8 噬菌斑的 cDNA 文库。将 1×10^8 个 cDNA 文库的 λ gt10 噬菌斑转染至硝酸纤维素膜上。在 39 $^{\circ}$ C, 50% 甲酰胺, 5 $\times$ SSC, 0.1%SDS, 5 $\times$ De haldt's , 0.1mg/ml 鲑精 D N A 中, 将该膜与 32 P - 标记的小鼠 S D F - 1 α cDNA (S E Q I D N o . 9 中所示, 与实施例 1 中所用的相同) 杂交, 在 50 $^{\circ}$ C, 用 0.3M NaCl , 30mM 柠檬酸钠, 0.1%SDS 洗涤并放射自显影。得到 40 个阳性克隆。

实施例 4 : 分离阳性克隆

用常规方法（参见 Cell Technology Experimental Protocol. pp8, Shuujun - sha 出版），从 9 个阳性克隆中制备噬菌体 DNA。用 NotI 消化噬菌体 DNA。用琼脂糖凝胶电泳测量插入 cDNA 的长度。8 个克隆的长度是 1.9kb，1 个克隆的长度是 3.5kb。因此认为这两种类型的克隆几乎是来自 Northern 分析结果的、全长的人 SDF-1 α 和 SDF-1 β cDNA。

从 1.9kb 的 8 个克隆中，捡出一个克隆的 cDNA，3.5kb 长的克隆用 NotI 消化，然后经琼脂糖电泳，切出片段，将其亚克隆至质粒 pBluescript 的 NotI 位点。

实施例 5：制备限制酶图谱并测序

制备人 SDF-1 (1.9kb) 的限制酶图谱（示于图 1）。将每个限制酶片段 cDNA 亚克隆至 pBluescript，然后确定每个插入片段两末端约 300bp 的核苷酸序列。组合这些序列，确定全长的核苷酸序列（SEQ ID No. 3 中所示）。

确定来自全长 cDNA 核苷酸序列的开放阅读框架和氨基酸序列，结果列于 SEQ ID No. 1。与已知的信号肽进行比较，将得到的 N-末端的 30-40 个氨基酸与已知的信号肽比较，推测出本发明多肽的信号肽，并在 SEQ ID No. 4 列出了所得的序列（参见 Von Heuane, G. Nucleic Acids Res. 14, 4683, (1986)）。用与上述相同的方法，得到 3.5kb 克隆的限制酶图谱（列于图 2），全长的核苷酸序列（列于 SEQ ID No. 7），开放阅读框架（列于 SEQ ID No. 6），氨基酸序列（列于 SEQ ID No. 5）以及含信号肽的序列（列于 SEQ ID No. 8）。

3.5kb 克隆与 1.9kb 克隆的推测的氨基酸序列彼此很相似。因此，

将1.9kb克隆命名为S D F - 1 α , 3.5kb克隆命名为S D F - 1 β 。

用荧光测定仪(由 Applied Biosystem Inc.提供), 经循环序列法确定核苷酸序列。由DNA测序仪(Model 373, 由 Applied Biosystem Inc.提供)完成核苷酸序列的阅读。

在数据库(GENBANK 和 EMBL用于DNA检索, NBRF和SWISSPROT用于氨基酸序列检索)中, 用计算机检索S D F - 1 α 和1 β 核苷酸序列与推测的氨基酸序列的同源性。确认本发明的cDNA编码新多肽。

实施例6: 构建用于制备表达载体的质粒载体

使用表达载体pUC-SR α ML-1(在欧洲专利申请No.559428中公开了该载体本身和其制备方法)的衍生物。插入下列两种片段以构建该衍生物: 即分别在多克隆位点中的PstI和SacI之间插入片段T 7

5' GTAATACGACTCACTATAGGGGAGAGCT 3' (SEQ ID NO. 12)

3' ACGTCATTATGCTGAGTGATATCCOCTC 5' (SEQ ID NO. 13)

在SpeI和KpnI位点之间插入片段S P 6

5' CTAGTCTATAGTGTACCTTAATCGTGGGTAC 3' (SEQ ID NO. 14)

3' AGATATCACAGTGGATTAGCAC 5' (SEQ ID NO. 15)

用PstI和SacI消化pUC-SR α ML1载体, 将所得的消化物经琼脂糖凝胶电泳制备, 并回收约4.1kbp的片段, 此后经BAP(细菌碱性磷酸酶)处理除去5' -末端的磷酸基团。将磷酸化的DNA片段T 7与从pUC-SR α ML1得到的约4.1kbp片段相连接, 并使其成环状。另外用SpeI和KpnI消化所得的载体, 使所得的消化产物进行琼脂糖凝胶电泳制备, 并回收约4.1kbp的片段, 此后经BAP(细菌碱性磷酸酶)处理除去5' -末端的磷酸基团。将磷酸化的DNA片段S P 6与由此制备的约4.1kbp片段相连, 使其形成环状。用该方法构建的质粒载体称

为pUC-SR α ML2 (见图3)。

实施例7：构建表达载体

对于hSDF-1 α ,合成引物X, Y和YH。引物X, Y和YH的序列如下：

引物X

5'- A ATA TAG TCG ACC ACC ATG AAC GCC AAG GTC GTG GTC GTG CTG G-3'
(SEQ ID NO. 16)

引物Y

5'-CGG CGG ACT AGT TTA CTT GTT TAA AGC TTT CTC CAG G-3' (SEQ ID NO. 17)

引物YH

5'- GCC GCC ACT AGT TTA GTG GTG GTG GTG GTG GTG CTT GTT TAA AGC TTT CTC
CAG G-3' (SEQ ID NO. 18)

用由此合成的寡核苷酸X和Y作为引物,将hSDF-1 α 质粒进行PCR扩增。由此得到的PCR片段含有一个位于5'与起始密码子相连的序列,即相当于本领域专业人员熟知的Kozak序列和编码由hSDF-1 α 蛋白质组成的蛋白质分子的cDNA。用SalI和SpeI消化该PCR片段,并分离纯化所得的消化物,然后插入至实施例6中制备的pUC-SR α ML2的SalI-SpeI位点以得到表达载体pUC-SR α ML2-hSDF-1 α A。

另外,用合成的寡核苷酸X和YH作为引物,将hSDF-1 α 质粒进行PCR反应。由此得到的PCR片段含有位于5'端与起始密码子相邻的序列,即相当于本领域专业人员熟知的Kozak序列,和编码由hSDF-1 α 蛋白质和与其c-末端相连的6个组氨酸(His)残基组成的蛋白质分子的cDNA。用SalI和SpeI消化该PCR片段,分离并纯化所得的消化产物,然后插入至实施例6中制备的pUC-SR α ML2的SalI-SpeI位点以得到表达载体pUC-SR α ML2-hSDF-1 α B。

对于hSDF-1 β ，合成引物Z和ZH。引物Z和ZH的序列如下：

引物Z：

5'-CGG CGG ACT AGT TCA CAT CTT GAA CCT CTT GTT TAA AGC -3' (SEQ ID NO. 19)

引物ZH：

5'-GCC GCC ACC AGT TCA GGG GGG GGG GGG GGG CAT CTT GAA CCT CTT GTT
TAA AGC -3' (SEQ ID NO. 20)

用由此合成的寡核苷酸X和Z作为引物，将hSDF-1 β 进行PCR反应。由此得到的PCR片段含有位于5'端与起始密码子相邻的序列，即相当于本领域专业人员熟知的Kozac序列，和编码由hSDF-1 β 蛋白质组成的蛋白质分子的cDNA。用SalI和SpeI消化该PCR片段，分离并纯化所得的消化产物，然后插入至实施例6中制备的pUC-SR α ML2的SalI-SpeI位点以得到表达载体pUC-SR α ML2-hSDF-1 β A。

另外，用合成的寡核苷酸X和ZH作为引物，将hSDF-1 β 质粒进行PCR扩增。由此得到的PCR片段含有位于5'端与起始密码子相邻的序列，即相当于本领域专业人员熟知的Kozac序列和编码由hSDF-1 β 蛋白质和与其C-末端相连的6个组氨酸(His)残基组成的蛋白质分子的cDNA。用SalI和SpeI消化该PCR片段，分离并纯化所得的消化产物，然后插入至实施例6中制备的pUC-SR α ML2的SalI-SpeI位点以得到表达载体pUC-SR α ML2-hSDF-1 β B。

分别将由此构建的pUC-SR α ML2-hSDF-1 α A，pUC-SR α ML2-hSDF-1 α B，pUC-SR α ML2-hSDF-1 β A和pUC-SR α ML2-hSDF-1 β B质粒转染至大肠杆菌菌株DH5中，从100ml所得转化体的培养物中回收，然后用CsCl密度梯度离心纯化两次。

实施例8：在COS细胞中表达

用二乙氨基乙基 (DEAE) 葡聚糖方法 (J. Immunology, Vol. 136, P. 4291, 1986), 分别将质粒 DNA 制品 pUC-SR α ML2, pUC-SR α ML2-hSDF-1 α A, pUC-SR α ML2-hSDF-1 α B, pUC-SR α ML2-hSDF-1 β A 和 pUC-SR α ML2-hSDF-1 β B 导入 COS-7 细胞 (Cell, Vol. 23, P. 175, 1981)。

即, 将约 1.8×10^6 个 COS-7 细胞接种至含 50 ml 液体培养基 (Dulbecco's 改良 MEM 培养基并补充了 10% 去除补体的胎牛血清) 的 225 cm² 容积的烧瓶 (由 Corning 生产) 中。在二氧化碳培养箱中 (37℃, 5% CO₂) 过夜培养后, 随后除去培养上清液, 向每个烧瓶中加入 12 ml DNA 混合液 (Dulbecco's 改良 MEM 培养基, 并补充了各 15 μ g 的每种质粒 DNA, 50 mM Tris-HCl 缓冲液 (PH 7.4) 和 400 μ g/ml 的 DEAE-葡聚糖), 然后在 5% CO₂ 气体中, 于 37℃ 培养 3 小时。此后, 用 15 ml 氯喹溶液代替 DNA 混合液 (Dulbecco's 改良 MEM 培养基, 并且添加了 150 μ M 氯喹和 7% 去除补体的胎牛血清), 接着再培养 3 小时。

除去氯喹溶液后, 将上述的液体培养基 (50 ml) 加至每个烧瓶中, 然后在 5% CO₂ 空气中, 于 37℃ 培养 72 小时, 发现在每个烧瓶中, 细胞几乎都长成单层。除去培养上清液后, 用无血清的液体培养基 (商标: SFM-101; 可从 Nissui Pharmaceutical Co. Ltd. 得到) 洗涤每个烧瓶中的细胞, 然后加入 25 ml 相同的无血清液体培养基, 然后继续培养 72 小时。此后, 回收所得的培养上清液, 并经膜滤器 (商标: STERIVEX-GS; 可从 Millipore Corp. 得到) 过滤以除去细胞和细胞碎片。将由此得到的培养上清液于 4℃ 贮存以备后用。期望已用含 hSDF-1 α 和 β cDNA 插入片段的质粒转化的 COS

细胞培养上清液含有 相当于hSDF-1 α 和 β 的多肽的表达并分泌的成熟蛋白质部分。

实施例 9：表达的确证

用离心浓缩滤器（商标：Centricon-10；可从 Millipore Corp. 得到）将实施例 8 中得到的各 2 ml 的每种转化 COS 细胞的培养上清液浓缩至体积为 100 μ l。将各 1 μ l 由此浓缩的每种样品与相同体积用于 SDS-PAGE（十二烷基磺酸钠聚丙烯酰胺凝胶电泳）的载样缓冲液（0.125 M Tris - HCl 缓冲液（pH 6.8），4% 十二烷基磺酸钠和 30% 甘油）混合，然后在 90 $^{\circ}$ C，将混合物处理 3 分钟，然后进行 SDS-PAGE 电泳。

如果 hSDF-1 α B 和 β B 在蛋白质的 C 末端导入有 His 六聚体，则不仅将其相应的 COS 细胞培养上清液而且将其纯化产物均经 SDS-PAGE 分析。

用金属螯合物亲和层析法（Biotechnology, Vol. 9, P. 273, 1991），利用 His 与各种过渡性的金属离子可形成复合化合物的功能，完成蛋白质的纯化。即，将由 COS 细胞得到的培养上清液（350 ml）与一定量（该数量可使盐的终浓度为 1 M）的氯化钠水溶液混合，使所得的混合物流过用 4 ml 锌结合的螯合琼脂糖（Sephacrose）（商标：Chelating Sepharose Fast-Flow；可从 Pharmacia 购买）填装的柱，以便将蛋白质吸附在树脂上。用含 1 M 氯化钠水溶液（40 ml）的 50 mM 磷酸盐缓冲液（pH 7.0）洗涤该柱，用含 1 M 氯化钠水溶液和 0.4 M 咪唑的 50 mM 磷酸盐缓冲液（pH 7.0）洗脱保留在柱上的蛋白质。然后，将所得洗脱液浓缩至 100 μ l 体积，将部分浓缩样品经 SDS-PAGE 分析。用 SDS 10/20 梯度凝胶完成 SDS-PAGE 分析，分别检

测相当于hSDF-1 α 和hSDF-1 β 分子量的产物。

序列表

(1) 一般资料:

(i) 申请人

(A) 名称: Ono Pharmaceutical Co., Ltd.

(B) 街道: 1-5, Doshomachi 2-chome

(C) 城市: Chuo-ku, Osaka-shi

(D) 州: Osaka

(E) 国家: Japan

(F) 邮政编码 (ZIP): 541

(ii) 发明题目: 新多肽及编码它们的 DNA

(iii) 序列数: 20

(2) SEQ ID No: 1 的资料:

(i) 序列特征:

(A) 长度: 89 个氨基酸

(B) 类型: 氨基酸

(D) 几何结构: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID No: 1

```
Met Asn Ala Lys Val Val Val Val Leu Val Leu Val Leu Thr Ala Leu
1          5          10          15
Cys Leu Ser Asp Gly Lys Pro Val Ser Leu Ser Tyr Arg Cys Pro Cys
20          25          30
Arg Phe Phe Glu Ser His Val Ala Arg Ala Asn Val Lys His Leu Lys
35          40          45
Ile Leu Asn Thr Pro Asn Cys Ala Leu Gln Ile Val Ala Arg Leu Lys
50          55          60
```

Asn Asn Asn Arg Gln Val Cys Ile Asp Pro Lys Leu Lys Trp Ile Gln
65 70 75 80
Glu Tyr Leu Glu Lys Ala Leu Asn Lys
85

(2) SEQ ID No: 2 的资料:

(i) 序列特征:

(A) 长度: 267个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 几何结构: 线性

(ii) 分子类型: cDNA到mRNA

(xi) 序列描述: SEQ ID No: 2;

ATGAACGCCA AGGTCGTGGT CGTGCTGGTC CTCGTGCTGA CCGCGCTCTG CCTCAGCGAC	60
GGGAAGCCCG TCAGCCTGAG CTACAGATGC CCATGCCGAT TCTTCGAAAG CCATGTTGCC	120
AGAGCCAACG TCAAGCATCT CAAAATTCTC AACACTCCAA ACTGTGCCCT TCAGATTGTA	180
GCCCGGCTGA AGAACACAA CAGACAAGTG TGCATTGACC CGAAGCTAAA GTGGATTGAG	240
GAGTACCTGG AGAAAGCTTT AAACAAG	267

(2) SEQ ID No: 3 的资料:

(i) 序列特征:

(A) 长度: 1856个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 几何结构: 线性

(ii) 分子类型: cDNA到mRNA

(xi) 序列描述: SEQ ID No: 3;

TCTCCGTCAG CCGCATTGCL CGCTCGGCGT CCGGCCCCCG ACCCGTGCTC GTCCGCCCCG	60
CCGCCCCGCC GCGCGGCCA TGAACGCCAA GGTCGTGGTC GTGCTGGTCC TCGTGCTGAC	120
CGCGCTCTGC CTCAGCGACG GGAAGCCCGT CAGCCTGAGC TACAGATGCC CATGCCGATT	180
CTTCGAAAGC CATGTTGCCA GAGCCAACGT CAAGCATCTC AAAATTCTCA ACACTCCAA	240
CTGTGCCCTT CAGATTGTAG CCCGGCTGAA GAACAACAAC AGACAAGTGT GCATTGACCC	300
GAAGCTAAAG TGGATTGAGG AGTACCTGGA GAAAGCTTTA AACAACTAAG CACAACAGCC	360
AAAAAGGACT TTCCGCTAGA CCCACTCGAG GAAACTAAA ACCTTGTGAG AGATGAAAGG	420
GCAAAGACGT GGGGGAGGGG GCCTTAACCA TGAGGACCAG GTGTGTGTGT GGGGTGGGCA	480
CATTGATCTG GGATCGGGCC TGAGGTTTGC AGCATTGAGA CCCTGCATTT ATAGCATACG	540
GTATGATATT GCAGCTTATA TTCATCCATG CCCTGTACCT GTGCACGTTG GAACTTTTAT	600
TACTGGGGTT TTTCTAAGAA AGAAATTGTA TTATCAACAG CATTTTCAAG CAGTTAGTTC	660
CTTCATGATC ATCACAATCA TCATCATCTT CATTCCTATT TTTTAAATCA ACGAGTACTT	720
CAAGATCTGA ATTTGGCTTG TTTGGAGCAT CTCCTCTGCT CCCCTGGGGA GTCTGGGCAC	780
AGTCAGGTGG TGGCTTAACA GGGAGCTGGA AAAAGTGTC TTTCTTCAGA CACTGAGGCT	840
CCCGCAGCAG CGCCCCCTCC AAGAGGAAGG CCTCTGTGGC ACTCAGATAC CGACTGGGGC	900
TGGGGCGCCG CCACTGCCTT CACCTCCTCT TTCAAACCTC AGTGATTGGC TCTGTGGGCT	960
CCATGTAGAA GCCACTATTA CTGGGACTGT CTCAGAGACC CCTCTCCCAG CTATTCCTAC	1020
TCTCTCCCCG ACTCCGAGAG CATGCTTAAT CTTGCTTCTG CTTCTCATTT CTGTAGCCTG	1080
ATCAGCGCCG CACCAGCCGG GAAGAGGGTG ATTGCTGGGG CTCGTGCCCT GCATCCCTCT	1140
CCTCCCAGGG CCTGCCCCAC AGCTCGGGCC CTCTGTGAGA TCCGTCTTTG GCCTCCTCCA	1200
GAATGGAGCT GGCCCTCTCC TGGGGATGTG TAATGGTCCC CCTGCTTACC CGCAAAAGAC	1260
AAGTCTTTAC AGAATCAAT GCAATTTTAA ATCTGAGAGC TCGCTTGAGT GACTGGGTTT	1320
GTGATTGCCT CTGAAGCCTA TGTATGCCAT GGAGGCACTA ACAAACTCTG AGGTTTCCGA	1380
AATCAGAAGC GAAAAAATCA GTGAATAAAC CATCATCTTG CCACTACCCC CTCCTGAAGC	1440
CACAGCAGGG GTTCAGGTTC CAATCAGAAC TGTTGCCAAG GTGACATTTC CATGCATAGA	1500
TGCGATCCAC AGAAGGTCCCT GGTGGTATTT GTAACTTTTT GCAAGGCATT TTTTATATA	1560
TATTTTGTG CACATTTTTT TTTACGATTC TTAGAAAAC AATGTATTT CAAAATATAT	1620
TTATAGTCGA ACAAGTCATA TATATGAATG AGAGCCATAT GAATGTCAGT AGTTTATACT	1680
TCTCTATTAT CTCAACTAC TGGCAATTTG TAAAGAAATA TATATGATAT ATAAATGTGA	1740
TTGCAGCTTT TCAATGTTAG CCACAGTGTA TTTTTTCACT TGTACTAAAA TTGTATCAAA	1800
TGTGACATTA TATGCACTAG CAATAAAATG CTAATTGTTT CATGGTAAAA AAAAAA	1856

(2) S E Q I D N o : 4 的资料 :

(i) 序列特征：

(A) 长度：1856个碱基对

(B) 类型：核酸

(C) 链型：单链

(D) 几何结构：线性

(ii) 分子类型：cDNA到mRNA

(V) 来源：

(A) 生物：Homo Sapiens

(H) 细胞系：FLEB14

(ix) 特征：

(A) 名称／关键词：C D S

(B) 位置：80..349

(C) 鉴别方法：利用与某些其它模式的相似性

(ix) 特征：

(A) 名称／关键词：信号肽

(B) 位置：80..142

(C) 鉴别方法：利用与已知序列或已建立的一致序列的相似性

(ix) 特征：

(A) 名称／关键词：成熟肽

(B) 位置：143..346

(C) 鉴别方法：利用与已知序列或已建立的一致序列的相似性

(xi) 序列描述：SEQ ID No : 4 :

TCTCCGTCAG CCGCATTGCC CGCTCGGGCGT CCGGCCCCCG ACCCGTGCTC GTCCGCCCCG 60
CCGCCCCGCC GCGCGCGCC ATG AAC GCC AAG GTC GTG GTC GTG CTG GTC CTC 112
Met Asn Ala Lys Val Val Val Val Leu Val Leu
-21 -20 -15
GTG CTG ACC GCG CTC TGC CTC AGC GAC GGG AAG CCC GTC AGC CTG AGC 160
Val Leu Thr Ala Leu Cys Leu Ser Asp Gly Lys Pro Val Ser Leu Ser
-10 -5 1 5
TAC AGA TGC CCA TGC CGA TTC TTC GAA AGC CAT GTT GCC AGA GCC AAC 208
Tyr Arg Cys Pro Cys Arg Phe Phe Glu Ser His Val Ala Arg Ala Asn
10 15 20
GTC AAG CAT CTC AAA ATT CTC AAC ACT CCA AAC TGT GCC CTT CAG ATT 256
Val Lys His Leu Lys Ile Leu Asn Thr Pro Asn Cys Ala Leu Gln Ile
25 30 35
GTA GCC CGG CTG AAG AAC AAC AAC AGA CAA GTG TGC ATT GAC CCG AAG 304
Val Ala Arg Leu Lys Asn Asn Asn Arg Gln Val Cys Ile Asp Pro Lys
40 45 50
CTA AAG TGG ATT CAG GAG TAC CTG GAG AAA GCT TTA AAC AAG TAAGCACAAC 356
Leu Lys Trp Ile Gln Glu Tyr Leu Glu Lys Ala Leu Asn Lys
55 60 65
AGCCAAAAG GACTTTCCGC TAGACCCACT CGAGGAAAC TAAAACCTTG TGAGAGATGA 416
AAGGGCAAAG ACGTGGGGGA GGGGGCCTTA ACCATGAGGA CCAGGTGTGT GTGTGGGGTG 476
GGCACATTGA TCTGGGATCG GGCCTGAGGT TTGCAGCATT TAGACCCCTGC ATTTATAGCA 536
TACGGTATGA TATTGCAGCT TATATTCATC CATGCCCTGT ACCTGTGCAC GTTGGAACTT 596
TTATTACTGG GGTTTTTCTA AGAAGAAAT TGTATTATCA ACAGCATTTT CAAGCAGTTA 656
GTTCCCTTCAT GATCATCACA ATCATCATCA TTCTCATTTT CATTTTTTAA ATCAACGAGT 716
ACTTCAAGAT CTGAATTGG CTGTGTTGGA GCATCTCCTC TGCTCCCTG GGGAGTCTGG 776
GCACAGTCAG GTGGTGGCTT AACAGGGAGC TGGAAAAGT GTCCTTTCTT CAGACACTGA 836
GGCTCCCGCA GCAGCGCCCC TCCCAAGAGG AAGGCCTCTG TGGCACTCAG ATACCGACTG 896
GGGCTGGGGC GCCGCCACTG CCTTCACCTC CTCTTTCAA CCTCAGTGAT TGGCTCTGTG 956
GGCTCCATGT AGAAGCCACT ATTACTGGGA CTGTCTCAGA GACCCCTCTC CCAGCTATTC 1016
CTACTCTCTC CCCGACTCCG AGAGCATGCT TAATCTTGCT TCTGCTTCTC ATTTCTGTAG 1076
CCTGATCAGC GCCGCACCAG CCGGGAAGAG GGTGATTGCT GGGGCTCGTG CCCTGCATCC 1136
CTCTCTCCC AGGGCCTGCC CCACAGCTCG GGCCCTCTGT GAGATCCGTC TTTGGCCTCC 1196
TCCAGAATGG AGCTGGCCCT CTCCTGGGGA TGTGTAATGG TCCCCCTGCT TACCCGCAA 1256
AGACAAGTCT TTACAGAATC AAATGCAATT TTAAATCTGA GAGCTCGCTT GAGTGACTGG 1316
GTTTGTGATT GCCTCTGAAG CCTATGTATG CCATGGAGGC ACTAACAAC TCTGAGGTTT 1376
CCGAAATCAG AAGCGAAAA ATCAGTGAAT AAACCATCAT CTTGCCACTA CCCCCTCCTG 1436
AAGCCACAGC AGGGGTTTCTG GTTCCAATCA GAACTGTTGG CAAGGTGACA TTTCCATGCA 1496

TAGATGCGAT CCACAGAAGG TCCTGGTGGT ATTTGTAACCT TTTTGCAAGG CATTMTTTT	1556
TATATATTTT TGTGCACATT TTTTTTTACG ATTCTTTAGA APACAAATGT ATTTCAAAAT	1616
ATATTTATAG TCGAACAAAGT CATATATATG AATGAGAGCC. ATATGAATGT CAGTAGTTTA	1676
TACTTCTCTA TTATCTCAAA CTACTGGCAA TTTGTAAAGA AATATATATG ATATATAAAT	1736
GTGATTGCAG CTTTTCAATG TTAGCCACAG TGTATTTTTT CACTGTACT AAAATTGTAT	1796
CAAATGTGAC ATTATATGCA CTAGCAATAA AATGCTAATT GTTTCATGGT AAAAAAAAAA	1856

(2) S E Q I D N o : 5 的资料 :

(i) 序列特征 :

(A) 长度 : 9 3 个氨基酸

(B) 类型 : 氨基酸

(D) 几何结构 : 线性

(ii) 分子类型 : 蛋白质

(xi) 序列描述 : S E Q I D N o : 5

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

(2) S E Q I D N o : 6 的资料 :

(i) 序列特征 :

(A) 长度 : 279个碱基对

(B) 类型 : 核酸

(C) 链型 : 单链

(D) 几何结构 : 线性

(ii) 分子类型：cDNA到mRNA

(xi) 序列描述：SEQ ID No：6：

```
ATGAACGCCA AGGTCGTGGT CGTGCTGGTC CTCGTGCTGA CCGCGCTCTG CCTCAGCGAC      60
GGGAAGCCCG TCAGCCTGAG CTACAGATGC CCATGCCGPT TCTTCGAAAG CCATGTTGCC      120
AGAGCCAACG TCAAGCATCT CAAAATTCTC AACACTCCAA ACTGTGCCCT TCAGATTGTA      180
GCCCCGCTGA AGAACAACAA CAGACAAGTG TGCATTGACC CGAAGCTAAA GTGGATTGAG      240
GAGTACCTGG AGAAGCTTT AACCAAGAGG TTCAAGATG      279
```

(2) SEQ ID No：7 的资料：

(i) 序列特征：

(A) 长度：3526个碱基对

(B) 类型：核酸

(C) 链型：单链

(D) 几何结构：线性

(ii) 分子类型：cDNA到mRNA

(xi) 序列描述：SEQ ID No：7：

```
TCTCCGTCAG CCGCATTGCC CGCTCGGCGT CCGGCCCCCG ACCCGTGCTC GTCCGCCCCG      60
CCGCCCCGCC GCGCGGCCA TGAACGCCAA GGTGCTGGTC GTGCTGGTCC TCGTGCTGAC      120
CGCGCTCTGC CTCAGCGACG GGAAGCCCGT CAGCCTGAGC TACAGATGCC CATGCCGATT      180
CTTCGAAAGC CATGTTGCCA GAGCCAACGT CAAGCATCTC AAAATTCTCA ACACTCCAAA      240
CTGTGCCCTT CAGATTGTAG CCCGGCTGAA GAACAACAAC AGACAAGTGT GCATTGACCC      300
GAAGCTAAAG TGGATTGAGG AGTACCTGGA GAAAGCTTTA AACCAAGAGG TCAAGATGTG      360
AGAGGGTCAG ACGCCTGAGG AACCCCTACA GTAGGAGCCC AGCTCTGAAA CCAGTGTTAG      420
GGAAGGGCCT GCCACAGCCT CCCCTGCCAG GGCAGGGCCC CAGGCATTGC CAAGGGCTTT      480
GTTTTGCACA CTTTGCCATA TTTTCACCAT TTGATTATGT AGCAAAATAC ATGACATTTA      540
```

TTTTTCATTT	AGTTTGATTA	TTCAGTGTCA	CTGGCGACAC	GTAGCAGCTT	AGACTAAGGC	600
CATTATTGTA	CTTGCCCTTAT	TAGAGTGTCT	TTCCACGGAG	CCACTCCTCT	GACTCAGGGC	660
TCCTGGGTTT	TGTATTCTCT	GAGCTGTGCA	GGTGGGGAGA	CTGGGCTGAG	GGAGCCTGGC	720
CCCATGGTCA	GCCCTAGGGT	GGAGAGCCAC	CAAGAGGGAC	GCCTGGGGGT	GCCAGGACCA	780
GTCAACCTGG	GCAAAGCCTA	GTGAAGGCTT	CTCTCTGTGG	GATGGGATGG	TGGAGGGCCA	840
CATGGGAGGC	TCACCCCTT	CTCCATCCAC	ATGGGAGCCG	GGTCTGCCTC	TTCTGGGAGG	900
GCAGCAGGGC	TACCCCTGAGC	TGAGGCAGCA	GTGTGAGGCC	AGGGCAGAGT	GAGACCCAGC	960
CCTCATCCCC	AGCACCTCCA	CATCCTCCAC	GTTCTGCTCA	TCATTCTCTG	TCTCATCCAT	1020
CATCATGTGT	GTCCACGACT	GTCTCCATGG	CCCCGCAAAA	GGACTCTCAG	GACCAAAGCT	1080
TTCATGTAAA	CTGTGCACCA	AGCAGGAAAT	GAAAATGTCT	TGTGTTACCT	GAAACACTG	1140
TGCACATCTG	TGTCTTGTGT	GGAATATTGT	CCATTGTCCA	ATCCTATGTT	TTTGTTCAAA	1200
GCCAGCGTCC	TCCTCTGTGA	CCAATGTCTT	GATGCATGCA	CTGTTCCCCC	TGTGCAGCCG	1260
CTGAGCGAGG	AGATGCTCCT	TGGGCCCTTT	GAGTGCAGTC	CTGATCAGAG	CCGTGGTCCT	1320
TTGGGGTGAA	CTACCTTGGT	TCCCCCACTG	ATCACAAAAA	CATGGTGGGT	CCATGGGCAG	1380
AGCCCAAGGG	AATTCGGTGT	GCACCAGGGT	TGACCCGAGA	GGATTGCTGC	CCCATCAGTG	1440
C TCCCTCACA	TGTCAGTACC	TTCAAAC TAG	GGCCAAGCCC	AGCACTGCTT	GAGGAAAACA	1500
AGCATTCACA	ACTTGTTTTT	GGTTTTTAAA	ACCCAGTCCA	CAAAATAACC	AATCCTGGAC	1560
ATGAAGATTC	TTTCCCAATT	CACATCTAAC	CTCATCTTCT	TCACCATT TG	GCAATGCCAT	1620
CATCTCCTGC	CTTCCTCCTG	GGCCCTCTCT	GCTCTGCGTG	TCACCTGTGC	TTGGGGCCCT	1680
TCCCACAGGA	CATTTCTCTA	AGAGAACAAT	GTGCTATGTG	AAGAGTAAGT	CAACCTGCCT	1740
GACATTTGGA	GTGTTCCCTT	CCCCTGAGG	GCAGTCGATA	GAGCTGTATT	AAGCCACTTA	1800
AAATGTTTAC	TTTTGACAAA	GGCAAGCACT	TGTGGGTTTT	TGTTTTGT TT	TTTATT CAGT	1860
CTTACGAATA	CTTTTGCCCT	TTGATTAAAG	ACTCCAGTTA	AAAAAAATTT	TAATGAAGAA	1920
AGTGGAAAAC	AAGGAAGTCA	AAGCAAGGAA	ACTATGT AAC	ATGTAGGAAG	TAGGAAGTAA	1980
ATTATAGTGA	TGTAATCTTG	AATTGTA ACT	GTTCTGTGAAT	TTAATAATCT	GTAGGGTAAT	2040
TAGTAACATG	TGTTAAGTAT	TTTCATAAGT	ATTTCAAATT	GGAGCTTCAT	GGCAGAAGGC	2100
AAACCCATCA	ACAAAAATTG	TCCCTTAAAC	AAAAATTAAA	ATCCTCAATC	CAGCTATGTT	2160
ATATTGAAAA	AATAGAGCCT	GAGGGATCTT	TACTAGTTAT	AAAGATACAG	AACTCTTTCA	2220

AAACCTTTTG	AAATTAACCT	CTCACTATAC	CAGTATAATT	GAGTTTTCAG	TGGGGCAGTC	2280
ATTATCCAGG	TAATCCAAGA	TATTTTAAAA	TCTGTCACGT	AGAACTTGA	TGTACCTGCC	2340
CCCAATCCAT	GAACCAAGAC	CATTGAATTC	TTGGTTGAGG	AAACAAACAT	GACCCTAAAT	2400
CTTGACTACA	GTCAGGAAG	GAATCATTTTC	TATTTCTCCT	CCATGGGAGA	AAATAGATAA	2460
GAGTAGAAAC	TGCAGGGAAA	ATTATTTGCA	TAACAATTCC	TCTACTAACA	ATCAGCTCCT	2520
TCCTGGAGAC	TGCCCAGCTA	AAGCAATATG	CATTTAAATA	CAGTCTTCCA	TTTGCAAGGG	2580
AAAAGTCTCT	TGTAATCCGA	ATCTCTTTTT	GCTTTCGAAC	TGCTAGTCAA	GTGCGTCCAC	2640
GAGCTGTTTA	CTAGGGATCC	CTCATCTGTC	CCTCCGGGAC	CTGGTGCTGC	CTCTACCTGA	2700
CACTCCCTTG	GGCTCCCTGT	AACCTCTTCA	GAGGCCCTCG	CTGCCAGCTC	TGTATCAGGA	2760
CCCAGAGGAA	GGGSCCAGAG	GCTCGTTGAC	TGGCTGTGTG	TTGGGATTGA	GTCTGTGCCA	2820
CGTGTATGTG	CTGTGGTGTG	TCCCCCTCTG	TCCAGGCACT	GAGATACCAG	CGAGGAGGCT	2880
CCAGAGGGCA	CTCTGCTTGT	TATTAGAGAT	TACCTCCTGA	GAAAAAGCT	TCCGCTTGGG	2940
GCAGAGGGGC	TGAATAGCAG	AAGGTTGCAC	CTCCCCAAC	CTTAGATGTT	CTAAGTCTTT	3000
CCATTGGATC	TCATTGGACC	CTTCCATGGT	GTGATCGTCT	GACTGGTGTT	ATCACCGTGG	3060
GCTCCCTGAC	TGGGAGTTGA	TGCCTTTTCC	CAGGTGCTAC	ACCCCTTTTCC	AGCTGGATGA	3120
GAAATTTGAGT	GCTCTGATCC	CTCTACAGAG	CTTCCCTGAC	TCATTCTGAA	GGAGCCCCAT	3180
TCCTGGGAAA	TATTCCTAG	AAACTTCCAA	ATCCCCTAAG	CAGACCACTG	ATAAAACCAT	3240
GTAGAAAATT	TGTTATTTTG	CAACCTCGCT	GGACTCTCAG	TCTCTGAGCA	GTGAATGATT	3300
CAGTGTTAAA	TGTGATGAAT	ACTGTATTTT	GTATTGTTTC	AAGTGCATCT	CCCAGATAAT	3360
GTGAAAATGG	TCCAGGAGAA	GGCCAATTCC	TATACGCAGC	GTGCTTTAAA	AAATAAATAA	3420
GAAACAACCT	TTTGAGAAAC	AACAATTTCT	ACTTTGAAGT	CATACCAATG	AAAAAATGTA	3480
TATGCACTTA	TAATTTTCCT	AATAAAGTTC	TGTACTCAAA	TGTAAA		3526

(2) (2) S E Q I D N o : 8 的资料 :

(i) 序列特征 :

- (A) 长度 : 3526个碱基对
- (B) 类型 : 核酸
- (C) 链型 : 单链
- (D) 几何结构 : 线性

(ii) 分子类型：cDNA到mRNA

(Vi) 来源：

(A) 生物：Homo Sapiens

(B) 细胞系：FLEB14

(ix) 特征：

(A) 名称/关键词：CDS

(B) 位置：80..361

(C) 鉴别方法：利用与某些其它模式的相似性

(ix) 特征：

(A) 名称/关键词：信号肽

(B) 位置：80..142

(C) 鉴别方法：利用与已知序列或已建立的一致序列的相似性

(ix) 特征：

(A) 名称/关键词：成熟肽

(B) 位置：143..358

(C) 鉴别方法：利用与已知序列或已建立的一致序列的相似性

(xi) 序列描述：SEQ ID No：8：

```
TCTCCGTCAG CCGCATTGCC CGCTCGGCGT CCGGCCCCCG ACCCGTGCTC GTCCGCCCCG      60
CCGCCCCGCC GCGCGCGCC ATG AAC GCC AAG GTC GTG GTC GTG CTG GTC CTC      112
               Met Asn Ala Lys Val Val Val Val Leu Val Leu
               -21 -20                               -15
GTG CTG ACC GCG CTC TGC CTC AGC GAC GGG AAG CCC GTC AGC CTG AGC      160
Val Leu Thr Ala Leu Cys Leu Ser Asp Gly Lys Pro Val Ser Leu Ser
-10                -5                1                5
```

TAC AGA TGC CCA TGC CGA TTC TTC GAA AGC CAT GTT GCC AGA GCC AAC Tyr Arg Cys Pro Cys Arg Phe Phe Glu Ser His Val Ala Arg Ala Asn 10 15 20	208
GTC AAG CAT CTC AAA ATT CTC AAC ACT CCA AAC TGT GCC CTT CAG ATT Val Lys His Leu Lys Ile Leu Asn Thr Pro Asn Cys Ala Leu Gln Ile 25 30 35	256
GTA GCC CGG CTG AAG AAC AAC AAC AGA CAA GTG TGC ATT GAC CCG AAG Val Ala Arg Leu Lys Asn Asn Asn Arg Gln Val Cys Ile Asp Pro Lys 40 45 50	304
CTA AAG TGG ATT CAG GAG TAC CTG GAG AAA GCT TTA AAC AAG AGG TTC Leu Lys Trp Ile Gln Glu Tyr Leu Glu Lys Ala Leu Asn Lys Arg Phe 55 60 65 70	352
AAG ATG TGAGAGGGTC AGACGCCTGA GGAACCCTTA CAGTAGGAGC CCAGCTCTGA Lys Met	408
AACCACTGTT AGGGAAGGGC CTGCCACAGC CTCCCCTGCC AGGGCAGGGC CCCAGGCATT	468
GCCAAGGGCT TTGTTTTCGA CACTTTGCCA TATTTTCACC ATTTGATTAT GTAGCAAAAT	528
ACATGACATT TATTTTTCAT TTAGTTTGAT TATTCAGTGT CACTGGCGAC ACGTAGCAGC	588
TTAGACTAAG GCCATTATTG TACTTGCCCTT ATTAGAGTGT CTTTCCACGG AGCCACTCCT	648
CTGACTCAGG GCTCCTGGGT TTTGTATTCT CTGAGCTGTG CAGGTGGGGA GACTGGGCTG	708
AGGGAGCCTG GCCCCATGGT CAGCCCTAGG GTGGAGAGCC ACCAAGAGGG ACGCCTGGGG	768
GTGCCAGGAC CAGTCAACCT GGGCAAAGCC TAGTGAAGGC TTCTCTCTGT GGGATGGGAT	828
GGTGGAGGGC CACATGGGAG GCTCACCCCC TTCTCCATCC ACATGGGAGC CGGGTCTGCC	888
TCTTCTGGGA GGGCAGCAGG GCTACCCCTGA GCTGAGGCAG CAGTGTGAGG CCAGGGCAGA	948
GTGAGACCCA GCCCTCATCC CGAGCACCTC CACATCCTCC ACGTTCCTGT CATCATCTC	1008
TGTCTCATCC ATCATCATGT GTGTCCACGA CTGTCTCCAT GGGCCCGCAA AAGGACTCTC	1068
AGGACCAAAG CTTTCATGTA AACTGTGCAC CAAGCAGGAA ATGAAAATGT CTTGTGTTAC	1128
CTGAAACAC TGTGCACATC TGTGTCTTGT GTGGAATATT GTCCATTGTC CAATCCTATG	1188
TTTTTGTTC AAGCCAGCGT CCTCCTCTGT GACCAATGTC TTGATGCATG CACTGTTCCC	1248
CCTGTGCAGC CGCTGAGCGA GGAGATGCTC CTTGGGCCCT TTGAGTGCAG TCCTGATCAG	1308
AGCCGTGGTC CTTTGGGGTG AACTACCTTG GTTCCCCCAC TGATCACAAA AACATGGTGG	1368
GTCCATGGGC AGAGCCCCAAG GGAATTCCGT GTGCACCAGG GTTGACCCCA GAGGATTGCT	1428
GCCCCATCAG TGCTCCCTCA CATGTCAGTA CCTTCAAACCT AGGGCCAAGC CCAGCACTGC	1488
TTGAGGAAAA CAAGCATTCA CAACTGTTTT TTGGTTTTTA AAACCCAGTC CACAAAATAA	1548
CCAATCCTGG ACATGAAGAT TCTTTCCCAA TTCACATCTA ACCTCATCTT CTTCAACATT	1608

TGGCAATGCC ATCATCTCCT GCCTTCCTCC TGGGCCCTCT CTGCTCTGCG TGTCACCTGT	1668
GCTTCGGGCC CTTCCACAG GACATTTCTC TAAGAGAACA ATGTGCTATG TGAAGAGTAA	1728
GTCAACCTGC CTGACATTTG GAGTGTTCCT CTCCCACTGA GGGCAGTCGA TAGAGCTGTA	1788
TTAAGCCACT TAAAATGTTT ACTTTTGACA AAGGCAAGCA CTTGTGGGTT TTTGTTTTGT	1848
TTTTCATCA GTCTTACGAA TACTTTTGCC CTTTGATTAA AGACTCCAGT TAAAAAAAT	1908
TTTAATGAAG AAAGTGGAAA ACAAGGAAGT CAAAGCAAGG AAATATGTA ACATGTAGGA	1968
AGTAGGAAGT AAATTATAGT GATGTAATCT TGAATTGTAA CTGTTCGTGA ATTTAATAAT	2028
CTGTAGGTA ATTAGTAACA TGTGTTAAGT ATTTTCATAA GTATTTCAAA TTGGAGCTTC	2088
ATGGCAGAAG GCAAACCCAT CAACAAAAT TGTCCCTTAA ACAAATAA AAATCCTCAA	2148
TCCAGCTATG TTATATTGAA AAAATAGAGC CTGAGGGATC TTTACTAGTT ATAAAGATAC	2208
AGAACTCTTT CAAAACCTTT TGAATTAAC CTCTCACTAT ACCAGTATAA TTGAGTTTTT	2268
AGTGGGGCAG TCATTATCCA GGTAATCCAA GATATTTTAA AATCTGTCAC GTAGAACTTG	2328
GATGTACCTG CCCCCAATCC ATGAACCAAG ACCATTGAAT TCTTGGTTGA GGAAACAAAC	2388
ATGACCCTAA ATCTTGACTA CAGTCAGGAA AGGAATCATT TCTATTTCTC CTCCATGGGA	2448
GAAAATAGAT AAGAGTAGAA ACTGCAGGGA AAATTATTG CATAACAATT CCTCTACTAA	2508
CAATCAGCTC CTTCTGGAG ACTGCCCAGC TAAAGCAATA TGCATTTAAA TACAGTCTTC	2568
CATTGCAAG GGAAAAGTCT CTTGTAATCC GAATCTCTTT TTGCTTTTGA ACTGCTAGTC	2628
AAGTGCCTCC ACGAGCTGTT TACTAGGGAT CCCTCATCTG TCCCTCCGGG ACCTGGTGCT	2688
GCCTCTACCT GACACTCCCT TGGGCTCCCT GTAACCTCTT CAGAGGCCCT CGCTGCCAGC	2748
TCTGTATCAG GACCCAGAGG AAGGGGCCAG AGGCTCGTTG ACTGGCTGTG TGTGGGATT	2808
GAGTCTGTGC CACGTGTATG TGCTGTGGTG TGTCCCCCTC TGTCCAGGCA CTGAGATACC	2868
AGCGAGGAGG CTCCAGAGGG CACTCTGCTT GTTATTAGAG ATTACCTCCT GAGAAAAAAG	2928
CTTCCGCTTG GAGCAGAGGG GCTGAATAGC AGAAGGTTGC ACCTCCCCCA ACCTTAGATG	2988
TTCTAAGTCT TTCCATTGGA TCTCATTGGA CCCTTCCATG GTGTGATCGT CTGACTGGTG	3048
TTATCACCGT GGGCTCCCTG ACTGGGAGTT GATCGCCTTT CCCAGGTGCT ACACCCTTTT	3108
CCAGCTGGAT GAGAATTGTA GTGCTCTGAT CCCTCTACAG AGCTTCCCTG ACTCATCTG	3168
AAGGAGCCCC ATTCTGGGA AATATTCCTT AGAACTTCC AAATCCCCTA AGCAGACCAC	3228

TGATAAAACC ATGTAGAAA TTTGTTATTT TGCAACCTCG CTGGACTCTC AGTCTCTGAG	3288
CAGTGAATGA TTCAGTGTTA AATGTGATGA ATACTGTATT TTGTATTGTT TCAAGTGCAT	3348
CTCCCAGATA ATGTGAAAAT GGTCCAGGAG AAGGCCAATT CCTATACGCA GCGTGCCTTTA	3408
AAAAATAAAT AAGAAACAAC TCTTTGAGAA ACAACAATTT CTACTTTGAA GTCATACCAA	3468
TGAAAAATG TATATGCACT TATAATTTTC CTAATAAAGT TCTGTACTCA AATGTAAA	3526

(2) S E Q I D N o : 9 的资料 :

(i) 序列特征 :

(A) 长度1797个碱基对

(B) 类型 : 核酸

(C) 链型 : 单链

(D) 几何结构 : 线性

(ii) 分子类型 : cDNA到mRNA

(xi) 序列描述 : S E Q I D N o : 9 :

GACCACTTTC CCTCTCGGTC CACCTCGGTG TCCTCTTGCT GTCCAGCTCT GCAGCCTCCG	60
GCGCGCCCTC CCGCCACGC CATGGACGCC AAGGTCGTCG CCGTCTGGC CCTGGTGCTG	120
GCCGCGCTCT GCATCAGTGA CGGTAAACCA GTCAGCCTGA GCTACCGATG CCCCTGCCGG	180
TTCTTCGAGA GCCACATCGC CAGAGCCAAC GTCAAGCATC TGAAAATCCT CAACACTCCA	240
AACTGTGCCC TTCAGATTGT TGCACGGCTG AAGAACAACA ACAGACAAGT GTGCATTGAC	300
CCGAAATTAA AGTGGATCCA AGAGTACCTG GAGAAAGCTT TAAACAAGTA AGCACAACAG	360
CCCAAAGGAC TTTCAGTAG ACCCCCGAGG AAGGCTGACA TCCGTGGGAG ATGCAAGGGC	420
AGTGGTGGGG AGGAGGGCCT GAACCTGGC CAGGATGGCC GCGGGACAG CACTGACTGG	480
GGTCATGCTA AGGTTTGCCA GCATAAAGAC ACTCCGCCAT AGCATATGGT ACGATATTGC	540
AGCTTATATT CATCCCTGCC CTCGCCCCTG CACAATGGAG CTTTATAAC TGGGGTTTTT	600
CTAAGGAATT GTATTACCCT AACCAGTAG CTTTCATCCCC ATTCTCCTCA TCCTCATCTT	660
CATTTTAAAA AGCAGTGATT ACTTCAAGGG CTGTATTGAG TTGCTTTGG AGCTTCTCTT	720
TGCCCTGGGG CCTCTGGGCA CAGTTATAGA CGGTGGCTTT GCAGGGAGCC CTAGAGAGAA	780
ACCTTCCACC AGAGCAGAGT CCGAGGAACG CTGCAGGGCT TGTCTGCAG GGGGCGCTCC	840
TGCACAGATG CCTTGTCTCTG AGTCAACACA AGATCCGGCA GAGGGAGGCT CCTTTATCCA	900

G TTCAGTGCC AGGGTCGGGA AGCTTCCTTT AGAAGTGATC CCTGAAGCTG TGCTCAGAGA	960
CCCTTTCCTA GCCGTTCCCTG CTCTCTGCTT GCCTCCAAC GCATGCTTCA TCTGACTTCC	1020
GCTTCTCACC TCTGTAGCCT GACGGACCAA TGCTGCAATG GAAGGGAGGA GAGTGATGTG	1080
GGGTGCCCCC TCCCTCTCTT CCCTTTGCTT TCCTCTCACT TGGGCCCTTT GTGAGATTTT	1140
TCTTTGGCCT CCTGTAGAAT GGAGCCAGAC CATCCTGGAT AATGTGAGAA CATGCCTAGA	1200
TTTACCCACA AAACACAAGT CTGAGAATTA ATCATAAACG GAAGTTTAAA TGAGGATTTG	1260
GACCTTGGTA ATTGTCCCTG AGTCCTATAT ATTTCAACAG TGGCTCTATG GGCTCTGATC	1320
GAATATCAGT GATGAAAATA ATAATAATAA TAATAATAAC GAATAGCCA GAATCTTGCC	1380
ATGAAGCCAC AGTGGGGATT CTGGGTTCCA ATCAGAAATG GAGACAAGAT AAAACCTGCA	1440
TACATTCTTA TGATCACAGA CGGCCCTGGT GGTTTTGGT AACTATTTAC AAGGCATTTT	1500
TTTACATATA TTTTGTGCA CTTTTTATGT TTCTTTGGAA GACAAATGTA TTTCAGAATA	1560
TATTTGTAGT CAATTCATAT ATTTGAAGTG GAGCCATAGT AATGCCAGTA GATATCTCTA	1620
TGATCTTGAG CTACTGGCAA CTTGTAAAGA AATATATATG ACATATAAAT GTATTGTAGC	1680
TTTCCGGTGT CAGCCACGGT GTATTTTTC ACTTGGGAATG AAATTGTATC AACTGTGACA	1740
TTATATGCAC TAGCAATAAA ATGCTAATTG TTTCATGCTG TAAAAAAAAA AAAAAAA	1797

(2) S E Q I D N o : 1 0 的资料 :

(i) 序列特征 :

(A) 长度 : 1 4 个碱基对

(B) 类型 : 核酸

(C) 链型 : 双链

(D) 几何结构 : 线性

(ii) 分子类型 : cDNA

(xi) 序列描述 : S E Q I D N o : 1 0

AATTCGCGGC CGCT

14

(2) S E Q I D N o : 1 1 的资料 :

(i) 序列特征 :

(A) 长度 : 1 0 个碱基对

(B) 类型：核酸

(C) 链型：双链

(D) 几何结构：线性

(ii) 分子类型：cDNA

(ix) 特征：

(A) 名称 / 关键词：misc-特征

(B) 位置：1

(D) 其它资料 / 标记 = 磷酸化的

(xi) 序列描述：SEQ ID No : 1 1

AGCGGCCGCG

10

(2) SEQ ID No : 1 2 的资料：

(i) 序列特征：

(A) 长度：2 8 个碱基对

(B) 类型：核酸

(C) 链型：双链

(D) 几何结构：线性

(ii) 分子类型：DNA (合成的)

(xi) 序列描述：SEQ ID No : 1 2

GTAATACGAC TCACTATAGG GGAGAGCT

28

(2) SEQ ID No : 1 3 的资料：

(i) 序列特征：

(A) 长度：2 8 个碱基对

(B) 类型：核酸

(C) 链型：双链

(D) 几何结构: 线性

(ii) 分子类型: DNA (合成的)

(xi) 序列描述 SEQ ID No: 13

CTCCCCTATA GTGAGTCGTA TTA CTGCA

28

(2) SEQ ID No: 14 的资料:

(i) 序列特征:

(A) 长度: 32 个碱基对

(B) 类型: 核酸

(C) 链型: 双链

(D) 几何结构: 线性

(ii) 分子类型: DNA (合成的)

(xi) 序列描述 SEQ ID No: 14

CTAGTCTATA GTGTCACCTA AATCGTGGGT AC

32

(2) SEQ ID No: 15 的资料:

(i) 序列特征:

(A) 长度: 23 个碱基对

(B) 类型: 核酸

(C) 链型: 双链

(D) 几何结构: 线性

(ii) 分子类型: DNA (合成的)

(xi) 序列描述 SEQ ID No: 15

CACGATTTAG GTGACACTAT AGA

23

(2) SEQ ID No: 16 的资料:

(i) 序列特征:

(A) 长度: 44 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 几何结构: 线性

(ii) 分子类型: DNA (合成的)

(xi) 序列描述: SEQ ID No: 16

AATATAGTCG ACCACCAT GA ACGCCAAGGT CGTGGTCGTG CTGG

44

(2) SEQ ID No: 17 的资料:

(i) 序列特征:

(A) 长度: 37 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 几何结构: 线性

(ii) 分子类型: DNA (合成的)

(xi) 序列描述: SEQ ID No: 17

CGGCGGACTA GTTTACTTGT TTAAAGCTTT CTCCAGG

37

(2) SEQ ID No: 18 的资料:

(i) 序列特征:

(A) 长度: 55 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 几何结构: 线性

(ii) 分子类型: DNA (合成的)

(xi) 序列描述: SEQ ID No: 18

GCCGCCACTA GTTTAGTGGT GGTGGTGGTG GTGCTTGTTT AAAGCTTTCT CCAGG 55

(2) SEQ ID No: 19 的资料:

(i) 序列特征:

(A) 长度: 39 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 几何结构: 线性

(ii) 分子类型: DNA (合成的)

(xi) 序列描述: SEQ ID No: 19

CGGCGGACTA GTTCACATCT TGAACCTCTT GTTTAAAGC

39

(2) SEQ ID No: 20 的资料:

(i) 序列特征:

(A) 长度: 57 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 几何结构: 线性

(ii) 分子类型: DNA (合成的)

(xi) 序列描述 SEQ ID No: 20

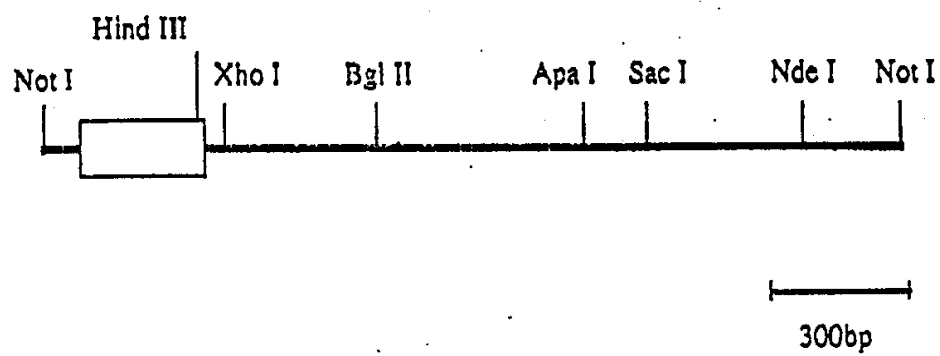
GCCGCCACTA GTTCAGTGGT GGTGGTGGTG GTGCATCTTG AACCTCTTGT TTAAAGC 57

附图的简要描述:

图 1...图 1 列出了人 SDF-1 α cDNA 的限制酶图谱。

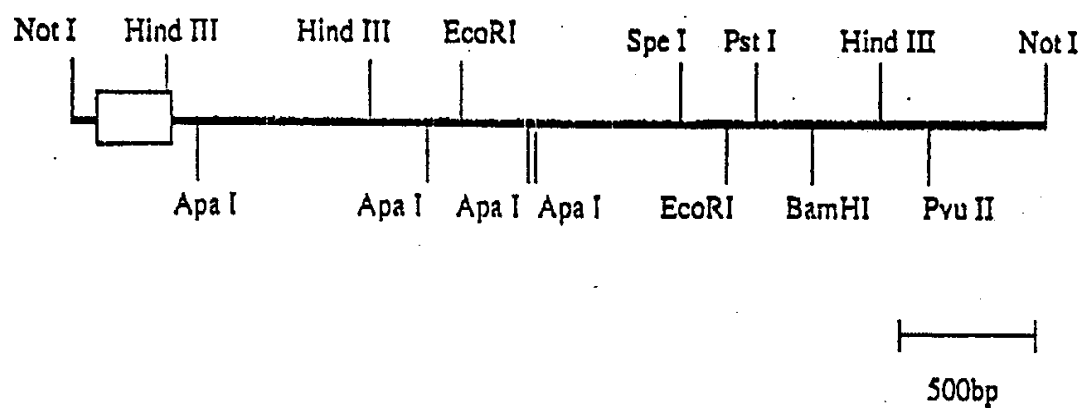
图 2...图 2 列出了人 SDF-1 β cDNA 的限制酶图谱。

说明书附图



Box是指开放的阅读框架

图 1



Box 是指开放的阅读框架

图 2

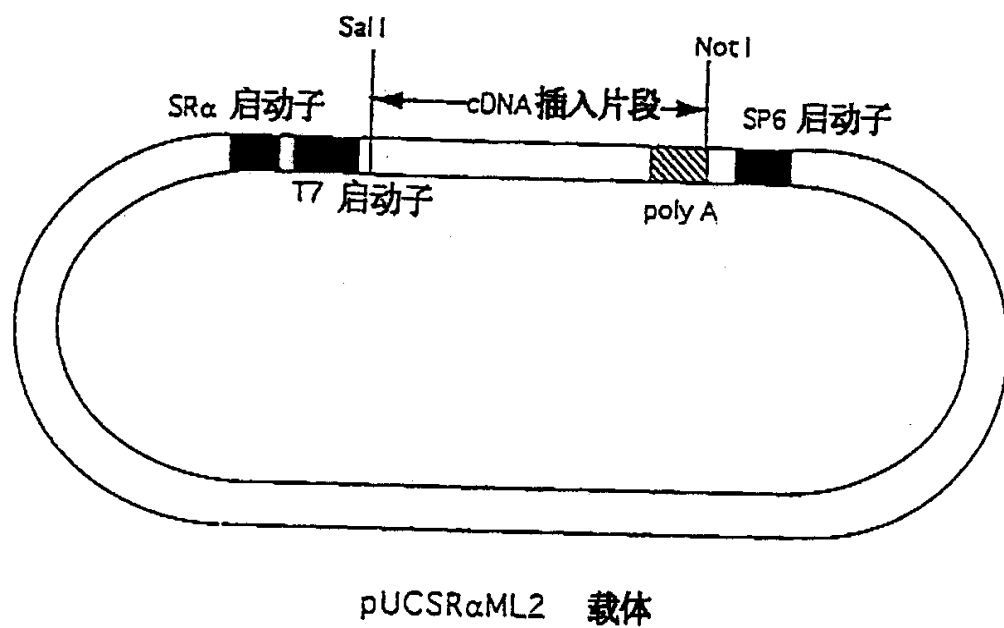


图 3