

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.

C12N 15/12 (2006.01)

C12N 15/62 (2006.01)

C12N 1/21 (2006.01)



# [12] 发明专利说明书

专利号 ZL 96194461.7

[45] 授权公告日 2006年2月8日

[11] 授权公告号 CN 1240836C

[22] 申请日 1996.6.6 [21] 申请号 96194461.7

[30] 优先权

[32] 1995.6.6 [33] US [31] 08/468,846

[86] 国际申请 PCT/US1996/009448 1996.6.6

[87] 国际公布 WO1996/039497 英 1996.12.12

[85] 进入国家阶段日期 1997.12.5

[71] 专利权人 人体基因组科学有限公司

地址 美国马里兰州

[72] 发明人 保罗·S·迈斯纳

丽贝卡·A·富德纳 卫颖飞

马克·D·亚当斯

审查员 李 岚

[74] 专利代理机构 永新专利商标代理有限公司

代理人 林晓红

权利要求书2页 说明书44页 附图8页

[54] 发明名称

转化生长因子  $\alpha$  HI

[57] 摘要

本发明公开了转化生长因子  $\alpha$  HI多肽和编码这种多肽的多核苷酸。本发明也提供了通过重组技术生产这种多肽的方法和这种多肽在治疗方面的用途，这些治疗包括刺激创伤愈合、治疗神经紊乱、治疗视觉障碍、治疗肾和肝脏功能障碍和刺激胚胎发生及血管形成。本发明也公开了抗这种多肽的拮抗剂和其作为治疗剂在治疗瘤形成疾病上的用途。本发明还公开了检测本发明多肽的改变的水平和编码本发明多肽之核酸序列中的突变的诊断测定方法。

1. 一种分离的多核苷酸，该多核苷酸包括选自如下一组的成员：

(a) 一种多核苷酸，该多核苷酸编码如图 1 所示的多肽；

(b) 一种多核苷酸，该多核苷酸编码包含图 1 的 1 至 380 位氨基酸的多肽；

(c) 一种多核苷酸，该多核苷酸编码包含图 1 的 1 至 316 位氨基酸的多肽；

(d) 一种多核苷酸，该多核苷酸编码包含图 1 的 267 至 316 位氨基酸的多肽；

(e) 一种多核苷酸，该多核苷酸编码包含图 1 的 40 至 316 位氨基酸的多肽。

2. 权利要求 1 的多核苷酸，其中所说的多核苷酸是 DNA。

3. 一种包含权利要求 2 的 DNA 的载体。

4. 一种由权利要求 3 的载体转化或转染过的宿主细胞。

5. 一种生产多肽的方法，该方法包括培养权利要求 4 的宿主细胞，使之表达由权利要求 2 的 DNA 编码的多肽，并分离所述的多肽。

6. 一种生产能够表达多肽的细胞的方法，该方法包括用权利要求 3 的载体对细胞进行转化、转染或转导。

7. 一种多肽，所说多肽包括选自如下一组的成员：

(a) 一种多肽，该多肽具有图 1 的推导的氨基酸序列；

- (b) 一种多肽，该多肽包含如图 1 所示的 1 至 380 位氨基酸；
- (c) 一种多肽，该多肽包含如图 1 所示的 267 至 316 位氨基酸；
- (d) 一种多肽，该多肽包含如图 1 所示的 40 至 316 位氨基酸；
- (e) 一种多肽，该多肽包含如图 1 所示的 1 至 316 位氨基酸。

8. 一种权利要求 7 的多肽的抗体。

9. 权利要求 7 的多肽在制备治疗需要 TGF $\alpha$ -HI 之患者的药物组合物中的应用。

10. 权利要求 1 的多核苷酸在制备用于诊断疾病或疾病的易感性的诊断剂中的应用。

11. 权利要求 7 的多肽在制备用于诊断疾病或疾病的易感性的诊断剂中的应用。

## 转化生长因子 $\alpha$ HI

本发明涉及最新鉴别的多核苷酸、由这些多核苷酸编码的多肽、这些多核苷酸和多肽的用途以及这些多核苷酸和多肽的生产方法。本发明的多肽已被推定性地鉴定为转化的生长因子 $\alpha$ 同系物。更具体地说，本发明的多肽已被推定性地鉴定为转化生长因子 $\alpha$ -HI，下文有时称为“TGF $\alpha$ -HI”。本发明也涉及抑制这些多肽的作用的方法。

细胞生长和分化似乎是由多种刺激、抑制和协同因子及激素起始、促进、保持和调节的。细胞自动调节机制的变化和/或障碍似乎是与生长相关的疾病(包括瘤形成)的根本原因。生长模式(modular)因子和各种病理和生理过程有关，这些过程包括：信号转导、细胞通讯、生长和发育、胚胎发生、免疫应答、造血细胞的存活和分化、炎症、组织修复和改造、动脉粥样硬化和癌变。表皮生长因子(EGF)、转化生长因子 $\alpha$ (TGF $\alpha$ )、 $\beta$ -动物纤维素(betacellulin)、双调蛋白和牛痘生长因子等都是各种细胞在正常生理条件或对外源刺激进行应答的情况下生产的生长和分化模式蛋白，并且这些因子也是 EGF 家族的成员。

这些肽生长因子通过自分泌和旁分泌机制影响创伤细胞。它们也对很多组织(如皮肤、角膜和胃肠道)的正常创伤愈合起重要作用，并且所有这些因子实质上享有氨基酸序列的同源性，这种同源性包括三个链内二硫键的保守替换。此外，此家族的所有因子都和分子量为 170,000 的跨膜糖蛋白受体结合，并且激活受体细胞质区的酪氨酸激

酶活性(Buhrow, S. A. 等, 生物化学杂志 258: 7824-7826(1983))。

这种受体由许多典型的细胞表达, 这些细胞包括皮肤角质化细胞、血管内皮细胞和 GI 道的上皮细胞。这些肽生长因子由几种和创伤愈合有关的细胞合成, 这些细胞包括血小板、角质化细胞和激活的巨噬细胞。这些生长因子既与某些细胞的生长及分化的刺激作用(例如, 瘤形成)有关, 也和其它类型细胞的抑制作用有关。

$\beta$ -动物纤维素是 32-kDa 糖蛋白, 这种糖蛋白可能是由一个更大的跨膜前体经蛋白水解断裂加工而成的。 $\beta$ -动物纤维素的羧基末端区和大鼠生长因子 $\alpha$ 的羧基末端区具有 50%的序列相似性。 $\beta$ -动物纤维素是视网膜色素上皮细胞和血管平滑肌细胞中有力的促细胞分裂剂。

双调蛋白是一种双向功能的细胞生长调节因子, 其对新生物细胞的 DNA 合成显示出有力的抑制活性, 却促进某些正常细胞的生长。已经指出双调蛋白具有包括治疗创伤和癌症在内的各种广泛用途。例如, 双调蛋白在体外对几种上皮来源的人类癌细胞系具有有力的抗增殖作用。如美国专利申请 5,115,096 所示, 双调蛋白也诱导人类前皮成纤维细胞的增殖。

TGF $\alpha$ 具有多向性生物作用。TGF $\alpha$ 某些成员的产生是通过许多产生瘤的转化的成纤维细胞合成的(Ciardiallo 等, 细胞生物化学杂志 42: 45-57(1990)), 以及通过多种肿瘤(包括肾部、胸部和鳞状癌、黑素瘤和成胶质细胞瘤)合成(Derynck, R. 等, 癌研究 47:707-712(1987))。通过分析转基因小鼠(其中, 肿瘤细胞表达高水平的 TGF $\alpha$ ), 有直接的证据表明 TGF $\alpha$ 的表达可能是正常细胞转化成其致瘤相应物的促进因子。TGF $\alpha$ 的转基因动物表现出各种致瘤性损伤, 致瘤性损伤取决于

小鼠株系及调节 TGF $\alpha$ 表达之启动子的选择(Sandgren, 等, 细胞, 61:1121-1135(1990))。

TGF $\alpha$ 也在正常的胚胎发育和成年人的生理过程中具有作用(Derynck, R. 癌研究进展, 58: 27-5(1992))。TGF  $\alpha$ 在许多组织(包括皮肤、脑、胃肠粘膜和激活的巨噬细胞)中表达。因此, TGF $\alpha$ 是控制上皮细胞生长的重要因子并且对创伤愈合具有作用(Schreiber 等, 科学, 232: 1250-1253(1986))。

本发明的多肽已经被推定性地鉴定为转化的生长因子 TGF $\alpha$ -HI。由于氨基酸序列和人类 TGF $\alpha$ 同源, 所以进行了这种鉴定。

按照本发明的一个方面, 本发明提供了新的成熟多肽及其生物学活性的并且在诊断上或治疗上有用的片段、类似物和衍生物。本发明的多肽是人类来源的。

按照本发明的另一个方面, 本发明提供了编码本发明的多肽的分离的核酸分子, 包括 mRNA、DNA、cDNA、基因组 DNA 以及其类似物和其生物学活性的并且在诊断上或治疗上有用的片段和衍生物。

按照本发明的另一个方面, 本发明提供了通过重组技术生产这种多肽的方法, 该方法包括培养含有编码本发明多肽的核酸序列的重组原核和/或真核宿主细胞。

按照本发明的另一个方面, 本发明提供了将这些多肽或编码这些多肽的多核苷酸用于治疗的方法, 所述治疗方法例如, 刺激创伤愈合以便在创伤和 AIDS 痴呆后恢复正常的神经功能、治疗视觉障碍、靶向某些细胞、治疗肾和肝脏紊乱和促进毛发发囊发育、刺激血管形成以便治疗烧伤、溃疡和角膜切口以及刺激胚胎发生。

按照本发明的另一个方面，本发明提供了核酸探针，这种核酸探针包含长度足以特异性地与本发明的核酸序列杂交的核酸分子。

按照本发明的另一个方面，本发明提供抗这些多肽的抗体。

按照本发明的另一个方面，本发明提供了本发明多肽的激动剂。

按照本发明的另一个方面，本发明提供了所说多肽的拮抗剂，其可用于抑制这些多肽的作用，例如，治疗角膜炎症、瘤形成(例如，肿瘤、癌症)及牛皮癣。

按照本发明的另一个方面，本发明提供了检测与本发明的多肽过量表达及编码这种多肽之核酸序列中的突变有关的疾病的诊断测定方法。

按照本发明的另一个方面，本发明提供了将这些多肽或编码这些多肽的多核苷酸体外用于与科学研究、DNA合成及DNA载体的人工合成有关的目的的方法。

从本文的教导中，本领域技术人员会清楚本发明的这些和其它方面。

下列附图用来说明本发明的实施方案，而无意于用它们来限制本发明权利要求所包括的范围。

图1描述了TGF $\alpha$ -HI的cDNA序列和相应的推导的氨基酸序列。使用了氨基酸标准单字母缩写。推定的信号序列有下划线，推定的可溶性部分有双下划线。

图2是人类双调蛋白、人类 $\beta$ -动物纤维素、人类表皮生长因子、人类heregulin和人类TGF $\alpha$ -HI(第五排)之间的比较的氨基酸序列同源性的图解。阴影区指保守EGF基元，其显示出在本发明的多肽中是保守的。

按照本发明的一个方面，本发明提供了一种分离的核酸(多核苷酸)，这种核酸编码具有图 1(SEQ ID NO: 2)的推导的氨基酸序列的成熟多肽。

可以从人脑或早期脑组织中得到编码本发明多肽的多核苷酸。本发明的多核苷酸在八周龄胚胎的 cDNA 文库中发现。其在结构上和 TGF $\alpha$  基因家族相关。其包含一个编码 380 个氨基酸残基之多肽的开放读框，其显示出与许多 TGF $\alpha$  基因家族的成员的同源性；这些成员包括 TGF $\alpha$  本身以及其他成员，例如双调蛋白和 cripto。此外，出现在所有成员的特征性基元中的 6 个半胱氨酸残基在 TGF $\alpha$ -HI 中是保守的。

如图 1(SEQ ID NO: 2)所示的本发明的全长多肽具有推定的信号序列，这种信号序列包含图 1(SEQ ID NO: 2)的 1 位氨基酸至 39 位氨基酸，其帮助所述多肽从细胞分泌。所述多肽被进一步加工，加工中图 1(SEQ ID NO: 2)的 40 位氨基酸至 266 位氨基酸从所述多肽上裂解下来，因为这一氨基酸骨架是推定的前体序列。并且，317 位氨基酸至 380 位氨基酸代表推定的跨膜部分，其被认为对指导所述多肽到特定的靶位点以显示下文描述的生物学功能是必需的。跨膜部分也可以从所述多肽中切除，这样，本发明多肽推定的可溶部分包含图 1(SEQ ID NO: 2)的 267 位氨基酸至 316 位氨基酸。

本发明的多核苷酸可以是 RNA 形式或是 DNA 形式，其中 DNA 包括 cDNA、基因组 DNA 和合成 DNA。这种 DNA 可以是双链或单链，如果是单链，其可以是编码链或非编码(反义)链。这种编码成熟多肽的编码序列可以和图 1(SEQ ID NO: 1)所示的编码序列相同；由于遗传密码的冗余性或简并性，这种编码序列也可以是一种不同的编码



序列，其可以和图 1 的 DNA(SEQ ID NO: 1)编码相同的成熟多肽。

编码图 1(SEQ ID NO: 2)的成熟多肽的多核苷酸可以包括：仅仅是编码成熟多肽的编码序列；编码成熟多肽的编码序列和附加的编码序列，如前导或分泌序列或蛋白原序列；编码成熟多肽的编码序列(和可有可无的附加的编码序列)和非编码序列，如内含子或成熟多肽编码序列 5'和/或 3'的非编码序列。

这样，“编码多肽的多核苷酸”这一术语包括只含有多肽编码序列的多核苷酸以及还含有附加的编码和/或非编码序列的多核苷酸。

本发明还涉及上文描述的多核苷酸变体，这种变体编码具有图 1(SEQ ID NO: 2)的推定的氨基酸序列的多肽的片段、类似物和衍生物。这种多核苷酸变体可以是天然产生的多核苷酸等位变体或是非天然产生的多核苷酸变体。

这样，本发明包括编码如图 1(SEQ ID NO: 2)所示的相同成熟多肽的多核苷酸，以及编码如图 1(SEQ ID NO: 2)所示的成熟多肽的片段、衍生物和类似物的多核苷酸变体。这些核苷酸变体包括缺失变体、取代变体、添加或插入变体。

正如上文所表明的，所述多核苷酸可以具有一种编码序列，该序列是图 1(SEQ ID NO: 1)中所示的编码序列的天然产生的等位变体。本领域已知等位变体是多核苷酸序列的另一种形式，其可以具有一个或多个核苷酸的取代、缺失或添加，而实质上不改变所编码的多肽的功能。

本发明也包括这样的多核苷酸，其中所说的成熟多肽的编码序列可以在相同的阅读框架中与帮助从宿主细胞表达和分泌多肽的多核苷酸序列融合，所述的多核苷酸序列如作为分泌序列在控制多肽从细

胞转运中起作用的前导序列。具有前导序列的多肽是前蛋白(preprotein), 并且可以具有由宿主细胞切割形成成熟形式的多肽的前导序列。这种多核苷酸也编码蛋白原(proprotein), 这种蛋白原是添加有附加的 5' 氨基酸残基的成熟蛋白。具有原序列(prosequence)的成熟蛋白是蛋白原, 是一种无活性的蛋白形式。一旦切除原序列, 留下的就是有活性的成熟蛋白。

这样, 本发明的多核苷酸例如可以编码一种成熟蛋白或编码具有原序列的蛋白质或编码既有原序列又有前序列(presequence)(前导序列)的蛋白质。

本发明的多核苷酸也可以具有在读框中与标记序列融合的编码序列, 所述的标记序列使得可以纯化本发明的多核苷酸。在细菌宿主的情况下, 所述的标记序列可以是由 pQE-9 载体提供的六组氨酸标记, 其提供来纯化融合进标记的成熟多肽, 或例如当使用哺乳动物细胞(如 COS-7 细胞)时, 标记序列可以是血细胞凝集素(HA)标记。所说的 HA 标记相应于来源于流感血细胞凝集素蛋白质的一种表位(Wilson, I. 等, 细胞, 37:767(1984))。

术语“基因”是指和产生多肽链有关的 DNA 区段; 其包括在编码区前面的和后面的区域(前导区和尾随序列)以及在各个编码区段(外显子)之间的间插序列(内含子)。

全长 TGF $\alpha$ -HI 基因的片段可以用作 cDNA 文库的杂交探针, 用来分离全长基因和与此基因有高度的序列类似性或类似生物活性的其它基因。这种类型的探针优选地是具有至少 30 个碱基, 并且可以含有, 例如 50 个或更多的碱基。所说的探针也可以用于鉴别相应于全长转录物的 cDNA 克隆和基因组克隆或含有包含调节和启动子区、

外显子和内含子的完整 TGF $\alpha$ -HI 基因克隆。筛选的实例包括通过利用已知 DNA 序列合成寡核苷酸探针来分离基因的编码区。具有互补于本发明的基因序列之序列的标记可以用来筛选人类 cDNA、基因组 DNA 或 mRNA 文库，以确定探针和哪些文库成员杂交。

本发明还涉及与以上所描述的序列杂交的多核苷酸(条件是两个序列之间具有至少 70%，优选地具有至少 90%，更优选地具有至少 95%的相同性)。本发明特别涉及在严格条件下与以上所述的多核苷酸杂交的多核苷酸。如本文所使用的，术语“严格条件”指仅在序列间具有至少 95%，优选地具有至少 97%的同源性时杂交才可以发生。在一个优选的实施方案中，与以上所述的多核苷酸杂交的多核苷酸编码这样一种多肽，其实质上保持与由图 1(SEQ ID NO: 1)的 cDNA 编码的成熟多肽相同的生物学功能或活性。

此外，所述多核苷酸可以具有至少 20 个碱基，优选地是至少 30 个碱基，更优选地是至少 50 个碱基，其与本发明的多核苷酸杂交，并且具有如上文所述的相同性，可以保留或不保留活性。例如，这种多核苷酸可以用作 SEQ ID NO: 1 多核苷酸的探针，例如用于回收多核苷酸或作为诊断探针或作为 PCR 引物。

这样，本发明涉及与编码图 1(SEQ ID NO: 2)多肽之多核苷酸具有至少 70%相同性，优选地至少 90%相同性并且更优选地至少 95%相同性的多核苷酸及其片段(这种片段具有至少 30 个碱基，优选地至少 50 个碱基)和这些多核苷酸编码的多肽。

本发明还涉及具有图 1(SEQ ID NO: 2)的推导的氨基酸序列的多肽，以及这种多肽的片段、类似物和衍生物。

术语“片段”、“衍生物”和“类似物”，当有关图 1(SEQ ID

NO: 2)的多肽时,指基本上保持与这样的多肽相同的生物功能或活性的多肽。这样,类似物包括蛋白原,这种类似物可以由蛋白原部分切除进而生产活性的成熟多肽。

本发明的多肽可以是重组多肽,天然多肽或合成多肽,优选地是重组多肽。

所说的图 1(SEQ ID NO: 2)的多肽的片段、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取代(优选地是保守氨基酸残基取代)并且取代的氨基酸残基可以是也可以不是由遗传密码子编码的氨基酸残基;或(ii)这样一种,其中一个或多个氨基酸残基包含取代基;或(iii)这样一种,其中成熟多肽与另一种化合物融合,所述化合物如增加多肽半衰期的化合物(例如聚乙二醇);或(iv)这样一种,其中附加氨基酸与成熟多肽融合,例如前导或分泌序列或用来纯化成熟多肽的序列或原序列。通过本文的阐述,可以认为这样的片段、衍生物以及类似物在本领域技术人员知识范围之内。

本发明优选地是提供分离形式的多肽和多核苷酸,并且优选地是将所述多肽和多核苷酸纯化成为同质性的。

术语“分离的”意指所述的物质脱离了其原始环境(例如,天然环境,如果其是天然产生的)。例如,一种存在于活的动物中的天然产生的多核苷酸或多肽不是分离的,但是与天然系统中某些或全部共存的物质分开的相同的多核苷酸或多肽是分离的。这样的多核苷酸可以是载体的一部分,和/或这样的多核苷酸或多肽可以是组合物的一部分,其仍然是分离的,这是因为这种载体或组合物不是其天然环境的一部分。

本发明的多肽包括图 1(SEQ ID NO: 2)的多肽(特别是成熟多肽)以及和图 1(SEQ ID NO: 2)的多肽具有至少 70%的相似性(最好是 70%的相同性),更优选地是 90%的相似性(最好是 90%的相同性),最优选地是 95%的相似性(最好是 95%相同性)的多肽,也包括这些多肽的部分,这种多肽的部分通常包含至少 30 个氨基酸,并且优选地是至少 50 个氨基酸。

如本领域所熟知的,两个多肽之间的“相似性”是通过比较一个多肽和另一个多肽序列的氨基酸序列和其保守氨基酸取代来确定的。

本发明多肽的片段或部分通过肽合成可以用于生产相应的全长多肽;因此,此片段可以用作生产全长多肽的中间体。本发明多核苷酸的片段或部分可以用于合成本发明的全长多核苷酸。

本发明也涉及包含本发明的多核苷酸的载体和用本发明的载体经基因工程生产的宿主细胞,以及经重组技术生产本发明的多肽的方法。

宿主细胞是用本发明的载体经基因工程(转导、转化或转染)产生的,所说的载体可以是克隆载体或表达载体。该载体可以是例如,质粒、病毒颗粒和噬菌体等形式。工程宿主细胞可以在改良的适于激活启动子、选择转化体或扩增本发明的基因的常规营养培养基中培养。培养条件,例如温度和 pH 值等,是以前用于表达选择的宿主细胞的那些,对普通技术人员是显而易见的。

本发明的多核苷酸可以用来经重组技术生产多肽。这样,例如,多核苷酸可以包含在各种用于表达多肽的表达载体的任何一种中。这样的载体包括染色体、非染色体和合成 DNA 序列,例如 SV 40 衍生

物；细菌质粒；噬菌体 DNA；杆状病毒；酵母质粒；从质粒和噬菌体 DNA 组合衍生的载体、病毒 DNA(如牛痘、腺病毒、家禽痘病毒、和假狂犬病病毒)。然而，任何其它载体也可以使用，只要其在宿主中可复制和稳定。

可以用多种方法将合适的 DNA 序列插入到载体中。一般来说，用本领域已知的方法将 DNA 序列插入到适当的限制性核酸内切酶位点。这样的方法和其它方法被认为在本领域技术人员的知识范围内。

在表达载体中的所说的 DNA 序列是可操作地连接到适当的指导 mRNA 合成表达控制序列(启动子)上的。这样的启动子的代表性例子可以提到的是: LTR 或 SV 40 启动子，大肠杆菌的 *lac* 或 *trp*、噬菌体  $\lambda$   $P_L$  启动子，和已知在原核或真核细胞或它们的病毒中控制基因表达的其它启动子。所说的表达载体也包含用于翻译起始和转录终止的核糖体结合位点。该载体也可以包含供扩增表达的合适的序列。

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特征，例如真核细胞培养物的二氢叶酸还原酶或新霉素抗性，或例如大肠杆菌中的四环素和氨苄青霉素抗性。

包含以上所述的适当的 DNA 序列以及适当的启动子或控制序列的载体可以用于转化适当的宿主，以使其能够表达蛋白质。

作为合适宿主的代表性例子，这里可以提到的是：细菌细胞，如大肠杆菌、链霉菌、鼠伤寒沙门氏菌；真菌细胞，如酵母；昆虫细胞，如 *Drosophila S2* 和 *Spodoptera Sf9*；动物细胞，如 CHO、COS 或 Bowes 黑素瘤；腺病毒；植物细胞等。通过本文的阐述，对适当的宿主的选择在本领域技术人员的知识范围之内。

更具体地说，本发明也包括重组构建体，该构建体包含以上广泛描述的一种或多种序列。该构建体包含载体，如质粒或病毒的载体，该载体已正向或反向插入了本发明的序列。在这一实施方案的更为理想的情况下，构建体还包含可操作连接到所述序列上的调节序列，包括，例如，启动子。大量适合的载体和启动子是本领域技术人员已知的，并且是通过商业途径可获得的。以举例的方式给出下列载体：细菌：pQE70、pQE60、pQE-9(Qiagen)、pBS、pD10、phagescript、psiX174、pbluescript SK、pbsks、pNH8A、pNH16a、pNH18A、pNH46A(Stratagene)、ptrc99a、pKK223-3、pKK233-3、pDR540、pRITS(Pharmacia)；真核：pWLNEO、pSV2CAT、pOG44、pXT1、pSG(Stratagene)、pSVK3、pBPV、pMSG、pSVL(Parmacia)。然而，任何其它质粒或载体都可以使用，只要它们在宿主中可复制和稳定。

可以用 CAT(氯霉素转移酶)载体或其它带有选择性标记的载体从任何所需的基因选择启动子区。两种合适的载体是 pKK232-8 和 pCM7。特别提到的细菌启动子包括 lacI、lacZ、T3、T7、gpt、 $\lambda P_R$ 、 $P_L$ 、和 trp。真核生物启动子包括 CMV 立即早期、HSV 胸苷激酶、早期和晚期 SV40、得自逆转录病毒的 LTRs 和小鼠金属硫蛋白-I。对适当的载体与启动子的选择在本领域普通技术人员的水平之内。

在另一个实施方案中，本发明涉及包含以上所述构建体的宿主细胞。所说的宿主细胞可以是高等真核细胞(如哺乳动物细胞)，或低等真核细胞(如酵母细胞)，或宿主细胞可以是原核细胞(如细菌细胞)。可以由磷酸钙转染、DEAE-葡聚糖介导的转染，或电穿孔有效地将构建体引入到宿主细胞中(Davis, L., Dibner, M., Battey, I., 分子生物学中的基本方法(1986))。

宿主细胞中的构建体可以用来以常规方式生产由重组序列编码的基因产物。此外，本发明的多肽可以由常规肽合成器合成生产。

成熟蛋白质可以在哺乳动物细胞、酵母、细菌或其它细胞中在适当的启动子控制下表达。采用来源于本发明的 DNA 构建体的 RNA，无细胞翻译系统也可以用来生产这种蛋白质。Sambrook 等，分子克隆：实验室手册，再版，冷泉港, N.Y., (1989)(本文一并参考)描述了与原核和真核宿主一起使用的合适的克隆和表达载体。

高等真核生物编码本发明的多肽的 DNA 的转录被插入到载体中的增强子序列增强。增强子是 DNA 的顺式作用元件，通常约 10 至 300 bp，作用在启动子上增加其转录。例子包括复制起点后侧 100 至 270 bp 上的 SV 40 增强子、细胞肥大病毒早期启动子增强子、复制起点后侧上的多形瘤增强子以及腺病毒增强子。

一般来说，重组表达载体包括复制起点和允许宿主细胞转化的选择性标记(例如，大肠杆菌的氨苄青霉素抗性基因和啤酒糖酵母 TRP1 基因)以及源于高表达基因的指导下游结构序列转录的启动子。这样的启动子可以是来自编码糖酵解酶(例如 3-磷酸甘油酸激酶(PGK))、 $\alpha$ -因子、酸性磷酸酶或热休克蛋白质等的操纵子。异源结构序列以合适的方式(phase)与翻译起始和终止序列装配，优选地，与能够指导翻译的蛋白质分泌进周质空间或细胞外培养基的前导序列装配。异源序列可以也可以不编码融合蛋白，这种蛋白质包括赋予所需特征的 N-末端鉴别肽，所需特征例如，表达的重组产物稳定或简化纯化步骤。

通过在具有功能性启动子的可操作读框(reading phase)中插入带有合适的翻译起始和终止信号的编码所需蛋白质的结构 DNA 序列来构建用于细菌的有用的表达载体。所说载体包含一个或多个表型选择



性标记和复制起点，以在宿主来保证保持载体和需要时提供扩增。适合转化的原核宿主包括大肠杆菌、枯草芽孢杆菌、鼠伤寒沙门氏菌、假单胞菌属、链霉菌属和葡萄球菌属的各种(当然其它的也可以选择来使用)。

作为一个代表性的但不是限制性的例子，用于细菌的有用的表达载体可以包含源于市售质粒(包含众所周知的克隆载体 pBR322(ATCC 37017)的遗传元件)的选择性标记和细菌复制起点。这样的市售载体包括，例如，pKK223-3(Pharmacia Fine 化学品公司，Uppsala，瑞典)和 GEM1(Promega Biotec，Madison，WI，美国)。这些 pBR322 “骨架”片段与适当的启动子和待表达的结构序列组合。

在合适的宿主菌株转化和合适的宿主菌株生长至适当的细胞密度之后，用合适的方法(例如温度变换或化学诱导)诱导选择的启动子，将细胞培养另外一段时间。

典型地经离心收获细胞，经物理或化学方法破碎细胞，保持所形成的粗产物用于进一步的纯化。

可以经任何常规的方法破碎用于表达蛋白质的微生物细胞，所述方法包括冻融循环、超声处理、机械破碎或使用细胞裂解剂，这些方法是本领域技术人员熟知的。

各种哺乳动物细胞培养系统也可以用于表达重组蛋白质。哺乳动物表达系统的例子包括由 Gluzman(细胞，23:175(1981))描述的猴肾成纤维细胞 COS-7 细胞系和其它能够表达相容载体的细胞系，例如，C127、3T3、CHO、HeLa 和 BHK 细胞系。哺乳动物表达载体包含复制起点、适合的启动子和增强子，也可以包含任何必需的核糖体结合位点、聚腺苷酸化位点、剪接供体和受体位点、转录终止序列和 5'

侧翼非转录序列。得自 SV40 剪接和聚腺苷酸化位点的 DNA 序列可以用来提供所需的非转录遗传元件。

所说的多肽可以用多种方法从重组细胞培养物回收和纯化，所述方法包括硫酸铵或乙醇沉淀、酸抽取、阴离子或阳离子交换层析、磷纤维素层析、疏水相互作用层析、亲和性层析、羟基磷灰石层析和植物凝集素层析。需要在完成成熟蛋白质的构型中可以使用蛋白质再折叠步骤。最后，可以使用高效液相层析(HPLC)作为最后的纯化步骤。

本发明的多肽可以是天然纯化的产物，或化学合成方法的产物，或经重组技术从原核或真核宿主(例如细菌、酵母、高等植物、培养的昆虫和哺乳动物细胞)生产的。依据重组生产方法中使用的宿主，本发明的多肽可以是糖基化的或可以是非糖基化的。本发明的多肽也可以包括起始甲硫氨酸氨基酸残基。

本发明的多核苷酸和多肽可以用作人类疾病的治疗和诊断的研究试剂和材料。

本发明的多肽可以用于确定受体的特征。目前，EGF 家族的受体包括 4 个 EGF 受体，称为 EGFR1、EGFR2、EGFR3、EGFR4。EGFR2 也可以称为 ERB-2，此分子在各种诊断和治疗适应症方面是有用的 (Prigent, S. A.和 Lemoine, N. R., 生长因子研究进展, 4:1-24(1992))。TGF $\alpha$ -HI 多肽可能是一种或几种这些受体以及鉴定的新 EGF 类型受体的配体。TGF $\alpha$ -HI 的使用有助于这些受体的鉴定、定性和克隆。例如，EGF 受体基因代表禽成红细胞增多病病毒的 v-erb-B 癌基因的细胞同系物。EGF 受体的过量表达或该蛋白质激酶调节片段的缺失可以导致细胞的致癌转化 (Manjusri, D.等, 人类细胞因子, 364 和

381(1991))。

本发明的多肽也可以用于因创伤或其它破坏性的病理(如 AIDS 痴呆、老年性痴呆等等)降低的神经功能的恢复或提高。已经发现 TGF $\alpha$ 和其同系物是大部分脑中的 EGF /TGF $\alpha$ 受体的最丰富的配体(Kaser 等, 脑研究和分子脑研究: 16: 316-322,(1992))。与 EGF(其只存在于较小的分散的区域中)相比, TGF $\alpha$ 在脑的各种区域中似乎有广泛的分布, 表明 TGF $\alpha$ 可能在脑组织中具有生理作用。脑中这些 TGF $\alpha$ 的众多受体位点说明 TGF 在促进正常脑细胞分化和发挥功能方面具有重要作用。因此, 在神经功能降低的情况下, 施用本发明的多肽可以刺激脑并提高适当的生理功能。

TGF $\alpha$ -HI 和其可溶性形式也可以用于治疗视觉障碍, 例如, 角膜炎症。各种实验表明 TGF $\alpha$ -HI 基因家族的成员和这种病理有关。最近的论文概括了一些和这些生长因子在眼疾病中所起的作用相关的数据(Mann 等, 细胞 73:249-261(1993))。最近的实验显示缺乏 TGF $\alpha$ 基因的小鼠由于白细胞和其它细胞对眼物质 propria 的渗透显示出角膜炎症。

此外, TGF $\alpha$ 生长因子对其靶细胞的特异性可以用作破坏靶细胞的机制。例如, 可以通过各种方法将 TGF $\alpha$ -HI 或其可溶性形式与毒素分子(例如灭活靶细胞的放射药物)偶合。这些生长因子-毒素融合物杀死靶细胞(和在某些情况下通过各种“旁观者”效应杀死邻区细胞)。Mesri 等(生物化学杂志 268: 4853-62(1993))发表了这种毒素融合基因的最新例子。可以将 TGF $\alpha$ -HI 和相关的分子在脂质体中形成胶囊并将其和识别并结合肿瘤或细胞特异性抗原的抗体结合, 从而提供“导向”细胞的方法。

因为 EGF 家族成员在转化的细胞中显示出抗增值作用，所以 TGF $\alpha$ -HI 可以以同样的方式用作抗致瘤的化合物。对于体内使用，主题多肽可以以各种方式施用，这些方式包括但不限于：注射、灌输、局部地、肠胃外等等。可以以任何生理上可接受的载体进行施用，这些载体包括：磷酸缓冲液盐水、盐水、无菌水等等。

因为已经发现在肾中有这些生长因子的表达，所以 TGF $\alpha$ -HI 也可以用于治疗某些肾功能障碍。这样，这些因子对此器官正常生理的维持是必需的。

因为 TGF $\alpha$ 和其同系物及肝细胞生长因子在部分肝切除和急性肝细胞坏死之后，触发肝细胞的再生，所以所说的治疗也和肝脏的再生或肝脏机能障碍有关(Masuhara, M.等, 肝脏学 16:1241-1249(1992))。

与 TGF $\alpha$ -HI 有关的有重要意义的治疗涉及创伤愈合。本发明的组合物可以用于治疗实质上包括所有的皮创伤、角膜创伤和身体上皮排列穴器官伤害在内的各种创伤。适于治疗的创伤包括由外伤(如烧伤、擦伤、刀伤)以及外科手术(如外科切口和皮肤移植)导致的创伤。适于用本发明的多肽治疗的其它病症包括慢性疾病(如慢性溃疡, 糖尿病溃疡)和其它非愈合(营养)病症。

TGF $\alpha$ -HI 或其可溶性片段可以掺入到生理上可接受的载体中以便用于感染的区域。所说的载体的特性变化很大，这种变化依赖于预期施用的位点。对于在皮肤上施用，乳油和软膏基质通常是优选的；合适的基质包括：羊毛脂、Silvadene(Marion)(特别是对于治疗烧伤)、Aquaphor(Duke 实验室, South Norwalk, Conn.)等等。如果需要，将包含 TGF $\alpha$ -HI 的组合物掺入绷带和其它的创伤包扎物中以便使伤口持续地暴露在所说的肽下。气溶胶的应用也找到了用途。

TGF $\alpha$ -HI 在治疗组合物中的浓度不是关键的但其应该足以诱导上皮细胞增值。可以局部地对感染区施用所说的组合物，一般作为滴眼剂施用到眼部或作为乳油、软膏或洗剂施用到皮肤上。对于眼部，需要经常治疗，通常以 4 小时或更少的间隔施用。在皮肤上，在愈合的过程中需要在感染区持续保持治疗组合物，每天施用治疗组合物 2 至 4 次或更多次。

本发明多肽的使用量随施用方式而变化，其它活性化合物等的使用量通常在 1 $\mu$ g-100 $\mu$ g 的范围内。本发明多肽可以和生理上可接受的载体(如盐水、磷酸缓冲液盐水等等)一起使用。所述化合物的使用量根据体外细胞的反应和实验动物对实验多肽或含有实验多肽的制剂的反应由经验确定。

TGF $\alpha$ -HI 或其可溶性片段可以用于血管形成、骨吸收、免疫应答和染色体接合及神经元效应器功能的调节。TGF $\alpha$ -HI 也可以用于花生四烯酸级联的调节。

TGF $\alpha$ -HI 或其可溶性片段也可以用于与终末分化相关的应用中。许多 TGF $\alpha$ 因子和它们的同系物在它们的靶细胞中诱导终末分化。可以通过施用所说的因子并诱导靶细胞死亡在体内利用这种特性。这种用药方式是考虑到和医学上不希望的细胞类型(如癌症和其它增值障碍(如炎症、牛皮癣等等))的过度增值有关的紊乱。除了体内施用，也有许多允许体外施用的情况。例如，可以通过用所说的生长因子和/或其衍生物治疗所说的细胞以从骨髓中清除不需要的细胞群。

所说的应用也涉及脱毛、毛发损失和其它影响毛发发囊发育的皮肤病症。若干证据表明这些情况和 TGF $\alpha$ 生长因子有关。经基因工程操作 TGF $\alpha$ 基因中含有无效突变的“剔除(knockout)”小鼠显示出和毛

发合成的数量及质量有关的畸变。此外，小鼠作图研究已经表明一些影响毛发生长的突变位于  $TGF\alpha$  基因基因座上(Mann 等, 细胞 73: 249-261(1993))。  $TGF\alpha$ -HI 或其衍生物的局部或全身应用可以用于治疗脱发及毛发损失的病症，这些权利要求在本发明的范围之内。

通过全身临床施用  $TGF\alpha$ -HI 生长因子，某些疾病的病理情况已部分地或完全地改善。所说的施用可以以基因治疗的形式进行，或通过施用由  $TGF\alpha$ -HIDNA 重组构建体或肽化学合成法合成的肽或蛋白质来完成(Woo 等, 蛋白质工程 3:29-37(1989))。

本发明提供了筛选化合物以鉴定本发明多肽的激动剂或拮抗剂化合物的方法。例子是将表达  $TGF\alpha$ -HI 受体的哺乳动物细胞或膜制剂和潜在的化合物一起培养，检测所说的化合物由该受体产生第二信号的能力以确定其是否是有效的激动剂。第二信使系统包括但不限于：cAMP 鸟苷酸环化酶、离子通道或磷酸肌醇水解作用。有效的拮抗剂由上文描述的方法确定，其中检测到的拮抗剂化合物和所说的受体结合但不能引起第二信使反应，进而阻断  $TGF\alpha$ -HI 与受体结合。

另一种鉴定对本发明多肽之受体特异性的潜在的拮抗剂的方法是竞争测定法，这种方法包括分离过量表达本发明多肽之受体的原生质膜，例如，人类 A431 癌细胞。将在介质(体积大约是 10ml)中系列稀释的含有 10nM  $^{125}I$ - $TGF\alpha$ -HI 的实验样品加入到含有潜在的拮抗剂化合物的 5ml 原生质膜中，并在 4°C 下培养 4 小时。将反应混合物稀释并立即通过微孔滤器。然后迅速洗涤滤器。并在  $\gamma$  计数器中测量结合的放射性。然后测量结合的  $TGF\alpha$ -HI 的量。在不存在所说的化合物的情况下进行对照分析以确定拮抗剂是否能降低结合的  $TGF\alpha$ -HI 的量。

潜在的拮抗剂化合物包括抗体或某种情况下和所说的多肽结合的寡肽。另外，潜在的拮抗剂可以是一种密切相关的蛋白质，这种蛋白质和所说的多肽的失活形式的受体结合，进而阻止了本发明多肽的活性。

另一个潜在的拮抗剂化合物是采用反义技术制备的反义构建体。反义技术可以通过三螺旋形成或反义 DNA 或 RNA 来控制基因的表达，这两种方法都基于多核苷酸和 DNA 或 RNA 的结合。例如，编码本发明的成熟多肽的多核苷酸序列的 5' 编码部分可以用来设计长度约 10 至 40 个碱基对的反义 RNA 寡核苷酸。一种 DNA 寡核苷酸被设计成与转录所涉及的基因区互补(三螺旋-参见 Lee 等，核酸研究，6:3073(1979); Cooney 等，科学，241:456,(1988); 和 Dervan 等，科学，251 :1360(1991))，进而阻止本发明多肽的转录和生产。反义 RNA 寡核苷酸在体内和 mRNA 杂交，并阻断 mRNA 分子翻译成为本发明的多肽(反义-Okano, J. Neurochem., 56 :560(1991); 寡脱氧核苷酸作为基因表达的反义抑制剂(CRC 出版社, Boca Raton, FL(1988))。以上描述的寡核苷酸可以传送到细胞中，以便可以体内表达反义 RNA 和 DNA，抑制本发明多肽的生产。

拮抗剂化合物包括小分子，这些小分子和本发明的多肽结合并在受体位点阻断其活性，这样，就阻断了正常的生物活性。小分子也可以和所说的多肽的受体结合，结果阻止了多肽和受体的结合。小分子的例子包括但不限于小肽或类肽分子。

拮抗剂可以用来治疗瘤形成，例如，癌和肿瘤。已知小鼠中肿瘤细胞分泌或产生 EGF 家族成员的抑制作用造成肿瘤的衰退。

本发明多肽的拮抗剂也可以治疗性地用于治疗某些皮肤病症，例

如，牛皮癣。已经发现所说的生长因子家族成员提高的表达水平在取自疾病(如牛皮癣损伤)的皮肤活组织检查中将再提高(Cook, et al., 癌研究, 52:3224-3227(1992))。所说的拮抗剂和药学可接受的载体(如下文描述的)一起用于组合物中。

本发明的多肽和激动剂或拮抗剂可以与合适的药物载体组合使用。这样的组合物包含治疗有效量的所说多肽和药学上可接受的载体或赋形剂。这样的载体包括但不限于盐水、缓冲盐水、右旋糖、水、甘油、乙醇和它们的组合物。其配方应该适宜于施用的方式。

本发明也提供了药物包或试剂盒，它们包含一个或多个填装有本发明的药物组合物的一种或多种成份的容器。这种容器中还可以包括管理药物和生物制品制造、使用或销售的政府机构规定形式的告示，这一告示反映了制造、使用或销售人类使用品的政府机构的同意。此外，本发明的多肽或组合物可以与其它治疗化合物结合使用。

所述药物组合物可以以方便的方式施用，所述方式例如口服、局部、静脉内、腹膜内、肌内、皮下、鼻内或真皮内途径施用。所说的药物组合物以治疗和/或预防特定疾病的有效量施用。一般来说，它们以至少大约 10 微克/千克体重的量施用，在大多数情况下，它们以不超过每天大约 8 毫克/千克体重的量施用。在大多数情况之下，考虑用药途径和病症等因素，剂量从每日大约 10 微克/千克到 1 毫克/千克体重。

多肽和多肽形式的激动剂和拮抗剂，可以依据本发明通过体内表达这样的多肽来利用，这常被称作“基因治疗”。

这样，例如，可以体外对患者细胞用编码多肽的多核苷酸(DNA 或 RNA)进行基因工程操作，用工程细胞向被治疗的患者提供所说的



多肽。这样的方法是本领域众所周知的，并且由本文的描述也显而易见。例如，可以用包含编码本发明的多肽的 RNA 的逆转录病毒颗粒对细胞进行基因工程操作。

同样地，可以通过例如本领域已知的方法体内对细胞进行基因工程操作，以便体内表达多肽。例如，将包装(packaging)细胞用含有编码本发明多肽的 RNA 的逆转录病毒质粒载体转导，以便包装细胞能产生含有兴趣基因的传染性病毒颗粒。可以将这种生产细胞施用给患者以便体内将细胞基因工程化并体内表达所说的多肽。经本发明的描述，通过这种方式施用本发明多肽的这些或其它方法对本领域技术人员是清楚的。

可以获得上文描述的反转录病毒质粒载体的反转录病毒包括但不限于：莫洛尼氏鼠白血病病毒、脾坏死病毒、反转录病毒如劳氏肉瘤病毒、Harvey 肉瘤病毒、禽类白血病病毒、长臂猿猩猩白血病病毒、人类免疫缺陷病毒、腺病毒、骨髓增殖肉瘤病毒、和乳房肿瘤病毒。在一个实施方案中，反转录病毒质粒载体来自莫洛尼氏鼠白血病病毒。

所述载体包含一个或多个启动子。可以使用的合适的启动子包括但不限于：反转录病毒 LTR；SV 40 启动子；和人类巨细胞病毒(CMV)启动子(Miller, 等, 生物技术, Vol. 7, No. 9, 980-990(1989)描述); 或其它任何启动子(例如真核细胞启动子，如包括但不限于组蛋白、pol III 和  $\beta$ -肌动蛋白启动子)。使用的其它病毒启动子包括但不限于：腺病毒启动子、胸苷激酶(TK)启动子和 B19 细小病毒启动子。通过本文的描述，合适的启动子的选择在本领域技术人员的知识范围内。

编码本发明多肽的核酸序列在合适的启动子控制下。可以使用的

合适的启动子包括但不限于：腺病毒启动子(如腺病毒主要晚期启动子)；或异源启动子(如巨细胞病毒(CMV)启动子)；呼吸合胞体病毒(RSV)启动子；可诱导的启动子(如 MMT 启动子、金属硫蛋白启动子)；热休克启动子；清蛋白启动子；ApoAI 启动子；人类珠蛋白启动子；病毒胸苷激酶启动子(如单纯疱疹胸苷激酶启动子)；反转录病毒 LTRs(包括上文描述的修饰的反转录病毒 LTRs)； $\beta$ -肌动蛋白启动子；和人类生长激素启动子。所说的启动子也可以是控制编码所述多肽之基因的天然启动子。

使用反转录病毒质粒载体转导包装细胞系以便形成生产细胞系。可以被转染的包装细胞的例子包括但不限于：PE501、PA317、 $\psi$ -2、 $\psi$ -AM、PA12、T19-14X、VT-19-17-H2、 $\psi$ CRE、 $\psi$ CRIP、GP+E-86、GP+envAml2 和 DAN 细胞系(Miller, 人类基因治疗, Vol. 1, pgs. 5-14(1990)描述的, 其全部内容本文一并参考)。可以通过本领域任何已知的方法用所述载体转导包装细胞。这些方法包括但不限于：电穿孔、使用脂质体和  $\text{CaPO}_4$  沉淀。另外, 反转录病毒质粒载体可以包囊在脂质体中, 或和脂类偶合, 然后施用到宿主中。

生产细胞系产生传染性的反转录病毒载体颗粒, 这种颗粒包含编码所述多肽的核酸序列。然后可以使用这些反转录病毒载体颗粒体内或体外转导真核细胞。转导的真核细胞将表达编码所述多肽的核酸序列。可被转导的真核细胞包括但不限于：胚胎干细胞、胚胎癌细胞、以及造血干细胞、肝细胞、成纤维细胞、成肌细胞、角质化细胞、内皮细胞和支气管上皮细胞。

本发明也涉及用本发明的基因作为诊断剂。本发明的基因的突变形式的检测使得可以诊断源于本发明的多肽表达不足

(underexpression)的疾病或疾病的易感性,例如,不适当(improper)的伤口治愈、不适当的神经功能、视觉障碍、肾和肝功能障碍、毛发发囊发育、血管形成和胚胎形成。

可以用各种技术在 DNA 水平上检测具有本发明的人基因突变的个体。可以从患者的细胞(如血液,尿,唾液,组织活组织检查和尸体解剖材料)获得用于诊断的核酸。基因组 DNA 可以直接用于检测,或在分析前用 PCR 酶促扩增(Saiki 等,自然,324:163-166(1986))。RNA 或 cDNA 也可以用于相同的目的。作为一个例子,可以用与编码本发明多肽的核苷酸互补的 PCR 引物鉴别和分析其突变。例如,可以通过与正常基因型比较扩增产物大小上的改变来检测缺失或插入。点突变可以经扩增的 DNA 与放射性标记的 RNA 或放射性标记的反义 DNA 序列杂交鉴别。经核糖核酸酶 A 消化,或从熔点温度的不同辨别完美配对的序列和错配双链体。

可通过直接的 DNA 测序方法揭示参照基因和具有突变的基因间的序列差异。此外,克隆的 DNA 区段可以用作探针以检测特异性 DNA 区段。当和 PCR 结合时,这种方法的敏感性大大提高。例如,将测序引物和双链的 PCR 产物或由改良的 PCR 方法产生的单链模板分子一起使用。通过常规的放射标记核苷酸方法或具有荧光标记的自动测序方法确定序列。

基于 DNA 序列差异的遗传试验可以通过检测在有或没有变性剂时凝胶上 DNA 片段电泳迁移率的改变完成。小的序列缺失和插入可以由高分辨率凝胶电泳显示出。不同序列的 DNA 片段可以在变性甲酰胺梯度凝胶上的区别,其中不同的 DNA 片段的迁移按照其特定的熔点或部分解链温度而停滞在凝胶的不同位置(参见,例如,Myers 等,

科学, 230: 1242 (1985))。

也可以由核酸酶保护测定法揭示特定位置上的序列变化, 所述测定法如核糖核酸酶保护和 S1 保护以及化学裂解方法(例如, Cotton 等, PNAS, 美国, 85: 4397-4401(1985))。

这样, 可以由以下方法检测特定 DNA 序列, 所述方法例如杂交、核糖核酸酶保护、化学裂解、直接 DNA 测序、或使用限制酶(例如, 限制片段长度多形性(RFLP))和基因组 DNA 的 Southern 印迹法。

除很多常规凝胶电泳和 DNA 测序法之外, 也可用原位分析检测突变。

因为相对于正常组织样品所说的多肽的过量表达可以检测某些疾病(例如, 瘤形成、皮肤机能障碍、视觉障碍和炎症)的存在, 所以, 本发明也涉及用于检测各种组织中本发明的多肽之改变的水平诊断分析方法。用于检测得自宿主的样品中本发明多肽水平的分析方法对本领域的技术人员是公知的, 所说的方法包括: 放射免疫测定、竞争结合测定、Western 印迹分析, 优选地是 ELISA 测定。ELISA 测定最初包括制备本发明多肽之抗原的特异性抗体, 优选地是单克隆抗体。此外, 制备单克隆抗体的报道抗体。将可检测的试剂和报道抗体结合, 所说的试剂如放射性、荧光或在这一实例中是辣根过氧化物酶。由宿主获得样品, 并将其在与样品中蛋白质结合的固体支持物(如聚苯乙烯皿)中温育。通过和非特异性蛋白质(如牛血清清蛋白)一起温育, 包被皿中任何自由的蛋白质结合位点。接下来, 将单克隆抗体在聚苯乙烯皿中保温, 在此期间单克隆抗体与附着于皿上的任何本发明多肽结合。用缓冲液将所有未结合的单克隆抗体洗掉。此时, 将和辣根过氧化物酶连接的报道抗体放入皿中, 结果导致报道抗体和任何结

合到本发明多肽上的单克隆抗体结合。然后将未结合的报道抗体洗掉。接着向皿中加入过氧化物酶底物，和标准曲线比较，在给定时间内产生的颜色的量即是给定体积的患者样品中存在的蛋白质的量。

也可以使用竞争测定确定宿主样品中本发明多肽的水平。这种测定方法包括分离过量表达本发明多肽受体的原生质膜。然后将含有已经标记的本发明多肽的实验样品加入到原生质膜中，并将其温育一段时间。将可能含有本发明多肽的宿主样品加入到反应混合物中。然后使反应混合物通过被迅速洗涤的滤器，并测定结合的放射性以确定所说的受体的竞争量，从而确定样品中本发明多肽的量。

因为很多类型的癌细胞在瘤形成或增殖的过程中正调节(upregulate)TGF $\alpha$ 家族的各种成员，所以可以将 TGF $\alpha$ -HI 的特异性抗体用于癌症的诊断和治疗中。这些抗体和 TGF $\alpha$ -HI 结合并失活 TGF $\alpha$ -HI。TGF $\alpha$ -HI(和/或其家族成员)的单克隆抗体在临床上用于某些机能障碍的诊断和治疗，所说的机能障碍包括(但不限于)增生的和致瘤性的生长异常。致瘤性组织对生长因子的表达的上调形成了检测感染患者血液中生长因子增加的各种血清测定的基础。一般，这种测定不仅用于诊断性确诊(setting)，也用于预后性确诊(以便在外科手术和化疗之后检测潜隐的肿瘤细胞的存在)。

此外，在受体结合测定中可以利用标记的 TGF $\alpha$ -HI 或利用 TGF $\alpha$ -HI 受体本身的抗体检测表达 TGF $\alpha$ -HI 受体的恶性细胞。依据 TGF $\alpha$ -HI 受体的存在和密度区别细胞，进而提供预测所说的细胞对 TGF $\alpha$ -HI 的生物活性之敏感性的方法。

本发明的序列对染色体鉴定也是有价值的。该序列特异地靶向位于单个的人染色体上的特定位置并能与之杂交。此外，现在需要鉴定

染色体上的特定位点。目前，仅有少数几种以实际的序列数据（重复多态性）为基础的染色体标记试剂可以用于标记染色体的位置。本发明的 DNA 染色体作图是将这些序列和疾病相关基因相关联的重要的第一步。

简而言之，通过从 cDNA 制备 PCR 引物（优选 10—25bp）便可以把序列定位到染色体上。采用 3' 未翻译区域的计算机分析可以快速选择引物，其中引物不应跨越超过基因组 DNA 上的一个外显子，否则使得扩增方法复杂化。然后采用这些引物用于 PCR 筛选含有单个的人染色体的体细胞杂交体。只有那些含有与该引物对应的人基因的杂交体才会生产扩增片段。

体细胞杂交体的 PCR 作图是将一个特定的 DNA 定位于特定的染色体上的快速程序。根据本发明采用同样的寡核苷酸引物，用来自于特定染色体或者大基因组克隆集合体的一组片段、按照类似方式可以实现亚定位（sublocalization）。可以类似地用于对染色体作图的其它的作图策略包括原位杂交，用标记的经流式分选的染色体进行预筛选以及通过杂交进行预选，从而构建出染色体特异性的 cDNA 文库。

cDNA 克隆与一个中期染色体涂片的荧光原位杂交（FISH）可以用来实现一步法准确染色体定位。该技术可以采用 50 或 60 个碱基长短的 cDNA。关于该技术的综述参阅 Verma 等，人类染色体：基本技术手册，Pergamon 出版社，纽约（1988）。

一旦一个序列已定位到一个准确的染色体位置，则染色体上该序列的物理位置可与遗传图谱数据相关联。这些数据例如可在 V.McKusick, 人类的孟德尔遗传中找到（可以通过 Johns Hopkins 大学 Welch 医学文库联机得到）。然后通过连锁分析（物理相邻基因的共

遗传性) 来鉴定基因与已定位到相同染色体区域上的疾病之间的关系。

接下来需要测定在受影响的和未受影响的个体之间 cDNA 或基因组序列中的差异。如果突变是在一些或所有的受影响个体中观察到的、但是又没有在任何一个正常个体中被观察到的话, 那么该突变可能是疾病的病原体。

根据物理作图和遗传作图技术目前的分辨率, 一个被准确定位到与疾病有关的染色体区域的 cDNA 可以是 50—500 个潜在的病原 (causative) 基因中的一种 (其中假定有 1 兆碱基的作图分辨率且每 20kb 为一个基因)。

多肽、其片段或该多肽的其它衍生物或类似物、或者表达上述物质的细胞可以用作生产其抗体的免疫原。这些抗体可以是例如多克隆抗体或者单克隆抗体。本发明也包括嵌合, 单链和人源化的抗体, 以及 Fab 片段或 Fab 表达文库的产物。本领域已知的多种方法可以用于生产这些抗体和片段。

针对相应于本发明的序列的多肽而产生的抗体可以通过将该多肽直接注射入动物体内或者通过将该多肽向动物给药来得到, 其中的动物优选非人类。然后, 如此得到的抗体会结合到该多肽上。通过这种方式, 即使是仅仅编码该多肽的一个片段的序列也可用于产生能结合整个天然多肽的抗体。然后, 该抗体可以用于从表达该多肽的组织中分离这种多肽。

为了制备单克隆抗体, 可以采用任何一种通过连续的细胞系培养生产抗体的技术。例子包括杂交瘤技术 (Kohler 与 Milstein, 1975, 自然, 256:495-497), 三体杂交瘤技术, 人 B 细胞杂交瘤技术 (Kozbor 等,

1983, 今日免疫学, 4:72) 以及生产人单克隆抗体的 EBV-杂交瘤技术 (Cole, 等, 1985, 单克隆抗体与癌症治疗, Alan R. Liss, Inc., pp. 77-96)。

可以将用于生产单链抗体的技术 (美国专利 4,946,778) 进行修改从而生产出针对本发明免疫原性多肽制品的单链抗体。也可以使用转基因小鼠来表达针对本发明免疫原性多肽制品的人源化抗体。

本发明将参照下面的实施例进一步加以说明; 但是, 应当了解本发明并不局限于这些实施例。除非另作声明的以外, 所有的份或量均为重量。

为了利于理解以下的实施例, 现叙述一些经常出现的方法和/或术语。

“质粒”通过一个在前的小写 p 和/或跟随几个大写字母和/或数字加以命名。本文中的起始质粒或者可以通过商业途径得到或在不受限制的基础上公众可得到, 或者可以根据已公开的方法从可得到的质粒中构建出来。此外, 对于与那些所述等价的质粒是本领域已知的并且对本领域普通技术人员是显而易见的。

DNA 的“消化”是指用一种仅对 DNA 上的某些序列起作用的限制性酶催化裂解 DNA。本文所采用的多种限制性酶可以通过商业途径得到, 并且其反应条件、辅因子和其它使用要求对本领域普通技术人员是已知的。为了分析目的, 通常把 1  $\mu$ g 的质粒或 DNA 片段与溶于约 20  $\mu$ l 缓冲溶液的约 2 单位的酶一起使用。为了分离用于质粒构建的 DNA 片段, 通常在一个更大的体积内用 20 至 250 单位的酶消化 5 至 50  $\mu$ g 的 DNA。针对具体的限制性酶而言, 合适的缓冲溶液和底物的量是由生产者规定的。通常采用在 37°C 下约 1 小时的温育时间,



但是该时间可以根据产品供应者的指示而变化。在消化后，反应混合物直接在聚丙烯酰胺凝胶上进行电泳以分离出所需的片段。

采用由 Goeddel, D. 等, 核酸研究, 8:4057 (1980)所述的 8%聚丙烯酰胺凝胶进行裂解片段的大小分离。

“寡核苷酸”或指一种单链多脱氧核苷酸，或指可以通过化学合成的两条互补的多脱氧核苷酸链。这些合成的寡核苷酸不具有 5' 磷酸，因此如果不在一种激酶存在下以 ATP 添加一个磷酸时，该寡核苷酸将不会连接到另一个寡核苷酸上。合成的寡核苷酸将连接到未被去磷酸化的片段上。

“连接”是指在两个双链核酸片段之间形成磷酸二酯键的过程 (Maniatis, T., 等, 出处同上, p.146)。除非另行提供的以外，采用已知的缓冲液和条件、每 0.5  $\mu$ g 约等摩尔量的待连接 DNA 片段 10 单位 T4 DNA 连接酶 (“连接酶”) 来实现连接。

除非另有说明，按 Graham, F.和 Van der Eb, A., 病毒学, 52:456-457 (1973)所述的方法进行转化。

### 实施例 1

#### 细菌表达和纯化可溶性形式的 TGF $\alpha$ -HI

利用 PCR 寡核苷酸引物起始扩增编码 TGF $\alpha$ -HI 的 DNA 序列，所说的引物相当于加工过的 TGF $\alpha$ -HI 蛋白质的 5' 序列(减去信号肽序列)和 TGF $\alpha$ -HI 基因 3' 的载体序列。将相应于 TGF $\alpha$ -HI 的附加核苷酸分别加入到 5' 和 3' 序列中。所说的 5' 寡核苷酸引物具有序列 5' CCCGGATCCGCACGAGACATACCTTGTCCG 3' (SEQ ID NO: 3)，此引物含有 BamHI 限制性内切酶位点(粗体)，接着由加工过的蛋白质密码子的假定的末端氨基酸开始的 TGF $\alpha$ -HI 编码序列的 21 个核

苷酸。所说的 3' 序列 5' GGGAAAGCTTTTAATACTGAAATCGTAC AGGAC 3' (SEQ ID NO: 4)包含和 Hind III 位点互补的序列, 并接着 TGF $\alpha$ -HI 的 23 个核苷酸。所说的限制性内切酶位点相应于细菌表达载体 pQE-9 的限制性内切酶位点(Qiagen, Chatsworth 公司, CA, 91311)。pQE-9 编码抗生素抗性(Amp<sup>r</sup>)、细菌的复制起点(ori)、IPTG 可调节的启动子操纵子(P/O)、核糖体结合位点(RBS)、6-组氨酸标记和限制性内切酶位点。然后用 BamHI 和 HindIII 消化 pQE-9。将扩增的序列连接到 pQE-9 中并将其插入到带有编码组氨酸标记和 RBS 之序列的框架中。然后通过 Sambrook 等(分子克隆: 实验手册, 冷泉港实验室出版社, (1989))描述的方法, 用连接混合物转化大肠杆菌菌株 M15/rep 4(Qiagen, 公司)。M15/rep4 包含质粒 pREP4 的多拷贝, 这种质粒表达 lacI 阻遏物同时赋予卡那霉素抗性(Kan<sup>r</sup>)。通过转化体在 LB 平板上的生长能力鉴定转化体, 并筛选出氨苄青霉素/卡那霉素抗性菌落。通过限制分析分离并确认质粒 DNA。将含有所需构建体的菌落在补充了 Amp(100 $\mu$ g/ml)和 Kan(25 $\mu$ g/ml)的 LB 液体培养基中过夜(O/N)生长。将 O/N 培养物以 1: 100 至 1: 250 的比率用于接种大体积培养物。细胞生长至 0.4 和 0.6 之间的光密度 600(O.D.<sup>600</sup>)时。加入 IPTG(异丙基-B-D-thiogalacto 吡喃糖苷)至最终浓度为 1mM。IPTG 通过失活 lacI 阻遏物, 清除 P/O, 导致增加基因表达。将细胞再培养 3 至 4 小时。通过离心收获细胞。将细胞沉淀溶解在 6M 盐酸胍液剂中。澄清后, 在允许含有 6-组氨酸标记的蛋白质紧密结合的条件下, 在镍-螯合物柱中通过层析从溶液中纯化溶解的 TGF $\alpha$ -HI(Hochuli, E. 等, 色谱法杂志 411: 177-184(1984))。将 TGF $\alpha$ -HI(85%纯度)以 6M 盐酸胍 (pH5.0)从柱中洗脱下来, 为了复性, 调至 3M 盐酸胍、100mM 磷酸

盐钠、10mM 谷胱甘肽(还原型)、2mM 谷胱甘肽(氧化型)。在溶液中温育 12 小时后, 将蛋白质对 10mM 磷酸钠透析。

## 实施例 2

利用杆状病毒表达系统克隆和表达可溶性 TGF $\alpha$ -HI

利用 PCR 寡核苷酸引物扩增编码 TGF $\alpha$ -HI 蛋白质的 DNA 序列, 所述引物相应于表达图 1 的第一个氨基酸至活性区的最后一个氨基酸的基因的 5' 和 3' 序列:

使用了三套引物:

第一套引物是:

5' CGCGGATCCGCCATCATGGGCGCCGCAGCCGC 3' (SEQ ID NO: 5)和

5' GCGTCTAGACTAGTATAGAACACTGTAGTCC 3' (SEQ ID NO: 6);

第二套引物是:

5' CGCGGATCCAGTTTATATTGGAAACCACATGCC 3' (SEQ ID NO: 7)和

5' GCGTCTAGACTAATAGAGAATACTAAAGTC 3' (SEQ ID NO: 8);

这些引物用于表达推定的活性(可溶性)区。

所有 5' 引物都具有 BamHI 限制性内切酶位点(粗体), 接着类似真核细胞有效翻译起始信号的核苷酸(Kozak, M., 分子生物学杂志, 196: 947-950(1987))(翻译起始密码子是 “ATG ” )。

所说的 3' 引物序列包含限制性核酸内切酶 XbaI 的切割位点, 并具有和 TGF $\alpha$ -HI 基因的 3' TGF $\alpha$ 可溶性区互补的核苷酸。用市售试剂盒(“GeneClean,” 生物 101 公司, La Jolla, Ca.)从 1%琼脂糖凝胶分离扩增的序列。将所说的片段用核酸内切酶 BamHI 和 XbaI 消化, 并在 1%琼脂糖凝胶上再次纯化。此片段命名为 F2。

使用载体 pA2(由 pVL941 载体修饰而成, 见下文讨论)由杆状病

毒表达系统表达 TGF $\alpha$ -HI 蛋白质(综述参见: Summer, M. D. 和 Smith, G. E. 1987, 杆状病毒载体和昆虫细胞培养过程的方法手册, 得克萨斯农业的试验站公报 1555)。这种表达载体包含苜蓿银纹夜蛾多核型多角病毒(AcMNPV)强多角体蛋白启动子, 其后面是限制性核酸内切酶的识别位点。猴病毒(SV)40 的聚腺苷酸化位点用于有效的聚腺苷酸化。为了容易地选择重组病毒, 把大肠杆菌的  $\beta$ -半乳糖苷酶基因以与多角体蛋白启动子同样的方向插入, 其后是多角体蛋白基因的聚腺苷酸化信号。多角体蛋白序列的两侧是用于细胞介导的共转染的野生型病毒 DNA 的同源重组的病毒序列。可以用很多其它的杆状病毒载体代替 pRG1, 所说的载体如 pAc373、pVL941 和 pAcIM1(Luckow, V. A. 和 Summer, M. D., 病毒学, 170: 31-39)。

用限制酶 BamHI 和 XbaI 消化所说的质粒, 然后由本领域已知的方法利用牛肠磷酸酶使质粒去磷酸化。然后用市售试剂盒 (“GeneClean”, BIO 101 公司, La Jolla, Ca.)从 1%琼脂糖凝胶分离 DNA。这种载体 DNA 命名为 V2。

用 T4 DNA 连接酶连接 F2 片段和所说的去磷酸化质粒 V2。然后转化大肠杆菌 HB101 细胞(Stratagene 克隆系统, 11011 North Torrey Pines Road La Jolla, Ca. 92037)并利用酶 BamHI 和 XbaI 鉴别含有带 TGF $\alpha$ -HI 基因的所说质粒(pBac TGF $\alpha$ -HI)的细菌。通过 DNA 测序确认所克隆的片段的序列。

用脂转染法(Felgner 等, 美国科学院学报, 84 :7413-7417(1987))将 5 $\mu$ g 质粒 pBacTGF $\alpha$ -HI 和 1.0 $\mu$ g 市售的线性杆状病毒 (“BaculoGold™杆状病毒 DNA”, Pharmingen, San Diego, CA.)共转染。

将 1 $\mu$ g BaculoGold™病毒 DNA 和 5 $\mu$ g 质粒 pBacTGF $\alpha$ -HI 在含有 50 微升无血清 Grace' s 培养基 (Life Technologies 公司, Gaithersburg, MD) 的微滴板的无菌孔中混合。在添加 10 微升脂转染试剂和 90 微升 Grace' s 培养基后, 将混合并于室温下培养 15 分钟。然后将转染混合物逐滴添加到 Sf9 昆虫细胞(ATCC CRL1711)中, 所说细胞接种在具有 1ml 无血清 Grace' s 培养基的 35mm 组织培养平板上。来回摇动平板以混合新添加的溶液。然后将平板在 27 $^{\circ}$ C 下培养 5 小时。5 小时后, 从平板除去转染溶液, 添加 1 微升补充有 10%胎牛血清的 Grace' s 昆虫培养基。把平板放回到温箱, 在 27 $^{\circ}$ C 下连续培养 4 天。

4 天后, 收集上清液, 用类似 Summers 和 Smith(同上)所述的方法进行噬斑测定。一点改变是使用具有“Blue Gal”的琼脂糖凝胶(Life Technologies 公司, Gaithersburg), 其使得易于分离染蓝的噬斑。(“噬斑测定”的详尽的描述也可以在 Life Technologies 公司(Gaithersburg)发布的昆虫细胞培养和杆状病毒学用户指南 9-10 页中找到)。

四天后, 将系列稀释的病毒添加到细胞中, 用 Eppendorf 吸管尖端挑取染蓝的噬斑。然后将包含重组病毒的琼脂悬浮于含 200 微升 Grace' s 培养基的 Eppendorf 管中。经简短离心除去琼脂, 将含重组杆状病毒的上清液用于感染接种到 35 毫米培养皿中的 Sf9 细胞。4 天后, 收获这些培养皿中的上清液, 于 4 $^{\circ}$ C 贮存。

使 Sf9 细胞在补充有 10%热灭活 FBS 的 Grace' 培养基中生长。在感染复数(MOI)2 下用重组杆状病毒 V- TGF $\alpha$ -HI 感染细胞。6 小时后, 除去培养基, 用 SF900II 培养基(减甲硫氨酸和半胱氨酸)(Life Technologies 公司, Gaithersburg)替代。42 小时后, 添加 5 $\mu$ Ci  $^{35}$ S 甲

硫氨酸和  $5\mu\text{Ci } ^{35}\text{S}$  半胱氨酸(Amersham)。在经离心收获之前,将细胞进一步培养 16 小时,标记蛋白质经 SDS-PAGE 和放射自显影显示出。

### 实施例 3

#### 在 COS 细胞中表达重组 TGF $\alpha$ -HI

质粒 TGF $\alpha$ -HIHA 的表达来源于载体 pcDNA3/Amp(Invitrogen),其包含: 1)SV 40 复制起点, 2)氨苄青霉素抗性基因, 3)大肠杆菌的复制起点, 4)CMV 启动子, 其后为多接头区、SV 40 内含子和聚腺苷酸化位点。将编码完整的 TGF $\alpha$ -HI 前体的 DNA 片段和在读框中融合进 3' 末端的 HA 标记克隆到所说的载体的多接头区, 因此, 重组蛋白质的表达在 CMV 启动子控制之下。HA 标记相应于来自流感血细胞凝集素蛋白质的表位, 这一点以前已经描述过(I. Wilson, H. Niman, R. Heighten, A. Cherenson, M. Connolly, and R. Lerner, 1984, 细胞 37:767,(1984))。HA 和靶细胞蛋白的融合, 使得用识别 HA 表位的抗体很容易测定重组蛋白质。

质粒构建策略描述如下:

用两种引物在克隆的原始的 EST 上通过 PCR 方法构建编码 TGF $\alpha$ -HI 的 DNA 序列, 所说的引物是: 5' 引物 5' **CGCGGATCCGCCATCATGGTGCTGTGGGAGTCC** 3' (SEQ ID NO:12), 包含 BamHI 位点(粗体), 之后是由起始密码子开始的 TGF $\alpha$ -HI 编码序列的 18 个核苷酸; 3' 引物 5' GCGCTCGAGGTA TAGAACACTGTAGTCC 3' (SEQ ID NO:13), 包含互补于 XhoI 位点、TGF $\alpha$ 区的最后 19 个核苷酸和 XhoI 位点的序列。pcDNA3/Amp 载体包含 BamHI/XhoI 克隆位点, 此位点可将 PCR 插入片段带入后接终止密码子的 3' HA 标记读框中。因此, PCR 产物包含 BamHI 位

点, 936 个碱基对编码序列和 XhoI 位点。用 BamHI 和 XhoI 限制性内切酶消化 PCR 扩增的 DNA 片段和载体 pcDNA3/Amp 并连接。将连接混合物转化到大肠杆菌菌株 SURE 中(可从 Stratagene 克隆系统, La Jolla, CA 92037 得到), 将转化的培养物接种在氨苄青霉素培养基平板上, 并筛选抗性克隆。从转化体中分离质粒 DNA, 并用限制分析检测是否存在正确的片段。为了表达重组 TGF $\alpha$ -HI, 通过 DEAE-DEXTRAN 方法用表达载体转染 COS 细胞(J. Sambrook, E. Fritsch, T. Maniatis, 分子克隆: 实验手册, 冷泉港实验室出版,(1989))。通过放射标记和免疫沉淀法检测 TGF $\alpha$ -HIHA 蛋白质的表达(E. Harlow, D. Lane, 抗体: 实验手册, 冷泉港实验室出版,(1988))。将细胞在转染后的两天用 <sup>15</sup>S-半胱氨酸标记 8 小时。然后收集培养物培养基并用去垢剂(RIPA 缓冲液(150 mM NaCl, 1% NP-40, 0.1% SDS, 1% NP-40, 0.5% DOC, 50mM Tris, pH 7.5)(Wilson, I. 等, Id. 37:767(1984)) 裂解细胞。用 HA 特异性单克隆抗体沉淀细胞裂解物和培养物培养基。在 15% SDS-PAGE 凝胶上分析蛋白质沉淀。

#### 实施例 4

##### 经由基因治疗的表达

成纤维细胞是通过皮肤活组织检查从一个研究对象中得到的。将得到的组织放置在组织培养基上并且分割成小块。将小的组织块放置在组织培养瓶的湿表面上, 其中每个瓶中放置约 10 块组织。将瓶颠倒放置, 盖紧并于室温下保留过夜。在室温下过 24 小时后反转瓶, 组织块固定在瓶底部, 加入新鲜培养基(例如含有 10%FBS, 青霉素和链霉素的 Ham' s F12 培养基), 然后于 37°C 下温育大约 1 周。这时加入新鲜培养基, 随后每隔几天更换一次培养基。再培养 2 周后,

出现了一单层成纤维细胞。该单层细胞经胰蛋白酶消化并放入更大的瓶中。

用 EcoRI 和 Hind III 消化旁侧有 Moloney 鼠肉瘤病毒长末端重复的 pMV-7 (Kirschmeier, P.T.等,DNA, 7:219-25 (1988)), 接下来用牛肠磷酸酶进行处理。线性载体在琼脂糖凝胶上分级分离并使用玻璃珠加以纯化。

采用分别与 5' 和 3' 末端序列对应的 PCR 引物扩增编码本发明多肽的 cDNA。5' 引物包含一个 EcoRI 位点, 3' 引物含有一个 Hind III 位点。在 T4 DNA 连接酶存在下, 将等量的 Moloney 鼠肉瘤病毒的线性化骨架与扩增的 EcoRI 和 Hind III 片段加在一起, 在适于两个片段连接的条件下保持所得到的混合物。将该连接混合物用于转化细菌 HB101, 然后为了证实该载体具有插入正确的感兴趣的基因而将细菌涂布在含有卡那霉素的琼脂上。

将兼嗜性 (amphotropic) pA317 或 GP+am12 包装细胞在含有 10% 小牛血清 (CS)、青霉素和链霉素的 Dulbecco' s 改良 Eagles 培养基 (DMEM) 的组织培养板上进行培育, 直至达到汇合密度。然后将含有基因的 MSV 载体加入培养基中并用该载体转导包装细胞。包装细胞随即生产含有该基因的具有感染性的病毒颗粒 (现在包装细胞被称作生产细胞)。

向转导的生产细胞中加入新鲜的培养基, 接下来从一个铺满生产细胞的 10cm 平板中收集培养基。含有感染性病毒颗粒的培养基经微孔过滤器过滤以除去脱附 (detached) 的生产细胞, 然后利用该培养基去感染成纤维细胞。从成纤维细胞的亚铺满平板中除去培养基并迅速地代之以来自于生产细胞的培养基。除去该培养基并代之以新鲜的



培养基。如果该病毒的滴度高的话，那么实质上所有的成纤维细胞均被感染并且无需选择。如果滴度非常低，那么就需要采用一个具有如 neo 或 his 这样的可选择性标记物的逆转录病毒。

然后将工程化的成纤维细胞注射入宿主，它或单独注射或在一个 cytodex 3 微载体珠上已培育至铺满之后再注射。

根据以上的教导，本发明的许多改进和变化都是可能的，因此在附加的权利要求的范围内可以另外的方式实施本发明。

## 序 列 表

## (1)一般信息:

(i)申请人: MEISSNER, ET AL.

(ii)发明名称: 转化生长因子 $\alpha$ HI

(iii)序列数: 8个

(iv)通讯地址:

(A)收信人: CARELLA, BYRNE, BAIN, GILFILLAN,  
CECCHI, STEWART & OLSTEIN

(B)街道: 6 BECKER FARM ROAD

(C)城市: ROSELAND

(D)州: 新泽西州

(E)国家: 美国

(F)邮码: 07068

(v)计算机可读形式:

(A)介质类型: 3.5 英寸磁盘

(B)计算机: IBM PS/2

(c)操作系统: MS-DOS

(D)软件: WORD PERFECT 5.1

(Vi)当前申请的数据:

(A)申请号: 08/468,846

(B)申请日: 1995.6.6

(C)分类号:

(Vii)在先申请的数据

(A)申请号:

(B)申请日:

(viii)律师/代理人信息:

(A)姓名: FERRARO, GREGORY D.

(B)登记号: 36,134

(c)卷号/文档号: 325800-465

(ix)电信信息:

(A)电话: 201-994-1700

(B)传真: 201-994-1744

## (2)SEQ ID NO: 1 的信息:

## (i)序列特征:

(A)长度: 1565 个碱基对

(B)类型: 核酸

(C)链型: 单链

(D)拓扑结构: 线性

(ii)分子类型: cDNA

(xi)序列描述: SEQ ID NO: 1

```

TGGGCGCGCG GCTGGATGCC CCCGGCCTGC GGCTCCCTGC GCTTCCCGCC GTGGAGGGGC 60
ACCACTCATG GCGCGCCGAG CCGCTGAGGC GCGGCTCCGG CTGCCGCGG CGCCTCCGCT 120
CGCCTTCTGC TGCTACACGT CCGTGCTTCT GCTCTTCGGC TTCTCTCTGC CCGGGAGCCC 180
CGCGTCCAAC CAGCCCCCGG GTGGTGGCGG CCGCAGCGGC GGGGACTGTC CCGGCGGCAA 240
AGGCAAGAGC ATCAACTGCT CAGAATTAAA TGTGAGGGAG TCTGACGTAA GAGTTTGTGA 300
TGAGTCATCA TGTAATAATG GAGGAGTCTG TAAAGAAGAT GGAGATGGTT TGAATGTGC 360
ATGCCAATTT CAGTGCCATA CAAATTATAT TCTGTCTGT GGATCAAATG GGGACACTTA 420
TCAAAATGAA TGCTTTCTCA GAAGGGCTGC TTGTAAGCAC CAGAAAGAGA TAACAGTAAT 480
AGCAAGAGGA CCATGCTACT CTGATAATGG ATCTGGATCT GGAGAAGGAG AAGAGGAAGG 540
GTCAGGGGCA GAAGTTCACA GAAAACACTC CAAGTGTGGA CCTGCAAAT ATAAAGCTGA 600
GTGTGATGAA GATGCAGAAA ATGTTGGGTG TGTATGTAAT ATAGATTGCA GTGGATACAG 660
TTTTAATCCT GTGTGTGCTT CTGATGGGAG TTCCATAAC AATCCCTGTT TTGTTGAGA 720
AGCATCTTGT ATAAAGCAAG AACAAATTGA TATAAGGCAT CTTGGTCATT GCACAGATAC 780
AGATGACACT AGTTTGTGGG GAAAGAAAGA TGATGGACTA CAATATCGAC CAGATGTGAA 840
AGATGCTAGT GATCAAAGAG AAGATGTTTA TATTGGAAAC CACATGCCTT GCCCTGAAA 900
CCTCAATGGT TACTGCATCC ATGGAAAATG TGAATTCATA TATTCTACTC AGAAGGCTTC 960
TTGTAGATGT GAATCTGGCT ACACTGGACA GCACTGTGAA AAGACAGACT TTAGTATTCT 1020
CTATGTAGTG CCAAGTAGGC AAAAGCTCAC TCATGTTCTT ATTGCAGCAA TTATTGGAGC 1080
TGTACAGATT GCCATCATAG TAGCAATTGT AATGTGCATA ACAAGAAAAT GCCCCAAAA 1140
CAATAGAGGA CGTCGACAGA AGCAAAACCT AGGTCATTTT ACTTCAGATA CGTCATCCAG 1200
AATGGTTTAA ACTGATGACT TTTATATGTA CACTGACCAT GTGATGTACA TTTATTATGT 1260
CTTTTTTTAA AGAATGGAAA TATTTATTTT AGAGGGCTTA TTTTGGACA TTTTATGT 1320
AGTACTGTTG GCTCGTATTT AGAATATTCA GCTACGACAG TTTTGGACTG TTTAGTAGTC 1380
TTTTTTTTAT GTTTTTAAAT ACAGAAATTG CTTTCACAAA TTTGTACCAC ATGGTAAATC 1440
TAAGACTTGT TCTTTACCCA TGGAAATGTA TATTTTTGCA AAGATGGACT ACTTCACAAA 1500
TGGTATAAAA GTCATATCCA CTTCTCCAC AATGACCACA GCAAATGACC AAGCATGAAC 1560
TAAAG 1565

```

## (2)SEQ ID NO: 2 的信息:

## (i)序列特征:

(A)长度: 380 个氨基酸

(B)类型: 氨基酸

(C)链型:

(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 Gly Ala Ala Ala Ala Glu Ala Pro Leu Arg Leu Pro Ala Ala
          -35                      -30                      -25
Pro Pro Leu Ala Phe Cys Cys Tyr Thr Ser Val Leu Leu Leu Phe
          -20                      -15                      -10
Ala Phe Ser Leu Pro Gly Ser Arg Ala Ser Asn Gln Pro Pro Gly
          -5                        1                        5
Gly Gly Gly Gly Ser Gly Gly Asp Cys Pro Gly Gly Lys Gly Lys
          10                        15                        20
Ser Ile Asn Cys Ser Glu Leu Asn Val Arg Glu Ser Asp Val Arg
          25                        30                        35
Val Cys Asp Glu Ser Ser Cys Lys Tyr Gly Gly Val Cys Lys Glu
          40                        45                        50
Asp Gly Asp Gly Leu Lys Cys Ala Cys Gln Phe Gln Cys His Thr
          55                        60                        65
Asn Tyr Ile Pro Val Cys Gly Ser Asn Gly Asp Thr Tyr Gln Asn
          70                        75                        80
Glu Cys Phe Leu Arg Arg Ala Ala Cys Lys His Gln Lys Glu Ile
          85                        90                        95
Thr Val Ile Ala Arg Gly Pro Cys Tyr Ser Asp Asn Gly Ser Gly
          100                       105                       110
Ser Gly Glu Gly Glu Glu Glu Gly Ser Gly Ala Glu Val His Arg
          115                       120                       125
Lys His Ser Lys Cys Gly Pro Cys Lys Tyr Lys Ala Glu Cys Asp
          130                       135                       140
Glu Asp Ala Glu Asn Val Gly Cys Val Cys Asn Ile Asp Cys Ser
          145                       150                       155
Gly Tyr Ser Phe Asn Pro Val Cys Ala Ser Asp Gly Ser Ser Tyr
          160                       165                       170
Asn Asn Pro Cys Phe Val Arg Glu Ala Ser Cys Ile Lys Gln Glu
          175                       180                       185
Gln Ile Asp Ile Arg His Leu Gly His Cys Thr Asp Thr Asp Asp
          190                       195                       200
Thr Ser Leu Leu Gly Lys Lys Asp Asp Gly Leu Gln Tyr Arg Pro
          205                       210                       215

```

Asp Val Lys Asp Ala Ser Asp Gln Arg Glu Asp Val Tyr Ile Gly		
	220	230
Asn His Met Pro Cys Pro Glu Asn Leu Asn Gly Tyr Cys Ile His		
	235	245
Gly Lys Cys Glu Phe Ile Tyr Ser Thr Gln Lys Ala Ser Cys Arg		
	250	260
Cys Glu Ser Gly Tyr Thr Gly Gln His Cys Glu Lys Thr Asp Phe		
	265	275
Ser Ile Leu Tyr Val Val Pro Ser Arg Gln Lys Leu Thr His Val		
	280	290
Leu Ile Ala Ala Ile Ile Gly Ala Val Gln Ile Ala Ile Ile Val		
	295	305
Ala Ile Val Met Cys Ile Thr Arg Lys Cys Pro Lys Asn Asn Arg		
	310	320
Gly Arg Arg Gln Lys Gln Asn Leu Gly His Phe Thr Ser Asp Thr		
	325	335
Ser Ser Arg Met Val		
	340	

## (2)SEQ ID NO: 3 的信息:

## (i)序列特征:

(A)长度: 30 个碱基对

(B)类型: 核酸

(C)链型: 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: 3

CCCGGATCCG CACGAGACAT ACCTTGTCGG

30

## (2)SEQ ID NO: 4 的信息:

## (i)序列特征:

(A)长度: 32 个碱基对

(B)类型: 核酸

(C)链型: 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  
 GGGAAGCTTT TAATACTGAA ATCGTACAGG AC 32

(2)SEQ ID NO: 5 的信息:

(i)序列特征:

(A)长度: 32 个碱基对

(B)类型: 核酸

(C)链型: 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: 5

CGCGGATCCG CCATCATGGG CGCCGCAGCC GC 32

(2)SEQ ID NO: 6 的信息:

(i)序列特征:

(A)长度: 31 个碱基对

(B)类型: 核酸

(C)链型: 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: 6

GCGTCTAGAC TAGTATAGAA CACTGTAGTC C 31

(2)SEQ ID NO: 7 的信息:

(i)序列特征:

(A)长度: 33 个碱基对

(B)类型: 核酸

(C)链型: 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: 7

---

CGCGGATCCA GTTTATATTG GAAACCACAT GCC

33

(2)SEQ ID NO: 8 的信息:

(i)序列特征:

(A)长度: 30 个碱基对

(B)类型: 核酸

(C)链型: 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: 8

GCGTCTAGAC TAATAGAGAA TACTAAAGTC

30

```

-60          -40          -20
TGGGGCGCGGCTGGATGCCCCCGGCTGGGGCTCCCTGGCGTTCCCCGCCGTCACAGGGGC
-----+-----+-----+-----+-----+-----+-----+-----+-----+
ACCCGGCGCGCACCTACGGGGCGCGACCGCCGAGGGACCGCGAAGGGCGGCAGGTCCCCCG
0          20          40
ACCAGTCATGGGGCGCGCAGCCGCTGAGGGCGCGCTCCCGGCTGCCGTGCCGGCCCTCCGCT
-----+-----+-----+-----+-----+-----+-----+-----+-----+
TGGTCAGTACCCGGCGGTGGCGGACTCCGGCGGAGGCCCGACGGACGGCGCGGAGGGCGA
M G A A A A E A P L R L P A A P P L
60          80          100
CGCCTTCGTGCTACACGTCGGTGTCTTCTGCTCTTCCGCTTCTCTCTGCCCGGGAGCCCG
-----+-----+-----+-----+-----+-----+-----+-----+-----+
GCGAAGACGACGATGTGACGCCACGAAGACGAGAGCGGAAGCGGAGAGAGAGAGAGAGAGAGAG
A F C C Y T S V L L L F A F S L P G S R
120          140          160
CGCGTCCAACGACCCCGGTGGTGGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCAA
-----+-----+-----+-----+-----+-----+-----+-----+-----+
GCGCAGGTTGGTGGGGGGCCACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGTT
A S N Q P P G G G G S G G S G G D C P G G K
180          200          220

```

下接图1B





上接图1A

```

AGGCAAGAGCATCAACTGCTCAGAAATTAATGTGAGGGAGTCTGACGTAAGAGTTTGTA
-----+-----+-----+-----+-----+-----+-----+-----+-----+
TCCGTTCTCGTAGTTGACGAGTCTTAATTTACACTCCCTCAGACTGCATTCATAAACACT
G K S I N C S E L N V R E S D V R V C D      260
240

TGAGTCATCATGTAATAATGGAGGAGTCTGTAAGAAGATGGAGATGGTTTGAAATGTGC
-----+-----+-----+-----+-----+-----+-----+-----+-----+
ACTCAGTAGTACATTTATACCTCCTCAGACATTTCTTCTACCTCTACCAAACTTTACACG
E S C C K Y G G V C K E D G D G L K C A      320
300

ATGCCAATTCAGTGCCATACAAATTAATATTCCTGTCTGTGGATCAAAATGGGACACTTA
-----+-----+-----+-----+-----+-----+-----+-----+-----+
TACGGTTAAAGTCACGGTATGTTTAATAATAAGACAGACACCTAGTTTACCCCTGTGAAT
C Q F Q C H T N Y I P V C G S N G D T Y      380
360

TCAAAATGAATGCTTTCTCAGAAGGGCTGTTGTAAGCACCCAGAAAGAGATAACAGTAAT
-----+-----+-----+-----+-----+-----+-----+-----+-----+
AGTTTACTTACGAAAGAGTCTTCCCGACGAAACATTCGGTGGTCTTTCTCTATTGTCATTA
Q N E C F L R R A A C K H Q K E I T V I      420
400

AGCAAGAGGACCATGCTACTCTGATAAATGGATCTGGATCTGGAGAAGGAGAAGGAAGG

```

下接图1C







上接图1D

```

CTATGTAGTGGCCAAAGTAGGCAAAAAGCTCAGCTCATGTTCTTATTGCAGCAATATTGGAGC
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
GATACATCACGGTTTCATCCGTTTTTCGAGTGAGTACAAGAATAACGTCGTTAATAACCCTCG
Y V V P S R Q K L T H V L I A A I I G A
1020 1040 1060

TGTACAGATTGCCCATCATAGTAGCAATTGTAATGTGCGATAACAAGAAAAATGCCCCAAAAA
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
ACATGTCTAACGGTAGTATCATCGTTAACATTACACCGTATTGTTCTTTACGGGGTTTTT
V Q I A I I V A I V M C I T R K C P K N
1080 1100 1120

CAATAGAGGACGTCGACAGAAAGCAAAACCCTAGGTCATTTTACTTCAGATACGTCATCCAG
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
GTTATCTCCCTGCAGCTGCTTCGTTTGGATCCAGTAAATGAAGTCTATGCCAGTAGGTC
N R G R R Q K Q N L G H F T S D T S S R
1140 1160 1180

AATGGTTTAAAC TGATGACTTTTATATGTACACTGACCATGTGATGTACATTTATTATGT
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
TTACCAAATTTGACTACTGAAAAATATACATGTGACTGGTACACTACATGTAATAATACA
M V *
1200 1220 1240

CTTTTTTTAAAGAATGGAAATATTATTTCAGAGGCCCTTATTTTGGACATTTTATTAGTGT

```

下接图1E

图 1E

上接图1E

```

-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
GAAAAAATTTCTTACCTTTATAAATAAAGTCTCCGGAAATAAAAACCTGTAAAAATCACA
1260 1280 1300
AGTACTGTTGGCTGATTTAGAAATATTCAGCTACGACAGTTTTGGACTGTTTAGTAGTC
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
TCATGACAACCGAGCATAAAATCTTATAAAGTCGATGCTGTCAAAAACCTGACAAAATCATCAG
1320 1340 1360
TTTGTGTTTATGTTTTTAAATACAGAAATTTGCTTTCACAAAATTTGTACCACATGGTAATTC
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
AAACAAAATACAAAAATTTATGTCCTTAAACGAAAGTGTTTAAACAATGGTGTACCATTAAAG
1380 1400 1420
TAAGACTTGTCTTACCCATGGAATGTAATAATTTTGGCAAAGATGGACTACTTCACAAA
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
ATTCTGAACAAGAAATGGGTACCCTTACATTAATAAAAACGTTTCTACCTGATGAAGTGT
1440 1460 1480
TGGTTATAAAGTCATATCCACITCTTCCACAATGACCACAGCAAATGACCCAAGCATGAAC
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
ACCAATATTTTCAGTATAGGTGAAGAAGGTGTTACTGGTGTCTGTTTACTGGTTCGTTACTTG

```

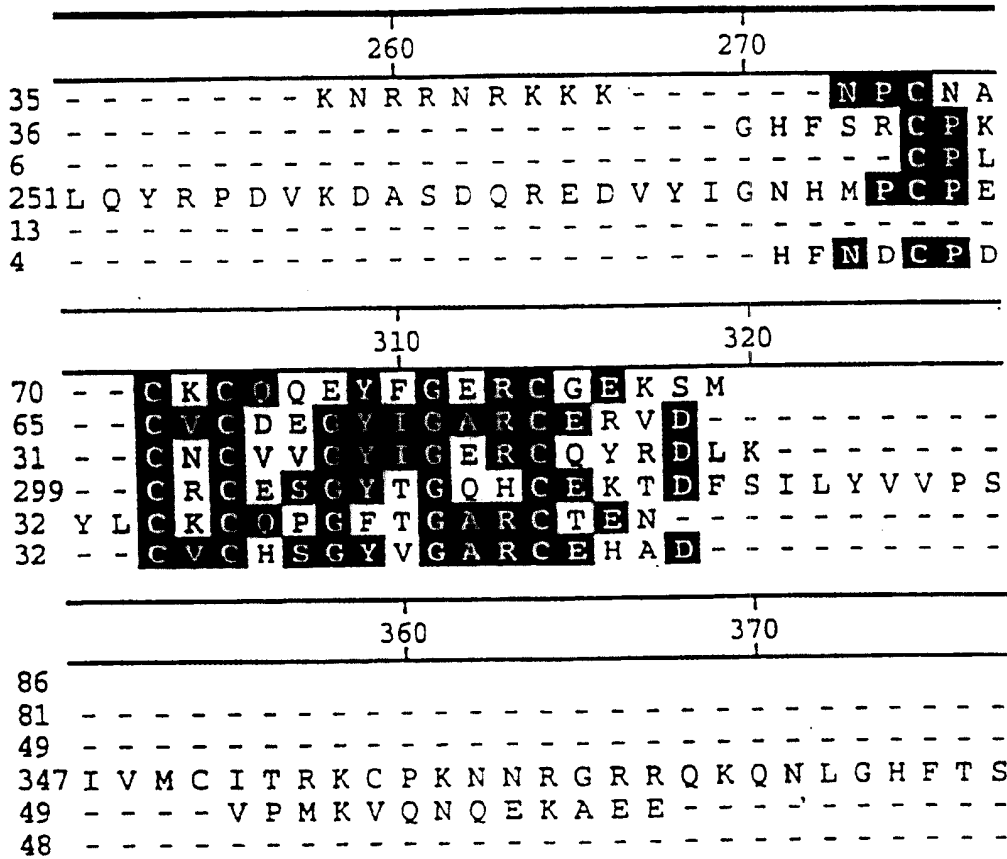
```

TAAAG
-----
ATTTC

```

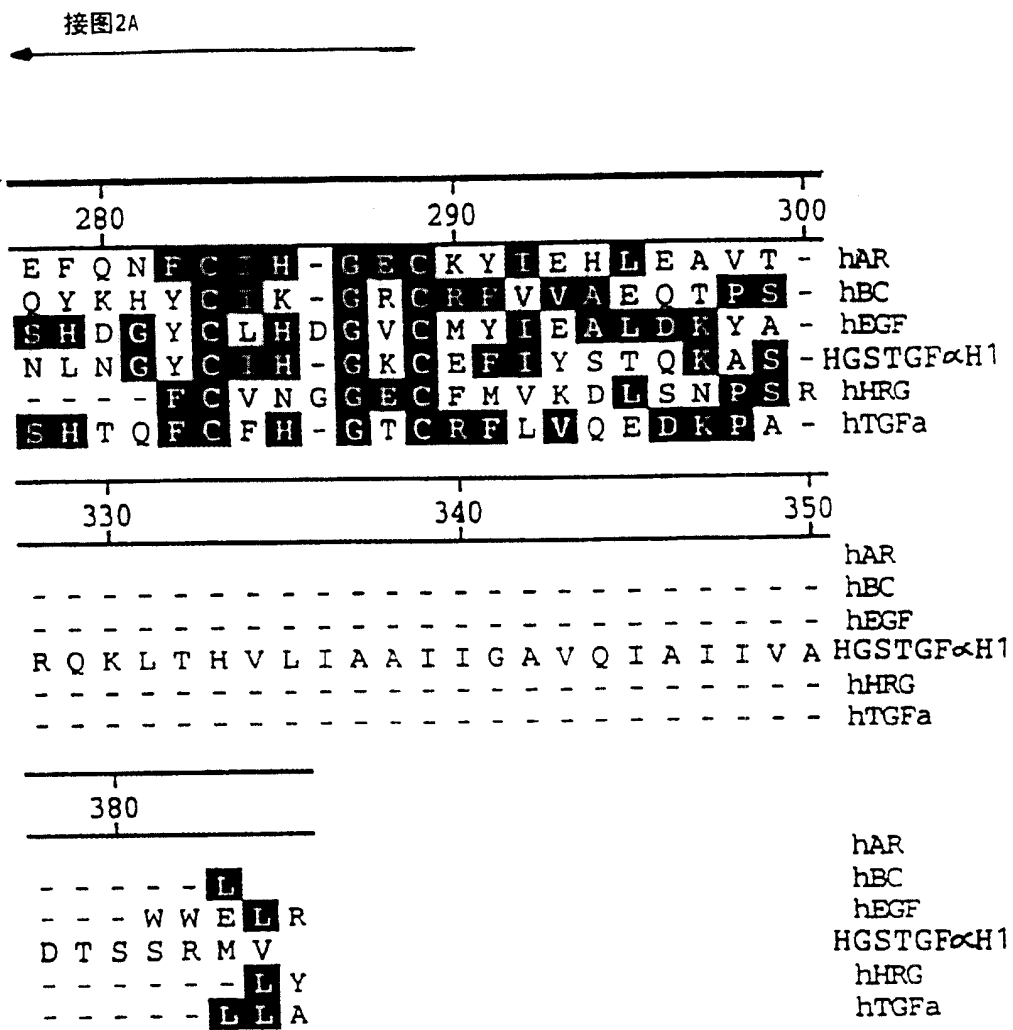


接图2B



- hAR = 人双调蛋白
- hBC = 人 $\beta$ -动物纤维素
- hEGF = 人表皮生长因子
- hURG = 人 heregulin
- htGF = 人转化生长因子

图2A



修饰‘修饰#1’：与共有序列准确配对的残基以阴影装饰（具有实心黑框）。

图2B