

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.



# [12] 发明专利说明书

专利号 ZL 00819051.8

*C07K 14/47 (2006.01)*  
*A61K 47/48 (2006.01)*  
*C12N 15/12 (2006.01)*  
*C12N 5/10 (2006.01)*  
*C07K 16/18 (2006.01)*  
*G01N 33/53 (2006.01)*

[45] 授权公告日 2007 年 1 月 31 日

[11] 授权公告号 CN 1297566C

[51] Int. Cl. (续)

*C12Q 1/68 (2006.01)*

*A61K 38/17 (2006.01)*

[22] 申请日 2000.12.15 [21] 申请号 00819051.8

[30] 优先权

[32] 1999.12.16 [33] US [31] 60/172,370

[32] 2000.12.7 [33] US [31] 09/732,357

[86] 国际申请 PCT/US2000/033901 2000.12.15

[87] 国际公布 WO2001/044291 英 2001.6.21

[85] 进入国家阶段日期 2002.8.16

[73] 专利权人 舍林股份公司

地址 德国柏林

[72] 发明人 理查德·哈金斯 底波拉·帕克斯  
戈登·帕里 道格拉斯·W·施奈德  
雷娜特·施泰因布雷歇尔

[56] 参考文献

WO9850073A1 1998.11.12 C07K14/435

WO9845442A2 1998.10.15 C07K14/47

US5781969A 1999.2.16 C12N15/12

审查员 韩世炜

[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 4 页 说明书 69 页 附图 10 页

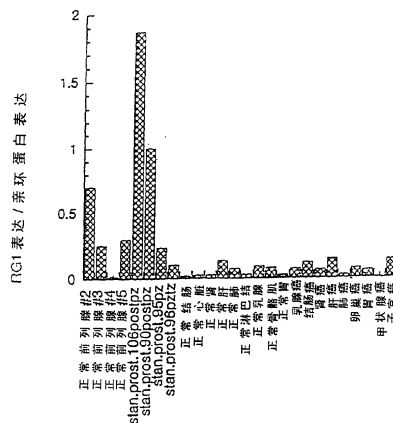
[54] 发明名称

编码新 RG1 多肽的 DNA

[57] 摘要

本发明涉及新的人胞外基质多肽，称为 RG1，编码所述多肽的多核苷酸，生产所述多肽的方法，用于表达所述多肽的表达载体和基因工程化宿主细胞。本发明还涉及使用所述多核苷酸和多肽用于研究、诊断和治疗应用的方法。

mRNA在人组织中的表达



1.一种分离的抗体或抗体片段,其特异性结合选自以下一组的多肽:

(a) 一种包含图 2(SEQ ID NO:2)所示序列 28—46 位氨基酸的多肽;

(b) 一种包含图 2(SEQ ID NO:2)所示序列 77—91 位氨基酸的多肽;

(c) 一种包含图 2(SEQ ID NO:2)所示序列 188—210 位氨基酸的多肽;

(d) 一种包含图 2(SEQ ID NO:2)所示序列 263—274 位氨基酸的多肽;

(e) 与 (a), (b), (c) 或 (d)的多肽具有至少 70%相同性的多肽。

2.权利要求 1 的抗体,其中抗体特异性结合氨基酸序列 PLGGESICSAGAPAKYSIT (SEQ ID NO:8)。

3.权利要求 1 的抗体,其中抗体特异性结合氨基酸序列 HSSDYSMWRKNQYVS (SEQ ID NO:10)。

4.权利要求 1 的抗体,其中抗体特异性结合氨基酸序列 DAGTDSGFTFSSPNFATIPQDTV (SEQ ID NO:11)。

5.权利要求 1 的抗体,其中抗体特异性结合氨基酸序列 NEIVDASAVPET (SEQ ID NO:12)。

6.权利要求 1 的抗体,其中抗体是多克隆抗体。

7.权利要求 1 的抗体,其中抗体是单克隆抗体。

8.一种免疫缀合物,其包含缀合于一治疗剂的分离的抗体或抗体片段,所述抗体或抗体片段特异性结合选自以下一组的多肽:

(a) 一种包含图 2(SEQ ID NO:2)所示序列 28—46 位氨基酸的

多肽；

(b) 一种包含图 2(SEQ ID NO:2)所示序列 77—91 位氨基酸的多肽；

(c) 一种包含图 2(SEQ ID NO:2)所示序列 188—210 位氨基酸的多肽；

(d) 一种包含图 2(SEQ ID NO:2)所示序列 263—274 位氨基酸的多肽；

(e) 与 (a), (b), (c) 或 (d)的多肽具有至少 70%相同性的多肽。

9.权利要求 8 的免疫缀合物，其中治疗剂是胞毒剂。

10. 权利要求 9 的免疫缀合物，其中胞毒剂选自蓖麻毒蛋白，阿霉素，道诺红霉素，紫杉醇，溴化乙锭，丝裂霉素，表鬼白毒吡喃葡萄糖苷，tenoposide，长春花新碱，长春花碱，秋水仙素，二羟炭疽菌素，放线霉素 D，白喉毒素，假单胞杆菌外毒素 (PE) A，PE40，相思豆毒蛋白，糖皮质激素和放射性同位素。

11. 权利要求 8 的免疫缀合物，其中抗体片段选自 Fv，F(ab')<sub>2</sub> 和 F(ab')<sub>2</sub> 片段。

12. 权利要求 8 的免疫缀合物在制备用于选择性破坏表达图 2(SEQ ID NO:2)所示多肽的细胞的治疗剂中的应用。

13. 权利要求 8 的免疫缀合物在制备用于治疗与 RG1 表达相关的人体疾病的治疗剂中的应用。

14. 特异性裂解编码多肽的 RNA 的核酶在制备用于治疗与 RG1 表达相关的人体疾病的治疗剂中的应用，所述多肽包含选自以下的成员：

(a) 一种包含图 2(SEQ ID NO:2)所示序列 28—46 位氨基酸序列的多肽，或其生物活性或免疫活性片段；

(b) 一种包含图 2(SEQ ID NO:2)所示序列 77—91 位氨基酸序

列的多肽，或其生物活性或免疫活性片段；

(c) 一种包含图 2(SEQ ID NO:2)所示序列 188—210 位氨基酸序列的多肽，或其生物活性或免疫活性片段；

(d) 一种包含图 2(SEQ ID NO:2)所示序列 263—274 位氨基酸序列的多肽，或其生物活性或免疫活性片段；

(e) 一种与 (a)，(b)，(c) 或 (d) 多肽有至少 70%相同性的多肽。

15. 互补于编码具有如下氨基酸序列的多肽或其一部分的多核苷酸在制备用于治疗人体中与 RG1 非适当表达相关的疾病的治疗剂中的应用：

(a) 图 2(SEQ ID NO:2)所示序列 28—46 位氨基酸序列；

(b) 图 2(SEQ ID NO:2)所示序列 77—91 位氨基酸序列；

(c) 图 2(SEQ ID NO:2)所示序列 188—210 位氨基酸序列；

(d) 图 2(SEQ ID NO:2)所示序列 263—274 位氨基酸序列；

(e) 与 (a)，(b)，(c) 或 (d) 的序列有至少 70%相同性的氨基酸序列。

16. 权利要求 1 的抗体或抗体片段在制备用于通过检测多肽而诊断与 RG1 非适当表达相关的疾病的诊断剂中的应用，所述多肽包含选自以下一组的成员：

(a) 一种包含图 2(SEQ ID NO:2)所示序列的第 28—46 位氨基酸的多肽；

(b) 一种包含图 2(SEQ ID NO:2)所示序列的第 77—91 位氨基酸的多肽；

(c) 一种包含图 2(SEQ ID NO:2)所示序列的第 188—210 位氨基酸的多肽；

(d) 一种包含图 2(SEQ ID NO:2)所示序列的第 263—274 位氨基酸的多肽；

(e) 一种与 (a), (b), (c) 或 (d) 的多肽由至少 70%相同性的多肽。

17. 选自以下一组的多核苷酸或与选自以下一组的成员有至少 70%相同性的多核苷酸在制备用于诊断与 RG1 非适当表达相关的疾病的诊断剂中的应用:

(a) 一种编码多肽, 或其生物活性或免疫活性片段的多核苷酸, 所述多肽包含图 2 (SEQ ID NO:2) 所示序列 28-46 位氨基酸序列;

(b) 一种编码多肽, 或其生物活性或免疫活性片段的多核苷酸, 所述多肽包含图 2 (SEQ ID NO:2) 所示序列 77-91 位氨基酸序列;

(c) 一种编码多肽, 或其生物活性或免疫活性片段的多核苷酸, 所述多肽包含图 2 (SEQ ID NO:2) 所示序列 188-210 位氨基酸序列;

(d) 一种编码多肽, 或其生物活性或免疫活性片段的多核苷酸, 所述多肽包含图 2 (SEQ ID NO:2) 所示序列 263-274 位氨基酸序列;

(e) 与 (a), (b), (c) 或 (d) 的多核苷酸互补的多核苷酸。

18. 权利要求 1 的抗体在制备用于诊断与图 2 (SEQ ID NO:2) 所示多肽相关的转移瘤的诊断剂中的应用。

19. 权利要求 18 的应用, 其中抗体是标记的, 以直接或间接产生可以用放射标记的化合物, 酶, 生色化合物和荧光物可检测的信号。

## 编码新 RG1 多肽的 DNA

本申请要求 1999 年 12 月 16 日提请的美国临时专利申请 No.60/172370 的优先权，该申请在此并入参考。

### 发明背景

本发明涉及新鉴别的多核苷酸和多肽；这种多核苷酸和多肽的变体和衍生物；产生这种多核苷酸，多肽，其变体和衍生物的方法；直接抗这种多肽，其变体和衍生物的抗体；及这种多核苷酸，多肽，变体，衍生物和抗体的应用。本发明尤其涉及新的人体胞外基质多肽（称为 RG1），编码这些多肽的多核苷酸，直接抗这些多肽的抗体，和封阻 RG1 表达的反义多核苷酸。

### 发明背景

前列腺癌是男性经常发生的疾病，在 45 岁以上的男性中有大约 1/3 发现患有此病。主要有遗传和环境两方面的因素，大多数情况是这两种因素组合所致的。关于家族性癌症的研究提示遗传因素在所有前列腺癌的大约 5-10% 中起作用，大约 45% 的情况中男性年龄小于 55 岁。

有这样的迹象即前列腺癌是以多级疾病发生的，一种先期损害是前列腺上皮内瘤形成（PIN）。这种疾病早期是雄激素依赖性的，而后是非激素依赖性的。临床经常检测到一种已知是良性前列腺增生的增殖性前列腺疾病，但也许不是癌症发生的一个时期。然而，其通常与前列腺癌相关。前列腺的癌症通常是多病灶的，一般是缓慢生长的，及是异质的。晚期癌症通常转移至淋巴结和骨。

前列腺癌通常是通过物理检查和通过前列腺特异性抗原 (PSA) 的血清水平加以诊断的。前列腺根除术是治疗局部疾病的一种选择。进展的转移疾病目前是通过睾丸切除术导致的雄激素消融, 或用 GnRH (促性腺激素释放激素) 处理, 及通过抗雄激素疗法进行治疗的。然而, 进展性疾病几乎总是成为对激素有抗性, 而且不能治愈进行性疾病。另外, 基本前列腺切除术和雄激素消融疗法均具有严重的副作用。包括与前列腺根除术相关的尿失禁和阳痿, 与雄激素消融疗法相关的骨折和骨质疏松症。

因此, 需要一种新的治疗早期和晚期前列腺癌的方法。还明显需要新的诊断制剂, 尤其能区别疾病进程的制剂, 因为这能明显影响治疗方案。例如, 如果疾病已经进展超过了前列腺并已经转移至淋巴结, 就不能进行前列腺根除术, 因为这对进展已经没有作用, 但却具有明显不期望的副作用。可以检测体内转移的制剂具有重要意义。

已经证明在前列腺癌中特异性蛋白表达发生变化, 包括在晚期前列腺癌中 p53 表达异常, TGF- $\beta$ 受体水平降低, E-钙粘着蛋白, C-Cam (一种细胞粘着分子), 和一些整联蛋白水平降低。在晚期雄激素非依赖性肿瘤中, 癌基因 bcl-2 的表达明显提高, 而且 bcl-2 高水平表达的患者预后相对不好。前面提及的基因表达中的变化是经充分论证的; 表达中无变化已经鉴别, 这已经证明是致病原因。因此, 这可用于鉴别新的蛋白质, 其表达与前列腺肿瘤的存在和发生相关, 可以用作诊断和治疗前列腺癌的分子标记。

本发明揭示了胞外基质蛋白的超家族的一种新的同源物。这种称为 RG1 的同源物是在前列腺组织中表达的, 并可以在前列腺肿瘤中过表达。

胞外基质是胶原蛋白和弹性蛋白的复合网眼，包埋于由蛋白聚糖和糖蛋白组成的粘弹性基质中。基质存在一个三维支架，分离组织区室，介导细胞吸附和确定组织结构(Bissel等，生物治疗杂志 99: 31-68, 1982; Carlson等，美国科学院院报 78:2403-2406, 1981)。基质发挥大分子滤膜的作用(Hay, E.D., 胞外基质的细胞生物学, 纽约, Plenum 出版社, 1982)，并还影响细胞分化，细胞有丝分裂和形态发生(Gospodarowicz, D. 癌症研究 38:4155-171, 1978)。正常细胞和基质之间的生物化学作用可以在瘤形成中改变，这可以影响肿瘤增殖。肿瘤细胞可以不同方式与基质相互作用。首先，肿瘤细胞可以通过特异性血浆膜受体附着于基质(Terranova等，癌症研究 42:2265-2269, 1982)。其次，基质的降解是通过由肿瘤细胞和宿主提供的酶级联介导的(Eisen等，生物化学生物生理学学报 151:637-645, 1968)。再者，在肿瘤的分化区域内，肿瘤细胞可以合成并积累基质，或诱导宿主细胞积累过量的基质(Brownstein等，癌症 40:2979-2986, 1977)。

RG1 示出与由 Mindin/F-spondin 基因编码的胞外基质蛋白的超家族同源。此基因家族由在氨基末端附近的两个保守的 spondin 结构域 FS1 和 FS2，和在羧基末端的至少一个血小板 spondin 1 型重复 (TSR1) 联合的 (Shimeld, S.M., *Mol.Bio.Evol.* 15(9):1218-1223, 1998)。TSR 基序最初是在脊椎动物胞外基质蛋白中发现的 (Bornstein, P., 细胞生物学杂志 130:503-506, 1995)，并接着在一些其它胞外基质蛋白中也已经发现。有一些迹象表明 TSR's 介导细胞粘附，并在肿瘤发生中起关键作用。例如，已经证实 TSR's 中包含的血小板 spondin 的蛋白酶解片段，及具有相当于血小板

spondin 的 TSR 区域的序列的合成肽，促进肿瘤细胞附着和转移（Prater 等，细胞生物学杂志 112:1031-1040，1991；Tuszynski 和 Nicosia, *BioEssays*18:71-76, 1996），具有抗血管生成活性（Tolsma 等，细胞生物学杂志 122:497-511，1993），并抑制血小板聚集和黑素瘤转移（Tuszynski 等，细胞生物学杂志 116:209-217，1992）。

目前这个超家族的成员包括 *Caenorhabditis elegans* 中的一个具有，*Drosophila* 中的一个单一基因和脊椎动物中的多个基因。在 *C.elegans* 中，基因 F10E7.4 编码除了 FS1 和 FS2 结构域之外的 5 个 TSR's（Higashijima 等，生物学进展 192:211-227，1997）。在 *Drosopnila* 中，称为 M-spondin（mspo）的家族成员含有 FS1 和 FS2 结构域和一个单一的 TSR（Umemiya 等，生物学进展 186:165-176，1997）。M-spondin 基因编码一种分泌的蛋白，其位于肌附着位点并似乎具有胞外基质蛋白的功能，支持肌-内突附着。脊椎动物中的家族成员包括分离自 zebrafish（Mindin 1 和 Mindin 2，F-spondin 1 和 F-spondin2），大鼠 F-spondin，*Xenopus* F-spondin 和大鼠 Mindin 的基因。Mindin 1 和 Mindin 2 彼此密切相关，并具有类似于 *Drosophila* M-spondin 的基因结构。Mindin 1 和 Mindin2 基因均编码除了 FS1 和 FS2 之外的一个单一 TSR（Higashijima 等，生物学进展 192:211-227，1997）。Zebrafish F-spondin 和 F-spondin2，大鼠 F-spondin（Klar 等，细胞 69:95-110，1992）和 *Xenopus* F-spondin（Altaba 等，美国科学院院报 90:8268-8272）基因均具有相似结构，编码除了 FS1 和 FS2 结构域之外的 6 个拷贝的 TSR。在脊椎动物中，Mindin/F-spondin 超家族可以分为两类：与原始的大鼠 F-spondin 和 Mindin 基因密切相关那些基因，和与

*Drosophila* M-spondin 基因密切相关的那些基因。脊椎动物的 Mindin 和 F-spondin 基因均编码蛋白质，所述蛋白质最初是在胚胎发育期间由神经管的底板表达的。

近来从文昌鱼中已经分离了一个单一的 F-spondin 相关的基因，AmphiF-spondin (Shimeld, S.M., *Mol.Biol.Evol.*15(9):1218-1223, 1998)。基于分子系统发生，AmphiF-spondin 是编码 6 个 TSR's 的脊椎动物 F-spondin 基因的特殊亚组密切相关的。AmphiF-spondin 编码 3 个 TSR's 和两个纤连蛋白 III 型重复，其中一个与来自结肠癌缺失 (DCC) 的纤连蛋白 III 型重复非常相同。这个蛋白质的表达在大多数中枢神经系统中发现，并非限于所述脊椎动物 Mindin 和 F-spondin 蛋白的中线。

这些数据提示胞外基质蛋白如新的 RG1 蛋白，是用于癌症诊断和治疗的良好候选物，这种新的 RG1 蛋白是 Mindin/F-spondin 超家族的同源物。

## 发明概要

本发明提供了一种多核苷酸序列，其特异地编码称为 RG1 的一种新蛋白质。RG1 多肽示出与大鼠 Mindin 胞外基质蛋白同源。其含有在 N 末端的一个疏水性序列，在 C 末端的两个 spondin 结构域 (FS1 和 FS2)，和血小板 spondin 1 型重复。RG1 与大鼠 Mindin 有 89.7% 的相似性。在此称为 rg1 并如图 1 所示的多核苷酸序列 (SEQ ID NO:1)，编码图 2 所示的 RG1 的氨基酸序列 (SEQ ID NO:2)。

本发明的一方面提供了多肽，其已经修饰为与胞外基质蛋白的 Mindin 家族同源的新的蛋白质，同源性是通过将图 2 所示氨基酸序列 (SEQ ID NO:2) 和其它胞外基质蛋白已

知氨基酸序列进行对比示出的。

本发明的另一方面提供了编码这种多肽的多核苷酸，尤其编码在此称为 RG1 的多肽的多核苷酸。

根据本发明的一方面，提供了编码 RG1 的分离的多核苷酸，包括 mRNAs, cDNAs, 及在本发明的另外的实施方案中提供了它们的有生物学，诊断，临床或治疗用途的变体，类似物或衍生物，或其片段，包括变体，类似物和衍生物的片段。

本发明这个方面的实施方案中尤为优选的是多核苷酸的自然发生的等位基因，其编码称为 RG1 的多肽的变体。

根据本发明的这个方面，提供了人源的称为 RG1 的新多肽，以及其有生物学，诊断或治疗用途的片段，变体和衍生物，前述片段和类似物的变体和衍生物。

本发明这个方面的实施方案中尤为优选的是由 rg1 多核苷酸的自然发生的等位基因变体编码的 RG1 的变体。

本发明的另一方面提供了一种产生前述多肽，多肽片段，变体和衍生物，前述变体和衍生物和类似物的片段的方法。在本发明的一个优选的实施方案中，提供了产生前述 RG1 多肽的方法，包括在宿主中表达人 RG1 的条件下，培养可表达掺入其中的外源性编码 RG1 的多核苷酸的宿主细胞，然后回收表达的多肽。

本发明的另一方面提供了利用前述多肽和多核苷酸，进行生物学，临床和治疗研究的产物，组合物，程序和方法。

根据本发明此方面的一些优选实施方案中，提供了通过测定 RG1 多肽或编码 RG1 的 mRNA 评估细胞中 RG1 表达的产物，组合物和方法；及分析 rg1 基因中遗传变异和畸变如缺陷的产物，组合物和方法。

根据本发明的这个及其它方面的一些优选实施方案中，提供了与 rg1 序列杂交的探针。

本发明的另一方面提供了对 RG1 多肽或其片段具有高度选择性的抗体，其可用于诊断和/或检测 RG1 表达的方法中，RG1 表达与前列腺癌相关。在根据本发明此方面的一些优选实施方案中，抗体是标记的，由此产生可检测的信号。尤为优选的是用放射标记，酶，生色团或荧光素标记的抗体。

本发明的另一方面提供了与治疗剂缀合的抗体，以施用于在体外的细胞，源于体内的细胞和体内细胞，或多细胞机体。尤为优选的是胞毒性治疗剂。在一些优选的实施方案中，将这种缀合的抗体施用于病人，以治疗特征在于 RG1 活性或表达的疾病，如前列腺癌。

本发明的另一方面提供了可用于刺激免疫应答的肽和抗独特型抗体。

本发明的另一方面提供了核酶和互补于 rg1 多核苷酸和多核苷酸（即反义多核苷酸），以施用于在体外的细胞，源于体内的细胞和体内细胞，或多细胞机体。尤为优选的是将此反义分子施用于病人，以治疗如前列腺癌或早期前列腺增生等疾病，通过降低 RG1 活性水平而减轻症状。

本领域技术人员通过以下详细论述可以理解本发明的其它方面，特点，优点和目的。然而应知道以下阐述和特异的实施例，及所述的优选实施方案，只是举例说明本发明。通过阅读以下论述和本发明的其它部分，本领域技术人员在本发明精神和所示范围内，可以对本发明加以各种变化和修改。

## 附图简述

图 1: 编码生物活性或免疫活性形式 RG1 的 rg1 的多核苷酸序列 (SEQ ID NO:1)。

图 2: RG1 的推导的氨基酸序列 (SEQ ID NO:2), 单下划线处为 F-spondin 结构域, 双下划线处为血小板 spondin 结构域。

图 3: RG1 与大鼠 Mindin 序列的氨基酸序列对比。RG1 序列在上部。

图 4: RG1 的多核苷酸和推导的氨基酸序列。

图 5: 通过基于 Taqman 的 PCR 分析 rg1 mRNA 在人体组织中表达。来自人体肿瘤和正常组织的 RNA 是通过标准方法分离的。检测 rg1 mRNA 表达的引物和探针是用 Perkin Elmer' s 引物表达软件设计的, 并通过合成遗传学合成。在人前列腺组织中检测 Rg1 mRNA。在其它组织中可以检测到较低水平表达的 rg1 mRNA, 例如在肝脏中。

图 6: 纯化由 LNCaP 细胞分泌的天然 RG1 蛋白。使用抗合成的 RG1 肽序列 (3C, SEQ ID NO:10; 见实施例 4) 产生的抗血清进行 Western 印迹分析, 以检测 LNCaP 细胞分泌的天然 RG1 蛋白。浓缩的 LNCaP 细胞条件培养基的 Q-琼脂糖凝胶层析的洗脱级分: (L) 层析柱充填物, (F) 层析柱流量, (1-12) 经过盐梯度的洗脱级分。预测的 RG1 分子量为大约 36kD, 然而观测到细菌表达的 RG1, BHK 表达的 RG1 和 LNCaP 表达的 RG1 在 PAGE 上均在大约 45kD 迁移 (L, 级分 6-9)。

图 7: 在人体前列腺组织中 RG1 表达的免疫组织化学染色。前列腺组织得自 Stanford 大学医学院的泌尿系。用载体 ABC-AP 试剂盒 (AK5002) 进行染色。用载体红色底物

试剂盒（SK-5100）观测染色，并用苏木精对染。结果示出在腺体结构中外周网眼膜明显染色。

## 发明详述

### 词解

本发明的说明书，实施例和所附权利要求中所使用的术语，除非特别说明均具有以下含义。

“RG1”是指具有图 2 所示氨基酸序列（SEQ ID NO:2）的多肽；其变体，类似物，衍生物和片段，及变体，类似物和衍生物的片段。术语“片段”，“衍生物”和“类似物”当是指图 2 的多肽（SEQ ID NO:2）时，是指基本保留与图 2 的多肽（SEQ ID NO:2）相同的生物和/或免疫学活性的多肽。

“rg1”是指具有图 1（SEQ ID NO:1）所示氨基酸序列的多核苷酸，和编码具有图 2（SEQ IN NO:2）所示 RG1 的氨基酸序列的多肽的多核苷酸；及指编码 RG1 变体，类似物，衍生物和片段，和变体，类似物和衍生物的片段的多核苷酸。Rg1 还指由 RNA 组成的这种多核苷酸，以及是编码图 2（SEQ ID NO:2）所示多肽序列的多核苷酸的补体的多核苷酸。

“多核苷酸”一般是指任何多核糖核苷酸或多脱氧核糖核苷酸，其可以是未修饰的 RNA 或 DNA 或修饰的 RNA 或 DNA。因此，例如本发明使用的多核苷酸是指单链和双链 DNA，单链和双链区混合的 DNA，单链和双链的 RNA，和单链和双链区混合的 RNA，包含 DNA 和 RNA 的杂交分子，其可以是单链的，或更典型是双链的或单链和双链区的混合物。另外，本发明使用的多核苷酸是指包含 RNA 或 DNA 或 RNA 和 DNA 的三链区域。这种区域中的链可以来自相

同分子或来自不同分子。这个区域可以包括所有一或多个分子，但更典型地只包含一些分子的一个区域。三螺旋区域的分子之一通常是寡核苷酸。

本发明所用术语“多核苷酸”包括含有一或多个修饰的碱基的上述 DNAs 或 RNAs。因此，“多核苷酸”在此是指由于稳定性或其它原因骨架经修饰的 DNAs 或 RNAs。另外，本发明多核苷酸也指包含不同寻常的碱基如肌苷，或修饰的碱基如氘标记的碱基的 DNAs 或 RNAs，所述碱基只是举例而非全部。

应意识到本领域技术人员可以根据多种使用目的，对 DNA 和 RNA 加以多种修饰。本发明所用术语“多核苷酸”包含这种经化学，酶学或代谢修饰形式的多核苷酸，以及病毒和细胞特有的化学形式的 DNA 和 RNA，所述细胞包括简单和复杂的细胞。

本发明所用的“多肽”包括以下所述的所有多肽。多肽的结构已经熟知，并可见于本领域无数的教材和其它出版物所述。文中此术语是指任何肽或蛋白质，其包含在一个线性链中通过肽键彼此连接的两或多个氨基酸。如本发明所使用，此术语指本领域通常也例如称为肽，寡肽和寡聚物的短链，及指本领域通常称为蛋白质的较长的链，其中它们具有许多类型。

应意识到多肽通常含有除了通常称为天然发生的 20 个氨基酸之外的氨基酸，而且许多氨基酸包括末端氨基酸在给定的多肽中可以是修饰的，通过自然程序如糖基化和其它翻译后修饰，或通过本领域熟知的化学修饰方法进行修饰。甚至多肽中天然发生的常见修饰也太多，以至于在此不能一一列举，但在基本的教材和更详细的专论以及长篇

研究文献中有充分阐述，而且这些内容为本领域技术人员所熟知。在已知的修饰中，可以存在于本发明多肽中的是乙酰化，酰化，ADP-核糖基化，酰胺化，黄素的共价附着，血红素部分共价附着，多核苷酸或多核苷酸衍生物的共价附着，脂质或脂质衍生物的共价附着，phosphatidylinositol的共价附着，交联，环化，二硫键形成，脱甲基化，共价交联形成，胱氨酸形成，焦谷氨酸形成，甲酰化， $\gamma$ -羧化，糖化，糖基化，GPI 锚形成，羟基化，碘化，甲基化，十四烷酰化，氧化，蛋白酶解，磷酸化，异戊二烯化，外消旋化，硒化，硫酸化，将氨基酸通过转移 RNA 介导加入蛋白质如精氨酰化，和遍在蛋白化。

这些修饰是本领域技术人员熟知的，并见于科学文献中的详细阐述。一些特别普通的修饰例如糖基化，脂质附着，硫酸化，谷氨酸刺激的 $\gamma$ -羧化，羟基化，和 ADP-核糖基化在最基本的教材中加以论述，如 I.E.Creighton, 蛋白质结构和分子学性质，第二版，W.H.Freeman 和 Company, 纽约，1993。关于这方面的一些详细论述例如见于 Wold, F., 蛋白质的翻译后共价修饰，B.C.Johnson 编辑，科学出版社，纽约，pp1-12, 1983; Seifter 等，酶学方法 182:626-646, 1990, 和 Rattan 等，蛋白质分析：翻译后修饰和老化，Ann.N.Y.Acad.Sci.663:48-62, 1992。

如所熟知的及如上所述，应意识到多肽并不总是完全线性的。例如多肽可以是由于遍在化结果而分支的，及可以有或无分支的环形，通常是由于翻译后的结果，包括天然形成的和不是天然发生的人为操纵所致。环形的，分支的和分支环形的多肽，可以通过非翻译天然程序及通过完全合成方法合成的。

修饰可以在多肽中的任何部位发生，包括肽骨架，氨基酸侧链和氨基或羧基末端。事实上，通过共价修饰封阻多肽中的氨基团或羧基团或这二者，在天然发生的和合成的多肽中是常见的，而且这种修饰可以存在于本发明的多肽中。例如，在蛋白酶解之前在大肠杆菌中产生的多肽的氨基末端残基，几乎总是 N-乙酰甲硫氨酸。

多肽中发生的修饰通常是其产生方式的作用。例如，通过在宿主中表达克隆的基因产生的多肽，修饰的性质和程度大部分是通过宿主细胞翻译后修饰的能力和肽氨基酸序列中存在的修饰信号确定的。例如熟知的糖基化通常不在细菌宿主如大肠杆菌中发生。因此，当需要糖基化时，多肽应在糖基化宿主中表达，一般为真核细胞。昆虫细胞通常完成与哺乳动物细胞相同的翻译后糖基化，为此已经产生了昆虫细胞表达系统，以有效表达具有天然糖基化模式的哺乳动物蛋白质。对其它修饰也需要考虑相似的问题。

应意识到相同类型修饰在给定的多肽中的一些位点可以相同或不同程度存在。同样，给定的多肽可以含有许多类型的修饰。

如本文使用，术语多核苷酸涵盖了所有这种修饰，尤其在通过在宿主细胞中表达多核苷酸而合成的多肽中存在的那些修饰。

本文所用的术语“编码多肽的多核苷酸”涵盖了这样的多核苷酸，其包括编码本发明多肽，尤其具有图 2(SEQ ID NO:2)所示氨基酸序列的 RG1 多肽的序列。该术语涵盖了这样的多核苷酸，其包括一个编码多肽的连续区域或不连续区域（例如由内含子中断），及附加的区域。

“生物学活性”是指天然发生的 RG1 多肽的结构性，

调节性或生物化学功能。

“免疫学活性”是指天然的，重组的或合成的 RG1 或其片段，在适当的动物或细胞中诱导特异性免疫应答，及与特异性抗体结合的能力。

“寡核苷酸”是指相对短的多核苷酸。该术语通常是指单链的脱氧核糖核苷酸，但也可以指单链或双链的核糖核苷酸，RNA:DNA 杂交体和双链的 DNAs。寡核苷酸，如单链的 DNA 探针寡核苷酸，通常是通过化学方法合成，如在自动寡核苷酸合成仪上合成。然而，可以通过各种其它方法产生寡核苷酸，包括体外重组 DNA 方法和通过在细胞和有机体中表达 DNA 的方法。“寡核苷酸”或“寡聚物”或多核苷酸“片段”，“部分”或“节段”是指具有至少大约 10 个核苷酸，及多如大约 60 个核苷酸，优选大约 15—30 个核苷酸，更优选大约 20—25 个核苷酸的多核苷酸序列。

“天然发生的 RG1”是指由未进行遗传工程化的人体细胞产生的 RG1，特别是期望是产生自多肽翻译后修饰的各种 RG1 形式，所述修饰包括但非限于乙酰化，羧基化，糖基化，磷酸化，脂质化，酰化和裂解。

多核苷酸或多肽的“变体”是指分别与参比多核苷酸或多肽不同的多核苷酸或多肽。这种含义的变体如以下和在本发明揭示的其它章节所述。

(1)与另一个参比多核苷酸的多核苷酸序列不同的多核苷酸。通常不同之处是限定的，以便参比的多核苷酸序列和变体是非常相似的，并在许多区域是相同的。

如下所述，变体的多核苷酸序列中的变化可以是沉默的。即它们可以不改变由多核苷酸编码的氨基酸。在改变是限于这种类型的沉默改变之处，变体将编码与参比序列

具有相同氨基酸序列的多肽。如下所述，变体的多核苷酸序列中的变化可以改变由参比多核苷酸编码的多肽的氨基酸序列。这种多核苷酸变化可以使由参比序列编码的多肽中产生氨基酸取代，添加，缺失，融合和截短，如下所述。

(2)与另外的参比多肽氨基酸序列不同的多肽。通常不同之处是限定的，以便参比多肽和变体的序列是非常相似的，并在许多区域是相同的。变体和参比多肽的氨基酸序列可以由于一或多个取代，添加，缺失，融合和截短而不同，它们可以以任何组合存在。编码这些相同或相似多肽的重组变体可以是合成的，或通过利用“遗传密码”丰余性而选择的。可以进行各种密码子取代如产生各种限制位点的沉默变化，以最佳化克隆入质粒或病毒载体中或在特殊的原核或真核系统中表达。也可以进行诱变以修饰多肽的性质，改变配体结合亲和性，链间亲和性，或多肽降解或周转率。

“等位基因变体”是指 rg1 多核苷酸的另一种形式。等位基因得自突变即多核苷酸序列发生变化，并通常产生其结构或功能可以或不改变的变化的 mRNAs 或多肽。任何给定的基因可以没有，或有一或多种等位基因形式。产生等位基因的突变一般是由于核苷酸的天然缺失，添加或取代产生的。这些类型的变化均可以单独或组合发生，或在给定的序列中发生一或多次。

“衍生物”是指分别衍生自天然发生的 rg1 或 RG1 的多核苷酸或多肽，通过在人体蛋白内不正常发生的化学修饰如遍在化，标记（例如用放射性核素，各种酶促修饰标记），pegylation（用聚乙二醇衍生），或通过氨基酸如鸟氨酸插入或取代（或取代编码这种氨基酸的核苷酸）而衍生。

“缺失”是指多核苷酸或氨基酸序列中分别缺少一或多个多核苷酸或氨基酸残基的变化。

“插入”或“添加”是指多核苷酸或氨基酸序列中的变化，与天然发生的多核苷酸或氨基酸序列相比，分别添加了一或多个多核苷酸或氨基酸残基。

“取代”是指分别用不同的多核苷酸或氨基酸置换一或多个多核苷酸或氨基酸。

优选地，氨基酸取代是将一个氨基酸用具有相似结构和/或化学性质的另一种氨基酸置换，如用异亮氨酸或缬氨酸置换亮氨酸，用谷氨酸置换天冬氨酸，或用丝氨酸置换苏氨酸，即保守氨基酸置换。插入或缺失典型地在大约 1—5 个氨基酸范围内。可以通过在多肽中用重组 DNA 方法系统地产生氨基酸插入，缺失或取代，并分析所得重组变体的活性，对发生的变化加以实验性测定。

“片段”是指一种多肽，其氨基酸序列与前述 RG1 多肽及其变体或衍生物的氨基酸序列部分而非全部相同。

多肽的“片段”“部分”或“节段”是指至少大约 5 个氨基酸，通常至少大约 7 个氨基酸，典型地至少大约 9—13 个氨基酸，及在各种实施方案中至少大约 17 或更多个氨基酸的一段氨基酸序列。

“重组体”或“重组的 DNA 分子”是指不是天然发生的，或是通过人为组合序列的两个其它分离的节段而产生的多核苷酸序列。“重组产生的”是指通常通过化学合成方法，或通过人为操纵分离的多核苷酸节段，例如通过基因工程技术所达到的人为组合。这种操纵通常是将一个密码子用编码相同或保守氨基酸的丰余密码子置换，同时典型地导入或除去一个序列识别位点。或者，与具有所需功能

的多核苷酸节段连接在一起，以产生一个遗传本体，其包含在通常的天然形式中未发现的所需功能组合。限制酶识别位点，调节序列，控制序列，或其它有益的特点可以通过设计掺入其中。“重组的 DNA 分子”包括克隆和表达载体。“重组体”还可以指编码多肽的多核苷酸，是用重组 DNA 技术制备的。

“分离的”是指“通过人为”从其天然环境中改变，即如果是天然发生的，将其从原始环境中改变或分离。例如，以其天然状态天然存在于生活动物中的天然发生的多核苷酸或多肽，不是“分离的”，但从其天然状态的共存物中分离的相同多核苷酸或多肽是“分离的”。例如，关于多核苷酸，术语分离是指从在其中其天然发生的染色体和细胞中分离的。

多核苷酸和多肽可以存在于组合物中，如将多核苷酸或多肽例如导入细胞中的培养基配方或溶液中，及存在于进行化学或酶促反应的组合物或溶液中，例如其不是天然发生的组合物，其中保留本文所用的此术语含义范围内的分离的多核苷酸或多肽。

“基本纯化”和“基本同源”是交替用于阐述 RG1 多肽，或其片段，或编码其的多核苷酸节段，其中这种多肽或多核苷酸分离自天然涵盖其的组分中。当从涵盖其天然状态的天然污染物中分离时，RG1 多肽或其片段，或编码其的 DNA 节段是基本没有天然相关的组分。因此，化学合成的或在不同于其天然起源的细胞的细胞系统中合成的多肽，基本没有其天然相关的组分。

“同源”这一术语当用于阐述多核苷酸时，是指两个多核苷酸或其指定的序列，当进行最佳序列对比时具有适当

的核苷酸插入或缺失，至少 70%核苷酸，通常为大约 75%—99%，优选至少大约 98—99%的核苷酸是相同的。

“相似性”这一术语当用于阐述多肽时，是通过将一个多肽的氨基酸序列和保守的氨基酸取代与另一个多肽的序列进行对比而测定的。

“聚合酶链反应”或“PCR”是指一种方法，其中如 1987 年 7 月 28 日发布的美国专利 No.4683195 所述扩增特异的 DNA。通常地，需要获得相应的或远处多肽片段末端的序列信息，这样可以设计寡核苷酸引物；这些引物指向一定方向，并与扩增模板的相反链的序列相同或相似。两种引物的 5' 末端核苷酸与扩增物的末端一致。PCR 可用于扩增来自总基因组 DNA 的特异的 DNA 序列，转录自总细胞 RNA 的 cDNA，质粒序列等（见 Mullis 等，冷泉港 Symp.Quant.Biol., 51:263, 1987; Erlich 编辑，PCR 技术，Stockton 出版社，NY，1989）。

“严格性”典型是指在大约  $T_m$ （解链温度）-5° C（低于探针  $T_m$  5° C）至大约 20—25° C  $T_m$  范围内。正如本领域技术人员所意识到，严格杂交可用于鉴别或检测相同的多核苷酸序列，或鉴别或检测相似的或相关的多核苷酸序列。本文所用术语“严格条件”是指杂交只在序列之间存在至少 95%优选至少 97%相同性的情况下发生。

本文所用术语“杂交”应包括“将多核苷酸链与其互补链通过碱基配对连接的任何方法”（Coombs, J., 生物技术词典，Stockton 出版社，纽约，N.Y., 1994）。

“治疗有效量”是指减轻疾病症状或状况的多肽或其抗体，拮抗剂，或抑制剂，包括反义分子和核酶的数量。这种化合物的治疗效力和毒性，可以在细胞培养物或实验动

物中通过标准药理学方法确定，例如  $ED_{50}$ （半数治疗有效量）和  $LD_{50}$ （半数致死量）。治疗作用和毒性作用之间的剂量比是治疗指数，并可以  $ED_{50}/LD_{50}$  表示。

“治疗”是指治疗患者所患疾病，这种疾病与前列腺肿瘤生长相关，并包括这样的疾病，其中需要降低患者的 RG1 水平。

### 发明详述

本发明涉及新的 RG1 多肽, rg1 多核苷酸, 和直接抗 RG1 多肽的抗体及其它物质, 如以下更详细的论述。本发明尤其涉及新的 RG1 多肽和编码这些 RG1 多肽的多核苷酸, 并特别涉及具有图 2(SEQ ID NO:2)所示氨基酸序列的 RG1, 和具有图 1(SEQ ID NO:1)所示多核苷酸序列的 rg1。本发明还涵盖了 RG1 变体。一种优选的 RG1 变体与图 2(SEQ ID NO:2)所示多肽序列有至少 70%相似性（优选至少 70%相同性），优选具有至少 90%相似性（优选至少 90%相同性），更优选具有至少 95%相似性（优选至少 95%相同性），并也包括这种多肽的一部分，多肽的这部分一般含有至少 30 个氨基酸及优选至少 50 个氨基酸。

预测的 RG1 多肽的编码序列起自图 1（SEQ ID NO:1）所示核苷酸序列的 5' 末端的 296 个碱基对。RG1 含有胞外基质蛋白的 Mindin/F-spondin 超家族的 3 个特征结构域：两个 spondin 结构域（FS1 和 FS2），分别包含 31—103 和 138—221 位氨基酸，和一个血小板 spondin 结构域，包含 278—330 位氨基酸。

本发明部分基于图 3 所示的 RG1 和大鼠 Mindin 之间的结构同源性，大鼠 Mindin 是胞外基质蛋白家族的另一个成

员。RG1 的氨基酸序列与大鼠 Mindin 具有接近 89.7%的相似性。

本发明还部分基于 RG1 的表达分布图，通过其在前列腺组织文库中表达及在前列腺肿瘤文库中过表达证实。通过基于 PCR 的 Taqman 分析，在来自正常和肿瘤组织的组织样品中分析 mRNA 表达，可以见到这种组织分布图。这种分析方法表明编码 RG1 的 mRNA 与其它组织相比，在前列腺组织中是过表达的。

### 多核苷酸

根据本发明的一方面，提供了编码 RG1 多肽的分离的多核苷酸，所述多肽具有图 2(SEQ ID NO:2)所示推导的氨基酸序列。

使用所提供的信息，如图 1 (SEQ ID NO:1) 所示的多核苷酸序列，本发明编码 RG1 多肽的多核苷酸可以使用标准克隆和筛选方法获得，如使用来自人体组织细胞的 mRNA 作为起始物克隆 cDNA 的那些方法。本发明举例说明了图 1(SEQ ID NO:1)所示多核苷酸序列是在得自人体前列腺组织的 cDNA 克隆中发现的。通过 mining Incyte' s LifeSeq 数据库，鉴别 rg1 是在前列腺组织中表达的基因。核苷酸序列是使用 Incyte 为研究数据库而提供的“蛋白质功能”工具，通过对数据库进行解读研究而鉴别的。核苷酸序列是在解读的数据库中细胞粘着分子类别中发现的，并描述为是 f-spondin 的同源物。对 rg1 多核苷酸序列在数据库的一系列文库中的分布进行的电子 Northern 分析表明，rg1 在前列腺组织中高水平表达，在一些其它组织文库中以较低水平表达。

在将数据库中 rg1 克隆系列装配成连续的多核苷酸序

列，并对连续的序列进行编辑后，在预装配的多核苷酸中鉴别全长编码序列。这个编码蛋白质的序列与大鼠 mindin 同源。

Lncyte 克隆 1640796, 1712252 和 1880265 得自 lncyte 以进行实验研究，克隆 3360733 经鉴别含有大多数 5' 核苷酸序列。这个克隆是经过充分测序的，并含有预测的 RG1 蛋白的全部编码序列。这个序列适于图 1 (SEQ ID NO:1)。

本发明的多核苷酸可以是 RNA 形式如 mRNA，或 DNA 形式，例如包括 cDNA 和基因组 DNA，通过克隆获得或通过化学合成方法或通过组合方法或本发明所述方法产生。DNA 可以是双链或单链的。单链的 DNA 可以是编码链（也称为有义链），或者可以是非编码链（也称为反义链）。

编码多肽的序列可以与图 1 (SEQ ID NO:1) 所示多核苷酸的编码序列相同。其也可以是由于遗传密码的冗余（简并）而具有不同序列的编码图 2 (SEQ ID NO:2) 的多核苷酸。

本发明编码图 2 (SEQ ID NO:2) 所示多肽的多核苷酸，可以包括但非限于多肽自身的编码序列；多肽的编码序列，及另外的非编码序列，包括例如但非限于内含子和非编码 5' 和 3' 序列，如在转录，mRNA 加工（例如剪接和聚腺苷酸化信号）起作用的转录的，未翻译序列，或编码如提供其它功能的其它氨基酸的另外的编码序列。因此，多肽例如可以融合于标记序列，如易于纯化融合多肽的肽。在本发明的一些优选的实施方案中，标记序列是一种 6-组氨酸肽，如 pTrcHisB 载体 (Invitrogen, Carlsbad, CA) 提供的标签，其中许多是可以商购的。如 Gentz 等 (美国科学院报 86:821-824, 1989) 所述，6-组氨酸可以常规纯化融合蛋白。

多核苷酸可以编码一种多肽，这种多肽是加上另外的氨基或羧基末端氨基酸，或多肽内部的氨基酸（例如当活性形式具有一个以上多肽链时）的多肽。这种序列可以在将多肽从前体加工为最终形式中起作用，可以易于多肽运输，可以延长或缩短多肽半衰期，或可以易于操纵多肽进行分析或产生等等。由于一般是在原位的情况下，另外的氨基酸可以通过蛋白酶从多肽中除去。

本发明还涉及所述多核苷酸的变体，其编码具有图 2(SEQ ID NO:2)所示推导的氨基酸序列的多肽的配对，类似物和衍生物。多核苷酸的变体可以是天然发生的变体如天然发生的等位基因变体，或者可以是未知是天然发生的变体。多核苷酸的这种非天然发生的变体可以通过诱变方法产生，包括用于多核苷酸，细胞或有机体的那些方法。

变体中在这点上是由于多核苷酸取代，缺失或添加而不同于前述多核苷酸的变体。取代，缺失或添加可以包含一或多个多核苷酸。变体可以是在编码区或非编码区或二者中改变的。编码区中的变化可以产生保守的或非保守的氨基酸取代，缺失或添加。

在本发明特别优选的实施方案中，在这点上编码具有图 2(SEQ ID NO:2)所示 RG1 的氨基酸序列的多肽的多核苷酸；其变体，类似物，衍生物和片段，及变体，类似物和衍生物的片段。

特别优选的是编码 RG1 变体，类似物，衍生物和片段，及片段的变体，类似物和衍生物的多核苷酸，其具有图 2(SEQ ID NO:2)所示 RG1 多肽的氨基酸序列，其中一些，少量，5-10，1-5，1-3，2，1 个或没有氨基酸残基是任何组合的取代，缺失或添加的。尤为优选的是不改变 RG1 多肽的性质

和活性的沉默取代，添加和缺失。同样尤为优选的是保守取代。最优选的是编码具有图 2(SEQ ID NO:2)所示没有取代的氨基酸序列的多肽的多核苷酸。

本发明另外优选的实施方案是这样的多核苷酸，其与编码具有图 2(SEQ ID NO:2)所示氨基酸序列的 RG1 多肽的多核苷酸具有至少 70% 相同性，及与这种多核苷酸互补的多核苷酸。或者，最优选的是包含与编码 RG1 多肽的多核苷酸有至少 80% 相同性的区域的多核苷酸，及其互补多核苷酸。在这点上，尤为优选的是具有至少 90% 相同性的多核苷酸，特别优选的是具有至少 95% 相同性的多核苷酸。另外，更为优选的是具有至少 97%，98% 和 99% 相同性的多核苷酸。

此外，特别优选的是编码基本保留与图 1 (SEQ ID NO:1) 所示多核苷酸序列编码的多肽相同生物活性的多肽的多核苷酸。

本发明还涉及与上述序列杂交的多核苷酸。在这点上，本发明特别涉及在严格杂交条件下，与上述多核苷酸杂交的多核苷酸。

如关于本发明多核苷酸的其它分析所揭示，例如上述的本发明多核苷酸可以用作 cDNA 和基因组 DNA 的杂交探针，以分离编码 RG1 的全长 cDNAs 和基因组克隆，及分离与 rg1 基因具有高度序列相似性的其它基因的 cDNA 和基因组克隆。这种探针通常包含至少 15 个碱基。优选地，这种探针具有至少 30 个碱基，并可以具有至少 50 个碱基。

例如，可以使用合成的寡核苷酸探针筛选文库分离 rg1 基因的编码区，所述探针是用已知 DNA 序列设计的。例如，具有与本发明多核苷酸的序列互补的序列的标记的寡核苷

酸，可以用于筛选 cDNA 或基因组 DNA 的文库，以鉴别与探针杂交的克隆。

总而言之，本发明的多核苷酸可以编码一种多肽，一种加上引导序列的多肽（可以称为前多肽）。

应意识到本发明还涉及编码多肽片段的多核苷酸，与编码多肽片段的多核苷酸杂交尤其在严格条件下杂交的多核苷酸，及多核苷酸如 PCR 引物，以扩增编码多肽片段的多核苷酸。优选的是相当于优选的多肽片段的那些多核苷酸，如下述。

### 多肽

本发明还涉及具有图 2(SEQ ID NO:2)所示推导的氨基酸序列的 RG1 多肽。

本发明还涉及这些多肽的片段，类似物和衍生物。术语片段，衍生物和类似物当指图 2(SEQ ID NO:2)所示多肽时，是指基本保留与这种多肽相同生物活性的多肽。

本发明的多肽可以是重组多肽，天然多肽或合成的多肽。在一些一些的实施方式中，是重组的多肽。

图 2(SEQ ID NO:2)所示多肽的片段，衍生物或类似物可以是 (i) 其中一或多个氨基酸残基用保守或非保守的氨基酸残基取代（一些保守的氨基酸残基），而且这种取代的氨基酸残基可以是或不是由遗传密码编码的，或 (ii) 其中一或多个氨基酸残基包括一个取代基，或 (iii) 其中多肽与另一种化合物融合，如提高多肽半衰期的化合物（例如聚乙二醇），或 (iv) 其中另外的氨基酸与多肽融合，如引导序列或分泌序列或用于纯化多肽的序列。通过本文教导，认为这种片段，衍生物和类似物在本领域技术范围内。

在这点上本发明特别一些的实施方式是具有图 2(SEQ

ID NO:2)所示 RG1 的氨基酸序列的多肽，其变体，类似物，衍生物和片段，及片段的变体，类似物和衍生物。

优选的变体是由于保守氨基酸取代而不同于参照物的那些变体。这种取代是用另一种相似特性的氨基酸取代多肽中给定的氨基酸。典型可见的保守取代是脂肪族氨基酸 Ala, Val, Leu, 和 Ile 的彼此置换，羟基 Ser 和 Thr 的互换，酸性 Asp 和 Glu 的交换，酰胺基 Asn 和 Gln 之间的取代，碱性 Lys 和 Arg 的交换及芳香族 Phe 和 Tyr 之间的置换。

在这点上尤为优选的是具有图 2(SEQ ID NO:2)所示 RG1 多肽的氨基酸序列的变体，类似物，衍生物和片段，及片段的变体，类似物和衍生物，其中一些，少量，5-10，1-5，1-3，2，1 个氨基酸残基或没有氨基酸残基是任何组合的取代，缺失或添加的。尤为优选的是不改变 RG1 多肽性质和活性的沉默取代，添加和缺失。同样尤为优选的是保守取代。最优选的是具有图 2(SEQ ID NO:2)所示没有取代的氨基酸序列的多肽。

本发明的多肽和多核苷酸优选是以分离形式提供的，优选纯化至同质性。

本发明的多肽还包括图 2(SEQ ID NO:2)所示多肽，以及与图 2(SEQ ID NO:2)所示多肽具有至少 70%相似性（优选 70%相同性），优选具有至少 90%相似性（优选 90%相同性），更优选具有至少 95%相似性（优选 95%相同性）的多肽，还包括这种多肽的一部分，多肽的这部分一般含有至少 30 个氨基酸，优选至少 50 个氨基酸。

如本领域已知两个多肽之间的“相似性”，是将一个多肽的氨基酸序列及其保守的氨基酸取代与另一个多肽的序列相对比而测定的。

通过肽合成方法，本发明多肽的片段或一部分可以用于产生相应的全长多肽；因此，片段可以用作产生全长多肽的中间物。

### 片段

本发明在这个方面优选的实施方案是这样的多肽，其包含 RG1 的片段，尤其图 2(SEQ ID NO:2)所示 RG1 的片段，及图 2(SEQ ID NO:2)所示 RG1 的变体和衍生物的片段。

在这点上，片段是具有一种氨基酸序列的多肽，所述氨基酸序列与前述 RG1 多肽及其变体或衍生物的氨基酸序列部分而非全部相同。

在这点上，片段是一种多肽，其具有的氨基酸序列与前述 RG1 多肽及其变体或衍生物的氨基酸序列部分而非全部相同。

这种片段可以是“无标准”的，即不是其它氨基酸或多肽的一部分或与其融合，或者可以包含于较大的多肽中，形成较大多肽的一部分或一个区域。当包含于较大的多肽中时，所述片段优选形成一个连续的区域。然而，一些片段可以包含于一个较大的多肽中。例如，一些优选的实施方案涉及包含于一种前体多肽中的本发明 RG1 多肽的一个片段，所述前体多肽是设计的以在宿主中表达，并具有融合于 RG1 片段氨基末端的异源的前多肽和多肽原区域，及融合于此片段羧基末端的另外的一个区域。因此，所述片段是指衍生自 RG1 的融合多肽或融合蛋白的一部分。

本发明多肽片段的代表性实施例可以具有大约 25—331 个氨基酸。

文中“大约”是指在一个末端或两个末端具有比所述范围多或少一些，少量，5，4，3，2，和 1 个氨基酸。例

如，文中大约 331 个氨基酸是指一种多肽片段，其具有 25 个加上或减去一些，少量，5，4，3，2，和 1 个氨基酸至 331 个氨基酸加上或减去一些，少量 5，4，3，2，和 1 个氨基酸残基，即在大如 25 个减去一些氨基酸至 331 个加上一些氨基酸的范围，及小如 25 个加上一些氨基酸至 331 个减去一些氨基酸的范围内。

在这方面优选的是在一或两个末端在所述范围内加上或减去如 5 个氨基酸。尤为优选的是在一或两个末端在所述范围内加上或减去如 3 个氨基酸。更优选的是在一或两个末端在所述范围内加上或减去如 1 个氨基酸，或不添加或缺失氨基酸。最优选的是大约 25—331 个氨基酸的片段。

本发明优选的片段是 RG1 的截短的突变体。RG1 的截短突变体包括图 2 (SEQ ID NO:2) 所示序列的变体或衍生物，除了缺失一个连续系列的残基（即一个连续区域，部分或一部分），包括图 2 (SEQ ID NO:2) 所示序列的氨基末端，或者一个连续系列的残基，包括羧基端残基，或在双截短体中是缺失两个连续系列的残基，一个包括氨基端，一个包括羧基端。优选的截短片段是具有上述大小的片段。

尤为优选的是具有 RG1 生物学和/或免疫学活性的片段。这种片段包括那些含有 RG1 预知结构性结构域的片段，其涵盖了至少 31—103，138—221 和 278—330 个氨基酸的片段，或那些用于产生抗体的片段，如实施例 4 所述。

在这方面一些优选的区域示于图 2 (SEQ ID NO:2)，并包括但非限于通过分析图 2 (SEQ ID NO:2) 所示氨基酸序列鉴别的上述类型的区域。

在这方面尤为优选的是那些包含 RG1 区域的片段，所述区域组合了一些如上述的结构特点。尤为优选的是两个

spondin 和一个 thrombospondin 区域，它们分别含有大约 31—103，138—221 和 278—330 个氨基酸，具有胞外基质蛋白的 Mindin/spondin 超家族的特点。这种区域可以包含在较大的多肽内，或者其自身就是本发明优选的片段。应意识到文中术语“大约”一般是指关于上述片段所作的解释。

另外优选的区域是那些介导 RG1 活性的区域。最优选的是具有 RG1 的化学，生物学或其它活性的片段，包括那些具有相似或改良活性，或具有降低的非所需活性的片段。在这方面特别优选的是这样的片段，其含有的区域在序列，位置或序列和位置是同源的，以激活相关多肽的区域，如 Mindin 家族的其它蛋白质，包括 RG1。

#### 载体，宿主细胞和表达系统

本发明还涉及包括本发明多核苷酸的载体，用本发明载体遗传工程化的宿主细胞，和通过重组方法产生本发明多肽。这种方法见于 Sambrook 等，分子克隆实验手册，冷泉港实验室出版社，Plainview, N.Y., 1989 和 Ausubel, F.M. 等，分子生物学通用方法，John Wiley&Sons, 纽约，N.Y., 1989。

可以对宿主细胞进行遗传工程化，以掺入多核苷酸并表达本发明的多核苷酸。例如，可以通过熟知的感染，转导，转染，转介和转化方法，将多核苷酸导入宿主细胞中。可以将多核苷酸单独或与其它多核苷酸一起导入。这种其它的多核苷酸可以单独导入，与本发明的多核苷酸一起导入，或与本发明的多核苷酸结合导入。

因此，例如可以使用标准的在哺乳动物中进行共同转染和选择的方法，将本发明的多核苷酸转染入具有另一种单独的编码可选择标记的多核苷酸中。在这种情况下，多

核苷酸通常是稳定地掺入宿主细胞基因组中。

或者，多核苷酸可以与含有可选择标记的载体结合以在宿主中增殖。可以通过前述方法将载体构建体导入宿主细胞中。通常地，质粒载体在沉淀物如磷酸钙沉淀，或在与荷电脂质的复合物中作为 DNA 导入。也可以使用电穿孔方法将多核苷酸导入宿主。如果载体是病毒，可以在体外将其包装或导入包装细胞中，包装的病毒可以转导入细胞中。本领域技术人员熟知许多适于产生多核苷酸和将多核苷酸导入细胞的方法。这种方法参见上述 Sambrook 等所述，这只是详细阐述这些方法的许多实验手册中的一个实例。根据本发明的这个方面，载体例如可以是质粒载体，单链或双链的噬菌体载体，单链或双链的 RNA 或 DNA 病毒载体。通过熟知的将 DNA 或 RNA 导入细胞中的方法，可以将这种载体作为多核苷酸优选 DNA 导入细胞中。在噬菌体和病毒载体的情况下，通过熟知的感染和转导的方法也可以并优选将载体作为包装的或衣壳化病毒导入细胞中。病毒载体可以有复制能力或复制缺陷的。在后一情况中，病毒增殖通常只在互补的宿主细胞中发生。

优选的载体是那些表达本发明多核苷酸和多肽的载体。这种载体通常包含可操纵地连于被表达的多核苷酸的顺式作用控制区，以在宿主中有效地表达。适当的反式作用因子是由宿主，互补载体提供的，或由导入宿主的载体自身提供的。

在这方面的一些优选的实施方案中，载体提供特异性表达。这种特异性表达可以是可诱导表达或只在特定类型的细胞中表达，或者是可诱导和细胞特异性的。特别优选的可诱导载体是可以诱导的以通过环境因子表达的载体，

所述环境因子是易于操纵的如温度和营养添加剂。本领域技术人员熟知适于本发明这方面的一些载体，包括用于原核细胞和真核细胞宿主的组成型和可诱导表达载体，并可以常规应用。

遗传工程化的宿主细胞可以在常规营养培养基中培养，所述培养基可以适当修饰，例如激活启动子，选择转化体或扩增基因。本领域技术人员意识到预先选择的用于宿主细胞以进行表达的培养条件如温度，pH 等，应适于表达本发明的多肽。

可以使用许多表达载体表达本发明的多肽。这种载体包括染色体型，附加型和病毒衍生的载体，例如衍生自细菌质粒，噬菌体，酵母染色体因子，病毒如杆状病毒，乳多空病毒，如 SV40，痘苗病毒，腺病毒，牛痘病毒，pseudorabies 病毒，逆转录病毒和甲病毒属如 Sindbis 病毒，和衍生自上述组合的载体，如那些衍生自质粒和噬菌体遗传因子如粘粒和噬菌粒的载体，所有这些载体根据本发明均可用于进行表达。通常地，任何适于保持，增殖或表达多核苷酸以在宿主中表达多肽的载体均可以用于表达。

通过任何熟知和常规的方法，均可以将适当的 DNA 序列掺入载体中。通常将表达的 DNA 序列与表达载体连接是通过用一或多种限制性内切酶切割 DNA 序列和表达载体，然后用 T4 DNA 连接酶将限制性片段连接而进行的。本领域技术人员熟知并可以常规使用进行限制和连接的方法。在这方面和使用另一种方法构建表达载体的适当方法也为本领域技术人员所熟知，并详见于 Sambrook 等所述。

表达载体中的 DNA 序列可操纵地连于适当的表达控制序列，包括例如指导 mRNA 转录的启动子，这种启动子代

表性包括噬菌体  $\lambda$  PL 启动子，大肠杆菌 lac, trp, tac 和 trc 启动子，SV40 早期和晚期启动子及逆转录病毒 LTRs 等，这里只列举了熟知的启动子的一少部分。应意识到适用于本发明这个方面的许多未提及的启动子也是本领域技术人员熟知的，并可以通过所列举的方式应用。

通常地，表达构建体含有转录起始和终止位点，并在转录区域含有进行翻译的核糖体结合位点。由构建体表达转录物的编码部分，在起始处包括一个翻译起始 AUG，终止密码子几乎位于翻译的多肽的末端。

另外，构建体可以含有调节以及增强表达的控制区域。通常地，根据许多目前通用的方法，这种区域可以通过控制转录操纵，如阻抑物结合位点和增强子等。

进行增殖和表达的载体一般包括可选择标记。这种标记也可以是适于扩增的，或者载体可以含有另外的标记。在这方面，表达载体优选含有一或多个可选择标记基因，以提供表型性状选择转化的宿主细胞。优选的标记对真核细胞培养物而言，包括二氢叶酸还原酶，新霉素，嘌呤霉素或潮霉素抗性基因，对大肠杆菌和其它细菌培养物而言，包括四环素，theomycin，卡那霉素或氨苄青霉素抗性基因。

含有适当所述 DNA 序列以及适当启动子，及其它适当控制序列的载体，可以使用熟知的适于表达所需多肽的方法导入适当的宿主中。适当的宿主代表性例如包括细菌细胞，如大肠杆菌细胞，链霉菌细胞和鼠伤寒杆菌细胞；真菌细胞如酵母细胞；昆虫细胞如果蝇 S2 和草地夜蛾 Sf9 细胞；动物细胞如 CHO, COS 和 Bowes 黑素瘤细胞；和植物细胞优选昆虫细胞 BT1-TN-5B1-4。本领域技术人员熟知大量表达构建体的宿主，并可以根据本发明的揭示选择表达

本发明多肽的宿主。

可以使用各种哺乳动物细胞培养系统进行表达。哺乳动物表达系统例如包括猴肾成纤维细胞 COS-7 系 (Gluzman 等, 细胞 23:175, 1991)。其它能表达可相容载体的细胞系例如包括 C127, 3T3, CHO, HeLa, 人肾 293 和 BHK 细胞系。在哺乳动物宿主细胞中, 可以使用许多基于病毒的表达系统。在将腺病毒用作表达载体的情况中, 编码 RG1 的多核苷酸序列可以连接入腺病毒转录/翻译复合物中, 所述复合物由晚期启动子和三联前导序列组成。在病毒基因组的一个非必需 E1 或 E3 区域中插入, 产生一个能在感染的宿主细胞中表达 RG1 的活性病毒 (Logan 和 Shenk, 美国科学院院报 81:3655-59, 1984)。另外, 转录增强子如 Rous 肉瘤病毒 (RSV) 增强子, 可以用于提高在哺乳动物宿主细胞中的表达。

更特别地, 本发明还包括重组的构建体, 如表达构建体, 其包含上述一或多个序列。此构建体包含一个载体, 如质粒或病毒载体, 其中已经插入本发明的序列。序列可以正向或反向插入。在这方面的一些优选的实施方案中, 构建体还包含调节序列, 例如包括可操纵地与序列连接的启动子。本领域技术人员已知大量适当的载体和启动子, 而且有许多可商购的载体适用于本发明。

以下举例提及了一些可商购的载体。优选在细菌中使用的载体是购自美国 Qiagen 公司的 pQE70, pQE60 和 pQE-9; 购自 Stratagene 公司 (LaJolla, CA) 的 pBS 载体, Phagescript® 载体, Bluescript® 载体, pNH8A, Pnh16a, pNH18A, pNH46A; 和购自 Pharmacia Biotech 公司 (Piscataway, N.J.) 的 ptrca, pK223-3, pKK233-3, Pdr540,

pRIT5。最优选的是购自 Invitrogen 的 pTrcHisB 载体。优选的真核细胞载体是购自 Stratagene 的 pWLNEO, pSV2CAT, Pog44, PXTI 和 pSG; 和购自 Pharmacia Biotech PSVK3, pBPV, pMSG 和 pSVL。最优选的是购自 Promega 的 pClneo 载体。这些载体只是例举了许多可商购载体的一少部分, 而且是本领域技术人员熟知的可用于本发明的载体。应意识到根据本发明可以使用适于例如在宿主中导入, 维持, 增殖或表达本发明多核苷酸或多肽的任何其它质粒或载体。

启动子区域可以使用载体选自任何所需的基因, 所述载体含有一个缺失启动子区域的报道转录单位, 如氯霉素乙酰转移酶 (cat) 转录单位, 下游的导入候选启动子片段的限制位点; 即可包含一个启动子的片段。正如所熟知的, 在 cat 基因上游的限制位点将含有启动子的片段导入载体中可以激活 CAT, 这可以通过标准的 CAT 分析检测。两个这种载体是 pKK232-B 和 pCM7。因此, 表达本发明多核苷酸的启动子不仅包括熟知的和易于获得的启动子, 还包括通过上述方法使用报道基因易于获得的那些启动子。

已知的适于表达本发明多核苷酸和多肽的细菌启动子是大肠杆菌 lacI 和 lacZ 启动子, T3 和 T7 启动子, T5 tac 启动子, λPR, PL 启动子, trp 启动子和衍生自 trp 和 lac 启动子的 trc 杂合启动子。已知的适于这方面的真核启动子是 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, 逆转录病毒 LTRs 启动子, 如 Rous 肉瘤病毒 (RSV) 启动子和金属硫蛋白启动子, 如小鼠金属硫蛋白-1 启动子。

在宿主细胞中表达的适当载体和启动子的选择是本领

域技术人员所熟知的，而且表达载体的构建，将载体导入宿主中并在宿主中表达的必需方法也是本领域技术人员所常规使用的。

通常地，重组载体包括复制起点，一个衍生自高度表达的基因的启动子，以指导下游结构序列转录，和一个可选择的标记，以在暴露于载体后分离含有载体的细胞。

本发明还涉及含有上述构建体的宿主细胞。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，或者宿主细胞可以是原核细胞，如细菌细胞。宿主细胞中的构建体可以以常规方式使用，以产生由重组序列编码的基因产物。

多肽可以在适当启动子控制下，在哺乳动物细胞，酵母，细菌或其它细胞中表达。没有细胞的翻译系统也可用于产生这种蛋白质，通过使用衍生自本发明的 DNA 构建体的 RNAs 进行。用于原核和真核细胞宿主的适当克隆和表达载体如 Sambrook 等所述。

通过高等真核细胞转录编码本发明多肽的 DNA，可以通过将一个增强子序列插入载体中而增强。增强子是 DNA 的顺式作用因子，通常大约 10—300 个 bp，在给定类型的宿主细胞中提高启动子的转录活性。增强子例如包括 SV40 启动子，其位于在 100—270bp 的复制起点的后侧，巨细胞病毒早期启动子增强子，位于复制起点后侧的多形瘤增强子，和腺病毒增强子。

使用标准方法，将编码本发明的异源结构序列的本发明的多核苷酸插入载体中，使其可操纵地连于启动子以进行表达。多核苷酸是定位的，以使转录起始位点接近位于 5' 端至核糖体结合位点。核糖体结合位点在 5' 端至 AUG 之

间,启动表达的多肽翻译。通常地,在起始密码子通常为 AUG 起始处,及核糖体结合位点与起始 AUG 之间没有其它的开放读框。同样,在多肽的末端有一个翻译终止密码子,而且有一个聚腺苷酸化信号,和接近在转录区 3' 末端的一个转录终止信号。

为将翻译的蛋白质分泌入内质网腔,壁膜间隙或胞外环境中,可以将适当的分泌信号掺入表达的多肽中。这个信号可以是多肽内源性的,或者可以是异源信号。多肽可以以修饰的形式表达,如以融合蛋白的形式,而且不仅可以包括分泌信号,也可以包括其它的异源功能区。因此,例可以在多肽的 N 末端加入如一个另外氨基酸区域尤其荷电的氨基酸,以在纯化期间或在随后的运用和贮存期间改良在宿主细胞中的稳定性和持续性。同样,也可以在多肽中加入特殊的区域以易于纯化。这种区域在最后制备多肽之前可以除去。在多肽中加入肽部分以产生分泌或排泄,改良稳定性或易于纯化等方法是本领域技术人员熟知并常规使用的。例如,当需要将大量的 RG1 导入抗体中时,需要直接高水平表达易于纯化的融合蛋白的载体。这种载体包括但非限于多功能的大肠杆菌克隆和表达载体,如 Bluescript® (Stratagene), 其中 rg1 编码序列可以在框架中以氨基端 Met 和随后的 7 个 $\beta$ -半乳糖苷酶残基的序列连接于载体中,以便产生杂合蛋白; pIN 载体 (Van Heede 和 Shuster, 生物化学杂志 264:5503-5509, 1989) 等。PtrcHis 载体 (Invitrogen, Carlsbad, CA) 可以用于表达外源多肽作为含有多组氨酸 (6 $\times$ His) 标记的融合蛋白以迅速纯化。在这种系统中产生的蛋白质设计为包括切割位点如肠激酶裂解位点,以便相应的克隆的多肽可以从融合肽组分中随

意释放。

在转化适当的宿主菌株及宿主菌株生长至适当的细胞密度后，可诱导启动子如果存在则可以通过适当的方法（例如温度变换或暴露于化学诱导剂）诱导，并将细胞另外培养一段时间。

然后通过离心收获细胞，通过物理或化学方法破坏，将所得粗提物保留以进一步纯化。

用于表达蛋白质的微生物细胞，可以通过任何常规方法破坏，包括冷冻-融化循环，超声，机械破坏，或使用细胞裂解剂，这些方法是本领域技术人员所熟知的。

RG1 多肽可以通过熟知的方法从重组细胞培养物中回收和纯化，这些方法包括硫酸铵或乙醇沉淀，酸性提取，阴离子或阳离子交换层析，磷酸纤维素层析，疏水作用层析，亲和层析，羟磷灰石层析和凝集素层析。优选使用高效液相层析（HPLC）进行纯化。当多肽在分离和或纯化期间变性时，可以应用熟知的重折叠蛋白质的方法以再生活性的构象。本领域熟知的各种纯化蛋白质的其它方法包括那些 Deutscher, M., 酶学方法, 第 182 卷, 科学出版社, San Diego, 1982; 和 Scopes, R., 蛋白质纯化原理和实践, Springer-Verlag, 纽约, 1982。

或者，本发明的多肽可以使用固相方法通过直接合成肽而产生（Stewart 等，固相肽合成，W.H.Freeman Co., San Francisco, 1969; Merrifield, J., *J.Am.Chem.Soc.* 85:2149-2154, 1963）。可以使用人工或自动方法进行体外蛋白质合成。自动合成例如可以通过使用 Applied Biosystems 431A 肽合成仪（Perkin Elmer, Foster City, Calif），根据生产者的指导进行。RG1 的各种片段可以是单独或组合使用化学方法化

学合成的，以产生全长的分子。

本发明的多肽包括自然纯化的产物，化学合成的产物，及通过重组方法从真核或原核宿主中产生的产物，所述宿主例如包括细菌，酵母，高等植物，昆虫和哺乳动物细胞。依赖于重组产生方法中的宿主，本发明的多肽可以是糖基化的或可以是非糖基化的。另外，本发明的多肽也可以包括一个初始形式的甲硫氨酸残基，在一些情况中是宿主介导的结果。

### RG1 多肽及编码其的多核苷酸的使用

Rg1 多核苷酸和 RG1 多肽根据本发明可以有許多用途，尤其使用 RG1 的化学和生物学性质的那些用途。另外可以用于诊断和治疗细胞增殖性疾病，如前列腺癌。本发明的这些方面通过以下揭示得以进一步例证，并在说明书的主要部分内进一步阐述。

使用本发明多核苷酸和多肽序列的基本原理，部分基于 RG1 和其它胞外基质分子之间的化学和结构同源性，及基于 RG1 在前列腺组织中比在其它组织中优先表达。RG1 可以用于诊断和治疗与前列腺组织异常生长相关的病理状态，功能失调或疾病。这些包括但非限于癌症和转移性肿瘤生长。

Rg1 序列可以用作 DNA 探针，及用作反义和核酶疗法的靶位，或作为产生反义多核苷酸的模板。

RG1 多肽可以用于产生 RG1 的抗体，以用于检测细胞和组织中多肽的水平，及将药物定向于原发和转移瘤。

RG1 多肽可用于刺激对含有 RG1 细胞的免疫应答。

编码 RG1 的多核苷酸可用于检测细胞和组织中编码 RG1 的多核苷酸水平的诊断分析中。

在与 RG1 表达相关的病变中，如前列腺癌，抑制 RG1 的表达和活性是有益的。RG1 表达可以通过施用反义寡核苷酸或核酶而抑制。或者，可以施用特异识别 RG1 多肽活性相关区域的抗体，以治疗与 RG1 活性相关的疾病或病理状态。

### 多核苷酸分析

本发明还涉及使用 rg1-相关的多核苷酸以检测互补多核苷酸，例如作为诊断试剂。与疾病相关的 rg1 多核苷酸的检测，提供了一种体外或体内诊断工具，可以补充或限于诊断由于 RG1 组织特异性表达所致的疾病或对这种疾病的易感性。

编码 RG1 的基因中有突变的个体，可以在 DNA 水平通过各种方法检测。进行诊断的多核苷酸样品可以得自病人的细胞，如取自血液，尿液，唾液，组织活检和尸检物。基因组 DNA 可以直接用于检测，或者在分析之前通过使用 PCR 进行酶促扩增 (Saiki 等, 自然 324:163-166, 1986)。RNA 或 cDNA 也可以以相同方式使用。例如，互补于编码 RG1 的多核苷酸序列的 PCR 引物，可以用于鉴别和分析 rg1 的表达和突变。例如，缺失和插入可以通过扩增产物的大小与正常基因型不同而检测。点突变可以通过将扩增的 DNA 与放射标记的 rg1 RNA 或者放射标记的 rg1 反义 DNA 序列杂交而鉴别。通过 RNase A 消化或通过解链温度不同，可以将完全匹配的序列从错配的双链体中区分出。

参照基因和具有突变的基因之间的序列差异还可以通过直接 DNA 测序证明。另外，克隆的 DNA 节段可以用作探针，以检测特异性 DNA 节段。这种方法的敏感性可以通过适当使用 PCR 或其它的扩增方法而明显增强。例如，测序

引物与双链的 PCR 产物或通过修改的 PCR 产生的单链的模板分子一起使用。序列测定是使用放射标记的多核苷酸通过常规方法，或使用荧光标记通过自动测序方法进行的。

基于 DNA 序列差异的遗传学试验，可以通过检测有或无变性剂的凝胶中 DNA 片段的电泳迁移率的变化进行。通过高解离凝胶电泳可以观测到少量序列缺失和插入。不同序列的 DNA 片段可以在变性甲酰胺梯度凝胶上区分，其中不同 DNA 片段的迁移率根据其特异性解链或部分解链温度，在凝胶中在不同位置减速（见例如 Myers 等，科学 230:1242, 1985）。

在特异位置的序列变化也可以通过核酸酶保护分析证实，如 RNase 和 S1 保护或者通过化学裂解方法证实（例如 Catton 等，美国科学院院报 85:4397-4401, 1985）。

因此，检测特异性 DNA 序列可以通过以下方法进行，如杂交，RNase 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态性（RFLP）和基因组 DNA 的 Southern 印迹分析）。

除了更常规的凝胶电泳和 DNA 测序之外，突变也可以通过原位分析进行检测。

### 多肽分析

本发明还涉及诊断分析如定量分析和检测细胞和组织 and 体液中 RG1 多肽水平的诊断分析，包括测定正常和异常水平。因此例如根据本发明的检测与在正常对照组织样品中相比 RG1 多肽过表达的诊断分析，可以用于检测瘤形成，例如前列腺癌。这种诊断试验可用于检测转移瘤生长。可用于测定来自宿主的样品中多肽如本发明的 RG1 多肽的水平的分析方法，是本领域技术人员所熟知的。这种分析方

法包括放射性免疫分析（RIA），竞争性结合分析，Western 印迹分析和酶联免疫吸附测定（ELISA），及荧光激活的细胞分选（FACS）。这些方法当中优选 ELISA。ELISA 分析开始包括制备一种 RG1 的特异性抗体，优选单克隆抗体。另外，通常要制备一种结合单克隆抗体的报道抗体。将报道抗体粘附于一种可检测试剂，如放射性，荧光或酶促试剂，在这个实施例中是辣根过氧化物酶。

为进行 ELISA，从宿主中取样品，并在结合样品中多肽的固体支持物上温育，如在聚苯乙烯培养皿上。然后将培养皿上的任何游离多肽结合位点，通过用非特异性蛋白质如牛血清白蛋白温育而覆盖。接着，将单克隆抗体在培养皿中，在单克隆抗体与附着于聚苯乙烯培养皿的任何 RG1 多肽附着的期间温育。用合成仪将未结合的单克隆抗体洗去。将连接于辣根过氧化物酶的报道抗体置于培养皿中，导致报道抗体与结合 RG1 的任何单克隆抗体结合。然后洗去未附着的报道抗体。然后在培养皿中加入过氧化物酶活性试剂，包括比色底物。通过原始和二级抗体连接于 RG1 的固定的过氧化物酶，产生有色的反应产物。在给定的时间内产生的颜色数量，表明样品中存在的 RG1 多肽数量。定量结果典型地是通过参考标准曲线获得的。

可以使用竞争性分析，其中将特异于 RG1 的抗体附着于固相支持物，并将标记的 RG1 和取自宿主的样品穿经此固相支持物，检测到的附着于固相支持物的标记数量与样品中 RG1 的数量相关。

这些及其它方法见于 Hampton 等（血清学方法实验手册，APS 出版社，St Paul, Minn, 1990）和 Maddox 等（实验方法杂志 158:12111, 1983）所述。

## 抗体

本发明还涉及特异结合 RG1 的抗体，在此称为 RG1 抗体。RG1 在前列腺组织中的过表达及其细胞表面位置，代表了进行筛选，诊断，预后，追踪分析和显影方法的理想标记的特性。另外，这些特性表明 RG1 可以是一些疗法的理想靶位，如定向抗体疗法，免疫疗法和基因疗法。如本文所用术语“特异结合”是指抗体与多肽之间的相互作用，其中这种相互作用依赖于多肽上存在特殊的结构（即抗原性决定簇或表位）；换言之，抗体识别并结合特异性多肽结构而不是一般的蛋白质。

RG1 多肽，其片段或其它衍生物，或其类似物，或表达它们的细胞，可以用作免疫原以产生抗体（Harlow，抗体，冷泉港出版社，NY（1989））。这些抗体例如可以是多克隆或单克隆抗体。本发明还包括嵌合的，单链的，人化的和人抗体，以及 Fab 片段，或 Fab 表达文库的产物。可以使用本领域已知的各种方法产生这种抗体和片段。

抗相当于本发明序列的多肽产生的抗体，可以通过将多肽直接注射入动物体内，或通过将多肽施用于动物优选非人动物而获得。这样获得的抗体然后结合其自身多肽。以这种方式，甚至只编码多肽片段的序列也可用于产生结合整个天然多肽的抗体。这种抗体然后可用于从表达那种多肽的组织中分离多肽。

为制备单克隆抗体，可以使用任何通过连续细胞系培养产生抗体的方法。例如包括杂交瘤技术（Kohler 和 Milstein，自然 256:495-497，1975），人 B 细胞杂交瘤技术（Kozbor 等，现代免疫学 4:72，1983）和产生人单克隆抗体的 EBV-杂交瘤技术（Cole 等，单克隆抗体和癌症，Alan R.Liss, Inc.,

77-96, 1985)。

另外, 可以使用产生“嵌合”抗体的技术, 将小鼠抗体基因与人抗体基因相剪接, 获得具有适当抗原特异性和生物学活性的分子 (Morrison 等, 美国科学院院报 81:6851-6855, 1984; Neuberger 等, 自然 312:604-608, 1984; Takeda 等, 自然 314:452-454, 1985)。或者, 生产单链抗体的技术 (美国专利 No.4946778) 可以适于生产 RG1 特异性单链抗体。

另外, “人”抗体可以使用美国专利 No.5877397 和 5569825 所述的方法生产, 所述专利在此并入参考。

抗体还可以通过在淋巴细胞群中体内诱导产生, 或通过筛选重组的免疫球蛋白文库或高特异性结合剂而产生, 如 Orlandi 等 (美国科学院院报 86:3833-3837, 1989) 和 Winter 和 Milstein (自然 349:293-299, 1991) 所述。

也可以产生含有 RG1 特异性结合位点的抗体片段。例如, 这种片段包括但非限于  $F(ab')_2$  片段, 其可以通过胃蛋白酶消化抗体分子而产生, 和 Fab 片段, 其可以通过还原  $F(ab')_2$  片段的二硫键而产生。或者, 可以构建 Fab 表达文库以迅速和简便地鉴别具有所需特异性的单克隆 Fab 片段 (Huse 等, 科学 256:1270-1281, 1989)。

在此所述 RG1 的氨基酸序列可以用于选择 RG1 多肽特异性区域, 以产生抗体。本领域技术人员应意识到抗体定向的 RG1 多肽的区域和表位可以随着具体应用而变化。例如, 用于免疫分析中以检测前列腺细胞上膜结合的 RG1 的抗体, 应定向于 RG1 多肽上的可及表位。表现免疫原性记过的 RG1 多肽的区域, 以及其它区域和结构域, 可以使用本领域已知的各种其它方法轻易鉴别, 如 Chou-Fasman,

Gamier-Robson, 或 Jameson-Wolf 分析。含有这些残基的片段特别适于产生抗 RG1 抗体。特别有用的片段包括但不限于序列 PLGGESICSAGAPAKYSIT (SEQ ID NO:8); HSSDYSMWRKNQYVS (SEQ ID NO:10); DAGTDSGFTFSSPNFATIPQDTV (SEQ ID NO:11), 和 NEIVDSASVPET (SEQ ID NO:12)。这些区域产生的单克隆抗体见于实施例 4 所述。

本发明的 RG1 抗体可以特别用于前列腺癌的诊断分析, 显影方法学, 和治疗方法中。本发明提供了用于检测 RG1 多肽和诊断前列腺癌的各种免疫学分析。这种分析通常包括一或多个能识别和结合 RG1 多肽的 RG1 抗体。最优选的抗体将选择性结合 RG1 并不结合 (或微弱结合) 非 RG1 多肽。这些分析包括本领域熟知的各种免疫学分析, 包括但不限于各种类型的免疫学分析, 酶联免疫吸附测定等。另外, 本发明还提供了能检测前列腺癌的免疫学显影方法, 包括但分析适用于标记的 RG1 抗体的放射闪烁照相显影方法。这种分析可以临床用于检测, 监测和预后前列腺癌。

上述抗体可以用于分离或鉴别表达多肽的克隆, 或纯化本发明的多肽, 通过将抗体附着于固体支持物以通过亲和层析分离和/或纯化。

另外, 可以使用细胞分选和纯化方法, 将 RG1 抗体用于分离 RG1 阳性细胞。特别地, RG1 抗体可以用于从异种移植的肿瘤组织中, 从培养的细胞中等, 使用基于抗体的细胞分选或嵌合纯化方法, 分离前列腺癌细胞。本发明 RG1 抗体的其它应用包括产生模拟 RG1 多肽的抗独特型抗体。

RG1 多肽可以用于监测前列腺癌或肿瘤转移的存在与否。各种生物学样品中, 包括血清, 前列腺和其它组织活

检样品，是否存在这种含有 RG1 的细胞，可以用 RG1 抗体检测。另外，RG1 抗体可以用于各种显影方法学中，如用 Tc-99m（或其它同位素）缀合的抗体进行免疫闪烁照相。例如，可以使用一种类似于最近阐述的使用 In-111 缀合的抗 PSMA 抗体的显影方法，以检测复发的和转移的前列腺癌（Sodee 等，临床核酸方法 21:756-766，1997）。

本发明的 RG1 抗体可以用可检测的标志标记或缀合于另一种分子，如胞毒剂，及用于将另一种分子定向于 RG1 阳性细胞（Vitetta, E.S.等，免疫毒素疗法，DeVita, Jr., V.T. 等编辑，癌症：肿瘤学原理和实践，第 4 版，J.B.Lippincott Co., Philadelphia, 2624-2636, 1993）。胞毒剂例如包括但不限于蓖麻毒蛋白，阿霉素，道诺红霉素，紫杉醇，溴化乙锭，丝裂霉素，表鬼臼毒吡喃葡萄糖苷，tenoposide，长春花新碱，长春花碱，秋水仙素，二羟炭疽菌素 dione，放线菌素 D，diphtheria 毒素，假单胞杆菌外毒素（PE）A，PE40，相思豆毒蛋白和糖皮质激素及其它化疗剂，以及放射性同位素。适当的可检测标记包括但不限于放射性同位素，荧光化合物，生物发光化合物，化学发光化合物，金属螯合剂或酶。适当的放射性同位素包括以下物质：锶-124，锶-125，砷-74，钡-103，钡-140，铍-7，铋-206，铋-207，镅-109，镅-115m，钙-45，铈-139，铈-141，铯-144，铯-137，铬-51，钴-56，钴-57，钴-58，钴-60，钴-64，铟-169，镱-152，Gadolinium-153，金-195，金-199，Hafnium-175，Hafnium-181，铪-11，碘-123，碘-131，铱-192，铁-55，铁-59，氩-85，铅-210，镁-54，汞-197，汞-203，钨-99，钨-147，镓-237，镍-63，铌-95，钨-185+191，钨-103，钨

—195m, 镨—143, 钼—147, 钼—233, 镭—2226, 铷—186, 铷—86, 钇—103, 钇—106, 铈—44, 铈—46, 硒—75, 银—110m, 银—11, 钠—22, 铈—85, 铈—89, 铈—90, 硫—35, Tantalum—182, 铟—99m, 碲—125, 碲—132, 铊—170, 铊—204, 铋—228, 铋—232, 锡—113, 钛—44, 钨—185, 钒—48, 钒—49, 镱—169, 钇—88, 钇—90, 钇—91, 锌—65 和镉—95。

### 前列腺癌的免疫治疗

本发明提供了多种治疗前列腺癌的免疫治疗方法, 包括抗体疗法, 体内接种, 和源于体内免疫治疗方法。在一个方法中, 本发明提供了 RG1 抗体, 其可以系统施用以治疗前列腺癌。例如, 可以将未缀合的 RG1 抗体导入患者体内, 这样抗体与前列腺癌细胞上, 内或与之缔合的 RG1 结合, 并引起细胞和肿瘤的破坏, 这是通过可包括以下机制引起的: 补体引起的细胞溶解, 抗体依赖性细胞毒性, 改变 RG1 的生理功能, 和/或抑制配体结合或信号转导途径。缀合于毒性剂如蓖麻毒蛋白或放射性同位素的 RG1 抗体, 也可以治疗性用于将毒性剂定向递送于携带 RG1 的前列腺肿瘤细胞, 并从而破坏肿瘤细胞。

施用 RG1 抗体进行的前列腺癌免疫治疗, 可以遵循已经成功用于其它类型癌症的多种方法的教导, 包括但非限于结肠癌 (Arlen 等, *Crit.Rev.Immunol.*18:133-138, 1998), 多发性肌瘤 (Ozaki 等, *血液* 90:3179—3186, 1997; Tsunenari 等, *血液* 90: 2437—2444, 1997), 胃癌 (Kasprzyk 等, *癌症研究* 52: 2771—2776, 1996), 白血病 (Zhong 等, *白血病研究* 20: 581—589, 1996), 结肠直肠癌 (Moun 等, *癌症研究* 54: 6160—6166, 1994; Velders 等, *癌症研究* 55:

4398—4403, 1995) 和乳腺癌 (Shepard 等, 临床免疫学杂志 11: 117—127, 1991)。

本发明还提供了含有 RG1 多肽或其片段的疫苗。本领域熟知在疫苗中使用肿瘤抗原以产生肿瘤或细胞介导的免疫性, 以用于抗肿瘤治疗中, 而且已经使用人 PSMA 和啮齿动物 PAP 免疫原用于前列腺癌中 (Houdge 等, 癌症研究杂志 63:231-237, 1995; Fong 等, 免疫学杂志 159:3113-3117, 1997)。这种方法可通过应用 RG1 多肽, 或其片段, 或编码 RG1 的核酸分子, 和能表达和适当呈递 RG1 免疫原的重组载体而易于实践。

例如, 可以使用病毒基因输送系统以输送编码 RG1 的核酸分子。根据本发明的此方面可以使用的各种病毒基因输送系统包括但非限于牛痘, 禽痘, 金丝雀痘, 腺病毒, 流感病毒, 脊髓灰质炎病毒, 腺伴随病毒, 慢病毒和新培斯病毒 (Restifo, 现代免疫学见解 8:658-663, 1996)。也可以通过将编码 RG1 或其片段的裸 DNA 导入患者 (即通过肌肉注射) 中诱导抗肿瘤应答, 而应用非病毒输送系统。在一个实施方案中, 可以应用全长的人 *rg1* cDNA。在另一个实施方案中, 可以应用人 *rg1* cDNA 片段。在另一个实施方案中, 可以应用编码特异性 T 淋巴细胞 (CTL) 表位的 *rg1* 核酸分子。CTL 表位可以使用特异程序确定 (例如 Epimer, Brown 大学), 以鉴别 RG1 多肽中能与特异的 HLA 等位基因最佳结合的肽。

也可以使用各种源于体内的方案。一种方法包括使用树突细胞, 将 RG1 多肽作为抗原呈递给患者免疫系统。树突细胞表达 MHC 的 I 和 II 型分子, B7 共同刺激因子, 和 IL-12, 并因此特别专用于抗原呈递细胞。在前列腺癌中, 将

用前列腺特异性膜抗原 (PSMA) 肽脉冲的自身树突细胞用于 I 期临床试验中, 以刺激前列腺癌患者的免疫系统 (Tjoa 等, 前列腺 28:65-69, 1996; Murphy 等, 前列腺 29:371-380, 1996)。可以使用树突细胞将 RG1 多肽呈递给 MHC 的 I 和 II 型分子邻近的 T 细胞。在一个实施方案中, 自身树突细胞是用能结合 MHC 分子的 RG1 多肽脉冲的。在另一个实施方案中, 树突细胞是用完整的 RG1 多肽脉冲的。然而另一个实施方案包括使用本领域已知的各种载体, 对树突细胞中 rg1 基因的过表达加以遗传工程化, 所述载体例如是腺病毒 (Arthur 等, 癌症基因治疗 4:17-25, 1997), 逆转录病毒 (Henderson 等, 癌症研究 56:3763-3770, 1996), 慢病毒, 腺伴随病毒, DNA 转染 (Ribas 等, 癌症研究 57:2865-2869, 1997), 和肿瘤衍生的 RNA 转染 (Ashley 等, *J.Exp.Med.*186:1177-1182, 1997)。

抗独特型抗 RG1 抗体也可用于抗癌治疗中, 作为疫苗诱导对表达 RG1 多肽的细胞的免疫应答。特别是本领域熟知产生抗独特型抗体的方法, 并可方便地适于产生抗独特型抗 RG1 抗体, 其模拟 RG1 多肽上的一个表位 (见例如 Wagner 等, 杂交瘤 16:33-40, 1997; Foon 等, 临床研究杂志 96:334-342, 1995; Herlyn 等, 癌症免疫学免疫治疗 43:65-76, 1996)。这种抗独特型抗体可以用于抗独特型治疗中, 与现有的直接抗肿瘤抗原的其它抗独特型抗体一样使用。

遗传免疫接种方法可以用于产生对表达 RG1 的癌细胞的预防性或治疗性激素和细胞免疫应答。使用本文所述的编码 RG1 的 DNA 分子, 可以将包含编码 RG1 多肽/免疫原的 DNA 和适当调节序列的构建体, 直接注射至个体的肌肉

或皮肤中，这样肌或皮肤的细胞吸收构建体并表达编码的 RG1 多肽/免疫原。RG1 多肽/免疫原可以表达为细胞表面多肽或是分泌的。RG1 多肽/免疫原的表达地址抗前列腺癌的预防性或治疗性激素和细胞免疫应答产生。可以使用本领域已知的各种预防性和治疗性遗传免疫接种方法（参见 [www.genweb.com](http://www.genweb.com) 网站上公布的信息和参考资料）。

### 反义寡核苷酸，反义载体和核酶

互补于 rg1 的反义多核苷酸可以是合成制备的。这种寡核苷酸可以用或不用脂质体输送至细胞中，脂质体可以有助于反义寡核苷酸摄取入细胞中。

或者，可以使用衍生自逆转录病毒，腺病毒，疱疹病毒或痘苗病毒，或各种细菌质粒的表达载体，构建和输送表达反义 rg1 的重组载体。见例如 Sambrook（如前）和 Ausubel（如前）所述的方法。

包含全长 cDNA 序列和/或其调节因子的多核苷酸，使研究人员可以使用 rg1 多核苷酸在有义链（Yousoufian 和 Lodish, 分子细胞生物学 13:98-104, 1993）或反义链（Eguchi 等, *Annu.Rev.Biochem.*60:613-652, 1991）中作为研究工具以调节基因功能。目前本领域熟知这种技术，而且有义或反义寡聚物，或较大片段可以从沿着编码或控制区的各种位置中设计。

编码 RG1 的基因可以通过用表达载体转染细胞或组织而关闭，所述表达载体高水平表达所需的 rg1 多核苷酸片段。这种构建体用不可翻译的有义或反义链充填细胞。即使在没有整合入 DNA 的情况下，这种载体也可以持续转录 RNA 分子，直至所有拷贝均由内源核酸酶破坏。瞬时表达用一个非复制载体可以持续一个月或更长时间，而且如果

适当的复制因子是载体系统的一部分可以持续更长时间。

如上所述，对基因表达的修饰可以通过设计反义分子，DNA 或 RNA 控制 rg1 的区域，即启动子，增强子和内含子。优选衍生自转录起始位点的寡核苷酸，例如在引导序列-10 和+10 区域之间的寡核苷酸。反义分子也可以设计为通过阻止转录物结合核糖体而阻断 mRNA 的翻译。相似地，抑制可以通过使用“三螺旋”碱基配对方法达到。三螺旋配对包含双螺旋充分打开以结合聚合酶，转录因子或调节分子的努力。新近的使用 3 螺旋 DNA 的治疗进展参见 Gee, J.E. 等揭示 (Huber 和 Car, 分子和免疫学方法, Futura 出版公司, Mt.Kisco, N.Y., 1994)。

核酶是能催化 RNA 特异性裂解的酶性 RNA 分子 (美国专利 No.4987071; WO93/23057)。核酶的作用机制包括核酶分子与互补的靶 RNA 序列特异性杂交，随后内切核酸酶性裂解。在本发明范围内的是遗传工程化的锤头基序核酶分子，其能特异及有效地催化编码 RG1 的 RNA 内切核酸酶性裂解。在任何潜在的 RNA 靶内的特异性核酶裂解位点，最初是通过扫描靶分子的包括以下序列 GUA, GUU 和 GUC 的核酶裂解位点而鉴别的。一旦鉴别，可以对相当于含有靶基因裂解位点的区域的在 15—20 个核糖核苷酸之间的短 RNA 序列，评价其使寡核苷酸不可操纵的二级结构特点。候选靶分子是否适当也可以使用核糖核酸酶保护分析，通过测试与互补寡核苷酸杂交的能力而加以评价 (Irie 等，药物学进展 40:207-257, 1997)。

本发明的反义分子和核酶可以通过本领域已知的任何合成 RNA 分子的方法制备。这些包括化学合成寡核苷酸方法，如固相亚磷酰胺化学合成方法。或者，RNA 分子可以通过

体外和体内转录产生，或通过编码 RG1 的 DNA 序列产生。这种 DNA 序列可以掺入各种具有适当 RNA 聚合酶启动子如 T7 或 SP6 的载体中。或者，可以将组成型或可诱导地合成反义 RNA 的反义 cDNA 构建体导入细胞系，细胞或组织中。

可以对 RNA 分子加以修饰以提高胞内稳定性和半衰期。可能的修饰包括但非限于在该分子的 5' 和/或 3' 末端添加侧翼序列，或在分子的骨架内使用硫代磷酸酯或 2' O-甲基而不是磷酸二酯酶键合。通过包含非常规的碱基如肌苷和 queosine 以及乙酰基-，甲基-，硫代-及相似修饰形式的腺嘌呤，胞苷，鸟嘌呤，胸腺嘧啶，和尿嘧啶，也可以提高稳定性，这些碱基一般不易被内源性内切核酸酶识别。

将反义载体导入细胞或组织的方法包括前述的那些方法，这些方法均等地适用于体内，体外和源于体内的治疗中。就源于体内的治疗而言，是将反义载体导入取自患者的细胞中并克隆繁殖，以自体移植入同一患者中，如在此并入参考的美国专利 No.5399493 和 5437994 所述。本领域熟知通过转染和通过基于脂质体或其它脂质或基于非脂质的制剂进行输送的方法。

#### 鉴别结合 RG1 的制剂的试验

本发明还涉及可以用于鉴别结合 RG1 的制剂的试验和方法。特别地，鉴别结合 RG1 的制剂，可以通过 RG1 配体或其它制剂或组分结合 RG1 的能力和/或抑制/刺激 RG1 活性的能力进行。

或者，结合 RG1 多肽的制剂可以使用酵母双杂交系统或结合捕获试验加以鉴别。在酵母双杂交系统中，将一个

编码融合蛋白的表达单位导入酵母细胞中并表达，所述融合蛋白是由两个亚单位转录因子的一个亚单位和 RG1 多肽组成。将此细胞进一步修饰以含有（1）一个编码可检测标记的表达单位，所述标记的表达需要两个亚单位转录因子，和（2）一个编码融合蛋白的表达单位，所述融合蛋白由转录因子的另一个亚单位和克隆的 DNA 节段组成。如果克隆的 DNA 节段编码一种结合 RG1 多肽的蛋白质，则表达导致 RG1 和编码的蛋白之间的相互作用。这使转录因子的两个亚单位紧密结合，重组为转录因子。这使可检测标记表达。酵母双杂交系统特别用于筛选 cDNA 文库，所述 cDNA 编码 RG1 的细胞结合配体的节段。

可用于上述试验的 RG1 多肽包括但非限于一种分离的 RG1 多肽，RG1 多肽的片段，一个已经改变而表达 RG1 多肽的细胞，或已经改变而表达 RG1 多肽的细胞的组分。另外，RG1 多肽可以是完整的多肽或是 RG1 多肽的指定片段。本领域技术人员将意识到只要可以测定 RG1 多肽的结合剂，例如通过改变分子量或活性进行测定，就可以使用本试验。

用于鉴别一种制剂/细胞区室是否结合 RG1 多肽的方法，主要基于使用的 RG1 多肽的性质。例如，可使用凝胶阻滞试验测定一种制剂是否结合 RG1 或其片段。或者，可采用免疫检测和生物芯片技术与 RG1 一起使用。技术人员可方便地应用本领域熟知方法测定一种特殊制剂是否与 RG1 多肽结合。

可以对制剂和细胞区室进一步测试其调节 RG1 活性的能力，使用无细胞测定系统或细胞测定系统。由于 RG1 多肽的活性是更加限定的，可以应用基于鉴别活性的功能测

定。

如本文所用，当制剂降低 RG1 活性时，称之为拮抗 RG1 活性。优选的拮抗剂将选择性拮抗 RG1，而不影响任何其它细胞蛋白。另外，优选的拮抗剂将 RG1 活性降低 50%以上，更优选降低 90%以上，最优选消除所用 RG1 活性。

在上述方法中测定的制剂可以上随机选择的或理性选择或设计的。如本文所用，当制剂是随机的而不考虑 RG1 多肽的特异性序列而选择的时候，将此制剂称为随机选择的。随机选择的制剂例如是使用化学文库或肽组合文库，或有机物或植物提取物的生长肉汤。

如本文所用，当制剂是在考虑靶位点的序列和/或其与制剂功能相关的构象的非随机基础上选择的，则称此制剂为理性选择或设计的。制剂可以利用由组成 RG1 多肽的肽序列加以理性选择或理性设计。例如，一个理性选择的肽制剂可以是这样的一种肽，其氨基酸序列等同于 RG1 多肽的片段。

本发明方法中测试的制剂可以是例如肽，抗体，寡核苷酸，小分子和维生素衍生物，以及碳水化合物。技术人员可易于意识到对所述筛选方法中使用的制剂的结构性质没有限制。本发明的一类的制剂是肽制剂，其氨基酸序列是基于 RG1 多肽的氨基酸序列选择的。

肽制剂可以使用标准固相（或溶液相）肽合成方法制备的，如本领域已知的方法。另外，编码这些肽的 DNA 可以是使用可商购的寡核苷酸合成仪器合成的，及使用标准的重组生产系统重组生产的。如果包括无基因编码的氨基酸，使用固相肽合成生产是必要的。

本发明的另一类制剂是与 RG1 的关键位置免疫反应的

抗体。如上所述，抗体是通过用肽免疫接种适当的哺乳动物对象而获得的，所述肽含有作为抗原性区域的抗体定向的 RG1 多肽的那些部分。这种制剂可以用于竞争性结合研究，以鉴别第二代抑制剂以及封阻 RG1 活性。

本发明方法中细胞提取物试验可以是例如细胞或组织的水相提取物，细胞或组织或部分纯化的细胞组分的有机提取物。技术人员可易于意识到对本发明的筛选方法中使用的细胞提取物的来源没有限制。

结合 RG1 多肽的制剂，如 RG1 抗体，可以用于调节 RG1 的活性，将抗癌剂定向于适当的哺乳动物细胞，或鉴别封阻与 RG1 相互作用的制剂。表达 RG1 的细胞可以使用结合 RG1 的制剂定向或鉴别。

RG1 结合剂怎样使用依赖于 RG1 结合剂的性质。例如，RG1 结合剂可以用于：将缀合的毒素，如白喉毒素，霍乱毒素，蓖麻毒蛋白或假单胞杆菌外毒素，输送至 RG1 表达细胞；调节 RG1 活性；直接杀死 RG1 表达细胞；或筛选鉴别竞争性结合剂。例如，RG1 抑制剂可以用于直接已知 RG1 表达细胞的生长，而 RG1 结合剂可以用作诊断剂。

### 药物组合物及施用

本发明还涉及药物组合物，其可以包含 rg1 多核苷酸，RG1 多肽，抗体，兴奋剂，拮抗剂，或抑制剂，单独或与至少一种其它制剂如稳定化合物组合，其可以在任何无菌的，生物相容的药物载体中施用，所述载体包括但非限于盐水，缓冲盐水，葡萄糖和水。任何这些分子均可以单独施用于患者，或其它制剂，药物或激素组合在药物组合物中施用，其中其与赋形剂或药物适当载体混合。在本发明的一个实施方案中，药物适当的载体是药物学惰性的。

本发明还涉及药物组合物的施用。这种施用是经口服或非肠道施用的。非肠道施用的方法包括局部施用，动脉内（直接施用于肿瘤），肌内，皮下，髓内，鞘内，心室内，静脉内，腹膜内或鼻内施用。除了活性配料之外，这些药物组合物可以含有适当的药物适当的载体，包括赋形剂和辅助剂，其促进将活性化合物加工为可作为药物施用的制品。关于配方和施用技术的进一步阐述见于最近版本的Remington's 药理学（Ed.Maack 出版公司，Easton, Pa.）。

口服施用的药物组合物，可以使用本领域熟知的药物适当的载体，以适于口服的剂量配制。这种载体可以使药物组合物配制为片剂，药丸，糖衣丸，胶囊，液体，凝胶，糖浆，悬浮液等，由患者摄取。

口服的药物制品可以通过将活性化合物与固体赋形剂组合，随意地研磨为混合物，及如果需要，加入适当的辅助剂之后，将此混合物加工为细粒以获得片剂或糖衣丸。适当的赋形剂是碳水化合物或蛋白质充填剂如糖，包括乳糖，蔗糖，甘露醇或山梨糖醇；来自谷物，小麦，水稻，马铃薯或其它植物的淀粉；纤维素如甲基纤维素，羟丙甲基纤维素，或羧甲基纤维素钠；和树脂包括阿拉伯树胶和西黄蓍胶；和蛋白质如明胶和胶原蛋白。如果需要，可以加入分解剂或增溶剂，如交联的聚乙烯吡咯烷酮，琼脂，藻酸或藻酸盐，如藻酸钠。

糖衣丸是用适当的包衣如浓缩的糖溶液包被的，其还可以含有阿拉伯树胶，滑石，聚乙烯吡咯烷酮，聚羧乙烯凝胶，聚乙二醇和/或二氧化钛，紫胶漆溶液，和适当的有机溶剂或溶剂混合物。可以在片剂或糖衣丸中加入染料或色素，以鉴别产物或确定活性化合物的数量，即剂量。

可以口服的药物制品包括由明胶组成的推入契合的胶囊，以及由明胶和包衣如甘露醇或山梨糖醇组成的柔软的密封的胶囊。推入契合的胶囊可以含有活性配料，混合充填剂或结合剂如乳糖或淀粉，润滑剂如滑石或硬脂酸镁，及任选的稳定剂。在柔软的胶囊中，活性化合物可以溶解或悬浮于适当的液体中，如脂肪油，液体石蜡或液态聚乙二醇中，有或没有稳定剂。

非肠道施用的药物配方包括活性化合物的水溶液。就注射而言，本发明的药物组合物可以在水溶液中配制，优选在生理适当的缓冲液如 Hank' s 溶液，Ringer' s 溶液或生理缓冲盐水中配制。水相注射悬浮液可含有提高悬浮液粘度的物质，如羧甲基纤维素钠，山梨糖醇或右旋糖苷。另外，活性化合物的悬浮液可以制备为适当的油性注射悬浮液。适当的脂质溶剂或运载体包括脂肪油如芝麻油，或合成的脂肪酸酯，如乙基油酸酯或甘油三酯，或脂质体。任选地，悬浮液还可以含有适当的稳定剂或提高化合物溶解性的制剂，以可以制备高度浓缩的溶液。

就局部或鼻腔施用而言，在配方中使用可以透过特殊屏障的适当渗透剂。这种渗透剂是本领域已知的。

### 试剂盒

本发明还涉及药物包装和试剂盒，其包含一或多个充填了一或多种前述本发明组合物的配料的容器。所使用的容器应参照负责药品或生物制品生产，应用或销售的政府部门的建议，所提的建议必需被用于临床的生产，应用或销售部门接受。另外，药物成分可以与其他有治疗作用的化合物联合使用。

### 生产和贮存

本发明的药物组合物可以透过本领域已知的方式生产，例如透过常规的混合，消散，成粒，包衣，研磨，乳化，胶囊化，截留或冻干程序。

药物组合物可以以盐的形式提供，并可以用许多酸形成，包括但非限于盐酸，硫酸，乙酸，乳酸，酒石酸，马来酸，琥珀酸等。盐在水性或其它质子溶剂中往往是更可溶的，相当于游离碱基形式。在其它情况中，优选的制品可以是冻干的粉末，1mM—50mM 组氨酸，0.1%—2%蔗糖，2%—7%甘露醇，pH4.5—5.5，在使用之前用缓冲液将它们组合。

在已经制备药物组合物之后，所述组合物包含在适当载体中配制的本发明的化合物，将它们置于适当的容器中，并标示，以在指定条件下处理。就施用 RG1 而言，这种标示包括施用数量，频率和方法。

### 治疗有效量

适用于本发明的药物组合物包括这样的组合物，其中活性配料是有效量的，以达到想要的目的，即特征在于通过 RG1 表达治疗特殊的疾病。本领域技术人员熟知确定有效量的方法。

就任何化合物而言，治疗有效量可以在细胞培养试验如肿瘤形成细胞中，或在动物模型中初步评估，一般使用小鼠，兔，狗或猪。动物模型还用于达到所需的浓度范围和施用途径。这样的信息然后可以用于确定在人体内的有效剂量和施用途径。

治疗有效量是指改善症状或状态的蛋白质或其抗体，拮抗剂或抑制剂的数量。这种化合物的治疗效力和毒性可以

通过标准药理学程序在细胞培养物或试验动物中确定，例如  $ED_{50}$ （半数治疗有效量）和  $LD_{50}$ （半数致死量）。治疗效力和毒性之间的剂量比是治疗指数，并可以以  $ED_{50}/LD_{50}$  表示。优选呈现高治疗指数的药物组合物。得自细胞培养试验和动物研究的数据用于明确阐述用于人体的剂量范围。这种化合物的剂量优选在流通浓度范围内，包括  $ED_{50}$  及很小的或没有毒性。根据应用的剂型，患者的敏感性及其施用途径，剂量可以在此范围内变化。

精确的剂量是根据治疗的患者的个体差异选择的。调节剂量和施用方法以提供足够水平的活性组分，或保持所需效力。需加以考虑的其它因素包括病情程度，例如肿瘤大小和位置；患者的年龄，体重和性别；饮食，施用时间和频率，药物组合，反应敏感性，和对治疗的耐受性/应答。根据特殊配方的半衰期和清除率，长效药物组合物可以每 3—4 天，每周，或每两周施用一次。

根据施用途径，标准剂量可以在 0.1—100000ug 之间变化，最多总剂量为大约 1g。文献中提供了关于特殊剂量和施用方法的指导。见美国专利 No.4657760；5206344；或 5225212。本领域技术人员针对多核苷酸可应用与针对蛋白质或其抑制剂不同的配方。相似地，根据特殊的性别，条件，位置等施用特异的多核苷酸或多肽。

本发明还通过以下实施例加以进一步详述。实施例只是通过参考特异的实施方案例证了本发明。这些实施例例证了本发明的某些方面，无限制本发明范围之意。

如果没有特别说明，所有的实施例均使用本领域熟知的标准技术进行。以下实施例的常规分子生物学技术可以如

标准实验手册所述进行，如 Sambrook 等，分子克隆实验手册，第 2 版；冷泉港实验室出版社，冷泉港，N.Y.，1989。

### 实施例 1：鉴别人 rg1 多核苷酸

rg1 是作为在前列腺中表达的基因通过淘选 Incyte' s LifeSeq 数据库鉴别的。核苷酸序列是对数据库的注释研究，使用 Incyte 为研究数据库而提供的“蛋白质功能”工具鉴别的。核苷酸序列是在注释的数据库中的细胞粘着分子中发现的，并阐述为同源的 f-spondin。对数据库中一系列文库中的 rg1 多核苷酸序列的分布进行的电子 Northern 分析表明，rg1 在前列腺文库中高水平表达，在其它许多组织文库中以较低水平表达，包括正常的组织和肿瘤组织。

在将数据库中 rg1 克隆系列装配为连续的多核苷酸系列，并对此连续的系列进行编辑后，在预定装配的多核苷酸中鉴别一个全长的编码序列。这个序列编码与 f-spondin 和 Mindin-2 同源的蛋白质。

Incyte 克隆 1640796，1712252 和 1880265 得自 Incyte，以进行试验研究，克隆 3360733 经鉴别含有最多的 5' 核苷酸序列。对这个克隆进行充分测序，含有推定的 RG1 蛋白的全部编码序列。这个序列示于图 1(SEQ ID NO:1)。

### 实施例 2：rg1 mRNA 表达

对得自正常组织和肿瘤组织和细胞系的各种样品中 rg1 mRNA 的表达，使用 Taqman 分析 (Perkin-Elmer) 通过半定量 PCR 测定。前列腺正常的，良性的和肿瘤组织样品得自史丹佛医科大学泌尿系，所述样品已经根据修改的 Gleason 分级系统分级。通过标准方法从这些样品中分离 RNA。来

自其它肿瘤和正常组织的 RNA 是商购的, 包括从 Clontech 和 Biochain 商购的。前列腺肿瘤细胞系 (PC-3, LNCaP 和 DU145) 得自美国典型培养物保藏中心, 并通过标准方法使用含有血清的培养基培养增殖。在裸鼠中建立来自这些细胞系的异种移植的肿瘤, 并在植入后大约 4—6 周, 从这些小鼠中收集肿瘤。通过标准方法从这些肿瘤中分离 RNA。

进行基于 Taqman 的 PCR 分析, 使用引物 CGC GCA TAG CTC CGA CTA C (SEQ ID NO:3) 和 GCC GCG TCC GCA AAG (SEQ ID NO:4), 和 Taqman 探针: 6-FAM-AGG AAG AAC CAG TAC GTC AGT AAC GGG CTG-Tamra (SEQ ID NO:5)。

这些引物和探针是用 Perkin Elmer 的引物表达软件设计的, 并通过合成遗传学合成。PCR 反应进行 30—40 次循环, 并用前列腺 RNA 量化, 以产生相对比的标准曲线。此分析表明 rg1 mRNA 在前列腺中最高丰余检测, 在一些其它组织中明显水平较低 (见图 5)。

### 实施例 3: 在 BHK 细胞中克隆和表达 RG1

RG1 编码区得自 Incyte 质粒 3360733。将此编码序列在标准 PCR 反应中 (100ul), 使用 1×Pfu Turbo 聚合酶缓冲液 (Stratagene, La Jolla, CA) /200uM dNTPs/0.2uM 寡核苷酸引物/2.5U Pfu Turbo 聚合酶 (Stratagene), 经 PCR 扩增, 使用引物 SST115 (5' -TCCCTCTAGAGCCACCATGGAAAA CCCAGCCCGGC-3' ) (SEQ ID NO:6), 和 SST113 (5' -AAGGCATCACGTGTTAGACGCAGTTATCAGGGAC G-3' ) (SEQ ID NO:7)。PCR 扩增条件如下: 95° C 3 分钟, (在 95° C 15 秒, 在 60° C 30 秒, 在 72° C 2 分钟)

×35 次，72° C 7 分钟。将所得 PCR 扩增产物用 QIAquick PCR 层析柱 (Qiagen, Valencia, CA) 纯化，并用 XbaI 和 PmaI 限制酶纯化，产生一个 1010bp 的片段，其是用 BIO 101 GeneClean 试剂盒 (Vista, CA) 从 1%琼脂糖凝胶中纯化的。将此纯化的片段连接于 (使用 Epicentre 快速连接试剂盒 (Epicenter, Madison, WI)) 用 XbaI 和 PmaI 消化的非致细胞病变的 Sindbis 表达载体 pSINrep21 (Agapov 等, 1998, PNAS95:12989—12994) 中，并转化入 DH5 $\alpha$ 感受态细胞中 (Life Technologies, Gaithersburg, CA)，及在含有氨苄青霉素 (100ug/ml) 的 LB 琼脂平板上选择。将一个这种氨苄青霉素抗性集落生长于含有氨苄青霉素的 LB 培养基中，通过序列分析示出其含有插入的 RG1 编码序列。这个质粒称为 pPEG6。

使用 Lipofectamine Plus 试剂 (Life Technologies, Gaithersburg, MD)，根据产生导致指导，将 2ug 的 PEG6 用于转染  $1-3 \times 10^5$  个牛仓鼠肾 (BHK) 细胞。在转染后，将细胞在加上胎牛血清的 DMEM 中温育 24—48 小时，此时细胞从 1 分裂为 10，通过加入嘌呤霉素 (终浓度为 2.5ug/ml) 和含有血清的 DMEM，开始选择含有质粒的细胞。在细胞铺满后 (加入嘌呤霉素 4—5 天后)，将细胞用 PBS 冲洗，，分裂 1—10，加入具有血清的 DMEM 和 5ug/ml 嘌呤霉素。在另外 2—3 天后，将培养基用没有血清的 DMEM 和 5ug/ml 嘌呤霉素置换，生长 2—3 天后，使用 RG1 抗体通过 Western 分析检测培养基中 RG1 蛋白的存在情况。检测的 RG1 蛋白的水平为 1ug/ml。

实施例 4: 抗体产生

兔多克隆抗血清是对抗衍生自 RG1 蛋白序列的 5 种合成的多肽序列而产生的。选择这些序列是因为推定其位于蛋白质表面，以产生更易于识别表面表位的抗血清。将半胱氨酸残基用氨基丁酸 (Abu) 置换以助于合成。所述 5 种多肽的特异氨基酸序列，在 RG1 蛋白上的位置，和名称如下所示。

名称	位置	氨基酸序列
C	28-46	PLGGESICSAGAPAKYSIT (SEQ ID NO:8)
2C	46-64	TFTGKWSQTAFPKQYPLFR(SEQ ID NO:9)
3C	77-91	HSSDYSMWRKNQYVS(SEQ ID NO:10)
4C	188-210	DAGTDSGFTFSSPNFATIPQDTV (SEQ ID NO:11)
5C	263-274	NEIVDSASVPET(SEQ ID NO:12)

将肽与匙孔瓣血蓝蛋白 (KLH) 通过另外的羧基端半胱氨酸常规偶联，以用作免疫原。相似地，制备牛血清白蛋白 (BSA) 缀合物以通过 ELISA 分析抗血清滴定。

将两种动物用每种肽免疫接种。初始的免疫接种是在 Freund's 完全佐剂中 (0.5mg 肽/动物) 进行的，随后间隔 3 周肌肉注射 Freund's 不完全佐剂中以 0.25mg 肽/动物攻击。定期取从实验动物中取血液，通过 ELISA 测定抗特异性 BSA-肽缀合物的抗体滴定，并与免疫前血清对比。抗 1C 和 3C 肽的抗血清示出是活性的。抗 2C 肽的抗血清不识别 RG1 多肽。抗 4C 和 5C 肽的抗血清未测试。

抗 RG1 的人单克隆抗体是通过免疫接种抗 RG1 肽和在大肠杆菌中表达的 6-组氨酸标记的 RG1 融合蛋白的转基因小鼠产生的。将这些动物的脾细胞与骨髓瘤细胞融合，以产生杂交瘤细胞。将所得杂交瘤通过 ELISA 筛选产生直接抗 RG1 多肽和蛋白的抗体的那些杂交瘤。

### 实施例 5: 对抗体进行 Western 印迹分析

通过 Western 印迹测试抗血清的 RG1 特异性。RG1 特异性抗血清（上述抗 1C 和 3C 序列产生的那些）进行测试在 COS 细胞中 RG1 瞬时表达，从 LNCaP 细胞中分泌天然 RG1 和从转染的幼仓鼠肾细胞（BHK）中产生的 RG1 情况。RG1-抗血清还进一步测试制备自以下物质的裂解物：LNCaP 肿瘤，LNCaP 细胞，PC3 肿瘤，PC3 细胞和人前列腺肿瘤的一些临床样品。将细胞和组织在去污剂缓冲液中裂解。在沸腾 5 分钟后，将 10ul 的每种裂解物加样于 12%SDS-聚丙烯酰胺凝胶上以解离蛋白质。然后将分离的蛋白质移至硝基纤维素膜上。RG1 抗体的结合特异性通过在存在同源和异源肽的情况下的结合加以验证。RG1 特异性抗血清可以检测除了 PC-3 细胞和 PC-3 肿瘤之外的所有样品中的蛋白质。

### 实施例 6: 纯化分泌自 LNCaP 细胞的天然 RG1 蛋白

生长于培养物中的 LNCaP 细胞通过 Western 印迹分析示出分泌天然 RG1 蛋白。为纯化此天然的蛋白质，将细胞在没有血清的培养基中生长 48 小时。收集此没有血清的条件培养基，离心以除去所有细胞，并通过超滤浓缩将近 50 倍。然后将浓缩的培养基用 20mM 乙酸钠（pH6.5）缓冲液稀释 10 倍，加样于 Q-琼脂糖阴离子交换层析柱上。层析柱洗脱由 NaCl 梯度（每分钟 5%）组成，同时收集 2.0ml 级分。将在大约 75mM NaCl 洗脱的 RG1 蛋白质通过 Western 印迹和 SDS PAGE 测定。天然 RG1 蛋白在略低于在细菌中表达的 6 组氨酸-RG1 融合蛋白的分子量泳动，推测上因为

其没有融合蛋白。

#### 实施例 7: 对 RG1 表达进行免疫组织化学染色

RG1 蛋白的表达是由 LifeSpan 生物科学公司在多种人体组织中测定的, 包括肾, 肺, 胰腺, 肌, 脑和前列腺。另外的前列腺组织得自史丹佛大学泌尿系, 并由 Berlex 测试。将组织切片用标准方法脱蜡。将多克隆抗体 RG1-3C 用作初级抗体, 检测系统由载体 ABC-AP 试剂盒 (AK5002) 和载体红色底物试剂盒 (Sk5002) 组成。作为阴性对照, 在没有初级抗体的情况下进行染色。

在上述说明书中提及的所有出版物和专利在此均并入参考。本发明结合特异的实施方案已经进行阐述, 但本领域技术人员应意识到在不偏离本发明精神和范围的条件下可以对本发明加以修改和用等价物代替。另外, 对本发明可进行许多修改以适应特殊情况, 材料, 组合物, 加工, 加工步骤等。所有这些修改应包含在所附权利要求范围内。

## 序 列 表

<110> Markins, Richard  
 Parkes, Deborah  
 Parry, Gordon  
 Schneider, Douglas  
 Steinbrecher, Renate

<120> 编码新RG1多肽的DNA

<130> 51791AUSM1

<140>

<141>

<150> 60/172,370

<151> 1999-12-16

<160> 12

<170> PatentIn Ver. 2.0

<210> 1

<211> 1785

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> (296)..(1291)

<400> 1

```

agzaagggggt gcggcagcac tgccagggga agaggggtgat ccgacccggg gaaggtcgct 60
gggcaggggcg agttgggaaa gcggcagccc ccgccgcccc cgcagccctt tctcctcctt 120
tctcccacgt cctatctgcc tctcgtctgga ggccaggccg tgcagcatcg aagacaggag 180
gaactggagc ctcatctggcc ggcccggggc gccggcctcg ggcttaata ggagctccgg 240
gctctggctg ggacccgacc gctgccggcc gcgctcccgc tgctcctgcc gggtg atg 298
                                     Met
                                     1

gaa aac ccc agc ccg gcc gcc gcc ctg ggc aag gcc ctc tgc gct ctc 346
Glu Asn Pro Ser Pro Ala Ala Ala Leu Gly Lys Ala Leu Cys Ala Leu
                    5                      10                      15

ctc ctg gcc act ctc ggc gcc gcc ggc cag cct ctt ggg gga gag tcc 394
Leu Leu Ala Thr Leu Gly Ala Ala Gly Gln Pro Leu Gly Gly Glu Ser
                20                      25                      30

atc tgt tcc gcc gga gcc ccg gcc aaa tac agc atc acc ttc acg ggc 442
Ile Cys Ser Ala Gly Ala Pro Ala Lys Tyr Ser Ile Thr Phe Thr Gly
                35                      40                      45

```

aag tgg agc cag acg gcc ttc ccc aag cag tac ccc ctg ttc cgc ccc	490
Lys Trp Ser Gln Thr Ala Phe Pro Lys Gln Tyr Pro Leu Phe Arg Pro	
50 55 60 65	
cct gcg cag tgg tct tcg ctg ctg ggg gcc gcg cat agc tcc gac tac	538
Pro Ala Gln Trp Ser Ser Leu Leu Gly Ala Ala His Ser Ser Asp Tyr	
70 75 80	
agc atg tgg agg aag aac cag tac gtc agt aac ggg ctg cgc gac ttt	586
Ser Met Trp Arg Lys Asn Gln Tyr Val Ser Asn Gly Leu Arg Asp Phe	
85 90 95	
gcg gag cgc gcc gag gcc tgg gcg ctg atg aag gag atc gag gcg gcg	634
Ala Glu Arg Gly Gln Ala Trp Ala Leu Met Lys Glu Ile Glu Ala Ala	
100 105 110	
ggg gag gcg ctg cag agc gtg cac gcg gtg ttt tcg gcg ccc gcc gtc	682
Gly Glu Ala Leu Gln Ser Val His Ala Val Phe Ser Ala Pro Ala Val	
115 120 125	
ccc agc gcc acc ggg cag acg tcg gcg gag ctg gag gtg cag cgc agg	730
Pro Ser Gly Thr Gly Gln Thr Ser Ala Glu Leu Glu Val Gln Arg Arg	
130 135 140 145	
cac tcg ctg gtc tcg ttt gtg gtg cgc atc gtg ccc agc ccc gac tgg	778
His Ser Leu Val Ser Phe Val Val Arg Ile Val Pro Ser Pro Asp Trp	
150 155 160	
ttc gtg gcc gtg gac agc ctg gac ctg tgc gac ggg gac cgt tgg cgg	826
Phe Val Gly Val Asp Ser Leu Asp Leu Cys Asp Gly Asp Arg Trp Arg	
165 170 175	
gaa cag gcg gcg ctg gac ctg tac ccc tac gac gcc ggg acg gac agc	874
Glu Gln Ala Ala Leu Asp Leu Tyr Pro Tyr Asp Ala Gly Thr Asp Ser	
180 185 190	
ggc ttc acc ttc tcc tcc ccc aac ttc gcc acc atc ccg cag gac acg	922
Gly Phe Thr Phe Ser Ser Pro Asn Phe Ala Thr Ile Pro Gln Asp Thr	
195 200 205	
gtg acc gag ata acg tcc tcc tct ccc agc cac ccg gcc aac tcc ttc	970
Val Thr Glu Ile Thr Ser Ser Ser Pro Ser His Pro Ala Asn Ser Phe	
210 215 220 225	
tac tac cca cgg ctg aag gcc ctg cct ccc atc gcc agg gtg aca ctg	1018
Tyr Tyr Pro Arg Leu Lys Ala Leu Pro Pro Ile Ala Arg Val Thr Leu	
230 235 240	
gtg cgg ctg cga cag agc ccc agg gcc ttc atc cct ccc gcc cca gtc	1066
Val Arg Leu Arg Gln Ser Pro Arg Ala Phe Ile Pro Pro Ala Pro Val	
245 250 255	
ctg ccc agc agg gac aat gag att gta gac agc gcc tca gtt cca gaa	1114
Leu Pro Ser Arg Asp Asn Glu Ile Val Asp Ser Ala Ser Val Pro Glu	
260 265 270	

acg ccg ctg gac tgc gag gtc tcc ctg tgg tgg tcc tgg gga ctg tgc 1162  
 Thr Pro Leu Asp Cys Glu Val Ser Leu Trp Ser Ser Trp Gly Leu Cys  
 275 285

gga ggc cac tgt ggg agg ctc ggg acc aag agc agg act cgc tac gtc 1210  
 Gly Gly His Cys Gly Arg Leu Gly Thr Lys Ser Arg Thr Arg Tyr Val  
 290 295 300 305

cgg gtc cag ccc gcc aac aac ggg agc ccc tgc ccc gag ctc gaa gaa 1258  
 Arg Val Gln Pro Ala Asn Asn Gly Ser Pro Cys Pro Glu Leu Glu Glu  
 310 315 320

gag gct gag tgc gtc cct gat aac tgc gtc taa gaccagagcc ccgcagcccc 1311  
 Glu Ala Glu Cys Val Pro Asp Asn Cys Val  
 325 330

tggggccccc cggagccatg ggggtgtcggg ggctcctgtg caggctcatg ctgcaggcgg 1371

ccgagggcac aggggggttc gcgctgctcc tgaccgaggc gaggccgagc cgaccatctc 1431

tgcactgaag ggcctctctg tggccggcac gggcattggg aaacagcctc ctcttttccc 1491

aaccttgctt cttaggggccc cccgtgtccc gtctgctctc agcctcctcc tcttgagga 1551

taaagtcate cccaaggctc cagctactct aaattatgct tccttataag ttattgctgc 1611

tccagcagat tgccttcat cgtccagggg cctggctccc acgtggttgc agatacctca 1671

gacctggtgc tctaggctgt gctgagccca ctctcccgag ggcgcacca agcggggggc 1731

acttgagaag tgaataaatg gggcggtttc ggaagcgtca aaaaaaaaaa aaaa 1785

<210> 2  
 <211> 331  
 <212> PRT  
 <213> Homo sapiens

<400> 2  
 Met Glu Asn Pro Ser Pro Ala Ala Ala Leu Gly Lys Ala Leu Cys Ala  
 1 5 10 15

Leu Leu Leu Ala Thr Leu Gly Ala Ala Gly Gln Pro Leu Gly Gly Glu  
 20 25 30

Ser Ile Cys Ser Ala Gly Ala Pro Ala Lys Tyr Ser Ile Thr Phe Thr  
 35 40 45

Gly Lys Trp Ser Gln Thr Ala Phe Pro Lys Gln Tyr Pro Leu Phe Arg  
 50 55 60

Pro Pro Ala Gln Trp Ser Ser Leu Leu Gly Ala Ala His Ser Ser Asp  
 65 70 75 80

Tyr Ser Met Trp Arg Lys Asn Gln Tyr Val Ser Asn Gly Leu Arg Asp  
 85 90 95

Phe Ala Glu Arg Gly Glu Ala Trp Ala Leu Met Lys Glu Ile Glu Ala  
 100 105 110  
 Ala Gly Glu Ala Leu Gln Ser Val His Ala Val Phe Ser Ala Pro Ala  
 115 120 125  
 Val Pro Ser Gly Thr Gly Gln Thr Ser Ala Glu Leu Glu Val Gln Arg  
 130 135 140  
 Arg His Ser Leu Val Ser Phe Val Val Arg Ile Val Pro Ser Pro Asp  
 145 150 155 160  
 Trp Phe Val Gly Val Asp Ser Leu Asp Leu Cys Asp Gly Asp Arg Trp  
 165 170 175  
 Arg Glu Gln Ala Ala Leu Asp Leu Tyr Pro Tyr Asp Ala Gly Thr Asp  
 180 185 190  
 Ser Gly Phe Thr Phe Ser Ser Pro Asn Phe Ala Thr Ile Pro Gln Asp  
 195 200 205  
 Thr Val Thr Glu Ile Thr Ser Ser Ser Pro Ser His Pro Ala Asn Ser  
 210 215 220  
 Phe Tyr Tyr Pro Arg Leu Lys Ala Leu Pro Pro Ile Ala Arg Val Thr  
 225 230 235 240  
 Leu Val Arg Leu Arg Gln Ser Pro Arg Ala Phe Ile Pro Pro Ala Pro  
 245 250 255  
 Val Leu Pro Ser Arg Asp Asn Glu Ile Val Asp Ser Ala Ser Val Pro  
 260 265 270  
 Glu Thr Pro Leu Asp Cys Glu Val Ser Leu Trp Ser Ser Trp Gly Leu  
 275 280 285  
 Cys Gly Gly His Cys Gly Arg Leu Gly Thr Lys Ser Arg Thr Arg Tyr  
 290 295 300  
 Val Arg Val Gln Pro Ala Asn Asn Gly Ser Pro Cys Pro Glu Leu Glu  
 305 310 315 320  
 Glu Glu Ala Glu Cys Val Pro Asp Asn Cys Val  
 325 330

<210> 3  
 <211> 19  
 <212> CNA  
 <213> 人工序列

<220>  
 <223> 人工序列描述: 引物

<400> 3  
 cgcgcatagc tccgactac

19

<210> 4  
 <211> 15  
 <212> DNA  
 <213> 人工序列

<220>  
 <223> 人工序列描述: 引物

<400> 15  
 gccgcgctccg caaag

<210> 5  
 <211> 30  
 <212> DNA  
 <213> 人工序列

<220>  
 <223> 人工序列描述: 探针

<400> 5 30  
 aggaagaacc agtacgtcag taacgggctg

<210> 6  
 <211> 36  
 <212> DNA  
 <213> 人工序列

<220>  
 <223> 人工序列描述: 引物

<400> 6 36  
 tccctctaga gccaccatgg aaaaccccag cccggc

<210> 7  
 <211> 35  
 <212> DNA  
 <213> 人工序列

<220>  
 <223> 人工序列描述: 引物

<400> 7 35  
 aaggcatcac gtgtagacg cagttatcag ggacg

<210> 8  
 <211> 19  
 <212> PRT  
 <213> 人工序列

<220>

<223> 人工序列描述: 肽

<400> 8

Pro Leu Gly Gly Glu Ser Ile Cys Ser Ala Gly Ala Pro Ala Lys Tyr  
1 5 10 15

Ser Ile Thr

<210> 9

<211> 19

<212> PRT

<213> 人工序列

<220>

<223> 人工序列描述: 肽

<400> 9

Thr Phe Thr Gly Lys Trp Ser Gln Thr Ala Phe Pro Lys Gln Tyr Pro  
1 5 10 15

Leu Phe Arg

<210> 10

<211> 15

<212> PRT

<213> 人工序列

<220>

<223> 人工序列描述: 肽

<400> 10

His Ser Ser Asp Tyr Ser Met Trp Arg Lys Asn Gln Tyr Val Ser  
1 5 10 15

<210> 11

<211> 23

<212> PRT

<213> 人工序列

<220>

<223> 人工序列描述: 肽

<400> 11

Asp Ala Gly Thr Asp Ser Gly Phe Thr Phe Ser Ser Pro Asn Phe Ala  
1 5 10 15

Thr Ile Pro Gly Asp Thr Val  
20



## 图1

1 AGAAAGCGGT GCGGCAGCAC TGCCAGGGGA AGAGGGTGAT CCGACCCGGG  
 51 GAAGGTCGCT GGGCAGGGCG AGTTGGGAAA GCGGCAGCCC CCGCCGCCCC  
 101 CGCAGCCCCT TCTCCTCCTT TCTCCACGT CCTATCTGCC TCTCGCTGGA  
 151 GGCCAGGCCG TGCAGCATCG AAGACAGGAG GAACTGGAGC CTCATTGGCC  
 201 GGGCCGGGGC GCGGCCCTCG GGCTTAAATA GGAGCTCCGG GCTCTGGCTG  
 251 GGACCCGACC GCTGCCGGCC GCGCTCCCGC TGCTCCTGCC GGGTGATGGA  
 301 AAACCCAGC CCGGCCGCCG CCCTGGGCAA GGCCCTCTGC GCTCTCCTCC  
 351 TGGCCACTCT CGGCGCCGCC GGCCAGCCTC TTGGGGGAGA GTCCATCTGT  
 401 TCCGCCGGAG CCGCGGCCAA ATACAGCATC ACCTTCACGG GCAAGTGGAG  
 451 CCAGACGGCC TTCCCAAGC AGTACCCCTT GTTCCGCCCC CCTGCCAGT  
 501 GGTCTTCGCT GCTGGGGGCC GCGCATAGCT CCGACTACAG CATGTGGAGG  
 551 AAGAACCAGT ACGTCAGTAA CGGGCTGCGC GACTTTGCGG AGCGCGGCGA  
 601 GGCTTGGGCG CTGATGAAGG AGATCGAGGC GGCGGGGGAG GCGCTGCAGA  
 651 GCGTGCACGC GGTGTTTTTCG GCGCCCGCCG TCCCAGCGG CACCGGGCAG  
 701 ACGTCGSCGG AGCTGGAGGT GCAGCGCAGG CACTCGCTGG TCTCGTTTGT  
 751 GGTGCGCATC GTGCCAGCC CCGACTGGTT CGTGGGCGTG GACAGCCTGG  
 801 ACCTGTGCCA CGGGGACCGT TGGCGGGAAC AGGCGGCGCT GGACCTGTAC  
 851 CCCTACGACG CCGGGACGGA CAGCGGCTTC ACCTTCTCCT CCCCCAAGTT  
 901 CGCCACCATC CCGCAGGACA CGGTGACCGA GATAACGTCC TCCTCTCCCA  
 951 GCCACCCGGC CAACTCCTTC TACTACCCAC GGCTGAAGGC CCTGCCTCCC  
 1001 ATCGCCAGGG TGACACTGGT GCGGCTGCCA CAGAGCCCCA GGGCCTTCAT  
 1051 CCCTCCCGCC CCAGTCCTGC CCAGCAGGGA CAATGAGATT GTAGACAGCG  
 1101 CCTCAGTTC AGAAACGCCG CTGGACTGCG AGGTCTCCCT GTGGTCTGCC  
 1151 TGGGGACTGT GCGGAGGCCA CTGTGGGAGG CTCGGGACCA AGAGCAGGAC  
 1201 TCGCTACGTC CGGGTCCAGC CCGCCAACAA CGGGAGCCCC TGCCCCGAGC  
 1251 TCGAAGAAGA GGCTGAGTGC GTCCCTGATA ACTGGGTCTA AGACCAGAGC

## 图1续

1301 CCCGCAGCCC CTGGGGCCCC CCGGAGCCAT GGGGTGTCGG GGGCTCCTGT  
1351 GCAGGCTCAT GCTGCAGGCG GCCGAGGGCA CAGGGGGTTT CGCGCTGCTC  
1401 CTGACCGCGG TGAGGCCGCG CCGACCATCT CTGCACTGAA GGGCCCTCTG  
1451 GTGGCCGGCA CGGGCATTGG GAAACAGCCT CCTCCTTTCC CAACCTTGCT  
1501 TCTTAGGGGC CCCCCTGTCC CGTCTGCTCT CAGCCTCCTC CTCCTGCAGG  
1551 ATAAAGTCAT CCCCAAGGCT CCAGCTACTC TAAATTATGT CTCCTATAA  
1601 GTTATTGCTG CTCCAGGAGA TTGTCCTTCA TCGTCCAGGG GCCTGGCTCC  
1651 CACGTGGTTG CAGATACCTC AGACCTGGTG CTCTAGGCTG TGCTGAGCCC  
1701 ACTCTCCCGA GGGCGCATCC AAGCGGGGGC CACTTGAGAA GTGAATAAAT  
1751 GGGGCGGTTT CGGAAGCGTC AAAAAAAAAA AAAAA

## 图2

1 MENPSPAAAL GKALCALLLA TLGAAGQPLG GESICSAGAP AKYSITETGK  
51 WSQTAEPKOY PLFRPPAOWS SLLGAAHSSD YSMWRKNOYV SNGLEDEAEE  
101 GEAWALMKEI EAAGEALQSV HAVFSAPAVP SGTGQTSDEL EVORRHSLYS  
151 FVVRIVPSPD WFGVDSLDL CDGDRWREQA ALDLYPYDAG TDSGETESSP  
201 NEATIPQDTV TEITSSSPSH PANSFYYPRL KALPPIARVT LVRLRQSPRA  
251 FIPPAPVLPS RDNEIVDSAS VPETPLDCEV SLWSSWGLCG GHCGRIGTKS  
301 FTEYVRVQPA NNGSPCPELE EEAECVPDNC V





# 图4续

```

721 GCAGCGCAGGCACTCGCTGGTCTCGTTTGTGGTGCGCATCGTGCCCAGCCCCGACTGGTT 780
-----+-----+-----+-----+-----+-----+-----+-----+
CGTCGCGTCCGTGAGCGACCAGAGCAAAACACCACGCGTAGCACGGGTCGGGGCTGACCAA

b   O R R H S L V S F V V R I V P S P D W F -

841 CGTGGGCGTGGACAGCCTGGACCTGTGCGACGGGGACCGTTGGCGGGAACAGGCGGGCCT 840
-----+-----+-----+-----+-----+-----+-----+-----+
GCACCCGCACCTGTCGGACCTGGACACGCTGCCCTGGCAACCGCCCTGTCCGCCGCGA

b   V G V D S L D L C D G D R W R E Q A A L -

841 GGACCTGTACCCCTACGACGCCGGGACGGACAGCGGCTTCACTTCTCCTCCCCAACTT 900
-----+-----+-----+-----+-----+-----+-----+-----+
CCTGGACATGGGGATGCTGCGGCCCTGCCTGTGCGCGAAGTGAAGAGGAGGGGTTGAA

b   D L Y P Y D A G T D S G F T F S S P N F -

901 CGCCACCATCCCCGAGGACACGGTGACCGAGATAACGTCCCTCCTCTCCAGCCACCCGGC 960
-----+-----+-----+-----+-----+-----+-----+-----+
GCGGTGGTAGGGCGTCTGTGCCACTGGCTCTATTGCAGGAGGAGAGGGTCGGTGGGCCG

b   A T I P Q D T V T E I T S S S P S H P A -

961 CAACTCCTTCTACTACCCACGGCTGAAGGCCCTGCCTCCCATCGCCAGGGTGACACTGGT 1020
-----+-----+-----+-----+-----+-----+-----+-----+
GTTGAGGAAGATGATGGGTGCCGACTTCCGGGACGGAGGGTAGCGGTCCCACTGTGACCA

b   N S F Y Y P R L K A L P P I A R V T L V -

1021 GCGGCTGCGACAGAGCCCCAGGGCCCTTCATCCCTCCCCGCCCCAGTCCTGCCAGCAGGGA 1080
-----+-----+-----+-----+-----+-----+-----+-----+
CGCCGACGCTGTCTCGGGTCCCGGAAGTAGGGAGGGCGGGGTGAGGACGGGTGCTCCCT

b   R L R Q S P R A F I P P A P V L P S R D -

1081 CAATGAGATTGTAGACAGCGCCTCAGTTCCAGAAACGCGCTGGACTGCGAGGTCTCCCT 1140
-----+-----+-----+-----+-----+-----+-----+-----+
GTTACTCTAAACATCTGTGCGGGAGTCAAGGTCTTTGCGGGACCTGACGCTCCAGAGGGA

b   N E I V D S A S V P E T P L D C E V S L -

1141 GTGGTCTCCTGGGACTGTGCGGAGGCCACTGTGGGAGGCTCGGGACCAAGAGCAGGAC 1200
-----+-----+-----+-----+-----+-----+-----+-----+
CACCAGCAGGACCCCTGACACGCCTCCGGTGACACCCCTCCGAGCCCTGGTTCTCGTCCCTG

b   W S S W G L C G G H C G R L G T K S R T -

1201 TCGCTACGTCCGGTCCAGCCCCCAACAACGGGAGCCCTGCCCGAGCTCGAAGAAGA 1260
-----+-----+-----+-----+-----+-----+-----+-----+
AGCGATGCAGGCCCAGGTCGGGGCGTTGTTGCCCTCGGGGACGGGGCTCGAGCTTCTTCT

b   R Y V R V Q P A N N G S P C P E L E E E -

1261 GGCTGAGTGCCTCCCTGATAACTGCGTCTAAGACCAGAGCCCCGAGCCCCCTGGGGCCCC 1320
-----+-----+-----+-----+-----+-----+-----+-----+
CCGACTCACGCAGGGACTATTGACGCAGATTCTGGTCTCGGGGCGTCGGGGACCCCGGGG

b   A E C V P D N C V *

1321 CCGGAGCCATGGGGTGTGCGGGGCTCCTGTGCAGGCTCATGCTGCAGGCGGCCGAGGGCA 1380
-----+-----+-----+-----+-----+-----+-----+-----+
GGCCTCGGTACCCACAGCCCCCGAGGACACGTCCGAGTACGACGTCCGCCGGCTCCCGT

1381 CAGGGGTTTTCCGCTGCTCCTGACCCGGTGAGGCCCGCCGACCATCTCTGCACTGAA 1440
-----+-----+-----+-----+-----+-----+-----+-----+
GTCCCCAAGCCCGACCAAGACTGGCGCCACTCCGGCCGGCTGGTAGAGACGTGACTT

1441 GGGCCCTCTGGTGGCCGGCACGGCCATTGGGAAACAGCCTCCTCCTTTCCCAACCTTGCT 1500
-----+-----+-----+-----+-----+-----+-----+-----+
CCCCGGAGACCACCGCCCGTCCCCTAACCTTTGTGCGGAGGAGAAAGGGTTGGAACGA

```

### 图4续

```
TCTTAGGGGCCCCCGTGTCCCGTCTGCTCTCAGCCTCCTCCTCCTGCAGGATAAAAGTCAT
1501 -----+-----+-----+-----+-----+-----+-----+ 1560
AGAATCCCCGGGGGCACAGGGCAGACGAGAGTCGGAGGAGGAGGACGTCCTATTTTCAGTA

CCCCAAGGCTCCAGCTACTCTAAATTATGTCTCCTTATAAGTTATTGCTGCTCCAGGAGA
1561 -----+-----+-----+-----+-----+-----+-----+ 1620
GGGGTCCGAGGTCGATGAGATTTAATACAGAGGAATATTCAATAACGACGAGGTCCTCT

TTGTCCTTCATCGTCCAGGGCCCTGGCTCCCACGTGGTTGCAGATACCTCAGACCTGGTG
1621 -----+-----+-----+-----+-----+-----+-----+ 1680
AACAGGAAGTAGCAGGTCCCCGGACCGAGGGTGCACCAACGTCTATGGAGTCTGGACCAC

CTCTAGGCTGTGCTGAGCCCACTCTCCCGAGGGCGCATCCAAGCGGGGGCCACTTGAGAA
1681 -----+-----+-----+-----+-----+-----+-----+ 1740
GAGATCCGACACGACTCGGGTGAGAGGGCTCCCGCGTAGGTTCGCCCCCGGTGAACTCTT

GTGAATAAATGGGGCGGTTTCGGAAGCGTC
1741 -----+-----+-----+-----+ 1770
CACTTATTTACCCCGCCAAAGCCTTCGCAG
```



# 图6

由LNCaP细胞分泌的天然RG1蛋白的纯化

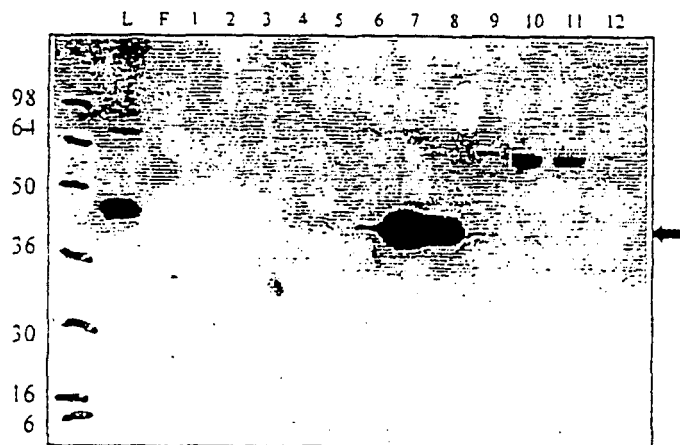


图7

RG1表达的免疫组分染色

