

[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 02805420.2

C07K 14/00 (2006.01)
C07K 2/00 (2006.01)
C07K 4/00 (2006.01)
C07K 5/00 (2006.01)
C07K 7/00 (2006.01)
A61K 38/00 (2006.01)

[45] 授权公告日 2009年5月13日

[11] 授权公告号 CN 100486992C

[51] Int. Cl. (续)

A61P 35/00 (2006.01)

A61P 43/00 (2006.01)

[22] 申请日 2002.2.22 [21] 申请号 02805420.2

[30] 优先权

[32] 2001.2.23 [33] US [31] 60/270,951

[86] 国际申请 PCT/US2002/005185 2002.2.22

[87] 国际公布 WO2002/067970 英 2002.9.6

[85] 进入国家阶段日期 2003.8.22

[73] 专利权人 斯克里普斯研究学院

地址 美国加利福尼亚州

[72] 发明人 P·施梅尔 K·瓦卡苏吉

M·弗里兰德

[56] 参考文献

Cloning and nucleotide sequence of the structural gene encoding for human tryptophanyl - tRNA synthetase. FROLOVA et al. Gene, Vol. 109 No. 2. 1991

Inhibition of Retinal Angiogenesis by Peptides Derived from Thrombospondin - 1. SHAFIEE et al. Investigative Ophthalmology & Visual Science, Vol. 41 No. 8. 2000

审查员 周洋

[74] 专利代理机构 中国专利代理(香港)有限公司

代理人 温宏艳 孟凡宏

权利要求书 3 页 说明书 54 页 附图 5 页

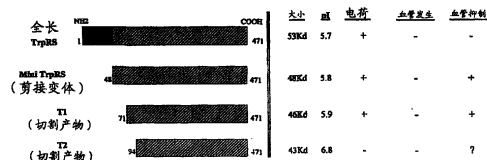
[54] 发明名称

用于调节血管发生的色氨酸 - tRNA 合成酶衍生的多肽

[57] 摘要

一种分离的、衍生自色氨酸 - tRNA 合成酶的水溶性多肽可用于抑制血管发生。该多肽基本由氨基酸残基序列：见右下式或其血管发生抑制片段组成。该分离的多肽的大小不超过约 45 千道尔顿。也提供了该多肽的使用方法。

人TrpRS构建体概要



*注：制备了四种蛋白中每一种的突变体，其中DLT(205-207)被BLR替代

```

1 MNSRPSALL ELNRSIATGQ ELVRSIDKGM ASKRDIDRVA KGLVSLRQVY KDAAGQVYCA DCPQGNFAPT MSRSPQVYEA
      ↓
81 KEDFVDFWTV QTFDANGIDY DELIVRFQSS KIDKELIRRI ERATGQRFER YLSRGTFFER KGMHQVLDAY ESRKDFYIAT
      ↓
141 GQGFSSRAM YVELIFFIFT KMLQVFNVP LVIQVTDKX VLRKDLQADQ AYGDVAENAK DIIACGFDIR KTFIFSDLDY
      ↓
241 MNSRSGFTDN VYKIQGVZFP NQVRSIFQFT DSDCKIKLRF PAIQDAPRFB MEFQIQRDR TDQCLLPCA IQDQFFRNF
      ↓
321 RDAVRIEHP KVALLESYFF FALQGAQEL SLSPSSSIF LDTFAQIKY KVNRHAPRGG RDTRESRQF GSGCDDVDF
      ↓
401 NYLFFLEDD DELEQIRSDY TSGANLQEL KVALLEVLQF LIAHQARQK EYDREIVRF MFRKLSDFD Q

```

*注：所有都是重组构建体并且有N末端Met和C末端KLAALAEHHHHHH

```

SAGIDVDKL IVRFGSSKID KELNRIERA TQQRPHHFLR RGIFFSHRDM
NQVLDAYEK KFFLYLTGRQ PSEBAMVGH LIPPIFTKNL QDVFVFLVI
QMTDDEKYLW KDLTLDQAVQ DAVENRSDI ACQFDINKPT IFSDLDMGN
SSGFYKVVK IQKHVIFNIV KGIQGTSDS CIGKISFPAL QAAPFSSNF
EQIFRRTDI QCLIPCAIDQ DPFYRTRDV APRIGYFKEA LLSHFFFPAL
QGAQTMSAS DPNSSIFLID TAKIKTKVN KHAFSGORDT IERHRQPGN
CDDVDSFMYL TFFLEDDDKL EQIRKDYTSG AMLTGLKKA LIEVLQPLIA
EHQAREKVT DEIVKHPMT RLSDFDQ (SBQ ID NO:12)

```

1. 一种分离的、水溶性多肽，基本由氨基酸残基序列

SAKGIDYDKL IVRFGSSKID KELINRIERA TGQRPHHFLR RGIFFSHRDM
 NQVLDAYENK KPFYLYTGRG PSSEAMHVGH LIPFIFTKWL QDVFNVPLVI
 QMTDDEKYLW KDLTLDQAYG DAVENAKDII ACGFDINKTF IFSDDLDMGM
 SSGFYKNVVK IQKHVTFNQV KGIFGFTDSD CIGKISFPAL QAAPSFNSNF
 PQIFRDRTDI QCLIPCAIDQ DPYFRMTRDV APRIGYPKPA LLHSTFFPAL
 QGAQTKMSAS DPNSSIFLTD TAKQIKTKVN KHAFSGGRDT IEEHRQFGGN
 CDVDVSEMYL TPFLEDDDKL EQIRKDYTSG AMLTGELKKA LIEVLQPLIA
 EHQARRKEVT DEIVKEFMTF RKLSEDFQ (SEQ ID NO:12)

或其血管发生抑制性片段组成；所述分离的多肽的大小不超过约 45 千道尔顿。

2. 根据权利要求 1 的分离的多肽，该多肽是能够抑制眼新血管形成的血管发生抑制性片段。

3. 根据权利要求 1 的分离的多肽，该多肽是能够抑制眼新血管形成的血管发生抑制性片段并且包含氨基酸残基信号序列 HVGH (SEQ ID NO: 10) 和 KMSAS (SEQ ID NO: 11) 中的至少一种。

4. 根据权利要求 1 的分离的多肽，该多肽是能够抑制眼新血管形成的血管发生抑制性片段并且包含氨基酸残基信号序列 HVGH (SEQ ID NO: 10)。

5. 根据权利要求 1 的分离的多肽，该多肽是能够抑制眼新血管形成的血管发生抑制性片段并且大小小于约 43 千道尔顿。

6. 一种具有 SEQ ID NO: 7 的氨基酸残基序列的分离的多肽。

7. 一种具有与选自下组的多核苷酸的序列至少 95% 同一的核苷酸序列的分离的多核苷酸：SEQ ID NO: 6 的多核苷酸；可以与 SEQ ID NO: 6 的多核苷酸杂交的多核苷酸；编码 SEQ ID NO: 7 的多肽的多核苷酸；编码 SEQ ID NO: 12 的多肽的多核苷酸；编码 SEQ ID NO: 7 的多肽表位的多核苷酸；以及可以与编码 SEQ ID NO: 7 的多肽表位的多核苷酸杂交的多核苷酸。

8. 一种包含权利要求 7 的分离的多核苷酸的重组载体。

9. 一种包含权利要求 8 的载体的重组宿主细胞。

10. 一种表达权利要求1的多肽的重组宿主细胞。

11. 一种水溶性多肽在制备用于抑制患者中眼新血管形成的药物中的用途，该多肽基本由氨基酸残基序列

SAKGIDYDKL IVRFGSSKID KELINRIERA TGQRPHHFLR RGIFFSHRDM
 NQVLDAYENK KPFYLYTGRG PSSEAMHVGH LIPFIPTKWL QDVFNVPLVI
 QMTDDEKYLW KDLTLDQAYG DAVENAKDII ACGFDINKTF IFSDLDYMG
 SSGFYKNVVK IQKHVTFNQV KGIFGFTDSD CIGKISFPAI QAAPSFSNSF
 PQIFRDRTDI QCLIPCAIDQ DPYFRMTRDV APRIGYPKPA LLHSTFFPAL
 QGAQTKMSAS DPNSSIFLTD TAKQIKTKVN KHAFSGGRDT IEEHRQFGGN
 CDVDVSFMYL TFFLEDDDKL EQIRKDYTSG AMLTGELKKA LIEVLQPLIA
 EHQARRKEVT DEIVKEFMTP RKLSEDFQ (SEQ ID NO: 12)

或其眼新血管形成抑制性片段组成。

12. 一种可注射的血管抑制性组合物，该组合物包含基本由氨基酸残基序列

SAKGIDYDKL IVRFGSSKID KELINRIERA TGQRPHHFLR RGIFFSHRDM
 NQVLDAYENK KPFYLYTGRG PSSEAMHVGH LIPFIPTKWL QDVFNVPLVI
 QMTDDEKYLW KDLTLDQAYG DAVENAKDII ACGFDINKTF IFSDLDYMG
 SSGFYKNVVK IQKHVTFNQV KGIFGFTDSD CIGKISFPAI QAAPSFSNSF
 PQIFRDRTDI QCLIPCAIDQ DPYFRMTRDV APRIGYPKPA LLHSTFFPAL
 QGAQTKMSAS DPNSSIFLTD TAKQIKTKVN KHAFSGGRDT IEEHRQFGGN
 CDVDVSFMYL TFFLEDDDKL EQIRKDYTSG AMLTGELKKA LIEVLQPLIA
 EHQARRKEVT DEIVKEFMTP RKLSEDFQ (SEQ ID NO: 12)

或其血管发生抑制性片段组成的多肽，以及药理学可接受的水性赋形剂，该组合物包含的多肽浓度为每毫升水性赋形剂至少 0.1 mg。

13. 根据权利要求12的血管抑制性组合物，其中该多肽具有 SEQ ID NO: 7 或 SEQ ID NO: 12 的氨基酸残基序列并且以每毫升水性赋形剂约 0.1-0.5 mg 的浓度存在。

14. 一种用于抑制眼新血管形成的试剂盒，包含包装于适当密封容器中的足够用于至少单剂量给药的量的权利要求1的多肽；至少一个适用于玻璃体内注射的自我密封的小于约 33 号的注射器针头；以及

至少一个精确校准的注射器。

15. 权利要求 14 的试剂盒，进一步包含打印的信息材料，它描述了组合物、其给药方法，以及政府规章可能要求的安全性和有效性信息。

用于调节血管发生的色氨酰-tRNA 合成酶衍生的多肽

优先权要求

本申请要求 2001 年 2 月 23 日提交的美国临时申请系列号 60/270,951 的优先权。

政府权利

本发明的进行受到了美国政府国家卫生部, GM23562 基金的政府性支持, 因此美国政府对本发明有某些权利。

发明领域

本发明涉及包含截短的 tRNA 合成酶多肽的组合物, 以及编码所述截短的 tRNA 合成酶多肽的核酸。也公开了制备和使用所述组合物的方法。

发明背景

氨基酰-tRNA 合成酶催化 tRNA 分子的氨基酰化, 它们是翻译过程中解码遗传信息的古老蛋白酶。在高等真核动物中, 9 种氨基酰-tRNA 合成酶与至少 3 种其它多肽缔合, 形成超分子多酶复合物 (Mirande et al., Eur. J. Biochem. 147: 281~89 1985)。每一种真核 tRNA 合成酶由核心酶和附加于核心酶的氨基末端或羧基末端的其它结构域组成, 该核心酶与 tRNA 合成酶的原核对应物密切相关 (Mirande, Prog. Nucleic Acid Res. Mol. Biol. 40: 95-142 1991)。

在大多数情况下, 附加的结构域似乎是用于多酶复合物的组装。然而, 额外结构域的存在并不严格与合成酶缔合为多酶复合物相关。

哺乳动物 TrpRS 分子具有氨基末端附加的结构域。在正常人细胞中, 可以检测到两种形式的 TrpRS: 由全长分子 (SEQ ID NO: 1 的氨基酸残基 1-471) 组成的主要形式和截短的次要形式 ("Mini-TrpRS"; SEQ ID NO: 3 的氨基酸残基 1-424)。次要形式是通过前 mRNA 的可变剪接而去除氨基末端结构域形成的 (Tolstrup et al., J. Biol. Chem. 270: 397-403 (1995))。已经确定出 Mini-TrpRS 的氨基末端是全

长 TrpRS 分子 48 位的甲硫氨酸氨基 (id.)。或者, 截短的 TrpRS 可以通过蛋白水解产生 (Lemaire et al., Eur. J. Biochem. 51: 237-52 (1975))。例如, 牛 TrpRS 在胰腺中高表达, 并且分泌到胰液中 (Kisselev, Biochimie 75: 1027-39 (1993)), 从而导致截短的 TrpRS 的产生。这些结果表明, 截短的 TrpRS 可以具有除氨基酰化 tRNA 以外的功能 (id.)。

血管发生, 或从已经存在的血管增殖出新的毛细血管, 是胚胎发育、其后的生长和组织修复的基本过程。血管发生是血管树发育和分化, 以及各种基本的生理过程, 包括胚胎发生、身体生长、组织和器官修复和再生、黄体和子宫内膜的周期性生长、以及神经系统发育和分化的先决条件。在女性生殖系统中, 血管发生存在于发育中的黄体中, 存在于排卵后的黄体中, 并且存在于胎盘中以建立和维持妊娠。血管发生还作为身体修复过程, 如伤口和骨折愈合过程的一部分而存在。血管发生也是肿瘤生长的一个因素, 这是因为肿瘤为了生长必须连续刺激新毛细血管的生长。血管发生是人实体癌生长的一个必须部分, 异常的血管发生与其它疾病如类风湿性关节炎、牛皮癣和糖尿病视网膜病相关 (Folkman, J. and Klagsbrun, M., Science 235: 442-447 (1987))。

一些因子参与血管发生。酸性和碱性成纤维细胞生长因子分子都是内皮细胞和其它细胞类型的有丝分裂原。尽管功能不清楚, Angiotropin 和血管生成素可以诱导血管发生 (Folkman, J., Cancer Medicine, pp. 153-170, Lea and Febiger Press (1993))。用于血管内皮细胞的高度选择性的有丝分裂原是血管内皮生长因子或 VEGF (Ferrara, N., et al., Endocr. Rev. 13: 19-32, (1992))。

大多数导致视力高度丧失的疾病是通过眼部新血管形成而造成这一后果的; 年龄相关的黄斑退化 (ARMD) 影响了 1200-1500 万 65 岁以上的美国人, 并且由于脉络膜 (视网膜下) 新血管形成的直接作用导致了他们当中 10-15% 的死亡。在 65 岁以上的美国人中, 视力丧失的最主要原因是糖尿病; 在美国, 1600 万人患有糖尿病, 每年有 40,000 人发生眼部并发症, 通常是由于视网膜新血管形成造成的。尽管激光凝固法可以有效预防高危糖尿病患者一个亚组的严重视力丧失, 视网膜病的总体 10 年发病率仍然基本没有改变。对于由于 ARMD 或炎症

性眼疾病如眼部组织胞浆菌病导致的脉络膜新血管形成的患者，光凝固法在预防视力丧失方面是无效的，仅有少数例外。尽管最近开发了有希望暂时减少以前治疗无效的脉络膜新血管形成患者的视力丧失的非破坏性光动力学治疗，仅 61.4%每 3-4 个月进行一次治疗的患者改善或稳定了视力，而安慰剂治疗组为 45.9%。

在正常成人中，血管发生是严格调节的，并且限于伤口愈合、妊娠和子宫周期。血管发生由特异性血管发生分子如碱性和酸性成纤维细胞生长因子 (FGF)、血管内皮生长因子 (VEGF)、血管生成素、转化生长因子 (TGF)、肿瘤坏死因子 α (TNF- α) 和血小板来源的生长因子 (PDGF) 激活。血管发生可以被抑制性分子如干扰素 α 、血小板反应蛋白-1、制管张素和内皮抑素而抑制。这些天然存在的刺激剂和抑制剂的平衡控制了正常情况下静止的毛细血管系统。当该平衡紊乱时，例如在某些疾病状态下，毛细血管内皮细胞被诱导增殖、迁移和最终分化。

血管发生在多种疾病，包括癌症和眼新血管形成中其了核心作用。也已经证明多种肿瘤的持续生长和转移依赖于新的宿主血管反应于肿瘤来源的血管发生因子而生长到肿瘤中。反应于多种刺激的新血管增殖是大多数眼病和眼盲，包括增殖性糖尿病视网膜病 (PDR)、ARMD、虹膜红变性青光眼、间质性角膜炎和早发性视网膜病中的主要发现。在这些疾病中，组织破坏可以刺激血管发生因子的释放，导致毛细血管增殖。VEGF 在虹膜新血管发生和新血管性视网膜病中起主要作用。尽管有报道明确显示了眼内 VEGF 水平和缺血性视网膜病性眼部新血管发生之间的关系，FGF 很可能起了作用。已知碱性和酸性 FGF 存在于正常成人视网膜中，尽管可检测的水平并不总是与新血管形成相关。这很可能是由于 FGF 非常紧密结合于细胞外基质的带电成分，并且不容易以能够被标准的眼内液体测定检测到的自由扩散的形式存在。

血管发生反应最常见的途径包括在增殖中的血管内皮细胞和细胞外基质之间形成整联蛋白介导的信息交换。这一类称作整联蛋白的粘附受体表达为异二聚体，在所有细胞上都具有 α 和 β 亚基。一种整联蛋白， $\alpha_v\beta_3$ 是该家族中最混栖的成员，并且允许内皮细胞与多种细胞外基质成分相互作用。该整联蛋白的肽和抗体拮抗剂通过选择性诱导

增殖中的血管内皮细胞的凋亡抑制血管发生。存在两种细胞因子依赖性血管发生途径，并且可以通过它们对不同血管细胞整联蛋白 $\alpha_v\beta_3$ 和 $\alpha_v\beta_5$ 的依赖性而确定。具体地，碱性 FGF 和 VEGF 诱导的血管发生分别依赖于 $\alpha_v\beta_3$ 和 $\alpha_v\beta_5$ ，因为每种整联蛋白的抗体拮抗剂在兔角膜和鸡绒毛尿囊膜 (CAM) 模型中选择性阻断这些血管发生途径中的一种。阻断所有 α_v 整联蛋白的肽拮抗剂抑制 FGF 和 VEGF 刺激的血管发生。尽管正常的血管没有显示出任何一种整联蛋白， $\alpha_v\beta_3$ 和 $\alpha_v\beta_5$ 选择性展示在活动的血管性眼病患者的组织中的血管上。尽管只有 $\alpha_v\beta_3$ 可以在 ARMD 患者的组织中持续观察到， $\alpha_v\beta_3$ 和 $\alpha_v\beta_5$ 都存在于 PDR 患者的组织中。全身性给予整联蛋白的肽拮抗剂可阻断视网膜血管发生的小鼠模型中的新血管形成。

因此，抗血管发生剂在治疗视网膜退化以防止这些向性因子和生长因子的破坏作用方面起作用。血管发生试剂具有促进需要的血管形成从而通过增加到细胞的血流而延缓视网膜退化的作用。

发明概述

色氨酰-tRNA 合成酶衍生的比天然存在形式短的多肽具有趋化因子活性并且可用于研究；诊断；预后和治疗用途。在一种实施方案中，这些 tRNA 合成酶衍生的多肽可用于调节血管内皮细胞功能，具体地，用于抑制血管发生，特别是眼新血管形成。

这些截短的色氨酰-tRNA 合成酶衍生的多肽具有氨基末端截短，但还可以包含 Rossmann 折叠核苷酸结合域。这些多肽可以调节血管内皮细胞功能。

一种优选的截短的色氨酰-tRNA 合成酶衍生的多肽包括基本由 SEQ ID NO: 1 的氨基酸残基 94-471 及其血管发生抑制性片段组成的多肽，所述片段特别是 SEQ ID NO: 10 和 SEQ ID NO: 11 所示信号序列的括号标出的片段或包含这些信号序列中的至少一个的片段。在一种优选实施方案中，截短的 tRNA 合成酶多肽是哺乳动物的，更优选是人的。

在另一实施方案中，本发明包括具有与选自下组的多核苷酸的序列至少 95% 同一的核苷酸序列的分离的多核苷酸：SEQ ID NO: 6 的多核苷酸；可以与 SEQ ID NO: 6 的多核苷酸杂交的多核苷酸；编码 SEQ

ID NO: 7 的多肽的多核苷酸; 编码 SEQ ID NO: 12 的多肽的多核苷酸; 编码 SEQ ID NO: 7 的多肽表位的多核苷酸; 以及可以与编码 SEQ ID NO: 7 的多肽表位的多核苷酸杂交的多核苷酸。本发明还包括一种包含编码上述任意色氨酰-tRNA 合成酶衍生的多肽的分离核酸分子的重组表达载体。另一方面是包含所述重组表达载体的宿主细胞。

本发明还提供了包含截短的色氨酰-tRNA 合成酶衍生的多肽和药物适用赋形剂的组合物和剂型。所述组合物适于眼内, 如玻璃体内、视网膜下等, 以及全身给药, 如经皮、经粘膜、肠道或肠胃外给药。

在另一种实施方案中, 本发明提供了一种通过给予血管发生抑制量的多肽和适当的生理相容的赋形剂或载体而治疗新生血管性眼病的方法, 所述眼病如年龄相关的黄斑变性、糖尿病眼部并发症、虹膜红变性青光眼、早发性视网膜病、角膜炎、局部缺血、视网膜病(如镰刀状细胞)、病理性近视、眼部组织胞浆菌病、翼状胬肉、punitate 内脉络膜病等。

附图简述

在附图中,

图 1 表示包含信号序列 (SEQ ID NO: 10 和 SEQ ID NO: 11) 的色氨酰-tRNA 合成酶多肽 (SEQ ID NO: 1), 该信号序列显示于方框中, 也包含在截短形式 (SEQ ID NO: 1 的氨基酸残基 94-471) 中。

图 2 是说明小鼠模型中视网膜血管发育的显微照片。

图 3 是下文中实施例 3 报道的数据的图示。

图 4 是下文中实施例 4 报道的数据的图示。

图 5 是说明 TrpRS 的一个片段 (T2) 在小鼠模型的视网膜中的结合位点的显微照片。

发明详述

定义

“截短的 tRNA 合成酶多肽”表示短于相应的全长 tRNA 合成酶的多肽。

“TrpRS”表示色氨酰-tRNA 合成酶。

“细胞培养物”包括培养基和培养的细胞。

短语“从细胞培养物分离多肽”包括分离从培养基分离可溶或分泌型多肽，以及从培养的细胞分离整合膜蛋白。

“细胞提取物”包括去除了细胞的培养基，特别是消耗培养基。应该理解，包含目的 DNA 或蛋白的细胞提取物表示从表达含有目的蛋白或 DNA 的细胞获得的均质制备物或无细胞制备物。

“质粒”是一种自主、自我复制的染色体外 DNA 分子，在前面用小写字母“p”表示和/或在后面用大小字母和/或数字表示。此处的起始质粒是可购得的，以不受限制的基础公共可得的，或可以根据公开的从可能的质粒构建。此外，与本领域公知的质粒等同的质粒对本领域技术人员是显而易见的。

“消化”DNA 是指用仅仅作用于 DNA 中某些序列的限制酶对 DNA 进行催化性切割。此处用到的各种限制酶是可购得的，所使用的它们的反应条件、辅因子和其它要求是本领域技术人员公知的。对于分析目的，一般在大约 20 μ l 缓冲液中使用 1 μ g 质粒或 DNA 片段和大约 2 单位酶。对于从质粒构建体中分离 DNA 片段的目的，一般在更大的体积中用 2-250 单位酶消化 5-50 μ g DNA 片段。用于特定限制酶的适当缓冲液和底物量是由制造商指定的。37 $^{\circ}$ C 下约 1 小时的温育时间是常用的，但可根据供应商的说明而改变。消化后在聚丙烯酰胺凝胶上直接电泳以分离所需片段。在此用本领域中使用的标准单字母符号(A, T, C, G, U)表示各种 DNA 和 RNA 片段中的核苷酸。

“多核苷酸”在本发明中可以是 RNA 形式或 DNA 形式，DNA 包括 cDNA、基因组 DNA 和合成 DNA。DNA 可以是双链或单链，如果是单链，则可以是编码链或非编码（反义）链。编码成熟多肽的编码序列可以与 SEQ ID NO: 6 所示的编码序列相同或可以是不同的编码序列，由于遗传密码的冗余或简并，所述编码序列编码与 SEQ ID NO: 7 相同的成熟多肽序列。

术语“编码多肽的多核苷酸”包括仅含多肽的编码序列的多核苷酸和含其它编码和/或非编码序列多核苷酸。

“寡核苷酸”是指单链多核苷酸或可以化学合成的两条互补的多核苷酸链。所述合成寡核苷酸没有 5'磷酸，因此只要不在激酶存在下通过 ATP 加入磷酸，就不会与另一寡核苷酸连接。合成寡核苷酸将连接于没有去磷酸化的片段。

“氨基酸残基”是指作为多肽的一部分的氨基酸。此处所述的氨基酸残基优选是“L”构型。然而，“D”同分异构形式的残基可取代任一个L-氨基酸残基，只要多肽还保留所需功能特性。NH₂指多肽氨基端的游离氨基，COOH指多肽羧基端的游离羧基。与标准多肽命名法(描述于 *J. Biol. Chem.*, 243: 3552-59(1969)并在 37 C.F.R. §§1.822 之下采用)一致，氨基酸残基的缩写见下表。

表 1
对应表
符号

1 字母	3 字母	氨基酸
Y	Tyr	酪氨酸
G	Gly	甘氨酸
F	Phe	苯丙氨酸
M	Met	甲硫氨酸
A	Ala	丙氨酸
S	Ser	丝氨酸
I	Ile	异亮氨酸
L	Leu	亮氨酸
T	Thr	苏氨酸
V	Val	缬氨酸
P	Pro	脯氨酸
K	Lys	赖氨酸
H	His	组氨酸
Q	Gln	谷氨酰胺
E	Glu	谷氨酸
Z	Glx	Glu 和/或 Gln
W	Trp	色氨酸
R	Arg	精氨酸
D	Asp	天冬氨酸
N	Asn	天冬酰胺
B	Asx	Asn 和/或 Asp

C	Cys	半胱氨酸
X	Xaa	未知或其它

此处用式子表示的所有氨基酸残基序列都是按常规的氨基末端到羧基末端的左到右方向。此外，短语“氨基酸残基”具有广泛的定义，包括表 1 所列氨基酸和修饰以及不常见的氨基酸，如 37 C.F.R. §§1.822 所提到的那些，在此引入作为参考。在氨基酸残基序列起点或末端的短线表示与其它序列或一个或多个氨基酸残基，或氨基末端基团如 NH₂，或羧基末端基团如 COOH 之间的肽键。

在肽或蛋白中，氨基酸的适当保守性取代是本领域技术人员公知的，并且一般可以不改变得到的分子的生物活性而制备。本领域的技术人员了解，概言之，在多肽的非关键区进行的单氨基酸取代基本不改变生物活性（见，例如，Watson et al. *Molecular Biology of the Gene*, 4th Edition, 1987, The Benjamin/Cummings Pub. Co., p. 224）。

所述取代优选根据表 2 所示进行：

表 2

原有残基	保守性取代
Ala (A)	Gly; Ser
Arg (R)	Lys
Asn (N)	Gln; His
Cys (C)	Ser
Gln (Q)	Asn
Glu (E)	Asp
Gly (G)	Ala; Pro
His (H)	Asn; Gln
Ile (I)	Leu; Val
Leu (L)	Ile; Val
Lys (K)	Arg; Gln; Glu
Met (M)	Leu; Tyr; Ile
Phe (F)	Met; Leu; Tyr
Ser (S)	Thr
Thr (T)	Ser
Trp (W)	Tyr
Tyr (Y)	Trp; Phe
Val (V)	Ile; Leu

其它取代也是允许的，并且可以根据经验确定或根据本领域的公知保守性取代而进行。

“补充性 (complementing) 质粒”描述了一种将核酸递送到包装细胞系，用于稳定整合到细胞基因组中的染色体中的质粒载体。

“递送质粒”表示携带或递送编码治疗性基因，或编码治疗性产物或其前体的基因，或调节基因或其它因子的核酸的质粒载体，所述核酸当体内递送或递送到细胞系，例如，但不限于包装细胞系以增殖治疗性病毒载体时，导致治疗作用。

此处描述了多种载体。例如，使用一种载体将特定的核酸分子递送到包装细胞系中，用于稳定整合到染色体中。这些类型的载体此处一般确定为补充性质粒。此处描述的另一种载体在细胞系（如包装细胞系）中携带或递送核酸分子或将核酸分子携带或递送到细胞系中，用于增殖治疗性病毒载体，这些载体此处一般称作递送载体。此处描述的第三种“类型”的载体用于携带编码治疗性蛋白或多肽或调节蛋白或作为有需要的受试者中特定细胞或细胞类型的调节序列的核酸分子；这些载体此处一般确定为治疗性病毒载体或重组腺病毒载体或病毒 Ad 来源的载体，并且以包裹含有用于表达治疗基因的表达盒的病毒核酸的病毒颗粒形式存在。

“DNA 或核酸同源物”是指包含预选的保守核苷酸序列，如编码治疗性多肽的序列的核酸。术语“基本同源的”表示具有至少 80%，优选至少 90%，最优选至少 95% 的同源性，或具有比上述低的同源性或同一性但具有保守的生物活性或功能。

术语“同源性”或“同一性”通常可以互换使用。在这一点上，可以通过例如用 GAP 计算机程序比较序列信息而确定同源性或同一性程度。GAP 程序利用 Needleman 和 Wunsch, *J. Mol. Biol.* 48:443 (1970) 的比对方法，该方法由 Smith 和 Waterman, *Adv. Appl. Math.* 2: 482 (1981) 修改。简言之，GAP 程序确定相似性为比对的符号（即核苷酸或氨基酸）中相似符号数目除以两个序列中较短的一个的符号总数。GAP 程序的优选缺省参数可以包括：(1) 一元比较矩阵（同一的值为 1，不同一的值为）以及 Gribskov 和 Burgess, *Nucl. Acids Res.* 14: 6745 (1986) 的加权比较矩阵，如 Schwartz 和 Dayhoff, eds., *Atlas of Protein Sequence and Structure*, National Biomedical

Research Foundation, pp. 353-358 (1979) 的描述; (2) 每个空位 3.0 的罚分和每个空位中每个符号 0.10 的罚分; 以及 (3) 末端空位无罚分。可以使用如 Pearson 和 Lipman, Proc. Natl. Acad. Sci. USA 85: 2444 (1988) 中的缺省参数, 用公知的计算机算法如 "FAST A" 程序确定两个核酸分子是否具有至少 80%, 85%, 90%, 95%, 96%, 97%, 98% 或 99% 同一的核苷酸序列。或者可以使用国家生物信息数据库的 BLAST 函数确定同一性。“同一性”本身具有本领域公知的含义并且可以用公知技术计算 (见, 例如 Computational Molecular Biology, Lesk, A. M., ed., Oxford University Press, New York, (1988); Smith, D. W., ed., Biocomputing: Informatics and Genome Projects, Academic Press, New York, (1993); Griffin, A. M., and Griffin, H. G., eds., Computer Analysis of Sequence, Data, Part I, Humana Press, New Jersey, (1994); von Heinje, G., Sequence Analysis in Molecular Biology, Academic Press, (1987); 和 Gribskov, M. and Devereux, J., eds., Sequence Analysis Primer, M Stockton Press, New York, (1991))。尽管存在许多方法测量两个多核苷酸或多肽序列之间的同一性, 术语“同一性”是本领域技术人员公知的 (Carillo, H. & Lipton, D., SIAM J. Applied Math. 48: 1073 (1988))。通常用于确定两个序列之间的同一性或相似性的方法包括, 但不限于公开于 Martin J. Bishop, ed., Guide to Huge Computers, Academic Press, San Diego, (1994), 和 Carillo, H. & Lipton, D., SIAM J. Applied Math. 48: 1073 (1988) 中的方法。确定同一性和相似性的方法汇编于计算机程序中。用于确定两个序列之间的同一性和相似性的优选计算机程序方法包括, 但不限于 GCG 程序包 (Devereux, J., et al., Nucleic Acids Research 12(I): 387 (1984)), BLASTP, BLASTN, PASTA (Atschul, S. F., et al., J. Molec. Biol. 215: 403 (1990))。

术语“同一性”表示测试和参考多肽或多核苷酸之间的比较。例如, 测试多肽可以确定为任何 90% 或更多同一于参考多肽的任何多肽。此处用到的术语“90% 同一于”是指相对于参考多肽为 90-99.99% 的百分比同一性。90% 或更高水平的同一性表示, 假定比较了测试和参考多肽 100 个氨基酸的长度。测试多肽中不超过 10% (即 100 个中的 10 个)

的氨基酸与参考多肽不同。可以在测试和参考多核苷酸之间进行相似的比较。这种不同可以表示为在氨基酸序列全长上随机分布的点突变，或者它们可以在一个或多个不同长度的位置上成簇，直到允许的最大数目，如 10/100 个氨基酸不同（约 90% 同一性）。不同定义为核酸或氨基酸取代或缺失。

术语“基因治疗”是指将异源 DNA 转移到患有所述治疗针对的疾病或状况的哺乳动物，特别是人的某些细胞、靶细胞。以某种方式将 DNA 导入选定的靶细胞中，使得异源 DNA 表达，并且使其编码的治疗性产物产生。或者，异源 DNA 可以某种方式介导编码治疗性产物的 DNA 表达，它可以编码一种产物，例如肽或 RNA，它们以某种方式直接或间接介导治疗产物的表达。基因治疗也可以用于编码基因产物的核酸，替代缺陷基因或补充导入的哺乳动物或细胞中基因产物的产生。导入的核酸可以编码治疗性化合物，如生长因子抑制剂，或肿瘤坏死因子或其抑制剂，如其受体，正常情况下它们在哺乳动物宿主中不产生或不以治疗有效量产生或不在治疗有用的时间产生。编码治疗性产物的异源 DNA 可以在导入患病宿主的细胞之前进行修饰，以便增强或改变产物或其表达。

“异源 DNA”是编码正常情况下在其所表达的细胞中不体内产生的 RNA 和蛋白的 DNA，或通过影响转录、翻译或其它可调节的生化过程而介导内源 DNA 表达或编码介导物的 DNA。异源 DNA 也可以称作外源 DNA。本领域技术人员所了解或认为对表达 DNA 的细胞来说是异源或外源的任何 DNA 在此处都包括在异源 DNA 的范围中。异源 DNA 的例子包括，但不限于编码可追踪标记蛋白，如赋予药物抗性的蛋白的 DNA，编码治疗有效的物质，如抗癌剂、酶和激素的 DNA，和编码其它类型的蛋白，如抗体的 DNA。由异源 DNA 编码的抗体可以被分泌或表达在异源 DNA 所导入的细胞表面。此处，“异源 DNA”或“外源 DNA”是指与存在于相应的野生型腺病毒中的对应 DNA 分子相比没有精确的方向和位置的 DNA 分子。它也可以称作来自于另一生物体或物种（即外源）或来自另一 Ad 血清型的 DNA 分子。

“治疗有效的 DNA 产物”是指一种由 DNA 编码的产物，当将 DNA 导入宿主时，表达一种有效改善或消除症状，即先天或获得性疾病的表现或治愈所述疾病的产物。一般地，编码所需异源 DNA 的 DNA 被克

隆到质粒载体中，通过常规方法，如磷酸钙介导的 DNA 摄取或微注射而导入生产细胞，如包装细胞。在生产细胞中扩增后，含有异源 DNA 的载体被导入选定的靶细胞。

“表达或递送载体”是指外源或异源 DNA 可以插入其中用于在适当宿主细胞中表达的任何质粒或病毒，即由 DNA 编码的蛋白或多肽在宿主细胞系统中合成。能够指导编码一种或多种蛋白的 DNA 片段（基因）表达的载体此处称作“表达载体”。它也包括允许使用逆转录酶从 mRNA 克隆 cDNA（互补 DNA）的载体。

“基因”是一种核酸分子，其核苷酸序列编码 RNA 或多肽。基因可以是 RNA 或 DNA。基因可以包括在编码区前面和后面的区（前导和尾）以及各个编码片段（外显子）之间的间插序列（内含子）。

用于核酸分子、多肽、或其它生物分子的“分离的”表示该核酸或多肽已经从获得多肽或核酸的遗传环境中分开。它也可以表示从天然状态发生了改变。例如，如这里所使用的该术语，天然存在于活动物中的多核苷酸或多肽不是“分离的”，但从其天然状态的共存物质中分开的相同多核苷酸或多肽是“分离的”。因此，在重组宿主细胞中产生和/或包含在其中的多肽或多核苷酸被认为是分离的。“分离的多肽”或“分离的多核苷酸”也包括从重组宿主细胞或天然来源部分或基本纯化的多肽或多核苷酸。例如，重组产生的化合物形式可以通过 Smith 和 Johnson, *Gene* 67: 31-40 (1988) 描述的一步法基本纯化。术语“分离的”和“纯化的”有时互换使用。所述多核苷酸可以是载体的一部分，和/或所述多核苷酸或多肽可以是组合物的一部分但仍然是分离的，因为所述载体或组合物不是其天然环境的一部分。

“分离的多核苷酸”表示该核酸不含在天然存在的生物体基因组中（如果有）位于编码目的核酸的基因侧翼的基因的编码序列。分离的 DNA 可以是单链或双链的，可以是基因组 DNA, cDNA, 重组杂交 DNA, 或合成 DNA。它可以与天然 DNA 序列相同，或者通过一个或多个核苷酸的缺失、添加或取代而与其不同。

当提到从生物细胞或宿主制备的制备物时，“分离的”或“纯化的”表示任何含有所提到的 DNA 或蛋白的细胞提取物，包括目的 DNA 或蛋白的粗提取物。例如，对于蛋白，可以根据各个技术或一系列制备或生化技术获得纯化制备物，目的 DNA 或蛋白可以在这些制备物中

以各种纯度存在。这些程序可以包括，例如，但不限于，硫酸铵分级分离、凝胶过滤、离子交换层析、亲和层析、密度梯度离心和电泳。

“基本纯的”或“分离的”DNA或蛋白制备物表示不含天然存在物质的制备物，所述DNA或蛋白在天然状态下通常与所述物质缔合。“基本纯的”应该理解为表示“高度”纯化的制备物，其含有至少95%的DNA或目的蛋白。

“包装细胞系”是一种提供缺失的基因产物或其等同物的细胞系。

“腺病毒颗粒”是病毒的最小结构或功能单位。病毒可以指单颗粒，颗粒的集合或病毒基因组。腺病毒(Ad)颗粒是相对复合的，可以拆分成多个亚结构。

“转录后调节元件(PRE)”是一种存在于未剪接的病毒或细胞信使RNA，即不含内含子的信息中的调节元件。示例包括，但不限于人肝炎病毒、土拨鼠肝炎病毒、TK基因和小鼠组蛋白基因。PRE可以置于聚腺苷酸序列之前或异源DNA序列之后。

“假型”描述了产生具有修饰的衣壳蛋白或具有来自与载体本身血清型不同的血清型的衣壳蛋白的腺病毒载体。一个例子是产生含Ad38纤维蛋白的腺病毒载体颗粒。这可以通过在表达不同纤维蛋白的包装细胞系中产生腺病毒载体而完成。

“目的启动子”可以是诱导型或组成型。诱导型启动子将仅仅在其它分子存在下才起始转录；组成型启动子不需要任何其它分子就可以调节基因表达。可调节或诱导型启动子也可以描述为RNA聚合酶结合和起始的速度或程度是由外源刺激调节的启动子。所述刺激包括，但不限于多种化合物或组合物、光、热、压力和化学能量可调节的启动子。可诱导的、可抑制的和可阻遏的启动子被认为是可调节的启动子。此处优选的启动子是在眼细胞，特别是光受体细胞选择性表达的启动子。

“受体”是指特异性结合于其它分子的生物活性分子。术语“受体蛋白”可以更具体用于表示特异性受体的蛋白性质。

“重组的”是指由基因工程形成的任何后代。它可以用于描述通过质粒在包装细胞中的重组形成的病毒。

“转基因”或“治疗性核酸分子”包括编码RNA或多肽的DNA和

RNA 分子。这些分子可以是“天然的”或天然来源的序列；它们也可以是“非天然的”或“外源的”，即天然或重组来源的。此处可以与术语“治疗性核酸分子”互换使用的术语“转基因”通常用于描述由病毒载体携带并转导到宿主细胞中的异源或外来（外源）基因。治疗性核酸分子包括反义序列或可以转录为反义序列的核苷酸序列。所有的治疗性核苷酸序列（或转基因）都包括在所述治疗性序列所递送到的细胞或细胞核中产生所需作用的核酸分子。例如，治疗性核酸分子可以包括编码用于递送到不能产生功能性蛋白的细胞中的功能性蛋白的核苷酸序列。

“眼的玻璃体”是指填充眼晶体后腔室的物质（即房水或玻璃体）。

“启动子区”是指基因的 DNA 的一部分，它控制与其可操作性连接的 DNA 的转录。启动子区包括 RNA 聚合酶识别、结合和转录起始所足够的 DNA 的特定序列。启动子区的该部分称作启动子。此外，启动子区包括调节 RNA 聚合酶的该识别、结合和转录起始活性的区。这些序列可以顺式作用或可以反应于反式作用因子。取决于调节的性质，启动子可以是组成型的或受调节的。

“可操作性连接的”表示序列或片段已经共价连接到一段单链或双链 DNA 片段中，从而一个片段上的控制序列可以控制表达或复制，或对其它片段进行这些控制。然而，两个片段并不必须是连续的。

“包装”是指能够将本发明的多肽、多克隆抗体、或单克隆抗体置于其中进行固定的固体基质或物质，如玻璃、塑料（如聚乙烯、聚丙烯或聚碳酸酯）、纸、箔等。因此，例如，包装可以是用于容纳毫克量的目的多肽的玻璃管瓶或它可以使微克量的目的多肽或抗体可操作性粘附（即连接的）从而能够分别免疫结合于抗体或抗原的微量滴定板孔。

“使用说明”一般包括描述试剂浓度或至少一种测定方法的参数，如待混合的试剂和样品的相对量、试剂样品混合的维持时间、温度、缓冲条件等的有形表达。

本发明上下文中的“诊断系统”也包括能够传递含有本发明的多肽或抗体分子的免疫复合物的形成的信号的标记或指示工具。

此处使用的“复合物”是指特异性结合反应如抗体-抗原或受体-

配体反应的产物。示例性的复合物是免疫反应产物。

以其各种语法形式使用的“标记”和“指示工具”表示直接或间接参与可检测信号的产生以指示复合物存在的单个原子和分子。任何标记或指示装置可以连接于或掺入作为本发明的抗体或单克隆抗体组合物的一部分的被表达蛋白、多肽或抗体分子，或独立使用，所述原子或分子可以单独使用或与其它试剂结合使用。这些标记本身是临床诊断化学公知的并且组成本发明的一部分，其程度为它们被其它新蛋白方法和/或系统使用。

讨论

图1所示多肽是 SEQ ID NO: 1 的氨基酸残基 94-471 (如 SEQ ID NO: 12)，以及具有 SEQ ID NO: 7 的氨基酸序列的多肽，或由 SEQ ID NO: 6 的 cDNA 编码的多肽，它们构成本发明的部分。除了上述多肽的变体，本发明还包括多核苷酸的变体。包括的多核苷酸变体可以是天然存在的多核苷酸等位变体或非天然存在的多核苷酸变体。因此，本发明包括编码如 SEQ ID NO: 7, SEQ ID NO: 12 所示多肽的多核苷酸，和由 SEQ ID NO: 6 的 cDNA 编码的多肽，以及所述多核苷酸的变体，该变体编码 SEQ ID NO: 7 和 SEQ ID NO: 12 的血管发生抑制性片段、衍生物或类似物。所述核苷酸变体包括缺失变体、取代变体和添加或插入变体。

如上文所述，多核苷酸可以具有作为 SEQ ID NO: 6 所示编码序列的等位变体的编码序列。如本领域所知，等位变体是具有一个或多个核苷酸的取代、缺失或添加的多核苷酸序列，它基本不改变所编码的多肽的功能。

此处用到的术语 T1 是指具有 SEQ ID NO: 13 的氨基酸序列的多肽和具有 SEQ ID NO: 5 的氨基酸序列的含 His₆ 标记的多肽。此处使用的术语 T2 是指具有 SEQ ID NO: 12 的氨基酸序列的多肽和具有 SEQ ID NO: 7 的氨基酸序列的含 His₆ 标记的多肽。此处用到的术语 TrpRS 是指具有 SEQ ID NO: 1 的残基 1-471 的氨基酸序列和具有 SEQ ID NO: 1 的氨基酸序列的含 His₆ 标记的多肽。

本发明还包括多核苷酸，其中成熟多肽的编码序列可以与辅助宿主细胞表达和分泌多肽的多核苷酸序列，例如作为控制多肽从细胞运送的分泌序列的前导序列在同一阅读框架中融合。具有前导序列的多

肽是一种前蛋白并且可以具有由宿主细胞切除以形成多肽的成熟形式的前导序列。多核苷酸也可以编码一种前蛋白，即成熟蛋白加额外的5'氨基酸残基。具有前序列(prosequence)的成熟蛋白是一种前蛋白并且是蛋白的无活性形式。一旦前序列被切除，则保留了活性成熟蛋白。

因此，例如，本发明的多核苷酸可以编码成熟蛋白，或编码具有前序列的蛋白，或编码具有前序列和前置序列(presequence)(前导序列)的蛋白。

本发明的多核苷酸也可以具有框内融合于允许本发明的多肽纯化的标记序列的编码序列。在细菌宿主的情况下，标记序列可以是由pQE-9载体提供的六组氨酸标记，从而提供融合于标记的成熟多核苷酸的纯化，或，当使用哺乳动物宿主如COS-7细胞时，标记序列可以是例如血凝素(HA)标记。HA标记相当于来源于流感血凝素蛋白的表位(Wilson, I., et al., Cell, 37: 767(1984))。

本发明进一步涉及与上文描述的序列杂交的多核苷酸，如果序列间存在至少50%，优选70%的同一性。本发明具体涉及在严格条件下与上文描述的多核苷酸杂交的多核苷酸。此处使用的术语“严格性条件”表示仅仅在序列间具有至少95%，优选至少97%的同一性时才发生杂交。与上文描述的多核苷酸杂交的多核苷酸在优选实施方式中编码保留与SEQ ID NO: 6的cDNA编码的成熟多肽基本相同的生物功能或活性的多肽。

当提到SEQ ID NO: 7, SEQ ID NO: 12的多肽或由SEQ ID NO: 6的多核苷酸编码的多肽时，术语“片段”、“衍生物”和“类似物”表示保留了与所述多肽基本相同的血管抑制(即血管发生抑制)功能或活性的多肽部分。因此，“类似物”包括可以通过前蛋白部分的切除而激活，从而产生血管抑制活性的成熟多肽的前蛋白。

本发明的多肽可以是重组多肽、天然多肽或合成多肽，优选重组多肽。

SEQ ID NO: 7, SEQ ID NO: 12的血管发生抑制性片段、衍生物或类似物，或由SEQ ID NO: 6的多核苷酸编码的多肽可以是(i)其中的一个或多个氨基酸残基被保守或非保守氨基酸残基(优选保守氨基酸残基)取代，并且所述被取代的氨基酸残基可以或不可以是由遗传密码编码的氨基酸残基，或(ii)其中的一个或多个氨基酸残基包含

取代基，或 (iii) 其中的多肽与另一种化合物，如增加多肽半衰期的化合物（例如聚乙二醇）融合，或 (iv) 其中额外的氨基酸与多肽融合，如前导或分泌序列或用于纯化多肽或前蛋白序列的序列。根据此处的教导，所述片段、衍生物和类似物被认为是在本领域技术人员知识范围内。

本发明的多肽和多核苷酸优选以分离形式提供，优选纯化为均质。

本发明也包括含有本发明的多核苷酸的载体，用本发明的载体基因工程化的宿主细胞和通过重组技术产生本发明的多肽。

宿主细胞是用本发明的载体基因工程化的（转导或转化或转染），该载体可以是克隆载体或表达载体。载体可以是，例如，质粒、病毒颗粒、噬菌体等形式。工程化宿主细胞可以在经改进而适于激活启动子、选择转化体或扩增 tRNA 合成酶多肽基因的常规营养培养基中培养。培养条件，如温度、pH 等是以前用于选择宿主细胞进行表达的条件，对于本领域技术人员是显而易见的。

本发明的多核苷酸可以用于通过重组技术产生相应的多肽。因此，例如，多核苷酸序列可以包含任何表达载体中，特别是用于表达多肽的载体或质粒。所述载体包括染色体、非染色体和合成 DNA 序列，如 SV40 的衍生物；细菌质粒；噬菌体 DNA；酵母质粒；来源于质粒和噬菌体 DNA、病毒如牛痘病毒、腺病毒、禽痘病毒和假狂犬病病毒 DNA 的组合的载体。优选的载体是 pET20b。然而，只要在宿主中可以复制和存活，可以使用任何其它的质粒或载体。

如上所述，适当的 DNA 序列可以通过多种程序插入载体中。概言之，通过本领域公知的程序将 DNA 序列插入适当的限制性内切酶位点。所述程序和其它程序被认为是在本领域技术人员知识范围之内。

表达载体中的 DNA 序列与适当的表达控制序列（启动子）可操作性连接，以指导 mRNA 合成。所示启动子的代表性示例包括 LTR 或 SV40 启动子、大肠杆菌 lac 或 trp 启动子、噬菌体 λ P_L 启动子和已知控制基因在原核或真核细胞或它们的病毒中表达的其它启动子。表达载体也包含用于翻译起始的核糖体结合位点和转录终止子。载体也可以包含用于扩增表达的适当序列。

此外，表达载体优选包含提供用于选择转化的宿主细胞的表型性

状如二氢叶酸还原酶或真核细胞培养物的新霉素抗性，或大肠杆菌四环素或氨基青霉素抗性的基因。

含有上述适当 DNA 序列和适当启动子或控制序列的载体可以用于转化适当宿主以允许该宿主表达蛋白。适当宿主的代表性示例包括细菌细胞，如大肠杆菌、伤寒沙门氏菌、链霉菌；真菌细胞如酵母；昆虫细胞，如果蝇和 Sf9；动物细胞如 CHO、COS 或 Bowes 黑色素瘤；植物细胞等。根据此处的教导，适当宿主的选择被认为是本领域技术人员知识范围内的。

更具体地，本发明也包括含有一种或多种上述广泛定义的序列的重组构建体。该构建体包括载体，如质粒或病毒载体，其中以正向或反向插入了本发明的序列。在本实施方式的优选方面，该构建体包含调节序列，包括，例如，与序列可操作性连接的启动子。大量适当的载体和启动子是本领域技术人员公知的，并且是可购得的。通过举例提供以下载体：细菌：pQE70，pQE-9 (Qiagen)，pBs，phagescript，PsiX174，pBluescript SK，pBsKS，pNH8a，pNH16a，pNH18a，pNH46a (Stratagene)；pTrc99A，pKK223-3，pKK233-3，pDR540，PRIT5 (Pharmacia)。真核：pWLneo，pSV2cat，pOG44，pXT1，pSG (Stratagene) pSVK3，pBPV，PMSG，pSVL (Pharmacia) 和 pET20B。在一种优选实施方式中，载体是 pET20B。然而，可以使用任何其它质粒或载体，只要它们在宿主中可复制并且可存活。

可以使用 CAT (氯霉素转移酶) 载体或具有可选择标记的其它载体从任何需要的基因选择启动子区。两种适当的载体是 pKK232-8 和 pCM7。特别命名的细菌启动子包括 lacI，lacZ，T3，T7，gpt， λ Pr，PL 和 trp。真核启动子包括 CMV 立即早期、HSV 胸苷激酶、早期和晚期 SV40、逆转录病毒的 LTRs、和小鼠金属硫蛋白-I。适当载体和启动子的选择在本领域普通技术人员的水平之内。

在另一实施方案中，本发明涉及含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，或宿主细胞可以是真核细胞，如细菌细胞。将构建体导入宿主细胞可以通过磷酸钙转染、DEAE-介导的转染或电穿孔而实现 (Davis, L., Diber, M., Battey, I., Basic Methods in Molecular Biology, 1986)。

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。或者，本发明的多肽可以通过常规肽合成仪合成产生。

蛋白可以在适当启动子的控制下，在哺乳动物细胞、酵母、细菌或其它细胞中表达。也可以利用无细胞的翻译系统，使用本发明的 DNA 构建体来源的 RNA 产生所述蛋白。用于原核和真核宿主的适当克隆和表达载体描述于 Sambrook, et al., *Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor, N.Y., (1989), 其公开内容在此引入作为参考。

通过将增强子序列插入载体中可以增加编码本发明的多肽的 DNA 的转录。增强子是 DNA 的顺式作用元件，通常约 10-约 300 个碱基对 (bp)，它作用于启动子以增加其转录。其示例包括复制起点晚期侧的 SV40 增强子 (bp100-270)、巨细胞病毒早期启动子增强子、复制起点晚期侧的多瘤病毒增强子、和腺病毒增强子。

一般地，重组表达载体将包括复制起点和允许宿主细胞转化的可选择标记，如大肠杆菌氨苄青霉素抗性基因和啤酒糖酵母 TRP1 基因，以及来源于高表达基因，用于指导下游结构序列转录的启动子。所述启动子可以来源于编码糖分解酶如 3-磷酸甘油激酶 (PGK)、 α 因子、酸磷酸酶、或热激蛋白等的操纵子。异源结构序列与转录起点和终止序列一起组装在适当噬菌体中，优选地，与能够指导翻译后蛋白分泌到周质间隙和胞外基质的前导序列一起组装。任选地，异源序列可以编码包含赋予所需特征，如表达的重组产物的稳定化或简单纯化的 N 端鉴定肽。

在转化适当的宿主株和使宿主株生长到适当的细胞密度后，通过适当方法（如温度改变或化学诱导）对所选择的启动子去阻遏，培养细胞一段额外的时期。

一般通过离心、物理或化学方法破坏而收获细胞，保留得到的粗提取物用于进一步纯化。

在蛋白表达中使用的微生物细胞可以通过常规方法破坏，包括冻融循环、超声处理、机械破坏或使用细胞裂解剂。

也可以使用多种哺乳动物细胞培养系统表达重组蛋白。哺乳动物表达系统的例子包括如 Gluzman, *Cell*, 23: 175 (1981) 的描述小鼠肾成纤维细胞的 COS-7 系，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、适当启动子和增强子, 以及任何必要的核糖体结合位点、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。来源于 SV40 病毒基因组的 DNA 序列, 例如, SV40 起点、早期启动子、增强子、剪接和聚腺苷酸化位点也可以用于提供需要的非转录遗传元件。

通过以前使用的方法, 包括硫酸铵或乙醇沉淀、酸提取、阴离子或阳离子交换层析、磷酸纤维素层析、疏水相互作用层析、亲和层析、羟磷灰石层析和凝集素层析从重组细胞培养物回收和纯化多肽。优选在纯化中使用低浓度(约 0.1-5 mM)钙离子(Price, et al. J. Biol. Chem., 244: 917(1969))。如果必要, 可以使用蛋白再折叠步骤, 从而是成熟蛋白的构型完整。最后, 可以使用高效液相层析(HPLC)进行最终的纯化步骤。

本发明的多肽可以是天然纯化产物, 或化学合成程序的产物, 或通过重组技术从原核或真核宿主(例如培养的细菌、酵母、高等植物、昆虫和哺乳动物细胞)产生。根据重组生产程序, 本发明的多肽可以用哺乳动物或其它真核生物的碳水化合物糖基化的, 或者可以是非糖基化的。

本发明的多肽可以经修饰而改进稳定性并通过本领域公知的方法增加效力。例如, L-氨基酸可以被 D-氨基酸取代, 氨基端可以乙酰化, 或羧基端可以修饰, 如用乙胺封端(Dawson, D.W., et al., Mol. Pharmacol., 55: 332-338(1999))。

根据本发明, 本发明的多肽也可以通过体内表达所述多肽而用作基因治疗。

此处教导的多种可以用于基因治疗的病毒载体包括腺病毒、肝炎病毒、牛痘病毒、腺伴随病毒(AAV)或优选 RNA 病毒如逆转录病毒。优选地, 逆转录病毒是鼠或鸟类逆转录病毒衍生物, 或是慢病毒载体。优选的逆转录病毒载体是慢病毒载体。可以插入单个外源基因的逆转录病毒载体的例子包括, 但不限于: 莫洛尼鼠白血病毒(MoMuLV), Harvey 鼠肉瘤病毒(HaMuSV)、鼠乳腺肿瘤病毒(MuMTV), SIV, BIV, HIV 和劳斯肉瘤病毒(RSV)。许多额外的逆转录病毒载体可以掺入多个基因。所有这些载体都可以转移或掺入选择性标记的基因, 使得可以

鉴定和产生被转导的细胞。通过将感兴趣的锌指来源的 DNA 结合多肽序列与另一编码特定靶细胞上受体的配体的基因一起掺入病毒载体，例如，将载体制备为靶特异性的。可以通过插入例如编码蛋白的多核苷酸可以将逆转录病毒载体制备为靶特异性的。优选的导向是通过使用导向于逆转录病毒载体的抗体而实现的。本领域技术人员将了解或能够容易确定可以插入逆转录病毒基因组从而允许含有锌指-核苷酸结合蛋白多核苷酸的逆转录病毒载体的靶特异性递送的特定多核苷酸序列，而不需要过多的实验。

由于重组逆转录病毒是缺陷的，为产生感染性载体颗粒，它们需要辅助。例如，可以通过使用含编码 LTR 内的调节序列控制下的所有逆转录病毒结构基因的质粒的辅助细胞系而提供该辅助。这些质粒缺失了使得包装机制识别用于包衣壳作用的 RNA 转录物的核苷酸序列。缺失包装信号的辅助细胞系包括，但不限于，例如 $\phi 2$, PA317 和 PA 12。这些细胞系产生空毒粒，因为没有包装基因组。如果将逆转录病毒载体导入包装信号完整，但结构基因被其它目的基因替代的细胞中，载体可以被包装并且可以产生载体毒粒。然后可以用由该方法产生的载体毒粒感染组织细胞系，如 NIH 3T3 细胞系，产生大量嵌合逆转录病毒毒粒。

另一种用于编码锌指来源的 DNA 结合多肽的多核苷酸的导向递送系统是胶体分散系统。胶体分散系统包括大分子复合物、纳米胶囊、微球、珠和基于脂质的系统，包括水包油乳液、胶束、混合胶束和脂质体。本发明的优选胶体系统是脂质体。脂质体是用于体内和体外递送载体的人工膜液泡。已经发现大小为 0.2-0.4 μm 的大的单层液泡 (LUV) 可以包裹很大比例的含大分子的水性缓冲液。RNA、DNA 和完整的毒粒可以包裹在水内部，并以生物活性形式递送到细胞 (Fraley, et al., Trends Biochem. Sci., 6: 77, (1981))。除哺乳动物细胞外，已经用脂质体在植物、酵母和细菌细胞中递送多核苷酸。为了使脂质体成为有效的基因转移载体，应该存在以下特征：(1) 以高效率包裹基因，而不损害其生物功能；(2) 与非靶细胞相比，优先和大量结合于靶细胞；(3) 以高效率将液泡中的水性成分递送到靶细胞的细胞质；和 (4) 准确和有效表达遗传信息 (Mannino, et al., Biotechniques, 6: 682, (1988))。

脂质体的组成通常是磷脂，特别是高相转变温度磷脂的组合，通常与类固醇，特别是胆固醇组合。也可以使用其它磷脂或其它脂质。脂质体的物理特征取决于 pH、离子强度和二价阳离子的存在。

用于脂质体制备的脂质的例子包括磷脂酰化合物，如磷脂酰甘油、磷脂酰胆碱、磷脂酰丝氨酸、磷脂酰乙醇胺、鞘磷脂、脑苷脂、神经节苷脂。特别有用的是二酰基磷脂酰甘油，其中脂部分含有 14-18 个碳原子，特别是 16-18 个碳原子，并且是饱和的。说明性的磷脂包括卵磷脂酰胆碱、二棕榈酰磷脂酰胆碱和二硬脂酰磷脂酰胆碱。

脂质体导向的分类是基于解剖和机械因素。解剖分类是基于选择性的水平，例如器官特异性、细胞特异性和细胞器特异性。机械性导向可以根据是被动还是主动而区分。被动的导向使用脂质体的天然倾向将网状内皮系统 (RES) 的细胞分布于含有毛细血管窦的器官。另一方面，主动导向包括通过将脂质体偶联于特异性配体如单克隆抗体、糖、糖脂或蛋白，或通过改变脂质体的组成或大小而改变脂质体，从而完成导向于天然存在的定位位点之外器官和细胞类型。

可以通过多种途径修饰被导向的递送系统的表面。对于脂质体导向的递送系统，可以将脂基团掺入脂质体的脂双侧从而维持导向配体与脂质体双侧的稳定缔合。可以用多种连接基团将脂质链连接于导向配体。

概言之，结合于导向递送系统表面的化合物将是允许导向递送系统寻找所需细胞和在其上“归巢”的配体和受体。配体可以是结合于另一种化合物，如受体的任何目的化合物。

概言之，结合于特异性效应分子的表面膜蛋白称作受体。在本发明中，抗体是优选的受体。抗体可以用于将脂质体导向于特异性细胞表面配体。例如，某些抗原在肿瘤细胞上特异性表达，称作肿瘤相关抗原 (TTAs)，可以用于将含抗体-锌指-核苷酸结合蛋白的脂质体直接导向于恶性肿瘤。由于可以根据其作用的细胞类型区分锌指-核苷酸结合蛋白基因产物，定向递送系统提供了与随机注射的非特异性脂质体相比的显著改进。许多程序可以用于将多克隆抗体或单克隆抗体连接于脂质体双层。抗体导向的脂质体可以包括单克隆或多克隆抗体或其片段，如 Fab，或 $F(ab')_2$ ，只要它们有效结合于靶细胞上的抗原性表位。脂质体也可以导向于表达激素或其它血清因子的受体的细胞。

本领域技术人员可以获得适于将核酸导入靶细胞的多种病毒和非病毒方法。原发性肿瘤细胞的遗传操作是本领域公知的。可以使用基因治疗领域公知的一种或多种技术完成细胞的遗传修饰 (Mulligan, R. C. *Human Gene Therapy*, 5(4): 543-563 (1993))。病毒转导方法可以包括使用包含驱动或抑制具有唾液酰转移酶活性的蛋白表达的核酸序列的重组 DNA 或 RNA 病毒感染靶细胞。用于本发明的适当 DNA 病毒包括, 但不限于腺病毒 (Ad)、腺伴随病毒 (AAV)、肝炎病毒、牛痘病毒或脊髓灰质炎病毒。用于本发明的适当 RNA 病毒包括, 但不限于逆转录病毒或新培斯病毒。本领域技术人员也可以理解, 存在的一些所述 DNA 和 RNA 病毒可以适用于本发明。

腺病毒载体可用于将基因转移到真核细胞, 以研究真核基因表达, 用于疫苗开发, 和用于动物模型。Ad 介导的基因治疗也已经用于人, 如用于将囊性纤维化跨膜传导调节物 (CFTR) 基因转移到肺。将重组 Ad 给予不同体内组织的途径包括, 例如气管内滴注、肌肉注射、外周静脉注射和立体定位接种到脑。然后, 本领域技术人员可以广泛获得腺病毒载体, 并且适用于本发明。

最近已经将腺伴随病毒 (AAV) 导入作为基因转移系统, 在基因治疗中有潜在应用。已经报道了野生型 AAV 表现出高水平感染性、广泛的宿主范围和整合到宿主细胞基因组中的特异性。1 型单纯疱疹病毒 (HSV-1) 作为载体系统是有吸引力的, 特别是用于神经系统, 因为它具有嗜神经特性。痘病毒家族的牛痘病毒也已经开发用作表达载体。上述载体中的每一种对于本领域技术人员都是可以广泛获得的, 并且适用于本发明。

逆转录病毒能够感染大比例的靶细胞并整合到细胞基因组中。逆转录病毒相对其它病毒较早被开发为基因转移载体, 并且首先成功用于基因标记和将腺苷脱氨基酶 (ADA) cDNA 转导到人淋巴细胞中。优选的逆转录病毒包括慢病毒。在优选实施方案中, 逆转录病毒选自 HIV, BIV 和 SIV。

已经用于或提出用于基因治疗的“非病毒”递送技术包括 DNA-配体复合物、腺病毒-配体-DNA 复合物、DNA 的直接注射、磷酸钙沉淀、基因枪技术、电穿孔、脂质体和脂转染。任何上述方法对本领域技术人员来说都是可广泛获得的, 并且适用于本发明。本领域技术人员可

以获得其它适当方法，应该理解，可以使用任何可得到的转染方法完成本发明。本领域技术人员已经利用了几种所述方法，获得了不同程度的成功。可以通过将分离的 DNA 分子包裹在脂质体颗粒中并使脂质体颗粒与靶细胞膜接触而完成脂转染。脂质体是自我组装的胶体颗粒，其中由磷脂酰丝氨酸或磷脂酰胆碱等两亲性分子组成的脂双层包裹了一部分周围的介质，使得脂双层围绕了亲水性内容物。可以构建单层或多层的脂质体，使得内容物含有所需的化学物质、药物，或如本发明，含有分离的 DNA 分子。

细胞可以体内、离体或体外转染。细胞可以作为从患者分离的原代细胞或衍生自原代细胞的细胞系进行转染，并且对于最终给予该细胞的患者来说，并不必须是自体的。离体或体外转染后，细胞可以植入宿主。为了在靶细胞中获得本发明的核酸的转录，使用了能够在靶细胞内驱动基因表达的转录调节区。转录调节区可以包含启动子、增强子、沉默子或阻遏子元件，并且与本发明的核酸功能上相关。优选地，适用于本发明的转录调节区包括但不限于人巨细胞病毒 (CMV) 介导的立即早期增强子/启动子、SV40 早期增强子/启动子、JC 多瘤病毒启动子、白蛋白启动子、PGK 和与 CMV 增强子偶联的 α -肌动蛋白启动子。

可以使用本领域技术人员可广泛获得的标准重组技术构建本发明的载体。所述技术可以参见普通分子生物学参考文献如 Sambrook, et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press (1989), D. Goeddel, ed., *Gene Expression Technology, Methods in Enzymology series, Vol. 185*, Academic Press, San Diego, CA (1991), 和 Innis, et al. *PCR Protocols: A Guide to Methods and Applications* Academic Press, San Diego, CA (1990)。

将本发明的多肽或核酸体内给予靶细胞可以使用本领域技术人员公知的多种技术完成。

本发明的载体可以以含有常规药物可接受载体、佐剂和赋形剂的剂量单位制剂进行口服、肠胃外、通过吸入喷雾、直肠或局部给药。此处使用的术语肠胃外包括皮下、静脉、肌内、胸骨内、输注技术或腹膜内给药。可以通过将药物与常温下为固体，但在直肠温度下为液

体，因此将在直肠中融化并释放药物的可可油和聚乙二醇等适当非刺激性赋形剂相混合而制备用于直肠给药药物的栓剂。

用本发明的载体和/或本发明的组合物治疗紊乱或疾病的剂量方案是以多种因素为基础的，包括疾病类型、年龄、体重、性别、患者的医学状况、状况的严重性、给药途径和使用的具体化合物。因此，剂量方案可以很大程度改变，但可以使用标准方法常规确定。

本发明的药物活性化合物可以根据药学的常规方法加工，以产生给予患者，包括人和哺乳动物的药物组合物。对于口服给药，药物组合物可以是例如液体、眼部嵌入物、胶囊、片剂、悬浮液。药物组合物优选以含有一定量活性试剂的剂量单位形式制备。例如，它们可以含有约 10^3 - 10^{15} 个病毒颗粒，优选 10^6 - 10^{12} 个病毒颗粒。用于人或其它哺乳动物的适当每日剂量可以根据患者的状况和其它因素而很大程度改变，但也可以采用常规方法确定。可以通过与适当的药物可接受载体如盐水、右旋糖或水一起作为组合物注射活性试剂而给药。

尽管本发明的核酸和/或载体可以作为唯一的活性药物制剂给药，它们也可以与本发明的一种或多种载体或其它试剂联合使用。当联合给药时，可以将治疗剂配制为同时或不同时给药的独立组合物，或者可以将治疗剂作为单个组合物给药。

也可以与适当的药物载体联合使用本发明的多肽。所述组合物包含治疗有效量的蛋白和药物可接受载体或赋形剂。所述载体包括但不限于盐水、缓冲盐水、右旋糖、水、甘油、乙醇及其组合。制剂应该适合于给药方式。

本发明也提供了装有本发明的药物组合物的一种或多种成分的一种或多种容器的药物包装或试剂盒。与该容器相关的可以是调控药物或生物产品的制造、使用或销售的政府部门规定形式的说明，该说明反映了制造、使用或销售部门关于人类用药的批准。此外，本发明的多肽可以与其它治疗性化合物联合使用。

药物组合物可以通过方便的方式给药，如眼内、眼药水和全身途径。给予患者的 tRNA 合成酶衍生的多肽的量和剂量方案将取决于许多因素，如给药方式、治疗的状况的性质、受治疗的受试者的体重和处方医生的判断。一般说来，多肽以至少约每日 $10 \mu\text{g}/\text{kg}$ 体重治疗有效剂量给药。优选地，该剂量为每日约 $10 \mu\text{g}/\text{kg}$ 体重- $1 \text{mg}/\text{kg}$ 体重，

考虑给药的频率、给药途径、症状等。

可以用血管抑制性 trpRS 治疗对抗内源和外源性血管发生因子的血管发生活性，并且防止进一步的生长或甚至抑制实体瘤，因为血管发生和新血管形成是实体瘤生长的必要步骤。所述治疗也可以用于治疗类风湿性关节炎、牛皮癣和糖尿病视网膜病，它们的特征均在于异常的血管发生。

提供了含治疗有效量和浓度的重组腺病毒递送载体的组合物，用于将治疗性基因产物递送到表达特定受体的细胞。这些细胞包括眼。特别关注的是眼的光受体细胞。可以通过任何实现与光受体接触的方法实现给药。为到达光受体细胞，优选的给药方式包括，但不限于视网膜下注射或玻璃体内注射。

也可以将重组病毒组合物配制为适用于植入眼前房或后房，优选玻璃体腔的，以持续释放制剂形式存在，如吸附于生物可降解的支持物，包括胶原海绵，或在脂质体中。持续释放的制剂可以配制用于多剂量给药，使得在选定的时间段，如一个月或到约一年中，给予几个剂量。因此，例如，可以制备脂质体，使得在一次注射中总共给予单剂量的约 2 倍-约 5 倍或更多倍。

载体可以配制于用于眼内，优选玻璃体内以约 0.05 ml-0.15 ml，优选约 0.05-0.1 ml 的体积给药的眼科学可接受载体中。

组合物可以提供于密封的无菌管瓶中，其中含有一定量的活性试剂，在眼内给药时以约 50-150 μl 的体积递送足量的病毒颗粒至光受体，该体积中含有至少约 10^7 ，更优选至少约 10^8 个斑形成单位。因此，一般地，该管瓶含约 0.15 ml 组合物。

为制备所述组合物，将病毒颗粒透析至适当的眼科学可接受载体或病毒可浓缩和/或与其混合。得到的混合物可以是溶液、悬浮液或乳状液。此外，病毒颗粒可以配制为组合物中的唯一药物活性成分或可以与其它用于治疗特定疾病的其它试剂组合。

为通过眼内注射或通过眼药水给药，适当的载体包括，但不限于生理盐水、磷酸缓冲盐水 (PBS)、平衡盐溶液 (BSS)、林格氏乳酸盐溶液，和含有增稠剂和增溶剂，如葡萄糖、聚乙二醇和聚丙二醇及其混合物的溶液。脂质体悬浮液也可以适于作为药物可接受载体。它们可以根据本领域技术人员公知的方法进行制备。适当的眼科学可接受

载体是公知的。用于眼科学用途的溶液或混合物可以用适当的盐配制为 0.01%-10%的等渗溶液, pH 约 5-7 (见, 例如美国专利 5,116,868, 它描述了用于局部施用的眼科用溶液的典型组合物)。所述 pH 调节为约 7.4 的溶液含有, 例如 90-100 mM 氯化钠、4-6 mM 磷酸氢二钾、4-6 mM 磷酸氢二钠、8-12 mM 柠檬酸钠、0.5-1.5 mM 氯化镁、1.5-2.5 mM 氯化钙、15-25 mM 乙酸钠、10-20 mM D,L-β 羟基丁酸钠和 5-5.5 mM 葡萄糖)。

可以用保护活性试剂抗身体的迅速清除的载体制备组合物, 例如时间释放制剂或包衣。所述载体包括控释制剂, 例如, 但不限于微胶囊化的递送系统, 和生物可降解的、生物相容性聚合物, 如乙烯乙酸乙烯酯、聚酞、聚乙醇酸、聚原酸酯、聚乳酸和其它类型的可以直接置于眼前房或后房或玻璃体腔的植入物。该组合物也可在颗粒中给药, 如 ELVAX® 颗粒 (乙烯-乙酸乙烯酯共聚物树脂, Dupont)。

脂质体悬浮液, 包括定向于组织的脂质体, 也可以适用于药物可接受的载体。例如, 可以通过本领域技术人员公知的方法制备脂质体制剂 (见, 例如 Kimm et al. *Bioch. Bioph. Acta* 728: 339-398(1983); Assil et al. *Arch Ophthalmol.* 105: 400(1987); 和美国专利 4,522,811)。也可以将病毒颗粒包裹到脂质体系统的水相中。

活性物质或试剂也可以与其它不损害所需作用的活性物质混合, 或与补充所需作用或具有其它作用的物质, 包括粘弹性物质混合, 所述物质如以 HEALON® 为商标出售 (Pharmacia, Inc) 的透明质酸, 该透明质酸是大约 300 万个透明质酸钠级分的高分子量 (MW) 分子 (见, 例如美国专利 5,292,362, 5,282,851, 5,273,056, 5,229,127, 4,517,295 和 4,328,803), 以 VISCOAT® 为商标出售的树脂 (可以从 Alcon Surgical, Inc. 得到), 该树脂是含氟的 (甲基) 丙烯酸酯, 如 1H, 1H, 2H, 2H-十七氟癸基甲基丙烯酸酯 (见, 例如美国专利 5,278,126, 5,273,751 和 5,214,080); 以 ORCOLON® 为商标出售的树脂 (Optical Radiation Corporation, 见, 例如美国专利 5,273,056), 和甲基纤维素、透明质酸甲酯、聚丙烯酰胺和聚甲基丙烯酸酯 (见, 例如美国专利 No. 5,273,751)。粘弹性物质的量一般为缀合物重量的约 0.5-5.0%, 优选 1-3%, 并且用于包被和保护受治疗的

组织。组合物可以包含染料，如亚甲基蓝或其它惰性染料，使得当注射到眼中时可以看到该组合物。也可以包含其它活性试剂。

组合物可以容纳于安瓿、一次性注射器或由玻璃、塑料或其它适当材料制备的多或单剂量管瓶。所述容纳的组合物可以在试剂盒中提供。具体地，此处提供包括管瓶、安瓿、或其它容器和试剂盒，优选具有递送约 0.100 ml 的足量组合物的一次性管瓶，和一次性针头，优选自我密封的 25-33 号或更小的针头。

最后，组合物可以包装为含有包装材料，一般是管瓶、含有本发明的多肽的眼科学可接受组合物、和说明组合物的治疗用途的标签的制品。

此处提供了用于实施本发明的试剂盒。该试剂盒含有一个或多个容器，如密封管瓶、用于单剂量给药的足够组合物、以及一个或多个针头，如自我密封的 25-33 号或更小的针头，优选 33 号或更小的针头、以及精确校准的注射器或其它适用于玻璃体内注射的精确校准的递送装置。

组合物的给药优选通过眼内注射，但如果足量的化合物能与玻璃体腔接触，其它给药方式可以是有效的。可以玻璃体内注射、房水注射或注射到眼外层，如结膜下注射或眼球筋膜下注射或如果使用穿透性制剂，通过局部施用到角膜，从而实现眼内注射。

对于特定患者，可以根据个体需要和管理及监测重组病毒给药的从业者的判断随时间调整特定的剂量方案。此处阐述的浓度和用量范围仅仅是示范性的，不限制所要求保护的方法的范围。

以下的实施例是用于说明本发明的具体实施方式及其多种用途。它们仅仅是用于说明和解释目的，但不作为限制。

实施例 1. 制备无内毒素的重组 TrpRS

按如下方法制备无内毒素的重组人 TrpRS。制备编码全长 TrpRS (SEQ ID NO: 1 的氨基酸残基 1-471) 或基本由 SEQ ID NO: 1 的残基 94-471 (即全长 TrpRS 的残基 94-471) 组成的，称作 T2 的截短的 TrpRS (SEQ ID NO: 12)，以及基本由 SEQ ID NO: 1 的残基 71-471 组成的，此后称作 T1 的截短的 TrpRS (SEQ ID NO: 13) 的质粒。每一种质粒也编码含 6 个组氨酸残基的 C 末端标记 (如 SEQ ID NO: 1 的氨基

酸残基 472-484) 和起始的甲硫氨酸残基。含 His₆ 标记的 T1 具有 SEQ ID NO: 5 的氨基酸序列, 而含 His₆ 标记的 T2 具有 SEQ ID NO: 7 的氨基酸序列。

将上述质粒导入大肠杆菌菌株 BL 21 (DE 3) (Novagen, Madison, WI)。按类似方法制备也编码 C 末端 6 个组氨酸残基标记的人成熟 EMAPII 备用。通过用异丙基 β-D-硫代吡喃半乳糖苷处理细胞 4 小时而诱导重组 TrpRS 的过表达。然后裂解细胞, 根据制造商建议的程序在 HIS·BIND® 镍亲和柱 (Novagen) 上从上清液中纯化蛋白。纯化后, 用含有 1 μM ZnSO₄ 的磷酸缓冲盐水 (PBS) 温育 TrpRS 蛋白, 然后去除游离的 Zn²⁺ (Kisselev et al., Eur. J. Biochem. 120: 511-17 (1981))。

通过采用 Triton X-114 进行相分离, 从蛋白样品中去除内毒素 (Liu et al., Clin. Biochem. 30: 455-63 (1997))。用 E-TOXATE® 凝胶-凝结测定 (Sigma, St. Louis, MO) 确定蛋白样品含有每 mL 少于 0.01 单位的内毒素。通过标准 Bradford 测定 (Bio-Rad, Hercules, CA), 使用牛血清白蛋白 (BSA) 为标准确定蛋白浓度。

实施例 2. 用 PMN 弹性蛋白酶切割人 TrpRS

检验了用 PMN 弹性蛋白酶切割人 TrpRS。在蛋白酶: 蛋白的比例为 1: 3000 的条件下, 用 PMN 在 PBS (pH 7.4) 中处理 TrpRS 0、15、30 或 60 分钟。切割后, 在 12.5% SDS-聚丙烯酰胺凝胶上分析样品。PMN 弹性蛋白酶切割由 DNA SEQ ID NO: 2 的核苷酸 3428-4738 编码的约 53kDa 的全长 TrpRS 产生了约 46kDa 的主要片段 (SEQ ID NO: 5, 具有 C 末端组氨酸标记的 T1) 和约 43kDa 的次要片段 (SEQ ID NO: 7, 具有 C 末端组氨酸标记的 T2)。

用重组 TrpRS 蛋白的羧基末端 His₆ 标记的抗体进行的蛋白印迹分析发现, 两种片段均具有羧基末端的 His₆ 标记。因此, 仅仅是两种 TrpRS 片段的氨基末端被截短。使用 ABI 494 型测序仪, 通过 Edman 降解测定了 TrpRS 片段的氨基末端序列。这些片段的测序表明, 氨基末端序列为 S-N-H-G-P (SEQ ID NO: 8) 和 S-A-K-G-I (SEQ ID NO: 9), 表明主要和次要 TrpRS 片段的氨基末端残基分别位于全长 TrpRS 的 71 和 94 位。图 1 概括了这些人 TrpRS 构建体。信号序列 -HVGH- (SEQ ID NO: 10) 和 -KMSAS- (SEQ ID NO: 11) 示于方框中。

在血管发生测定中分析主要和次要 TrpRS 片段的血管抑制活性。在这些测定中使用了均具有 C 末端组氨酸标记 (SEQ ID NO: 1 的氨基酸残基 472-484) 的主要和次要 TrpRS 片段 SEQ ID NO: 5 和 SEQ ID NO: 7 的重组形式。两种 TrpRS 片段都能够抑制血管发生。

实施例 3. Trp-RS 的截短片段表现出对视网膜血管发生的有效血管抑制作用

在出生后小鼠视网膜血管发生模型检验了衍生自色氨酸 tRNA 合成酶 (TrpRS, 53 kd; SEQ ID NO: 1) 的截短形式的血管抑制活性。Friedlander et al. Abstracts 709-B84 and 714-B89, IOVS 41(4): 138-139 (March 15, 2000) 报道了小鼠中分期进行的出生后视网膜血管发生。本发明提供了通过使用该分期的视网膜血管形成而测定血管发生抑制的方法。

将无内毒素的重组 Mini-TrpRS (含组氨酸标记的 TrpRS 的 48kDa 剪接变体; SEQ ID NO: 3) 和 T2 (含组氨酸标记的 TrpRS 的 43kDa 切割产物; SEQ ID NO: 7) 制备为重组蛋白。在出生后 (P) 7 或 8 天将这些蛋白玻璃体内注射到新生 Balb/C 小鼠, 在 P12 或 P13 收获视网膜。用 IV 型胶原抗体和荧光素偶联的第二抗体显现视网膜完整制备物中的血管。通过共聚集显微镜检查, 基于注射的蛋白对深层、外部血管丛的形成的作用评估抗血管发生活性。用解剖显微镜 (SMZ 645, Nikon, Japan) 进行玻璃体内注射和视网膜分离。在出生后第 7 天 (P7) 小鼠中用薄刀片建立眼睑裂缝, 从而暴露眼球, 用于注射 T2 (5 pmol) 或 TrpRS (5 pmol)。用安装有 32 号针头的注射器 (Hamilton Company, Reno, NV) 注射样品 (0.5 μ l)。在中纬线和角膜缘之间进行注射, 在注射过程中通过直接显现而监测针头尖端的定位, 从而确定其在玻璃体腔中。从研究中排除具有针头诱导的晶状体或视网膜损伤的眼。注射后, 使眼睑复位, 关闭裂缝。

出生后 12 天 (P12), 麻醉动物, 取出眼球。在 4% 低聚甲醛 (PFA) 中放置 10 分钟后, 通过边缘切口切除玻璃体。通过在冰上的甲醇中浸渍 10 分钟而制备用于染色的分离的视网膜, 然后在 50% 胎牛血清 (Gibco, Grand Island, NY) 中用 20% 正常山羊血清 (The Jackson Laboratory, Bar Harbor, ME) 在 PBS 中在冰上封闭 1 小时。通过用在封闭缓冲液中 1: 200 稀释的兔抗小鼠 IV 型胶原抗体 (Chemicon,

Temecula, CA) 4℃下对视网膜染色 18 小时而特异性显现血管。将 ALEXA FLUOR®594 偶联的山羊抗兔 IgG 抗体 (Molecular Probes, Eugene, OR) (在封闭缓冲液中 1: 200 稀释) 与视网膜一起 4℃下温育 2 小时。用缓慢退色的封固介质 M (Molecular Probes, Eugene, OR) 封固视网膜。

基于 P8-P12 之间形成的深层外部视网膜血管层 (次生层) 评估血管抑制活性。通过内层血管网 (初生层) 的外观评估正常发育和毒性症状。用于本实施例中的蛋白构建体都没有产生对初生层的任何副作用。

图 2 提供了 T2 抑制小鼠视网膜次生的深部血管网的血管形成的图示。在图 2 中, A 行表示暴露于 TrpRS 的视网膜血管网, B 行表示暴露于 Mini-TrpRS 的视网膜血管网, C 行表示暴露于本发明的多肽 T2 的视网膜的血管网。第 1 (左) 列表示初生的表面血管网, 第二列表示次生的深层血管网。如图 2 所证明, 没有任何多肽影响初生的血管网, 而仅仅 T2 显著抑制了次生的深层血管网的血管形成。

大多数用 PBS 处理的眼表现出正常的视网膜血管发育, 但在约 8.2% (n=73) 处理的眼中观察到了外部血管层的完全抑制。在 28% (n=75) mini-TrpRS (0.5 mg/ml) 处理的眼中观察到了外部血管网的完全抑制。较小的、截短形式 (T2) 是强得多的血管发生抑制剂, 以剂量依赖方式作用, 用 0.1 mg/ml T2 处理后 14.3% (n=14) 得到了完全抑制, 0.25 mg/ml 处理后抑制 40% (n=20), 0.5 mg/ml 处理后抑制 69.8% (n=53)。图 3 中图示了 0.5 mg/ml 处理的数据。小鼠视网膜提取物含有与人 mini-TrpRS 相同的表观分子量和免疫反应性的蛋白, 这是通过 SDS-PAGE 和蛋白印迹分析得到的。全长小鼠和人 TrpRS 具有约 88% 的氨基酸同一性并分别含有 475 和 471 个氨基酸。TrpRS 的截短形式, 特别是 T2, 对视网膜血管发育具有有效的血管抑制作用。

实施例 4. Matrigel 血管发生测定

根据 Brooks et al. *Methods Mol. Biol.*, 129: 257-269 (1999) 和 Eliceiri et al. *Mol. Cell*, 4: 915-924 (1999) 中描述的方法使用小鼠 matrigel 血管发生测定检验 T2 (SEQ ID NO: 7) 的血管抑制活性。根据以下改进进行。无胸腺 wehi 小鼠皮下植入 400 μl 生长因子耗尽的 matrigel (Becton Dickinson, Franklin Lakes, NJ), 其中

含有 20nM VEGF。开始通过在 matrigel 栓中加入 2.5 μ l T2 检测 T2 的血管抑制活性。通过在栓中加入多种浓度的 T2 测定效力。第 5 天，小鼠静脉注射荧光素标记的内皮结合性凝集素 Griffonia (Bandeiraea) Simplicifolia I，异凝集素 B4 (Vector Laboratories, Burlingame, CA)，切除 matrigel 栓。将栓在 RIPA 缓冲液 (10 mM 磷酸钠, pH 7.4, 150 mM 氯化钠, 1% Nonidet P-40, 0.5% 脱氧胆酸钠, 0.1% 十二烷基硫酸钠) 磨碎后，通过分光光度计分析对每个栓的荧光素含量进行定量。

实施例 5. T2 结合在视网膜中的定位

为评估注射到视网膜中的 T2 的摄取和定位，在出生后第 7 天 (P7) 将荧光素标记的 (ALEXA®488, Molecular Probes, Inc., Eugene OR) T2 注射到眼的玻璃体中。在 P8 和 P12 收获眼球，并且在 4% PFA 中固定 15 分钟。进一步分离视网膜和粘附的非视网膜组织，并且 4°C 下置于 4% PFA 中过夜，然后在干冰上包埋于介质中 (TISSUE-TEK® 0. C. T., Sakura FineTechnical Co., Japan)。用 PBS 对冻干的切片 (10 微米) 进行再水合，并且用含 5% BSA、2% 正常山羊血清的 PBS 封闭。按上文所述用抗小鼠 IV 型胶原抗体显现血管。用含 VECTASHIELD® 的 DAPI 核染料 (Vector Laboratories, Burlingame, CA) 将组织封固在盖玻片中。

或者，4°C 下用溶于缓冲液的 200 nM 荧光素标记的全长 TrpRS 或荧光素标记的 T2 温育未染色的视网膜过夜。在 PBS 中洗涤切片 6 次，每次 5 分钟，然后用 1 μ g/ml DAPI 温育 5 分钟，用于显现核。在用荧光素标记的 T2 温育之前，通过 4°C 下温育 1 μ M 未标记的 T2 8 小时而进行用未标记的 T2 预封闭。用多光子 BioRad MRC1024 共聚焦显微镜检查视网膜。用共聚焦辅助软件 (BioRad, Hercules, CA) 从一组 Z 系列图像中产生 3 维血管图像。

小鼠 Matrigel 栓测定中的血管抑制效力。我们检验了 T2 (SEQ ID NO: 7)，从而确定它是否具有血管抑制活性，尽管它已经失去了氨基酰化活性。用小鼠 matrigel 测定检验 T2 的体内血管抑制活性。VEGF₁₆₅ 诱导了血管发育到小鼠 matrigel 栓中。当与 VEGF₁₆₅ 一起向 matrigel 中加入 T2 时，以剂量依赖性方式阻断了血管发生，IC₅₀ 为 1.7 nM，如图 4 所示。

荧光素标记的 T2 定位于视网膜血管。为显现 T2 (SEQ ID NO: 7) 的眼内定位, 我们检验了出生后第 7 天玻璃体内注射荧光素标记的 T2 后的分布。第二天分离视网膜, 切片, 并用共聚焦显微镜观察。通过用荧光素标记 (ALEXA®594) 的抗 IV 型胶原抗体对标记的 T2 处理的眼进行共染色而证实定位 (数据未显示)。注射荧光素标记的 T2 (在 P12) 后 5 天, 仍然可以看见标记的 T2 的绿色荧光 (图 5A)。在这些视网膜中, 在 P12 没有观察到次生血管层, 表明荧光素标记的 T2 保留了与未标记 T2 相似的血管抑制活性。在 P7 注射荧光素标记的全长 TrpRS 的视网膜在 12 天内发育出了次生血管层, 但没有观察到血管染色 (图 5B)。在图 5 中, 荧光素标记的蛋白是绿色的, 胶原标记的血管是红色的, 核是蓝色的。

为进一步评估标记的 T2 的抑制特性, 用荧光素标记的 T2 对正常新生视网膜的横切片进行染色。在这些条件下, 荧光素标记的 T2 仅结合于血管 (图 5C)。结合是特异性的, 因为它可以通过用未标记的 T2 预温育而阻断 (数据未示出)。当将荧光素标记的全长 TrpRS 施用于视网膜时, 没有观察到视网膜血管染色 (图 5D), 这与全长酶的血管抑制活性缺乏一致。

如图 5 所示, 荧光素标记的 T2 是血管抑制性的, 并且定位于视网膜血管。在出生后 7 天 (P7), 注射 (0.5 μ l, 玻璃体内) 荧光素标记的 T2 (图 5A) 或全长 TrpRS (图 5B)。在 P8 收获视网膜, 用抗 IV 型胶原抗体和 DAPI 核染料染色, 标记的 T2 (图 5 中指向血管的上箭头) 定位于初生表面血管网中的血管 (1°)。注意次生深层血管网完全不存在 (2°)。尽管初生 (1°) 和次生 (2°) 血管层都存在于荧光素标记的全长 TrpRS 注射的眼中 (图 5B 中的箭头), 没有观察到标记。

在一组独立的实验中, 用荧光素标记的 T2 (图 5C) 或荧光素标记的全长 TrpRS (图 5D) 对冷冻的 P15 视网膜切片进行染色, 并且在共聚焦扫描激光显微镜中成像。标记的 T2 选择性定位于血管, 并且显示为穿过初生和次生视网膜血管层的位于图 5C 标记 “2°” 下的亮绿色血管。全长 TrpRS 没有观察到染色 (图 5D)。

全长 TrpRS 含有独特的 NH₂ 端结构域并且缺乏血管抑制活性。去除该完整结构域的部分或全部得到了具有血管抑制活性的蛋白。负责 T2 的血管抑制活性的结构似乎包含在 Rossmann 折叠核苷酸结合域的核心

内。可以通过可变剪接或蛋白水解而去除的 NH₂ 端结构域可以调节 TrpRS 的血管抑制活性,可能是通过暴露了血管抑制所必须的结合位点而达到的,该位点在全长 TrpRS 中不能达到。

VEGF 诱导的血管发生在小鼠 matrigel 模型中完全由 T2 所抑制,它是新生视网膜中的生理性血管发生。令人感兴趣的是,在 VEGF 刺激的血管发生中观察到了 TrpRS 片段的体外、在 CAM 中和在 matrigel 模型中的最有效抗血管发生作用。新生小鼠视网膜血管发生结果与 VEGF 刺激的血管发生和 TrpRS 片段的血管抑制作用之间的联系是一致的;该系统中的视网膜血管发生可以由 VEGF 所驱动。此外,在视网膜模型中观察到的抑制对新发生中的血管是特异的;该处理不改变预先存在的(在注射时)初生血管层的血管。尽管 T2 的血管抑制活性机制是未知的,T2 在视网膜内皮血管系统的特异性定位和 T2 对新发生中的血管的选择性作用提示,T2 可能通过在增殖中或迁移中的细胞上表达的内皮细胞受体起作用。对 T2 的血管抑制活性的进一步理解需要相关细胞受体的鉴定。

在干扰素- γ 刺激时产生血管抑制性 mini TrpRS 的多种细胞类型也产生血管抑制性因子如 IP-10。因此,这些结果提出了 TrpRS 在血管发生的正常、生理相关途径中的作用的可能性。另一种遍在的细胞蛋白-pro-EMAPII (p43) 具有两种明显无关的与此处报道的 TrpRS 相似的作用。Pro-EMAPII 通过与哺乳动物氨基酰基 tRNA 合成酶的多合成酶复合物缔合而辅助蛋白翻译。它经加工和分泌为 EMAPII,已经提出了 EMAPII 在肺发育中作为血管抑制性介质的作用。

因此,T2 可以用于正常或病理状况下观察到的生理相关的血管发生性重塑。在正常血管发生中,T2 可以辅助建立生理上重要的无血管区,该区存在于一些器官,如视网膜中央凹的无血管区。如果抑制全长 TrpRS 的切割,可以存在病理性血管发生,导致血管过度生长。

在眼病中,新血管形成可以导致视力的严重丧失。这些患者潜在可以从血管发生的抑制中获益。血管内皮生长因子与视网膜中的新血管形成和黄斑水肿相关,但认为其它的血管发生刺激物也在视网膜血管发生中起作用。我们已经观察到 VEGF 刺激的血管发生和 TrpRS 片段的有效血管抑制活性,使得这些分子可以用于视力下降和其它增殖性视网膜病的治疗。文献中还没有报道如同本发明的 T2 中那样,在 70%

的时间中完全抑制血管发生(图 5)的抗血管发生试剂。TrpRS 片段的另一优点是它们是天然存在的,因此,是潜在的非免疫原性抗血管发生剂。因此,可以通过基于靶细胞-或病毒载体的治疗递送这些分子。因为许多患有新血管性眼病的患者具有相关的全身性局部缺血性疾病,需要用直接置入眼中的基因工程化细胞或病毒载体进行抗血管发生治疗。

除了血管发生性视网膜病的治疗,本发明的 TrpRS 片段,特别是 T2 及其血管发生抑制性片段也可以通过防止肿瘤的血管形成而抑制实体瘤生长。本发明的 TrpRS 片段阻断 VEGF 诱导的内皮细胞体外增殖和趋化,因此可用于治疗任何涉及不利的内皮血管增殖和血管形成的病理。

序列表

<110> P. 施梅尔
K. 瓦卡苏吉
M. 弗里兰德
斯克里普斯研究学院

<120> 用于调节血管发生的色氨酸-tRNA合成酶衍生的多肽

<130> TSRI-813.1PC

<150> 60/270,951

<151> 2001-02-23

<160> 13

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 484

<212> PRT

<213> 人工序列

<220>

<223> 重组人trpRS

<400> 1

```

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

```

```

180          185          190
Ile Gln Met Thr Asp Asp Glu Lys Tyr Leu Trp Lys Asp Leu Thr Leu
195          200          205
Asp Gln Ala Tyr Gly Asp Ala Val Glu Asn Ala Lys Asp Ile Ile Ala
210          215          220
Cys Gly Phe Asp Ile Asn Lys Thr Phe Ile Phe Ser Asp Leu Asp Tyr
225          230          235
Met Gly Met Ser Ser Gly Phe Tyr Lys Asn Val Val Lys Ile Gln Lys
245          250          255
His Val Thr Phe Asn Gln Val Lys Gly Ile Phe Gly Phe Thr Asp Ser
260          265          270
Asp Cys Ile Gly Lys Ile Ser Phe Pro Ala Ile Gln Ala Ala Pro Ser
275          280          285
Phe Ser Asn Ser Phe Pro Gln Ile Phe Arg Asp Arg Thr Asp Ile Gln
290          295          300
Cys Leu Ile Pro Cys Ala Ile Asp Gln Asp Pro Tyr Phe Arg Met Thr
305          310          315          320
Arg Asp Val Ala Pro Arg Ile Gly Tyr Pro Lys Pro Ala Leu Leu His
325          330          335
Ser Thr Phe Phe Pro Ala Leu Gln Gly Ala Gln Thr Lys Met Ser Ala
340          345          350
Ser Asp Pro Asn Ser Ser Ile Phe Leu Thr Asp Thr Ala Lys Gln Ile
355          360          365
Lys Thr Lys Val Asn Lys His Ala Phe Ser Gly Gly Arg Asp Thr Ile
370          375          380
Glu Glu His Arg Gln Phe Gly Gly Asn Cys Asp Val Asp Val Ser Phe
385          390          395          400
Met Tyr Leu Thr Phe Phe Leu Glu Asp Asp Asp Lys Leu Glu Gln Ile
405          410          415
Arg Lys Asp Tyr Thr Ser Gly Ala Met Leu Thr Gly Glu Leu Lys Lys
420          425          430
Ala Leu Ile Glu Val Leu Gln Pro Leu Ile Ala Glu His Gln Ala Arg
435          440          445
Arg Lys Glu Val Thr Asp Glu Ile Val Lys Glu Phe Met Thr Pro Arg
450          455          460
Lys Leu Ser Phe Asp Phe Gln Lys Leu Ala Ala Ala Leu Glu His His
465          470          475          480
His His His His

```

<210> 2

<211> 4877

<212> DNA

<213> 人工序列

<220>

<223> pET20B中的重组人mini-TrpRS

<221> CDS

<222> (3428)...(4738)

<400> 2

tggcgaatgg gacgcgcctt gtagcggcgc attaagcgcg gcggggtgtgg tggttacgcg 60
 cagcgtgacc gctacacttg ccagcgcctt agcgcgcctt cctttcgcctt tcttcccttc 120
 ctttctcgcc acgttcgccc gctttccccc tcaagctcta aatcgggggc tcccttttagg 180
 gtcccgatgt agtgctttac ggcacctcga ccccaaaaaa cttgattagg gtgatgggtc 240
 acgtagtggg ccatcgccct gatagacggt ttttcgcctt ttgacgttgg agtccacggt 300
 ctttaatagt ggactcttgt tccaaactgg aacaacactc aaccctatct cggctctatc 360
 ttttgattta taagggattt tgccgatttc ggccatttgg ttaaaaaatg agctgattta 420
 acaaaaattt aacgcgaatt ttaacaaaat attaacgttt acaatttcag gtggcacttt 480
 tcggggaaat gtgcgcggaa ccctatttgg tttatttttc taaatacatt caaatatgta 540
 tcgcgctcatg agacaataac cctgataaat gcttcaataa tattgaaaaa ggaagagtat 600
 gagtattcaa catttcgctg tcgcccttat tccctttttt gcggcatttt gccttccctg 660
 ttttgctcac ccagaaacgc tggtgaaagt aaaagatgct gaagatcagt tgggtgcacg 720
 agtgggttac atcgaactgg atctcaacag cggttaagatc cttgagagtt ttcgccccga 780
 agaacgtttt ccaatgatga gcacttttaa agttctgcta tgtggcgcgg tattatcccg 840
 tattgacgcc gggcaagagc aactcggctc ccgcatacac tattctcaga atgacttggg 900
 tgagtaactca ccagtcacag aaaagcatct tacggatggc atgacagtaa gagaattatg 960
 cagtgctgccc ataaccatga gtgataaacac tgcggccaac ttacttctga caacgatcgg 1020
 aggaccgaag gagctaaccg cttttttgca caacatgggg gatcatgtaa ctgccttga 1080
 tcgttgggaa ccggagctga atgaagccat accaaaacgc gagcgtgaca ccacgatgcc 1140
 tgcagcaatg gcaacaacgt tgcgcaaac attactggc gaactactta ctctagcttc 1200
 ccggcaacaa ttaatagact ggatggaggc ggataaagtt gcaggaccac ttctgcgctc 1260
 ggcccttcgc gctggctggg ttattgctga taaatctgga gccgggtgagc gtgggtctcg 1320
 oggtatcatt gcagcactgg ggcagatgg taagccctcc cgtatcgtag ttatctacac 1380
 cagcggggag caggcaacta tggatgaacg aaatagacag atcgcctgaga taggtgcctc 1440
 actgattaag cattggtaac tgtcagacca agtttactca tatatacttt agattgattt 1500
 aaaacttcat ttttaattta aaaggatcta ggtgaagatc ctttttgata atctcatgac 1560
 caaaatccct taacgtgagt tttcgttcca ctgagcgtca gaccccgtag aaaagatcaa 1620
 aggatcttct tgagatcctt tttttctgcg cgtaatctgc tgcttgcaaa caaaaaaac 1680
 accgctacca gcggtggttt gtttgccgga tcaagagcta ccaactcttt tccggaagg 1740
 aactggcttc agcagagcgc agataccaaa tactgtcctt ctagtgtagc cgtagttagg 1800
 ccaccacttc aagaactctg tagcacccgc tacatacctc gctctgctaa tcctgttacc 1860
 agtggctgct gccagtgccg ataagtcgtg tottaccggg ttggactcaa gacgatagtt 1920
 accggataag gcgcagcggg cgggctgaac ggggggttcg tgcacacagc ccagcttggg 1980
 gcgaacgacc tacaccgaac tgagatacct acagcgtgag ctatgagaaa gcgccacgct 2040
 tcccgaaggg agaaaaggcg acaggtatcc ggtaaagcgc agggctcgaa caggagagcg 2100
 cccgagggag cttccagggg gaaacgcctg gtatctttat agtcctgtcg ggtttcgcca 2160
 cctctgactt gagegtcgat ttttgtgatg ctctcaggg gggcggagcc tatggaaaa 2220
 cgccagcaac gcggcctttt tacggttctt ggcccttttg tggccttttg ctccatggtt 2280
 ctttctcgcg ttatcccctg attctgtgga taaccgtatt accgcctttg agtgagctga 2340
 taccgctcgc cgcagccgaa gcaccgagcg cagcagatca gtgagcgagg aagcggaga 2400
 tggcctgatg cggatatttc tccttacgca tctgtcgggt atttcacacc gcataatagg 2460
 tgcactctca gtacaactcg ctctgatgcc gcatagttaa gccagtatac actccgctat 2520
 cgctacgtga ctgggtcatg gctgcgcccc gacacccgcc aacacccgct gacgcgccct 2580
 gacgggcttg tctgctcccg gcatccgctt acagacaagc tgtgaccgct tccgggagct 2640
 gcatgtgtca gaggttttca ccgtcatcac cgaaacgcgc gaggcagctg cggtaaaagct 2700
 catcagcgtg gtcgtgaagc gattcacaga tgtctgcctg ttcatccgcy tccagctcgt 2760
 tgagtttctc cagaagcgtt aatgtctggc ttctgataaa gcgggccatg ttaagggcgg 2820
 tttttctctg tttggctact gatgcctccg tghtaagggg atttctgttc atgggggtaa 2880
 tgataccgat gaaaacgagag aggatgctca cgatacgggt tactgatgat gaacatgccc 2940
 ggttactgga acgttgtgag ggtaaacacac tggcgggatg gatgcggcgg gaccagagaa 3000
 aaactactca gggctcaatgc cagcgtcttc ttaatacaga tgtagggtgt ccacagggta 3060
 gccagcagca tcctgcgatg cagatccgga acataatggt gcagggcgct gacttccgcy 3120
 tttccagact ttacgaaaca cggaaaccca agaocattca tgttgttgct caggtcgcag 3180

```

acgttttgca gcagcagtcg cttcacgttc gctcgcgtat eggtgattca ttctgctaac 3240
cagtaaggca accccgccag cctagccggg tctcaacga caggagcacg atcatgcgca 3300
cccgaggcca ggaccaacg ctgcccgaga tctcgatccc gcgaaattaa tagactcac 3360
tatagggaga ccacaacggt ttcctctag aaataatfff gtttaactff aagaaggaga 3420
tatacat atg agc tac aaa gct gcc gcg ggg gag gat tac aag gct gac 3469
Met Ser Tyr Lys Ala Ala Ala Gly Glu Asp Tyr Lys Ala Asp
      1             5             10

tgt cct cca ggg aac cca gca cct acc agt aat cat ggc cca gat gcc 3517
Cys Pro Pro Gly Asn Pro Ala Pro Thr Ser Asn His Gly Pro Asp Ala
 15             20             25             30

aca gaa gct gaa gag gat ttt gtg gac cca tgg aca gta cag aca agc 3565
Thr Glu Ala Glu Glu Asp Phe Val Asp Pro Trp Thr Val Gln Thr Ser
             35             40             45

agt gca aaa ggc ata gac tac gat aag ctc att gtt cgg ttt gga agt 3613
Ser Ala Lys Gly Ile Asp Tyr Asp Lys Leu Ile Val Arg Phe Gly Ser
             50             55             60

agt aaa att gac aaa gag cta ata aac cga ata gag aga gcc acc ggc 3661
Ser Lys Ile Asp Lys Glu Leu Ile Asn Arg Ile Glu Arg Ala Thr Gly
             65             70             75

caa aga cca cac cac ttc ctg cgc aga ggc atc ttc ttc tca cac aga 3709
Gln Arg Pro His His Phe Leu Arg Arg Gly Ile Phe Phe Ser His Arg
 80             85             90

gat atg aat cag gtt ctt gat gcc tat gaa aat aag aag cca ttt tat 3757
Asp Met Asn Gln Val Leu Asp Ala Tyr Glu Asn Lys Lys Pro Phe Tyr
 95             100             105             110

ctg tac acg ggc cgg ggc ccc tct tct gaa gca atg cat gta ggt cac 3805
Leu Tyr Thr Gly Arg Gly Pro Ser Ser Glu Ala Met His Val Gly His
             115             120             125

ctc att cca ttt att ttc aca aag tgg ctc cag gat gta ttt aac gtg 3853
Leu Ile Pro Phe Ile Phe Thr Lys Trp Leu Gln Asp Val Phe Asn Val
             130             135             140

ccc ttg gtc atc cag atg acg gat gac gag aag tat ctg tgg aag gac 3901
Pro Leu Val Ile Gln Met Thr Asp Asp Glu Lys Tyr Leu Trp Lys Asp
             145             150             155

ctg acc ctg gac cag gcc tat ggc gat gct gtt gag aat gcc aag gac 3949
Leu Thr Leu Asp Gln Ala Tyr Gly Asp Ala Val Glu Asn Ala Lys Asp
             160             165             170

atc atc gcc tgt ggc ttt gac atc aac aag act ttc ata ttc tct gac 3997
Ile Ile Ala Cys Gly Phe Asp Ile Asn Lys Thr Phe Ile Phe Ser Asp
             175             180             185             190

ctg gac tac atg ggg atg agc tca ggt ttc tac aaa aat gtg gtg aag 4045

```

```

Leu Asp Tyr Met Gly Met Ser Ser Gly Phe Tyr Lys Asn Val Val Lys
      195                200                205

att caa aag cat gtt acc ttc aac caa gtg aaa ggc att ttc ggc ttc 4093
Ile Gln Lys His Val Thr Phe Asn Gln Val Lys Gly Ile Phe Gly Phe
      210                215                220

act gac agc gac tgc att ggg aag atc agt ttt cct gcc atc cag gct 4141
Thr Asp Ser Asp Cys Ile Gly Lys Ile Ser Phe Pro Ala Ile Gln Ala
      225                230                235

gct ccc tcc ttc agc aac tca ttc cca cag atc ttc cga gac agg acg 4189
Ala Pro Ser Phe Ser Asn Ser Phe Pro Gln Ile Phe Arg Asp Arg Thr
      240                245                250

gat atc cag tgc ctt atc cca tgt gcc att gac cag gat cct tac ttt 4237
Asp Ile Gln Cys Leu Ile Pro Cys Ala Ile Asp Gln Asp Pro Tyr Phe
      255                260                265                270

aga atg aca agg gac gtc gcc ccc agg atc gcc tat cct aaa cca gcc 4285
Arg Met Thr Arg Asp Val Ala Pro Arg Ile Gly Tyr Pro Lys Pro Ala
      275                280                285

ctg ttg cac tcc acc ttc ttc cca gcc ctg cag ggc gcc cag acc aaa 4333
Leu Leu His Ser Thr Phe Phe Pro Ala Leu Gln Gly Ala Gln Thr Lys
      290                295                300

atg agt gcc agc gac cca aac tcc tcc atc ttc ctc acc gac acg gcc 4381
Met Ser Ala Ser Asp Pro Asn Ser Ser Ile Phe Leu Thr Asp Thr Ala
      305                310                315

aag cag atc aaa acc aag gtc aat aag cat gcg ttt tct gga ggg aga 4429
Lys Gln Ile Lys Thr Lys Val Asn Lys His Ala Phe Ser Gly Gly Arg
      320                325                330

gac acc atc gag gag cac agg cag ttt ggg ggc aac tgt gat gtg gac 4477
Asp Thr Ile Glu Glu His Arg Gln Phe Gly Gly Asn Cys Asp Val Asp
      335                340                345                350

gtg tct ttc atg tac ctg acc ttc ttc ctc gag gac gac gac aag ctc 4525
Val Ser Phe Met Tyr Leu Thr Phe Phe Leu Glu Asp Asp Asp Lys Leu
      355                360                365

gag cag atc agg aag gat tac acc agc gga gcc atg ctc acc ggt gag 4573
Glu Gln Ile Arg Lys Asp Tyr Thr Ser Gly Ala Met Leu Thr Gly Glu
      370                375                380

ctc aag aag gca ctc ata gag gtt ctg cag ccc ttg atc gca gag cac 4621
Leu Lys Lys Ala Leu Ile Glu Val Leu Gln Pro Leu Ile Ala Glu His
      385                390                395

cag gcc cgg cgc aag gag gtc acg gat gag ata gtg aaa gag ttc atg 4669
Gln Ala Arg Arg Lys Glu Val Thr Asp Glu Ile Val Lys Glu Phe Met

```

```

      400              405              410
act ccc egg aag ctg tcc ttc gac ttt cag aag ctt gcg gcc gca ctc 4717
Thr Pro Arg Lys Leu Ser Phe Asp Phe Gln Lys Leu Ala Ala Ala Leu
415              420              425              430

gag cac cac cac cac cac cac tgagatccgg ctgctaacaa agccccgaaag 4768
Glu His His His His His His
      435

gaagctgagt tggctgctgc cacegctgag caataactag cataaccctc tggggcctct 4828
aaacgggtct tgaggggttt tttgctgaaa ggaggaacta tatccggat 4877

<210> 3
<211> 437
<212> PRT
<213> 人工序列

<220>
<223> pET20B中的人TrpRS

<400> 3
Met Ser Tyr Lys Ala Ala Ala Gly Glu Asp Tyr Lys Ala Asp Cys Pro
 1              5              10              15
Pro Gly Asn Pro Ala Pro Thr Ser Asn His Gly Pro Asp Ala Thr Glu
      20              25              30
Ala Glu Glu Asp Phe Val Asp Pro Trp Thr Val Gln Thr Ser Ser Ala
      35              40              45
Lys Gly Ile Asp Tyr Asp Lys Leu Ile Val Arg Phe Gly Ser Ser Lys
      50              55              60
Ile Asp Lys Glu Leu Ile Asn Arg Ile Glu Arg Ala Thr Gly Gln Arg
      65              70              75              80
Pro His His Phe Leu Arg Arg Gly Ile Phe Ser His Arg Asp Met
      85              90              95
Asn Gln Val Leu Asp Ala Tyr Glu Asn Lys Lys Pro Phe Tyr Leu Tyr
      100             105             110
Thr Gly Arg Gly Pro Ser Ser Glu Ala Met His Val Gly His Leu Ile
      115             120             125
Pro Phe Ile Phe Thr Lys Trp Leu Gln Asp Val Phe Asn Val Pro Leu
      130             135             140
Val Ile Gln Met Thr Asp Asp Glu Lys Tyr Leu Trp Lys Asp Leu Thr
      145             150             155             160
Leu Asp Gln Ala Tyr Gly Asp Ala Val Glu Asn Ala Lys Asp Ile Ile
      165             170             175
Ala Cys Gly Phe Asp Ile Asn Lys Thr Phe Ile Phe Ser Asp Leu Asp
      180             185             190
Tyr Met Gly Met Ser Ser Gly Phe Tyr Lys Asn Val Val Lys Ile Gln
      195             200             205
Lys His Val Thr Phe Asn Gln Val Lys Gly Ile Phe Gly Phe Thr Asp
      210             215             220
Ser Asp Cys Ile Gly Lys Ile Ser Phe Pro Ala Ile Gln Ala Ala Pro
      225             230             235             240
Ser Phe Ser Asn Ser Phe Pro Gln Ile Phe Arg Asp Arg Thr Asp Ile

```

```

                245                250                255
Gln Cys Leu Ile Pro Cys Ala Ile Asp Gln Asp Pro Tyr Phe Arg Met
                260                265                270
Thr Arg Asp Val Ala Pro Arg Ile Gly Tyr Pro Lys Pro Ala Leu Leu
                275                280                285
His Ser Thr Phe Phe Pro Ala Leu Gln Gly Ala Gln Thr Lys Met Ser
                290                295                300
Ala Ser Asp Pro Asn Ser Ser Ile Phe Leu Thr Asp Thr Ala Lys Gln
305                310                315                320
Ile Lys Thr Lys Val Asn Lys His Ala Phe Ser Gly Gly Arg Asp Thr
                325                330                335
Ile Glu Glu His Arg Gln Phe Gly Gly Asn Cys Asp Val Asp Val Ser
                340                345                350
Phe Met Tyr Leu Thr Phe Phe Leu Glu Asp Asp Asp Lys Leu Glu Gln
                355                360                365
Ile Arg Lys Asp Tyr Thr Ser Gly Ala Met Leu Thr Gly Glu Leu Lys
                370                375                380
Lys Ala Leu Ile Glu Val Leu Gln Pro Leu Ile Ala Glu His Gln Ala
385                390                395                400
Arg Arg Lys Glu Val Thr Asp Glu Ile Val Lys Glu Phe Met Thr Pro
                405                410                415
Arg Lys Leu Ser Phe Asp Phe Gln Lys Leu Ala Ala Ala Leu Glu His
                420                425                430
His His His His His
                435

```

<210> 4

<211> 4811

<212> DNA

<213> 人工序列

<220>

<223> 重组人TrpRS的切割产物T1

<221> CDS

<222> (3428)...(4672)

<400> 4

```

tggcgaatgg gacgcgccct gtagcggcgc attaagcgcg gcgggtgtgg tggttacgcg 60
cagcgtgacc gctacaottg ccagcgcctt agcgcgcgct cctttcgcct tcttccttc 120
ctttctcgcg acgttcgcgc gctttccccc tcaagctcta aatcgggggc tcccttagg 180
gttccgattt agtgctttac ggcacctcga ccccaaaaaa cttgattagg gtgatggttc 240
acgtagtggg ccategcocct gatagacggg ttttcgcocct ttgacgttgg agtccacggt 300
ctttaatagt ggactcttgt tccaaactgg aacaacactc aaccctatct cggctctattc 360
ttttgattta taagggattt tgccgatttc ggcctattgg ttaaaaaatg agctgattta 420
acaaaaatth aacgcgaatt ttaacaaaat attaacgttt acaatttcag gtggcacttt 480
tcggggaaat gtgocgggaa cccctatttg tttatttttc taaatacatt caaatatgta 540
tccgctcatg agacaataac cctgataaat gcttcaataa tattgaaaaa ggaagagtat 600
gagtattcaa catttcocgtg tcgcccttat tccctttttt ggggcatttt gccttctctg 660
ttttgctcac ccagaaacgc tggtgaaagt aaaagatgct gaagatcagt tgggtgcacg 720
agtgggttac atcgaactgg atctcaacag cggttaagatc cttgagagtt ttcgccccga 780
agaacgtttt ccaatgatga gcacttttaa agttctgcta tgtggcgcgg tattatcccc 840

```

```

tattgacgcc gggcaagagc aactcggctcg ccgcatacac tattctcaga atgacttgggt 900
tgagtactca ccagtcacag aaaagcatct tacgggatggc atgacagtaa gagaattatg 960
cagtgtctgcc ataaccatga gtgataaacac tgcggccaac ttacttctga caacgatcgg 1020
aggaccgaag gagctaaccg cttttttgca caacatgggg gatcatgtaa ctgccttga 1080
tcgttgggaa ccggagctga atgaagccat accaaacgac gagcgtgaca ccacgatgcc 1140
tgcagcaatg gcaacaacgt tgcgcaaac attaactggc gaactactta ctctagcttc 1200
ccggcaacaa ttaatagact ggatggaggc ggataaagtt gcaggaccac ttctgctctc 1260
ggcccttccg gctggctgggt ttattgtctga taaatctgga gccggtgagc gtgggtctctc 1320
cggtatcatt gcagcactgg ggccagatgg taagccctcc cgtatcgtag ttatctacac 1380
gacgggggag caggcaacta tggatgaacg aaatagacag atcgcctgaga taggtgcctc 1440
actgattaag cattggtaac tgtcagacca agtttactca tatatacttt agattgattt 1500
aaaacttcat ttttaattta aaaggatcta ggtgaagatc ctttttgata atctcatgac 1560
caaaatccct taacgtgagt tttcgtcca ctgagcgtca gaccccgtag aaaagatcaa 1620
aggatcttct tgagatcctt tttttctgcy cgtaatctgc tgccttgcaaa caaaaaaac 1680
accgtacca gcggtgggtt gtttgcggga tcaagageta ccaactcttt ttccgaaggt 1740
aactggcttc agcagagcgc agataccaaa tactgtcctt ctagtgtagc cgtagttag 1800
ccaccacttc aagaactctg tagcaccgcc tacatacctc gctctgctaa tctgttacc 1860
agtggctgct gccagtgccg ataagtctg tcttaccggg ttggactcaa gacgatagtt 1920
accggataag ggcagcgggt cgggctgaac ggggggttcg tgcacacagc ccagcttggg 1980
gcgaacgacc tacaccgaac tgagatacct acagcgtgag ctatgagaaa gcgccacgct 2040
tcccgaaggg agaaaggcgg acaggtatcc ggtaagcggc agggctcggaa caggagagcg 2100
cacgagggag cttccagggg gaaacgcctg gtatctttat agtctgtctg ggttctcgca 2160
cctctgactt gagctcagat tttgtgatg ctcgtcaggg gggcggagcc tatggaaaaa 2220
cgccagcaac cgggcctttt tacggttctt ggcccttttg tggccttttg ctcacatggt 2280
ctttctctgc ttatcccctg attctgtgga taaccgtatt accgcctttg agtgagctga 2340
taccgctcgc cgcagccgaa cgaccgagcg cagcagatca gtgagcagag aagcggaga 2400
gcgcctgatg cggtatcttc tcttacgca tctgtcgggt atttcacacc gcataatag 2460
tgcactctca gtacaatctg ctctgatgcc gcatagttaa gccagtatac actccgctat 2520
cgctacgtga ctgggtcatg gctgcgcccc gacaccgcc aaçaccgct gacgcgccct 2580
gcgggcttg tctgctccc gcatccgctt acagacaagc tgtgaccgto tccgggagct 2640
gcatgtgtca gaggttttca ccgtcatcac cgaacgcgc gaggcagctg cggtaaagct 2700
catcagcgtg gtcgtgaagc gattcacaga tgtctgcctg tcatcccg tccagctcgt 2760
tgagtttctc cagaagcgtt aatgtctggc ttctgataaa gcgggcccag ttaagggcgg 2820
tttttctctg tttggctact gatgcctccg tgtaaggggg atttctgttc atgggggtaa 2880
tgataccgat gaaacgagag aggatgctca cgatacgggt tactgatgat gaacatgcc 2940
ggttactgga acgttgtgag ggtaaacaaac tggcgggtatg gatgcggcgg gaccagagaa 3000
aaatcactca gggctcaatgc cagcgtctcg ttaatacaga tgtagggtgtt ccacagggta 3060
gccagcagca tctctcgatg cagatccgga acataatggt gcagggcgct gacttcccg 3120
tttccagact ttacgaaaca cggaaaccga agaccattea tgttgttgcct caggtcgcag 3180
acgttttgca gcagcagctc cttcacgttc gctcgcgtat cgggtgattca ttctgctaac 3240
cagtaaggca accccgccag cctagccggg tctcaacga caggagcacg atcatgcgca 3300
cccgtggcca ggaccacaac cgtcccgaga tctcgatccc gcgaaattaa tacgactcac 3360
tatagggaga ccacaacggt ttccctctag aaataatttt gtttaacttt aagaaggaga 3420
tatacat atg agt aat cat ggc cca gat gcc aca gaa gct gaa gag gat 3469
Met Ser Asn His Gly Pro Asp Ala Thr Glu Ala Glu Glu Asp
1 5 10

ttt gtg gac cca tgg aca gta cag aca agc agt gca aaa ggc ata gac 3517
Phe Val Asp Pro Trp Thr Val Gln Thr Ser Ser Ala Lys Gly Ile Asp
15 20 25 30

tac gat aag ctc att gtt cgg ttt gga agt agt aaa att gac aaa gag 3565
Tyr Asp Lys Leu Ile Val Arg Phe Gly Ser Ser Lys Ile Asp Lys Glu

```

	35	40	45	
cta ata aac cga ata gag aga gcc acc ggc caa aga cca cac cac ttc				3613
Leu Ile Asn Arg Ile Glu Arg Ala Thr Gly Gln Arg Pro His His Phe				
	50	55	60	
ctg cgc aga ggc atc ttc ttc tca cac aga gat atg aat cag gtt ctt				3661
Leu Arg Arg Gly Ile Phe Phe Ser His Arg Asp Met Asn Gln Val Leu				
	65	70	75	
gat gcc tat gaa aat aag aag cca ttt tat ctg tac acg ggc cgg ggc				3709
Asp Ala Tyr Glu Asn Lys Lys Pro Phe Tyr Leu Tyr Thr Gly Arg Gly				
	80	85	90	
ccc tct tct gaa gca atg cat gta ggt cac ctc att cca ttt att ttc				3757
Pro Ser Ser Glu Ala Met His Val Gly His Leu Ile Pro Phe Ile Phe				
	95	100	105	110
aca aag tgg ctc cag gat gta ttt aac gtg ccc ttg gtc atc cag atg				3805
Thr Lys Trp Leu Gln Asp Val Phe Asn Val Pro Leu Val Ile Gln Met				
	115	120	125	
acg gat gac gag aag tat ctg tgg aag gac ctg acc ctg gac cag gcc				3853
Thr Asp Asp Glu Lys Tyr Leu Trp Lys Asp Leu Thr Leu Asp Gln Ala				
	130	135	140	
tat ggc gat gct gtt gag aat gcc aag gac atc atc gcc tgt ggc ttt				3901
Tyr Gly Asp Ala Val Glu Asn Ala Lys Asp Ile Ile Ala Cys Gly Phe				
	145	150	155	
gac atc aac aag act ttc ata ttc tct gac ctg gac tac atg ggg atg				3949
Asp Ile Asn Lys Thr Phe Ile Phe Ser Asp Leu Asp Tyr Met Gly Met				
	160	165	170	
agc tca ggt ttc tac aaa aat gtg gtg aag att caa aag cat gtt acc				3997
Ser Ser Gly Phe Tyr Lys Asn Val Val Lys Ile Gln Lys His Val Thr				
	175	180	185	190
ttc aac caa gtg aaa ggc att ttc ggc ttc act gac agc gac tgc att				4045
Phe Asn Gln Val Lys Gly Ile Phe Gly Phe Thr Asp Ser Asp Cys Ile				
	195	200	205	
ggg aag atc agt ttt cct gcc atc cag gct gct ccc tcc ttc agc aac				4093
Gly Lys Ile Ser Phe Pro Ala Ile Gln Ala Ala Pro Ser Phe Ser Asn				
	210	215	220	
tca ttc cca cag atc ttc cga gac agg acg gat atc cag tgc ctt atc				4141
Ser Phe Pro Gln Ile Phe Arg Asp Arg Thr Asp Ile Gln Cys Leu Ile				
	225	230	235	
cca tgt gcc att gac cag gat cct tac ttt aga atg aca agg gac gtc				4189
Pro Cys Ala Ile Asp Gln Asp Pro Tyr Phe Arg Met Thr Arg Asp Val				
	240	245	250	

```

gcc ccc agg atc ggc tat cct aaa cca gcc ctg ttg cac tcc acc ttc 4237
Ala Pro Arg Ile Gly Tyr Pro Lys Pro Ala Leu Leu His Ser Thr Phe
255                260                265                270

ttc cca gcc ctg cag ggc gcc cag acc aaa atg agt gcc agc gac cca 4285
Phe Pro Ala Leu Gln Gly Ala Gln Thr Lys Met Ser Ala Ser Asp Pro
                275                280                285

aac tcc tcc atc ttc ctc acc gac acg gcc aag cag atc aaa acc aag 4333
Asn Ser Ser Ile Phe Leu Thr Asp Thr Ala Lys Gln Ile Lys Thr Lys
                290                295                300

gtc aat aag cat gcg ttt tct gga ggg aga gac acc atc gag gag cac 4381
Val Asn Lys His Ala Phe Ser Gly Gly Arg Asp Thr Ile Glu Glu His
                305                310                315

agg cag ttt ggg ggc aac tgt gat gtg gac gtg tct ttc atg tac ctg 4429
Arg Gln Phe Gly Gly Asn Cys Asp Val Asp Val Ser Phe Met Tyr Leu
                320                325                330

acc ttc ttc ctc gag gac gac gac aag ctc gag cag atc agg aag gat 4477
Thr Phe Phe Leu Glu Asp Asp Asp Lys Leu Glu Gln Ile Arg Lys Asp
335                340                345                350

tac acc agc gga gcc atg ctc acc ggt gag ctc aag aag gca ctc ata 4525
Tyr Thr Ser Gly Ala Met Leu Thr Gly Glu Leu Lys Lys Ala Leu Ile
                355                360                365

gag gtt ctg cag ccc ttg atc gca gag cac cag gcc cgg cgc aag gag 4573
Glu Val Leu Gln Pro Leu Ile Ala Glu His Gln Ala Arg Arg Lys Glu
                370                375                380

gtc acg gat gag ata gtg aaa gag ttc atg act ccc cgg aag ctg tcc 4621
Val Thr Asp Glu Ile Val Lys Glu Phe Met Thr Pro Arg Lys Leu Ser
                385                390                395

ttc gac ttt cag aag ctt gcg gcc gca ctc gag cac cac cac cac cac 4669
Phe Asp Phe Gln Lys Leu Ala Ala Ala Leu Glu His His His His His
                400                405                410

cac tgagatccgg ctgctaacaa agcccgaag gaagctgagt tggctgctgc 4722
His
415

caccgctgag caataactag cataaccct tggggcctct aaacgggtct tgaggggttt 4782
tttgctgaaa ggaggaacta tatccgat 4811

```

<210> 5

<211> 415

<212> PRT

<213> 人工序列

<220>

<223> 重组人TrpRS的切割产物T1

<400> 5

```

Met Ser Asn His Gly Pro Asp Ala Thr Glu Ala Glu Glu Asp Phe Val
 1          5          10          15
Asp Pro Trp Thr Val Gln Thr Ser Ser Ala Lys Gly Ile Asp Tyr Asp
          20          25          30
Lys Leu Ile Val Arg Phe Gly Ser Ser Lys Ile Asp Lys Glu Leu Ile
 35          40          45
Asn Arg Ile Glu Arg Ala Thr Gly Gln Arg Pro His His Phe Leu Arg
 50          55          60
Arg Gly Ile Phe Phe Ser His Arg Asp Met Asn Gln Val Leu Asp Ala
 65          70          75          80
Tyr Glu Asn Lys Lys Pro Phe Tyr Leu Tyr Thr Gly Arg Gly Pro Ser
          85          90          95
Ser Glu Ala Met His Val Gly His Leu Ile Pro Phe Ile Phe Thr Lys
          100          105          110
Trp Leu Gln Asp Val Phe Asn Val Pro Leu Val Ile Gln Met Thr Asp
          115          120          125
Asp Glu Lys Tyr Leu Trp Lys Asp Leu Thr Leu Asp Gln Ala Tyr Gly
          130          135          140
Asp Ala Val Glu Asn Ala Lys Asp Ile Ile Ala Cys Gly Phe Asp Ile
          145          150          155          160
Asn Lys Thr Phe Ile Phe Ser Asp Leu Asp Tyr Met Gly Met Ser Ser
          165          170          175
Gly Phe Tyr Lys Asn Val Val Lys Ile Gln Lys His Val Thr Phe Asn
          180          185          190
Gln Val Lys Gly Ile Phe Gly Phe Thr Asp Ser Asp Cys Ile Gly Lys
          195          200          205
Ile Ser Phe Pro Ala Ile Gln Ala Ala Pro Ser Phe Ser Asn Ser Phe
          210          215          220
Pro Gln Ile Phe Arg Asp Arg Thr Asp Ile Gln Cys Leu Ile Pro Cys
          225          230          235          240
Ala Ile Asp Gln Asp Pro Tyr Phe Arg Met Thr Arg Asp Val Ala Pro
          245          250          255
Arg Ile Gly Tyr Pro Lys Pro Ala Leu Leu His Ser Thr Phe Phe Pro
          260          265          270
Ala Leu Gln Gly Ala Gln Thr Lys Met Ser Ala Ser Asp Pro Asn Ser
          275          280          285
Ser Ile Phe Leu Thr Asp Thr Ala Lys Gln Ile Lys Thr Lys Val Asn
          290          295          300
Lys His Ala Phe Ser Gly Gly Arg Asp Thr Ile Glu Glu His Arg Gln
          305          310          315          320
Phe Gly Gly Asn Cys Asp Val Asp Val Ser Phe Met Tyr Leu Thr Phe
          325          330          335
Phe Leu Glu Asp Asp Asp Lys Leu Glu Gln Ile Arg Lys Asp Tyr Thr
          340          345          350
Ser Gly Ala Met Leu Thr Gly Glu Leu Lys Lys Ala Leu Ile Glu Val
          355          360          365
Leu Gln Pro Leu Ile Ala Glu His Gln Ala Arg Arg Lys Glu Val Thr
          370          375          380
Asp Glu Ile Val Lys Glu Phe Met Thr Pro Arg Lys Leu Ser Phe Asp

```



```

cctctgactt gagcgtcgat ttttgtgatg ctcgtcaggg gggcggagcc tatggaaaaa 2220
cgccagcaac ggggcctttt tacggttcct ggccttttgc tggccttttg ctcacatggt 2280
ctttcctgcg ttatcccctg attctgtgga taaccgtatt accgcctttg agtgagctga 2340
taccgctcgc cgcagccgaa cgaccgagcg cagcagagca gtgagcgagg aagcgggaaga 2400
gcgcctgatg cggtatatttc tccttacgca tctgtgcggt atttcacacc gcatatatgg 2460
tgcactctca gtacaatctg ctctgatgcc gcatagttaa gccagtatac actccgctat 2520
cgctacgtga ctgggtcatg gctgcgcccc gacacccgcc aacacccgct gacgcgcctt 2580
gacgggcttg tctgtccccc gcatecggctt acagacaagc tgtgaccgct tccgggagct 2640
gcatgtgtca gaggttttca ccgtcatcac cgaaacgcgc gaggcagctg cggtaaagct 2700
catcagcgtg gtcgtgaagc gattcacaga tgtctgcctg ttcacccgcg tccagctcgt 2760
tgagtttctc cagaagcgtt aatgtctggc ttctgataaa gccggccatg ttaagggcgg 2820
tttttctctg tttggctact gatgcctccg tgtaaggggg atttctgttc atgggggtaa 2880
tgataccgat gaaacgagag aggatgctca cgatacgggt tactgatgat gaacatgccc 2940
ggttactgga acgttgtgag ggtaaacaaac tggcgggatg gatgcggcgg gaccagagaa 3000
aaatcactca gggtaaatgc cagcgccttcg ttaatacaga tgtaggtgtt ccacagggta 3060
gccagcagca tcctgcgatg cagatccgga acataatggt gcagggcgct gacttccgcg 3120
tttcagact ttacgaaaca cggaaacega agaccattca tgttgttget caggtcgcag 3180
acgttttgca gcagcagtcg cttcacgttc gctcgcgtat cggtgattca ttctgctaac 3240
cagtaaggca acccccggag cctagccggg tcctcaacga caggagcacg atcatgcgca 3300
cccgtaggca ggacccaacg ctgcccgaga tctcgatccc gcgaaattaa tacgactcac 3360
tatagggaga ccacaacggt ttccctctag aaataatttt gtttaacttt aagaaggaga 3420
tatacat atg agt gca aaa ggc ata gac tac gat aag ctc att gtt cgg 3469
Met Ser Ala Lys Ile Asp Tyr Asp Lys Leu Ile Val Arg
1 5 10

ttt gga agt agt aaa att gac aaa gag cta ata aac cga ata gag aga 3517
Phe Gly Ser Ser Lys Ile Asp Lys Glu Leu Ile Asn Arg Ile Glu Arg
15 20 25 30

gcc acc ggc caa aga cca cac cac ttc ctg cgc aga ggc atc ttc ttc 3565
Ala Thr Gly Gln Arg Pro His His Phe Leu Arg Arg Gly Ile Phe Phe
35 40 45

tca cac aga gat atg aat cag gtt ctt gat gcc tat gaa aat aag aag 3613
Ser His Arg Asp Met Asn Gln Val Leu Asp Ala Tyr Glu Asn Lys Lys
50 55 60

cca ttt tat ctg tac acg ggc cgg ggc ccc tct tct gaa gca atg cat 3661
Pro Phe Tyr Leu Tyr Thr Gly Arg Gly Pro Ser Ser Glu Ala Met His
65 70 75

gta ggt cac ctc att cca ttt att ttc aca aag tgg ctc cag gat gta 3709
Val Gly His Leu Ile Pro Phe Ile Phe Thr Lys Trp Leu Gln Asp Val
80 85 90

ttt aac gtg ccc ttg gtc atc cag atg acg gat gac gag aag tat ctg 3757
Phe Asn Val Pro Leu Val Ile Gln Met Thr Asp Asp Glu Lys Tyr Leu
95 100 105 110

tgg aag gac ctg acc ctg gac cag gcc tat ggc gat gct gtt gag aat 3805
Trp Lys Asp Leu Thr Leu Asp Gln Ala Tyr Gly Asp Ala Val Glu Asn
115 120 125

```

gcc aag gac atc atc gcc tgt ggc ttt gac atc aac aag act ttc ata	3853
Ala Lys Asp Ile Ile Ala Cys Gly Phe Asp Ile Asn Lys Thr Phe Ile	
130 135 140	
ttc tct gac ctg gac tac atg ggg atg agc tca ggt ttc tac aaa aat	3901
Phe Ser Asp Leu Asp Tyr Met Gly Met Ser Ser Gly Phe Tyr Lys Asn	
145 150 155	
gtg gtg aag att caa aag cat gtt acc ttc aac caa gtg aaa ggc att	3949
Val Val Lys Ile Gln Lys His Val Thr Phe Asn Gln Val Lys Gly Ile	
160 165 170	
ttc ggc ttc act gac agc gac tgc att ggg aag atc agt ttt cct gcc	3997
Phe Gly Phe Thr Asp Ser Asp Cys Ile Gly Lys Ile Ser Phe Pro Ala	
175 180 185 190	
atc cag gct gct ccc tcc ttc agc aac tca ttc cca cag atc ttc cga	4045
Ile Gln Ala Ala Pro Ser Phe Ser Asn Ser Phe Pro Gln Ile Phe Arg	
195 200 205	
gac agg acg gat atc cag tgc ctt atc cca tgt gcc att gac cag gat	4093
Asp Arg Thr Asp Ile Gln Cys Leu Ile Pro Cys Ala Ile Asp Gln Asp	
210 215 220	
cct tac ttt aga atg aca agg gac gtc gcc ccc agg atc ggc tat cct	4141
Pro Tyr Phe Arg Met Thr Arg Asp Val Ala Pro Arg Ile Gly Tyr Pro	
225 230 235	
aaa cca gcc ctg ttg cac tcc acc ttc ttc cca gcc ctg cag ggc gcc	4189
Lys Pro Ala Leu Leu His Ser Thr Phe Phe Pro Ala Leu Gln Gly Ala	
240 245 250	
cag acc aaa atg agt gcc agc gac cca aac tcc tcc atc ttc ctc acc	4237
Gln Thr Lys Met Ser Ala Ser Asp Pro Asn Ser Ser Ile Phe Leu Thr	
255 260 265 270	
gac acg gcc aag cag atc aaa acc aag gtc aat aag cat gcg ttt tct	4285
Asp Thr Ala Lys Gln Ile Lys Thr Lys Val Asn Lys His Ala Phe Ser	
275 280 285	
gga ggg aga gac acc atc gag gag cac agg cag ttt ggg ggc aac tgt	4333
Gly Gly Arg Asp Thr Ile Glu Glu His Arg Gln Phe Gly Gly Asn Cys	
290 295 300	
gat gtg gac gtg tct ttc atg tac ctg acc ttc ttc ctc gag gac gac	4381
Asp Val Asp Val Ser Phe Met Tyr Leu Thr Phe Phe Leu Glu Asp Asp	
305 310 315	
gac aag ctc gag cag atc agg aag gat tac acc agc gga gcc atg ctc	4429
Asp Lys Leu Glu Gln Ile Arg Lys Asp Tyr Thr Ser Gly Ala Met Leu	
320 325 330	
acc ggt gag ctc aag aag gca ctc ata gag gtt ctg cag ccc ttg atc	4477

```

Thr Gly Glu Leu Lys Lys Ala Leu Ile Glu Val Leu Gln Pro Leu Ile
335                340                345                350

gca gag cac cag gcc cgg cgc aag gag gtc acg gat gag ata gtg aaa 4525
Ala Glu His Gln Ala Arg Arg Lys Glu Val Thr Asp Glu Ile Val Lys
                355                360                365

gag ttc atg act ccc cgg aag ctg tcc ttc gac ttt cag aag ctt gcg 4573
Glu Phe Met Thr Pro Arg Lys Leu Ser Phe Asp Phe Gln Lys Leu Ala
                370                375                380

gcc gca ctc gag cac cac cac cac cac tgagatccgg ctgctaacaa 4623
Ala Ala Leu Glu His His His His His His
                385                390

agccccgaaag gaagctgagt tggctgctgc caccgctgag caataactag cataaccct 4683
tggggcctct aaacgggtct tgaggggttt tttgctgaaa ggaggaacta tatccggat 4742

```

<210> 7

<211> 392

<212> PRT

<213> 人工序列

<220>

<223> 重组人TrpRS的切割产物T2

<400> 7

```

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

```

```

Thr Asp Ile Gln Cys Leu Ile Pro Cys Ala Ile Asp Gln Asp Pro Tyr
 210                215                220
Phe Arg Met Thr Arg Asp Val Ala Pro Arg Ile Gly Tyr Pro Lys Pro
225                230                235                240
Ala Leu Leu His Ser Thr Phe Phe Pro Ala Leu Gln Gly Ala Gln Thr
                245                250                255
Lys Met Ser Ala Ser Asp Pro Asn Ser Ser Ile Phe Leu Thr Asp Thr
                260                265                270
Ala Lys Gln Ile Lys Thr Lys Val Asn Lys His Ala Phe Ser Gly Gly
 275                280                285
Arg Asp Thr Ile Glu Glu His Arg Gln Phe Gly Gly Asn Cys Asp Val
 290                295                300
Asp Val Ser Phe Met Tyr Leu Thr Phe Phe Leu Glu Asp Asp Asp Lys
305                310                315                320
Leu Glu Gln Ile Arg Lys Asp Tyr Thr Ser Gly Ala Met Leu Thr Gly
                325                330                335
Glu Leu Lys Lys Ala Leu Ile Glu Val Leu Gln Pro Leu Ile Ala Glu
                340                345                350
His Gln Ala Arg Arg Lys Glu Val Thr Asp Glu Ile Val Lys Glu Phe
                355                360                365
Met Thr Pro Arg Lys Leu Ser Phe Asp Phe Gln Lys Leu Ala Ala Ala
 370                375                380
Leu Glu His His His His His His
385                390

```

```

<210> 8
<211> 5
<212> PRT
<213> 人 (Homo sapiens)

```

```

<400> 8
Ser Asn His Gly Pro
 1                5

```

```

<210> 9
<211> 5
<212> PRT
<213> 人 (Homo sapiens)

```

```

<400> 9
Ser Ala Lys Gly Ile
 1                5

```

```

<210> 10
<211> 4
<212> PRT
<213> 人 (Homo sapiens)

```

```

<400> 10
His Val Gly His

```

1

<210> 11
 <211> 5
 <212> PRT
 <213> 人 (Homo sapiens)

<400> 11
 Lys Met Ser Ala Ser
 1 5

<210> 12
 <211> 378
 <212> PRT
 <213> 人 (Homo sapiens)

<400> 12
 Ser Ala Lys Gly Ile Asp Tyr Asp Lys Leu Ile Val Arg Phe Gly Ser
 1 5 10 15
 Ser Lys Ile Asp Lys Glu Leu Ile Asn Arg Ile Glu Arg Ala Thr Gly
 20 25 30
 Gln Arg Pro His His Phe Leu Arg Arg Gly Ile Phe Phe Ser His Arg
 35 40 45
 Asp Met Asn Gln Val Leu Asp Ala Tyr Glu Asn Lys Lys Pro Phe Tyr
 50 55 60
 Leu Tyr Thr Gly Arg Gly Pro Ser Ser Glu Ala Met His Val Gly His
 65 70 75 80
 Leu Ile Pro Phe Ile Phe Thr Lys Trp Leu Gln Asp Val Phe Asn Val
 85 90 95
 Pro Leu Val Ile Gln Met Thr Asp Asp Glu Lys Tyr Leu Trp Lys Asp
 100 105 110
 Leu Thr Leu Asp Gln Ala Tyr Gly Asp Ala Val Glu Asn Ala Lys Asp
 115 120 125
 Ile Ile Ala Cys Gly Phe Asp Ile Asn Lys Thr Phe Ile Phe Ser Asp
 130 135 140
 Leu Asp Tyr Met Gly Met Ser Ser Gly Phe Tyr Lys Asn Val Val Lys
 145 150 155 160
 Ile Gln Lys His Val Thr Phe Asn Gln Val Lys Gly Ile Phe Gly Phe
 165 170 175
 Thr Asp Ser Asp Cys Ile Gly Lys Ile Ser Phe Pro Ala Ile Gln Ala
 180 185 190
 Ala Pro Ser Phe Ser Asn Ser Phe Pro Gln Ile Phe Arg Asp Arg Thr
 195 200 205
 Asp Ile Gln Cys Leu Ile Pro Cys Ala Ile Asp Gln Asp Pro Tyr Phe
 210 215 220
 Arg Met Thr Arg Asp Val Ala Pro Arg Ile Gly Tyr Pro Lys Pro Ala
 225 230 235 240
 Leu Leu His Ser Thr Phe Phe Pro Ala Leu Gln Gly Ala Gln Thr Lys
 245 250 255
 Met Ser Ala Ser Asp Pro Asn Ser Ser Ile Phe Leu Thr Asp Thr Ala
 260 265 270

Lys Gln Ile Lys Thr Lys Val Asn Lys His Ala Phe Ser Gly Gly Arg
 275 280 285
 Asp Thr Ile Glu Glu His Arg Gln Phe Gly Gly Asn Cys Asp Val Asp
 290 295 300
 Val Ser Phe Met Tyr Leu Thr Phe Phe Leu Glu Asp Asp Asp Lys Leu
 305 310 315 320
 Glu Gln Ile Arg Lys Asp Tyr Thr Ser Gly Ala Met Leu Thr Gly Glu
 325 330 335
 Leu Lys Lys Ala Leu Ile Glu Val Leu Gln Pro Leu Ile Ala Glu His
 340 345 350
 Gln Ala Arg Arg Lys Glu Val Thr Asp Glu Ile Val Lys Glu Phe Met
 355 360 365
 Thr Pro Arg Lys Leu Ser Phe Asp Phe Gln
 370 375

<210> 13

<211> 401

<212> PRT

<213> 人 (Homo sapiens)

<400> 13

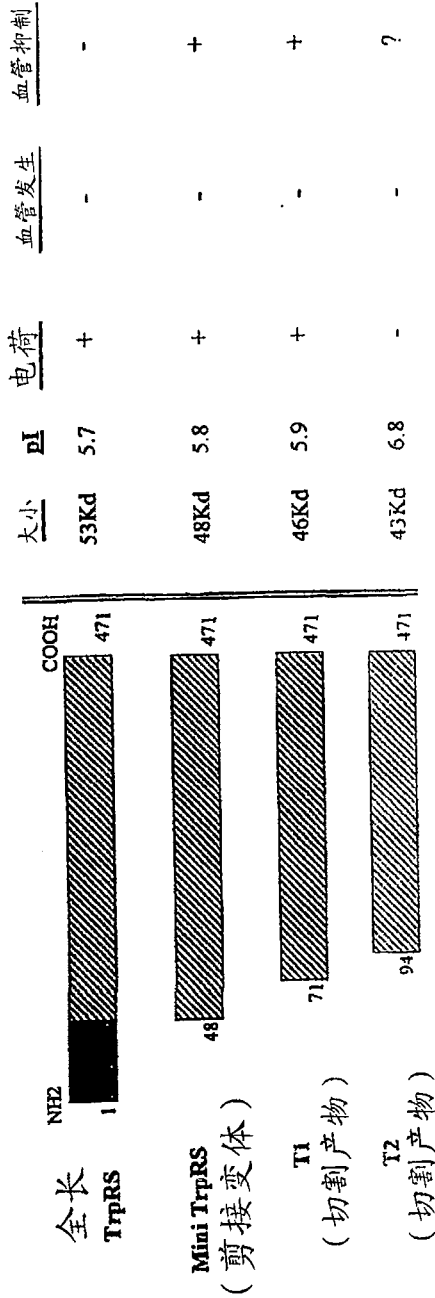
Ser Asn His Gly Pro Asp Ala Thr Glu Ala Glu Glu Asp Phe Val Asp
 1 5 10 15
 Pro Trp Thr Val Gln Thr Ser Ser Ala Lys Gly Ile Asp Tyr Asp Lys
 20 25 30
 Leu Ile Val Arg Phe Gly Ser Ser Lys Ile Asp Lys Glu Leu Ile Asn
 35 40 45
 Arg Ile Glu Arg Ala Thr Gly Gln Arg Pro His His Phe Leu Arg Arg
 50 55 60
 Gly Ile Phe Phe Ser His Arg Asp Met Asn Gln Val Leu Asp Ala Tyr
 65 70 75 80
 Glu Asn Lys Lys Pro Phe Tyr Leu Tyr Thr Gly Arg Gly Pro Ser Ser
 85 90 95
 Glu Ala Met His Val Gly His Leu Ile Pro Phe Ile Phe Thr Lys Trp
 100 105 110
 Leu Gln Asp Val Phe Asn Val Pro Leu Val Ile Gln Met Thr Asp Asp
 115 120 125
 Glu Lys Tyr Leu Trp Lys Asp Leu Thr Leu Asp Gln Ala Tyr Gly Asp
 130 135 140
 Ala Val Glu Asn Ala Lys Asp Ile Ile Ala Cys Gly Phe Asp Ile Asn
 145 150 155 160
 Lys Thr Phe Ile Phe Ser Asp Leu Asp Tyr Met Gly Met Ser Ser Gly
 165 170 175
 Phe Tyr Lys Asn Val Val Lys Ile Gln Lys His Val Thr Phe Asn Gln
 180 185 190
 Val Lys Gly Ile Phe Gly Phe Thr Asp Ser Asp Cys Ile Gly Lys Ile
 195 200 205
 Ser Phe Pro Ala Ile Gln Ala Ala Pro Ser Phe Ser Asn Ser Phe Pro
 210 215 220
 Gln Ile Phe Arg Asp Arg Thr Asp Ile Gln Cys Leu Ile Pro Cys Ala
 225 230 235 240
 Ile Asp Gln Asp Pro Tyr Phe Arg Met Thr Arg Asp Val Ala Pro Arg

```

                245                250                255
Ile Gly Tyr Pro Lys Pro Ala Leu Leu His Ser Thr Phe Phe Pro Ala
                260                265                270
Leu Gln Gly Ala Gln Thr Lys Met Ser Ala Ser Asp Pro Asn Ser Ser
                275                280                285
Ile Phe Leu Thr Asp Thr Ala Lys Gln Ile Lys Thr Lys Val Asn Lys
                290                295                300
His Ala Phe Ser Gly Gly Arg Asp Thr Ile Glu Glu His Arg Gln Phe
305                310                315                320
Gly Gly Asn Cys Asp Val Asp Val Ser Phe Met Tyr Leu Thr Phe Phe
                325                330                335
Leu Glu Asp Asp Asp Lys Leu Glu Gln Ile Arg Lys Asp Tyr Thr Ser
                340                345                350
Gly Ala Met Leu Thr Gly Glu Leu Lys Lys Ala Leu Ile Glu Val Leu
                355                360                365
Gln Pro Leu Ile Ala Glu His Gln Ala Arg Arg Lys Glu Val Thr Asp
                370                375                380
Glu Ile Val Lys Glu Phe Met Thr Pro Arg Lys Leu Ser Phe Asp Phe
385                390                395                400
Gln

```

人TrpRS构建设体概要



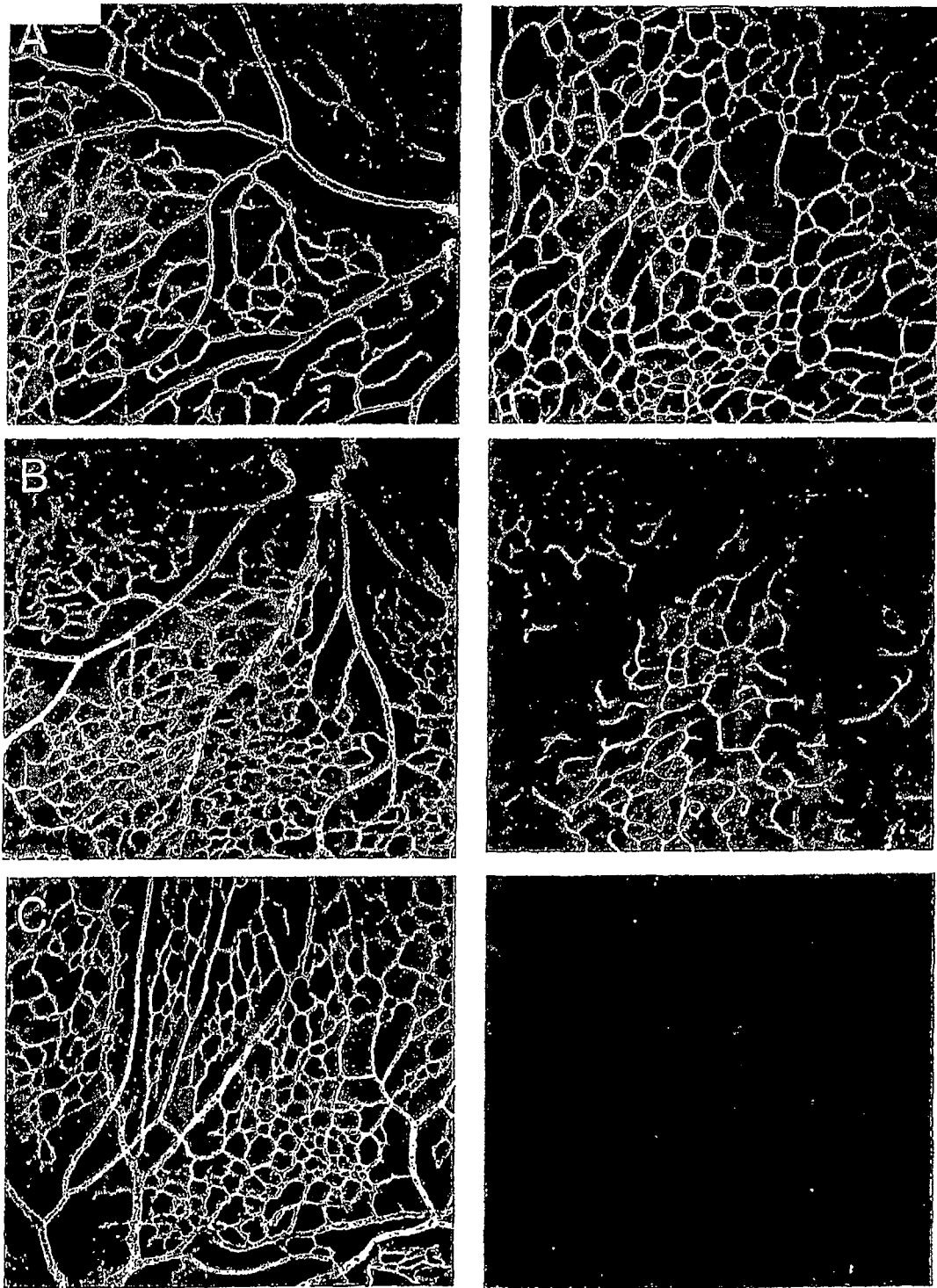
• 注：制备了四种蛋白中每一种的突变体，其中DLT (205-207) 被ELR替代

```

1 MPNSEPASLL ELFNSTATQG ELVRSLLKAGN ASKDEIDSAV KMLVSLKMSY KAAAGEDYKA DCPPGNPAPT SNHGPDATEA
81 EEDFVDPWTV QTSSAKGIDY DKLIVRFGSS KIDKELINRI ERATGQRPFH FLRRGIFFSH RDMNQVLDAY ENKPPFYLYT
161 GRGFSSEARM VGHLLIPFIET KWLQDVFNVP LVIQMTDDEK YLWKDLTLTLDQ AYGDAVENAK DIIACGFDIN KTFIFSDLDY
241 MGNSSGFYKN VVKIQKHVTF NOVKGIFGFT DSDCIGKISF PAIQAAPSES NSFPQIFRDR TDIQCLIPCA IDQDPYFRMT
321 RDVAPRIGYP KPALLHSTFF PALQGAQTKM SASDPNSSIF LTDTAKQIKT KVNKHAFSGG RDTIEHRQF GGNCDDVDSF
401 NYLTFFFLEDD DKLEQIRKDY TSGAMLTGEL KRALIEVIQP LIAEHQARRK EVTDEIVKEF MTPRKLISDFD Q
    
```

• 注：所有都是重组构建设体并且有N末端Met和C末端KLAALAEHHHHHH

图 1



初生层

次生层

图 2

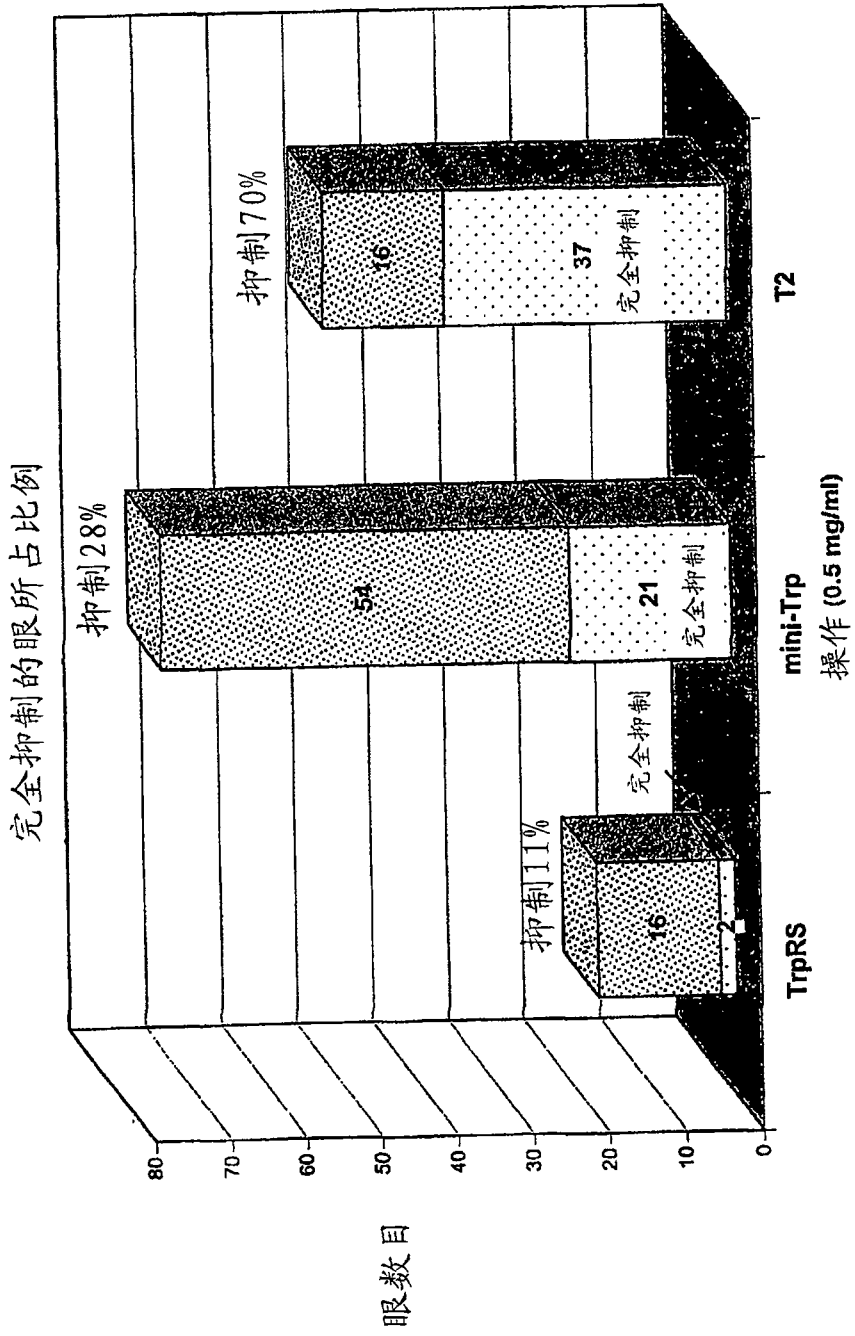


图 3

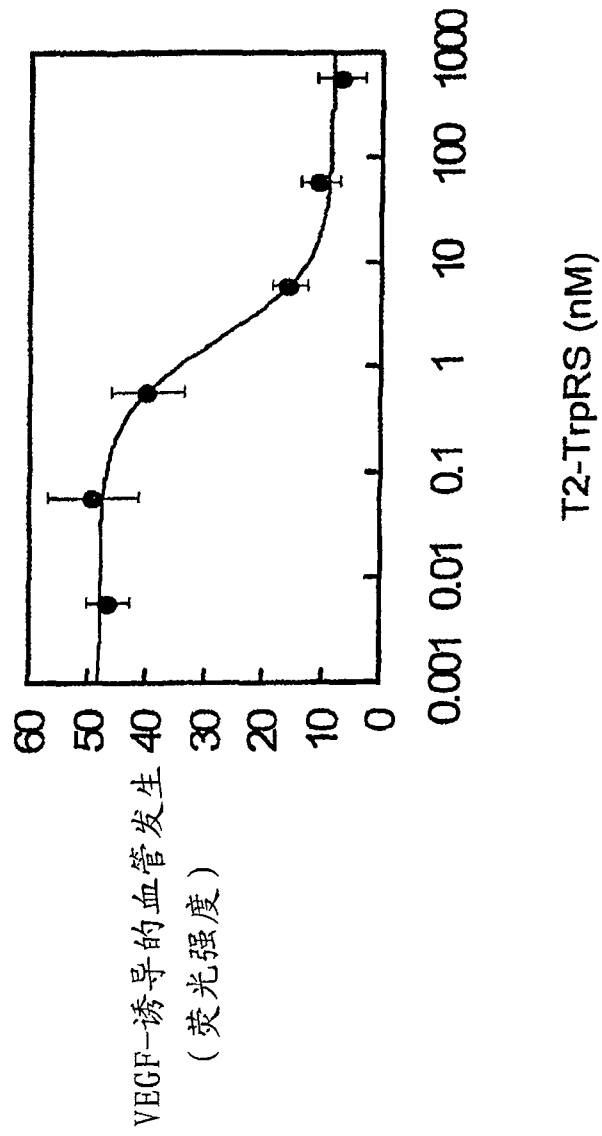


图 4

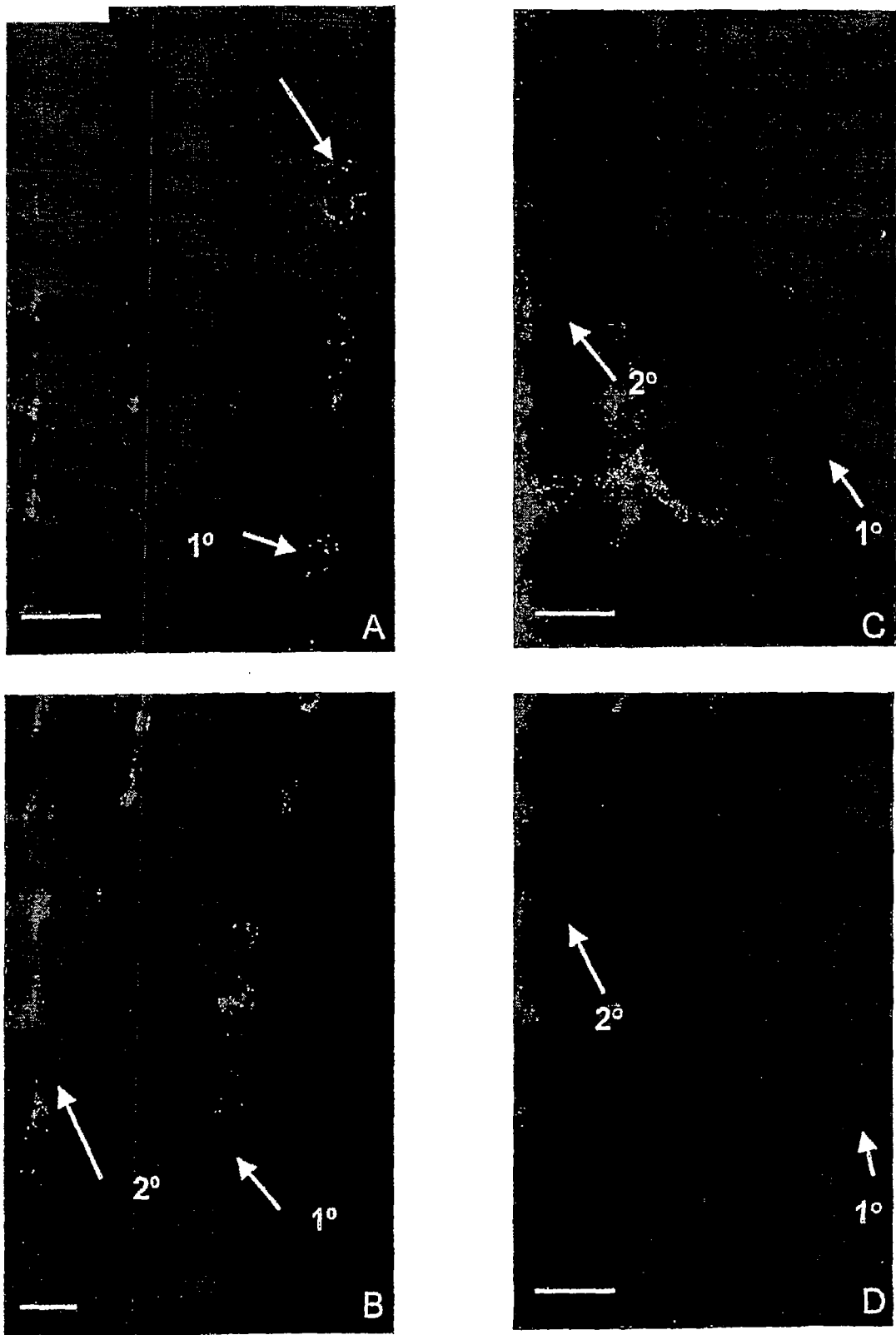


图 5