



(12) 发明专利

(10) 授权公告号 CN 1774500 B

(45) 授权公告日 2011.03.02

(21) 申请号 200480010147.3

WO 0129208 A1, 2001.04.26, 全文.

(22) 申请日 2004.02.13

US 2002065213 A1, 2002.05.30, 全文.

(30) 优先权数据

60/448,179 2003.02.14 US

C. B. BREWER. Cytomegalovirus Plasmid

Vectors for Permanent Lines of Polarised

Epithelial cell. METHODS IN CELL BIOLOGY

43. 1994, (43), 233-245.

(85) PCT申请进入国家阶段日

2005.10.14

D. R. THOMSEN. PROMOTER REGULATORY

REGION OF THE MAJOR IMMEDIATE EARLY GENE OF

HUMAN CYTOMEGALOVIRUS. PROC. NATL. ACAD.

SCI 81. 1984, (81), 659-663.

(86) PCT申请的申请数据

PCT/US2004/004407 2004.02.13

(87) PCT申请的公布数据

WO2004/074439 EN 2004.09.02

审查员 罗霄

(73) 专利权人 比奥根艾迪克 MA 公司

地址 美国马萨诸塞州

(72) 发明人 威廉·P·西斯克 霍利·普伦蒂斯

(74) 专利代理机构 北京市柳沈律师事务所

11105

代理人 巫肖南 封新琴

(51) Int. Cl.

C12N 5/02 (2006.01)

C07H 21/04 (2006.01)

(56) 对比文件

US 6194176 A1, 2001.02.27, 全文.

权利要求书 2 页 说明书 18 页 序列表 10 页

附图 13 页

(54) 发明名称

用于瞬时或稳定表达外源分子的表达盒和载体

(57) 摘要

本发明提供了用于表达多核苷酸的表达盒和含有所述表达盒的载体。所述表达盒包括启动子/增强子, 插入区, 以及聚腺苷酸化信号结构域。本发明还提供了表达系统以及所述表达盒及载体的使用方法。

1. 表达盒,其中含有:
人 CMV 立即早期 1(hCMV IE1) 启动子 / 增强子区域;
目的多核苷酸 ;和
变体人生长激素 (hGHv) polyA 信号结构域,和
长度可变的间插序列 (VLIVS),所述 VLIVS 含有剪接供体和剪接受体位点,并含有 hCMV IE1 基因的内含子 A

其中所述信号结构域保留了发出转录终止信号和 / 或稳定 mRNA 的能力,其中所述 hGHv polyA 信号结构域在高度严谨条件下与含有 SEQ ID NO :18 或 19 所示核苷酸序列的 ssDNA 杂交。

2. 权利要求 1 所述的表达盒,其中所述的 hCMV IE1 启动子 / 增强子区域含有 SEQ ID NO :1 中 x_1 到 x_2 所示的序列,其中 x_1 表示在 SEQ ID NO :1 中第 1- 第 20 位之间的核苷酸, x_2 表示在 SEQ ID NO :1 中第 715- 第 720 位之间的核苷酸。

3. 权利要求 1 所述的表达盒,其含有 SEQ ID NO :1 所示的序列。

4. 权利要求 1 所述的表达盒,其含有与 SEQ ID NO :1 所示的序列互补的序列。

5. 权利要求 1 所述的表达盒,其中所述的 hCMV IE1 启动子 / 增强子区域含有与 SEQ ID NO :1 的核苷酸 1-719 或其互补序列具有至少 100% 同一性的序列。

6. 权利要求 1 所述的表达盒,其含有 SEQ ID NO :1 的第 1- 第 1867 位核苷酸。

7. 权利要求 1 所述的表达盒,其中所述 VLIVS 含有 hCMV IE1 基因的内含子 A,该基因具有在所述内含子 A 的剪接受体和剪接供体之间的缺失。

8. 权利要求 1 所述的表达盒,其中所述的 VLIVS 含有 SEQ ID NO :1 中 x_3 到 x_4 所示的序列,其中 x_3 表示在 SEQ ID NO :1 中第 715- 第 720 位之间的核苷酸, x_4 表示在 SEQ ID NO :1 中第 1236- 第 1254 位之间的核苷酸。

9. 权利要求 1 所述的表达盒,还包含其它启动子 / 增强子元件或调节区。

10. 权利要求 9 所述的表达盒,其中所述启动子 / 增强子元件是 SV40 启动子 / 增强子元件。

11. 权利要求 9 的表达盒,其中所述调节区是 SV40 聚腺苷酸化区。

12. 权利要求 1 的表达盒,还包含选择标记。

13. 权利要求 12 的表达盒,其中所述选择标记选自以下物质组成的组:二氢叶酸还原酶, GPF, 新霉素, Hygro, Zeocin[™], 单纯疱疹病毒胸苷激酶, 腺嘌呤磷酸核糖转移酶, 次黄嘌呤 - 鸟嘌呤磷酸核糖转移酶, 嘌呤霉素 N- 乙酰转移酶或腺苷脱氨酶。

14. 权利要求 12 的表达盒,其中所述选择标记是二氢叶酸还原酶。

15. 权利要求 1 所述的表达盒,其中所述目的多核苷酸编码治疗剂。

16. 权利要求 1 所述的表达盒,其中所述目的多核苷酸相对于启动子位于反义方向。

17. 权利要求 1 所述的表达盒,其中所述目的多核苷酸是异源多核苷酸。

18. 权利要求 1 所述的表达盒,其中所述目的多核苷酸还包括转录控制元件。

19. 权利要求 18 所述的表达盒,其中所述转录控制元件选自转录终止信号,聚腺苷酸化结构域和下游增强子元件组成的组。

20. 权利要求 1 所述的表达盒,其中所述目的多核苷酸编码用于诊断或治疗的多肽。

21. 权利要求 1 所述的表达盒,其中所述目的多核苷酸编码用作疫苗的抗原多肽。

22. 权利要求 1 所述的表达盒,其中所述 hGHv polyA 信号结构域含有 SEQ ID NO:19 的至少 100 个连续核苷酸。

23. 权利要求 22 所述的表达盒,其中所述 hGHv polyA 信号结构域包含 SEQ ID NO:19。

24. 权利要求 1 所述的表达盒,其为裸露的核酸构建体形式。

25. 表达载体,其中含有权利要求 1-24 中任一项所述的表达盒。

26. 宿主细胞,其中含有权利要求 25 所述的表达载体。

27. 宿主细胞,其中含有权利要求 1-24 中任一项所述的表达盒。

28. 制备多肽的方法,包括使权利要求 26 或 27 的宿主细胞增殖并使目的多核苷酸表达。

29. 权利要求 26 或 27 的宿主细胞在制备用于向需要治疗的动物递送治疗剂的递送物质中的用途。

30. 权利要求 26 或 27 的宿主细胞在制备用于基因治疗的药物组合物中的用途。

31. 遗传载体 pV40,如图 2 所示。

32. 遗传载体 pV60,如图 6 所示。

33. 遗传载体 pV70,如图 3 所示。

34. 遗传载体 pV80,如图 10 所示。

35. 遗传载体 pV90,如图 11 所示。

36. 遗传载体 pXLC. 1,如图 4 所示。

37. 遗传载体 pXLC. 2,如图 5 所示。

38. 遗传载体 pXHC. 1,如图 7 所示。

39. 遗传载体 pXHC. 3,如图 8 所示。

40. 遗传载体 pXHC. 5,如图 9 所示。

用于瞬时或稳定表达外源分子的表达盒和载体

技术领域

[0001] 本发明涉及表达载体和表达盒,更具体地说涉及用于在生物体中表达外源分子的方法、组合物和系统。

背景技术

[0002] 将核酸分子,多肽,肽,和小分子导入靶细胞及组织不仅被用作治疗递送系统,还被用于在体外制备治疗分子。该方法的适用性随着对宿主细胞以及细胞分裂、分化及表达机理的分子生物学的进一步理解而增加。

[0003] 发明概述

[0004] 现已发现:由 CMV 启动子驱动、并由变体人生长激素基因(hGHv)的 polyA 结构域终止的转录比不含有这两种元件或其中之一的其它表达载体更有效。因此,本发明提供了用于表达目的多核苷酸的表达盒及表达载体。本发明所述表达盒包括调控元件的组合,所述调控元件的组合能够提供有效的转录,有效的转录终止,以及转录产物的增加的 mRNA 稳定性。实施方案中,所述表达盒包括人巨细胞病毒启动子/增强子,克隆位点或目的多核苷酸,以及 hGHv 聚腺苷酸化信号结构域。任选地可存在长度可变的间插序列(variable length intervening sequence)。

[0005] 本发明提供表达盒,其中包括人 CMV 立即早期 1(hCMV IE1)启动子/增强子区域,目的多核苷酸,以及变体人生长激素(hGHv)polyA 信号结构域或其变体。所述 polyA 信号结构域或其变体的长度为至少 100 个核苷酸,并含有序列 AATAAA,并且与 hGH polyA 信号结构域具有至少 92% 的同一性。

[0006] 本发明还提供了含有本发明所述表达盒的表达载体,以及含有本发明所述表达盒或表达载体的宿主细胞。

[0007] 本发明还提供了表达盒,其中包含人 CMV 立即早期 1(hCMV IE1)启动子/增强子区域,含有剪接供体和剪接受体位点的长度可变的间插序列(VLIVS),目的多核苷酸,及变体人生长激素(hGHv)polyA 信号结构域或其变体。所述 polyA 信号结构域或其变体的长度为至少 100 个核苷酸,其中含有序列 AATAAA,并与 hGHv polyA 信号结构域具有至少 92% 的同一性。

[0008] 本发明还提供了表达载体,其中包含人 CMV 立即早期 1(hCMV IE1)启动子/增强子区域,其中含有剪接供体位点和剪接受体位点的长度可变的间插序列(VLIVS),克隆位点,hGH 聚腺苷酸化区域,以及选择标记。在本发明的一个方面,所述 hCMV IE1 启动子/增强子区域位于所述克隆位点的上游(5'),所述 hGH 聚腺苷酸化区域位于该克隆位点的下游(3')。

[0009] 本发明还包括在体内递送目的药剂的方法。该方法包括:将含有表达盒的组合物递送至受试者,所述表达盒包含:hCMV IE1 启动子/增强子区域;含有剪接供体和剪接受体位点的长度可变的间插序列;编码所述目的药剂的多核苷酸;以及人生长激素(hGHv)polyA 信号结构域或其变体。

[0010] 本发明还包括表达系统。所述表达系统包括经由本发明表达盒转染或转化的宿主细胞,其中所述宿主细胞在使目的多核苷酸表达的条件下培养;以及回收所述目的药剂。

[0011] 本发明的一个或多个实施方案的详述见下文的附图和说明书。本发明的其它特点、目的和优势可以从下面的说明和附图以及权利要求部分明显地看出。本申请中引用的所有文献在此引入作为参考。

附图说明

[0012] 图 1 是 pV10 载体的质粒图谱。通过 PCR 制备巨细胞病毒立即早期 1 (CMV IE1) 启动子 / 内含子 (IVS), 并将其克隆入 HindIII 和 BamHI 位点。另外还标明了氨苄西林抗性基因 β 内酰胺酶 (bla)。

[0013] 图 2 是 pV40 的质粒图谱。标明了巨细胞病毒立即早期 1 (CMV IE1) 启动子, 内含子 (IVS), 长约 600 个碱基对的 hGHv 聚腺苷酸化信号结构域 (polyA), 以及氨苄西林抗性基因 β 内酰胺酶 (bla)。

[0014] 图 3 是质粒 pV70 的示意图。图中标出了 CMV IE1 启动子, 含有缺失连接 (deletion junction) 以及剪接供 (SD) 和剪接受体 (SA) 位点的 IVS, hGHvpolyA 信号结构域, 和 bla 基因。所述缺失连接表示的是 BspE1' // HpaI 的平端连接结果。

[0015] 图 4 是制备 pXLC. 1 载体的示意图。所述 pXLC. 1 载体是使用表达载体 pV70 和含有轻链编码序列的 PCR 产物构建而成的。用 BamHI 将 PV70 线性化, 用 BamHI 消化所述 PCR 产物, 然后连接这两者生成所述 pXLC. 1 载体。

[0016] 图 5 是制备 pXLC. 2 载体的示意图。

[0017] 图 6 图示了 pV60 载体以及 pXHC 载体的制备。所述 pXHC 载体是使用表达载体 pV60 和含有大部分重链编码序列 (不包括 N- 末端的 52 个氨基酸) 构建而成的。用 BamHI 将 PV60 线性化, 用 BamHI 消化所述 PCR 产物, 然后将二者连接在一起。

[0018] 图 7 是 pXHC. 1 载体的示意图。

[0019] 图 8 是 pXHC. 3 载体的示意图。

[0020] 图 9 是图示了将 dhfr 盒添加到 pXHC. 3 得到的 pXHC. 5 载体。所述表达盒的调控元件源自于 pSI (Promega, Genbank 登录号 #U47121), 其中含有 SV40 启动子 / 增强子, 人工内含子和 SV40 晚期聚腺苷酸化序列。

[0021] 图 10 是载体 pV80 的示意图。标出的是: 含有剪接供体 (SD) 和剪接受体 (SA) 位点的巨细胞病毒立即早期 1 (CMV IE1) 启动子 / 内含子 (IVS) 片段, hGHv 聚腺苷酸化信号结构域 (polyA), 氨苄西林抗性基因, β 内酰胺酶 (bla), SV40 启动子 / 增强子, 人工内含子和 SV40 晚期聚腺苷酸化序列。

[0022] 图 11A 和 11B 是载体 pV90 的示意图及相应的标注 (annotated) 序列。图 11A 是载体图谱。标出的是: 含有剪接供体 (SD) 和剪接受体 (SA) 位点的巨细胞病毒立即早期 1 (CMV IE1) 启动子 / 内含子 (IVS) 片段, hGHv 聚腺苷酸化信号结构域 (polyA), 氨苄西林抗性基因, β 内酰胺酶 (bla), SV40 启动子 / 增强子, 人工内含子和 SV40 晚期聚腺苷酸化序列。所述载体不含 dhfr 表达盒中的 NotI 位点。图 11B 是 pV90 载体的标注序列。图 11B 中, SEQ IDNO:19 中的第 1275-1866 位核苷酸代表长度约 600 个核苷酸的 hGHv polyA。

[0023] 图 12 是 pSI-DHFR. 2 的载体图谱。SV40 启动子驱动 dhfr 基因。

[0024] 图 13 图示了经转染的合并物的相对比产率。3 个合并物用来分析 pCMV-hGHvPA-SEAP 和 pSEAP2, 一个合并物用于分析 pUC18。

[0025] 图 14 是显示所述分离株的相对比产率的一对图表。每一种构建体的前 20 种分离株按排列顺序给出。

[0026] 图 15 给出的是 Genbank 登录号 No. K00470 中的序列 (SEQ ID NO :18)。

[0027] 发明详述

[0028] 本发明涉及表达盒以及含有该表达盒的载体。所述表达盒含有能够驱动真核宿主中的转录的转录调控区域以及转录终止区域。本发明所述表达盒和载体提供了强的转录起始和终止以及转录产物的增强的 mRNA 稳定性。

[0029] 本发明提供了来自任一株巨细胞病毒的启动子和任选的增强子元件, 例如本申请所述的或者例如美国专利 No. 5, 658, 759 等文献中所述的巨细胞病毒毒株, 这些文献的内容在本申请引入作为参考。例如, 可用于本发明所述表达盒的合适的 CMV 立即早期启动子区域可以从 CMV- 启动的 β - 半乳糖苷酶表达载体 CMV β 获得 (MacGregor et al., Nucl. Acids Res. 17 :2365(1989))。

[0030] 如本申请中进一步所述, 所述 hGHv 聚腺苷酸化信号结构域提供了强的转录终止信号, 以及提高的 mRNA 转录物稳定性。所述调控 / 表达元件可以通过例如, 目的多核苷酸或克隆位点 (例如, 多克隆位点)。自 hGHv 聚腺苷酸化信号结构域分离。

[0031] 本发明的一个方面, 提供了其中含有表达盒的多核苷酸, 该表达盒包括巨细胞病毒 (CMV) 转录调控区域, 长度可变的间插序列 (例如, 来自 CMV 的内含子 A), 目的多核苷酸, 和聚腺苷酸化信号结构域。本发明进一步涉及用于从宿主细胞制备和回收异源多肽的方法和表达载体。

[0032] 另一方面, 本发明的表达盒包括, 可操作地连接的如下序列 (i)CMV 主要立即早期 1 (IE1) 启动子 / 增强子区域和长度可变的间插序列 (例如, 内含子 A 的衍生物), (ii) 目的多核苷酸, 和 (iii)hGHv 聚腺苷酸化信号结构域。所述术语“可操作地连接的”是指并置 (juxta position), 其中的各成分间的相互关系使得它们可以其预定方式发挥作用 (例如, 功能性地连接)。因此, 例如, 可操作地连接于目的多核苷酸的启动子 / 增强子以这样一种方式与后者连接, 即使得能够在与活化所述启动子 / 增强子表达相容的条件下完成所述目的多核苷酸的表达。

[0033] 在本发明的具体实施方案中, 所述表达盒包括 SEQ ID NO :1 中的约第 1- 约第 1867 位核苷酸所示的序列 (例如, 约第 1- 第 1865, 1866, 1867, 1868, 或 1869 位核苷酸)。SEQ ID NO :1 中第 1-1867 位核苷酸所示表达盒包含大量不同的结构域, 例如 CMV IE1 启动子 / 增强子区域, 所述结构域具有 SEQID NO :1 中约 x_1 - 约 x_2 所示的序列, 其中 x_1 是第 1-20 位核苷酸中的核苷酸, x_2 是约第 715-720 位核苷酸 (例如, SEQ ID NO :1 中约第 1-719 位核苷酸) 中的核苷酸。所述表达盒的另一结构域包括长度可变的间插序列 (VLIVS), 其中含有剪接供体和剪接受体位点。所述 VLIVS 可以为长度至少 50bp (例如, 长度至少 100, 150, 200, 或 250bp), 而且可以包括来自本领域任一已知来源的剪接供体和受体。参见, 例如, Varani et al., Annu Rev BiophysBiomol Struct 27 :407-45(1998) 和 Koning, Eur J Bi° Chem 219 :25-42(1994)。合适的插入结构域可以包括任意 CMV 毒株基因组的内含子 A 的全部, 或可以包括较小的片段, 这种较小片段含有连接在一起的含剪接供体位点的 5' 序列和含剪接

受体位点的 3' 序列。例如,所述 VLIVS 包括 SEQ ID NO:1 中约 x3- 约 x4 的核苷酸,其中 x3 是第 715-720 位核苷酸中的核苷酸, x4 是第 1236-1254 位核苷酸(例如, SEQ ID NO:1 中第 719-1236 位核苷酸)中的核苷酸。位于 CMV IE1 启动子 / 增强子之后的间插序列的大小可以变化,可多达为 SEQ ID NO:1 中的 317 个核苷酸。例如,从 pV40 和 pV70(分别见例如图 2 和 3)所示的 IVS 序列缺失 317 个核苷酸,来制备 SEQ ID NO:1 的 VLIVS。因此,在本发明的另一方面,所述表达盒包括 SEQ ID NO:1 的约第 1- 约第 1254 位核苷酸的序列(例如,CMV IE1 启动子 / 增强子和间插序列)。多克隆位点可以存在于 IVS 区域(例如,包括 BamHI 位点和 NotI 位点的 SEQ ID NO:1 的第 1255-1272 位核苷酸)之后(即,下游)。可以使用本领域技术人员已知技术,在所述表达盒中构建不同的或另外的限制酶切位点。所述表达盒还可以进一步包括 polyA 结构域。

[0034] 所述 polyA 信号结构域源自于 hGHv 基因,该基因的 3' UTR 序列可变,例如,从等位基因到等位基因。hGHv 基因的一种等位基因的描述见 Genbank 登录号 No. K00470(SEQ ID NO:18),另一序列为图 11B 中描述的 SEQ ID NO:19,该序列对应于 SEQ ID NO:18 中第 2032-2625 位核苷酸(参见图 15)。polyA 信号结构域的非天然存在的变体可以通过诱变技术制备,包括适于多核苷酸、细胞或生物体的技术。hGHv 基因的 polyA 变体包括 polyA 信号结构域,该结构域不同于野生型 hGHv polyA 信号结构域但是仍保留了发出转录终止信号和 / 或稳定 mRNA 的能力。例如,所述聚腺苷酸化信号结构域可以包括 SEQ ID NO:18 或 19 所示的 hGHv 聚腺苷酸化信号结构域序列。分子生物学领域的技术人员还应理解,所述序列的长度无需约为 600 个核苷酸。相反,包含任何这样的 polyA 序列结构域,其包含 hGHv 基因的至少 100nt(例如,至少 200,300,400,500,或 600nt)的连续核苷酸序列(包含标准(canonical)AATAAA 位点)。此外,本发明还包括与上述序列之间的差异达 8%(例如,与 SEQ ID NO:18 或 19 或其特征结构域具有 92%同一性)的序列。例如,长度为 100nt 的多核苷酸,其与 SEQ ID NO:1 的第 1-1867 位核苷酸具有 95%同一性,并含有可保持终止转录的能力的 AATAAA 序列。

[0035] 本发明的另一方面中,提供了含有表达盒的载体。在本申请中,“载体”是指一经导入受体细胞(例如,细菌如大肠杆菌(E. coli)),即能够自主复制的核酸分子(DNA 或 RNA)。质粒、病毒和噬菌体是载体的实例。从表达载体“表达”的过程是已知的,包括细胞酶的使用和从目的多核苷酸制备表达产物的过程(processes)。表达载体是指能够介导克隆的多核苷酸在宿主细胞中表达的载体,所述宿主细胞与用于载体复制或增殖的细胞可以相同或不同类型。许多哺乳动物表达载体能够在常见细菌(受体细胞)中增殖,但在哺乳动物细胞(宿主细胞)中表达目的多核苷酸而不能在细菌中表达。

[0036] 本发明涉及那些能够允许多核苷酸在真核(例如,哺乳动物,更具体地为,人,鼠,猿,牛,猪,啮齿动物,或绵羊细胞)细胞中有效转录和翻译的载体的设计和使用。本发明所述载体包括上述表达盒,其中含有能提供有效转录终止和 mRNA 稳定性的聚腺苷酸化信号结构域。

[0037] 本发明所述载体包括:用于接纳目的多核苷酸的克隆位点;足以允许被插入宿主细胞中克隆位点的多核苷酸的转录的转录调控元件(例如,CMV IE1 启动子 / 增强子区域);足以允许所述多核苷酸的 RNA 转录物在宿主细胞中翻译的翻译元件,以及(如果需要)足以允许所述载体在宿主细胞或用于所述载体的增殖的另一受体细胞中进行复制的复制元

件。本发明所述载体能介导在宿主细胞中的所述瞬时或稳定表达。

[0038] 在具体的实施方案中,本发明载体包括(1)SEQ ID NO:1所示序列;(2)与SEQ ID NO:1所示序列互补的序列;(3)与SEQ ID NO:1至少80%(优选至少90%;95%;98%或99%)相同的序列或其互补序列;或(4)载体,其含有SEQ ID NO:1中的约第1-约第1867位核苷酸并且含有目的多核苷酸和/或选择标记。

[0039] 含有SEQ ID NO:1的载体具有大量不同的结构域和编码区域。例如,具有SEQ ID NO:1中约 x_1 -约 x_2 所示序列的CMV IE1启动子/增强子区域存在于所述载体中,其中 x_1 是第1-20位核苷酸中的核苷酸, x_2 是约第715-720位核苷酸(例如,SEQ ID NO:1中约1-719位核苷酸)中的核苷酸。本发明表达载体的另一结构域包括含有剪接供体和剪接受体位点的长度可变的间插序列(VLIVS)。例如,所述IVS包括SEQ ID NO:1的约 x_3 -约 x_4 ,其中 x_3 包括第715-720位核苷酸中的核苷酸, x_4 包括第1236-1254位核苷酸(例如,SEQ ID NO:1中大约第719-1236位核苷酸)中的核苷酸。所述表达载体的多克隆位点包括SEQ ID NO:1的第1255-1272位核苷酸(例如,BamHI位点和NotI位点)。可以使用本领域技术人员已知技术在所述表达载体中构建不同的或其它的限制酶切位点。所述表达载体还可以进一步包括polyA信号结构域。所述polyA信号结构域是hGHv polyA信号结构域或hGH polyA信号结构域的其它变体。例如,polyA信号结构域包括SEQ ID NO:19所示的hGHv polyA信号结构域序列。另外,本发明载体中还含有一种或多种选择标记。例如,SEQ ID NO:1包括二氢叶酸还原酶(dhfr)基因(例如,SEQ ID NO:1中约第2568-约第3132位核苷酸)。除上述的那些元件之外,本发明载体可以包括另外的启动子/增强子元件和调控区域(例如,聚腺苷酸化结构域)。所述另外的调控元件和聚腺苷酸化结构域可以位于选择标记或目的多核苷酸的侧翼(例如,紧密邻接于其5'和3'端)。例如,所述含有SEQ ID NO:1的载体包括SEQ ID NO:1中约第2568-约第3132位核苷酸所示的二氢叶酸还原酶(dhfr)基因。所述dhfr基因侧接SV40启动子/增强子元件和SV40聚腺苷酸化区域(例如,分别为SEQ ID NO:1的约第1868-约第2210位核苷酸和约第3144-约第3440位核苷酸)。

[0040] 选择标记的具体实例是编码下述蛋白质的选择标记,所述蛋白质可赋予针对细胞抑制药物或杀细胞药物的抗性,所述蛋白质例如,能赋予抗氨甲蝶呤抗性的DHFR蛋白质(Wigler et al., Proc. Natl. Acad. Sci. USA 77:3567(1980); O'Hare et al., Proc. Natl. Acad. Sci. USA 78:1527(1981));能赋予抗霉酚酸(mycophenolic acid)抗性的GPF蛋白(Mulligan & Berg, Proc. Natl. Acad. Sci. USA 78:2072(1981)),能赋予抗氨基糖苷G-418抗性的新霉素抗性标记物(Colberre-Garapin et al., J. Mol. Biol. 150:1(1981));能赋予抗潮霉素抗性的Hygro蛋白(Santerre et al., Gene 30:147(1984));和ZeocTM抗性标记(可购自Invitrogen)。此外,单纯疱疹病毒胸苷激酶(Wigler et al., Cell 11:223(1977)),次黄嘌呤-鸟嘌呤磷酸核糖转移酶(hypoxanthine-guanine phosphoribosyltransferase)(Szybalska & Szybalski, Proc. Natl. Acad. Sci. USA 48:2026(1962)),以及腺嘌呤磷酸核糖转移酶(Lowy et al., Cell 22:817(1980))可以分别应用于 tk^- , $hgprt^-$ 或 $aprt^-$ 细胞中。其它选择标记编码嘌呤霉素N-乙酰转移酶或腺苷脱氨酶。

[0041] 对两个或多个核酸分子来说,当使用比较算法或者通过人工比对及目测进行检测,在比较窗比较并比对以得到最大对应度窗口时,术语“相同”或百分比“同一性”是指两个或多个序列或亚序列是相同的或者具有特定比例的相同核苷酸。这些定义也指序列的互补

序列（例如，SEQ ID NO:1 所示序列的互补序列或其含有表达盒的片段）。例如，所述表达盒及其片段包括与 SEQ ID NO:1 部分序列（例如，SEQ ID NO:1 的 1-719, 1-1254 位核苷酸等）具有至少约 80%，约 90%，约 95%，约 97%，约 98% 或约 99% 核苷酸序列同一性的那些序列。因此，如果序列与 SEQ ID NO:1 的全长序列或其结构域之间具有所需的序列同一性，那么该序列就可以分别作为本发明的表达盒或结构域而起作用。

[0042] 就序列比较而言，通常将一段序列用作参照序列，将受试序列与之进行比较。当使用序列比较算法时，将受试及参照序列输入电脑，设定亚序列坐标（如有必要），并设定序列算法的程序参数。可以使用默认的程序参数，或者也可以设定替换参数。然后，根据设定的参数或默认的程序参数，用所述序列比较算法计算受试序列相对于参照序列的序列同一性百分比。在本申请中，“比较窗”，包括参照选自具有下列数目之一连续核苷酸的任一片段：25-600 个，通常约 50- 约 200 个，更通常为约 100- 约 150 个，其中可以在将序列与具有同等数目个连续核苷酸的参照序列进行最佳比对之后，对这两个序列进行比较。用于比较的序列比对法已为本领域所熟知。用于比较的序列最佳比对可以用以下方法进行：例如，Smith&Waterman, Adv. Appl. Math. 2:482(1981) 的局部同源性算法，Needleman&Wunsch, J. Mol. Biol. 48:443(1970) 的同源性比对算法，Pearson&Lipman, Proc. Natl. Acad. Sci. USA 85:2444(1988) 的相似性搜索算法，这些算法的电脑化运行（GAP, PILEUP, BESTFIT, FASTA, 和 TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI），或通过人工比对以及目测检查。

[0043] 适于测定序列同一性百分比（即，基本上的相似性或同一性）的算法的一个实例是 BLAST 算法，其描述见 Altschul, J. Mol. Biol. 215:403-410, 1990。用于执行 BLAST 分析的软件公众可以从 the National Center for Biotechnology Information (on the World Wide Web at ncbi.nlm.nih.gov/) 获得。这种算法包括：首先通过鉴定询问序列中的短字长 W (short words of length W)，鉴定出高得分序列对 (HSP)，它们或者相互匹配，或者当与数据库序列中等长的字符比对时满足一些正值的阈值得分 (positive-valued threshold score) T。“T” 是指邻近字符得分阈值 (neighborhood word score threshold)。这些起始邻近击中字符串 (initial neighborhood word hits) 作为种子 (seed) 启动检索以发现含有它们的更长的 HSP。然后，沿每一序列的双向延伸得到的击中字符串，只要能增加累积的比对得分即可。使用参数 M (成对匹配残基的奖励得分；永远 > 0) 和 N (错配残基的处罚得分；永远 < 0) 计算核苷酸序列的累积得分。当达到下述情况时击中字符串的双向延伸停止：累积比对得分比其最大获得值低数量 X；由于一或多个阴性 - 得分残基比对的累积，所述累积得分降至零或零以下；或者到达任一序列末端。BLAST 算法参数 W、T 和 X 决定比对的灵敏度和速度。一个实施方案中，为了测定一种核酸序列是否落于本发明范围之内，使用 BLASTN 程序（对于核苷酸序列），引入字长 (W) 11 的缺省值 (default)，期望值 (E) 10，M = 5，N = 4，以及对两条链进行比较。对于氨基酸序列而言，BLASTP 程序使用下列缺省参数：字长 (W) 3，期望值 (E) 10，以及 BLOSUM62 得分矩阵（参见，例如，Henikoff, Proc. Natl. Acad. Sci. USA 89:10915, 1989）。

[0044] 另外还使用 BLAST 算法对两个序列之间的相似度进行了统计分析（参见，例如，Karlin, Proc. Natl. Acad. Sci. USA 90:5873-5787, 1993）。BLAST 算法提供了一种相似度测算方法是最小求和概率 (smallest sum probability) (P(N))，该方法提供了两个核苷酸

或氨基酸序列之间随机发生匹配的概率的指标。例如,如果受试核酸和参照核酸之间比较中的最小求和概率小于约 0.1,更优选小于约 0.01,以及最优选小于约 0.001,那么就认为该核酸与参照序列相似。

[0045] 本发明此外还包括能够与 SEQ ID NO:1 中约第 1-1867 位核苷酸所示多核苷酸序列特异性杂交的多核苷酸。术语“选择性(或特异性)杂交”是指在严谨杂交条件下,分子与具体参照多核苷酸的结合、双螺旋或杂交。术语“严谨杂交条件”是指这样的条件,在该条件下探针主要与其亚序列杂交,所述目标亚序列通常在核酸的复杂混合物中,但是所述探针不与其它序列杂交。严谨条件是序列-依赖性的,因环境的不同而不同,例如,取决于探针的长度。较长的序列可以在较高温度特异性杂交。核酸杂交的全面指导见 Tijssen, *Techniques in Biochemistry and Molecular Biology—Hybridization with Nucleic Probes*, “Overview of principles of hybridization and the strategy of nucleic acid assays”(1993)。通常,在一定离子强度和 pH 时,严谨条件被选择为比具体序列的热解链温度 (T_m) 低约 5-10°C。 T_m 是指这样的温度(在一定的离子强度、pH 和核酸浓度),在该温度达平衡时,互补于靶序列的探针的 50% 与所述靶序列杂交(由于靶序列过量存在,在 T_m 达平衡时,50% 的探针被结合)。严谨条件中的盐浓度小于约 1.0M 钠离子、通常约 0.01-1.0M 钠离子浓度(或其它盐),pH 为 7.0-8.3,温度对于短探针(例如,10-约 50 个核苷酸)为至少约 30°C 而对于长探针(例如,大于约 50 个核苷酸)为至少约 60°C。严谨条件还可以通过加入去稳定剂例如甲酰胺来实现。对于选择性或特异性杂交而言,阳性信号(例如,本发明核酸的鉴定)是背景杂交的约 2 倍。对于长探针而言,用于鉴定本发明范围内基本相同的核酸的“严谨”杂交条件包括:在含 50% 甲酰胺、5x SSC 和 1% SDS 的缓冲液中、42°C -65°C 进行杂交,65°C 用 0.2x SSC 和 0.1% SDS 洗涤。但是,本领域普通技术人员显然明白,杂交条件可以根据序列组成进行调整。代表性的“中度严谨杂交条件”包括:37°C、在包含 40% 甲酰胺、1M NaCl 和 1% SDS 的缓冲液中杂交,然后在 45°C、用 1x SSC 洗涤。阳性杂交应该至少是背景信号的两倍。本领域普通技术人员很容易认识到可以使用其它杂交和洗涤条件以提供类似严谨度的条件。因此,本发明的表达盒可以包括在高度严谨条件下与含 SEQ ID NO:18 或 19 所示核苷酸序列的 ssDNA 杂交的 hGHv polyA 信号结构域。

[0046] 所述表达盒可以以裸核酸构建体的形式使用。可选地,所述表达盒可作为核酸载体(例如,表达载体如上述表达载体)的一部分导入。所述载体包括质粒和病毒载体。载体可以包括位于所述表达盒侧翼的序列,包括与诸如哺乳动物基因组序列的真核基因组序列或病毒基因组序列同源的序列。这样就可以通过同源重组酶将所述表达盒导入真核细胞或病毒的基因组。例如,包含与病毒序列侧接的表达盒的质粒载体可用于制备适于将所述表达盒递送至脊椎动物的病毒载体,所述脊椎动物包括鱼、鸟类或哺乳动物细胞。所采用技术已为本领域技术人员所熟知。

[0047] 所述术语“目的多核苷酸”意欲包括至少能够被转录的核酸分子。所述分子相对于启动子可为有或反义方向。可以采用本领域熟知的技术,使用反义构建体抑制细胞中基因的表达。所述目的多核苷酸可以包括异源多核苷酸。术语异源多核苷酸包括任一基因。异源多核苷酸通常源自与表达盒或载体中与该异源多核苷酸可操作连接的调控元件所不同的物种,或者如果来自相同的来源,那么就是其原始形态的经修饰的基因。因此,可操作地连接于启动子的异源多核苷酸的来源与所述启动子的来源不同,或者,如果它们的来源

相同,那么就是其原始形态的经修饰的启动子。对异源多核苷酸的修饰可以通过以下方法进行,例如,通过用限制性内切酶处理 DNA 制得能够可操作连接于所述启动子的 DNA 片段。另外还可使用定点诱变对异源多核苷酸进行修饰。异源多核苷酸还可以包括标记基因(例如,编码 β -半乳糖苷酶或绿色荧光蛋白的基因)或者其产物能够调节其它基因表达的基因。因此,用作 mRNA、tRNA 和 rRNA 的多核苷酸也在该定义的范畴之内。所述异源基因可以是野生型基因的等位变体,或者其可以是突变基因。mRNA 可任选包括位于 5' 和 / 或 3' 端转录但不翻译的侧翼区域的一部分或全部,该区域与翻译的编码序列天然或非天然相连。

[0048] 目的多核苷酸可以任选更进一步包括通常与所转录出的分子相连的相关转录控制元件,例如转录终止信号,聚腺苷酸化结构域和下游增强子元件。所述目的多核苷酸可以编码治疗产物或者作为治疗产物的模板,所述治疗产物可以是例如肽,多肽,蛋白质或者核糖核酸。所述目的多核苷酸通常是编码多肽产物的 DNA 序列(例如 cDNA 或基因组 DNA),所述多肽产物诸如酶(例如 β -半乳糖苷酶);血液衍生物;激素;细胞因子;白介素;干扰素;TNF;生长因子(例如 IGF-1);可溶性受体分子(例如,可溶性 TNF 受体分子);神经递质或其前体;营养因子诸如 BDNF, CNTF, NGF, IGF, GMF, aFGF, bFGF, NT3 和 NT5;载脂蛋白诸如 ApoAI 和 ApoAIV;肌营养不良蛋白(dystrophin)或小肌营养不良蛋白(minidystrophin);肿瘤-抑制型蛋白诸如 p53, Rb, Rap1A, DCC 和 k-rev;凝血相关因子(factors involved in coagulation)例如因子 VII, VIII 和 IX;或者完整或部分天然或人工免疫球蛋白(例如 Fab 和 ScFv,或者克隆的 IgG 的轻链或重链)。

[0049] 目的多核苷酸还可以包括用作制备反义分子的模板,其在靶细胞中的转录使得细胞 mRNA 的基因表达或转录能够得以控制。可以用本领域已知技术,所述分子可例如在靶细胞中被转录为细胞 mRNA 的互补 RNA,从而能够阻遏细胞 mRNA 翻译为蛋白质。具体而言,反义分子可以用于在炎性细胞因子或代谢细胞因子引发的关节炎和组织丢失治疗过程中阻遏这些细胞因子的翻译。

[0050] 所述目的多核苷酸序列通常编码诊断或治疗用的多肽。可以使用含有本发明所述表达盒的各种宿主细胞(例如, COS 细胞或 CHO 细胞或其衍生物),在生物反应器中在体外制备所述多肽。可选地,本发明所述表达盒和 / 或载体可用于基因递送,蛋白质递送,和 / 或基因治疗。

[0051] 本发明还可以用于表达毒性因子和多肽。后者具体而言可以为细胞毒素(诸如白喉毒素,假单胞菌毒素和蓖麻毒素 A),能诱导针对外界因子(例如胸苷激酶和胞嘧啶脱氨酶)的敏感性的产物,或者可选能够诱导细胞死亡的因子(例如 Grb3-3 和抗 -ras ScFv)。

[0052] 治疗用途意指能够实现消除疾病或病症,治愈疾病或病症,以及 / 或者减轻疾病或病症严重程度的用途。诊断用途包括使用下述分子,所述分子能够测定分子与疾病过程之间的因果关系或相关程度或提供相关信息,或者测定疾病或病症是否存在。诊断试剂并不能直接促进疾病或病症的改善。

[0053] 另外,目的多核苷酸还可以编码可用作疫苗的抗原多肽。抗原多肽或核酸分子源自于病原生物体例如,细菌或病毒。例如,抗原多肽包括存在于病原生物体的基因组或基因产物中的抗原决定簇,所述病原生物体例如,病毒性出血性败血症(viral haemorrhagic septicemia),细菌性肾病,弧菌病(vibriosis)和疔病(furunculosis)的病原生物体。抗原性的多肽可以选自例如丙型肝炎病毒基因组的区域和基因产物。

[0054] 在本申请中,“分离的”这一词汇用来描述分子或组合物诸如本发明的载体或表达盒或目的多核苷酸时,意指所述分子或组合物与至少一种其它化合物例如蛋白质、DNA、RNA 或其它污染物分离开来,所述分子或组合物在体内或其天然存在状态下与所述其它化合物相结合。因此,当目的多核苷酸同与其天然结合的任何其它成分已经分离开来的时候,那么就可以称这种目的多核苷酸为分离的。但是,分离的组合物也可以是基本上纯的。分离的组合物可以处于均质状态 (homogeneous state)。它可以是干燥的 / 冻干的状态或水溶液。纯度和均质性可以用分析化学技术测定,例如,聚丙烯酰胺凝胶电泳 (PAGE),琼脂糖凝胶电泳或高压液相色谱 (HPLC)。

[0055] 在本申请中,术语“核酸分子”和“多核苷酸”可以交换使用,包括寡核苷酸 (即,短多核苷酸)。另外它们还可以指合成的和 / 或非天然存在的核酸分子 (例如,包括核苷酸类似物或经过修饰的主链残基或键)。另外,该术语还可指单链或双链形式的脱氧核糖核苷酸或核糖核苷酸寡核苷酸。该术语包括其中含天然核苷酸类似物的核酸。该术语还包括带有合成主链的核酸-样结构。本发明提供的 DNA 主链类似物包括磷酸二酯,硫代磷酸酯,二硫代磷酸酯,甲基-膦酸酯,氨基磷酸酯,烷基磷酸三酯,氨基磺酸酯,3'-硫缩醛,甲叉 (甲亚氨基 (methyylimino)),3'-N-氨基甲酸酯,吗啉代氨基甲酸酯,和肽核酸 (PNA);参见 Oligonucleotides and Analogues, F.Eckstein 编著, a Practical Approach, IRL Press at Oxford University Press(1991);Antisense Strategies, Annals of the New York Academy of Sciences, Volume 600, Baserga 和 Denhardt 编著 (NYAS 1992);Milligan(1993) J. Med. Chem. 36:1923-1937;Antisense Research and Applications(1993, CRC Press)。PNA 含有非离子主链,例如 N-(2-氨基乙基 (aminoethyl)) 甘氨酸单元。硫代磷酸酯键的描述见 W097/03211;W0 96/39154;Mata(1997) Toxicol. Appl. Pharmacol. 144:189-197。该术语包括的其它合成主链包括甲基-膦酸酯键或交替的甲基膦酸酯和磷酸二酯键 (Strauss-Soukup(1997) Biochemistry 36:8692-8698), 和苯甲基-膦酸酯键 (Samstag(1996) Antisense Nucleic Acid Drug Dev 6:153-156)。

[0056] 在本申请中,“重组”是指合成的或者体外操作的多核苷酸 (例如,“重组多核苷酸”),使用重组多核苷酸在细胞或其它生物系统中制备产物的方法,或者由重组多核苷酸编码的多肽 (“重组蛋白”)。重组多核苷酸包括不同来源的核酸分子,其被连接入表达盒或表达载体,所述载体用于表达例如融合蛋白或者通过多肽的可诱导型或组成型表达制备的那些蛋白质 (例如,本发明表达盒或载体可操作地连接于异源的多核苷酸,例如多肽编码序列)。

[0057] 在常见的表达系统中,从异源多核苷酸制备多肽的过程要么不受调控,要么通过下述方式实现调控:对自可操作地连接于编码异源多肽的多核苷酸上游的转录启动子的转录进行调节。但是,为了提供正确的转录终止和 mRNA 稳定性,还必须在下游进行适当调控。本发明的一个方面,在本发明表达盒或载体中目的多核苷酸的下游提供了人生长激素变体 (hGHv) 聚腺苷酸化 (polyA) 信号结构域。所述 hGHv polyA 信号结构域含有源自人生长激素基因序列的序列。所述 hGHv 聚腺苷酸化信号结构域序列在真核细胞中提供强的转录终止,并且提供增强了的 mRNA 稳定性。该 hGHv 聚腺苷酸化信号结构域具有独特的优势,胜过包括使用 CMV 启动子 / 增强子的表达盒和 / 或载体在内的现有的表达盒和 / 或载体。

[0058] 另外,还可能存在翻译元件,并意图包括允许 RNA 转录物翻译为蛋白质所必需的

专门序列（诸如核糖体结合位点和起始密码子）包括在内。翻译元件还可以包括共有序列、前导序列、剪接信号等，用于促进或提高翻译程度，或者提高表达产物的稳定性。例如，所述 hGHv 聚腺苷酸化信号结构域提供提高了的 mRNA 稳定性。本发明所述载体可以拥有辅助转录区域，例如内含子，聚腺苷酸化信号，Shine/Dalgarno 翻译信号和 Kozak 共有序列（Shine et al., Proc. Natl. Acad. Sci. (U. S. A.) 71:1342-1346(1974); Kozak, Cell 144:283-292(1986)）。

[0059] 术语“复制元件”意图包括允许载体在载体受体细胞中复制所必需的专门序列（例如复制起点）。通常，所述载体含有足以允许载体在受体细胞中自主稳定复制的至少一个复制起点。

[0060] 为了有助于本发明载体的筛选和保持，可以在所述载体中引入一或多种选择标记（例如能够赋予抗生素抗性或者使细胞能在极限培养基上或在毒性代谢产物存在的条件下生长的多核苷酸）。

[0061] 另一实施方案中，本发明涉及含有上述构建体（例如，本发明所述的表达盒或载体）的宿主细胞。本发明所述表达盒可以通过下述方式对宿主细胞进行重组修饰：转染或者转化宿主细胞以表达所需的目多核苷酸。在本申请中，术语“重组修饰”是指将本发明的表达盒或载体导入活细胞或表达系统。通常，其中含有目多核苷酸的表达盒存在于载体（例如，质粒）中。表达系统包括已导入了目多核苷酸（其产物将被表达）的活宿主细胞，如本申请所述。

[0062] 宿主细胞是指表达盒（包括其中含有表达盒的载体）能够在其中增殖并且多核苷酸编码的产物能够被表达的细胞。宿主细胞还包括目的宿主细胞或其衍生物的任一后代。应理解所有后代并不会完全等同于亲代细胞，因为在复制过程中有突变发生。但是，当使用“宿主细胞”这一术语时是将这样的后代包括在内的。可用于本发明的宿主细胞包括细菌细胞、真菌细胞（例如，酵母）、植物细胞和动物细胞。例如，宿主细胞可以是较高等的真核细胞例如哺乳动物细胞等，或者较低等的真核细胞例如酵母等，或者所述宿主细胞可以是原核细胞例如细菌细胞等。将构建体导入宿主细胞可以通过下列方式实施：磷酸钙转染，DEAE-葡聚糖介导的转染，或电穿孔（Davis, L., Digner, M., Battey, L., Basic Methods in Molecular Biology(1986)）。作为合适宿主的代表性实例，可以列出的有：真菌细胞，诸如酵母；昆虫细胞诸如果蝇 S2 和斜纹夜蛾（*Spodoptera*）Sf9；动物细胞例如 CHO, COS 或 Bowes 黑素瘤；植物细胞等。通过本申请中的教导，可以认为合适宿主的选择在本领域技术人员掌握的范围之内。

[0063] 可用于本发明的宿主细胞是真核宿主细胞（例如，哺乳动物细胞）。在本发明的一方面，所述宿主细胞是适合于在细胞培养中生长的哺乳动物产物细胞。工业化生产中常用的所述细胞的实例有 CHO, VERO, BHK, HeLa, CV1（包括 Cos; Cos-7），MDCK, 293, 3T3, C127, 骨髓瘤细胞系（具体是鼠的），PC12 和 W138 细胞。中国仓鼠卵巢（CHO）细胞被广泛用于制备数种复合体重组蛋白，例如细胞因子，凝血因子和抗体（Brasel et al., Blood 88:2004-2012(1996); Kaufman et al. J. Biol. Chem. 263:6352-6362(1988); McKinnon et al., J. Mol. Endocrinol. 6:231-239(1991); Wood et al., J. Immunol. 145:3011-3016(1990)）。二氢叶酸还原酶（DHFR）-缺陷的突变体细胞系（Urlaub et al., Proc. Natl. Acad. Sci. USA 77:4216-4220(1980)）是所选的（of choice）CHO 宿主细胞系，因为有

效的 DHFR 可选择且可扩增的基因表达系统使得能够在这些细胞中实现高水平的重组蛋白质表达 (Kaufman, MethEnzymol 185:527-566 (1990))。此外,这些细胞易于以可贴壁或悬浮培养方式操作,并且表现出相对较好的遗传稳定性。CHO 细胞和表达于其中的重组蛋白质已经被广泛地研究,已经被管理机构批准用于临床生产。此外,还可以使用源自上述任一细胞系并具有所需表型的宿主细胞。例如,衍生的宿主细胞包括已经被选择性用于培养所需表型(例如,通过阳性和/或阴性筛选方法)的 CHO 细胞(例如, DG44 细胞系)。

[0064] 本发明的一方面,提供了用于体外生产目的多核苷酸所编码药剂的表达系统。如本申请所述,所述目的多核苷酸可以编码具有药用、医用、营养和/或工业价值的多肽。例如,所述的多核苷酸可以编码基于多肽 (polypeptide-based) 的药物。通常所述多肽可以作为细胞外产物表达。例如,可以用本发明所述表达盒和/或载体制备的多肽包括,但是不局限于, Flt3 配体, CD40 配体, 促红细胞生成素, 血小板生成素, 降钙素, Fas 配体, NF- κ B 受体活化剂的配体 (RANKL), TNF- 相关的凋亡-诱导型配体 (TRAIL), ORK/Tek, 胸腺基质 (thymic stroma)- 衍生的淋巴细胞生成素 (lymphopoietin), 粒细胞集落刺激因子, 粒细胞-巨噬细胞集落刺激因子, 肥大细胞生长因子, 干细胞生长因子, 表皮生长因子, RANTES, 生长激素, 胰岛素, 促胰岛素生成素 (insulinotropin), 胰岛素-样生长因子, 甲状旁腺激素, 干扰素(例如, 干扰素 β), 神经生长因子, 胰高血糖素, 白介素 1-18, 集落刺激因子, 淋巴毒素- β , 肿瘤坏死因子, 白血病抑制因子, 制瘤素 (oncostatin)-M, 细胞表面分子 Elk 和 Hek 的各种配体(例如 eph- 相关激酶的配体, 或 LERKS), 以及抗体的轻链或重链。

[0065] 任一上述蛋白质的受体也可以用发明的方法和组合物表达, 包括两种形式的肿瘤坏死因子受体(称为 p55 和 p75), 白介素-1 受体 (1 型和 2 型), 白介素-4 受体, 白介素-15 受体, 白介素-17 受体, 白介素-18 受体, 粒细胞-巨噬细胞集落刺激因子受体, 粒细胞集落刺激因子受体, 制瘤素-M 和白血病抑制因子的受体, NF- κ B 的受体活化剂 (RANK), TRAIL 受体, BAFF 受体, 淋巴毒素 β 受体, TGF β 受体 I 型和 2 型, 以及含有死亡结构域的受体, 诸如 Fas 或凋亡-诱导型受体 (Apoptosis-Inducing Receptor) (AIR)。

[0066] 可以用本发明表达盒和/或载体表达的其它蛋白质包括簇分化抗原(称为 CD 蛋白质), 例如, Leukocyte Typing VI (Proceedings of the VIth International Workshop and Conference; Kishimoto, Kikutani et al., eds.; Kobe, Japan, 1996) 中公开的那些, 或者后来的研讨会 (workshop) 公开的 CD 分子。所述分子的实例包括 CD27, CD30, CD39, CD40, 及其配体 (CD27 配体, CD30 配体和 CD40 配体)。这其中的一些是 TNF 受体家族的成员, 其中包括 41BB 和 OX40; 所述配体通常是 TNF 家族的成员 (41BB 配体和 OX40 配体也是); 因此, 所述 TNF 和 TNFR 家族的成员也可以使用本发明表达。

[0067] 另外, 具有酶促活性的多肽也可以按照本发明来表达。实例包括金属蛋白酶-去整合蛋白 (disintegrin) 家族成员, 各种激酶, 葡糖脑苷脂酶 (glucocerebrosidase), 过氧化物歧化酶 (superoxide dismutase), 组织纤溶酶原激活物, 因子 VIII, 因子 IX, 载脂蛋白 E, 载脂蛋白 A-I, 球蛋白, IL-2 拮抗剂, α -1 抗胰蛋白酶, TNF- α 转化酶 (TACE), 以及众多的其它酶。具酶促活性蛋白的配体也可以使用本发明表达盒和载体表达。

[0068] 本发明的组合物和方法还可用于表达其它类型的重组蛋白质和多肽, 包括免疫球蛋白分子或其部分以及嵌合抗体(例如, 具有与鼠抗原结合区域偶联的人恒定区的抗体)或其片段。已知多种技术可对编码免疫球蛋白分子的 DNA 进行操作以产生编码重组蛋白

质的 DNA,所述重组蛋白质诸如单链抗体,高亲合力抗体,或其它以基于抗体的多肽(参见,例如,Larrick et al.,Biotechnology 7:934-938(1989);Reichmann et al.,Nature 332:323-327(1988);Roberts et al.,Nature 328:731-734(1987);Verhoeyen et al.,Science239:1534-1536(1988);Chaudhary et al.,Nature 339:394-397(1989))。克隆的人源化抗体包括能特异性结合淋巴毒素 β 受体和整联蛋白诸如 VLA-1、VLA-4、和 $\alpha v \beta 6$ 的抗体,所述抗体可以是激动剂或拮抗剂。

[0069] 各种融合蛋白质也可以使用本发明的方法和组合物表达。所述融合蛋白质的实例包括:表达为与部分免疫球蛋白分子的融合物的蛋白质,表达为带有拉链部分(zipper moiety)的融合蛋白的蛋白质,以及新的多功能蛋白质诸如细胞因子与生长因子的融合蛋白质(例如,GM-CSF 和 IL-3, MGF 和 IL-3)。WO 93/08207 和 WO 96/40918 描述了一种被称为 CD40L 的分子的各种可溶性寡聚形式的制备方法,包括分别免疫球蛋白融合蛋白质和拉链(zipper)融合蛋白质;其中描述的技术适用于其它蛋白质。

[0070] 目的多核苷酸一旦被表达,那么所述表达产物(例如,蛋白质或多肽)可以使用本领域常规技术纯化。例如,当所述目的多核苷酸编码含有纯化标记的融合多肽时,可以使用可特异性结合所述标记的抗体来纯化。在一方面,将编码标记分子的寡核苷酸连接在编码所需多肽的目的多核苷酸的 5' 或 3' 末端;所述寡核苷酸可以编码聚 His(例如 Hexa His),或其它“标记”诸如 FLAG, HA(流感病毒血凝素(hemagglutinin Influenza virus)),或可得到其商品化抗体的 myc。在表达所述多肽时,通常将所述标记与多肽融合在一起,从用作将目的多肽自宿主细胞中的亲和纯化的方法。例如,可以通过使用抗标记抗体作为亲和基质的柱层析,来完成亲和纯化。任选地,随后可以用各种方法例如蛋白水解切割,将标记从纯化后的多肽除去。

[0071] 本发明所述的表达盒和载体可用于将目的多核苷酸稳定地转移入宿主细胞。稳定转移意指目的多核苷酸在宿主中连续地保持。本发明所述的表达盒或载体还可以用于将目的多核苷酸在宿主细胞中的瞬时表达。瞬时转染的宿主细胞会在细胞复制和生长期间丢失外源性 DNA。

[0072] 本发明的表达盒可用于将治疗剂递送至需要治疗的人或动物。可选地,本发明所述表达盒可用于将编码体内发挥疫苗作用的潜在免疫原性多肽的药剂递送至受试者(例如,人),具体是用于家养动物的免疫接种,包括食用(foodstock)动物,例如鱼,猪,马,牛,犬及猫科。

[0073] 本发明所述表达盒可以作用裸露的核酸构建体直接施用,该构建体通常含有与宿主细胞基因组同源的侧翼序列。可以用包括生物弹射击(biolistic)转化法和脂质转染法(lipofection)在内的几种已知技术,来提高脊椎动物对裸露核酸构建体的摄取。

[0074] 可选地,所述表达盒可作为载体的一部分导入,所述载体包括质粒载体或病毒载体。

[0075] 通常,将所述表达盒或载体与可药用的载体或稀释剂混合来制备药物组合物。合适的载体和稀释剂包括等渗盐溶液包括,例如,磷酸盐缓冲的盐水。可以将其中含有所述表达盒或载体的组合物配制用于各种施用形式,包括,例如经肌肉施用。

[0076] 当其中含有所述表达盒或载体的组合物以可注射形式使用时,通常将其与可药用的载体混合制成可直接注射入待治疗位点的可注射剂型。所述可药用的载体或稀释剂可以

是,例如,无菌的等渗(isotonic)溶液。其中含有所述表达盒或载体的所述组合物还可以配制成为口服活性形式。

[0077] 所用的实际剂型可由本领域技术人员轻易确定,并且随待施用物质的性质和施用途径变动。

[0078] 所用物质的剂量可以根据各种参数来调整,尤其是根据所用物质,受试者的年龄、体重、状态,所用施用方式和所需的临床治疗方案来调整。医师应能根据任一具体受试者和疾病确定所需的施用途径和剂量。

[0079] 实施例

[0080] pV10 载体的构建

[0081] pV10 构建的其它细节概括于图 1。从用人巨细胞病毒株 AD169 转染的人二倍体成纤维细胞中分离出基因组 DNA(ATCC No. VR-538),然后用其作为 PCR 扩增 CMV 立即早期基因 1 启动子/增强子区域(CMV IE1P/E)的模板(CMV IE1 基因的 5' UTR 的细节参见图 11(B)(SEQ ID NO:1))。扩增启动子,所用的引物在 5' 末端含有 HindIII 位点(tttAAGCTTGACATTGATTATTGACTAG;SEQ ID NO:2;下划线部分为限制酶切位点)并在 3' 末端含有 BamHI 位点(ttttGGATCCCTGTCAAGGACGGTGACTGC;SEQ ID NO:3;下划线部分为限制酶切位点)。位于限制酶切位点之前的末端"t"核苷酸包含在寡核苷酸设计中,以促进限制性内切酶消化,并在随后的克隆步骤中被除去。

[0082] 所有 PCR 反应全部在 DNA engine PTC-200Pelier Thermal Cycler(MJResearch, Watertown,MA)中完成。反应总体积为 100 μ l:1X NEB Vent 聚合酶缓冲液(10mM KCl,20mM Tris pH 8.8,25 $^{\circ}$ C,10mM(NH₄)SO₄,2mMMgSO₄,0.1% Triton X-100),2.5mM dNTP's,2 单位 Vent DNA 聚合酶(New England Biolabs,Beverly,MA),每种引物各 1 μ g,1 μ g 分离自 CMV 感染的细胞的基因组 DNA 作为模板。反应的条件如下:99 $^{\circ}$ C、1 分钟,55 $^{\circ}$ C、30 秒,75 $^{\circ}$ C、1.5 分钟,共 15 个循环。得到的片段用限制性内切酶 BamHI 和 HindIII(New England Biolabs)消化后,亚克隆入经 BamHI 和 HindIII 消化的克隆载体 pUC19。对插入片段的序列进行分析,发现与克隆的 CMV IE1 启动子/增强子区域的公开序列一致。

[0083] pV40 载体的构建

[0084] pV40 的构建概括于图 2。从分离自人成纤维细胞的基因组 DNA 通过 PCR 扩增出含有 polyA 信号的 hGHv 基因的 3' UTR。5' 引物的(+)链(TTTTGGATCCCTGCCCCGGGTGGCATCC;SEQ ID NO:20)含有末端的 BamHI 限制酶切位点,3' 引物的(-)链含有末端的 EcoRI 位点(TTTTGAATTCATGAGAGGACAGTGCCAAGC;SEQ ID NO:21)。PCR 反应条件与 pV10 构建中所述的相同。得到的 PCR 片段用 BamHI 和 EcoRI 消化后,用凝胶纯化,然后连接入经 BamHI 和 EcoRI 消化的载体 pV10。得到的质粒命名为 pV40,用限制性内切酶分析加以确证。随后测序表明:该 hGHv 基因的 3UTR(SEQ ID NO:19)和公开的 Genbank 登录号 No. K00470 中的 hGHv 基因序列(SEQ ID NO:18)有少数几个核苷酸不同。这些差异中的至少一些是由于等位基因突变引起的。

[0085] pV70 载体的构建

[0086] 用 BspEI 和 HpaI 消化所述 pV40 载体以便从内含子 A 区域中去除 317 个核苷酸的片段(参见,例如,图 2 和 3)。pV60 通过平端连接入(blunt endligation into)pV40 的 dhfr 编码区域的 BspEI-HpaI 制得。所述 pV70 表达载体含有调控转录的人巨细胞病毒主要

立即早期 1 (hCMV IE1) 启动子 / 增强子区域。其中还含有 hCMV IE15' UTR 和内含子 A, 而所述内含子包括 317 个碱基对缺失。为了终止转录, 所述载体含有人生长激素变体聚腺苷酸化 (hGHv poly A) 区域, SEQ ID NO:19。pV40、pV60 和 pV70 的构建下文有详细描述, 进一步描述见附图。

[0087] A. pXLC. 1 的产生

[0088] 将含有抗体轻链编码序列的 PCR 产物用 BamHI 消化后, 克隆入表达载体 pV70 的单一 BamHI 位点 (图 3)。将所述轻链编码区域插入单一的 BamHI 位点, 该位点位于 hCMV IE1 内含子 3' 末端的 5'UTR 和 hGHv poly A 区域的 5' 末端之间。所述质粒命名为 pXLC. 1。图 4 图示的是产生 pXLC. 1 载体的示意图。

[0089] B. neo 表达盒 -pXLC. 2 的添加

[0090] 将新霉素转移酶 (neo) 表达盒导入 pXLC. 1 用作选择标记 (图 5)。所述 neo 表达盒被制备为名为 pGT-N28 的商品化质粒 (New England Biolabs, catalog#307-28) 的 BamHI/EcoRI 片段。该质粒中, 所述 neo 基因是由磷酸甘油酸激酶 (PGK) 启动子驱动, 终止于 PGK 聚-腺苷酸化位点。使用下列衔接子寡聚物 (adaptor aligo) 将 neo 表达盒 5' BamHI 尾转化为 NarI 尾端:

[0091] EcoRI

[0092] BamHI 相容 (compatible) GATCGATGAATTCGG (SEQ ID NO:4)

[0093] CTACTTAAGCCGC (SEQ ID NO:5) NarI 相容

[0094] 利用这些接头, 添加了新的 EcoRI 位点。首先将所述衔接子连接至 BamHI/EcoRI 切出的 neo 片段, 然后将经转化的 NarI/EcoRI 片段克隆入经 NarI 和 EcoRI 消化后的 pXLC. 1。这样, 就将所述 neo 表达盒插入了所述质粒的轻链序列 3' 末端。该质粒命名为 pXLC. 2 (图 5)。

[0095] 重链表达载体 -pXHC. 5 的构建

[0096] A. 重链的 RT-PCR 扩增

[0097] 通过 RT-PCR 反应扩增得到重链, 使用 5' PCR 引物 TTTTGGATCCATGTACTGGGTGAAGCAG (SEQ ID NO:6), 其中斜体序列是加入的带有 BamHI 位点的接头区域, 加下划线的碱基对应于重链编码序列中的第二个甲硫氨酸。使用 3' PCR 引物, GCCCGATCCTCATTTACCCGAGACAG (SEQ ID NO:7), 其还含有添加的具有 BamHI 位点的接头区域 (斜体字母) 以及对应于重链编码序列并含终止密码子的序列 (下划线部分)。获得了 1268 个碱基对的预期 PCR 产物。

[0098] B. pXHC 的构建

[0099] 因为使用的 5' PCR 引物是与编码区域中的第二个 ATG 密码子而不是与 PCR 产物所含编码区域中的起始 ATG 杂交, 所以该重链编码区域是不完整的。将含有所述不完整重链的 BamHI 片段克隆入质粒 pV60 (图 6)。除了在内含子的缺失位置处含有 dhfr 编码区外, 该载体与 pV70 (上述) 完全相同。将所述重链编码区域插入单一的 BamHI 位点, 该位点位于 hCMV IE1 内含子 3' 末端的 5' 非翻译序列与 hGHv 变体 polyA 区域的 5' 末端之间。将这种含有所述不完整重链的质粒命名为 pXHC。

[0100] C. 重链编码序列的填充 (completion)-pXHC. 1

[0101] 将上述重链序列中缺失的编码区插入 pXHC 制备得到质粒 pXHC. 1 (图 7)。为了实

现这一点,使用含有抗体编码序列的质粒作为模板,通过 PCR 制备得到一种片段。使用的 5' PCR 引物为:

[0102] PstI BamHI

[0103] TTTTCTGCAGTCACCGTCCTTGACACGGGATCCATGGACTGGACCTTGGAGGG (SEQ ID NO :8)。斜体序列对应于 BamHI 位点 5' 侧的 pXHC 序列,粗体序列对应于从重链序列中起始密码子前的两个碱基处开始的序列。所用的 3' 引物为 CTGAGGAGACGGTGACCAGGTCCCTTGGCCCC (SEQ ID NO :9)。该引物与第一外显子(重链可变区)的末端杂交。正如所料得到了 445 个碱基对的 PCR 产物,用 PstI 和 StuI 切割该产物。将该 PstI/StuI 片段克隆入经 PstI 和 StuI 切割的 pXHC 以制备 pXHC. 1。

[0104] D. 内含子 dhfr 的去除 -pXHC. 3

[0105] pXHC. 1 含有 hCMV IE1 内含子中的 dhfr 基因。由于这种结构在扩增时可能引发问题,从该内含子除去 dhfr 基因,并将表达盒插入到重链表达盒的 3' (图 8)。第一步是去除内含子中的 dhfr 基因。这一步通过下述方法完成:从 pXHC. 1 中切割出重链编码序列作为 PstI/EcoRI 片段,然后将其克隆入经同样酶切割的 pXLC. 1 质粒。通过这一克隆步骤容易地将所述质粒中的轻链置换为重链 (switched the light chain for the heavy chain)。得到的质粒与 pXHC. 1 相同,除了含 dhfr 基因的内含子被具有上述 pXLC. 1 缺失的内含子取代外。该重链质粒命名为 pXHC. 3 (图 8)。

[0106] E. dhfr 表达盒的插入 -pXHC. 5

[0107] dhfr 构型 (configuration) 改变的第二步是将 dhfr 表达盒插入重链表达盒的 3' (图 9)。所述 dhfr 表达盒源自于质粒 pSI-DHFR 的 BglIII/BamHI 片段上,并含有 SV40 早期启动子、dhfr 基因和 SV40poly A 区域。使用下列衔接子寡聚物,将该片段克隆入位于 pXHC. 3 中 hGHv polyA 区域 3' 末端的 EcoRI 位点:

[0108] SalI

[0109] EcoRI 相容 AATTCGTCGACA (SEQ ID NO :10)

[0110] GCAGCTGTCTAG (SEQ ID NO :11) BamHI/BglIII 相容

[0111] 首先将该衔接子连接到经 EcoRI 切割的质粒,然后将所述 BglIII/BamHI dhfr 表达盒连接到连接了衔接子的 (adapted) 质粒。该质粒命名为 pXHC. 5 (图 9)。

[0112] pXLC. 2 和 pXHC. 5 的定性

[0113] 使用限制性内切酶对这两种质粒进行分析,验证插入片段的存在和方向。另外,对质粒的编码区进行测序,证实 PCR 或克隆步骤过程中没有任何突变的累积。

[0114] 通过下述方法验证功能性 neo 选择标记的存在:将 pXLC. 2 转染入 CHO 细胞,证明抗 G418 的抗性。当将 pXLC. 2 和 pXHC. 5 质粒共-转染入适应无血清条件的 (serum free adapted) DG44CHO 宿主细胞时,证明能够进行双重选择的能力。在双重选择培养基 (含 400mg/ml G418 的 α -MEM 10% dFBS) 中,菌落从共-转染物 (co-transfection) 长出。同样条件下,任一单独的选择培养基 (either selection alone) (α -MEM 或 400mg/ml G418) 能杀死未转染的细胞。

[0115] 载体的构建:pV80 和 pV90 载体

[0116] 从重链表达载体 pXHC. 5 制备 pV80 (图 10)。使用 BamHI 去除 pXHC. 5 中的重链编码序列。连接该主链使其在 BamHI 位点处再环化。

[0117] 在 pV80 载体上做两处改变制备 pV90 (图 11)。pV80 构建体中第 3166 位的 NotI 位点 (位于 dhfr 编码区的末端, 参见所附序列) 被破坏。为了实现这一点, 用 NotI 消化该质粒, 用 Klenow 聚合酶将突出端“补平 (fill-in)”。因此, 再连接后的质粒就失去了第 3166 位的 NotI 位点。然后, 用 BamHI 消化所述载体, 导入 NotI 接头, 在克隆位点创建新的 NotI 位点, 该 NotI 接头是下列 14-mer 自身退火制得的: GATCCGCGGCCGCG (SEQ ID NO:12)。当一起退火时, 所述接头序列为:

[0118] NotI

[0119] BamHI 相容的 GATCCGCGGCCGC (SEQ ID NO:13)

[0120] CGCCGCGCCTAGG (SEQ ID NO:14) BamHI 相容的

[0121] 该克隆步骤在 NotI 位点的两侧各自再建 BamHI 位点。这些 BamHI 位点可用于用该载体产生的稳定细胞系的基因分析。通过序列分析验证 pV90 的克隆位点序列。

[0122] 在 pV80 载体中, 可以将多核苷酸 (例如, 编码序列) 克隆入 BamHI 位点 GGATCCCTGCCCGGGT (SEQ ID NO:15)。所述粗体序列代表 BamHI 位点。在 pV90 载体中, 可以将多核苷酸 (例如, 编码序列) 克隆入 BamHI 或 NotI 位点 GGATCCGGCGGCCGCGGATCCCTGCCCGGGT (SEQ ID NO:16)。其中, 粗体序列表示 BamHI 位点, 下划线序列表示 NotI 位点。为了取得使用 NotI 位点时的最佳结果, 应该在 PCR 引物中起始密码子之前添加“C”以实现与 Kozak 序列的最佳匹配 (例如, GGATCC GCGGCCG C ATG (SEQ ID NO:17))。

[0123] pV80 和 pV90 中的限制酶切位点

[0124] 除了 NotI 限制酶切位点外 pV80 和 pV90 完全相同: pV80 在第 3166 位具有单一 NotI 位点 (dhfr 编码区的末端), pV90 在第 1260 位具有单一 NotI 位点 (克隆位点)。

[0125] 其中的 2 个或少于 2 个被发现共同限制酶切位点 (common restriction sites of which 2 or fewer were found) 包括表 1 中列出的位点。

[0126] 表 1:

[0127]	AatI (1)	2274	ApaI (1)	1733	AspEI (2)
	1382, 4807	BamHI (1)	1254		
[0128]	BsgI (1)	1494	BsiXI (1)	3404	ClaI (1) 3404
	EgeI (1)	3585			
[0129]	HindIII (2)	2293, 6059	HpaI (1)	3308	KasI (1) 3585
	Kpn2I (1)	1121			
[0130]	KpnI (2)	1927, 3144	NarI (1)	3585	NdeI (2) 254,
	3637	NheI (1)	2557		
[0131]	NotI (1)	see above	PstI (2)	1231, 2331	PvuI (2) 3544,
	444	OSacI (2)	585, 2819		
[0132]	SacII (2)	673, 1258	SmaI (2)	1271, 3161	SpeI (1)
	20StuI (1)	2274			
[0133]	XbaI (1)	3150	XhoI (1)	3127	XmaI (2) 1271, 3161

[0134] 未发现下列常见酶切位点:

[0135] 表 2

[0136] AatII AfaI AfeI AgeI AluI ApaLI AspHI AspI AvaI

AvaII BsiWI DpnI

[0137] DpnII DraI DraII DraIII EaeI EagI EarI EcoRI EcoRV
FspI HaeII HaeIII

[0138] HincI HpaI NaeI NcoI NdeII NspI PacI SalI SphI
XhoII XmaIII

[0139] 报道物基因的克隆

[0140] 将其中含有 CMV IE1、内含子 A 片段和 hGHv polyA 构建体的表达盒 / 载体与以基于商品化 SV40 的高表达载体进行比较。

[0141] 使用分泌的碱性磷酸酶 (SEAP) 作为报道物来比较所述质粒。基于 SV40 的表达载体 pSEAP2- 对照 (Cat#6052-1, Clontech) 能够在 SV40 早期启动子和 SV40 增强子的控制下表达 SEAP。所述 SEAP 编码序列后接 SV40 晚期聚腺苷酸化信号以确保真核细胞中的 SEAP 转录物能得到正确有效的加工。合成的转录阻遏物 (transcription blocker) (TB) 由邻接的聚腺苷酸化和转录暂停位点组成, 其位于 MCS 上游, 能够减少背景转录。所述载体整合大量以下特性: 通过提高 SEAP 表达的效率提高 SEAP 敏感性, 或改善所述载体的使用 (utility)。这些包括: 增加 Kozak 共有序列翻译起始位点; 去除 SV40 小-t 内含子, 该内含子能够导致隐蔽剪接以及一些基因和 / 或细胞类型中表达的减少; 将 SV40 的早期聚腺苷酸化信号转换为晚期聚腺苷酸化信号, 这通常会引起 mRNA 水平增加 5 倍; 扩大的多克隆位点 (MCS); 压缩 (compact) 质粒大小; 除去 SEAP mRNA 3' 非翻译区的外源序列。Genbank 登录号 U89938。

[0142] 为了制备 pCMV-hGHvPA-SEAP 质粒, 用 PCR 方法从 pSEAP2- 对照质粒中提取 SEAP 编码序列, 并将其克隆入 pV110 载体。pV110 质粒是 pV90 的衍生物 (无 dhfr 表达盒, 不同的多聚接头, 除此以外都完全相同)。通过测序对所述 pCMV-hGHvPA-SEAP 质粒构建体进行验证。

[0143] 用于转染的宿主是二氢叶酸还原酶 (dhfr) 缺陷的中华仓鼠卵巢细胞系 DG44 (Urlaub et al., Cell 33, 405-412 (1983))。用编码二氢叶酸还原酶 (dhfr) 的质粒共 - 转染 CMVSEAP 和 pSEAP2- 对照报道质粒, 以便能够筛选出稳定的转染体 (pSI-DHFR. 2, 图 12)。每一转染含有 50 μ g 报道质粒和 5 μ g pSI-dhfr. 2。所有 DNA 都使用 Megaprep 试剂盒 (Qiagen) 制备。在转染之前, 用 EtOH 沉淀 DNA, 然后用 70% EtOH 洗涤, 干燥, 重悬于 HEBS (20mM Hepes, pH 7.05, 137mM NaCl, 5mM KCl, 0.7mM Na₂HPO₄, 6mM 葡萄糖), 然后在转染前进行定量。用包括 pUC18 在内的阴性对照 (ATCC No. 37253) 作为报道物对照, 用无 DNA 转染 (no DNA transfection) 的作为转染对照 (表 3)。

[0144] 0.28kV 和 950mF 条件下使用 0.4cm 样品池 (BioRad), 通过在 0.8ml HEBS 中进行电穿孔, 转染细胞和 DNA。每次转染都使用 5E6 细胞。电穿孔脉冲后, 让细胞在样品池中于室温保温 5-10 分钟。然后, 将其转移到离心管 (含有 10ml 含核苷及 10% dFBS 的 α MEM) 中, 然后 1K RPM 沉淀 5 分钟。将重悬的沉淀物接种于含有 α MEM (含 10% dFBS 但无核苷) 的 6-孔平板中, 在湿润温箱中, 于 5% CO₂、36°C 保温直至菌落形成。

[0145] 表 3 转染试验

[0146]

报道质粒 (各 50 μ g)	DHFR 质粒 (各 50 μ g)	转染编号
pSEAP2- 对照	pSIDHFR. 2	3
pCMVSEAP	p SIDHFR. 2	3
pUC18	pSIDHFR. 2	1
无 DNA	无 DNA	1

[0147] 转染约 2 周后,在只含有 pSIDHFR. 2 质粒的转染物中有菌落形成。稳定转染体以合并物 (pool) 或者分离株方式进行分析。试验测定比繁殖力,其中把培养基换为新鲜的培养基,24 小时后从培养基取样,进行细胞计数。将产品滴度标准化为 24 小时的试验结束时的细胞数目,然后将产率表示为每个细胞的 SEAP 活性。

[0148] SEAP 试验. 使用 the Great EscAPe™ SEAP Reporter System3 (clontech) 对经调节的培养基 (conditioned medium) 进行分析。该试验使用荧光底物检测经调节的培养基中的 SEAP 活性。按照 96 孔格式 (format) 使用试剂盒,除了下列内容不同外,按照厂商说明书进行。把试剂盒中的试验缓冲液换为 1.5M 二乙醇胺,0.75mM $MgCl_2$, 15mM L- 同型精氨酸 (homoarginine), 和 10% Emerald II (Cat#9761, Applied Biosystems Biosystems)。所有标准物和样本都用新鲜的培养基而不用提供的稀释缓冲液来稀释。将只在 60 分钟后读数 1 次改为以 10-20 分钟的间隔时间多次读数,然后使用这些读数结果将 SEAP 活性表示为每分钟相对荧光单位 (RFU/min)。平板读数器 (Cytofluor II, PerSeptive Biosystems) 使用的发射滤波器 (emission filter) 为 460nm 而非推荐的 449nm。

[0149] 根据试剂盒提供的标准物,将 RFU/min 值相对于标准曲线标准化。由于提供的标准物未定量,所以所有值全都是相对值。将这些相对值相对于细胞数目和孵育期进行标准化,以产生相对比繁殖力 (每个细胞每日的 SEAP 活性)。

[0150] 合并物. 菌落出现后,收集并且合并来自每次转染的细胞。将合并物以每孔 $\sim 2 \times 10^5$ 个细胞种植入 6- 孔平板。第二天,将培养基换为 2ml 新鲜培养基。24 小时后,进行细胞计数,使用培养基的样本测定 SEAP 活性。合并物试验的结果图示于图 13。

[0151] 分离株. 从转染物“挑取 (picking)”菌落获得分离株。通过用设定在 50ml 的 P200Pipetman™ 从菌落直接吸取完成“挑取”。首先将吸出的菌落转移到 48 孔平板,然后在有足够数目的细胞时再转移到 6 孔平板。在将近铺满至铺满细胞密度时,使用上述 24- 小时试验测定比产率 (图 14)。

[0152] 如图 13 和 14 概括的那样,基于 CMV IE 启动子和 hGHv polyA 信号结构域的组合的表达载体比自称具有高表达能力的商品化载体要好得多。

[0153] 虽然已经描述了本发明的许多实施方案。但是,显然还可以做各种改变而不脱离本发明的实质和范围。因此,其它实施方案也在下列权利要求的范围之内。

序列表

<110> 比奥根艾迪克公司 (Biogen Idec Inc.)

<120> 用于瞬时表达或稳定表达外源分子的表达盒和载体

<130>13751-013W01

<150>US 60/448, 179

<151>2003-02-14

<160>21

<170>FastSEQ for Windows Version 4.0

<210>1

<211>6069

<212>DNA

<213> 人 (Homo sapiens)

<400>1

agcttgacat tgattattga ctagttatta atagtaatca attacggggt cattagttca	60
tagcccatat atggagttcc gcgttacata acttacggta aatggcccgct ctggctgacc	120
gccaacgac ccccgcccat tgacgtcaat aatgacgtat gttcccatag taacgccaat	180
agggactttc cattgacgtc aatgggtgga gtattttacgg taaactgccc acttggcagt	240
acatcaagtg tatcatatgc caagtacgcc ccctattgac gtcaatgacg gtaaatggcc	300
cgcctggcat tatgcccagt acatgacctt atgggacttt cctacttggc agtacatcta	360
cgtattagtc atcgctatta ccatgggtgat gcggttttgg cagtacatca atgggcgtgg	420
atageggttt gactcacggg gatttccaag tctccacccc attgacgtca atgggagttt	480
gtttttggcac caaaatcaac gggactttcc aaaatgtcgt aacaactccg ccccatgac	540
gcaaatgggc ggtaggcgtg tacggtggga ggtctatata agcagagctc gtttagtgaa	600
ccgtcagatc gcctggagac gccatccacg ctgttttgac ctccatagaa gacaccggga	660
ccgatccagc ctccgcggcc gggaacggtg cattggaacg cggtattccc gtgccaagag	720
tgacgtaagt accgcctata gagtctatag gccaccccc ttggcttctt atgcatgcta	780
tactgttttt ggcttggggc ctatacacc ccgttctc atgttatagg tgatggtata	840
gcttagccta taggtgtggg ttattgacca ttattgacca ctcccctatt ggtgacgata	900
ctttccatta ctaatccata acatggctct ttgccacaac tctctttatt ggctatatgc	960
caatacactg tccttcagag actgacacgg actctgtatt ttacaggat ggggtctcat	1020
ttattattta caaattcaca tatacaacac caccgtcccc agtgcccgca gtttttatta	1080

aacataacgt	gggatctcca	cgcgaaatctc	gggtacgtgt	tccggaacgg	tggagggcag	1140
tgtagtctga	gcagtactcg	ttgctgccgc	gcgcgccacc	agacataata	gctgacagac	1200
taacagactg	ttcctttcca	tgggtctttt	ctgcagtcac	cgctccttgac	acgggatccg	1260
cggccgcgga	tccctgcccg	ggtggcatcc	ctgtgacccc	tccccagtg	ctctcctggt	1320
cgtggaaggt	gctactccag	tgcccaccag	ccttgtccta	ataaaaattaa	gttgcacat	1380
tttgtttgac	taggtgtcct	tgtataatat	tatgggggtgg	aggcgggtgg	tatggagcaa	1440
ggggcaggtt	gggaagacaa	cctgtagggc	cttcagggtc	tattgggaac	caggctggag	1500
tgcagtggca	cgatcttggc	tcgctgcaat	ctccgcctcc	tgggttcaag	cgattctcct	1560
gcctcagtct	cccgaatagt	tgggattcca	ggcatgcacg	accaggetca	gctaattttt	1620
gtattttttg	tagagacggg	gtttcaccat	attggccagt	ctggctctcca	tctcctgacc	1680
tcaggtaatc	cgcccgcctc	ggcctcccaa	attgctggga	ttacagggtat	gagccactgg	1740
gcccttcctt	gtcctgtgat	tttaaaataa	ttataaccagc	agaaggacgt	ccagacacag	1800
catgggctac	ctggccatgc	ccagccagtt	ggacatttga	gttgtttgct	tggcactgtc	1860
ctctcatgaa	ttcgtcgaca	gatctgcgca	gcaccatggc	ctgaaataac	ctctgaaaga	1920
ggaacttgg	taggtacctt	ctgaggcgga	aagaaccagc	tgtggaatgt	gtgtcagtta	1980
gggtgtggaa	agtcgccagg	ctcccccagc	ggcagaagta	tgcaaagcat	gcattctcaat	2040
tagtcagcaa	ccaggtgttg	aaagtcccca	ggctccccag	caggcagaag	tatgcaaagc	2100
atgcatctca	attagtcagc	aaccatagtc	ccgcccctaa	ctccgcccct	cccgccccta	2160
actccgcca	gttccgcca	ttctccgccc	catggctgac	taattttttt	tatttatgca	2220
gaggccgagg	ccgcctcggc	ctctgagcta	ttccagaagt	agtgaggagg	cttttttgg	2280
ggcctaggct	tttgcaaaaa	gcttgattct	tctgacacaa	cagtctcgaa	cttaagctgc	2340
agaagttagt	cgtgaggcac	tgggcaggta	agtatcaagg	ttacaagaca	ggtttaagga	2400
gaccaataga	aactgggctt	gtcgagacag	agaagactct	tgcgtttctg	ataggcacct	2460
attggtctta	ctgacatcca	ctttgccttt	ctctccacag	gtgtccactc	ccagttcaat	2520
tacagctctt	aaggctagag	tacttaatac	gactcactat	aggctagcat	ggttcgacca	2580
ttgaactgca	tcgtcgccgt	gtcccaaaat	atggggattg	gcaagaacgg	agacctaccc	2640
tggcctccgc	tcaggaacga	gttcaagtac	ttccaaagaa	tgaccacaac	ctcttcagtg	2700
gaaggtaaac	agaatctgg	gattatgggt	aggaaaacct	ggttctccat	tcctgagaag	2760
aatcgacctt	taaaggacag	aattaatata	gttctcagta	gagaactcaa	agaaccacca	2820
cgaggagctc	atcttcttgc	caaaagtttg	gatgatgcct	taagacttat	tgaacaaccg	2880
gaattggcaa	gtaaagtaga	catggttttg	atagtcggag	gcagttctgt	ttaccaggaa	2940
gccatgaatc	aaccaggcca	cctcagactc	tttgtgacaa	ggatcatgca	ggaatttgaa	3000
agtgcacagt	ttttcccgaa	aattgatttg	gggaaatata	aacttctccc	agaataccca	3060
ggcgtcctct	ctgaggcca	ggaggaaaaa	ggcatcaagt	ataagtttga	agtctacgag	3120
aagaaagact	aactcgagaa	ttcacgcgtg	gtacctctag	agtcgacccg	ggcggccggc	3180
cgcttcgagc	agacatgata	agatacattg	atgagtttgg	acaaaccaca	actagaatgc	3240
agtgaaaaaa	atgcttttatt	tgtgaaatit	gtgatgctat	tgcttttatt	gtaaccatta	3300
taagctgcaa	tacaacaagt	aacaacaaca	attgcattca	ttttatgttt	caggttcagg	3360
gggaggtgtg	ggaggttttt	taaagcaagt	aaaacctcta	caaagtgtgt	aaaatcgata	3420

aggatctgtc	gacgaattca	ctggccgtcg	ttttacaacg	tcgtgactgg	gaaaaccctg	3480
gcgttaccca	acttaategc	cttgacgac	atcccccttt	cgccagctgg	cgtaatagcg	3540
aagaggcccc	caccgategc	ccttcccaac	agttgcgag	cctgaatggc	gaatggcgcc	3600
tgatgcggta	ttttctcctt	acgcattctgt	gcggtatttc	acaccgcata	tgggtgactc	3660
tcagtacaat	ctgctctgat	gccgcatagt	taagccagcc	ccgacacccg	ccaacacccg	3720
ctgacgcgcc	ctgacgggct	tgtctgtctc	cggcatccgc	ttacagacaa	gctgtgaccg	3780
tctccgggag	ctgcatgtgt	cagaggtttt	caccgtcatt	accgaaacgc	gcgagacgaa	3840
agggcctcgt	gatacgcta	tttttatagg	ttaatgtcat	gataataatg	gtttcttaga	3900
cgtcagggtg	cacttttcgg	ggaaatgtgc	gcggaacccc	tatttgttta	tttttctaaa	3960
tacattcaaa	tatgtatccg	ctcatgagac	aataaccctg	ataaatgctt	caataatatt	4020
gaaaaaggaa	gagtatgagt	attcaacatt	tccgtgtcgc	ccttattccc	ttttttgcgg	4080
cattttgcct	tcctgttttt	gtcaccacg	aaacgctgg	gaaagtaaaa	gatgctgaag	4140
atcagttggg	tgcacgagtg	ggttacatcg	aactggatct	caacagcgg	aagatccttg	4200
agagttttcg	ccccgaagaa	cgttttccaa	tgatgagcac	ttttaagtt	ctgctatgtg	4260
gcgcggtatt	atcccgatt	gacgccgggc	aagagcaact	cggtcgccgc	atacactatt	4320
ctcagaatga	cttggttgag	tactcaccag	tcacagaaaa	gcattcttac	gatggcatga	4380
cagtaagaga	attatgcagt	gttgccataa	ccatgagtga	taacactg	gccaacttac	4440
ttctgacaac	gateggagga	ccgaaggagc	taaccgcttt	tttgacacac	atgggggatc	4500
atgtaactcg	ccttgatcgt	tgggaaccgg	agctgaatga	agccatacca	aacgacgagc	4560
gtgacaccac	gatgcctgta	gcaatggcaa	caacgttgcg	caaactatta	actggcgaa	4620
tacttactct	agcttcccgg	caacaattaa	tagactggat	ggaggcggat	aaagttgcag	4680
gaccacttct	gcgctcgccc	cttccggctg	gctggtttat	tgtgataaaa	tctggagccg	4740
gtgagcgtgg	gtctcgcggt	atcattgcag	cactggggcc	agatggtaag	ccctcccgt	4800
tcgtagttaa	ctacacgacg	gggagtcagg	caactatgga	tgaacgaaat	agacagatcg	4860
ctgagatagg	tgcctcactg	attaagcatt	ggtaactgtc	agaccaagtt	tactcatata	4920
tacttttagat	tgatttaaaa	cttcattttt	aatttaaaa	gatctaggtg	aagatccttt	4980
ttgataatct	catgacaaaa	atcccctaac	gtgagttttc	gttccactga	gcgtcagacc	5040
ccgtagaaaa	gatcaaagga	tcttcttgag	atcctttttt	tctgcgcgta	atctgtgtct	5100
tgcaaacaaa	aaaaccaccg	ctaccagcgg	tggtttgttt	gccggatcaa	gagctaccaa	5160
ctctttttcc	gaaggtaact	ggcttcagca	gagcgcagat	accaaatact	gtccttctag	5220
tgtagccgta	gttaggccac	cacttcaaga	actctgtagc	accgcctaca	tacctcgtc	5280
tgttaatect	gttaccagtg	gtgtgtgcca	gtggcgataa	gtcgtgtctt	accgggttgg	5340
actcaagacg	atagttaccg	gataaggcgc	agcggtcggg	ctgaacgggg	ggttcgtgca	5400
cacagcccag	cttgagcga	acgacctaca	ccgaactgag	atacctacag	cgtagctat	5460
gagaaagcgc	cacgcttccc	gaaggagaaa	aggcggacag	gtatccggta	agcggcaggg	5520
tcggaacagg	agagcgcacg	aggagcttc	cagggggaaa	cgcttggtat	ctttatagtc	5580
ctgtcgggtt	tcgccacctc	tgacttgagc	gtcgattttt	gtgatgtctg	tcaggggggc	5640
ggagcctatg	gaaaaacgcc	agcaacgcgg	cctttttacg	gttcctggcc	ttttgtgtgc	5700
cttttgccta	catgttcttt	cctgcgttat	ccccgtattc	tgtggataac	cgtattaccg	5760

cctttgagtg agctgatacc gctcgccgca gccgaacgac cgagcgcagc gagtcagtga	5820
gcgaggaagc ggaagagegc ccaatacgca aaccgcctct ccccgcgctg tggccgattc	5880
attaatgcag ctggcacgac aggtttcccg actggaaagc gggcagtgag cgcaacgcaa	5940
ttaatgtgag ttagctcact cattagggcac ccaggtttt acactttatg cttccggctc	6000
gtatgttgtg tggaattgtg agcggataac aatttcacac aggaaacagc tatgaccatg	6060
attacgcca	6069

<210>2

<211>28

<212>DNA

<213> 人工序列

<220>

<223> 引物

<400>2

tttaagcttg acattgatta ttgactag	28
--------------------------------	----

<210>3

<211>30

<212>DNA

<213> 人工序列

<220>

<223> 引物

<400>3

ttttggatcc ctgtcaagga cggtgactgc	30
----------------------------------	----

<210>4

<211>15

<212>DNA

<213> 人工序列

<220>

<223> 衔接子寡核苷酸

<400>4

gatcgatgaa ttcgg	15
------------------	----

<210>5

<211>13

<212>DNA

<213> 人工序列

<220>

<223> 衔接子寡核苷酸

<400>5

ctacttaagc cgc

13

<210>6

<211>28

<212>DNA

<213> 人工序列

<220>

<223> 引物

<400>6

ttttggatcc atgtactggg tgaagcag

28

<210>7

<211>28

<212>DNA

<213> 人工序列

<220>

<223> 引物

<400>7

gcccggatcc tcatttaccc ggagacag

28

<210>8

<211>53

<212>DNA

<213> 人工序列

<220>

<223> 引物

<400>8

ttttctgcag tcaccgtcct tgacacggga tccatggact ggaccttgga ggg 53

<210>9

<211>33

<212>DNA

<213> 人工序列

<220>

<223> 引物

<400>9

ctgaggagac ggtgaccagg gtccttggc ccc 33

<210>10

<211>12

<212>DNA

<213> 人工序列

<220>

<223> 衔接子寡核苷酸

<400>10

aattcgtcga ca 12

<210>11

<211>12

<212>DNA

<213> 人工序列

<220>

<223> 衔接子寡核苷酸

<400>11

gcagctgtct ag 12

<210>12

<211>14

<212>DNA

<213> 人工序列

<220>

<223> 合成制备的寡核苷酸

<400>12

gatccgcggc cgcg

14

<210>13

<211>13

<212>DNA

<213> 人工序列

<220>

<223> 合成制备的寡核苷酸

<400>13

gatccgcggc cgc

13

<210>14

<211>14

<212>DNA

<213> 人工序列

<220>

<223> 合成制备的寡核苷酸

<400>14

cgccggcgcc tagg

14

<210>15

<211>16

<212>DNA

<213> 人工序列

<220>

<223> 合成制备的寡核苷酸

<400>15

ggatccctgc ccgggt

16

<210>16

<211>30

<212>DNA

<213> 人工序列

<220>

<223> 合成制备的寡核苷酸

<400>16

ggatccgcgg ccgcggatcc ctgcccgggt

30

<210>17

<211>18

<212>DNA

<213> 人工序列

<220>

<223> 合成制备的寡核苷酸

<400>17

ggatccgcgg ccgcatg

18

<210>18

<211>2660

<212>DNA

<213> 人 (Homo sapiens)

<400>18

gaattcagca ctgaatcatg ccagaaaccc ccgcaatcta ttggctgtgc tttggcccct	60
tttcccaaca cacacattct gtctggtggg tggaggggaa acatgcgggg aggaggaaag	120
gaataggata gagagtggga tggggtcggt aggggtctca aggactggcc taccctgaca	180
tccttctccg cgttcagggt ggccaccatg gcctgtgtgc agagggcacc cacgtgaccc	240
ttaaagagag gacaagttag gtggtatctc tggctgacat tctgtgcaca accctcaca	300
cgctggtgat ggtgggaagg gaaagatgac aagtcagggg gcatgatccc agcatgtgtg	360
ggaggagctt ctaaattatc cattagcaca agcccgtcag tggccccagg cctaaacatg	420
cagagaaaca ggtgaggaga agcagcgaga gagaaggggc cagggtataaa aaggggccac	480

aagagaccag	ctcaaggatc	ccaaggccca	actccccgaa	ccactcaggg	tcctgtggac	540
agctcactag	cggcaatggc	tgcaggtaag	cgccccataa	atcccccttg	cacaatgtgt	600
cctgagggga	gaggcggcgt	cctgtagatg	ggacgggggc	actaacctc	aggtttgggg	660
cttatgaatg	ttagctatcg	ccatctaagc	ccagtatttg	gccaatctct	gaatgttctt	720
ggtccctgga	ggaggcagag	agagagagag	agaaaaaaaa	aaccagctc	ctggaacagg	780
gagagcgctg	gcctcttgct	ctccagctcc	ctctgttgcc	tccggtttct	ccccaggctc	840
ccggacgtcc	ctgctcctgg	cttttggcct	gctctgcctg	tcctggcttc	aagagggcag	900
tgccttccca	accattccct	tatccagget	ttttgacaac	gctatgctcc	gcgcccgtcg	960
cctgtaccag	ctggcatatg	acacctatca	ggagtittgta	agctcttggg	taatgggtgc	1020
gcttcagagg	tggcaggaag	gggtgaattt	cccccgctgg	gaagtaatgg	gaggagacta	1080
aggagctcag	ggttggtttt	tgaagtgaag	atgcaggcag	atgagcatac	gctgagttag	1140
gttcccagaa	aagtaacaat	gggagcaggt	ctccagcata	gaccttggtg	ggcggtcctt	1200
ctcctaggaa	gaagcctata	tcctgaagga	gcagaagtat	tcattcctgc	agaacccccca	1260
gacctccctc	tgcttctcag	agtcctattc	aacaccttcc	aacagggtga	aaacgcagca	1320
gaaatctgtg	agtggatgcc	ttctccccag	gtgggatggg	gtagacctgt	ggtcagagcc	1380
cccgggcagc	acagccactg	ccggctcctc	ccctgcagaa	cctagagctg	ctccgcatct	1440
ccctgctgct	catccagtca	tggctggagc	ccgtgcagct	cctcaggagc	gtcttcgcca	1500
acagcctggt	gtatggcgcc	tcggacagca	acgtctatcg	ccacctgaag	gacctagagg	1560
aaggcatcca	aacgctgatg	tgggtgaggg	tggcaccagg	atccaatcct	ggggccccac	1620
tggcttccag	ggactgggga	gagaaacact	gctgccctct	ttttagcagt	caggcgctga	1680
cccaagagaa	ctcaccgtat	tcttcatttc	ccctcgtgaa	tcctccaggc	ctttctctac	1740
aacctggagg	ggagggagga	aaatggatga	atgagagagg	gagggaacag	tgcccaagcg	1800
cttggcctct	ccttctcttc	cttcactttg	cagaggctgg	aagatggcag	cccccgact	1860
gggcagatct	tcaatcagtc	ctacagcaag	tttgacacaa	aatcgacaaa	cgatgacgca	1920
ctgctcaaga	actacgggct	gctctactgc	ttcaggaagg	acatggacaa	ggtcgagaca	1980
ttcctgcgca	tcgtgcagtg	ccgctctgtg	gagggcagct	gtggcttcta	gctgccccggg	2040
tggcatecct	gtgacccctc	cccagtgcct	ctcctggctg	tggaaggtgc	tactccagtg	2100
cccaccagcc	ttgtcctaata	aaaattaagt	tgcattcattt	tgtttgacta	ggtgtccttg	2160
tataatatta	tggggtggag	gcgggtggta	tggagcaagg	ggccagggtg	ggaagacaac	2220
ctgtagggcc	ttcagggtct	attcggaac	caggctggag	tgcagtggca	gtcttggtc	2280
gctgcaatct	ccgcctcctg	ggttcaagcg	attctcctgc	ctcagctctc	cgaatagtgt	2340
cgattccagg	catgcaagac	caggctcagc	taatttttgt	atttttggta	gagacgggggt	2400
ttaccatat	tggccagtct	ggtctccatc	tcctgacctc	aggtaatccg	cccgcctcgg	2460
cctcccaaat	tgctgggatt	acaggtatga	gccactgggc	ccttccctgt	cctgtgattt	2520
taaaataatt	ataccagcag	aaggacgtcc	agacacagca	tgggctacct	ggccatgccc	2580
agccagttgg	acatttgagt	tgtttgcttg	gcactgtcct	ctcatgcatt	gggtccactc	2640
agtagatgct	tgttgaattc					2660

<210>19

<211>594

<212>DNA

<213> 人 (Homo sapiens)

<400>19

ctgcccgggt ggcatccctg tgacccctcc ccagtgcctc tcctggtcgt ggaaggtgct	60
actccagtgc ccaccagcct tgctctaata aaattaagtt gcatcatttt gtttgactag	120
gtgtccttgt ataataattat ggggtggagg cgggtggat ggagcaaggg gccaggttgg	180
gaagacaacc ttagggcct tcagggtcta ttcgggaacc aggctggagt gcagtggcag	240
tcttgctcg ctgcaatctc cgctcctgg gttcaagcga ttctcctgcc tcagtctccc	300
gaatagttag gattccagge atgcaagacc aggctcagct aatttttgta tttttggtag	360
agacgggggt tcacatatt ggccagtctg gtctccatct cctgacctca ggtaatccgc	420
ccgctcggc ctcccaaatt gctgggatta caggtatgag ccactgggcc ctccctgtc	480
ctgtgatttt aaaataatta taccagcaga aggacgtcca gacacagcat gggctacctg	540
gccatgcca gccagttgga catttgagtt gtttgcttgg cactgtcctc tcat	594

<210>20

<211>27

<212>DNA

<213> 人工序列

<220>

<223> 引物

<400>20

ttttgatcc ctgcccgggt ggcatcc	27
------------------------------	----

<210>21

<211>30

<212>DNA

<213> 人工序列

<220>

<223> 引物

<400>21

ttttgaattc atgagaggac agtgccaagc	30
----------------------------------	----

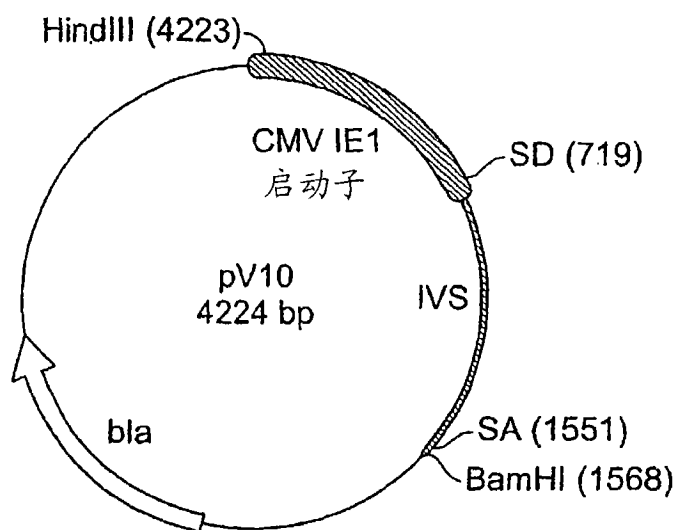


图 1

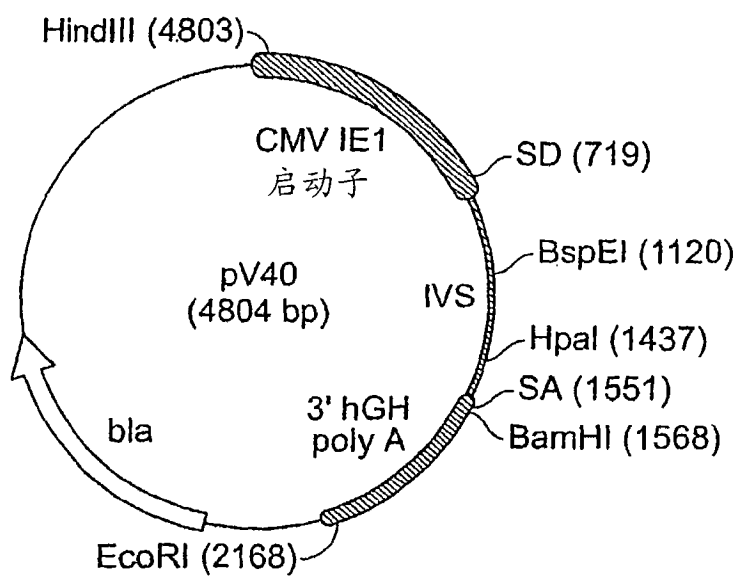


图 2

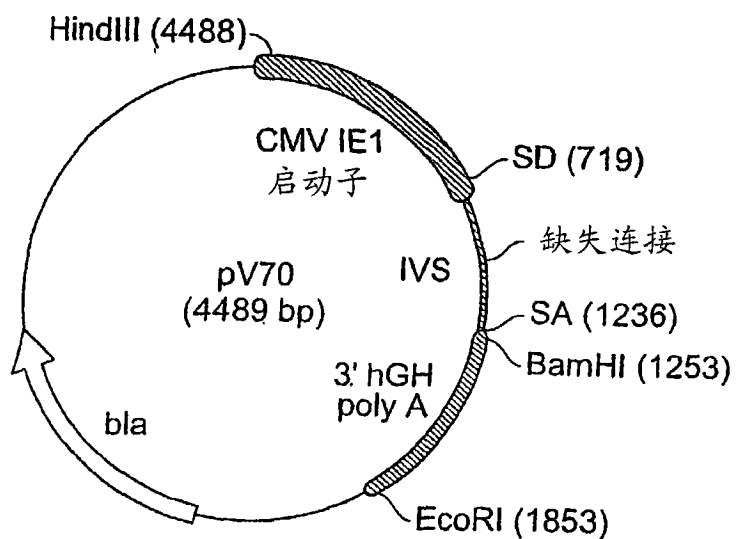


图 3

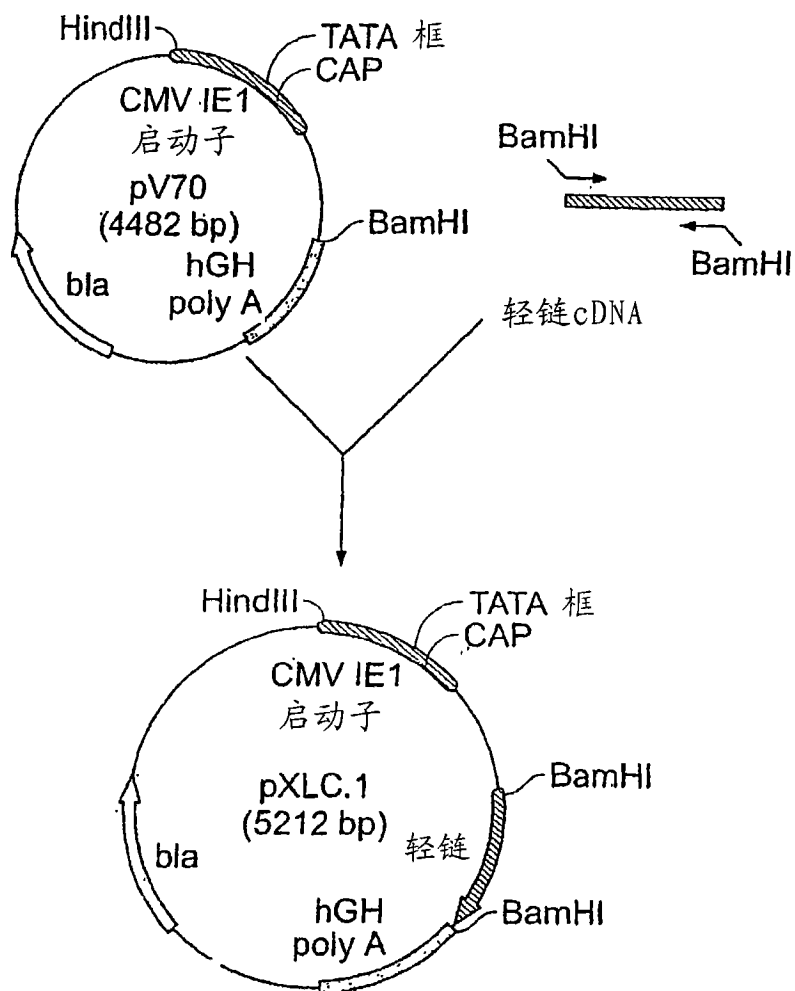


图 4

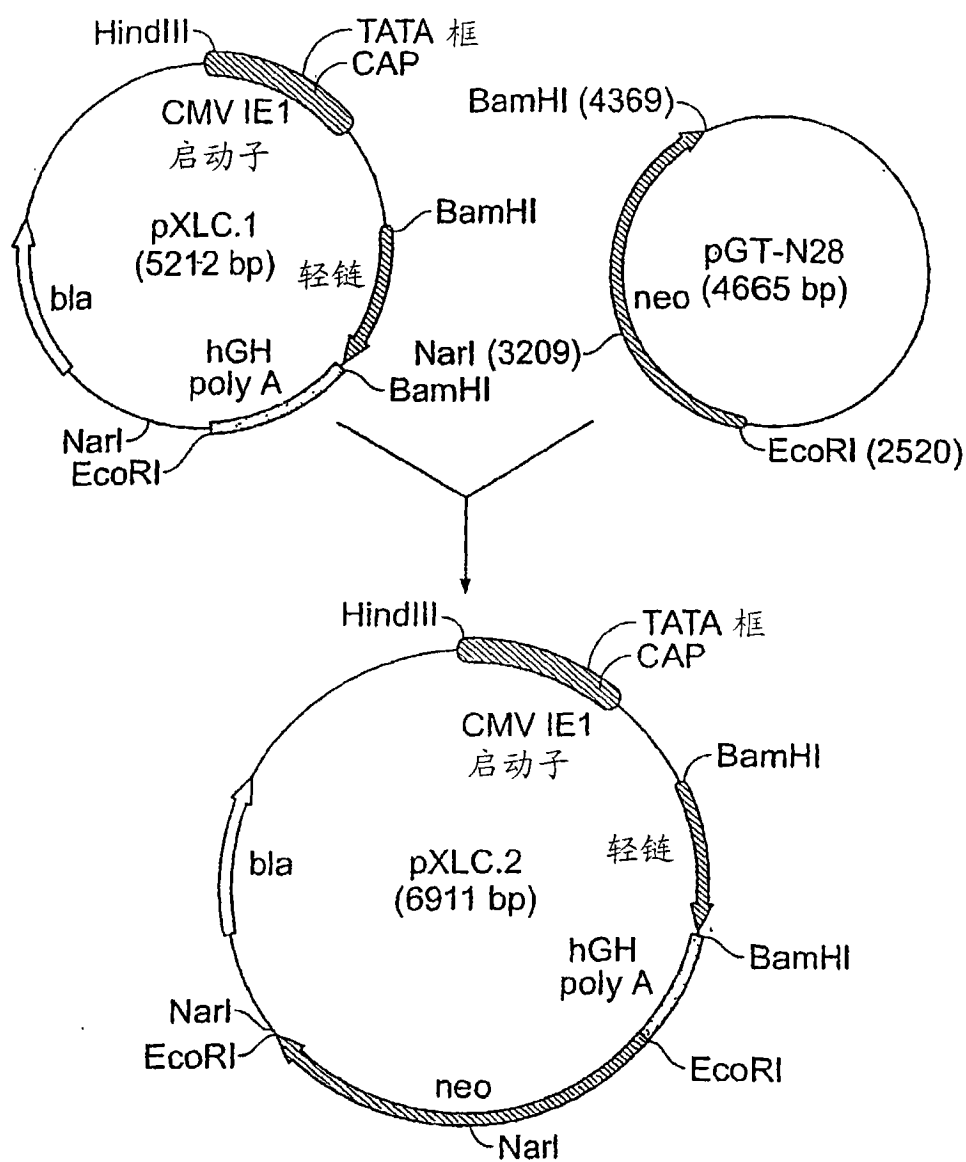


图 5

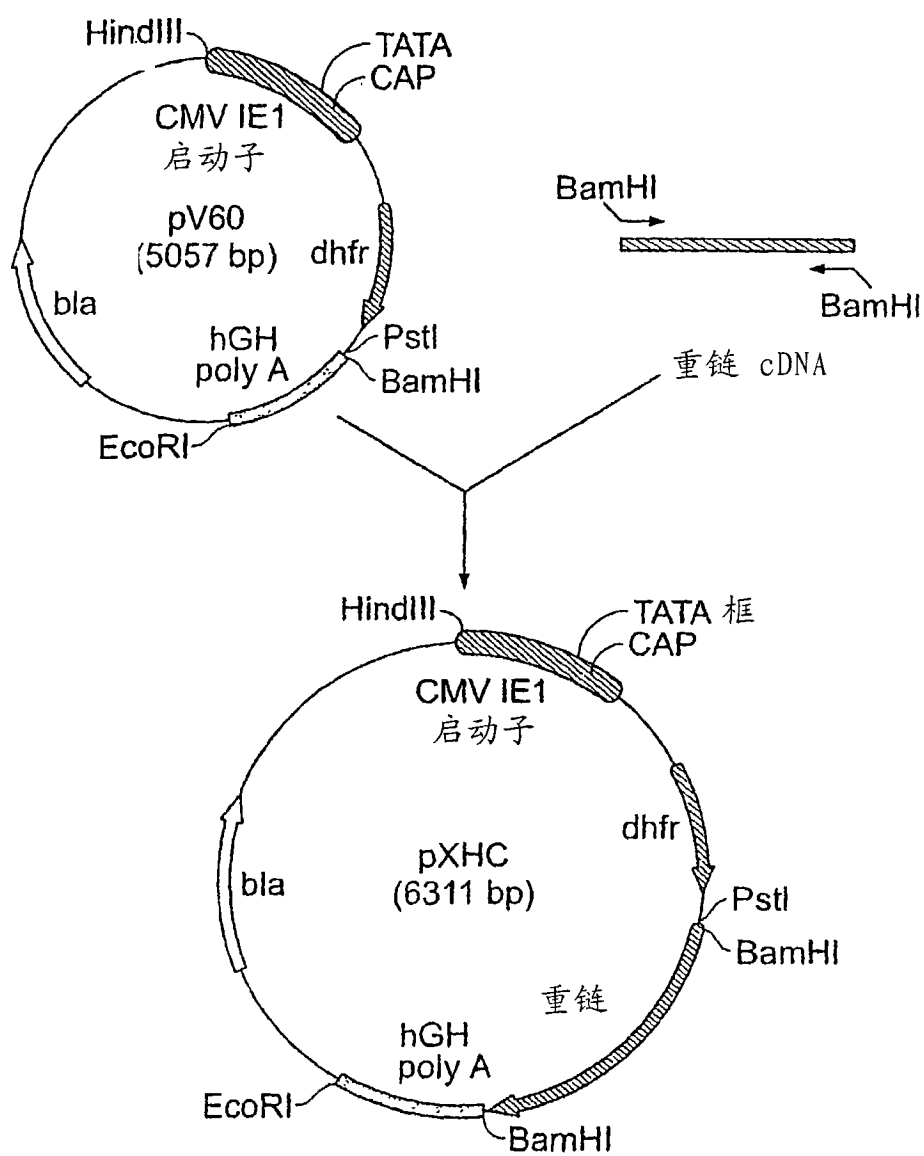


图 6

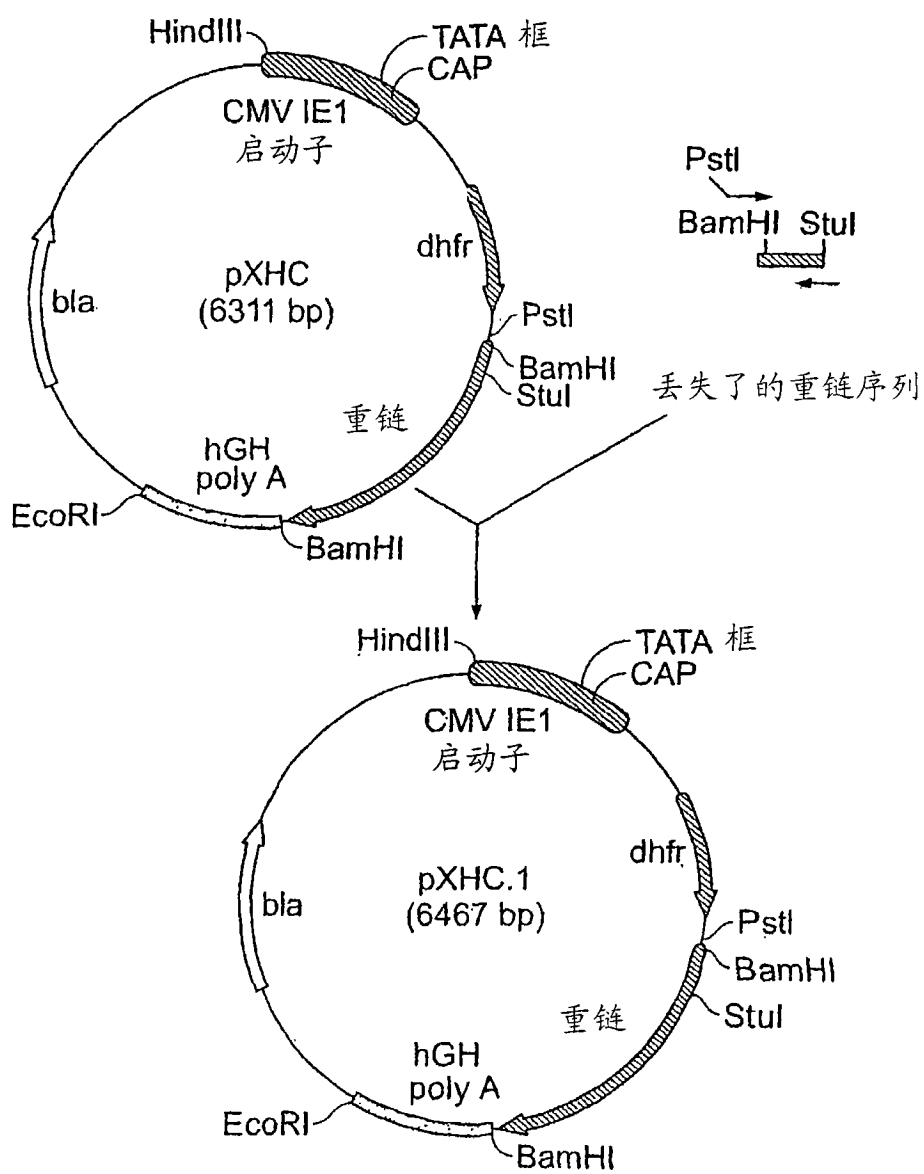


图 7

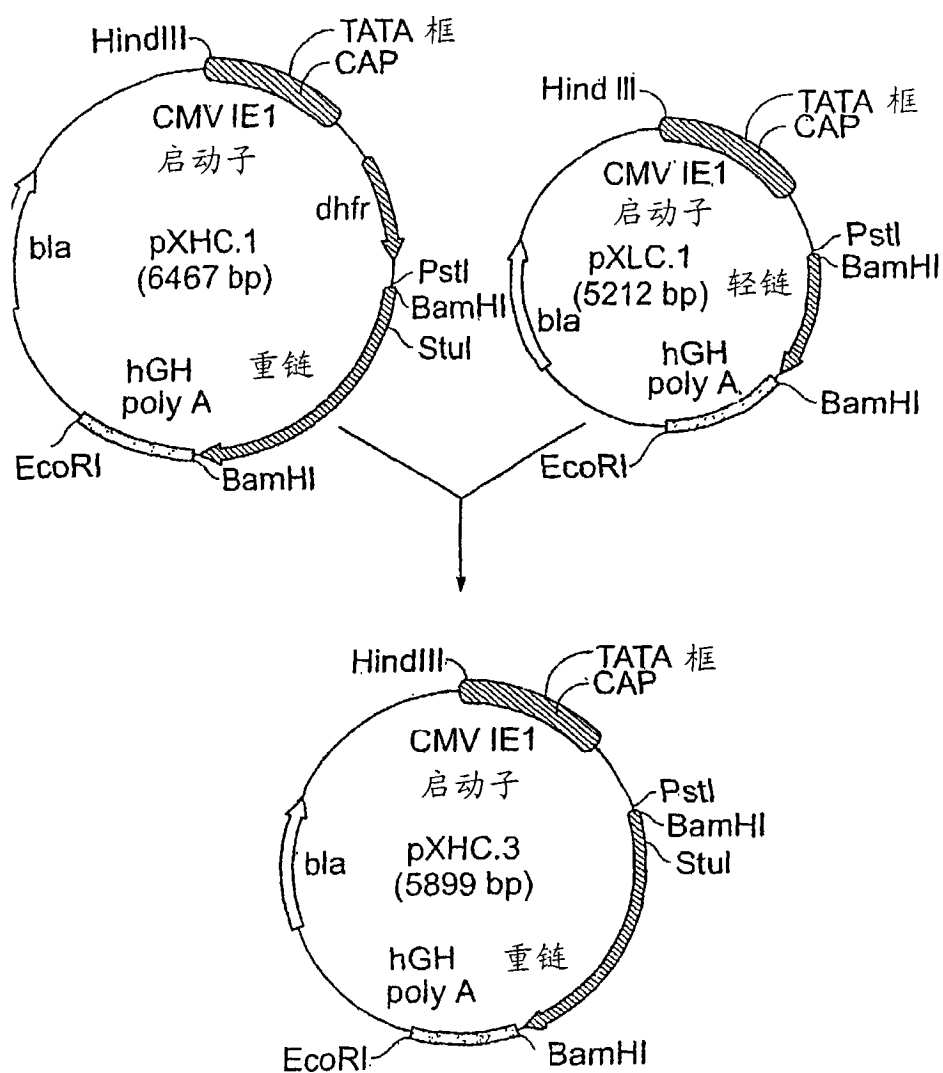


图 8

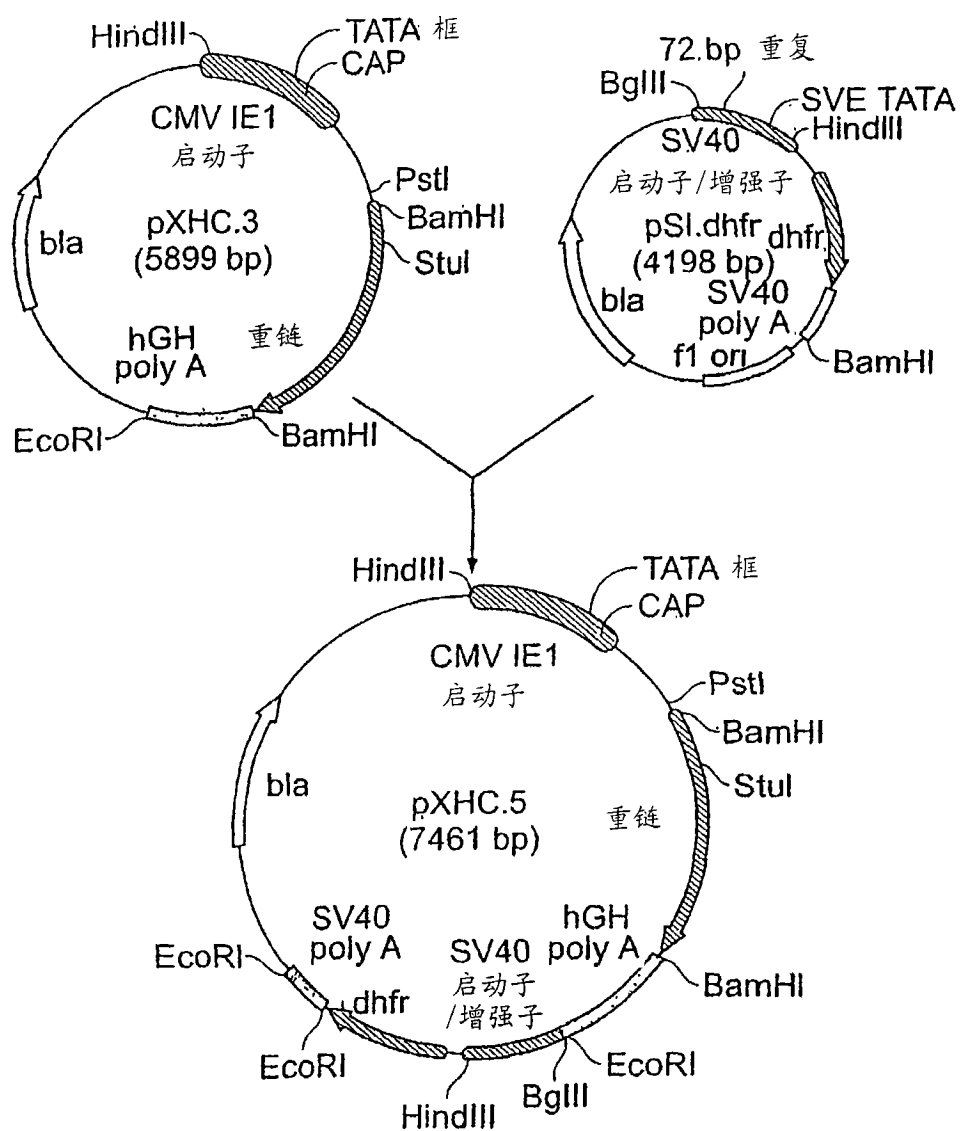


图 9

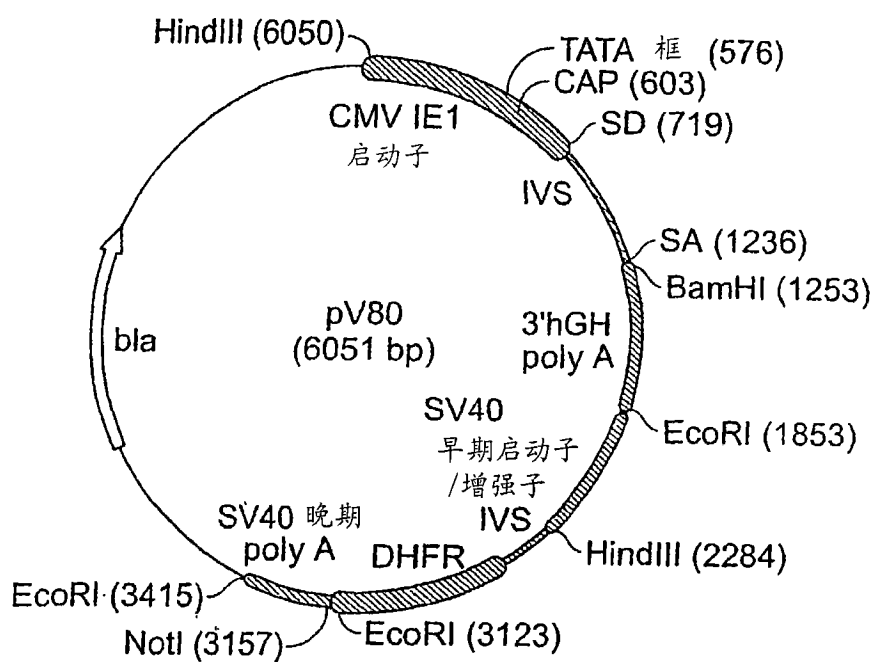


图 10

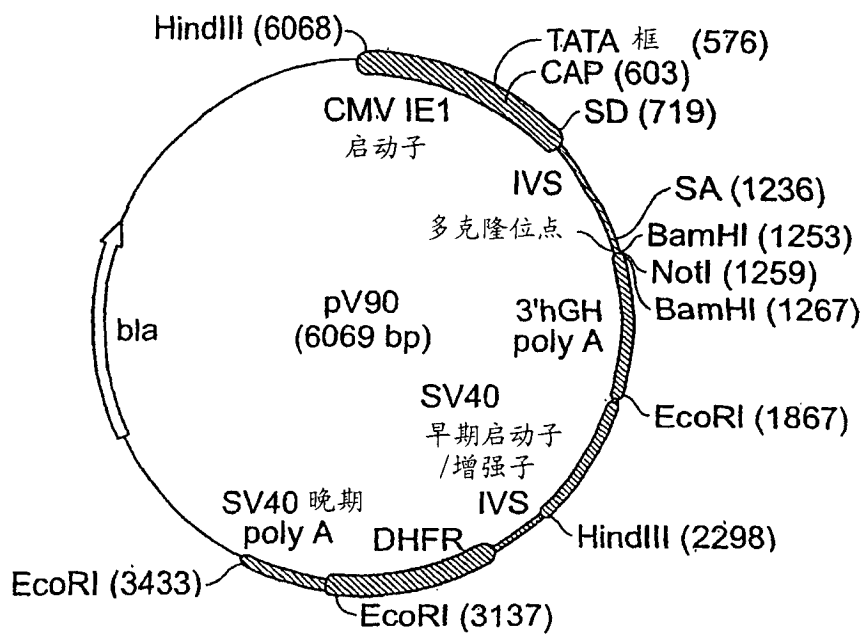


图 11A

标注的Pv90序列

```

1   AGCTTGACATTGATTATAGTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGAGTTCGCGGTTA
87  CATAACTTACGGTAAATGGCCGCTGGCTGACCGCCCAACGACCCCGCCCATGACGTCAATAATGACGTATGTTCCCATAGTA
173 ACGCCAATAGGGACTTTCATTGACGTCAATGGGTGGAGTATTTACGGTAACTGCCCATTTGGCAGTACATCAAGTGTATCATAT
259 GCCAAGTACGCCCTATTGACGTCAATGACGGTAAATGGCCGCTGGCATTATGCCAGTACATGACCTTATGGGACTTTCCTA
345 CTGGGCAGTACATCTACGATTAGTCATCGCTATTACCATGGTGATGCGGTTTGGCAGTACATCAATGGGCGTGGTAGCGGTTT
431 GACTCACGGGATTTCCAAGTCTCCACCCATTGACGTCAATGGGAGTTTGT.TTGGCACCAAAATCAACGGGACTTTCAAAATG
TATA box(576)
517 TCCTAACAACTCCGCCCATTTGACGCAATGGGCGGTAGGCGGTACGGTGGGAGGTCTA.TATAAGCAGAGCTCGTTTATGTAA
CAP(603)
601 CCGTCAGATCGCCTGGAGAGCGCATCCACGCTGTTTGGACCTCCATAGAGACACCGGGACCGATCCAGCCTCGCGGCGGGGAA
687 GGTGCATTGGAACGCGGATTCGCCGTGCAAGA
sp1:ce' donor(719)
720 GTGACGTAAGTACCGCCTATAGAGTCTATAGGCCACCCCTTGGCTTCTTATGGATGCTATACTGTTTGGCTCCCGTCTATA
806 CACCCCGCTTCCCTCATGTTATAGGTGATGGTATAGCTTAGCCTATAGGTGTTGGGTTATTGACCATTTAGGCACTCCCTATTG
978 GTGACGATACCTTCCATTACTAATCCATAACATGGCTCTTTGGCACAACCTCTTTTATGGCTATATGCCAATACACTGTCTTCA
978 GAGACTGACACGGACTCTGTATTTTACAGGATGGGGTCTCATTTATTTTACAAATTCACATATACAAACACCGCTCCCACT
1064 GCCCGCAGTTTATTAACAATAACGTGGGATCTCCACGCAATCTGGGGTACGTGTTCCGGAACGGTGGAGGGCAGTGTAGTCTG
1150 AGCAGTACTCGTTGCTGCCGCGCGCCACAGACATAATAGCTGACAGACTAACAGACTGTTCCTTTCCATGGGCTTTTCTGCA
BamHI(1253)BamHI(1267)
多克隆位点
剪接受体(1236) NotI(1259)
1236 G TCACCGTCTT TGACAGC GATCCGCGCGCGGATCCTGCGCGGGTGGCATCCTGTGACCCCTCCCACTGCC
BamHI NotI BamHI polyadenylation site (1359)
1312 TCTCCTGGTGGTGAAGGTGCTACTCCAGTGCCACACAGCCTTGCTTAATAAAATTAAGTTGCATCATTTTGT
1386 TTGACTAGGTGCTTGTATATAATTATGGGGTGGAGGCGGGTGGTATGGAGCAAGGGGACAGTTGGGAAGACA
1460 ACCCTGTAGGGCTTTCAGGGTCTATTGGGAACAGGCTGGAGTGACGTGGCAGCATCTTGGCTCGCTGCAATCTC
1534 CGCCTCCTGGGTTCAAGCGATTCTCTGCTCAGTGTCCGCAATAGTTGGGATTCACAGGCATGCACGACAGGCG
1608 TCAGCTAATTTTGTATTTTGGTAGAGACGGGGTTTCAACATATTGGCCAGTCTGGTCTCCATCTCCCTGACCT
1682 CAGGTAATCCGCCCGCTCGGCTCCCAAATTGCTGGGATTACAGGTATGAGCCACTGGGCTTCCCTGTCTCT
1756 GTGATTTTAAATAATTATACAGCAGAGGACGCTCCAGACACAGCATGGGCTACCTGGCCATGCCAGCCAGT
SalI(1873)
SV40早期启动子/增强子(1868)
EcoRI(1867)
1830 TGGACATTTGAGTTGTTTGGTGGCACTGTCTCTCATGAATTCGTGACAGATCTGCGCAGCACCATGGCCTGAAATAA
1910 CCTCTGAAAGAGGAACCTTGGTTAGGTACCTTCTGAEGCGGAAAGAACAGCTGTGGAATGTGTGTCAGTTAGGCTGTGGAAGTCC
1996 CCAGGCTCCCGCAGCAGGCAAGTATGCAAAGCATGCTCAATAGTCAGCAACAGGTGTGGAAGTCCCGAGGCTCCCAAGC
2002 AGGCAGAGATGCAAGCATGCTCAATAGTCAGCAACCATAGTCCCGCCCTAACCTCCGCCCATCCGCCCCCTAACCTCCGC
2168 CCAGTCCCGCCCATCTCCGCCCATGGCTGACTAATTTTATTTATTCAGAGGCGGAGGCCGCTCGGCTCTGAGCTATTCT
HindIII(2298)
2254 CAGAAGTAGTGAGGAGGCTTTTGGAGGCTTAGGCTTTTGCAAAAGCTTGATTCTCTGACACAACAGTCTCGAACTTAAGCTG
2340 CAGAAGTTGGTGGTGGGCACTGGGCAAGTATCAAGGTTACAAGACAGGTTTAAAGGAGACCAATAGAAACTGGGCTTGTGCA
2426 GACAGAGAAGACTCTTGGCTTCTGATAGGCACCTATTGGTCTTACTGACATCCACTTGGCTTCTCTCCACAGGTGTCCACTCC
DHFR coding sequence (2568)
2512 CAGTTCAATTACAGCTCTTAAGGCTAGAGTACTTAATACGACTCACTATAGGctagCATGGTTCGACCATTTGAAGTGCATCG
2594 TCGCCGTGTCCTCAAAATATGGGGATTGGCAAGAACCGAGACCTACCTGGCTCCGCTCAGGAACGAGTTCAAG
2668 TACTTCCAAAGATGACCAACCTCTTCAAGTGAAGGTAAACAGAATCTGGTATTATGGGTAGGAAACCTG
2742 GTTCTCCATCTCTGAGAAGATCGACCTTTAAAGGACAGAAATTAATATAGTTCTCAGTAGAGAACTCAAAGAAC
2816 CACCACGAGGAGCTCATTTCTTGGCAAAAGTTGGATGATGCTTAAGACTTATTGAACAACCGGAATTGGCA
2990 AGTAAAGTAGACATGGTTGGATAGTCCGAGGCGAGTCTGTTTACCAGGAAGCCATGAATCAACCAGGCCACCT
2964 CAGACTCTTTGTGACAAGGATCATGAGGAATTTGAAAGTGACAGGTTTCCAGAAATTTGATTTGGGGAAT
3038 ATAACTTCTCCAGAAATACCGAGGCTCTCTGAGGTCAGGAGGAAAAAGGCATCAAGTATAAGTTGAA
EcoRI(3137) SalI(3161)
3112 GTCTACGAGAAGAAAGACTAACTCGAGAAATTCACGCTGGTACCTCTAGAGTCGACCCGGGCGGCGCCGCTTCTGAGCAGAC

```

图 11B-1

3195 ATGATAAGATACATTGATGAGTTTGGACAAACCAACTAGAAATGCAATGAAAAAATGCTTTATTTGTGAAATTTGTGATGCTAT
 聚腺苷酸化位点 (3308)
 3281 TGCTTTATTTGTAACCATTAAGCTGCAATAAACAGTTAACAAACAATTGCATTATTTTATGTTTCAGGTTTCAGGGGGAGG
 EcoRI (3433)
 3367 TGTGGGAGGTTTTTAAAGCAAGTAAACCTCTACAAATGTGGTAAATCGATAAGGATCTGTGACGAATTCACCTGGCCGTCGTT
 SalI (3427)
 3453 TTACAACGTCGTGACTGGGAAACCCCTGGCGTTACCCAACCTAATCGCCTTGACGACATCCCCCTTTCGCCAGCTGGCGTAATAG
 3539 C6AAGAGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGCGCCTGATGCGGTATTTCTCCTTACGC
 3625 ATCTGTGCGGTATTTACACCCGATATGGTGCACCTCAATACAATCTGCTCTGATGCCGATAGTTAAGCCAGCCCGACACCCG
 3711 CCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTCTCCGCGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCAT
 3797 GTGTGASAGGTTTTACCGTCATCACCAGAACGCGCGAGACGAAAGGGCTCGTGATACGCTATTTTATAGGTTAATGTCATGA
 3883 TAATAATGTTTTCTTAGCGTCAGGTGGCACTTTTCGGGAAATGTGCGGGAACCCCTATTTGTTATTTTCTAAATACATTCA
 3969 AATATGTATCCGCTCATGAGACAATAACCTGATAAATGCTTCAATAATATTGAAAAGGAAGATATGAGTATTTCAACATTTCCG
 4055 TGTGCGCCTTATTCCTTTTTTGGCGCATTTGCTTCTGTTTTGCTCAGCCAGAAACGCTGGTGAAGTAAAAGATGCTGAAG
 4141 ATCAGTTGGGTGCAGAGTGGTTACATCGAATGGAATCTCAACAGCGGTAAAGTCTTGAGAGTTTTGCGCCCGAAGAACGTTTT
 4227 CCAATGATGAGCACTTTAAAGTTCTGCTATGTGGCGCGGTATTTATCCGATTTGACGCGGGCAAGAGCAACTCGGTGCGCGCAT
 4313 ACACATATTTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAGCATCTTACGGATGGCATGACAGTAAGAGAAATTATGCA
 4399 GTGCTGCCATAACCATGAGTGATAACACTGCGGCCAATTAATCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTG
 4485 CACAACATGGGGATCATGTAATCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAACGACGAGCGGTGACACCAC
 4571 GATGCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAAAGACT
 4657 GGAATGGAGGCGGATAAAGTTGACGAGCACTTCTGCGCTCGGCCCTCCGCTGCGTGGTTTATTGCTGATAAATCTGGAGCGCGT
 4743 GAGCGTGGGTCTCGCGGTATCATTGACGACCTGGGCGCAGATGGTAAGCCCTCCGATCGTAGTTATCTACACGACGGGAGTCA
 4829 GGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCTCACTGATTAAAGCATTGGTAACTGTCAGACCAAGTTTACT
 4915 CATATATACCTTAGATTGATTTAAACTTCATTTTAAATTTAAAGGATCTAGGTGAATATCTTTTGATAATCTCATGACCAAA
 5001 ATCCCTTAACGTGAGTTTTGTTCCACTGAGCGTCAGACCCGTAAGAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTCTGCG
 5087 CGTAATCTGCTGCTTGCAAAACAAAAACCCGCTACACGCGGTTGTTGTTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGA
 5173 AGGTAACCTGGCTTCAGCAGAGCGCAGATACCAAACTACTGCTCTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTA
 5259 GCACCGCTACATACCTCGCTCTGTAATCTGTTACAGTGGCTGCTGCCAGTGGCGATAAGTGTGCTTACCGGTTGGACTC
 5345 AAGACGATAGTTACCGATAAGGCGCAGCGGTGCGGCTGAACGGGGGTTGCTGACACAGCCAGCTTGGAGCGAACGACCTACA
 5431 CCGAATCTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCAGCTTCCCGAAGGGAGAAAGCGGACAGGTATCCCGTAAGCGGC
 5517 AGGCTCGGAACAGGAGAGCGCAGGAGGAGCTTCCAGGGGAAACGCTGGTATCTTTATAGTCTGTCGGGTTTCGCCACTCTG
 5603 ACTTGAGCGTCTGATTTTGTGATGCTGTCAGGGGGGCGGAGCCTATGGAATAACCGCAGCAACCGCGCTTTTACGGTTCTCGG
 5689 CTTTTGCTGGCTTTTGTCTACATGTTCTTCTGCTTATCCCTGATCTGTGGATAACCGTATTACGGCTTTGAGTGAGCT
 5775 GATACCGCTCGCCGAGCCGAACGAGCGCAGCGAGTCACTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAACCGCCTCT
 5861 CCCCGCGCTTGGCCGATTCATTAATGACGCTGGCAGCAGGTTTCCGACTGGAAGCGGGCAGTGAGCGCAACGCAATTAAATG
 5947 TGAGTTAGCTCACTATTAGGCACCCAGGCTTACACTTTATGCTTCCGCTCGTATGTTGTGGAATTGTGAGCGGATAACAA
 6033 TTTCACACAGGAACAGCTATGACCATGATTACGCCA

图 11B-2

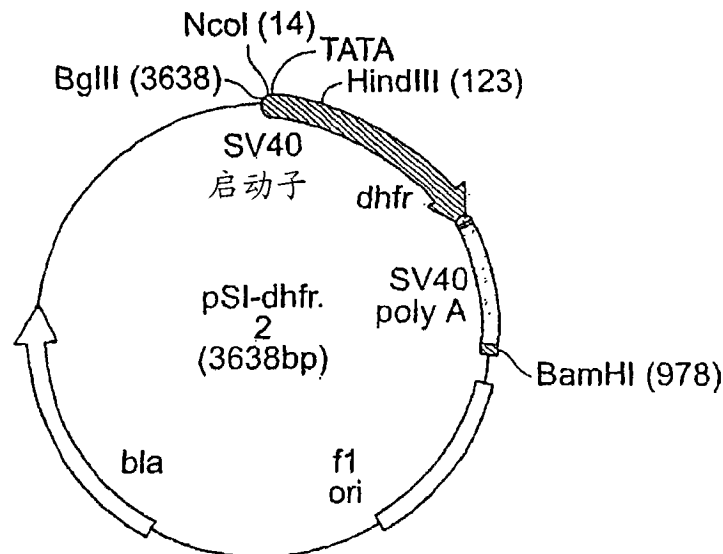


图 12

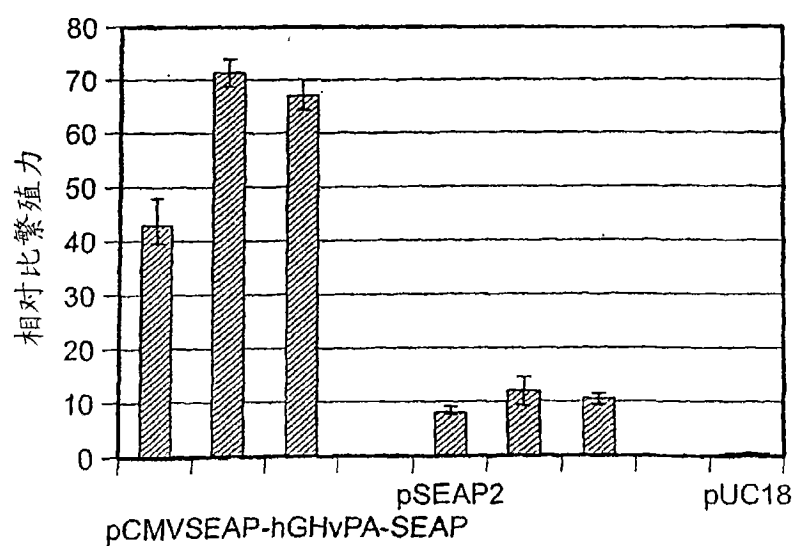


图 13

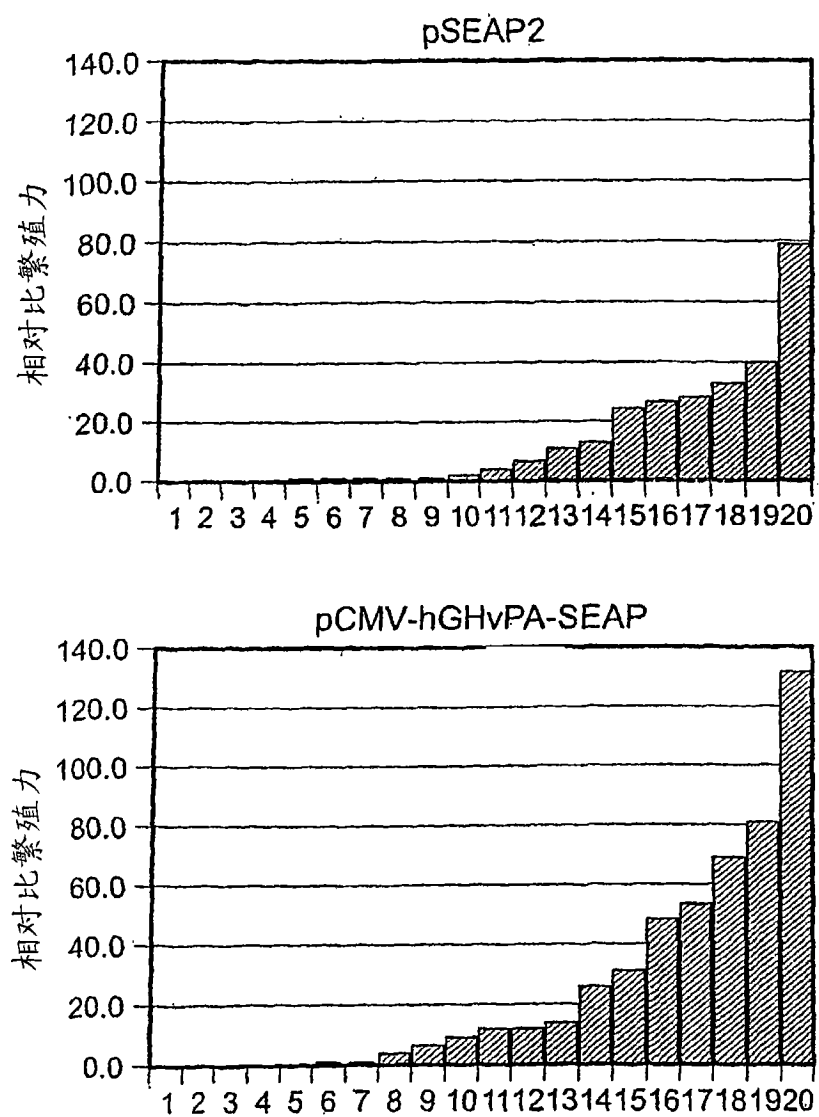


图 14

2660 bp DNA 线性 (SEQ ID-NO:18)
 定义生长激素突变前体 (GH-V) 基因 完整的cds
 ACCESSION K00470 VERSION K00470.1 G1:183174
 sig 肽 连接 (556..565,837..904) / 基因 ="GH-V"
 mat 肽 连接 (905..997,1208..1327,1419..1583,1834..2028) / 基因 ="GH-V"
 / 产物 ="生长激素突变体"
 exon 837..997 / 基因 ="GH-V" / 号=2
 exon 1208..1327 / 基因 ="GH-V" / 号=3
 exon 1419..1583 / 基因 ="GH-V" / 号=4
 exon 1834..2136 / 基因 ="GH-V" / 号=5
 poly A 信号 2118..2123 / 基因 ="GH-V"
 重复区 2234..2505 / rpt 家族 ="Alu" 碱基计数 593 a 719 c 753 g 595 t
 gaattcagca ctgaatcatg cccagaaacc ccgcaatcta ttggctgtgc tttggccctt 61
 tttcccaaca cacacattct gtctggtggg tggaggggaa acatgcgggg agggagaaag 121
 gaataggata gagagtggga tgggtcgggt aggggtctca aggactggcc tatcctgaca 181
 tccttctccg cgttcagggt ggccaccatg gccgtgtgcc agagggcacc cacgtgacc 241
 ttaagagagag gacaagtgg gtggtatctc tggctgacat tctgtgcaca accctcacaa 301
 cgctggtgat ggtgggaagg gaaagatgac aagtcagggg gcatgatccc agcatgtgtg 361
 ggaggagctt claaattatc cattagcaca agcccgtcag tggcccagg cctaaacatg 421
 cagagaaaca ggtgaggaga agcagcgaga gagaaggggc caggtataaa aagggccac 481
 aagagaccag ctcaaggatc ccaaggccca actccccgaa ccactcaggg tcctgtggac 541
 agctcactag cggcaatggc tgcaggtaag cggccctaaa atcccttgg cacatgtgt 601
 cctgagggga gaggcggcgt cctgtatgat ggacgggggc actaaccttc aggttgggg 661
 cttatgaatg ttatgtatcg ccactaagc ccagtatttg gccaatctct gaattgtcct 721
 ggtccctgga ggaggcagag agagagagag agaaaaaaa aaccagctc ctggaacagg 781
 gagagcgctg gcctcttctc ctccagctcc ctctgttgc tccggttct cccaggtc 841
 ccgacgtcc ctgctcctgg cttttggctt gctctgctg tcctggcttc aagagggcag 901
 tgccctccca accattccct tatccaggct tttgacaac gctatgctc gcgcccgtc 961
 cctgtaccag ctggcatatg acacctatca ggagtgtgt agctcttgg taatgggtgc 1021
 gcttcagagg tggcaggaaag ggggtgaattt ccccgctgg gaagtatgg gagagagcta 1081
 aggagctcag ggttgtttc tgaagtgaat atgcaggcag atgagcata gctgagtgg 1141
 gtccccagaa aagtaacaat gggagcagg ctccagcata gaccttggg ggcggtcct 1201
 ctctagagaa gaagcctata tctgaaggga gcagaagat tcatctctgc agaaccacca 1261
 gacctccctc tgcctctcag agtctattcc aacaccttcc aacagggtga aaacgcagca 1321
 gaaatctgtg agtggatgcc ttctcccag gtgggatggg gtagaectgt ggtcagagcc 1381
 cccggggcagc acagccactg ccggtccttc cctgcagaa cctagagctg ctccgcatct 1441
 cctgctgct catccagtc tggctggagc ccgtgcagct cctcaggagc gtcttcgcca 1501
 acagcctggt gtatggcgcc tcggacagca acgtctatcg ccacctgaag gacctagagg 1561
 aaggcatcca aacgtgatg tgggtgaggg tggcaccagg atccaatcct ggggcccac 1621
 tggcttccag ggactgggga gagaaacact gctgccctct ttttagcagt caggcgctga 1681
 cccaagagaa ctccacgtat tcttcattc ccctcgtgaa tctccaggc ctttctctac 1741
 aacctggagg ggagggggga aatggatga atgagagagg gagggaaacag tgcccaagcg 1801
 cttggcctct ccttctcttc ctccactttg cagaggctgg aagatggcag ccccggaact 1861
 gggcagatct tcaatcagtc ctacagcaag ttgacacaa aatcgacaa cgtatgacga 1921
 ctgctcaaga actacgggct gctctactgc ttccaggaag acatggacaa ggtcgagaca 1981
 ttctctgcga tcgtgcagtg ccgctctgtg gagggcagct gtggcttcta gctgcccggg 2041
 tggcatccct gtgacccctc ccagtgctt ctctggtcg tggaaagtgc tactccagtg 2101
 cccaccagcc ttgtcctaat aaaattaaat tgcattatit tgtttgacta ggtgtccttg 2161
 tataatatta tggggtggag gcgggtggtg tggagcaagg ggccagggtt ggaagacaac 2221
 ctgtaggggc ttccagggtct attcgggaac caggctggag tgcagtggca gtcttggctc 2281
 gctgcaatct ccgctcctg ggttcaagcg attctcctgc ctgagctccc cgaatagtgt 2341
 cgattccagg catgcaagac caggctcagc taatttttgt atttttggta gagacgggg 2401
 ttaccataat tggccagtct ggtctccatc tctgacctc aggtaatccg cccgctcgg 2461
 cctcccaaat tgctgggatt acaggtatga gccactgggc ccttcctgt cctgtgattt 2521
 taaaataatt ataccagcag aaggacgtcc agacacagca tgggctacct ggccatgccc 2581
 agccagttgg acatttgagt tgtttgcttg gcactgtcct ctcatgcatt ggttccactc 2641
 agtagatgct tgttgaattc

图 15