(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 103080321 A
(43)申请公布日 2013.05.01

(21)申请号 201180043099.8

(22)申请日 2011.07.06

(30)优先权数据
2010-154782 2010.07.07 JP

(85)PCT申请进入国家阶段日
2013.03.07

(86)PCT申请的申请数据
PCT/JP2011/065916 2011.07.06

(87)PCT申请的公布数据
WO2012/005378 EN 2012.01.12

(71)申请人 第一三共株式会社
地址 日本东京都

(72)发明人 西宫大祐 井上龙也

(74)专利代理机构 中国专利代理(香港)有限公司 72001
代理人 孔青 权陆军

(51)Int.Cl.
C12N 15/85(2006.01)

权利要求书2页 说明书36页
序列表36页 附图17页

(54)发明名称

具有提高外源基因表达活性的DNA元件

(57)摘要

公开了使用新颖的DNA元件在哺乳动物细胞中稳定获得外源基因的高表达的方法。更特别地，公开了通过改变已引入外源基因表达单元的基因位点周围的染色质结构来增强转录激活的DNA元件。

CN 103080321 A

1. 由序列表中的 SEQ ID NO: 1 代表的多核苷酸序列组成的多核苷酸。

2. 由序列表中的 SEQ ID NO: 2 代表的多核苷酸序列组成的多核苷酸。

3. 由序列表中的 SEQ ID NO:3 代表的多核苷酸序列组成的多核苷酸。

4. 由序列表中的 SEQ ID NO: 4 代表的多核苷酸序列组成的多核苷酸。

5. 由序列表中的 SEQ ID NO: 5 代表的多核苷酸序列组成的多核苷酸。

6. 包含由序列表中的 SEQ ID NO: 1-5 中任一个所代表的多核苷酸序列的至少 3000 个连续核苷酸的多核苷酸。

7. 包含由序列表中的 SEQ ID NO: 1-5 中任一个所代表的多核苷酸序列的至少 2000 个连续核苷酸的多核苷酸。

8. 包含由序列表中的 SEQ ID NO: 1-5 中任一个所代表的多核苷酸序列的至少 1500 个连续核苷酸的多核苷酸。

9. 由与权利要求 1-8 中任一项的多核苷酸的多核苷酸序列具有 95% 或更高同源性的多核苷酸序列组成的多核苷酸。

10. 由与权利要求 1-8 中任一项的多核苷酸的多核苷酸序列具有 99% 或更高同源性的多核苷酸序列组成的多核苷酸。

11. 由多核苷酸序列组成的多核苷酸，所述多核苷酸序列含有权利要求 1-10 中任一项的多核苷酸的多核苷酸序列的两种或多种序列。

12. 由两种或多种类型的选自权利要求 1-10 中任一项的多核苷酸的多核苷酸组成的多核苷酸。

13. 包含权利要求 1-12 中任一项的多核苷酸的多核苷酸序列的外源基因表达载体。

14. 权利要求 13 的外源基因表达载体，其中由所述外源基因编码的蛋白是多聚体蛋白。

15. 权利要求 13 的外源基因表达载体，其中由所述外源基因编码的蛋白是异多聚体蛋白。

16. 权利要求 15 的外源基因表达载体，其中由所述外源基因编码的蛋白是抗体或其功能片段。

17. 转化细胞，其中已引入权利要求 13-16 中任一项的外源基因表达载体。

18. 权利要求 17 的转化细胞，其中所述细胞为源自哺乳动物的培养细胞。

19. 权利要求 18 的转化细胞，其中所述源自哺乳动物的培养细胞为选自 COS-1 细胞、293 细胞和 CHO 细胞的细胞。

20. 权利要求 17-19 中任一项的转化细胞，其中由所述外源基因编码的蛋白是多聚体蛋白。

21. 权利要求 20 的转化细胞，其中由所述外源基因编码的蛋白是异多聚体蛋白。

22. 权利要求 21 的转化细胞，其中由所述外源基因编码的蛋白是抗体或其功能片段。

23. 用于生产蛋白的方法，所述方法特征在于包括培养权利要求 17-22 中任一项的转化细胞，并从得到的培养产物中获得由所述外源基因编码的蛋白。

24. 用于在已引入外源基因或外源基因表达载体的转化细胞中提高外源基因表达的方法，所述方法特征在于使用权利要求 1-12 中任一项的多核苷酸或权利要求 13-16 中任一项的外源基因表达载体。

25. 权利要求 1-12 中任一项的多核苷酸用于在转化细胞中提高外源基因表达的用途。

具有提高外源基因表达活性的 DNA 元件

技术领域

[0001] 本发明涉及转化的哺乳动物宿主细胞，其分泌外源蛋白的能力已通过利用具有 DNA 元件的外源基因表达载体而提高，并涉及利用宿主细胞生产外源蛋白的方法。

背景技术

[0002] 由于遗传重组技术的发展，蛋白质药物产品例如治疗性蛋白和抗体药物的市场快速增长。特别是抗体药物具有高特异性，即使将其给予人体也不引起不利的免疫反应，因此，一直在对其积极进行开发。

[0003] 可使用微生物、酵母、昆虫、动物或植物细胞、转基因动物或植物细胞等等，作为在其中产生由抗体药物代表的蛋白质药物的宿主细胞。为了使蛋白质药物具有生物学活性或免疫原性，翻译后修饰例如折叠或糖基化是必要的，因此不能进行复杂的翻译后修饰的微生物或具有不同聚糖结构的植物不适合作为以生物反应器操作的宿主细胞。考虑到培养的哺乳动物细胞例如来自于与人亲缘很近的物种的 CHO 细胞具有与人相似的聚糖结构，而且安全，并且可以使用所述细胞进行翻译后修饰，目前标准是使用这类细胞。

[0004] 在将培养的哺乳动物细胞用作宿主细胞的情况中，与微生物等（见非专利文件 1）相比存在生长率低、生产率低、成本高等问题。此外，为了在临床试验中使用蛋白质药物产品，有必要给予大量产品。因此，其生产能力的缺乏也是世界性难题。因此，为了提高外源基因在培养的哺乳动物细胞中的生产率，迄今已进行了启动子、增强子、药物选择标记、基因扩增和培养工程技术等等的大量研究。然而，目前的情形是尚未建立能始终如一地提高基因表达的系统。作为外源蛋白低生产率的的原因之一，考虑到“位置效应”（见非专利文件 2）。当将外源基因引入宿主细胞时，其被随机地整合到宿主染色体基因组，并且外源基因的转录受到外源基因整合区域周围 DNA 的很大影响。然而，位置效应受到例如外源基因的插入位点、拷贝数、结构等因素的影响，非常难以控制染色体中的插入位点。

[0005] 为了解决所述问题，业已鉴别了调控多核苷酸序列（也叫 DNA 元件），例如基因座控制区 (LCR)、支架 / 基质附着区 (S/MAR)、绝缘子、遍在染色质开放元件 (UCOE) 和抗阻遏物 (STAR 元件）等（见非专利文件 3-6）。LCR 不需要在内源基因的基因座打开染色质结构。然而，LCR 是转录调控元件，其具有打开外源基因被整合位点的 DNA 周围染色质结构的能力及当其与外源基因表达单元一起使用时重塑大范围染色质的能力，并且认为其需要富含 AT 的区域（见非专利文件 7）。

[0006] 常将 LCR 代表的上述 DNA 元件与启动子联合使用，已知与只用启动子的情况相比，在 DNA 元件与启动子联合使用的情况中，外源基因表达水平增加。然而，至今报道了极少 DNA 元件类型，并且促进外源基因表达增强的各种机制也相互不同。另外，即使 DNA 元件和启动子联合使用，在 DNA 元件和启动子控制下也不能生产足量的治疗性蛋白。因此，不能说已获得能增加外源蛋白生产率的 DNA 元件的充足知识。

[0007] 因此，本发明目标是提供增加在蛋白质药物产品中待用的外源蛋白产量的方法，所述方法使用具有在宿主细胞例如培养的哺乳动物细胞中提高外源基因表达的高活性的

DNA 元件。

[0008] 引文列表

非专利文献

NPL 1: Florian M. Wurm. (2004) Production of recombinant protein therapeutics in cultivated mammalian cells (在培养哺乳动物细胞中生产重组蛋白治疗药物). Nat. Biotechnol. 22(11): 1393-1398

NPL 2: Ted H. J. Kwaks 和 Arie P. Otte. (2006) Employing epigenetics to augment the expression of therapeutic proteins in mammalian cells (采用表观遗传学来增加哺乳动物细胞中治疗性蛋白的表达). TRENDS in Biotechnol. 24(3): 137-142

NPL 3: Pierre-Alain Girod, Duc-Quang Nguyen. 等 (2007) Genome-wide prediction of matrix attachment regions that increase gene expression in mammalian cells (在哺乳动物细胞中增加基因表达的基质附着区的基因组范围预测). Nat. Methods. 4(9): 747-753

NPL 4: Adam C. Bell, Adam G. West, Gary Felsenfeld (2001) Insulators and Boundaries: Versatile Regulatory elements in the Eukaryotic Genome (绝缘子和边界:真核基因组中多用途调控元件), Science 291: 447-450

NPL 5: Steven Williams, Tracey Mustoe. 等 (2005) CpG-island fragments from the HNRPA2B1/CBX3 genomic locus reduce silencing and enhance transgene expression from the hCMV promoter/enhancer in mammalian cells (来自 HNRPA2B1/CBX3 基因组位点的 CpG- 岛片段在哺乳动物细胞中降低沉默并提高自 hCMV 启动子 / 增强子的转基因表达). BMC Biotechnol. 5(17): 1-9

NPL 6: Arie P. Otte, Ted H. J. Kwaks. 等 (2007) Various Expression-Augmenting DNA elements Benefit from STAR-Select, a Novel High Stringency Selection System for Protein Expression (各种增强表达的 DNA 元件受益于 STAR-Select 这种用于蛋白质表达的新型高严格性选择系统). Biotechnol. Prog. 23: 801-807

NPL 7: Qiliang Li, Kenneth R. Peterson, Xiangdong Fang 和 George Stamatoyannopoulos, (2002) Locus control regions (基因座控制区), Blood 100(9): 3077-3086

发明概述

技术问题

如上所述,仍然没有许多类型的作为调控多核苷酸序列的 DNA 元件,另外,其中只有极少 DNA 元件在提高外源基因表达中高度有效。本发明的目标是提供利用 DNA 元件在哺乳动物细胞中稳定实现高表达的方法等等,所述 DNA 元件通过伴随其中已引入外源基因表达单元的基因座周围的染色质结构的变化来提高转录激活。

[0009] 问题的解决方案

本发明人为了解决上述问题进行了深入的研究,结果发现待表达的外源蛋白在培养的哺乳动物细胞中的生产率和分泌可通过使用一种或多种特定类型的 DNA 元件来提高,由此完成了本发明。

[0010] 换句话说，本发明包括以下发明。

[0011] (1) 由序列表中的 SEQ ID NO: 1 代表的多核苷酸序列组成的多核苷酸。

[0012] (2) 由序列表中的 SEQ ID NO: 2 代表的多核苷酸序列组成的多核苷酸。

[0013] (3) 由序列表中的 SEQ ID NO: 3 代表的多核苷酸序列组成的多核苷酸。

[0014] (4) 由序列表中的 SEQ ID NO: 4 代表的多核苷酸序列组成的多核苷酸。

[0015] (5) 由序列表中的 SEQ ID NO: 5 代表的多核苷酸序列组成的多核苷酸。

[0016] (6) 包含序列表中的 SEQ ID NO: 1-5 中任一个所代表的多核苷酸序列的至少 3000 个连续核苷酸的多核苷酸。

[0017] (7) 包含序列表中的 SEQ ID NO: 1-5 中任一个所代表的多核苷酸序列的至少 2000 个连续核苷酸的多核苷酸。

[0018] (8) 包含序列表中的 SEQ ID NO: 1-5 中任一个所代表的多核苷酸序列的至少 1500 个连续核苷酸的多核苷酸。

[0019] (9) 由与 (1)-(8) 中任一项的多核苷酸的多核苷酸序列具有 95% 或更高同源性的多核苷酸序列组成的多核苷酸。

[0020] (10) 由与 (1)-(8) 中任一项的多核苷酸的多核苷酸序列具有 99% 或更高同源性的多核苷酸序列组成的多核苷酸。

[0021] (11) 由多核苷酸序列组成的多核苷酸，所述多核苷酸序列含有 (1)-(10) 中任一项的多核苷酸的多核苷酸序列的两种或多种序列。

[0022] (12) 由两种或多种类型的选自 (1)-(10) 中任一项的多核苷酸的多核苷酸组成的多核苷酸。

[0023] (13) 包含 (1)-(12) 中任一项的多核苷酸的多核苷酸序列的外源基因表达载体。

[0024] (14) (13) 的外源基因表达载体，其中由外源基因编码的蛋白是多聚体蛋白。

[0025] (15) (14) 的外源基因表达载体，其中外源基因编码的蛋白是异多聚体蛋白。

[0026] (16) (15) 的外源基因表达载体，其中外源基因编码的蛋白是抗体或其功能片段。

[0027] (17) 转化细胞，其中已引入 (13)-(16) 中任一项的外源基因表达载体。

[0028] (18) (17) 的转化细胞，其中所述细胞是源自哺乳动物的培养细胞。

[0029] (19) (18) 的转化细胞，其中所述源自哺乳动物的培养细胞为选自 COS-1 细胞、293 细胞和 CHO 细胞的细胞。

[0030] (20) (17)-(18) 中任一项的转化细胞，其中由外源基因编码的蛋白为多聚体蛋白。

[0031] (21) (20) 的转化细胞，其中所述外源基因编码的蛋白为异多聚体蛋白。

[0032] (22) (21) 的转化细胞，其中所述外源基因编码的蛋白为抗体或其功能片段。

[0033] (23) 生产蛋白质的方法，所述方法特征在于包括培养 (17)-(22) 任一项的转化细胞并从得到的培养产物中获得由外源基因编码的蛋白。

[0034] (24) 在已引入外源基因或外源基因表达载体的转化细胞中提高外源基因表达的方法，所述方法特征在于使用 (1)-(12) 中任一项的多核苷酸或 (13)-(16) 中任一项的外源基因表达载体。

[0035] (25) (1)-(12) 中任一项的多核苷酸用于在转化细胞中提高外源基因表达的用途。

[0036] 发明的有利作用

根据本发明，通过使用 DNA 元件将外源基因表达载体引入到哺乳动物宿主细胞，可显著提高治疗性蛋白、抗体等的外源基因的表达。

[0037] 附图简述

图 1 显示通过扩增 GADPH 区域来证实对其进行 ChIP-on-chip 的样品是用抗乙酰化组蛋白 H3 抗体经特异性染色质免疫沉淀的图表。

[0038] 图 2 为已插入 DNA 元件的 SEAP 表达载体的示意图。

[0039] 图 3 为显示在不含 DNA 元件或含 DNA 元件 A2、A7、A18、B5 或 C14 的稳定表达的 CHO 细胞系中，在 CMV 启动子控制下 SEAP 表达的图表。证实了 DNA 元件 A2、A7、A18、B5 和 C14 对提高表达的作用。

[0040] 图 4 包含两个图表，显示在不含 DNA 元件或含 DNA 元件 A2 或 A7 的稳定表达的 CHO 细胞系中，在 EF-1α 或 SV40 启动子控制下的 SEAP 表达。证实了 DNA 元件 A2 和 A7 对提高表达的作用。

[0041] 图 5 是已插入 DNA 元件的抗体表达（抗体基因 X 重链和轻链共表达）载体的示意图。

[0042] 图 6 包含两个图表，显示在不含 DNA 元件或含 DNA 元件 A7 的稳定表达的 CHO 细胞系中，在 CMV 或 EF-1α 启动子控制下的抗体分泌水平（通过 ELISA 法测量）。证实了 DNA 元件 A7 提高表达的作用。

[0043] 图 7 是显示 DNA 元件 A2 和相关序列的序列长度的表格。

[0044] 图 8 包含两个图表，显示在不含 DNA 元件或含 DNA 元件 A2 或相关序列的稳定表达的 CHO 细胞系中的 SEAP 表达。证实了 DNA 元件 A2 和相关序列提高表达的作用。

[0045] 图 9 是显示 DNA 元件 A7 和相关序列的序列长度的表格。

[0046] 图 10 包含三个图表，显示在不含 DNA 元件或含 DNA 元件 A7 或相关序列的稳定表达的 CHO 细胞系中的 SEAP 表达。证实了 DNA 元件 A7 和相关序列提高表达的作用。

[0047] 图 11 为显示 DNA 元件 A18 和相关序列的序列长度的表格。

[0048] 图 12 是显示在不含 DNA 元件或含 DNA 元件 A18 或相关序列的稳定表达的 CHO 细胞系中 SEAP 表达的图表。证实了 DNA 元件 A18 和相关序列提高表达的作用。

[0049] 图 13 是显示 DNA 元件 B5 和相关序列的序列长度的表格。

[0050] 图 14 是显示在不含 DNA 元件或含 DNA 元件 B5 或相关序列的稳定表达的 CHO 细胞系中 SEAP 表达的图表。证实了 DNA 元件 B5 和相关序列提高表达的作用。

[0051] 图 15 是显示 DNA 元件 C14 和相关序列的序列长度的表格。

[0052] 图 16 包含三个图表，显示在不含 DNA 元件或含 DNA 元件 C14 或相关序列的稳定表达的 CHO 细胞系中的 SEAP 表达。证实了 DNA 元件 C14 和相关序列提高表达的作用。

[0053] 图 17 是显示在不含 DNA 元件或含 DNA 元件 A2、A7、A18、B5 或 C14 的稳定表达的 HEK293 细胞系中 SEAP 表达的图表。证实了 DNA 元件 A2、A7、A18、B5 和 C14 在 HEK293 细胞中提高表达的作用。

[0054] 图 18 是显示基于 DNA 元件 A2、A7 或 A18 全长序列的起点和终点的核苷酸的视图。

[0055] 图 19 是显示基于 DNA 元件 B5 或 C14 全长序列的起点和终点的核苷酸的视图。

[0056] 实施方案说明

本发明在下文中将参考实施例进行具体描述。然而，这些实施例不限制本发明的技术范围。将在发明实施例中使用的质粒、限制酶、DNA 修饰酶等为市售产品，并可按照通用程序使用。另外，本领域技术人员亦熟知或可从文献中获得用于以下方面的程序：DNA 克隆、多核苷酸序列测定、宿主细胞的转化、经转化宿主细胞的培养、从获得的培养液中分离抗体、抗体的纯化等等。

[0057] 本文所用术语“基因”不仅包括 DNA，也包括其 mRNA、cDNA 及其 RNA。

[0058] 本文所用术语“多核苷酸”亦以与核酸相同的意思使用，并亦包括 DNA、RNA、探针、寡核苷酸和引物。

[0059] 本文所用术语“多肽”和“蛋白”无差别地使用。

[0060] 本文所用术语“基因表达”指从基因转录 mRNA 的现象和 / 或从 mRNA 翻译蛋白质的现象。

[0061] 本文所用术语“外源基因”指人工引入宿主细胞的基因。

[0062] 本文所用术语“外源蛋白”指由外源基因编码的蛋白。

[0063] 本文所用术语“基因表达单元”指在转录阅读框方向具有至少启动子区域、外源基因和转录终止子区域（多聚 (A) 添加信号）的多核苷酸。

[0064] 本文所用术语“提高外源基因表达的活性”，指通过创造对含有外源基因的基因表达单元周围的 DNA 的转录和翻译有利的环境，并显著提高转录和翻译效率，来提高宿主细胞中外源蛋白的产量。

[0065] 本文所用术语“DNA 元件”指具有提高外源基因表达活性的多核苷酸，其中所述多核苷酸位于基因表达单元附近或在含有基因表达单元的外源基因表达载体中。

[0066] 本文所用术语“抗体功能片段”，指具有抗原结合活性的部分抗体片段，其包括 Fab、$F(ab')_2$ 等。然而，所述术语不限于这些分子，只要片段对抗原具有结合亲和力即可。

[0067] 1. 用于提高外源基因表达的 DNA 元件

如实施例 1 所示，可通过利用乙酰化组蛋白 H3 和基因组 DNA 的相互作用来获得本发明的 DNA 元件。一般而言，认为组蛋白 (H3 和 H4) 的乙酰化与转录激活相关，并且一直在提倡两种主要理论。一种理论是组蛋白乙酰化与核小体构象改变有关，以此方式使得组蛋白尾部乙酰化，从而成为电中和的，导致 DNA 组蛋白相互作用减弱 (Mellor J. (2006) Dynamic nucleosomes and gene transcription（动态核小体和基因转录）. Trends Genet. 22(6): 320-329)。另一个理论是组蛋白乙酰化与各种转录因子的募集相关 (Nakatani Y. (2001) Histone acetylases - versatile players（组蛋白乙酰基转移酶 - 多面手）. Genes Cells. 6(2): 79-86)。在任何一个理论中，有高可能性的是，组蛋白乙酰化与转录激活相关，并且通过用抗乙酰化组蛋白 H3 抗体实施染色质免疫沉淀 (ChIP)，可能浓缩与乙酰化组蛋白 H3 相互作用的 DNA 元件。

[0068] 在本发明中，A2 是用于提高外源基因表达的 DNA 元件的实例。A2 位于人 15 号染色体的 80966429-80974878 区域，为具有 62.2% AT 含量的 8450 bp 的多核苷酸序列。A2 的多核苷酸序列由序列表中的 SEQ ID NO: 1 代表。

[0069] A7、A18、B5 和 C14 是相似的 DNA 元件的实例。A7 位于人 11 号染色体的 88992123-89000542 区域，为具有 64.52% AT 含量的 8420 bp 的多核苷酸序列。A7 的多核苷酸序列由序列表中的 SEQ ID NO: 2 代表。

[0070] A18 位于人 4 号染色体的 111275976-111284450 区域，为具有 62.54% AT 含量的 8475 bp 的多核苷酸序列。A18 的多核苷酸序列由序列表中的 SEQ ID NO: 3 代表。

[0071] B5 位于人 1 号染色体的 143034684-143043084 区域，为具有 66.37% AT 含量的 8401 bp 的多核苷酸序列。B5 的多核苷酸序列由序列表中的 SEQ ID NO: 4 代表。

[0072] 最后，C14 位于人 11 号染色体的 46089056-46097482 区域，为具有 63.81% AT 含量的 8427 bp 的多核苷酸序列。C14 的多核苷酸序列由序列表中的 SEQ ID NO: 5 代表。

[0073] 在本发明中，可通过使用由报告基因例如 SEAP 编码的蛋白活性作为指标来测定 DNA 元件提高外源基因表达的活性。与不存在 DNA 元件的情况相比，在存在 DNA 元件的情况中，当报告蛋白活性增加时（优选 2 倍或更高，更优选 4 倍或更高，再优选 5 倍或更高），可确定 DNA 元件具有提高外源基因表达的活性。甚至在活性增加 2 倍或更多的情况中，也预期这将会降低细胞培养规模和细胞培养时间，因此可能增加产量和降低细胞培养成本。如果产量增加，那么可能稳定提供外源蛋白用作药物。此外，如果细胞培养成本降低，则外源蛋白用作药物的成本降低，并且待给予外源蛋白的患者的财务负担也会降低。

[0074] 在本发明中，以上任何一种 DNA 元件可以单独使用，也可使用一种类型的 DNA 元件的 2 个或多个拷贝。或者，也可以联合使用 2 种或多种不同类型的上述 DNA 元件。

[0075] A2、A7、A18、B5 和 C14 为本发明使用的 DNA 元件的优选实例。

[0076] 将用于本发明的 DNA 元件可以是这样的多核苷酸序列，其包含与由 SEQ ID NO: 1-5 代表的任何一个多核苷酸序列具有 80% 或更高同源性的多核苷酸序列，并具有提高外源基因表达的活性。80% 或更高的同源性优选为 90% 或更高的同源性，更优选 95% 或更高的同源性，最优选 99% 或更高的同源性。可以用例如 FASTA 或 BLAST 程序在例如日本 DNA 数据库等中检索多核苷酸序列的同源性。

[0077] 将用于本发明的 DNA 元件可以是这样的 DNA 元件，其在严格条件下与这样的由多核苷酸序列组成的多核苷酸杂交，并具有提高外源基因表达的活性：所述多核苷酸与由选自 SEQ ID NO: 1-5 代表的多核苷酸序列的多核苷酸序列组成的多核苷酸互补。

[0078] 本文所用术语“严格条件”指其中形成所谓的特异杂交物 (hybrid) 但不形成非特异杂交物的条件。例如，例示性严格条件为这样的条件，在此条件下由具有高同源性的多核苷酸序列组成的核酸互补链杂交，包含具有较低同源性的多核苷酸序列的核酸的互补链不杂交，所述具有高同源性的多核苷酸序列即与选自 SEQ ID NO: 1-5 代表的多核苷酸序列的多核苷酸序列具有 80% 或更高、优选 90% 或更高、更优选 95% 或更高、最优选 99% 或更高同源性的多核苷酸序列。更具体的是，可例示的条件为这样的条件，在此条件下，钠盐浓度为 15-750 mM、优选 50-750 mM、更优选 300-750 mM，温度为 25-70℃、优选 50-70℃、更优选 55-65℃，甲酰胺浓度为 0-50%、优选 20-50%、更优选 35-45%。另外，作为严格条件，可例示在杂交后洗涤滤纸的条件，其中钠盐浓度通常为 15-600 mM、优选 50-600 mM、更优选 300-600 mM，温度为 50-70℃、优选 55-70℃、更优选 60-65℃。

[0079] 本领域技术人员可以参考分子克隆 (Molecular Cloning) (Sambrook, J. 等，Molecular Cloning: a Laboratory Manual 第二版，Cold Spring Harbor Laboratory Press, 10 Skyline Drive Plainview, NY (1989)) 等容易地获得这样的同源基因。另外，可以通过 FASTA 搜索或 BLAST 搜索以相同方式确定上述多核苷酸序列的同源性。

[0080] 可通过本技术领域已知方法引入突变（缺失、取代和 / 或添加）到上述多核苷酸

序列，所述方法例如 Kunkel 法或缺口双链法或基于该方法。例如，可以使用利用定点诱变法的突变导入试剂盒（例如，Mutant-K （由 TaKaRa Bio, Inc. 制造）、Mutant-G （由 TaKaRa Bio, Inc. 制造）或体外 LA PCR 诱变系列试剂盒（由 TaKaRa Bio, Inc. 制造））等。这样的突变多核苷酸也可作为本发明的 DNA 元件使用。

[0081] 可以使用包含由序列表中 SEQ ID NO: 1-5 任一个所代表的多核苷酸序列的至少 3000 或至少 2000 个连续核苷酸的部分片段作为本发明 DNA 元件。所述部分片段的实例包括 :A2 的部分片段 A2-1 到 A2-17，A7 的部分片段 A7-1 到 A7-18，A18 的部分片段 A18-1 到 A18-4，B5 的部分片段 B5-1 到 B5-6，以及 C14 的部分片段 C14-1 到 C14-14。然而，DNA 元件不限于这些部分片段，只要其具有提高外源基因表达的活性即可。

[0082] 在本发明中，可以单独使用上述任何一个部分片段，也可以使用一种类型部分片段的 2 个或多个拷贝。或者，可联合使用 2 种或多种不同类型的部分片段。另外，可以联合使用上述任何 DNA 元件的全长序列和部分片段。在以上联合中，全长序列和部分片段可以源自相同 DNA 元件或源自不同 DNA 元件。

[0083] 关于 A2 的各个片段的多核苷酸序列 :A2-1 对应于序列表中 SEQ ID NO: 1 的 1-3000 位核苷酸的多核苷酸序列 ;A2-2 对应于序列表中 SEQ ID NO: 1 的 2801-5800 位核苷酸的多核苷酸序列 ;A2-3 对应于序列表中 SEQ ID NO: 1 的 5401-8450 位核苷酸的多核苷酸序列 ;A2-4 对应于序列表中 SEQ ID NO: 1 的 701-2700 位核苷酸的多核苷酸序列 ;A2-5 对应于序列表中 SEQ ID NO: 1 的 701-2200 位核苷酸的多核苷酸序列 ;A2-6 对应于序列表中 SEQ ID NO: 1 的 701-3700 位核苷酸的多核苷酸序列 ;A2-7 对应于序列表中 SEQ ID NO: 1 的 2001-5000 位核苷酸的多核苷酸序列 ;A2-8 对应于序列表中 SEQ ID NO: 1 的 4001-7000 位核苷酸的多核苷酸序列 ;A2-9 对应于序列表中 SEQ ID NO: 1 的 1-3700 位核苷酸的多核苷酸序列 ;A2-10 对应于序列表中 SEQ ID NO: 1 的 2001-5800 位核苷酸的多核苷酸序列 ;A2-11 对应于序列表中 SEQ ID NO: 1 的 2801-7000 位核苷酸的多核苷酸序列 ;A2-12 对应于序列表中 SEQ ID NO: 1 的 701-5800 位核苷酸位核苷酸的多核苷酸序列 ;A2-13 对应于序列表中 SEQ ID NO: 1 的 2001-7000 位核苷酸的多核苷酸序列 ;A2-14 对应于序列表中 SEQ ID NO: 1 的 2801-8450 位核苷酸的多核苷酸序列 ;A2-15 对应于序列表中 SEQ ID NO: 1 的 1-5800 位核苷酸的多核苷酸序列 ;A2-16 对应于序列表中 SEQ ID NO: 1 的 701-7000 位核苷酸的多核苷酸序列 ;和 A2-17 对应于序列表中 SEQ ID NO: 1 的 2001-8450 位核苷酸的多核苷酸序列。

[0084] 关于 A7 的各个片段的多核苷酸序列 :A7-1 对应于序列表中 SEQ ID NO: 2 的 601-3600 位核苷酸的多核苷酸序列 ;A7-2 对应于序列表中 SEQ ID NO: 2 的 3601-8420 位核苷酸的多核苷酸序列 ;A7-3 对应于序列表中 SEQ ID NO: 2 的 5401-8420 位核苷酸的多核苷酸序列 ;A7-4 对应于序列表中 SEQ ID NO: 2 的 3401-6400 位核苷酸的多核苷酸序列 ;A7-5 对应于序列表中 SEQ ID NO: 2 的 1501-4500 位核苷酸的多核苷酸序列 ;A7-6 对应于序列表中 SEQ ID NO: 2 的 4401-7400 位核苷酸的多核苷酸序列 ;A7-7 对应于序列表中 SEQ ID NO: 2 的 2401-5400 位核苷酸的多核苷酸序列 ;A7-8 对应于序列表中 SEQ ID NO: 2 的 1-3600 位核苷酸的多核苷酸序列 ;A7-9 对应于序列表中 SEQ ID NO: 2 的 1501-5400 位核苷酸的多核苷酸序列 ;A7-10 对应于序列表中 SEQ ID NO: 2 的 2401-6400 位核苷酸的多核苷酸序列 ;A7-11 对应于序列表中 SEQ ID NO: 2 的 3401-7400 位核苷酸的多核苷酸序列 ;

A7-12 对应于序列表中 SEQ ID NO: 2 的 4401-8420 位核苷酸的多核苷酸序列 ;A7-13 对应于序列表中 SEQ ID NO: 2 的 1-5400 位核苷酸的多核苷酸序列 ;A7-14 对应于序列表中 SEQ ID NO: 2 的 1501-6400 位核苷酸的多核苷酸序列 ;A7-15 对应于序列表中 SEQ ID NO: 2 的 2401-7400 位核苷酸的多核苷酸序列 ;A7-16 对应于序列表中 SEQ ID NO: 2 的 3401-8420 位核苷酸的多核苷酸序列 ;A7-17 对应于序列表中 SEQ ID NO: 2 的 1-6400 位核苷酸的多核苷酸序列,和 A7-18 对应于序列表中 SEQ ID NO: 2 的 1501-7400 位核苷酸的多核苷酸序列。

[0085] 关于 A18 的各个片段的多核苷酸序列 :A18-1 对应于序列表中 SEQ ID NO: 3 的 1-5040 位核苷酸的多核苷酸序列 ;A18-2 对应于序列表中 SEQ ID NO: 3 的 1001-6002 位核苷酸的多核苷酸序列 ;A18-3 对应于序列表中 SEQ ID NO: 3 的 2001-7000 位核苷酸的多核苷酸序列 ;A18-4 对应于序列表中 SEQ ID NO: 3 的 3000-7000 位核苷酸的多核苷酸序列。

[0086] A2、A7 和 A18 的各个片段的起点和终点示于图 18 中。

[0087] 关于 B5 的各个片段的多核苷酸序列 :B5-1 对应于序列表中 SEQ ID NO: 4 的 1-4001 位核苷酸的多核苷酸序列 ;B5-2 对应于序列表中 SEQ ID NO: 4 的 1-3200 位核苷酸的多核苷酸序列 ;B5-3 对应于序列表中 SEQ ID NO: 4 的 2491-5601 位核苷酸的多核苷酸序列 ;B5-4 对应于序列表中 SEQ ID NO: 4 的 5373-8401 位核苷酸的多核苷酸序列 ;B5-5 对应于序列表中 SEQ ID NO: 4 的 901-4001 位核苷酸的多核苷酸序列 ;和 B5-6 对应于序列表中 SEQ ID NO: 4 的 4001-7000 位核苷酸的多核苷酸序列。

[0088] 关于 C14 的各个片段的多核苷酸序列 :C14-1 对应于序列表中 SEQ ID NO: 5 的 960-4015 位核苷酸的多核苷酸序列 ;C14-2 对应于序列表中 SEQ ID NO: 5 的 1987-5014 位核苷酸的多核苷酸序列 ;C14-3 对应于序列表中 SEQ ID NO: 5 的 4020-7119 位核苷酸的多核苷酸序列 ;C14-4 对应于序列表中 SEQ ID NO: 5 的 960-8141 位核苷酸的多核苷酸序列 ;C14-5 对应于序列表中 SEQ ID NO: 5 的 960-6011 位核苷酸的多核苷酸序列 ;C14-6 对应于序列表中 SEQ ID NO: 5 的 4939-8141 位核苷酸的多核苷酸序列 ;C14-7 对应于序列表中 SEQ ID NO: 5 的 960-5014 位核苷酸的多核苷酸序列 ;C14-8 对应于序列表中 SEQ ID NO: 5 的 2994-7119 位核苷酸的多核苷酸序列 ;C14-9 对应于序列表中 SEQ ID NO: 5 的 4020-8141 位核苷酸的多核苷酸序列 ;C14-10 对应于序列表中 SEQ ID NO: 5 的 1-5014 位核苷酸的多核苷酸序列 ;C14-11 对应于序列表中 SEQ ID NO: 5 的 1987-7119 位核苷酸的多核苷酸序列 ;C14-12 对应于序列表中 SEQ ID NO: 5 的 2994-8141 位核苷酸的多核苷酸序列 ;C14-13 对应于序列表中 SEQ ID NO: 5 的 960-7119 位核苷酸的多核苷酸序列 ;C14-14 对应于序列表中 SEQ ID NO: 5 的 1987-8141 位核苷酸的多核苷酸序列。

[0089] B5 和 C14 各个片段的起点和终点亦示于图 19 中。

[0090] 2. 多核苷酸的获取

在本发明中,可通过如下所述通用程序获得含有外源基因的多核苷酸,所述外源基因编码其产量将会增加的外源蛋白,这一点在后面阐述。例如,可使用基于外源基因片段合成的 DNA 探针,通过筛选源自表达外源基因的细胞或组织的 cDNA 文库,来分离所述多核苷酸。mRNA 可以通过本技术领域常用的方法来制备。例如,用胍试剂、酚试剂等处理细胞或组织,从而获得总 RNA,此后通过使用寡聚 (dT) 纤维素柱或含有 Sepharose 2B 作为载体的聚 U-Sepharose 柱等等的亲和柱法,或通过分批处理法,获得多聚 (A)+RNA (mRNA)。同样,

多聚 (A)+RNA 可以通过蔗糖密度梯度离心法等进一步分级分离。然后，用由此获得的 mRNA 作为模板，同时亦使用寡聚 dT 引物和逆转录酶，合成单链 cDNA。从由此获得的单链 cDNA，用 DNA 聚合酶 I、DNA 连接酶、RNA 酶 H 等合成双链 cDNA。用 T4DNA 聚合酶使由此合成的双链 cDNA 具有平端，接着与衔接子（例如 EcoI 衔接子）连接，进行磷酸化等，将所得 DNA 掺入到诸如 λgt11 等 λ 噬菌体中以实现体内包装，藉此可制备 cDNA 文库。也可使用 λ 噬菌体以外的质粒载体来制备 cDNA 文库。此后，可自 cDNA 文库选择含有靶 DNA 的克隆（阳性克隆）。

[0091] 在待用于增加蛋白产量或增加含有外源基因的多核苷酸产量的上述 DNA 元件分离自基因组 DNA 情况下，或在含有启动子和终止子区域的多核苷酸分离自基因组 DNA 的情况下，参照通用程序 (Molecular Cloning (1989), Methods in Enzymology 194 (1991) 自待用作收集源的生物细胞系提取基因组 DNA，选择并分离多核苷酸。可以参照例如 Cryer 等的方法 (Methods in Cell Biology, 12, 39-44 (1975)) 或 P. Philippsen 等的方法 (Methods Enzymol., 194, 169-182 (1991)) 进行基因组 DNA 的提取。

[0092] 也可以通过例如 PCR 法 (PCR Technology. Henry A. Erlich, Atockton press (1989)) 获得靶 DNA 元件或含有外源基因的多核苷酸。在用 PCR 法扩增多核苷酸时，使用 20- 到 30- 聚体的合成的单链 DNA 作为引物，并用基因组 DNA 作为模板。在确认基因的多核苷酸序列后使用扩增的基因。可以使用基因组 DNA 文库例如细菌人工染色体 (BAC) 作为 PCR 模板。

[0093] 在另一方面，可以通过以下方法获得含有其序列未知的外源基因的多核苷酸：(a) 按照通用程序制备基因文库；和 (b) 从制备好的基因文库中选择所期需多核苷酸并扩增所述多核苷酸。基因文库可如下制备：通过用合适的限制酶将染色体 DNA 部分消化，以使染色体 DNA 片段化，让所获得的片段与合适的载体连接，并将载体引入到合适的宿主，所述染色体 DNA 通过通用程序自用作采集源的生物细胞系获得。基因文库亦可如下制备：从细胞提取 mRNA，从 mRNA 合成 cDNA，让 cDNA 与合适的载体连接，将载体引入到合适的宿主。亦可使用用于基因文库制备的通常称作载体的质粒、噬菌体载体、粘粒等，作为在所述制备中使用的载体。可使用适于上述载体类型的宿主，作为待转化或转染的宿主。用含有对外源基因特异的序列的标记探针，通过菌落杂交法、噬菌斑杂交法等，从上述基因文库选择含有外源基因的多核苷酸。

[0094] 另外，也可以通过完全化学合成生产含有外源基因的多核苷酸。例如，可通过以下方法来合成基因：制备两对互补寡聚核苷酸并退火的方法；由 DNA 连接酶连接几种退火的 DNA 链的方法；制备几种部分配对的多核苷酸并且由 PCR 填充空隙的方法；等等。

[0095] 可通过常规技术例如双脱氧法 (Sanger 等, Proc. Natl. Acad. Sci., USA, 74, 5463-5467, (1977)) 等进行多核苷酸序列的测定。另外，也可以用市购序列测定试剂盒等容易地进行上述多核苷酸序列的测定。

[0096] 3. 外源基因表达载体，元件载体

提供以下载体作为本发明的外源基因表达载体：含有一种类型的上述 DNA 元件，一种类型的上述 DNA 元件的两个或多个拷贝，或两个或多个不同类型的上述 DNA 元件的组合，并进一步含有外源基因表达单元的载体。当用上述外源基因表达载体在宿主细胞中表达外源基因时，DNA 元件可紧邻基因表达单元的上游或下游，或可位于远离基因表达单元的位置。

另外，可以使用一种含有多个所述 DNA 元件的外源基因表达载体。附带说一下，DNA 元件可以以相对于基因表达单元的正向或反向插入。

[0097] 另外，亦包括以下载体作为待用于本发明的载体：含有一种类型的上述 DNA 元件，一种类型的上述 DNA 元件的两个或多个拷贝，或两种或多种不同类型的上述 DNA 元件组合，并且不含有基因表达单元（下文中也称作“元件载体”）的载体。所述元件载体可以与上述含有 DNA 元件的外源基因表达载体或不含 DNA 元件而仅包含外源基因表达单元的外源基因表达载体组合使用。与单独使用外源基因表达载体的情形相比，通过让元件载体与外源基因表达载体共存，外源基因的表达增强，因此，上述载体组合也包括在本发明的外源基因表达载体中。

[0098] 编码外源蛋白的基因不受特别限制，然而，其实例包括：报告基因，例如分泌性碱性磷酸酶 (SEAP)、绿色荧光蛋白 (GFP) 和荧光素酶；各种酶基因，例如 α - 淀粉酶基因和 α - 半乳糖苷酶基因；各种干扰素基因，其为药学上有用并有生理活性的蛋白，例如 α 干扰素和 γ 干扰素基因；各种白细胞介素例如 IL-1 和 IL-2 基因；各种细胞因子基因，例如红细胞生成素 (EPO) 基因和粒细胞集落刺激因子 (G-CSF) 基因；生长因子基因；和抗体基因。可以通过任何方法获得这些基因。

[0099] 本发明在有关高度疏水蛋白和由于复合构成而难于分泌和生产的蛋白上特别有效。因此，上述外源蛋白包括多聚体蛋白，例如作为抗体或其功能片段的异多聚体。“抗体功能片段”指具有抗原结合活性的抗体的部分片段，包括 Fab、$F(ab')_2$、Fv、scFv、双抗体、线性抗体、自抗体片段形成的多特异性抗体等。抗体的功能片段还包括 Fab'，其为通过在还原条件下处理 $F(ab')_2$ 获得的抗体可变区中的单价片段。然而，这些功能片段不限于这些分子，只要所述片段对抗原具有结合亲和力即可。另外，这些功能片段不仅包括通过用合适的酶处理抗体蛋白全长分子获得的片段，还包括在合适的宿主细胞中使用遗传修饰的抗体基因生产的蛋白。

[0100] 基因表达单元在转录阅读框方向具有至少启动子区域、外源基因和转录终止子区域（多聚 (A) 添加信号）。本文可使用的启动子可以是组成型表达启动子或诱导型表达启动子。组成型表达启动子的实例包括各种天然启动子，例如 SV40 早期启动子、腺病毒 E1A 启动子、CMV （巨细胞病毒）启动子、EF-1α （人延伸因子 -1α）启动子、HSP70 启动子、MT 启动子、RSV 启动子、UBC 启动子和肌动蛋白启动子；和人工（融合）启动子，例如 SRα 启动子和 CAG 启动子。另外，多聚 (A) 添加序列可以是具有引起从启动子转录的转录终止的活性的序列，可以是与启动子相同或不同的基因的序列。

[0101] 需要使用强启动子来提高外源蛋白的生产。然而，当试图使用高活性启动子生产难以折叠的蛋白或难以分泌的蛋白时，所述蛋白反而可能不能分泌。这是因为当以超过进行翻译的核糖体和进行折叠和分泌的内质网容量的量生产时，过量生产的蛋白在细胞内变性、积累和泛素化，然后由蛋白体降解。因此，优选适当地选择这样的启动子，其可以达到这样程度的表达水平：得到的蛋白不变性或聚集，或得到的蛋白量不超过分泌容量。或者，通过调节（例如降低）启动子活性来使用启动子。在多聚体蛋白中，形成异多聚体的分子易受上述效应的影响，特别是诸如为异四聚体的抗体等分子。抗体具有相互缔合的两条重链分子和两条轻链分子，因此，为了恰当地缔合所述分子，其表达水平是重要因素。

[0102] 另外，本发明的外源基因表达载体和元件载体可以各自含有选择标记用于选

择转化体。通过使用例如赋予对药物抗性的抗药标记，可以选择转化体，所述药物例如浅蓝菌素 (cerulenin)、金担子素 (aureobasidin)、Zeocin、刀豆氨酸 (canavanine)、放线菌酮 (cycloheximide)、潮霉素 (hygromycin)、嘌呤霉素 (puromycin)、杀稻瘟素 (blasticidin)、四环素 (tetracycline)、卡那霉素 (kanamycin)、氨苄西林 (ampicillin) 或新霉素 (neomycin)。另外，当使用以下基因作为标记时，亦可选择转化子：赋予对例如乙醇等溶剂抗性、对甘油、盐等的渗透压抗性、对诸如铜离子等金属离子抗性的基因等等。

[0103] 本发明的外源基因表达载体和元件载体各自可以是不掺入到染色体 DNA 的载体。一般而言，将外源基因表达载体转染到宿主细胞，此后随机掺入到染色体中。然而，通过使用源于哺乳动物病毒例如猿猴病毒 40(SV40)、乳头瘤病毒 (BPV、HPV) 或 EBV 的组成型成分，载体可作为在转染宿主细胞中自主复制的游离型载体使用。例如，可以使用含有 SV-40 来源的复制起点 (oriP) 及编码为反式作用因子的 SV40 大 T 抗原的序列的载体、含有 EBV 来源的 oriP 和编码 EBNA-1 的序列的载体等等。可以通过提高外源基因表达的活性来表示 DNA 元件的作用，而与载体类型或其到染色体中的掺入存在或不存在情况无关。

[0104] 4. 转化细胞

本发明的转化细胞是已经引入含有上文条目“1”所述的 DNA 元件的上文条目“3”所述的外源基因表达载体的转化细胞。作为外源基因表达载体，可以引入仅含有 DNA 元件的外源基因表达载体 (A)，或可以联合引入含有 DNA 元件的外源基因表达载体以及在上文条目“3”所述的元件载体 (B)。或者，可以联合引入不含有 DNA 元件的外源基因表达载体和元件载体 (C)。

[0105] 可以按照例如以下方法来实施在使用以上 (B) 或 (C) 组合的宿主细胞中的外源基因的表达：Girod 等的方法 (Biotechnology and Bioengineering, 91, 2-11 (2005)) 和 Otte 等的方法 (Biotechnol. Prog., 2007, 23, 801-807 (2007))。

[0106] 待转化的宿主细胞的实例包括真核细胞，其优选实例包括哺乳动物细胞，更优选的实例包含源自人、小鼠、大鼠、仓鼠、猴或牛的细胞。所述哺乳动物细胞实例包括 COS-1 细胞、293 细胞和 CHO 细胞 (CHO-K1、DG44、CHO dhfr-、CHO-S)，然而，宿主细胞不限于这些。

[0107] 在本发明中，可以使用任何方法来将表达载体引入到宿主细胞中，只要所述方法允许引入的基因稳定存在于宿主细胞并且在其中适当地表达。通常使用的方法实例包括磷酸钙法 (Ito 等，c(1984) Agric. Biol. Chem., 48, 341)、电穿孔法 (Becker, D. M. 等，1990; Methods. Enzymol., 194, 182-187)、原生质球法 (Creggh 等，Mol. Cell. Biol., 5, 3376 (1985))、乙酸锂法 (Itoh, H. (1983) J. Bacteriol. 153, 163-168) 和脂转染法。

[0108] 5. 生产外源蛋白的方法

在本发明中，可通过以下方法来生产外源蛋白：培养上文条目“4”所述的转化细胞，其中已经通过已知方法使用上文条目“3”所述的载体引入编码外源蛋白的基因；从所得培养产物中收集蛋白质，接着纯化蛋白。本文所用术语“培养产物”指除培养物上清液外还有培养细胞或细胞匀浆。附带说一下，不仅可以选择单体蛋白，还可以选择多聚体蛋白作为可以使用上文条目“4”所述的转化细胞生产的外源蛋白。在其中生产由多个不同亚基形成的异多聚体蛋白的情况中，必需分别将多个编码这些亚基的基因引入到上文条目“4”所述的宿主细胞中。

[0109] 可以按照培养宿主细胞的常规方法来实施培养转化细胞的方法。

[0110] 在其中转化细胞为哺乳动物细胞的情况中，细胞在例如 37℃和 5% 或 8% CO_2 培养时间约 24-1000 小时的条件下培养。可以通过分批培养、补料分批培养、连续培养等在静置、振摇、搅拌或通气条件下进行培养。

[0111] 可以通过 SDS-PSGE、Western 分析、ELISA 等从上述培养产物（培养液）中进行编码外源蛋白的基因表达产物的确认。为了分离和纯化生产的蛋白，可以使用常规蛋白分离和纯化方法。在完成培养后，在细胞中生产靶蛋白的情况下，使用超声匀浆器、弗氏压碎器、Manton-Gaulin 匀浆器、Dinomil 等使细胞匀浆化，由此获得靶蛋白。另外，在细胞外生产靶蛋白的情况下，使用培养液本身，或通过离心等方法去掉细胞。此后，用有机溶剂萃取等收集靶蛋白，然后可以通过使用以下技术来分离和纯化收集的靶蛋白：例如各种色谱技术（疏水色谱法、反相色谱法、亲和色谱法、离子交换色谱法等）、使用分子筛的凝胶过滤和使用聚丙烯酰胺凝胶电泳等等，它们可按照需要单独或组合使用。

[0112] 上述培养方法和纯化方法仅仅是实例，所述方法不限于此。可以通过已知的氨基酸分析技术例如使用 Edman 降解法的自动氨基酸序列测定，来确认纯化的基因产物的氨基酸序列。

[0113] 6. 生产抗体蛋白的方法

作为将使用上文条目“5”所述的生产方法生产的异多聚体蛋白，可例示抗体蛋白。抗体蛋白是包含两个重链多肽分子和两个轻链多肽分子的四聚体蛋白。因此，为了获得保持抗原结合亲和力状态下的所述抗体蛋白，必需将重链和轻链基因两者引入到上文条目“4”所述的转化细胞。在这种情况下，重链和轻链基因表达单元可以存在于同一个表达载体或不同表达载体上。

[0114] 作为将在发明中生产的抗体，可例示通过用所期需抗原免疫诸如兔、小鼠或大鼠等实验动物而制备的抗体。另外，亦可例示通过使用上述抗体作为起始材料获得的嵌合抗体和人源化抗体作为本发明中将生产的抗体。另外，使用遗传修饰动物或噬菌体展示方法获得的人抗体也包括在本发明将生产的抗体中。

[0115] 将用于抗体生产的抗体基因不限于具有特异性多核苷酸序列的抗体基因，只要待从抗体基因转录和翻译的重链多肽和轻链多肽的组合具有结合给定抗原蛋白的活性即可。

[0116] 另外，不必要的是，抗体基因编码抗体全长分子，而可以使用编码抗体功能片段的基因。所述编码其功能片段的基因可以通过遗传修饰编码抗体蛋白全长分子的基因获得。

[0117] 7. 其它外源蛋白的生产方法

除上述抗体外，将使用本发明生产方法生产的外源蛋白的实例还包括：各种源于人或非人的蛋白、其功能片段及其修饰产物。所述蛋白等等的实例包括肽激素，例如心房钠尿肽 (ANP)、脑钠素 (BNP)、C 型利钠肽 (CNP)、加压素、生长抑素、生长激素 (GH)、胰岛素、催产素、生长素释放肽、瘦蛋白、脂连蛋白、肾素、降钙素、骨保护素和胰岛素样生长因子 (IGF)；细胞因子类，例如白介素、趋化因子、干扰素、肿瘤坏死因子（例如 TNF-α、TNF-β 和 TNF 超家族）、神经生长因子类（例如 NGF)、细胞生长因子类（例如 EGF、FGF、PDGF、HGF 和 TGF)、造血生长因子类（例如 CSF、G-CSF 和红细胞生成素）和脂肪因子 (adipokine)；受体，例如 TNF 受体；酶，例如溶菌酶、蛋白酶、蛋白水解酶和肽酶；其功能片段（具有原始蛋白部分或全部生物学功能的片段）；和包含这些蛋白中任何一种的融合蛋白。然而，所述蛋白不限于

此。

实施例

[0118] 本发明在下文中将参考实施例进行具体描述。然而，这些实施例不限制本发明的技术范围。本发明实施例中待使用的质粒、限制性内切酶、DNA 修饰酶等等为市购产品，并且可以按照通用程序使用。另外，用于 DNA 克隆、多核苷酸序列测定、宿主细胞转化、经转化宿主细胞的培养、从得到的培养产物中收集蛋白质、蛋白质纯化等等的程序亦为本领域技术人员所熟知，或可参见文献。

[0119] 实施例 1

提取 DNA 元件

(1-1) 使用抗 - 乙酰化组蛋白 H3 抗体的染色质免疫沉淀

用 EZ ChIP (Upstate) 按照以下程序实施使用抗 - 乙酰化组蛋白抗体的 ChIP。顺便说一下，除非另外指出，否则使用 Upstate 产品作为在以下程序中使用的抗体、缓冲液等等。

[0120] 首先，使用 GIBCO (注册商标)FreeStyleTM 293 培养基 (Invitrogen) 在 37℃和 8% CO_2 条件下培养 293F 细胞 (Invitrogen)，接着离心 (1000 rpm, 5 分钟，室温)，藉此收集生长期细胞。在含有 1% 甲醛的培养基中固定 2 x 10^7 细胞 10 分钟后，将 10x 甘氨酸加入其中，接着室温孵育 5 分钟。离心 (3000 rpm, 5 分钟，4℃) 后，移除上清，加入 PBS 至细胞沉淀中以悬浮细胞。然后，再次离心细胞悬液以移除 PBS，此后将 SDS 裂解缓冲液加入至细胞沉淀中以悬浮和裂解细胞。在超声匀浆器 (BRANSON) 中对每个通过细胞裂解获得的样品进行 DNA 片段化，同时用冰水冷却样品，向其中加入含有蛋白酶抑制剂混合物和蛋白 G- 固定的琼脂糖的稀释缓冲液。在 4℃让得到的混合物旋转 1 小时，接着离心，然后收集上清液。随后，向其中加入 10 μg 标准兔 IgG 或 α - 乙酰组蛋白 H3 抗体，接着 4℃旋转过夜。向得到的溶液中加入 G 蛋白 - 固定的琼脂糖，并且在 4℃旋转所得混合物 1 小时，接着离心，然后收集沉淀。用低盐免疫复合洗涤缓冲液 (Low Salt Immune Complex Wash Buffer) 对由此获得的沉淀洗涤两次，用高盐免疫复合洗涤缓冲液洗涤两次，用 LiCl 免疫复合洗涤缓冲液洗涤两次，最后用 TE 缓冲液洗涤 4 次。然后向其中加入洗脱缓冲液 (含有 20 μl 1 M 碳酸氢钠、10 μl SDS 和 170 μl 无菌水)。30 分钟后，离心混合物，收集上清液。

[0121] 随后，将 5 M 氯化钠加入上清液中，在 65℃将所得混合物加热过夜。然后向其中加入 RNA 酶 A，37℃孵育所得混合物 30 分钟。然后向其中加入 0.5 M EDTA、1 M Tris-HCl 和蛋白酶 K，在 45℃让所得混合物孵育 2 小时。

[0122] 最后，向其中加入通过用蛋白酶 K 处理获得的溶液的 5 倍量的试剂 A、B 和 C，接着用旋转过滤器 (Spin filter) 离心 (10000 rpm，30 秒，室温)，藉此纯化染色质 - 免疫沉淀的 DNA。

[0123] (1-2) 微阵列分析

通过使用 GenomePlex 全基因组扩增 (WGA) 试剂盒 (Sigma)，扩增在 (1-1) 中获得的每个 ChIP 样品。程序按照试剂盒附带的 Sigma 方案进行。

[0124] 为了证实 ChIP，通过使用 320 ng 由 WGA 扩增的每一种 DNA 作为模板，亦使用以下引物和 SYBR (注册商标) 预混 Ex Taq™ (Perfect Real Time) (TAKARA)，通过 PCR 方法扩增甘油醛 -3- 磷酸脱氢酶 (GAPDH) 内部基因 (95℃ 5 秒和 60℃ 20 秒 x 45 个循环)。附

带说一下，GAPDH 是待用作确定通过 ChIP 是否富集 DNA 元件的阳性对照的管家基因，并且用连接到 EZ ChIP (Upstate) 的引物来实施 PCR 方法。

5′-TACTAGCGGTTTTACGGGCG-3′

5′-TCGAACAGGAGGAGCAGAGAGCGA-3′

[0125] 如图 1 所示，确定了 GAPDH 在用抗 - 乙酰化组蛋白 H3 抗体进行免疫沉淀的样品中特异性扩增。对每个通过 WGA 扩增的 DNA 样品进行微列阵分析 (NimbleGen) 以实施染色质免疫沉淀芯片 (ChIP-on-chip)。“ChIP-on-chip”是用于通过对在 (1-1) 中富集的每个 DNA 进行微阵列分析来鉴别每个 DNA 元件的技术。

[0126] (1-3) 提取 DNA 元件

基于在 (1-2) 中获得的 ChIP-on-chip 分析的结果，提取 5 个具有 62% 或更多 AT 含量的序列。

[0127] A2 :染色体 15 (80966429-80974878)

A7 :染色体 11 (88992123-89000542)

A18 :染色体 4 (111275976-111284450)

B5 :染色体 1 (143034684-143043084)

C14 :染色体 11 (46089056-46097482)

实施例 2

用分泌性碱性磷酸酶 (SEAP) 表达作为 DNA 元件表达效果的指标

(2-1) 构建 SEAP 表达载体

通过使用 pSEAP2- 对照 (Clontech) 作为模板，由 PCR 方法 (94℃ 30 秒和 68℃ 2 分钟 x 40 个循环) 使用以下引物和 KOD-plus- (TOYOBO) 来扩增 SEAP 基因。

5′-AAAGCTAGCATGCTGCTGCTGCTGCTGCTGGGCC-3′

5′-AAAAGATCTTCATGTCTGCTCGAAGCGGCCGGCCGC-3′

[0128] 随后通过琼脂糖凝胶电泳分离扩增的 SEAP 片段并从凝胶切割，接着用 QIAquick 凝胶提取试剂盒 (Qiagen) 纯化。将由此获得的 DNA 片段用作插入片段。用限制性内切酶 NheI 和 BglII 消化该插入片段，并且用限制性内切酶 NheI 和 BamHI 消化载体 pIRES hyg3 (Clontech)。分别对得到的 DNA 片段进行琼脂糖凝胶电泳来分离靶片段，并且从凝胶切割靶片段，接着纯化。然后，进行连接反应和转化。用 LigaFast 快速 DNA 连接体系 (Promega) 实施连接反应。转化如下进行。首先，解冻冷冻的感受态细胞 JM109 (TAKARA)，并且将连接反应后获得的 10 μl 溶液加入至解冻的细胞溶液，让得到的混合物置于冰上 30 分钟。此后，向混合物施用热激 (42℃，45 秒)，让混合物在冰上冷却 5 分钟。向该细胞悬液中加入 1 ml LB 培养基，让得到的混合物在 37℃振摇 1 小时。然后，将混合物铺板在含有 0.1 mg/ml 氨苄青霉素的 LB 平板上，让该平板在 37℃孵育 14-16 小时。此后，通过碱裂解，从 LB 平板培养的菌落收集靶质粒。最后，测定通过碱裂解获得的质粒中的 SEAP 的多核苷酸序列，藉此构建 pCMV/SEAP ires Hygro。

[0129] (2-2) 克隆 DNA 元件

随后，用来自含有对应于各 DNA 元件的多核苷酸序列的细菌人造染色体 (BAC) 的 BAC 亚克隆试剂盒 (Gene Bridges)，将在实施例 1 中提取的各 DNA 元件克隆至 (2-1) 中获得的

SEAP 表达载体。

[0130] 首先，用限制性内切酶 SpeI 消化 (2-1) 中获得的 pCMV/SEAP ires Hygro 数小时，接着乙醇沉淀，用无菌水溶解沉淀。通过使用经 SpeI 消化的载体作为模板，使用以下引物和 KOD-plus- (TOYOBO) 实施 PCR 方法 (94℃ 15 秒、55℃ 30 秒和 68℃ 10 分钟 x 30 个循环)。

A2D:

5′-GGAAATTGAGAAGTATCATTCACAACAGTACCACAAACATGAAATAAATGTGGAT

CCTATTAATAGTAATCAATTACG-3′

A2R:

5′-CTCATTCTGTGGGTTGTCATTTCACTTCCTTGATGCTATCCTTTCAAGCAAAATC

CTAGTCAATAATCAATGTCAACG-3′

A7D:

5′-CTTATTTTCTAAGTAGTATAGACTTAATTGTGAGAACAAAATAAAAACTTGGATC

CTATTAATAGTAATCAATTACG-3′

A7R:

5′-CTCTTCCCATTCTCATTTGAATCTACTTCAAAAGGTTTACCATACTAAGACCTAG

TCAATAATCAATGTCAACG-3′

A18D:

5′-CGCCTGTAATCCCAGCACTTTGGGAGGCTGAGGCGGGTGGATCACCTGAGGTCGA

TCCTATTAATAGTAATCAATTACG-3′

A18R:

5′-CATACAGAAGCCAGTTTGAACTGAGACCTCACTCCATTTCTTACAAGTTATGCCC

TAGTCAATAATCAATGTCAACG-3′

B5D:

5′-ACCGTTTTATATTGTTTAAGCATTTCCTAGACATATTTGGCTACAAATCTAGATC

CTATTAATAGTAATCAATTACG-3′

B5R:

5′-GATCTTAGGGGGGCTGATTATATAAAACAATAGAAATGTAGTCTTAGATGAAACC

TAGTCAATAATCAATGTCAACG-3′

C14D:

5′-CACAAAGTTCACTGTCAAGGCCAGGTGATGAGGCCCACACATGCCCGGACCTTGA

TCCTATTAATAGTAATCAATTACG-3′

C14R:

5′-CAAAACCTCATCTCTACTGAAAATAGAAAATTAGCTGGGCGTGGTGGCAGGTGCC

CTAGTCAATAATCAATGTCAACG-3′

[0131] 该扩增用反应溶液的一部分通过琼脂糖凝胶电泳确认后，对其余的反应溶液进行乙醇沉淀。用无菌水溶解沉淀，将得到的溶液用作用于转化的 DNA。

[0132] 随后，制备用于转化的大肠杆菌（*Escherichia coli*）。

[0133] 对应于在实施例 1 中提取的 5 个序列的 BAC 克隆如下。

提取的序列	对应的 BAC 克隆
A2	RP11－152F13
A7	RP11－643G5
A18	RP11－115A14
B5	RP11－640M9
C14	RP11－702F3

[0134] 将 10 μl 解冻的上述 BCA(Advanced GenoTechs Co.) 接种到 1 ml 培养基（含有 15 μg/ml 终浓度的氯霉素）中，在 37℃孵育过夜。将 30 μl 培养液转移至 1.4 ml 培养基（含有 15 μg/ml 终浓度的氯霉素），在 37℃孵育 2 小时。重复离心两次并且重复用无菌水洗涤两次，让细胞悬浮在 20 μl 无菌水中。向冷却的杯 (0.1 cm) 中加入 1 μl pRED/ET (Gene Bridges) 和大肠杆菌，接着电穿孔 (1350 V, 10 μF)。然后，向其中加入 1 ml SOC 培养基，让得到的混合物在 30℃孵育 70 分钟。将 100 μl 培养液铺板在 LB 平板上（分别含有 3 μg/ml 终浓度的四环素和 15 μg/ml 终浓度的氯霉素），在 30℃孵育过夜。第二天将每个由此获得的菌落接种到 1 ml 培养基（分别含有 3 μg/ml 终浓度的四环素和 15 μg/ml 终浓度的氯霉素）中，在 30℃孵育过夜。将 30 μl 培养液转移至 1.4 ml 培养基（分别含有 3 μg/ml 终浓度的四环素和 15 μg/ml 终浓度的氯霉素）中，30℃孵育 2 小时。然后，向其中加入 50 μl 10% L-阿拉伯糖，在 37℃进一步孵育 1 小时。此后，用无菌水重复洗涤两次，将悬浮在 30 μl 无菌水中的大肠杆菌和 1 μl 用于转化的 DNA 加入至冷却的杯 (0.1 cm)，接着电穿孔 (1350 V, 10 μF)。然后，向其中加入 1 ml SOC 培养基，让得到的混合物在 37℃孵育 90 分钟。将全部量的培养液铺板在 LB 平板（含有 100 μg/ml 氨苄西林）上，孵育平板。此后，通过碱裂解获得靶质粒。最后，确认获得的质粒的序列及其限制性酶切位点，藉此构建靶质粒。载体构建体如图 2 所示。

[0135] (2-3) 利用 SEAP 表达作为指标进行评估

利用宿主细胞 CHO-K1 (ATCC) 和转染试剂 Lipofectamine 2000(Invitrogen) 评估每个在 (2-2) 中构建的质粒。

[0136] 转染后 2 天开始进行约 2 周的用 800 μg/ml 潮霉素的抗生素筛选，藉此建立稳定表达的多克隆细胞系。在测量的前一天对由此建立的细胞系进行培养基更换，将给定数量的细胞接种至 24-孔板 (IWAKI)。细胞铺板 24 小时后，收集培养上清液，测量 SEAP 的活性。用 SensoLyte™ pNPP 分泌型碱性磷酸酶报告物测定来测量培养上清液中的 SEAP 活性。

[0137] 测量结果如图 3 所示。当将不具有元件的对照的 SEAP 活性标准化为 1 时，具有 DNA 元件 A2、A7、A18、B5 或 C14 的稳定表达的 CHO 细胞系的培养上清液中的 SEAP 活性，显示为对照的 5 倍或更多倍数的数值。基于该结果，确定所有 5 种类型的 DNA 元件显著提高 SEAP 表达。附带说一下，以上 5 种类型 DNA 元件的多核苷酸序列分别由序列表中的 SEQ ID NOS :1-5 代表。

[0138] 实施例 3

待联合使用的启动子的概述

实施例 2 中评估 DNA 元件使用的载体的启动子为 CMV 启动子，因此在实施例 3 中研究 DNA 元件与其它通用启动子的联合使用。

[0139] (3-1) 用 EF-1α 和 SV40 启动子构建 SEAP 表达载体

通过使用 pSEAP2- 对照 (Clontech) 作为模板，由 PCR 方法 (94℃ 30 秒和 68℃ 2 分钟 x 40 个循环）用 (2-1) 中所述的引物和 KOD-plus- 来扩增 SEAP 基因。以与 (2-1) 中相同的方法制备扩增的 SEAP 作为插入片段。用限制性内切酶 NheI 和 BglII 消化该插入片段，并用限制性内切酶 NheI 和 BamHI 消化载体 pIRES puro3 (Clontech)，以与 (2-1) 中相同的方法构建 pCMV/SEAP ires Puro。

[0140] 随后，通过使用 pEF1/V5-His A (Invitrogen) 作为模板，由 PCR 方法 (94℃ 15 秒、60℃ 30 秒、68℃ 2 分钟 x 30 个循环）使用以下引物和 KOD-plus- 来扩增 EF-1α 启动子。

5′-AAAACTAGTCAGAGAGGAATCTTTGCAGCTAATGGACC-3′

5′-AAAGATATCCCTAGCCAGCTTGGGTGGTACCAAGC-3′

[0141] 通过用上文构建的 pCMV/SEAP ires Puro 作为载体，用限制性内切酶 SpeI 和 EcoRV 对载体和启动子进行消化，并按照 (2-1) 中所述方法构建 pEF/SEAP ires Puro。

[0142] 同样地，通过使用 pcDNA3.1+ (Invitrogen) 作为模板，由 PCR 方法 (94℃ 15 秒、60℃ 30 秒、68℃ 1 分钟 x 30 个循环）利用以下引物和 KOD-plus- 来扩增 SV40 启动子。

5′-AAAACTAGTCTGTGGAATGTGTGTCAGTTAGGGTG-3′

5′-AAAGATATCAGCTTTTTGCAAAAGCCTAGGCCTC-3′

[0143] 通过使用上文构建的 pCMV/SEAP ires Puro 作为载体，用限制性内切酶 SpeI 和 EcoRV 对载体和启动子进行消化，按照 (2-1) 所述方法构建 pSV40/SEAP ires Puro。

[0144] (3-2) 克隆 DNA 元件 A2 或 A7

随后，用在 (3-1) 中构建的 pEF/SEAP ires Puro 和 pSV40/SEAP ires Puro 作为基本结构进行 DNA 元件 A2 或 A7 的克隆。

[0145] 首先，用限制性内切酶 SpeI 消化 pEF/SEAP ires Puro 和 pSV40/SEAP ires Puro 数小时，接着乙醇沉淀，用无菌水溶解沉淀。通过利用经 SpeI 消化的各种载体作为模板，由 PCR 方法 (94℃ 15 秒、55℃ 30 秒、68℃ 10 分钟 x 30 个循环）利用以下引物和 KOD-plus- 来制备用于转化的 DNA。

A2 (EF/D)：

5′-GGAAATTGAGAAGTATCATTCACAACAGTACCACAAACATGAAATAAATGTGCTA

GTCAGAGAGGAATCTTTGCAGC-3′

A2 (SV40/D)：

5′-GGAAATTGAGAAGTATCATTCACAACAGTACCACAAACATGAAATAAATGTGCTA

GTCTGTGGAATGTGTGTCAGTTAG-3′

[0146] A2 (EF 和 SV40/R)：

5′-CTCATTCTGTGGGTTGTCATTTCACTTCCTTGATGCTATCCTTTCAAGCAAAATT

TTAAAACTTTATCCATCTTTGCA-3′

A7 (EF/D):

5′-CTTATTTTCTAAGTAGTATAGACTTAATTGTGAGAACAAAATAAAAACTTGCTAG

TCAGAGAGGAATCTTTGCAGC-3′

A7 (SV40/D):

5′-CTTATTTTCTAAGTAGTATAGACTTAATTGTGAGAACAAAATAAAAACTTGCTAG

TCTGTGGAATGTGTGTCAGTTAG-3′

A7 (EF 和 SV40/R):

5′-CTCTTCCCATTCTCATTTGAATCTACTTCAAAAGGTTTACCATACTAAGAACTAG

TTTTAAAACTTTATCCATCTTTGCA-3′

通过使用由此制备的用于转化的 DNA 和用 pRed/ET 转染的 BAC，将 DNA 元件 A2 或 A7 克隆至 (3-1) 所述的载体。载体结构如图 2 所示。附带说一下，按照 (2-2) 所述方法实施程序。

[0147] (3-3) 用 SEAP 表达作为指标进行评估

用宿主细胞 CHO-K1 (ATCC) 和转染试剂 Lipofectamine 2000 (Invitrogen) 评估在 (3-2) 中构建的每个质粒。

[0148] 转染后 2 天开始进行约 2 周的用 8 μg/ml 嘌呤霉素的抗生素筛选，藉此建立稳定表达的多克隆细胞系。在测量前一天对由此建立的细胞系进行培养基更换，将给定数量的细胞接种至 24- 孔板 (IWAKI)。细胞铺板 24 小时后，收集培养上清液，测量 SEAP 的活性。用 SensoLyte™ pNPP 分泌型碱性磷酸酶报告物测定 (ANASPEC) 来测量培养上清液中的 SEAP 活性。

[0149] 测量结果如图 4 所示。当将不具有元件的对照的 SEAP 活性标准化为 1 时，DNA 元件 A2 或 A7 显示出提高表达的作用，使得在使用 EF-1α 启动子情况下，SEAP 活性为对照的 2 倍或更高，而在使用 SV40 启动子情况下，SEAP 活性为对照的 4 倍或更高。基于该结果，确定当与通用启动子联合使用时，这些 DNA 元件显示出提高外源基因表达的作用。

[0150] 实施例 4

利用抗体表达作为指标进行评估

(4-1) 构建人轻链表达载体 pEF6KCL

通过使用质粒 pEF6/V5-HisB (Invitrogen) 作为模板，由 PCR 方法使用以下引物和 KOD-plus- 来获得位于 2174 位（紧邻 BGHpA 下游）和 2958 位 (SmaI) 之间的 DNA 片段（含有 f1 复制起点和 SV40 启动子和起点的 DNA 片段，下文中称作“片段 A”，片段 A 的多核苷酸序列由序列表中的 SEQ ID NO: 6 代表）。

5′-CCACGCGCCCTGTAGCGGCGCATTAAGC-3′

5′-AAACCCGGGAGCTTTTTGCAAAAGCCTAGG-3′

[0151] 获得的片段 A 和含有编码人 K 链分泌信号、人 K 链恒定区和人多聚 (A) 添加信号的 DNA 序列的 DNA 片段（下文中称作“片段 B”）通过重叠 PCR 连接。用限制性内切酶 KpnI

和 SmaI 消化由此得到的连接了片段 A 和片段 B 的 DNA 片段，让得到的片段与用限制性内切酶 KpnI 和 SmaI 消化的质粒 pEF6/V5-HisB (Invitrogen) 连接，藉此构建具有信号序列、克隆位点、人 K 链恒定区和 EF-1α 启动子下游区的人多聚 (A) 添加信号序列的人轻链表达载体 pEF6KCL。

[0152] 将通过切割 pEF6KCL（其通过上述用限制性内切酶 KpnI 和 SmaI 的方法获得）获得的 DNA 片段与用 KpnI 和 SmaI 消化的 pEF1/myc-HisB (Invitrogen) 连接，接着转化、碱裂解及其序列确认，藉此构建质粒 pEF1KCL。

[0153] (4-2) 人重链表达载体 pEF1FCCU 的构建

用限制性内切酶 NheI 和 PmeI 消化含有编码人 IgG1 信号序列和恒定区氨基酸序列的 DNA 序列的 DNA 片段（该 DNA 片段的多核苷酸序列由序列表中的 SEQ ID NO: 7 代表），让得到的片段与用限制性内切酶 NheI 和 PmeI 消化的质粒 pEF1KCL 连接，藉此构建具有信号序列、克隆位点、人重链恒定区和 EF-1α 启动子下游区的人多聚 (A) 添加序列的人重链表达载体 pEF1FCCU。

[0154] (4-3) 构建单一人源化抗体基因 X 表达载体（人源化抗体基因 X/pEF_LHN#）

通过连接在 (4-1) 或 (4-2) 中构建的 L- 链或 H- 链表达载体，构建单一人源化抗体表达载体 (pEF_LHN（缺乏可变区））。

[0155] 通过 PCR 方法将限制性内切酶 SalI 位点添加至从 pEF1KCL 启动子的上游区至多聚 (A) 下游区的基因表达单元的两端。然后进行琼脂糖凝胶电泳，从凝胶切割所期需的 DNA 片段，并纯化该 DNA 片段，藉此制备插入片段。通过用限制性内切酶 SalI 消化在 (4-2) 中构建的 pEF1FCCU，在位于基因表达单元上游区的 SalI 位点使载体线性化。然后，让线性化载体与以上插入片段连接，接着转化、碱裂解、序列确认，藉此构建单一人源化抗体表达载体 (pEF_LHN（缺乏可变区））。

[0156] 随后，将以下寡核苷酸引入至载体 pEF_LHN（缺乏可变区）的 AatII 位点。

5′-CGCGGCCGCACTAGTGACGT-3′

5′-CACTAGTGCGGCCGCGACGT-3′

[0157] 将各个寡核苷酸稀释至 5 皮摩尔，并且通过利用 T4 多核苷酸激酶 (TAKARA) 让反应在 37℃进行 1 小时。然后，向其中加入 10x 缓冲液 (TAKARA)，室温下在 96℃进行 1 分钟退火。连接这些寡核苷酸和用限制性内切酶 AatII 消化的载体 pEF_LHN，接着转化、碱裂解和序列确认，藉此构建 pEF_LHN#（缺乏可变区）。

[0158] 通过将人源化抗体基因 X 的可变区整合至以上构建的通用载体 (pEF_LHN#（缺乏可变区）），来完成人源化抗体基因 X 表达单一载体（人源化抗体基因 X/pEF_LHN#）的构建。

[0159] 首先，通过利用以下引物和 KOD-plus-，由 PCR 方法 (94℃ 15 秒、55℃ 30 秒和 68℃ 1 分钟 x 30 个循环）扩增人源化抗体基因 X 的 L- 链可变区。

[0160] L- 链可变区：

5′-AAACATATGGCGACATCCAGATGAC-3′

5′-AAACGTACGCTTGATCTCCACCTTGG-3′

用限制性内切酶 NdeI 和 BsiWI 消化扩增的 L- 链可变区片段和通用载体 (pEF_LHN#

(缺乏可变区)),接着琼脂糖凝胶电泳,从凝胶切割所期需的片段、纯化、连接反应、转化、碱裂解和序列确认,藉此将 L- 链可变区整合至载体。以同样的方法,通过利用以下引物和 KOD-plus-,由 PCR 方法 (94℃ 15 秒、55℃ 30 秒和 68℃ 1 分钟 x 30 个循环) 扩增人源化抗体基因 X 的 H- 链可变区。

[0161] H- 链可变区 :

5′-AAAGCTGAGCCAGGTGCAGCTGCAGG-3′

5′-AAAGCTGAGCTCACGGTCACCAGGGTTC-3′

用限制性内切酶 BlpI 消化扩增的 H- 链可变区片段和具有插入其中的 L- 链可变区的通用载体,接着琼脂糖凝胶电泳,从凝胶切割所期需片段、纯化、连接反应、转化、碱裂解和序列确认,藉此将 H- 链可变区整合至载体,并构建单一人源化抗体基因 X 表达载体 (人源化抗体基因 X/pEF_LHN#)。

[0162] (4-4) 构建单一人源化抗体基因 X 表达载体 (人源化抗体基因 X/pCMV_LHN#)

通过利用在 (4-3) 中构建的单一人源化抗体基因 X 表达载体 (人源化抗体基因 X/pEF_LHN#) 作为基本载体结构,按照以下程序通过置换启动子构建另一个单一人源化抗体基因 X 表达载体 (人源化抗体基因 X/pCMV_LHN#)。

[0163] 通过利用 pIRES puro3 作为模板,由 PCR 方法 (94℃ 30 秒和 68℃ 3 分钟 x40 个循环) 使用以下引物和 KOD-plus- 来扩增 CMV 启动子片段。

[0164] H- 链上游 :

5′-CTTTTGCAAAAAGCTTCGCGTTACATAACTTACGGTAAATGGCC-3′

5′-TTCATGGTGGCGCTAGCCCGCAGATATCGATCCGAGCTCGGTA-3′

L- 链上游 :

5′-TGACGTCGACAAGCTTCGCGTTACATAACTTACGGTAAATGGCC-3′

5′-CTGGATGTCGCCATATGCGCCGGAGATCCACAGCAGCAGGGAGATGAACACCTGG

GTCTGCAGCACCATGGTGGCGCTAGCCCGCAGATATCGATCCGAGCTCGGTA-3′

向 PCR 反应溶液中加入限制性内切酶 DpnI,让反应在 37℃进行 1 小时,接着用 miniElute 反应净化试剂盒 (Qiagen) 纯化,藉此制备用于 In-Fusion 的样品。同时,用限制性内切酶 HindIII、NheI、NdeI 和 FseI 消化人源化抗体基因 X/pEF_LHN#,接着琼脂糖凝胶电泳,藉此从得到的片段中分离两个大片段。从凝胶切割每个片段,从凝胶提取 DNA,藉此制备用于 In-Fusion 的样品。将所有用于 In-Fusion 的样品放在一起,使用 In-Fusion™ AdvantagePCR 克隆试剂盒 (TAKARA) 进行克隆,接着转化、碱裂解和序列确认,藉此构建单一人源化抗体基因 X 表达载体 (人源化抗体基因 X/pCMV_LHN#)。

[0165] (4-5)DNA 元件 A7 的克隆

A7 选自 5 种类型的确定具有增强 SEAP 表达作用并且克隆至抗体表达载体的 DNA 元件。

[0166] 以与 (2-2) 相同的方法,通过使用用限制性内切酶 NotI 消化的每个人源化抗体基因 X 表达单一载体 (人源化抗体基因 X/pEF_LHN# 和人源化抗体基因 X/pCMV_LHN#) 作为模板,通过使用以下引物和 KOD-plus- 进行的 PCR 方法 (94℃ 15 秒、55℃ 30 秒和 68℃ 11 分

钟 x30 个循环）制备用于转化的 DNA。

[0167] 人源化抗体基因 X/pEF_LHN# D：

5′-CTCTTCCCATTCTCATTTGAATCTACTTCAAAAGGTTTACCATACTAAGACTCGA

GGCACTAGTGACGTCAGGTGGCACT-3′

人源化抗体基因 X/pEF_LHN# R：

5′-CTCTTCCCATTCTCATTTGAATCTACTTCAAAAGGTTTACCATACTAAGAGCACT

AGTGACGTCAGGTGGCACTTTTCGG-3′

人源化抗体基因 X/pCMV_LHN# D：

使用人源化抗体基因 X/pEF_LHN# D。

[0168] 人源化抗体基因 X/pCMV_LHN# R：

使用人源化抗体基因 X/pEF_LHN# R。

[0169] 通过使用以上制备的用于转化的 DNA 和用 pRed/ET 转染的 BAC，将 DNA 元件 A7 克隆至 (4-3) 和 (4-4) 中所述的单一人源化抗体基因 X 表达载体。载体构建体如图 5 所示。附带说一下，按照 (2-2) 所述的方法实施程序。

[0170] (4-6) 使用抗体表达作为指标进行评估

使用宿主细胞 CHO-K1 (ATCC) 和转染试剂 Lipofectamine 2000 (Invitrogen) 评估在 (4-5) 中构建的每个质粒。

[0171] 转染后 2 天开始进行约 2 周的用 800 μg/ml 遗传霉素 (Roche) 的抗生素筛选，藉此建立稳定表达的多克隆细胞系。在测量前一天对由此建立的细胞系进行培养基更换，将给定数量的细胞接种至 24- 孔板 (IWAKI)。细胞铺板 24 小时后，收集培养上清液，通过 ELISA 方法测量培养上清液中抗体的表达水平。附带说一下，如下进行 ELISA。向用 50 ng/ 孔抗 -κ 轻链包被的 96- 孔板中每孔加入 100 μl 无细胞培养上清液，让板在 37℃孵育 1 小时。随后，移除样品（培养上清液），用 200 μl PBS- 吐温 (0.05%) 洗涤每个孔。然后，向每个孔中加入 100 μl HRP- 标记的抗 - 人 IgG (Fc)，让板在 37℃再孵育 1 小时。此后，移除 HRP- 标记的抗 - 人 IgG (Fc)，用 PBS- 吐温 (0.05%) 洗涤每个孔。然后，用 POD 底物 ABTS 试剂盒 (Nacalai) 显色，测量在 405 nm 测量波长处的吸光度。为了稀释抗 -κ 轻链、抗 - 人 IgG (Fc) 和样品，使用 PBS- 吐温 (0.05%)。通过使用连续稀释至 12 ng、6 ng、3 ng、1.5 ng、0.75 ng、0.375 ng 和 0.1875 ng 的人 IgG 作为标准品，计算样品的浓度。

[0172] 结果如图 6 所示。已证实当在抗体表达载体中使用 EF-1α 启动子或 CMV 启动子时，与不具有元件的对照相比，具有 DNA 元件 A7 的样品具有更高的提高抗体产生的作用。

[0173] 实施例 5

表现出提高外源基因表达的活性的序列长度

(5-1) 克隆具有不同序列长度的 DNA 元件

基于实施例 2 中使用的序列长度，构建含有每个 DNA 元件但是具有不同序列长度的载体。

[0174] 基于每个 DNA 元件 A2、A7、A18、B5 和 C14 的全长设计的具有不同序列长度的 DNA 元件详情分别示于图 7、9、11、13、15、18 和 19。用限制性内切酶 SpeI 消化 (2-1) 中所述的

pCMV/SEAP ires Hygro 达数小时，接着乙醇沉淀，用无菌水溶解沉淀。通过使用经 SpeI 消化的载体作为模板，由 PCR 方法（94℃ 15 秒、55℃ 30 秒、68℃ 10 分钟 x 30 个循环）使用以下引物和 KOD-plus-，来制备用于转化的 DNA。通过使用由此制备的用于转化的 DNA 和相应的用 pRed/ET 转染的 BAC，将每个具有不同序列长度的 DNA 元件克隆至 (2-1) 所述的 pCMV/SEAP ires Hygro 中。载体构建体如图 2 所示。附带说一下，按照 (2-2) 所述方法实施程序。

A2-1D:

5′-CATGCACAGATTAGCCATTTAGTACTTACTAAATCAAACTCAATTTCTGAAGTCT

AGTTATTAATAGTAATCAATTACG-3′

A2-1R:

5′-CTCATTCTGTGGGTTGTCATTTCACTTCCTTGATGCTATCCTTTCAAGCAAAATT

CAATAATCAATGTCAACGCGTATAT-3′

A2-2D:

5′-ACACTGGTCAAAGGGACAGGTCATTGTTATGCTGGCAATGCAGGCTGCTGAAAAC

TAGTTATTAATAGTAATCAATTACG-3′

A2-2R:

5′-ACTGTAGCTTCTTATTTTTTACCTGCAGTGCATTCCTGTAAAAGTAGTGTGGAGT

CAATAATCAATGTCAACGCGTATAT-3′

A2-3D:

5′-CTGGAAATTGAGAAGTATCATTCACAACAGTACCACAAACATGAAATAAATGTGC

TAGTTATTAATAGTAATCAATTACG-3′

A2-3R:

5′-CCAAGCTTGTCCAACCGCGGCCTGCAGGCTGCATGCAGCCTGTGAAGGCTTTGAT

CAATAATCAATGTCAACGCGTATAT-3′

A2-4D:

5′-TCAATCATTTATCAATTTTATCTTCAAAGTCCCTCACTTCAGGGAGATGATATAC

TAGTTATTAATAGTAATCAATTACG-3′

A2-4R:

5′-ATATATAAAAGTTCATGTATATATAAAATCATGCAATACACGGCCTTTTGTGACT

CAATAATCAATGTCAACGCGTATAT-3′

A2-5D:

5′-CGCATAAAAGGAAAAGCATCCTTAAAATAAACACCATCAATGGCTCCTCGGTGGC

TAGTTATTAATAGTAATCAATTACG-3′

A2-5R:

[0175] 使用 A2-4R。

A2-6D:

5′-GGGAGGCTACAGCTTGCCTCTCTAACCACTAAAAGGCATGACCCTCCTCAAAGCT

AGTTATTAATAGTAATCAATTACG-3′

A2-6R:

[0176] 使用 A2-4R。

A2-7D:

5′-TCTGGCTTCCCTGGGCCACGCTGGAAGAAGAATTGTCTTGCGCCACACATAAAAC

TAGTTATTAATAGTAATCAATTACG-3′

A2-7R:

5′-AGCTGATTTTTACGTTAAATGTAACATGTAAAGAAATATATGTGTGTTTTTAGAT

CAATAATCAATGTCAACGCGTATAT-3′

A2-8D:

5′-GTGAAGAGGAGGAGATGTCAAAATTCAAAGTCTTAAATGATGTAGTTTTAAGTAC

TAGTTATTAATAGTAATCAATTACG-3′

A2-8R:

5′-ATGACACTTGATATTGTTGTTTATATTGCTGGTTAGTATGTGCCTTCATTTACCT

CAATAATCAATGTCAACGCGTATAT-3′

[0177] A2-9D：

使用 A2-6D。

[0178] A2-9R：

使用 A2R。

[0179] A2-10D：

使用 A2-2D。

[0180] A2-10R：

使用 A2-7R。

[0181] A2-11D：

使用 A2-8D。

[0182] A2-11R：

使用 A2-2R。

[0183] A2-12D：

使用 A2-2D。

[0184] A2-12R：

使用 A2-4R。

[0185] A2-13D：

使用 A2-8D。

[0186] A2-13R：

使用 A2-7R。

[0187] A2-14D：

使用 A2D。

[0188] A2-14R：

使用 A2-2R。

[0189] A2-15D：

使用 A2-2D。

[0190] A2-15R：

使用 A2R。

[0191] A2-16D：

使用 A2-8D。

[0192] A2-16R：

使用 A2-4R。

[0193] A2-17D：

使用 A2D。

[0194] A2-17R：

使用 A2-7R。

[0195] A7-1D：

5′-AAAAACAAAACTGGAGTAAACAAGATGAATTGTTTTAATAGAGGCACTGTATTAC

TAGTTATTAATAGTAATCAATTACG-3′

A7-1R:

5′-ATACAATGTTCCATGTATTCTGTGCCTGAACCTATGCAGCTGATGTAGCTGAAGT

CAATAATCAATGTCAACGCGTATAT-3′

A7-2D:

5′-GATCTTATTTTCTAAGTAGTATAGACTTAATTGTGAGAACAAAATAAAAACTTGC

TAGTTATTAATAGTAATCAATTACG-3′

A7-2R:

5′-TGTTGTTTTCAGCCACTAAGTTTGAGGTGATTTGTTCTGGCAGTCCTAGGAAACT

CAATAATCAATGTCAACGCGTATAT-3′

A7-3D：

使用 A7-2D。

[0196] A7-3R：

5′-AGCCTACACTACCCTTTGCAGCCTTTGGTAACTATCCTTCTGCTGTCTACCTCCT

CAATAATCAATGTCAACGCGTATAT-3′

A7-4D:

5′-AGGAGCTCCTGAATGAAGGACATCACTCAGCTGTGTTAAGTATCTGGAACAATAC

TAGTTATTAATAGTAATCAATTACG-3′

A7-4R:

5′-GACATAAAATGTAAGATATGATATGCTATGTAAGATATGATACCTGCCTTAAAAT

CAATAATCAATGTCAACGCGTATAT-3′

A7-5D:

5′-CACTGCTTGATACTTACTGTGGACTTTGAAAATTATGAATGTGTGTGTGTGTGTC

TAGTTATTAATAGTAATCAATTACG-3′

A7-5R:

5′-CAATTACATTCCAGTGATCTGCTACTTAGAATGCATGACTGAACTCCTGGGTGGT

CAATAATCAATGTCAACGCGTATAT-3′

A7-6D:

5′-TTATTTTGAAGAGAAACTCCTGGTTCCCACTTAAAATCCTTTCTTGTTTCCAAGC

TAGTTATTAATAGTAATCAATTACG-3′

A7-6R:

5′-AAGCAGTGTGTGTTTACCTGCATGTGTATGTGAATTAACTCTGTTCCTGAGGCAT

CAATAATCAATGTCAACGCGTATAT-3′

A7-7D:

5′-ATTGCATGTTCTCATTTATTTGTGGGATGTAAAAATCAAAACAATAGAACGTATC

TAGTTATTAATAGTAATCAATTACG-3′

A7-7R:

5′-TTGGGAGGCCGCAGCTGGTAGATCACTTGAGGCCACGAATTTGACACCAGCAGGT

CAATAATCAATGTCAACGCGTATAT-3′

A7-8D：

使用 A7-1D。

[0197] A7-8R：

使用 A7R。

[0198] A7-9D：

使用 A7-7D。

[0199] A7-9R：

使用 A7-5R。

[0200] A7-10D：

使用 A7-4D。

[0201] A7-10R：

使用 A7-7R。

[0202] A7-11D：

使用 A7-6D。

[0203] A7-11R：

使用 A7-4R。

[0204] A7-12D：

使用 A7-2D。

[0205] A7-12R：

使用 A7-6R

A7-13D：

使用 A7-7D。

[0206] A7-13R：

使用 A7R。

[0207] A7-14D：

使用 A7-4D。

[0208] A7-14R：

使用 A7-5R。

[0209] A7-15D：

使用 A7-6D。

[0210] A7-15R：

使用 A7-7R。

[0211] A7-16D：

使用 A7-2D。

[0212] A7-16R：

使用 A7-4R。

[0213] A7-17D：

使用 A7-4D。

[0214] A7-17R：

使用 A7R。

[0215] A7-18D：

使用 A7-6D。

[0216] A7-18R：

使用 A7-5R。

[0217] A18-1：

5′-ATCCCCTGCTCTGCTAAAAAAGAATGGATGTTGACTCTCAGGCCCTAGTTCTTGA

TCCTATTAATAGTAATCAATTACG-3′

A18-1R：

使用 A18R。

A18-2D:

5′-CTAAAGTGCTGGGATTACAGGCATAAGCCACCGTGCCCGGCTGGAGCATTGGGAT

CCTATTAATAGTAATCAATTACG-3′

A18-2R:

5′-ACTACTTACACATTTCGAGTTTTAAATAAGGCGTTCAATATAGAGTGAACACCTA

GTCAATAATCAATGTCAACG-3′

A18-3D:

5′-CAGGCATAAGCCACCGCACCCGGCCACCCCTTACTAATTTTTAGTAACGTCGATC

CTATTAATAGTAATCAATTACG-3′

A18-3R:

5′-CTGATTGACTTTGACCTCTGCTTTCCAACTTTGCCCCAAAGAAAGTTAGTCACCT

AGTCAATAATCAATGTCAACG-3′

[0218] A18-4D：

使用 A18-3D。

[0219] A18-4R：

5′-TTCAATGAAACAAGCTCTGTGAGGCTCATTTGTACCCATTTTGTTCAGTACTGCC

TAGTCAATAATCAATGTCAACG-3′

B5-1D:

5′-ACATACCCAGAGACACTGAGAGAGACAGACAGACAGTAAACAGAGGAGCACGATC

CTATTAATAGTAATCAATTACG-3′

B5-1R:

使用 B5R。

[0220] B5-2D：

5′-GCTCAATTGTATCTTATGAAAACAATTTTTCAAAATAAAACAAGAGATATGATCC

TATTAATAGTAATCAATTACG-3′

B5-2R：

使用 B5R。

B5-3D:

5′-CCTGTGCTGAATACCGTCTGCATATGTATAGGAAAGGGTTAACTCAGCAGGGATC

CTATTAATAGTAATCAATTACG-3′

B5-3R:

5′-TATGTGAATGGAAATAAAATAATCAAGCTTGTTAGAATTGTGTTCATAATGACCC

TAGTCAATAATCAATGTCAACG-3′

[0221] B5-4D：

使用 B5D。

[0222] B5-4R：

5′-GAAAGTCTACAATTTTTTCAGTTTAAAATGGTATTTATTTGTAACATGTACCCTA

GTCAATAATCAATGTCAACG-3′

B5-5D：

使用 B5-1D。

[0223] B5-5R：

5′-CAAAGATGAAGGATGAGAGTGACTTCTGCCTTCATTATGTTATGTGTTCATATCC

TAGTCAATAATCAATGTCAACG-3′

B5-6D:

5′-CAGTGAATTATTCACTTTGTCTTAGTTAAGTAAAAATAAAATCTGACTGTGATCC

TATTAATAGTAATCAATTACG-3′

B5-6R:

5′-GAACAGACAGGTGAATGAGCACAGAGGTCATTTGTAAACCGTTTGTGGTTAGCCT

AGTCAATAATCAATGTCAACG-3′

C14-1D:

5′-CTTTTTGGCTTCTGTGTTTAAGTTATTTTTCCCCTAGGCCCACAAACAGAGTCGA

TCCTATTAATAGTAATCAATTACG-3′

C14-1R:

5′-AACCTTGGAAAAATTCTGTTGTGTTTAGAAGCATGTACCAATCTATCACTCCTAG

TCAATAATCAATGTCAACG-3′

C14-2D:

5′-CTATTCACTGTCTGTAGGATGAAAAAGTTAATAACACCCTGAGAGGTTTCGATCC

TATTAATAGTAATCAATTACG-3′

C14-2R:

5′-CCTTAGATTAGTTTATTGTATTTTTTATCAGCTACTATAAGGTTTACACACCCTA

GTCAATAATCAATGTCAACG-3′

C14-3D:

5′-CAAGACCCTCAAAATTCAAAAATTTCCTTTATCTTGCTGTAGCACCTCCTGCGAT

CCTATTAATAGTAATCAATTACG-3′

C14-3R:

5′-GGAGGGGATAGGAAGGGGATGAGGCCTAACAGGTTGATGATCTAGGCTTTACCTA

GTCAATAATCAATGTCAACG-3′

C14-4D:

5′-CTCAAAAAGGAGATAATTCCAGCCCCTCGCCTTAAAGAATCCCTATCAAGTGATC

CTATTAATAGTAATCAATTACG-3′

C14-4R：

使用 C14-1R。

C14-5D:

5′-CGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCACGCCGTTGGATCC

TATTAATAGTAATCAATTACG-3′

[0224] C14-5R：

使用 C14-1R。

[0225] C14-6D：

使用 C14-4D。

[0226] C14-6R：

5′-TTAACTTTTTCATCCTACAGACAGTGAATAGTAAAGCTTTCTGTGAAGACATACC

CTAGTCAATAATCAATGTCAACG-3′

C14-7D：

使用 C14-2D

C14-7R：

使用 C14-1R。

[0227] C14-8D：

使用 C14-3D。

[0228] C14-8R：

5′-AAATTATTTCCTGGTGGGCAATATTAGAATATGGGGAATGTTTGCTTCTGAGCCT

AGTCAATAATCAATGTCAACG-3′

C14-9D：

使用 C14-4D。

[0229] C14-9R：

使用 C14-3R。

[0230] C14-10D：

使用 C14-2D。

[0231] C14-10R：

使用 C14R。

[0232] C14-11D：

使用 C14-3D。

[0233] C14-11R：

使用 C14-2R。

[0234] C14-12D：

使用 C14-4D。

[0235] C14-12R：

使用 C14-8R。

[0236] C14-13D：

使用 C14-3D。

[0237] C14-13R：

使用 C14-1R。

[0238] C14-14D：

使用 C14-4D。

[0239] C14-14R：

使用 C14-2R。

[0240] 关于 A2 的各个片段的多核苷酸序列：A2-1 对应于序列表中 SEQ ID NO: 1 的 1-3000 位核苷酸的多核苷酸序列；A2-2 对应于序列表中 SEQ ID NO: 1 的 2801-5800 位核苷酸的多核苷酸序列；A2-3 对应于序列表中 SEQ ID NO: 1 的 5401-8540 位核苷酸的多核苷酸序列；A2-4 对应于序列表中 SEQ ID NO: 1 的 701-2700 位核苷酸的多核苷酸序列；A2-5 对应于序列表中 SEQ ID NO: 1 的 701-2200 位核苷酸的多核苷酸序列；A2-6 对应于序列表中 SEQ ID NO: 1 的 701-3700 位核苷酸的多核苷酸序列；A2-7 对应于序列表中 SEQ ID NO: 1 的 2001-5000 位核苷酸的多核苷酸序列；A2-8 对应于序列表中 SEQ ID NO: 1 的 4001-7000 位核苷酸的多核苷酸序列；A2-9 对应于序列表中 SEQ ID NO: 1 的 1-3700 位核苷酸的多核苷酸序列；A2-10 对应于序列表中 SEQ ID NO: 1 的 2001-5800 位核苷酸的多核苷酸序列；

A2-11 对应于序列表中 SEQ ID NO: 1 的 2801-7000 位核苷酸的多核苷酸序列 ;A2-12 对应于序列表中 SEQ ID NO: 1 的 701-5800 位核苷酸的多核苷酸序列 ;A2-13 对应于序列表中 SEQ ID NO: 1 的 2001-7000 位核苷酸的多核苷酸序列 ;A2-14 对应于序列表中 SEQ ID NO: 1 的 2801-8450 位核苷酸的多核苷酸序列 ;A2-15 对应于序列表中 SEQ ID NO: 1 的 1-5800 位核苷酸的多核苷酸序列 ;A2-16 对应于序列表中 SEQ ID NO: 1 的 701-7000 位核苷酸的多核苷酸序列 ;和 A2-17 对应于序列表中 SEQ ID NO: 1 的 2001-8450 位核苷酸的多核苷酸序列。

[0241] 关于 A7 的各个片段的多核苷酸序列 :A7-1 对应于序列表中 SEQ ID NO: 2 的 601-3600 位核苷酸的多核苷酸序列 ;A7-2 对应于序列表中 SEQ ID NO: 2 的 3601-8420 位核苷酸的多核苷酸序列 ;A7-3 对应于序列表中 SEQ ID NO: 2 的 5401-8420 位核苷酸的多核苷酸序列 ;A7-4 对应于序列表中 SEQ ID NO: 2 的 3401-6400 位核苷酸的多核苷酸序列 ;A7-5 对应于序列表中 SEQ ID NO: 2 的 1501-4500 位核苷酸的多核苷酸序列 ;A7-6 对应于序列表中 SEQ ID NO: 2 的 4401-7400 位核苷酸的多核苷酸序列 ;A7-7 对应于序列表中 SEQ ID NO: 2 的 2401-5400 位核苷酸的多核苷酸序列 ;A7-8 对应于序列表中 SEQ ID NO: 2 的 1-3600 位核苷酸的多核苷酸序列 ;A7-9 对应于序列表中 SEQ ID NO: 2 的 1501-5400 位核苷酸的多核苷酸序列 ;A7-10 对应于序列表中 SEQ ID NO: 2 的 2401-6400 位核苷酸的多核苷酸序列 ;A7-11 对应于序列表中 SEQ ID NO: 2 的 3401-7400 位核苷酸的多核苷酸序列 ;A7-12 对应于序列表中 SEQ ID NO: 2 的 4401-8420 位核苷酸的多核苷酸序列 ;A7-13 对应于序列表中 SEQ ID NO: 2 的 1-5400 位核苷酸的多核苷酸序列 ;A7-14 对应于序列表中 SEQ ID NO: 2 的 1501-6400 位核苷酸的多核苷酸序列 ;A7-15 对应于序列表中 SEQ ID NO: 2 的 2401-7400 位核苷酸的多核苷酸序列 ;A7-16 对应于序列表中 SEQ ID NO: 2 的 3401-8420 位核苷酸的多核苷酸序列 ;A7-17 对应于序列表中 SEQ ID NO: 2 的 1-6400 位核苷酸的多核苷酸序列 ;和 A7-18 对应于序列表中 SEQ ID NO: 2 的 1501-7400 位核苷酸的多核苷酸序列。

[0242] 关于 A18 的各个片段的多核苷酸序列 :A18-1 对应于序列表中 SEQ ID NO: 3 的 1-5040 位核苷酸的多核苷酸序列 ;A18-2 对应于序列表中 SEQ ID NO: 3 的 1001-6002 位核苷酸的多核苷酸序列 ;A18-3 对应于序列表中 SEQ ID NO: 3 的 2001-7000 位核苷酸的多核苷酸序列 ;和 A18-4 对应于序列表中 SEQ ID NO: 3 的 3000-7000 位核苷酸的多核苷酸序列。

[0243] 关于 B5 的各个片段的多核苷酸序列 :B5-1 对应于序列表中 SEQ ID NO: 4 的 1-4001 位核苷酸的多核苷酸序列 ;B5-2 对应于序列表中 SEQ ID NO: 4 的 1-3200 位核苷酸的多核苷酸序列 ;B5-3 对应于序列表中 SEQ ID NO: 4 的 2491-5601 位核苷酸的多核苷酸序列 ;B5-4 对应于序列表中 SEQ ID NO: 4 的 5373-8401 位核苷酸的多核苷酸序列 ;B5-5 对应于序列表中 SEQ ID NO: 4 的 901-4001 位核苷酸的多核苷酸序列 ;和 B5-6 对应于序列表中 SEQ ID NO: 4 的 4001-7000 位核苷酸的多核苷酸序列。

[0244] 关于 C14 的各个片段的多核苷酸序列 :C14-1 对应于序列表中 SEQ ID NO: 5 的 960-4015 位核苷酸的多核苷酸序列 ;C14-2 对应于序列表中 SEQ ID NO: 5 的 1987-5014 位核苷酸的多核苷酸序列 ;C14-3 对应于序列表中 SEQ ID NO: 5 的 4020-7119 位核苷酸的多核苷酸序列 ;C14-4 对应于序列表中 SEQ ID NO: 5 的 960-8141 位核苷酸的多核苷酸序列 ;

C14-5 对应于序列表中 SEQ ID NO: 5 的 960-6011 位核苷酸的多核苷酸序列 ;C14-6 对应于序列表中 SEQ ID NO: 5 的 4939-8141 位核苷酸的多核苷酸序列 ;C14-7 对应于序列表中 SEQ ID NO: 5 的 960-5014 位核苷酸的多核苷酸序列 ;C14-8 对应于序列表中 SEQ ID NO: 5 的 2994-7119 位核苷酸的多核苷酸序列 ;C14-9 对应于序列表中 SEQ ID NO: 5 的 4020-8141 位核苷酸的多核苷酸序列 ;C14-10 对应于序列表中 SEQ ID NO: 5 的 1-5014 位核苷酸的多核苷酸序列 ;C14-11 对应于序列表中 SEQ ID NO: 5 的 1987-7119 位核苷酸的多核苷酸序列 ;C14-12 对应于序列表中 SEQ ID NO: 5 的 2994-8141 位核苷酸的多核苷酸序列 ;C14-13 对应于序列表中 SEQ ID NO: 5 的 960-7119 位核苷酸的多核苷酸序列 ;C14-14 对应于序列表中 SEQ ID NO: 5 的 1987-8141 位核苷酸的多核苷酸序列。

[0245] 各个片段在全长序列上的起点和终点示于图 18 和 19。

[0246] (5-2) 评估具有不同序列长度的 DNA 元件

使用宿主细胞 CHO-K1 (ATCC) 和转染试剂 Lipofectamine 2000 (Invitrogen) 来评估在 (5-1) 中构建的每个质粒。

[0247] 以与 (2-3) 相同的方法，转染后用潮霉素进行抗生素筛选，藉此建立稳定表达的多克隆细胞系。在测量前一天对由此建立的细胞系进行培养基更换，将给定数量的细胞接种至 24- 孔板。细胞铺板 24 小时后，收集培养上清液，测量 SEAP 的活性。

[0248] 测量结果如图 8、10、12、14 和 16 所示。证实不仅全长 DNA 元件，而且具有比全长更短的序列长度的克隆亦具有提高表达的作用。基于该结果，证实 DNA 元件 A2、A7、A18、B5 和 C14 即使在其具有比全长更短的序列长度的情况下也具有提高外源基因表达的活性。然而，当序列长度为全长时它们表现出最高的作用。

[0249] 实施例 6

使用除 CHO 细胞系外的宿主细胞的作用

使用 CHO 细胞系作为在实施例 2-5 中评估的细胞系。然而，在实施例 6 中，选择 HEK293 细胞系作为除 CHO 细胞系外的细胞系。在 37℃、5% CO_2 存在下，用含有 10%FCS 的 DMEM 培养基 (Invitrogen) 对 HEK293 细胞系进行静置培养，在转染前一天向 6 孔板中接种给定数目的细胞。为了评估在 (3-2) 中构建的含有各 DNA 元件的 SEAP 表达载体，使用各质粒和转染试剂 Lipofectamine 2000 (Invitrogen) 实施转染。转染后 2 天开始进行约 2 周的用潮霉素的抗生素筛选，藉此建立稳定表达的多克隆细胞系。在测量前一天对由此建立的细胞系进行培养基更换，将给定数量的细胞接种至 24- 孔板。细胞铺板 24 小时后，收集培养上清液，测量 SEAP 的活性。用 SensoLyte™ pNPP 分泌型碱性磷酸酶报告物测定来测量培养上清液中的 SEAP 活性 (ANASPEC)。

[0250] 测量结果如图 17 所示。以与实施例 3 相同的方法，证实在 HEK293 细胞系中每个元件在提高外源基因 (SEAP) 表达方面亦高度有效。

[0251] 工业适用性

通过使用本发明的 DNA 元件将外源基因表达载体引入到哺乳动物宿主细胞，使提高治疗性蛋白、抗体等的外源基因生产率成为可能。

[0001]

序列表

<110> DAIICHI SANKYO COMPANY, LIMITED

<120> 具有提高转基因表达作用的 DNA 元件

<130> DSPCT-FP1124

<150> JP2010-154782
<151> 2010-07-07

<160> 7

<170> PatentIn version 3.4

<210> 1
<211> 8450
<212> DNA
<213> 人

<400> 1
attttgcttg aaaggatagc atcaaggaag tgaaatgaca acccacagaa tgagagataa 60

tttttgcaaa tcatgtatct gataagggac ctgtagtcag aatatgcaaa gaacccttac 120

aattcaataa gacaacccaa tttaaaaaca ggcaaaggat gtgaataggc atttctccaa 180

agatacggaa aaacggccaa taagcacata aaaagatgct caaaatcatt tgccatttgg 240

gaaatgcaat caaaaccaca atgaggtatc acttcacgcc cattagggtg gctatagatc 300

agaaagtcag ataacatgtg ttggcaagca catggaaaca ctgaagtcct tacacactgc 360

tggtaggaat gtaaaatggt gcagccactg tggaaaacag ttttccaatt tctcaaaatg 420

ttaaacacag ttatcataca cccaagcaat tctactctta ggtatatacc caagagaaat 480

gaaaacatat gtcttcacca gaacttgctg ttcacagcag cattatgcat aatagaccaa 540

aagtggaaac aactcaactg cccatcaact ggtgaatgga taagtaaaat gtgatgtaac 600

cagtcattgg actgtcattc attaataaaa agaacaaggt actgattcat gttctaacat 660

gagtgaatct tgaaaacact atgctaaatt aaagaagcca gtcacaaaag gccgtgtatt 720

[0002]

gcatgatttt atatatacat gaacttttat atatatataa ttatatatat tatatataat 780

tttatatata taaatttcta tatataaata tataaaatca tatatatgat atatattttt 840

tcatatacat catatatatt tacaaaaatt atatatcata tatcatatga tatatgagat 900

atatatcatg atatatatga tatatgatat atatcatatg agatatatga tatcatgaga 960

tatatgatat catatgatat atatgatata gatatcatat gatatatata taatatatat 1020

atgatagata tattatatat gatagatatg atagatatca tattatatat gatagatatg 1080

atagatatca tattatatat gatagatata gatatcatat tatatatgat agatatgata 1140

gatatcatat tatatatgat agatatgata gatatcatat tatatatgat agatatgata 1200

gatatcatat tatatatgat agatatgata gatatcatat tatatatgat agatatgata 1260

gatatcatat tatatatgat agatatgata gatatcatat tatatatgat agatatgata 1320

gatatcatat tatatatgat agatatgata gatatcatat tatatatgat agatatgata 1380

gatatcatat tatatatgat agatatgata gatatcatat tatatatgat agatatgata 1440

gatatcatat tatatatgat atcatatata taccacatac atcatatata catcatatat 1500

acatcatata tatcatacat atatatgaac tttccagaat aggtatatca ataaagacag 1560

gaagtataca agtggttgcc acagcctgag aggagcaggg aatggtgagt gactgctaat 1620

ggatatggca ctttttttgg ggggtgatga aaatgttctg gtcagacaat ggcaattaca 1680

aaactgtata cacacgaaaa accaaagaat cacacacttt aaaagggagg atttagctcg 1740

gcatggtggc atgcgcctgt actcccagtt actcgggagg ctgaagcagg actgcttaga 1800

gcccaggact tcaaggctgc agcgagctat gatcgctcca ctgcactcca acaaggatga 1860

cagtgcgaga cccgttttct aaataataat aataataata ataataaata acccaaggta 1920

cccagttcac atgcaaaacc actggtaaac ataaattatc tccaagtaat ctagaaagaa 1980

aatgagcaca taagacgtct tctaaaaaca cacatatatt tctttacatg ttacatttaa 2040

[0003]

cgtaaaaatc agctatgcag aagttacatg aacattttat gttggaaagg taaatgacta 2100

ttattaatac agaatggtta agtacattta tgtttttatg tacaaacgca taaaaggaaa 2160

agcatcctta aaataaacac catcaatggc tcctcggtgg tcacaaaaca aaatcctcac 2220

acctttgtct tccttcacaa ttgagcttta tccacctttt caggcttatc tcccattatt 2280

acctgacaca aacttgggtg ggccagagtt tccactgacc atcccccgac tattcatcca 2340

acactatgtt cactgcctcc cattcctgac catttgcctt ttgtcttcaa ctaattctgg 2400

ggacgttttg tccaaataaa tgatccatat tcttgaaggc tggaatcaag tcctattaca 2460

aatatatttt ctcaccctct ccagagcata gcaacccagc atctactggc ctctcacagc 2520

tctaaccatc cacaacccta agctggcttc tcatcaaacg ggtacttttc accacccaaa 2580

ttcaattaat tcactcttac aataatgaag aatagtcgcc tacagcctac cttttccagc 2640

cttgattcaa tcatttatca attttatctt caaagtccct cacttcaggg agatgatata 2700

tcagctttca cccagagtcc taaagaaaac agcactcttg ccaatgacat agtgccacct 2760

agtggcaaca taaggtaaat cacagtggca gtagaaggat ctccacacta cttttacagg 2820

aatgcactgc aggtaaaaaa taagaagcta cagtactgtt tggcaggaca atttgtttca 2880

tacgtgcata ctatcgccct gactaaatta actcgcaagt cttacaggta ttatttgttt 2940

tcagttccat gcacagatta gccatttagt acttactaaa tcaaactcaa tttctgaagt 3000

gtcttacacc aatatattca tgcacatatg gttaaaattt tccttgagga tctatcatgt 3060

gagagtgtgg cttattataa caagtaaaca gaacaaataa atacaaaatg aaaagaaatc 3120

gtatgattta ctcgcatata agggagcttg ttgtggatta agtttcatga cccaggacac 3180

tgaaacagaa atggaataaa tgagaataaa attaaaagtt gtcatcaaaa atatagaagc 3240

catctaaaga cctaggtgtc aagcatagct ctatgagtac aatcccgtgc ctgagattac 3300

catatgccca gctgtatgct atacactaag agatttagga aggaagcggg gtcagggatt 3360

[0004]

gaccccagac tccatctttt caagtgggga agaaagatct tccgattgaa aaataaaggc 3420

aaaaaaggct tcaccgtcac agaagtttca acaaccaaca ggatatttaa aacagttatc 3480

aaagcaaaac cattgtatgt tcacttacat ttttacatag tccctcaaac tcacaaaatg 3540

ctgtttactc agggacttct tccggtctta ctagggagcc tggaaagtga cgggaggatt 3600

gcaagggacc actagaaccc tcttcctcaa ttccccttct ctgagaaggg aggctacagc 3660

ttgcctctct aaccactaaa aggcatgacc ctcctcaaag ttaatagccg gattccctga 3720

tagatatttt cactaaatga attctcataa aactctcact aagatttaga gaaggcttcc 3780

agggttgaat tcctgaacat taagaacagc atgtttttta aaagtttaac ttggtgattg 3840

gaccaggact tcatctaggc tatgaatgct cagaatggta ggtcctttac caaacagctt 3900

gagtttgtgt ataaagtgat ctcatcctct taagagtcag agaaacagaa ccaagcgact 3960

tcactataat ttgatctgag gaagtttctt actcacaata ggtaaatgaa ggcacatact 4020

aaccagcaat ataaacaaca atatcaagtg tcattcacac atgcaaaaaa cagacaaaat 4080

cccaaactct gtgttctaac aaatcgcaaa aacctcacta acaataaatt gaaatgacca 4140

aatgtttgga ctgaaaagca atgccttggt agcctagcca tgcctaactc aaataacaga 4200

accatctcga tgttaaaatc ctcacagatc aagctgtgta tgtctcgggt caagacttcg 4260

ccaaaaagca gtgagcacac acttaagagg gaaaaaatct acctcagcct cctaaatgca 4320

atcatctcta cacgagttgc aggccccaag cttcaacgtg ttctgctgga caacgcagta 4380

gaaagctgac aagcaggtgg ccttcccaca ctgactgaac cacctccatg cccatgtcca 4440

ttcattttct tgcccacccc atgtgctata acagacctcc tggctcaggg cactctttcc 4500

ttcctgactg ccttcactta atgactttgt acttttaggt gcaaaaatta tctgcagaaa 4560

tccacactga aaaccaagct tgagaaaggc agcaataacc aacattttta caagaagaac 4620

aaggtcaata tcaagcccat cagattcaaa tagcaagcat ggatgaaaat gaaagattga 4680

[0005]

aaggcttgag tgccttctta atgtattaaa tatccattta atttacaatt aagctcactg 4740

tgctcactgg ccttttaatc agctttccag gtcctgctca gacttgccta ggacatggga 4800

atgaaagaac ctatacattt atggaccaat ctaccttaac taacttgtca agtgttcctg 4860

catcaagcag aagaaacatc agtgaaactg atacaggaat taacccccttg ttaatccata 4920

aaacttaaag gagcgggatc caatcttctg gcttccctgg gccacgctgg aagaagaatt 4980

gtcttgcgcc acacataaaa tacacgaaca ctaataatag ctgctaagct ttaaaaaaat 5040

tgcaaaaaag gaaaatctca taatttttg tttgttgtga ggtggagcct cactctgtca 5100

cccaggccgg agtgcagtgg caccatcttg gctcactgca acctctgcct cctgggttca 5160

agccattctc ctgcctcagc ctcccgagta gctgggatga taggcgtgtg ccaccatgcc 5220

cagctaattt tcgtattttt agtagagacg gggtttcacc atgttggcca ggctggtctc 5280

aaactcctga cctcaggtga tccacccacc tcggcctccc aaagtgctgg gattacaggt 5340

gtgagccacc gtgcccggcc aatgttttaa gaacgtttac gaatttgtat tgggccacat 5400

tcaaagcctt cacaggctgc atgcagcctg caggccgcgg ttggacaagc ttggattaga 5460

gaaatctaca gagacaaact agtgacttag tagccctctg atagctcatg atttgcaaga 5520

aacttaggat gactatgtgt aaagaccaca aacatcaatt taactgaatg gttcccgcca 5580

cactggaatg aggaagctga gcaaactcag aggactctaa gaaagggctg atgtcatctg 5640

aactgttcgg aattataaac tcctctaaac atgtttcaaa gccagaactt gtaggagttg 5700

ttctgataca cggattaaaa gagggatgac aaagtgtctg tcccccacac tggtcaaagg 5760

gacaggtcat tgttatgctg gcaatgcagg ctgctgaaaa gaatgtatct gtcaaaagta 5820

atcaaagtaa tgaccccaga aggctccaga aacagactgg taaattcagg ttgctttcag 5880

acttccacaa tgctggcaca caaggggaaa gacaaaacta acatttacag agcattatat 5940

ttgatattac atttaatccc cattaaaaag atactatttc ccgtttcact agtgaaaaag 6000

[0006]

ttgatctttc aaaggttaaa ttatttaaca ccaaggtcaa agggtaagtt ggagagacca 6060

gattcaaacc cagtctgaca ttaaaacatg tgttttcccc ccacatcgtc tcctgctaat 6120

aacctcaaat ctaaaaactg acttgcccta caccttgagc cccatcctac aaactctccc 6180

tgacgttatt aattcagctg tcactgtgca cctacaacgt gccagacacc atactcctca 6240

acactctgta ggcacagaag gaacagataa aaatccctac cttcatagat attattctag 6300

gggtaacaca ggtaaataaa acattaaaat agttttcaca tagtagcaaa ttccatatag 6360

caaaataaaa cagaagaagg aatagcaaat gagggagatg ccctcttaaa catggtgctg 6420

agggaaggcc tccctgagaa agatatcatt taccccaaaa ataaaaaagc aagtaataga 6480

aaaaacaggt aaaaggtgtt ctagacactt aaacctgcca cattgagaac tcagggttct 6540

gatgcaaaac ctcgctgcat agaatgcatt aacttatttt tatacattta aacaaacaaa 6600

ctctacttaa gaactgtgtt ctaaaggaag gagcatatta caggaaggca atttttggtc 6660

agagtagaca cacttaaaaa ctaaacctat tgaaagacca agaacaactg aaagtctttg 6720

ctttgtcaga tttttgacca aaaggaaaat taaagaaaca caccgtgccc atccaatgat 6780

ttcaccaagg aattttaaga gagaaaatcc tacttcttcc tcacccagta gccagtgaaa 6840

tgactgagca aattcacaag ttcactgggg ctgctttcat gtaacacagg gacaacacat 6900

gacagacaca gtggaaccct acaggttgcc tagtatttga aagactgtga agaggaggag 6960

atgtcaaaat tcaaagtctt aaatgatgta gttttaagta tgttcagcaa tttcaccact 7020

cagtagtaaa gccagctaca gttgaaagca atcagaaatt tgaggggtgt gaaataagca 7080

gaagcacaga agttaaggat ttgtattctt cccacatttt ccactttatt ttatactgct 7140

gagaaaaaac aaatttaata gttttctgct gtataagaga gacacattca ctttatgtca 7200

cagtaagagt cactcaattt taatacaact atctcaatgt ataaattaac attctccccc 7260

ctgcccacac atagtaagtc tcttatgatg ttgctgatta gagaagcaaa agttgccgct 7320

[0007]

acaattctct tcctgcattt taatataaac aatcatcagt cttttcttca tagagtgcag 7380

tgtgggcact atcatcagaa tgtaccagca ctgggtgtgc aaagtttaca aagattagca 7440

agagcaaaag tgttgagatt tttgaaattc atgctgctgc aaagaagtat gtaaaaactc 7500

actcaccata gaggaccaca cagaaactca ggcatgaagt tatatggctg tgtgagtggt 7560

ttgggagaag gaacggaaag cacttccacc aacctatatg cctgagcaaa ttaatgcaaa 7620

acctcagaag ctacaaaaaa gtttatctac ctaaattaaa attggtgtcc acagcagtag 7680

ccagcaaaat gcctgcgaag cgcaaagtgg taaatatttt agggtctgta ggtcatatgg 7740

tctctgttaa acaatatgta aatgaatggg tgtggctgtg ttccaataaa acttcattta 7800

taaaaagagg cagcatggta catccagtca gcaagctata atgtaccaac ccccggtcta 7860

acactaacca aatacctctt aataagccaa agaaactgtg tcctcttagg ccggaagcgg 7920

tggctcacac ctataatccc agcattttgg gaggccgagg cggggagatc acctgaggtc 7980

aggagtttga gaccatcctg gccaacatgg tgaaacccta tttctactaa aaatacaaaa 8040

attagccagg cgtgctggcg ggcgcctgta atgccaacta ctggggaggc tgaagcacga 8100

gaatcgcttg aacccaggag gcagaggttg cagcgagcct agatcacgcc attgcactcc 8160

agcctgggca acaagagaga aactccgtct caaaaaaaaa aaaggaaata aaagtataca 8220

aagtgaaaac aaagaaatta aactgccctt atttgccagt gacattactg tctatgcaca 8280

aaattccaaa aatctacaaa aaagcttcta gtactaaaaa tgagtttagc aaggttgtag 8340

aatccaaggt cagcatataa cataaaatca ccttcctata tactagcaat caccaactgg 8400

aaattgagaa gtatcattca caacagtacc acaaacatga ataaatgtg 8450

<210> 2
<211> 8420
<212> DNA
<213> 人

[0008]

<400> 2

tcttagtatg gtaaaccttt tgaagtagat tcaaatgaga atgggaagag agaaaaggga 60

gagaagcaac ataagaaatc tcttttaagg aattttatat agagagaaac agaggaatca 120

gttgatagtt ggaaattatt ttaaagaaaa tgggttattt taaagaaaaa aggtattaca 180

acatgtttgc actattgtgg gaataatcaa gttgagacag aaaattattt tttaaggaag 240

agtctaattg ctgaagtgaa agagaatgaa tgagaccctg tgcataagtg tgatcagata 300

ggagcatgta cagctcaagt aagaacagga agaaagagac aataaacatg tacagatagg 360

atgggctggt cgatgtggtg gtgaaaagac atgcgagtta ttactgatta cttctatttc 420

cccagtgaaa taggaagcca ggttcataaa ccaaaatgaa gaggagcgag gcagtattgg 480

aagttcagga aaagtaatag gtgtaaaaat atgtaaagta gaattaccag ggagtatgaa 540

gatacatttc caattaagga tgaagaattt aaagtgaggc cagccaatac ccctgctttg 600

cttcagctac atcagctgca taggttcagg cacagaatac atggaacatt gtatttaaat 660

agggcctgga ttttacaaaa gtaacacaat gaagaagaga gatgcaaggc tatttgaggg 720

tgtttgtggg agagattgta aaatattagc taagtaagaa ggggactgca aattttagtg 780

gtataaagga atgaggaaaa gtgtaaatac agtggggtca aagaatgttt ggagccaagg 840

cactagaggc aattagctga aaatgtaggt gattattggt gagtgacatg gtttaaatga 900

aaagtataga agggtacaat tatccatcat gaaaagttct agggtacaac taagatctga 960

gtagctgaag tagaatgaaa gtagaatgga cctttccata tccagccagg ttcagtgaca 1020

gaaggttagg aaacaaatta taaaccactt gagagaacat atcccctaag ttgtttttgc 1080

tatttttctt tcagcatata tttgttggaa tgccaactat gttcagttca attaatatgg 1140

gcttcttaaa taagggctcc agcactggat aatcctgcca tttattttga tacattccat 1200

cctgctgctc agatctattg gcatctacag gatgtctttt gagaagatgg gcattcacat 1260

[0009]

ccctatgtcc tagcaaattt ccaactcaga aaaccacatt aggcttctct atatatcttc 1320

caactatttc aatggaaaat acaattctct gatttcttcc tatgatattt atcaaagaga 1380

atggtgcctg ccagttctag ggtgggggaa ctcaatacaa atcaccaacc tttagatgac 1440

accctgtctt caaagtgctt tcaaagtctg gcagaaaaaa agtacccagt ggctataaga 1500

ccacccagga gttcagtcat gcattctaag tagcagatca ctggaatgta attggctagt 1560

gagttcattt tactcttctc ttcttggtca catgttaccg cccttgtacc ctgcacgttc 1620

tctttcccag acttacaaag catgttctct tgaattcgtt ctcttttaa attcacacag 1680

tcttaatgat tcttctttca caagagtctt tcactcttac aattcagttc aagtcatcca 1740

catgcttatt atgagcaagg gtctgggact taggggaaaa gggaataaaa agatgaatga 1800

aatgtgatcc ctgcagtcca agagcttgct gtgaaaaagg aagtttggct tacattgcct 1860

ccctaatccc ttggctaggc cagaacagaa tattgtctaa aacctcctca cgtcagcagt 1920

cctctggggt ggtgactgga agtagaattt aaacaaaaat ataattgaca cataataatt 1980

gtgcatactt atagggtaca atctgatgtt tcgatatgtg tttaaatggg tgcattgtgt 2040

aatgatcaaa ttgaggtaat ttatccacca ccttgaagag agatttttca atattctcat 2100

tgcgaagaag caggaatttt tagcagacaa ctgagatgct tcttgttcac actaagtcat 2160

tctgacgatg gatttacata acttgttgtt ttttttgtgt gtgtgttttt gagacagagt 2220

cttactttgt cgactaggct gaagtgcagt ggcacaatct cggctcactg caacctccac 2280

ctcccgggtt caaacgattc tcctgcctca gcctcctgag tagctgggat tacaggtgca 2340

tgcaactagg cctggctaat ttttatattt ttaatacaga tgggatttca ccatgttggc 2400

cctgctggtg tcaaattcgt ggcctcaagt gatctaccag ctgcggcctc ccaaagtgca 2460

gggattacag gtgtgagaca ccaagcctgg tacatttaca tttcttatct ggatctttcc 2520

tttagtaagt gctaaggaat cctacttccc ccaatatttt ttcctatttc aatgttttag 2580

[0010]

catgtatcat gttactactt tgcagacatt tgattttccc ctttgtttac tgtaaagtat 2640

atttttatag cctttgtaat agaagtattc taaaatctgc ctgcaaccta tctttctgac 2700

tctgcatttt agggaataat tctctgttgt ggaatgaaaa aaaaaacaga gcctgtggag 2760

tcagagatct catttcaaat tatagttatc cctaggaata aatctgagtg acaggtagta 2820

tagtataata ataagtataa agctatggtt aaggaaaact caacaacctt atctgtaaat 2880

tgggatgaca acagcctacg tcaaaaaaat gtgaaggtaa atgagataat gtaaggctga 2940

tacttagtaa gcaatttaaa aacacccaaa aaactattgc catgattact ctacttactc 3000

tatttctcta tgctccaggc aaatgaacta ctaatgaccc aggggtcctt ccccattctc 3060

ttcttcacaa ggaaatattc tctctctgtg tgctgtttat taaaatctac tgccccttt 3120

agaagccttt ccagatcatc ccatggccaa gaacgatcgc tgcttcctct tctttacata 3180

cagatgtttt tctcctgctt gacaattatt tttgtgcaat tattttcctt ttgattgtgt 3240

ttttaatgtc ccccccaccc cacaattttc cagactgttt gctccacgag agaggagacc 3300

atcatctctg tgctcaccgt tgtatgacca gtatcctgag gagtggctgt tacataatta 3360

catcaggcac tcaataaaaa tttgatgaat aaacactgga ttttaaggca ggtatcatat 3420

cttacatagc atatcatatc ttacatttta tgtccctcac ataaatacca cagagtgaag 3480

tatatgacag ataaggtcat ttctcttgat aagtacatag tccagtctga aacagatatg 3540

ccaaaaaaaa acaaaactgg agtaaacaag atgaattgtt ttaatagagg cactgtatta 3600

gtttcctagg actgccagaa caaatcacct caaacttagt ggctgaaaac aacaaaaatt 3660

tattgtctca cagttataga tgttagaagt ataaaattaa ggtgtcagtg ggattggttc 3720

cttctggggg ctgtggaaga gaatctgtcc caagccttca cactgtaaag tacagtactg 3780

gagggatagg acttcaactt gctctatctc agatagagag gagccatttg ttgtgaattg 3840

agaagagggg tatgttgaat ccataataag cacataaaaa cttggctggt tcataggaga 3900

[0011]

agtaacatgt ttccagctct agtaaaaaac aaattgaagt ggcctataaa aaggtacaga 3960

gtacgacaga atgaaaaata aatgaacaag aatacagaga ggatgtggta aattatcatg 4020

tttccctaat atgttattgg acactaaatg gtattagaat tatttatcaa taataattct 4080

aaactgttgc aattgaaaga atatattaag tggtgttata tgagaagtgc cagggcattc 4140

tcatttctgt ccaatgggag aaacattttc gtttgagacc tccgtgaata atacagtctt 4200

ttagttagga gagctgcatt ttgagtggtg caggcagaat ggcgatctct cacccacaca 4260

aacactaaga tagagagaga cagagacaga gacagagaca gcagagagag acagagaaag 4320

gaagtacagg tactcagata gagataagcc atttcttgac attaagaaat aaagtagaat 4380

ccattggagg gaaataaaac tgcctcagga acagagttaa ttcacataca catgcaggta 4440

aacacacact gcttgatact tactgtggac tttgaaaatt atgaatgtgt gtgtgtgtgt 4500

gtgtgtacat tcagccctcc atatccatgg attttgcatt cacagattca accaaccatg 4560

aattaaaaac atttggaaat aacaaacatt aaaatataac aatacaacaa taaaaataat 4620

acaaataaaa aatatagtgt aacaactgtt tacatagcat gtatgttgta ttaagtagta 4680

taaatctaga gattacttaa tgtataccag aggatgcata ggctatatgc aaatactatg 4740

ccactttaaa ctgataagaa cagatactaa acttcatctt agccaaaagt cagagaaaca 4800

atataactat gccatttttac ataagggact tgagctgagc atcctcagat ttcagtatct 4860

ttggagttcc tggaaacaat tccttgtttt atatatatat atgtgtgtgt atatatatat 4920

atatatatac acacatatat atatatatat atatatgata gctactgagt gacaggtgat 4980

attataccat accacttgtc actcagtagc tgtatatgca tatgtatata tatacatata 5040

catatatgtg tgtatgtgta tgtgtgtgtg tgtgtgtgtg tgtgtgtatg ctgtctttcc 5100

tcggtatcac agggaattgg agatatatat attctttca gtacaaaaaa aattgaacac 5160

agatgggtat ggtaccagaa cagaaggtaa agacacatga aaaaaatttg caacaacatg 5220

[0012]

aatggaactg gagatcatta tttgaggaga aataatccag gcacagaaaa acaagcattt 5280

tattatttta ggtgaaagac aaacatttta ttttaggtga aataatccag gcacagaaag 5340

acaaacattg catgttctca tttatttgtg ggatgtaaaa atcaaaacaa tagaacgtat 5400

ggaggtagac agcagaagga tagttaccaa aggctgcaaa gggtagtgta ggctttgagg 5460

gtgaggtggg gatggttatt gggtacaaaa aatagttaga aagaataaat aatatctagt 5520

atttaatagc acaacaggtt gactatagtc aaaataacat aattgtacaa tttaaatatg 5580

aaattaaata tatatacaag actagaacac caagttgaat gactccagct tgcgaaaccc 5640

acattgatca ccatgcttgc cccaagggaa gctgtacaat gtctggctcg tccagaaccc 5700

catcatttat cactagcaat ctattgtcca taatcatgtt taaattaata gcattttaaa 5760

ggtacaaata ttttttaaaa aacaaataat tatttaattc gcctttaaa agcttttaa 5820

aaacgttttt aaaaactttt ttaaagtcct gaggactatt ttctttaaag tgctcagtta 5880

cagagctcca tatattgggc tatgatagcc ttacctgatt cttgccaaga atctagtgcc 5940

cagaaaatgc aaatacaaag taagcaactg aaaaataaac aaataagttg gaggtatgct 6000

acctgttgaa atatgaccta gcgcaaacac ctatgccact tgcttatgaa atcatatagg 6060

ttttcggtgt gcagttttga ctgaatgagg gagtttacgc tggaccacaa gggggccccc 6120

ctgtcaataa cgtactccat ttgtgtatta agtcaaaaat gaaatggaag agaaaagaaa 6180

catcgatgac cccaagtctc tttaattgaa tggaggtaaa agggaaacaa cgaatgagaa 6240

aagtactctg ccctttaag aatcttgcat tcacattcct gatgaagtta tttttcctcc 6300

tctcactgat tcccatttca ctctattaca tagcaccgtg ttccccagga gctcctgaat 6360

gaaggacatc actcagctgt gttaagtatc tggaacaata aatatactag tttcaatgtc 6420

taggctatgg gtattccttt ttactgaagg tatgacatat agctgcccag gcctgactaa 6480

attaatagta ataataatta ataatggcaa atttttattc tattaagtta cttggcttga 6540

[0013]

cttgtagaaa tagcaacatt catctgaaat gccccctcct acacttatgt ctaaggacaa 6600

atcccacata caccacagat aacttcattt tacatgtttt attctgttac caaactaaat 6660

ttttatcata tagtctgttg ctcactgaac tcttcagtaa ttctcaacat accatgtaaa 6720

gcattaagca cagttccaac acagagcaaa tgagcaataa ctgttagtta ttataacatt 6780

attatgtgtt ttcagtgcat taaaccactg gtctgatacc tagcccaaca ttctattaaa 6840

ccacataatc cagttgaata atatatgata atataataaa atggcgataa gtgctaaata 6900

tccagataga aacacagatg gaatcagaca gctttcccaa gaaatagaga aaatagtaga 6960

taggcgatct aggcctaagc actctaagca gaagctaagt tatcacagga tatcttggca 7020

atctgtggca cgtgaaccct tttcttctgg agtctggaac tatgttgcaa ctctcacttt 7080

ctccctatct agagactcag tttgttccct tgtgattatc agcagttgag aaatccttag 7140

accttctgaa aggactactt tttaaattta tatatataat atttaaaata catatcttta 7200

tatataatat atatttaaat atataatatt taaattaata tatatttaaa tatataatat 7260

ttaaattaat atatatttaa ataaataaat ttatatttaa atatataata attaaaatat 7320

atttttaatg aacagagagt aaaggattat tttgaagaga aactcctggt tcccacttaa 7380

aatcctttct tgtttccaag tttttcaaat ggagccctct ttaccagctt gccccctcag 7440

agataagctg ttcccctact tattcagatc tgagatctga aaacattcct tttcctgtga 7500

gttcagctag gacaaagatg gagctttttg ataaaatttg gcaaacacat tttttaaaga 7560

tgaaaatttt taaaaattga aaaaaaaaca tttatagaaa gagacttcta atccaaattt 7620

aacttctcaa actatgtttt gaccggctag cataatgttt cagtctttct ggagaatgcc 7680

ccttgaaact gttttcttct acacaacttc ctcctttcct ttgactttcc tgctctggaa 7740

gggaagaaca ggaagaggac agatcaaatt actcaagagg aaggacaaga aataaggaac 7800

caaattatca acaattggag aaagaaagct gatgtcagta tcatttcata tatgattatg 7860

[0014]

tcagagtcag gtggataagc caatcctgtt gaatagcata cttttcctgc tactcctgaa 7920

gggtaaagag gtctttctct tacaaagccg tcctagctag taatcttaca ggtgcaaaaa 7980

gcttgttttc atgttatttc ttagtaactc aaaatacctc taaagttata catattatga 8040

aagtactaca gtcacagtgc tgagaaaagg agtaaataag acaatgtata taaaaacact 8100

tggctcagcc cctggctctg tggttgataa atattaagtt agtattcatt attattataa 8160

tttccaaaga gtccattaaa agatatagaa gaagggaggc agcaataaca ctaagagaaa 8220

attccattat ctccaactat ttatcctcta gcccaaaata attgccatta gaaagagcaa 8280

ctttaacaaa aattttaagt tgcaatagat gttcaacttt aaatccatcc cagaaaaatt 8340

tctaaccaaa ggagcataga agatttgatc ttattttcta agtagtatag acttaattgt 8400

gagaacaaaa taaaaacttg 8420

<210> 3
<211> 8475
<212> DNA
<213> 人

<400> 3
gcataacttg taagaaatgg agtgaggtct cagttcaaac tggcttctgt atgacttcaa 60

agccaaagtc agcaacttag aaggcaaaaa ttataattta gttggcaaat acgagaaaag 120

gtcagaaaca catgaaatga agctcaatag gaacacttac agggtagcag ggtagtagcc 180

tagggaaaaa agtcagacac taaaattgtt taaataggta agttcaaggg acaggtaaag 240

accttagtgg gtaagaagcc aatcagcaga cgaactgcaa gcaagcactg tctctctttc 300

ccttctgtct cctcttgtag taactgacca caattaaggc tgcctagggg aataatgaag 360

taatcctcct attatcagca atggtctgat ccagtgccag gcaccacaga caacttggtg 420

ttcagagaag atccttcaag atgaacaaag ggtcaaaata aaaaattcta gaagagagaa 480

gactgatcac aatttaatgt aaggcttgga aggaactgat ctctaccttc cttaacatct 540

[0015]

caagaacttc ctcagattca ttggatgttg agtgtgtgtg agtctagtag aaaaatgaat 600

ttttgtttct taacttggat atgtgattag gatgttaata attaagtctg ggctaatatt 660

gaaggtatct tatgatgggc ttcttaaagc attgatcaca aagactgcat gttcataaac 720

tgagctgcac ttgttaggat tctagatgtt tgaaatttct tgtgttattt tggtctcaga 780

tttctagaca aattttctca aattcctatt tcacttttg acatatcatg agtgactcaa 840

atgtttgccc ttgagtcgga aaacacccag cattaggaat aggcacataa acataatact 900

tcaagcttca gatttaagct caattataaa gtgtttaaag gctgtgctga tagttcttct 960

gagtagaatt cctacaacta tgggtttgtc tataataaaa tgttcactct atattgaacg 1020

ccttatttaa aactcgaaat gtgtaagtag taataaagaa aatatgtcct cctgtaacca 1080

aagctaggac cgattacatg ttcacttgac tgacagatac aatcacctat attaggagca 1140

atcagcactt ccttacaaac taacaacttg agatgtagtg ttcccattgg ctatgaagat 1200

tttctttatt tactcagaat agtctgtagg atctgccagc tgcccctgat tataccagct 1260

gcacccaatg atcacagtga acattatttt acattctaaa taactggtgc aaggtgagcc 1320

atggttttct gagtttccta tcacctttgt gtttcaggtc ctcaaatgtt aatttgtaaa 1380

gctgctgttt caggcaaaac taacaaaatt agcatctaat caataaccat actatgtcca 1440

cccatatcct ataacacaga agtaggggaa gagtgagaaa ggtggaagtg gagaaataga 1500

ggcccaaaaa gaaagtttta tcacaggaat atctagatgt cttctgggat tgtctgttaa 1560

agagctgtga cactcatata aatgcagaat tactctcttt cttccttgtt ggttagaagg 1620

ccaagggtgc catggtaata ctaccaaaca tatatcaaag cttggcagga aaaatggtac 1680

cttcagaaat tttataatct gatatcaaat aggtcaagaa atataataaa actagtttct 1740

ttggtttcct tagaaacctg gaaaacttta aattagaaac ttagaaagct ttaaatcaga 1800

ctttgtagtt aaaaaaggaa attttagttc cttccagcat tagaattccg tgattctctg 1860

[0016]

actctgagcc tggattaaat ctagcccagc tgagtggaaa cttaagtaac tagctggttg 1920

cctttagtga tcttccactt tatggctgct tccgcctaag aagttcatca tcgtgactta 1980

ctttctttgg ggcaaagtcg tgactaactt tctttggggc aaagttggaa agcagaggtc 2040

aaagtcaatc agaaatggga caaactcact tcctactgcc tggtgaaggg gccattttca 2100

gtagcccctt ttcaagatta gtttcattca agatttgata agctgttttg actttactat 2160

agatcttatt atccatgtca gttaagttta tgcttccact aaatctatct gaattcaaaa 2220

ggtaaaaagc taatgctcag tcttatcaga tttatcttat ttattaatag aatgtggatt 2280

tttttaagca tataacaata atagtaatga taggaccata aatgtggatg gctctttaca 2340

agtcactaac attacataaa ttcctcaaca acacactctg aggccataac aaacttttag 2400

aaataacaca attggctacg gaactccagc catctagctt catgggctcc cactttaatt 2460

tcaaaacaac agaactgtgc acattcattt acatgattag ggcagagctt aactgtatct 2520

catgtagcac ctacatcatt cttcagacaa acttattgcc ttttacagac aagaaaactg 2580

gggctcaaaa aaggacttgc ttataactgg ctaataaaga ggaactctgg gttcaaagtg 2640

agtccaattc tttcttccac ccacagcttc tgctaaagtc attacagaaa tgcatagagc 2700

agttcttcca cgttattgct taggtttcta aagagcagtg acctaataca acatgctcta 2760

taatttatta ctgatttaac tatttcacta aggattcact tttaactttt aacttgtaaa 2820

tatgtctaat aaacaccact gaaatagcaa cctctttctt catggccttg tggttgtaaa 2880

gcaagctagt aatatatgtc tgtggatttg tgctaataaa gttctataca cctcattaat 2940

tccacaaatc ctactgggta tttcttatct gccagatcct acgctaggta ctggatacac 3000

agtactgaac aaaatgggta caaatgagcc tcacagagct tgtttcattg aaaagcagag 3060

agatacacac taatcaacaa attaatagta acacactacg atgtgttttg aaggaaaatt 3120

agagcatcaa agagacggtg ttagcaggtg gaggggagct cttttagatg gagaatgaga 3180

[0017]

atgcctccct aaagacatgg gaataaattg agatcacaaa aaatgagaaa tagccagcct 3240

tgagaagagc agaaggaaga acattcaaag gaaaagaaag tgcatactgg aaagcctgaa 3300

cactagagtt tggtgtatgt aaggagctga gcaatggtca cttgtgtgat aagatgtgtg 3360

gatgtggggt ggggggcagg ggtgagtccc acgcagctct taagtgtgtc ctcagactcc 3420

tgtggtttcc atcagccaca acctgaataa ctgtgtggta atccaaaaat gattacagat 3480

taaacatata aaaatatcat tacacccata gtacctaagc caaggacaca gtattctatc 3540

ttttcaatga agatctgcat gaagtaaaat tattatatat aattttaggt attgatatag 3600

atacatcagt ggatagatat agatatgtgt ctctggtata gaaaaaagtt ttaaagggat 3660

attaaaagtt cttatcttgc agggttgaag attgtggcaa ctttcatttc tttttaattt 3720

taagaaaaaa gtggtattat gggggattag catgtttgtg ggtatatgta tatttttaat 3780

taaaaaataa acaacaaaat gaaaacgttt ttcttctatg aaagcctaat aagaagaaat 3840

ttcagctgtt ttaacttagg gagctaaaaa catcaaatcc aagaatgttc tctggaactg 3900

agctcaatac atttttattt gagtaagaat tggatacatt tccatcccct tggggctcca 3960

gtctgtcaat attttacttt tcagcgataa aaagacacat gtagataatc acagtgacct 4020

cagtaacttt ccttctctta tttaagttta ttttatttct atcgtagttt tccctgttaa 4080

agatttttc tttttgctta catatataat tttagagaat aacaatgcac acacaaaaaa 4140

ttcctcttgt tctgctagac ctggactttt tctctaatat atatctccat tttttgtctt 4200

ttttcagacg tattttggaa gcaaaggaga gaattgctat atagctgact tcctcttctc 4260

atcaacagtg ttttaacagt ttttaagcaa aagtcagctt tgtttatcta agatttttt 4320

tgctggcatt taacctaccc ctgcctcccc tttcccaagt ccacttcagc caacctctca 4380

ttcgacaggt accaccctct aacataactg aaataatgtc taccattact ggatcttgct 4440

agcaaagaat ctcaaatttt cccacttggt tgtaaattat tttgtaatct ctagtgttta 4500

[0018]

aggtgcgctt gtcctatcta atcccctccc tggcaggaca ccttacagaa cctaccccctt 4560

acactagtca ttaagcacca tcagggacgg atggctgtgt cactggtctg tttggtattc 4620

cctactgatc ctaccatgtg gtgattatct atgacttccc taatccctgg ctgccttagc 4680

tgggactggc tgacatgctt ctcaggttgc cgctggcttt acagtccttt actgcccatg 4740

ccactttgga gataggcagg gctagtactt ttctatataa gcccccaaac ttgactttgt 4800

gtttcacagt aggtgaaaaa gttgggtctc ttttctttta cttttctttc cacaagatga 4860

taaagctagg ggaagcctgt ggacatggtt tatttctgca actgcaatga ttgattggtg 4920

cttcctgctg cttacttcct aaactttgtg ctcagtgtca gatccctagc agtttctatc 4980

ccctgctctg ctaaaaaaga atggatgttg actctcaggc cctagttctt tttaattaaa 5040

ttgtattttt gttatcatta ttattattat tattttgaga tggggtctta ctctgtcgcc 5100

caggctgaag tgcagtggtg caatcacagc tcactgtttt agcctcctga gtagctggga 5160

ctacaagcgt catgccacca tgcttctttt taatttttta aaatggtttt ctgccttcaa 5220

ttctaagcac ttctcaattg taaccaagag ataatacttt ttatgaattc ttaaagttat 5280

caacagatac tcaaagtttt agcaaagtct aaatgatatt aagcttgtcc ttattgccca 5340

agtgacttca atgactattt gttaattgca accaagggtc atttttaaa tgaatatata 5400

ttattattat atatataata ttaaggtcct caaataccta aaagtttagc aaaatctaaa 5460

taatattgtg catattcttt tattactgta ttagtccgtt ttcatgttgc tgataaagac 5520

atacccaaga ctgggcaatt tacaaaagaa agaggttcac tggactcaca gttccacgtg 5580

gctggggagg cctcacaatc acggcagctt acgggattgt tgagaaatga cacttctcaa 5640

gctggggcta aactatctct gtggtagttg ttctgattca agtattgaat tggtttttt 5700

tgttttttt gagatggagt ttcgttcttg ttgcccaggc tggagtgcaa tggcacgatc 5760

tcagctcacc gcaacctctg cctcccgggt tcaagtgatt ctcctgcttc agcctcccaa 5820

[0019]

gtagctggga ctacaggcat gagccaccac acccagctaa ttttgtattt ttagtagaga 5880

catggtttct ccatgttggt caggctggtc tcaaactccc aacctcaggt gatccacctg 5940

ccttggcctc ctaaagtgct gggattacag gcataagcca ccgtgcccgg ctggagcatt 6000

ggtatataaa agctgcctag gtaactctaa cctttggccc catacatctg aaggatacct 6060

acaatgcacc tgaaaaatgc aactgaaaca gtagttccct gggaccacac actcagaaag 6120

ggggtgtatc aggagatcta gggaccagga gggtggaaga cctaaggcag cactacagat 6180

gatggagaaa aacccactgg ggaggggcga tcctaacctt gagaatcact gagatcatgc 6240

agaagtattt gatcctacag cattaatatt gtattgtatt gtattagtat atatatatag 6300

tgtatatata tagtattagt atatatattg tattgtatta gcatatatat actaattgta 6360

ttgtattgta tttatatata tagtattgta ttagtatata tatacagtat atatgtatat 6420

atactaatac aatgtactaa tacaatacaa taccatatat atatacacta acacaataca 6480

attagtatat atatatatat atatactaat acaatacaat actatatata tactaataca 6540

atatatacat atatactcac caagacatat tagtggtctg atgtctggct gccacactca 6600

tcttctacct tcagctctgc tctaccaaat atcatttgtt tctgggatct ttgcagtcca 6660

aggaacttca tccttgatat cccacccctt actaactttt ttttttttt ttttttttga 6720

gacggagtct cgctgtgtca cccaggctgg agtgcagtgg tgtgatctcg gctcactgca 6780

agctccacct cctgggatca caccattctc ctgcctcagc ctcccaagta gctgggacta 6840

caggtgcccg ccaccacacc aggctaatgt tttaccgtgt tagcaaggat ggtctcgatc 6900

tcctgacctc atgatccatc cgccttggcc tcctaaagtg ctgggattac aggcataagc 6960

caccgcaccc ggccaccccct tactaatttt tagtaacgtc caaggattaa aggaaatttg 7020

ccttacctat ttaacaggaa tcaacagggt taatctcact ccctttctaa aaataattta 7080

taaacattgc agacaatctc atctatccct gtctaaactg tgtggaatta ctgccattta 7140

[0020]

atgtaatcag tctactcatt tagtttgcct aaggaatttt tgaaaaaaca gttaaatgaa 7200

tgacttaatg gaataaccag gaagttgaag tctccaatag taagaatgaa ctcttgctct 7260

ctggataatc aaatgggtcc ttcctccttc aggtagatca tgccatttcc tcacttacac 7320

tgaacaggta aacaacataa ttactgactt caacttctag ttaattcctt cttttatcac 7380

tgagtatcct ttggctggga gttttgttgg ctatgctgcc atttttttcta gttatcacag 7440

tcctataaca taccaatcct tcaatataac tcatctttaa attgtggttt taccttctca 7500

agaagttatt aattatgcca gtgctaaatc ttctaaaatg attgttgact tgttgattag 7560

cccccatgca attcccctct cccgtcccctc agcacgtaag gaatggccct ttgcttactt 7620

ccacagatcc ttaaatctac cagttagaag ctaatagcct acctctctac caggaaggaa 7680

ctgtgggctg gaacataata catgttgact tataatttct tagaaaattg tgtgagaaac 7740

atcaaactcc tgattccagg atatgccaaa gacacatcat taaaaagcaa aacaaaacaa 7800

aacaaacctc atttgacgtt gctagtagtg gcatatttca tcaagatcag ctcaaataaa 7860

tagaagtgag attttcacac aaattagact gtagtgcttt ttttttaac ttatctttac 7920

catatgattt ttaacggtaa aaaaaatcgt ttgagatatt agatgtataa tatttatcat 7980

ccaattactt cattagttca atctttttttc aatggcgctc ctgcatctga gaataaggtc 8040

agaaaatttc atgttctgat ttcatgctga ttttcagaag aaaaatgtta gttttgtata 8100

gaataaccca tcctaagaaa tacatttctt attatatttc ttatcttata tttcttagga 8160

caatgagcta ttcaaagggt gatgataacc agcaccatca gtcagcatta tctaagaata 8220

agaatctgtg tttctacata cagacctcct aaaaaggaac ctacacttaa caggattccc 8280

caggcaattt ggatgcacat taaagcttga gcaacactgc attagaaagt tagttttcca 8340

tcacaaaaac agtaacaaaa ggaatataaa gtaagttact ttaataatat aagaagaggg 8400

gcaggccggg cgcagtggct cacgcctgta atcccagcac tttgggaggc tgaggcgggt 8460

[0021]

ggatcacctg aggtc 8475

<210> 4
<211> 8401
<212> DNA
<213> 人

<400> 4
tttcatctaa gactacattt ctattgtttt atataatcag ccccccctaag atcaacatgt 60

ccacattttt tggcaaagac aaagcctact gatttcagga tcattatttt ccttttttcaa 120

aagcacaaac ccaaactgag aaataaatca agagaaattc tccttttttc tatgctaatt 180

tagaagtaga gtctttattt cttttcaaac ccaaagagaa tcagacatac aatatgaatt 240

tatctacttt cgcttgctca gactgagagg aaagattaat attttcaggc tgttagtcaa 300

aactgttcat tcaaatatta tttaataaaa tccaagaacc agctaaaaag tcgcttaagc 360

taagaaacct tcaccagcct catgggaaat tgtgtacagt tttccactag aatagcctat 420

aaatgcttac tgaaaatgtc taagttcata tcttggtaac taacatttta attcaatctg 480

cagaataata tatgcttctt tagtgctaag atatgaatat tagaggcatt ctttcttaaa 540

atttctattt agttatactt tcacaaataa ctatataata ttaaaattct gcatgtggca 600

taaaacatat tttaatggag aaggtaatgt gtagggagtt tatttctgtt tgctattaga 660

acttgtgttt attcttggtt aaaaaaaactg cagattacaa catagaaaaa aacaaaagta 720

tgttgtatat ctcttacagt agaagataaa gagtagttct aaatttagaa aggaaaaata 780

aatatacaca gtgaaaatat gtgtcagtga gatgttaatc aaagatcaac tattgctgag 840

accagcaata ttaaatccct gcacaattac tcatattata atgagaattt taaaaagaaa 900

atatgaacac ataacataat gaaggcagaa gtcactctca tccttcatct ttgtattccc 960

aattcaggaa gctggtatag tatcttcatt ataattacta ttcaacaaac atttgtaaaa 1020

[0022]

tgaatgaata aggaatgaat gatgagaaaa atgataaaca tctccctctg tctcctggga 1080

gttaactgca ctactttctt ttaaatttaa ttaatcctca atgtccttgt aaaatagcca 1140

aagggaaaat gtatttacat tactctaaat attgatgcaa tctacaaaaa gtgttaaaca 1200

acttcctcaa agtaaataaa atgttcacaa tccagctagg ataaaaggat ttaaatcatt 1260

tcctaggtag agggctttca attagagccc ctgctgcatt aaccatggga actcatctca 1320

ctctcttcat gatggagccc tgagtgttgc tgctaatctg tactctacca ttctaatgct 1380

tttaaggttc cttttcagcc cttcctcctc gtaatccaca aatactgaga ccaaggcatt 1440

ttttgggtca gtcctaattt caagcattct atcctgccct ccccaaatga actcacactt 1500

attagaccat atgttcctat attagttcag gaaggggggaa aaaatgttaa tcacacttgt 1560

atataagaga tcatagaaaa acagtttact aacctgtgaa aataccattc attctctgtt 1620

tacctctggt ccacagctaa gcaatcagca ggatataaat gtaccctatg ttcactattc 1680

agtattcata agtatactac ttatgaattg gaaatctgac acaacattta catgacctaa 1740

ttttgaaaat ttaaaatagt gtaaggcccc taggcttaat tttacagggg aaagattaaa 1800

gggacacaag caaacatata ttctctctct gtgctgtggg acactggtaa ttttttgact 1860

taaaatattt gatacttaaa atgccaaact tctacatttc tgcagtaaca aggcagttat 1920

catattgaat accatttctt tctctccagt aagtagagtt aatattagca catgaactga 1980

aaatattaag tgattataaa aacgtccaaa taaattcatt aaaatttagc ttggcaaaat 2040

gttagtttca tgttcttggt agaagtcctt ttatatttat attcaaatga aatgaacaat 2100

ttacaagcaa aggaaatggc atcaaatatt tcacaccctg cctcccaagg tgtattgatt 2160

catgcttttt gctcagatct aggtttctcc actcaggaaa agaggagaat gtacccatac 2220

ttgggaaaac aagtttccga tggcacagct ttgatcaaac agcaaaattc tatccatcta 2280

tgtattgcca tctgacagta tgacaaatgg tcccatgtgc gatattcaca ctgcattgca 2340

[0023]

gtcaaacctg taagtcaaag gatatgaaat aatagtaact atacattaag cacagaagaa 2400

aatgaaacaa acaaaaaggt tttaaaccaa ccaaaaatat gtcttatttt ggatgttcta 2460

tatgttctta cattctctca ggtcttttgt gtcattatga acacaattct aacaagcttg 2520

attattttat ttccattcac atattacagg caacaagctg aaaaagtaga acggggtgta 2580

gagagacagg acaaagtaca gattagggct tgaagtgccc ctgaccagtc gacagcaacc 2640

acatggaata atgactcatg tgcattaatg atcacactaa atgatatttg ttttttacc 2700

tagtccttca actgacagct taaagaactt caggttgttc tgattcttga gcctcctcta 2760

cagcttcaga gaggactttc attttatttt ggatcaaatg ctccacaact agttgaaact 2820

ggaattaaat tttatatgaa gttcctagat gatttaaagc tgtaagaaga agaataatga 2880

atcataagaa aacttgctgc tacagatatc aaaaaggaat gttaccatcc ctcatgctaa 2940

tccttttcat tttaaataaa caggatctaa aaaaaataat gctgggaagt cctaaccaca 3000

tcaagaatgc ctcagatcag tgacccaggg aaccttccag aatggatgaa atagacccaa 3060

agctgaattc acctaatttt agggccaaaa acccaaaaaa caaaacaaga ccaaaaaaat 3120

cttcagatac tgggagaaca aatctcaatt gctcaattgt atcttatgaa aacaattttt 3180

caaaataaaa caagagatat ttaagattca ttaagttctt gtcatttcaa attttaagaa 3240

aaatattttc taatggaatt acatatattt atatgattct tctagttata tccatggtaa 3300

taaatactct tttcagttgg aaataaaacc catttgtgct atattattag ggaaaatatc 3360

tacataaatt agtttttaat ttaactaaag tctatctttt gaattcataa gcataaaatt 3420

ttaaccactt gcaaaattta taacacactt aaggtagtca gatgccttgt caagtagttt 3480

aacaaaagtg attttcacct gtttgtttta ataacagtgc atcgatttta tgaaaatcag 3540

gcatgccctc gggtcctaac aaagtatacg aagctgaatg gatctatgcc aaatatgcca 3600

gattttactt tctgagtctg attttatact tctgtcctct ttcttaccac atggcttcca 3660

[0024]

gtatcactta cagactaacc cttcaaaagg agaaggctaa gttactaaca tttggaaggc 3720

ttatgaaagt gaagcatagt tatgagccag caatgttttt atttagggaa tgtgtgcaaa 3780

ccatacactt aagcaagctc tggggaatga gagttggggg gaatcaactc ttttatttgc 3840

taattggtat ttcctttaaa agatagagtt cttccagatt ttaactgtgt taatagttac 3900

tctagaaaaa ttggagattt gtgtgcatat attttatgtt gtaaacagac acatacccag 3960

agacactgag agagacagac agacagtaaa cagaggagca ctaaccacaa acggtttaca 4020

aatgacctct gtgctcattc acctgtctgt tccccacctt gccttttata gcaactatag 4080

caacagccat gagagtcatt gtggaaagaa ataaaataaa attaaaaaat cctggaagct 4140

tgtaaagaat gtgagcaaag gggaggaagt tgtgaaaaaa atgaataaag ggcaccgatc 4200

cagagtattg aagaaggcag agtggagagc ctagtaatga gtatctggta ccccagtatc 4260

ctctcccaca gaatctgtac agctctccgt ttatgacagt ttaaacttaa tttaaattat 4320

caaacagaca ctttcctcaa acatataaat gatgaggcag ttcattcagg ctgtatgtat 4380

aaagttgttc cagccacctt tttctaatgg cttctctata tcttttacat ggagacaatg 4440

agagatttgc ttaggacaat ttgactgtaa tttagaagta ggaaatggga agtatttgta 4500

tcttctttgc ctaactcaca ttagttactc aagtaagcat ttcttccgtt attgcatttt 4560

cctgattaca agttttatgt tttctctaaa acacatatca aaagaaatgt cctaagcact 4620

atgcaggggg aagccatgac atttatccac cactgtcagc aaaaacatga acttagccct 4680

caacagaata tttcacttca ttctagtgtc acctctgcgt cacctgcact ggagtcacca 4740

cttgcctgtt gggtaagacc aggatgcacc gctgaaataa aaaggggtca gacaatacaa 4800

gaaaagccag tagaaattgc caaatgtatc agaatacaca caggctttct aaggatatgg 4860

cccaagagga aggctctaga gcccaccctg aaacaggatt tttgacttca cagataaatt 4920

atttaatttt caataacaca attcaattaa agaaagggaa atacaaggct aaacaaataa 4980

[0025]

gaaatgaaga caaaaaccca acctttcaaa tctaaagaaa ataatctgtt ttaaagacac 5040

agatgaagat caggaaccca aaacagaaga aaggaaaggc aattaacgct ggcatctgat 5100

aacaacgaaa agtatggagt ctggagaatc gctagactct aaaaattata aaggtttaga 5160

cttggacttt gtacactgaa gaaaagaaaa ctgcatgcat ttatactgac caatgtacac 5220

tattgctgct ttttaacttt tgtgtatatg tagggtagat tttttttaa gtgaaagcaa 5280

gcttattaag aaagtaaaag aataaaaagg tggcttctcc ataggcagaa aactagcgta 5340

gtttttttat tagaaattgt tattcaataa tagtacatgt tacaaataaa taccatttta 5400

aactgaaaaa attgtagact ttcaaatcag ttagggtggt caccctaaaa aagggcattt 5460

tttccccttа gtctccttgt tcatgttgct cacaacaaga aatgggctaa tgctatgaat 5520

aataataaca aacactgcct tctgtcaggc cctgtgctga ataccgtctg catatgtata 5580

ggaaagggtt aactcagcag gtcttgtttg cccagactct gtacatttcc aagaaaggtc 5640

tgcctttagg actggtcctt ggccagctcc tggagaatga gctctcagct tttagaaaat 5700

tctatctgct aagaatagtt ttgcatgtct caggtcttgg gccacaaaat atcagtttaa 5760

tcagatggtt tatgttaaca agtatgattt atggcaaaca tagatctcta atctccattt 5820

ctctctcata tatctatatt tatctatcca tatatatgta cctatatata tcaaatatga 5880

agatatgttt atagcaattg catataaata gagagatagt atgtagtagg aagagagaca 5940

tagatattat tcttcatttt agaatgttat cttggtatgt ttaaaaggaa aaacttaaga 6000

tgtgttgcaa ttgcagtatg agtttcaggt atgtacatgt tatgtgtgtg tgtgagagac 6060

acacacaaac acatttcaaa catgttttat gtttaagctc aatattcaaa cacagaaata 6120

taacatctat tcttaatatg ttttatgtaa gtacagcagc agcattatta aatactgtat 6180

ttctatggtg attgaaaatt agtaggcaga gaatttttgt aatggttctt aataatttтt 6240

gtaatagtaa atgattactt tttgtttagt atagttttat aatctataca tgaataaagt 6300

[0026]

ggatatttct attcatatag aaatgtgatt tactctcatg tacttatcta catgctaaaa 6360

ccataagtta tcaattttag ttctgtgcca aggcactttt actgaataaa aataatcagc 6420

taattttata ttttcctgat tcaaatttat atgcccgtgt aatgttccgg ggtttttttt 6480

tttaatttct gtaaatcaga atattcagat gttgaaaaag tctttgcctt cagatttaaa 6540

agataccttt gaaatgtagc atatcccaaa atgcaaccca gaggctggca atgtcaacat 6600

ttttctgttt taaaaaacct cttatgaaaa ctattgccat actaaatttt ttacttgctg 6660

atgacttaca gctggaaagg attctgtaca tataagacat caaatattga ggatactgga 6720

acttttaaat taatggcaaa gaaagtcaac aaaggaagtt catatgaaat caaactagta 6780

atatgattac aaaaaaaaaa gtttaaaatt tttcttggcc ccagtcttat catttctgag 6840

ccaaatacaa ttctatcgaa atcacctgaa actgaaatca ccattctagg ctggttttcc 6900

cataaagatg gactgctcca aaaagaggaa tcaagaaaga atttggctca cagtgaatta 6960

ttcactttgt cttagttaag taaaaataaa atctgactgt taactacaga aatcatttca 7020

aattctgtgg tgataataaa gtaatgacca cttttcagct ggagggacta acttcttttt 7080

ttttttgct gcatatatag ctgtggtaca ttttaatgtg aaatgatgac tgcatcagct 7140

tatatccatg gagcagattt tagcattcag cttgggtctc ccagtcaata tctacgagtc 7200

tcttcttaag gagatcgatg acacagatac atacagacta acaaatgtga taccaataat 7260

caagaattca ctcagttaag attttgccca ctgatttcca cacaagaaac ctagaattta 7320

ctagattctt gtgcctgtga ggctccactc atttccctga atcacaaaag ctacagagta 7380

tttagataga aatataccta ctcttaacat gaaccatttt aaatatatgt attactgtgt 7440

ccacaggagt acactttaaa gcagggactt cactcttcaa tctctccaat cacgtgttac 7500

ctaaagtggc atgtggttcc ctaaagctta ataactgaca ttgccttaaa aaaggggttt 7560

gcttcccgac taatgtggaa aaagtctgaa aaatgatttt aaatctttca ctaaatttct 7620

[0027]

catttggtca cgtggaggaa aatgatttca ccaaatagat actctcatta atttttaat 7680

gtaatttatc aaagaaatga aatatttaga taaattccag atttccccca ccatgagctt 7740

ctccgaaagt atactccatc acagactgct cactaagaag ctctactgca gtcaaagtga 7800

ccgaatttaa ggggacataa tgactacttc tgctacacag aaacattatc catctctaac 7860

acttccctat gagatggaag acggacttct aatcaggtac cagagagggc tctgccaact 7920

tcagggcttt gatgaataag aatggttgag agcgctcatc ataaatgaat tcagtataac 7980

tgagtgagaa agtgagagaa ccagagaaat aaatcctcat gtagaaaatt taggggtatg 8040

aaatgccaaa tgccagttaa ccaaagcttt ctttgtcata aagcaacttc tataaaaatt 8100

gctgaaaata aattcttcat ggctcaatgt gaatcagtaa tttccatctc tattacactg 8160

ttgtttaccc aaaaactatt tttaatgact aagactcaga gtttgccaga gtgttttcca 8220

caaaacaact gttttgagat actccagatc tgtaatcaag taagtctgaa aaaccccaaa 8280

tacctcactc acctcttgga tatgcataaa gcacactaat atataacgtt ctaaaaagcc 8340

aatcattaaa accgttttat attgtttaag catttcctag acatatttgg ctacaaatct 8400

a 8401

<210> 5
<211> 8427
<212> DNA
<213> 人

<400> 5
gcacctgcca ccacgcccag ctaattttct attttcagta gagatgaggt tttgccatgt 60

tggccaggct ggtctcgaac tcttgacctc aggtgatcca cccgcctcag cctaccaaag 120

agctgggatt acaggcgtga gccaccgcgc ctggccatat taacaaattt taaatcacaa 180

ctatgtgggg ggggaggcta gtattattac agcagattgg tttgctatat aaacaagtac 240

tttaaaaaat atttcttggg ccaggcgtgg tggctcacgc ctgtaatccc agcactttgg 300

[0028]

gaggccgagg tgggcagatc acttgaggcc aagagttaag agaccagcct ggccaacatg 360

gtgaaacccc atctctacta aaaatataac aattagccag gcatggaggt gcatgcctgt 420

aattccagct gctcgagagg ctgaggcatg agaactgctg gatcctggga ggcagaggtt 480

gcagtgagct gatattgcgc cactgcactc catatccagc ctgggcaaca tggcaagact 540

ccgtctcaaa taaataaata aataaataaa taaaactaaa ggcagagttt tcttaaataa 600

acatggtagc cctcagcaac aatattgtaa gaactcctcg caagagaaaa agctggaata 660

agatactggc taagcaagta agaaaggcac tgccctgctt ctgcatacat tcaaactaag 720

acatatacat tgcagcttac acttacattt tccaatatcc ccaggcatcc ctttcccttc 780

tcaaacagcc aaaaggaacc agccatgcaa ataaaaatac aagttcaaga gcctaaaaga 840

agtcagtgtc ctaaaagaga aaattaatgt aaagaattaa gatttttga aactacactt 900

tctttctggg gctgtttact ggcctccaat acatcaatcc tgtaacactg tgaactacag 960

tgatagattg gtacatgctt ctaaacacaa cagaattttt ccaaggttac atacactgta 1020

acaaaagggg cattttgcag catcttattt tccttaatca actagtttgg atattctaac 1080

agtgcaaaca ttgtaaacaa taaattttca ttacctttg aactttctga agtcaaccaa 1140

aggcttgtgg tatggatgca atgagtacta gacaggcaga gctgaatact agtcaaaata 1200

ttcagttact ggtgtgatag tccttttggg ggcatacatc acttagggag aaactgaggt 1260

gcaaggacat tttacacaca gcaaaaacat tctcaggaat ttgtcacatc attaccataa 1320

gccaaaaatc tcaaggtctt agaacagcct gagcttctga tcaaattata ttgtaaaaag 1380

agaggaaaaa aatgtgaagc gtgctatttt ttaaaataac agtaactact actactgctg 1440

ctgctgctaa ttctaaacgt ttactgagcc cttattatgt gccaagcacc gtgctaggta 1500

cggtcataga ttttaacaat taatccctgt aacaaccctc tgatattagt taataaaatt 1560

aaagtagaat cctcaccaaa aaaatttaaa ctttccaaat aaaaatataa ataaattatt 1620

[0029]

aaagacattt cacctctttc tctgcctcag actacatttt caagtattaa atttacacta 1680

aaaccacatt tattttcagg aattccagtt aaagcgtaca gatattcaag atgttgacaa 1740

ttattacaga agaatcacag aactctgaaa ttaaatactg gcacagaaaa ccttccatcc 1800

aaccttacgg aacaactatc cccattttaa aaaaaaagga acagcatata tatcaggctt 1860

gataataaga ggcttctcat gcccacacta gcaatgaatg atgccataat tataaagaga 1920

cctgtatcgc cacatgcata aaaataattt acatctgcta agtcaagttt tcaatatatt 1980

attttgtgtg taaaccttat agtagctgat aaaaaataca ataaactaat ctaaggtaaa 2040

ctaaaacact aggttgtttc tgaagactca ctttagaatt tgagcagcat aataatcata 2100

atattagtaa tcaaactact tagcagaaag ttcttagagg gctgggaagc tgtgtataat 2160

aaaatggagc agacaagaag gaagggtttt ccgtactgtt taaatcaact acaggtccca 2220

gcatgcagtg ctctaatctg aagttaagca aaaactgcaa tgcatactgg gacttgtagt 2280

aagtaaacca cgttatcaca gcaagtttca agaaagtctg aactatctag cacaatttga 2340

ctatatctta ttatcagagt ctaatcaaat ttaaatcaaa tttgtatgtt ctctgatgtg 2400

gcacacagtt tctctagcac ataccggaaa aagtatcaat atttagacca acattttcac 2460

attagaaaaa tcttacgtag gagaagcaca gaaaaaaatg ctgaaaaagc aaaaaaactt 2520

gatgaataaa aaatataatt tttgaaatag ttttttaaag tttgaatgga tccatttcaa 2580

cattctctaa tcctccccca caaaaagttt aattgttttg gccgggcgcg gtggctcacg 2640

cctgtaatcc caacacttta ggaggctgag gcgggtgaat tacgagatca agagatcgag 2700

accatcctgg ccaacatggt gaaaccatct ctactaaaaa tacaaaaatt agttgggcgt 2760

ggtggcgcac gcctgtagtc ccagctactc aggaggctga gacaggagaa ttgcttgaac 2820

ctgggaggtg gaggctgcag tgagctaata tcgcaccact gcactccagc ctggtgacag 2880

tgtgagattc attctcaaaa aaaaaaaaaa aaaaagttta attgttttaa caggttgctt 2940

[0030]

tttaacaatt attcaagatg tattttataa ataatttttc ttgaagaaaa ttctcagaag 3000

caaacattcc ccatattcta atattgccca ccaggaaata attttttttag taatacgcac 3060

acaccccatc acaaaaacaa acaaaaaaca ctgaagttct gcttttgtca agtccttact 3120

caatatttat gccctccatt cctcacctct aattccctac acacacacac acacacgcac 3180

acatccccac acacacacgc ttctacaaag aacacttaga aaaacagtat tccaactaca 3240

agcccacttc tctcatccac tgacctcttc tgaaaacaca aaagattttt taagctatca 3300

gtaacacgtc caaacacaag ctgataagtt tgagctagaa tttacatata tacagttgct 3360

acacaccctc ctatttttctg caagtctgtg gaaggaggct gggaaagaac taagtgcaat 3420

ctgcatcagg aggcctaaca caggtggtgg gttattttca ggcaacagca ccttcacaaa 3480

catgttttgg aatatagtcc aagaaattcc taacaaggaa agataagctg gcacacaaat 3540

ttaacgcaat ccagctaaaa atcatctgca acacatgcta ctacatttca ccataaaagt 3600

gacgggctac tataaaggat ttgaagcttc gtcaatacaa catactgtcc ataaggccag 3660

agatagcagt tgccatggtt actataccca cttttatcag gaaattactg tcattacccc 3720

aaagttttgg gtacttattt aaaatttaaa aaaaacacac acaatttagg gttctgactg 3780

ttaattgagt gaaataatca actactgttt gatttgtaag tatgtcgctt tggagatgca 3840

catggttaac aatacttgga tctgcagcag aaaaaaaatc aattcctttc tgctgctcct 3900

tctcctcaag tactgacagt ttgtattctc aatgcagcca aaacaataaa acaaaaccca 3960

tcttttggc ttctgtgttt aagttatttt tcccctaggc ccacaaacag agtcaaaata 4020

aagcctagat catcaacctg ttaggcctca tccccttcct atccctctca tactggttca 4080

ctttcttgac tacttagaaa aggcagaaaa catttctgta actgattcca aagtatagaa 4140

aagaatagtt gccttcaact gagatatttt caccaaagtc ttttttattt actttttttt 4200

taaggcaggg agaggggaga gacttgcagg gtactgaaag ggagaagtgg aggagtattc 4260

[0031]

aaattgccac acaagtctag tgtaagaaag ttgctttaga agagtccaaa ggatggctga 4320

acctcacata taatttctaa aagctttgga agagttcacc ataattttaa gactgaattg 4380

agggacaagt aatagaaaag ttattcataa agtctacttc aacattttta caaaagataa 4440

ctattcaaaa atttaacaca catataagaa ttatacgaaa gcctacaaaa tagtatggcc 4500

acatatacac acaaacatac aaagtagaaa acataagcta tttaagaaat aattatctac 4560

aataaattca atgcaatgtt aacatattat ctcttttta aaaaatcgca aagcagcaaa 4620

aacatacacc tgagaaaatt aatgtgatca aaacgttaaa gaattcttag gcctataaaa 4680

aaagcccatg tacaaaagct cctgagaagt caacataaat cattaatatt tcccagcaca 4740

aaataatatg aaaattcaaa catgtttcaa gaaatcagtt ctagatatag atataaaaga 4800

attccattaa aggtcagaga cctaaaactt taattccttc ccttctctgt ttgaatagta 4860

attaaataca aaagccttca gcaataaaat actaaggata caaaatttaa aagcacatta 4920

atataagctt aacttcagta tgtcttcaca gaaagcttta ctattcactg tctgtaggat 4980

gaaaaagtta ataacaccct gagaggtttc atttttatct aaacagttaa gtgtttttct 5040

caccgttcac agaagcaagt ttctatattt actttctaaa gggggcaatt tcaaaagaat 5100

agtcacttct aaaatttaag atactatacc ttttgatagg ctcataaaca cagggttcct 5160

aattatctat attttacttt aaaatgtttc tattccaaat ttgtgagcag agtttataag 5220

aaagctgaaa ctcaaggctt taaacttttg ggttattttt acacaaaaat atttcagtgc 5280

actcctctag atttgagtag tcatttcctt gtgcatcctt ctaaaataga aaaacaaaaa 5340

tgatatatcc atatatacct aatactaaca catacagata tacatctttt tcactgtgaa 5400

acaagcttga aagctttagg cagtaagaat ttttcagaaa gttagcagag tcagtcaaaa 5460

cattcaaaac ttgaaccatg acatctgtta ctctgtcaat aagagtctat agaagaatca 5520

gggaacttac atactcacta aaatcaacta ctatcacatc acatcaatgg agaaatgaag 5580

[0032]

aaaaactgta ataggggaca tacaattcac aggatcttca aaagggaaaa tgatcttttt 5640

ttttttttta aattatgaga aactgactag gcagcatttt ttcaaaagca gcttcaaaac 5700

tataacaaag acatttttgg taaccacagc agtatttaaa aaacaaaaat ttaggccggg 5760

cgtggtggct cacgcctata atcccagcac tttgggaggc caaggcaggt ggatcacctg 5820

agtcaggagt tcaagaccag cctgaccaac atggtgatac cccgtctcta ctcaaaatac 5880

aaaacttagc cgggcgtagt ggcggacacc tctataatca cagctactca ggaggctgag 5940

aggcaggaga atcgcttgaa cctgggaggc agaggttgca gtgagccgag atcacgccgt 6000

tgcactccag cctgggaaac agagcgagac tccgtctcaa aaaataaaaa aataaaaaaa 6060

ctatagtgtc cagggtgcac tttaaatgta ttactttctc aactgatatg gaaaaagtta 6120

gcatttaaag acagaagctt ctgtccatgt attaattagt tacctatctc aacaacttaa 6180

tatctgcatg ctttcttacc atttatgaag aacttttata tgtattatct catttggtct 6240

tactgagaaa acagtatttt gcctacaaaa tagacaaaat tcaaagcaga tttatcaaac 6300

tttctagcat ccccaaattt ttaaaacttc gacacaaaac tttacaagca accacagtgg 6360

catgatattt tcagtgataa tcaattcacc taacactaac agagtttcaa aggaccatgt 6420

gctataaatg ctatgaaact gttaaagtag ctatattcat ctttatgcag ttactgttac 6480

atcaacaatg acctaccact gatacaactt gacttacagt tcaagaatct cagtctttgc 6540

aggctaactt aagtacatca accatatgta tttataaagc cgagtgccta aaaattgatc 6600

tatattagaa tcatagtctg taaatccgag gggaaaaaac tacaagaagt ctaaaatttt 6660

ttcaacacac tatacccctt tccaaaatct caactactct atatcctatt tgtattaata 6720

ttatagggat gataacaagg cttaaagccc taaatcatac caactacttt tgtttataac 6780

aattacaaat aatttttaa aatacatgct caacatccca ctcatcaaca caagactaat 6840

tccccttcca aataaaataa ttctaaacag tgctctgtac caagggccag aatccttata 6900

[0033]

ctatccgcaa tcgcacatct actttgtaca gtcaaagact tcactttcaa gtagcaaaca 6960

ttatttatga atggaatttt taaatggact tactcaaaat ctttctggaa ctttaaggtg 7020

ttaatcctgt tgcttagctg aagctaagca gagctgtaat aagtagcaag accctcaaaa 7080

ttcaaaaatt tcctttatct tgctgtagca cctcctgctg gatagcattt agagatcttc 7140

atgtaagcag aagaagagta tttcagaggc agctccttcc agaagactga ataggaaaaa 7200

ggatggaccc ttcaaagcta aaagaaatag gccccatcca tcacttatac cttctaaaaa 7260

tacaatttag cccaggtagg tgtctttttc atctattact actccagttc cacaaagact 7320

tgcctcagtc caaaatacaa catgcttaaa taaagcctgc aaaattgtct aaaaactaag 7380

ttaaaaagca ttcaatagca cccaagcaaa acactttatt atgggcagcc aagcaatgtc 7440

agtcaaactg taaatactat tatgttacca aaagcaaaag tctgatgtta aaaaaaaaaa 7500

aaaaaaagcc cctggaatat tcgtaacatg ttagccagat gtttgtgttt tgagaacttt 7560

gtgcactatt actatgctct tcacttaagg atagttgtac atctacaaac gttttaagta 7620

cagaaatttt tttataaaca ttagcataac tgtacacaaa atttcctctt tgccatgaaa 7680

agataggtcc tgggatttga aaatgtattt ttcagacatt tttaatgacc ccctaaaata 7740

aactagtttt aagcccacaa caccgattcc ataaacaagt aaagacagaa gaagagaata 7800

agaaggaact taccaaaatt aaaatgaata atagtatttc cagtaaaaat gtagtaacag 7860

tttccaacaa tgctgtaaac caaataaatt gtgaaactta aaaaaggaag gaggggccca 7920

gtcttcaaag accaaaagca aagctgacct atttatttct attgcttaga gtgaacacca 7980

gatgtaaaca aatatcataa acactgaaaa gtacgcttac atggtttagc ctcaatttca 8040

gtacccttac caggccctca ataaagctac agatgttggt gagaactcgc tcaaaaagga 8100

gataattcca gcccctcgcc ttaaagaatc cctatcaagt gaacctgtga aaagacttcc 8160

ttcccagagt gcacaactgc tttaaaaaaa aaaaactttc atcagcccaa attaatctga 8220

[0034]

ttctaatatt caactatcca ttatttatat ataaatgttc ttccctctct aactttccca 8280

gctcgagcat ctacattcct gacaccgact attagcaaaa atgcacaact ccttccccag 8340

ctatggggca aatctttgaa atctgaaaca cagccacaaa gttcactgtc aaggccaggt 8400

gatgaggccc acacatgccc ggacctt 8427

<210> 6
<211> 1704
<212> DNA
<213> 人工序列

<220>
<223> 片段A

<400> 6
ggtaccaccc aagctggcta ggtaagcttg ctagcgccac catggtgctg cagacccagg 60

tgttcatctc cctgctgctg tggatctccg gcgcatatgg cgatatcgtg atgattaaac 120

gtacggtggc cgcccccctcc gtgttcatct tcccccccctc cgacgagcag ctgaagtccg 180

gcaccgcctc cgtggtgtgc ctgctgaata acttctaccc cagagaggcc aaggtgcagt 240

ggaaggtgga caacgccctg cagtccggga actcccagga gagcgtgacc gagcaggaca 300

gcaaggacag cacctacagc ctgagcagca ccctgaccct gagcaaagcc gactacgaga 360

agcacaaggt gtacgcctgc gaggtgaccc accagggcct gagctccccc gtcaccaaga 420

gcttcaacag gggggagtgt taggggcccg tttaaacggg tggcatccct gtgacccctc 480

cccagtgccct ctcctggccc tggaagttgc cactccagtg cccaccagcc ttgtcctaat 540

aaaattaagt tgcatcattt tgtctgacta ggtgtccttc tataatatta tggggtggag 600

gggggtggta tggagcaagg ggcaagttgg gaagacaacc tgtagggcct gcggggtcta 660

ttgggaacca agctggagtg cagtggcaca atcttggctc actgcaatct ccgcctcctg 720

ggttcaagcg attctcctgc ctcagcctcc cgagttgttg ggattccagg catgcatgac 780

[0035]

caggctcacc taatttttgt ttttttggta gagacggggt ttcaccatat tggccaggct 840

ggtctccaac tcctaatctc aggtgatcta cccaccttgg cctcccaaat tgctgggatt 900

acaggcgtga accactgctc cacgcgccct gtagcggcgc attaagcgcg gcgggtgtgg 960

tggttacgcg cagcgtgacc gctacacttg ccagcgccct agcgcccgct cctttcgctt 1020

tcttcccttc ctttctcgcc acgttcgccg gctttccccg tcaagctcta aatcgggggc 1080

tccctttagg gttccgattt agtgctttac ggcacctcga ccccaaaaaa cttgattagg 1140

gtgatggttc acgtagtggg ccatcgccct gatagacggt ttttcgccct ttgacgttgg 1200

agtccacgtt ctttaatagt ggactcttgt tccaaactgg aacaacactc aaccctatct 1260

cggtctattc ttttgattta taagggattt tgccgatttc ggcctattgg ttaaaaaatg 1320

agctgattta acaaaaattt aacgcgaatt aattctgtgg aatgtgtgtc agttagggtg 1380

tggaaagtcc ccaggctccc cagcaggcag aagtatgcaa agcatgcatc tcaattagtc 1440

agcaaccagg tgtggaaagt ccccaggctc cccagcaggc agaagtatgc aaagcatgca 1500

tctcaattag tcagcaacca tagtcccgcc cctaactccg cccatcccgc ccctaactcc 1560

gcccagttcc gcccattctc cgccccatgg ctgactaatt ttttttattt atgcagaggc 1620

cgaggccgcc tctgcctctg agctattcca gaagtagtga ggaggctttt ttggaggcct 1680

aggcttttgc aaaaagctcc cggg 1704

<210> 7
<211> 1120
<212> DNA
<213> 人工序列

<220>
<223> 人 IgG1 信号+人 IgG1 恒定区

<400> 7
tgctagcgcc accatgaaac acctgtggtt cttcctcctg ctggtggcag ctcccagatg 60

[0036]

ggtgctgagc caggtgcaat tgtgcaggcg gttagctcag cctccaccaa gggcccaagc 120

gtcttccccc tggcaccctc ctccaagagc acctctggcg gcacagccgc cctgggctgc 180

ctggtcaagg actacttccc cgaaccccgtg accgtgagct ggaactcagg cgccctgacc 240

agcggcgtgc acaccttccc cgctgtcctg cagtcctcag gactctactc cctcagcagc 300

gtggtgaccg tgccctccag cagcttgggc acccagacct acatctgcaa cgtgaatcac 360

aagcccagca acaccaaggt ggacaagaga gttgagccca aatcttgtga caaaactcac 420

acatgcccac cctgcccagc acctgaactc ctgggggggac cctcagtctt cctcttcccc 480

ccaaaaccca aggacaccct catgatctcc cggaccccctg aggtcacatg cgtggtggtg 540

gacgtgagcc acgaagaccc tgaggtcaag ttcaactggt acgtggacgg cgtggaggtg 600

cataatgcca agacaaagcc ccgggaggag cagtacaaca gcacgtaccg ggtggtcagc 660

gtcctcaccg tcctgcacca ggactggctg aatggcaagg agtacaagtg caaggtctcc 720

aacaaagccc tcccagcccc catcgagaaa accatctcca aagccaaagg ccagccccgg 780

gaaccacagg tgtacaccct gccccatcc cgggaggaga tgaccaagaa ccaggtcagc 840

ctgacctgcc tggtcaaagg cttctatccc agcgacatcg ccgtggagtg ggagagcaat 900

ggccagcccg agaacaacta caagaccacc cctcccgtgc tggactccga cggctccttc 960

ttcctctaca gcaagctcac cgtggacaag agcaggtggc agcagggcaa cgtcttctca 1020

tgctccgtga tgcatgaggc tctgcacaac cactacaccc agaagagcct ctccctgtct 1080

cccggcaaat gagatatcgg gcccgtttaa acgggtggca 1120

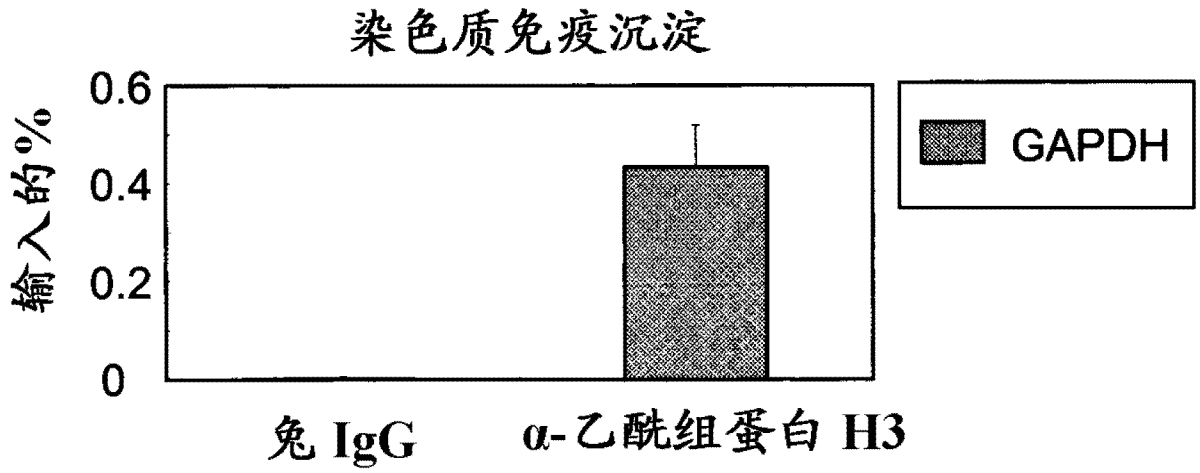


图 1

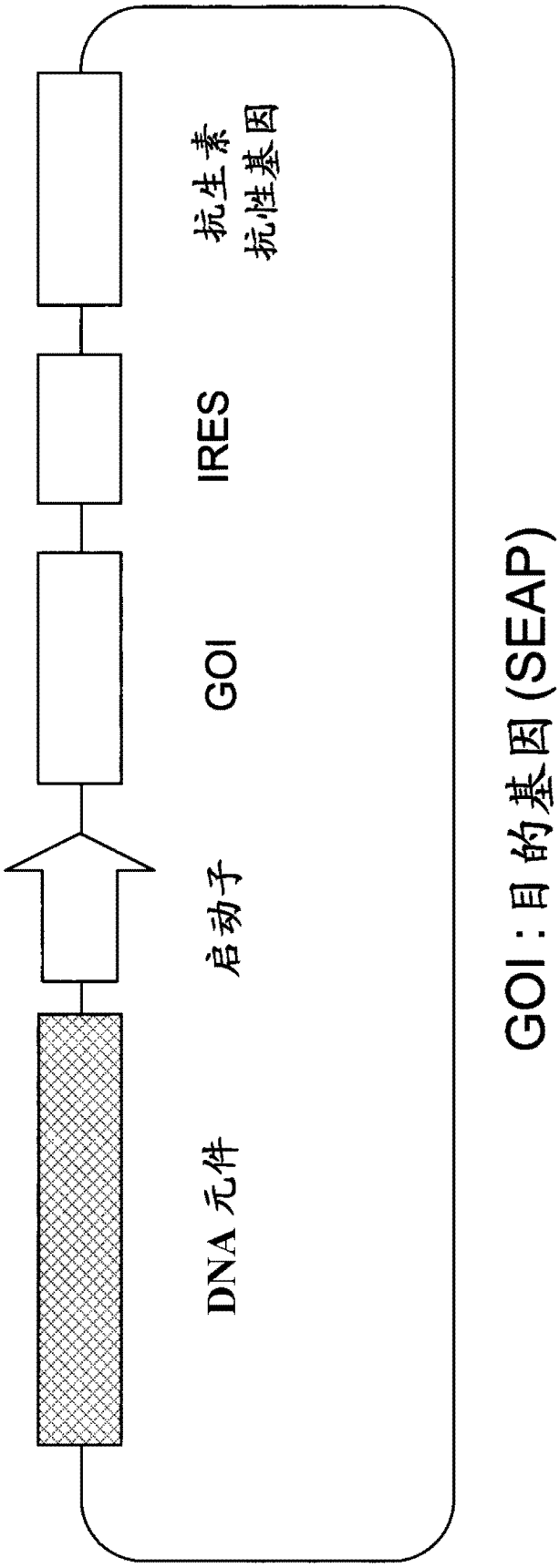


图 2

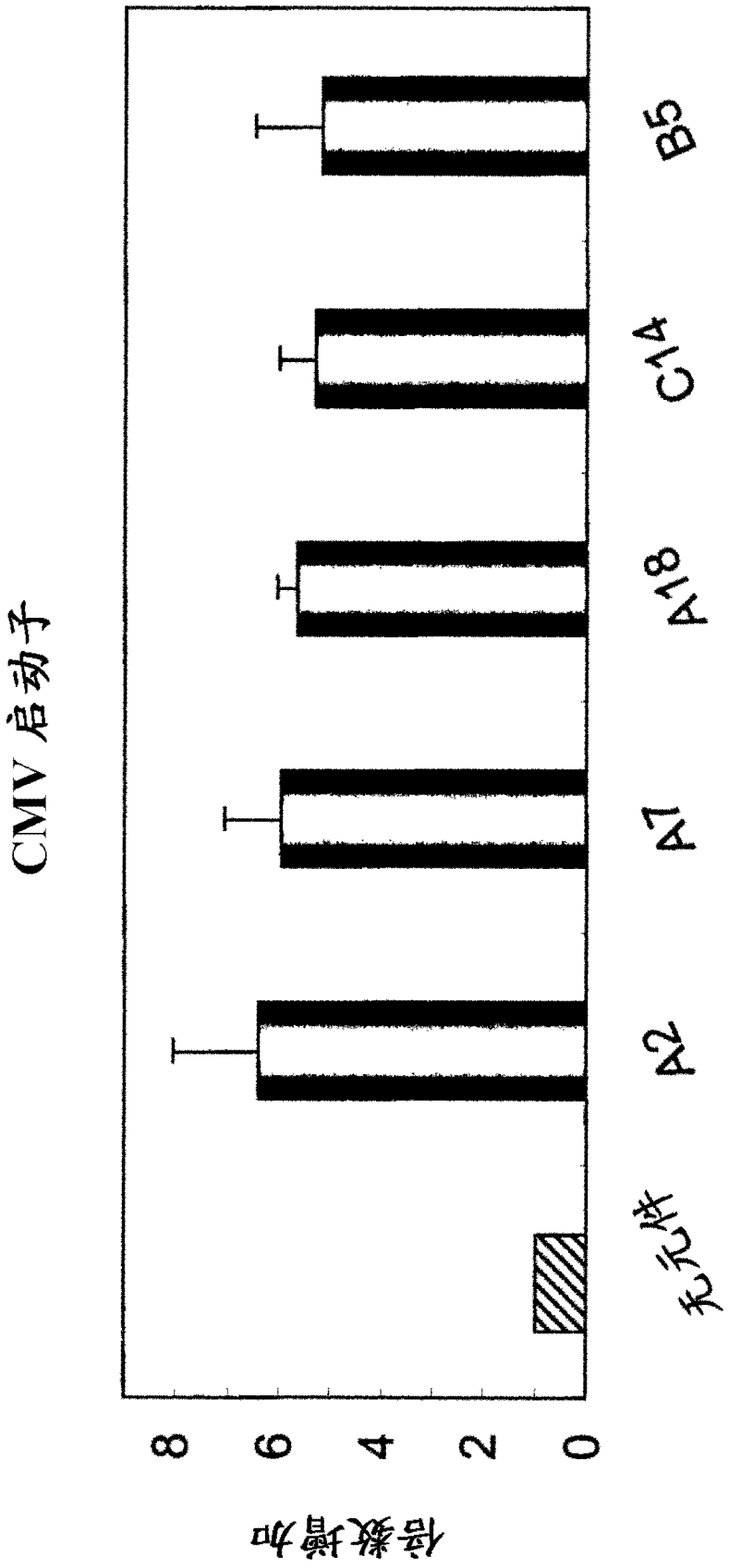


图 3

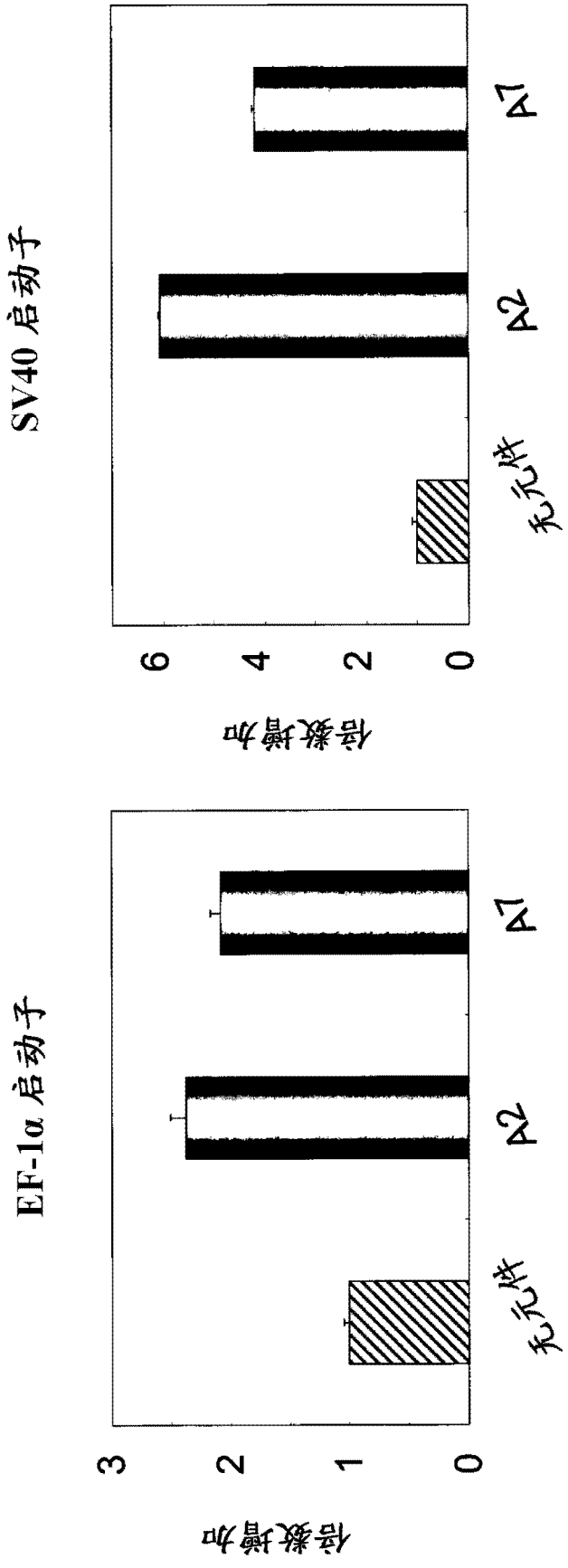


图 4

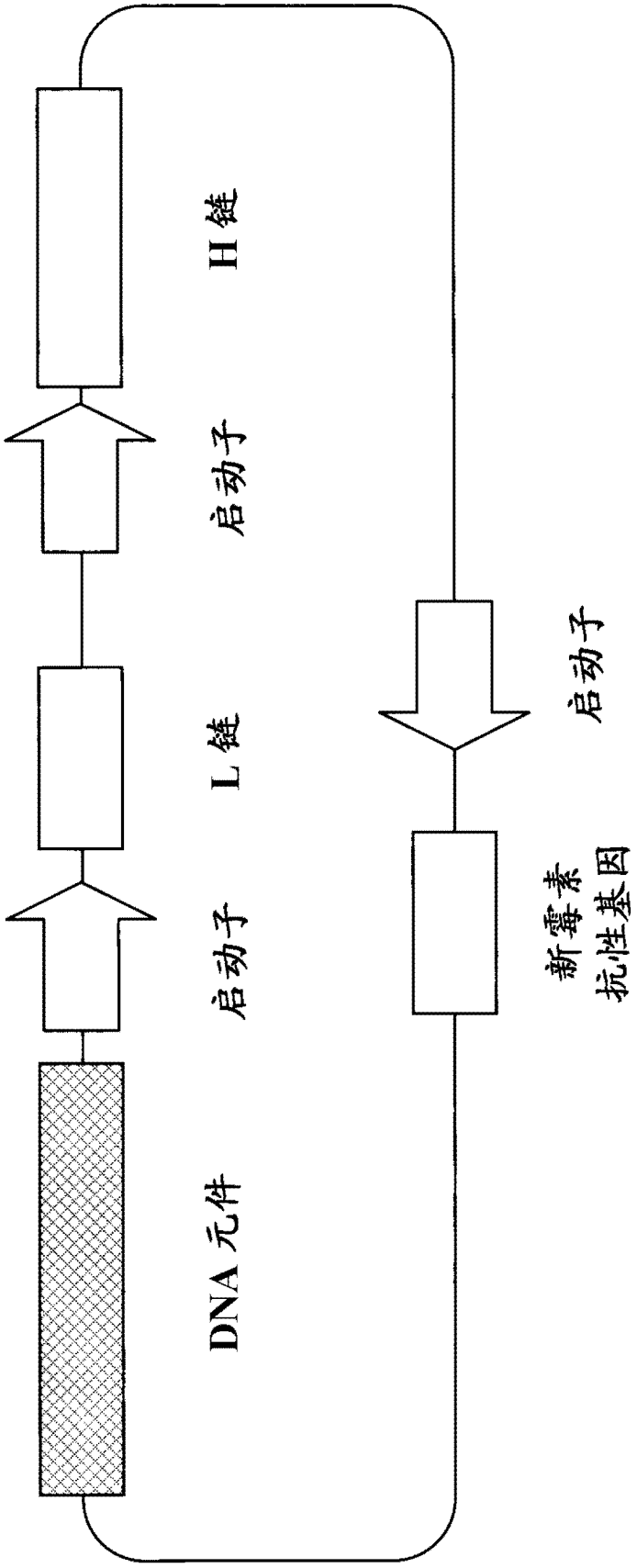
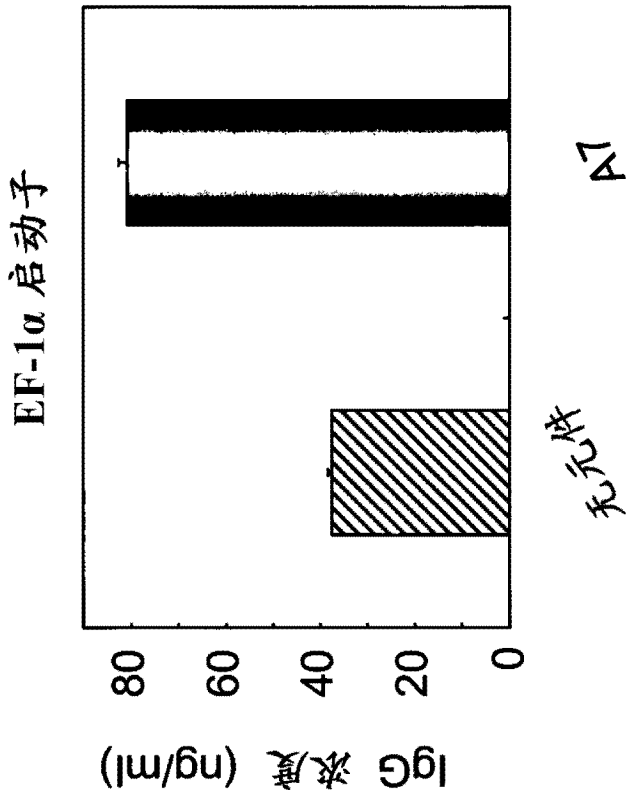


图 5


EF-1α 启动子
IgG 浓度 (ng/ml)
0
20
40
60
80
无元件
A7

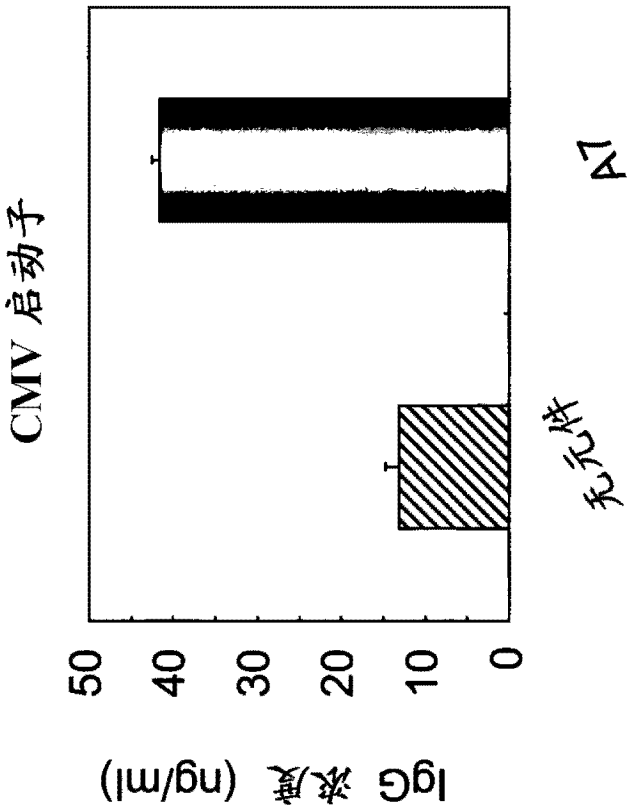

CMV 启动子
IgG 浓度 (ng/ml)
0
10
20
30
40
50
无元件
A7

图 6

ID	起始位置	终止位置	长度(bp)
A2	80966429	80974878	8450
A2-1	80966429	80969428	3000
A2-2	80969229	80972228	3000
A2-3	80971829	80974878	3050
A2-4	80967129	80969128	2000
A2-5	80967129	80968628	1500
A2-6	80967129	80970128	3000
A2-7	80968429	80971428	3000
A2-8	80970429	80973428	3000
A2-9	80966429	80970128	3700
A2-10	80968429	80972228	3800
A2-11	80969229	80973428	4200
A2-12	80967129	80972228	5100
A2-13	80968429	80973428	5000
A2-14	80969229	80974878	5650
A2-15	80966429	80972228	5800
A2-16	80967129	80973428	6300
A2-17	80968429	80974878	6450

图 7

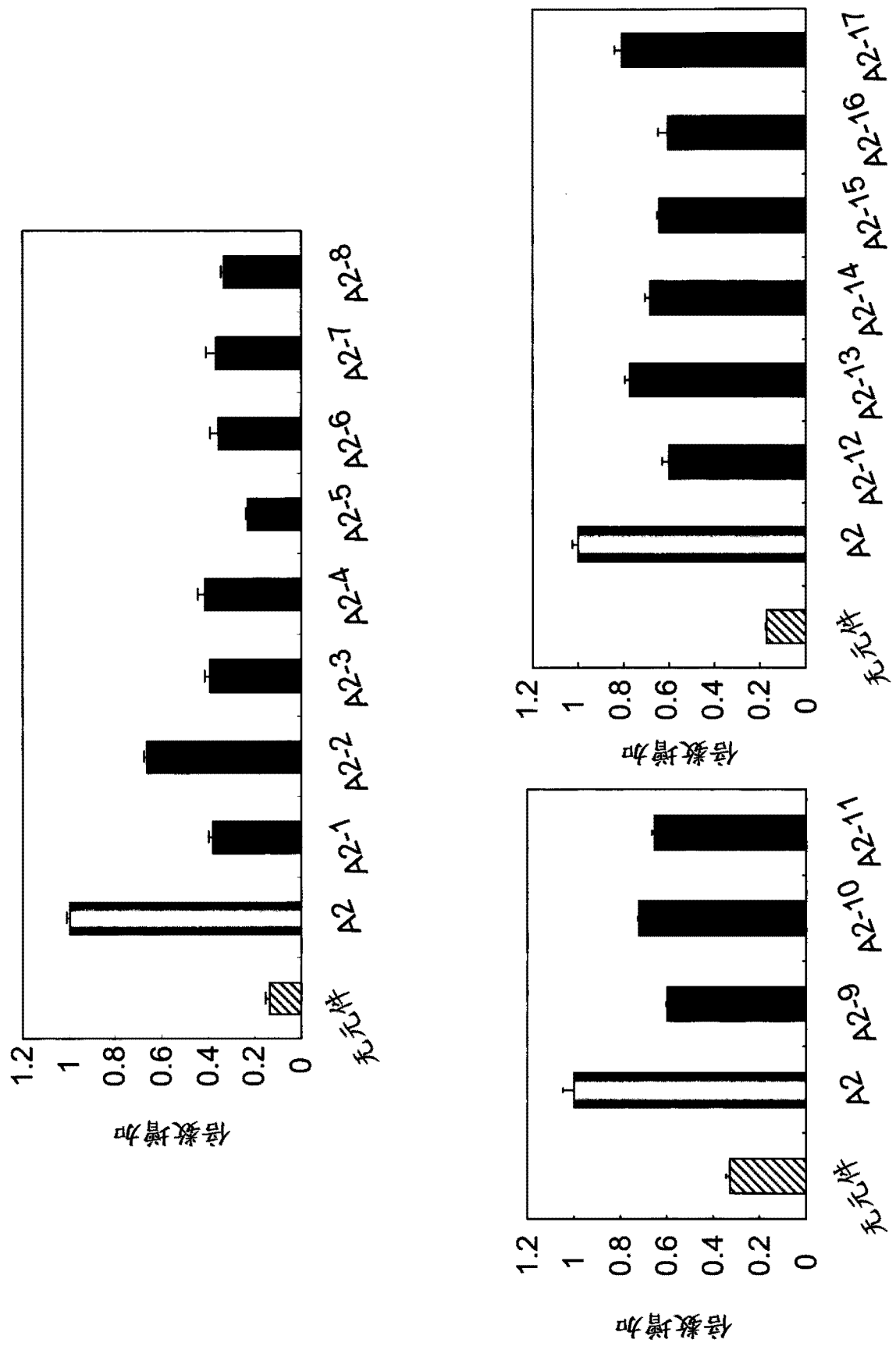

倍数增加
1.2
1
0.8
0.6
0.4
0.2
0
无元件
A2
A2-1
A2-2
A2-3
A2-4
A2-5
A2-6
A2-7
A2-8
A2-9
A2-10
A2-11
A2-12
A2-13
A2-14
A2-15
A2-16
A2-17

图 8

ID	起始位置	终止位置	长度(bp)
A7	88992123	89000542	8420
A7-1	88992723	88995722	3000
A7-2	88995723	89000542	4820
A7-3	88997523	89000542	3020
A7-4	88995523	88998522	3000
A7-5	88993623	88996622	3000
A7-6	88996523	88999522	3000
A7-7	88994523	88997522	3000
A7-8	88992123	88995722	3600
A7-9	88993623	88997522	3900
A7-10	88994523	88998522	4000
A7-11	88995523	88999522	4000
A7-12	88996523	89000542	4020
A7-13	88992123	88997522	5400
A7-14	88993623	88998522	4900
A7-15	88994523	88999522	5000
A7-16	88995523	89000542	5020
A7-17	88992123	88998522	6400
A7-18	88993623	88999522	5900

图 9

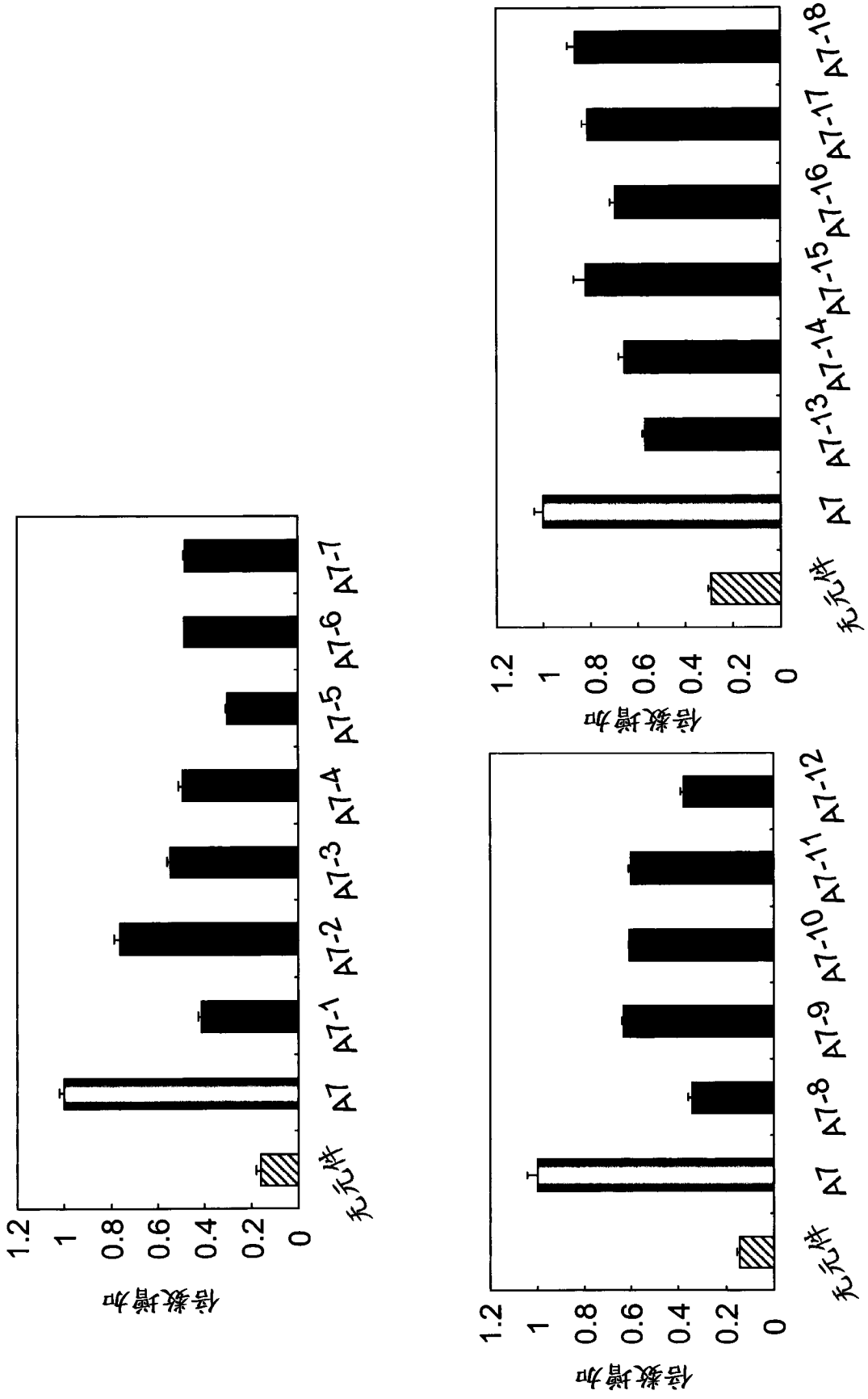

倍数增加
1.2
1
0.8
0.6
0.4
0.2
0
无元件
A7
A7-1
A7-2
A7-3
A7-4
A7-5
A7-6
A7-7
A7-8
A7-9
A7-10
A7-11
A7-12
A7-13
A7-14
A7-15
A7-16
A7-17
A7-18

图 10

ID	起始位置	终止位置	长度(bp)
A18	111275976	111284450	8475
A18-1	111275976	111281015	5040
A18-2	111276976	111281977	5002
A18-3	111277976	111282975	5000
A18-4	111278975	111282975	4001

图 11

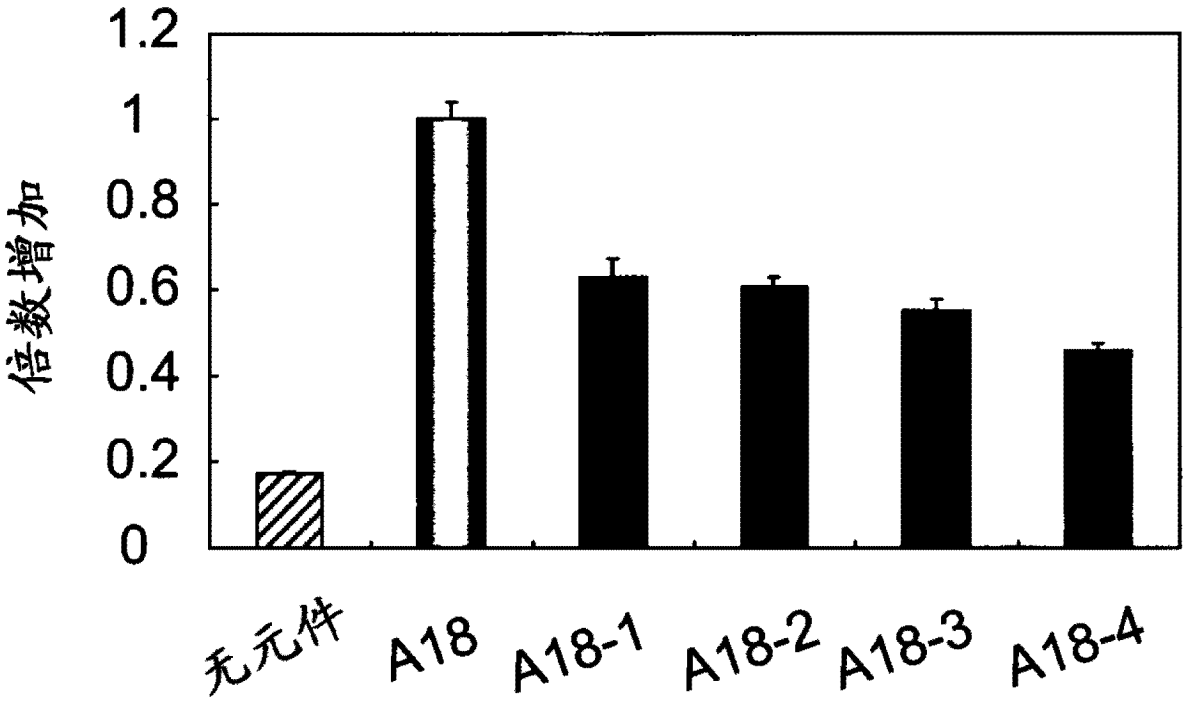


图 12

ID	起始位置	终止位置	长度(bp)
B5	143034684	143043084	8401
B5-1	143034684	143038684	4001
B5-2	143034684	143037883	3200
B5-3	143037174	143040284	3111
B5-4	143040056	143043084	3029
B5-5	143035584	143038684	3101
B5-6	143038684	143041683	3000

图 13

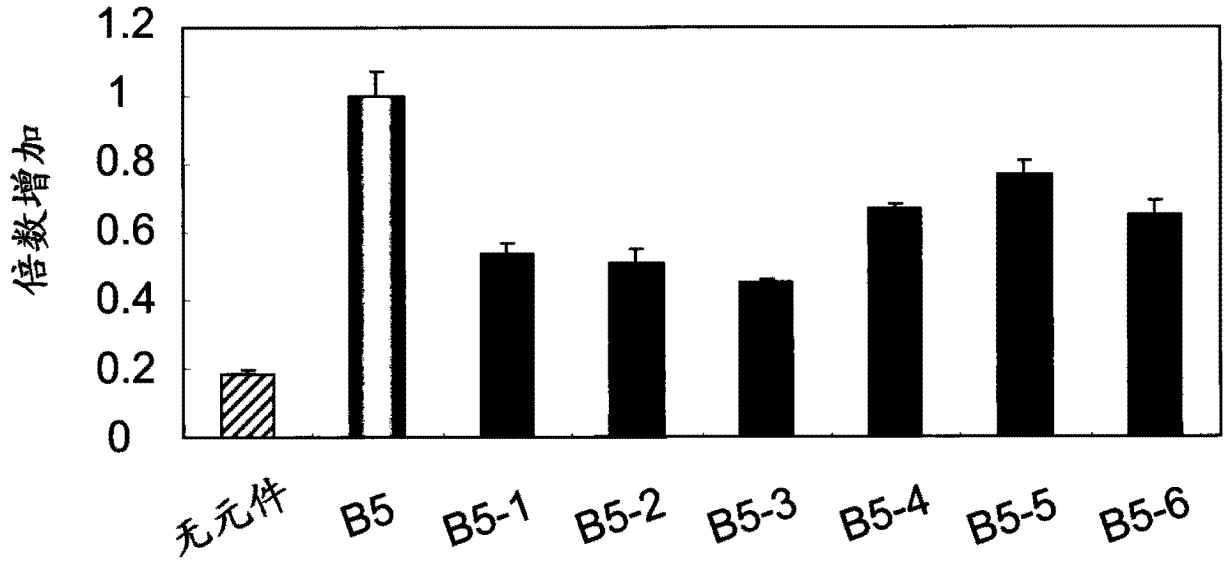


图 14

ID	起始位置	终止位置	长度(bp)
C14	46089056	46097482	8427
C14-1	46090015	46093070	3056
C14-2	46091042	46094069	3028
C14-3	46093075	46096174	3100
C14-4	46090015	46097196	7182
C14-5	46090015	46095066	5052
C14-6	46093994	46097196	3203
C14-7	46090015	46094069	4055
C14-8	46092049	46096174	4126
C14-9	46093075	46097196	4122
C14-10	46089056	46094069	5014
C14-11	46091042	46096174	5133
C14-12	46092049	46097196	5148
C14-13	46090015	46096174	6160
C14-14	46091042	46097196	6155

图 15

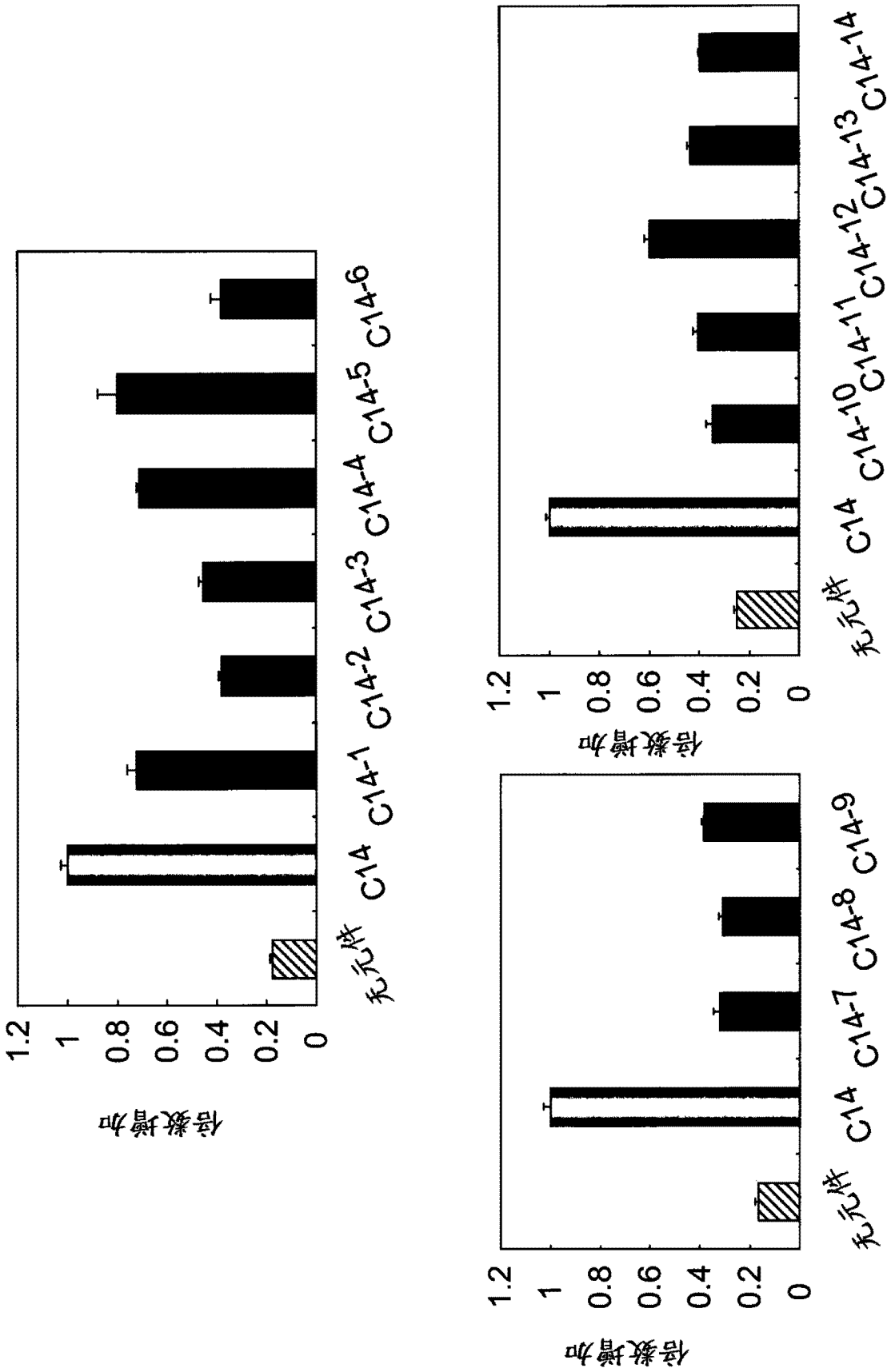

倍数增加
1.2
1
0.8
0.6
0.4
0.2
0
无元件
C14
C14-1
C14-2
C14-3
C14-4
C14-5
C14-6
倍数增加
1.2
1
0.8
0.6
0.4
0.2
0
无元件
C14
C14-10
C14-11
C14-12
C14-13
C14-14
倍数增加
1.2
1
0.8
0.6
0.4
0.2
0
无元件
C14
C14-7
C14-8
C14-9

图 16

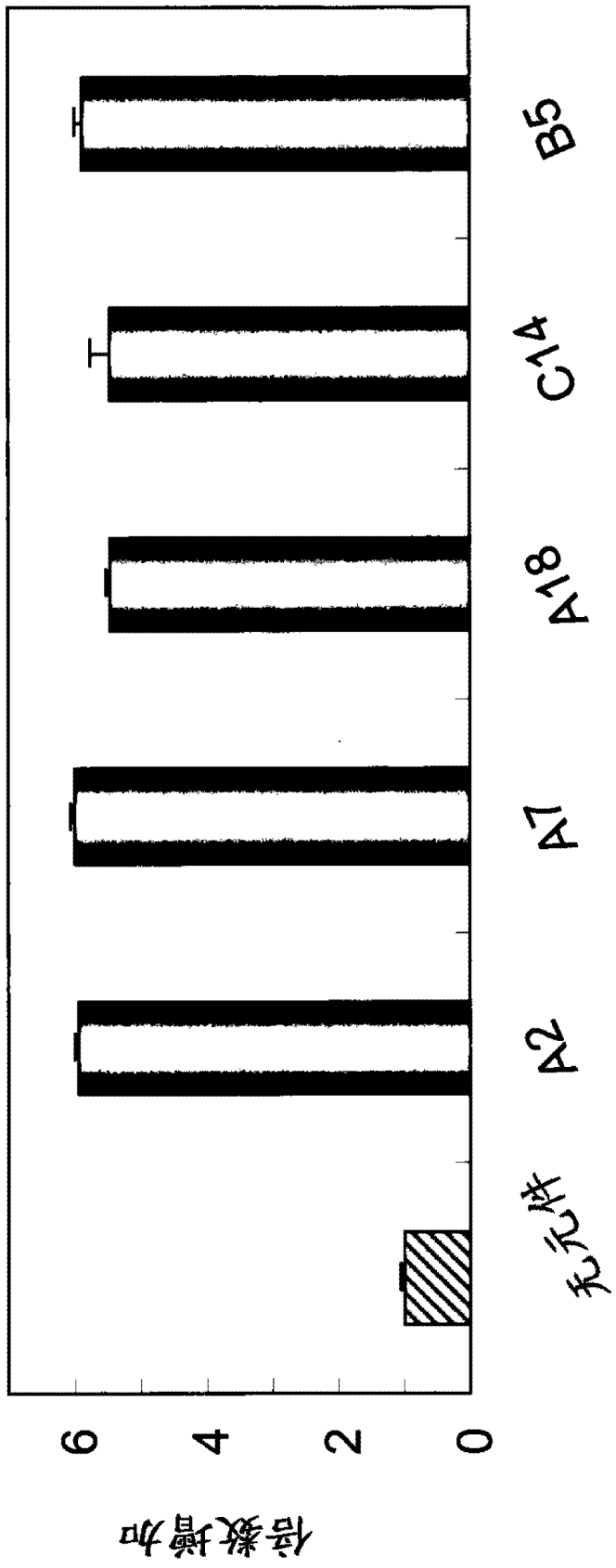


图 17

基于 DNA 元件 A2 全长序列的起点和终点

A2	起点	终点
A2-1	1	3000
A2-2	2801	5800
A2-3	5401	8450
A2-4	701	2700
A2-5	701	2200
A2-6	701	3700
A2-7	2001	5000
A2-8	4001	7000
A2-9	1	3700
A2-10	2001	5800
A2-11	2801	7000
A2-12	701	5800
A2-13	2001	7000
A2-14	2801	8450
A2-15	1	5800
A2-16	701	7000
A2-17	2001	8450

基于 DNA 元件 A7 全长序列的起点和终点

A7	起点	终点
A7-1	601	3600
A7-2	3601	8420
A7-3	5401	8420
A7-4	3401	6400
A7-5	1501	4500
A7-6	4401	7400
A7-7	2401	5400
A7-8	1	3600
A7-9	1501	5400
A7-10	2401	6400
A7-11	3401	7400
A7-12	4401	8420
A7-13	1	5400
A7-14	1501	6400
A7-15	2401	7400
A7-16	3401	8420
A7-17	1	6400
A7-18	1501	7400

基于 DNA 元件 A18 全长序列的起点和终点

A18	起点	终点
A18-1	1	5040
A18-2	1001	6002
A18-3	2001	7000
A18-4	3000	7000

图 18

基于 DNA 元件 B5 全长序列的起点和终点

	起点	终点
B5	1	8401
B5-1	1	4001
B5-2	1	3200
B5-3	2491	5601
B5-4	5373	8401
B5-5	901	4001
B5-6	4001	7000

基于 DNA 元件 C14 全长序列的起点和终点

	起点	终点
C14	1	8427
C14-1	960	4015
C14-2	1987	5014
C14-3	4020	7119
C14-4	960	8141
C14-5	960	6011
C14-6	4939	8141
C14-7	960	5014
C14-8	2994	7119
C14-9	4020	8141
C14-10	1	5014
C14-11	1987	7119
C14-12	2994	8141
C14-13	960	7119
C14-14	1987	8141

图 19