



(12) 发明专利申请

(10) 申请公布号 CN 104093834 A

(43) 申请公布日 2014. 10. 08

(21) 申请号 201280069001. 0

(22) 申请日 2012. 12. 05

(30) 优先权数据

61/567, 675 2011. 12. 07 US

(85) PCT国际申请进入国家阶段日

2014. 08. 04

(86) PCT国际申请的申请数据

PCT/IB2012/056977 2012. 12. 05

(87) PCT国际申请的公布数据

W02013/084157 EN 2013. 06. 13

(71) 申请人 格兰马克药品股份有限公司

地址 瑞士拉绍德封

(72) 发明人 D·吕舍尔 C·埃比谢尔-古米

P·莫莱迪 M·博特施格尔

(74) 专利代理机构 北京市中咨律师事务所

11247

代理人 史文静 黄革生

(51) Int. Cl.

C12N 15/09 (2006. 01)

权利要求书4页 说明书36页

序列表38页 附图10页

(54) 发明名称

表达盒

(57) 摘要

本发明涉及用于表达编码多肽的多核苷酸序列的表达盒。

1. 一种表达盒,其包含启动子、编码多肽的多核苷酸序列和表达增强元件,其中表达增强元件包含真核甘油醛 3- 磷酸脱氢酶 (GAPDH) 启动子下游的非翻译性基因组 DNA 序列,其中由多核苷酸序列编码的多肽不是 GAPDH,并且其中真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列始于跨越核苷酸位置约 +1 至核苷酸位置约 +7000 的区域内部,其中核苷酸位置是相对于 GAPDH mRNA 转录起始的,并且其中真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列的长度是约 100 至约 15000 个核苷酸。

2. 根据权利要求所述 1 的表达盒,其中表达盒还包含真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列,其中真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列始于跨越约真核 GAPDH 启动子的 5' 末端至核苷酸位置约 -3500 的区域内部,其中核苷酸位置是相对于 GAPDH mRNA 转录起始,并且其中真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列的长度是约 100 至约 15000 个核苷酸。

3. 一种表达盒,其包含启动子、编码多肽的多核苷酸序列和真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列,其中由多核苷酸序列编码的多肽不是 GAPDH,并且其中真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列始于跨越约真核 GAPDH 启动子的 5' 末端至核苷酸位置约 -3500 的区域内部,其中核苷酸位置是相对于 GAPDH mRNA 转录起始的,其中真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列的长度是 100 至约 15000 个核苷酸,条件是表达盒不包含真核 GAPDH 启动子或其片段。

4. 根据权利要求 3 所述的表达盒,其中表达盒还包含真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列,其中真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列始于跨越核苷酸位置约 +1 至核苷酸位置约 +7000 的区域内部,其中核苷酸位置是相对于 GAPDH mRNA 转录起始的,并且其中真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列的长度是约 100 至约 15000 个核苷酸。

5. 根据权利要求 1 至 4 中任一项所述的表达盒,其中真核 GAPDH 启动子下游和 / 或上游的非翻译性基因组 DNA 序列未与编码多肽的多核苷酸序列有效连接。

6. 根据权利要求 1 至 4 中任一项所述的表达盒,其中表达盒还包含多聚腺苷化位点。

7. 根据权利要求 1 或 4 所述的表达盒,其中真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列的长度是约 10 个核苷酸并且在最大情况下延伸至 IFF01 基因的倒数第二内含子或延伸至其部份。

8. 根据权利要求 1 或 4 所述的表达盒,其中真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列始于真核 GAPDH 多聚腺苷化位点的下游并且其中真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列的长度是至少约 100 个核苷酸并且在最大情况下延伸至 IFF01 基因的倒数第二内含子。

9. 根据权利要求 2 所述的表达盒,其中真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列的长度是至少约 100 个核苷酸并且在最大情况下延伸至 NCAPD2 基因的起始密码子。

10. 根据权利要求 2 所述的表达盒,其中真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列的长度是至少约 100 个核苷酸并且在最大情况下延伸至 NCAPD2 基因的倒数第三内含子。

11. 根据权利要求 3 所述的表达盒,其中真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列的长度是至少 100 个核苷酸并且在最大情况下延伸至 NCAPD2 基因的起始密码子。

12. 根据权利要求 3 所述的表达盒,其中真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列的长度是至少 100 个核苷酸并且在其最大情况下延伸至 NCAPD2 基因的倒数第三内含子。

13. 根据权利要求 1 至 4 中任一项所述的表达盒,其中真核 GAPDH 启动子下游和 / 或上游的非翻译性基因组 DNA 序列是哺乳动物源的。

14. 根据权利要求 13 所述的表达盒,其中真核 GAPDH 启动子下游和 / 或上游的非翻译性基因组 DNA 序列是啮齿类源或人源的。

15. 根据权利要求 1 或 4 所述的表达盒,其中真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列包含选自 SEQ ID NOs:8 和 21 的核苷酸序列或其片段。

16. 根据权利要求 1 或 4 所述的表达盒,其中真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列包含与选自 SEQ ID NOs:8 和 21 的核苷酸序列或其片段互补的核苷酸序列。

17. 根据权利要求 1 或 4 所述的表达盒,其中真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列包含与选自 SEQ ID NOs:8 和 21 的核苷酸序列或其片段相同至少 80% 的核苷酸序列。

18. 根据权利要求 2 或 3 所述的表达盒,其中真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列包含选自 SEQ ID NO:7、9、10、11、12、13、14、15、16、20、22、23、24、25、26、27 和 28 的核苷酸序列或其片段。

19. 根据权利要求 18 所述的表达盒,其中选自 SEQ ID NOs:7、9、10、11、12、13、14、15、16、20、22、23、24、25、26、27 和 28 的核苷酸序列或其片段包含 5 个或更少的核酸修饰。

20. 根据权利要求 2 或 3 所述的表达盒,其中真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列包含与选自 SEQ ID NO:7、9、10、11、12、13、14、15、16、20、22、23、24、25、26、27 和 28 的核苷酸序列或其片段互补的核苷酸序列。

21. 根据权利要求 2 或 3 所述的表达盒,其中真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列包含与选自 SEQ ID NO:7、9、10、11、12、13、14、15、16、20、22、23、24、25、26、27 和 28 的核苷酸序列或其片段相同至少 80% 的核苷酸序列。

22. 根据权利要求 1 至 4 中任一项所述的表达盒,其中启动子和编码多肽的多核苷酸序列是有效连接的。

23. 根据权利要求 1 或 4 所述的表达盒,其中真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列以与编码多肽的多核苷酸序列相同的方向定向。

24. 根据权利要求 1 或 4 所述的表达盒,其中真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列相对于编码多肽的多核苷酸序列以相反方向定向。

25. 根据权利要求 2 或 3 所述的表达盒,其中真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列以与编码多肽的多核苷酸序列相同的方向定向。

26. 根据权利要求 2 或 3 所述的表达盒,其中真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列相对于编码多肽的多核苷酸序列以相反方向定向。

27. 根据权利要求 1 至 4 中任一项所述的表达盒,其中启动子选自 SV40 启动子、MPSV 启动子、小鼠 CMV、人 tk、人 CMV、大鼠 CMV、人 EF1 α 、中国仓鼠 EF1 α 、人 GAPDH、杂合启动子,所述杂合启动子包括 MYC、HYK 和 CX 启动子。

28. 根据权利要求 1 至 4 中任一项所述的表达盒,其中多肽选自抗体、抗体片段或抗体

衍生物。

29. 根据权利要求 6 所述的表达盒,其中多聚腺苷化位点选自 BGH poly(A) 和 SV40poly(A)。

30. 根据权利要求 1 至 4 中任一项所述的表达盒,还包含选自额外的启动子、增强子、转录控制元件和选择标记的遗传元件。

31. 根据权利要求 30 所述的表达盒,其中遗传元件是选择标记,其中在编码选择标记的多核苷酸序列中所含的 CpG 位点的含量是 45 或更小。

32. 表达载体,其包含根据权利要求 1 至 29 中任一项所述的表达盒。

33. 表达载体,其依次包含:

- a) 真核 GAPDH 启动子上游和 / 或下游的非翻译性基因组 DNA 序列
- b) 启动子
- c) 编码多肽的多核苷酸序列
- d) 多聚腺苷化位点
- e) 增强子
- f) 真核 GAPDH 启动子下游和 / 或上游的非翻译性基因组 DNA 序列,或
- a) 真核 GAPDH 启动子上游和 / 或下游的非翻译性基因组 DNA 序列
- b) 增强子
- c) 启动子
- d) 编码多肽的多核苷酸序列
- e) 多聚腺苷化位点
- f) 真核 GAPDH 启动子下游和 / 或上游的非翻译性基因组 DNA 序列,或
- a) 增强子
- b) 真核 GAPDH 上游和 / 或下游的非翻译性基因组 DNA 序列
- c) 启动子
- d) 编码多肽的多核苷酸序列
- e) 多聚腺苷化位点
- f) 真核 GAPDH 下游和 / 或上游的非翻译性基因组 DNA 序列,

其中包含增强子是任选的,并且其中由多核苷酸序列编码的多肽不是 GAPDH,并且其中真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列始于跨越核苷酸位置约 +1 至核苷酸位置约 +7000 的区域内部,其中核苷酸位置是相对于 GAPDH mRNA 转录起始的,并且其中真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列的长度是约 100 至约 15000 个核苷酸并且其中真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列始于跨越约真核 GAPDH 启动子的 5' 末端至核苷酸位置约 -3500 的区域内部,其中核苷酸位置是相对于 GAPDH mRNA 转录起始的,并且其中真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列的长度是约 100 至约 15000 个核苷酸,条件是如果 a) 或 b) 是真核 GAPDH 上游的非翻译性基因组 DNA 序列,则 f) 是真核 GAPDH 下游的非翻译性基因组 DNA 序列,并且如果 a) 或 b) 是真核 GAPDH 下游的非翻译性基因组 DNA 序列,则 f) 是真核 GAPDH 上游的非翻译性基因组 DNA 序列。

34. 根据权利要求 32 或 33 所述的表达载体,其中表达载体还包含选自额外的启动子、增强子,转录控制单元、复制起点和选择标记的遗传元件。

35. 根据权利要求 32 或 33 所述的表达载体,其中表达载体还包含复制起点和选择标记,其中表达载体的编码复制起点和选择标记的多核苷酸序列中所含的 CpG 位点的含量是 200 或更小。

36. 宿主细胞,其包含根据权利要求 1 至 31 中任一项所述的表达盒或根据权利要求 32 至 35 中任一项所述的表达载体。

37. 根据权利要求 1 至 31 中任一项所述的表达盒或根据权利要求 32 至 35 中任一项所述的表达载体,用于作为用于治疗疾病的药物。

38. 根据权利要求 1 至 31 中任一项所述的表达盒或根据权利要求 32 至 35 中任一项所述的表达载体,用于基因疗法。

39. 用于表达多肽的体外方法,包括用根据权利要求 1 至 31 中任一项所述的表达盒或根据权利要求 32 至 35 中任一项所述的表达载体转染宿主细胞并回收多肽。

40. 根据权利要求 39 所述的方法,其中将表达盒或表达载体是稳定转染。

41. 根据权利要求 39 所述的方法,其中将表达盒或表达载体是瞬时转染。

42. 根据权利要求 1 至 31 中任一项所述的表达盒或根据权利要求 32 至 35 中任一项所述的表达载体的用途,用于从哺乳动物宿主细胞表达异源多肽。

表达盒

[0001] 相关申请

[0002] 本申请要求 2011 年 12 月 7 日递交的 US 临时申请号 61/567,675 的权益 ;所有这些通过引用以其整体并入本文。

发明领域

[0003] 本发明涉及用于表达编码多肽的多核苷酸序列的表达盒。本发明是还涉及包含表达盒的载体和宿主细胞以及表达盒用于从宿主细胞产生多肽的用途。

背景技术

[0004] 用于产生重组多肽的表达系统是现有技术中熟知的并且由例如, Marino MH(1989)Biopharm, 2:18-33 ;Goeddel DV 等人, (1990)Methods Enzymol 185:3-7 ;Wurm F 和 Bernard A(1999)Curr Opin Biotechnol 10:156-159 描述。用于药物应用中的多肽优选地在哺乳动物细胞如 CHO 细胞、NSO 细胞、SP2/0 细胞、COS 细胞、HEK 细胞、BHK 细胞等中产生。用于此目的的表达载体的必需元件通常选自包含原核复制起点和原核选择标记, 任选地真核选择标记的原核质粒增殖单元, 例如大肠杆菌 (E. coli), 和用于表达一个或多个目的结构基因的一个或多个表达盒, 所述表达盒每一者包含启动子、编码多肽的多核苷酸序列和任选地转录终止子, 所述转录终止子包含多聚腺苷化信号。为了在哺乳动物细胞中瞬时表达, 可以包括哺乳动物复制起点, 如 SV40ori 或 OriP。作为启动子, 可以选择组成型或诱导型启动子。为了优化的转录, 可以在 5' 非翻译区中包含 Kozak 序列。为了 mRNA 加工, 尤其 mRNA 剪接和转录终止, 取决于结构基因的组织状况 (外显子 / 内含子组织状况), 可以包括 mRNA 剪接信号, 以及多聚腺苷化信号。基因的表达以瞬时方式或使用稳定的细胞系进行。生产细胞系中多肽的稳定和高度表达的水平对于产生重组多肽的整个过程至关重要。在过去数年间, 对生物分子如蛋白质并且特别对抗体或抗体片段的需求已经显著地增加。高成本和不良产率已经是生物分子可用性方面的限制因素并且它已经成为发展以工业规模增加所需生物分子的产量的稳健工艺的主要挑战。因而仍需要改善表达载体的效率以在重组多肽生产中实现高表达。

[0005] 发明简述

[0006] 本发明一般地涉及可以用来在重组多肽生产中实现表达增加的表达系统如表达盒和表达载体。在一个方面, 本公开提供表达盒, 其包含启动子、编码多肽的多核苷酸序列和真核甘油醛 3-磷酸脱氢酶 (GAPDH) 启动子下游的非翻译性基因组 DNA 序列, 其中由多核苷酸序列编码的多肽不是 GAPDH, 并且其中真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列始于跨越从核苷酸位置约 +1 至核苷酸位置约 +7000 的区域内部, 其中核苷酸位置是相对于 GAPDH mRNA 转录起始的, 并且其中真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列的长度是约 100 至约 15000 个核苷酸。

[0007] 在又一个方面, 本公开提供表达盒, 其包含启动子、编码多肽的多核苷酸序列和真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列, 其中由多核苷酸序列编码的多肽不

是 GAPDH, 并且其中真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列始于跨越大约真核 GAPDH 启动子的 5' 末端至核苷酸位置约 -3500 的区域内部, 其中核苷酸位置是相对于 GAPDH mRNA 转录起始的, 其中真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列的长度是 100 至约 15000 个核苷酸, 条件是表达盒不包含真核 GAPDH 启动子或其片段。

[0008] 在又一个方面, 本公开提供包含表达盒的表达载体和包含表达盒或表达载体的宿主细胞, 其中所述表达载体包含表达盒。

[0009] 在其他方面, 本公开提供一种用于表达多肽的体外方法, 所述方法包括用表达盒或表达载体转染宿主细胞并回收多肽, 和表达盒或表达载体用于从哺乳动物宿主细胞表达异源多肽的用途。

[0010] 附图简述

[0011] 图 1 显示报道基因表达构建体 (REP), 其由小鼠巨细胞病毒启动子 (mCMV)、含有第一内含子的 Ig 供体接纳体片段 (IgDA)、IgG1 抗体轻链 (IgG1LC)、源自脑心肌炎病毒的内部核糖体进入位点 (IRES)、IgG1 抗体重链 (IgG1HC)、绿色荧光蛋白 (GFP) 和猿猴病毒 40 多聚腺苷化信号 (poly(A)) 组成。

[0012] 图 2 显示在转染后第 5 日 IgG1 抗体在 CHO-S 细胞中的瞬时表达 (将 IgG 滴度的均值对两次独立转染作图)。使用 GAPDH_A 和 GAPDH_B 载体 (GAPDH_A 和 GAPDH_B)、无 GAPDH 上游元件和下游元件的相同载体 (A 和 B) 和作为对照的 pGLEX41 载体 (pGLEX41), 转染细胞。使用 Octet 仪器 (Fortebio, Menlo, CA, 美国) 确定上清液中积累的 IgG1 抗体的浓度。

[0013] 图 3 显示 IgG1 抗体在 HEK293EBNA 细胞中的表达。使用 GAPDH_A 和 GAPDH_B 载体 (GAPDH_A 和 GAPDH_B) 和作为对照的 pGLEX41 载体 (pGLEX41), 转染细胞。将上清液收获并在转染后第 10 日使用 Octet 仪器分析。数据代表每个载体在旋转管 (tubespin) 中的 N = 3 个独立转染。

[0014] 图 4 显示使用细胞汇集物在分批生产时的表达水平研究。使用 GAPDH_A 和 GAPDH_B 载体 (GAPDH_A(1)、GAPDH_A(2)、GAPDH_B(1) 和 GAPDH_B(2))、无 GAPDH 上游元件和下游元件的相同载体 (A(1) 和 A(2)) 和作为对照的 pGLEX41 载体 (pGLEX41), 转染细胞并且产生稳定细胞的汇集物。在 7 日培养后, 使用 Octet 仪器对上清液分析上清液中积累的抗体。给出每种汇集物的 IgG 滴度均值 (μ g/ml)。数据代表每种汇集物 N = 2 个批次。

[0015] 图 5 显示在通过稳定转染和有限稀释法产生的群体上的表达水平研究。使用 GAPDH_A 和 GAPDH_B 载体 (GAPDH_A 和 GAPDH_B)、无 GAPDH 上游元件和下游元件的相同载体 (A 和 B) 和作为对照的 pGLEX41 载体 (pGLEX41), 转染细胞。在转染后 14 日读取来自稳定转染的克隆和小量汇集物所表达的 GFP 荧光的均值。将细胞在 96 孔板中选择压力下培育。数据代表每种载体 N = 48 个克隆或小量汇集物。

[0016] 图 6 显示培养基添加物胰岛素和 PMA (佛波醇 12-肉豆蔻酸酯 13-乙酸酯, 一种佛波醇酯) 对上清液中 IgG1 抗体表达的影响。用 GAPDH_A 载体 (GAPDH_A) 和作为对照的 pGLEX41 载体 (pGLEX41) 转染后, 将细胞稀释于 PowerCHO2 培养基、4mM Gln、+/- 胰岛素和 PowerCHO2、4mM Gln、PMA +/- 胰岛素中。与标准培养基相比, 对 pGLEX41 (实心棒) 或 GAPDH_A (空心棒) 没有观察到表达差异。

[0017] 图 7 显示人 GAPDH 基因座的概览。GAPDH 基因侧翼为基因 NCAPD2 和 IFFO1。

[0018] 图 8 显示人 GAPDH 基因、GAPDH 上游元件和下游元件和为分析 GAPDH 上游片段化

研究所产生的片段的细节。引入 NruI 限制性位点以促进克隆步骤并且该位点不是基因组 5' GAPDH 上游序列的部分（因此使用星号突出显示它）。片段的大小是：片段 1 (SEQ ID NO:9):511bp、片段 2 (SEQ ID NO:10):2653bp、片段 3 (SEQ ID NO:11):1966bp、片段 4 (SEQ ID NO:12):1198bp、片段 8 (SEQ ID NO:13):259bp、片段 9 (SEQ ID NO:14):1947bp、片段 11 (SEQ ID NO:15):1436bp 和片段 17 (SEQ ID NO:16):1177bp。

[0019] 图 9 显示 GAPDH 上游元件和下游元件片段化的表达结果。在转染后第 10 日在 CHO 细胞中以瞬时转染方式获得表达结果。使用 Octet 仪器进行定量。载体 pGLEX41 充当阴性对照。pGLEX41-amp^r 还是显示无 GAPDH 侧翼元件的载体基础表达的阴性对照。pGLEX41- 上 / 下含有全长侧翼区（上游区和下游区）并充当阳性对照。pGLEX41- 上仅含有上游侧翼区并且 pGLEX41- 下仅含有下游侧翼区。全部其他构建体均含有图 8 中描述的片段。片段 2 和片段 3 按照与 IgG1LC 和 IgG1HC 相同的方向克隆或相对于 IgG1LC 和 IgG1HC (AS) 以相反方向克隆。

[0020] 图 10 显示在转染后第 8 日 IgG1 抗体在 CHO-S 细胞中的瞬时表达（将 IgG 滴度的均值对三次独立转染作图；误差条：SD+/-）。使用具有中国仓鼠 GAPDH 上游元件与小鼠 CMV (A_GAPDH_UP) 或中国仓鼠 GAPDH 启动子 (A_GAPDH_UP_PR) 组合的载体，转染细胞。转染仅具有小鼠 CMV (A) 或中国仓鼠 GAPDH 启动子 (A_PR) 的质粒作为对照。使用 Octet QK 仪器 (Fortebio, Menlo, CA, 美国) 确定上清液中积累的 IgG1 抗体的浓度。

[0021] 发明详述

[0022] 本公开涉及表达盒和表达载体，所述表达盒和表达载体包含启动子、编码多肽的多核苷酸序列和真核甘油醛 3- 磷酸脱氢酶 (GAPDH) 启动子下游的非翻译性基因组 DNA 序列，其中由多核苷酸序列编码的多肽不是 GAPDH，并且其中真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列始于跨越核苷酸位置约 +1 至核苷酸位置约 +7000 的区域内部，其中核苷酸位置是相对于 GAPDH mRNA 转录起始的，并且其中真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列的长度是约 100 至约 15000 个核苷酸。

[0023] 本公开还涉及一种表达盒，其包含启动子、编码多肽的多核苷酸序列和真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列，其中由多核苷酸序列编码的多肽不是 GAPDH，并且其中真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列始于跨越大约真核 GAPDH 启动子的 5' 末端至核苷酸位置约 -3500 的区域内部，其中核苷酸位置是相对于 GAPDH mRNA 转录起始的，其中真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列的长度是 100 至约 15000 个核苷酸，条件是表达盒不包含真核 GAPDH 启动子或其片段。

[0024] 如本文所用的术语“表达盒”包含多核苷酸序列，所述多核苷酸序列编码待表达的多肽和控制其表达的序列如启动子以及任选地增强子序列，包括顺式作用转录性控制单元的任何组合。控制基因表达（即其转录和转录产物翻译）的序列常称作调节单元。调节单元的大多数部分位于基因的编码序列上游并与之有效连接。表达盒也可以含有包含多聚腺苷化位点的下游 3' 非翻译区。本发明的调节单元与待表达的基因即转录单位有效连接，或由间插 DNA 例如异源基因的 5' - 非翻译区分隔。优选地，表达盒侧翼为一种或多种合适的限制性位点以便可以将表达盒插入载体中和 / 或将其从载体切除。因而，本发明的表达盒可以用于构建表达载体，尤其哺乳动物表达载体。本发明的表达盒可以包含一个或多个例如两个、三个或甚至更多个在真核 GAPDH 启动子或其片段下游的非翻译性基因组 DNA 序列，

和 / 或一个或多个例如两个、三个或甚至更多个在真核 GAPDH 启动子或其片段上游的非翻译性基因组 DNA 序列。如果本发明的表达盒包含多于一个在真核 GAPDH 启动子或其片段下游和 / 或上游的 DNA 序列,则这些 DNA 序列可以是直接连接的,即可以包含接头序列,例如含有与 5' 末端和 3' 末端连接并允许轻易依次克隆所述序列或其片段的限制性位点的接头序列。备选地,真核 GAPDH 启动子或其片段下游和 / 或上游的 DNA 序列可以不是直接连接的,即可以随间插 DNA 序列一起克隆。

[0025] 如本文所用的术语“编码多肽的多核苷酸序列”包括编码表达所述多肽的基因、优选地异源基因的 DNA。

[0026] 术语“异源编码序列”、“异源基因序列”、“异源基因”、“重组基因”或“基因”互换使用。这些术语指这样的 DNA 序列,所述 DNA 序列编码设法在宿主细胞中、优选地在哺乳动物细胞中表达并收获的重组蛋白产物,尤其重组异源蛋白产物。基因的产物可以是多肽。异源基因序列天然地不存在于宿主细胞中并且源自相同或不同物种的生物并且可以是基因修饰的。

[0027] 术语“蛋白质”和“多肽”互换地用来包括通过相邻残基的 α -氨基和羧基之间的肽键彼此连接的一串氨基酸残基。

[0028] 如本文所用的术语“非翻译性基因组 DNA 序列”包括构成生物的遗传信息的 DNA。几乎全部生物的基因组是 DNA,唯一例外是具有 RNA 基因组的一些病毒。大多数生物中的基因组 DNA 分子被组织成称作染色体的 DNA-蛋白质复合体。染色体的大小、数目和基因组 DNA 的性质在不同生物之间不同。病毒 DNA 基因组可以是单链或双链、线型或环状的。全部其他生物具有双链 DNA 基因组。细菌具有单个环状染色体。在真核生物中,大多数基因组 DNA 作为大小不同的多条线形染色体位于核内(核 DNA)。真核细胞额外地在线粒体中含有基因组 DNA,并且在植物和低等真核生物中,在叶绿体中含有基因组 DNA。这种 DNA 通常是环状分子并且在这些细胞器内部作为多个副本存在。非翻译性基因组 DNA 序列通常不与启动子有效连接并且因此不翻译。它可以含有不翻译的基因,因此含有编码例如不表达的蛋白质的基因。

[0029] 如本文所用的术语“真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列”对应于真核 GAPDH 启动子 3' 的非翻译性真核基因组 DNA。真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列通常在核苷酸位置约 +1 处,优选地在核苷酸位置 +1 处开始,其中核苷酸位置是相对于 GAPDH mRNA 转录起始的,即相对于编码 GAPDH 的真核基因的转录起始(transcription start)的起点(origin)。真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列通常具有与真核 GAPDH 启动子相同的来源,例如,如果该 GAPDH 启动子是人源的,则人 GAPDH 启动子下游的非翻译性基因组 DNA 序列也是人源的并对应于人 GAPDH 启动子下游天然存在的人类基因组 DNA 序列。

[0030] 如本文所用的术语“真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列”对应于真核 GAPDH 启动子 5' 的非翻译性真核基因组 DNA。真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列通常在约真核 GAPDH 启动子的 5' 末端处的核苷酸位置处,优选地在紧邻真核 GAPDH 启动子 5' 末端之后的核苷酸位置处开始。真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列通常具有与真核 GAPDH 启动子相同的来源,例如,如果该 GAPDH 启动子是人源的,则人 GAPDH 启动子上游的非翻译性基因组 DNA 序列也是人源的并对应于人 GAPDH 启动子上游

天然存在的人类基因组 DNA 序列。

[0031] 如非另外具体说明,则真核 GAPDH 启动子、真核 GAPDH 启动子下游或上游的非翻译性基因组 DNA 序列和如本文中所示的其他 DNA 序列的位置是相对于 GAPDH mRNA 转录起始的,例如是相对于真核 GAPDH 的转录起始的起点的。

[0032] 术语“真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列延伸至”或“真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列延伸至”用来限定真核 GAPDH 启动子上游和 / 或下游的非翻译性基因组 DNA 序列的长度从起始至特定遗传元件的长度的延伸,例如至内含子的延伸。这种延伸包括编码该遗传元件例如内含子或其部份的 DNA 序列的全长。

[0033] 人、大鼠和小鼠的真核 GAPDH 启动子和 GAPDH 启动子上游和 / 或下游的真核基因组 DNA 可以在 NCBI 公共数据库中找到(人、小鼠、大鼠和中国仓鼠 GAPDH 基因的条目分别是基因 ID 2597(mRNA:NM_002046.3)、14433(mRNA:NM_008084.2)、24383(mRNA:NM_017008.3)和 100736557(mRNA:NM_001244854.2;美国国家生物技术信息中心(NCBI):<http://www.ncbi.nlm.nih.gov/>)并且示意地在图 7 和 8 中针对人 GAPDH 基因显示。

[0034] 通常认为真核 GAPDH 启动子相对于 GAPDH mRNA 转录起始从约 bps-500 伸展至约 +50。人 GAPDH 启动子位于第 12 号染色体上。基于片段化研究,Graven 等人(Graven 等人,(1999)Biochimica et Biophysica Acta,147:203-218)认为人 GAPDH 启动子相对于 GAPDH mRNA 转录起始从 bps-488 伸展至 +20。根据 NCBI 公共数据库,人 GAPDH 启动子相对于如 NCBI 公共数据库所定义的 GAPDH mRNA 转录起始从 bps-462 伸展至 +46。如非另外具体说明,如本文提到的人 GAPDH 启动子相对于 GAPDH mRNA 转录起始从位置 -462 伸展至位置 +46,这对应于从 SEQ ID NO:17 的 bps 4071 伸展至 4578 的序列。

[0035] 如本文中提到的针对人、小鼠和大鼠源 GAPDH 基因、IFF01 基因和 NCAPD2 基因的 DNA 所用的编号对应于 NCBI 公共数据库中针对这些基因所用的编号(<http://www.ncbi.nlm.nih.gov/>)。

[0036] 如本文所用的术语“启动子”限定一般位于基因上游的调节性 DNA 序列,所述调节性 DNA 序列通过指导 RNA 聚合酶与 DNA 结合并启动 RNA 合成介导转录的起始。

[0037] 如本文所用的术语“增强子”限定一种核苷酸序列,该核苷酸序列作用为使激动基因转录成为可能,所述基因转录与基因的身份、该序列相对于基因的位置或序列的取向无关。本发明的载体任选地包含增强子。

[0038] 术语“功能性连接”和“有效连接”互换使用并且指两个或更多个 DNA 区段、尤其待表达基因序列和控制其表达的那些序列之间的功能关系。例如,如果启动子和 / 或增强子序列(包括顺式作用转录性控制元件的任何组合)在适宜的宿主细胞或其他表达系统中刺激或调节编码序列的转录,则它与编码序列有效连接。与转录的基因序列有效连接的启动子调节序列是在物理上与转录的序列邻接的。

[0039] “取向”指给定 DNA 序列中的核苷酸顺序。例如,DNA 序列相对于另一个 DNA 序列以相反方向的取向是这种情况,其中与从其中获得该序列的 DNA 中的参照点相比时,该序列相对于另一个序列的 5' 至 3' 顺序反转。这类参考点可以包括源 DNA 中转录其他指定 DNA 序列的方向和 / 或含有该序列的复制型载体的复制起点。

[0040] 如本文所用的术语“表达载体”包括分离和纯化的 DNA 分子,所述分子在转染入适宜的宿主细胞时在宿主细胞内部提供重组基因产物的高水平表达。除编码重组或基因产物

的 DNA 序列之外,表达载体还包含宿主细胞系中将 DNA 编码序列有效转录成 mRNA 并将 mRNA 有效翻译成蛋白质所需要的调节性 DNA 序列。

[0041] 如本文所用的术语“宿主细胞”或“宿主细胞系”包括能够在培养物中生长并表达所需重组产物蛋白质的任何细胞、尤其哺乳动物细胞。

[0042] 如本文所用的术语“片段”包括相应核苷酸序列的一部分,例如真核 GAPDH 启动子下游和 / 或上游的非翻译性基因组 DNA 序列的一部分或编码特定遗传元件 (如启动子) 的核苷酸序列的一部分。真核 GAPDH 启动子下游和 / 或上游的非翻译性基因组 DNA 序列的片段可以保留生物活性并因此改变 (例如增加) 与启动子有效连接的编码序列的表达模式。真核 GAPDH 启动子下游和 / 或上游的非翻译性基因组 DNA 序列的片段范围从至少约 100 至约 3000bp、优选地从约 200 至约 2800bp、更优选地从约 300 至约 2000bp 核苷酸、尤其从约 500 至约 1500bp 核苷酸。为了在本发明的表达盒中克隆真核 GAPDH 启动子下游和 / 或上游的非翻译性基因组 DNA 序列的片段,通常将含有允许便利克隆的限制性位点的接头序列与所述片段的 5' 末端和 3' 末端连接。

[0043] 如本文所用的术语“核苷酸序列同一性”或“相同核苷酸序列”包括在比对序列并引入空位 (如果需要) 以实现最大序列同一性百分数后,候选序列中与例如真核 GAPDH 启动子下游和 / 或上游的非翻译性基因组 DNA 序列的核苷酸序列相同的核苷酸的百分数。因此,可以通过常用来比较两个核苷酸序列的核苷酸位置中相似性的标准方法确定序列同一性。通常,相对于真核 GAPDH 启动子下游和 / 或上游的非翻译性基因组 DNA 序列,候选序列的核苷酸序列同一性是至少 80%、优选地至少 85%、更优选地至少 90% 并且最优选地是至少 95%、尤其 96%,更特别地是 97%,甚至更特别地是 98%、最特别地是 99%,包括例如,80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99% 和 100%。

[0044] 如本文所用的术语“CpG 位点”包括其中胞嘧啶核苷酸在线性碱基序列中沿其长度紧邻鸟嘌呤核苷酸出现的 DNA 区域。“CpG”是“-C- 磷酸酯 -G-”的速记,即,由仅一个磷酸酯分隔的胞嘧啶和鸟嘌呤;磷酸酯将 DNA 中的任何两个核苷连接在一起。“CpG”记号用来区分这种线性序列与胞嘧啶和鸟嘌呤的 CG 碱基配对。

[0045] 如本文所用的术语“备选密码子选择”包括使用编码相同氨基酸的备选密码子以避免 CpG 序列基序。这包括优选地使用不具有内部 CpG 位点的密码子 (例如编码丙氨酸并含有 CpG 位点的 GCG,可由 GCT、GCC 或 GCA 替换) 以及避免产生新 CpG 位点的两个密码子连接。

[0046] 在涉及 DNA 序列长度和涉及相对于 GAPDH mRNA 转录起始例如相对于真核 GAPDH 转录起始起点的核苷酸位置时,本文所用的术语“约”包括偏离所述值最多 $\pm 50\%$ 、通常最多 $\pm 10\%$ 的值,例如“约 3000 个核苷酸”包括 2700 至 3300 个核苷酸、优选地 2900 至 3100 个核苷酸、更优选地 2995 至 3005 个核苷酸的值,“约 100 个核苷酸”包括 50 至 150 个核苷酸、优选地 90 至 110 个核苷酸、更优选地 95 至 105 个核苷酸的值,“约 15000 个核苷酸”包括 13500 至 16500 个核苷酸、优选地 14500 至 15500 个核苷酸、更优选地 14990 至 15010 个核苷酸、最优选地 14995 至 15005 个核苷酸的值,“约 200 个核苷酸”包括 150 至 250 个核苷酸、优选地 190 至 210 个核苷酸、更优选地 195 至 205 个核苷酸的值,“约 8000 个核苷酸”包括 7200 至 8800、优选地 7500 至 8500 个核苷酸、更优选地 7990 至 8010 个核苷酸、最优

选地 7995 至 8005 个核苷酸的值,“约 500 个核苷酸”包括 450 至 550 个核苷酸、优选地 475 至 525、更优选地 490 至 510、最优选地 495 至 505 个核苷酸的值,“约 5000 个核苷酸”包括 4500 至 5500 个核苷酸、优选地 4750 至 5250、更优选地 4990 至 5010、最优选地 4995 至 5005 个核苷酸的值,“约 1000 个核苷酸”包括 900 至 1100 个核苷酸、优选地 950 至 1050、更优选地 990 至 1010、最优选地 995 至 1005 个核苷酸的值,“约 4500 个核苷酸”包括 4050 至 4950 个核苷酸、优选地 4250 至 4750、更优选地 4490 至 4510、最优选地 4495 至 4505 个核苷酸的值,“约 1500 个核苷酸”包括 1350 至 1650 个核苷酸、优选地 1450 至 1550、更优选地 1490 至 1510、最优选地 1495 至 1505 个核苷酸的值,“约 4000 个核苷酸”包括 3600 至 4400 个核苷酸、优选地 3800 至 4200、更优选地 3990 至 4010、更优选地 3995 至 4005 个核苷酸的值,“约 2000 个核苷酸”包括 1800 至 2200 个核苷酸、优选地 1900 至 2100、更优选地 1990 至 2010、最优选地 1995 至 2005 个核苷酸的值,“约 3500 个核苷酸”包括 3150 至 3850 个核苷酸、优选地 3300 至 3700、更优选地 3490 至 3510、最优选地 3495 至 3505 个核苷酸的值,“约 2700 个核苷酸”包括 2430 至 2970 个核苷酸、优选地 2600 至 2800、更优选地 2690 至 2710、最优选地 2695 至 2705 个核苷酸的值,“约 3300 个核苷酸”包括 2970 至 3630 个核苷酸,优选地 3100 至 3500、更优选地 3290 至 3310、最优选地 3295 至 3305 个核苷酸的值,“约 3200 个核苷酸”包括 2880 至 3520 个核苷酸、优选地 3000 至 3400、更优选地 3190 至 3210、最优选地 3195 至 3205 个核苷酸,约 +7000 或约位置 +7000 包括位置 +6300 至 +7700、优选地位置 +6700 至 +7300、更优选地位置 +6990 至 +7010、最优选地位置 +6995 至 +7005,约 +1 或约位置 +1 包括位置 -10 至 +10、优选地位置 -5 至 +5、更优选地位置 -1 至 +2,约 -3500 或约位置 -3500 包括位置 -3150 至 -3850、优选地位置 -3300 至 -3700、更优选地位置 -3490 至 -5010、最优选地位置 -3495 至 -3505。

[0047] 对于与 SEQ ID 号的序列中的位置相关的本文所涉及的或本文所用的针对人、小鼠和大鼠源的 GAPDH 基因、IFF01 基因和 NCAPD2 基因的 DNA 所用的编号,术语“约”包括偏离最多 ± 500 bp、优选地 ± 100 bp、更优选地 ± 10 bp、最优选地 ± 5 bp 的值。

[0048] 在一个实施方案中,本公开提供一种表达盒,其包含启动子、编码多肽的多核苷酸序列和真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列,其中由多核苷酸序列编码的多肽不是 GAPDH,并且其中真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列始于跨越核苷酸位置约 +1 至核苷酸位置约 +7000 的区域内部,其中核苷酸位置是相对于 GAPDH mRNA 转录起始的,并且其中真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列的长度是约 100 至约 15000 个核苷酸。

[0049] 在一个实施方案中,真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列的长度是至少约 100 个核苷酸并且在其最大情况下延伸至 IFF01 基因的倒数第二内含子或延伸至其部份。在一个实施方案中,真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列的长度是至少约 100 个核苷酸并且在其最大情况下延伸至 IFF01 基因的最末内含子。

[0050] 人 IFF01 基因位于第 12 号染色体 (NCBI 基因 ID :25900) 约 bp 6665249 至 bp 6648694 的人 DNA 中。在一个实施方案中,在其最大情况下延伸至人中 IFF01 基因的最末内含子的真核 GAPDH 启动子下游非翻译性基因组 DNA 序列的长度在其最大情况下伸展至人中编码 IFF01 基因的第 12 号染色体的约 bp 6650677 (位置 +7021)。在一个实施方案中,在其最大情况下延伸至人中 IFF01 基因的倒数第二内含子的真核 GAPDH 启动子下游非翻译性

基因组 DNA 序列的长度在其最大情况下伸展至人中编码 IFF01 基因的第 12 号染色体的约 bp 6657230 (位置 +13574)。在显示第 12 号染色体 (NCBI 基因 ID :25900) 的 bp 6657230 至 bp 6639125 的 SEQ ID NO:17 中包含在其最大情况下分别伸展至人中 IFF01 基因的最末内含子和人中 IFF01 基因的倒数第二内含子的真核 GAPDH 启动子下游非翻译性基因组 DNA 序列。伸展至最末内含子的真核 GAPDH 启动子下游非翻译性基因组 DNA 序列伸展至如 SEQ ID NO:17 所示的核苷酸序列的约 bp 11553, 并且伸展至倒数第二内含子的真核 GAPDH 启动子下游非翻译性基因组 DNA 序列伸展至如 SEQ ID NO:17 所示核苷酸序列的约 bp 18106。

[0051] 小鼠 IFF01 基因 (NCBI 基因 ID :320678) 位于第 6 号染色体的约 bp125095259 至 125111800 的小鼠 DNA 中。在一个实施方案中, 在其最大情况下伸展至小鼠中 IFF01 基因的最末内含子的真核 GAPDH 启动子下游非翻译性基因组 DNA 序列的长度在其最大情况下伸展至小鼠中编码 IFF01 基因的第 6 号染色体的约 bp 125109211 (位置 +6391)。在一个实施方案中, 在其最大情况下伸展至小鼠中 IFF01 基因的倒数第二内含子的真核 GAPDH 启动子下游非翻译性基因组 DNA 序列的长度在其最大情况下伸展至小鼠中编码 IFF01 基因的第 6 号染色体的约 bp 125103521 (位置 +12081)。在显示第 6 号染色体 (NCBI 基因 ID :320678) 的 bp 125103521 至 bp 125119832 的 SEQ ID NO:18 中包含在其最大情况下分别伸展至小鼠中 IFF01 基因的最末内含子和倒数第二内含子的真核 GAPDH 启动子下游非翻译性基因组 DNA 序列。伸展至小鼠中 IFF01 基因的最末内含子的真核 GAPDH 启动子下游非翻译性基因组 DNA 序列伸展至如 SEQ ID NO:18 所示的核苷酸序列的约 bp 10622, 并且伸展至小鼠中 IFF01 基因的倒数第二内含子的真核 GAPDH 启动子下游非翻译性基因组 DNA 序列伸展至如 SEQ ID NO:18 所示核苷酸序列的约 bp 16312。

[0052] 大鼠 IFF01 基因 (NCBI 基因 ID :362437) 位于第 4 号染色体的约 bp1612649669 至 bp161282150 的大鼠 DNA 中。在一个实施方案中, 在其最大情况下伸展至大鼠中 IFF01 基因的最末内含子的真核 GAPDH 启动子下游非翻译性基因组 DNA 序列的长度在其最大情况下伸展至大鼠中编码 IFF01 基因的第 4 号染色体的约 bp 161280937 (位置 +5154)。在一个实施方案中, 在其最大情况下伸展至大鼠中 IFF01 基因的倒数第二内含子的真核 GAPDH 启动子下游非翻译性基因组 DNA 序列的长度在其最大情况下伸展至大鼠中编码 IFF01 基因的第 4 号染色体的约 bp 161279451 (位置 +6640)。

[0053] 在显示第 4 号染色体 (NCBI 基因 ID :362437) 的 bp 161279451 至 bp161290508 的 SEQ ID NO:19 中包含在其最大情况下分别伸展至大鼠中 IFF01 基因的最末内含子和倒数第二内含子的真核 GAPDH 启动子下游非翻译性基因组 DNA 序列。伸展至 IFF01 基因的最末内含子的真核 GAPDH 启动子下游非翻译性基因组 DNA 序列伸展至如 SEQ ID NO:19 所示的核苷酸序列的约 bp 9572, 并且伸展至 IFF01 基因的倒数第二内含子的真核 GAPDH 启动子下游非翻译性基因组 DNA 序列伸展至如 SEQ ID NO:19 所示核苷酸序列的约 bp 11058。

[0054] 中国仓鼠 IFF01 基因 (NCBI 基因 ID :100753382) 位于约 bp3577293 至 3593683 的中国仓鼠 DNA 中。在一个实施方案中, 在其最大情况下伸展至中国仓鼠中 IFF01 基因的最末内含子的真核 GAPDH 启动子下游非翻译性基因组 DNA 序列的长度在其最大情况下伸展至中国仓鼠中编码 IFF01 基因的约 bp 3579014 (位置 +6883)。在一个实施方案中, 在其最大情况下伸展至中国仓鼠中 IFF01 基因的倒数第二内含子的真核 GAPDH 启动子下游非翻译性基因组 DNA 序列的长度在其最大情况下伸展至中国仓鼠中编码 IFF01 基因的约 bp 3585061 (位

置 +12930)。该染色体位置仍未在 NCBI 数据库中注释并且当前的序列信息含有许多未知的碱基。因此对界限的精确注释可以随更精确序列信息的可获得性而变化。

[0055] 在显示 bp3567932 至 bp3585061 的 SEQ ID NO:29 中包含在其最大情况下分别延伸至中国仓鼠中 IFF01 基因的最末内含子和倒数第二内含子的真核 GAPDH 启动子下游非翻译性基因组 DNA 序列。延伸至 IFF01 基因的最末内含子的真核 GAPDH 启动子下游非翻译性基因组 DNA 序列伸展至如 SEQ ID NO:29 所示的核苷酸序列的约 bp 11083, 并且延伸至 IFF01 基因的倒数第二内含子的真核 GAPDH 启动子下游非翻译性基因组 DNA 序列伸展至如 SEQ ID NO:29 所示核苷酸序列的约 bp 17130。

[0056] 在又一个实施方案中, 真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列始于真核 GAPDH 多聚腺苷化位点处, 例如始于编码真核 GAPDH 多聚腺苷化位点的第一核苷酸。优选地, 真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列始于真核 GAPDH 多聚腺苷化位点的下游, 例如紧邻编码真核 GAPDH 多聚腺苷化位点的最末核苷酸之后开始。甚至更优选的真核 GAPDH 启动子下游非翻译性基因组 DNA 序列始于真核 GAPDH 多聚腺苷化位点的下游并且真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列的长度是至少约 100 个核苷酸并且在其最大情况下延伸至 IFF01 基因的倒数第二内含子。

[0057] 在一个实施方案中, 真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列始于跨越核苷酸位置约 +3881 至核苷酸位置约 +5000 的区域内部, 优选地始于跨越核苷酸位置约 +3931 至核苷酸位置约 +5000 的区域内部, 更优选地始于跨越核苷酸位置约 +4070 至核苷酸位置约 +5000 的区域内部, 其中核苷酸位置是相对于 GAPDH mRNA 转录起始的。

[0058] 始于例如本发明中使用的真核 GAPDH 多聚腺苷化位点下游的真核 GAPDH 启动子下游非翻译性基因组 DNA 序列通常始于核苷酸位置约位置 +3931 处, 优选地始于核苷酸位置约 +4070 处, 其中核苷酸位置是相对于 GAPDH mRNA 转录起始的。

[0059] 在人中, 人 GAPDH 多聚腺苷化位点下游的非翻译性基因组 DNA 序列始于约核苷酸位置 +3931 处 (相对于对应如 SEQ ID NO:17 中所显示的 bp8463 的 GAPDH mRNA 转录起始)。优选地, 如果 GAPDH 多聚腺苷化位点下游的非翻译性基因组 DNA 序列来自人, 则 GAPDH 多聚腺苷化位点下游的非翻译性基因组 DNA 序列始于约 +3931 (相对于 GAPDH mRNA 转录起始; 其对应如 SEQ ID NO:17 中所显示的 bp 8463), 并且其长度是约 3357bp, 对应于从如 SEQ ID NO:17 中所显示的约 bp8463 至约 bp11819 的序列, 更优选地它始于约 +4070 处 (相对于对应如 SEQ ID NO:17 中所显示的 bp 8602 的 GAPDH mRNA 转录起始) 并且其长度是约 3218bp, 对应于从如 SEQ ID NO:17 中所显示的约 bp8602 至约 bp11819 的序列。

[0060] 在又一个实施方案中, 真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列包含选自 SEQ ID NOs:8 和 21 的核苷酸序列或其片段。

[0061] 在又一个实施方案中, 真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列包含与选自 SEQ ID NOs:8 和 21 的核苷酸序列或其片段互补的核苷酸序列。

[0062] 在又一个实施方案中, 真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列包含与选自 SEQ ID NOs:8 和 21 的核苷酸序列或其片段相同至少 80% 的核苷酸序列。

[0063] 在一些实施方案中, 选自 SEQ ID NOs:8 和 21 的核苷酸序列或其片段包含 5 个或更少、优选地 4 个或更少、更优选地 3 个或更少、最优选地 2 个或更少的、特别是 1 个核酸修饰, 其中核酸修饰优选地是核酸置换。

[0064] 在又一个实施方案中,真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列的长度优选地是约 200 至约 8000 个核苷酸、更优选地约 500 至约 5000 个核苷酸、甚至更优选地约 1000 至约 4500 个核苷酸、最优选地约 1500 至约 4000 个核苷酸、特别地约 2000 至约 3500 个核苷酸、更特别约 2700 至约 3300 个、甚至更特别约 3200 个、最特别地 3218 个核苷酸。如本文定义的真核 GAPDH 启动子下游非翻译性基因组 DNA 序列的长度不包括对非翻译性基因组 DNA 序列添加的任何接头序列。

[0065] 在又一个实施方案中,真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列以与编码多肽的多核苷酸序列相同的方向定向。

[0066] 在又一个实施方案中,真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列以相对于编码多肽的多核苷酸序列相反方向定向。

[0067] 在一些实施方案中,包含启动子、编码多肽的多核苷酸序列和真核 GAPDH 启动子下游非翻译性基因组 DNA 序列的表达盒还包含真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列,其中真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列始于跨越约真核 GAPDH 启动子的 5' 末端至核苷酸位置约 -3500 的区域内部,其中核苷酸位置是相对于 GAPDH mRNA 转录起始,并且其中真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列的长度是约 100 至约 15000 个核苷酸。

[0068] 在另一个实施方案中,该表达盒包含启动子、编码多肽的多核苷酸序列和真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列,其中由多核苷酸序列编码的多肽不是 GAPDH,并且其中真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列始于跨越约真核 GAPDH 启动子的 5' 末端至核苷酸位置约 -3500 的区域内部,其中核苷酸位置是相对于 GAPDH mRNA 转录起始的,其中真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列的长度是 100 至约 15000 个核苷酸,条件是表达盒不包含真核 GAPDH 启动子或其片段。

[0069] 在一些实施方案中,该表达盒还包含真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列,其中真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列始于跨越核苷酸位置约 +1 至核苷酸位置约 +7000 的区域内部,其中核苷酸位置是相对于 GAPDH mRNA 转录起始的,并且其中真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列的长度是约 100 至约 15000 个核苷酸。在这些实施方案中,使用的真核 GAPDH 启动子下游非翻译性基因组 DNA 序列例如如上文描述。

[0070] 在一些实施方案中,真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列的长度优选地是约 200 至约 8000 个核苷酸、更优选地约 500 至约 5000 个核苷酸、甚至更优选地约 1000 至约 4500 个核苷酸、最优选地约 1500 至约 4000 个核苷酸、特别地约 2000 至约 3500 个核苷酸、更特别约 2700 至约 3300 个、甚至更特别约 3200 个、最特别地 3158 个核苷酸长度。如本文定义的真核 GAPDH 启动子上游非翻译性基因组 DNA 序列的长度不包括对非翻译性基因组 DNA 序列添加的任何接头序列。

[0071] 在又一个实施方案中,真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列的长度是至少约 100 个核苷酸并且在其最大情况下延伸至 NCAPD2 基因的起始密码子。在又一个实施方案中,真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列的长度是至少约 100 个核苷酸并且在其最大情况下延伸至 NCAPD2 基因的倒数第三内含子。在又一个实施方案中,真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列的长度是至少约 100 个核苷酸并且在其最

大情况下延伸至 NCAPD2 基因的倒数第二内含子。在又一个实施方案中,真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列的长度是至少约 100 个核苷酸并且在其最大情况下延伸至 NCAPD2 基因的最末内含子。

[0072] 人 NCAPD2 基因 (NCBI 基因 ID :9918) 位于第 12 号染色体约 bp6603298 至 bp6641132 的人 DNA 中。在一个实施方案中,在其最大情况下延伸至人中 NCAPD2 基因的最末内含子的真核 GAPDH 启动子上游非翻译性基因组 DNA 序列的长度在其最大情况下伸展至人中编码 NCAPD2 基因的第 12 号染色体的约 bp 6640243 (相对于 GAPDH 基因转录起始的位置 -3414,对应 SEQ ID NO:17 中 bp 1119)。

[0073] 在一个实施方案中,在其最大情况下延伸至人中 NCAPD2 基因的倒数第二内含子的真核 GAPDH 启动子上游非翻译性基因组 DNA 序列的长度在其最大情况下伸展至人中编码 NCAPD2 基因的第 12 号染色体的约 bp6639984 (相对于 GAPDH 基因转录起始的位置 -3673,对应 SEQ ID NO:17 中 bp860)。

[0074] 在一个实施方案中,在其最大情况下延伸至人中 NCAPD2 基因的倒数第三内含子的真核 GAPDH 启动子上游非翻译性基因组 DNA 序列的长度在其最大情况下伸展至人中编码 NCAPD2 基因的第 12 号染色体的约 bp6639125 (相对于 GAPDH 基因转录起始的位置 -4532 ;其对应 SEQ ID NO:17 中的 bp 1)。

[0075] 在显示第 12 号染色体 (NCBI 基因 ID :9918) 的 bp 6657230 至 bp6639125 的 SEQ ID NO:17 中包含在其最大情况下分别延伸至人中 NCAPD2 基因的最末内含子、倒数第二内含子和倒数第三内含子的真核 GAPDH 启动子上游非翻译性基因组 DNA 序列。

[0076] 小鼠 NCAPD2 基因 (基因 ID :68298) 位于第 6 号染色体的约位置 125118025 至 125141604 的小鼠 DNA 中。在一个实施方案中,在其最大情况下延伸至小鼠中 NCAPD2 基因的最末内含子的真核 GAPDH 启动子上游非翻译性基因组 DNA 序列的长度 (据估计具有转录起始上游的 500bp 长度) 在其最大情况下伸展至小鼠中编码 NCAPD2 基因的第 6 号染色体的约 125118607。

[0077] 在一个实施方案中,在其最大情况下延伸至小鼠中 NCAPD2 基因的倒数第二内含子的真核 GAPDH 启动子上游非翻译性基因组 DNA 序列的长度在其最大情况下伸展至小鼠中编码 NCAPD2 基因的第 6 号染色体的约 bp 125118880。

[0078] 在一个实施方案中,在其最大情况下延伸至小鼠中 NCAPD2 基因的倒数第三内含子的真核 GAPDH 启动子上游非翻译性基因组 DNA 序列的长度在其最大情况下伸展至小鼠中编码 NCAPD2 基因的第 6 号染色体的约 bp 125119832。

[0079] 在显示第 6 号染色体 (NCBI 基因 ID :68298) 的 bp125103521 至 bp125119832 的 SEQ ID NO:18 中包含在其最大情况下分别延伸至小鼠中 NCAPD2 基因的最末内含子、倒数第二内含子和倒数第三内含子的真核 GAPDH 启动子上游非翻译性基因组 DNA 序列。伸展至最末内含子的真核 GAPDH 启动子上游非翻译性基因组 DNA 序列伸展至如 SEQ ID NO:18 所显示的核苷酸序列的约 bp 1226 (相对于小鼠 GAPDH mRNA 转录起始的 -3006)。伸展至倒数第二内含子的真核 GAPDH 启动子下游非翻译性基因组 DNA 序列伸展至如 SEQ ID NO:18 所显示的核苷酸序列的约 bp953 (相对于小鼠 GAPDH mRNA 转录起始的 -3279)。伸展至倒数第三内含子的真核 GAPDH 启动子下游非翻译性基因组 DNA 序列伸展至如 SEQ ID NO:18 所显示的核苷酸序列的约 bp 1 (相对于小鼠 GAPDH mRNA 转录起始的 -4231)。

[0080] 大鼠 NCAPD2 基因 (基因 ID :362438) 位于第 4 号染色体的约位置 161288671 至 161310417 的真核 DNA 中。在一个实施方案中,在其最大情况下延伸至大鼠中 NCAPD2 基因的最末内含子的真核 GAPDH 启动子上游非翻译性基因组 DNA 序列的长度在其最大情况下伸展至大鼠中编码 NCAPD2 基因的第 4 号染色体的约 bp 161289191。在一个实施方案中,在其最大情况下延伸至大鼠中 NCAPD2 基因的倒数第二内含子的真核 GAPDH 启动子上游非翻译性基因组 DNA 序列的长度在其最大情况下伸展至大鼠中编码 NCAPD2 基因的第 4 号染色体的约 bp 161289446。在一个实施方案中,在其最大情况下延伸至大鼠中 NCAPD2 基因的倒数第三内含子的真核 GAPDH 启动子上游非翻译性基因组 DNA 序列的长度在其最大情况下伸展至大鼠中编码 NCAPD2 基因的第 4 号染色体的约 bp161290508。

[0081] 在显示第 4 号染色体 (NCBI 基因 ID :362438) 的 bp161279451 至 bp161290508 的 SEQ ID NO:19 中包含在其最大情况下分别延伸至大鼠中 NCAPD2 基因的最末内含子、倒数第二内含子和倒数第三内含子的真核 GAPDH 启动子上游非翻译性基因组 DNA 序列。延伸至最末内含子的真核 GAPDH 启动子上游非翻译性基因组 DNA 序列伸展至如 SEQ ID NO:19 所显示的核苷酸序列的约 bp 1318 (相对于大鼠 GAPDH mRNA 转录起始的 -3101)。延伸至倒数第二内含子的真核 GAPDH 启动子下游非翻译性基因组 DNA 序列伸展至如 SEQ ID NO:19 所显示的核苷酸序列的约 bp 1063 (相对于大鼠 GAPDH mRNA 转录起始的位置 -3356)。延伸至倒数第三内含子的真核 GAPDH 启动子下游非翻译性基因组 DNA 序列伸展至如 SEQ ID NO:19 所显示的核苷酸序列的约 bp 1 (相对于大鼠 GAPDH mRNA 转录起始的位置 -4418)。

[0082] 中国仓鼠 NCAPD2 基因 (基因 ID :100753087) 位于约位置 3544184 至 3569879 的真核 DNA 中。该染色体位置在 NCBI 数据库上不可获得。在一个实施方案中,在其最大情况下延伸至中国仓鼠中 NCAPD2 基因的最末内含子的真核 GAPDH 启动子上游非翻译性基因组 DNA 序列的长度在其最大情况下伸展至中国仓鼠中的约 3569380bp。在一个实施方案中,在其最大情况下延伸至中国仓鼠中 NCAPD2 基因的倒数第二内含子的真核 GAPDH 启动子上游非翻译性基因组 DNA 序列的长度在其最大情况下伸展至中国仓鼠中的约 3569131bp。在一个实施方案中,在其最大情况下延伸至中国仓鼠中 NCAPD2 基因的倒数第三内含子的真核 GAPDH 启动子上游非翻译性基因组 DNA 序列的长度在其最大情况下伸展至中国仓鼠中的约 3567932bp。

[0083] 在显示 bp3567932 至 bp3585061 的 SEQ ID NO:29 中包含在其最大情况下分别延伸至中国仓鼠中 NCAPD2 基因的最末内含子、倒数第二内含子和倒数第三内含子的真核 GAPDH 启动子上游非翻译性基因组 DNA 序列。延伸至最末内含子的真核 GAPDH 启动子上游非翻译性基因组 DNA 序列伸展至如 SEQ ID NO:29 所显示的核苷酸序列的约 bp 1449 (相对于中国仓鼠 GAPDH mRNA 转录起始的 -2752)。延伸至倒数第二内含子的真核 GAPDH 启动子下游非翻译性基因组 DNA 序列伸展至如 SEQ ID NO:29 所显示的核苷酸序列的约 bp 1200 (相对于中国仓鼠 GAPDH mRNA 转录起始的位置 -3001)。延伸至倒数第三内含子的真核 GAPDH 启动子下游非翻译性基因组 DNA 序列伸展至如 SEQ ID NO:29 所显示的核苷酸序列的约 bp 1 (相对于中国仓鼠 GAPDH mRNA 转录起始的位置 -4200)。

[0084] 在一些实施方案中,真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列通常始于跨越核苷酸位置约 -500 至核苷酸位置约 -3500 区域内部,优选地始于跨越核苷酸位置约 -576 至核苷酸位置约 -3500 区域内部,其中核苷酸位置是相对于 GAPDH mRNA 转录起始

的。

[0085] 在一些实施方案中,真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列通常始于核苷酸位置约位置 -500 处,优选地始于核苷酸位置约 -576 处,其中核苷酸位置是相对于 GAPDH mRNA 转录起始的。

[0086] 在人中,人 GAPDH 启动子上游的非翻译性基因组 DNA 序列始于约核苷酸位置 -463 处(相对于对应如 SEQ ID NO:17 中所显示的 bp4533 的 GAPDH mRNA 转录起始)。优选地,如果 GAPDH 启动子上游的非翻译性基因组 DNA 序列来自人,则该 GAPDH 启动子上游的非翻译性基因组 DNA 序列始于约 -500 处(相对于 GAPDH mRNA 转录起始;其对应如 SEQ ID NO:17 中所显示的 bp 4533)。更优选地,如果 GAPDH 启动子上游的非翻译性基因组 DNA 序列来自人,则该 GAPDH 启动子上游的非翻译性基因组 DNA 序列始于约 -576 处(相对于 GAPDH mRNA 转录起始;其对应如 SEQ ID NO:17 中所显示的 bp 4533),并且其长度是约 3158bp,对应于从如 SEQ ID NO:17 中所显示的约 bp800 至约 bp3957 的序列。

[0087] 在又一个实施方案中,真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列包含选自 SEQ ID NOs:7、9、10、11、12、13、14、15、16、20、22、23、24、25、26、27 和 28 的核苷酸序列或其片段,优选地选自 SEQ ID NOs:7、9、10、11、12、13、14、15 和 16 的核苷酸序列或其片段,或选自 SEQ ID NOs:20、22、23、24、25、26、27、28 和 16 的核苷酸序列或其片段。更优选地是选自 SEQ ID NOs:10、12、15 和 16 的核苷酸序列或其片段,更优选地是选自 SEQ ID NOs:10、12、15 和 16 的核苷酸序列或其片段,其中包含 SEQ ID NOs:10 和 / 或 16 的核苷酸序列相对于编码多肽的多核苷酸序列以相反方向定向,并且包含 SEQ ID NOs:12 和 / 或 15 的核苷酸序列以与编码多肽的多核苷酸序列相同的方向定向。同等更优选地是选自 SEQ ID NOs:23、25、28 和 16 的核苷酸序列或其片段,更优选地是选自 SEQ ID NOs:23、25、28 和 16 的核苷酸序列或其片段,其中包含 SEQ ID NOs:23 和 / 或 16 的核苷酸序列相对于编码多肽的多核苷酸序列以相反方向定向,并且包含 SEQ ID NOs:25 和 / 或 28 的核苷酸序列以与编码多肽的多核苷酸序列相同的方向定向。

[0088] 在又一个实施方案中,真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列包含与选自 SEQ ID NOs:7、9、10、11、12、13、14、15、16、20、22、23、24、25、26、27 和 28 的核苷酸序列或其片段互补的核苷酸序列,优选地与选自 SEQ ID NOs:7、9、10、11、12、13、14、15 和 16 的核苷酸序列或其片段互补的核苷酸序列或与选自 SEQ ID NOs:20、22、23、24、25、26、27、28 和 16 的核苷酸序列或其片段互补的核苷酸序列。更优选地是与选自 SEQ ID NOs:10、12、15 和 16 的核苷酸序列或其片段互补的核苷酸序列。同等地更优选地是与选自 SEQ ID NOs:23、25、28 和 16 的核苷酸序列或其片段互补的核苷酸序列。

[0089] 在又一个实施方案中,真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列包含与选自 SEQ ID NOs:7、9、10、11、12、13、14、15、16、20、22、23、24、25、26、27 和 28 的核苷酸序列或其片段相同至少 80% 的核苷酸序列,优选地与选自 SEQ ID NOs:7、9、10、11、12、13、14、15 和 16 的核苷酸序列或其片段相同至少 80% 的核苷酸序列或与选自 SEQ ID NOs:20、22、23、24、25、26、27、28 和 16 的核苷酸序列或其片段相同至少 80% 的核苷酸序列。更优选地是与选自 SEQ ID NOs:10、12、15 和 16 的核苷酸序列或其片段相同至少 80% 的核苷酸序列,更优选地是与选自 SEQ ID NOs:10、12、15 和 16 的核苷酸序列或其片段相同至少 80% 的核苷酸序列,其中包含 SEQ ID NOs:10 和 / 或 16 的核苷酸序列相对于编码多肽的多核苷酸序列以

相反方向定向,并且包含 SEQ ID NOs:12 和 / 或 15 的核苷酸序列以与编码多肽的多核苷酸序列相同的方向定向。同等更优选地是与选自 SEQ ID NOs:23、25、28 和 16 的核苷酸序列或其片段相同至少 80% 的核苷酸序列,更优选地是与选自 SEQ ID NOs:23、25、28 和 16 的核苷酸序列或其片段相同至少 80% 的核苷酸序列,其中包含 SEQ ID NOs:23 和 / 或 16 的核苷酸序列相对于编码多肽的多核苷酸序列以相反方向定向,并且包含 SEQ ID NOs:25 和 / 或 28 的核苷酸序列以与编码多肽的多核苷酸序列相同的方向定向。

[0090] 在一些实施方案中,选自 SEQ ID NOs:7、9、10、11、12、13、14、15、16、20、22、23、24、25、26、27 和 28 的核苷酸序列或其片段包含 5 个或更少、优选地 4 个或更少、更优选地 3 个或更少、最优选地 2 个或更少的、特别是 1 个核酸修饰,其中核酸修饰优选地是核酸置换。

[0091] 在一些实施方案中,选自 SEQ ID NOs:7、9、11、14、20、22、24 和 27 的核苷酸序列或其片段包含 5 个或更少、优选地 4 个或更少、更优选地 3 个或更少、最优选地 2 个或更少的、特别是 1 个核酸修饰,其中核酸修饰优选地是核酸置换。

[0092] 在一些实施方案中,选自 SEQ ID NOs:7、9、11、14 的核苷酸序列或其片段在第 16 位置处相对于 SEQ ID NOs:7、9、11、14 的核苷酸序列的起始包含一个核酸置换。优选地,在第 16 位置处的 G 相对于核苷酸序列起始替换为 T。

[0093] 在一些实施方案中,选自 SEQ ID NOs:20、22、24 和 27 的核苷酸序列或其片段在第 13 位置处相对于 SEQ ID NOs:20、22、24 和 27 的核苷酸序列的起始包含一个核酸置换。优选地,在第 13 位置处的 G 相对于核苷酸序列起始替换为 T。

[0094] 在又一个实施方案中,真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列以与编码多肽的多核苷酸序列相同的方向定向。

[0095] 在又一个实施方案中,真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列相对于编码多肽的多核苷酸序列以相反方向定向。

[0096] 在优选实施方案中,如上文描述,表达盒包含启动子、编码多肽的多核苷酸序列和真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列和真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列。优选地,真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列和真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列的起源是相同的,即,是相同物种的。更优选地,真核 GAPDH 启动子下游非翻译性基因组 DNA 序列、真核 GAPDH 启动子上游非翻译性基因组 DNA 序列和宿主细胞的起源是相同的,即,是相同物种的,例如真核 GAPDH 启动子下游非翻译性基因组 DNA 序列、真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列和宿主细胞的起源来自相同的哺乳动物,例如来自人。

[0097] 在一些实施方案中,如果真核 GAPDH 启动子下游和 / 或上游的非翻译性基因组 DNA 序列是来自一个物种的非翻译性基因组 DNA 序列,则表达盒的启动子不是来自相同物种的 GAPDH 启动子。

[0098] 在一些实施方案中,如果真核 GAPDH 启动子下游和 / 或上游的非翻译性基因组 DNA 序列是人源下游和 / 或上游的非翻译性基因组 DNA 序列,则表达盒的启动子不是人 GAPDH 启动子。

[0099] 在一些实施方案中,表达盒的启动子不是 GAPDH 启动子。

[0100] 在一个实施方案中,如果表达盒包含启动子、编码多肽的多核苷酸序列和真核甘油醛 3- 磷酸脱氢酶 (GAPDH) 启动子下游的非翻译性基因组 DNA 序列,其中由多核苷酸序列

编码的多肽不是 GAPDH, 并且其中真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列始于跨越核苷酸位置约 +1 至核苷酸位置约 +7000 的区域内部, 其中核苷酸位置是相对于 GAPDH mRNA 转录起始的, 并且其中真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列的长度是约 100 至约 15000 个核苷酸并且其中表达盒还包含真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列, 其中真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列始于跨越约真核 GAPDH 启动子的 5' 末端至核苷酸位置约 -3500 的区域内部, 其中核苷酸位置是相对于 GAPDH mRNA 转录起始的, 并且其中真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列的长度是约 100 至约 15000 个核苷酸, 表达盒的启动子可以是真核 GAPDH 启动子, 优选地是哺乳动物 GAPDH 启动子, 更优选地是啮齿类或人 GAPDH 启动子。在这个实施方案中, 始于跨越约真核 GAPDH 启动子的 5' 末端至核苷酸位置约 -3500 的区域内部的真核 GAPDH 启动子上游非翻译性基因组 DNA 序列优选地刚好位于真核 GAPDH 启动子的上游, 更优选地在这个实施方案中, 表达盒包含天然存在的基因组 DNA 序列, 所述基因组 DNA 序列包含真核 GAPDH 启动子并延伸至核苷酸位置约 -3500, 其中核苷酸位置是相对于 GAPDH mRNA 转录起始的。

[0101] 在一些实施方案中, 真核 GAPDH 启动子下游和 / 或上游的非翻译性基因组 DNA 序列是哺乳动物源的, 例如真核 GAPDH 启动子是哺乳动物 GAPDH 启动子, 并且哺乳动物 GAPDH 启动子下游和 / 或上游的非翻译性基因组 DNA 序列如本文所述那样使用。

[0102] 在一些实施方案中, 真核 GAPDH 启动子下游和 / 或上游的非翻译性基因组 DNA 序列是啮齿类或人源的, 例如真核 GAPDH 启动子是啮齿类或人 GAPDH 启动子, 并且啮齿类或人 GAPDH 启动子下游和 / 或上游的非翻译性基因组 DNA 序列如本文所述那样使用。

[0103] 优选地, 真核 GAPDH 启动子下游和 / 或上游的非翻译性基因组 DNA 序列选自人、大鼠或小鼠源, 更优选地来自人源或小鼠源, 最优选地来自人源。

[0104] 在一些实施方案中, 真核 GAPDH 启动子下游和 / 或上游的非翻译性基因组 DNA 序列未与编码多肽的多核苷酸序列有效连接。

[0105] 在一些实施方案中, 表达盒包含多聚腺苷化位点。优选地, 多聚腺苷化位点选自 SV40poly(A) 和 BGH(牛生长激素)poly(A)。

[0106] 在一些实施方案中, 表达盒的启动子和编码多肽的多核苷酸序列有效连接。

[0107] 在一些实施方案中, 表达盒的启动子选自 SV40 启动子、人 tk 启动子、MPSV 启动子、小鼠 CMV、人 CMV、大鼠 CMV、人 EF1 α 、中国仓鼠 EF1 α 、人 GAPDH、杂合启动子, 所述杂合启动子包括 MYC、HYK 和 CX 启动子。

[0108] 在一些实施方案中, 由表达盒编码的多肽可以是非糖基化多肽和糖基化多肽。糖基化多肽指具有至少一条寡糖链的多肽。

[0109] 非糖基化蛋白的例子是例如非糖基化激素; 非糖基化酶; 神经生长因子 (NGF) 家族、上皮生长因子 (EGF) 和成纤维细胞生长因子 (FGF) 家族的非糖基化生长因子以及激素和生长因子的非糖基化受体。

[0110] 糖基化蛋白的例子是激素和激素释放因子、凝血因子、抗凝血因子、激素或生长因子的受体、神经营养因子细胞因子和它们的受体、T 细胞受体、表面膜蛋白、转运蛋白、归巢受体、地址素、调节蛋白、抗体、嵌合蛋白 (如免疫黏附素) 和任何糖基化蛋白的片段。优选地, 多肽选自抗体, 抗体片段或抗体衍生物 (例如 Fc 融合蛋白和特殊抗体样式, 如双特异性抗体)。如本文所用的抗体片段包括, 但不限于 (i) 结构域; (ii) 由 VL、VH、CL 或 CK 和 CH1

结构域组成的 Fab 片段,包括 Fab' 和 Fab'-SH;(iii) 由 VH 和 CH1 结构域组成的 Fd 片段;(iv) 由单个可变域组成的 dAb 片段 (Ward ES 等人, (1989) Nature, 341(6242):544-6);(v) F(ab')₂ 片段,一种包含两个连接的 Fab 片段的双价片段;(vi) 单链 Fv 分子 (scFv),其中 VH 结构域和 VL 结构域由允许这两个结构域缔合以形成抗原结合位点的肽接头连接 (Bird RE 等人, (1988) Science, 242(4877):423-6;Huston JS 等人, (1988) Proc Natl Acad Sci U S A, 85(16):5879-83);(vii) “双抗体”或“三抗体”,通过基因融合所构建的多价或多特异性片段 (Holliger P 等人, (1993) Proc Natl Acad Sci U S A, 90(14):6444-8;Holliger P 等人, (2000) Methods Enzymol, 326:461-79);(viii) scFv,与 Fc 区融合的双抗体 (diabody) 或结构域抗体和 (ix) 与相同或不同抗体融合的 scFv。

[0111] 在一些实施方案中,表达盒还包含遗传元件,所述遗传元件选自额外的启动子、增强子、转录控制元件和选择标记,优选地在动物细胞中表达的选择标记。转录控制元件例如是 Kozak 序列或转录终止子元件。

[0112] 在一个实施方案中,遗传元件是一种选择标记,其中在编码选择标记的多核苷酸序列中含有的 CpG 位点的含量是 45 或更小,优选地是 40 或更小,更优选地是 20 或更小,特别地是 10 或更小,更特别地是 5 或更小,最特别地是 0(已经彻底移除 CpG 位点)。

[0113] 在又一个方面,本公开提供表达载体,优选地包含如上文所述表达盒的哺乳动物表达载体。

[0114] 在一些实施方案中,表达载体包含至少两个独立的转录单位。具有两个独立转录单位的表达载体也称作双基因载体。其例子是其中第一转录单位编码抗体重链或其片段且第二转录单位编码抗体轻链的载体。另一个例子是其中两个转录单位编码蛋白质(如酶)的两个不同亚基的双基因载体。然而,还可能的是本发明的表达载体包含多于两个独立转录单位,例如三、四或甚至更多个独立转录单位,所述独立转录单位各自包含编码不同多肽链的不同核苷酸序列。其例子是具有 4 个独立转录单位的载体,所述独立转录单位各自含有编码由 4 个不同亚基组成的酶的一个亚基的不同核苷酸序列。

[0115] 在一些实施方案中,表达载体还包含选自额外启动子、增强子,转录控制单元、复制起点和选择标记的遗传元件。

[0116] 在一些实施方案中,表达载体还包含复制起点和选择标记,其中表达载体的编码复制起点和选择标记的多核苷酸序列中所含的 CpG 位点的含量是 200 或更小,优选地是 150 或更小,特别地是 100 或更小,更特别地是 50 或更小,最特别地是 30 或更小。

[0117] 常使用的任何选择标记如胸苷激酶(tk)、二氢叶酸还原酶(DHFR)、嘌呤霉素、新霉素或谷氨酰胺合成酶(GS)可以用于本发明的表达盒或表达载体。优选地,本发明的表达载体还包含有限数目的有有限制性位点用于插入分泌本发明异源蛋白的表达盒。在仅特别用于瞬时/附加型表达使用时,本发明的表达载体还可以包含在真核宿主细胞中用于自主复制/附加型维持的复制起点如 EB 病毒(EBV)或 SV40 病毒的 oriP 起点,但是可以缺少选择标记。还可能在缺少促进载体复制的相关因子的细胞中瞬时表达。

[0118] 携带表达盒的表达载体还可以包含编码荧光标记的表达盒、编码 ncRNA 的表达盒、编码抗凋亡蛋白的表达盒或编码增加分泌途径容量的蛋白质的表达盒。

[0119] 在又一个方面,本公开提供一种表达载体,其依次包含:

[0120] a) 真核 GAPDH 启动子上游和/或下游的非翻译性基因组 DNA 序列

- [0121] b) 启动子
- [0122] c) 编码多肽的多核苷酸序列
- [0123] d) 多聚腺苷化位点
- [0124] e) 增强子
- [0125] f) 真核 GAPDH 启动子下游和 / 或上游的非翻译性基因组 DNA 序列, 或
- [0126] a) 真核 GAPDH 启动子上游和 / 或下游的非翻译性基因组 DNA 序列
- [0127] b) 增强子
- [0128] c) 启动子
- [0129] d) 编码多肽的多核苷酸序列
- [0130] e) 多聚腺苷化位点
- [0131] f) 真核 GAPDH 启动子下游和 / 或上游的非翻译性基因组 DNA 序列, 或
- [0132] a) 增强子
- [0133] b) 真核 GAPDH 上游和 / 或下游的非翻译性基因组 DNA 序列
- [0134] c) 启动子
- [0135] d) 编码多肽的多核苷酸序列
- [0136] e) 多聚腺苷化位点
- [0137] f) 真核 GAPDH 下游和 / 或上游的非翻译性基因组 DNA 序列,
- [0138] 其中包含增强子是任选的, 并且其中由多核苷酸序列编码的多肽不是 GAPDH, 并且其中真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列始于跨越核苷酸位置约 +1 至核苷酸位置约 +7000 的区域内部, 其中核苷酸位置是相对于 GAPDH mRNA 转录起始的, 并且其中真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列的长度是约 100 至约 15000 个核苷酸并且其中真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列始于跨越约真核 GAPDH 启动子的 5' 末端至核苷酸位置约 -3500 的区域内部, 其中核苷酸位置是相对于 GAPDH mRNA 转录起始的, 并且其中真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列的长度是约 100 至约 15000 个核苷酸, 条件是如果 a) 或 b) 是真核 GAPDH 上游的非翻译性基因组 DNA 序列, 则 f) 是真核 GAPDH 下游的非翻译性基因组 DNA 序列, 并且如果 a) 或 b) 是真核 GAPDH 下游的非翻译性基因组 DNA 序列, 则 f) 是真核 GAPDH 上游的非翻译性基因组 DNA 序列。
- [0139] 在一些实施方案中, 本公开提供一种表达载体, 其依次包含:
- [0140] a) 真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列
- [0141] b) 启动子
- [0142] c) 编码多肽的多核苷酸序列
- [0143] d) 多聚腺苷化位点
- [0144] e) 增强子
- [0145] f) 真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列, 其中包含增强子是任选的。
- [0146] 在又一个方面, 本公开提供一种表达载体, 其依次包含:
- [0147] a) 真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列
- [0148] b) 增强子
- [0149] c) 启动子
- [0150] d) 编码多肽的多核苷酸序列

- [0151] e) 多聚腺苷化位点
- [0152] f) 真核 GAPDH 下游的非翻译性基因组 DNA 序列,
- [0153] 其中包含增强子是任选的。
- [0154] 在又一个方面,本公开提供一种表达载体,其依次包含:
- [0155] a) 增强子
- [0156] b) 真核 GAPDH 上游的非翻译性基因组 DNA 序列
- [0157] c) 启动子
- [0158] c) 编码多肽的多核苷酸序列
- [0159] e) 多聚腺苷化位点
- [0160] f) 真核 GAPDH 下游的非翻译性基因组 DNA 序列,
- [0161] 其中包含增强子是任选的。
- [0162] 在又一个方面,本公开提供一种表达载体,其依次包含:
- [0163] a) 真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列
- [0164] b) 启动子
- [0165] c) 编码多肽的多核苷酸序列
- [0166] d) 多聚腺苷化位点
- [0167] e) 增强子
- [0168] f) 真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列,其中包含增强子是任选的。
- [0169] 在又一个方面,本公开提供一种表达载体,其依次包含:
- [0170] a) 真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列
- [0171] b) 增强子
- [0172] c) 启动子
- [0173] d) 编码多肽的多核苷酸序列
- [0174] e) 多聚腺苷化位点
- [0175] f) 真核 GAPDH 上游的非翻译性基因组 DNA 序列
- [0176] 其中包含增强子是任选的。
- [0177] 在又一个方面,本公开提供一种表达载体,其依次包含:
- [0178] a) 增强子
- [0179] b) 真核 GAPDH 下游的非翻译性基因组 DNA 序列
- [0180] c) 启动子
- [0181] d) 编码多肽的多核苷酸序列
- [0182] e) 多聚腺苷化位点
- [0183] f) 真核 GAPDH 上游的非翻译性基因组 DNA 序列,
- [0184] 其中包含增强子是任选的。
- [0185] 表达载体的真核 GAPDH 上游非翻译性基因组 DNA 序列、增强子、启动子、编码多肽的多核苷酸序列、多聚腺苷化位点和真核 GAPDH 启动子下游非翻译性基因组 DNA 序列例如如上文描述。
- [0186] 在又一个方面,本公开提供包含如上文描述的表达盒或表达载体的宿主细胞。该宿主细胞可以是人或非人细胞。优选的宿主细胞是哺乳动物细胞。哺乳动物宿主细胞的优

选例子包括而不限于人胚肾细胞 (Graham FL 等人, J. Gen. Virol. 36:59-74)、MRC5 人成纤维细胞、983M 人黑色素瘤细胞、MDCK 犬肾细胞、RF 培养的从 Sprague-Dawley 大鼠分离的大鼠肺成纤维细胞、B16BL6 鼠黑色素瘤细胞、P815 鼠肥大细胞瘤细胞、MT1 A2 鼠乳房腺癌细胞、PER:C6 细胞 (Leiden、荷兰) 和中国仓鼠卵巢 (CHO) 细胞或细胞系 (Puck 等人, 1958, J. Exp. Med. 108:945-955)。

[0187] 在一个具体的优选实施方案中, 宿主细胞是中国仓鼠卵巢 (CHO) 细胞或细胞系。合适的 CHO 细胞系例如包括 CHO-S (Invitrogen, Carlsbad, CA, 美国)、CHO K1 (ATCC CCL-61)、CHO pro3-、CHO DG44、CHO P12 或 dhfr⁻CHO 细胞系 DUK-BII (Chasin 等人, PNAS 77, 1980, 4216-4220)、DUXBI 1 (Simonsen 等人, PNAS 80, 1983, 2495-2499)、或 CHO-K1SV (Lonza, 巴塞尔, 瑞士)。

[0188] 在又一个方面, 本公开提供用于表达多肽的体外方法, 所述方法包括用如上文描述的表达盒或表达载体转染宿主细胞并回收多肽。该多肽优选地是异源的, 更优选地是人多肽。

[0189] 根据本发明为了将表达盒或表达载体转染入宿主细胞, 可以使用任何转染技术, 如本领域熟知的那些, 例如电穿孔法、磷酸钙共沉淀法、DEAE-葡聚糖转染法、脂质体转染法, 适当时用于给定宿主细胞类型。应当指出, 用本发明表达盒或表达载体转染的宿主细胞应当解释为是瞬时或稳定转染的细胞系。因此, 根据本发明, 本发明表达盒或表达载体可以按附加型方式维持, 即瞬时转染的, 或可以稳定整合在宿主细胞的基因组中, 即稳定转染的。

[0190] 瞬时转染以不采用针对具有选择标记的载体的任何选择压力为特征。在转染后通常持续 2 直至 10 日的瞬时表达实验中, 转染的表达盒或表达载体作为附加型元件维持并也不整合至基因组中。即, 转染的 DNA 通常不整合至宿主细胞基因组中。宿主细胞倾向于丢失转染的 DNA 并且当培养瞬时转染的细胞汇集物时在群体中长满转染的细胞。因此表达在紧邻转染后的时间段中最强烈并且随时间减少。优选地, 将本发明的瞬时转染子理解为在选择压力不存在的情况下在细胞培养物中维持直至转染后 2 至 10 日时间的细胞。

[0191] 在本发明的优选实施方案中, 宿主细胞例如 CHO 宿主细胞用本发明的表达盒或表达载体稳定转染。稳定转染意指新引入的外来 DNA 如载体 DNA 被掺入基因组 DNA 中, 通常借助随机非同源重组事件。可以通过选择其中在整合入宿主细胞的 DNA 之后载体序列已经被扩增的细胞系, 增加载体 DNA 的拷贝数以及同时增加基因产物的量。因此, 可能的是当暴露于用于扩增基因的选择压力进一步增加时, 这种稳定整合在 CHO 细胞中产生双倍微小染色体。另外, 稳定转染可以导致丢失与重组基因产物表达不直接相关的载体序列部分, 例如当基因组整合时变得多余的细菌拷贝数控制区。因此, 转染的宿主细胞已经将表达盒或表达载体的至少部分或不同部分整合至基因组中。

[0192] 在又一个方面, 本公开提供如上文描述的表达盒或表达载体用于从哺乳动物宿主细胞表达异源多肽的用途, 特别是如上文描述的表达盒或表达载体用于从哺乳动物宿主细胞体外表达异源多肽的用途。

[0193] 可以根据本领域技术人员已知的方法实施蛋白质的表达和回收。

[0194] 为了表达多肽, 使用如上文描述的表达盒或表达载体的真核 GAPDH 启动子下游和 / 或上游非翻译性基因组 DNA 序列和如上文描述的宿主细胞并且它们通常是相同来源的。

令人惊讶地已经发现如果表达盒或表达载体的真核 GAPDH 启动子下游和 / 或上游的非翻译性基因组 DNA 序列和宿主细胞是不同来源的,例如,如果在 CHO 细胞中使用人的真核 GAPDH 启动子下游和 / 或上游的 DNA 序列,则获得表达的增加。

[0195] 在又一个方面,本公开提供如上文描述的表达盒或表达载体制备用于治疗疾病的药物的用途。

[0196] 在又一个方面,本公开提供如上文描述的表达盒或表达载体,其用作治疗疾病的药物。

[0197] 在又一个方面,本公开提供如上文描述的表达盒或表达载体,其用于基因疗法。

实施例

[0198] 实施例 1 :表达载体的克隆 :

[0199] I. 材料和方法

[0200] I. 1 质粒构建体

[0201] I. 1. 1. LB 培养板

[0202] 将 500ml 水与 16g LB 琼脂 (Invitrogen, Carlsbad, CA, 美国) 混合并煮沸 (1 升 LB 含有 10g 胰蛋白胨、5g 酵母提取物和 10g NaCl)。在冷却下来后,将相应的抗生素添加至溶液,所述溶液随后铺板 (氨苄青霉素平板为 100 μ g/ml 和卡那霉素平板为 50 μ g/ml)。

[0203] I. 1. 2. 聚合酶链反应 (PCR)

[0204] 全部 PCR 均使用 50 μ l 终体积中的 1 μ l dNTP (每种 dNTP 为 10mM;Invitrogen, Carlsbad, CA, 美国)、2 单位 **Phusion®** DNA 聚合酶 (Finnzymes Oy, Espoo, 芬兰)、25nmol 引物 A (Mycrosynth, Balgach, 瑞士)、25nmol 引物 B (Mycrosynth, Balgach, 瑞士)、10 μ l 5X HF 缓冲剂 (7.5mM MgCl₂, Finnzymes, Espoo, 芬兰)、1.5 μ l 二甲基亚砜 (DMSO, Finnzymes, Espoo, 芬兰) 和 1-3 μ l 模板 (1-2 μ g) 进行。表 1 中列出使用的全部引物。

[0205] 通过在 98°C 初始变性 3 分钟启动 PCR,随后是 35 个循环:在 98°C 变性 30 秒,在引物特异性温度 (根据 CG 含量) 退火 30 秒和在 72°C 延长 2 分钟。在 72°C 进行最终延长 10 分钟,之后冷却和保持在 4°C。

[0206] 表 1 :PCR 中使用的引物的总结。GAPDH :甘油醛 3- 磷酸脱氢酶序列,5' :上游序列,3' :下游序列。引物 GlnPr1172 中引入 "T" (加下划线) 以避免引物二聚体形成。

[0207]

引物	引物序列	扩增的序列	Seq ID
GlnPr 1171	ATTATTCGCGATGGCTCCTGGCA TCTCTGGGACCGAGGC	5' GAPDH	SEQ ID No: 1
GlnPr 1172	ATCGTCGCGAAGCTTGAGATTGT CCAAGCAGGTAGCCAG		SEQ ID No: 2
GlnPr 1173	AGCAAGTACTTCTGAGCCTTCA GTAATGGCTGCCTG	3'GAPDH	SEQ ID No: 3
GlnPr 1174	TGGCAGTACTAAGCTGGCACCA CTACTTCAGAGAACAAG		SEQ ID No: 4

[0208] I. 1. 3. 限制性消化

[0209] 对于全部限制性消化,将约 1 μ g 质粒 DNA(用 NanoDrop, ND-1000 分光光度计(Thermo Scientific, 威名顿, DE, 美国)定量)混合入 10-20 单位的每种酶、4 μ l 相应 10X NEBuffer(NEB, Ipswich, MA, 美国)中,并且将体积用无菌 H₂O 补足至 40 μ l。在不进一步明示的情况下,将消化物在 37°C 温育 1 小时。

[0210] 在每次制备性消化主链后,添加 1 单位牛小肠碱性磷酸酶(CIP;NEB, Ipswich, MA, 美国)并将混合物在 37°C 温育 30 分钟。

[0211] 如果在 NEBuffer 3(NEB, Ipswich, MA, 美国)中进行消化,则在添加 CIP 之前将缓冲液改变成 NEB 缓冲液 4,因为这种酶在这个缓冲液中具有强烈活性并且还可以消化在外末端处的一些核苷酸。

[0212] I. 1. 4. PCR 纯化和琼脂糖凝胶电泳

[0213] I. 1. 4. 1. PCR 净化

[0214] 为允许消化,在限制性消化之前,使用 Macherey Nagel Extract II 试剂盒(Macherey Nagel, Oensingen, 瑞士),遵循制造商的手册,使用 40 μ l 洗脱缓冲液净化全部 PCR 片段。这种方案还用于改变 DNA 样品的缓冲液。

[0215] I. 1. 4. 2. DNA 提取

[0216] 对于凝胶电泳,使用 UltraPure™ 琼脂糖(Invitrogen, Carlsbad, CA, 美国)和 50X Tris 乙酸 EDTA 缓冲液(TAE, pH 8.3;Bio RAD, Munich, 德国),制备 1% 凝胶。为了 DNA 染色,将 1 μ l Gel Red Dye(Biotum, Hayward, CA, 美国)添加至 100ml 琼脂糖凝胶。作为大小标志物,使用 2 μ g 的 1kb DNA 梯(NEB, Ipswich, MA, 美国)。电泳以 125 伏运行约 1 小时。

[0217] 将目的条带从琼脂糖凝胶切下并使用试剂盒 Extract II(Macherey-Nagel, Oensingen, 瑞士),遵循制造商手册,使用 40 μ l 洗脱缓冲液进行纯化。

[0218] I. 1. 5. 连接

[0219] 对于每个连接,将 4 μ l 插入物与 1 μ l 载体、400 单位连接酶(T4DNA 连接酶,NEB, Ipswich, MA, 美国),1 μ l 10X 连接酶缓冲液(T4DNA 连接酶缓冲液;NEB, Ipswich, MA, 美国)在 10 μ l 体积中混合。将混合物在室温温育 1-2 小时。

[0220] I. 1. 6. 连接产物转化入感受态细菌中

[0221] 为了克隆 pGLEX41-[REP] 并对于用含有标准复制起点的 pCR-Blunt 载体产生的构建体,使用 TOP10 (One **Shot**[®] TOP 10 感受态大肠杆菌 ;Invitrogen, Carlsbad, CA, 美国)。

[0222] 为了含有 R6K 复制起点的质粒的复制起始,需要表达 pir 序列编码的 π 蛋白。 π 蛋白由 One **Shot**[®] PIR1 感受态大肠杆菌 (Invitrogen, Carlsbad, CA, 美国) 表达。这些细菌用于含有 R6K 序列的全部载体。

[0223] 为了用连接产物转化感受态细菌,在冰上融化 25-50 μ l 细菌 5 分钟。随后,将 3-5 μ l 连接产物添加至感受态细菌并在冰上温育 20-30 分钟,随后在 42°C 热休克 1 分钟。随后,每管添加 500 μ l S.O.C 培养基 (Invitrogen, Carlsbad, CA, 美国) 并在 37°C 在搅拌下温育 1 小时。最后,将细菌涂布在含氨苄青霉素 (Sigma-Aldrich, St. Louis, MO, 美国) 的 LB 平板上并在 37°C 温育过夜。为了在 pCR-Blunt 载体中克隆,使用含卡那霉素 (Sigma-Aldrich, St. Louis, MO, 美国) 的平板。

[0224] I. 1. 7. 小规模 (小量) 和中等规模 (中等量) 质粒制备

[0225] I. 1. 7. 1. 小量制备

[0226] 为了小量制备,将转化细菌的菌落在 2.5ml LB 和氨苄青霉素或卡那霉素中在 37°C, 200rpm 培育 6-16 小时。遵循提供的手册,用大肠杆菌质粒纯化试剂盒 (QuickPure, Macherey Nagel, Oensingen, 瑞士) 提取 DNA。

[0227] 通过测量在 260nm 处的吸光度并评估必须在 1.8 和 2 之间的 OD260nm/OD280nm 比率,用 NanoDrop ND-1000 分光光度计 (Thermo Scientific, 威名顿, DE, 美国),定量来自小量制备的质粒 DNA 一次。在发送样品到 Fasteris SA (Geneva, 瑞士) 以便序列证实之前,进行对照消化。

[0228] 对于 BAC 提取,使用 QuickPure 试剂盒 (Macherey Nagel, Oensingen, 瑞士),同时对方案作出以下修改:用含有 pBACe3.6 载体的细菌接种 10ml LB 和氯霉素 (12.5 μ g/ml) (Sigma-Aldrich, St. Louis, MO, 美国)。在振动平台上 37°C 温育过夜后,将培养物以 13 300rpm 离心 5 分钟,之后重悬于 500 μ l A1 缓冲液中。添加 500 μ l A2 裂解缓冲液并且将溶液在室温温育 5 分钟。随后,将该溶液用 600 μ l A3 缓冲液中和并以 13 300rpm 离心 10 分钟。将上清液加载在柱上,并且从这个步骤往后,使用 QuickPure 小量制备试剂盒的标准方案。

[0229] I. 1. 7. 2. 中等量制备

[0230] 为了中等量制备,将转化的细菌在 200 至 400ml LB 和氨苄青霉素 (或卡那霉素) 中于 37°C 培育过夜。随后,将培养物以 725g 离心 20 分钟并且使用商业试剂盒 (NucleoBond Xtra Midi ;Macherey Nagel, Oensingen, 瑞士),遵循制造商的手册中提供的低拷贝质粒方案纯化质粒。

[0231] 用 NanoDrop ND-1000 分光光度计 3 次定量来自中等量制备的质粒-DNA,用限制性消化物证实并最终发送测序 (Fasteris SA, Geneva, 瑞士)。

[0232] II. 结果和讨论

[0233] II. 1. 克隆 GAPDH 表达盒上游和下游的 DNA 区域 (5' 和 3' GAPDH)

[0234] 在 Imagen (柏林, 德国) 订购 BAC 克隆 RPCIB753F11841Q。这个克隆在含有氯霉

素抗性基因的 pBACe3.6 载体主链中含有人 GAPDH 序列。在通过小量制备提取 DNA 后,通过 Nanodrop 确定载体浓度是 27ng/ μ l。

[0235] 使用 27ng 纯化的克隆 RPCIB753F11841Q 作为模板,通过 PCR 扩增启动子上游和多聚腺苷化位点下游紧邻 GAPDH 表达盒周围的 DNA 序列。用引物 GlnPr1171 (SEQ ID NO:1) 和 GlnPr1172 (SEQ ID NO:2) 扩增启动子上游的 3kb 片段,产生具有 SEQ ID No. 5 的扩增子。由于引物 GlnPr1172 (SEQ ID NO:2) 相对于模板序列携带碱基变化 (G 至 T),源自这个 PCR 反应的全部序列也将携带这个碱基变化。该变化位于相对于 GAPDH 基因转录起始的位置 -3721 内 (SEQ ID NO:17 的 bp812,相对于 SEQ ID NO:5 起始的位置 23)。用引物 GlnPr1173 (SEQ ID NO:3) 和 GlnPr1174 (SEQ ID NO:4) 扩增多聚腺苷化位点下游的 3kb 片段,产生具有 SEQ ID NO:6 的扩增子 (表 1)。用于这些 PCR 的退火温度是 72°C。

[0236] 将 5' 和 3' GAPDH 片段 (SEQ ID NOs:5 和 6) 克隆在市售 PCR 产物克隆载体 pCR-Blunt (pCR-Blunt, PCR Zero Blunt 克隆试剂盒, Invitrogen) 中。将连接产物转化至 TOP10 感受态细菌中并涂布在卡那霉素 LB-琼脂平板上。放大菌落并通过小量制备法分离质粒。进行对照消化以鉴定阳性克隆,其产生 pCR-Blunt-5' GAPDH 和 pCR-Blunt-3' GAPDH 构建体。

[0237] II. 2. 制备编码报道蛋白 GFP 和重组 IgG1 单克隆抗体的 DNA 片段 (LC-IRES-HC-IRES-GFP)

[0238] 本工作中使用的报道基因构建体 (REP) 以多顺反子基因方式组成: IgG1 单克隆抗体轻链 (LC)-IRES-IgG1 单克隆抗体重链 (HC)-IRES-绿色荧光蛋白 (GFP)。源自脑心肌炎病毒的内部核糖体进入位点 (IRES) (Gurtu 等人, Biochem Biophys Res Commun. ; 229(1):295-298, 1996)) 的存在允许以下 3 种肽的翻译: IgG1 单克隆抗体轻链 (LC)、IgG1 单克隆抗体重链 (HC) 和 GFP (图 1)。转染的细胞将因此分泌 IgG1 单克隆抗体并以依赖性方式积累胞内 GFP。然而,多顺反子 mRNA 在真核细胞中不常见并且它们的翻译不是非常有效的,导致 IgG1 和 GFP 表达的滴度相对低。

[0239] 使用限制性酶 NheI 和 BstBI (BstBI 在 65°C 使用) 消化含有 REP 构建体的载体。将含有该表达构建体的 REP 片段切下,纯化和用于进一步的克隆步骤。

[0240] II. 3. 表达载体的克隆

[0241] 将载体 pGLEX41 (一种源自 pcDNA3.1 (+) (Invitrogen, Carlsbad, CA) 的表达载体) 用于产生稳定细胞系。使用它作为初始主链,所述初始主链已经被修饰以产生具有和不具有 GAPDH 序列的第二代载体 A 和 B。对于全部载体,使用相同的启动子-内含子组合 (mCMV 和编码第一内含子 (IgDA) 的供体-接纳体片段) (Gorman 等人, (1990) Proc Natl Acad Sci USA, 87:5459-5463)。

[0242] 中间载体 pGLEX41-HM-MCS-amp^r 的克隆:

[0243] 新一代载体的开发始自 pGLEX41。使用限制性酶 NruI 和 BspHI 切割这种载体以释放氨苄青霉素抗性盒。将主链片段用 CIP 处理并通过凝胶电泳纯化。已经从 GeneArt 订购编码密码子优化 (以便在大肠杆菌中表达) 形式的氨苄青霉素抗性基因 (包含 bla 启动子) 的 DNA 片段。使用限制性酶 NruI 和 BspHI (与针对主链所用相同的酶), 从 GeneArt 克隆载体 #1013237 切下插入物, 将其纯化并克隆入主链中。

[0244] 通过限制性消化分析小量制备物。克隆 pGLEX41-HM-MCS-amp^r#2 具有期望的限

制性图谱并且通过测序证实了正确片段的整合。

[0245] 中间载体 pGLEX41-MCS-R6K-ampⁱA 的克隆：

[0246] 为了交换载体 pGLEX41-HM-MCS-ampⁱA#2 的 pUC 复制起点，使用 PvuI 和 BspHI 消化该载体。将主链片段用 CIP 处理并纯化。新插入片段含有 R6K 复制起点和修饰的 SV40poly(A) 序列作为表达盒的部分。已经消除 SV40poly(A) 周围不必要的细菌或病毒主链序列（见下表 2）。已经从 GeneArt 订购插入片段；使用酶 PvuI 和 BspHI（与针对主链所用的酶相同），从 GeneArt 克隆载体 #1013238 切下该片段，将其纯化并克隆入主链片段中。制备小量制备物并通过序列分析验证。克隆 pGLEX41-MCS-R6K-ampⁱA#1 具有正确的序列。

[0247] 表 2：不同载体中 CpG 的含量

[0248]

	表达载体中的 CpG 含量		
	pGLEX41	密码子优化的载体：“A”	CpG 减少的载体“B”
氨苄青霉素抗性	49	43	19
嘌呤霉素抗性	93	36	1
遗传霉素抗性	74	51	0
复制起点	45	9	9
总计：	261	139	29

[0249] 中间载体 pGLEX41-MCS-R6K-ampⁱB 的克隆

[0250] 使用限制性酶 BspHI 和 CIP，打开载体 pGLEX41-MCS-R6K-ampⁱA#1 以释放氨苄青霉素抗性。新的插入片段含有经优化以便在大肠杆菌中表达的氨苄青霉素抗性密码子，但是，已经替换可以通过备选密码子选择予以消除的全部 CpG 序列（见上表 2）。这个片段在 GeneArt 订购。为了释放插入片段，使用 BspHI 消化 GeneArt 克隆载体 #1016138。通过凝胶电泳纯化插入片段和主链片段后，将它们连接并转化至 PIR1 细菌中。将小量制备物直接发送测序。pGLEX41-R6K-MCS-ampⁱB#1 具有正确序列并且用于进一步克隆步骤。

[0251] 在 pGLEX41 衍生的表达载体中克隆报道基因构建体

[0252] 为了在表达载体 pGLEX41-MCS-R6K-ampⁱA 和 pGLEX41-MCS-R6K-ampⁱB 中克隆报道基因构建体 REP，使用限制性酶 NheI 和 ClaI 切割载体。使用限制性酶 NheI 和 BstBI（在 65℃）打开表达载体 pGLEX41-HM-MCS。在消化后，全部载体主链用 CIP 处理并且通过凝胶电泳纯化主链。将主链与编码报道基因构建体 REP 的 NheI/BstBI（BstBI 与 ClaI 相容）片段连接。将连接产物转化至 PIR1 或 TOP 10 感受态细菌中并涂布在氨苄青霉素 LB-琼脂平板上。放大菌落并通过小量制备法分离质粒。通过小量制备物的限制性消化鉴定阳性克隆并且由 Fasteris SA 后续测序证实。

[0253] 在 pGLEX41 衍生的表达载体中添加 GAPDH 侧翼序列

[0254] 这个段落的全部限制性消化在 80 μl 终体积中进行并在 37℃ 温育过夜。

[0255] 使用限制性酶 NruI，将 5' GAPDH 序列（SEQ ID NO:7）从 pCR-blunt-5' GAPDH 切下并在使用 NruI 作线性化并用 CIP 处理以避免再环化的表达载体 pGLEX41-R6K-ampⁱA-[REP]

和 pGLEX41-R6K-ampB-[REP] 中连接。在扩增（通过转化连接产物所获得的）PIR1 菌落后，通过限制性消化分析小量制备物。克隆 pGLEX41-R6K-ampA-5' GAPDH-[REP]#2 和 pGLEX41-R6K-ampB-5' GAPDH-[REP]#1 在限制性分析中显示预期大小的条带，随后通过测序证实并用于其他克隆步骤。将这些新载体随后用 ScaI 打开并用 CIP 处理。使用相同酶，将 3' GAPDH 片段 (SEQ ID NO:8) 从 pCR-Blunt-3' GAPDH 切下并连接至两个主链中，以产生 pGLEX41-R6K-ampA-GAPDH-[REP] 和 pGLEX41-R6K-ampB-GAPDH-[REP] 表达载体。

[0256] 克隆 pGLEX41-R6K-ampA-GAPDH-[REP]#2 和 pGLEX41-R6K-ampB-GAPDH-[REP]#8 的对照消化物在限制性分析中显示预期大小的条带。随后通过测序证实 3' GAPDH 片段以正确方向插入 (Fasteris)。

[0257] II. 4. 抗性载体的克隆

[0258] 克隆抗性载体的起点是载体 pGLEX-MCS-R6K-ampA#1。对于抗性基因的表达，弱启动子足够用，将 mCMV 启动子替换为 SV40 启动子。编码抗性基因的基因从 GeneArt SA (雷根斯堡, 德国) 订购且将其为在中国仓鼠 (嘌呤霉素 :puroA 以及新霉素 :neoA) 中的表达而优化或通过选择性密码子选择降低 CpG 含量 (嘌呤霉素 :puroB 以及新霉素 :neoB)。

[0259] pGLEX-R6K-AmpA-PuroA/PuroB 的克隆：

[0260] 为了在表达盒中克隆嘌呤霉素抗性，使用限制性酶 NruI 和 Xba I 打开载体 pGLEX41-MCS-R6K-ampA#1，随后用 CIP 处理。插入片段从 GeneArt 订购并且作为 GeneArt 克隆载体 #1013239 中的插入物提供。它含有 SV40 启动子和密码子优化的嘌呤霉素抗性基因（针对 CHO 细胞的密码子选择）。使用酶 NruI 和 XbaI (与针对主链所用相同的酶)，从 GeneArt 克隆载体切下插入物，将其纯化并克隆入主链片段中。制备小量制备物并通过限制性消化进行分析。克隆 pGLEX-MCS-R6K-ampA-puroA#1 显示正确的图谱并可以通过测序证实。

[0261] 通过交换嘌呤霉素抗性基因的编码区，同时保留 SV40 启动子，将这个载体用于载体 pGLEX-MCS-R6K-ampA-puroB 的克隆。新插入片段含有密码子优化形式的嘌呤霉素基因，其中已经替换因备选密码子选择而可消除的全部 CpG 序列。该片段已经从 GeneArt 订购并且在克隆载体 #1016139 中输送。为了释放插入片段，使用限制性酶 Xba I 和 NotI 消化 GeneArt 载体。将插入片段通过凝胶电泳纯化并克隆至 pGLEX-MCS-R6K-ampA-puroA 的主链中，随后通过使用 XbaI 和 NotI 限制性消化释放嘌呤霉素可读框，然后进行 CIP 处理。通过序列分析直接证实所产生的载体 pGLEX-MCS-R6K-ampA-puroB#1。

[0262] 载体 pGLEX-R6K-ampA-NeoA 和 pGLEX-R6K-ampA-NeoB 的克隆

[0263] 为了在表达盒中克隆新霉素抗性基因，使用限制性酶 Xba I 和 NotI 打开载体 pGLEX-R6K-puroA#1，随后用 CIP 处理。插入片段从 GeneArt 订购并且作为 GeneArt 克隆载体 #1013242 (neoA) 和 #1026894 (neoB) 中的插入物提供。它们分别含有针对 CHO 细胞密码子选择进行密码子优化的新霉素抗性基因和 CpG 缩减形式的新霉素抗性基因。使用酶 XbaI 和 NotI (与针对主链所用相同的酶)，从 GeneArt 克隆载体切下插入物，将其纯化并克隆入主链片段中。制备小量制备物并通过测序证实克隆。

[0264] 载体 pGLEX-R6K-ampB-NeoB 和 pGLEX41-R6K-ampB-puroB 的克隆：

[0265] 将载体 pGLEX41-R6K-puroB#1 使用限制性酶 BspHI 打开并随后作 CIP 处理。插入片段含有经密码子优化以便在大肠杆菌中表达的氨苄青霉素抗性基因，同时已经替换因

备选密码子选择而可消除的全部 CpG 序列。这个片段已经在 GeneArt 订购并在克隆载体 #1016138 中抵达。为了释放插入片段,使用 BspHI 消化 GeneArt 克隆载体。通过凝胶电泳纯化插入片段和主链片段后,将它们连接并转化至 PIR1 细菌中。将小量制备物直接发送测序并且其可以得到证实 (pGLEX41-amp^rB-R6K-puroB#1)。

[0266] 通过使用限制性酶 BspHI 打开 pGLEX-R6K-neoB-amp^rA 以产生主链片段,进行克隆,产生载体 pGLEX-R6K-neoB-amp^rB。使用相同的限制性酶组合对 pGLEX-R6K-amp^rB-hygroB 的消化产生了编码 amp^rB 的插入片段。将 amp^rB 插入物克隆至 pGLEX-R6K-neoB-amp^rA 主链中。

[0267] II.5 添加人 GAPDH 基因上游和下游的序列至抗性载体中

[0268] 用 NruI 消化载体 pCR-blunt-5' GAPDH 以获得 5' GAPDH 插入物 (3164bp)。将编码抗性基因的载体用 NruI 消化,随后用 CIP (牛小肠磷酸酶, NEB, Ipswich, MA) 处理以制备主链片段。将 4 个不同主链片段 (pGLEX-R6K-neoA-amp^rA、pGLEX-R6K-neoB-amp^rB、pGLEX-R6K-puroA-amp^rA 和 pGLEX-puroB-amp^rB) 与 3164bp 的 5' GAPDH 插入物连接并转化至 PIR1 感受态细菌中。使用 ApaI 对小量制备物的限制性消化允许鉴定出克隆 pGLEX-R6K-neoB-amp^rB-5' GAPDH#5、pGLEX-R6K-neoA-amp^rA-5' GAPDH#6、pGLEX-R6K-puroA-amp^rA-5' GAPDH#16 和 pGLEX-puroB-amp^rB-5' GAPDH#5。

[0269] 这些中间载体随后用限制性酶 ScaI 切割并用 CIP 处理以制备主链用于连接。使用 ScaI 切割携带第二插入片段 pCR-Blunt-3' GAPDH 的载体以释放插入片段 (3224bp), 即 GAPDH 下游侧翼区。将 4 个不同主链分子与纯化的 3224bp 插入片段连接并转化至 PIR1 感受态细胞中。通过限制性消化分析小量制备物。显示预期大小的限制性片段的克隆是 pGLEX-R6K-neoB-amp^rB-GAPDH#8、pGLEX-R6K-neoA-amp^rA-GAPDH#1、pGLEX-R6K-puroA-amp^rA-GAPDH#1 和 pGLEX-puroB-amp^rB-GAPDH#4。随后通过测序分析证实这些克隆 (Fasteris, Geneva, 瑞士)。

[0270] II.1.5. 为转染所克隆的质粒的中等量制备

[0271] 为了具有足够量的质粒,使用 Macherey Nagel 试剂盒 (NucleoBond Xtra Midi; Macherey Nagel, Oensingen, 瑞士) 制备中等量制备物。通过限制性消化和测序证实后,将质粒线性化并用于 CHO-S 细胞中转染。表 3 汇总了以中等量制备获得的质粒 DNA 批次、已经被制备用于转染的线性化 DNA 制备物、用于线性化的酶的浓度和来自 Fasteris SA 的证实各个质粒的身份和序列信息的序列档案。在用于转染之前,全部中等量制备物均通过测序证实。

[0272] 表 3: 克隆的质粒的总结。DNA 中等量制备物和线性化中等量制备物的浓度 (采用相应的酶)。GSC 号编码各个质粒并允许确定相关的测序档案。

[0273]

质粒	中等量制备物的 浓度($\mu\text{g/ml}$)	用于线性 化的酶	线性化质粒的 浓度($\mu\text{g/ml}$)	Glenmark 质粒代码
pGLEX41-R6K-AmpiA- [REP]-GAPDH	1538	EcoRV	1019	GSC 2774
pGLEX41-R6K-AmpiB- [REP]-GAPDH	1243	EcoRV	1233	GSC 2775
pGLEX-R6K-AmpiA-neoA- GAPDH	890	AseI	766	GSC 2776
pGLEX-R6K-AmpiB-neoB- GAPDH	594	AseI	979	GSC 2777
pGLEX-R6K- AmpiA- puroA- GAPDH	917	AseI	859	GSC 2778
pGLEX-AmpiB-puroB- GAPDH	869	AseI	1049	GSC 2779
pGLEX41-[REP]	2119	BspHI	868	GSC 2239
pGLEX41-R6K-AmpiA- [REP]	865	BspHI	779	GSC 2240
pGLEX41-R6K-AmpiB- [REP]	1751	BspHI	806	GSC 2249
pGLEX-R6K-AmpiA-neoA	890	BspHI	764	GSC 2214
pGLEX-R6K-AmpiB-neoB	767	BspHI	654	GSC 2244
pGLEX-R6K-AmpiA-puroA	708	BspHI	659	GSC 2220
pGLEX-R6K-AmpiB-puroB	574	BspHI	746	GSC 2213

[0274] 实施例 2 :用表达载体转染细胞 :

[0275] 1. 材料和方法

[0276] CHO-S 细胞和 HEK293 细胞

[0277] 哺乳动物细胞是表达蛋白质的优选宿主,因为它们能够将重组蛋白正确折叠、装

配和转录后修饰。使用 CHO 细胞系,原因是它们经充分表征并且不作为大多数对人致病的病毒的宿主,这使它们成为相对安全的稳定产生治疗性蛋白的宿主。将中国仓鼠卵巢细胞 (CHO-S, Invitrogen, Carlsbad, CA, 美国) 在补充有 4mM L-谷氨酰胺 (Applichem, 德国) 并在摇床 (200rpm, 具有 2.5cm 环状划桨) 中温育的 PowerCHO-2CD 培养基 (Lonza, Verviers, 比利时) 中于 37℃、5% CO₂ 和 80% 湿度悬浮培养。

[0278] 使用 HEK293 细胞,原因是它们易于转染并允许快速产生重组蛋白直至较低克数量。所用的细胞是 HEK293-EBNA 细胞 (ATCC, 马纳萨斯, VA) 并且在 Ex-cell 293 培养基 (Sigma-Aldrich, St. Louis, MI) 中常规悬浮培养。

[0279] 使用接种密度 0.5×10^6 个活细胞 /ml, 每 3-4 日在新鲜培养基中常规实施 CHO-S 和 HEK293EBNA 细胞的继代培养。使用 10ml 培养基, 在含有允许气体交换的可透性滤器的 50ml 生物反应器管 (Tubespın Bioreactor50 ;TPP, Trasadingen, 瑞士) 中培养细胞。使用台盼蓝细胞排斥法 (cell exclusion method), 用 Countess 自动化细胞计数器 (Invitrogen, Carlsbad, CA, 美国) 确定细胞活力和浓度。对于 CHO-S 细胞使用 PCV 管 (TPP, Trasadingen, 瑞士), 通过细胞压积法 (packed cell volumn, PCV) 确定证实细胞浓度。

[0280] 细胞压积 (PCV)

[0281] PCV 方法基于微量 PCV 管 (PCV 细胞压积管 ;TPP, Trasadingen, 瑞士) 中以 5000rpm 离心特定体积的培养液体 1 分钟。在离心期间, 细胞沉淀在管基部处的有刻度的毛细管中。随后通过相对于所离心的细胞培养液的量评估沉淀物体积确定细胞压积的百分数。例如, 1% PCV 表明 $10 \mu\text{l}$ 细胞沉淀物存在于 1ml 培养液中。

[0282] 对于细胞的常规细胞计数, 将 $200 \mu\text{l}$ 每份样品移入 PCV 管中并且用尺 (“易读取” 测量装置 ;TPP, Trasadingen, 瑞士) 读取相应沉淀物的体积 (以 μl 计)。将这个体积放大 5 倍以使该值针对 1ml 并且随后使用细胞特异性相关系数, 乘以该值以获得活细胞浓度的估计值 (以 “百万个细胞 /ml” 计)。

[0283] “自动化” 细胞计数

[0284] 在样品与相同量的台盼蓝混合时, 用 **Countess[®]** 自动化细胞计数器 (Invitrogen, Carlsbad, CA, 美国) 确定细胞浓度和活力。随后将溶液移入 **Countess[®]** 室载玻片中, 之后由该仪器读取。这种仪器允许 Neubauer 室的自动读出, 在校正后, 所述自动读出确定细胞活力和死细胞及活细胞的浓度。

[0285] 流式细胞术分析

[0286] 流式细胞术是一项分析单独细胞的多个参数的技术。这项技术允许定量和定性分析表型上彼此不同的细胞, 例如死细胞与活细胞 (根据细胞的大小和颗粒性)。它还允许对表达目的蛋白如 GFP 的细胞定量。通过无菌抽吸 $300 \mu\text{l}$ 样品从培养物收集细胞并且用配备 488nm 处发射的气冷氩激光的荧光相关细胞分选 (FACS) Calibur 流式细胞仪 (Becton, Dickinson and Company, Franklin Lakes, NJ, 美国) 分析。用 CellQuest 软件进行分析。使用 530/30-nm 带通滤波器 (band pass filter), 用 FL-1 检测到 GFP 发射。

[0287] 在第一门中, 在 SSC/FSC 点图中按线性标度从分析排除细胞残片以及死细胞。随后, 活细胞的 GFP 荧光按对数标度在直方图中显示。荧光分布的中位值用来评估所分析的细胞群体的 GFP 表达水平。

[0288] IgG 定量方法 :OCTET QK

[0289] OCTET QK 系统 (FortéBio, Menlo Park, CA, 美国) 执行抗体、蛋白质、肽、DNA 和其他生物分子的无标记物定量并提供生物分子结合相互作用的动力学表征。结合速率 (nm) 和样品的积累 IgG1 浓度 ($\mu\text{g/ml}$) 之间的相关性允许用校正曲线对 IgG 滴度定量。

[0290] 将细胞样品以 300g 离心 5 分钟。随后将上清液用 Octet 缓冲液 (对 IgG1 抗体而言 1/5) 在 96 孔平板中稀释, 之后, 使用蛋白 A 生物传感器 (Protein A DIP and READ™ Biosensor, Forté Bio, USA), 以 Octet 分析以获得每孔的抗体浓度。

[0291] 使用 JetPEI 瞬时转染

[0292] 使用聚乙烯亚胺 (PEI ; JetPEI, Polyplus-transfection, Illkirch, 法国) 进行 CHO-S 和 HEK293EBNA 细胞的瞬时转染和稳定转染。PEI 是可以与带负电荷的分子如 DNA 复合的阳离子聚合物。带正电荷的 DNA-PEI 复合物与带负电荷的细胞表面结合并且由胞吞作用内化。它达到溶酶体区室, 从这里它由裂解释放至核。DNA-PEI 复合物的高转染效率因 PEI 保护 DNA 免于溶酶体降解的能力所致。根据制造商提供的手册转染细胞。

[0293] 全部质粒在稳定转染之前均线性化 (重悬在 $100\mu\text{l}$ Tris-EDTA, pH7.5 中的 $100\mu\text{g}$ DNA)。对于瞬时转染, 直接使用来自中等量制备 DNA 的环状质粒。在这项研究中, 在 50ml 生物反应器管中维持瞬时转染并且不添加抗生素。

[0294] 通过共转染一个表达载体和两个抗性载体 (分别编码嘌呤霉素或新霉素抗性) 获得了表达 IgG1 和 GFP 的稳定 CHO-S 克隆。

[0295] 稳定汇集物和小量汇集物的选择

[0296] 通过分析胞内 GFP 表达, 借助流式细胞术 (BD FACS Calibur 细胞计数器, #1293) 在转染后 24 小时确定转染效率。如果 GFP 阳性细胞的百分数高于 20%, 将转染的细胞用选择培养基稀释并分配至 96 孔平板中 (对于产生分离的稳定小量汇集物的有限稀释法) 或在 T-Flask 中 (以产生稳定汇集物)。使用的选择培养基是 PowerCHO-2, 4mM 谷氨酰胺, 补充有不同浓度的遗传霉素和嘌呤霉素。

[0297] 在转染后 7 日, 通过添加选择培养基至细胞更新选择严格性。一旦 96 孔平板中的集落汇合, 使用荧光读数仪读取平板。

[0298] 使用无抗生素的 PowerCHO-2, 4mM L- 谷氨酰胺, 将 T-Flask 中的汇集物扩充至旋转管规模。用 Countess 自动化细胞计数器 (Invitrogen, Carlsbad, CA, 美国) 评价它们的活力和浓度。一旦细胞密度允许, 则通过以密度 0.5×10^6 细胞 /ml 在 50ml 生物反应器管中的 10ml 培养基内接种细胞 (在摇床 (200rpm) 中于 5% CO_2 , 37°C 和 80% 湿度温育), 对每个汇集物启动种子培养。通过以 0.5×10^6 细胞 /ml 在生长培养基中接种细胞 (通过 PCV 分析确定细胞浓度), 将每个种子培养物一周 2 次传代。种子培养物用于接种全部生产过程 (批次)。

[0299] 对于下一个 4-5 周, 一周一次一式两份接种生产过程。如上文对克隆群体所述, 通过 FACS 和 IgG 表达评价汇集物稳定性。

[0300] 生产过程 (分批发酵)

[0301] 使用用于接种的种子培养物, 以 0.5×10^6 个细胞 /ml 的浓度接种细胞汇集物的批量 (batch run), 并且随后将细胞在进料培养基中培养 7 日。在第 4 日和第 8 日, 将 $200\mu\text{l}$ 细胞以 300g 离心 5 分钟并且使用 Octet, 对上清液分析积累的 IgG。此外, 通过 FACS 分析

每个批次的 GFP 表达。

[0302] 2. 结果

[0303] 2.1 CHO 细胞中的瞬时表达：

[0304] 这项研究中所比较的载体主要区别在于它们的主链。全部载体的整个表达盒（启动子、第一内含子、表达构建体，poly(A)）是完全相同的。载体源自如实施例 1 中所述的载体 pGLEX41。在一个载体中，氨苄青霉素抗性基因经密码子优化以便在大肠杆菌中表达并且将细菌主链缩减至最少：pGLEX41-R6K-Amp^rA-[REP]（简言之 A）。在第二载体中，氨苄青霉素抗性基因经密码子优化以便在大肠杆菌中表达，但是通过使用备选密码子（当可能时），避开全部 CpG 序列：这个载体称作 pGLEX41-R6K-Amp^rB-[REP]（简言之 B）。第三修改包括使用克隆在载体 A 和 B 的表达盒上游和下游的 GAPDH 侧翼序列，产生载体 pGLEX41-R6K-Amp^rA-[REP]-GAPDH（简言之 GAPDH_A）和 pGLEX41-R6K-Amp^rB-[REP]-GAPDH（简言之 GAPDH_B）。

[0305] 进行 CHO-S 细胞（Invitrogen）的瞬时转染以比较在不同质粒主链的环境下表达的报道蛋白的表达水平。转染（一式两份）在 50ml 生物反应器管（TPP, Trasadingen, 瑞士）中使用 10ml 最终培养基体积进行并在转染后第 5 日由 Octet 分析（图 2）。

[0306] 具有经修正的主链的全部载体（A 和 B）显示比对照载体 pGLEX41 略微更高的表达水平。在载体 A 和载体 B 之间仅存在少量差异。这是预期的，因为主链中的唯一差异是不应当对瞬时表达产生影响的氨苄青霉素抗性。

[0307] 最惊人的观察结果是 GAPDH 序列对表达的积极影响。与不携带 GAPDH 序列的质粒相比，采用携带 GAPDH 侧翼序列的质粒时，获得 2 倍更高的表达水平。对于构建体 A 和 B 均是如此。与 pGLEX41 载体相比，可以观察到 3 倍更高的表达。如果考虑质粒的尺寸，这甚至更令人惊讶。与载体 GAPDH-A (13436bp) 相比，载体 A (7048bps) 的尺寸几乎是其一半。因此，假定瞬时转染过程期间递送的 DNA 的量对于全部质粒是相同的，则仅半数摩尔量的 GAPDH-A 被递送至核。

[0308] 2.2 HEK293 细胞中的瞬时表达

[0309] 进行 HEK293EBNA 细胞的瞬时转染以比较在不同质粒主链的环境下表达的报道蛋白的表达水平。转染（一式两份）在 50ml 生物反应器管（TPP, Trasadingen, 瑞士）中使用 10ml 终末培养基体积进行并在转染后第 10 日由 Octet 分析（图 3）。

[0310] 图 3 中显示的结果表明 HEK293EBNA 细胞中的表达显著增加，这可以使用 GAPDH 侧翼区获得。GAPDH-B 载体显示表达增加三倍，而 GAPDH-A 载体显示甚至更高的 5 倍表达增加。这些载体不含有 oriP 元件并且可能因此具有甚至更高滴度的潜能。

[0311] 2.3 稳定 CHO 细胞系中的表达

[0312] 建立稳定的转染细胞

[0313] 通过共转染表达载体和编码抗性基因的载体，随后采用抗生素介导的选择压力，产生稳定的群体。在转染后 14 日移除选择压力。这些步骤导致产生稳定的小量汇集物和稳定的汇集物，将所述汇集物在生产过程中以定期间隔培养，以比较不同构建体的报道蛋白（IgG1 抗体和 GFP）的表达水平和表达的稳定性。

[0314] 在生产过程中用细胞汇集物进行的报道蛋白表达研究

[0315] 通过稳定转染产生汇集物。在选择过程期间（转染后第一个 14 日），通过 FACS 分

析汇集物。可以在这个时间范围内观察到 GFP 阳性细胞部分的增加,连同培养物活力的增加。14 日后从汇集物移除抗生素介导的选择压力。使用这种方法,不能获得用“B”质粒转染的细胞汇集物。一旦细胞可以在 50ml 生物反应器管中培养,则测定生成的汇集物的表达水平。一式两份进行批次。通过 FACS 分析细胞的 GFP 表达,并且在表达 8 日后通过 Octet 测定上清液中 IgG 的积累。

[0316] 可以在汇集物的 IgG 滴度和 GFP 表达之间观察到成比例关系。因此,图 4 中仅显示 IgG 数据。与载体 pGLEX41 或不含有 GAPDH 序列的相同载体相比,用含有 GAPDH 序列的载体转染的全部汇集物均显示更高的表达(A 和 A-GAPDH 之间 2.8 倍。在 B 和 B-GAPDH 之间不能得出结论,因为无 B 汇集物存活)。用 A-GAPDH 和 B-GAPDH 进行的转染引起比 pGLEX41 转染更高的 IgG 表达(分别是 2.7 倍和 3.5 倍)(对于批次 -2 而言)。因此在汇集物中,GAPDH 侧翼序列看起来有利于蛋白质的产生。

[0317] 最后,用 B-GAPDH 载体进行的转染引起比用 A-GAPDH 进行的转染更高的 IgG 表达(1.25 倍)。因此,抗性基因中 CpG 减少看起来也有利于蛋白质的稳定产生。

[0318] 在克隆群体上的表达水平研究

[0319] 将细胞转染并在 96 孔平板中分配于选择培养基中以获得克隆或寡克隆群体。在 7 日后,通过向细胞添加选择培养基更新选择压力。转染后 14 日,通过使用 ELISA- 平板读数仪评估 GFP 的表达。图 5 中显示结果。

[0320] 证实了细胞汇集物中所获得的结果,用含有 GAPDH 侧翼序列的载体转染的细胞比不含有 GAPDH 上游和下游序列的相同主链表达显著更多的 GFP(倍数从 1.7 至 2 倍)或比作为对照使用的其他载体表达显著更多的 GFP(pGLEX41 :2.5 倍)(图 5)。此外,含有带抗性序列的载体(B)的群体比相应载体(A)引起更高的表达,其中所述抗性序列已经减少 CpG,所述相应载体仅已经被密码子优化(A 和 B 之间 1.5 倍;B 和 B-GAPDH 之间 1.2 倍)。

[0321] 可以从表达研究得出几个结论。首先,GAPDH 上游和下游序列允许比作为基准使用的标准载体(pGLEX41)更高的表达。另外,用不含 GAPDH 序列的相同载体主链转染细胞时获得较低的表达水平证实,对表达的有益影响与插入的 GAPDH 侧翼序列相关。此外,表达质粒和选择质粒中 CpG 数目的减少看起来也略微有利于表达。

[0322] 实施例 3:用新设计的载体转染的 CHO-S GMP 细胞的瞬时表达水平

[0323] 已经在文献中描述,GAPDH 启动子的 5' 区域携带潜在的胰岛素应答元件及佛波醇酯应答元件(Alexander-Bridges 等人,(1992)Advan Enzyme Regul, 32:149-159)。佛波醇酯应答元件(-1040-1010bp)位于通常称作 GAPDH 启动子的上游(-488-+20)。在稳定 H35 肝细胞瘤细胞系中所进行的缺失研究中,作者不能够展示缺失碱基对 -1200 至 -488 的显著影响(相对于转录起始点)。因此,佛波醇酯应答元件可能无法与从 GAPDH 启动子驱动的表达功能性联系。然而,进行瞬时转染实验以评价胰岛素和 PMA(佛波醇-12-肉豆蔻酸酯-13-乙酸酯,最常见佛波醇酯)在使用含有 GAPDH 侧翼元件的质粒时观察到的瞬时表达和稳定表达增加方面的贡献。

[0324] 为了获得无胰岛素的生长培养基,从粉状培养基制备 PowerCHO2 并且不添加胰岛素。PMA 从 Sigma(St. Louis, MO) 购买并且以终浓度 1.6 μ M(对应于 Alexander-Bridges 对 H35 肝细胞瘤细胞系所使用的浓度)投入 PowerCHO2(+/- 胰岛素)中。

[0325] 如先前所描述在 50ml 生物反应器管(Tubespins, TPP, Trasadingen, 瑞士)中进行

转染。为了避免存在 OptiMEM(Life technologies, Carlsbad, CA) 提供的胰岛素, 将转染培养基更换成补充有 4mM Gln 和 25mM HEPES 的 RPMI1640 (PAA, Pasching, Austria)。在转染后, 将细胞分配在 12 孔平板中并添加 1ml 四种不同的培养基 (PowerCHO2, 4mM Gln, +/- 胰岛素; PowerCHO2, 4mM Gln, 1.6 μ M PMA, +/- 胰岛素)。再次, 使用 (实施例 2 中描述的) 利用两个 IRES 表达 IgG1 和 GFP 的报道基因构建体。这个载体允许验证转染效率。在全部 4 种不同的培养基制备物中发现转染细胞的百分数和活力相似。

[0326] 如 6 中所显示图, 未在这个实验期间观察到胰岛素耗尽和 / 或 PMA 添加的明显影响。在用于表达的全部培养基中均获得相似的滴度。

[0327] 这表明存在于 GAPDH 基因侧翼序列上游中的潜在佛波醇酯应答元件和胰岛素应答元件不影响瞬时转基因表达。

[0328] 实施例 4: 在启动子上游和 polyA 位点下游的 GAPDH 表达盒侧翼的 DNA 的片段化分析, 旨在研究对报道基因表达的影响

[0329] 人 GAPDH 基因座位于人基因组的第 12 号染色体上。据称 GAPDH 在哺乳动物来源的全部细胞中具有组成型活性, 因为该酶是葡萄糖代谢中的关键角色。在该启动子的上游, GAPDH 基因侧翼为 NCAPD2, 即伸展多于 30000bp 的基因。在多聚腺苷化位点下游, GAPDH 基因侧翼为 IFF01 (详见图 7)。

[0330] GAPDH 和该启动子以及侧翼区在不同物种之间均保守 (见表 4)。

[0331] 表 4: 人、大鼠和小鼠 GAPDH 侧翼区之间的高度同源区段。使用克隆管理者 9 (ScieED, Cary, NC, USA) 作出分析。编号分别相对于上游侧翼元件或下游侧翼元件 (分别是序列 ID NO:7 和序列 ID NO:8) 的第一碱基。对于小鼠, 用于比对的序列是序列 ID No 18 的碱基 532-3731 (上游) 和 8164-11364 (下游), 并且对于大鼠, 是序列 ID No 19 的碱基 719-3918 (上游) 和 8495-11058 (下游)。

[0332]

上游区域				下游			
同源性序列[大鼠]		同源性序列[小鼠]		同源性序列[大鼠]		同源性序列[小鼠]	
>80 %	>90 %	>80 %	>90 %	>80 %	>90 %	>80 %	>90 %
161-249	279-331	15-69	278-329	1608-1764	1706-1764	1614-1671	1904-2061
256-338	554-623	159-249	546-626	1894-2067	1912-2061	1888-2072	2927-3071
515-659		273-342				2918-3082	
2296-2349		515-647					
2381-2513		1143-1223					
2736-2818		1957-2009					
		2029-2080					
		2375-2485					
		2730-2821					

[0333] 啮齿类和人之间 DNA 同源性的比较显示 38% 的最低 DNA 保守。保守 DNA 片段在启动子区域或编码基因的区域外部的存在表示可能存在针对细胞的选择压力以维持该 DNA 序列或以仅允许某些 / 微小变化。在我们的特定情况下, GAPDH 侧翼区可能对细胞是重要的, 因为它们维持 GAPDH 基因的高表达水平。将再次选择 DNA 序列中导致表达下降的变化。

[0334] 为了评价上游和下游 GAPDH 元件对观察到的表达增加的贡献, 产生仅含有上游 GAPDH 侧翼区 (SEQ ID NO:7)、上游 GAPDH 侧翼区片段或下游 GAPDH 侧翼区 (SEQ ID NO:8)

的构建体。报道分子 IgG1 类型抗体由 IRES 构建体（轻链-IRES-重链）表达，因此避免多个质粒的共转染。图 8 中显示关于片段化 GAPDH 上游片段的细节。使用上游 GAPDH 侧翼区的以下片段：片段 1 (SEQ ID NO:9)、片段 2 (SEQ ID NO:10)、片段 3 (SEQ ID NO:11)、片段 4 (SEQ ID NO:12)、片段 8 (SEQ ID NO:13)、片段 9 (SEQ ID NO:14)、片段 11 (SEQ ID NO:15)、片段 17 (SEQ ID NO:16)。

[0335] 使用的上游 GAPDH 侧翼区 (SEQ ID NO:7) 含有 NruI 限制性位点的 2x3 (总计 6 个) 核苷酸，其中三个核苷酸与基因组 DNA 在其 5' 处连接并且三个核苷酸与基因组 DNA 在其 3' 处连接。使用的下游 GAPDH 侧翼区 (SEQ ID NO:8) 含有 ScaI 限制性位点的 2x3 (总计 6 个) 核苷酸，其中三个核苷酸与基因组 DNA 在其 5' 处连接并且三个核苷酸与基因组 DNA 在其 3' 处连接。在 SEQ ID NO:20 (无限制性位点的上游 GAPDH 侧翼区) 和 SEQ ID NO:21 (无限制性位点的下游 GAPDH 侧翼区) 中显示无相应限制性位点核苷酸的上游 GAPDH 侧翼区和下游 GAPDH 侧翼区。使用的上游 GAPDH 侧翼区片段各自在其 5' 和 / 或其 3' 末端处含有相应限制性位点的与基因组 DNA 连接的 3 个核苷酸 (片段 1 在其 5' 末端处含有 NruI 限制性位点的 3 个核苷酸；片段 2 在其 3' 末端处含有 NruI 限制性位点的 3 个核苷酸；片段 3 在其 5' 末端处含有 NruI 限制性位点的 3 个核苷酸；片段 4 在其 3' 末端处含有 NruI 限制性位点的 3 个核苷酸；片段 8 在其 3' 末端处含有 NruI 限制性位点的 3 个核苷酸；片段 9 在其 5' 末端处含有 NruI 限制性位点的 3 个核苷酸并在其 3' 末端处含有 NruI 限制性位点的 3 个核苷酸；片段 11 在其 3' 末端处含有 NruI 限制性位点的 3 个核苷酸)。片段 17 不含限制性位点的核苷酸。不含相应限制性位点核苷酸的上游 GAPDH 侧翼区片段在 SEQ ID NO:22 (不含限制性位点的片段 1)、SEQ ID NO:23 (不含限制性位点的片段 2)、SEQ ID NO:24 (不含限制性位点的片段 3)、SEQ ID NO:25 (不含限制性位点的片段 4)、SEQ ID NO:26 (不含限制性位点的片段 8)、SEQ ID NO:27 (不含限制性位点的片段 9)、SEQ ID NO:28 (不含限制性位点的片段 11) 中显示。

[0336] 在转染后第 10 日，使用 Octet (Fortebio, Menlo, CA, 美国) 评估上游和下游 GAPDH 元件对表达的影响以便对上清液中分泌型 IgG1 的量定量 (见图 9)。与 pGLEX41-amp^rA 主链中使用的改进的新载体设计相比，原始载体 pGLEX41 产生较低表达结果 (80%)。与原始 pGLEX41 主链相比，这种新设计包括大肠杆菌中氨苄青霉素抗性所需的 amp^rA 基因的密码子优化、不同的复制起点 (R6K 替代 pUC 复制起点) 和消除细菌源的不必要接头序列 (或间隔序列)。两种载体具有大致相同的尺寸。

[0337] 令人惊讶地，与不含上游和下游序列的相同载体相比，包括上游元件 (SEQ ID NO:7) 和下游元件 (SEQ ID NO:8) 的 pGLEX41-amp^rA (在显示表达结果的图 9 中命名为 pGLEX41-up/down) 产生更高的表达 (1.5 倍数)。如果考虑尺寸的差异 (上 / 下游片段增加质粒的尺寸大约 6000bp) 和因此转染期间所输送质粒拷贝的差异，基于每个质粒，这种影响甚至可能是更重要的。

[0338] 仅含有上游片段 (up) 的载体显示与原始表达构建体 pGLEX41-amp^rA 相似的表达水平。与原始表达构建体 pGLEX41-amp^rA 相比，仅含有下游片段 (down) 的载体显示表达显著增加 (1.2 倍)。如果上游片段和下游片段均存在，则可以观察到表达的进一步增加。这通过上游片段的片段化来证实。与 pGLEX41-amp^rA 相比，片段 9 和启动子近端片段 8 不显示任何表达差异。片段 1、11 和 17 显示表达的增加。对片段 4 观察到最高增加。应

当强调的是,启动子近端片段 8 不显示任何影响。因此,表达的增加不能由先前公开的序列解释 (Alexander-Bridges 等人, (1992) *Advan Enzyme Regul*, 32:149-159), Graven 等人, (1999) *Biochimica et Biophysics Acta* 147:203-218)。

[0339] 有趣地,片段 2 和片段 3 导致表达显著减少。这是出乎意料的,特别地鉴于以下事实:以相反方向克隆的这些片段(图 9 中的反义(AS))不造成这种影响。就片段 1、8、9、11 和 17 而言,对以有义方向或反义方向整合的片段未观察到表达差异(数据未显示)。片段 11,虽然是片段 2 的部分,但不显示这种影响。因此,看起来有损表达的序列元件应当至少部分在为了获得片段 11 而在片段 2 中缺失的 BstBI-BstBI 片段上。

[0340] 此外,负向元件(至少部分地)位于 BstBI-BstBI 片段上的假设由片段 3(其包含 BstBI-BstBI 片段)和片段 1 之间所观察到的表达增加支持。

[0341] 尽管看起来容易定位具有负向影响的片段(BstBI-BstBI),从这项研究看,不太清楚对片段 2 和片段 3 观察到的这种负向影响怎样被完整上游片段中存在的序列元件补偿。可能的是,这种负向影响由针对片段 1 和片段 4 观察到的微小正向影响平衡(但是片段 1 的表达增加小于片段 4)。然而,与负向影响(0.4 倍)相比,观察到的片段 4 的正向影响(1.25 倍)看起来较不重要。另外,与完整 GAPDH 上游侧翼区相比,片段 9,作为不含 BstBI-BstBI 片段的完整上游区域(但是,片段 9 包含了作为片段 2 和片段 3 的部分并可能对表达具有负向影响的 EcoRV-BstBI 片段),不显示增加的表达。

[0342] 仅可能推测关于已观察到的影响背后的机制。随片段 2 和片段 3 所观察到的对表达的负向影响的方向依赖性排除了未鉴定的可读框表达(例如 ncRNA 表达),因为不存在可能触发仅一个方向表达的周围的启动子。表达降至基础水平以下的事实显示不仅不存在正向影响(例如增强子活性),反而存在方向依赖性负向影响。

[0343] 总之,如果两种侧翼区即上游区域和下游区域均存在于表达质粒中,则观察到 CHO 细胞中令人惊讶的瞬时表达增加。虽然片段 4 看起来对表达具有显著的正向影响,但是不能鉴定到负责所观察到的全部表达增加的单一片段。表达载体 pGLEX41-amp^rA(up/down)的表达增加看起来是两种区域即上游侧翼区和下游侧翼区的汇总影响。

[0344] 实施例 5:克隆中国仓鼠 GAPDH 基因上游的非翻译性基因组 DNA 序列及中国仓鼠启动子。

[0345] 1.1 克隆中国仓鼠 GAPDH 基因上游的非翻译性基因组 DNA 序列至表达载体中

[0346] 通过 PCR 从 CHO-S 细胞(Life Technologies)的基因组 DNA 扩增中国仓鼠 GAPDH 基因上游的非翻译性基因组 DNA 序列。如实施例 1 中所述提取基因组 DNA。使用小鼠 CMV 启动子或中国仓鼠 GAPDH 启动子,制备用于表达实施例 1 中描述的报道基因构建体[REP]的构建体。

[0347] 为了克隆与小鼠 CMV 启动子组合的中国仓鼠 GAPDH 基因上游基因组 DNA 序列,使用实施例 1 中描述的 PCR 方案,将引物 GlnPr1896 和 GlnPr1897 用于扩增 3kb 片段(SEQ ID No 29 的 bp672 至 bp 3671)并产生具有 SEQ ID No 30 的扩增子。该扩增子含有中国仓鼠 GAPDH 基因上游的基因组 DNA 序列和通过引物引入的 5' 和 3' 限制性位点。

[0348] 为了克隆与中国仓鼠 GAPDH 启动子组合的中国仓鼠 GAPDH 基因上游基因组 DNA 序列,使用引物 GlnPr1902 和 GlnPr1905 以扩增含有基因组 DNA 序列的 3508bp 片段,所述基因组 DNA 序列包含中国仓鼠 GAPDH 基因上游的基因组 DNA 序列和 GAPDH 启动子(SEQ ID No 29

的 bp 672 至 bp 4179), 产生具有 SEQ ID No 31 的扩增子。在第二 PCR 中, 将 GlnPr1901 和 GlnPr1902 用于扩增仅含有启动子区的 508bp 片段 (SEQ ID No 29 的 bp3672 至 bp 4179), 产生 SEQ ID No 32。使用引物 GlnPr1903 和 GlnPr1904 扩增 (实施例 1 中描述的) 载体 “A” 中所用的内含子。

[0349] 使用具有 SEQ ID NO:32 的扩增子和具有内含子序列的扩增子作为模板, 用引物 GlnPr1904 和 GlnPr1901 进行第一融合 PCR。该扩增子含有中国仓鼠 GAPDH 启动子、内含子和通过引物引入的 5' 和 3' 限制性位点。表 5 中显示全部引物。

[0350] 使用具有 SEQ ID No. 31 的扩增子和具有内含子序列的扩增子作为模板, 用引物 GlnPr1905 和 GlnPr1904 进行第二融合 PCR。该扩增子含有中国仓鼠 GAPDH 基因上游的基因组 DNA 序列、中国仓鼠 GAPDH 启动子、内含子和通过引物引入的 5' 和 3' 限制性位点。

[0351] 在 1% 琼脂糖凝胶上纯化后, 将目的条带切下并使用试剂盒 “NucleoSpin Gel and PCR Clean-up” (Macherey Nagel, Oensingen, 瑞士) 纯化。使用 Zero Blunt PCR 克隆试剂盒 (Invitrogen, Carlsbad, CA, 美国), 将纯化的片段克隆至质粒 pCR_Blunt 中。将连接产物转化入感受态大肠杆菌 TOP10 (One **Shot**[®] TOP10 感受态大肠杆菌; Invitrogen, Carlsbad, CA, 美国) 中并通过限制性分析小量制备物进行分析。这产生了含有中国仓鼠 GAPDH 基因上游基因组 DNA 序列的质粒 pCR_blunt[CHO- 上游 GAPDH]、含有中国仓鼠 GAPDH 基因上游基因组 DNA 序列和 GAPDH 启动子及来自载体 “A” 的内含子的 pCR_Blunt[CHO- 上游 GAPDH_GAPDH 启动子], 以及含有 GAPDH 启动子和来自载体 “A” 的内含子的 pCR_Blunt[CHO-GAPDH 启动子]。

[0352] 为了评价扩增子的对分泌型基因表达的影响, 使用 (在实施例 1 中描述的) 载体 “A”。如先前所描述, 这个载体中使用的表达盒含有编码分泌型 IgG1 和 GFP 的多顺反子基因 (见实施例 1)。转染的细胞将因此分泌 IgG1 单克隆抗体并以依赖性方式积累胞内 GFP。

[0353] 为了释放含有中国仓鼠 GAPDH 基因上游基因组 DNA 序列的 3kb 插入片段, 使用限制性酶 NaeI 消化质粒 pCR_Blunt[CHO- 上游 GAPDH]。将这个插入物克隆在 “A” 的主链中, 所述主链使用限制性酶 NruI 消化并进行 CIP 处理 (CIP; NEB, Ipswich, MA, 美国)。使用 T4DNA 连接酶 (T4DNA 连接酶, NEB, Ipswich, MA, 美国), 将主链和插入物连接在一起并随后转化至感受态大肠杆菌 PIR1 中。挑取克隆用于小量制备物制备和后续限制性分析。将所产生的质粒称作 “A_GAPDH_UP”, 通过测序分析证实并且使用 NucleoBond Xtra Midi 试剂盒 (Macherey Nagel, Oensingen, 瑞士) 以中等量制备规模产生。

[0354] 为了使用中国仓鼠 GAPDH 启动子克隆表达构建体, 通过使用限制性酶 NheI 和 NruI 消化, 从质粒 pCR_Blunt[CHO- 上游 GAPDH_GAPDH 启动子] 和 pCR_Blunt[CHO-GAPDH 启动子] 释放插入片段。将所产生的片段克隆在载体 “A” 的主链中, 使用相同的酶打开并进行 CIP 处理。在用 T4DNA 连接酶连接并转化入感受态大肠杆菌 PIR1 后, 挑取克隆用于小量制备物限制性分析。将所产生的质粒称作 “A_GAPDH_UP_Prom” (具有中国仓鼠 GAPDH 上游非翻译性基因组 DNA 序列和所述启动子的质粒) 和 “A_PR” (仅具有所述启动子的质粒), 通过测序分析证实并且使用试剂盒 NucleoBond Xtra Midi (Macherey Nagel, Oensingen, 瑞士) 以中等量制备规模产生。

[0355] 2. 评估中国仓鼠 GAPDH 基因上游非翻译性基因组 DNA 序列对报道基因构建体表达的影响

[0356] 在旋转管生物反应器中使用 10ml 培养基体积转染 CHO-S 细胞（如实施例 2 中所述）。将转染的细胞在摇床中于 37℃、5% CO₂ 和 80% 湿度温育，并以 200rpm 搅拌。使用带蛋白 A 生物传感器（FortéBio, Menlo Park, CA, 美国）的 Octet QK 系统，对细胞的上清液分析 IgG1 表达。图 10 中显示结果。

[0357] 与小鼠 CMV 启动子 (A) 相比，含有 GAPDH 启动子的质粒的表达水平（“A_PR”）减少 50%，这表明中国仓鼠 GAPDH 启动子不如病毒启动子强。与仅具有 GAPDH 启动子的构建体（“A_PR”）相比，含有与中国仓鼠 GAPDH 启动子组合的中国仓鼠 GAPDH 基因上游非翻译性基因组 DNA 序列的质粒（“A_GAPDH_UP_Prom”）显示表达增加 2 倍。含有中国仓鼠 GAPDH 基因上游非翻译性基因组 DNA 序列和小鼠 CMV 启动的质粒（“A_GAPDH_UP”）显示最高表达和超过仅含有小鼠 CMV 启动子的质粒（“A”）多于 40% 的增加。这证实中国仓鼠 GAPDH 基因上游的非翻译性基因组 DNA 序列对报道蛋白表达具有增强子效应。

[0358] 表 5：实施例 5 中克隆所用的引物

[0359]

引物	SEQ ID No	序列	方向	限制性位点
GlnPr 1896	SEQ ID No 33	TACGGCCGGCTTCACTGTACAGTGGCACAT	正向	NaeI
GlnPr 1897	SEQ ID No 34	TCAGGCCGGCCGTGGTTCTTCGGTAGTGAC	反向	NaeI
GlnPr 1901	SEQ ID No 35	TACTCGCGAAGAAGATCCTCAACTTTCCACAGCC	正向	NruI
GlnPr 1902	SEQ ID No 36	GTTCATAAACGAGCTCTGCTATTTATAGGAAGTGGGGTG	反向	/
GlnPr 1903	SEQ ID No 37	CACCCCAGTTCCTATAAATAGCAGAGCTCGTTTAGTGAAC	正向	/
GlnPr 1904	SEQ ID No 38	CGCTAGCACCGGTCGATCGA	反向	NheI
GlnPr 1905	SEQ ID No 39	TACTCGCGATTCACTGTACAGTGGCACATAC	正向	NruI

[0001]

序列表

<110> 格兰马克药品股份有限公司(Glenmark Pharmaceuticals SA)
<120> 表达盒
<130> GBR2001/PC1
<150> US 61/567,675
<151> 2011-12-07
<160> 39
<170> PatentIn 版本3.3
<210> 1
<211> 39
<212> DNA
<213> 人工
<220>
<223> 引物 1
<400> 1
attatttcggc atgctctctg gcactctctg gacgagggc 39

<210> 2
<211> 39
<212> DNA
<213> 人工
<220>
<223> 引物 2
<400> 2
atcgtcgcga agcttagat tgcctcaagca ggtagccag 39

<210> 3
<211> 36
<212> DNA
<213> 人工
<220>
<223> 引物 3
<400> 3
agcagagact tatagacat caatamiggc tgcctg 36

<210> 4
<211> 39
<212> DNA
<213> 人工
<220>
<223> 引物 4
<400> 4
tgctagact aagctggcac caatactca gagaacaag 39

<210> 5
<211> 3179
<212> DNA
<213> 人工
<220>
<223> 5' GAPDH 片段
<400> 5
atctcgcga agcttagat tgcctcaagca ggtagccag gacgagggc agcagaaga 60
ccatccactg gtagcgaagc tagcttagc ggccagagac agactgggc ctccctctct 120
gagtagatt gttctctct cttttttaa taagtcttc atgcttttc taggtctag 180
gtaccagct ctgctctca cagctcaga caatgactt gtcacacag agcccgccg 240
tactacctt cagctctca acaccagca ggcagcttc aaaaagaaa caaagtgt 300
cttctcaagt gatgctca gtaggaagc tatgatgct cgcctgttc ccggcgaga 360
agccacacag ctgggtgca gaggctgtt ttcctagga ctgctggg ggctgtgt 420
gtagtctc tcccccacg ccagagaagc gctctcctg tacagcttg atttatct 480
ttctgtcag tctggatg tctctctt tctctgctt ctatttttc accatcttg 540
tgactcagct tttcttatt cttttatc ttgcataga tcttcagca gagatgacg 600
aagacagac acccagaaa acactctca ttctcagag atggctctg agccacgat 660
cctaggaagt ctgtctctg ctctcctgt caggtatcc tgcagggtg ctggaattc 720
gaattctgt tccctgtat aatatttgt tctctctt ttttaaaaa aaaaaggcc 780
ggcactgtg gtcacgct gtaaccag cactttgca taccagcgc ggtgatac 840
ctgagtagg gattcgaga ccagctgac caatgtag aacccacc tatctaaaa 900
ataaaaaat agccagcgt attggctgc gctgtatc ccagctacc aagagctga 960
ggcagagaa tctctgaac ccagagcgc agcttgtgt gacccgaat cacacattg 1020
cctccagct tggcaaca tagcaacct cactctaaa ttaaaaaaa aatgctaca 1080
cgtcttttaa aatcaagga tttctctaa atagctaaa ctgaatgag ttgagctgt 1140

[0002]

icaactilgg aalatagtt igccaatcic ctgthtict aagaataaa tgtttttata	1200
tacttttaga catttttttc taagcttgic ttgttticat ctttacatt agccagttt	1260
catgcagcag agagaggtt atcagtcag agagagaga gtagccag agtctaggg	1320
ctgtcccg gaigcagat gagcttcg cccgtcact gcaccttc cctctcaac	1380
cctagacc tgcacatga ccagacgcc tctcggga gaattatga ggtctagc	1440
tccagatcag tgcctctgaa cgggggcaa ttgtctgc cagagatat ctgacaacac	1500
ctgggctcg tttgtgtc atagctata gggagaat gttaccagca ttgtggga	1560
gagccagg atgtgctca acatctgca gtgcacaga tggccctca acaagaatc	1620
acacgccca caatgcaat agctcagc ttgagaaac ctgtctaga ccaagggtt	1680
cttctgccc tgtctcac ccccccac ctgtgttc ctatccat ctccaaggt	1740
tgtcagaga cggcccggt ctgtggaag ttcaatcat gttccctcc agcttgcag	1800
gagacaagac ctgtctcca gcatctca ttgtcccg gtctcagag ggcgtggt	1860
atgtcagg cggctgcc cttgaagct ggcacct ctccagctc tcaagttc	1920
ttgtctagt cacttcga ccggagct gtgtggt gtctgacca cactgtgc	1980
ttgtctga tgcacaga acctacgt agtctgtc ctgtgact ggtctctc	2040
tttgggga gaagacga atgaanaa caagatga ggcagagg aagatcac	2100
ttgggtgt caggacatg gaggctct taatctct aatgagaa ggttaaggt	2160
ctgttgca ggaatagct agtcaggt gggcttct cctctccc acccaccac	2220
tttccatt aggaacttc ttgtctgt ctgtctac tgtctggt gttcatgt	2280
aanagccc acccaccat ccagtcag ccagcagg acacagct gctctggt	2340
ccagcagg aaagcaatc agaaagcag gtcaggac tgggtctc tgggtctt	2400
ttagcaag attatacca ggcagctaa actagcac cgttttat ctgaaggg	2460
agggctag tctcagct ttgtggca gaaagagg cgggacct cctccatg	2520
acctgtat ggcacagg ccccccac cctctctc cttaagtg ttacacac	2580
tccatcca cactacaa cctgcccic tgcctcta ccagaagt ggtccctg	2640
tggagggg caggagctt gttccacc ttgtccaa acctcttc ccactttc	2700
cctctctt actaccct cctccatc cctccagg agccatga aggagctc	2760
tgtctctg ctgctgca cgtccagg cgggagga ctccagct cagctccc	2820
ttctcccc aggttgatg gaagaaag caactgtc ctctctag ccagcagc	2880
ccaggttt ccaggatgc cttgtggg ggcctctg cccccacc cctctctc	2940
ctcagctg ggcctccac cggagagag cgtgtgca caccagcc ggtgtgtg	3000
cctaatat caggccagg ctacaggct gaggacac gtgacctc gtgcagaa	3060
ctccctcc cctcaagc gctccagc cctctctc ctccagcc ccagtcac	3120
gtccagtg ccagccag gctcagtc cagagatgc agagcacc gcaataat	3179
<210> 6	
<211> 3238	
<212> DNA	
<213> 人工	
<220>	
<223> 3'-GAPDH 片段	
<400> 6	
tggcagtaet aagctggac cactactca gagaacaagg cttttctc tctgtctc	60
agtcaggc tatctgtgt ttgcacaa ttgaagaag tttctgtg gcagccag	120
ggagctgac aggtgagga agtcaggct cgcactggc ttgacgtg actgttgt	180
ggagctcag ctgagctga gctcagcg gaaattcag ctggacct gcagctga	240
ggtctgagc agtctctc ttgtctct tgcctctg tttatctg gaacatgt	300
gcagccct aaggtctca agcaggatc atctagtaa accaagtc taaacctg	360
ccnaggcg taaggactat ataatgtta aaatcgga aaatgcca cctgcctg	420
tttgaggaa gatgaactga gatgtctg ggtactat ttcatcat gttctagg	480
gaacttggt agggcaagg cgtgtgct ggactaggt ccagacct ggtctgca	540
ctnaggct cagtgtctt ggtcagta tccaggct caattgac ggtgtctc	600
ctagtggga caaagtaca taccctga gacgcgac gctgtaac ccagactt	660
ggagccaa ggtgtgtg taccctgag ttaggttt gagaacagc tggcaaat	720
gtgaacatc cgtctctc aatatcaaa aatcagca ggtctatg cactgcta	780
tgtccagc tacaggcat ctgaagcag aatctgtt gacccagg ggcagagct	840
gcgtgagct gacccacac catgcact cagctagg aacagatg gactccat	900

[0003]

ctcaaaaaa	aaaaagtae	ctaccicaga	gticaaata	gtgaatatta	ggaagtgtt	960
gagacagiga	caccaaaagig	cacataaaat	actgcacgt	ttaattatta	ttaaagaatc	1020
catttgaatg	tcagctaac	acagcictct	atacggagc	atigtgaac	gcatctccc	1080
agcttctca	ggcttttcca	agatccagg	acacgttag	ctgttggtt	cagtgtatga	1140
cagacagga	ggaagacat	ctttagcga	tactaaaca	gagacctga	gagacatac	1200
acctcgccac	acatgcctg	agcttcacct	tctatgcat	tctgacgtg	ccaggagag	1260
aagagctccc	acctccctc	aaaacactg	gccatcccg	ggcactaag	ctcttllaaa	1320
gcacggcacc	tcacgaggg	agggccacg	ccacataac	tcaccttgc	aggaggacg	1380
cgtgaacacg	tggaccata	caggacaa	gtgcctcgc	cagcccaac	gcctctgac	1440
gctacaggg	acagagccc	tctccagtg	cgtagctgc	agagccaa	cttagctcc	1500
agctgcatc	tattccatc	cagtgctgg	gccagttag	accagtggg	aagacagga	1560
gtgtgtccg	gacacaggg	atggaggaaa	gtccacat	tcacttctt	gatacgtga	1620
caaggtgag	ggccgaatc	gtctggcag	catllaaag	atggggaat	agcagacac	1680
cacgtgtga	ggcaggagag	cccaactgt	ggtagaat	gcccagaat	gtagggcca	1740
agcttagctc	cagacaccc	agagccctg	agaagccag	actgaggag	aaagcctag	1800
ggagggcgc	cccgtcccc	agggaccgc	ctgtgcaga	gtgcagctg	atgtccctt	1860
ctgtcagcc	ccacctctg	ctctgttag	ctccctgtg	caggccctc	ggtagaagg	1920
ggagggcag	tctctatct	atggagctg	cagatagac	atccgcatg	gagctctag	1980
ctcgtctgg	cggcgccgc	ggctgcata	ggggacgcg	agtagaaga	ggagacttc	2040
ttagaaaaa	cacagttgt	caacttggg	gcaggagga	atccggaag	cggaggaac	2100
tctctctct	gtgtctcac	ctttggcag	ccgttagaa	gtaccgaag	ccaggccgg	2160
gcgggggat	ggcaggagag	cggcaggac	tgaactctt	ccaaacccac	cttagacgg	2220
aaatggccc	cgcctgtgtc	tgggaactc	agaggctag	gtacggcatg	atgtctcag	2280
ctgtaattc	cagcacttg	ggagccagag	gggggggt	cacagctca	gggttcaag	2340
gcaagcttg	ccactatgt	gaaccccat	ctctactaa	aatgaaaaa	ttagccaggt	2400
gtgtgacgc	gcaccttga	gcttaggca	agataatc	ttagacctg	gagccagag	2460
ttgtatgag	ccaaaaaaa	aaaaaaaga	aaatgcaga	gtacagttag	gagagaagct	2520
gttagcttc	cagacatcg	accttaggc	cttgcttta	tgctatctg	cagtatttt	2580
gtattttta	aaaattact	tcttcttgc	gggttaact	acacaggctc	aaatgcata	2640
atcatgccc	tgaatttaga	taaaaatcg	ccccacac	ctttctgtt	cttgccagt	2700
tttaactgc	cttaacccg	gggaacccc	agagctggg	gcttgggag	gtttcagccc	2760
tccgtctatg	aattggcata	gtctatccca	ctgcaagg	agagagctgt	gggtctggc	2820
cagcccccac	aggtaactc	caaaaggcag	ccccacaga	agatgtgac	cagtcatg	2880
ctgagggtct	ctaggctgt	gtttcaacct	ttctccccc	tggtccccc	tggaaagccc	2940
catgcccag	ggagggcct	accgtctcc	agactgggt	atgagccag	ggcagctc	3000
cactgcacg	tcaggccgc	gcctcatgt	gcacatctc	atgtactgt	gcagtgccc	3060
gtcatgtcg	ttcttggcg	tggccaact	cagcttggt	ggggaggca	ggccatcgg	3120
ggtaacagg	tggcgttcc	agcgtctcg	ttgctcccg	caggaggcca	acagccaa	3180
agcagtgct	ggccggggg	cccaggcgc	catlacgaa	ggctcagaag	tacttgct	3238
<210> ?						
<211> 3164						
<212> DNA						
<213> 人工						
<220>						
<223> 上游 CAP1 侧翼区						
<400> ?						
agaagctga	gattgtcca	gcaggtagc	agagagccc	atcagccaag	aaaccatca	60
ctggtacga	aggcagctg	tggggcgag	accagactg	gcccctccc	ctgcagtga	120
ttgtttctt	ctttttttt	aaatcagtt	ttctgcttt	ttctaggtc	taggtaacc	180
ctctgcttt	ctacagctc	agacatgac	ttgttcacc	cagagcccg	ccttactac	240
cgtggcttc	caaacacca	gcggcagct	ccccaaaga	aaccnaagt	tgtttctca	300
agtagagat	ccagtacga	agtagatg	ctccgcctg	ttcccgccg	agaagacaa	360
cagctagggt	gcaggggct	gtttccata	ggactgtcg	cggggcccg	agtagatg	420
ctctgccca	ctgcgcaga	agggctctc	ctgtacagt	tggatttat	ttctctctg	480
cagtgaggga	ttgtctcat	tgtctctga	tatctatct	ttaccatct	ttgtatcca	540
gctttttct	attctttta	ttcttgcct	agatcttca	gcagagatga	cagaagacga	600
gacaccaag	aaacaactc	ccattctag	agactcggt	cgcaggaca	gactctagga	660

[0004]

```

agtcgttcc tgcctccct gtcagggta tccgtaggc tgacctggaa ttcaaatct 720
gtttccctg taataattt gtcgtctct ttttttana aaaaaaaag gcggggcct 780
gtggctcag cctgtaatc cagcacctg cgtaccacg ggggtggat aacctgaggt 840
aggcggtgc agccagcct gaccacatg gagaacccc aicctacta aaaaataaaa 900
attagccggc cgtatggcg tggccctgta atccagcta ctcaaggcg taggcaggga 960
gaatgcctg aaccagagg cggaggtgt agtagccga aatcacaca ttgcctcca 1020
gtttggcna caatagcga cctccatct aattaaaaa aaaaatgct acacgtctt 1080
taaatgcna ggccttctt taattagcc taactgaact ggttgaagt gcttcaact 1140
tggatatai gttgcaat ctcctgttt tciaatgat aatgttttt atatacttt 1200
agacatttt tcttaagctt gttttgttt cctcttccc attagcccg tttcatgag 1260
cagagagagg gtatcacg caagagaga tagtgagcc cagagctta ggcctgtcc 1320
cggatagca gatgagct cgtcccgct acgtcaact ttccctctt aacctgiga 1380
cccgcacag tgacagaca gctctcgg ggagattat gtagtcta ggtccagat 1440
cagtgctct gaaccggcg caattttgt tgcagagga cactgcaa cactggggc 1500
ctgttttgt gctatagct atagggagc aatgaccca gctttgtg gaagagcca 1560
gggtgtggc tcaatctct gcatgaca ggaaggccc taacacaga atcacagcg 1620
ccacatgtc atagctga cagtgagaa aacctgctt agaccaagg ttgctttct 1680
ccgttgctt ccccccac cactctgtt tcccaatc cctctcaaa gttggcagc 1740
agacggcct aggtctgtg aagtcagat cagatcccc ttagctctg taggagaca 1800
gaactgtct ctagcttcc tcaatttcc cgggtctga gaggcgtga gctatgctc 1860
aggcggctg cccccgaag cctgcgacc cctctcagc tctcaagtc ttctctgtc 1920
agtaccttc gaaccggcg ctgtgagct gctgtctga ccacatggt gctctgtc 1980
tcatgacac agcacctac gtagtagt cctccgggc actgagctc cctcttccg 2040
ggagagaca gtaataaaa atgacagca taggcagcg ggaagatca cgttggtg 2100
gtgcaggagc atggaggtc tctaatgct ctcaatgaa aaggttaac ggtctgtt 2160
gcaggaatg ctgagtcga gttgggctt cctccatcc ccccccac cctttcac 2220
attaggacc ttctgctt gctttgcta ctctgtct gttgttatt gtaaaagcc 2280
ccacacacc atgcagtg cagccagag agaacagc ctgctctg gtccagag 2340
gaaagcaat cccagaagg caggctcag gactggag ctgtgggtc ttittaaag 2400
aagattata ccaggcagc taacttagc aaccgctt tagctaga ggcaggggc 2460
tggctcagc ttatgtgg ccagcaaga ggcggggat ccccttcca tgcactgct 2520
gatggcnaa ggccaccca ccccccct tcttcaaa gttttaga cctctcctc 2580
ccacatcac aaactggc cctgtccct ctaccaga aatggtccc ctgtggagg 2640
ggcaggaga cctgttcca cctgtgccc aagacctt ttcccttt ttcccttc 2700
ttgctcacc ctgcttca tatccctcg ctagctagt gaaaggagt tctgtctc 2760
tggctgctt gacgttcca gggggggag ggaattccc cctcagctc cgtcttgc 2820
cccagctgg atggaatga agccactg tctctccc taggcagc agccacagg 2880
tttcaggag tgccttgt gaggctctt gggcctccc cagcatctt gctctcgc 2940
tggggccca gccggagag agcctgtgt gacacagg cgggattgt ctgcttcat 3000
tatcagctc agctacag gctgcagg atctgact tctgtcga aacctcccc 3060
tcccccca gccctccc gacctctt cctctcag ccccagtg ccagtgcga 3120
gtgccagcc caggctcgg tccagagat gccaggacc atcg 3164

```

<210> 8
 <211> 3224
 <212> DNA
 <213> 人工
 <220>
 <223> 下游 GAPDH 侧翼区

```

<400> 8
actaagctgg caccactact taagagaaca aggcctttt cctctctgc tccagtccta 60
ggctatctgc ttttgccaa acatggaaga agctatctg tggcagccc cagggaagct 120
gacagtgga ggaagtcag cctgcactg gctctgagc ctgactggt agtgagctc 180
agcctggagc tagctgag cgggaattc cagcttgcc tccagctg taggtcttg 240
agcagtgct ctattgttt ctgtgcttc gttcttalc taggacatc gtggcagcc 300
ctlaaggctc taagcagga tcaatctagg taaccaagt acciaaacc aigcccaagg 360
cggtaaggac tatataagt ttaaaatag gtaaaatgc ccaacttga tagtttagg 420

```

[0005]

```

gaagatgaac tgagatgigt caggigact taatccacc aacgacctha ggggaacttg 480
ggtaggagca aggcgigtg ciggagacta ggtccagac ectggctctg ccaatgaacg 540
gtcagctgc ttgggcagt taatccagg ccacacttg cagctgigt tabctagtgg 600
agacaaagt acataacctg gttagagcgg caagccctga acccagacaa ttgggaggg 660
caagtggtgt gtatccctg aggtcaggag tttagacta gctggccaa catgtgaaa 720
ctcgtctct actaaatla caaatacag ctaggctta tggcactgt ctatagtcc 780
agtiacaggc atgtgaagc agagaaatg ctgtacccc ggaagcagag gtgtcagtga 840
gttgagaca caccactga cttcagctla ggcacagag tatgagact catctaaaa 900
aaaaaaaaa taatcactc aggttcaaa ctatgaata ttaggagtg cttagagag 960
tgacacaaa gtgcacata aatctctgct agttctata ttatanaaa atccattga 1020
atgtcagctt aacacagct ctatccaga ggcatttga accgctctc cccagctct 1080
ccagctttt ccagaaatc gggacactg agctgttgg tctcagtga tgacagac 1140
ggaggaagca catctttagc tgatactta acagagacc tgagagaca taacccggg 1200
cacacatga ttgagctta cttctctgt catctgagc tgaccaggag agcaagact 1260
cccactccc ttcaaaacg tgtccactc ccggcacta agctctttt aaagcaggg 1320
actccagga gggaggcca cagcacata caatcaact ggcaggtaga cagctgagc 1380
acttgacca tagcaggac aaggtgcctc ggcagccccc aagccctct ggcctgaca 1440
ggacagagc cctctccag ctgctgtgc tgcagagctt atgtctagc tccagctga 1500
tttatctca cttcagctc tggctcagt agtaccagc ttgaagacg tgaatggct 1560
ccgacacaa ggaatgggg aaaggtccca catcacatt ctgatactt ggaacagtg 1620
agggcgcca atcgtcttg cagctttta aagaatggga agtagcagc acccagcgt 1680
gaagcagga gagctccac tgtgttgga atggccccc aatgtaggc caagcctag 1740
ctcagacac cccagagcc ttgagagcc aagctgagg gagaagcct gaggagagc 1800
cgctcagtc cccagagcc ggcctgtgc agctgtgag ctatgttc cctctgtca 1860
gcccacctt ctgctctgt gagctccct ctgcagagc ctgctgtga agggagagc 1920
agctctctt ctataggc tgtcagatg aatctcga tgggtctct cagctctgt 1980
ttggggggc ggcggctgc tggcgggac cgtctgaa ggaagagct tttagaaaa 2040
aaacaccagt tgtacactt gggcagcca ggaacttga agacggagc catctctct 2100
ctgtgtctt cacccttgc cagctctga gaagtcagg aagcagagc gggcgaggc 2160
gtgtcgagg ggcggcagg aactgaact tctccaaac caccctgac gggaaatgg 2220
cccagctgt gtcttgga ctacagagc gagttaggc atgagctc atgctgtaa 2280
ttcagcact ttggagcca gagtggtg gactcagc ttaggttgc aagcaagcc 2340
tgccacatc ggtgaaccc catctctat aaaaatgca aaattagca ggtgtgtg 2400
cgggcctct ggaagctgag gcaagataa caatgaacc ttggagagg aggtgtagt 2460
gagcaaaaa aaaaaaaa agaatagct gaagtcacg taggagaga gtgtgtgag 2520
ctcagccc ctgactctg ggccttgct ttatgtct atgcagtat ttgtgatt 2580
ttaaaattc acttcttgt tgcgttga ctacacag gtaaatga caatcttg 2640
ccatgactt agataaaat ctgcccac accctctg tcccttga gtlttaac 2700
tgctctaac cagggaacc accagagtg gtgctctg gagttctag cctccctg 2760
atgaatgac atagctcat caactgcaa gggagagag tgtgggtct ggcagcccc 2820
accagctac tccaaaggc cagccctca gcaagatg accagctat tgcctgagg 2880
tctctgggc tgtgttcaa ctttctccc cgtgtgtc ccttggaag cccatgccc 2940
agggagggc cttaccgtc tccagctgg gtatgagc ggcggaggt ctccatctg 3000
acgtccagc cgccttcat gctcacatc tcaatgact cgtcaggtg ccgttcatg 3060
ctgtcttgg ccgtgcca cttcagctt ggtgggag gcaagctat cgggttacc 3120
aggtggctt cacagcctt ctgttccc cgcaggagg caaacagc aagagagtg 3180
gtggggcgg ggcctcagg agcattact gaagctag aagt 3224

```

<210> 9
 <211> 511
 <212> DNA
 <213> 人工

<220>
 <223> 片段 1

```

<400> 9
cgaagctga gattgtcaa gcaagtagc agagagccc atcagccaag aaacatcaa 50
ctgtactga agcagcttg tggggcagg accagactgg gcccctctt cctcagtgta 120
ttgtttctt ctctttttt aaatcagtt ttcttgcct ttctaggttc taggtaccag 180

```

[0006]

```

ceictggcctt ctacagccic agacaatgac ttgttcacac cagagcccc cgttactacc 240
cgtcggcatic caaacaccca gcagcgagct tccaaaaaga aacccaagt tgtcttcica 300
agtatgagt ccagtggga aggtatgat ctcacgcttg tccccgctg agaaggcaca 360
cagctagggt gcagaggct ggcttcata gaactcttg cggggcctg agtgtatg 420
ctctgcccc ctg-cgcaga agggcctct cgttacagct tggatttat tctctctg 480
cgttgggga ttgtctact tgtctctga t 511

<210> 10
<211> 2653
<212> DNA
<213> 人工

<220>
<223> 片段 2

<400> 10
atctatitit taccatctt tgtgactcag ctititctia ttctttaat tcttgcata 60
gatctttcag cagatgac agaagacgag acaccaaga aaacaactcc cattctcaga 120
gcctggctc gcagccacag atctcaggaa gtctgttctt gctctcctg tgcaggtat 180
ccttagggt gactggaa tcaattctg ttctctgtt aaataattg tctgtctctt 240
tttttaaaa aaaaaaagg cgggcctg tggctcagc ctgtaatcc agcatttgc 300
gataccaagg cgggtgata accctaggta gggagtctga gaccagctg accaactgg 360
agaaaccca tctctactaa aattaaaaa ttacccggc gtattggct gcgctctaa 420
tccagctac tcaagagct gaggcaggag aatgcctga acccagagc ggaggttga 480
gtgagcgaa atcaacctt tgcctccag ctgggcgaa aatagcgaa ctctctctc 540
aatlaaaaa aaatgctta cagctctt aaatgcag gcttctctt aaatagctt 600
aacctgctg ctctgctg ctctacttt ggaatctg ttgcaatc tcttctt 660
ctaatgata aatgttctt tatacttctt gactttttt cttaagctt tcttctt 720
atcttctca ttacccagt tctatgcag agagagagg ttatcagtc agagagat 780
gagtggccc aggtctctg ggcctctcc gggatggag atgacttcc tgcctctca 840
ctgccactt tccctctca accctggac ctgctagt gaccagacg cctctctgg 900
gagaattatg cgtgctctg gctcagac agtctctg aaccggagg aatttctt 960
gccagaggc atctgaca accctgggc tgttctgt tcatagcta tagggaga 1020
agtctaccg cattgttgg agaggctag ggaigtgtt caacatctg cgtgctcag 1080
gaiggccct caacaagaa tccacggct cacaatgta atagcttca agttgagaa 1140
acctgctct gacaaaggc tgcctctg cgtctctc acccaacct cactgtgtt 1200
ccttaatcc atctcaag gtggcagca gacggccca ggtctgtga agttcagat 1260
agtatcccc cagctctg agagacag accgtctc cagcttctt cattgtctc 1320
ggctctcag agggctgag ctatctgca ggcggctg cctctgagg ctgctcacc 1380
ctctcagct cctcaagct tctctctga gtcacttgc aaccggagg tctgagctg 1440
ctgtctgac cactctggt cctctctgt catgacaa gcacactac tcatagctg 1500
tccctggca ctgagctcc tcttctggg gagaagacg taatgaaaa tgacaagct 1560
gaggcagag ggaagatcc gcttgggtg tgcaggaca tggaggtgt cttaatgct 1620
tcaatgaaa aggttaacg gtcctgttg caggaatag tgaatcag gttggcttc 1680
ctcactccc ccccccacc ccttctcca ttaggacct tctgtcttg ctctgtctc 1740
ctgtcttg gtgtcttg tgaagccc gacccaaca tgcagtggc agcagacga 1800
ggacacgct tggctctgg tccagcagc aaaggcaac ccagaaagg aggtctagg 1860
actggatct tgggtgtct ttttaagcaa agattatcc caggcagct aaacttgc 1920
accgtcttt agctagagg gcaggggct ggtctcagt tatgttggc cagcaagag 1980
gccgggac cctctccat gcactctg atggccaag gccacccac cccacctct 2040
tcttaccag tgttcagac cctccctcc cactctcca aacttggcc tctgctctc 2100
taccagaga atgattccc tgtggaggg gtcagggac ctgttccac cgtgtgcca 2160
agactcttt tccactttt tctctctct tgaactccc tgcctcaat atccccgc 2220
gagcragtg aaaggagtc cctgctctt ggtctcgtt cagtctcag gtcgggagg 2280
gacttctgc ctactctc gctctctc caggctgga tggaaatgaa ggcacactg 2340
ctctctctt aggcagaca gcccaggt ttcaggagt gcttctgtg gaggctctg 2400
ggccctccc agcatctg tctctgctt gggccctag cccgagaga gccgtgtg 2460
cacacaggg caggatgct tgcctaat atcagctca ggtacaggg ctgagaca 2520
ctgtgacct cctgagaa accctccc cccctcaag cgcctccc agctctctc 2580

```

[0007]

ctctccaggc cccacagtcg cagtgcccg tgcccagcc aggcctcggt cccagagatg	2640
ccaggagcca tag	2653
<210> II	
<211> 1966	
<212> DNA	
<213> 人工	
<220>	
<223> 片段 3	
<400> II	
cgagcttga gattgtcca gcaggtagcc agagagccc atcagctaac aaacatcca	60
ctggtacgta aggcagcttg tggggcgag accagactgg gcccacct cctgcagtga	120
ttgtttctt ctcttttt aaatcaggt ttccgcctt ttctaggct taggtaccag	180
cctcagctt ctacagctc agacaatgac ttgtccac cagagcccg cgtactacc	240
cgtcgcctc caaacacca gcagcgagct tccaaaaga aaccctaat tgtcttcca	300
agtgatgagt ccagtagga aggtatgat ctccgcctg ttcccggcg agnaggaca	360
cagtagggt gtaggggtt ggttccata ggaccctg cggggcctg agtgatag	420
ctctcccca ctgcgcaga agggccttc ctgtacagt tggatttat ttctcttg	480
cggtaggga ttgtctact tgtctctga tctctttt ttaaccatc ttgtactca	540
gcttttctt attccttaa ttcttgcat agatcttca gcagagatg cagaagaca	600
gacaccaag aaacacac ccatctcag agctcggtt cgcaggaca gacttagga	660
agctcttc tctctctt gtcaggga tctgtaggg tgactggaa ttgaattct	720
gttcccttg taatatat tctctctt tttttttaa aaaaaaaa gggggcact	780
gtgctcag cctgaatc cagctctt cgtactcag ggggtggt aacttaggt	840
aggagttg agaccgct gacacatg gagaaccc atctctaca aaataaaa	900
attagccgg cgtatggc tgcctctga atccagct ctcaggagc tgaggagga	960
gaatgcctg aaccagag cggaggtgt agtagcga atcacaca ttgcactca	1020
gcttggcaa caatagcaa cctcatctc aaattaaaa aaaaatgct aacgtctt	1080
taaaatgca gcttctct taatagac taactgact ggttagct gcttcaat	1140
tgaatatat gtttccat ctcttggt tctaatgat aaatgttt atatacttt	1200
agacatttt tctaatgt gcttggtt cacttctc attagccag tttaicgag	1260
cagagagag gtaacagtg cagagagga tgagtagcc cagagctca gggctgtc	1320
cggatggca gatgactc ctgcctgc atgcactt ttccctctc aactctga	1380
cctgcacag tgaccagca gctctctg ggaaattat gcagctca ggtccagat	1440
cagtctct gaaccggg caatttgt tgccagga cactgacaa cactgggg	1500
ctgtttgt gctatgct atagggaag aatctaca gatttggt gaagagcca	1560
ggatgggc taacatctt gcagtcaca ggaagccc taacaaaaga atcaatggc	1620
ccacaatgc aatagctca cagtagaa aactgcct agaccaagg ttgctttg	1680
ctgtgctt caccaccc caatctgt tcttaatc cactctaaa gtttgtag	1740
agacggcc agctcttg aattcagt aatctccc tccagctg caggagaca	1800
gactgtct ccagcttc tcttgctc cggctctga gaggctga gctatgtc	1860
agcggctg cccctgaag ctgcgcac cctctcag tctcaagc ttctctgt	1920
agtcacctt gaaccgag ctgtagct gctctctga ccac	1966
<210> 12	
<211> 1198	
<212> DNA	
<213> 人工	
<220>	
<223> 片段 4	
<400> 12	
tgtgctctt gctgcatg caacagaca ctactcagt agtctctt gggcactgag	60
ctctctttt gggggagaa gacagtaat aaaaatgaca agcatgagc agagggaag	120
atcagcttg ggtgtgcg gacatggag gtctcttaa tgcctcaat gagaaggt	180
tacgtctct ggttcagg atagctagt cagagtggt gctctctca ctcccacac	240
ccacccttt caccatagg gactcttg cttgtctt gtaactctg tctgggtgt	300
catttgaaa agccgcact aactatgca gtggagctc cagagagca cagctggct	360
ctggctcca gcaggaagg caatccaga aagcagggt caggcttg agtctgtg	420
gtcttttta agcaagat atccagcc aggttaact tagcaccgg ctcttagct	480
gaagagcag ggcctgtgt cagttatg tggccagca aagagccg ggtccctt	540
ccatgcacc tctgatgg ccaaggcac ccaaccac cctctctt aaagtgtt	600

[0008]

```

agcaccctcc caccaccac tcaaaaacct ggcctcctgc cctctatcca gaagaatga 660
tccccctggg gaggggcag gggaccctgt cccacctgtt gcccagacc tcttttccca 720
ctttttccct ctctctgact caccctgccc tcaataccc ccggcgacgc cagtgaagg 780
gagtccttgg ctcttggctc gcttcacagt cccaggcgg gaggagactt cccctctac 840
gtccctctct tgc-cccagg ctggaaggaa tgaaggcac actgtctctc tccctaggca 900
gcacagccca caggtttcca ggagtgccct tggggaggg ctctggcttc ccaccagcca 960
tctgtctctc cccctggggc cccagcccg agagagcctc tggtcacac agggccggga 1020
tgtctgccc taattatcag gtccaggcta cagggttgcg ggcacatgtg acctctcgtg 1080
cagaaacctc cccctccccc tcaagccgac tcccgagcct cctctctctc caggccccc 1140
gtgcccagtg cccagtgccc agctccaggtc tgggtccag agatgccagg agccatcg 1198

<210> 13
<211> 259
<212> DNA
<213> 人工

<220>
<223> 片段 8

<400> 13
ctctgggccc cccaccagcc atctctgctt cggcttgggg cccaccagcc gaggagccg 60
ctggtgcaca caggccgggg attgtctgcc ctaattatca ggtccaggtt acagggtctc 120
aggacatgtt gaccttcgtt gacgaacctt cccctcccc ctcaagcgcg ctcccgagcc 180
tcttctctct cccagccccc agtgcccagt gcccagtgcc cagccagggc ctgggtccca 240
gagatgcacg gggccatcg 259

<210> 14
<211> 1947
<212> DNA
<213> 人工

<220>
<223> 片段 9

<400> 14
cgaagcttga gattgtccaa gcaggtagcc agagagtgcc atcagccaag aaaccatcca 60
ctgttagcta agcagacctg tggggccgag accagacttg gccctccctt cctgacatga 120
tttgtttctt ctctcttttt aaatcaggtt tctctgcttt tcttaggttc taggtaccag 180
ctctgggttt ctacagcttc agacaatgac ttgttcacac sagagccctg ccgtactacc 240
cgtcgcatc caaaccccca gcagcgagct tccaaaaaga aacccaagtt tgtcttctca 300
agatgatagt ccagtggagg aggtatgagc ctcccgcttg tcccgagccg agaaaggaca 360
cagctagagt gcagagggct agtttccata ggaactgtcg cggggacctg agttagatg 420
ctctgcccc ctgcgcaga agggcctctc ctgtacagct tggattttat ttctctgtg 480
cgggtatgga ttgtctcact tgttctctga taictatatt ttacacatct ttgtactca 540
gcttttctct attctcttaa ttctttgcat agacttttca gcagagatga cagaagacga 600
gacacccaag aaacaacctc ccatcttcag agcatcggtt cgcagccaca gatcttagga 660
agctgttccc tgcctccctt gtgcaggata tccgttaggg tgcactggaa ttcaaccagg 720
aggctgtgag ctggctgtcg tgaccacact gggtcccttg ctgtcatgac aacagcacac 780
tagctcagta gtgtcccttg ggcactgagc tccctctttg cggggagagag acagtaatga 840
aaaatgacaa gcattgagga gagggggaag tcaagcttgg gtgttgacgg agcatggagg 900
tgccttcaat gctctcaatg agaaagggtt aacggctctg gttgcaggaa tagctgagtc 960
agaggtaggg ctctctccac tccccacccc caccctcttc accatlaggg acctcttgc 1020
cttgcctctg ctactctgct ctgggtgggc attgtgaaaa gcccgacca accatgccag 1080
tggcagcag acgaggacac agcttggctc tgggtccag caggaaggc aatccagaa 1140
aggcagggtc agggactgga gtccgtggg tcttttttaa gcaagatta tcaccaggca 1200
ggctaaactt agcaaccgct ttttagctag aaggccaggg gcttgggtgc aggttatgct 1260
ggccagcaaa agaggccctg galcccccct ccaigcactt gctgatggg caaggccacc 1320
cccccacac ccttctctta caagtgatca gaccttccc atcccacact cacaacctg 1380
gacctctgac ctctaccag aagaatggat ccccttgggg agggggcagg ggaactgttc 1440
ccaccgtgtg cccaagactt ettttccac ttlttccctc ttcttgactc accttgcct 1500
caatacccc ccggcgagcc agtgaagggg agtcccttgc tcttggctcg cctgcacgtc 1560
ccaggccggg gaggacttc cgccttcacg tccgtcttt cgcctcaggg tggatggaat 1620
gaagagcaca ctgtctctct ccttaggcag cacagccac aggttttcag gattgcttt 1680
gtgggagccc tctggccccc caccagccat ctgtctctcc gcttggggcc ccagcccgga 1740

```

[0009]

gagagcgcgt ggtagacaca gggccgggat tgcctgcctt aattatcagg tccaggctaa 1800
aggcctgcag gacatcgtga ccttcctgic agaaccttc cctccctctt caagcugcgt 1860
cccagacctc cttctctccc aggcctccag tgcctcagtc ccagtgcga gccaggcctt 1920
cggtcccgaa gatgccagga gccatcg 1947

<210> 15
<211> 1436
<212> DNA
<213> 人工

<220>
<223> 片段 11

<400> 15
atctattttt tcacctctt tggactcag ctctttctta ttctttaat tctttgata 60
gatctttcag cagagatgac agagacgag acacccaaga aaacaaactc cattctcaga 120
gcctggcttc gaggcacag atcttaggaa gtcgttctt gtcctccctg tgcagggtat 180
cctgtagggt gacctggagt tcgaaccgga ggcctggagc tggctgtctg gaccacatg 240
gtgcctctgc tgcctgaca acagcacact acgtcagtag tgcctccctg gactgagct 300
ccctctttgc ggggagaga cagtaatgaa aaatgacag catgaggcag aggggaagt 360
cagctttggc tggtaggga gcatggaggt gctcttaag ctctaatga gaaagggtta 420
acgtctctgc ttgcaggagt agctgagta gagggtgggc ttcttccact ccccaacccc 480
accccttcta ccttaggga cctctctgac ttgccttgc tactctgctc tgggttgcta 540
ttgtgaaag ccgcaccaa catgccagt ggcagccaga cagggacaa gcctggctct 600
gggtcccgcc aggaaggca atccagaaa ggcaggctca gggactggag tctgtgggt 660
gctttttaag caagattat caccaggcag gctaaacta gaaacggct ttagctaga 720
agggcaggag gctgtgta gcttatgct ggcagcaaa gaggccggg atccctctc 780
catgcactg ctgattggcc aaggccccc caccacccc cctctctac aagtgttcag 840
caccctcca tccacacac acaaacctgg cctctgccc tctaacaga agaatggatc 900
cctgttgga ggggcagg gacctgtcc cactgtgtc caagacctc ttttccact 960
tttccctct tcttgacta cctgccttc aataccccc ggcagacca gtgaaaggga 1020
gtctctgct cctggctgc ctgcagctc caggggggag agggacttc gctctcagt 1080
ccgcctctc gccccaggt ggaaggatg aaaggcaac tgcctctc cctaggcagc 1140
acagccaca ggtttccag agtgccttg tggaggcct ctgagcccc accagcact 1200
ctgctctcc cctgggccc cagcccgag agatctctg ctgcacacag ggcaggatt 1260
gtctgccta attatcagt ccaggctaca ggcctgcagg acatctgagc ctctcgtga 1320
gaaacctccc cctcccttc aagtgcctc ccgagcttc ttactctca ggcctccagt 1380
gccagtgc cagtcccg ccaggctc ggtccagag atgcaggag ccatcg 1436

<210> 16
<211> 1177
<212> DNA
<213> 人工

<220>
<223> 片段 17

<400> 16
atctattttt tcacctctt tggactcag ctctttctta ttctttaat tctttgata 60
gatctttcag cagagatgac agagacgag acacccaaga aaacaaactc cattctcaga 120
gcctggcttc gaggcacag atcttaggaa gtcgttctt gtcctccctg tgcagggtat 180
cctgtagggt gacctggagt tcgaaccgga ggcctggagc tggctgtctg gaccacatg 240
gtgcctctgc tgcctgaca acagcacact acgtcagtag tgcctccctg gactgagct 300
ccctctttgc ggggagaga cagtaatgaa aaatgacag catgaggcag aggggaagt 360
cagctttggc tggtaggga gcatggaggt gctcttaag ctctaatga gaaagggtta 420
acgtctctgc ttgcaggagt agctgagta gagggtgggc ttcttccact ccccaacccc 480
accccttcta ccttaggga cctctctgac ttgccttgc tactctgctc tgggttgcta 540
ttgtgaaag ccgcaccaa catgccagt ggcagccaga cagggacaa gcctggctct 600
gggtcccgcc aggaaggca atccagaaa ggcaggctca gggactggag tctgtgggt 660
gctttttaag caagattat caccaggcag gctaaacta gaaacggct ttagctaga 720
agggcaggag gctgtgta gcttatgct ggcagcaaa gaggccggg atccctctc 780
catgcactg ctgattggcc aaggccccc caccacccc cctctctac aagtgttcag 840
caccctcca tccacacac acaaacctgg cctctgccc tctaacaga agaatggatc 900
cctgttgga ggggcagg gacctgtcc cactgtgtc caagacctc ttttccact 960
tttccctct tcttgacta cctgccttc aataccccc ggcagacca gtgaaaggga 1020

[0010]

gtccctggct cctggctgc ctgcargtcc caggcgggg agggacttc genctcagt	1080
cccgctcttc gccccaggt ggaiggaag aaaggacac tgcctcttc cclaggcag	1140
acagcccaca ggtttccagg agtgccttg tgggaag	1177
<210> 17	
<211> 18106	
<212> DNA	
<213> 人 (Homo sapiens)	
<400> 17	
gtgagcagca caggacacit caagccctgt tgggttcagg gcaggcctaa acatctgcc	60
gccttgggca gctacgggt ctggcagtc cttcccgct cttcttatcc ccagctgggt	120
tgcacccaaa ttgcagagt gacctaaag cagatcttg tctccagttc tttttttatt	180
acccccaaa cacaacaaa gcagcatct atccaatct ttttttttgc ttttttaag	240
ttttttttt ttagagcagt tttaggttca aagcaaat gacagagaag taccgggagt	300
tcccttctac ccttggccc tacacatcac agccttccc accttcaaa tcttgcacca	360
gggtggcaca ttgtttacg ctgaacctac acttaccac tatctctaa agttatgggt	420
tactttggag ttcaatgca gtaatgcat gtacccaca ttgtagtct atacagaga	480
gtttactgc ctacaaatc ccttgcctc cacttattc tctctctc cccacaccc	540
ctgatcttt tactgttgc atcactttgt ctcttcaga atgtatcat ggaatgatc	600
ggatggagc cttctacat tggcttctta gtaatgtgc ttttaaggct caatgtctc	660
catggccttg ttctttttt atcagaagta acgttttca ggcctgtct gaatctcct	720
ttctccctcc aggtataat agatgaatt gacagagaag ttggggctg tcataccaga	780
ggtttggatg gaataagga gcttgagatt ggcgaagcag gtacagag agcgcacaa	840
gtcaagaac catcactg tactaagc agcctgtgc ggcagacca gactggccc	900
tccctctct cagtatttg ttctcttct ttttttaaa cacttttcc tgcctttct	960
aggtctagg taccagctc tggcttctac agcctcagac aatgacttg tcaaccaga	1020
gtccggcgt actccctgc ggcacccaa caccagcag cagcttcca aaagaaacc	1080
caaatgtgtc ttctcaagt atgactcag tgggaaggt atgatctcc cgcctgttc	1140
cggccagaa ggcacacag taggttcag aggtctggt tccataggac ctgttcggg	1200
ggcttgagt tagatctct gccccactgc cgcagaagg cctctctgt acagcttga	1260
ttttattct tctgtcgtg gtggatgtt cttacttgt cttgataic tatttttca	1320
ccaacttgt gactcagtt ttcttttct cttaattct ttgcatgat ctttcagag	1380
aga'gacaga agacagaca ccaagaaa caactccat tctcagaca tggctcga	1440
ggcacagtc ctaggagtc tgtctctgc cttctgtgc aggtatctt gtagggtac	1500
ctggattcag aattctgtt ccttgtaaa atatttgtt gtcttttt tttaaaaaa	1560
aaaaagccc ggcactgtg ctacagctg taatccagc actttgcgt actaagcag	1620
gtgataacc tgggttggg agttcgagc cagctgacc aacatggaa aacctcatc	1680
ctactaaaa taaaaatta gccggcgta ttggcgtgc cctgtatcc cagctactca	1740
agagctgag gcaggaatc cgtctgaac cagagccga gtttgtatg agctgaatc	1800
acaccattgc actcagctt gggcaaatc agcgaacct cacttcaat taaaaaaaa	1860
atgcctcac gctctttaa atgcaagct ttctcttaa ttgactaac tgaactggt	1920
tgaactgct caacttggg ataatgttt gccaatcct ttgtttcta atgaataat	1980
gttttttat acttttagc atttttctt aagctgtct ttgttctc ttctacatta	2040
gtccagttc atgcagaga gaggagttc ttactcaga gacagatag ttgacctaga	2100
gtctagggc ctgtccggg atgcagatg agcttctgc cccgtactg ccactttcc	2160
cctctaaccc tctggacct gcaactgac cagacagct ccttgggag aattatgag	2220
tgcctaggt ccagatcagt gcttctaac tggggcctt ttgtctgct agagacatc	2280
tgcacaccc tgggctgtt ttgttgtca tactctatg gggagaaag ctactcagat	2340
ttgtggagc agccagga ttgtgtcaa cactctgag tgcacagat ggcctctca	2400
caagaaatca cagggccac aatgtcata gcttccagt tgaagaaac tgccttagc	2460
caagggtgc ttctgtcgt gtgtctcac ccaacccac ttgttctc taatccatc	2520
tcaaaaggt ggcagagac cggccagcc tcttggagt tcatatcat atcccttca	2580
gtctcagag agacagacc tgcctcagc catctctat ttgttccgg ttgcagagg	2640
gtatgagca ttctcagga ggtctgccc ctgaagctg cgcacccctc tccagctct	2700
caagtctct ctgtcagtc acttctaac tggagctgt gacttggct tctgaaccac	2760
actgtacct ttgtctcat gacacagca cactactca gtactctcc ctggcactg	2820
agctcctct ttggggggg aagacagta tgaataatg cagcatgag gcagggaga	2880

[0011]

```

agatcagctt tgggtgggtc aggcacatgg aggtgtcttt aatgtctca atgagaagg 2940
gttaagggtc ctgggttcag gaatagttga gtacagagtg ggccttcttc cacttcccc 3000
ccccacctt ttaccattt gggaccttct tgccttgctc ttgtactct gtcttggtg 3060
gtatgttga aaagcccgca ccaacctgc cagtgccgc cagacagga cacagcttg 3120
ctctgggtc cagaggaa ggcaatccc gaaggcagg gtacggact gggtctctg 3180
gggtgtttt taagcaaga ttatcaccag gtaggctaaa cttagcaac ggcttttag 3240
tagaaggcca ggggtctgt gtacgttat gttaggcag caaagagcc cggatcccc 3300
ctccatgca ctgtctgag ggccaaggcc acccaccgc acccacttc ttacaagtgt 3360
tcagaccct cccatccac actcacaaac ctggccctct gccctctac cagaagatg 3420
gatccccgt ggggggggc aggggacct ttccacgt gtcccaaga cctcttttc 3480
cacttttcc ctctcttga ctaccctgc ctcaatata cctcgggca gctagtaaa 3540
gggtctctt ctctcttgc tgcctgac gtccagggc ggggaagac ttccgctc 3600
acgtccgt ctctgccc ggttgatgg aatgaagcc acactgtct tctctagg 3660
cagcacgcc cacagttc caggatgct ttgtggag gctctggc cccaccgc 3720
ctctgtct ctccctagg gctccagcc gtagagacc ctctgtcgc acaggccg 3780
gatgtctc cttaattat aggtccagc tacagggtg caggacatg tgactctg 3840
tgcaaaacc tccccccc cctcaagcc ctccagcc ctctcttc tccaggccc 3900
cagtcacag tcccagtc cagccagc ctctcttc agatgtca gggcagga 3960
gatggagg ggaagtgg ggttggaag gaacacgg ccccgccc agcccttg 4020
gcccctct ggccttgc tggcagtc ggtctcta ctccagacc tccagagaa 4080
gtcccaac ttccgctt ctacgttt gaagaaga aaggaggg ggcagccg 4140
gtcagccc gacggctg ggtccccc tcaattcc cctctcag gtcccaag 4200
tctctgtt tcatcaag gttagggc cccgtctt gactctct tctctgtg 4260
ccacagtc agtctggg accagacc atacctcc atcgccaa tctcagtc 4320
ttcccccct cgtggggc caccagtc gtgtgtcc agtgaaca ggcgctg 4380
gaaaaaaa agcgggga agtagggc cgtctatg cgttttgc ggcacgta 4440
gtcagccct caagacctt ggttggaat gctagacc ggcagagc gggctcag 4500
tcaccgctg ccccgccc cccgttct ataatagc cccagacc cccgttgc 4560
tctctgtc tctgttga cgtcagcc caietctt tgcctcca ggtgaagc 4620
ggcagaga aaccggag gctaggaag gctgaagg ggcagagc ggcagacc 4680
ggtatgtt ggcgctgc ggggtggc cggcgccct cgcattga gggggggg 4740
gagacgtg tgcggcgg gctggcag gggccgtt ggggaggg agggaggg 4800
tgtgtcgg cggggccc tagggtca ctgtcttc ctccagca gccagccc 4860
atcgtcga cccaaggg aagggaag tggagtaa cgggtaggt cggggggc 4920
tggggggcc tgggttgc cggcccca accgctta cgaacctg ggtctcgg 4980
tcttgcgt ctatgggg caggtagct gtccccc caggagctc aaggtagc 5040
ctggaccg ggcagccc gcaccagc ttgggccc ttgtagct cgccttgc 5100
ggccatct cccagacc cttcccca gtccagaa acaggagtc cttactcc 5160
ccgagatc cagccggc ccttagtg gggagctt ctctcttc gctctgct 5220
gggtcagtg tgcagaga gcccctccc cagggctc ggcagctag gcccggat 5280
gtatgtgc agcgggtc tccctgct gactgtgc ctgggtga ggcgcgca 5340
tgttcaac ggaaggaa tgaatggc gctttaga aagcctgc gtagtacc 5400
ctgctctt gctcagtg gttagtgc gtgtgggg gaagttagt gtaggggc 5460
tagctgct gattctct cgggtgat ctctctag attatctt gtaaatca 5520
agaatgggt ttatgggt cctctgtg cccctccg caggagtg gtgtgtg 5580
catgtgca agcgggga agttagtca tggtagtg gaaagaca ttccaccg 5640
aaatggcc ctctgggt ggcctctc tctagctg gttacctc cggccccc 5700
ctccctgc cagctagc ttgaccgc ccaaggcc aggtgtaa tctaccgg 5760
aggtatgggt gtctgggc ctggggaa ctgctctt cctatctg tctccgaa 5820
accagctc caccgacc ctgtctgag gttaataa gctgtgac ttctgtag 5880
tggggcct ggtgggct ctctccac cctctccc acacatgc actacctg 5940
gtccacct ctgtcttg gaaagagc aggaagca ggaacttg caatcaag 6000
ccttggaat aggggtta aatacagc cctctctc cccgccc agtctgtc 6060
ccttttag gaggacta gagaagggt ggcctgcc tctcagta attctgac 6120

```

[0012]

```

tttactcttg ccttttgagt ttgatgaige tgaatglaaa agcgatttct ccttaaggag 6180
tgcagctgag ctaggcagca gcaagcattc ctgggctggc atagtagggg gttgaatacc 6240
atgtacaaag ctgtgcccac gacgtgggtt ggcagtgccc cacatggcgg ctctctctgg 6300
aaggccttcg tatgactggg aggtgtgggc agccctggag ccttcagttg cagccatgcc 6360
ttaagccagg ccagccttgc aggggaagtc aaggagata aaattcaacc tcttggggcc 6420
tcctgggggt aaggagatgc tgcattcgcc ctcttaattg gggagtgccc tagggctgat 6480
cacatatctt gggggagcct cccctcccca tgccttcttg cctctgtctt cttagatttg 6540
gtctatfagg ggccttggtc ccaggcgctg ctlttaactc tggtaangtg gattatgttg 6600
ccatcaatga ccccttcaat gacctcaact acatggtagg tcttatatgg tgaagcccaa 6660
agolgtgttg gaggagcca ctgggtgtat gggcagccc ttatatacct caagtattcc 6720
cccaggttla catgtlccaa taigaattca ccaatggcaa attctatggc accgttaagg 6780
ctgagaacgg gaagctgttc atcaatggaa atccatcac cacttccag gattgagtg 6840
aagacagat ggaagaaatg tgccttgggg aggcacatag gattgtgttg ctcccttggg 6900
tatatgttaa ccttgttccc ctcaataagg tctgtccccc atctccccc caccaccata 6960
ggcgagatcc ctccaaaac agtggggggc atgttggggc tgaatagctc gtaggagcca 7020
ctggctcttt caccaccaig gagaagcttg ggttagctgc aggaagccc gcaagagggc 7080
aagctgcttc agccttgcaa agctagagcc aggttctata agtggctgtt tctctgtgtt 7140
aggtcatttt ggaaggggga gcaaaagggt tcatcactc tgcacctctt gctgatgcc 7200
ccatgttctt catgggttg acccaatgaa agtatgaaa cagctcgaag atcatcaggt 7260
gaggagcca gggccgttg agaagcgccc agccttggaac ctatgtgaca agtcccccig 7320
acttgctccc cgtctctctt tcttttgoug aaatgccttc tgcacacaa actgtttagg 7380
acctctggcc aggtctatcc atgacacatt tggtatcttg gaaggtctca tggtatgaa 7440
gttgggaat ggaactgagg ctccacatt tctatccaa gactggctcc tctctgggg 7500
ggctgcgtgc aaccctgggg ttggggcttc tgggacagg ctltcccaaa atttcttllc 7560
aaggtgggga ggaagttaga ggggtgagt gggagtagg ctgagggggt tcatctcttt 7620
tgcagaccac agtccatgcc atcactgcca cccagaagac tgtggatgg cctccggga 7680
aactgtggcg tgatggcgcc ggggtctccc agaacatcat cctgtctctt actggcgctg 7740
ccaaggtgt gggcaggttc atccctttagc tgaacgggaa gctcacaggc atggccttcc 7800
gtgtccccc tgcacagtg tcatgttggt acctgacctg cgtctatgaa aaacttgcca 7860
aataatgaa catcaagag gtatgtgaag agggctcgga gggcccccct aagggatcc 7920
tgggtacac tgaacacag gttgtctct ctgacttcaa cagcagacc cactctcca 7980
cctltagcc tgggtctggc atlgccctca acgacacatt tgtcaagcnc atttcttgtt 8040
agtgtgttg ggcagagac tggctcttaa aaatgtcagg gctctggccc ctctgttggt 8100
tggctcagaa aaaggccct gacacctctt tctatcttt aggtatgaa acgaatttg 8160
ctacagcaac aggtgtgttg acctcttggc ccaatggccc tccaaggagt aagacccctg 8220
gaccaccgc cccagcaga gcaacagggc aagagagaga cctctacttg tggggatcc 8280
ctgcacact cagtcccca ccacactgaa tctcccttc tccagttagc catgttagcc 8340
ccttgagag gggagggccc tagggagcgc cactttgta tgtaccatca ataaaglac 8400
ctgtctcaa cagttactt gtctgtctt atctaggtt ctggggcaga gggagggaa 8460
gttggtttg tctcaaggc agacattctt gtggggagg gaactgttat gtctctca 8520
gactgaggt aggcctcca aacagcttg ctgtctga gaaccatttg ctltccgttc 8580
agactgttg agtctacag gaagctgcca ccaactatc agagaaacag gcttttctt 8640
ctctctctc cagctctagg ctatctgtg ttggcaaac atggagaag ctatctgtg 8700
ggagcccca ggaagctga cagtgagc aagtacggc tgcactggx ctctgactt 8760
gactgttag tggagctag ctggagctg agctgaagc ggaattcca gcttggctc 8820
cgcagctgtg aggtcttag cagctctct atgtcttt gtgcctctg gtcttatctg 8880
aggacatct ggcagcccc taaggcttc aagcagatt catctagta aacaagtac 8940
ctaaacat gcccaaggc glaaggacta tataatggtt aaatctgtt aaatgccc 9000
acctgcata gtttgagga agatgaatg agatgtgta ggtgactta ttctcatat 9060
cgtcttagg ggaacttgg tagggcaag gcgtgtagt gggacttag tccagabccc 9120
tggctctgcc actgaacgcc tcatgtctt tggcagta ctcccgggc tcaatttga 9180
cgtgtctta ctatgtgag acaaaaglac atactctgt agagcgcca cgtctgaac 9240
ccagcactt tggagggca agtgggtgt atcacttag gtcaggagt tgaacagag 9300
ctggcaaca tggtaactt cgtttctac taaaatata aaatctagct aggtttatg 9360
gcacatgct atagtccag ctacaggat gctgaagag gagaatgct tgcacccgg 9420

```

[0013]

```

aggcagagggc tgcagtgagc tggagaccaca ccaatggcact ccagccciagg caacagagta 9480
tggacttcca tctcaaaaaa aaaaaaagla cctaccicag agttcaaaact agtgaatatt 9540
aggaaigtgt tggagacagt acaccaaggt gcaacaalaa taccgctcag ttctattatt 9600
attaaagaat ccatitgaat gtcagctcaa cacagccccc tataccgagg catgtgaae 9660
cgcatctccc cagttctccc aggtttttcc aagaatcagg gacactgtag cctgttggtc 9720
tcagtgtatg acagacacgg aggaagcaca tctttagttg atacttaaac agagaccctg 9780
aggcacata caccgcgcca cacatgcata gaggcttccc ttctctgta ttctgcagtg 9840
accagagagc caagagctcc cactccctt caaacaatg tgcctatccc gggcactaag 9900
gcctctttaa agcagggcac ctccagaggc gagggcacaa gccacataca ctccacctgg 9960
caggtaggca gcgtgagcac gtgaccata caggagacaa ggtgcacagg ccagctccaa 10020
cgcccttggc cgtgacagc gacagaagcc ctctccagct gctgtgtgtg cagagggcat 10080
ggttagcttc cagctgcat ctattccact ccagtgctg ggcagttag caccagtgtg 10140
gaagacagtg agctggctcc ggaacacagg gatggagaa aggtccaca ttacatttc 10200
tgatactggc scaaggtgag gggccgcaat cgtctcgca gcatitlaa gatgggaag 10260
tagcagcac ccacgctga aggcagagc gcccacag ttgtggaat gcccacagaa 10320
tgtagggcc aagctagct ccagacccc cagagccctg gagaagcaca gatgagga 10380
gaagcttga gggagagagc cccagctccc caggagccgg cctgtgtcag agctgagct 10440
gatgtccc tctgtcagc ccaacctctt gctctcga gctccctgt gcaggccct 10500
cggtgcag gggagagcag gtctctatct catggagctg tggatgaga catgcgctc 10560
gagttctca gctcgttg gggggggcgg cggctcga agcgggagc cagtgaagc 10620
aggagcttt ctgaaaaaa acatcagtg tcaacttgg ggcaggcagg aatctgaag 10680
acggagacca ctctctctcc tctgtctca cctcttgca gccctgaga agtaccgaa 10740
gcgagggcag ggcgcgggga tggcgaggga gggcagga ctgaacttc tcaaaccca 10800
ccctgacagg gaaatggccc ccgctgtgt ctgggaact cagagctga ggtcaggcat 10860
gatgtctac gctgtaat ccagccttt gggagcaga ggcgggtga tcaagcgtc 10920
aggagttcaa ggcagcttg gccacatgg tgaaccca ttctactaa aaatgcaaa 10980
attagccagg tctgtggcg ggcgcttgg agctgaggg aagataaca ctgaacctg 11040
ggaggggag gtgttagtga gccaaaaaa aaaaaaaag aaatagctga agtcacagta 11100
ggagagaagc tctgagcct ccagcctct gacttaggg ccttgcttt atgtctatc 11160
gcagtatttt tctgatttt aaaaattac ttcttcttg cgtgttaact tacacagggt 11220
caaatgcaca aatcagggc ctgaatttag aaaaaact gcccacaaa cctctctgt 11280
ccttgccagt tttaaacgt ccttaacca ggggaacca cagagctgg ggccttgga 11340
ggtttagcc ctccgtcat gaattgacat agctatcca actgccagg gagagagctg 11400
tgggtctgg ccagccccc caggtaact ccaagggca gccccacag aagaigtac 11460
ccagtcattg cctgagggt tctgggctg tgttcaaac ttctctccc ctgtctccc 11520
ctggaagcc ccatgccag gggagggccc taccgtctc cagactgggt gatgagccg 11580
cggcaggtct ccatctgac gtccaggcc cgttcaagc tgcacatc catgtactg 11640
tgcagtgcc ggttcatgt gttcttgcc gtggcaact ccagctgtg tgggcaggc 11700
agggcctag ggttaacag gtgggttca cagcctct gtggcccg ccaggagcc 11760
aacagccaa gacagtggt tgggcgggg gccaggcag ccttactga aggtcagat 11820
tttaaaata accagacca ggcagaagg agcaalaca ccatgttt ctctgcttt 11880
cttgaaatc ttgtctatg gaggatga aggtactg tgcagctgt gtcttllaa 11940
gactgataag agtggtgag aagtcagtg tggcgggat ggaagcagtc cccagctcc 12000
cagggcaca caggttcta tggggccgg ctacccctc atcccctga ctgcctggt 12060
cctggctgt cctctactg catgtgagc tccgttaag gttagcaca ggtttcttc 12120
atctggcag accagacct cgaagggaat taactctc tggacccgt gccatacag 12180
cttggtaac tctgaaggg ggcacatcc ttatcatt ctctcttat cactttatt 12240
tctaccta aagcatttt cagacactaa aaggaagact cccataitg aggtgccatt 12300
gattctgaga cctgtctga ttctgaat gtcacagtg gaaacacgg ggtctagaat 12360
caagaatc cgtctctcc cccgggccc atctgcagc ccacacttg aggcagcgg 12420
agggcaggc cccagcagc ccaactcag tgcagggac agtcagctc ctctcttgt 12480
ttgggtaca aaggcaggc gccgaaggc gaagcaaat gaacttctt ggaatgacg 12540
ttctgactg cctgagtga gagaacggc aggalaccag ctgcaaat gttgttga 12600
cagagaggt ggaacttag ggaagaggg aaacagttc caggccaga cagcagat 12660

```

[0014]

```

ggtcaggaca gtagatttg ggttagatg gggctggct ggttagaag ccttagtgg 12720
ggacgcttc ctagcaggct tggatccaca gctggggcc cccagtagtg gtaggagag 12780
gacacgtga cgttcaacc caggggaaa atacagtgc accagccct gttcagatca 12840
gctgcctaca ccccaattc tccaccttgg aaagaccga gttagggag ctaggtttg 12900
gaacctaac tcaggagac tgggagaag gaaatgacc aaatatttc tactgaata 12960
caagctagt acgccaca gtgtctctg catagtcta tgcctcact taatcaca 13020
attacagca ggaacaagg caggaggtg acgtcatgt caggagatg ccttagaca 13080
ccaccttga aatgcaggt tgcctgttg aaacacttg ggggtggan gtagagag 13140
agagaggtg gtagaggtg ggggagaag tgcctcgc ccaacatct gggcttgg 13200
tttacctgc tccatttag tgaacctgag agatcaagg tcacaagat aaagcacct 13260
tggtacttc tgaagtcag aagtcctct aaggttctc ctaatgaga agtcccaaa 13320
ccacagcca ggaacagta cctgcttgg gctgttagg aacagcca aatagagga 13380
ggtagcagt gggcaagta gtagagttc atctgtatt acagctgct cccatcatt 13440
gcattacag ctgagctca cctctgtca gatcggcgc ggtattgat tctcacaga 13500
tcagaacc tacttaagt gaactgtga tggagggat ctaggcttg tctcttct 13560
gagatctaa tgccttaga tctgtacca tctccatca cccagatg gacagctta 13620
gttgggaaa acagttcat gcttctcct gattctcat tatgttga tgtataat 13680
ttcatata tattaaag taataaat agaaaagag tgcacagcc ggttgcgtg 13740
gctatgctt gtaatccag caatttggg gggagagtg atagatcc tgaagtcgg 13800
agtcagagc cagcttga acatggaga aacccatct ctacaaaa taacaaata 13860
gccagggtg tgggcacat cctgaatcc cagctactg ggggttgag gcaggagat 13920
tgttagacc cggagggcg aggtgcagt gacccaagt tgcctatg cctccagcc 13980
tgggccaac gacaaaat cgtctctaa aaataaatt aaattttt aaaaatga 14040
ataagtcac caataaagt aatgttga taatccaa atgattccc cggatcagt 14100
ggaaaactg tcttccaa accagctct ggtacaaa agttagggg ctgtgctta 14160
aagctatgc actggagag agggaggtg agatgattg aaataatit tttttttt 14220
ttcttgaga tggatcttg ctctgacc caggcttgg tgcagtgga caatccag 14280
tcacgagc ctaacctcc cagattccg caattctct gctcagctt cccaagtag 14340
tggattaca ggtgcacc acaatgctg gctactttt gttttttt tagagaggg 14400
gttctatg tgggtgca ggtgtgtc aactctga cctcaagta tctgcggg 14460
ttggctccc aaagtctg gttatagc gagagccac agctcagcc aaaaatgt 14520
taaaagcat gattcattt aagataat tttgtctt tccagatg ttttaccg 14580
ggacttgag gtcaggga gactaggg agatggagc ctagatctc tctcagaa 14640
gatgcgcga caaggtccc tagttgtg aggtcaggt gaacacaa accitaaat 14700
caccatcaa aagaggaga caaatccg gtcaggcag agggagaa acgtagccc 14760
ggtagaggg aggttggga gactggtt ccttgcctt gctttagct agacacaa 14820
atcaggaa cctctatc cggcgaag aaacaggt tgaicagga gaggagat 14880
tgtgtgtg tgggagaa gattatag gaaataat tctgactt cttctcaac 14940
actccatg tagagcaat aagataag tcaagcgg gctgtgtg tcaagctgt 15000
aatccagca ctttgaag ccaggggg tgaacaga gttcaggga ttagacat 15060
ctgctaac atgttgaac cctgtctc taataaata caaaaaaa gctgggcta 15120
gttgcaggc ccttagtgc cagctactt gaggctgag gtagagaa ggttgaac 15180
tggaggtg agcttgcgt gactgagt cgcgcactg cctccagcc tggcgaca 15240
agcgagcta cgttcaaa aataaata ataatat atagaaag aataagta 15300
ctctctct atccactgt ttaatttcc aggttccg gtaacaaag tgaacttga 15360
tcgaatata taatagaa tctccaga ataacatt cttatgtt aaattgag 15420
ctgtctgag tggctgaa aatctcact ctctctg gctgggtg ctacagcta 15480
taatccagc actttggag gtagagtg gctgactg aggtcagg atcagata 15540
tctgtctaa caggtgaa cccatctt actaaaaa acaaaagaa ttagccggc 15600
gtatggcg ggcctgtg tccagatc tcaggagct gaggaggag aatggctga 15660
accagagag tggagctgc agtgagga gtaacgca ctgcactca gcttggta 15720
cctagcaga ctctctcg aaaaaaaa aaaaacct cttctatca gaccagaa 15780
tgaaccttc cttgtcca agatccaa ctgtctat tgcctcca ttgttact 15840
aggagcata tcatgaga gattgagt ggtgtatg cgttccgt gttcaagaa 15900
cctttattt tactataa tggcctaaa gcaagagt agtgatct gcaatttga 15960

```

[0015]

```

taigtgaag agaagccaa atagtctcc ttaagigna aaggtaaag ttttaactt 16020
aalaagaaa agaaccgat gcigagilg gtaagcicta cagtaagaal gaalcitct 16080
tcogtgaat tgaagaag gaaaagaaa ttigtccag tttgtctg acactcgaa 16140
ctgcaanagg taccgccac ggcacgcta agtgcttngt taagalagan aagcattga 16200
atgttggtt gggagcatg aacagaaca cgtccagtt gacagcaact ggcaccaagc 16260
taccaggag tcagtctat ctlltgcttc agcctccac tgggtgtctt ggaacatg 16320
cccgaagat aagggggac tagtatgaa tctccatgt gggctgaag aaggtctag 16380
acgcagaaa ggcctcatg ctctcccaa gtcagagct cgggaagga ccactccac 16440
agccaggag agtcacggg ggtccgtcc atccctctc ccacagcaac agcaacctc 16500
attctgtgt tccacaatc ctctctctt ttttgtatt ttgagacgg agcttgtct 16560
tgtctccag gctgggtgc agtggtgga tctggctca ctgaagctc cgtctccgg 16620
gttcagcca ttctctgcc tcaactccc gattagctg gactacagt gctgcctac 16680
acggcagct aattttttt ttigtattt ttctagaga cgggtttca cctgttagc 16740
caggttgct togtctctt gacctctga tctctccgc tggctccc aagtctgtg 16800
gattacagg gtagccacc gcgcccgc tccacaatc ctttgtcag ctcttaagc 16860
ctagatttg agctanaga ccagctcac ttgctctag gacctctag aagttactc 16920
aagcttggt aaaaagcatg gagattctc tttatctca aagctgagt ggaagaaac 16980
ggaagctcc agtcatctc tgaagccat ggcactgata ggttaataa taataaaca 17040
gtcttacta tctacatc gctatgacc agacatgct ctgaagcta tataagatt 17100
attctatg atctctaaa caacctata ggtgagtat tattaattc ctattttac 17160
aaataaaaag gctggcaca gggctcatg ctgttaacc caacacttg tgggcccag 17220
ttaggaggac tcttgaac cagaagttg agggcagctt ggaagacata gtagacctt 17280
gttctctga aaaaaatac aaaaattagc cgggtgtgt agcacagcc tgggtctca 17340
gtacttggt aggtgaggt agggagttg ctgtgacca gaagatcac gctgcgttg 17400
gcttgagtg acagagcag accctgtc aaaaaaaga agaaagaaag aagaagaaa 17460
aaagaacag gtaggggtt ggcacatgg ctcatgtc taattctagt acttgagg 17520
gctgaggag gaggatcat taagccnag agtlaagac ctgcttgga aacacagga 17580
gacacctct ctacaaaa atacaaata aaattgctg ggtatgttg cacacactg 17640
tggccccc taccagaag gctgggttg gaggatcat ccagctagg caacagatg 17700
aaacagatg agactgtc tcaaaaaa aagacagaa agacaagga gagaaggag 17760
gaggataga gggggggg aagagagaa gagagagaa ggaaggaag aaggaagaa 17820
agaaggaag gaaaggaag gaaagagg gctcagag gaacaagta ctattacaa 17880
gttcagggt gttgaagaa acagagctg gattgaacc gatctctg tgtttcaga 17940
gctgaggtc ttacagccc accagagga ctatagctg gactcatc ctctcccaa 18000
gaccttgct tttaggtcc tgcagcagg atagccac tgcctctga gccactca 18060
aaacctctg gacaccag ccttggtt cgtctctg ccttac 18106

```

<210> 18
 <211> 16312
 <212> DNA
 <213> 小鼠属物种 (Mus sp.)

```

<400> 18
gtgagtgct ctgttggtt cgggtcagt aaactctc tgcctccta agtctctga 60
gactctgtc tgaactttt gcttgctag ctctacta gctatgac caggtgact 120
aagaccgat cttaaltct ggtcccttt taaagcca ttcaittct ttttgttaa 180
catgcatgt tatttgctt gtgtatgtt ttgatgaag ggcctagc ccttggaat 240
tggattaga ctgttggtt ccttgagaac tgaacagg tctttgtat gacagcag 300
tgcctctac cactatgca cctctcaga cctagtag gcttgagat aggtcttgc 360
tatgagacc cagctctc agactaac agatccctt gctctgtt ctgttgata 420
gagtatgcc gcaatgtt gcactggct gctctttag agtatttca tgttaangc 480
aaaagtcgt aagtagact cccatagct ctgtccaca cccgatggc ctccaccca 540
tgacctctt gcttggtat acgtgtctt ggaagaaga acgtatga cagatcaga 600
tscaaagtc actatgga cacagtcca calacttacc acttgctca tactggcat 660
ttcaggact ttgagtcac cttaacccc accctggtt tggcatcgt tagtggat 720
cacagtcca gattgttct ttatgttct tcttcttca tggctattc ttlttaaga 780
ggagataat gtttcagcc ctgtctttaa ctctcttcc tccagctat aatagaaga 840
tttagcaga agctctgtc ctgtacacc agagccttg atgnaaga ggaattga 900

```

[0016]

```

acigccragg gaggcagcca ggcagcccg tcgccaaga aaccctccgc tggatagaga 960
ggatgcnaag cccacatggg gctcciatga gggatccca ccttttttt gtcttttgt 1020
ggtttgtttt tggttttgaa atggctccct tgcctttctt ttctagcttc aaggtacag 1080
cctctgactt ctgtagactc agacaatgac ttgtctarac ctacgccgc acgtaccaaa 1140
cttggtgttc cacaactca gcagcgcaag aagtcacaga ggaagccaa agttgtctc 1200
ttgagtgacg agtcacatga ggcaggtatg atacaccttc ctacacagat agctgccctg 1260
gtgggtcttt ccagtgggc ttgcctagct gtggatctat tcttcctggc catgtcttc 1320
tcactatgca gctggctga ctccacagc gtgcctgtgt cgtgtgtgtt tgtgtgtgt 1380
ttgtgtgtgt tgtgtgcaca cgtgtgtgtgt tgcgcacgag gcatgtgtt tcttttgaga 1440
tccattttgc cccatttta tgactcagct ttgttctta taattctttg aatagaactt 1500
tcagcagaga tgacagaaga agagacccc aagagaccca ccccaatcg tagagcalet 1560
ggcgaagac acaggtctta ggcgtgctt ttgctatc cgtccctgc cgtgcagggt 1620
atctatagg atgactttg attcagtc cccccctt tgtagaagt cttgttgtt 1680
gtgttttaa catgaaatg gctttgcat tgaggtctt cctttagag cctgtctcca 1740
ctttcccttt ctcgagccg gtgcctttg ctgttttta atgataaat gtttttag 1800
acttttaag atccgtctt ggtgtgtgt tccatctt agccctgagc tgtgtgtga 1860
tcttcagtc cctgtgtgt cctcatctt gcacaatga aactctgtg tgcagaatta 1920
tgggtttct agttcagc gacagttct cagcagtcg ggtactttg cccccagaa 1980
aacattggc aatacctaga gacttattt tgttgcctt gtggcgagg ggcgtcagt 2040
gaagcctta agatgaagg cactctatg cactttgtg cccataaac agcatgtcg 2100
gaglacaaa ctgagaaat gctgtctag acctaaggt ttgttctct gtgcgccca 2160
ccccctccc cgaagactg caatgactc agcctgcta atggcagaa ttatgaacc 2220
cactcagtc tatagtgac aagtcctat gctcagagc cagtcctta ctacataag 2280
ggctctgag gggctcagc tgaagccat gctgtgtgc caccatccc aagccttaa 2340
ctgtctagt cactggagc agggagcaca atgtctgtt ctgtgccat gacacagca 2400
cagtctgtc gaagtctcc ctacctctg gtgaggggc agtaatgata gctgacaggc 2460
gtgagcnaa aggaaggtg cactctgaa acactaggt aggcctggc tatggaagg 2520
ctcttgggt aggaagagg ttacgtgtg gcatagacc tggtaggaa ggcagaaag 2580
ggactctgaa ctgccaatg ctctctccc caccctgag ctgggcac cccacctt 2640
ccccctgtt tcatctgagc tgcctccgc tcatgggag cctggcagc ggcacacag 2700
ctgtcagct aacacagcc tggctctgg ggcacagga gtagagcaa tctggaagg 2760
caaggagcca agactagat tgggtgtcag cccagctgc tccctgctt ttacgaaag 2820
gtatcacca ggcagctaa acttagaat agctcttga gctagatga caggggctg 2880
gttcatgtt gcactggct agcaagagg cgcagagtc cccctgccc gcactctga 2940
cagtctccc aagccctcc acccttccc gccctctac ccttcaggt gaagaatcc 3000
gtctcgaa agacaanaa aaaccaaga atgccttct tcccttccc tgttgcaag 3060
aggctagggt ctccctgtc ctggctcag ggtataaat tctctctct gttctccc 3120
ctactgact taccctgtg tccacaggg cagcaaaaa agagagctc ctggctcag 3180
ggcagtcag cagtcctct gatggggagg gacttccgc ctacgtccc aactctcac 3240
cctgggcta cagtgggtg aaggggcagt gtctctaga ctggcgggg cagctctag 3300
gttcaggga ggaatactat aggggcctc tgaacctct ccaattcac ccttaagg 3360
gagctgccc ttaccggg gtccagctt aggttctca ggttaacta ggcagagtt 3420
tgtagctc aattatcag ttccaggt gtagggcct ctgacctg cgttagaat 3480
ctctttta agctctcc tggccccc ctggccca cagcgggtt cacaagcag 3540
gcctctgccc aaagacaga gccaggagg ggggggggc agataggaa atgggctcc 3600
tgaggctcc cggggagga aaagtcccc caccctgca ttcttctca ctctttttt 3660
tttcggggg ggggttgtt ggtctactc cgaagacaa cagagagaag atcttcaat 3720
tttcagcag ctttcaata atgggagag gtctgatgt gcagtgtcag ggcagccac 3780
acttctcat ttccctgtt ctccatttt acitgggaag cagratctag gtctgtggt 3840
cctgagtc ctgtgtcac actccaaga cctctgtcc taagtctat agctgtatt 3900
cctgagtc tatctggga cccatcccc ggtacctcc tgaagaggc aactctcgt 3960
cccccccc tatcagtcg ggcacacac gcttgggtg tgcacattt aaaaatagg 4020
cggctcnaa gaggggagg aggggaaag agagagccc agtactgc gctttacgg 4080
gtgcagtag ctacggctc tgcaccttg agctaggact gataagcag ggcgggggc 4140

```

[0017]

```

ggggggggg tcatcagctc cccccacaa tccggglicc tataaatacg gactgagaa 4200
ctccctggig ctctcigctc ctccctggtc cagagacggc cgaaccttct tctgacatgc 4280
caggtagaaa tccggagtg ggcgcagga gcccgggac agtcggaaac tgggaaggag 4320
agtgggactt gtacgggtct agggatgctg gtgcgaatg tgcgaacgg accagagtc 4380
cgrattgcag ggggggag atggagagac tgatggggcg cagggcgaga atggagatgg 4440
gggggggag gggactgcct gggtctcttc gggcacgct aatctcatct tcttctcttg 4500
cagctctctc ccttagacaa aatggtagaa gtgggttgta acgggtgagt tccagggcgg 4560
ggcctctctc cgttgcctac gcaggcttgg ctgacctggg ggtctcagag tatgtgggg 4620
aggtagatga gattggcgaa ggcggcagg agacctcaag ctacagcttc ggaactggcg 4680
atggctcgca ctggcgccc ccagctgctg cactctgggt aactctgcct ttgggggag 4740
gagcgcccg ggtcttaag tataggaac aacccacgc gcccttcag accatcccg 4800
taatcccgag tccgggctt ctctcttac ttctgcgcc tggagatga cgtccagga 4860
gggaagcccc tccccatct ccccttctct cggggctggg ggaatggct tctgaggtgc 4920
atacctttgc gcatcctc ccaggcttgg gcttcttca ggttaactgg ccgcggcat 4980
gtgcacacg ggaaggaat gaaigaacc ccgttatga atcttgctha ggccttctt 5040
cttctagct tctgactaac ctactctc tggctgggt ggaatgctt tctactgta 5100
ggcagatga tgaaggtt cgtgtctc gagagatc taccacaa tctgtctac 5160
cttattagcc ttaaaagccc ttgagctta ttgtctcgg aaataagcg tatctagat 5220
tatctctga aatcaagc ggaatcacg aggtctgctt gacctccaa cccagaggt 5280
agttatggc tagtgcagag ccgtggatg ggaatgag tcatgggtgt ttgaaaaga 5340
aatcttctac caaaaatgg ctctgtagt agcagccct tccatccct gacitctca 5400
tcacagctc gactgacc aggcctata ggcagagtg taaggtcat taagagatt 5460
gggtgtctct ggcctcaga atctgccc tctcccggt ccactctca gaaacagat 5520
ctctccctc cgcctgata tgggttana tttagccgt tgaacttct gactctggg 5580
cttgagggg ctctccccc tctccctct acacatct gttctcgg ctctcttct 5640
tgcccgaaa gaaaggtgt ttgcgaacg agcctcgga ttgggttgg aacccccca 5700
catgtttct cagtcttcc ccttagtctg agggacttgg aggaacagc tggcccccc 5760
ctgtctctct cagctgacc tttagccttg cccttgagc ttctctgta atgattcac 5820
aggtctgccc tctccgggg gttagcttg aagtcagcc atgcgggaa aaacttctca 5880
ggcatgagt gatggggtg atgtgcag ctgtacctc gactggggca aatgtccct 5940
tcttaagaga cttagaatga ctggaggag gtttgcggg caagaatca cctctggac 6000
agaaagaaa cctcacttt ataacctgc taaaaagcc ctgcagccc tgggtctc 6060
aaagatata aaattagct tcttggact tttaggggt ggaacagct atatttggg 6120
tgtacctca agcattcac tagctttat aaaggaat ctgaacaa catgaactt 6180
ctgagtata catctaat gtactgtct tgcataag ctgtacct ttgttggt 6240
cgtctatgc tnatggctc aggtctcna cactatggt gaaacagat tccacttga 6300
agaacatgag atagctggg gctactaca gacctagag gactctgat ctgactccc 6360
ctgttcttg tcttccagt ttggcgtat tgggccttg gtaactagg ctgcatttg 6420
cagtgcaaa gggagatg ttgcataa cgaaccttc atgactca actacatgt 6480
ctactgttc cgtatgact cactcaagg caaatcac ggaacagta aggcagaaa 6540
tgggaagctt gtaacacg ggaagccat cactcttc caggagcag acccaatca 6600
catcaaatgg ggtggcccg gtctgagta ttctgtggg tctactgtg tcttctcac 6660
catgagag ggcgggtta gggccgaa agctgaaggt gacggcac ctgatatgg 6720
tgaacctga aaacagaa ctgactga aatcaact ttctctta acagggcac 6780
ttgaagggt ggcacaaag ggtactatc tccgcccct ctgcagatc cccatgttt 6840
gtatgggtg tgaaccaga gaataagc aactacata agatgtcag caatgactc 6900
tgaaccaca actgttag cccctggc aaggtatc atgacaatt tggcatgtg 6960
gaaggctca tggatgtc gactggga gacagctat gacttctta tcttacctg 7020
ccatgttgg acccttctt gtaggtctc ctctggga gagggtgct gtgaggac 7080
tactcttg cccctgct tctacaca cagctatg catctgct acccagaa 7140
ctgtatgg cctctgga aagcttggc gtagtgcc tgggtctgc cagaacata 7200
tccctgata cactgtgt gccaggtg tgggaaggt cctccagag ctgaacgga 7260
agctacatgg catggcttc cgtgtacta ccccaatgt gctcctcgt gactagagt 7320
gccgctgga gaacctgta tctatggga gactgggt tctctctgt gtacagtga 7380
cttggacaa gcatagcat ttgggggtt gtcttata gcccaagat gatgacata 7440

```

[0018]

```

agaagtggt gaagcagca tctgagggc cactgaagg cacttgagg tacactgagg 7500
accagtggt ctcctggac ttcaacagca atccacac cccacac cccacac 7560
ctggcattgc tctcaatgac aacttgta agctcalle ctggatggt ggaatggaa 7620
acctgacttt taagagcaac tggaggttg gtgcctcgt gtgctagct cagaaaagaa 7680
acccaaacta acagtgtcc caattgttc taggtatgac aaagaatag gtacagcaa 7740
caggttggtg gacctcatg cctacatggc ccccaaggag taagaatcc tggaccacc 7800
acccagcaa ggaactgag caagagagg cctatccaa ctggcccc aacatgagc 7860
atctcctca caatttcat cccagaccc caataaca ggaaggact agggagcct 7920
cctactctc ttgaataca tcaataagt tctgtgacc caactctct tggtttaat 7980
gggtgttg ggaaggagg gtgtgtatg ggaagtgaag caggctaat caactttgt 8040
gcatglatcl ttatggctc tgaatcag ctgggtgctg gaagtcagg gtacatct 8100
atcaagtc cagccaggg cgcctgtct gaacagcga tatgttggc atggcagg 8160
gtgggttg acaatggag aggaatgca gtctgagct cactgttg acagcagc 8220
cttcagagc tcaactctc tgaagaagt cgggtctct ttatctgaa acaagctta 8280
attgtacc caactctgc agcagagct aactgtat aactgtga ggaagccag 8340
tacgtaaa cctatggia tcaatgaag gggctgtg cactgtga tctttgtt 8400
ctgtgtcag acagtactc agtctgtg aagtatca cactgtat gcttttaag 8460
gttcatct aatgttcaa catagctct ctggaggt ctgtatgac tccatagct 8520
aggacatc taacttttg gtttgtcag tgaagcct ggcgcacac cctcagctc 8580
tactggaga aacagccc ggcagcagc aactgtgag cccacccat gtgcggcag 8640
cagcttacc tctcagcgt cgtctggc cagcagagc atctccaa ggcagctc 8700
cagccttg tctcagca ggcagagt aacagctg agcagctg cagcagcc 8760
cacactcc aagcctta tagggagca agccttcc caagctat gatgaatc 8820
cttcacaga gtctcagc aagttagc cgtacaaa tcaagatc caagcata 8880
cagcagct cctgtatc attcttca aagtgaag ggcagctc caagtctc 8940
ggcagctc cgaagtgga gaagttagc ataccatc gtgagagc cagcagag 9000
tgtccaatg tcatggcgt tgaagagga ggtggggt ggcctagct cagacacc 9060
cagaacttg ggaaggag cccagggca ggttcatc agtcagcag aagaccca 9120
cctctgaga cccacccc tgcacagc agctctggg tgcagcac tggttctat 9180
ctcatgagc tatcagta gactctga tggagctc cactctact cggcgtg 9240
ggcgtctc taagggag cgaatgaag ggaagact tctagaaa aacacagc 9300
tgtcacttg gacaaagc gaagctga gacttcaag ctgtctct catggctc 9360
tgagtgag ggggtccc gcaaggct ggggggag aagacctc cctttcaac 9420
tgttgagg gcaagctc tctatctg ggggctc agagctcag agatagaa 9480
agtggaagt cagcagcgt agtgcctc tgaagcag tgggctta gctctgtc 9540
cagcgaat ctgtatgg gctggggt cctgaccc aacactct ggtcaagc 9600
atctctgc cccaaagt gctgggtg agccagat ctcttgaag catgacat 9660
atctcagc tgaatatt tggagatg ggttggtt ttgtgtct ttgagcag 9720
gctgtcagc attctgtc ggcctgaac tctataga ggtgagct gctgtact 9780
tgagtaact gctgtgtc tctatgct tgaatgag atatgcta gcaactga 9840
ctacgtagc ttctgtgt ttgtgtgt gtgtatgt ggtgtgtt ggcacatt 9900
tgtctgtc cgtcagag agctttaa tcatgaaat cctctgct ctgttcca 9960
agactgaac tgcagatg taccactc ctgtgttc agtaattt aaaaaatt 10020
tttaatgtg gtgtgtgt gtatgtgt ctgtgtgt gtgtgtgt gtatgtgt 10080
gtgtgtgt gtatgtgt gtgtgtgt atgtgtgt cgtatgtgt cgtgtgtgt 10140
gtgtgtgt atgtgtgt ctgtgtgt ctgtgtgt atgtgtgt gtgtgtgt 10200
atgtgtgt gtgtgtgt gtgtgtgt ctgtgtgt gtgtgtgt gtgtgtgt 10260
gtgtacag ggcacacac aggttagga agctctac cactaagca tactcgaa 10320
cccgatatt actcttacc aagctaac acaaaacc gtaattga taagatct 10380
tccccatc ttgtctct gctgtgtt ggttcaag cactcagat tggacact 10440
gtcagttca tctgtgtc aagcagga ggaagatt cacttgag cccactca 10500
ggcagctcc aagcagct ggaatgaag atgtgctg gctctgcc acagcact 10560
gggctatt cagcctcc tccctggag agagaccag cctgtcgg gcagcctt 10620
accgtccc gactgtgt atgagcgg ggcagctc cactcagc tccagccc 10680

```

[0019]

gcttcacgt gencatctcc atgtactcat gcagggtgag gtccatgtcg ttcttggtg 10740
 tggccagctc cagctcagt aaggtaggt caggagagc cagggttagc aaaccagc 10800
 aaggccctgg gggccagca gcicaggaga caaglaagct acaactgcct gtctcccca 10860
 ggalatggag gagggtggg gcaactggcg atctgacttc aggcctccac tgggttagg 10920
 caaggctca gacagtgtt cccgctcgg cctgtcttg ccaaccact cctactgca 10980
 cccaagatc aatcagctga tgtctgagtt ctcgttaaga acagaaagag cacaacaac 11040
 aglaaggatg gtccagtcg gtgcgggtcg gggagacgc aiccctccag cgtttactg 11100
 gctcaggcag cagggtcccg aaggcanacc taanlaana acccggtctc aaaaagggt 11160
 aaggagcca caccaggcg ctacagtgat aaggctcttg cgtactcgg aatgaatt 11220
 tgtctctaa cctccacac tgcacagtg catgtggaga gupagagac cacaacact 11280
 aataacaa ataaiaaat ttaghattt ggggctggg gagaagcgc agtggtag 11340
 agcaccgac tctctcca aaggctcaga gtcaaatc cagcaacc acgtgtgctc 11400
 acaaccatc gtaacagat ctgacgctt ctctgggt gtctgaagc agttacatg 11460
 tacttacat taataatan ataatctta aaaaaaaa aaaaattta gtaatttgt 11520
 tgtttgttt tggaaacag gtctctcat gtacccctgg ctgtctgga actaccta 11580
 tagaccagc tggctggaa ctacagaaag cccctgcctc tgcctccga gtgtggat 11640
 taagggtgt caccacact cccgtaagg cagcagttt taacctgta gtctacac 11700
 cctgggggt tgaatcag acactctga tctgtgatt taigtataa ttgttaacg 11760
 tagcaaat agttatgag cagcaagaa atagttiaag cctgggcca caacagctg 11820
 aagcactgt ttaaggctc cagctctgg aaggctgag accactgct taaggacag 11880
 caggccact cctgtctca tcttgcac tgagctcag ctaaatgaa gctaatcac 11940
 tgtgaatat tccaaagct cgtgcgaag gagggaact ggtgtctct gtttataga 12000
 ttctctcca catctctc tccctaat cctgtctcc cacaagag agacactc 12060
 ctctgagtg attcagggt aaatcccaa caaagaat gggggggg agtctctga 12120
 tcagaaag totccttc agagtgact ctgagcag gtttgagg agnaaaaa 12180
 gtaggctag gggaaagt atctaggca gagacaact gggctctc ttgtttgt 12240
 atacaagg caggccag ataggagaa tcagctgag gtggagat ttgttggt 12300
 gctgagcc aagacagg aggtcagtg gcttgaag atgtgaca gtgtggaa 12360
 gtacccgaa acaatgta actgacgg gtgggtgg gcccaggag cagacaagg 12420
 agacgccag gacatgca tttgggta tctgtgac agtgggtg taagccag 12480
 agagatgg tgaagga tctgttgc cagagagc tgcacatt gtgaaaaa 12540
 catgaactg agatcagag ctacagtc accccaca gaaatgca cagacatt 12600
 caggtcag atcagctgc acccagta atcaacgc acacacac acatattc 12660
 agtccaga gactgtca cccataata caaacaca cacaactat tacagtcag 12720
 agagcgtg taccctac ataccacc acacacac acacacac acacattt 12780
 acagtcga gggcgtgt caccctta taccacaa caaacaca cacaacat 12840
 attacagtc gagcagct gtacccata aataaaga taccctac tccctccc 12900
 acccaccct cattagtc tctaaaga taaggcga ctacagca tctctlaag 12960
 tctatagt caccctgatt taacacga aggggaagt gcaatctc gaagtgtt 13020
 tcaaaatc accccaat ttagacaa gtgtggga gggaggga gagtagata 13080
 ggaagatga agggcggg ccatggaag aaactcac caactctg catcactt 13140
 caataatgg gctgtctc tctctagc aaiaacct agataagg tcatggat 13200
 aaggctca cagatctc agggagcc aggaatct ctgtaccac cgtaccag 13260
 tggatctt ttgtttct cctaaaca ttctccag agaacaga agctataa 13320
 cttaggaca gttcaaca tctccaa tttttgga gggcgggg agatagag 13380
 tttagagag gcttctct tggacctg gcaatctg agtattct gtcccag 13440
 ctgtctca actacagag aicctctg ctgttctc actacaga ctaaagtg 13500
 caccacaca ctccattta tagtactg tatcaact aggtctac gctgaagt 13560
 agcagcat tctacatga gacacagtc cagccagc gtgtttct ctgtctac 13620
 acatctgt tactctagt tatttgtg cttaaggat aacacaga gggagcag 13680
 tctgtcac tgtctaga ctctactg gtgacgta ggaanaag gacaagct 13740
 taatctacc gtctaga gtygacac acccgggg gggacaga agagagag 13800
 agataga gggcagca gggagatg ggaatcatt tgcctcag caagctta 13860
 ctgtgagc aacatgct ggtgggat ttgggtta aaacaaag cttgtttt 13920
 cctcagta gcagccat ctgactgt ctggcacg cagaagat acttcaag 13980

[0020]

```

ttccctttca aaagcaaac agagigacc ccaiggcnc gggaaigrac cacagggtc 14040
gcaggagacc caatggcatt agctctcgc ctcacccctg caticagtc taagacatag 14100
agccctgttc actgtccaat gggaccccc ttcctgcagt cctgtcctgc agtcacagag 14160
gacaatctaa accccagccc caccacraft caiaaccgat gcccaaggc tcaaaaaa 14220
aaagaggagc acaccctcag gagctccaag gcigagccaa ggagacagga gtccctctgg 14280
gtgtacaggc acttagggtg ttcacagtaa laaaigccca cactacccca tctcaacctg 14340
ccagatcctg tcttaagagc cagcgcgcgc tcactctatc tgaatctctt aataaccttg 14400
ccgtgttacc agcctgtcgg gagagattat gaacatctct atiggacaaa taaagagaca 14460
ggccacccct tagcagggtt gaaaccctct ctaaccctaa caggttgagg caggaggatt 14520
gcccaagtt tgaatggcta cacagaatct tggctgggtt gggagcttag gatgtctag 14580
tagttagag catgcccctg tctgcagag gacccaagtt tatctccag cgtccacatc 14640
aggctacaaa acaccctga gcgccagctc ctggaaacct gatgccac tctgacctcc 14700
aaggagagag gggaggagc ggagggaagg aaggaggagc ggagggaagg aaggaggag 14760
ggaggaggag agagagagag aagagagagc agagagagag agagagagag gatggaggc 14820
tggatctcgc aggaaggcag gcaggcaaga tgaatgtgtg tgagagctta caacaaaaga 14880
ataaaagga atgaaggagc tagctagga acaccctatc agtggaggc ttactgraca 14940
gcatgaagac ctgagtttga tcttgaacct catctagaaa ataatctagg gtgaaagga 15000
tggctcagca gtttaagaca ctgtgtttgc tcttgaattt cccagcaacc acatggtagc 15060
tcacaacctg ggaactgatg cctctctctg gcaigcaggt gtacatacag atagaacct 15120
catatattaa ataaataaat aaataatna atcttaaaaa atanaataaa actagccata 15180
gtagaacaca ctgcaatcc caaagctaaq aatgtacaga cagacagaca gacagacaga 15240
cagacaaaag atgcttaggg ctccctggcc agctcagcct actagcaag tttaagacta 15300
gtgaagacc ctgtctcaat taaaaaaaa aaacaaaaa cctttaacca cactaaaga 15360
acactaccca ggtttgtctt ctggcctcca caettgtatg tgtacacatg tgcgctgca 15420
cgacacaca cacacacaca cacacacaca cacacacaa cacagagaa aggaggagga 15480
cccaggagag gataaaaagc tggtagtca aagctgctt aaagctctg cctaatctct 15540
atgcacagag agggctgggc agaagcacca ggagcttag gccaccttag actacaagac 15600
aagatcttga aaagaacag agagctcagg gcaggagga cactacaaag tgcagagag 15660
ctgcagttg tgaigcacac ttttaatccc aggagagcca cagtaacttc tccagagtga 15720
gttcaggac agccaggctt acagagaagc ctgtctctgc agggggaana gaaaaaaaa 15780
agaaagggg ggacagaaa gaaaagata gtcgtgggg aaatgatca cgggtaaag 15840
gcatgacca caagtcac aaacctaac ttgattctg tcaccccat ggtaaaagag 15900
aacctctcc cacagttgt tctctacct ctacacatgt gcatggactg cgtatctct 15960
ctgacactcc atccatcat ccatacatcc atccatcat ccatactgt tgtgtgtgt 16020
tgtgtgtgt tgtgtgtgt tgtgtgtgt ataaatanaa aggaaggagg gaggacagc 16080
aaaagatttg ttccatgtg tgcagagcc ggaatttga cccacatcta tgtctctca 16140
gatctgaag tctcaaggac ctgtctacgc taccacacct gagaggatca ctactgcaa 16200
cagatggaag ctggggcctt ggggttcagc gaggaggagg ccagcaacc acctacagc 16260
tggctctga gccctggacc tacagcttgc ctctgaggc cggcaccctt ac 16312

```

<210> 19
 <211> 11058
 <212> DNA
 <213> 大鼠鼠属物种(Rattus sp.)

```

<400> 19
gtgagtgac cggatcact ggccctgttg ggttcggtg taggaacgg aatcgctct 60
cccgaagtc ttagagatcc ctgtctggac ttagtagcct cactacagc gtagaacga 120
agtacttaa gcccagacct tttttttt tttttctga gctgggacc gaaccagag 180
cttgccttt gctaggcaag tgcctacca ctgagctaaa tcccaaccc caggccagac 240
cttttllana ggcaattcgh ttctctttt atttatltt tagttagca tgglatltt 300
ccatgttta tgtctgaa ggagccagt gccctggagt tggagccaa gacagttgt 360
gtgtctgaa actaacaca ggtcttacc actatgccgc cccccccc cctagctta 420
gtaaaggcca tttctactt gcatagggt ctgtctatgt agaccagat gccctaac 480
taacagat tccctgccc ctgtctcag tgtctgta gtaaggtgt accccatgt 540
tggcactag actctctt tagagtgct tcaatllaa agtaacatg agtccatc 600
tgccttccc acacctcat ggccttacc ccatcactt atctctgta cgtatctta 660
caatctcat tgaatggat gaagaaatga gatgacga tngatcca aagtcctc 720

```

[0021]

acigaaacca gggcaacatt ctggccacta gggcaacatt aggcactgta	780
ggtaacacca ctcccacacg tgggaggga tgaatcagtt ggaatcacag ttccagagtt	840
tggttcctta acttgctgta tcttccatgg ctgcttctta aataggagga taatgcttcc	900
gaccccgctt ttacattttt ctccaggctg taatagagga attgagcag aagcttcgtg	960
ctgtacacac cagaggcag gatggaatag aggcactgga aactggccaa ggagcgagcc	1020
agcgggacct gtcggcaag aaacctcccg ccggtgtag aggcacagcc acgctgggcc	1080
ccacaggcag tcccctctt tctttttttt ccttgtaggt ttgttttgg ttttgaatc	1140
atgtccctct gcatattcca gtcacaggc aacagccctt gacttcgta gactcagaca	1200
aigacttgtt cacacctaa cctggcgga ccaaaccccg tggccacaa actcagcaac	1260
ggaatgcca gaggaaagc aaagtgtct tctcaaggga cagctcagtt gaagatgta	1320
agataaccc tcccacagc acacatgctt gggggggg tccagccagg ctggcctggc	1380
tgtgaacca ttctccctg cagtgcttc cccactgac agtgcctgg ctgacttca	1440
ctctgtgtgt gtgtttgtg ttgtgtcca gatcgctt gcccatctt tatgactcag	1500
ctttgttcc taataattct tgaatagaa ttccagcga gatgacgaa gaagagcac	1560
ccagaaacac caccaccat cagagacat ctggcgaag acacagctcc taggactgc	1620
cttggctat tccatctct gctggcagg gtaacttact gaatgatga gaattcaag	1680
ccccccccc ccttggaaa tacttgttgc ctcctgttt ttacatgta aatgtcttt	1740
tggagcagg ccttctctt agagccctg actgtactt tccagatcc aatcaccatg	1800
ttttctgtg aataaatgtt ttatatact tttaagcatt ctggccgctt gctgttcc	1860
atttatagc agcagatag agtccagaga gagaaccca ggcactctga ggcactctg	1920
gacttcaag gacccgtgg tgcctactt ttgcacact gctgagatt gctgtagaga	1980
gtctgtgtt ttctagctt acataagcac ttctcagct ggggtactt tgcacccag	2040
aacacgttg gcaatctct gagaccagg ggggagaga tggctcagc attagagca	2100
ctactgccc ttccagaggt cctgagttca atccacga accacatgtt ggtccacag	2160
catctgaat gatctctgt gccctctcc ggtgtgcta aagacagct tatantaat	2220
aataaatct ttaaaaaa aactagaga cctattctt ctgccacct tggggcgag	2280
ggctgctgt agcaccagc tcaatcagag cctaggagg aagactcct catcatagc	2340
actccacaa cagcatgct ggcactgca actagagggt ctgctcaga cctgcgttt	2400
gttctgtgt tgcacaccc catccagaa gatggcaaa tgaatcagc tggcgtgtg	2460
cagatcat gcatatgct ctgttgtag caaggctac ctcccagag tcatctctc	2520
actatgtaa ggggtctct gctcagctc agccacagg tgtctatgt gctgcccac	2580
ttcccagc cctaatctg ctgactact tggagcaga ggcagactg ttgttccgc	2640
tgtatgaca acagcacgt gcagctgag cttccctac ctttgtat ggaagagta	2700
aigacagcg tgggcagaa gaagatgac ctccagaa ctaggttagg ggtgactgt	2760
ggaaggctc ttgagacct tgggtaggaa aagggtaac tggtagata ggcctcagt	2820
aggaagcag aaaaaggca gctgactgt agatgccgt ctcccagca ctcacttgg	2880
gccatctca cctttccac cctgttttg gccgtctcc tgcataaa aagcttggc	2940
atggaccaca cagctgtcag tctaacaca gcttggctc tgggctcca agaggttagg	3000
gcaatccag aagggcaag agccagact agatttggg tgcagccag cgtgtccct	3060
gcctttaag caaagttat caacaggca gtaaatca gcaataggt cttcagctag	3120
aagtcaggg gctgtctc aagttcact ggcctagca agagccca ggtccctc	3180
ggccgcacc tgcgtatgc ccgacctcc caatcaacc ctccacctt ttctagctc	3240
ctacccctg aaagcaaaa agtctcttc tccctcact gtcacagag gctagggtt	3300
gcctgtctg gctgggafa cttccctctt gcttcttcc cttctgata tccactgct	3360
ttcacaggg cagccaagg aagagagtc ttggctctg ggcactcag caggtcctt	3420
ggatggggg ggaattccg cctcagctc cagctctca ccttgggtt ggggtgggtg	3480
aaaggatag tgttccaat cctggcggtt caactcag ttccggagag ggaagccct	3540
tgacctctt caacccaat cctaacagg gatgttacc cagggcctc gttataggtt	3600
cagcagttaa actcaggaa gtttgtctt aatcaggtt ccagactgca gggatcccc	3660
gacctatgg gtagaatct cctttcaag cctctagca cttcttgtt ttcatctgc	3720
ggtacacag ccagactct ggcctagag atagagcca ggcgtggg agagagaaa	3780
atgggctgc gaaggcccc caccggaga aaagtcccc cccgtctct cccacagct	3840
ttctctctg tctttttt gggcgagg ggcgtgtg ttctactacc gaagacaa	3900
gagaagatc tcaactctc cttaagcttt tcaataatg ggaagttc gatggcag	3960

[0022]

```

cotlgggcag accacacatt ctgtccatt tccclgggtc ctgcagctct cagattctcc 4020
cattttatto gggagcagc ttcttgall ctgggtctcg gatgcctcg gggacactc 4080
caggactcc tctctctaa tccatagct gtaicctcg agtctctcc tgggaacct 4140
cctccgcca ctctccggc gggacacat cagctccct cccctctca gctcgagcc 4200
cacagcttg gtcgtgacg attcaaaa caggcgggt ccaaaaagag ggggggggg 4260
aatgagagag gccacgtac tggggcttt acggggcac gtatctcagg cctctggcc 4320
ctgagctgg gactgagc ggcagcggg agggggggg cggctctca gctccccc 4380
ccatccagtt cctataata cggactgcag cctccctgg tgcctctga tctctctgt 4440
cttagagaca gccctctct ctgttgagt gccagtgaa aaaaaagga 4500
ctggcccca gggggggg gacgaagga aataggant ggggggagg acgtgtacg 4560
agtttgggg cactgctcg agctccgaa gccagccca gctctcgat tgcagagat 4620
ggtagggag gtaagggg atggggggg aatggggg agtggggg ggggctgg 4680
cagctaatc tgaattct ctcccgagc ctgtctcat agacagagc gtgaggctg 4740
gtgagcgg gtagttcc gggggggc ccgtctgt ggcctgcg gtcgtctaa 4800
attggggct gctgttagc gtgggggg gtagtggg tggcgagat acccaagg 4860
acctcaaggt cagctctg acctgggaa ggtctgact taacgggac cagctgtg 4920
ctctcttgt acctccct ttgtgggg gggggggg gtctanaga gggagagc 4980
ccctgacct ggcgcacc cgtccagat ccgaccga atccccagc tgggtctt 5040
ttcttact tggcccg agagctcag tgcagagc gggggggc cccatctc 5100
ccctctca ggtctgtg cagggggg tgggtgct accttggc atttctaa 5160
agcttact atgttact gcgcagca ttgtcaac gggagggaa tgaatgac 5220
acgttaaga aatctctt cggctctt cactttag tgtgctaa ctctctat 5280
cctctcagct ggtggagc tctttatc ttaggcag gtagtgag gagggttc 5340
cgtctctca ttaggctct actataat cgtctctg tggcttca tggctctca 5400
gccttatt cctgggcat acttatat ctgattgt ctctggag cgaagtgt 5460
ttacagagt ctactggc tccagccc ggggtggg tagcaagc gtatgcga 5520
gggggggt gggggggg gctgctat gttgtctg aaaaaat ttctgcaca 5580
aaatgctcc ggtgttagc gcccctcc tccaggtct ctactctat cccagctag 5640
cactgacca aacctatg gccagatg aaggtctat aagagatg ggtctctg 5700
agctcagga tctacctt ttcccaac cactctag aaacagac tctctctc 5760
gcctgact ggggttaat ttgctgtc gactctct tctgggtg ctgagctgg 5820
ctctctgct tctctctc tccacacac ttgtctct gctctctt tttttttt 5880
tttttttg cggggaag caggtgtt gcaatgag ctctgcta ggtctgaa 5940
atcactggt ttctgact ttcccccag aaggtctg gaggagag gttgggtg 6000
cgtctctg tctctgac ctctgctt gctctgag ctctgagc gttggtg 6060
tctctctg ggtgttagc tggctcag cctgtgga accgtctc cagctctg 6120
tgggtgggt gatgaata agtgcagc ctgtctca gggggggg aacgtctc 6180
tcttaagga cttaaatg ctgtgttg cggggaag ctgtgtgt aatctctt 6240
tggacagga aggacctc acttatag tctataaa gctctgag cctgtgtg 6300
ctcaggaat ataaattg actcttgg actcttgg ggggaact tctatatt 6360
ggtgtctat ccaagctc aacgtctt attaggaat actctgga aacatgac 6420
tctctgct tctctctt gctctgac tgaatcag taagactg cctcttct 6480
tctctgct gctctgct gctctgag ctctcagc atgtgagc gctcttct 6540
ttctgaaga acatgctg gccaggtc gactcaac caggaggg tctctgac 6600
cctctctc cctctctt tctctcag ttgtctga tggagctt gctctcag 6660
gctctctt ctgtgaca agtgacat gtgtctca agacacct ctgtctc 6720
aactatag tctctgct cagctgac tctctcag gaaagtca aggtctc 6780
aagctgaga atggagct gctctcag ggaaccca tctctctt caggtctc 6840
gtatggaac atgcagag tctctcag agtggaag ggtctctt aatctctc 6900
tctctctc ctgtctga gctctgct acatcaat ggtctctt gctctgct 6960
atgtctga gctctgct gctctcag cctctgga ggtctctt aggtctc 7020
aagctgag taagggag cctctatg ggtctcct ccaactga gctctgct 7080
tgaatcac tctctcga cagctcag tgaaggtg gctctcag gctctctt 7140
cgtctctc cgtctgct cctctgtt tgtgtgtt gaaacagc aatctgac 7200
actctcga gctctcag aatctctt gaaaccca ctgtctc cctctcga 7260

```

[0023]

```

aggtcatcca igacaacttt ggcatcgttg aagggtcat ggatgtagg caatggagac 7320
agntcatgca igagtggacc ttittitlga gaigtccctt tgggtlagg agctgacctg 7380
caagacctca cccatigcct ctatgcttcc tagaccacag tccatgcat cactgacct 7440
cagaagactg tggatggccc ctctggaaag ctgtggcgig aiggcgtgg gscagcccag 7500
aacatcatcc ctgtaiccc tgggtctgcc aaggctgtgg gcaaggicat cccagagctg 7560
aacgggaagc tcaatggcat ggcttccgt gtctctacc ccaatgiate cgttgtggat 7620
ctgacatgcc gccggagaa accgttactt agggggagag ctggatttgt ttgtctatg 7680
gtatgttgtt gacttggggc aagagcagtc atcttgggt ttgccttac aaggccaagt 7740
aigatgcat caagaaggig gigaagcagg cggccgagg scactaaag gcatctctg 7800
gtacactga ggcacaggt gtctctgtg acitcaacag caactccat tcttccact 7860
tgtatgttg gctggcat gtctcgaag acaactlgt gaagtcatt tctgttatg 7920
tgggaccgc aagctggct ctggagga acigaaggt tgggtctc tgggtggcag 7980
cttgaagag aagcccaac taaccttgt ccaatctgt tctagttatg acaatgata 8040
tggctacagc aacgggttg tggacctat ggcttcatg gcttcaagg agtaagaac 8100
ctggaccac ccagccccc aagatcttg agagccagag agagccctc agttgtgag 8160
gagtcctat cccaacttag ccccaaac tagatctc cctcaact ccatccaga 8220
cccatatac acagggggg cctggggagc ctctctct ctggaatac atcaataag 8280
ttctgtcac cctctcttg tgtgtaatg cagtgttg tgggttatg gcagtgag 8340
gaggtccat tgggtttgt gcacatctt ttatggctc tggctccag ctgggtcta 8400
gaggtccag gtacattct aacagctgc cagctcagg cgtctgcca gacactgat 8460
gltctgcca gltgaggg tgggtctga cctggaaac gaugtgag tctltgaca 8520
gcactgact tccagcttg ctctcttga gaagatct tctcttat ctgaaaca 8580
aggttaatt taaccagct tcaagctc cagatgtaa attagtgaa gttgaagg 8640
tgccagbac agtaggact attgttcca tgaagaggga aggtccctg tgcctgtt 8700
acctgtctt agtgaggc actgaaggt gcttggctc tgggcagtg actagccag 8760
gtcactgtt ggttgggaat gaattgaag tgcattcac ggtgactga ctcaagcct 8820
tcattgttt caaggtcc attcaatgt cagctctga tgcctctcc tggagatcc 8880
tagctccca ggaccaggc catgtaac ttitgttt gtctgtgaca gccatggag 8940
catctctca gctgtcact gaggagac ccagagacc cctaacctg tgcagccac 9000
acccaagtgt ggcagagct tcatctcagt ggcagcttg cccagcact ggtcttcca 9060
aagccagac cagtctact gcagtgga ggtgggagc tagctgtgag cagtgtcag 9120
gtccagccc tcaaggctt cctaggagc aaggtccat ccaaggtac atgggaac 9180
tgtttacat ggtgtcctg gccagttag accagtaca aaccaagg aaacaggcag 9240
gtccataac cacttctt acaatggaa agggaggtt ccaagtgc tggcagcatt 9300
ccaaagatg agaagtaga gataccaca tggtagga gtaggagtg ccaatgtgt 9360
tggagtgtt acaggggtg gggcgagcc tagctccga caccacaga ccttgagag 9420
gacaagcag ggcaggtt catggttg cacaagat ggtctctgc agaccacac 9480
cctgcccca gtagctctg gggcaccag ccttgctc tatctatgg agctgtaga 9540
tgagactct cagtgggt cctcagctc gcttggggt ggcggcggt cgttaagcg 9600
gacgtagtg aaagcaggag actttctaga aagaatacc agttgtacc ttggggcaaa 9660
gcaggaaagc tctgagact tccaaactg ctctcttg ccttatctt tgggtgacg 9720
gggtatgaca ggcaggagc ggggagga ggaacttg ttltcaact gtggggaga 9780
acagtllgc tcttgggac tccagatca gaaagtct aagtcacagc atggagatg 9840
tccctagcc acatnaagc cctagctct tatcagaa taattgggt tgcgggggc 9900
tggtggact gtaggagtg agatgggg agatgggg tttagggg tgttctact 9960
tgtgtact gtgaccca oattctgg tgaagcaat cctctgtct tcaattgtt 10020
tttttttt tttaattt tttattat atgagtarac tgaactgtt tccagata 10080
ccagaagagg gcatcagat ccatlacaga tgggtgtgag ccaatgtg gttgtggga 10140
atgaactca ggaactctg aagagcagtc agtctctta aacgttagc catcttcca 10200
gcccctctg cttcaagt ggtggggg gtcacaaac cagttagct gtcaccata 10260
caatgaatt tgggggtcg ggtgggtg ggttttgt ttcttllag atgagtcag 10320
tgtatctct gctggctt gaactctca cagaggcag gctgtctag tactgtgc 10380
aatcaccct cctgtgct ctgtgact gaatgaggt tatgctca taagctgac 10440
ttaactgcca ttcttggat ttgttgtt tgtgttgt gttgttgt tggggacaa 10500

```

[0024]

```

ggtttactg ctgtccagaa cagcctaaaa ctcagacaa tctttctgc tccgcttcc 10560
aagatlgaga ctgcagacat gtaccgctag gccgtgltta gagatllcta aaaaaattt 10620
agtgtgtgtg tgtgtgtgtg tgtgtgtgtg tgtgtgtgtg tgtgtgtgtg tgtgtgtgtg 10680
tctgtgtgtg tctgtgtgtg tctgtgtgtg tctgtgtgtg tctgtgtgtg tctgtgtgtg 10740
ctgagctata ttccaaaacc ccagatttac tcttaacaa ggtcaaacag cttaactgtg 10800
atttgattt ggataatcct gtcctctat tctttgtc ctgcccagt ttggctctc 10860
agacctcag agtgggac tagtcgatt cactgacgg ccaagcagg gagaggaag 10920
ttcttctag gccaccta caggactcc caaggacag tgcacatana gattgacct 10980
ggtaactcc acaggccct ggggcctgt cagctctct cctgggaga cagccagcc 11040
atcaggcgc gctcttac 11058

```

```

<210> 20
<211> 3158
<212> DNA
<213> A

```

```

<400> 20
agcttggat tgtccagca ggtagccaga gacgcccac agccaagaa ccatccactg 60
gtacgttagg cagcctgtgc gggcgagacc agactggccc ctcctctct gcagtgatt 120
gttcttctt ctttlttaa tccagtttc ctgcttttc taggtctag gtaccagct 180
ctgcttcta cagctcaga caagacttt gtcacaccg agcccccgc tactaccgt 240
cgcatccaa acaccagca ggcagcttcc aaaaagaac ccaagttgt ctctccagt 300
gatgagcca gtgaggaag tatgatgtc cgcctgttc cgcctcaga aggcacacg 360
ctagggtga gaggcctgt tccatagga ctgtctcgg ggccttagt gtagatgtc 420
tgcccactg ccgcagaag gccctctctg tacagcttg atttatttc ttctgtcgg 480
tggttgattg tctcattgt tctcagat ctatttttc accatcttg tgactcagct 540
tttcttatt ctttaatic ttgcataga tctttcagc gagatgacg aagccagac 600
accagaana acaactcca ttctcagac atcggtctgc aggcacagat cctaggagt 660
ctgtctctgt cctcccgtg cagggtacc tgiagggtga cctggaatic gaattctgt 720
tccctgtaa aatattgtc tgcctcttt ttttaaaaa aaaaaggcc ggcactgtg 780
gtcacgctt gtaatccag cactttcga taccagggc ggtgataac ctgaggtagg 840
gagltcaga ccagctgac caacatggag aaacccacc tctactaaa aaaaaaatt 900
agccggcgt atggcgctg gccgtgacc ccagctacc aagagctga ggcaggagaa 960
tcgctgaac ccagagggg aggtttagt ggcgggaat cacaccatg cactccagct 1020
tggaacaaa tagcaacct ccactcaaa ttaaaaaaa aatgctaca cgtctttaa 1080
aatgaaggg ttctcttan atagcctaa ctgactcgg ttgagctgt tcaactttg 1140
aatatagtt tgcaacttc ctgttttct aatgaatana tgttttata tactttaga 1200
catttttc taugctgtc ttgtttct ctctcatt agcccgltt catgcagag 1260
agagaggtt atcagtcag agagagatga gtgagccag agtcttagg cgtctccgg 1320
gatggcagat gacttctg ccccgtaact gcaacttcc cctctcacc ctctggacc 1380
tgcacagtga ccagacgcc tctctggga gaattatga gtcctaggc tccagatcag 1440
tgtcttga cgggggcaa ttgtctgc cagagacat ctgacacac ctgggcctg 1500
ttgttgtc atagctaln ggggaagaat gctacagca ttgtggan ggyccagg 1560
atgtgtca acatctga gtcacagga tggccctcc acanagaac acagggcca 1620
caatgcaat agctcagc ttgaganaac ctgtctaga ccaagggtg ctctcgcg 1680
tggtctcac ccaccccca ctgtctcc ctatcccat ctccaaggt tggcagaga 1740
ccggccagg ctgttgga ttcatcat gactccctc agctctcag gacacagac 1800
ctgtctcca gattctca ttgtcccg gtctcagag ggcgtagct atgtcagg 1860
cggctgccc cctgaagct ggcacccct ctcagctcc tcaagtctc tctgtagt 1920
cactctgaa ccgagggctg tgagctggct gctgtacca cactgtgca tctgtgtca 1980
tgacaacag acactagtc agtagtctc ctgggcact gactctctc ttgcggga 2040
gaagacagta ataaaaatg acagctga ggcagaggg aagctacgc ttgggtgtg 2100
caggagcgt gaggtctct taatgtctc aatganaag gctaacggt ctgtgtca 2160
ggaatagct agtcagagt gggcttct ccactcccc acccaccct ttaccatt 2220
aggaacttc ttgcttct ctgtctctc tctctaggt gctattgtg aaaaagcgc 2280
accaactag ccagtcag ccagcaggg acacagctg gctctagtc ccagcagaa 2340
aggaatccc agaaagcag gctcaggaac tgggtctgt tgggtgttt ttangcaag 2400
attatccca ggcagctaa actagcac cgtttttag ctgaagggc agggacttg 2460

```

[0025]

```

tgccagggtta tgcctgggcca gcaaaagagc ccgggattccc cctcccatgc acctgcgat 2520
gggctcagagg caccaccacc caccacccttc ctacaaagtg ttacagaccr tccatccca 2580
cactacacaa cctggccctc tggcctctta cagaagaaat ggatccctcg tggaggggg 2640
caggggaccc gtccaccag tggcccaag acctcttttc ccacttkttc cctcttcitg 2700
atccaccctg cctcaaat ctcccggggc agccagigna agggagtcct tggctccgg 2760
ctgcctcga cgtccaccgg cggggaggga ctcccgccct cactcccgcc tcttggccc 2820
aggctggatg gaatgaagg cacactgict ctctccctag gcagccagc ccacaggttt 2880
ccaggagtc ctttgggga ggcctcagg ccccaccag ccctccctg ctcggctgg 2940
ggctccagcc cggagagagc cgtgggtgca cacagggcgg ggaattctcg cctcaatat 3000
caggctcagg ctacagggt gcaggacatc gtgaccttc gtgcagaaa ctccctctc 3060
ctcfaagcc gctcccgag cctcttctct ctccaggccc ccagtccca gtcctcagtg 3120
ccagctccag gctcgggtc cagagatgcc agggacca 3158

<210> 21
<211> 3218
<212> DNA
<213> 人

<400> 21
aagctggcac cactactica gagaacagg ccttttcctc tctctgctcc agtcttggc 60
tatctgtgtg tggcraaca tggaaagac tatctgtg gcagccccag gtaggtgac 120
aggttaggga atccagagct cgcactgggc ttgagcctg actggttact gtagcttagc 180
ctggagctga gctgcaggg gcaattccag ctggccctcc gcagcttga ggtcttggc 240
acgtctctta ttgctttcig tggcctcgtg tcttctcga ggaatcig gctagccct 300
aaggtctca agcaggatc atctaggtaa accagttac taaaacatg cccaagcgg 360
taagacttat ataagttta aaaatggta aaaatgccc cctgcctag tttagggaa 420
gaagactga gatgttcag ggtgacttat tccatcctc gtccttaggc gaacttggg 480
agggcgaag cgtgtagctg ggaacttagt ccagaccctt gctctcga ctgaacgct 540
cagttgcttt ggcagttac tccggggctt cactttgac gttgtctac ctatggaga 600
caaaagtata taccctggta gaggcgcac gctgtaac ccagatctt gggaggcca 660
gtaggtgta tccctgagc tccagatct gagccagcc tggccaact ggtgaactc 720
gtctctactt aaaaattaca aaacagcca ggtctcag ccagatgcta tagtccagc 780
cacaggcatg ctgaagcagg agaactgctt gaccccgga ggcagaggtt gcatgagct 840
gagaccacac cctgcacac cagcttagc aacagatct gagactcat ctcaaaaaa 900
aaaaaagtac ctactcaga gttaaaacta gtaataata ggaagtctt gacacagta 960
caccaaagtg cacaataaat ctctgccagt tcaattata taaagaatc cattgaatg 1020
ttagctcaat accgctctc ctaccagagc atgtgaact gcactctcc agctctcca 1080
ggctttcca agaactcagg acacttagc ctgttggctt cagtctatg cagacagga 1140
ggaagcact cttagctga tacttaaca gacacctga ggcacatac acccgagac 1200
acaigcatgg agcttcactt tctctgcat ttgtagtga ccaggagagc aagagctcc 1260
acctcccttc aaacacigt gccatccc ggcactaagc cctctttaa gacagcacc 1320
tccagaggg agggccacag ccacatac tccacttggc aggtggacag cgtgagcac 1380
tggacctag cagggaacag gtgcccgc cagcccaac gccctctgc cgtgacagg 1440
acagaagccc tctccagct cgtgtctgc agagccatg cgtgctccc agctgcatc 1500
tatccactc cagtgcctg gccagttagc accagtgtg aagacagta gctgtctcc 1560
gacaacagg atggagaaa ggtcccatc tccattctt gatactga caagtgagg 1620
ggccgaatc gctctgag catttaag atgggaagt agcagaacc cagctgaa 1680
ggcagggag ccccaactg ggtgaaatg gcccagaat ggttaggcca agcttagct 1740
cagacccc agagcttgg agagccaag ctgaggag agagcttag gtaggagcc 1800
cccgtccc agggccggc ctgttcaga gctgagctg atgttccc ctgtgcagc 1860
ccacctctg cctccttag ctccctctg ccagggctc ggtgcagg gtagggcag 1920
tctctctc atggagctg cagatgagc atcgatcg gactctag cctcgttgg 1980
ggcgccggc ggtctgtaa ggggaacc agigaagca ggaacttc laggaaaaa 2040
ccacagttg caacctggg gcagcaggc atcttaga cggacggac tctctctc 2100
gtgctcacc cctctgag ccgtgagaa gtactgga gtagggcagg gccgggat 2160
ggcgggggg cggcaggac tgaactctt ccaacccc cctgacagg aatggccc 2220
cgtctgtc ttgggaact agagcttag gtaggtag atggctcag cctgtaatt 2280
cagcacttg gaggcagag ccgggtgat cagagctca ggaattca gcaagctgg 2340

```

[0026]

ccacacaggt gacaccccat cctactaaa aatgcacaaa ttagccaggt gtgggtggg	2409
ggccttggg gctggagga agataatcac ttgaaacagg gaggggagg ttgtagtgg	2480
ccaaaaaaa aaaaaaaga aalagctgaa gtcacagtag gagagaagct gctgagctc	2520
cagcaccgag acctcaggcg clggctttaa tctctatcag cagatcttct gtagtcttct	2580
aaaattcact ttctgttgc ggtgtaactt acacaggctc aaatgcacaa atcatggccc	2640
tgaatttaga taaaaatcag ccccccacac acttttgttc ctggccaggt tttaaatgc	2700
ctctaacccg gggacccacc agagctgggt gccctggagc gcttcagcca tccgtcatg	2760
aatggacata gctcctccaa ctgcacaggc agagagctgt gggcttgggc cagcccccac	2820
aggtaatccc caaaggcgag ccccacagca agatgtgaaa cagtcattgc ctgagggtct	2880
ctgggctgt gtcccaacct ttctcccgcc tggctccccc tggaaagccc catgccagg	2940
ggaggcgcct accggtctcc agactgggtg aagagccggt ggcaggctcc caictgacg	3000
tccaggccgc gcttcagctt gcacatctcc atgtactcgt gtaggtgcgg gttcargtct	3060
ttcttgcccg tggccaactc cagctgggtt ggggcagcca gggccatcgg gattaacag	3120
tggcttcac agcgcctcgt ttgcccgcgc caggagccca acacgccaag agcagtggct	3180
gggcccgggg ccaggccagc cattactgaa ggtcaga	3218
<210> 22	
<211> 508	
<212> DNA	
<213> 人	
<400> 22	
agcttgagat tctcaagca ggtagccaga gaggccatc agccagaaa caatccatg	60
gtacgtacgg cagcctgtgc gggcagacc agacaggcc ctccctctct gaagtattt	120
gtttctctct ctttttttaa tcaatcttct ctgccttct taggtcttag gtaaccgct	180
ctggcttctc cagctcaga caatgacitl gtcacacag agcccccgc tatccctgt	240
cggctctcaa acaccacga cggagcttcc aaaaaaaac ccaagtgtt ctctcaagt	300
gatgagctca gtaggaagg taigatgct cgcctgtct caggccaga aggcacacg	360
ctagggtgca gaggctgggt ttccatagga ctgtctggc gggcctgagt gtagatgct	420
tgcaccatgt ccgcagaagg gctctctct tagatcttg attttattc ttctgtggg	480
tgtggatttg tctacttgt tctctgat	508
<210> 23	
<211> 2850	
<212> DNA	
<213> 人	
<400> 23	
atctattttt taacctcttt tggactcag ctctttctta tcccttaatt tctttgata	60
gatctttcag cagagatgac agagacagag acacccaaga aaatactcc cattctaga	120
gcctggcttc gaggccacag atcttaggaa gtcgttctt gctctccctg tgcagggtat	180
ctgttaggtt gacciggaat tcaatctcgt ttctcttctt aaataattg tctgtctct	240
tttttttaaa aaaaaaagg ccggccatgt tggctcagc ctgtatccc agcatttgc	300
gataccaagg cgggtggata acctgaggtt ggcagttcga gcccagctg accacatgt	360
agaaacccca tctctactaa aaaaaaaa ttacccgggc gctctgggtt ggcctgttaa	420
tcccagctac tcaagggtct gaggcaggag aaicgcctga acccagaggc ggcgttgta	480
gtgagccgaa atcacacatc tgcactccag ctgggcaac aatagcgaa ctccatctca	540
aataaaaaa aaatgctca cagctctttt aaatgcaag gctctctctt aaattagct	600
aactgaactg cgttagcttg ctctcaactt ggaatatag ttgccaatc tcttgcttt	660
ctaatgata atgtttttta tatcttttta gacatttttt cctaatcttg tcttgcttt	720
atctttcaca ttacccctgt ttcatgcagc agagagaggc ttatcagctc agagagat	780
gagtagccgc agatctctag gctctctcc ggcattggc agagcttcc tgcctctca	840
ctgcacctt tccctctca acctcaggac ctgcacagt gaccagacag cctctctgg	900
gagaattatg cagtgcctag gctccagatc agtctctctg aaccgggggc aatttgtct	960
gccagaggac atctgacac acctggggcc tgtttgttg tcatagctta taggggaaga	1020
atgtaccag catttgtggg aagagccag gcatgtgct caacatctg cagtgcacg	1080
gatggccctt caacaaagaa tcaacggcc caaatgtca atagctcaa agttgagaaa	1140
acctgtctca gaccaagggt tctttctgc cgtgtgcttc acccaccac cactgtgtt	1200
ccctaatccc atctcaagg gttgcagca gaccgctca ggtctgtgga agttcagatc	1260
atgattcctt cagctctgc agagacacag acctgtctcc cagcatctct cattgtctcc	1320
gggtctgcag agggctgag ctatgtctca gggggctgc cccctgaagc ctgcagacc	1380
ctctccagct cctcaagtct tctctgtgta gtaaccttg aaccggaggc tgtgagctg	1440

[0027]

```

ctgtctgac cacactggg cctctggtgt caigacaaca gcaactacg tcaatagtc 1560
tccctgggca ctgagctccc tcttctgggg gaggagacag taatgaaaa tgacaagcat 1560
gaggcagagg ggaagatcac gcttgggtgg tgcaggagca tggagtgct ctatagctc 1620
tcaatgagna aggttaacg gicctgggtt caggaatagc tgggtcagag gtggggcttc 1680
ctccactccc cccccccc cctttcacca ttaggagcct tcttgcttg ctctgtctc 1740
ctgtcttgg gttgtatg tgaatagccc gcaacaccca tgcctatgg agtcagagc 1800
ggacacagtc tggcttggg tccagagagg aaaggcaac cagaaaggc aggttcagg 1860
actggagtc ttgggtgct tttaagcaa agattatcac caggcagct aaacttagc 1920
accgctttt agctagaagc gcaggagcct gggtcaggt tatgttggc cagaaagag 1980
gtccgggata cccctccat gcactgtct atgggccaag gccacccac ccccccct 2040
tcttacaag ttctcagac cctccatcc cactctaca aacctggcc tctgctctc 2100
taccagaaga atggtcccc tgtgggagg ggcaggagac ctgtccac cgtgtgcca 2160
agacctttt tccacttt tcccttct tgcctaccc tgcctcaat atcccagc 2220
gagctcagtg aaaggagtc cctggctct ggctgcctg caggtccac ggcgggagg 2280
gacttcgac ctacgtcc gctcttccc ccaggctgga tggatgaaa ggcacactg 2340
ctctctccc aggcagcac gccacaggt tccagaggt gcttltggg gaggctctg 2400
ggccctcac agcatctg tctccgctt gggctccag cccgagaga gcgttggt 2460
cacataggg cggatctg tgcctaat atcagctca gctacaggc ctgaggata 2520
ctgtgactt cgtgcagaa cctccctct cccctcag cgtctctcc agctctctc 2580
ctctcagtc cccagtcg cagtcctag tgcctccc aggtctggt cccagagtg 2640
ccaggagcca 2660

<210> 24
<211> 1963
<212> DNA
<213> 人

<400> 24
agcttgagat tctcaagca ggtagccaga gaggccatc agccaagaa cctccactg 60
gtactaagg cagctctgc gggcagccc agactggcc ctcccctct gcagtgttt 120
gtttctctt ctttttaaa tcaatcttc ctgcttttc taggtctag gtaccagct 180
ctggcttcta cagctcaga caatgacttt gtcacccag agcccgccg taatccctg 240
tggctccaa acacccagc gagagctcc aaaaagaaa ccaagtgt ctctcaagt 300
gatgctcca gtgaggaag tatgctgc ccgctgttc ccggcgaga aggcacacg 360
ctagggtgca gaggctgtt tctatagga ctgtctggc gggctgagt gtatgctc 420
tgcctcactg ccgagagag gctctctctg taccgttg atttttttc tctgttggc 480
tgtggattg ctctactgt tctctgatat ctatttttc accatottt tgcctcagct 540
ttttctatt ctttaattc ttgcataga tctttcaga gcatgacag aagacagac 600
accaagaaa acactccca tctcagagc atggctcgc aggcacagat cctaggaggt 660
ctgttctgt cctccctg caggatctc ttaggggta cctagatct gaattctgt 720
tccctgttaa aatattgtc tgtctcttt ttttaaaaa aaaaaggcc gggctgtg 780
gtctcagct gtaatccag caatttgca taccagggc ggtgataac ctgagtagg 840
gagttcaga ctacgtgac caataggag aaacctcac tctactaaa aiaaaaaat 900
agccgggct atggcgctg gctgtaac ccagctact aagaggtga ggcaggaga 960
tgcctgaac ccagagggc aggttgtagt gaggcgaat cagacattg cactcagct 1020
tggcaacaa tagcaact ccatctcaa ttaaaaaaa atgcttaca cgtctttaa 1080
aatcaagc ttctcttaa attagctaa ctgaactgc ttgagctgt tcaacttgg 1140
aatatagtt tgcctctc ctgtttctt aaagataaa tgttttata tactttaga 1200
ctatttttc taagctgtc ttgtttcat ctctcatt agccaggtt catgagcag 1260
agagagggt atcagtgag agagagaiga gtgagccag agtctaggc cctgtcccg 1320
gatggcagat gacttctg cccgtact gcaccttc cctctcac ctctggacc 1380
tgcacagta ccagcagcc tcttggggg gaattatga gtctctagg tccagatcag 1440
tgttttgaa cgggggcaa tttgtctg cagaggacat ctgacacac ctgggctct 1500
tttgttgct atagctata ggggaagat gctaccaga ttgtgggaa gaggccagg 1560
atgtgctca acatctgca gtgcacaga tggccctca acaaagaac acagggcca 1620
caaigtcat agctctacg ttgagaaac ctgctctaga ccaagggtg ctctctctg 1680
tgtctctac ccccccac ctgtgttcc ctcaacct ctccaaagt tggcagcaga 1740
ccggccagg ctgttggaag ttcagatcat gctccctcc agctctcag gagacaagc 1800

```

[0028]

ctgtctccca gaattctcca ttgttccagg gtctgacagg ggcgtgagct atgtctcagg 1860
 cgggttgcgc cctgaagcct ggcgacccct ctccagctcc tcaagcttcc tcgtctgagt 1920
 caacttccaa ccggagcttg tgagctagct gtctgaccca cac 1963

<210> 25
 <211> 1195
 <212> DNA
 <213> 人

<400> 25
 tgggtccctc gtctctatga caacagcaca ctacgtcagt agtctccctt ggcacatgag 60
 ctccctcttt gggggagaa gacatgaatg aaaaatgaca agcatgaggg agagggagag 120
 atcagcttg ggtggtcag gacatggag gtctctttaa tgccttcaat gagaaggggt 180
 taacgtctct ggttcagga atagctgagt cagagtaggg gtctctccca ctccctcacc 240
 ccaccccttt caccattagg gacctcttg ccttgccttt gctactctgc tcgggttggt 300
 catgttgaaa agcccgacc aacctatcca gtggcagcca gacgaggaca cagcttggt 360
 ctgggtccca gagggaagc caatccaga aaggtaaggt taggactag agtcttgtag 420
 gtctctttaa agcaagatt ataccaggg aggttaact tagcaaccgg ctlttagcta 480
 gaaggcagg ggcctgggt caggttatgc tggccagaa aagagggcg ggaacctct 540
 ccctatgccc tgcgtaggg tcaaggccac cccacccac ccccttcttt acaagtgtt 600
 agcaccctcc catccacac tcaacaacct ggcctctgc cctcttacca gaagatgga 660
 tccctgtgg gaggggcag ggcacctgtt cccacctgt gcccaagac tcttttccca 720
 ctcttctctt ctcttgact caccctgccc tcaatatacc ccgggcagc cagtgaaggg 780
 gagtccctgg ctccctggct gcttgcagt cccaggcagg ggaaggactt ccgtctctac 840
 gtccgcctct tcgccaggg ctggatggaa tgaaggcac actgtctctc tccctaggca 900
 gcacagccca caggltcca gtagtgcttt tggggaggg ctctggggcc ccaccagcca 960
 tccgtctctc cgcctgggg cccagccgg agagagccgc tggtcacac agggccggga 1020
 ttgtctgccc taattatcag gtccaggcta cagggtctga ggaatctgt acctccgtg 1080
 cagaaccttc cccctcccc tcaagccgcc tccagagct ccttctcttc caggccctca 1140
 gtgccctgt cccagtgccc agcccaggcc tcggctccag agatgccagg agcca 1195

<210> 26
 <211> 256
 <212> DNA
 <213> 人

<400> 26
 cctctggccc cccaccagcc atctgtctct cgcctagggg cccaccgcc gagagagccg 60
 ctgtgtcaca caggccggg atgtctgccc cttaattatc ggtccaggct acaggctgc 120
 aggacatctg gaccttctgt gcagaacct cccctcccc ctcaagccgc ctcccgagcc 180
 tccctctct ctagggcccc agtgcctagt gccagtgcc cagccaggcc ctctgtccca 240
 gagatgccag gagcca 256

<210> 27
 <211> 1941
 <212> DNA
 <213> 人

<400> 27
 agcttgagat tgtccagca ggtagtcaga gacgccatc agccagaaaa ccatccatg 60
 gtacgttaag cagctgtgc gggcagacc agactggccc ctccctctctt gcagtattt 120
 gttcttctt ctctttttaa taactgttcc ctgccttttc taggttttag gtaccagct 180
 ctggttctta cagctcaga caatgacttt gtacaccag agccagcccg tactaccgt 240
 ccgcatccaa acaccagca ggcagctcc aaaaagaac ccaagtgtt ctctcaagt 300
 gatgagtcga gtgaggaagg taatgcttc cgcctgttcc ccggccagaa aggcacacag 360
 ctaggatgca gaggcttgt ttccatagga cctgtctgg ggccttgagt gtatgtctc 420
 tgcctcactg ccgcagaagc gctctctctg taacgttgg attttatct ttctgtcgg 480
 tgtgggattg tctcacttgh tctctgatat ctatttttcc accatctttg tgactcagct 540
 ttctcttatt cctttatct ttgcataga tctttcagca gcatgacag aagccagac 600
 accaagaaa acaactccca ttctcagacc atcggtctgc aggcacagat cctaggaagt 660
 ctgttctgt cctctctgt cagggtatcc tgtagggtga cctgaattc gaacggagg 720
 ctgtgagctg gctgtctga ccacactagt gctctgtctg taatgacac agcacctac 780
 gtctagctgt cctctgggc actgacttcc ctcttgggg ggaagagca gtaatgaaa 840
 atgacaagca taagcagag ggaagatca cgttgggtg gtccagagc atgaggtgc 900
 tcttantgt ctcaatgaa aagggttaac gctctgggtt gcaagatag ctgactcaga 960

[0029]

```

gggggggtt cctccatcc cccacccac cctttccac attagggaac ttcttgctt 1020
gtcttgcta ctctgtctg ggtgtcatt gggaaagcc cgcaccacc atgcagatg 1080
cagccagagc agcacacag cggctctgg gcccagcag gaaaggcaat cccagaagg 1140
caggtcagg gcctggagc cgtgggtgc tttttangc angattatc ccagctagg 1200
taactttag aaccggctt tagctagaag ggcaggggc tgggttagg ttatgtagg 1260
ccagcaaga ggcggggat cccctccca tgcactgtc gatgggcaa ggcacccca 1320
ccaccccc ttcttaccn gttttcaga ccttccacc ccacacacc aaacttggc 1380
ctctgtctc ctaccgaag aatgatccc ctgtgggag ggcagggga ccgttccca 1440
ccgtgtccc aagactctt ttccacttt tccctcttc ttgactacc ctgacctca 1500
tatccccgg cgcagccag gaaaggagt ccttggtcc tggctgctt gcactccca 1560
ggcagggag ggaattccc cctcagtc cgtctctgc ccaggttgg atggaaiga 1620
agcacactg tctctccc taggcagac agccacagg ttccaggag tgccttctg 1680
ggagctctt ggcccacc cagctactt gtcctccc tggggccca gcccggag 1740
agcctggtt gcacacagg ccggatgtt cgcctaat tatcagtc aggtacagg 1800
gtgcaggac atcgtgact tcgtgcaga aacctcccc tcccccaa ggcctccc 1860
gagctctct ctctccagg cctccagtc ccagtccca gtccagagc caggctcgg 1920
tccagagat gccaggacc a 1941

```

<210> 28
 <211> 1433
 <212> DNA
 <213> 人

```

<400> 28
atctatttt traccatct tgtactcag cttttctta ttctttaat tcttgata 60
gatttttcg cagagagac agagacagc acaccacga aaacacccc cattctcga 120
gcatggctc gaggcacag atctaggaa gtcgttctt gtcctcttg tgcaggat 180
cctgtaggt gactggaat tgaacccga gctgtgagt tggctgctt gacacactg 240
gtgctctgc tctatgacn acagcactt agtcagtag tgcctcttg gacagagct 300
cctctttgc gggagagag cagtaagaa aatgacagc catgagcag agggagat 360
cagcttggg tggtagagc catgaggtt gctttaatg ctctcaatg gaagaggta 420
aggtctctg ttgaggaat agctagaca gaggaggga ttctctact cccctcccc 480
acccttcca ccttagggc cctcttggc ttgctcttg tactatgct tgggtgcta 540
ttggaagag cccgaccaa ccatgccagt ggcagccga cagagacac gcttggctt 600
gggtccacc aggaaggca atccagaaa ggcaggctc gggatggag tctgtgggt 660
gttttttag ccaagttat cccaggcag gctaacctt gcaacggct tttagctaga 720
agggcaggg gctgtgcta gcttatgct ggcagcaaa gggccccag atccctccc 780
catgacctg ctgtgggac aaggccccc caccacccc cctctctac aagtgtcag 840
caccctcca tccacacac acaacctgg cctctgccc tctactaga agatggatc 900
cctgtggga ggggcaggg gactgttcc cactgtgic ccaagacct ttttccact 960
tttctctct tcttgctca ccttgcctc aatatcccc ggcagccca gtgaaggga 1020
gtctgtgct cctgtctgc ctgacgtcc caggcgggg aggaacttc gcccacgt 1080
ccgctcttc gcccaggct gcatggaat aaagccacc tctctcttc cctaggcag 1140
acagcttca gttttcagg agtgccttg tggaggcct tggggcccc accagctat 1200
ctgtctctg cctgggccc cagccggag agactcctg gtacacacg ggcaggatt 1260
gttgcctca attatcagt ccaggctca ggtgtcagg acatctgac ctctctgca 1320
gaaactccc cctcccttc aagcgcctc ccagctctc ttctctcca ggcctccat 1380
gcccagttc cagtgcagc cccagcttc gttccagag atccaggag tca 1440

```

<210> 29
 <211> 17136
 <212> DNA
 <213> 灰色鼠(Cricetulus griseus)

<220>
 <221> misc_feature
 <222> (4674)..(4912)
 <223> n 是 a、c、g 或 t

<220>
 <221> misc_feature
 <222> (11347)..(13938)
 <223> n 是 a、c、g 或 t

<220>
 <221> misc_feature

<222> (14757)..(14870)
 <223> n 是 a、c、g 或 t
 <400> 29
 gtgggtgtg cagggtgact ggccctatgg ggctccgtgt cagtgatcac agtctgtttt 60
 tggagtgtg tgggactctt aacttacgtc cgtgtgtggc tcactacaga catttaacta 120
 gagtgacttg agatracitl ctatgtccac ttcccggtcc taaaaaaaa atttatittg 180
 tatacattag tattttgcct gtgtgtatgt ccatgtgaag gtgcagatc tggaaattga 240
 gtgcaggca gtgtgagct gccacatggg tgcctgggac tgaacccagg tcttctgcag 300
 ggcagcagag tgcctctctt ctgcgcctca gtaaaagcag atttctagtt gagaataggt 360
 ctltctgtat agctctagct ggctttagat ttgttacctt gacacagctg gcccacattt 420
 cagattcccc tgcctctgac tctatgtggt tggtaggaaa ggcgtgagtc acatgtctgg 480
 cactagacta gctttgtgtt tagagcaaaa ctatgtatgt agagtcccg ttgttccctg 540
 tctctgccc catcatcttc ataccttgg gtactgtgat ctacacatct acagtgaact 600
 ggaatgaatg gtgtgagtga caatgaatga cagagcagta cccagagtc acagtgcaca 660
 ctgatactgg gtctctctga cagtggcaca tacttgcac caccagtcca tactggacat 720
 ttctatgacc ctatggccac atcacccttg tgggtgact cagttgtaat ttccacctat 780
 atagtgtttt caggttggct ttltacttag aatataacat ttatattatg ttcttttttt 840
 ggttttgagt ttgtttttc gagacaggtt ttctcttat tgccttggag gtgtgtctgg 900
 aacttgcctt ggagccagg cgtgtcttga acacacagag atcttccac ctctgctccc 960
 cgagtctgg gatlaaagc gtgcgcacac aatgcctcag catctttctg ttctctgttt 1020
 tttaagagg gagataatgt ttccagtgtt gtttttaact ttltctctgt aggttacctt 1080
 agatgaattt ggcgcagcgt ttccgctatg tcttaccaga gcatgtgatg gaatagagga 1140
 gcttgaacct ggcacagcag ggcgcagcga agccctgtct gcaaaagac cctctgtatg 1200
 tatgagagac aagccacac ggcctccca caaatcaccc ctgtcatlla ttgtttttt 1260
 tttttaaat caaatctcc tgcatttctt agtctcagg cgaagcttct tgaattctat 1320
 agatctcagc atgtactttg tcatacctaa gctctgagc accaacctgt atcttccaaa 1380
 cactcagcaa cgaagtcac agaagaagc caaagtgttc ttctcaagt atgactccag 1440
 tgaggttggt atgtctgccc ttcccgag cgtgtgtgtt gctttgggtt ctcttgggg 1500
 ctgtctgtgt tgtagagta ttcttctggt taatgtgttc cccactgtgc acgttggctc 1560
 accttgcag ctctatgttt taagatccac ttacacacat ctttgtgact agatttttgt 1620
 tctataatt ctttgaag aactttcag ggaatgagc gaagcagaga ctcccaagag 1680
 aacccccc atcccgat catctgccc aagacacag ttctaggagt tgcctacctt 1740
 attccctccc ctgtgtgca ggttacttla atgtatttt aattcatctt ctccctgtta 1800
 aaaaagtgc ctcttgttt ttatattgta aatgtcttt gcatctagag cttttctctt 1860
 agagcccttc tgcctgtgc tttaatttc caacaacac ttacacatc acctgttttt 1920
 ctatgaatna aigtttttat atacttttat tcttctgtgg tctgtttcat ctctccactt 1980
 agcccgattt caggcagcaa tagaggglla tcaaacaga gagaagatc acagccccc 2040
 gcatcttgag ctgtctctg ggtcttcatg ttctctcatl gattcatctt gaaacttga 2100
 tgagcttcca ctgtagaga ttctctcttc ttagctgtgg ggtactttct acccagagac 2160
 ctatttctgt ctatgctttt tggacacat aaagcttctt tctgtctag aacacccac 2220
 aacagtgtat caggagtga caagtgcac aacaatgtc tagactaac gttttcttcc 2280
 tctgtgtccc tcatccaga agatctatg actccagga gtacagcttt actatgggt 2340
 ctgcaggccc gtcacgccc ttaggggcaa gccatgccca ccttcccaa gtcttaate 2400
 tgcagctca ctggaacag gacacatga ttctgtgttc atgacaacag cactttgcca 2460
 tagaatgtct cctatctct taatgtgtgt gtgggggac agtaatgaca aacctaggt 2520
 agggggcaaa agggagaggt gacacttaga aacatgttt aggttagggc agaatatgg 2580
 aggggtctct gacactttt gattggaaag ggttaatgtt gctctgaaa cctctgttgg 2640
 aaggcagaaa agggcaggg ctgagtccc gcatgggac catttccatc ctctgcatcc 2700
 tgcctccggc tcatggaaag ctgggcatg gccacacag ctgtactgtt tggctctggg 2760
 gcccacagga ggttgggcaa tctcagaatg gcaaggagcc aggaatgtat ttgggttga 2820
 gccagcctg ctccctgctt tttaagcaaa gttatctcc aggcacgtta aacttagcaa 2880
 ttgggtcttt cagctaaag agcaggggtt tggctcag ttgtactgac ctatgcaaga 2940
 ggcacagga tccctctgct cagcactgtt ggtgagctc caagccttt cccgagagct 3000
 caggatccac ctttccaccc ctccctactc tttagaggag gaaccccttt tctcttccc 3060
 actttgttga ggggttggg gccagctgtt ttggtcttgg ggtataatca cccctacccc 3120
 ttctacttcc cctctctctc agacctcacc ctgcctccac gagggcagcc aagaaggag 3180

[0031]

```

agtcctcggc tgcaggccca gtaggcacgt cccaggacag gtagggactt ccgcccacac 3240
gtcagctctt ccgcccgggg gctgnagtgg gtgaaggagg cagtgtctcc tagcctgggg 3300
ggtgaacccc tcaggttccg agggagacag ctctgggagg ctctcttggc tctcccaacc 3360
caaccacaca ccagacatt gtcctacccc cggggcccca acctagacct taactaggga 3420
acacagaggc cagtttgtaa gctcaatta tgcaggacat ccgacctgtt ggctaggga 3480
gcgcccctcc aggcgcttc cctagcctcc tctggccct caccagcccag gcctctggcc 3540
caagaaagg aagtcggggt ggggatgga actgcaatg cgaaggccc ccgaggagg 3600
caaatgacc cctccgggc ctctctgtgt ccgagacttg ttttgcctg tgcactacc 3660
gaagaaccac gagaagatcc tcaacttlt caccagcttt gaataaagg gagagggtg 3720
cgggtgcagg tgggcacac acgcacttct gctcaaccg ccccccacc ccccgcttc 3780
tgtctctcc caggttctcc ccttttata gggcgagca ctcttagtc ctgggtctt 3840
ggaagctctt agtacacat ctctgtctt aagtcctag tctgtatcc ctgggtctta 3900
tctgtctcc cctaccggg tcactctcc agcgaagca tctcgttcc cctcccttc 3960
tcagcccga gccacacgt ttggtgtgt cacattcca aaacgagcg ggtcaaga 4020
gagaggttg agagtgccc agtgccacg ctactccgg ctctacagt gcactagct 4080
caggtctag cgtcttgag ctgactgg atggaigag gggcgagag cggggcag 4140
cgtctcggc gctccccc accccagtc ctataatcc ggaatgcgc cctcccggt 4200
gtctctgtt cctctctgt ctgagacag ccgcacttt ccgtcagtg ccaggigaa 4260
accgaggtt gggcgagag tggccggga cgttcgaaa cggggaagg ggcgtctag 4320
ccgagactg cggcgcttg ggcagctcc actgccgag cccggctcc gcatgcag 4380
ggcggaggg ggaaggatg ggggcagg cggggcag cgtacctct ctctctctc 4440
tctctctcc cagctctgt ccgagacgc aatggtaga gtggcgta accgtgagt 4500
tctgtctg gctaggtag ggtccgggt cccgtccgt cgtctatga ggtctaccc 4560
acccggggc tctcggag cgtggtag cggtagtg gcccgagcc ccaggagac 4620
ctcaagtc gtagccacc tctctcttg cgggatgag cagcgagag tccnnnnnn 4680
nnnnnnnn nnnnnnnn nnnnnnnn nnnnnnnn nnnnnnnn nnnnnnnn 4740
nnnnnnnn nnnnnnnn nnnnnnnn nnnnnnnn nnnnnnnn nnnnnnnn 4800
nnnnnnnn nnnnnnnn nnnnnnnn nnnnnnnn nnnnnnnn nnnnnnnn 4860
nnnnnnnn nnnnnnnn nnnnnnnn nnnnnnnn nnnnnnnn nnnnnnnn 4920
ggcacgggc cgtcggtgc ccgttttag gcatccatca tctcccaagg gcttcctta 4980
ggtgtgttg ccgcggcat gtgcaaacg ggaaggaat gaatgaacca ccgttagga 5040
acctccttc ggccttctc ctctctagc cgtgactaac ctctccacc cctcccggg 5100
tgagtcgccc tctgtactgt aagccagtg atgcaaggt tccgtctct cagagagct 5160
ctactcggc agctgtctc tattattag ctcaagcag cctcaagcc tcatctact 5220
tgacatag atatttttg tagattctt gagaatgaa gctgacttg agaggtctg 5280
tgtctctct cccagccca aggtggtgc tatggctag cgcggaggg gggagtgss 5340
gggagctga gtcaigtgg ttctgaagg aaatttcca ccacaaaag gctccggtg 5400
tagcatccc ttcccocat aacctctgt tccatcaca cctgaccca aacctgtag 5460
gccagactgt aaggtcact aagagattg agtctcagc cctcgaacc ctgcccctt 5520
ccccccca tctctggaa accagatct cccgctcca ccttaactg aggttatat 5580
tagcggctg accttcagt attggagtc tggcccca caccacttg ttgctctg 5640
tctgatttt tagctagaa attcaagtc ttgcanaa agagccagg gattagggt 5700
tggaaagctc agtgatttc tcagtcttc cctllaggg gaggacttg gaggagcag 5760
gtggccgac cctgtctca ctactctga cctllaact tgccttga gcttgatg 5820
gtgagtga cagttcttc ctgncagg ggtgagctt gaaacaggg caggtagaa 5880
caacttccc aggggtggg ggtagtgaat gccgtgtgc caccagggg cacttgca 5940
cctctggag actgaatg actggtggg ggttgaga aggtttgag cccatccc 6000
tctggagag gaaagtaac cccatttat ggcctctg taaaagcca gtcaacctc 6060
attgtccaa ggaagtaga cctctgggg ctctcaagg atagggtgt tctatattg 6120
ggcctgct ctacgcttc agccagctt attaaagaa attcataca aaactgaat 6180
tctctctc ttaatacta atagtgtgt ggtatccat taaaatgtt gctctgaca 6240
gtagctatg attctgtg gctctcaca gctatggac aactgaatc tctctctga 6300
aggcatgt tcactcagt actactata gacctagag tctctgacc ccttaacta 6360
cttttttt tctgctca gattggccg tattggacc ctggtacca gggctgctt 6420

```

[0032]

caettcttgcg aaagtggag tggttgcat caatgacccc ttcattgacc tcaactatat	6480
ggcttcacatg ttccagiatg acctacacca tggcaugltc aaaggtacag tcaaggctga	6540
gaatggaaag ctgtcaica acgggaagcg catcacatc ttccaggagc gagaaccgcg	6600
caactatcaa tgggtgaig ctgggcgca gtaigtigt gaactatcg gctctttaa	6660
caccatggag aaggtgggg cccacttga gggcggggc agagaggta tcactccgc	6720
cccttctgt gatgcacca tgtttgtat ggggtgaac caagaaagt atgacaactc	6780
cccaagahi gtcaggtag gaggcagag ggcgtggca aagtgggaa gcaggggcaa	6840
ggttcagggt gggcagcct cctaacctg ctctctctt cagcaatgc tctgaccca	6900
ccaactgctt agccccctg gccaaagta tccatgaca ctttggcatt gggaaagac	6960
tcatgtatg tagttcatct gtttctctct gccagatgt ggcgtgtgg tggggccct	7020
gcaagacctc acitccctgc tctgtctct tccagctcag gtcactgca tcaatgccc	7080
ccagagactt gttgattgac cttccggaa gctgtggct gattggcgt gggctgcca	7140
gaacatcac cctgaccca ctggcgtgc caaggtgtg gcaaaagta gccagagct	7200
gaagcggaag ctgactgga tggcttccg tgttctacc cccaagctg cgttgttga	7260
tcgacatgt cgtctggaga aacctgtatg tctgggtgg gctgagggt gtctctatg	7320
gtgggttgg ggttgtgta gtcacttga tttttgccc taataggcca agtatgaga	7380
catcaagag gttgtgaagc aggaactga gggccactg aagggctac tggctacac	7440
caagacacag gttgtctct ggaacttcaa cagtgctcc caacttcca ccttggatg	7500
tgggcctgg atgtcttca atgacaacti tgaatgctc atttccgtt atgtggatc	7560
agaaacagtc tttttaaagt aacttgggt ttttaccac ttgggttca acitaaaaag	7620
taaaccttc ctaactgccc tcccaatgt tctaggat gacaatgaa ttggctacag	7680
caacagatg gttgacctc tggctacat ggcctccaag gagaagaag cccacctgg	7740
accatccac ccagcaagga ctgagcaag agggagccc ttgttgtga gcagtccctg	7800
tccatnacc cccacacca tcatctctct cacagtctc atcccgacc cccagaata	7860
ggagggctt agggagccc actctcttga atacctcaa taagttcac tgaacctac	7920
tctctgccc ttcaaatga agttggggg agggagctg tgaactagca agattggaa	7980
tctctgtgt cactttcaa acagggcact agccactgc cagctcagg tctctgtct	8040
gaacagatga aaatcaact aagggtctt ggtctggga gngtgggg ttgacaatg	8100
gacagatgt atgtctttag aggttgggc tggactgat caagtccag ggtgtgtgt	8160
gtgttatctg caaacaaag tcaatgtgt ctggagcct tagttaaaa tgaagagatg	8220
cccaacatg taaaaatga tgcgcaggg gaagtgaata cactgtatc aacctgaac	8280
agctggctg taagccagca gcttcaata tgttcagggt tggctcctg gtctgggt	8340
ggtcactgt atccagttg ttaggaaatg ttgtcatgc ttgttagaa gacaacat	8400
ctcaggctg gtagtgtag tgggtcttc taatgccagc attccagca cggaaatgg	8460
agggcgagg ggcagcttg gaataacac ccagggcaga gttcccatc cggaaaagc	8520
aaaaaccga aagtgctaa aacatigaa caaagggtc caataatcc ttaattgct	8580
ttagagatc cactgtcagc ttggcatgg ctctatgag acatctgac ttgttccct	8640
gctttccag gtcaggagaa tgaatgccac agaacgttc ctcagatg ggtggagaa	8700
ccgggtccc tgaacccca cctcacac caatgtcagg agggagctt accttccct	8760
ccagagctg tctgcttca ggaacccgt ctacgccag aacctagtc tcttttgt	8820
gcatttccc ggggaaaagc tacacatcc ctgtacagca ggcaggagg cagctgtg	8880
ctcatggca cggacagaa ggcaggtac cttgcttcc tgcgttacc cgttccctt	8940
cgaagctta ctaggagca aagcttctc cagggttgc tgaigcgaa agcttgcga	9000
aagtcggga ctccagatg agtctgtgc aagtgctg ggcagttg caccagaac	9060
aaagtccag atagccaag agacagga ggtccacaa tccatttcc tacaaacga	9120
gaaggagga tcaaaagt tctggcagc ttcaaatg agagaatg caaacgtca	9180
gagtgaagaa ggcaggatg tccaaatg atgggaat ccaggaggg tggggccag	9240
ctatgtcca gaccccgag gacccgggc gaggagcag cccagggac aggttcatg	9300
cagtggaca aaagacccc aacctggct cgttgatc caggagcca ggccttggc	9360
ctatctctt ggaatata gatgacat ggcactcga gtctcatt tctcttggc	9420
ctggcagg gtcgtatgc gggccgag tggagcag agcttctca gaaaaaac	9480
ccagatga ctttgggca agcaggaagc caagagag ctcagtgc gctctctt	9540
ccatgctct atctcttgg gtcagatg atgcaggca ggaagtgg tgggaagct	9600
tgtcccatc cactctatg aggaacatt cctcttgg aggtctgag aggtcagag	9660
atcaaaatg tctcaagta cagatggag atgttctg agccacagt gttgtgac	9720

[0033]

caggccagc ttctgtcca gctgaagg ggtgggct gctcactat gtcccaag	9780
cagaccca acattggta caagcaatc ctcttgtct agagtatgt ggtgacaag	9840
caaggccc taaacctagt gaacatgtg cacactgtc ctagttaat gtaattggg	9900
gastggatg aggtttgt ttgagacgg gtctcgtat gcatctcgg ctggcttga	9960
atccacagc gagagacgg cctggtaatt gtggcaatc gtaggagac accgtaacca	10020
cgcctctaa gagctggct atcagggtg gtccaaggc atgtcaggc gacatggag	10080
agggtaaag ggaaggaa cgcacatgt agctctttc cctgtgctt tggcttgtc	10140
tgtttctct cggttggct acctaatga tatactctg ccatcaggc ctatgtatta	10200
gtaataato ctaatagcc ggtctccct acacaaacc acctgcctg gacacctag	10260
tgttggatg acagatgac accaccacac ctgcttact gtgttctt agattttacc	10320
tgttttgtt ggtttttta ttttaagaa agtttttgt attgtatcc aggtatgct	10380
caaatccag gcagctttt gcttcagct ccaaggctg ggttcagc caigtaccat	10440
catgctggg ttacacatt ttaataata attattttt attctgtt aattgtatt	10500
tgcctgata tatgtctgt taaagttat agagccttg gaactggag tacagacat	10560
cgttggta ctgggaatt aaccaggcc ctctggaag gcatccagt ttcttaacta	10620
atagccatc tctcagccc caggttttt ttttttttt ttactctgt ggtatgtgt	10680
gtgtgtcac acgacggcg atgtctgac atacaccag acacaggctc tcacacaggc	10740
tagacaagta ctctaccat gacttatat cccaaacc aaatttact ttacacaagg	10800
tcaattcct aaatcatgac ttgacttgg gataaaacc tgccttgta tcttctatt	10860
cttagccagc ttctgtctt cagacatgg ggcactggc ggtctgctt gaccaccag	10920
gcaggagag gaagctatc atctggcca cacttatgg tgcctcaaa ggcagctcc	10980
acttgagat gtacctgtt caccggcac tggctggga cgaaggcca tcagacagg	11040
tctttacgg tccccgact gctgtatcg ggcagctct taccgtccc cgcactgtt	11100
gaigaggcg cggcagctt ccaictgac gtccaggcc cgtttatgc tgcacatct	11160
catgtactc tgcagtgcc ggttcatgc attcttgct gtgcccagt ccagctcag	11220
aaaagtagg caggccagg tggagggtt agcaagcca aggcagacag gggattagag	11280
cccagctgg ccccgagac caggtagcca tggtaaaag ctgggcttc aaaagaagg	11340
aagcagatc aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa	11400
aaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa	11460
aaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa	11520
aaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa	11580
aaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa	11640
aaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa	11700
aaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa	11760
aaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa	11820
aaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa	11880
aaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa	11940
aaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa	12000
aaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa	12060
aaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa	12120
aaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa	12180
aaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa	12240
aaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa	12300
aaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa	12360
aaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa	12420
aaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa	12480
aaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa	12540
aaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa	12600
aaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa	12660
aaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa	12720
aaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa	12780
aaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa	12840
aaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa	12900
aaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa	12960

[0034]

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn	13020
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn	13080
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn	13140
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn	13200
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn	13260
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn	13320
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn	13380
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn	13440
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn	13500
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn	13560
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn	13620
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn	13680
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn	13740
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn	13800
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn	13860
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn	13920
nnnnnnnnnn nnnnnnnnnn acacacacac acacacacac acacacttac ttcctgagga	13980
agacgicagg gaacacaggc agatagagcc cactatccca gacacacaaa gtagagcaaa	14040
aatggctcct taagacacaa agcagcacaa ccaacagagg tctcttcagc cctatataic	14100
acctagtttt acacatigaa gggaaagatg taacattctag aacattttta acacgtcaga	14160
cccacaatc cagacttcag atgggagatc agacagtag aggtagatgg agagataga	14220
ggggccaggc ccatggagga aaactccac cgtctatcag ctttagcacc caataatggg	14280
cagctctacc tcagtaacac tgagagata aggccatac agacaaaagg aactttigcc	14340
acaaagctac caaggaact ttctttacct ttctctcaa gtcaattctt caggaaaggc	14400
agaaatcaca caacttcaga acaactttcc atccatgtgc agtatcata tttttatgtg	14460
tgtagatata gatgtcaltg tgtgtgtta caggggaag cctgatgtcg atgcaggaaa	14520
tctctctagg tggctcttat actctagatc tggtaacgg catlgtgtgt ccgtgaactt	14580
gatctgtaga ccagacggc ctcaaacata cagagatcta cctgtctcgt cctcccgat	14640
gtgggatta aaagcgtgtg cctcagccgg acgttggtag cgaatgctt taatccagc	14700
actcggaggc cagagccagg aggaatcttg tgagttctag accagccttg tctacannnn	14760
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn	14820
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn	14880
ggcagcctcc aaagccacag agaaaccccg tctcaaaaa caaaacaaaa ataaataaat	14940
aaataaataa aatctctggt cttcacctgc cggctcagaa tattttttaa gacaaggatt	15000
atgtagccca ggttggtctc aaacttgtaa tgtagctaa agtatgtgtg cctctgatcc	15060
ttctgctctt acctccaaag tgttaggact ataggctgc actaccacac ccgtttctat	15120
gtggcttagg atcacactca ggtaggcttt catgcacatt gggcaggcat tctaccaact	15180
gagttacaca gtccagccc aggatgtctc accagatcac ggtttacagt aggcacctgt	15240
ggctcagggt ttgcaccaca tggcagccc agtcaatgag ctaagtgaac actcatactt	15300
gaataaaggc accgaatctg aaaggggttg gtacaacca ggaggaaaa cggcggaaagc	15360
aataaagggc gtacacagcc aaaggctact ttccccaga caagtctcag ctggaaagcg	15420
cacacgggct gggatctaga aactcagaag ttctctgggt taaaaacaaa agccctggtt	15480
ttctctagga tagtatgttc atttgaigtg tgcctggaca ctgagaacac aggaataata	15540
cttccaaagt ttgtctcaaa atgtcagaca ggtgaacct ttacagggtt tgcaggagag	15600
ctgatacaag aggtctctcg tctctctcac ccatcagagt ccaagacaca gtctctcac	15660
agccctaccc aacctaacgt cctctcagcg tgacctctc cctacaaacc tgcctgcage	15720
ccagaaggcc agtcagagcc ccagctccat agccatttgt aaaggactcc ccaggctcaa	15780
aaacaaaggc gnaactctc caggagctca caggctgagc caaggaggca ggaatgtccc	15840
ggcacttagg gtgtccgggt gataatgccc aatagtatg atgtgtcaac ctgcagatc	15900
ctgtcttaag tgcacagtgt ctatctgatc atcttaataa cccagcctgc gggagagac	15960
atgcacatct cctctgaca aataagaggc caggcccaag cttagcatgg cagtacacat	16020
ctgtaacaca gaactcaaca ggcagagcag gaggattgac tgcctcttat tccaggccag	16080
tttggtctaa cagaagagac ctgtactgac gctaaagagc atggtctaga gtaagagca	16140
cttctgtctc ctggagagga cccaggtttg ctctcagaca cccatgtcag gctcacagtc	16200
gcctgtgaac accagctcca gggatctga tgccctcttc tggcctccag gggcagcgc	16260

[0035]

gllgagctc acacacacac acacacacac acacacacac acacacacac acacacacac 16320
 acagagtgca aagctgaagt gcagctacac cccalaatcc tagcnaatgg agacaaaagg 16380
 accagaggtt cgagcccaac ctgactacac agacagtlac ttgtaaaata atacatatat 16440
 acgaaatna acnaaccana aatnatgana cagcactca gatcaugga gccagttttg 16500
 gllgtacacg cctttaaacc cagcactcag gaggcagagg caggllgato tctgagtttg 16560
 aggcagcct agtctacaga atgagatcca ggaclacaca gagaacctct gctccacct 16620
 cctctccct aaaaaaaatt aagaaagaaa agaaagatcc agggcttgg aagcaatgga 16680
 teagtaggia aagccacaag ccccaagcc cacaacctca aatttgatcc ctgggacctc 16740
 catgtagaa ggagagaact cactgcagc tgcctctga cctccacac cacacctgg 16800
 caggacatg ctcctccgc cacatataa cacaatara tacaatata cacaatata 16860
 tacaatcatg taacagaaa agaagtagg aagacaaaag gttaattca agtggtaga 16920
 gagcaagat ttgaacctac atctatctg gctagatcc taggtctca agggccatt 16980
 agagctacca aaactgggt gggtgcact cctctgtca acaggtgaa gctctgtc 17040
 ttggctctg gctgggagg gccagccag cagcacacct ccagcttgg ttgtgagcc 17100
 tggaaacaga gccctgcca aacctctac 17130

<210> 30
 <211> 3020
 <212> DNA
 <213> 人工

<220>
 <223> 扩增子

<400> 30
 tacggccggc ttacgtgtac agtggcacat actgttacc acagtgtcat actggacatt 60
 ttcatgccc taaggccaca tcacctctgt ggtgtacac agttgtaatt tcacctata 120
 tagtgatttc aggttggctt ttactttaga atataacatt taatttatgt ttcttttttg 180
 gttttaggtt ttgtttttcg agacagggtt tctctgtatt gctttgagc ttgtcttga 240
 actgtctctg gagaccaggc tggcttcgaa ctacacaga tctccacac tctgctccc 300
 gagtctggg attaaaggcg tggccacca atgcccagc atctttctgt ttctctgttt 360
 ttaagaggga agataatgtt tccagtctg tttttaaatt ttctctgta ggtaccata 420
 gatgaatttg agcagaagct tgggcatgt cataccagag gcattggaagg aatagaggag 480
 ctgaaactg gccagaagg cagccagcaa gccctgctg ccagaanaac ctctctgtgt 540
 atgagagaca agccacacag gactccccc aaatcatccc tgcatttat ttgttttgt 600
 ttttaaaatc acatctccct gccatttata gctctaggc gacagctct gactctata 660
 gattcagaca atgactttgt cacacctag ccctcagcca ccaaccgta tctctcaac 720
 actcagcaac gaagtccaa gaagaaagcc aaagtgtct tttaagtag tagtccagt 780
 gagatggta tgcctgccc tccggcagc tgcctgtct ccttggagtc tcttggggc 840
 ttgcttggtt gttaggtcat tcttcttggt aatgtgtcc ccaatgtgca cgttggctca 900
 ccttgcagc tctatgttt aagatccact ttacacatc ttgtgactc agcttttgtt 960
 cctataatto ttgcaaga actttcagc gaagagagg aagaagagac tcccaagaga 1020
 accaccccca tccgagatc atctgcccga agacacaggt cttaggagt gccacacta 1080
 tcccatccc tctctgtcag gxtatctaa ttgallttta attcatctcc tcccttgtaa 1140
 aaatgttgc tcttgttttt taatagttaa atggcttttg catcaaggg ctttctctca 1200
 gagccatct gactgtctgt ttacatttc acaaacact taaccaatca ccatgtttc 1260
 taagaaana tgtttttata tacttttatt ctttgtgtgt ctgttctatc ttccactta 1320
 gccagattc aggcagcaat agagggttat caaacagag aagaatata ccagccagc 1380
 catctagag ctgtctctgg gcttctagt tcccatag attcatctg cacaactgt 1440
 gaganccac ttagagaat tctcattct cagcctgggg gtactttcta cccagagac 1500
 tattctgtc atagcctttt gggacacata aagcttccit ctgcataga acacccaca 1560
 acagtgtatc agagtgtag aagttgaca acnaigtct agacctagg tttcttctt 1620
 ctgtgtgctt catccagaa gatcatga ctccaggag tcagcttita ctatgggtc 1680
 tgcaggggc tccagccctt cagcggaag ccatgccac cctcccaag tcttcatct 1740
 gctagctac ttggaacagg agacactgt tctgtgtca tgacaacgc acattgccat 1800
 agaaatgct cctacctctt agtgtgtgg tggggaaaca gtaatgaca accatggca 1860
 ggaagcaaaa ggaagagcg cactcagaa acatgttta gtttaggca gaactatga 1920
 gaggtctct agactctttt atgggaagg gtaatgtct cctctgaac ctctgttga 1980
 agcagaaaa gggacaggc tgaatcccc cactgggac atttcaicc tctgcatct 2040

[0036]

```

gcccccgct catgnaagc cgggcaagg gccacacagc tctagtott ggccttgggg 2100
ccccangag gtagggcaat cccagaaagg caagagacca ggaatggatt tggggtgag 2160
ccccccgag ccccgccct taaagcaagc gtaacacca ggcagactaa acttagcaat 2220
taggtcttic agcnaaaga gcaggggcct ggtctcaagt tgcactgac tagcaaaagc 2280
gccccagat cccctcgcc agcacctgig gctgagrico caagccttc ccagagctc 2340
aggatccac ctttccacc tccctactct tcagaggagg aaccccttt ctccttcca 2400
cttgctggag ggggtcggg ccaggctggt ctggctggg gtataatgc ccctaccct 2460
tctacttic cctctctca gactcacc ctcctcacc agggagcca agnaagaga 2520
gtccctggct gcaggccag taggcacgt ccaggacggc gaggacttc agcctcagc 2580
tccagctct cgccttggg ctgagtgag tgaaggggc agtctctct agcctgggc 2640
gigaaccc caggctcga ggaggacgt ctggggagg tctttgctt cctccacc 2700
aaccccaac caggacttg tctcacc cgggcctca cctagactt aactaggaa 2760
cacagggcc agtttgtag tctaatat gcaggcct ccgactgtg gcttagggag 2820
cgcctccca ggcctcttc ctgactctt ctggccctt acagccagg ccttggcc 2880
nagaatgga agtggggg ggggagga ctgcaatgc gaaggcccc cgcaggggc 2940
aaagtacc cccccggc tttctgctt cagacttgt tttgctgt gtcactacc 3000
aaggaccag gcggcctga 3020

```

<210> 31
 <211> 3537
 <212> DNA
 <213> 人工

<220>
 <223> 扩增子

```

<400> 31
tactcgcat tcactgtaca ggggacata ctgicacca cagtgcata ctggacatt 60
tcagaccc caggccacat cccccgig gtgcacca gtglaatt caccctatat 120
agtatttca ggttggcct ttcctagaa tataacatt aaattatgt tcttttttg 180
ttttggttt tgttttcca gacagggtt cctgtatag cttagaggt tgtctggaa 240
cttgcctgg agaccaggt ggtctgac tcacagat cctccact ctgcctccg 300
agtctggga taaagggt ggcacacca tggccagca tcttttgtt tcttgtttt 360
taaaaggga gataatgt ccagtgctg ttttaactt tttctgtag gctaccatg 420
atgaattga gcagagctt cggcctgic ataccaggg catgattga atagaggagc 480
tgaacctg ccaagaggt agcagtag ccctgctgc caagaaccc ctgcttgga 540
tgagagcca gcccacagg atccccaca atccatccc gtcatttat ggtttttgt 600
tttaaaaca cctctccgt ccatcttag tctcaggcg acagctctg actctatat 660
attcagaca tgacttggc acactaagc ctgagcac caaccgcat cgtccaaac 720
ctcagcaag caagtcac agaaagcca aagtgctt ttaagtat ggtccagt 780
aggatgtat gctcgcct cccggagct gctgctgic ctggggtct ccttgggct 840
tactgttg taggtcatt ctcttgga atgttccc cactgtgac gttggctac 900
cttgagct ctatgttta agacactt tacactat ttgtgaca gctttgtc 960
ctataatct tgaacaga ctctcagg aaatagga agaagact ccaagagaa 1020
ccaccccat cgcagatc ctgcccga gacacaggt ttaggattg cctccatct 1080
tccatccct gctgtcagg gtatcttaa tgattttta tcatcttct ccttggtaa 1140
aatgttgc cctgtttt aatgtgaa tggcttgc attcaagg ctctctctg 1200
agccatctg actgtctt tcatctca acaacact aacactcac catgtttct 1260
agaataat gtttttat actttatc ttgctggt tctttctct tccacttag 1320
ccagattca ggcagcata gaggctat aaaccagaa gactatcac cagccagga 1380
atctgagc tgtcttgg tctctagt tctcattg tcaacttg acaactgat 1440
agactccat gttagaatt ctactctc agcttgggg tactttcac ccagagact 1500
atttctgta tagctttg gacacata agcttctc ctgcataga cccccaca 1560
cagtgtatc gtaggtaca agtgacaa caatgctc gactaggt tttctctc 1620
tgtgtctc atccagag agatctac tctcaggat cagctttac tatgggtct 1680
gcagggggt cagccctc agggcaagc cagccccc cccccagt cctaatct 1740
ctgactcat tgaataga gactctatt ctgctgta gacactga catggcata 1800
gaatgtct ctactctt ctgtgtgt ggggaacag taatgacaa ccataggag 1860
gggcaaaa ggaagaggt acctagaa catgtctg gttagggag aactatgg 1920
ggctctga gactctta tgggaagg ttaatgct cctgaacc tctgttgaa 1980

```

[0037]

ggcagaaaag ggcagaggct gaggccgcg actgggacca ttccatctct ctgcatccig	2040
ccccggctc atggaagcc tggcatggg ccacacagct gtcagtctg gctctggggc	2100
cccaaggagg tagggcaatc ccagatggc aaggagccag gactggaltt gggctgcagc	2160
ctagccigtct cctgcccctt taagcaagg ttatccaccg gccagctaan cttagcattt	2220
aggtctctca gctaaaagag cagggggctg gtctcaagtt gactgaact agcaagagg	2280
ccccggate cccctgccc gacccctgg ctgagctcc aagcccttc cagagctca	2340
ggattcacc ttccacctt ccttactct cagagggagc acccccttc tcttccac	2400
ttgtggggc gggctggggc cagctgttc tggctgggg tataatacc cctaccctt	2460
ctacttccc ctctctcag acctaccct gcctccaga gggcagcaa gaaaggagag	2520
tccctggctg caggccactt agcagctcc caggccggg agggacttc gccctcagt	2580
ccagctctc gccctgggc tgcagtggc gaaaggcca gttctctca gccctggcg	2640
tgcaccttc aggttcggg gaggacgct ctgggggct tctttgcct ctccaccca	2700
cccccaacc aggcattgt cctcaccg gggcccaac ctgacctta atgagggaac	2760
acagggcca gtttgtagt ctcaattatg caggccatc cgaactgtg cgtaggagc	2820
gctctccag gcccttccc tagctctct ctggcctca cagccaggc ctctggcca	2880
agaaatgaa gtggggggg gggctggac tgggaatgc aagggccccc gtagggcca	2940
aagtagccc tccgggctt ttctgtctc gaggctgtt ttgctgtg tctctccga	3000
agaaccaga gaagatctc aactttcca cagctttgc ataaaggga gaggctcgc	3060
ggtagctg tggcacacc gcaattctg tcaaccgca cccccccc cctgttctg	3120
tctctcca ggttctccc attttatgg ggcggcaat tttaggtcc tgggtcttg	3180
aagctctag tacacactct tctctctaa gtcctatg tctattccct cgtctctat	3240
ctgtcccca tccagggtc acctccacg cgaagcaat tcagttccc tccccctc	3300
agcccgagc ccacagttt ggtgctgca cattcaaaa acgaggcgg tcaaaaga	3360
ggggggggc aggtgcgag tggccagct actgcgct ttacgggtc actgactca	3420
ggctcagcg ccttgagct gtgctggat ggaagcgg ggcggagcg gggcagcg	3480
tctcgggc tcccaccac cccagttct ataaataga gactctgtt agtgaac	3537
<210> 32	
<211> 537	
<212> DNA	
<213> 人工	
<220>	
<223> 扩增子	
<400> 32	
taetcgcaa gaagatctc aactttcca cagctttgc ataaaggga gaggctcgc	60
ggtagctg tggcacacc gcaattctg tcaaccgca cccccccc cctgttctg	120
tcttccca ggttctccc attttatgg ggcggcaat tttaggtcc tgggtcttg	180
aagctctag tacacactct tctctctaa gtcctatg tctattccct cgtctctat	240
ctgtcccca tccagggtc acctccacg cgaagcaat tcagttccc tccccctc	300
agcccgagc ccacagttt ggtgctgca cattcaaaa acgaggcgg tcaaaaga	360
ggggggggc aggtgcgag tggccagct actgcgct ttacgggtc actgactca	420
ggctcagcg ccttgagct gtgctggat ggaagcgg ggcggagcg gggcagcg	480
tctcgggc tcccaccac cccagttct ataaataga gactctgtt agtgaac	537
<210> 33	
<211> 30	
<212> DNA	
<213> 人工	
<220>	
<223> 引物 GlnPr1896	
<400> 33	
tacggccgc ttacgtgac agtggcact	30
<210> 34	
<211> 30	
<212> DNA	
<213> 人工	
<220>	
<223> 引物 GlnPr1897	
<400> 34	
tcaggccgc cgtggtctt aggtagtag	30
<210> 35	
<211> 35	
<212> DNA	

[0038]

<213>	人工	
<220>		
<223>	引物 GlnPr1901	
<400>	35	
	tactcgcgaa gaagatccct aacttttcca cagcc	35
<210>	36	
<211>	40	
<212>	DNA	
<213>	人工	
<220>		
<223>	引物 GlnPr1902	
<400>	36	
	gttcactaaa cgagctctgc tatttatagg aactggggcg	40
<210>	37	
<211>	40	
<212>	DNA	
<213>	人工	
<220>		
<223>	引物 GlnPr1903	
<400>	37	
	caccccgatt cctataaata gcagagctcg tttagtgaac	40
<210>	38	
<211>	20	
<212>	DNA	
<213>	人工	
<220>		
<223>	引物 GlnPr1904	
<400>	38	
	cgcctagcacc ggtcgatcga	20
<210>	39	
<211>	31	
<212>	DNA	
<213>	人工	
<220>		
<223>	引物 GlnPr1905	
<400>	39	
	tactcgcgat tcactgtaca gggcacata c	31

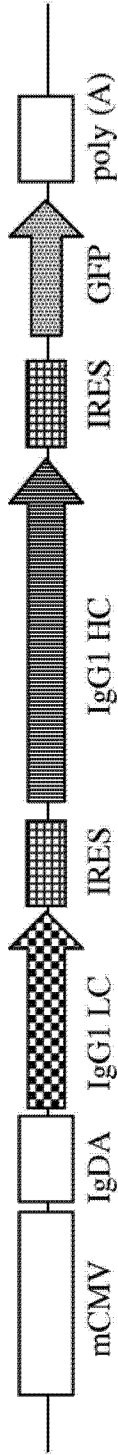


图 1

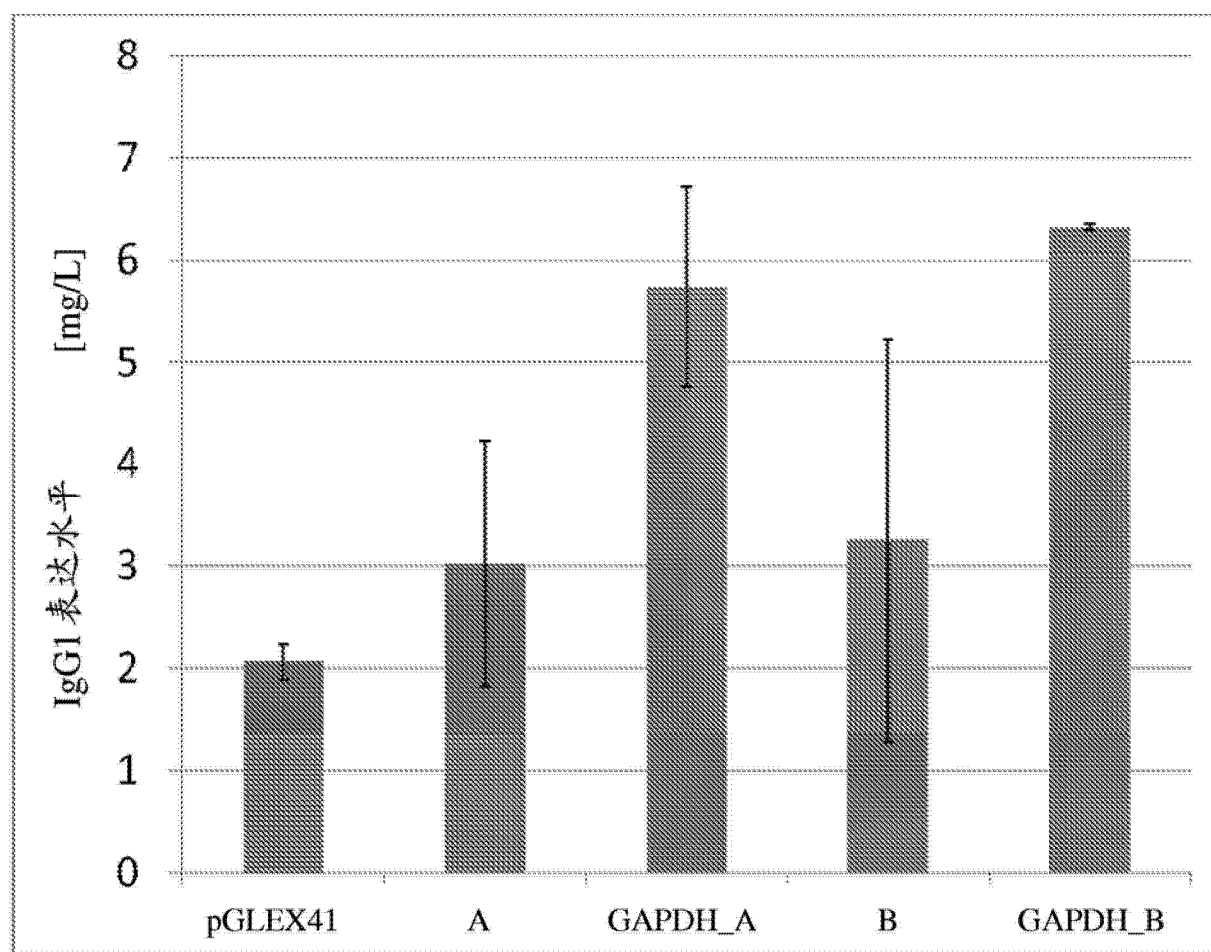


图 2

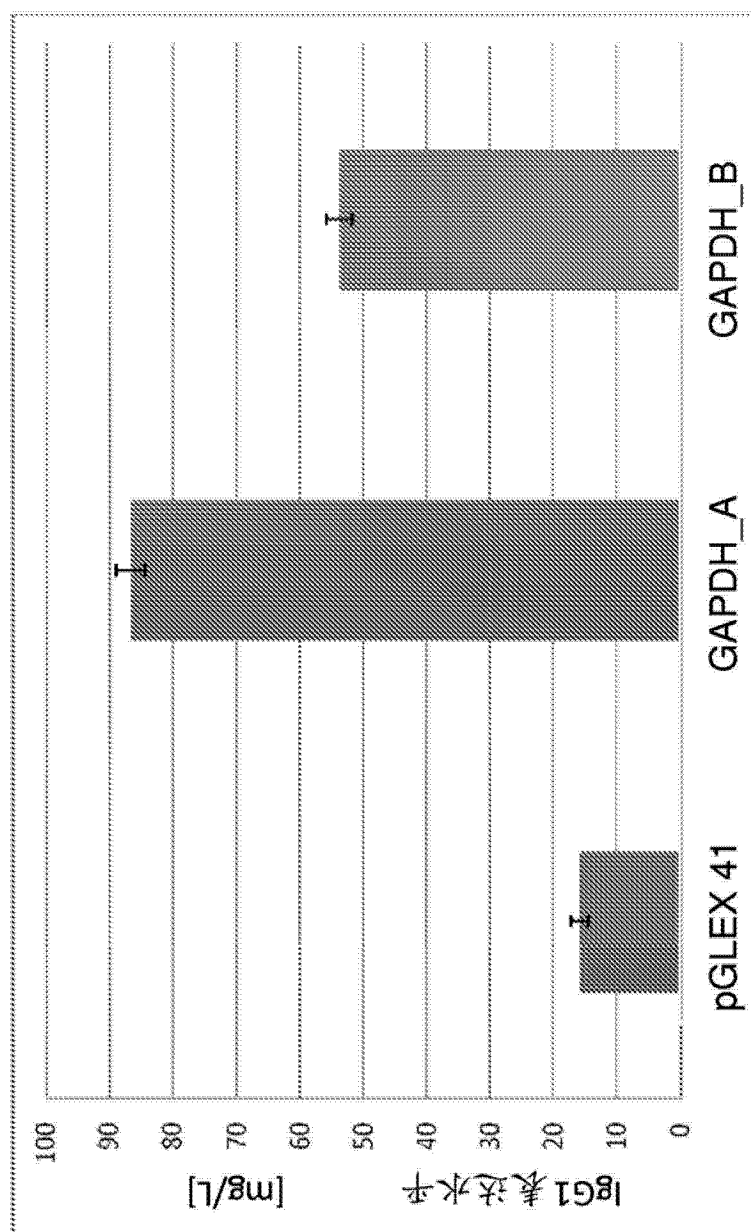


图 3

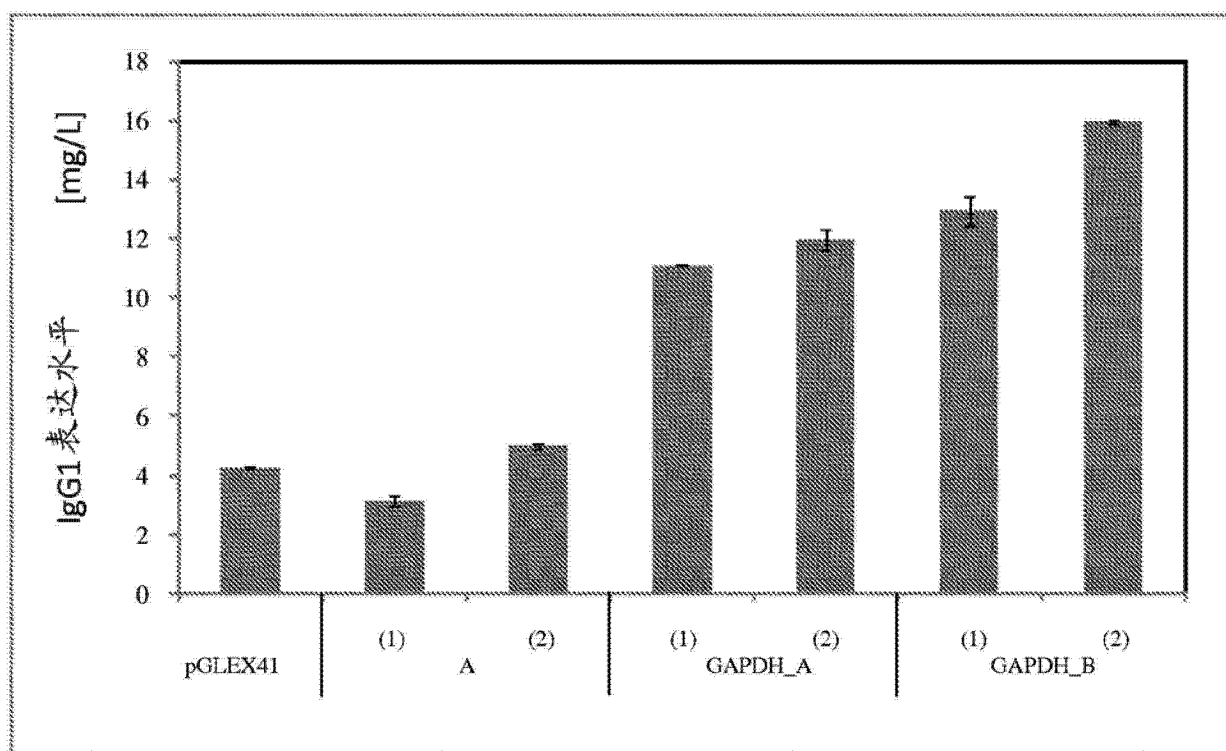


图 4

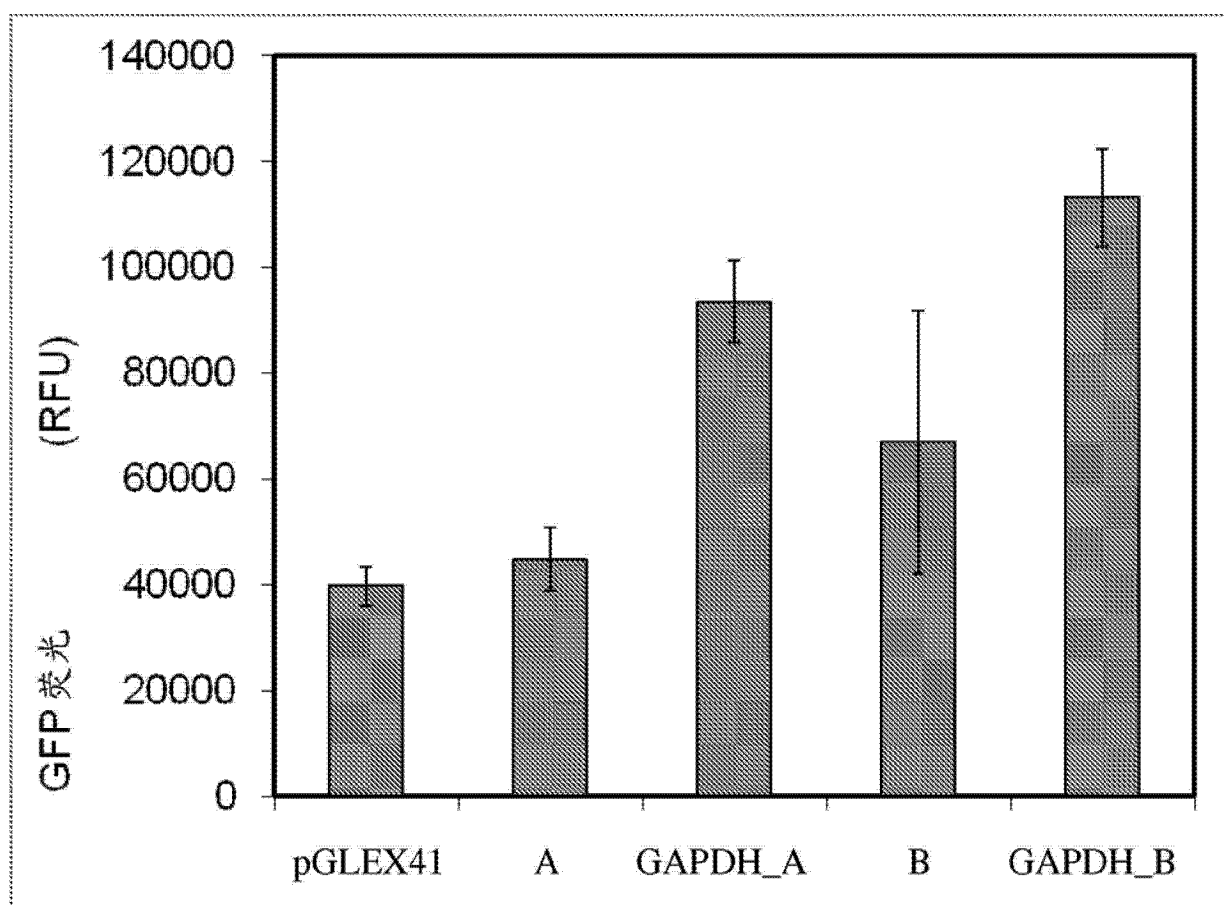


图 5

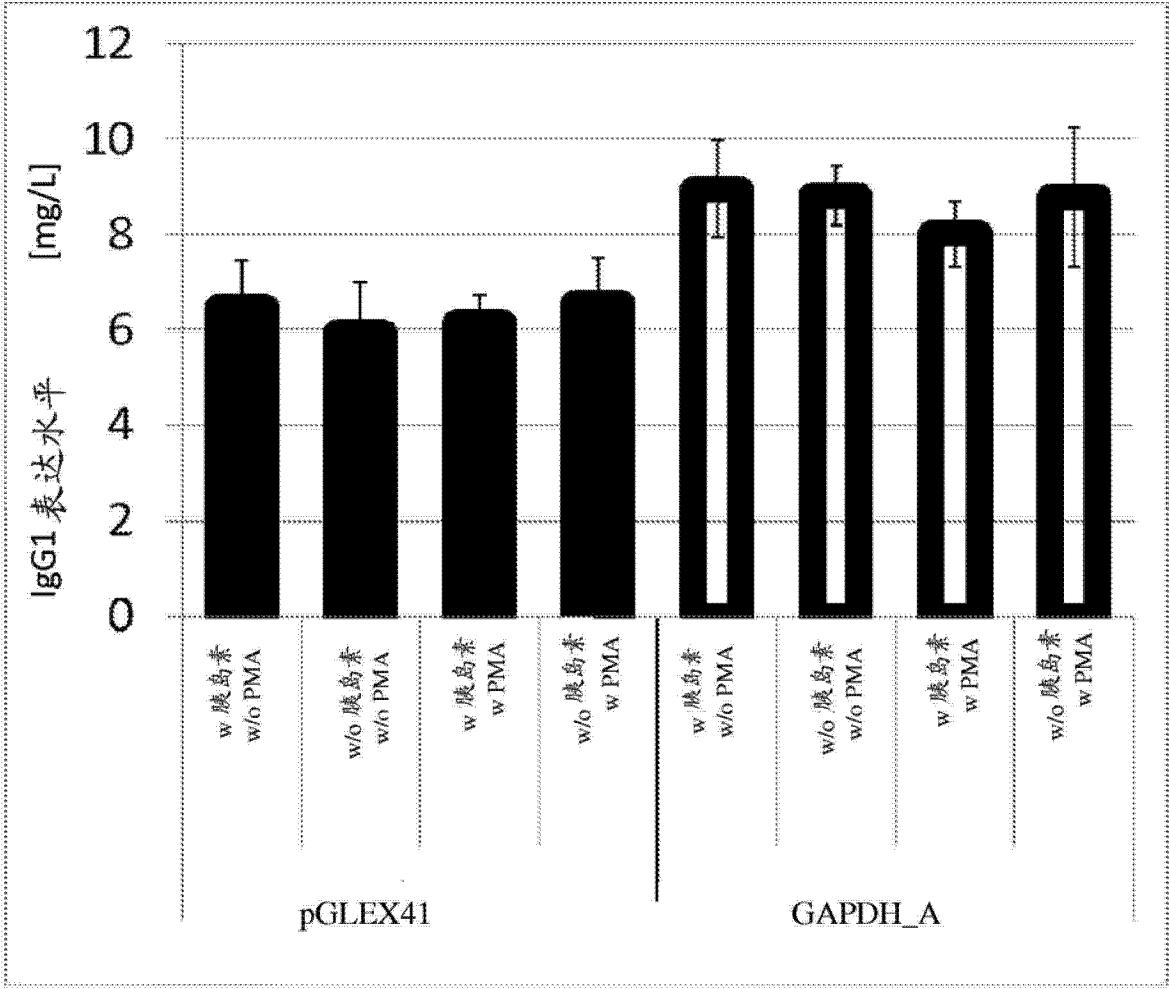


图 6

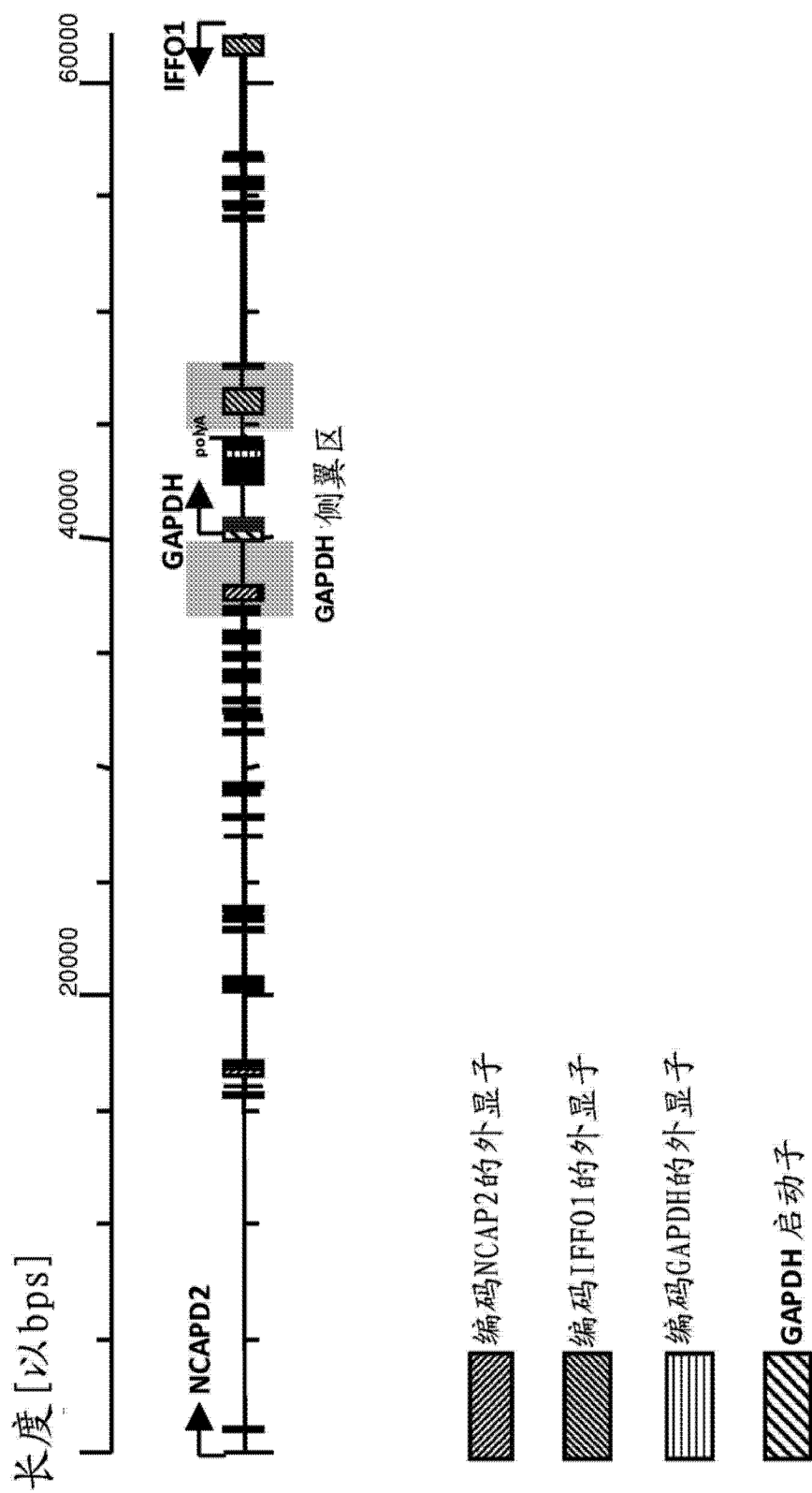


图 7

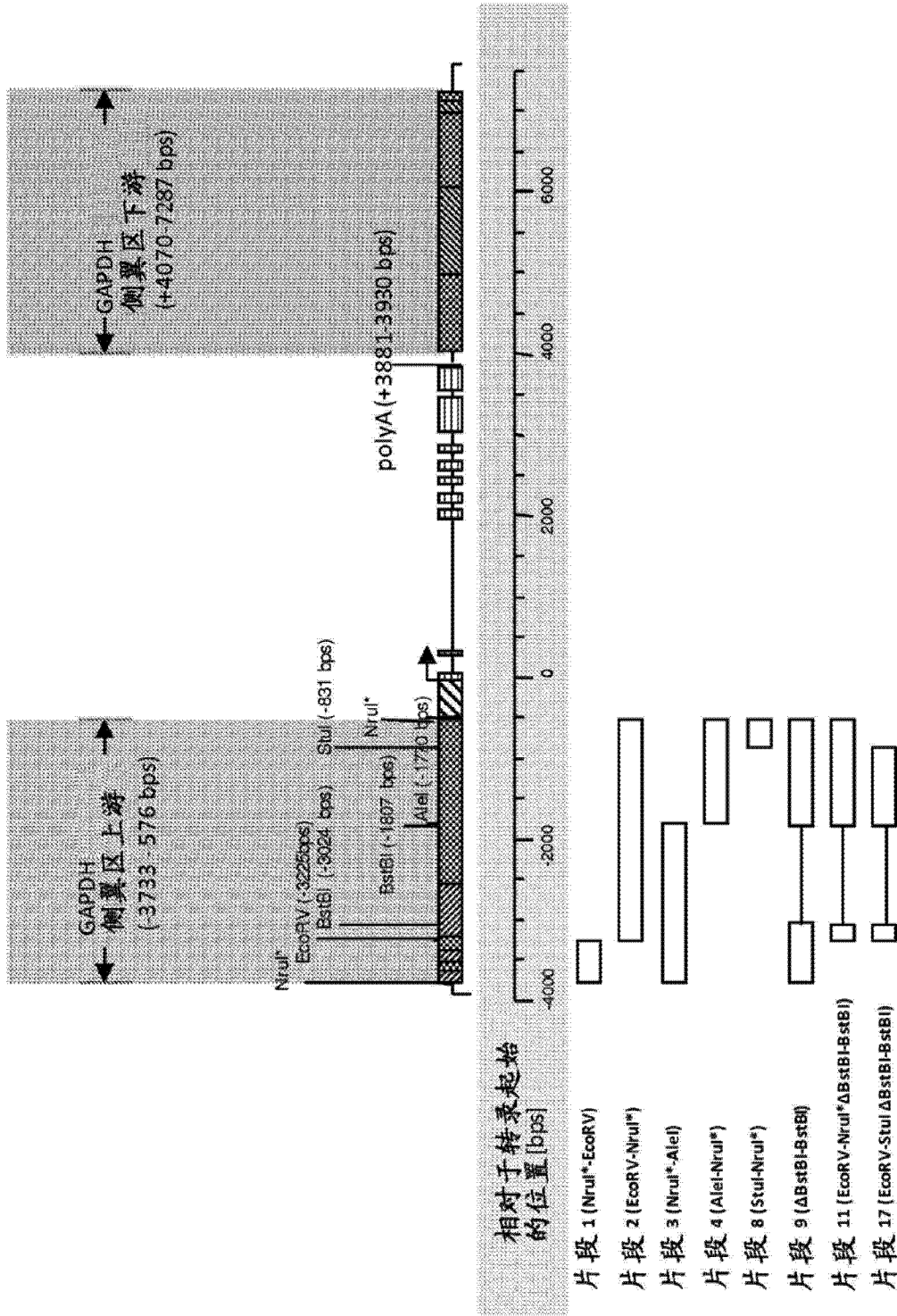


图 8

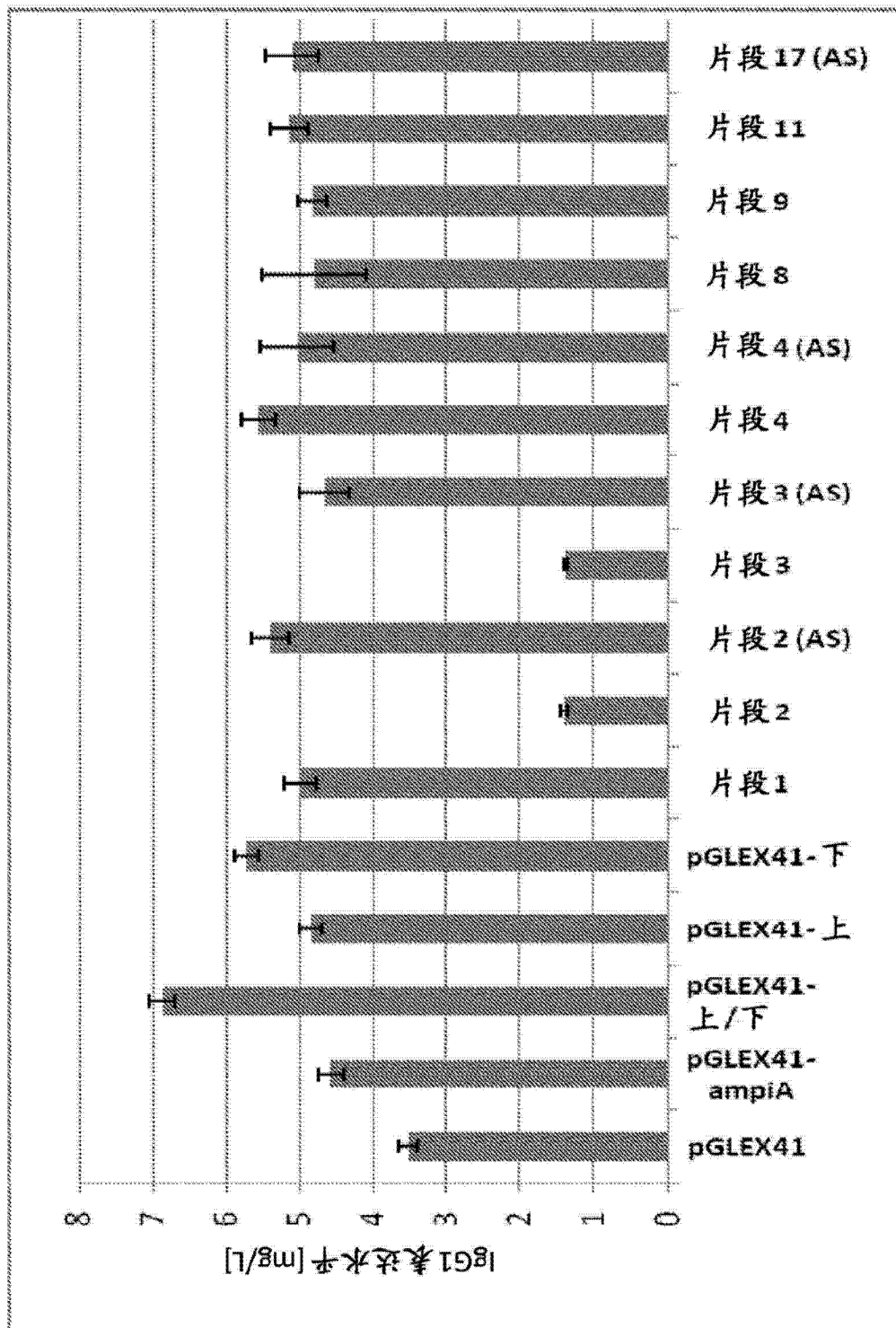


图 9

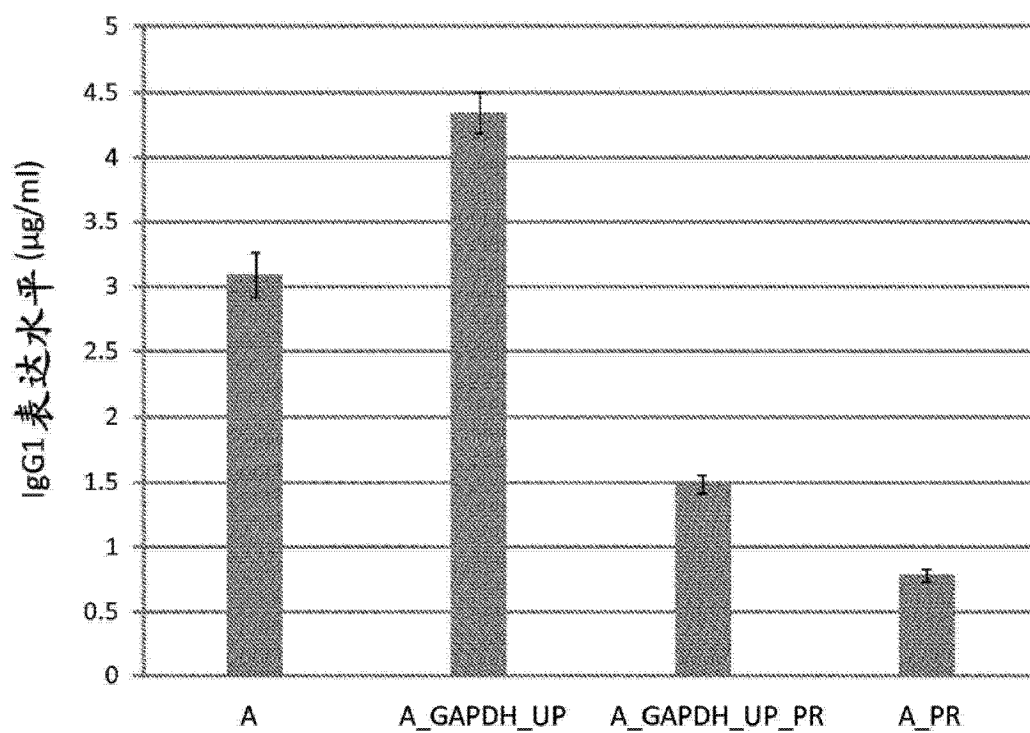


图 10