



(12) 发明专利

(10) 授权公告号 CN 107075549 B

(45) 授权公告日 2021.04.20

(21) 申请号 201580044100.7

(22) 申请日 2015.08.19

(65) 同一申请的已公布的文献号
申请公布号 CN 107075549 A

(43) 申请公布日 2017.08.18

(30) 优先权数据
62/039,416 2014.08.19 US

(85) PCT国际申请进入国家阶段日
2017.02.17

(86) PCT国际申请的申请数据
PCT/US2015/045793 2015.08.19

(87) PCT国际申请的公布数据
W02016/028838 EN 2016.02.25

(73) 专利权人 瑞泽恩制药公司
地址 美国纽约

(72) 发明人 D·德什潘德 D·布拉科夫 陈刚
J·凡德尔

(74) 专利代理机构 中国贸促会专利商标事务所
有限公司 11038
代理人 罗菊华

(51) Int.Cl.
C12P 21/06 (2006.01)
C12N 15/00 (2006.01)
C12N 5/07 (2006.01)
C12N 5/10 (2006.01)
C12N 5/02 (2006.01)
C12N 9/12 (2006.01)

审查员 汪豪杰

权利要求书2页 说明书20页

序列表27页 附图21页

(54) 发明名称

重组蛋白的高效选择性

(57) 摘要

本发明提供一种新的表达系统,所述表达系统包含促进糖蛋白的所需翻译后修饰的哺乳动物选择性标记。具体地讲,本发明包括通过采用以哺乳动物来源的GPT基因为基础的选择标记系统在哺乳动物细胞中实现最佳重组蛋白表达的方法和组合物。本发明包括在哺乳动物细胞中促进重组蛋白的选择性以及增加表达拷贝和蛋白产率的方法,和使用GPT表达系统的方法。



1. 一种分离的哺乳动物宿主细胞,所述宿主细胞包含外源性引入的核酸,所述核酸包含哺乳动物衣霉素(Tn)抗性基因、第一所关注基因GOI和至少一个调控元件,

其中所述Tn抗性基因可操作地连接到所述第一所关注基因GOI和所述至少一个调控元件,

其中所述第一所关注基因GOI编码抗体轻链或其抗原结合片段、抗体重链或其抗原结合片段、或Fc融合蛋白或其片段,并且

其中所述哺乳动物Tn抗性基因编码由SEQ ID NO:3或SEQ ID NO:4的氨基酸序列组成的蛋白。

2. 根据权利要求1所述的分离的哺乳动物宿主细胞,其中所述Tn抗性基因包含选自以下的核酸序列:SEQ ID NO:2或SEQ ID NO:12。

3. 根据权利要求1所述的分离的哺乳动物宿主细胞,其中所述至少一个调控元件选自启动子、核糖体结合位点或增强子。

4. 根据权利要求1所述的分离的哺乳动物宿主细胞,其中所述第一所关注基因GOI可操作地连接到启动子。

5. 根据权利要求1所述的分离的哺乳动物宿主细胞,其还包含外源性添加的第二所关注基因GOI。

6. 根据权利要求5所述的分离的哺乳动物宿主细胞,其中所述第二所关注基因GOI编码糖蛋白。

7. 根据权利要求6所述的分离的哺乳动物宿主细胞,其中所述糖蛋白选自抗体轻链或其抗原结合片段、抗体重链或其抗原结合片段、Fc融合蛋白或其片段、配体以及受体或其配体结合片段。

8. 根据权利要求1所述的分离的哺乳动物宿主细胞,其中所述细胞选自CHO-K1、COS-7、HEK293、肿瘤细胞、淋巴细胞或视网膜细胞。

9. 根据权利要求8所述的分离的哺乳动物宿主细胞,其中所述细胞为CHO-K1细胞。

10. 一种包含核酸的载体,其中所述核酸包含哺乳动物衣霉素(Tn)抗性基因、第一所关注基因GOI和至少一个调控元件,

其中所述Tn抗性基因可操作地连接到所述第一所关注基因GOI和所述至少一个调控元件,

其中所述第一所关注基因GOI编码抗体轻链或其抗原结合片段、抗体重链或其抗原结合片段、或Fc融合蛋白或其片段,

其中所述哺乳动物Tn抗性基因编码由SEQ ID NO:3或SEQ ID NO:4的氨基酸序列组成的蛋白。

11. 根据权利要求10所述的载体,其中所述Tn抗性基因包含选自以下的核酸序列:SEQ ID NO:2或SEQ ID NO:12。

12. 根据权利要求10所述的载体,其中所述至少一个调控元件选自启动子、核糖体结合位点或增强子。

13. 根据权利要求10所述的载体,其中所述第一所关注基因GOI可操作地连接到启动子。

14. 根据权利要求10所述的载体,其还包含第二所关注基因GOI。

15. 根据权利要求14所述的载体,其中所述第二所关注基因GOI编码糖蛋白。
16. 根据权利要求15所述的载体,其中所述糖蛋白选自抗体轻链或其抗原结合片段、抗体重链或其抗原结合片段、Fc融合蛋白或其片段、配体以及受体或其配体结合片段。
17. 一种方法,其包括:
 - (a) 提供哺乳动物宿主细胞群,
 - (b) 将根据权利要求10-16中任一项所述的载体通过转染引入步骤(a)的细胞群中,
 - (c) 在Tn存在下培养步骤(b)的细胞群,从而获得包含所述载体的所述核酸的细胞转染子。
18. 根据权利要求17所述的方法,其中所述Tn的浓度为至少1 μ g/mL。
19. 根据权利要求17所述的方法,其中所述Tn的浓度为1 μ g/mL、2.5 μ g/ml或5 μ g/mL。
20. 根据权利要求17所述的方法,其中所述细胞转染子通过在连续增加浓度的Tn存在下培养步骤(b)的细胞群来获得。
21. 根据权利要求20所述的方法,其中所述培养包括在1 μ g/mL的第一Tn浓度下培养,随后在2.5 μ g/ml或5 μ g/mL的第二Tn浓度下培养。
22. 根据权利要求17所述的方法,所述方法还包括从培养的细胞分离从所述第一所关注基因GOI表达的第一所关注蛋白POI。
23. 根据权利要求20所述的方法,所述方法还包括从培养的细胞分离从所述第一所关注基因GOI表达的第一所关注蛋白POI。
24. 根据权利要求17所述的方法,其中所述哺乳动物宿主细胞选自CHO-K1、COS-7、HEK293、肿瘤细胞、淋巴细胞或视网膜细胞。
25. 根据权利要求17所述的方法,其中所述哺乳动物宿主细胞为CHO-K1。
26. 一种制备重组的所关注蛋白POI的方法,其中所述方法包括:
 - a. 提供根据权利要求1-9中任一项所述的哺乳动物宿主细胞;和
 - b. 在Tn存在下培养所述细胞,以从所述所关注基因GOI表达所关注蛋白POI。
27. 根据权利要求26所述的方法,其中所述Tn的浓度为至少1 μ g/mL。
28. 根据权利要求26所述的方法,其中所述Tn的浓度为1 μ g/mL、2.5 μ g/ml或5 μ g/mL。
29. 根据权利要求26所述的方法,其中所述培养包括在连续增加浓度的Tn存在下培养。
30. 根据权利要求29所述的方法,其中所述培养包括在1 μ g/mL的第一Tn浓度下培养,随后在2.5 μ g/ml或5 μ g/mL的第二Tn浓度下培养。
31. 根据权利要求26所述的方法,其中所述哺乳动物宿主细胞选自CHO-K1、COS-7、HEK293、肿瘤细胞、淋巴细胞或视网膜细胞。
32. 根据权利要求26所述的方法,其中所述哺乳动物宿主细胞为CHO-K1。

重组蛋白的高效选择性

背景技术

[0001] 序列表

[0002] 本申请以引用方式并入以计算机可读形式提交的序列表,该序列表于2015年8月3日创建为文件8700W0_ST25.txt (75,769字节)。

技术领域

[0003] 本发明提供重组蛋白在哺乳动物细胞中以一致和高效的方式表达。具体地讲,本发明包括通过采用哺乳动物选择标记在哺乳动物细胞中实现改善的蛋白表达的方法和组合物。本发明包括在哺乳动物细胞中促进重组蛋白的选择性以及增加表达拷贝数和蛋白产率的方法,和使用此类表达系统的方法。

[0004] 相关领域说明

[0005] 细胞表达系统的开发是提供用于研究和治疗用途的可靠且高效的给定蛋白来源的重要目标。由于例如哺乳动物表达系统对重组蛋白进行适当的翻译后修饰的能力,哺乳动物细胞中的重组蛋白表达就制备治疗性蛋白而言通常是优选的。

[0006] 有多种载体可用于在哺乳动物宿主中表达,它们各自含有使得能够在细胞培养期间容易地分离表达重组蛋白的细胞的选择标记。在此类系统中利用选择性标记基因(SMG),原因是它们为表达所关注蛋白的细胞赋予选择性优势,然而,SMG必须就其表型中性、高效性和多能性等原因进行优化。

[0007] 尽管有许多宿寄SMG的载体和表达系统可用,但在哺乳动物系统中实现的重组蛋白的表达无论在数量还是质量还是这两者上通常都无法令人满意。分子的生物“指纹”(例如翻译后修饰,比如糖基化)在开发重组蛋白治疗剂中在限定分子的实用性和功效方面尤为重要(Cumming,D.A.,1990,Glycobiology,1(2):115-130)。对表达的所关注蛋白的生物性质不会有负面影响的SMG是特别有利的。

[0008] 大多数SMG为细菌来源的,且由于日渐担忧细菌抗生素抗性基因对环境细菌的横向转移风险,就用于哺乳动物系统而言带来其他缺点(Breyer,D.等人,2014,Critical Reviews in Plant Sciences 33:286-330)。取消使用细菌抗生素抗性基因可能对于消费者认可和减轻此类感知风险具有积极作用。

[0009] 基因工程化的自体细胞在临床上快速取得成功(参见例如Kershaw,M.H.等人,2013,Nature Reviews:Cancer 13:525-541)。对于人类自体细胞产物中的基因修饰的载体的选择和设计是至关重要的,特别是因为将非人类组分不希望地引入人类自体细胞可能对患者的安全具有严重后果(Eaker等人2013,Stem cells Trans.Med.2:871-883;最先在2013年10月7日在线发表于SCTMEXPRESS)。仅具有哺乳动物来源而非细菌来源的组分的载体系统就用于过继性免疫治疗的患者特异性T细胞而言将是有利的。

[0010] 因此,期望引入哺乳动物选择性基因,特别是在用于产生所关注的哺乳动物蛋白的表达系统中为转化的细胞提供表型或代谢优势的那些基因。此外,可靠地表达足够高水平的治疗性蛋白并对该治疗性蛋白进行适当且一致的翻译后修饰的细胞系是高度期望的。

因此,在本领域中需要改善的哺乳动物表达系统。

发明内容

[0011] 使用哺乳动物衣霉素 (Tn) 抗性基因作为哺乳动物表达系统中的选择性标记可增加转染子的效率和拷贝数。已观察到,可操作地连接到所关注基因的Tn抗性基因的使用对哺乳动物细胞群产生选择性压力,从而增加转染子(即所关注基因)的随机整合。应当理解,选择性标记系统可促进所需转染子的选择,然而本发明的方法赋予所关注基因的效率 and 随机整合出人意料的增加,以及赋予所需蛋白可靠的生物质量。因此本发明的组合物和方法使得可以实现对所表达蛋白的质量上良好的翻译后修饰的有利选择。

[0012] 在一个方面,本发明提供包括哺乳动物衣霉素 (Tn) 抗性基因的分离的细胞,所述基因编码与SEQ ID NO:3氨基酸序列具有至少93%同一性的蛋白,所述基因可操作地连接到所关注基因 (GOI) 和至少一个调控元件。

[0013] 在另一个方面,本发明提供制备重组的所关注蛋白 (POI) 的方法,其中该方法包括:提供编码核酸分子的哺乳动物宿主细胞,所述核酸分子包含 (i) 哺乳动物衣霉素 (Tn) 抗性基因和 (ii) 编码该POI的基因;在第一浓度的Tn存在下培养该细胞;分离表达Tn抗性基因的至少一个拷贝的细胞群;在增加浓度的Tn存在下培养该细胞群,其中增加Tn的浓度增加POI的产生;以及从细胞培养中分离该POI。

[0014] 在又一个方面,本发明提供使N-聚糖蛋白底物糖基化的方法,其中该方法包括:提供编码核酸分子的哺乳动物宿主细胞,所述核酸分子包含可操作地连接到编码需要糖基化的蛋白底物的基因的哺乳动物衣霉素 (Tn) 抗性基因;在第一浓度的Tn存在下培养该细胞;分离表达Tn抗性基因的至少一个拷贝的细胞群;在增加浓度的Tn存在下培养该细胞群,其中增加Tn的浓度增加POI的产生;以及从细胞培养中分离该蛋白底物。

[0015] 在所述方法的一些实施方案中,该Tn抗性基因可操作地连接到编码POI的基因,且该编码POI的基因可操作地连接到至少一个调控元件。

[0016] 在一些实施方案中,该Tn抗性基因外源性添加到细胞。在其他实施方案中,该Tn抗性基因编码与SEQ ID NO:3的氨基酸序列具有至少93%同一性的蛋白。在其他实施方案中,该Tn抗性基因编码与SEQ ID NO:3的氨基酸序列具有至少94%同一性的蛋白。在一些实施方案中,该Tn抗性基因编码与SEQ ID NO:4的氨基酸序列具有至少93%同一性的蛋白。在另外其他实施方案中,该Tn抗性基因编码与SEQ ID NO:4的氨基酸序列具有至少94%同一性的蛋白。

[0017] 在一些实施方案中,该哺乳动物Tn抗性基因包括中国仓鼠 (*Cricetulus griseus*) Tn抗性基因。在其他实施方案中,该哺乳动物Tn抗性基因包括人类Tn抗性基因。

[0018] Tn抗性基因还可包括选自以下的核酸序列:SEQ ID NO:2、SEQ ID NO:11、SEQ ID NO:12、SEQ ID NO:13、SEQ ID NO:14、SEQ ID NO:15、SEQ ID NO:16和SEQ ID NO:17。

[0019] 在前述发明的某些实施方案中,该哺乳动物Tn抗性基因包括与SEQ ID NO:2的核酸序列具有至少92%同一性的核酸序列。在一些实施方案中,该哺乳动物Tn抗性基因包括与SEQ ID NO:12的核酸序列具有至少92%同一性的核酸序列。

[0020] 在本发明的分离的细胞中提供可操作地连接到Tn抗性基因的至少一个调控元件,其中该调控元件包括但不限于启动子、核糖体结合位点和增强子。在又一个实施方案中,该

G0I可操作地连接到启动子。在另一个实施方案中,该G0I可操作地连接到核糖体结合位点,诸如IRES。

[0021] 在一些实施方案中,本发明的分离的细胞和方法还包括第二所关注基因(G0I),而该G0I编码所关注蛋白(P0I)。在一个实施方案中,该所关注基因(G0I)为外源性添加的G0I。在另一个实施方案中,该外源性添加的G0I为人类基因。在又一个实施方案中,该调控元件为外源性添加的调控元件。

[0022] 在其他实施方案中,该第一和/或第二G0I编码包括但不限于抗体重链、抗体轻链、抗原结合片段和/或Fc融合蛋白的P0I。

[0023] 在另一个实施方案中,该第一G0I和第二G0I独立地选自编码抗体轻链或其抗原特异性片段、抗体重链或其抗原特异性片段、Fc融合蛋白或其片段以及受体或其配体特异性片段的基因。在一个实施方案中,重组酶识别位点存在于第一G0I与第二G0I之间。在其他实施方案中,本发明还提供在第一G0I的5'的重组酶识别位点和相对于第二G0I在3'的重组酶识别位点。

[0024] 在又一个实施方案中,该G0I编码选自抗体轻链或其抗原结合片段、抗体重链或其抗原结合片段、Fc融合蛋白或其片段、配体以及受体或其配体结合片段的糖蛋白。

[0025] 本发明的分离的、非天然存在的细胞可衍生自真核细胞。在一个实施方案中,该细胞为哺乳动物细胞。在一些实施方案中,该分离的细胞为离体人类细胞。在其他实施方案中,该细胞选自:CHO(例如CHO K1、DXB-11 CHO、Veggie-CHO)、COS(例如COS-7)、淋巴细胞、干细胞、视网膜细胞、Vero、CV1、肾(例如HEK293、293 EBNA、MSR 293、MDCK、HaK、BHK21)、HeLa、HepG2、WI38、MRC 5、Colo25、HB 8065、HL-60、Jurkat、Daudi、A431(表皮的)、CV-1、U937、3T3、L细胞、C127细胞、SP2/0、NS-0、MMT细胞、肿瘤细胞以及衍生自前述细胞的细胞系。在某些实施方案中,本发明的分离的细胞为CHO-K1细胞、淋巴细胞、视网膜细胞或干细胞。

[0026] 在一个实施方案中,第一浓度的Tn为1 μ g/mL。在另一个实施方案中,增加浓度的Tn包括第二和第三浓度的Tn。

[0027] 在一些实施方案中,第二浓度大于第一浓度的Tn,而第三浓度大于第二浓度的Tn。在某些实施方案中,第二浓度的Tn为2.5 μ g/ml,而第三浓度为5 μ g/mL。

[0028] 在另外其他实施方案中,增加浓度的Tn包括第二浓度的Tn,其中该第二浓度的Tn为2.5 μ g/ml或5 μ g/mL。

[0029] 除非另有说明或上下文中显而易见,否则本发明的任何方面和实施方案可与本发明的任何其他方面或实施方案结合使用。

[0030] 通过审阅接下来的详细说明,其他目标和优点将变得显而易见。

附图说明

[0031] 图1说明用于将编码所关注基因(例如eGFP)的核酸序列引入细胞基因组中的克隆载体结构中的可操作表达盒的示意图。SV40启动子:猴病毒40启动子;GPT:GlcNAc-1-P转移酶(例如CHO-GPT,SEQ ID NO:2;或hGPT,SEQ ID NO:12);IRES:内部核糖体进入位点;eGFP:增强的绿色荧光蛋白;SV40polyA:猴病毒40polyA。

[0032] 图2A至2C代表哺乳动物GPT氨基酸序列,即人类(GPT_HUMAN;UniProtKB登录号

Q9H3H5;SEQ ID NO:4)、恒河猴 (*Rhesus macaque*) (GPT_MACMU;UniProtKB登录号F6TXM3;SEQ ID NO:5)、黑猩猩 (GPT_PANTR;UniProtKB登录号H2R346;SEQ ID NO:6)、狗 (GPT_CANFA;UniProtKB登录号E2RQ47;SEQ ID NO:7)、豚鼠 (GPT_CAVPO;UniProtKB登录号E2RQ47;SEQ ID NO:8)、大鼠 (GPT_RAT;UniProtKB登录号Q6P4Z8;SEQ ID NO:9) 和小鼠 (GPT_MOUSE;UniProtKB登录号P42867;SEQ ID NO:10) 相较于中国仓鼠 (GPT_CRIGR;UniProtKB登录号P24140;SEQ ID NO:3) GPT氨基酸序列的比对。

[0033] 图3A和3B举例说明可如何使用本发明的方法和组合物实现蛋白优化。图3A描绘从用1 μ g/mL衣霉素 (Tn) 培养的第一细胞池中选择阳性细胞转染子的方法。随后,以增加浓度的衣霉素 (例如2.5 μ g/mL或5 μ g/mL) 培养第二细胞,以增强蛋白表达。图3B描绘从用1 μ g/mL衣霉素 (Tn) 培养的第一细胞池中选择阳性细胞转染子的方法,然后连续地增加后续细胞培养中的Tn浓度以便优化蛋白表达。

[0034] 图4A至4B经修饰的CHO细胞包括侧翼为lox位点的YFP基因。侧翼为lox位点的选择标记 (抗生素抗性基因和eGFP) 合并并在YFP位点并经由靶向整合用Cre重组酶置换YFP。随机整合体既表达YFP也表达eGFP。图4A:用Cre重组酶载体和包括eGFP的hpt表达载体转染细胞;但培养中无潮霉素。图4B:用Cre重组酶载体和包括eGFP的hpt表达载体转染细胞;存在400 μ g/mL潮霉素。

[0035] 图5A至5F显示代表衣霉素选择性的各种参数的FACS散点图。经修饰的CHO细胞包括侧翼为lox位点的YFP基因。侧翼为lox位点的选择标记 (抗生素抗性基因和eGFP) 合并并在YFP位点并经由靶向整合用Cre重组酶置换YFP。随机整合体既表达YFP也表达eGFP。图5A:用Cre重组酶载体和包括eGFP的CHO-GPT表达载体转染细胞;但培养中无衣霉素。图5B:用Cre重组酶载体和包括eGFP的CHO-GPT表达载体转染细胞;存在1 μ g/mL衣霉素。图5C:用Cre重组酶载体和包括eGFP的CHO-GPT表达载体转染细胞;存在2.5 μ g/mL衣霉素。图5D:用Cre重组酶载体和包括eGFP的人类GPT表达载体转染细胞;但培养中无衣霉素。图5E:用Cre重组酶载体和包括eGFP的人类GPT表达载体转染细胞;存在1 μ g/mL Tn。图5F:用Cre重组酶载体和包括eGFP的人类GPT表达载体转染细胞;存在2.5 μ g/mL Tn。

[0036] 图6A和6B显示GPT表达细胞池与非GPT表达池在其增强可操作地连接的GOI (诸如eGFP) 的表达的相对能力方面的比较。图6A如下说明通过PCR就细胞池所测得的CHO-GPT的基因拷贝相对数:无Tn选择的Pool-49细胞 (未添加外源性GPT);具有5 μ g Tn选择的Pool-49细胞 (无外源性GPT);Pool-1细胞天然表达较高量的GPT (数据未显示),并在无Tn选择的情况下进行测试;无Tn选择的Pool-78细胞 (无外源性GPT);表达外源性添加的hpt的CHO细胞和400 μ g/mL潮霉素选择;在1 μ g/mL Tn选择条件下表达外源性GPT的CHO细胞;从1 μ g/mL Tn选择池选择的表达外源性GPT的CHO细胞进一步在1 μ g/mL Tn中培养;从1 μ g/mL Tn选择池选择的表达外源性GPT的CHO细胞进一步在2.5 μ g/mL Tn中培养;从1 μ g/mL Tn选择池选择的表达外源性GPT的CHO细胞进一步在5 μ g/mL Tn中培养。图6B说明通过qPCR就相同的细胞池 (如图6A) 所测得的所关注基因eGPT的基因拷贝相对数目。

[0037] 图7A至7D如下说明由细胞培养所产生的Fc融合蛋白1FcFP1) 的糖型特征,图7A:使用标准方案的不表达GPT的CHO细胞 (批次B10002M410),相较于图7B:表达CHO-GPT且无Tn选择的CHO细胞 (批次110728)。图7C:表达CHO-GPT的且以1 μ g/mL Tn选择的CHO细胞 (批次110728-01),相较于图7D:表达CHO-GPT的且以5 μ g/mL Tn选择的CHO细胞 (批次110728-02)。

各色谱图如下指出含有唾液酸化残基的级分:0SA=0个唾液酸残基;1SA=1个唾液酸残基;2SA=2个唾液酸残基;3SA=3个唾液酸残基;4SA=4个唾液酸残基。

[0038] 图8说明从(A)批次B10002M410、(B)批次110728、(C)批次110728-01和(D)批次110728-02取样的Fc融合蛋白1(FcFP1)的叠加糖基化谱。从GPT批次中产生的各蛋白的糖基化谱(glycoprofile)与参比标准蛋白相容,且一致地产生主要的糖型种类。显然的是,相较于参比标准蛋白,在GPT批次中并未产生新的和独特的糖型种类。

具体实施方式

[0039] 在描述本发明的方法前,应当理解,本发明不限于所述的特定方法和实验条件,因为此类方法和条件可以变化。还应当理解,本文所用的术语仅用于描述特定实施方案的目的,而无意进行限制,因为本发明的范围将仅由所附权利要求书限制。

[0040] 如在该说明书和所附权利要求书中所用,除非上下文明确地相反指出,否则单数形式“一个”、“一种”和“该/所述”包括复数含义。因此,例如提及“一种方法”包括本文所述类型的和/或在阅读本公开后对本领域的技术人员将变得显而易见的一种或多种方法和/或一个或多个步骤。

[0041] 除非另有定义或另有规定,否则本文所用的所有技术和科学术语均具有与本发明所属领域的普通技术人员通常理解相同意义。

[0042] 尽管与本文中描述的那些方法和材料类似或等同的任何方法和材料均可用于本发明的实践或检验,但现在描述特定的方法和材料。本文提及的所有出版物均以引用方式整体并入本文。

[0043] 本领域中熟知的多种基因可向培养中的哺乳动物细胞赋予选择性表型。一般来讲,选择性标记基因表达蛋白,通常是在细胞培养中赋予对各种抗生素的抗性的酶。在某些选择条件下,可使表达荧光蛋白标记的细胞显现,并因此为可选择的。本领域的实例包括 β -内酰胺酶(*bla*; β -内酰胺抗生素抗性基因或*ampR*;氨基青霉素抗性基因)、*bls*(灭瘟素抗性乙酰转移酶基因)、潮霉素磷酸转移酶(*hpt*;潮霉素抗性基因)等。

[0044] 本文所述的方法依赖于使用衣霉素以及能让细胞耐衣霉素而得以在细胞培养中生长的酶(标记)。衣霉素(Tn)为抗生素的混合物,其作为细菌和真核生物N-乙酰葡萄糖胺转移酶的抑制剂,从而防止N-乙酰葡萄糖胺脂质中间体的形成以及新合成的糖蛋白的糖基化。(King,I.A.和Tabiwo,A.,1981,Effect of tunicamycin on epidermal glycoprotein and glycosaminoglycan synthesis in vitro.Biochem.J.,198(2):331-338)。Tn具有细胞毒性,因为其特异性抑制UDP-N-乙酰葡萄糖胺:磷酸多萜醇N-乙酰葡萄糖胺-1-P转移酶(GPT),一种催化多萜醇连接的寡糖类的生物合成起始步骤的酶。在衣霉素的存在下,在内质网(ER)中制备的天冬酰胺连接的糖蛋白不被N-连接的聚糖糖基化,并因此在ER内可能无法正确折叠且因而可被分解(Koizumi等人1999,Plant Physiol.121(2):353-362)。因此,Tn是在细菌和真核细胞中导致细胞凋亡的值得注意的未折叠蛋白反应(UPR)的诱导物。

[0045] 二磷酸尿苷GPT(也称为GlcNAc-1-P转移酶)的基因被鉴定为在某些细胞条件下过表达,以便于赋予对Tn的抗性(Criscuolo和Krag,1982,J Biol Chem,263(36):19796-19803;Koizumi等人,1999,Plant Physiology,第121卷,第353-361页)。编码GPT的基因(也描述为GenBank登录号M36899(SEQ ID NO:2))从耐Tn的中国仓鼠卵巢细胞系中分离出且编

码408个氨基酸的蛋白 (SEQ ID NO:3) (Scocca和Krag,1990,J Biol Chem 265 (33):20621-20626;Lehrman,M.等人,1988,J Biol Chem 263 (36):19796-803)。仓鼠GPT在酵母细胞(粟酒裂殖酵母(*S.pombe*))中过表达且在这些细胞中赋予Tn抗性;还提供方便的纯化GPT酶的来源(Scocca JR等人1995,Glycobiology,5 (1):129-36)。在杂交瘤细胞(表达IgG的B细胞,相比静止期B细胞)中分析了GPT的转录水平,尽管观察到产IgG的细胞并未表现出升高水平的GPT转录或活性,但在从静止期过渡到活化的B细胞中看到了GPT的少量增加。结论是,GPT水平可与B细胞中对LPS(抗原)刺激的增殖反应的早期产生相对应(Crick,D.C.等人1994,J Biol Chem 269 (14):10559-65)。

[0046] 此外,之前并不清楚,在有或无Tn存在下改变GPT的表达,在细胞表达系统中是否对蛋白产物的糖基化有影响并因而影响产物质量。应当理解,最佳和一致的糖基化在制备治疗性糖蛋白中至关重要的蛋白属性。

[0047] 本发明提供在哺乳动物细胞系统中利用哺乳动物Tn抗性基因GPT作为可调节的选择标记来制备重组蛋白的改进方法,而可操作地连接到GPT的所关注基因增加的拷贝数与GPT表达盒增加的随机整合到细胞中相关。

[0048] 本领域已认识到,制备治疗性蛋白特别是糖蛋白依赖于模拟此类蛋白的天然糖基化的哺乳动物型表达系统。(有关综述,参见Bork,K.等人,2009,J Pharm Sci.98 (10):3499-3508)。例如,某些糖蛋白的末端单糖(诸如N-连接的复合聚糖)通常被唾液酸占据。唾液酸化可影响糖蛋白的药代动力学性质,诸如吸收度、血清半衰期和清除率或糖蛋白的其他物理化学或免疫原性性质。过表达的重组糖蛋白通常具有不完整或不一致的糖基化。可靠的方法对于在哺乳动物细胞系中产生的治疗性糖蛋白的工艺一致性和质量至关重要。

[0049] 本发明还提供用于使重组蛋白糖基化的改进方法,即在哺乳动物细胞系统中制备糖蛋白的方法,以便提供一致质量产率的所需的蛋白。

[0050] 定义

[0051] DNA区当在功能上彼此相关时为可操作地连接的。例如,如果启动子能够参与序列的转录,则该启动子可操作地连接到编码序列;如果核糖体结合位点经定位成以便可以翻译,则该核糖体结合位点可操作地连接到编码序列。一般来讲,可操作地连接包括但不必须为邻近。就序列诸如分泌性前导序列而言,邻近且适当放置在阅读框中为典型的特征。增强生成的序列(诸如启动子)当与GOI功能上相关,例如其存在造成GOI的表达增加时,则可操作地连接到所关注基因(GOI)。

[0052] 因此,短语“可操作地连接的”(诸如在DNA表达载体构建体的上下文中)控制序列(例如启动子或操纵子或标记)适当地位于相对于编码序列的某一位置,使得该控制序列指导或允许由该编码序列编码的所关注多肽/蛋白的产生。例如,当选择标记为细胞在某些培养条件下存活所必需的时,该所关注基因可操作地连接到该选择标记基因,因为在没有可操作的选择标记蛋白的存在下,表达将不会发生。

[0053] 如本文所用的“启动子”是指足以指导其可操作地连接到的DNA序列转录的DNA序列,即其连接方式使得当适当的信号存在时,允许所关注基因和/或选择标记基因转录。基因的表达可置于本领域已知的任何启动子或增强子元件的控制下。

[0054] 在本发明的上下文中的“表达载体”可以是任何适合的载体,包括染色体、非染色体和合成的核酸载体(包括一组适合的表达控制元件的核酸序列)。此类载体的实例包括

SV40的衍生物、细菌质粒、嗜菌体DNA、杆状病毒、酵母质粒、衍生自质粒和嗜菌体DNA组合的载体、以及病毒核酸(RNA或DNA)载体。在一个实施方案中,编码Fc融合蛋白或多肽的核酸分子包括在裸DNA或RNA载体(包括例如线性表达元件(如描述于例如Sykes和Johnston,1997, Nat Biotech12,355-59中))、紧凑的核酸载体(如描述于例如US6,077,835和/或W000/70087中)或质粒载体诸如pBR322、pUC 19/18或pUC 118/119中。此类核酸载体及其用法是本领域熟知的(参见例如US5,589,466和US5,973,972)。

[0055] 如本文所用的“操纵子”是指这样的DNA序列,其引入基因中或基因附近的方式使得该基因可通过将阻遏蛋白与该操纵子结合而调控,并因此防止或允许GOI(即编码所关注多肽或蛋白的核苷酸)的转录。

[0056] 核糖体结合位包括“内部核糖体进入位”(IRES)或可包括5'帽。许多IRES序列是本领域熟知的。IRES代表翻译控制序列,其中IRES位点通常位于所关注基因的5',并使得以帽非依赖性方式翻译RNA。经转录的IRES可直接结合核糖体亚单位,以使得mRNA起始密码子的位置在核糖体中适当地取向以进行翻译。IRES序列通常位于mRNA的5' UTR中(起始密码子的正上游)。IRES在功能上取代对各种与真核生物翻译机制相互作用的蛋白因子的需求。

[0057] 术语“增强的”或“改善的”当用于描述蛋白表达时包括由本发明的表达系统或方法产生的蛋白(即基因产物)的量和/或质量一致性的增加。因此,这包括相比通常通过随机整合到基因组中所观察到的结果,例如相较于使用另一种选择性标记构建体的整合体池,在表达上增加至少约1.5倍至增加至少约3倍。因此,将所关注蛋白所观察到的表达增加倍数与在不存在包括GPT基因的本发明表达盒或细胞的情况下,或在存在包括不同选择性标记的表达盒或细胞的情况下,在基本上相同条件下测得的相同基因的表达水平进行比较。表达增加还可通过所产生的随机整合事件的数量来测量。增加的重组效率包括基因座重组能力的增强(例如,采用重组酶识别位点)。增强是指相比随机重组而言的可测量效率,随机重组通常为0.1%。在某些条件下,增强的重组效率是随机的大约10倍,或约1%。除非规定,否则受权利要求书保护的发明不限于特定的重组效率。表达增强也可通过所产生的基因拷贝数来测量,如通过定量聚合酶链式反应(qPCR)或其他熟知的技术来测量。

[0058] 增强或改善的产物还指更一致的质量,例如通过本发明的GPT表达系统所观察到的翻译后修饰。一致的质量包括例如在重复的生产线后具有所需的糖基化谱。一致性就质量而言是指均匀性和标准化程度,而重复的生产批次基本上无变化。在本文中教导了计算用以度量一致性的Z数。用于度量一致性的其他统计指标是本领域已知的。

[0059] 短语“选择压力”是施加给活的生物体(例如细胞)或系统(例如表达系统)的力或刺激,其改变该活的生物体或系统在给定环境中的行为和存活(例如存活能力)。

[0060] 短语“基因扩增”表示基因序列的相同拷贝的数目的增加。某些细胞过程的特征在于产生一种或多种特定基因的多个拷贝,其扩增该基因赋予细胞的表型,例如抗生素抗性。

[0061] 当关于表达盒采用短语“外源性添加的基因”或“外源性添加的GOI”时,该短语是指不存在于自然界中发现的细胞基因组内的任何基因,或整合至该基因组中(该基因组内的不同的基因座中)的另外的基因拷贝。例如,CHO基因组内的“外源性添加的基因”(例如选择性标记基因)可为并非在自然界中在特定CHO基因座内发现的仓鼠基因(即,来自仓鼠基因组中的另一基因座的仓鼠基因)、来自任何其他物种的基因(例如人类基因)、嵌合基因(例如人/小鼠),或者可以为并非在自然界中在CHO基因组内发现的仓鼠基因(即与来自仓

鼠基因组中的另一基因座的基因具有低于99.9%同一性的仓鼠基因),或并非在自然界中发现存在于CHO天然基因组内的任何其他基因。

[0062] 随机整合事件不同于靶向整合事件,而将基因插入细胞的基因组在随机整合事件中并非位点特异性的。靶向整合的实例是同源重组。随机(非同源)整合表示所产生的整合体的位置(基因座)并非已知的或指定的。随机整合被认为通过非同源末端接合(NHEJ)发生,然而并不限于该方法。

[0063] 选择效率表示表达选择性标记且适当时表达在选择性标记控制下的所关注蛋白的存活细胞的群体百分比。

[0064] 百分比同一性当描述Tn抗性蛋白时意在包括沿着邻接的同源区展示出所述同一性的同源序列,但在相比较的序列中不具有同源性的间隙、缺失或插入的存在不纳入百分比同一性的计算中。在该上下文中在解释“百分比同一性”的使用时,将涉及下列氨基酸序列比较:

[0065]

1	MWAFFPELPLPLLLVNLIGSLLGFVATVTLIPAFRSHFIAARLCGQDLNKLSQQQIPESQ	60	GPT_MOUSE
1	MWAFFPELPL--PLLVNLFGSLLGFVATVTLIPAFRSHFIAARLCGQDLNKLSRQQIPESQ	58	GPT_CRIG

[0066] 如本文所用,上述“GPT_CRIG”序列(对于中国仓鼠GPT)与小鼠同源物(“GPT_小鼠”)之间的“百分比同一性”测定不应包括仓鼠氨基酸10和11的比较,因为在比对中该仓鼠同源物并不具有同源的序列来进行比较(即小鼠GPT在该点具有插入,或该仓鼠同源物具有间隙或缺失,视具体情况而定)。因此,在上述比较中,百分比同一性比较将从5'端的“MWA”延伸至3'端的“ESQ”。在该事件中,小鼠同源物的差异仅在于:其在仓鼠GPT位置51具有“R”。因为该比较在一段60个碱基对中的58个邻接的碱基上进行,其中仅有一个氨基酸差异(其并非间隙、缺失或插入),所以两个序列(小鼠和仓鼠)之间从仓鼠GPT位置1至仓鼠GPT位置58有超过98%的同一性(因为“百分比同一性”并不包括间隙、缺失和插入的罚分)。虽然上述实例以氨基酸序列为基础,但是应当理解,核酸序列百分比同一性将以相同的方式来计算。

[0067] 术语“细胞”包括适合表达重组核酸序列的任何细胞。细胞包括原核生物和真核生物(单细胞或多细胞)的细胞、细菌细胞(例如,大肠杆菌(*E. coli*)、芽孢杆菌属菌种(*Bacillus spp.*)、链霉菌属菌种(*Streptomyces spp.*)等的菌株)、分枝杆菌细胞、真菌细胞、酵母细胞(例如,酿酒酵母(*S. cerevisiae*)、粟酒裂殖酵母(*S. pombe*)、巴斯德毕赤酵母(*P. pastoris*)、甲醇毕赤酵母(*P. methanolica*)等)、植物细胞、昆虫细胞(例如,SF-9、SF-21、杆状病毒感染的昆虫细胞、粉纹夜蛾(*Trichoplusia ni*)等)、非人动物细胞、哺乳动物细胞、人细胞或细胞融合物例如杂交瘤或四源杂交瘤。在某些实施方案中,细胞为人、猴、猿、仓鼠、大鼠或小鼠细胞。在其他实施方案中,细胞为真核细胞并且选自以下细胞:CHO(例如CHO K1、DXB-11 CHO、Veggie-CHO)、COS(例如COS-7)、视网膜细胞、Vero、CV1、肾(例如HEK293、293 EBNA、MSR 293、MDCK、HaK、BHK21)、HeLa、HepG2、WI38、MRC 5、Colo25、HB 8065、HL-60、Jurkat、Daudi、A431(表皮的)、CV-1、U937、3T3、L细胞、C127细胞、SP2/0、NS-0、MMT细胞、肿瘤细胞和来源于前述细胞的细胞系。在一些实施方案中,细胞包含一个或多个病毒基因,例如表达病毒基因的视网膜细胞(例如PER.C6®细胞)。

[0068] 短语“累积细胞密度”或“ICD”表示在一段时间内作为整体考虑的培养基中的细胞密度,以每毫升的细胞-天数来表示。在一些实施方案中,ICD以培养中约第12天的细胞来测

量。

[0069] “糖基化”或短语“使蛋白糖基化”包括形成糖蛋白，而寡糖附连到蛋白的天冬酰胺(Asn)残基(即N-连接的)或丝氨酸(Ser)/苏氨酸(Thr)残基(即O-连接的)的侧链。聚糖可以为单糖残基的均聚物或杂聚物，其可以为直链的或支链的。N-连接的糖基化已知主要在内质网中开始，而O-连接的糖基化显示出在ER或高尔基体中开始。

[0070] “N-聚糖蛋白”或“N-聚糖蛋白底物”包括含有或可接受N-连接的寡糖的蛋白。N-聚糖可由N-乙酰基半乳糖胺(GalNAc)、甘露糖(Man)、果糖(Fuc)、半乳糖(Gal)、神经氨酸(NANA)和其他单糖组成，然而N-聚糖通常具有共同的核心五糖结构，其包括：三个甘露糖和两个N-乙酰葡萄糖胺(GlcNAc)糖。带有连续氨基酸序列(即序列段(sequon))Asn-X-Ser或Asn-X-Thr的蛋白(其中X为除脯氨酸之外的任何氨基酸)可提供N-聚糖的附连位点。

[0071] 一般说明

[0072] 本发明至少部分地基于以下发现：在某些条件下可在细胞中产生重组蛋白，其中编码该蛋白的基因可操作地连接到Tn抗性基因GPT，且将产蛋白的细胞的选择设定为在细胞基因组中增加随机整合事件并因此增加所关注基因的拷贝数以及最终增加蛋白产量。

[0073] 本发明还至少部分地基于以下发现：产蛋白的细胞可经优化以表达具有一致和可靠的翻译后修饰的蛋白。GPT表达盒还可以如在表达构建体中诸如经由表达载体使用本领域已知的各种基因编辑技术而整合到细胞基因组中。包括GPT的表达载体可通过随机或靶向重组，诸如同源重组或由识别特定重组位点的重组酶所介导的重组(例如，Cre-lox介导的重组)整合到基因组中。

[0074] 可通过在整合位点在染色体DNA中引入断裂来促进真核细胞中的同源重组。模型系统已证明，如果在染色体靶序列中引入双链断裂，那么在基因靶向期间同源重组的频率会增加。这可通过将某些核酸酶靶向特定整合位点而实现。在靶基因座识别DNA序列的DNA结合蛋白是本领域已知的。基因靶向载体也用于促进同源重组。在不在于同源指导修复(homology directed repair)的基因靶向载体的情况下，细胞常常会通过非同源末端接合(NHEJ)(其可能在裂解位点处导致多个核苷酸的缺失或插入)来闭合该双链断裂。基因靶向载体构建和核酸酶选择在本发明所属领域的技术人员的技术范围内。

[0075] 在一些实例中，具有模块化结构并含有单独锌指域的锌指核酸酶(ZFN)在靶向序列中识别特定的3-核苷酸序列(例如靶向整合的位点)。一些实施方案可利用具有靶向多个靶序列的单独锌指域的组合的ZFN。

[0076] 转录活化因子样(TAL)效应子核酸酶(TALEN)也可用于位点特异性基因组编辑。TAL效应子蛋白DNA结合域通常与限制性核酸酶诸如FokI的非特异性裂解域组合使用。在一些实施方案中，将包含TAL效应子蛋白DNA结合域和限制性核酸酶裂解域的融合蛋白用于识别和裂解本发明基因座内的靶序列处的DNA(Boch J等人,2009Science 326:1509-1512)。

[0077] RNA引导的核酸内切酶(RGEN)是从细菌适应性免疫机制开发的可编程的基因组工程化工具。在该系统(成簇规律间隔短回文重复序列(CRISPR)/CRISPR相关性(Cas)免疫反应)中，蛋白Cas9当与两个RNA(其中一个引导靶选择)复合时形成序列特异性核酸内切酶。RGEN由组分(Cas9和tracrRNA)以及靶特异性CRISPR RNA(crRNA)组成。DNA靶裂解的效率以及裂解位点的位置均基于前间区序列邻近基序(PAM)的位置而变化，该基序是针对靶识别的额外要求(Chen,H.等人,J.Biol.Chem.2014年3月14作为手稿M113.539726在线发表)。

[0078] 另外其他同源重组方法可供技术人员使用,诸如具有精确DNA结合特异性的BuD衍生的核酸酶(BuDN)(Stella,S.等人Acta Cryst.2014,D70,2042-2052)。精确的基因组修饰方法基于与基因组内的独特靶序列相容可获得的工具来选择,以使得避免细胞表型被破坏。

[0079] 本发明提供细胞和方法以供稳定地将核酸序列(所关注基因)整合到哺乳动物细胞中,其中该核酸序列能够通过GPT序列整合来增强表达。本发明还提供组合物和方法以供与表达构建体(例如表达载体)相关地使用GPT,以及将外源性GPT添加到所关注的哺乳动物细胞中。本发明提供细胞和方法以供用于一致但稳健地制备糖蛋白特别是治疗性糖蛋白的方法。

[0080] GPT选择标记盒的构建

[0081] 本文提供包括可操作的GPT表达盒的表达载体。该表达盒包括必需的调控元件以允许和驱动哺乳动物GPT和所需基因产物的转录和翻译。

[0082] 还可开发本文所述的基因和调控序列的各种组合。还可开发的本文所述的适当序列的其他组合的实例包括含有本文所公开的GPT基因的多个拷贝的序列,或通过将所公开的GPT与其他核苷酸序列组合以实现最佳调控元件组合而衍生的序列。此类组合可连续地连接或排列以提供向所关注基因和调控元件取向的最佳的GPT间距。

[0083] 编码GPT的基因的同源序列已知存在于来自其他哺乳动物物种的细胞中(比如人类;参见图2)以及衍生自其他哺乳动物组织类型的细胞系中,且可通过本领域熟知的技术来分离。示例性哺乳动物GPT氨基酸序列的列表提供于图2中。可对SEQ ID NO:2和11-17所示的核苷酸序列进行核苷酸序列的改变,诸如密码子优化,以便允许SEQ ID NO:3-10所示的对应GPT蛋白的最佳表达。此外,可通过改变编码GPT的核苷酸序列,来改变SEQ ID NO:3-10所示的氨基酸序列。此类技术包括但不限于本领域熟知的定点或随机诱变技术。

[0084] 然后将所得的GPT变体就本文所述的GPT活性进行测试,例如对衣霉素的抗性进行测试。通过常规实验,将氨基酸序列与具有GPT活性的SEQ ID NO:3至少约93%相同、或至少约95%相同、或至少约96%相同、或至少约97%相同、或至少约98%相同的GPT蛋白分离出,且预期其表现出与SEQ ID NO:3相同的Tn抗性、选择性效率和翻译后益处。因此,GPT的哺乳动物同源物和GPT变体也可由本发明的实施方案涵盖。图2A至2C显示各种哺乳动物GPT氨基酸序列(即SEQ ID NO:3-10)的比对。哺乳动物GPT序列(核酸和氨基酸)在仓鼠、人类、小鼠和大鼠基因组之中为保守的。表1鉴别示例性哺乳动物GPT蛋白及其同源性程度。

[0085] 表1A:GPT同源物的氨基酸同一性

动物	SEQ ID NO	%人类同一性	%小鼠同一性	%大鼠同一性	%仓鼠同一性
仓鼠	3	93.87	96.08	96.08	-
小鼠	10	94.12	-	97.07	96.08
人类	4	-	94.12	93.63	93.87
大鼠	9	93.63	97.07	-	96.08

[0087] 表1B:代表性GPT同源物的核酸同一性

动物	SEQ ID NO	%仓鼠同一性
仓鼠	2	-

小鼠	11	92
人类	12	92
大鼠	13	94
猕猴	14	92
黑猩猩	15	92

[0089] 可使用本文所提供的GPT/衣霉素法来产生表达增加水平的所关注蛋白的细胞群。取决于蛋白被细胞加工的效率如何,表达的绝对水平将随特定蛋白而变化。

[0090] 因此,本发明还包括选自SEQ ID NO:2和11-17的表达GPT的核苷酸序列。本发明还涵盖表达GPT的核苷酸序列,该序列与选自SEQ ID NO:2和11-17的核苷酸序列至少92%相同、至少93%相同、至少94%相同、至少95%相同、至少96%相同、至少98%相同或至少99%相同。

[0091] 本发明包括含有SEQ ID NO:1、SEQ ID NO:2或SEQ ID NO:12的载体。包括哺乳动物GPT基因和任选的调控元件的载体包括供瞬时或稳定转染的载体。

[0092] 在一个实施方案中,GPT基因用于增强GOI的表达,如图1所示。图1显示与IRES序列和GPT选择性标记可操作地连接的GOI。GPT盒还包括启动子序列,例如SV40启动子,和多聚腺苷酸化(poly(A))序列,例如SV40poly(A)。

[0093] 表达增强盒(包括GPT和上游启动子)最佳地整合在细胞基因组中。使用本发明的方法,GOI在以增加浓度的Tn为基础的培养条件下在GPT表达盒内表达(图3A或图3B)。FACS读出结果(诸如图5B、5C、5E和5F中所示)举例说明稳定转染的细胞群中的表达分布,特别是使用哺乳动物Tn抗性选择标记CHO-GPT和hGPT在选择效率上的急剧增加。哺乳动物GPT表达还增强了所关注基因产物的表达,例如荧光蛋白eGFP的产生。相较于使用GPT在以一种浓度的Tn为基础的培养条件下在表达系统中表达的GOI,增加浓度的Tn的连续培养导致约2倍的增强表达,诸如图6B中所举例说明。

[0094] 本发明包括包含这样的GPT基因的哺乳动物细胞,其中该GPT基因为外源性的且通过本发明的方法整合到细胞基因组中。包括这样的GPT基因的细胞具有至少一个外源性添加的所关注基因(GOI),该基因在GPT基因的上游或下游。

[0095] 在多种实施方案中,可通过将GOI置于哺乳动物选择性标记GPT的控制下来增强GOI的表达。在其他实施方案中,可通过将GOI置于哺乳动物选择性标记GPT的控制下并提供包括大于0.5 μ g/mL Tn浓度的细胞培养条件,来增强GOI的随机整合事件。在一些实施方案中,细胞培养条件包括大于1 μ g/mL的Tn浓度。调控元件可以可操作地连接到GOI,其中GOI的表达—在GOI和GPT的选定距离下(在5'或3'方向)—保留增强GOI表达的能力,例如,超过通常由于随机整合事件所观察到的表达。在多种实施方案中,增强为至少约1.5倍至约2倍或更多。相较于随机整合或随机表达,表达的增强为约1.5倍或约2倍或更多。

[0096] 在另一个实施方案中,使用本发明的方法和组合物可得到一致地糖基化的蛋白。如表4中所示,经Tn处理的GPT/GOI重组蛋白批次使得重复批次具有同等的糖基化谱。因此,增强的蛋白表达诸如一致的糖基化谱可直接通过计算如本文所教导的Z数来比较。Z数方程式将色谱图上代表唾液酸(SA)部分的相对波峰数以及各波峰的相对形状和强度列入考虑。Z数以各波峰所占据的面积为基础并且可用作复合糖蛋白的一致性指标(参见例如图7A-7D、图8和实施例3,如本文所述)。

[0097] 蛋白表达优化还可就各GOI而实现,包括例如表达盒取向或密码子优化。蛋白优化还可通过在细胞培养法中改变递增的Tn浓度来实现。

[0098] 重组表达载体可包括编码蛋白的合成的或cDNA衍生的DNA片段,其可操作地连接到衍生自哺乳动物、病毒或昆虫基因的适合的转录和/或翻译调控元件。此类调控元件包括转录启动子、增强子、编码适合的mRNA核糖体结合位点的序列以及控制转录和翻译终止的序列,如本文详细描述。哺乳动物表达载体还可包括非转录元件,诸如复制起点、其他5'或3'侧翼非转录序列,以及5'或3'非翻译序列诸如剪接供体和受体位点。还可并入帮助识别转染子的另外的选择性标记基因(诸如荧光标记)。

[0099] 在另一个实施方案中,载体包括编码所关注蛋白的核酸分子(或所关注基因),其包括含有所述核酸分子(基因)的表达载体,其中该核酸分子(基因)可操作地连接到表达控制序列。

[0100] 提供包括所关注基因(GOI)的载体,其中该GOI可操作地连接到适合在哺乳动物宿主细胞中表达的表达控制序列。

[0101] 可用于本发明的有用的启动子包括但不限于SV40早期启动子区、包含在劳氏肉瘤病毒的3'长末端重复序列中的启动子、金属硫蛋白基因的调控序列、小鼠或人类巨细胞病毒IE启动子(Gossen等人,(1995)Proc.Nat.Acad.Sci.USA89:5547-5551)、花椰菜花叶病毒35S RNA启动子以及光合成酶核酮糖二磷酸羧化酶的启动子、来自酵母或其他真菌的启动子元件诸如Gal 4启动子、ADC(醇脱氢酶)启动子、PGK(磷酸甘油激酶)启动子、碱性磷酸酶启动子以及下列动物转录控制区,其表现出组织特异性并已用于转基因动物:弹性蛋白酶I、胰岛素、免疫球蛋白、小鼠乳腺肿瘤病毒、白蛋白、甲胎蛋白、 α 1-抗胰蛋白酶、 β -球蛋白和肌球蛋白轻链-2。

[0102] 本发明的核酸分子还可以可操作地连接到有效的poly(A)终止序列(例如SV40 poly(A))、用于大肠杆菌中质粒产物的复制起点和/或方便的克隆位点(例如多接头)。核酸还可包括与组成型启动子相反的可调节的诱导型启动子(可诱导的、可抑制的、发育性调节的),诸如CMV IE(技术人员将认识到,此类术语实际上为某些条件下的基因表达程度的描述语)。

[0103] 本发明提供制备所关注蛋白的方法,而提供包括所关注基因(GOI)的表达载体。此类表达载体可用于重组制备任何所关注蛋白。可用于转染脊椎动物细胞的表达载体中的转录和翻译控制序列可通过病毒来源而提供。例如,常用的启动子和增强子衍生自病毒,诸如多瘤病毒、腺病毒2、猴病毒40(SV40)和人类巨细胞病毒(CMV)。病毒基因组启动子、控制和/或信号序列可用于驱动表达,所提供的此类控制序列与所选择的宿主细胞相容。取决于在其中表达重组蛋白的细胞类型,还可使用非病毒细胞启动子(例如, β -球蛋白和EF-1 α 启动子)。

[0104] 衍生自SV40病毒基因组的DNA序列例如SV40起点、早期和晚期启动子、增强子、剪接和多聚腺苷化位点可用于提供对异源DNA序列的表达有用的其他基因元件。早期和晚期启动子是特别有用的,因为二者可容易地从SV40病毒作为还含有SV40病毒复制起点的片段得到(Fiers等人,Nature 273:113,1978)。也可使用较小或较大的SV40片段。通常,包括从位于SV40复制起点中的Hind III位点向BglII位点延伸的大约250bp序列。

[0105] 用于表达多个转录物的双顺反子表达载体先前已有描述(Kim S.K.和Wold B.J.,

Cell 42:129,1985;Kaufman等人1991,同上)并可与GPT表达系统组合使用。其他类型的表达载体也将是有用的,例如描述于美国专利第4,634,665号(Axel等人)和美国专利第4,656,134号(Ringold等人)中的那些。

[0106] 整合位点(例如重组酶识别位点)可置于编码POI的基因序列的5'或3'。适合的整合位点的一个实例为lox p位点。适合的整合位点的另一个实例为两个重组酶识别位点,例如选自lox p位点、lox和lox 5511位点。

[0107] 基因扩增盒及其表达载体

[0108] 先前已描述或本领域已知的有用调控元件也可包括在用于转染哺乳动物细胞的核酸中。图1举例说明GPT载体中的可操作盒,其还包括启动子序列、IRES序列、所关注基因和poly(A)序列。

[0109] 在本发明的上下文中的表达载体可以是任何适合的载体,包括染色体、非染色体和合成的核酸载体(包括一组适合的表达控制元件的核酸序列)。此类载体的实例包括SV40的衍生物、细菌质粒、嗜菌体DNA、杆状病毒、酵母质粒、衍生自质粒和嗜菌体DNA组合的载体、以及病毒核酸(RNA或DNA)载体。在一个实施方案中,编码抗体的核酸分子包含在裸DNA或RNA载体中,包括例如线性表达元件(如描述于例如Sykes和Johnston,Nat Biotech 12,355-59(1997)中)、紧凑的核酸载体(如描述于例如US 6,077,835和/或WO 00/70087中)或质粒载体诸如pBR322、pUC 19/18或pUC 118/119。此类核酸载体及其用法为本领域熟知的(参见例如US 5,589,466和US 5,973,972)。

[0110] 或者,表达载体可为适合在酵母系统中表达的载体。可使用任何适合在酵母系统中表达的载体。适合的载体包括例如包含组成型或诱导型启动子的载体,诸如酵母 α 因子、醇氧化酶和PGH(综述见于:F.Ausubel等人编Current Protocols in Molecular Biology, Greene Publishing and Wiley InterScience New York(1987)以及Grant等人,Methods in Enzymol 153,516-544(1987))。

[0111] 在某些实施方案中,载体包括编码所关注蛋白的核酸分子(或所关注基因),其包括含有所述核酸分子(基因)的表达载体,其中该核酸分子(基因)可操作地连接到适合在宿主细胞中表达的表达控制序列。

[0112] 表达控制序列经工程化以控制和驱动所关注基因的转录,及后续在各种细胞系统中的蛋白表达。质粒将可表达的所关注基因与包括所需调控元件比如启动子、增强子、选择性标记、操纵子等的表达控制序列(即表达盒)组合。在本发明的表达载体中,GPT和所关注蛋白诸如编码抗体的核酸分子可包括以下部分或与它们相关:任何适合的启动子、增强子、操纵子、阻遏蛋白、poly(A)终止序列和其他表达促进元件。

[0113] 所关注基因(例如编码抗体的核苷酸序列)的表达可置于本领域已知的任何启动子或增强子元件的控制下。此类元件的实例包括强表达启动子(例如,人类CMV IE启动子/增强子或CMV主要IE(CMV-MIE)启动子,以及RSV、SV40晚期启动子、SL3-3、MMTV、泛素(Ubi)、泛素C(UbC)和HIV LTR启动子)。

[0114] 在一些实施方案中,该载体包括选自以下的启动子:SV40、CMV、CMV-IE、CMV-MIE、RSV、SL3-3、MMTV、Ubi、UbC和HIV LTR。

[0115] 本发明的核酸分子还可以可操作地连接到有效的poly(A)终止序列、用于大肠杆菌中质粒产物的复制起点、作为选择性标记的抗生素抗性基因、和/或方便的克隆位点(例

如多接头)。核酸还可包括与组成型启动子相反的可调节的诱导型启动子(可诱导的、可抑制的、发育性调节的),诸如CMV IE(技术人员将认识到,此类术语实际上为某些条件下的基因表达程度的描述语)。

[0116] 选择性标记是本领域熟知的元件。在某些情况下,除了GPT之外,还可以使用另外的选择性标记,其中此类标记使细胞可见。可使用正选择或负选择。

[0117] 在一些实施方案中,该载体包括一种或多种编码绿色荧光蛋白(GFP)、增强的绿色荧光蛋白(eGFP)、氰基荧光蛋白(CFP)、增强的氰基荧光蛋白(eCFP)、黄色荧光蛋白(YFP)或增强的黄色荧光蛋白(eYFP)的选择性标记基因。

[0118] 就本发明的目的而言,真核细胞中的基因表达可使用强启动子紧密调节,该启动子通过操纵子来控制,而该操纵子继而由调节性融合蛋白(RFP)来调节。RFP基本上由转录阻断域和调节其活性的配体结合域组成。此类表达系统的实例描述于US20090162901A1中,其全文以引用方式并入本文。

[0119] 许多真核细胞和嗜菌体中的操纵子已研究透彻(Neidhardt编*Escherichia coli* and *Salmonella*; Cellular and Molecular Biology第2版第2卷ASM Press, Washington D.C. 1996)。这些包括但不限于大肠杆菌的LexA基因的操纵子区,其与LexA肽结合,以及乳糖和色氨酸操纵子,其与由大肠杆菌的LacI和trpR基因编码的阻遏蛋白结合。这些还包括来自 λ P_R和噬菌体P22ant/mnt基因的嗜菌体操纵子,其与由 λ cI和P22arc编码的阻遏蛋白结合。在一些实施方案中,当阻遏蛋白的转录阻断域为限制酶诸如NotI时,操纵子为该酶的识别序列。本领域的技术人员将认识到,操纵子必须位于启动子附近或位于启动子的3',以使得其能够通过该启动子控制转录。例如,美国专利第5,972,650号(以引用方式并入本文)规定了tetO序列在TATA盒的特定距离内。在具体实施方案中,该操纵子优选地置于启动子的正下游。在其他实施方案中,该操纵子置于启动子的10个碱基对内。

[0120] 在某些实施方案中,该操纵子选自:tet操纵子(tetO)、NotI识别序列、LexA操纵子、乳糖操纵子、色氨酸操纵子和Arc操纵子(AO)。在一些实施方案中,阻遏蛋白选自TetR、LexA、LacI、TrpR、Arc、 λ C1和GAL4。在其他实施方案中,该转录阻断域衍生自真核阻遏蛋白,例如衍生自GAL4的阻遏域。

[0121] 在示例性细胞表达系统中,细胞经工程化以表达四环素阻遏蛋白(TetR)且所关注蛋白置于启动子的转录控制下,而该启动子的活性受TetR调控。两个一前一后的TetR操纵子(tetO)置于载体中的CMV-MIE启动子/增强子的正下游。由这样的载体中的CMV-MIE启动子所指导的编码所关注蛋白的基因的转录在不存在四环素或在某些其他适合的诱导物(例如强力霉素)的情况下可被TetR阻断。在诱导物的存在下,TetR蛋白能结合tetO,因而使所关注蛋白的转录然后使翻译(表达)发生(参见例如美国专利第7,435,553号,其全文以引用方式并入本文)。

[0122] 另一种示例细胞表达系统包括调节性融合蛋白,诸如TetR-ER_{LBD}T2融合蛋白,其中该融合蛋白的转录阻断域为TetR而配体结合域为具有T2突变的雌激素受体配体结合域(ER_{LBD}) (ER_{LBD}T2; Feil等人(1997) *Biochem. Biophys. Res. Commun.* 237:752-757)。当tetO序列置于强CMV-MIE启动子的下游和附近时,来自CMV-MIE/tetO启动子的所关注核苷酸序列的转录在他莫昔芬的存在下被阻断,并通过移除他莫昔芬解除阻断。在另一个实例中,使用融合蛋白Arc2-ER_{LBD}T2(由通过15个氨基酸的接头所连接的两个Arc蛋白所组成的单链二聚

体和ER_{LBD}T2(同上)所组成的融合蛋白)涉以及Arc操纵子(AO),更具体地讲,CMV-MIE启动子/增强子正下游的两个一前一后的arc操纵子。细胞系可由Arc2-ER_{LBD}T2调节,其中表达所关注蛋白的细胞由CMV-MIE/ArcO2启动子驱动并可通过移除他莫昔芬来诱导。(参见例如US 20090162901A1,其以引用方式并入本文)。

[0123] 在一些实施方案中,本发明的载体包括CMV-MIE/TetO或CMV-MIE/AO2杂交启动子。

[0124] 本发明的载体还可将Cre-lox工具用于重组技术,以便有利于所关注基因的复制。Cre-lox策略需要至少两个组成部分:1) Cre重组酶,一种催化两个loxP位点之间的重组的酶;以及2) loxP位点(例如,由8-bp核心序列(重组在其中进行)和两个侧翼13-bp反向重复序列所组成的特异性34-碱基对bp序列)或突变lox位点。(参见例如Araki等人PNAS 92:160-4(1995);Nagy,A.等人Genesis26:99-109(2000);Araki等人Nuc Acids Res 30(19):e103(2002)和US20100291626A1,其全部以引用方式并入本文)。在另一种重组策略中,酵母衍生的FLP重组酶可与共有序列FRT一起使用(另见例如Dymecki,S.PNAS93(12):6191-6196(1996))。

[0125] 在另一个方面,基因(即编码本发明重组多肽的核苷酸序列)插入表达盒的GPT基因的上游或下游,且任选可操作地连接到启动子,其中该启动子连接的基因的5'侧翼为第一重组酶识别位点而3'侧翼为第二重组酶识别位点。此类重组酶识别位点允许表达系统的宿主细胞中Cre介导的重组。在一些情况下,第二启动子连接的基因在第一基因的下游(3')且3'侧翼为第二重组酶识别位点。在另外其他情况下,第二启动子连接的基因的5'侧翼为第二重组酶识别位点,而3'侧翼为第三重组酶识别位点。在一些实施方案中,重组酶识别位点选自loxP位点、lox511位点、lox2272位点和FRT位点。在其他实施方案中,重组酶识别位点为不同的。在又一个实施方案中,宿主细胞包括能表达Cre重组酶的基因。

[0126] 在一个实施方案中,载体包括编码本发明的抗体轻链或抗体重链的第一基因,以及编码本发明的抗体轻链或抗体重链的第二基因。

[0127] 在一些实施方案中,载体还包括X-盒结合蛋白1(mXBP1)基因,其能够通过内质网(ER)中的蛋白折叠所涉及到的基因表达的控制进一步增强蛋白产生/蛋白分泌。(参见例如Ron D和Walter P.Nat Rev Mol Cell Biol.8:519-529(2007))。

[0128] 任何细胞均适于表达本发明的重组核酸序列。用于本发明的细胞包括哺乳动物细胞,诸如非人类动物细胞、人类细胞或细胞融合物,比如杂交瘤或四源杂交瘤。在某些实施方案中,细胞为人、猴、仓鼠、大鼠或小鼠细胞。在其他实施方案中,细胞为真核细胞并且选自以下细胞:CHO(例如CHO K1、DXB-11CHO、Veggie-CHO)、COS(例如COS-7)、视网膜细胞、Vero、CV1、肾(例如HEK293、293 EBNA、MSR 293、MDCK、HaK、BHK21)、HeLa、HepG2、WI38、MRC 5、Colo25、HB 8065、HL-60、Jurkat、Daudi、A431(表皮的)、CV-1、U937、3T3、L细胞、C127细胞、SP2/0、NS-0、MMT细胞、肿瘤细胞和来源于前述细胞的细胞系。在一些实施方案中,细胞包含一个或多个病毒基因,例如表达病毒基因的视网膜细胞(例如PER.C6®细胞)。

[0129] 在又一个方面,本发明涉及重组的哺乳动物宿主细胞,诸如转染瘤,其产生免疫球蛋白,诸如抗体或双特异性分子。此类宿主细胞的实例包括工程化的哺乳动物细胞,诸如CHO或HEK细胞。例如,在一个实施方案中,本发明提供包括稳定整合到细胞基因组中的核酸的细胞,该核酸包括编码抗体表达的序列,而该抗体包括本发明的重组多肽。在另一个实施方案中,本发明提供包括非整合(即附加型)核酸(诸如质粒、粘粒、噬粒或线性表达元件)的

细胞,该核酸包括编码抗体表达的序列,而该抗体包括本发明的重组多肽。在其他实施方案中,本发明提供通过用包括本发明表达载体的质粒来稳定转染宿主细胞而产生的细胞系。

[0130] 因此,在一个方面,本发明提供包含以下部分的细胞:(a) 编码外源性添加的哺乳动物GPT基因的重组多核苷酸以及(b) 编码多亚基蛋白的多核苷酸。在一些实施方案中,外源性添加的GPT基因与SEQ ID NO:2的核酸序列90%相同,其非限制性实例提供于SEQ ID NO:11-17中,且多亚基蛋白为抗体。在其他实施方案中,细胞还含有外源性添加的GPT基因和调控元件。在一个实施方案中,细胞为哺乳动物细胞,诸如用于制备生物药品的CHO细胞。

[0131] 在另一个方面,本发明提供衍生自在之前方面所述的细胞的细胞系。所谓“衍生自”是指由各个细胞无性繁殖的并具有某些选定品质(诸如以给定效价产生活性蛋白的能力,或增殖至特定密度的能力)的细胞群。在一些实施方案中,衍生自具有编码哺乳动物GPT基因的重组多核苷酸和编码多亚基蛋白的多核苷酸的细胞的细胞系能够以至少3克每升(g/L)培养基、至少5g/L或至少8g/L的效价产生多亚基蛋白。在一些实施方案中,该细胞系可达到比由基本上相同但无编码GPT的重组多核苷酸的细胞所衍生的细胞系所能达到的累积细胞密度(ICD)大至少30%、大至少50%、大至少60%或大至少90%的累积细胞密度。

[0132] 提供了扩增GOI的方法。举例说明的方法将增加浓度的衣霉素施用于真核GPT表达系统,从而扩增可操作地连接到外源性添加的哺乳动物GPT基因的GOI的基因拷贝。

[0133] 所关注蛋白

[0134] 编码所关注基因的核酸序列可方便地整合到包括Tn抗性标记基因和IRES,且任选侧翼有重组酶识别位点的细胞中。可使用任何适合在哺乳动物细胞中表达的所关注蛋白,然而糖蛋白将尤其受益于本发明的方法。例如,所关注蛋白可为抗体或其抗原结合片段、双特异性抗体或其片段、嵌合抗体或其片段、ScFv或其片段、Fc标签蛋白(例如Trap蛋白)或其片段、生长因子或其片段、细胞因子或其片段、细胞表面受体的胞外域或其片段。

[0135] 具有天冬酰胺连接的(N-连接的)聚糖的糖蛋白普遍存在于真核细胞中。这些聚糖的生物合成及其向多肽的转化在内质网(ER)中发生。N-聚糖结构在ER和高尔基复合体中通过许多糖苷酶和糖基转移酶进一步修饰。使用本发明进行的蛋白制备针对天然N-聚糖结构的一致性,以消除免疫原性表位(“糖表位”(glycotope))。

[0136] 使用本发明的方法,重组蛋白批次表现出有利的特性。重复的蛋白生产批次的HPLC(以荧光检测)证明了糖蛋白具有统一的表达和糖基化模式,如本文图7-8中举例说明。提供了使N-聚糖蛋白底物糖基化的方法,其中提供编码核酸分子的哺乳动物宿主细胞,该核酸分子包含可操作地连接到编码需要糖基化的蛋白底物的基因的哺乳动物衣霉素(Tn)抗性基因;将该细胞在第一浓度的Tn存在下培养;将表达Tn抗性基因的至少一个拷贝的细胞群分离;将该细胞群在增加浓度的Tn存在下培养;以及从细胞培养中分离N-聚糖蛋白底物。蛋白底物的N-聚糖含量可就单糖、寡糖的存在通过本领域已知的任何方法来评估。

[0137] 聚糖连接的蛋白的详细结构分析可与蛋白的功能特征相关联。这种对蛋白糖基化进行表征的分析通常涉以及以下几个步骤:i) 酶或化学释放所连接的聚糖;ii) 经由芳族或脂族胺的还原性胺化或全甲基化使释放的聚糖衍生化;iii) 分析聚糖。许多分析糖基化模式的变化是技术人员已知的。糖蛋白可带有以特定数量占据各个位点的数种形式的糖型,并因此而言,其复杂性可能使其难以在某些生产方法中重现。糖型的类型和数量的一致性是可测量的且代表治疗性蛋白生产的期望结果。

[0138] 宿主细胞和转染

[0139] 用于本发明方法中的哺乳动物宿主细胞为真核宿主细胞,通常为哺乳动物细胞,包括例如CHO细胞和小鼠细胞。在一个实施方案中,本发明提供包括编码衍生自中国仓鼠(*Cricetulus griseus*)的Tn抗性标记蛋白的核酸序列(如SEQ ID NO:3所示)或其同源物或变体的细胞。在一些实施方案中,该细胞包括Tn抗性标记基因的多基因拷贝。在其他实施方案中,本发明提供编码衍生自人类(SEQ ID NO:4)、猕猴(SEQ ID NO:5)、黑猩猩(SEQ ID NO:6)、狗(SEQ ID NO:7)、豚鼠(SEQ ID NO:8)、大鼠(SEQ ID NO:9)或小鼠(SEQ ID NO:10)的Tn抗性标记蛋白的核酸序列。

[0140] 本发明包括用本发明的表达载体转染的哺乳动物宿主细胞。经转染的宿主细胞包括经表达载体转染的细胞,而该表达载体包括编码所关注蛋白或多肽的序列。所表达的蛋白将通常取决于所选的核酸序列而分泌到培养基中,但可能保持在细胞中或沉积在细胞膜中。各种哺乳动物细胞培养系统均可用于表达重组蛋白。适合的哺乳动物宿主细胞系的实例包括Gluzman (1981) Cell 23:175所述的猴肾脏细胞的COS-7细胞系,以及能够表达适当载体的其他细胞系,包括例如CV-1/EBNA (ATCC CRL 10478)、L细胞、C127、3T3、CHO、HeLa和BHK细胞系。针对特定选择或扩增流程而开发的其他细胞系也可与本文所提供的方法和组合物一起使用。在本发明的一个实施方案中,该细胞为称为K1的CHO细胞系(即CHO K1细胞)。为了实现重组蛋白高产量的目标,宿主细胞系应在适当的情况下预先适应生物反应器培养基。

[0141] 本领域已知数种转染方法,它们在Kaufman (1988) Meth. Enzymology 185:537中进行了综述。所选的转染方案将取决于宿主细胞类型和GOI性质,且可基于常规实验来选择。任何这样的方案的基本要求是首先将编码所关注蛋白的DNA引入适合的宿主细胞中,然后以相对稳定、可表达的方式鉴定和分离已包含异源DNA的宿主细胞。

[0142] 可用于将异源DNA引入哺乳动物细胞的某些试剂包括Lipofectin™试剂和Lipofectamine™试剂(Gibco BRL, Gaithersburg, Md.)。这两种试剂均为用于形成脂质-核酸复合物(或脂质体)的市售试剂,当应用于培养的细胞时,有利于核酸摄入细胞中。

[0143] 所选的转染方案和选择用于其中的元件将取决于所用的宿主细胞的类型。本领域的技术人员知道许多不同的方案和宿主细胞,并且可基于所用的细胞培养系统的要求来选择用于表达所需蛋白的适当系统。在另一个方面,本发明涉及编码多肽的表达载体,其包括但不限于抗体、双特异性抗体、嵌合抗体、ScFv、抗原结合蛋白或Fc融合蛋白。此类表达载体可用于使用本发明的方法和组合物进行的重组多肽制备。

[0144] 本发明的其他特征在以下对示例性实施方案的描述过程中将变得显而易见,这些实施方案用于说明本发明而无意限制本发明。

[0145] 实施例

[0146] 提供以下实施例是为了向本领域的普通技术人员描述如何构造和使用本发明的方法和组合物,而非旨在限制发明人所认为的其发明的范围。已努力确保有关所用的数字(例如量、温度等)的准确性,但应考虑某些实验误差和偏差。除非另有指出,否则份数为重量份,分子量为平均分子量,温度为摄氏度,而压力为大气压或接近大气压。

[0147] 实施例1表达GPT的转染细胞的选择效率。

[0148] 将经修饰的CHO K1细胞用含有CHO-GPT (SEQ ID NO:2)、人类GPT (SEQ ID NO:12)

的质粒载体或含有潮霉素磷酸转移酶 (Hpt, 潮霉素抗性基因) 的质粒载体转染; 例如, 将选择性标记基因 (CHO-GPT或hpt) 在其相应载体中经由IRES序列在转录上连接到下游eGFP基因。例如, 各质粒被构建成为在5'至3'方向上包含下列基因序列: Lox位点、SV40晚期启动子、CHO-GPT (或Hpt)、IRES、增强的绿色荧光蛋白 (eGFP) 和第二Lox位点。将纯化的重组质粒与表达Cre重组酶的质粒共同转染进从5'至3'在转录活性基因座处含有以下位点的经修饰的CHO宿主细胞: lox位点、YFP和第二lox位点。因此, 宿主CHO细胞可通过流式细胞术作为绿色阳性或黄色阴性细胞分离。当表达eGFP (转录上受GPT或hpt基因调控) 的重组质粒与表达Cre重组酶的质粒共同转染时, 由Cre重组酶介导的重组在含有lox位点的染色体基因座处导致GPT/eGFP盒的位点特异性整合, 并且发生YFP基因的取代 (即绿色阳性细胞)。若eGFP随机整合, 则将产生绿色阳性和黄色阳性两种细胞。

[0149] 将细胞群用0、1 μ g/ml、2.5 μ g/ml或5 μ g/ml衣霉素 (Tn) 或400 μ g潮霉素 (Hyg) 温育, 如表2所概述。通过荧光激活细胞分选 (FACS) 分析来测量观察到的重组群体 (ORP)。对细胞进行分选以定量各细胞群, 并针对仅表达GFP而不表达YFP的细胞计算选择效率 (图4或5)。

[0150] 比较耐Tn或Hyg的细胞池之间的选择效率 (表达GFP的存活细胞的群体百分比) (表2)。

[0151] 表2: 选择效率

[0152]	Hpt 或 GPT	Cre	选择剂 (μ g/ml Hyg 或 Tn)	选择效率% (总 GFP+)
	Hpt	+	-	1.35
	Hpt	+	+(400 Hyg)	98.8
	choGPT	+	-	0.89
	choGPT	+	+(1Tn)	86.9
[0153]	choGPT	+	+(2.5 Tn)	96.1
	hGPT	+	-	2.6
	hGPT	+	+(1Tn)	97
	hGPT	+	+(2.5 Tn)	96.7

[0154] 据观察, 衣霉素选择与潮霉素选择同样高效。CHO-GPT和人类GPT在1 μ g/ml或2.5 μ g/ml衣霉素存在下在整合体的选择方面均是高效的。

[0155] 实施例2基因产物的扩增。

[0156] 通过将增加浓度的衣霉素施用于GPT表达系统来进行增量选择。用含有如上CHO-GPT基因 (SEQ ID NO:2) 的质粒载体转染CHO K1细胞。该质粒以5'至3'方向含有第一Lox位点、SV40晚期启动子、CHO-GPT基因、IRES、eGFP和第二Lox位点。CRE-lox位点指导所关注基因整合到基因组中, 从而产生每个细胞具有至少一个GPT插入的稳定的转染细胞池。(如上所见, 由于随机整合, 可产生更多的整合体)。最初在1 μ g/ml衣霉素 (Tn) 存在下培养CHO细胞。然后从稳定的池 (称为细胞池2) 中选择转染子, 随后在1 μ g/ml、2.5 μ g/ml或5 μ g/ml Tn存在下扩增。进行数轮选择以鉴别能增强eGFP表达 (多个拷贝) 的细胞群。在2.5 μ g/ml或5 μ g/ml Tn存在下, 随机整合事件大大增加。

[0157] 使用标准qPCR法测量基因产物CHO GPT、eGFP或mGapdh(归一化对照)的拷贝数。来自进一步用2.5 μ g/ml Tn温育的1 μ g/ml Tn抗性池的细胞中的eGFP拷贝数为来自进一步用1 μ g/ml Tn温育的1 μ g/ml Tn抗性池的eGFP拷贝数的至少2倍。当经1 μ g/ml Tn处理的池进一步用5 μ g/ml Tn温育时,基因拷贝数进一步增加。eGFP基因拷贝数的增加与增加的CHO-GPT基因拷贝相关联。(参见图6A和6B)。

[0158] 为了确定基因拷贝数的增加是否会转变成增加的蛋白表达,针对经多轮Tn选择(即1、2.5或5 μ g Tn)处理过的表达GPT和eGFP的相同细胞池,通过FACS测量了平均荧光强度(MFI)(参见例如图6B中的实例7、8和9)。这些细胞池的eGFP表达的比较在表3中示出。

[0159] 就eGFP生成而言,接受第二轮5 μ g Tn选择的表达GPT的细胞池,相较于1 μ g Tn处理,产量恰好为2.5倍,相较于2.5 μ g Tn处理,产量为1.5倍(表3)。

[0160] 表3:eGFP蛋白产生

GPT 1 μ g池+第二Tn(μ g) 处理	MFI
1 μ g	1098
2.5 μ g	1867
5 μ g	2854

[0162] 不受任一理论的束缚,Tn浓度的递增以受控的方式放大了对细胞的选择压力,从而提高产量。

[0163] 如下所述,还将Tn抗性表达载体用于另外的实验,以检测Tn选择对糖基化模式的作用。

[0164] 实施例3示例性二聚蛋白的表达和糖基化谱。

[0165] 用含有GPT的表达载体转染表达“Trap”蛋白(Fc融合蛋白1,下文称为FcFP1)的CHO细胞。该质粒以5'至3'方向具有Lox位点、SV40晚期启动子、Tn抗性基因(CHO-GPT)、IRES eGFP、SV40polyA和第二Lox位点。使用1 μ g/mL Tn或5 μ g/mL Tn进行GPT选择性标记的选择。通过悬浮培养于无血清的生产培养基中来扩增所选的细胞池。通过FACS分析,由eGFP的表达确认了GPT转染。将从所选的池收集的团块送去进行GPT表达的拷贝数分析,并建立12天的生产率分析,以测定用不同浓度的衣霉素选择的池中的FcFP1表达水平。

[0166] 对FcFP1就其复合糖基化模式进行选择,其具有丰富的糖基化位点。为了测定糖基化谱,将表达FcFP1蛋白的细胞在标准方案(无Tn)或Tn处理条件(如表4所示)下进行细胞培养而扩增,然后分离并纯化蛋白。

[0167] 表4:FcFP1蛋白产生

蛋白	批号	处理
FcFP1 Trap	110728	无
FcFP1 Trap	110728-1	1 μ g/ml Tn
FcFP1 Trap	110728-2	5 μ g/ml Tn

[0169] 使用基于熟知的HPLC方法的色谱法和荧光邻氨基苯甲酸(AA)标签(Anumula和Dhume,Glycobiology,8(7):685-694,1998),就各批次糖蛋白进行详细的聚糖分析,以确定Tn对糖基化谱是否有负面影响。还将生产批次与代表治疗上可接受的蛋白批次的参照标准进行了比较。代表性聚糖分析如图7A-7D所示。各批次(相较于参照批次)一致地产生相同的波峰数目、相对形状和相对强度。各色谱图的重叠(图8)表明并无独特或异常的波峰露出。

[0170] 通过熟知的HPLC方法针对参照标准批次就FcFP1蛋白进行了寡糖性质分析。就FcFP1 trap蛋白批次测量了唾液酸化的量,并计算了各批次(3个重复)的Z数。Z数代表各批次之间变异的量度。Z数考虑相对波峰数以及各波峰的相对形状和强度。例如,图7A-7D中各0SA、1SA、2SA、3SA和4SA波峰的面积如表5经定量。

[0171] 表5:寡糖定量分析

蛋白批次	重复	0SA	1SA	2SA	3SA	4SA	OS 谱 (Z 数)
参考 B100002M410	1	6506.43	13388.34	11268.60	5176.21	1728.15	1.53
	2	5869.80	11932.32	10159.21	4196.10	1550.09	1.51
	3	6870.18	14536.84	12090.21	5200.58	1707.74	1.51
	平均值	6415.47	13285.83	11172.65	4857.63	1661.99	1.52±0.01
FcFP1 Trap 110728	1	6159.09	9394.92	7368.03	3074.66	675.48	1.34
	2	7530.49	12117.03	9589.08	2951.63	810.09	1.36
	3	5508.95	8580.56	6902.59	3794.81	630.79	1.34
	平均值	6399.51	10030.84	7953.23	3074.66	705.45	1.35±0.01
FcFP1 trap 110728-1	1	5330.22	8149.81	6539.33	2490.06	641.37	1.35
	2	5034.39	9009.42	7059.61	2698.05	812.21	1.40
	3	6222.44	10235.08	8428.04	3276.75	848.83	1.39
	平均值	5529.02	9131.44	7342.33	2821.62	767.47	1.38±0.03
FcFP1 trap 110728-2	1	6300.77	10001.93	8109.12	3000.96	790.99	1.36
	2	5999.09	9952.47	7968.58	2885.50	717.70	1.36
	3	4322.29	6176.33	5187.48	1742.26	458.52	1.32
	平均值	5540.72	8710.24	7088.39	2542.91	655.74	1.35±0.02

[0173] OS=寡糖;0SA=0个唾液酸残基;1SA=1个唾液酸残基;2SA=2个唾液酸残基;3SA=3个唾液酸残基;4SA=4个唾液酸残基

[0174]
$$Z \text{ 数} = \frac{(\text{面积 } 1SA + 2 * \text{面积 } 2SA + 3 * \text{面积 } 3SA + 4 * \text{面积 } 4SA)}{(\text{面积 } 0SA + \text{面积 } 1SA + \text{面积 } 2SA + \text{面积 } 3SA + \text{面积 } 4SA)}$$

[0175] 相较于参照批次,各批所计算的Z数在可接受范围内,因此应当理解,各蛋白批次达到与治疗分子相同的物质。因为Tn的存在已知对于N-连接的糖蛋白的糖基化具有负面作用,因此,鉴于由Tn产生的增加的选择压力条件,蛋白产生和产率是可靠和一致的这种情况出人意料。

[0176] 在不背离本发明的精神或实质的情况下,本发明可以体现为其他具体实施方案。

- [0001] 序列表
- [0002] <110> Regeneron Pharmaceuticals, Inc.
- [0003] Deshpande, Dipali
- [0004] Burakov, Darya
- [0005] Chen, Gang
- [0006] Fandl, James P.
- [0007] <120> 重组蛋白的高效选择性
- [0008] <130> 8700W0
- [0009] <150> US 62/039,416
- [0010] <151> 2014-08-19
- [0011] <160> 17
- [0012] <170> PatentIn 3.5版
- [0013] <210> 1
- [0014] <211> 6964
- [0015] <212> DNA
- [0016] <213> 人工序列
- [0017] <220>
- [0018] <223> 合成的
- [0019] <400> 1
- [0020] aagcttatac tcgagctcta gattgggaac cgggtctct cgaattcgag atctagttaa 60
- [0021] aacacgcggc cgctaactcag ccataccaca tttgtagagg ttttacttgc tttaaaaaac 120
- [0022] ctcccacacc tccccctgaa cctgaaacat aaaatgaatg caattgttgt tgtaacttg 180
- [0023] tttattgcag cttataatgg ttacaaataa agcaatagca tcacaaattt cacaaataaa 240
- [0024] gcattttttt cactgcatte tagttgtggt ttgtccaaac tcatcaatgt atcttateat 300
- [0025] gtctaccggt ataacttctg ataatgtata ctatacgaag ttagccggtg gggcccctct 360
- [0026] cttcatgtga gcaaaaggcc agcaaaaggc caggaaccgt aaaaaggccg cgttgctggc 420
- [0027] gtttttccat aggctccgcc cccctgacga gcatcacaaa aatcgacgct caagtcagag 480
- [0028] gtggcgaaac ccgacaggac tataaagata ccaggcgttt ccccctggaa gtcacctctg 540
- [0029] gcgctctcct gttccgacce tgccgcttac cggatactg tccgccttcc tcccttcggg 600
- [0030] aagcgtggcg ctttctcata gctcacgctg taggtatctc agttcgggtg aggtcgttcc 660
- [0031] ctccaagctg ggctgtgtgc acgaaccccc cgttcagccc gaccgctgcg ccttatccgg 720
- [0032] taactatcgt cttgagtcca acccgtaag acacgactta tcgccactgg cagcagccac 780
- [0033] tggtaacagg attagcagag cgaggtatgt aggcggtgct acagagttct tgaagtgggtg 840
- [0034] gcctaactac ggctacacta gaagaacagt atttggtatc tgcgctctgc tgaagccagt 900
- [0035] taccttcgga aaaagagttg gtagctcttg atccggcaaa caaaccaccg ctggtagcgg 960
- [0036] tggttttttt gtttgcaagc agcagattac gcgcagaaaa aaaggatctc aagaagatcc 1020
- [0037] tttgatcttt tctacggggt ctgacgctca gtggaacgaa aactcacggt aagggatttt 1080
- [0038] ggteatgggc gcgcctcata ctctgcagg catgagatta tcaaaaagga tcttcaccta 1140
- [0039] gatcctttta aattaaaaat gaagttttaa atcaatctaa agtatatatg agtaaacttg 1200
- [0040] gtctgacagt taccaatgct taatcagtga ggcacctatc tcagcgatct gtctatttcg 1260
- [0041] ttcatccata gttgcctgac tccccgtcgt gtagataact acgatacggg agggcttacc 1320

[0042]	atctggcccc	agtgctgcaa	tgataccgcg	agaccacgc	tcaccggctc	cagatttatac	1380
[0043]	agcaataaac	cagccagccg	gaagggccga	gcgcagaagt	ggtcctgcaa	ctttatccgc	1440
[0044]	ctccatccag	tctattaatt	gttgccggga	agctagagta	agtagttcgc	cagttaatag	1500
[0045]	tttgcgcaac	gttggtgcca	ttgctacagg	catcgtggtg	tcacgctcgt	cgtttggtat	1560
[0046]	ggcttcattc	agctccggtt	cccaacgatac	aaggcgagtt	acatgatccc	ccatgttgtg	1620
[0047]	caaaaaagcg	gtagctcct	tcggctctcc	gatcgttctc	agaagtaagt	tggccgcagt	1680
[0048]	gttatcactc	atggttatgg	cagcactgca	taattctctt	actgtcatgc	catccgtaag	1740
[0049]	atgcttttct	gtgactggtg	agtactcaac	caagtcattc	tgagaatagt	gtatgcggcg	1800
[0050]	accgagttgc	tcttgcccgg	cgtcaatacg	ggataatact	gcgccacata	gcagaacttt	1860
[0051]	aaaagtctc	atcattggaa	aacgttttct	ggggcgaaaa	ctctcaagga	tcttaccgct	1920
[0052]	gttgagatcc	agttcgatgt	aaccactcgc	tgcacccaac	tgatcttcag	catcttttac	1980
[0053]	tttaccagc	gtttctgggt	gagcaaaaac	aggaaggcaa	aatgccgcaa	aaaagggat	2040
[0054]	aagggcgaca	cggaaatgtt	gaatactcat	actcttctt	tttcaatatt	attgaagcat	2100
[0055]	ttatcagggt	tattgtctca	tgagcggata	catattttaa	tgtatttaga	aaaataaaca	2160
[0056]	aataggggtt	ccgcgcacat	ttccccgaaa	agtgccacct	gacgtcaggt	acacaacttc	2220
[0057]	gtatagcata	cattatacga	agttatggta	ccaagcctag	gcctccaaaa	aagcctctc	2280
[0058]	actacttctg	gaatagctca	gaggcagagg	cggcctcggc	ctctgcataa	ataaaaaaaaa	2340
[0059]	ttagtcagcc	atggggcgga	gaatgggcgg	aactgggcgg	agttaggggc	gggatgggcg	2400
[0060]	gagttagggg	cgggactatg	gttgctgact	aattgagatg	catgctttgc	atacttctgc	2460
[0061]	ctgctgggga	gcctggggac	tttccacacc	tggttgctga	ctaattgaga	tgcatgcttt	2520
[0062]	gcatacttct	gcctgtggg	gagcctggg	actttccaca	ccggatccac	catgtgggcc	2580
[0063]	ttcccgaggt	tgccgctgcc	gctgctggtg	aatttgctgc	gctcgtcgtc	gggatttgtg	2640
[0064]	gctactgtga	ccctcatccc	tgcttccgt	agccacttta	tcgccgcgcg	cctctgtggc	2700
[0065]	caggacctca	acaagctcag	ccggcagcag	atcccagaat	cccaggaggt	gatctgcggt	2760
[0066]	gctgttttcc	ttatcatcct	cttctgcttc	atcccttcc	ccttctgaa	ctgctttgtg	2820
[0067]	gaggagcagt	gtaaggcatt	ccccaccat	gaatttggg	ccctgatagg	tgccctcctt	2880
[0068]	gccatctgct	gcatgatctt	cctgggcttc	gctgatgatg	tactcaatct	gcgctggcgc	2940
[0069]	cataagctgc	tgctgcccac	agctgcctct	ctacctctcc	tcatggttta	cttactaac	3000
[0070]	tttgcaata	caaccattgt	ggtacceaa	cccttccgct	ggattcttgg	cctgcatttg	3060
[0071]	gacttgggaa	tctataacta	tgtctacatg	ggactgcttg	cggtgttctg	taccaatgcc	3120
[0072]	atcaacatcc	tagcaggaat	taatggccta	gaggctggtc	agtcactagt	catctctgct	3180
[0073]	tctatcattg	tcttcaacct	ggtagagctg	gaaggtgatt	atcgggatga	tcatgtcttt	3240
[0074]	tcctctact	tcatgatacc	atTTTTTTT	accaccttgg	gattgctata	ccataactgg	3300
[0075]	taccatcac	aggtgtttgt	gggagatacc	ttctgttatt	ttgctggcat	gacctttgcc	3360
[0076]	gtggtgggaa	tcttgggaca	cttcagcaag	accatgctac	tcttctttat	tccacaagtg	3420
[0077]	ttcaatttcc	tctactcgt	gcctcagctc	cttcacgcca	tcccctgcc	tcgacaccgc	3480
[0078]	ataccagac	tcaatccgaa	gacgggcaaa	ctggagatga	gctattccaa	gttcaagacc	3540
[0079]	aagaacctct	ctttcttggg	cacctttatt	ttaaaggtag	cagagcgct	ccagctagtg	3600
[0080]	acagttcacc	gaggcgagag	tgaggatggt	gccttcaactg	aatgtaacaa	catgaccctc	3660
[0081]	atcaacttgc	tactcaaaat	ctttgggcc	atacatgaga	gaaacctcac	actgctcctg	3720
[0082]	ctgcttttgc	agatcctgag	cagcgtctgc	accttctcca	ttcgatacca	gcttgtccga	3780
[0083]	ctcttctatg	atgtctgaac	gcgtcccccc	tctcctccc	ccccccctaa	cgttactggc	3840

[0084]	cgaagccgct	tggaataagg	ccggtgtgcg	tttgtctata	tgttattttc	caccatattg	3900
[0085]	ccgtcttttg	gcaatgtgag	ggccccgaaa	cctggccctg	tcttcttgac	gagcattcct	3960
[0086]	aggggtcttt	cccctctcgc	caaaggaatg	caaggtctgt	tgaatgtcgt	gaaggaagca	4020
[0087]	gttcctctgg	aagcttcttg	aagacaaaca	acgtctgtag	cgaccctttg	caggcagcgg	4080
[0088]	aacccccac	ctggcgacag	gtgcctctgc	ggccaaaagc	cacgtgtata	agatacacct	4140
[0089]	gcaaagcgg	cacaaccca	gtgccacggt	gtgagttgga	tagtttgga	aagagtcaaa	4200
[0090]	tggtctcct	caagctatt	caacaaggg	ctgaaggatg	cccagaaggt	acccattgt	4260
[0091]	atgggatctg	atctggggcc	tcggtgcaca	tgctttacat	gtgttttagtc	gaggttaaaa	4320
[0092]	aacgtctagg	ccccccaac	cacggggacg	tggttttct	ttgaaaaaca	cgattgctcg	4380
[0093]	aatcaccatg	gtgagcaagg	gcgaggagct	gttcaccggg	gtggtgccca	tcttggtcga	4440
[0094]	gctggacggc	gacgtaaagc	gccacaagtt	cagcgtgtcc	ggcgagggcg	agggcgatgc	4500
[0095]	cacctacggc	aagctgacct	tgaagttcat	ctgcaccacc	ggcaagctgc	ccgtgccctg	4560
[0096]	gcccacctc	gtgaccacct	tgacctacgg	cgtgcagtgc	ttcagccgct	accccgacca	4620
[0097]	catgaagcag	cacgacttct	tcaagtccgc	catgcccgaa	ggctacgtcc	aggagcgcac	4680
[0098]	catcttcttc	aaggacgacg	gcaactacaa	gacccgcgcc	gaggtgaagt	tcgagggcga	4740
[0099]	cacctggtg	aaccgcatcg	agctgaaggg	catcgacttc	aaggaggacg	gcaacatcct	4800
[0100]	ggggcacaag	ctggagtaca	actacaacag	ccacaacgtc	tacatcatgg	ccgacaagca	4860
[0101]	gaagaacggc	atcaaggtga	acttcaagat	ccgccacaac	atcgaggacg	gcagcgtgca	4920
[0102]	gctcggcgac	cactaccagc	agaacacccc	catcggcgac	ggccccgtgc	tgctgcccga	4980
[0103]	caaccactac	ctgagcacct	agtcggcct	gagcaaagac	cccaacgaga	agcgcgatca	5040
[0104]	catggtcctg	ctggagtteg	tgaccggcgc	cgggatcact	ctcggcatgg	acgagctgta	5100
[0105]	caagtaatcg	gcccctaatac	agccatacca	catttgtaga	ggttttactt	gctttaaaaa	5160
[0106]	acctcccaca	cctccccctg	aacctgaaac	ataaaatgaa	tgcaattggt	gttgtttaact	5220
[0107]	tgtttattgc	agcttataat	ggttacaaat	aaagcaatag	catcacaat	ttcacaata	5280
[0108]	aagcattttt	ttactgcat	tctagttgtg	gtttgtccaa	actcatcaat	gtatcttatac	5340
[0109]	atgtcggcgc	gttgacattg	attattgact	agttattaat	agtaatcaat	tacggggtca	5400
[0110]	ttagttcata	gcccataat	ggagttccgc	gttacataac	ttacggtaaa	tgccccgct	5460
[0111]	ggctgaccgc	ccaacgacct	ccgcccattg	acgtcaataa	tgacgtatgt	tcccatagta	5520
[0112]	acgccaatag	ggactttcca	ttgacgtcaa	tggttgaggt	atctacggta	aactgcccac	5580
[0113]	ttggcagtac	atcaagtgtg	tcatatgcca	agtaccccc	ctattgacgt	caatgacggt	5640
[0114]	aaatggcccc	cctggcatta	tgcccagtac	atgaccttat	gggactttcc	tacttgccag	5700
[0115]	tacatctacg	tattagtcac	cgctattacc	atggtgatgc	ggttttgga	gtacatcaat	5760
[0116]	gggcgtggat	agcggtttga	ctcacgggga	tttccaagtc	tccaccccat	tgacgtcaat	5820
[0117]	gggagtttgt	tttggcacca	aatcaacgg	gactttccaa	aatgtcgtaa	caactccgcc	5880
[0118]	ccattgacgc	aaatggcgg	taggcgtgta	cggtgggagg	tctatataag	cagagctctc	5940
[0119]	cctatcagtg	atagagatct	ccctatcagt	gatagagatc	gtcgacgttt	agtgaaccgt	6000
[0120]	cagatgcct	ggagacgcca	tccacgtgt	ttgacctcc	atagaagaca	ccgggaccga	6060
[0121]	tccagcctcc	gcgccggga	acggtgcatt	ggaacgcgga	ttccccgtgc	caagagtgc	6120
[0122]	gtaagtaccg	cctatagagt	ctataggccc	accccttg	cttcttatgc	atgctatact	6180
[0123]	gttttggtc	tggtgtctat	acacccccgc	ttctcatgt	tataggtgat	ggtatagctt	6240
[0124]	agcctatagg	tgtgggttat	tgaccattat	tgaccactcc	cctattggtg	acgatacttt	6300
[0125]	ccattactaa	tcataacat	ggctctttgc	cacaactctc	tttattgget	atatgccaat	6360

[0126]	acactgtcct tcagagactg acacggactc tgtatTTTTa caggatgggg tctcatttat	6420
[0127]	tatttcaaaa ttcacatata caacaccacc gtccccagtg cccgcagttt ttattaaaca	6480
[0128]	taacgtggga tctccacgcg aatctcgggt acgtgttccg gacatgggct cttctccggt	6540
[0129]	agcggcggag cttctacatc cgagccctgc tcccatgcct ccagcgactc atggtcgctc	6600
[0130]	ggcagctcct tgctcctaac agtggaggcc agacttaggc acagcacgat gccaccacc	6660
[0131]	accagtgtgc cgcacaaggc cgtggcggta gggatatgtg ctgaaaatga gctcggggag	6720
[0132]	cgggettga ccgctgacgc atttggaaga ctttaaggcag cggcagaaga agatgcaggc	6780
[0133]	agctgagttg ttgtgtctg ataagagtca gaggtaactc ccgttgccgt gctgttaacg	6840
[0134]	gtggagggca gtgtagctg agcagtactc gttgctgcc cgcgccac cagacataat	6900
[0135]	agctgacaga ctaacagact gttccttcc atgggtctt tctgcagtca ccgtcctga	6960
[0136]	cacg	6964
[0137]	<210> 2	
[0138]	<211> 1231	
[0139]	<212> DNA	
[0140]	<213> 中国仓鼠	
[0141]	<400> 2	
[0142]	caccatgtgg gccttcccgg agttgccgct gccgctgctg gtgaatttgt tcggctcget	60
[0143]	gctgggattt gtggctactg tgaccctcat ccctgcctc cgtagccact ttatcgccgc	120
[0144]	gcgcctctgt ggccaggacc tcaacaagct cagccggcag cagatcccag aatcccagg	180
[0145]	agtgatctgc ggtgctgtt tcttatcat cctcttctgc ttcatecct tccccttct	240
[0146]	gaactgctt gtggaggagc agtghtaaggc attccccac catgaattg tggccctgat	300
[0147]	aggtgccctc cttgccatct gctgcatgat cttcctgggc ttcgctgat atgtactcaa	360
[0148]	tctgcgctgg cgccataagc tgctgctgcc cacagctgcc tctctacct tctcatggt	420
[0149]	ttacttcaact aactttggca atacaacat tgtgttacc aagcccttc gctggattct	480
[0150]	tggcctgcat ttggacttgg gaatcctata ctatgtctac atgggactgc ttgcggtgtt	540
[0151]	ctgtaccaat gccatcaaca tcttagcagg aattaatggc ctagaggctg gtcagtcact	600
[0152]	agtcactctt gcttctatca ttgtcttcaa cctggtagag ctggaagggt attatcgga	660
[0153]	tgatcatgtc ttttccctct acttcatgat accattttt tttaccacct tgggattgct	720
[0154]	ataccataac tggtagccat cacaggtgt tgtgggagat accttctgt attttctgg	780
[0155]	catgacctt gccgtggtgg gaatcttgg acacttcagc aagaccatgc tactcttctt	840
[0156]	tattccaaa gtgttcaatt tctctactc gctgcctcag ctccttcag ccateccctg	900
[0157]	ccctcgacac cgcatacca gactcaatcc gaagacgggc aaactggaga tgagctattc	960
[0158]	caagttcaag accaagaacc tctcttctt gggcacctt attttaaagg tagcagagcg	1020
[0159]	cctccagcta gtgacagtc accgaggcga gactgaggat ggtgccttca ctgaatgtaa	1080
[0160]	caacatgacc ctcatcaact tgctactcaa aatctttgg cccatacatg agagaaacct	1140
[0161]	caactgctc ctgctgctt tgcagatcct gagcagcgt gtcaccttct ccattcgata	1200
[0162]	ccagcttgtc cgactcttct atgatgtctg a	1231
[0163]	<210> 3	
[0164]	<211> 408	
[0165]	<212> PRT	
[0166]	<213> 中国仓鼠	
[0167]	<400> 3	

[0168]	Met Trp Ala Phe Pro Glu Leu Pro Leu Pro Leu Leu Val Asn Leu Phe
[0169]	1 5 10 15
[0170]	Gly Ser Leu Leu Gly Phe Val Ala Thr Val Thr Leu Ile Pro Ala Phe
[0171]	20 25 30
[0172]	Arg Ser His Phe Ile Ala Ala Arg Leu Cys Gly Gln Asp Leu Asn Lys
[0173]	35 40 45
[0174]	Leu Ser Arg Gln Gln Ile Pro Glu Ser Gln Gly Val Ile Cys Gly Ala
[0175]	50 55 60
[0176]	Val Phe Leu Ile Ile Leu Phe Cys Phe Ile Pro Phe Pro Phe Leu Asn
[0177]	65 70 75 80
[0178]	Cys Phe Val Glu Glu Gln Cys Lys Ala Phe Pro His His Glu Phe Val
[0179]	85 90 95
[0180]	Ala Leu Ile Gly Ala Leu Leu Ala Ile Cys Cys Met Ile Phe Leu Gly
[0181]	100 105 110
[0182]	Phe Ala Asp Asp Val Leu Asn Leu Arg Trp Arg His Lys Leu Leu Leu
[0183]	115 120 125
[0184]	Pro Thr Ala Ala Ser Leu Pro Leu Leu Met Val Tyr Phe Thr Asn Phe
[0185]	130 135 140
[0186]	Gly Asn Thr Thr Ile Val Val Pro Lys Pro Phe Arg Trp Ile Leu Gly
[0187]	145 150 155 160
[0188]	Leu His Leu Asp Leu Gly Ile Leu Tyr Tyr Val Tyr Met Gly Leu Leu
[0189]	165 170 175
[0190]	Ala Val Phe Cys Thr Asn Ala Ile Asn Ile Leu Ala Gly Ile Asn Gly
[0191]	180 185 190
[0192]	Leu Glu Ala Gly Gln Ser Leu Val Ile Ser Ala Ser Ile Ile Val Phe
[0193]	195 200 205
[0194]	Asn Leu Val Glu Leu Glu Gly Asp Tyr Arg Asp Asp His Val Phe Ser
[0195]	210 215 220
[0196]	Leu Tyr Phe Met Ile Pro Phe Phe Phe Thr Thr Leu Gly Leu Leu Tyr
[0197]	225 230 235 240
[0198]	His Asn Trp Tyr Pro Ser Gln Val Phe Val Gly Asp Thr Phe Cys Tyr
[0199]	245 250 255
[0200]	Phe Ala Gly Met Thr Phe Ala Val Val Gly Ile Leu Gly His Phe Ser
[0201]	260 265 270
[0202]	Lys Thr Met Leu Leu Phe Phe Ile Pro Gln Val Phe Asn Phe Leu Tyr
[0203]	275 280 285
[0204]	Ser Leu Pro Gln Leu Leu His Ala Ile Pro Cys Pro Arg His Arg Ile
[0205]	290 295 300
[0206]	Pro Arg Leu Asn Pro Lys Thr Gly Lys Leu Glu Met Ser Tyr Ser Lys
[0207]	305 310 315 320
[0208]	Phe Lys Thr Lys Asn Leu Ser Phe Leu Gly Thr Phe Ile Leu Lys Val
[0209]	325 330 335

[0210]	Ala Glu Arg Leu Gln Leu Val Thr Val His Arg Gly Glu Ser Glu Asp
[0211]	340 345 350
[0212]	Gly Ala Phe Thr Glu Cys Asn Asn Met Thr Leu Ile Asn Leu Leu Leu
[0213]	355 360 365
[0214]	Lys Ile Phe Gly Pro Ile His Glu Arg Asn Leu Thr Leu Leu Leu Leu
[0215]	370 375 380
[0216]	Leu Leu Gln Ile Leu Ser Ser Ala Val Thr Phe Ser Ile Arg Tyr Gln
[0217]	385 390 395 400
[0218]	Leu Val Arg Leu Phe Tyr Asp Val
[0219]	405
[0220]	<210> 4
[0221]	<211> 408
[0222]	<212> PRT
[0223]	<213> 智人
[0224]	<400> 4
[0225]	Met Trp Ala Phe Ser Glu Leu Pro Met Pro Leu Leu Ile Asn Leu Ile
[0226]	1 5 10 15
[0227]	Val Ser Leu Leu Gly Phe Val Ala Thr Val Thr Leu Ile Pro Ala Phe
[0228]	20 25 30
[0229]	Arg Gly His Phe Ile Ala Ala Arg Leu Cys Gly Gln Asp Leu Asn Lys
[0230]	35 40 45
[0231]	Thr Ser Arg Gln Gln Ile Pro Glu Ser Gln Gly Val Ile Ser Gly Ala
[0232]	50 55 60
[0233]	Val Phe Leu Ile Ile Leu Phe Cys Phe Ile Pro Phe Pro Phe Leu Asn
[0234]	65 70 75 80
[0235]	Cys Phe Val Lys Glu Gln Cys Lys Ala Phe Pro His His Glu Phe Val
[0236]	85 90 95
[0237]	Ala Leu Ile Gly Ala Leu Leu Ala Ile Cys Cys Met Ile Phe Leu Gly
[0238]	100 105 110
[0239]	Phe Ala Asp Asp Val Leu Asn Leu Arg Trp Arg His Lys Leu Leu Leu
[0240]	115 120 125
[0241]	Pro Thr Ala Ala Ser Leu Pro Leu Leu Met Val Tyr Phe Thr Asn Phe
[0242]	130 135 140
[0243]	Gly Asn Thr Thr Ile Val Val Pro Lys Pro Phe Arg Pro Ile Leu Gly
[0244]	145 150 155 160
[0245]	Leu His Leu Asp Leu Gly Ile Leu Tyr Tyr Val Tyr Met Gly Leu Leu
[0246]	165 170 175
[0247]	Ala Val Phe Cys Thr Asn Ala Ile Asn Ile Leu Ala Gly Ile Asn Gly
[0248]	180 185 190
[0249]	Leu Glu Ala Gly Gln Ser Leu Val Ile Ser Ala Ser Ile Ile Val Phe
[0250]	195 200 205
[0251]	Asn Leu Val Glu Leu Glu Gly Asp Cys Arg Asp Asp His Val Phe Ser

[0252]	210	215	220
[0253]	Leu Tyr Phe Met Ile Pro Phe Phe Phe Thr Thr Leu Gly Leu Leu Tyr		
[0254]	225	230	235 240
[0255]	His Asn Trp Tyr Pro Ser Arg Val Phe Val Gly Asp Thr Phe Cys Tyr		
[0256]		245	250 255
[0257]	Phe Ala Gly Met Thr Phe Ala Val Val Gly Ile Leu Gly His Phe Ser		
[0258]		260	265 270
[0259]	Lys Thr Met Leu Leu Phe Phe Met Pro Gln Val Phe Asn Phe Leu Tyr		
[0260]		275	280 285
[0261]	Ser Leu Pro Gln Leu Leu His Ile Ile Pro Cys Pro Arg His Arg Ile		
[0262]		290	295 300
[0263]	Pro Arg Leu Asn Ile Lys Thr Gly Lys Leu Glu Met Ser Tyr Ser Lys		
[0264]		305	310 315 320
[0265]	Phe Lys Thr Lys Ser Leu Ser Phe Leu Gly Thr Phe Ile Leu Lys Val		
[0266]		325	330 335
[0267]	Ala Glu Ser Leu Gln Leu Val Thr Val His Gln Ser Glu Thr Glu Asp		
[0268]		340	345 350
[0269]	Gly Glu Phe Thr Glu Cys Asn Asn Met Thr Leu Ile Asn Leu Leu Leu		
[0270]		355	360 365
[0271]	Lys Val Leu Gly Pro Ile His Glu Arg Asn Leu Thr Leu Leu Leu Leu		
[0272]		370	375 380
[0273]	Leu Leu Gln Ile Leu Gly Ser Ala Ile Thr Phe Ser Ile Arg Tyr Gln		
[0274]		385	390 395 400
[0275]	Leu Val Arg Leu Phe Tyr Asp Val		
[0276]		405	
[0277]	<210> 5		
[0278]	<211> 408		
[0279]	<212> PRT		
[0280]	<213> 猕猴		
[0281]	<400> 5		
[0282]	Met Trp Ala Phe Ser Glu Leu Pro Met Pro Leu Leu Val Asn Leu Ile		
[0283]	1	5	10 15
[0284]	Val Ser Leu Leu Gly Phe Val Ala Thr Val Thr Leu Ile Pro Ala Phe		
[0285]		20	25 30
[0286]	Arg Gly His Phe Ile Ala Ala Arg Leu Cys Gly Gln Asp Leu Asn Lys		
[0287]		35	40 45
[0288]	Thr Ser Arg Gln Gln Ile Pro Glu Ser Gln Gly Val Ile Ser Gly Ala		
[0289]		50	55 60
[0290]	Val Phe Leu Ile Ile Leu Phe Cys Phe Ile Pro Phe Pro Phe Leu Asn		
[0291]		65	70 75 80
[0292]	Cys Phe Val Lys Glu Gln Cys Lys Ala Phe Pro His His Glu Phe Val		
[0293]		85	90 95

[0294]	Ala Leu Ile Gly Ala Leu Leu Ala Ile Cys Cys Met Ile Phe Leu Gly
[0295]	100 105 110
[0296]	Phe Ala Asp Asp Val Leu Asn Leu Arg Trp Arg His Lys Leu Leu Leu
[0297]	115 120 125
[0298]	Pro Thr Ala Ala Ser Leu Pro Leu Leu Met Val Tyr Phe Thr Asn Phe
[0299]	130 135 140
[0300]	Gly Asn Thr Thr Ile Val Val Pro Lys Pro Phe Arg Pro Ile Leu Gly
[0301]	145 150 155 160
[0302]	Leu His Leu Asp Leu Gly Ile Leu Tyr Tyr Val Tyr Met Gly Leu Leu
[0303]	165 170 175
[0304]	Ala Val Phe Cys Thr Asn Ala Ile Asn Ile Leu Ala Gly Ile Asn Gly
[0305]	180 185 190
[0306]	Leu Glu Ala Gly Gln Ser Leu Val Ile Ser Ala Ser Ile Ile Val Phe
[0307]	195 200 205
[0308]	Asn Leu Val Glu Leu Glu Gly Asp Cys Arg Asp Asp His Val Phe Ser
[0309]	210 215 220
[0310]	Leu Tyr Phe Met Ile Pro Phe Phe Phe Thr Thr Leu Gly Leu Leu Tyr
[0311]	225 230 235 240
[0312]	His Asn Trp Tyr Pro Ser Arg Val Phe Val Gly Asp Thr Phe Cys Tyr
[0313]	245 250 255
[0314]	Phe Ala Gly Met Thr Phe Ala Val Val Gly Ile Leu Gly His Phe Ser
[0315]	260 265 270
[0316]	Lys Thr Met Leu Leu Phe Phe Met Pro Gln Val Phe Asn Phe Leu Tyr
[0317]	275 280 285
[0318]	Ser Leu Pro Gln Leu Leu His Ile Ile Pro Cys Pro Arg His Arg Ile
[0319]	290 295 300
[0320]	Pro Arg Leu Asn Ile Lys Thr Gly Lys Leu Glu Met Ser Tyr Ser Lys
[0321]	305 310 315 320
[0322]	Phe Lys Thr Lys Ser Leu Ser Phe Leu Gly Thr Phe Ile Leu Lys Val
[0323]	325 330 335
[0324]	Ala Glu Ser Leu Arg Leu Val Thr Ile His Gln Ser Asp Thr Glu Asp
[0325]	340 345 350
[0326]	Gly Glu Phe Thr Glu Cys Asn Asn Met Thr Leu Ile Asn Leu Leu Leu
[0327]	355 360 365
[0328]	Lys Ile Phe Gly Pro Ile His Glu Arg Asn Leu Thr Leu Leu Leu Leu
[0329]	370 375 380
[0330]	Leu Leu Gln Ile Leu Gly Ser Ala Phe Thr Phe Ser Ile Arg Tyr Gln
[0331]	385 390 395 400
[0332]	Leu Val Arg Leu Phe Tyr Asp Val
[0333]	405
[0334]	<210> 6
[0335]	<211> 408

[0336] <212> PRT
 [0337] <213> 黑猩猩
 [0338] <400> 6
 [0339] Met Trp Ala Phe Ser Glu Leu Pro Met Pro Leu Leu Ile Asn Leu Ile
 [0340] 1 5 10 15
 [0341] Val Ser Leu Leu Gly Phe Val Ala Thr Val Thr Leu Ile Pro Ala Phe
 [0342] 20 25 30
 [0343] Arg Gly His Phe Ile Ala Ala Arg Leu Cys Gly Gln Asp Leu Asn Lys
 [0344] 35 40 45
 [0345] Thr Ser Arg Gln Gln Ile Pro Glu Ser Gln Gly Val Ile Ser Gly Ala
 [0346] 50 55 60
 [0347] Val Phe Leu Ile Ile Leu Phe Cys Phe Ile Pro Phe Pro Phe Leu Asn
 [0348] 65 70 75 80
 [0349] Cys Phe Val Lys Glu Gln Cys Lys Ala Phe Pro His His Glu Phe Val
 [0350] 85 90 95
 [0351] Ala Leu Ile Gly Ala Leu Leu Ala Ile Cys Cys Met Ile Phe Leu Gly
 [0352] 100 105 110
 [0353] Phe Ala Asp Asp Val Leu Asn Leu Arg Trp Arg His Lys Leu Leu Leu
 [0354] 115 120 125
 [0355] Pro Thr Ala Ala Ser Leu Pro Leu Leu Met Val Tyr Phe Thr Asn Phe
 [0356] 130 135 140
 [0357] Gly Asn Thr Thr Ile Val Val Pro Lys Pro Phe Arg Pro Ile Leu Gly
 [0358] 145 150 155 160
 [0359] Leu His Leu Asp Leu Gly Ile Leu Tyr Tyr Val Tyr Met Gly Leu Leu
 [0360] 165 170 175
 [0361] Ala Val Phe Cys Thr Asn Ala Ile Asn Ile Leu Ala Gly Ile Asn Gly
 [0362] 180 185 190
 [0363] Leu Glu Ala Gly Gln Ser Leu Val Ile Ser Ala Ser Ile Ile Val Phe
 [0364] 195 200 205
 [0365] Asn Leu Val Glu Leu Glu Gly Asp Cys Arg Asp Asp His Val Phe Ser
 [0366] 210 215 220
 [0367] Leu Tyr Phe Met Ile Pro Phe Phe Phe Thr Thr Leu Gly Leu Leu Tyr
 [0368] 225 230 235 240
 [0369] His Asn Trp Tyr Pro Ser Arg Val Phe Val Gly Asp Thr Phe Cys Tyr
 [0370] 245 250 255
 [0371] Phe Ala Gly Met Thr Phe Ala Val Val Gly Ile Leu Gly His Phe Ser
 [0372] 260 265 270
 [0373] Lys Thr Met Leu Leu Phe Phe Met Pro Gln Val Phe Asn Phe Leu Tyr
 [0374] 275 280 285
 [0375] Ser Leu Pro Gln Leu Leu His Ile Ile Pro Cys Pro Arg His Arg Ile
 [0376] 290 295 300
 [0377] Pro Arg Leu Asn Ile Lys Thr Gly Lys Leu Glu Met Ser Tyr Ser Lys

[0378]	305	310	315	320
[0379]	Phe Lys Thr Lys Ser Leu Ser Phe Leu Gly Thr Phe Ile Leu Lys Val			
[0380]		325	330	335
[0381]	Ala Glu Ser Leu Gln Leu Val Thr Val His Gln Ser Glu Thr Glu Asp			
[0382]		340	345	350
[0383]	Gly Glu Phe Thr Glu Cys Asn Asn Met Thr Leu Ile Asn Leu Leu Leu			
[0384]		355	360	365
[0385]	Lys Ile Leu Gly Pro Ile His Glu Arg Asn Leu Thr Leu Leu Leu Leu			
[0386]		370	375	380
[0387]	Leu Leu Gln Ile Leu Gly Ser Ala Ile Thr Phe Ser Ile Arg Tyr Gln			
[0388]	385	390	395	400
[0389]	Leu Val Arg Leu Phe Tyr Asp Val			
[0390]		405		
[0391]	<210> 7			
[0392]	<211> 408			
[0393]	<212> PRT			
[0394]	<213> 家犬			
[0395]	<400> 7			
[0396]	Met Trp Ala Phe Pro Glu Leu Pro Met Pro Leu Leu Val Asn Leu Val			
[0397]	1	5	10	15
[0398]	Gly Ser Leu Leu Gly Phe Val Ala Thr Val Thr Leu Ile Pro Ala Phe			
[0399]		20	25	30
[0400]	Arg Gly His Phe Ile Ala Ala His Leu Cys Gly Gln Asp Leu Asn Lys			
[0401]		35	40	45
[0402]	Thr Gly Arg Gln Gln Ile Pro Glu Ser Gln Gly Val Ile Ser Gly Ala			
[0403]		50	55	60
[0404]	Val Phe Leu Ile Ile Leu Phe Cys Phe Ile Pro Phe Pro Phe Leu Asn			
[0405]	65	70	75	80
[0406]	Cys Phe Met Glu Glu Gln Cys Lys Ala Phe Pro His His Glu Phe Val			
[0407]		85	90	95
[0408]	Ala Leu Ile Gly Ala Leu Leu Ala Ile Cys Cys Met Ile Phe Leu Gly			
[0409]		100	105	110
[0410]	Phe Ala Asp Asp Val Leu Asn Leu Arg Trp Arg His Lys Leu Leu Leu			
[0411]		115	120	125
[0412]	Pro Thr Ala Ala Ser Leu Pro Leu Leu Met Val Tyr Phe Thr Asn Phe			
[0413]		130	135	140
[0414]	Gly Asn Thr Thr Ile Val Val Pro Lys Pro Phe Arg Pro Ile Leu Gly			
[0415]	145	150	155	160
[0416]	Leu His Leu Asp Leu Gly Ile Leu Tyr Tyr Val Tyr Met Gly Leu Leu			
[0417]		165	170	175
[0418]	Ala Val Phe Cys Thr Asn Ala Ile Asn Ile Leu Ala Gly Ile Asn Gly			
[0419]		180	185	190

[0420]	Leu Glu Ala Gly Gln Ser Leu Val Ile Ser Ala Ser Ile Ile Val Phe
[0421]	195 200 205
[0422]	Asn Leu Val Glu Leu Glu Gly Asp Tyr Arg Asp Asp His Val Phe Ser
[0423]	210 215 220
[0424]	Leu Tyr Phe Met Ile Pro Phe Phe Phe Thr Thr Leu Gly Leu Leu Tyr
[0425]	225 230 235 240
[0426]	His Asn Trp Tyr Pro Ser Gln Val Phe Val Gly Asp Thr Phe Cys Tyr
[0427]	245 250 255
[0428]	Phe Ala Gly Met Thr Phe Ala Val Val Gly Ile Leu Gly His Phe Ser
[0429]	260 265 270
[0430]	Lys Thr Met Leu Leu Phe Phe Met Pro Gln Val Phe Asn Phe Leu Tyr
[0431]	275 280 285
[0432]	Ser Leu Pro Gln Leu Leu His Ile Ile Pro Cys Pro Arg His Arg Ile
[0433]	290 295 300
[0434]	Pro Arg Leu Asn Thr Lys Thr Gly Lys Leu Glu Met Ser Tyr Ser Lys
[0435]	305 310 315 320
[0436]	Phe Lys Thr Lys Ser Leu Ser Phe Leu Gly Asn Phe Ile Leu Lys Val
[0437]	325 330 335
[0438]	Ala Ala Ser Leu Gln Leu Val Thr Val His Gln Ser Glu Asn Glu Asp
[0439]	340 345 350
[0440]	Gly Ala Phe Thr Glu Cys Asn Asn Met Thr Leu Leu Asn Leu Leu Leu
[0441]	355 360 365
[0442]	Lys Val Leu Gly Pro Met His Glu Arg Asn Leu Thr Leu Leu Leu Leu
[0443]	370 375 380
[0444]	Leu Leu Gln Ile Leu Gly Ser Ala Val Thr Phe Ser Ile Arg Tyr Gln
[0445]	385 390 395 400
[0446]	Leu Val Arg Leu Phe Tyr Asp Val
[0447]	405
[0448]	<210> 8
[0449]	<211> 408
[0450]	<212> PRT
[0451]	<213> 豚鼠
[0452]	<400> 8
[0453]	Met Trp Ala Phe Ser Glu Val Pro Ile Pro Leu Leu Val Asn Leu Ile
[0454]	1 5 10 15
[0455]	Gly Ser Leu Leu Gly Phe Val Ala Thr Leu Thr Leu Ile Pro Ala Phe
[0456]	20 25 30
[0457]	Arg Gly His Phe Ile Ala Ala Arg Leu Cys Gly Gln Asp Leu Asn Lys
[0458]	35 40 45
[0459]	Thr Asn Arg Gln Gln Ile Pro Glu Ser Gln Gly Val Ile Ser Gly Ala
[0460]	50 55 60
[0461]	Val Phe Leu Ile Ile Leu Phe Cys Phe Ile Pro Phe Pro Phe Leu Asn

[0462]	65	70	75	80
[0463]	Cys Phe Val Lys Glu Gln Cys Lys Ala Phe Pro His His Glu Phe Val			
[0464]		85	90	95
[0465]	Ala Leu Ile Gly Ala Leu Leu Ala Ile Cys Cys Met Ile Phe Leu Gly			
[0466]		100	105	110
[0467]	Phe Ala Asp Asp Val Leu Asn Leu Arg Trp Arg His Lys Leu Leu Leu			
[0468]		115	120	125
[0469]	Pro Thr Ala Ala Ser Leu Pro Leu Leu Met Val Tyr Phe Thr Asn Phe			
[0470]		130	135	140
[0471]	Gly Asn Thr Thr Ile Val Val Pro Lys Pro Phe Arg Pro Val Leu Gly			
[0472]		145	150	155
[0473]	Leu His Leu Asp Leu Gly Ile Leu Tyr Tyr Val Tyr Met Gly Leu Leu			
[0474]		165	170	175
[0475]	Ala Val Phe Cys Thr Asn Ala Ile Asn Ile Leu Ala Gly Ile Asn Gly			
[0476]		180	185	190
[0477]	Leu Glu Ala Gly Gln Ser Leu Val Ile Ser Ala Ser Ile Ile Val Phe			
[0478]		195	200	205
[0479]	Asn Leu Val Glu Leu Gln Gly Asp Tyr Arg Asp Asp His Val Phe Ser			
[0480]		210	215	220
[0481]	Leu Tyr Phe Met Ile Pro Phe Phe Phe Thr Thr Leu Gly Leu Leu Tyr			
[0482]		225	230	235
[0483]	His Asn Trp Tyr Pro Ser Gln Val Phe Val Gly Asp Thr Phe Cys Tyr			
[0484]		245	250	255
[0485]	Phe Ala Gly Met Thr Phe Ala Val Val Gly Ile Leu Gly His Phe Ser			
[0486]		260	265	270
[0487]	Lys Thr Met Leu Leu Phe Phe Met Pro Gln Val Phe Asn Phe Leu Tyr			
[0488]		275	280	285
[0489]	Ser Leu Pro Gln Leu Leu His Ile Ile Pro Cys Pro Arg His Arg Ile			
[0490]		290	295	300
[0491]	Pro Arg Leu Asn Thr Lys Thr Gly Lys Leu Glu Met Ser Tyr Ser Lys			
[0492]		305	310	315
[0493]	Phe Lys Thr Asn Ser Leu Ser Phe Leu Gly Thr Phe Ile Leu Lys Val			
[0494]		325	330	335
[0495]	Ala Glu Arg Leu Gln Leu Val Thr Val His Arg Ser Glu Gly Glu Asp			
[0496]		340	345	350
[0497]	Gly Ala Phe Thr Glu Cys Asn Asn Met Thr Leu Ile Asn Leu Leu Leu			
[0498]		355	360	365
[0499]	Lys Ile Phe Gly Pro Ile His Glu Arg Asn Leu Thr Leu Leu Leu Leu			
[0500]		370	375	380
[0501]	Leu Leu Gln Ile Val Gly Ser Ala Val Thr Phe Ser Ile Arg Tyr Gln			
[0502]		385	390	395
[0503]	Leu Val Arg Leu Phe Tyr Asp Val			400

[0504]		405	
[0505]	<210>	9	
[0506]	<211>	410	
[0507]	<212>	PRT	
[0508]	<213>	褐家鼠	
[0509]	<400>	9	
[0510]	Met Trp Ala Phe Pro Glu Leu Pro Leu Pro Leu Pro Leu Leu Val Asn		
[0511]	1	5	10 15
[0512]	Leu Ile Gly Ser Leu Leu Gly Phe Val Ala Thr Val Thr Leu Ile Pro		
[0513]		20	25 30
[0514]	Ala Phe Arg Ser His Phe Ile Ala Ala Arg Leu Cys Gly Gln Asp Leu		
[0515]		35	40 45
[0516]	Asn Lys Leu Ser Arg Gln Gln Ile Pro Glu Ser Gln Gly Val Ile Ser		
[0517]		50	55 60
[0518]	Gly Ala Val Phe Leu Ile Ile Leu Phe Cys Phe Ile Pro Phe Pro Phe		
[0519]		65	70 75 80
[0520]	Leu Asn Cys Phe Val Glu Glu Gln Cys Lys Ala Phe Pro His His Glu		
[0521]		85	90 95
[0522]	Phe Val Ala Leu Ile Gly Ala Leu Leu Ala Ile Cys Cys Met Ile Phe		
[0523]		100	105 110
[0524]	Leu Gly Phe Ala Asp Asp Val Leu Asn Leu Arg Trp Arg His Lys Leu		
[0525]		115	120 125
[0526]	Leu Leu Pro Thr Ala Ala Ser Leu Pro Leu Leu Met Val Tyr Phe Thr		
[0527]		130	135 140
[0528]	Asn Phe Gly Asn Thr Thr Ile Val Val Pro Lys Pro Phe Arg Trp Ile		
[0529]		145	150 155 160
[0530]	Leu Gly Leu His Leu Asp Leu Gly Ile Leu Tyr Tyr Val Tyr Met Gly		
[0531]		165	170 175
[0532]	Leu Leu Ala Val Phe Cys Thr Asn Ala Ile Asn Ile Leu Ala Gly Ile		
[0533]		180	185 190
[0534]	Asn Gly Leu Glu Ala Gly Gln Ser Leu Val Ile Ser Ala Ser Ile Ile		
[0535]		195	200 205
[0536]	Val Phe Asn Leu Val Glu Leu Glu Gly Asp Tyr Arg Asp Asp His Val		
[0537]		210	215 220
[0538]	Phe Ser Leu Tyr Phe Met Met Pro Phe Phe Phe Thr Thr Leu Gly Leu		
[0539]		225	230 235 240
[0540]	Leu Tyr His Asn Trp Tyr Pro Ser Gln Val Phe Val Gly Asp Thr Phe		
[0541]		245	250 255
[0542]	Cys Tyr Phe Ala Gly Met Thr Phe Ala Val Val Gly Ile Leu Gly His		
[0543]		260	265 270
[0544]	Phe Ser Lys Thr Met Leu Leu Phe Phe Met Pro Gln Val Phe Asn Phe		
[0545]		275	280 285

[0546]	Leu Tyr Ser Leu Pro Gln Leu Phe Gln Ile Ile Pro Cys Pro Arg His
[0547]	290 295 300
[0548]	Arg Met Pro Arg Leu Asn Thr Lys Thr Gly Lys Leu Glu Met Ser Tyr
[0549]	305 310 315 320
[0550]	Ser Lys Phe Lys Thr Lys Ser Leu Ser Phe Leu Gly Thr Phe Ile Leu
[0551]	325 330 335
[0552]	Lys Val Ala Glu Ser Leu Arg Leu Val Thr Val His Arg Gly Glu Ser
[0553]	340 345 350
[0554]	Glu Asp Gly Ala Phe Thr Glu Cys Asn Asn Met Thr Leu Ile Asn Leu
[0555]	355 360 365
[0556]	Leu Leu Lys Val Phe Gly Pro Thr His Glu Arg Asn Leu Thr Leu Phe
[0557]	370 375 380
[0558]	Leu Leu Leu Leu Gln Val Leu Ser Ser Ala Val Thr Phe Ser Ile Arg
[0559]	385 390 395 400
[0560]	Tyr Gln Leu Val Arg Leu Phe Tyr Asp Val
[0561]	405 410
[0562]	<210> 10
[0563]	<211> 410
[0564]	<212> PRT
[0565]	<213> 小家鼠
[0566]	<400> 10
[0567]	Met Trp Ala Phe Pro Glu Leu Pro Leu Pro Leu Pro Leu Leu Val Asn
[0568]	1 5 10 15
[0569]	Leu Ile Gly Ser Leu Leu Gly Phe Val Ala Thr Val Thr Leu Ile Pro
[0570]	20 25 30
[0571]	Ala Phe Arg Ser His Phe Ile Ala Ala Arg Leu Cys Gly Gln Asp Leu
[0572]	35 40 45
[0573]	Asn Lys Leu Ser Gln Gln Gln Ile Pro Glu Ser Gln Gly Val Ile Ser
[0574]	50 55 60
[0575]	Gly Ala Val Phe Leu Ile Ile Leu Phe Cys Phe Ile Pro Phe Pro Phe
[0576]	65 70 75 80
[0577]	Leu Asn Cys Phe Val Glu Glu Gln Cys Lys Ala Phe Pro His His Glu
[0578]	85 90 95
[0579]	Phe Val Ala Leu Ile Gly Ala Leu Leu Ala Ile Cys Cys Met Ile Phe
[0580]	100 105 110
[0581]	Leu Gly Phe Ala Asp Asp Val Leu Asn Leu Arg Trp Arg His Lys Leu
[0582]	115 120 125
[0583]	Leu Leu Pro Thr Ala Ala Ser Leu Pro Leu Leu Met Val Tyr Phe Thr
[0584]	130 135 140
[0585]	Asn Phe Gly Asn Thr Thr Ile Val Val Pro Lys Pro Phe Arg Trp Ile
[0586]	145 150 155 160
[0587]	Leu Gly Leu His Leu Asp Leu Gly Ile Leu Tyr Tyr Val Tyr Met Gly

[0588]	165	170	175
[0589]	Leu Leu Ala Val Phe Cys Thr Asn Ala Ile Asn Ile Leu Ala Gly Ile		
[0590]	180	185	190
[0591]	Asn Gly Leu Glu Ala Gly Gln Ser Leu Val Ile Ser Ala Ser Ile Ile		
[0592]	195	200	205
[0593]	Val Phe Asn Leu Val Glu Leu Glu Gly Asp Tyr Arg Asp Asp His Ile		
[0594]	210	215	220
[0595]	Phe Ser Leu Tyr Phe Met Ile Pro Phe Phe Phe Thr Thr Leu Gly Leu		
[0596]	225	230	235
[0597]	Leu Tyr His Asn Trp Tyr Pro Ser Arg Val Phe Val Gly Asp Thr Phe		
[0598]	245	250	255
[0599]	Cys Tyr Phe Ala Gly Met Thr Phe Ala Val Val Gly Ile Leu Gly His		
[0600]	260	265	270
[0601]	Phe Ser Lys Thr Met Leu Leu Phe Phe Met Pro Gln Val Phe Asn Phe		
[0602]	275	280	285
[0603]	Leu Tyr Ser Leu Pro Gln Leu Phe His Ile Ile Pro Cys Pro Arg His		
[0604]	290	295	300
[0605]	Arg Met Pro Arg Leu Asn Ala Lys Thr Gly Lys Leu Glu Met Ser Tyr		
[0606]	305	310	315
[0607]	Ser Lys Phe Lys Thr Lys Asn Leu Ser Phe Leu Gly Thr Phe Ile Leu		
[0608]	325	330	335
[0609]	Lys Val Ala Glu Asn Leu Arg Leu Val Thr Val His Gln Gly Glu Ser		
[0610]	340	345	350
[0611]	Glu Asp Gly Ala Phe Thr Glu Cys Asn Asn Met Thr Leu Ile Asn Leu		
[0612]	355	360	365
[0613]	Leu Leu Lys Val Phe Gly Pro Ile His Glu Arg Asn Leu Thr Leu Leu		
[0614]	370	375	380
[0615]	Leu Leu Leu Leu Gln Val Leu Ser Ser Ala Ala Thr Phe Ser Ile Arg		
[0616]	385	390	395
[0617]	Tyr Gln Leu Val Arg Leu Phe Tyr Asp Val		
[0618]	405	410	
[0619]	<210> 11		
[0620]	<211> 1920		
[0621]	<212> DNA		
[0622]	<213> 小家鼠		
[0623]	<400> 11		
[0624]	gttgcctcct aagagcttct tgctggcag gagggagggt caggtcctag cgtcctagct 60		
[0625]	gggttttggt cccgctggcg ccggaatcct ctgcgggttg ggagccgcac tgccggctgc 120		
[0626]	cgagccacg ggattgttc tggcttacca gttagctgag taggcggcgg ggcggcggcc 180		
[0627]	accggagggt caccatgtgg gccttcccgg agttgccct gccgctgcc ctgctggtga 240		
[0628]	attgatcgg ctgcgtgttg ggattcgttg ctacagtcac cctcatcct gccttccgta 300		
[0629]	gccactttat cgccgcgcgc ctctgtggcc aggacctcaa caagctcagc cagcagcaga 360		

[0630]	tcccagagtc ccagggagtg atcagcggtg ctgttttctt tatcatcctc ttctgcttca	420
[0631]	tccctttccc ctctctgaac tgcttcgtgg aggagcagtg taaggcattc ccccaccatg	480
[0632]	aatttgtggc cctaataagg gcctctcttg ccatctgctg catgatcttc ctggggtttg	540
[0633]	ctgatgatgt cctcaatctc cgctggcgcc acaagctgct gctgcccaca gctgcctcac	600
[0634]	tacctctct catggtctac ttcacaaact ttggcaatac aaccatcgtg gtgcccgaagc	660
[0635]	ccttccgctg gattctgggc ctgcatttgg acttggggat cctgtactac gtctacatgg	720
[0636]	ggctgcttgc agtgttctgt accaatgcca tcaacatcct ggcgggcatt aatggcctag	780
[0637]	aggccggtca gtcactagtc atctctgctt ctatcattgt cttcaacctg gtggaactgg	840
[0638]	aagtgatta tcgagatgat catatctttt ccctttactt catgatacca tttttcttta	900
[0639]	ccacctggg actgctttac cacaactggg acccgctccg cgtgtttgtg ggagacacct	960
[0640]	tctgttactt tgcgggcatg acttttgcg tgggtgggat cttgggacac ttcagcaaga	1020
[0641]	ccatgctgct cttctttatg ccacaagtat tcaatttctt ctactcactg cctcagctct	1080
[0642]	tccatatcat ccctgcct cgacaccgga tgcccagact caacgcaaag acaggcaaac	1140
[0643]	tggaaatgag ctattccaag ttcaagacca agaacctctc tttctgggc acctttattt	1200
[0644]	taaagtagc agagaacctc cggttagtag cagttcacca aggtgagagt gaggacgggtg	1260
[0645]	ccttactga gtgtaacaac atgacctca tcaacttget acttaaagtc tttgggccta	1320
[0646]	tacatgagag aaacctcacc ctgctcctgc tgctcctgca ggtcctaagc agcgcgcca	1380
[0647]	ccttctccat tcgttaccac ctgctccgac tcttctatga tgtctgagct ccctgacagc	1440
[0648]	tgccctttac ctacagctc ccattggacc tcagccagga ccagcctctg tctggtccga	1500
[0649]	gatgacctc tggccagge ctgctgaca cttttgttct cagcttctgc catctgtgac	1560
[0650]	tactgatata ctggatggac accttctggt acttgaagtc cgctagtgtg accttgccca	1620
[0651]	ggcctttcat cttgccttgc cctcccttc tgtcccatct gcagcctcac caggtgggct	1680
[0652]	tgtagcctct attatgcaaa tattcgtagc tcagcttca gagcgctaac tctaaaggaa	1740
[0653]	ttacctgag ccttgagaga gaacctgggc tagggctaga gttagggcta catactccaa	1800
[0654]	ggtgacctca catttgacta tcaaatgaag tgttgtgatt gggaagcgtg gaggcagggc	1860
[0655]	catgtgctca gaacggtgac aataaaggac tgctttttac ttgttaaaaa aaaaaaaaaa	1920
[0656]	<210> 12	
[0657]	<211> 2150	
[0658]	<212> DNA	
[0659]	<213> 智人	
[0660]	<400> 12	
[0661]	aagtatccgt tcttgctgct ctttctttaa ttgcgtttcc agtactctct cggtgattct	60
[0662]	actcttgaac atagatgaa atttggaatc acacttctct tccacttcca tccccacct	120
[0663]	ctaataccca tattaataat ggcggcgccc gccttccgca gtaatggtt ttcagcgaac	180
[0664]	aagatccggg cggaacagt agataggcgg gtgcagcggg gcagaacata ggttgcctta	240
[0665]	gagaggttcc ccggtgtccc gacggcgct caagtcagag ttgctgggtt ttgctcagat	300
[0666]	tgggtgaggga agagcctgcc tgtggggagc ggccactcca tactgctgag gcctcaggac	360
[0667]	tgetgetcag cttgcccgtt acctgaagag gcggcggagc cgggccccctg accggtcacc	420
[0668]	atgtgggect tctcggaatt gcccatgcc ctgctgatca atttgatctg ctgctgctg	480
[0669]	ggatttgtgg ccacagtcac cctcatccc gccttccggg gccacttcat tgctgcgcgc	540
[0670]	ctctgtggtc aggacctcaa caaaaccagc cgacagcaga tcccagaatc ccaggagtg	600
[0671]	atcagcggtg ctgttttctt tatcatcctc ttctgcttca tccctttccc ctctctgaac	660

[0672]	tgctttgtga aggagcagtg taaggcattc cccaccatg aatttgtggc cctgataggt	720
[0673]	gccttccttg ccatctgctg catgatcttc ctgggctttg cggatgatgt actgaatctg	780
[0674]	cgctggcgcc ataagetgct gctacctaca gctgcctcac tacctctcct catggctctat	840
[0675]	ttaccaact ttggcaacac gaccattgtg gtgccaagc ctttccgcc gatacttggc	900
[0676]	ctgcatctgg acttgggaat cctgtactat gtctacatgg ggctgctggc agtgttctgt	960
[0677]	accaatgcca tcaatatcct agcaggaatt aacggcctag aggctggcca gtcactagtc	1020
[0678]	atctctgctt ccatcattgt cttcaacctg gtagagttgg aagggtattg tcgggatgat	1080
[0679]	catgtctttt cctctactt catgataccc tttttttca ccactttggg attgctctac	1140
[0680]	cacaactggt acccatcacg ggtgtttgtg ggagatacct tctgttactt tgctggcatg	1200
[0681]	acctttgccg tgggtggcat cttgggacac ttcagcaaga ccatgctact attcttcatg	1260
[0682]	ccccaggtgt tcaacttctt ctactcactg cctcagctcc tgcatatcat cccctgcctt	1320
[0683]	cgccaccgca taccagact caatatcaag acaggcaaac tggagatgag ctattccaag	1380
[0684]	ttcaagacca agagcctctc tttcttgggc acctttattt taaaggtggc agagagcctc	1440
[0685]	cagctggtga cagtacacca gagtgagact gaagatggtg aattcactga atgtaacaac	1500
[0686]	atgacctca tcaacttgct acttaaagtc cttgggcca tacatgagag aaacctcaca	1560
[0687]	ttgctcctgc tgctgctgca gatcctgggc agtgccatca cttcttccat tcgatatcag	1620
[0688]	ctcgttcgac tcttctatga tgtctgagtc cttgatcat tctcctttac ctcacagtct	1680
[0689]	ctaggattcc tgactcaggc tgacctctt ctctggctcc agactgcctc cttgccagg	1740
[0690]	cctctctcac tcttatact cctccagatt ttgttctcag cattttcctt tctctgtgat	1800
[0691]	cattggcatc ctgggcgtt cttgccctct gctgactact gattggattt tacctatggc	1860
[0692]	ttctgcaac ttgctactct ctccctctcc atccatctt tgcagcctca taggggtggga	1920
[0693]	tacagcagct tttttgcag ttatccacac tcacattca gagtctgac tctcaaggaa	1980
[0694]	ccactggttt ttgggataga acttgggcca gggctaggaa cacaggctcc acgggtgacat	2040
[0695]	gtcatttgat tgtaaattaa gtgttctgat tagtaagaac taagcagggg gccacatgct	2100
[0696]	ctcaatggag acaataaagt gttgtctttt tcttattgtt taaaaaaaa	2150
[0697]	<210> 13	
[0698]	<211> 1840	
[0699]	<212> DNA	
[0700]	<213> 褐家鼠	
[0701]	<400> 13	
[0702]	ggggctggc gccggaatcc tctgagtga gggagctgca ctgctggctg ccgaggcctc	60
[0703]	tggtttgttc ctggcttacc aagttagctg agtaggcggc ggagcggcgg cccccggagg	120
[0704]	gtcactatgt ggccctccc ggagttacct ctgccgctgc cgctgttggg gaatttgatc	180
[0705]	ggatcgctgt tgggatttgt ggctaccgtc accctcatcc ctgccttccg tagccacttt	240
[0706]	atgcccgcc gtctctgtgg ccaggacctc aacaagctca gccggcagca gatcccagag	300
[0707]	tcccaggag tgatcagcgg tgctgtttc cttatcatcc tcttctgctt catcccttc	360
[0708]	ccttctctga actgctttgt ggaggagcag tgtaaggcat tccccacca tgaatttgtg	420
[0709]	gcctgatag gtgcctcct tgctatctgc tgcatgatct tctgggttt tgctgatgac	480
[0710]	gtactcaatc tgcgatggcg tcataagctg ctgctccca cagctgcctc actacctctc	540
[0711]	ctcatggtct acttactaa ctttggcaat acaaccattg tgggtccgaa gcccttccgc	600
[0712]	tggattcttg gcctgcatth ggatttgggc atcctgtact atgtctacat gggactgctt	660
[0713]	gcagtgcttct gtaccaatgc cataacatc ctagegggaa ttaatggcct agaggctggt	720

[0714]	caatcactag	tcattctctgc	ttctattatt	gtcttcaacc	tggtggagct	ggaaggtgat	780
[0715]	tatcgggacg	atcatgtctt	ttccctctac	ttcatgatgc	catttttttt	taccaccttg	840
[0716]	ggattgctgt	accataactg	gtaccegtct	caggtgtttg	tgggagacac	cttctgttat	900
[0717]	tttgctggca	tgaccttgc	cgtgggtgga	atcttgggac	acttcagcaa	gaccatgctg	960
[0718]	ctcttcttta	tgccacaagt	attcaatttc	ctctactcac	tgccctcagct	ctttcagatc	1020
[0719]	atccctgcc	ctcgacaccg	tatgcccaga	ctcaatacga	agacaggcaa	actggagatg	1080
[0720]	agctattcca	agttcaagac	caagagcctc	tctttcttgg	gcacgtttat	tttaaaggta	1140
[0721]	gcagagagcc	tccggctggt	gacagttcac	cgaggggaga	gtgaggatgg	tgcccttact	1200
[0722]	gagtgaaca	acatgaccct	catcaacttg	ctacttaaag	tctttgggcc	tacacatgag	1260
[0723]	agaaacctca	cactgttctt	gctgctcctg	caggttctga	gcagcgtctg	caccttctcc	1320
[0724]	attcgttacc	agctcgtccg	actcttctat	gatgtctgag	ctccctgacg	actgcccttt	1380
[0725]	accacacagt	ctccattgga	cctcagccag	gaccacctc	tgtccgctcc	gaccgccttc	1440
[0726]	tggtccaggc	tcagcttctg	ccgtcatctg	tgactactga	catcctggat	ggactcctta	1500
[0727]	gtggacttga	cgccactag	ttggacttgc	ctatgcttcc	ttgagtttgc	tactccctcc	1560
[0728]	ctttctgcag	cctcaccagg	tgggctgta	gcatctttta	tgcaaatatt	catggctcaa	1620
[0729]	ctttcagaac	cctaactcta	aaggaatccc	ctgggcttgc	agagagaacc	tgggctaggg	1680
[0730]	ctagagttag	ggcaacatac	tccaaggtaa	cctcacatct	gactatcaaa	ttaagtgttc	1740
[0731]	tgattaggaa	gagcagaggc	agggccatgt	gctcagaatg	gtgacaataa	aggattgcct	1800
[0732]	tttacttgcc	aaaaaaaaa	aaaaaaaaa	aaaaaaaaa			1840
[0733]	<210>	14					
[0734]	<211>	5424					
[0735]	<212>	DNA					
[0736]	<213>	猕猴					
[0737]	<400>	14					
[0738]	tattaaaaat	ggcggctgcc	gccctccgca	gtaatagttg	ttcagcgaat	aaaatccggg	60
[0739]	cggaaacagt	aggtaggctg	gttcagtggtg	ggcagaacct	aggttgcctt	agagaggttc	120
[0740]	ttcaggtcc	cgagggcggc	tcaagtcaga	gttggttgat	tctgctcaga	ttggtgtggg	180
[0741]	aagagcctgc	ctgtggggag	cgccactcc	atactgctga	ggcctcagga	ctgctgctca	240
[0742]	gcttgccagt	tacctgaaga	ggcggcggag	ccgggccctt	gaccggtcac	catgtgggcc	300
[0743]	ttctcggaat	tgccatgcc	gctgctggtc	aatttgatcg	tctcgtctgt	gggatttgtg	360
[0744]	gccacagtca	ccctcatccc	agccttccgg	ggccacttca	ttgctgcgcg	cctctgtggt	420
[0745]	caggacctca	acaaaaccag	ccgacaacag	atgtgagcag	cggcacacgg	ttccgggcag	480
[0746]	ggggcaaggg	ctaaggaagg	agtggtctagg	gcagggggcg	ggaccggggt	gtttgaccac	540
[0747]	acgtgaaaac	tcagaactaa	cccaggcagc	ctggaactcg	gagaggtgat	gagcagaact	600
[0748]	tattcgcatt	gggaaagga	tggttaggga	accttgggta	tatcagggac	tctagcagtg	660
[0749]	gtgcttctct	ccctccgccc	ccctcaccac	ttccaaaat	aaaaaaccag	gaatgagaag	720
[0750]	accgcttggg	gttattgtaa	cacctgcact	agtgagtgac	cacacccctt	ttctctttc	780
[0751]	ccctcgcccc	cttgetgctg	ggccacagcc	cagaatccca	gggagtgate	agcgggtgctg	840
[0752]	ttttcttat	catctctctc	tgttctatcc	ctttcccctt	cctgaactgc	tttgtgaagg	900
[0753]	agcagtgtaa	ggcattcccc	caccatgaag	taagtgggtt	cgtgggggtt	gttgccctgtg	960
[0754]	gctgggacct	gggaggtacc	tgagagaatt	ggtgttattt	gggcttgtgg	ggaggggcta	1020
[0755]	agaaatgata	agaaaagaca	agaattctta	aaaggtgaaa	tgggagcagg	cttgagtcat	1080

[0756] ggacctgccc tagcctcccc cagtttgtgg ccctgatagg tgccctcctt gccatctgct 1140
 [0757] gcatgatctt cctgggcttt gcggatgatg tactgaatct gcgctggcgc cataagctgc 1200
 [0758] tgctaccac agctgcctca ctacctctcc tcatggteta tttcaccac tttggcaaca 1260
 [0759] cgaccattgt ggtgccgaag cccttccgcc cgattcttgg cctgcatctg gacttgggta 1320
 [0760] ggtagtctctg ccaactgctac tcctatggca cctacttcag ggcacccttc ctgggtgctt 1380
 [0761] acattctcct tcaagtgttc cttttctgtc tctgtgtctt cccagatcct ttctggtagc 1440
 [0762] ccttcatcct atcctctgtc ctaccactt ttctaaatcc tctccccta ggtggcacta 1500
 [0763] cttttcctac catctctccc ttcaggaatc ctgtactatg tctacatggg gctgctggca 1560
 [0764] gtgttctgta ccaatgcat caatataccta gcaggaatta atggcctaga ggctggccag 1620
 [0765] tcactagtaa tttctgcttc catcattgtc ttcaacctgg tagagttaga aggtaggtagg 1680
 [0766] gattgggggt ggggagagag aagtctgaat gttaaagggt tggcctgata tatgactttg 1740
 [0767] ggaaattcag ggaaaaaag caatatgcgt agtaattata gaagataagg gaggctactt 1800
 [0768] actttgcaaa taatgcagat ttattgaaag tgagaaagaa aaatagcagc cgtgtcattt 1860
 [0769] atagctggat tggcactaac agctaggcca tgatcttctc ccattgaata taacaattt 1920
 [0770] cacagaacct caacgttaca cagggtcatt ctgtgacat gatggaggaa gacaaaaact 1980
 [0771] cgaccctcc ctctataatc ctgtttgagc acagataaaa ccacaaaaac actgagcaac 2040
 [0772] ccacaaaatg gccaatgccc tctctctctg ttaacgtgag ccatgagcga ctgctgcggc 2100
 [0773] tttccaataa caactcagtt cctaccacct tttatttgt tttttgagac agggctctcc 2160
 [0774] tctgtcacc aggctggaga gcagtggcac gatcttagct cactgcatcc tctgactcaa 2220
 [0775] acgatcctcc tgccccagac tcccaggtag ctggggctac aggcatgcgc caccacacct 2280
 [0776] ggcaaathtt tgtatthttt gtagagacag gttttacca cgttgcttag gctggtcttg 2340
 [0777] aactcctggg ccgaagtgat ttgtcagcct tggcctcca aagtgtggg attacactct 2400
 [0778] tgagccactg tggccagcca gttcctacca cttcttagat aaacattaaa atgcttgatc 2460
 [0779] agagaattat tgttgtttc tttcttttt ctttttttt ttttttgag acggagtttc 2520
 [0780] actcttccc aggetggagt gcaatggtgc gatctcgct cactgcaacc tccacctccc 2580
 [0781] aggttcaggc gattctctg cctcagctc cctagtagct gggattacag gcatgtgcca 2640
 [0782] tcacgccag ctagtttgt attttagta gagatgggat ttctcatgt tggtaagct 2700
 [0783] ggtctcaaac tccagacctc aggtgatccg cccacctcg cctccaaag tgctgggatt 2760
 [0784] acaggtgta gccaccgac ccggccatga attacactg ctttctaaca gcaccaatc 2820
 [0785] cagagcaaaa ctcttactt cttttacct ctccaaaat acccaaaact gcaagccct 2880
 [0786] cctaactctc tcttactgag acattccgtg gttcccaatg gtgtgtggt tctgaagtct 2940
 [0787] ccctttttac aacaagtcac taaacctagc ttcgaactat agatgtgtt ctagtggct 3000
 [0788] ttggetgatg gacatcaaca aatgtttatt aaagctaagt actttttaa catgatgta 3060
 [0789] tttaaatctt gtaatggtt tatgtggcac atgttataat cagccctgtt ttacagatga 3120
 [0790] gtaaacagat ttagagaagt taaatgtgc atgatcaaga tcaaggtcac aaagctaaga 3180
 [0791] agtaaagttg gtgtccaaac tgacatcaga atgggctaaa ccaaatata gacagtaact 3240
 [0792] agtttgaag gctgcatgaa agaggtgaa tattgggaat tgccttgggt gacataaaag 3300
 [0793] gggattgag ttcttgaag tgacttgggt gaggtggatg atacagctgt aaacagaact 3360
 [0794] tagacaaaa taggacctg gtatgcagaa gaagtggta ttaattttcc cttctttctt 3420
 [0795] ccttgcttc aaagtgatt gtcgggatga tcatgtctt tcccttact tcatgatacc 3480
 [0796] ctttttttc accacttgg gattgctgta ccacaactg taagtaggc tatggataag 3540
 [0797] gggaaaagg gaaactacc cgaacacatg gcaaatggt cccttatcat aaccacctt 3600

[0798]	gtggtggaga ggttaaacct gtgcatacct ccatggaatt ttctgtgtct tcagttggtc	3660
[0799]	gtattctgaa atttctccct acccaacagt atttggggat gagtgcgtgg aggtcccagg	3720
[0800]	aatagatgaa ttcaggcct tggatcctgc agagttgtct cacaactgga gtctcctcta	3780
[0801]	agtcagaact agggtcagg ctagtacagt gcccataggg tgtgatgtga gagaaaggat	3840
[0802]	tggtaatgcc tcttgccact ggctcggatc ctctcccca cacaggtacc catcacgggt	3900
[0803]	gtttgtggga gacaccttct gttactttgc tggcatgacc tttgccgtgg tggcatctt	3960
[0804]	gggacacttc agcaagacca tgctactatt ctcatgccc caggtgttca acttctctta	4020
[0805]	ctcactgcct cagctcctgc atatcatccc ctgccctgc caccgatac ccaggtagcc	4080
[0806]	gctttggggc ttgaaatgga catcatagcc ttttacttg ggatatctaa tgccagcett	4140
[0807]	tacattgctg tgcaaaggga gtgggcccaa agaagggcta tttccatgtg agtaaccctt	4200
[0808]	tataacttcc aaagcacatt tatttgatc atctgatact cacagtggtt ctgataacag	4260
[0809]	caagcagcag agccagaaat agatctcagg ttgactccac attcaatgct ctctctattg	4320
[0810]	attagccaca gggaggagg ttcaaagt ggcccagtca catgaagctg tcttcccc	4380
[0811]	gcagactcaa tatcaagaca ggcaaactgg agatgagcta ttccaagttc aagaccaaga	4440
[0812]	gcctctctt cttgggcacc tttattttaa aggtaacagg gtaacaagga ggtaaggccc	4500
[0813]	taggetgcca tctgacctt gaggaatggg gaacctagc ctacatcaga tccaagggga	4560
[0814]	acttgaagc attaaataga tccacattcc taaagcatag gtattagctg aggttctctt	4620
[0815]	cacctctggt ccctccaggt ggcagagac ctccggctgg tgacaataca ccaaagtgat	4680
[0816]	actgaggatg gtgaattcac tgaatgtaac aacatgacct tcatcaactt gctacttaa	4740
[0817]	atctttgggc ccatacatga gagaaacctc acattgctcc tgctgctgct gcaggtgagg	4800
[0818]	atggggattg ggtttatacc tcttctctc ctttctccg tgattcttat tccagtccat	4860
[0819]	ttctcttgc agatcctggg cagtgcctc accttctcca ttcgatatca gctcgttcca	4920
[0820]	ctctctatg atgtctgagt cccttgatca ttgtcctta cctcacagtc tctaggattc	4980
[0821]	ctgactcagg ctgacctctc tctggcceca gactgectcc ttgcccaggc ctctctcact	5040
[0822]	cttcatactc ttccagattt tgttctcagc atttctttt ccctgtgac actggcatec	5100
[0823]	tgggcgtttc ttgcccccta ctgtctactg attggattt acttatgact ttctgcaact	5160
[0824]	tgctactctc cctctccatc ctgtctttgc agcctcacag ggtgggatac agaagtttt	5220
[0825]	ttttgcagc tatccacagt cacattcag agtctgact ctcaaggaa tactggtttt	5280
[0826]	tgggatagaa cttgggccag ggctaggac acaggtcca cagtgacctg ttatttgatt	5340
[0827]	gtaaattaag tgttctgatt agtaagaagt aagcagggg ccacatgctc tcaatggaga	5400
[0828]	caataaagtg ttgtctattt cttta	5424
[0829]	<210> 15	
[0830]	<211> 5894	
[0831]	<212> DNA	
[0832]	<213> 黑猩猩	
[0833]	<220>	
[0834]	<221> misc_feature	
[0835]	<222> (3408) .. (3408)	
[0836]	<223> n为a、c、g或t	
[0837]	<400> 15	
[0838]	aaaccgtagc tgcgtttccg ggaactgagt tgtgtttacc ttggcttccg actatgttgg	60
[0839]	caacagttt cctgcaagaa actggcgcgt ctccacacc tcgtcctcc tccccccc	120

[0840]	ctgcctttca atagccatct tcctggagcc ggaggcatcc cagattaagg gagaggtacg	180
[0841]	ggccctttaa gcttgacctg tggaggcgga cggagctaaa actgacgtgg aaccggaatg	240
[0842]	tgagcgggtg cagacacgtg gtacaaggag gcattcatct tggaaccggg caattggcat	300
[0843]	ttccgctctg ggtagtacat ctttaacata atgttaggga agtatccgtt cttggctgcc	360
[0844]	tttctttaat tgcgtttcca gtactctctc ggtgattcta ctcttgaaca taggatgaaa	420
[0845]	tttggaatca cacttctctt gcacttccat cccaccctc taatgcccatt attaaaaatg	480
[0846]	gcggccgccc ctttccgcag taatggttgt tcagcgaaca agatccgggc ggaaacagta	540
[0847]	gataggcggg tgcagcggg cagaacatag gttgccttag agaggttctc cgggtgtctcg	600
[0848]	aggcggctc aagttagagt tgttgggtt tgctcagatt ggtgtggaa gagcctgcct	660
[0849]	gtggggagcg gccactccat actgctgagg cctcaggact gctgctcagc ttgcccgtta	720
[0850]	cctgaagagg cggcggagcc gggcccctga ccggtcacca tgtgggcctt ctcggaattg	780
[0851]	cccatgccgc tgctgatcaa tttgatcgtc tcgctgctgg gatttgtgga cacagtcacc	840
[0852]	ctcatcccgg ctttccgggg ccacttcatt gctgcgcgc tctgtgtca ggacctcaac	900
[0853]	aaaaccagcc gacagcagat gtgagcagcg gcacacgggt ctgggcaggg ggcaagggt	960
[0854]	aaggaaggag tggctagggc agggcgggg accggggtgc ttgaccacac gtgaagactc	1020
[0855]	agaactaacc caggcagcct ggaactcgga gaggtgatga gcagaactta ctgcattgg	1080
[0856]	ggaaaggatg ggtagggacc cttgggtata tctgggactc tggcagtgtt gctttctcc	1140
[0857]	ctccgcccc ctcaccactt accagaataa aaaaccggga atgagaagac cactttgggt	1200
[0858]	tattgtaaca cctgcactag tgagtacca cgccccctt gctcttccc ctgccccct	1260
[0859]	tgctgctggg ccacagccca gaatcccagg gagtgatcag cgggtgctgtt ttccttatca	1320
[0860]	tcctctctg cttcatccct ttccccttc tgaactgctt tgtgaaggag cagtgtgaagg	1380
[0861]	cattccccca ccatgaagta agtgggttcg tgggggtgat tgcctgtggc tgggacctgg	1440
[0862]	gagtgacctg agagaattgg ggttatttgg gcttgtggg aggggctaag aaattatcag	1500
[0863]	aaaagacagg aattcttaaa aggtggaatg ggagcaggct tgagtcatgg acctgcccga	1560
[0864]	gccccccca gtttgtggcc ctgatagggt cctccttgc catctgctgc atgatcttc	1620
[0865]	tgggcttgc ggatgatgta ctgaatctgc gctggcgcca taaactgctg ctacctacag	1680
[0866]	ctgcctcact acctctctc atggtctatt tcaccaactt tggcaacacg accattgtgg	1740
[0867]	tgcccgaagc cttccgcccg atacctggcc tgcatctgga cttgggtagg tagtcctacc	1800
[0868]	actgetgcc ctatggcacc tacttcaggg aaccttctt ggtgctccac attctctcc	1860
[0869]	aagtgttct tttctgtctc tgtgtcttc cagatcctt ctggtagccc ttcatectat	1920
[0870]	cgtcgctct caccacttt ctaaaaatc ttaaatctc ctcccctagg tggcactact	1980
[0871]	tcttttcta ccattctcc ccgcaggaat cctgtactat gtctacatgg ggctgctggc	2040
[0872]	agtgttctgt accaatgcca tcaatcct agcaggaatt aacggcctag aggctggcca	2100
[0873]	gtcactagtc atttctgctt ccatcattgt cttcaacctg gtagagtgg aaggtagggtg	2160
[0874]	ggattggggg tggggagaga gaagtctgag cattaaggt gtggcctgat atatgacttt	2220
[0875]	gggaaattca gggaaaaaa gcaatatgtg tagtaattat agaagataag ggaagctact	2280
[0876]	tactttgcaa ataacaatgc agatttatta aaagtgaaga agaaaaatag cagccctgtc	2340
[0877]	atztatagct ggattggcac taatagctag gccatgatct tctcccattg aatataaaca	2400
[0878]	gtttcacaga accccaactg tacacagggt cattctgtga ccatgatgga gcaagactaa	2460
[0879]	aactagacc ctccctctgt aatcatgtt gagcacaggc aaaaccacaa gaacactgag	2520
[0880]	caaccacaa aatggccaag atcccctctc tcggctaaca tgagcgactg ctgctgctct	2580
[0881]	ccaataacaa ctcagttctt accacttctt ttttttttt ttgagacagg gtctcctct	2640

[0882]	gtcatgcagg ctggagagca gtggcgcaat cttagctcac tgcattcctct gactcaaacg 2700
[0883]	atctctctgc cccagcctcc caagtagctg gggctacagg catgtgccac cacacctggc 2760
[0884]	aaatTTTTgt atTTTTgtA gagacagggt ttccatgt tccctagct ggtcttgaac 2820
[0885]	tcctggactc aagtgatctg ccaggcctcc caaagtgtg ggattcactc ttgagccact 2880
[0886]	gtgcccagcc agttcctacc atttcttaa taaacattaa aatgcttgat catagaatta 2940
[0887]	ctcttgcttt cttttcttt cttttcttt tttttgaga cggagtttg ttcttgccca 3000
[0888]	ggccggagta caatggtgag atctcgctc accgcaacct ccgctccca ggttcaagcg 3060
[0889]	attctctgc ctcagcctcc ctagtagctg ggattacagg cacgtgccac cacgcccagc 3120
[0890]	taatTTTgta tttctagtag agacggggt tctccatgt ggtcaggctg gtctcgaact 3180
[0891]	cctgacctca ggtgatctgc ctgctcagc ctcccaaagt gctgggatta caggcgtgag 3240
[0892]	tcaccgcacc cggccatgaa ttactctgc tttctaacag caccagctc agagcaaaac 3300
[0893]	tactttctt caccctctc caaaataccc aaaacaaac ctactacaag cccctcctaa 3360
[0894]	caccctctta ctgagacatt ccgtggttc caatggtgtg tggtttnga agtctccctt 3420
[0895]	tttacaaca gtcatataac ctagcttga gctatagatg tgtttctgat ggtcttggct 3480
[0896]	gatgaacatc aacaagtgtt tattaaagct aagtactttt taaacactat cttatttaa 3540
[0897]	tcttgaatg gttttatgt gcagatgta taatcagccc tgttttacag atgagaaaac 3600
[0898]	aggcttagag aagtcaaatg tgtcatgatc aagatgaagg tcacaaagct aagaagtaa 3660
[0899]	gttggtatcc aaacttacat cagaatggc taaaccaa ttaagatag aactagtttg 3720
[0900]	gaaggctgca cgaagagggt ggaatattg gaattgcctt gggtagacata aaaggagtat 3780
[0901]	tgagttctta aaagtactt gggtaggtg gatgataaca gctgtaaaca gaacttagac 3840
[0902]	aaaaatagga ccaaggttg cagaggaagt gggattaac ttttcttct tcttctctg 3900
[0903]	cttccaaagg tgattgtcg gatgatcatg tcttttccct ctacttcatg ataccctttt 3960
[0904]	ttttaccac tttgggattg ctctaccaca actggtaagt aggcctgtg ataaggggac 4020
[0905]	aactacctga acacatggca aagatggccc ttatcataac ccaccttggt gtggtgaagc 4080
[0906]	taaacctgag catacctcta tggagtttc tgcgtctca gttggtagta ttctgaaatt 4140
[0907]	tctctctacc cagtagtagt taggatgag tgcgtggagg cccaggaat agttgaattc 4200
[0908]	aggccttggt atcctgcaga gttgctcac aactggagtc tcctctgagt cagaactagg 4260
[0909]	gtcagggcta gtccagtgc catagggtg gatgtgagag aagggttggt taatgcctct 4320
[0910]	tgccactggc tcggatctc tccccaca caggtagcca tcacgggtgt ttgtgggaga 4380
[0911]	tacctctgt tactttgtg gcatgacct tgcctggtg ggcattctg gacacttcag 4440
[0912]	caagaccatg ctactattct tcatgcccc ggtgttcaac ttctctact cactgcctca 4500
[0913]	gctctgcat atcatccct gccctgcca ccgataccc aggtagccgc tttggggctt 4560
[0914]	gaaatggaca tcatagcctt ttacttggg atatctaag ccagcctata catttgctgt 4620
[0915]	gcaaagggag tgggccccaa gaagggtat ttccatgta gtagccctt ataacttaca 4680
[0916]	aagcacattt atttgcataa tctgctacag tggttctgat aacagtaagc agcagagcca 4740
[0917]	gaaatgatc tcaggtgac tccacatca atgctctcc tattagccac agggaggagg 4800
[0918]	gttcaaatag tggcccagtc acatgaagct atcttcccc cgcagactca atatcaagac 4860
[0919]	aggcaactg gagatgagct attccaagt caagaccaag agcctctct tcttgggac 4920
[0920]	ctttatttta aagtaacag ggtaacaagg aggtaaggc ctaggctgac atcctgacct 4980
[0921]	tgaggaatgg ggaacctagt cctacatcag atccaaggg aacttgaag cattaatag 5040
[0922]	atccacattc ctaaagcata ggtattagct gaggttctct tcacctctg tccctccagg 5100
[0923]	tggcagagag cctccagctg gtgacagtac accagagtga gactgaagat ggtgaattca 5160

[0924]	ctgaatgtaa caacatgacc ctcatcaact tgctacttaa aatccttggg cccatacatg	5220
[0925]	agagaaacct cacattgctc ctgctgctgc tgcaggtag gatgggaatc gagtttatac	5280
[0926]	ctcctgtctt ccccttctgc gtgattctta ctccagtcca tttctctctg cagatcctgg	5340
[0927]	gcagtgccat caccttctcc attcgatata agctcgttcg actcttctat gatgtctgag	5400
[0928]	tcccttgatc attgtccttt acctcacagt ctctaggatt cctgactcag gctgacctct	5460
[0929]	ctctctggtc ccagactgcc tccttgccca ggcctctctc actcttcata ctctccaga	5520
[0930]	ttttgttctc agcattttcc tttctctgtg atcattggca tcctgggcgt ttcttgcct	5580
[0931]	ctactgacta ctgattggat ttacatag gctttctgag acttgctact ctctcctct	5640
[0932]	ccatcccatc tttgcagcct catagggtgg gatacagcag ctttttttgc agttatccac	5700
[0933]	actcacattt cagagtctg actctcaagg aaccactgg ttttgggata gaacttgggc	5760
[0934]	cagggttagg aacacaggct ccacggtag atgtcattg attgtaaatt aagtgttctg	5820
[0935]	attagtaaga actaagcagg gggccacatg ctctcaatgg agacaataaa gtgttctct	5880
[0936]	tttctattg tta	5894
[0937]	<210> 16	
[0938]	<211> 4557	
[0939]	<212> DNA	
[0940]	<213> 家犬	
[0941]	<400> 16	
[0942]	gtgaggaggc aagtgcggcg ggggacagcc gagggtgcgc gctggaggct cgcgggagtc	60
[0943]	ctgggggcgc ctcaattcag agttgggttt tgctcaggcc gctgtgggag gatccagcct	120
[0944]	gtgccgagcg gctgctctc cccgcggggg gctccgggct accgcccage tcgcccatta	180
[0945]	gccgaggcgg cggcagagcg gggcccctgg ctggtcatca tgtgggcctt cccggagttg	240
[0946]	ccgatcccgc tgctggtgaa tttggtcggc tcgctgctgg gatttgtggc cacggtcacc	300
[0947]	ctcatccccg ctttccgtgg ccacttcac gccgcgacc tctgtggcca ggacctcaac	360
[0948]	aaaaccggcc ggcagcagat gtgagcggg gcacccgggt ccggggaggg ggccggcagg	420
[0949]	gcaagggcgg gacctgggt gcctgacccc gcggacacgc agcgctaacc ccgacagacg	480
[0950]	ctgcgggctc tgggagacga agggcagcgc tggccaactc tgggaaggga tgttgcagta	540
[0951]	caggggaccc tcgggtgtat cagggactcc agcgtggtg cccttccacc ccccttccc	600
[0952]	gtagatcgtt gtaatgctt ctctagttag tgaccagcc cctctctc tccccgcc	660
[0953]	cctccctttg ctgctgggcc acagcccaga gtcccaggga gtgatcagcg gtgctgtttt	720
[0954]	ccttaccatc ctcttctgct tcatccctt ccccttctg aactgtttta tggaggagca	780
[0955]	gtgtaaagcc ttccccatc acgaagtaag tgggtgagtt gggggcggtt gcttggggct	840
[0956]	ggggcctggg agctacctgg gagagttgt gttattaggg tttgggtgga ggggctgagg	900
[0957]	aaggagcga gagacgggtg tttttgcaag atgatgtgg cataggctg agcggtagc	960
[0958]	tgcccagacc tccccagtt cgtggccctg ataggtgcgc tccttgccat ctgctgcatg	1020
[0959]	atcttctgg gctttgcgga cgatgtactg aatctgcgt ggcgccaca gctgctgctg	1080
[0960]	cctacagctg cctcgtacc tcttctatg gtctatttca ccaactttg caacacgacc	1140
[0961]	attgtggtgc ccaagcctt cggccgatt cttggcctgc atctggactt gggtagtag	1200
[0962]	ccctgtgact gacgtccctg tggccctac tttggggcacc ccttaccctg ggagataatc	1260
[0963]	tagcagagca tcattcctgg tgctccagat cctcttcaa gtgtcccat cttgttctg	1320
[0964]	tgtcttctca gatccgttct gttggtcctt cgtccaatcc tctgtcctca ccactttct	1380
[0965]	cagaagaata ttcttaagtc ctatttcta tggatggcac acttcttaet ctcttcttc	1440

[0966]	cccagggatc ctgtattatg tttacatggg gctgctggca gtgttctgta ccaatgceat	1500
[0967]	caatatccta gcaggaatta atggcctaga ggcaggccag tcgctagtta tttctgcttc	1560
[0968]	catcatcgtc ttcaatctgg tggagctgga aggtaggtga gagtgggagt ctgagtatta	1620
[0969]	aggaaactgc ctgatacctg gctttgggga attcaggaaa aaataaaagc aatatattaa	1680
[0970]	gattaaatgt aaagaaaaac agctctgtca ttgacagctg aattggcact aataggtagg	1740
[0971]	ccatggtctt ctgctgaaca taaacaattt cacagaactt cacaatcaga cgaggctact	1800
[0972]	ctatgtccat gatagagtaa agcaaacca gattcctcca taaacatgtc tgagtatage	1860
[0973]	cagaactgca ttttgtgcat cccacaaaa tgactaggat ctccctcttc tggctaaggt	1920
[0974]	gagcaattgc ttcttctga taacttggtt ctacttagag aaaactaaga tgctcataga	1980
[0975]	attacttcca ctgacagcac ccagcttgg gcaaaacttt gcctccttc tttctcccc	2040
[0976]	aaattactca aaacaatcct ataacacatc ttcttaatac ttccctactg aggcattccc	2100
[0977]	tggttaccta tgggtcgctg tctacagtgt ctctcttgtt acacgtcagt aaaccagct	2160
[0978]	ttgactgcag gtgtgtttct ggtggtcttt ggctgatgga tatcagtgtt tattaaaaca	2220
[0979]	aaatactctt aaagcattta aactttgtaa tgtggcaagt gttctcatga accatatttt	2280
[0980]	acagttgagg aaacagagg cgagagaatt taagtgtgc atgatcaagg tcacacagtt	2340
[0981]	agaaagtaaa gctagtattc aaacctgggc tgaatgatct aaaccaaatt gaagacagca	2400
[0982]	acttgattta ggaagggttc atgaatgagg tggaatatta ggaattgcct gagtgacaca	2460
[0983]	aaagaagtag tgagtctgg aatgggactt ggaagagtg gaaagtacag ctggggacag	2520
[0984]	aacttgagac agaaatagga cccagttatg cagggggaag taccttatca actcatcctt	2580
[0985]	ctttctttt cttcttccc tgcttccaaa ggtgattatc gggatgatca cgtcttttcc	2640
[0986]	ctctacttta tgataccctt ttttttacc accttgggat tgctctacca taactggtaa	2700
[0987]	gtgggccatg tgaacatgta gcaagtatgg tcctgttggc cctgacccaa ctctgttgg	2760
[0988]	agaggctaag cctgcgcaca cctgtattga gtgtttctg gatgcctagt tggtaatat	2820
[0989]	cttcaattac tctctacca gttgcagtta gagacaagt ctgtggagcc cccaagaaga	2880
[0990]	gatgaattca gggctttggg ttctggagcc ttgttggaa atctggagtt tcttccgggc	2940
[0991]	caggactaga gtcagggcta gtccagggtt cagggcgtgt aatatgagag aaaagactga	3000
[0992]	tagtgctcc tgccactggc tcagatcctc tccccacac aggtacccat cacaggtgtt	3060
[0993]	tgtgggagat accttctgtt actttgctgg catgacctt gccgtggtgg gcatcttggg	3120
[0994]	acacttcagc aagaccatgc tactcttctt catgccccag gtgttcaact tctctactc	3180
[0995]	actgcctcag ctctgcata tcataccctg cctcgcac cgcattcca ggtagccact	3240
[0996]	ttggggctta aaaggacat cttagctttt tcaattggga tgcataaagc cagccttctg	3300
[0997]	catctgctgt gtaaggggaa tgggccc aaa ggaggctct ttccgtgcaa ttagccctta	3360
[0998]	taaatgacag agcacattca cccacataat ctgatcagct ctgatcacac agtggtaagc	3420
[0999]	agagccgaa acagatcttc aggttgtctg attccactt cggctactct cctattaatt	3480
[1000]	gaccgcagtg tggagagttc ttggagtagt ggcccagtca cataaagctc tcttccccct	3540
[1001]	gcagactcaa taccaagaca ggcaaactgg agatgagcta ttccaagttc aagaccaaga	3600
[1002]	gcctctcttt cttgggcaac tttattctaa aggtaacag gtaacgaggt aaggctctag	3660
[1003]	gccaccatcc ggaattcagg gcctggggac cctcggttg catcagatcc aaggggagcc	3720
[1004]	tggaagcatg aagcagatcc cccattgctg aagcagagtt gaagtctct ccacctctgg	3780
[1005]	cccctccagg tagcagcgag cctgcagcta gtgacagtgc accagagtga gaatgaggat	3840
[1006]	ggtgccttca cggagtgtaa caacatgacg ctctcaact tgctccttaa ggttctcggg	3900
[1007]	cccatgcatg agagaaacct gactctgctc ctgctgctgc tccaggtgtg gtcagggag	3960

[1008]	ggctttgctg gctctggtct ccctttctcc atggctctga ctctggtgtg tttctttctc	4020
[1009]	ctcacagatc cttggcagtg ctgtcacctt ctccatccgg taccagcttg tccggctctt	4080
[1010]	ctacgatgtc tgagtcccc aatccttgcc cttcaactgca tagtctgcag ggttcctgac	4140
[1011]	tcaggcctgc ctctttctgg gccaggcacg cttccgggcc caggcctctc tcacctetta	4200
[1012]	ctttctcca gattttgtac ttagegattc cgttccgctg tgatcgacat cctgggcctg	4260
[1013]	tcttgccctg tactgactgt tgattggact ttgcctgtgg ctttcttcaa cttgctgctc	4320
[1014]	tcctctctta tcccatccct gggcctccc aaagtgggat actgtgcttt ttatgcagtt	4380
[1015]	atccaccact cggactctcg aggaatatgt tgggcctggg gatagaacce tggctgggga	4440
[1016]	gaggacaca ggctcgaaga tcacttgatt atttgacat aaattaagta ttctgattcg	4500
[1017]	taagagcaga ttggggggcc aggtgctccc agtggtgaca ataaagtgt gtctttt	4557
[1018]	<210> 17	
[1019]	<211> 5374	
[1020]	<212> DNA	
[1021]	<213> 中国仓鼠	
[1022]	<400> 17	
[1023]	caaggcagag cctaggttgc ttataaaac ctcttgggga agccccaggg cggttcaaat	60
[1024]	taagagttgt tgggttttgc cccgcctcgc atgtgaggag cggacactgc tcacggctga	120
[1025]	gacctcgggg ctgcttcca ccagttagct gagaaggctg cggagctgga acctctggcc	180
[1026]	actgccatg tgggccttct ctgaggtacc gattccgctg ctggtgaatt tgatcggctc	240
[1027]	gctgctggga tttgtggcca cgctcacct catcccggcc tttcgtggcc actttatcgc	300
[1028]	tgcgcgcctc tgtggccagg acctcaaca aaccaaccgg cagcagatgt gagcagtggc	360
[1029]	acacgggtgt cccgggcagg ggccaggggt gggcaaggca caggcgagct ctgaggtgct	420
[1030]	taaatgtgcg tacgaaccaa atctaactgg agttgtccgg gaccctggga ctcgatggcc	480
[1031]	agaagtggtt agcactgggg aatgctaagg aaggggacct ttgagtgaga acatccagcg	540
[1032]	gcgctgect cccccgcc cccactgcc tcccgtcca ctgctcccc gcctcactcc	600
[1033]	tgggaagatc ttttgggtca catggtttt gactaacca cgcccattc ttcttcttc	660
[1034]	tccacccct tgctgcgggg ccacagcca gaatcccagg gagtgatcag cggtgccgtt	720
[1035]	ttcctataca tctgttctg cttcatcccc ttccccttct tgaactgctt tgtgaaggag	780
[1036]	cagtgaagg ctttccccca ccatgaagta agtgggttcg tgggggcggt tgcttggggc	840
[1037]	ctgggaggtt cccgagagag ttggggttgt gtggatttga ggaggagga ctgaggacct	900
[1038]	agtggaaaag acagaaattt ttgaaagctt gaatggcagt aggcttgagt catgacctgc	960
[1039]	ccgagcctcc cccagtttgt ggccctcata ggtgcccttc ttgccatctg ctgcatgatt	1020
[1040]	ttcttgggct tcgcgacga tgtcctgaat ctacgtggc gccataagct gctgctgcc	1080
[1041]	acagctgat cactacctt tcttatgtc tattttacca actttggcaa cacaaccatt	1140
[1042]	gtggtacca agccctccg cccagttctt ggcttgatc tggatttggg tgagtatccc	1200
[1043]	tgctgtaca gccctgtgg cacttatctc aagtcacct cccccaaag gtgcccagca	1260
[1044]	gagaccctt cttgatgtc cacactccc tgtttttgt cgtccctgt gaatgctcag	1320
[1045]	gttctctctt gtgcctgtc attgtgtgtt ctgttttcag aataccgta gatcctttcc	1380
[1046]	tagctgtcac tgctttttat actatgtct gcaggatcc tgtactatgt ctacatgggg	1440
[1047]	ctcctggcag tgttctgtac caatgccatc aatatcctag caggaattaa cggcctagag	1500
[1048]	gccggccagt cattgtcat ctctgcttcc atcattgtct tcaacctggt ggagctgcaa	1560
[1049]	ggttgggtgg aagagagaga tctcagtgtt cagagaattg cctgatatat agctttgaga	1620

[1050] aaaggggggc ttatagaaga tagggaaagc tatttacttt gcaaataaca atgaagagtt 1680
[1051] acttgagtag gaggaagaaa aatagcagtc tgtcatttat agctggattg gcatgagtag 1740
[1052] ctagaccatg actgtttcct attggacata aatagtttca tagaacceca gcatgagaga 1800
[1053] ggggcgcct gaccgtggtg aaacaagaca gaaaccagac ttctcccact gtaatcatgt 1860
[1054] ctgaacaccg acaaaagcac aggaacaaag tcagcccaaa catcccttct tgtctaattg 1920
[1055] gagagggtat agcttctttg cagtaacaac tcagttctg ctacttctta agttgttcag 1980
[1056] tcagaaaatt acttctgctt tctgacatca ggcagtccag agcacaactt ttcttgcag 2040
[1057] cctcccaga accacttaa gtgaatccta ttgtaagtcc cttctaaca ctttagagt 2100
[1058] acctgccag catgaggcct tgggtacagt cccagtatc tctgtttgca tgcatgtaca 2160
[1059] catacccaca tgcacacact gaacttactt attgaaggta agcaatattt atttgcat 2220
[1060] tttgtgtg tgacaaaatt tctactatga attcagaata gccttgaatt cactatgtag 2280
[1061] cctaggcct cctcgaactt acagtataa tctgcctca gcttcttaag tgctaagatt 2340
[1062] caaggtatgc actaccaggc cagctaagaa agcaatttt aaactaggta tgggtggcaca 2400
[1063] catcactaat tctaactc tgggagactt gggcaggaag atcatgagtt tgagctcagc 2460
[1064] ctgggcactt ggtaagtctc tgtttctaga aataaaacat ggagtggta tacacacctg 2520
[1065] taatcccagc attcatgagg ctggggcagg aggatcacca caaggtcaag acctgcctgg 2580
[1066] gttacataag caagttcaag gccagcgtga actacgtagt gagaccctgc ctcaaacaaa 2640
[1067] caaataaata aataaacatg atcctgagtt tggttcccag tactcccc aataaatgaa 2700
[1068] atgaaatgaa agagctgggg aggcagctta gtgctaaggt ccaggaccc atgtgaaggc 2760
[1069] agctgcgtgt atgtgtatc agcctgttt catacactag ataataaac tggtttcaa 2820
[1070] acttaagtca gcatgtctgg acaaatgaa gactttaact tgtttttgac ggtttcatga 2880
[1071] cagtagtgag ctattgggaa ctgcctgggt accatcaaag gaataatgag gggctgggga 2940
[1072] tttagctcag cggcataagc gcctgccttg caagcaggca gtcagagtt tgatccccg 3000
[1073] taccgataaa aaggaaaaag acaaaaaaaa aaggaataat gagtttttga aggtggctcg 3060
[1074] ggtgagggga ggtggcaaca gagacagga tgcgacagac aaaatgaaga gcaggggaca 3120
[1075] taggggat ggggtgtcac ttttcttct ttgtccttg tttccaagg tgattaccg 3180
[1076] gatgatcatg tctttccct ctacttcatg ataccgttt tttcaccac cttgggattg 3240
[1077] ctgtatcata actggttaagg aggctgtgc tcagggaaaa ggaaaacaac taactgtca 3300
[1078] ttggacaaag atggtcctga tcttaacca gctcctgaaa gacagctga acttgcgcat 3360
[1079] acttttctc agtgtttct gggattcag ttgggtgatt gcctcccc gccccgttt 3420
[1080] tttttgaga cacctgtggc ctctcagtg ctaggccagt gctttactac tgagtcttg 3480
[1081] cctctagtat tctcaggtt gttctttct cagcagttg agacaagtgc tatggagccc 3540
[1082] caggaataat tatgggact tgcgttctgc agacttgeta gaccctctg tccgaactag 3600
[1083] gatcagggga gcatgtgtg tggctcacac ctgtaatcc agcactcagg agactgaggc 3660
[1084] aggaggatta ccatgagatc gagggcacc tcagctcaca tagtgactt gaggccagcc 3720
[1085] tggactacat agcgagactc ttgtctcaa aagaaaaaaa aagaaagaat aaaagaacag 3780
[1086] gggcttgagt tcttccagta cctagcctgt gtgatgtgag agaaaagact gtgatgcctc 3840
[1087] ttggcactgg cttggatcct tccccaca caggtacca tctcaggtgt ttgtgggaga 3900
[1088] caccttctgt tactttctg gcatgacctt tgccgtgta ggcacttgg gacacttctcag 3960
[1089] caagaccatg ctgcttctc tcatgccaca ggtgttcaac ttctctact cgctgcctca 4020
[1090] gctcctgcat atcatcccct gtccctgcc cgtataccc aggtagctgt ttgggggctg 4080
[1091] gaaagccttt ctactgggat gtctaacc aggctctaca tttgctgtgc aaagaatgtg 4140

[1092] ggcccatagg aaggctaact tttttcatgt aggtggccct ttaagtttac acagcacgtt 4200
[1093] tacttccata atctcattta atactcacag tagttctgat catagagtag taagcagcag 4260
[1094] agccagaaat agatctcact ccatgatcag tgtttttctt agtcattaac ggaagaaagt 4320
[1095] tttttgagta gtgaccagtg cacacgaagc tgtctttccc ctacagactc aataccaaga 4380
[1096] caggcaaaact ggagatgagc tattccaagt tcaagaccaa cagcctttct ttcttgggca 4440
[1097] cttttatfff aaaggtaaca aggtaacgag gaggtaaggc cccaggccac catcctgaac 4500
[1098] ttgggacatg ggggaccag gcctacatta gatctagagg gagcttggaa gcattaagca 4560
[1099] gagccctggt cctgacatac aggtattggt tgaagttttt ctgtctgtct ctgggtctctc 4620
[1100] taggtagcag agagactcca gctagtgaca gtgcaccgga gtgagggtga ggacggggcc 4680
[1101] ttcactgagt gtaacaacat gaccctcact aacttgctgc ttaaaatctt tgggcccata 4740
[1102] catgagagga acctcacatt gctcttctgt ctgctacagg tgagcctggg gtgagtttgt 4800
[1103] gcctcctcat gtccttttct ctatggttct tattctagtc cttttctctc tgcagatcgt 4860
[1104] gggcagtgct gtcaccttct ccattcgata ccagcttctc cgactcttct atgacgtttg 4920
[1105] agttcctgaa gattgcctc tgccacactg tctccagggg tcttgctcag gccagccagt 4980
[1106] ctggttctgt gggcctctcc caatcttcag tctccttcag atttattccc agcatttttc 5040
[1107] ataacctatg attatcaaca tctgagcca tttttgccct ccagcaacta ctaactggac 5100
[1108] tttgcctatg gcctccttca acttgccact ctccctacc atcacagcca gaggcttgat 5160
[1109] gtagcagctt ttatgcagat atccacaact cagcttccag agtcctcact ctcaaagaac 5220
[1110] atgctgggcc ttgagataga acctgagcta gggctagga cactggtgca aggggtgatt 5280
[1111] gatatttgat tataaattaa gtgttctgat tagtaagaca gaaggggagc ctgggtctcc 5340
[1112] caacggtgac aataaagtgt taccttttcc ttgt 5374

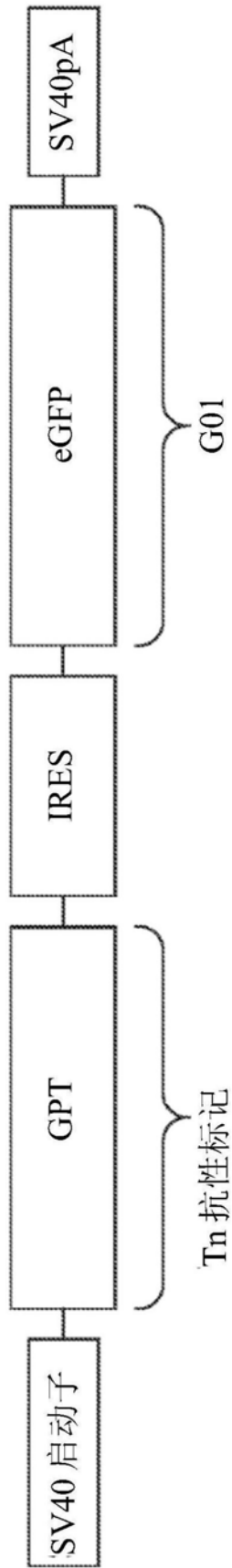


图1


```

*****
179 FCTNAINILAGINGLEAGOSLVISASIIVFNLVELEGDCRDDHVFSLYFMIPFFFTTLGL 238 GPT_HUMAN
179 FCTNAINILAGINGLEAGOSLVISASIIVFNLVELEGDCRDDHVFSLYFMIPFFFTTLGL 238 GPT_MACMU
179 FCTNAINILAGINGLEAGOSLVISASIIVFNLVELEGDCRDDHVFSLYFMIPFFFTTLGL 238 GPT_PANTR
179 FCTNAINILAGINGLEAGOSLVISASIIVFNLVELEGDCRDDHVFSLYFMIPFFFTTLGL 238 GPT_CANFA
179 FCTNAINILAGINGLEAGOSLVISASIIVFNLVELEGDCRDDHVFSLYFMIPFFFTTLGL 238 GPT_CAVPO
181 FCTNAINILAGINGLEAGOSLVISASIIVFNLVELEGDCRDDHVFSLYFMMPFFFTTLGL 240 GPT_RAT
181 FCTNAINILAGINGLEAGOSLVISASIIVFNLVELEGDCRDDHIFSLYFMIPFFFTTLGL 240 GPT_MOUSE
179 FCTNAINILAGINGLEAGOSLVISASIIVFNLVELEGDCRDDHVFSLYFMIPFFFTTLGL 238 GPT_CRIGR

*****
239 LYHNWYPSRVFVGDVDFCYFAGMTFAVVGILGHFSKTMLEFFMPQVFNFLYSLPQLLHIIP 298 GPT_HUMAN
239 LYHNWYPSRVFVGDVDFCYFAGMTFAVVGILGHFSKTMLEFFMPQVFNFLYSLPQLLHIIP 298 GPT_MACMU
239 LYHNWYPSRVFVGDVDFCYFAGMTFAVVGILGHFSKTMLEFFMPQVFNFLYSLPQLLHIIP 298 GPT_PANTR
239 LYHNWYPSQVDFVGDVDFCYFAGMTFAVVGILGHFSKTMLEFFMPQVFNFLYSLPQLLHIIP 298 GPT_CANFA
239 LYHNWYPSQVDFVGDVDFCYFAGMTFAVVGILGHFSKTMLEFFMPQVFNFLYSLPQLLHIIP 298 GPT_CAVPO
241 LYHNWYPSQVDFVGDVDFCYFAGMTFAVVGILGHFSKTMLEFFMPQVFNFLYSLPQLFQIIP 300 GPT_RAT
241 LYHNWYPSRVFVGDVDFCYFAGMTFAVVGILGHFSKTMLEFFMPQVFNFLYSLPQLFHIIP 300 GPT_MOUSE
239 LYHNWYPSQVDFVGDVDFCYFAGMTFAVVGILGHFSKTMLEFFIPOVFNFLYSLPQLLHAIP 298 GPT_CRIGR

*****
299 CPRHRIPRLNIKTGKLEMSYSKFKTKLSLFLGTFFILKVAESLQLVTVHQSETEDEGFTEC 358 GPT_HUMAN
299 CPRHRIPRLNIKTGKLEMSYSKFKTKLSLFLGTFFILKVAESLRLVTIHQSDTEDEGFTEC 358 GPT_MACMU
299 CPRHRIPRLNIKTGKLEMSYSKFKTKLSLFLGTFFILKVAESLQLVTVHQSETEDEGFTEC 358 GPT_PANTR
299 CPRHRIPRLNIKTGKLEMSYSKFKTKLSLFLGMFILKVAASLQLVTVHQSENEDEGAFTEC 358 GPT_CANFA
299 CPRHRIPRLNIKTGKLEMSYSKFKTNSLFLGTFFILKVAERLQLVTVHRSEGEDGAFTEC 358 GPT_CAVPO
301 CPRHRMPRLNIKTGKLEMSYSKFKTKLSLFLGTFFILKVAESLRLVTVHRGESEDEGAFTEC 360 GPT_RAT
301 CPRHRMPRLNAKTGKLEMSYSKFKTKNLSLFLGTFFILKVAENLRLVTVHOGESSEDEGAFTEC 360 GPT_MOUSE
299 CPRHRIPRLNPKTGKLEMSYSKFKTKNLSLFLGTFFILKVAERLQLVTVHRGESEDEGAFTEC 358 GPT_CRIGR

```

图2B

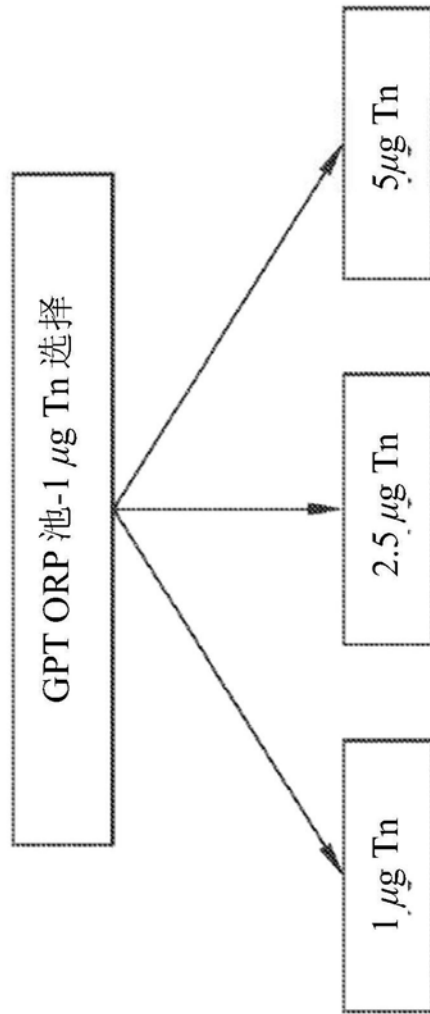


图3A

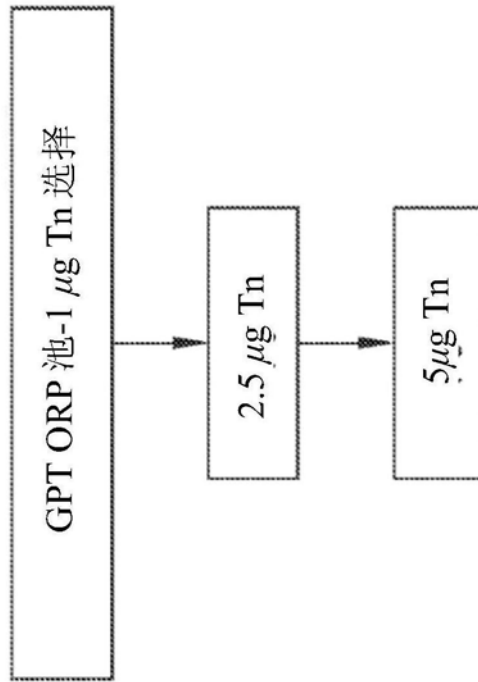


图3B

潮霉素选择效率

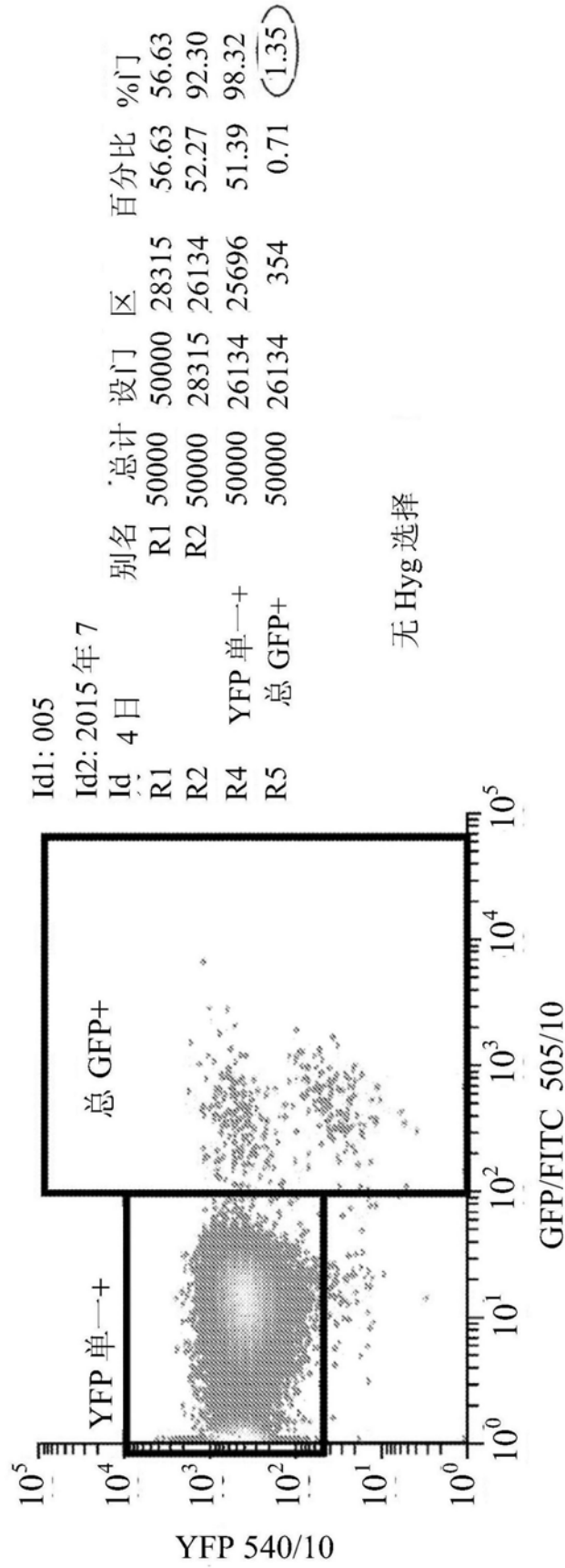


图4A

潮霉素选择效率

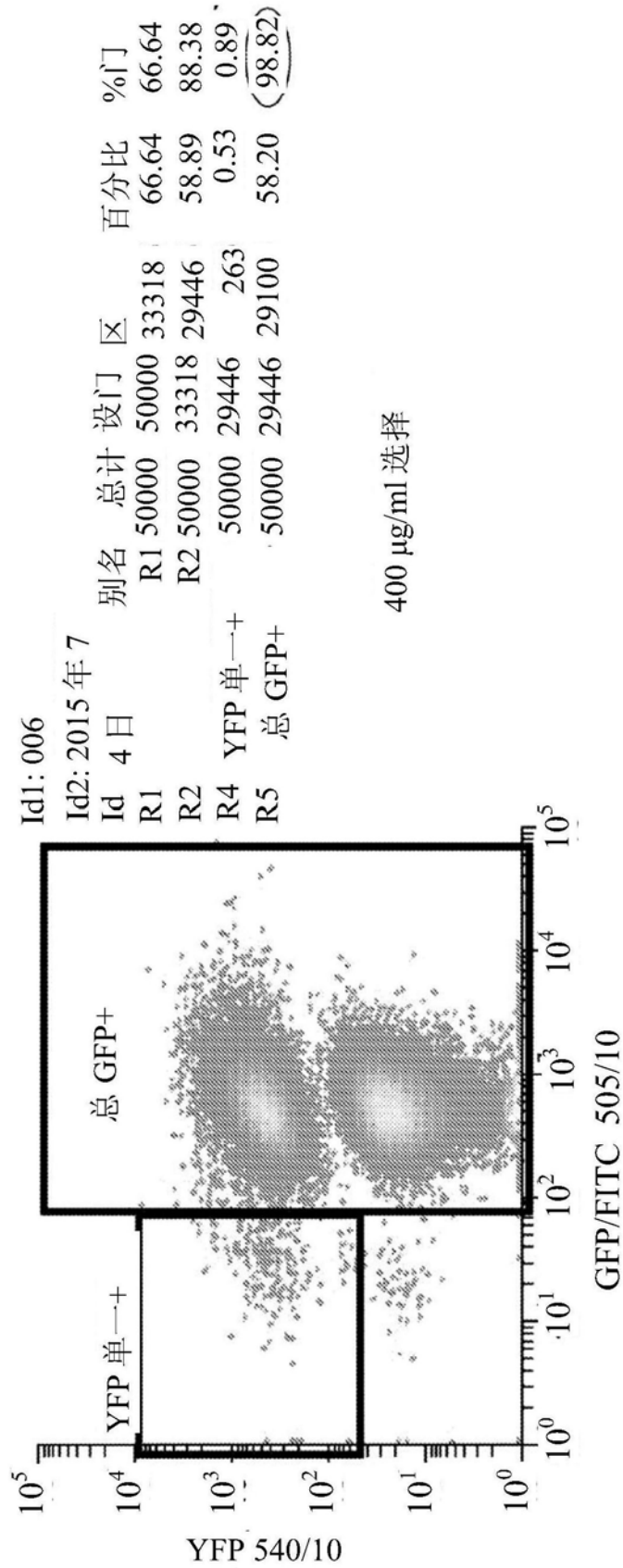


图4B

Tn 和 GPT 作为选择性标记的选择效率

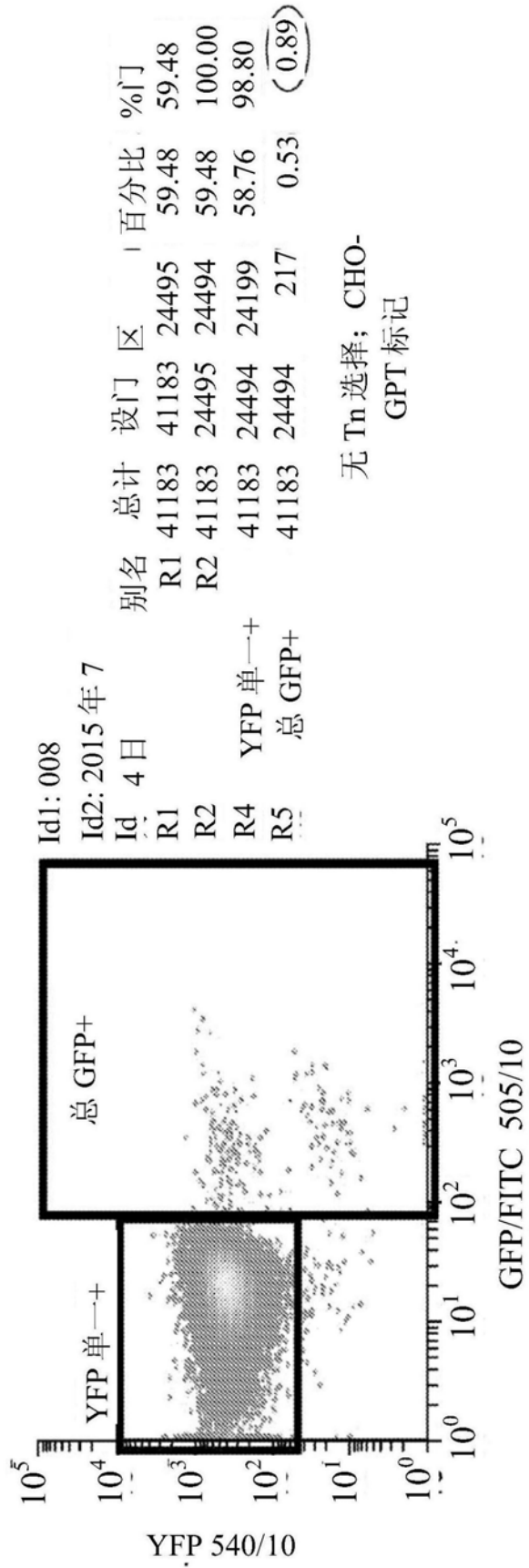


图5A

Tn 和 GPT 作为选择性标记的选择效率

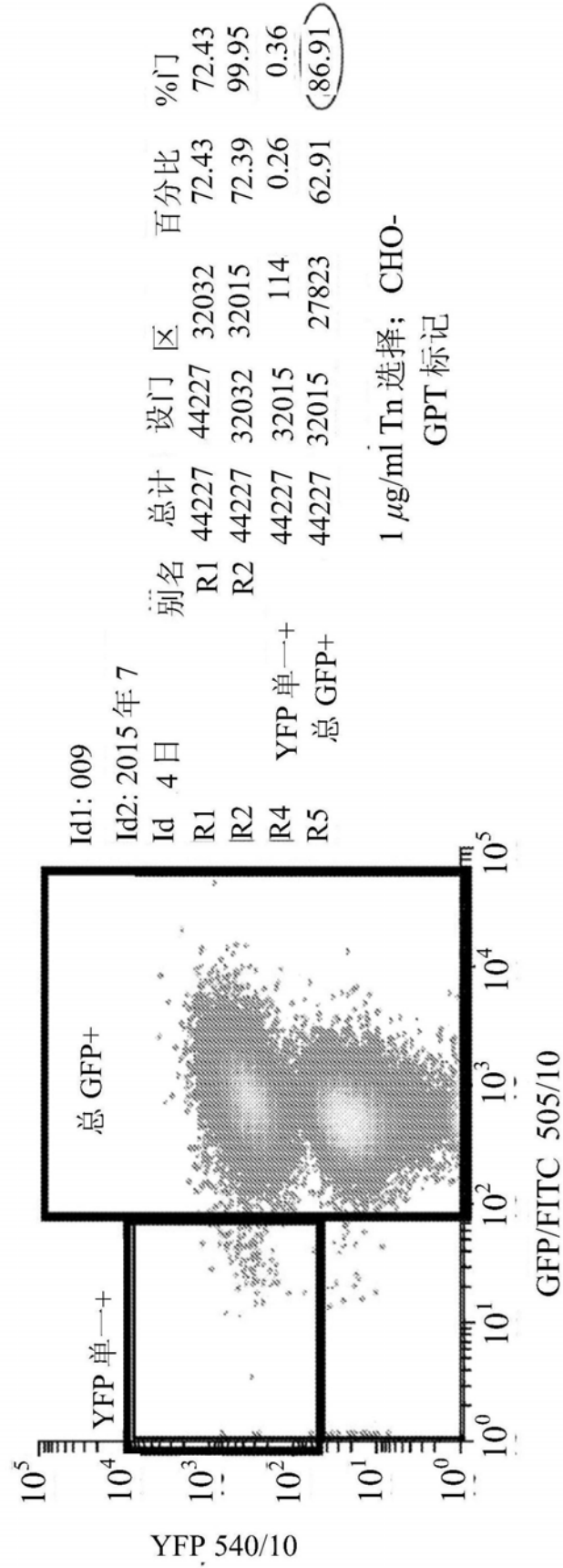


图5B

Tn 和 GPT 作为选择性标记的选择效率

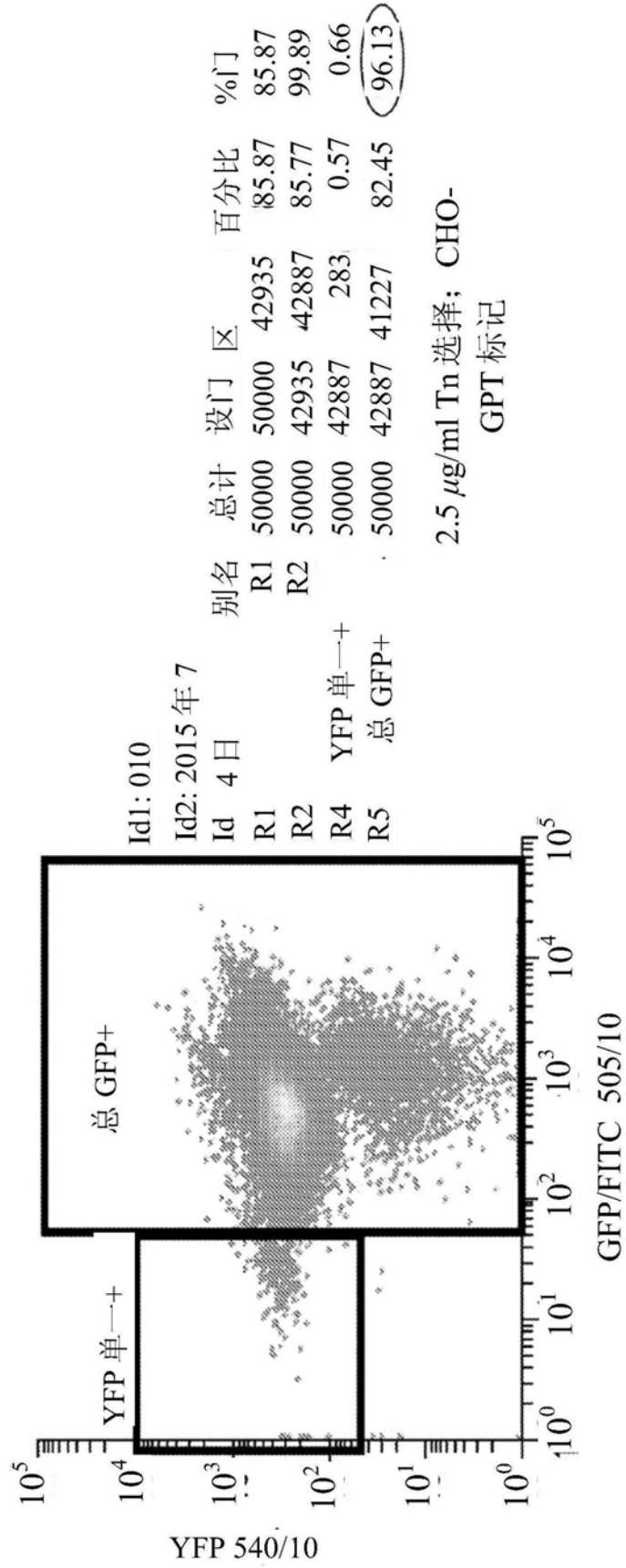


图5C

Tn 和 GPT 作为选择性标记的选择效率

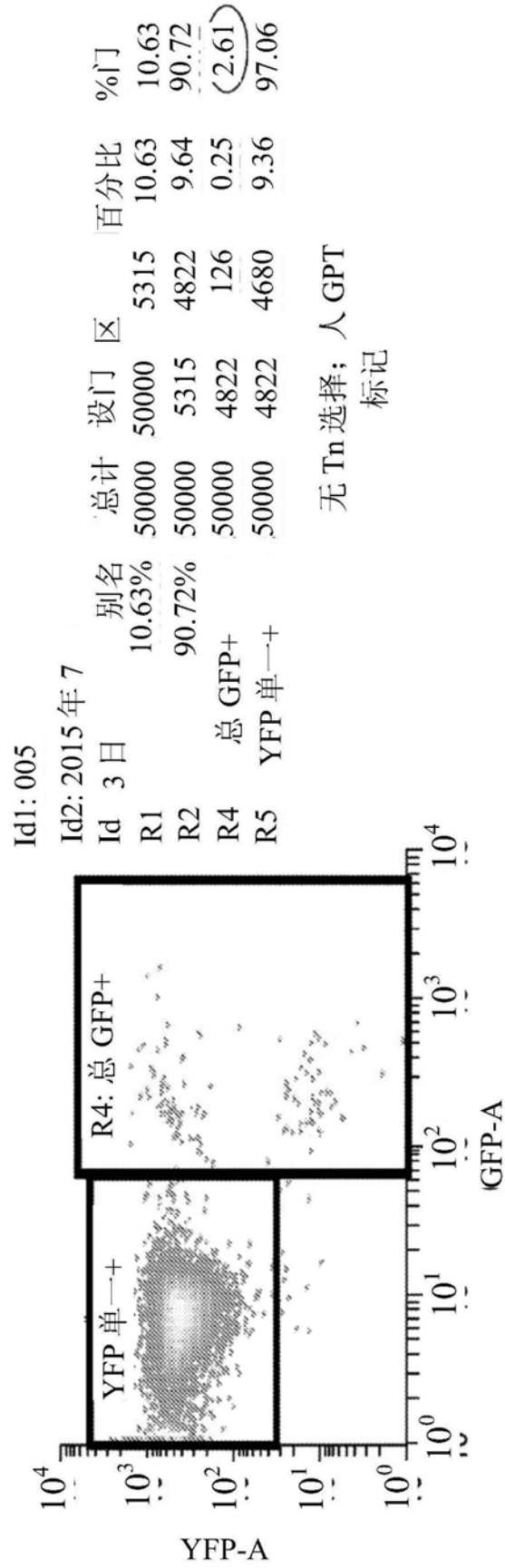


图5D

Tn 和 GPT 作为选择性标记的选择效率

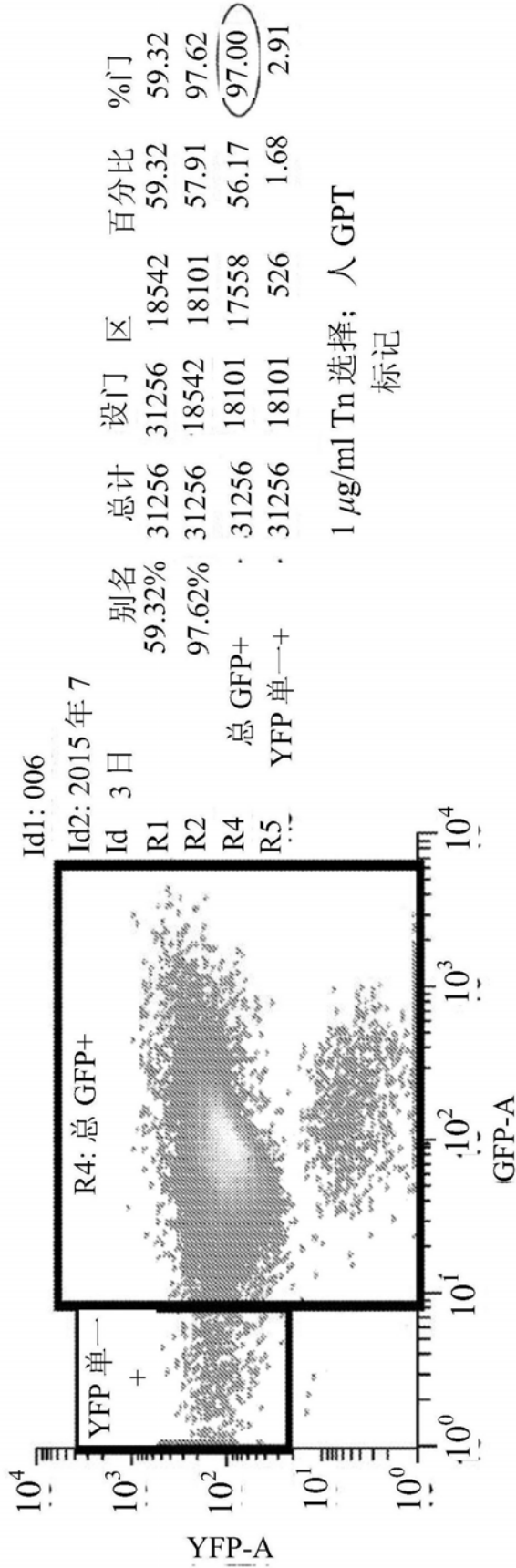


图5E

Tn 和 GPT 作为选择性标记的选择效率

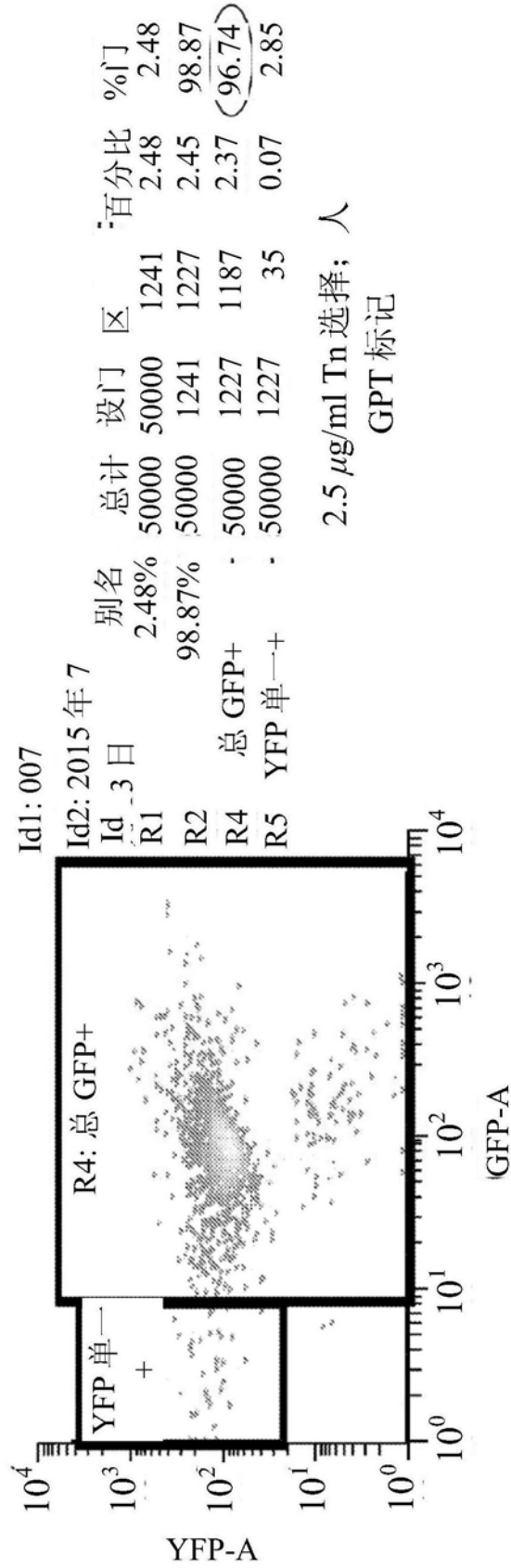


图5F

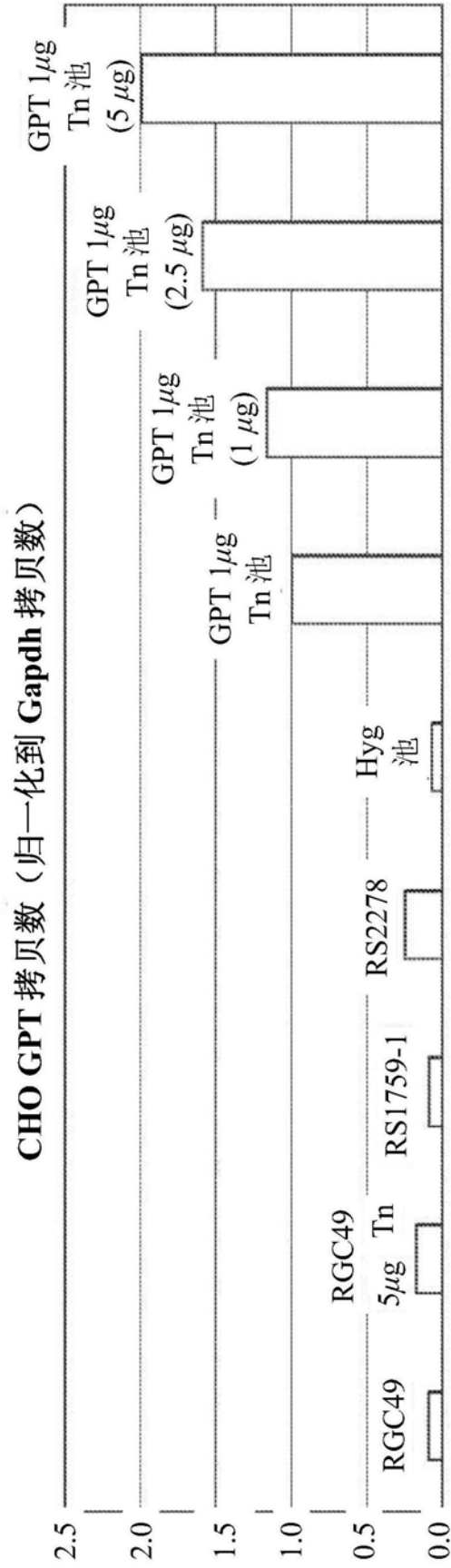


图6A

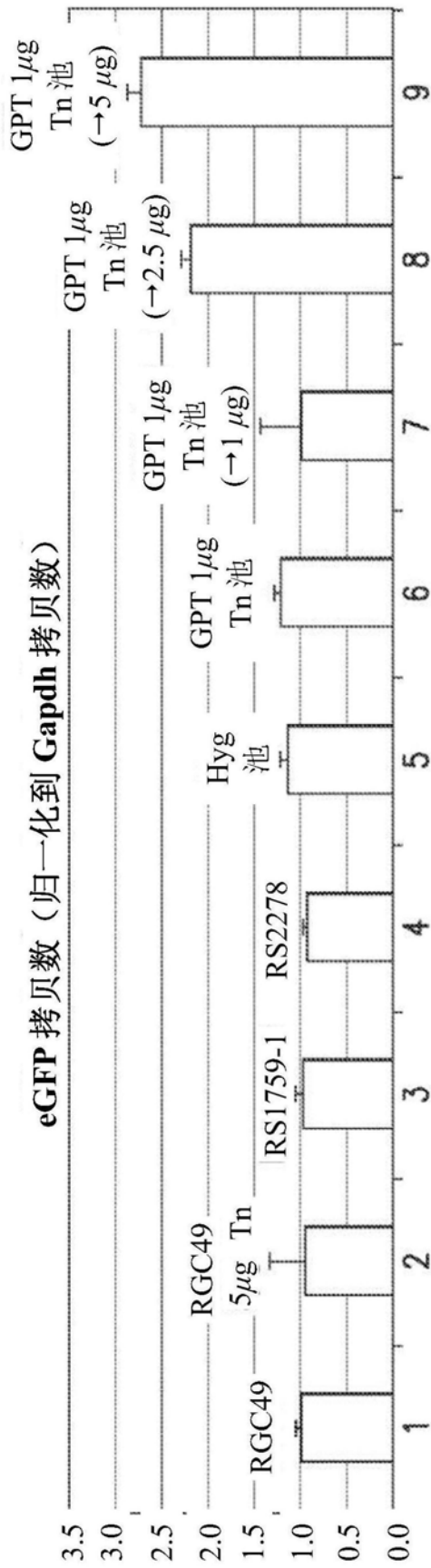


图6B

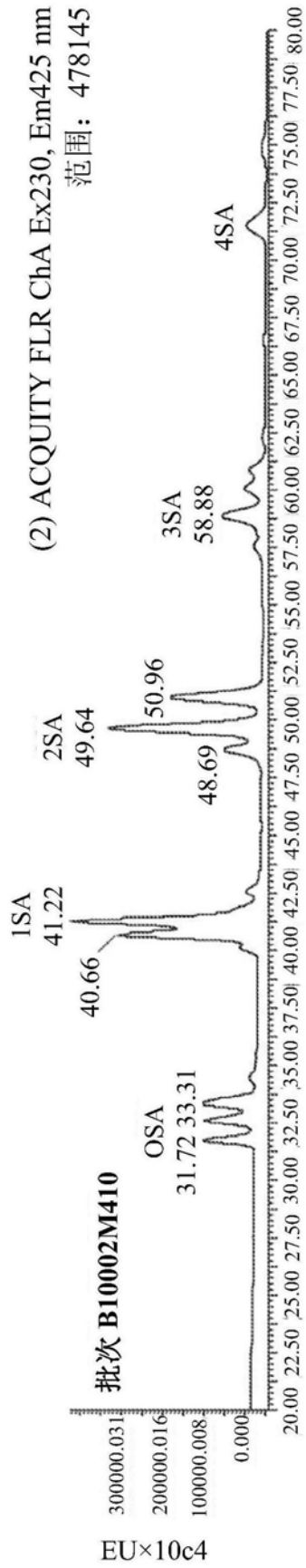


图7A

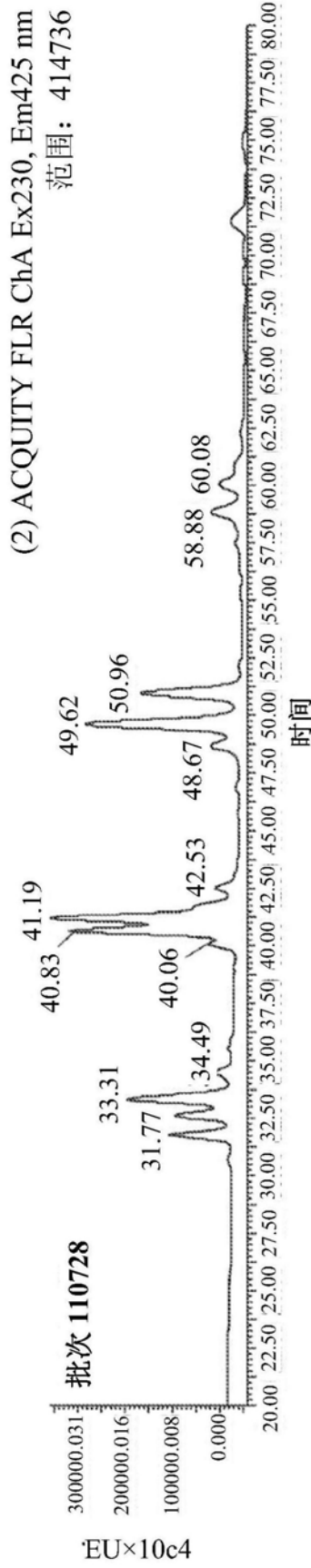


图7B

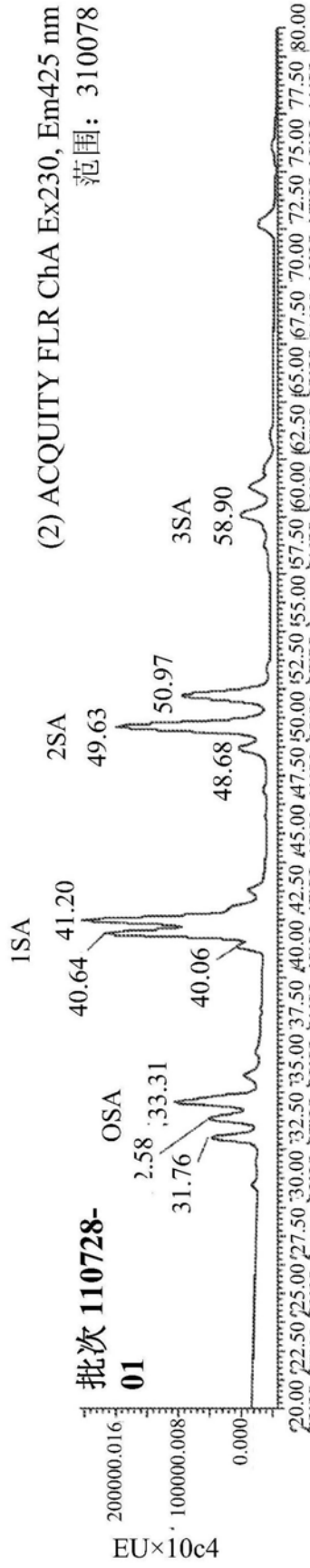


图7C

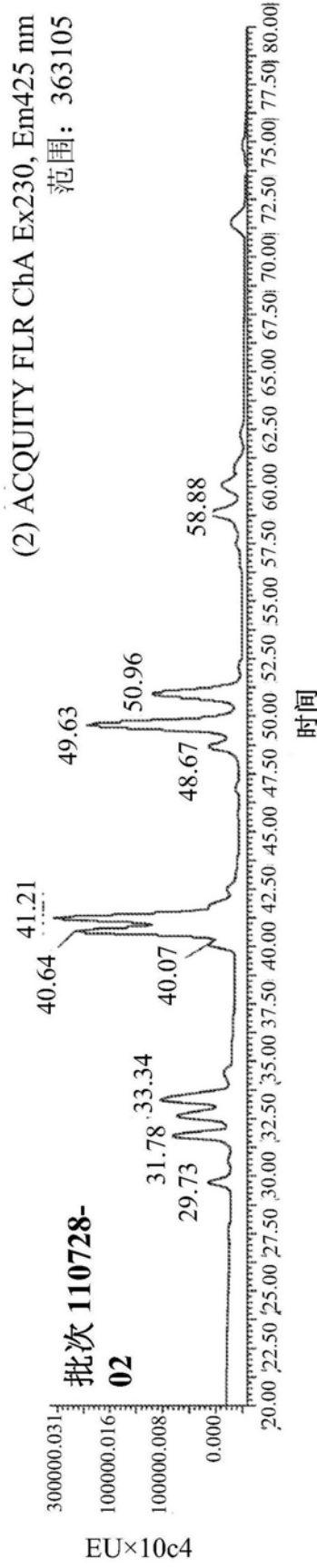


图7D

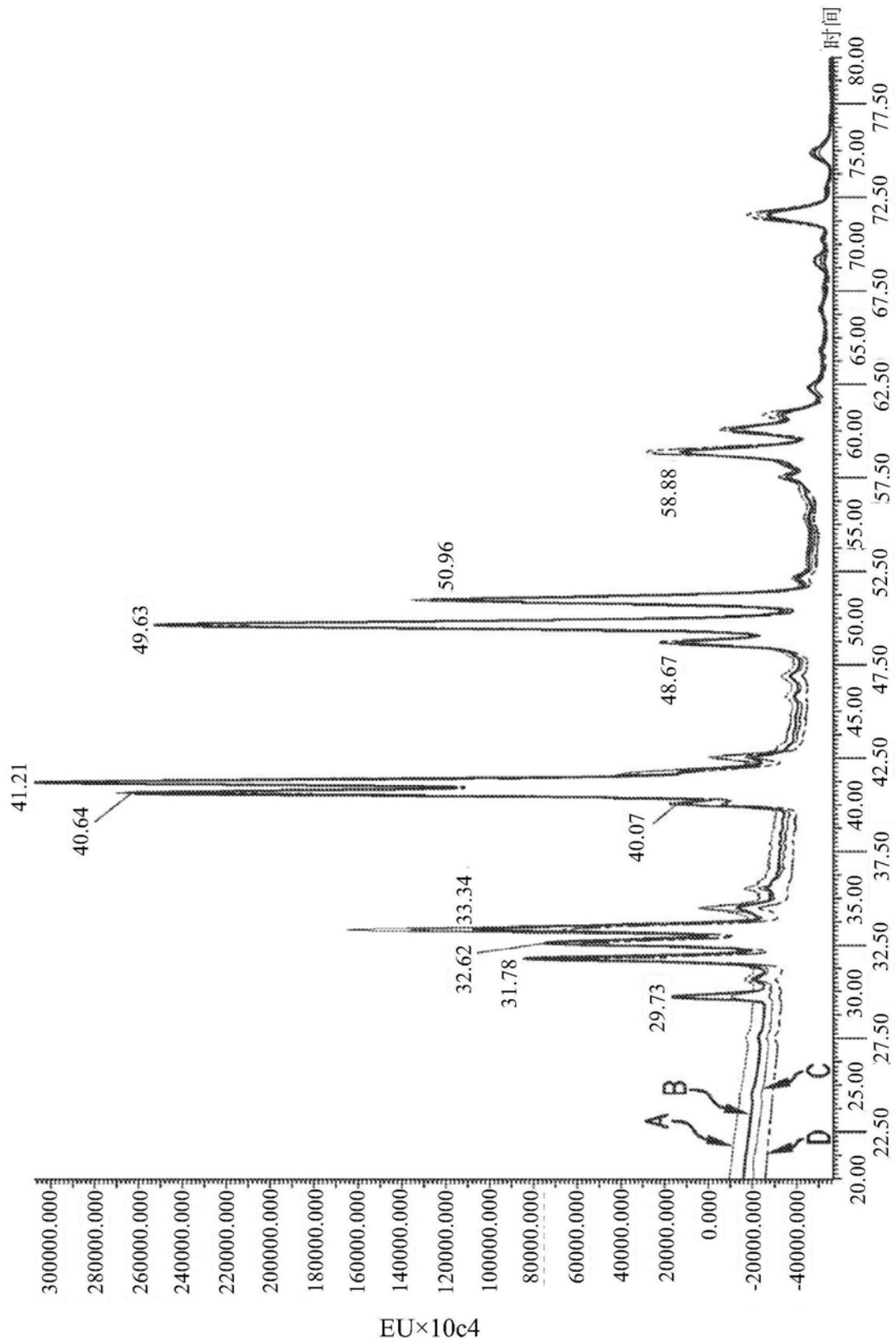


图8