(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 113396221 A
(43)申请公布日 2021.09.14

(21)申请号 202080009763.6

(22)申请日 2020.01.14

(30)优先权数据
62/793933 2019.01.18 US

(85)PCT国际申请进入国家阶段日
2021.07.16

(86)PCT国际申请的申请数据
PCT/IB2020/050267 2020.01.14

(87)PCT国际申请的公布数据
WO2020/148652 EN 2020.07.23

(71)申请人 詹森生物科技公司
地址 美国宾夕法尼亚州

(72)发明人 W·佩里

(74)专利代理机构 中国专利代理(香港)有限公司 72001
代理人 黄登高 初明明

(51)Int.Cl.
C12N 15/63(2006.01)
C12N 9/12(2006.01)
C12N 9/38(2006.01)
C12N 15/52(2006.01)
C12N 15/72(2006.01)

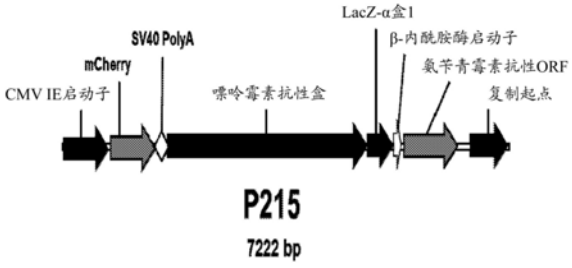
权利要求书2页 说明书41页
序列表13页 附图2页

(54)发明名称

作为非抗生素选择标记的β-半乳糖苷酶α肽及其用途

(57)摘要

本文提供了使用核酸构建体作为选择性标记的方法。该核酸构建体包含分离的β-半乳糖苷酶表达盒,该分离的β-半乳糖苷酶表达盒包含编码可操作地连接至启动子的β-半乳糖苷酶的氨基末端片段的核酸序列。还提供了包含该β-半乳糖苷酶表达盒的分离的载体、生成该分离的载体的方法以及包含该分离的载体的试剂盒。



CN 113396221 A

1.一种使用核酸构建体作为选择性标记的方法,所述方法包括:

a.使在乳糖操纵子中包含缺失的宿主细胞与所述核酸构建体接触,其中所述核酸构建体包含分离的β-半乳糖苷酶表达盒,所述分离的β-半乳糖苷酶表达盒包含编码能够操作地连接至启动子的β-半乳糖苷酶的氨基末端片段的核酸序列;以及

b.使所述宿主细胞在其中所述核酸构建体在所述宿主细胞中得以维持的条件下生长。

2.根据权利要求1所述的方法,其中β-半乳糖苷酶的所述氨基末端片段包含与SEQ ID NO:1具有至少75%同一性的氨基酸序列。

3.根据权利要求1或2所述的方法,其中β-半乳糖苷酶的所述氨基末端片段包含SEQ ID NO:1的氨基酸序列。

4.根据权利要求1-3中任一项所述的方法,其中所述核酸序列还包含复制起点。

5.根据权利要求4所述的方法,其中所述复制起点为高拷贝复制起点。

6.根据权利要求5所述的方法,其中所述高拷贝复制起点为pUC57复制起点。

7.根据权利要求6所述的方法,其中所述pUC57复制起点包含SEQ ID NO:19的核酸序列。

8.根据权利要求1-7中任一项所述的方法,其中所述分离的β-半乳糖苷酶表达盒还包含二聚体解离元件。

9.根据权利要求8所述的方法,其中所述二聚体解离元件包含具有位点特异性重组酶识别位点的核酸序列。

10.根据权利要求8或9所述的方法,其中所述二聚体解离元件还包含编码位点特异性重组酶的核酸序列。

11.根据权利要求8或9所述的方法,其中所述宿主细胞包含编码位点特异性重组酶的核酸序列。

12.根据权利要求8-11中任一项所述的方法,其中所述二聚体解离元件为ColE1二聚体解离元件。

13.根据权利要求12所述的方法,其中所述ColE1二聚体解离元件包含SEQ ID NO:20的核酸序列。

14.根据权利要求1-13中任一项所述的方法,其中所述宿主细胞包含LacZΔ15缺失。

15.根据权利要求1-14中任一项所述的方法,其中分离的载体包含所述分离的β-半乳糖苷酶表达盒。

16.根据权利要求15所述的方法,其中所述分离的载体的大小小于约1.5千碱基。

17.根据权利要求15或16所述的方法,其中所述分离的载体包含选自SEQ ID NO:9-13、17和18的核酸序列。

18.一种生成根据权利要求15-17所述的分离的载体的方法,其中所述方法包括:

a.使宿主细胞与所述分离的载体接触;

b.使所述宿主细胞在产生所述载体的条件下生长;

c.从所述宿主细胞分离所述载体。

19.根据权利要求18所述的方法,其中所述宿主细胞在基本培养基中生长。

20.根据权利要求19所述的方法,其中所述基本培养基包含乳糖作为唯一碳源。

21.根据权利要求20所述的方法,其中所述基本培养基包含约1%至约4%重量/体积

(w/v)乳糖。

22.根据权利要求21所述的方法,其中所述基本培养基包含约2%w/v乳糖。

23.一种试剂盒,所述试剂盒包含:

a.根据权利要求1-13中任一项所述的分离的β-半乳糖苷酶表达盒;以及

b.在乳糖操纵子中包含缺失的宿主细胞。

24.根据权利要求23所述的试剂盒,所述试剂盒还包含基本培养基,所述基本培养基包含乳糖作为唯一碳源。

25.根据权利要求23或24所述的试剂盒,其中载体包含所述分离的β-半乳糖苷酶表达盒。

26.根据权利要求23-25中任一项所述的试剂盒,其中所述宿主细胞包含所述LacZΔ15缺失。

27.根据权利要求26所述的试剂盒,其中所述宿主细胞选自大肠杆菌(E.coli)宿主细胞和酵母宿主细胞。

作为非抗生素选择标记的β-半乳糖苷酶α肽及其用途

技术领域

[0001] 本发明涉及包含非抗生素选择标记的分离的β-半乳糖苷酶表达盒。具体地，该分离的β-半乳糖苷酶表达盒包含可操作地连接至启动子的β-半乳糖苷酶的氨基末端片段。还提供了包含该β-半乳糖苷酶表达盒的分离的载体、产生该分离的载体的方法以及包含该分离的载体的试剂盒。

[0002] 以电子方式提交的参考序列表

[0003] 本申请包含序列表，该序列表作为ASCII格式的序列表经由EFS-Web以电子方式递交，文件名为"JBI6031USPSP1Seqlist1"并且创建日期为2019年1月17日，并且大小为48kb。经由EFS-Web提交的该序列表是本说明书的一部分并且全文以引用方式并入本文。

背景技术

[0004] 质粒载体通常含有在大肠杆菌(E.coli)中表达的基因，并且当通过转化或电穿孔将质粒引入到细胞中时，提供了从那些不含有质粒的细胞中鉴定或选择含有质粒的细胞的方法。最常用的选择性标记是赋予抗生素抗性的基因。然而，在几种情况下抗生素抗性基因是不期望的。当质粒用于形成生物制剂(诸如抗体)的制造细胞系时，通常移除或破坏抗生素抗性基因。对于基因疗法，抗生素抗性基因也是不期望的。虽然卡那霉素/新霉素抗性基因通常被FDA容许，但EU监管机构要严格得多。欧洲药典指出"除非另有证据和授权，否则用作选择性遗传标记的抗生素抗性基因(尤其对于临床上有用的抗生素)不包括在载体构建体中。其他用于重组质粒的选择技术是优选的("Gene transfer medical products for human use"，European Pharmacopei 7.0，2011年)。虽然当细胞系开发需要少量质粒时，对抗生素选择标记的破坏是可能的，但这些技术对于其中需要制造更多质粒的基因疗法应用是不切实际的。

[0005] 为了开发基于质粒的基因疗法以避免体内基因沉默，需要其中复制起点和选择标记的组合大小＜1kb的质粒载体。当与具有3kb或更大的质粒主链的传统质粒相比，质粒主链为1kb或更小时，治疗转基因在小鼠中表达更长久并以更高水平表达(Lu等人，Mol.Ther.，第20卷第11期，第2111-2119页，2012年)。已提出，未在体内表达的大段DNA诱导沉默。因此，具有较小质粒主链的质粒可能更有效。

[0006] 对于其中使用瞬时转染来制造治疗剂的应用，也需要较小的质粒。一个示例是腺病毒相关病毒载体的产生，其中使用质粒的大规模转染来生成临床材料。较小的质粒减少必须被转染的DNA的量，从而降低成本。

[0007] 因此，需要生成可用于基因疗法应用的包含选择性标记的较小质粒。

发明内容

[0008] 在一个总的方面，提供了使用核酸构建体作为选择性标记的方法。该方法包括(a)使在乳糖操纵子中包含缺失的宿主细胞与核酸构建体接触，其中核酸构建体包含分离的β-半乳糖苷酶表达盒，该分离的β-半乳糖苷酶表达盒包含编码可操作地连接至启动子的β-半

乳糖苷酶的氨基末端片段的核酸序列;以及(b)使宿主细胞在其中核酸构建体在该宿主细胞中得以维持的条件下生长。

[0009] 在另一个总的方面,提供了分离的β-半乳糖苷酶表达盒。该分离的表达盒包含编码可操作地连接至启动子的β-半乳糖苷酶的氨基末端片段的核酸序列。

[0010] 在某些实施方案中,β-半乳糖苷酶的氨基末端片段包含与SEQ ID NO:1具有至少75%同一性的氨基酸序列。在某些实施方案中,β-半乳糖苷酶的氨基末端片段包含SEQ ID NO:1的氨基酸序列。

[0011] 在某些实施方案中,核酸序列还包含复制起点。复制起点可例如为高拷贝复制起点。在某些实施方案中,高拷贝复制起点为pUC57复制起点。在某些实施方案中,pUC57复制起点包含SEQ ID NO:19的核酸序列。

[0012] 在某些实施方案中,分离的β-半乳糖苷酶表达盒还包含二聚体解离元件。二聚体解离元件可例如包含具有位点特异性重组酶识别位点的核酸序列。二聚体解离元件还可包含编码位点特异性重组酶的核酸序列。在某些实施方案中,宿主细胞包含编码位点特异性重组酶的核酸序列。二聚体解离元件可例如为ColE1二聚体解离元件。在某些实施方案中,ColE1二聚体解离元件包含SEQ ID NO:20的核酸序列。

[0013] 还提供了包含本发明的分离的β-半乳糖苷酶表达盒的分离的载体。在某些实施方案中,分离的载体的大小小于约1.5千碱基。在某些实施方案中,分离的载体包含选自SEQ ID NO:9-13、17和18的核酸序列。

[0014] 还提供了生成本发明的分离的载体的方法。该方法包括(a)使宿主细胞与分离的载体接触;(b)使宿主细胞在产生该载体的条件下生长;以及(c)从宿主细胞分离载体。

[0015] 在某些实施方案中,宿主细胞在基本培养基中生长。基本培养基可包含乳糖作为唯一碳源。在某些实施方案中,基本培养基包含约1%至约4%重量/体积(w/v)乳糖。在某些实施方案中,基本培养基包含约2%w/v乳糖。

[0016] 还提供了试剂盒,该试剂盒包含(a)本发明的分离的β-半乳糖苷酶表达盒;以及(b)在乳糖操纵子中包含缺失的宿主细胞。在某些实施方案中,试剂盒还包含基本培养基,该基本培养基包含乳糖作为唯一碳源。在某些实施方案中,载体包含分离的β-半乳糖苷酶表达盒。在某些实施方案中,宿主细胞包含LacZΔM15缺失。在某些实施方案中,宿主细胞选自大肠杆菌宿主细胞和酵母宿主细胞。

附图说明

[0017] 结合附图进行阅读时,能够更好地理解前述发明内容和以下对本专利申请的优选实施方案的详细说明。但是,应当理解,本专利申请不限于附图中示出的精确实施方案。

[0018] 图1示出了P215质粒的示意图。

[0019] 图2示出了P216质粒的示意图。

[0020] 图3示出了P217质粒的示意图。

[0021] 图4示出了P218质粒的示意图。

[0022] 图5示出了P219质粒的示意图。

[0023] 图6示出了P469-2质粒的示意图。

具体实施方式

[0024] 背景以及说明书全篇中引用或描述了各种出版物、文章和专利;这些参考文献中的每一者全文均以引用方式并入本文。本说明书中包括的对文件、行为、材料、装置、文章等的讨论旨在为本发明提供上下文。此类讨论并不是承认这些事项中的任一事项或全部事项均相对于所公开或受权利要求书保护的任何发明形成现有技术的一部分。

[0025] 除非另有定义,否则本文使用的所有技术和科学术语的含义与本发明所属领域的普通技术人员通常所理解的含义相同。否则,本文所用的某些术语具有本说明书中所述的含义。

[0026] 应该注意的是,除非上下文清楚地指明,否则如本文和所附权利要求中所用的单数形式"一个/一种"和"所述/该"包括复数引用物。

[0027] 除非另有说明,否则任何数值,例如本文所述的浓度或浓度范围,应理解为在所有情况下均由术语"约"修饰。因此,数值通常包括所述值±10%。例如,1mg/mL的浓度包括0.9mg/mL至1.1mg/mL。同样,1%至10%(w/v)的浓度范围包括0.9%(w/v)至11%(w/v)。除非上下文另有明确指示,否则如本文所用,使用的数值范围明确地包括所有可能的子范围、该范围之内的所有单个数值,包括此类范围之内的整数和该范围之内的分数。

[0028] 除非另有说明,否则在一系列元素之前的术语"至少"应理解为指代系列中的每个元素。本领域的技术人员将会认识到、或仅使用常规实验就能够确定本文所述发明的具体实施方案的多种等同物。此类等同物旨在由本发明所涵盖。

[0029] 如本文所用,术语"包含"、"包括"、"具有"或"含有"或其任何其他变型应被理解为是指包括指明的整数或整数组,但不排除任何其他整数或整数组,并且旨在是非排他性的或开放式的。例如,包含元素列表的组合物、混合物、过程、方法、制品或设备不一定仅限于那些元素,而是可以包括这些组合物、混合物、过程、方法、制品或设备未明确列出或固有的其他元素。此外,除非明确地指明相反,否则"或"是指包括性的或而不是指排他性的或。例如,条件A或B由以下任一项满足:A为真(或存在)而B为假(或不存在),A为假(或不存在)而B为真(或存在),以及A和B均为真(或存在)。

[0030] 如本文所用,多个列举的要素之间的连接术语"和/或"被理解为包含单个选项和组合选项两者。例如,在两个要素由"和/或"连接的情况下,第一种选项是指在没有第二个要素的情况下适用第一个要素。第二种选项是指在没有第一个要素的情况下适用第二个要素。第三种选项是指适合一起使用第一要素和第二要素。这些选项中的任一种均被理解为落在含义内,并且因此满足如本文所用的术语"和/或"的要求。多于一种选项的并行适用性也被理解为落在含义内,并且因此满足术语"和/或"的要求。

[0031] 如本文所用,在说明书和权利要求书中通篇使用的术语"由.....组成"表示包括任何列举的整数或整数组,但不能将附加的整数或整数组添加到指定的方法、结构或组成中。

[0032] 如本文所用,在说明书和权利要求书中通篇使用的术语"基本上由.....组成"表示包括任何列举的整数或整数组,并且任选地包括不会实质上改变指定方法、结构或组成的基本或新型特性的任何列举的整数或整数组。参见M.P.E.P.§2111.03。

[0033] 还应当理解,当提及优选发明的部件的尺寸或特性时,本文使用的"约"、"大约"、"大致"、"基本上"和类似的术语表示所描述的尺寸/特性不是严格的边界或参数,并且不排

除在功能上相同或相似的微小变化，如本领域普通技术人员所理解的。至少，包括数值参数的这种参考将包括使用本领域已接受的数学和工业原理(例如，舍入、测量或其他系统误差、制造公差等)的变化，不会改变最低有效数字。

[0034] 在两个或更多个核酸或多肽序列(例如氨基末端β-半乳糖苷酶肽和编码它们的多核苷酸；本文所述的分离的载体的核酸)的上下文中术语"相同"或"同一性"百分比是指当进行比较和比对以获得最大对应性时，相同的或具有相同的氨基酸残基或核苷酸的指定百分比的两条或更多条序列或子序列，如使用以下序列比较算法中的一种序列比较算法或通过目视检查所测量的。

[0035] 对于序列比对，通常将一个序列作为参考序列，将测试序列与其进行对比。当使用序列比对算法时，将测试序列和参考序列输入计算机中，指定子序列坐标(如果必要的话)，并指定序列算法的程序参数。然后，序列比较算法基于指定的程序参数来计算对于测试序列相对于参考序列的序列同一性百分比。

[0036] 用于比较的序列的最佳比对可例如通过Smith&Waterman的局部同源性算法进行(Adv.Appl.Math.第2卷：第482页(1981年))，通过使用Needleman&Wunsch的同源比对算法进行(J.Mol.Biol.第48卷：第443页(1970年))，通过搜索Pearson&Lipman的相似性方法进行(Proc.Nat′l.Acad.Sci.USA第85卷：第2444页(1988年))，通过这些算法(GAP，BESTFIT，FASTA和TFASTA，在威斯康星州遗传学软件包，遗传学计算小组，威斯康辛州，麦迪逊，科学街区575号(575 Science Dr.，Madison，WI)计算机化实施，或通过目测(大体参见Current Protocols in Molecular Biology，F.M.Ausubel等人编辑，《实验室指南》，格林尼出版协会和约翰威利父子公司的合资企业(1995年增补版)(Ausubel))。

[0037] 适用于确定序列同一性百分比和序列相似性的算法的示例是BLAST和BLAST 2.0算法，其分别描述于Altschul等人1990年的J.Mol.Biol.第215卷，第403至第410页和Altschul等人1997年的Nucleic Acids Res.第25卷：第3389-3402页。用于执行BLAST分析的软件可通过国家生物技术信息中心公开获得。该算法涉及首先通过识别查询序列中长度为W的短字词来识别高评分序列对(HSP)，当与数据库序列中相同长度的字词比对时，该短字词匹配或满足一些正值阈值分数T。T被称为邻近字词分数阈值(Altschul等人，出处同上)。这些初始邻近字词命中充当用于发起搜索以找到包含它们的较长HSP的种子。然后将字词命中沿每个序列在两个方向上延伸，只要可增加累积的比对分数。

[0038] 对于核苷酸序列，使用参数M(一对匹配残基的奖励分数；始终＞0)和N(错配残基的处罚分数；始终＜0)来计算累积分数。对于氨基酸序列，评分矩阵用于计算累积分数。字词命中在每个方向上的延伸在以下情况下停止：累积比对分数从其最大实现值下降数量X；累积分数由于一个或多个负分数残基比对的累积而变为零或更低；或到达任一序列的末端。BLAST算法参数W、T和X确定比对的灵敏度和速度。BLASTN程序(用于核苷酸序列)默认使用字长(W)11、期望值(E)10、M＝5、N＝-4以及两条链的比较。对于氨基酸序列，BLASTP程序默认使用字长(W)3、期望值(E)10以及BLOSUM62评分矩阵(参见Henikoff和Henikoff的Proc.Natl.Acad.Sci.USA第89卷：第10915页(1989年))．

[0039] 除了计算序列同一性百分比之外，BLAST算法还对两条序列之间的相似性进行统计分析(参见例如Karlin和Altschul的Proc.Nat'l.Acad.Sci.USA第90卷：第5873页至5787页(1993年))。BLAST算法提供的一种相似性量度是最小总和概率(P(N))，其提供了两条核

苷酸或氨基酸序列之间偶然发生匹配的概率的指示。例如，如果在测试核酸与参考核酸的比较中最小总和概率小于约0.1，更优选地小于约0.01，并且最优选地小于约0.001，则认为核酸类似于参考序列。

[0040] 两个核酸序列或多肽基本上相同的另一个指示是通过第一核酸编码的多肽与通过第二核酸编码的多肽免疫地交叉反应，如下所述。因此，多肽通常与第二多肽基本上相同，例如，其中两个肽仅通过保守置换而不同。两个核酸序列基本上相同的另一个指示是，两个分子在严格条件下彼此杂交。

[0041] 如本文所用，术语"分离的"意指生物组分(诸如核酸、肽、蛋白质或细胞)已经与组分天然存在的生物体的其他生物组分(即其他染色体和染色体外的DNA和RNA、蛋白质、细胞和组织)基本上分离、分开得到或从其中纯化出来。因此，已被"分离"的核酸、肽、蛋白质和细胞包括通过标准纯化方法和本文所述的纯化方法纯化的核酸、肽、蛋白质和细胞。"分离的"核酸、肽、蛋白质和细胞可为组合物的一部分，并且如果该组合物不是核酸、肽、蛋白质或细胞自身环境的一部分，则仍然是分离的。该术语还包括通过在宿主细胞中重组表达制备的核酸、肽和蛋白质以及化学合成的核酸。

[0042] 如本文所用，同义地称为"核酸分子"、"核苷酸"或"核酸"的术语"多核苷酸"是指任何多核糖核苷酸或多脱氧核糖核苷酸，其可为未修饰的RNA或DNA或者修饰的RNA或DNA。"多核苷酸"包括但不限于单链和双链的DNA、为单链区和双链区的混合物的DNA、单链和双链的RNA以及为单链区和双链区的混合物的RNA、包含可为单链或更典型地为双链或者为单链区和双链区的混合物的DNA和RNA的杂合分子。此外，"多核苷酸"是指包含RNA或DNA或RNA和DNA两者的三链区。术语多核苷酸还包括含有一个或多个修饰的碱基的DNA或RNA，以及具有出于稳定性或其它原因而被修饰的主链的DNA或RNA。"修饰的"碱基包括例如三苯甲基化的碱基和稀有碱基诸如肌苷。可以对DNA和RNA进行多种修饰；因此，"多核苷酸"包括通常天然存在的多核苷酸的化学修饰、酶修饰或代谢修饰形式，以及病毒和细胞特有的DNA和RNA的化学形式。"多核苷酸"也包括相对短的核酸链，通常被称为寡核苷酸。

[0043] 如本文所用，术语"载体"是复制子，其中可以可操作地插入另一核酸区段以引起该区段的复制或表达。

[0044] 如本文所用，术语"表达"是指基因产物的生物合成。该术语涵盖基因到RNA的转录。该术语还涵盖RNA到一个或多个多肽的翻译，并且还涵盖所有天然存在的转录后和翻译后修饰。表达的CAR可处于宿主细胞的细胞质内，处于胞外环境诸如细胞培养物的生长培养基中，或锚定至细胞膜。

[0045] 如本文所用，术语"可操作地连接"是指当核酸(例如，启动子和编码多肽的核酸)被置于结构或功能关系时核酸之间的连接。例如，如果核酸序列的一个片段和核酸序列的另一个片段相对于彼此定位在同一邻接核酸序列上并且具有结构或功能关系(诸如启动子或增强子相对于编码序列定位以便有利于编码序列的转录；核糖体结合位点相对于编码序列定位以便有利于翻译；或前序列或分泌前导序列相对于编码序列定位以便有利于前蛋白(例如，参与所编码的多肽的分泌的前蛋白)的表达)，则核酸序列的一个片段可以可操作地连接至核酸序列的另一个片段。在其他示例中，可操作地连接的核酸序列不是邻接的，而是以使得它们作为核酸或作为由核酸表达的蛋白质而彼此具有功能关系的方式定位。例如，增强子不必是邻接的。连接可通过在方便的限制位点处连接或通过使用合成的寡核苷酸衔

接子或接头来实现。

[0046] 如本文所用,术语"启动子"是指能够引发信使RNA中的基因序列转录的核酸序列,此类转录通过启动子上或附近的RNA聚合酶的结合而引发。

[0047] 如本文所用,术语"复制起点"或"复制的起点"是指质粒复制所必需的核酸序列。复制起点的示例包括但不限于pBR322复制起点、ColE1复制起点、pUC57复制起点、pMB1复制起点、pSC101复制起点和R6Kγ复制起点。复制起点可为高拷贝或低拷贝的。当存在于载体中时,高拷贝复制起点可导致每个细胞具有高拷贝数(例如150个至200个)的载体。当存在于载体中时,中等拷贝复制起点可导致每个细胞具有中等拷贝数(例如25个至50个)的载体。当存在于载体中时,低拷贝复制起点可导致每个细胞具有低拷贝数(例如1个至3个)的载体。

[0048] 如本文所用,术语"二聚体解离元件"是指有利于核酸序列(例如载体或质粒)的多聚体在体内转化成其中存在所述序列的单体的核酸序列。二聚体解离元件可包含核酸序列,该核酸序列包含位点特异性重组酶靶位点(例如LoxP靶位点、rfs靶位点、FRT靶位点、RP4 res靶位点、RK2 res靶位点和res靶位点)。二聚体解离元件可包含编码位点特异性重组酶(例如Cre重组酶、ResD重组酶、Flp重组酶、ParA重组酶、Sin重组酶、β重组酶、γδ重组酶、mpR重组酶和pSK41解离酶)的核酸序列。分离的载体/核酸的二聚体可通过作用于包含在二聚体解离元件内的靶DNA序列的酶来解离。该酶重组靶DNA序列。作为非限制性示例,由宿主细胞或包含二聚体解离元件的载体表达的酶XerC和XerD识别ColE1二聚体解离元件的cer靶位点,并且与若干附加辅因子一起作用以确保产生载体/核酸的单体。

[0049] 如本文所用,术语"肽"、"多肽"或"蛋白"可指由氨基酸组成的分子,并且可被本领域的技术人员识别为蛋白。本文使用氨基酸残基的常规单字母或三字母代码。术语"肽"、"多肽"和"蛋白质"在本文中可互换使用,指任何长度的氨基酸的聚合物。该聚合物可为直链或支链的,其可包含经修饰的氨基酸,并且其可夹杂有非氨基酸。该术语还涵盖已被天然修饰或通过干预修饰的氨基酸聚合物;天然修饰或干预修饰为例如二硫键形成、糖基化、脂化、乙酰化、磷酸化或任何其他操作或修饰,诸如与标记组分缀合。该定义还包括例如含有一种或多种氨基酸类似物(包括例如非天然氨基酸等)以及本领域已知的其他修饰的多肽。

[0050] 本文所述的肽序列根据通常的惯例书写,其中肽的N-末端区域在左侧,并且C-末端区域在右侧。虽然氨基酸的同分异构形式是已知的,但除非另外明确指明,它是所表示的氨基酸的L形式。

[0051] 多核苷酸、载体、宿主细胞及使用方法

[0052] 在一个总的方面,提供了使用核酸构建体作为选择性标记的方法。该方法包括(a)使在乳糖操纵子中包含缺失的宿主细胞与核酸构建体接触,其中核酸构建体包含分离的β-半乳糖苷酶表达盒,该分离的β-半乳糖苷酶表达盒包含编码可操作地连接至启动子的β-半乳糖苷酶的氨基末端片段的核酸序列;以及(b)使宿主细胞在其中核酸构建体在该宿主细胞中得以维持的条件下生长。

[0053] 在另一个总的方面,本发明涉及一种包含核酸序列的分离的β-半乳糖苷酶表达盒,该核酸序列编码可操作地连接至启动子的β-半乳糖苷酶的氨基末端片段。

[0054] 在某些实施方案中,β-半乳糖苷酶的氨基末端片段包含与SEQ ID NO:1具有至少75%同一性的氨基酸序列。在某些实施方案中,β-半乳糖苷酶的氨基末端片段包含与SEQ

ID NO:1具有至少75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%同一性的氨基酸序列。β-半乳糖苷酶的氨基末端片段可包含SEQ ID NO:1。

[0055] 在某些实施方案中,核酸序列还包含复制起点。复制起点可例如为高拷贝复制起点。在某些实施方案中,高拷贝复制起点为pUC57复制起点。在某些实施方案中,pUC57复制起点包含与SEQ ID NO:19具有至少90%、91%、92%、93%、94%、95%、96%、97%、98%或99%同一性的核酸序列。在某些实施方案中,pUC57复制起点包含SEQ ID NO:19的核酸序列。

[0056] 在某些实施方案中,分离的β-半乳糖苷酶表达盒还可包含二聚体解离元件。二聚体解离元件可例如包含具有位点特异性重组酶识别位点的核酸序列。位点特异性重组酶识别位点可例如选自LoxP位点、rfs位点、FRT位点、RP4 res位点、RK2 res位点和res位点。二聚体解离元件还可包含编码位点特异性重组酶的核酸序列。在某些实施方案中,宿主细胞包含编码位点特异性重组酶的核酸序列。位点特异性重组酶可例如选自Cre重组酶、ResD重组酶、Flp重组酶、ParA重组酶、Sin重组酶、β重组酶、γδ重组酶、tnpR重组酶和pSK41解离酶。

[0057] 二聚体解离元件可例如为ColE1二聚体解离元件。ColE1二聚体解离元件可包含与SEQ ID NO:20具有至少90%、91%、92%、93%、94%、95%、96%、97%、98%或99%同一性的核酸序列。在某些实施方案中,ColE1二聚体解离元件包含SEQ ID NO:20的核酸序列。

[0058] 在某些实施方案中,分离的载体包含本发明的分离的β-半乳糖苷酶表达盒。可使用根据本公开的本领域技术人员已知的任何载体,诸如质粒、粘粒、人工染色体(例如,细菌人工染色体(BAC)、酵母人工染色体(YAC)和/或衍生自P1的人工染色体(PAC)、转座子、噬菌体载体或病毒载体。在一些实施方案中,载体是重组表达载体,诸如质粒。载体可包括建立表达载体的常规功能的任何元件,例如启动子、核糖体结合元件、终止子、增强子、选择标记和复制起点。启动子可为组成型、诱导型或阻抑型启动子。能够向细胞递送核酸的多种表达载体在本领域中是已知的,并且在本文可用于产生β-半乳糖苷酶肽的氨基末端片段。常规克隆技术或人工基因合成可用于根据本发明的实施方案生成重组表达载体。

[0059] 在某些方面,分离的载体的大小小于约1.5千碱基。分离的载体的长度可例如为约700个碱基对、约800个碱基对、约900个碱基对、约1000个碱基对(约1千碱基)、约1100个碱基对(约1.1千碱基)、约1200个碱基对(约1.2千碱基)、约1300个碱基对(约1.3千碱基)、约1400个碱基对(约1.4千碱基)或约1500个碱基对(约1.5千碱基)。在某些实施方案中,分离的载体的大小小于约1千碱基。在某些实施方案中,分离的载体的大小小于约900个碱基对。在某些实施方案中,分离的载体的大小小于约800个碱基对。

[0060] 在某些实施方案中,分离的载体包含与选自SEQ ID NO:9-13、17和18的核酸具有至少90%、91%、92%、93%、94%、95%、96%、97%、98%或99%同一性的核酸序列。在某些实施方案中,分离的载体包含选自SEQ ID NO:9-13、17和18的核酸序列。

[0061] 还提供了生成本发明的分离的载体的方法。该方法包括(a)使宿主细胞与分离的载体接触;(b)使宿主细胞在产生该载体的条件下生长;以及(c)从宿主细胞分离载体。

[0062] 在某些实施方案中,宿主细胞在基本培养基中生长。基本培养基可包含乳糖作为唯一碳源。在某些实施方案中,基本培养基包含约1%至约4%重量/体积(w/v)乳糖。在某些

实施方案中，基本培养基包含约1％至约4％w/v、约1％至约3％w/v、约1％至约2％w/v、约1.5％至约4％w/v、约1.5％至约3％w/v、约1.5％至约2％w/v、约2％至约4％w/v、约2％至约3％w/v、约2.5％至约4％w/v、约2.5％至约35％w/v或约3％至约4％w/v乳糖。在某些实施方案中，基本培养基包含约2％w/v乳糖。

[0063] 在某些实施方案中，本发明涉及包含本发明的分离的载体的宿主细胞。鉴于本公开，本领域的技术人员已知的任何宿主细胞可用于比较本发明的分离的载体。合适的宿主细胞包括具有LacZΔM15缺失但乳糖生物合成途径的其余部分完整的细胞。在噬菌体Φ80整合的上下文中含有该突变(即Φ80lacZΔM15标记)的菌株在完整乳糖操纵子的上下文中含有该突变，并因此是合适的宿主。当用LacZ-α互补质粒转化时产生显著水平的β-半乳糖苷酶的LacZ基因的在氨基末端(N-末端)区域中具有不同缺失的其他宿主也可为合适的宿主。本发明的合适的宿主细胞可包括大肠杆菌宿主细胞或酵母宿主细胞。

[0064] 还提供了试剂盒，该试剂盒包含(a)本发明的分离的β-半乳糖苷酶表达盒；以及(b)在乳糖操纵子中包含缺失的宿主细胞。在某些实施方案中，载体包含分离的β-半乳糖苷酶表达盒。在某些实施方案中，宿主细胞包含LacZΔM15缺失。在某些实施方案中，宿主细胞可选自大肠杆菌宿主细胞或酵母宿主细胞。

[0065] 在某些实施方案中，试剂盒还包含基本培养基，该基本培养基包含乳糖作为唯一碳源。在某些实施方案中，基本培养基包含约1％至约4％重量/体积(w/v)乳糖。在某些实施方案中，基本培养基包含约1％至约4％w/v、约1％至约3％w/v、约1％至约2％w/v、约1.5％至约4％w/v、约1.5％至约3％w/v、约1.5％至约2％w/v、约2％至约4％w/v、约2％至约3％w/v、约2.5％至约4％w/v、约2.5％至约35％w/v或约3％至约4％w/v乳糖。在某些实施方案中，基本培养基包含约2％w/v乳糖。

[0066] 实施方案

[0067] 本发明提供了以下非限制性实施方案。

[0068] 实施方案1是一种使用核酸构建体作为选择性标记的方法，该方法包括：

[0069] a.使在乳糖操纵子中包含缺失的宿主细胞与核酸构建体接触，其中核酸构建体包含分离的β-半乳糖苷酶表达盒，该分离的β-半乳糖苷酶表达盒包含编码可操作地连接至启动子的β-半乳糖苷酶的氨基末端片段的核酸序列；以及

[0070] b.使宿主细胞在其中仅含有核酸构建体的宿主细胞在该宿主细胞中得以维持的条件下生长。

[0071] 实施方案2是根据实施方案1所述的方法，其中β-半乳糖苷酶的氨基末端片段包含与SEQ ID NO:1具有至少75％同一性的氨基酸序列。

[0072] 实施方案3是根据实施方案1或2所述的方法，其中β-半乳糖苷酶的氨基末端片段包含SEQ ID NO:1的氨基酸序列。

[0073] 实施方案4是根据实施方案1-3中任一项所述的组合物，其中核酸序列还包含复制起点。

[0074] 实施方案5是根据实施方案4所述的方法，其中复制起点为高拷贝复制起点。

[0075] 实施方案6是根据实施方案5所述的方法，其中高拷贝复制起点为pUC57复制起点。

[0076] 实施方案7是根据实施方案6所述的方法，其中pUC57复制起点包含SEQ ID NO:19的核酸序列。

[0077] 实施方案8是根据实施方案1-7中任一项所述的方法，其中分离的β-半乳糖苷酶表达盒还包含二聚体解离元件。

[0078] 实施方案9是根据实施方案8所述的方法，其中二聚体解离元件包含具有位点特异性重组酶识别位点的核酸序列。

[0079] 实施方案10为根据实施方案8或9所述的方法，其中二聚体解离元件还包含编码位点特异性重组酶的核酸序列。

[0080] 实施方案11为根据实施方案8或9所述的方法，其中宿主细胞包含编码位点特异性重组酶的核酸序列。

[0081] 实施方案12是根据实施方案8-11中任一项所述的方法，其中二聚体解离元件为ColE1二聚体解离元件。

[0082] 实施方案13是根据实施方案12所述的方法，其中ColE1二聚体解离元件包含SEQ ID NO:20的核酸序列。

[0083] 实施方案14是根据实施方案1-13中任一项所述的方法，其中宿主细胞包含LacZΔM15缺失。

[0084] 实施方案15是根据实施方案1-14中任一项所述的方法，其中分离的载体包含分离的β-半乳糖苷酶表达盒。

[0085] 实施方案16是根据实施方案15所述的方法，其中分离的载体的大小小于约1.5千碱基。

[0086] 实施方案17是根据实施方案15或16所述的方法，其中分离的载体包含选自SEQ ID NO:9-13、17和18的核酸序列。

[0087] 实施方案18是一种生成根据实施方案15-17中任一项所述的分离的载体的方法，其中该方法包括：

[0088] a.使宿主细胞与分离的载体接触；

[0089] b.使宿主细胞在产生该载体的条件下生长；

[0090] c.从宿主细胞分离载体。

[0091] 实施方案19是根据实施方案18所述的方法，其中宿主细胞在基本培养基中生长。

[0092] 实施方案20是根据实施方案19所述的方法，其中基本培养基包含乳糖作为唯一碳源。

[0093] 实施方案21是根据实施方案20所述的方法，其中基本培养基包含约1％至约4％重量/体积(w/v)乳糖。

[0094] 实施方案22是根据实施方案21所述的方法，其中基本培养基包含约2％w/v乳糖。

[0095] 实施方案23是一种分离的β-半乳糖苷酶表达盒，该分离的β-半乳糖苷酶表达盒包含编码可操作地连接至启动子的β-半乳糖苷酶的氨基末端片段的核酸序列。

[0096] 实施方案24是根据实施方案23所述的分离的β-半乳糖苷酶表达盒，其中β-半乳糖苷酶的氨基末端片段包含与SEQ ID NO:1具有至少75％同一性的氨基酸序列。

[0097] 实施方案25是根据实施方案23或24所述的分离的β-半乳糖苷酶表达盒，其中β-半乳糖苷酶的氨基末端片段包含SEQ ID NO:1的氨基酸序列。

[0098] 实施方案26是根据实施方案23-25中任一项所述的分离的β-半乳糖苷酶表达盒，其中核酸序列还包含复制起点。

[0099] 实施方案27是根据实施方案26所述的分离的β-半乳糖苷酶表达盒,其中复制起点为高拷贝复制起点。

[0100] 实施方案28是根据实施方案27所述的分离的β-半乳糖苷酶表达盒,其中高拷贝复制起点为pUC57复制起点。

[0101] 实施方案29是根据实施方案28所述的分离的β-半乳糖苷酶表达盒,其中pUC57复制起点包含SEQ ID NO:19的核酸序列。

[0102] 实施方案30是根据实施方案23-29中任一项所述的分离的β-半乳糖苷酶表达盒,其中分离的β-半乳糖苷酶表达盒还包含二聚体解离元件。

[0103] 实施方案31是根据实施方案30所述的分离的β-半乳糖苷酶表达盒,其中二聚体解离元件包含具有位点特异性重组酶识别位点的核酸序列。

[0104] 实施方案32为根据实施方案30或31所述的β-半乳糖苷酶表达盒,其中二聚体解离元件还包含编码位点特异性重组酶的核酸序列。

[0105] 实施方案33是根据实施方案30-32中任一项所述的分离的β-半乳糖苷酶表达盒,其中二聚体解离元件为ColE1二聚体解离元件。

[0106] 实施方案34是根据实施方案33所述的分离的β-半乳糖苷酶表达盒,其中ColE1二聚体解离元件包含SEQ ID NO:20的核酸序列。

[0107] 实施方案35是一种分离的载体,该分离的载体包含根据实施方案23-34中任一项所述的分离的β-半乳糖苷酶表达盒。

[0108] 实施方案36是根据实施方案35所述的分离的载体,其中分离的载体的大小小于约1.5千碱基。

[0109] 实施方案37是根据实施方案35或36所述的分离的载体,其中分离的载体包含选自SEQ ID NO:9-13、17和18的核酸序列。

[0110] 实施方案38是一种试剂盒,该试剂盒包含:

[0111] a.根据实施方案23-37中任一项所述的分离的β-半乳糖苷酶表达盒;以及

[0112] b.在乳糖操纵子中包含缺失的宿主细胞。

[0113] 实施方案39是根据实施方案38所述的试剂盒,该试剂盒还包含基本培养基,该基本培养基包含乳糖作为唯一碳源。

[0114] 实施方案40是根据实施方案38或39所述的试剂盒,其中载体包含分离的β-半乳糖苷酶表达盒。

[0115] 实施方案41是根据实施方案38-40中任一项所述的试剂盒,其中宿主细胞包含LacZΔM15缺失。

[0116] 实施方案42是根据实施方案41所述的试剂盒,其中宿主细胞选自大肠杆菌宿主细胞和酵母宿主细胞。

[0117] 实施例

[0118] 实施例1:经由β-半乳糖苷酶的α-互补的质粒选择而不是TOP10细胞中的抗生素选择

[0119] 材料

[0120] 细胞:One Shot Top10感受态细胞(Thermo-Fisher公司(Waltham,MA);目录号C404003)。NEB 5-α(New England Biolabs公司(Ipswich,MA);目录号C2987)。GT115

(InvivoGen公司(San Diego,CA);目录号GT115-21)。NEB Stable(New England Biolabs公司;目录号C3040H)。Stellar(Takara Bio USA公司(Mountain View,CA);目录号636766)。DH10B(Thermo-Fisher公司;目录号18297010)。Stbl3(Thermo-Fisher公司;目录号C737303)。Xli-blue(Agilent公司(Santa Clara,CA);目录号200236)。

[0121] 质粒:pUC19(Thermo-Fisher Scientific公司;目录号SD0061)。pBluescript II.KS(-)(Agilent公司(Santa Clara,CA);目录号212208)。克隆P215(SEQ ID NO:9)和P216(SEQ ID NO:10)。GWIZ-Luciferase(Genlantis Corporation公司(San Diego,CA);P030200)。P219(SEQ ID NO:13;图5)。P469-2(SEQ ID NO:17;图6)。

[0122] 培养基:M9+乳糖培养基(Teknova公司(Hollister CA);目录号M1348-04(平板)):0.3%KH_2PO_4、0.6%Na_2HPO_4、0.5%(85mM)NaCl、0.1%NH_4Cl、2mM $MgSO_4$、50mg/L L-亮氨酸、50mg/L异亮氨酸、1mM硫胺素、2%乳糖和1.5%琼脂。

[0123] M9+葡萄糖培养基(Teknova公司(Hollister CA);目录号M1346-04(平板)):0.3%KH_2PO_4、0.6%Na_2HPO_4、0.5%(85mM)NaCl、0.1%NH_4Cl、2mM $MgSO_4$、50mg/L L-亮氨酸、50mg/L异亮氨酸、1mM硫胺素、1%葡萄糖和1.5%琼脂。LB-羧苄青霉素(100)平板(Teknova公司(Hollister CA);目录号L1010)。LB平板(Teknova公司(Hollister CA);L1100)。LB+60μg/mL X-Gal,0.1mM IPTG(Teknova公司(Hollister CA);L1920)。SOC培养基(Thermo-Fisher公司;15544034)。LB肉汤(Thermo-Fisher公司;10855021);D-PBS,pH 7.1,不合Mg^{2+}且不含Ca^{2+}(ThermoFisher公司,14200-075)。

[0124] 结果

[0125] 不含抗生素选择标记的质粒对于基因疗法应用和治疗剂产品的细胞系开发是期望的。据报道,1kb或更小的质粒主链在体内递送给动物时可用于避免基因沉默。这些实验的目的是探索一种用于开发用于在大肠杆菌中选择含质粒细胞的小代谢选择标记的新策略。

[0126] 假设表达β-半乳糖苷酶的α肽的质粒可与TOP10细胞中的LacZΔ15等位基因互补,从而完成乳糖操纵子并允许细胞在以乳糖作为唯一碳源的基本培养基上生长。

[0127] 质粒pUC19和pBluescript II两者均表达β-半乳糖苷酶α肽融合蛋白。测试这些质粒是否能够与Top10宿主菌株中的lac突变互补并允许其在基本培养基上生长。

[0128] 为了测试pUC19和/或pBluescript II是否能够与TOP10细胞中的LacZΔ15突变互补,使用以下程序将这些质粒转化到细胞中。

[0129] 如下在无菌微量离心管中制备两种转化体混合物:1)1μL(100pg)pBluescript II质粒+50μL One Shot TOP10细胞;2)1μL(10pg)pUC19质粒+50μL One Shot TOP10细胞。将转化体混合物在冰上温育30分钟,然后在42℃下热休克30秒。热休克后,将转化体混合物在冰上温育1分钟。向转化体混合物中加入450μL SOC培养基,并将细胞在37℃并振荡下温育1小时。将含有细胞的转化体混合物离心,并将细胞重悬于500μL无菌D-PBS缓冲液中。将细胞离心并重悬两次以上。对于每个样品,在D-PBS中制备细胞的两个1:10系列稀释液。将200μL的每种稀释液涂覆到M9+乳糖平板上。另外将200μL的前两种稀释液涂覆到LB-羧苄青霉素(100)平板上。将平板在37℃下温育过夜。

[0130] 过夜温育后,有许多来自两种转化体的菌落接种到LB-羧苄青霉素(100)平板上;将这些平板保存在4℃下。未见来自任一种转化体的菌落接种到M9+乳糖平板上;将这些平

板在37℃下再温育24小时。在M9-乳糖平板上未见菌落。将细胞在30℃下再培养48小时。在这些平板上未见菌落,甚至在长时间温育后也未见菌落。

[0131] 表达LacZ-α融合肽的克隆载体都不能与TOP10宿主菌株中的Lac突变互补,以允许在含有乳糖作为唯一碳源的基本培养基中生长。

[0132] pUC19和pBluescript II克隆载体对LacZ-α肽融合蛋白的表达可能不足以与所测试的宿主菌株中的lac突变充分地互补。两种载体均产生通过多克隆区域转录的融合蛋白,并且此类融合蛋白对于与LacZΔ15突变互补来说可能是次优的。

[0133] 实施例2:在大肠杆菌中用作代谢选择标记的表达LacZ的质粒

[0134] 设计了两个具有中等启动子和强启动子的LacZ-α表达盒(分别为LacZYA和OmpF)。OmpF启动子序列基于Stavropoulos等人所使用的OmpF启动子(Stavropoulos和Strathdee,Genomics,第72卷第1期,第99-104页,2001年)。LacZYA启动子来源于pBluescript中的序列以及由lac阻遏蛋白结合的乳糖操纵子序列。

[0135] 对于LacZα肽的开放阅读框(ORF),Reddy(Reddy,Biotechniques,第37卷第6期,第948-952页,2004年)报道称质粒pUC19产生的β-半乳糖苷酶活性比pBluescript高约10倍。这些质粒具有驱动lacZα肽的同一启动子元件。然而,pBluescript具有比pUC19长得多的多接头,并且pUC19编码非lacZ C末端残基。这些差异中哪一种差异导致更高的pUC19β-半乳糖苷酶活性是未知的。Nishiyama等人发现60个氨基酸的N-末端α肽在其测定中具有最大的β-半乳糖苷酶活性(Nishiyama等人,Protein Sci.,第24卷第5期,第599-603页,2015年)。使用来自菌株MG1655的在残基60处截短的以下野生型LacZα区域:MTMITDSLAVVLQRRDWENPGVTQLNRLAAHPPFASWRNSEEARTDRPSQQLRSLNGEWR(SEQ ID NO:1)。

[0136] 终止子序列来源于Orosz等人所述的rmBT2终止子(Orosz等人,Eur.J.Biochem.,第201卷第3期,第653-659页,1991年)。

[0137] 通过基因合成在GeneWiz(South Plainfield,NJ)构建P215(SEQ ID NO:9)(图1)和P216(SEQ ID NO:10)(图2)质粒。质粒含有氨苄青霉素抗性盒和4.9kb转基因。

[0138] 结果

[0139] 不含抗生素选择标记的质粒对于基因疗法应用和治疗剂产品的细胞系开发是期望的。据报道,1kb或更小的质粒主链在体内递送给动物时可用于避免基因沉默。这些实验的目的是探索一种用于开发用于在大肠杆菌中选择含质粒细胞的小代谢选择标记的新策略。

[0140] 假设表达β-半乳糖苷酶的α肽的质粒可与Top10细胞中的LacZΔ15等位基因互补,从而完成乳糖操纵子并允许细胞在以乳糖作为唯一碳源的基本培养基上生长。

[0141] 在实施例1中,测试了表达lacZα融合肽的pUC19和pBluescript载体是否可与TOP10细胞互补并允许它们在含有乳糖的基本培养基上生长。这些实验不成功。

[0142] 基于以下假设:由这些载体编码的lacZα融合蛋白在与LacZΔ15突变互补方面是次优的,并且不以足够高的水平表达以实现在含乳糖基本培养基上生长,合成具有新型lacZα表达盒的载体。测试这些载体与LacZΔ15突变互补的能力。将十纳克(ng)的质粒P215和P216以及pBluescript II转化到50μL OneShot Top10细胞中。将细胞与DNA在冰上温育20分钟,在42℃下热休克30秒,并放回冰上温育1分钟。将450μL SOC加入细胞中,将细胞在37℃并振荡下温育1小时。取出250μL细胞并将剩余的细胞放回培养箱中。将所提取的细胞

用500μL的D-PBS洗涤两次，并在最后一次洗涤后重悬于200μL的D-PBS中。将50μL细胞接种在LB-羧苄青霉素(100)、M9+葡萄糖和M9+乳糖平板上，并将平板在37℃下温育。在热休克后4.5小时后，如上所述洗涤来自培养箱的剩余细胞，并将其接种到M9+葡萄糖和M9+乳糖平板上。将平板在37℃下温育过夜。

[0143] 接种在M9+葡萄糖上的转化体形成细胞菌苔，这表明Top10宿主细胞可以在这些平板上生长。接种在LB-羧苄青霉素(100)上的转化体也产生多个菌落。将LB-羧苄青霉素平板保存在4℃下。M9+乳糖平板保持在37℃下以再温育24小时。

[0144] 当接种在M9+乳糖平板上时，允许恢复一小时的转化体或允许恢复四小时的转化体均产生大量的菌落。在pBluescript II转化体上无菌落确认了来自实施例1的结果，表明pBluescript II无法通过LacZΔ15突变的互补产生足够的β-半乳糖苷酶以允许在乳糖基本培养基上生长。将平板保存在4℃下。

[0145] 在不存在抗生素选择的情况下，将天然质粒诸如ColE1有效维持在大肠杆菌宿主中，而pUC系列载体可在不存在选择的情况下以高速率从细胞中丢失(Summers，Molecular Microbiology，第29卷，第1137-1145页，1998年)。然而，考虑到P215和P216转化的细胞在基本培养基上相对于在富LB培养基上生长速率慢得多，如果质粒丢失的频率不太高，则质粒DNA纯化在不存在选择的情况下使细胞培养物在LB中生长更快且成本更低。包含β-半乳糖苷酶α-互补质粒的细胞容易与在LB-IPTG-XGAL平板上生长的不含质粒的细胞区分开，因为β-半乳糖苷酶水解XGAL(5-溴-4-氯-3-吲哚基-β-D-吡喃半乳糖苷)指示剂，使细胞变成蓝色。当这些细胞在不存在抗生素的情况下在LB培养基中生长时，该测定用于研究质粒丢失的频率。

[0146] 通过在LB-IPTG-XGAL平板上划线分离细胞来获得纯的细胞群，并且包含质粒的菌落变成蓝色。如所预期的，在平板上划线分离的大部分菌落是蓝色的。

[0147] 在获得纯的细胞群后，使细胞的连续培养物生长。选取单个蓝色菌落并使其在15mL管中的2mL LB培养基中生长。将培养物在37℃并振荡下温育过夜。

[0148] 将来自培养物的细胞划线分离到LB-IPTG-XGAL平板上，并将平板在37℃下温育过夜。再次划线的平板上的菌落是蓝色的。将单个菌落接种在250mL烧瓶中的50mL LB中，并在37℃并振荡下温育过夜。

[0149] 将50μL的过夜培养物的10^{-4}稀释液接种到LB-IPTG-XGAL平板上。将平板在37℃下温育过夜。取50mL培养物中的1μL稀释成50mL LB中的新培养物溶液(50,000倍稀释)。使培养物在37℃下生长过夜。

[0150] 温育过夜后，观察到平板上的所有菌落均为蓝色的。将前一天晚上50mL培养物的10^{-4}稀释液中的50μL接种在LB-IPTG-XGAL平板上。取前一天晚上50mL培养物中的1μL稀释成50mL LB中的新培养物溶液(50,000倍稀释)。使培养物在37℃下生长过夜。

[0151] 温育过夜后，在含有50μL的10^{-4}稀释液的平板上观察到约1000个菌落。P215转化的所有菌落均为蓝色的，在P216转化平板上仅观察到3个白色菌落。

[0152] 结果表明，质粒P215和P216即使在不存在选择的情况下也是稳定的。对于P215和P216，这些质粒分别为7.2kb和7.3kb。从单个菌落到50mL，然后以1:50,000稀释并生长汇合两次，表明细胞可在不存在选择的情况下生长至1.25×10^{8}升的体积，同时仍将质粒保留在大多数细胞中。当热休克后使细胞在SOC培养基中恢复一小时时相对于当恢复四小时时，转

化效率是相似的。

[0153] 所构建的α互补质粒与Top10细胞中的LacZΔ15突变互补，从而实现在以乳糖作为唯一碳源的基本培养基上生长。还发现这些质粒在不存在选择性压力的情况下在LB液体培养物中是稳定的。

[0154] 实施例3:减小β-半乳糖苷酶α互补质粒的大小

[0155] 在先前的实验中，已证明，来自P215和P216质粒的β-半乳糖苷酶α肽的表达可用作质粒上的选择标记，从而替代抗生素抗性基因。接下来，为了限定最小可能的复制子，探索限定质粒的哪些区域对于质粒在大肠杆菌中的选择和复制是必需的。

[0156] 结果

[0157] 使用标准克隆技术，从质粒P215中移除mCherry和嘌呤霉素抗性基因，以形成质粒P217(SEQ ID NO:11)(图3)。

[0158] 使用标准克隆技术从质粒P217中移除氨苄青霉素抗性基因。将所连接的DNA转化到50μL的TOP10细胞中，在冰上温育20分钟，热休克30秒，并在冰上再温育3分钟。温育后，将450μL的SOC培养基加入细胞中，并将细胞在37℃并振荡下温育1小时。将细胞制粒并用1mL的d-PBS洗涤3次。将细胞接种到M9-乳糖平板上并在37℃下温育两天。选取来自转化体的菌落并划线分离到LB-IPTG-XGAL平板上。每个克隆的所得菌落均为蓝色的。选取单个克隆(克隆P218(SEQ ID NO:12;图4))，并且DNA测序确认已经形成期望的缺失。

[0159] 为了进一步减小β-半乳糖苷酶选择盒的大小，rmBT2转录终止子(SEQ ID NO:7)有缺失。除了该序列可能不需要维持转录物稳定性之外，据报道来自pUC57/pMB1起点上游启动子的通读转录可通过增加通过该起点的复制引物区的转录来增加拷贝数(Panayotatos，Nucleic Acid Res.，第12卷第6期:第2641-2648页，1984年;Oka等人，Mol Gen Genet.，第172卷第2期，第151-159页，1979年)。

[0160] 使用标准克隆技术，获得缺失构建体P219的克隆(SEQ ID NO:13;图5)。通过DNA测序确认缺失。

[0161] 通过该工作阐明的最小β-半乳糖苷酶表达盒/复制起点盒(SEQ ID NO:18)为938bp。其实现了小于1kb的目标，以便避免与较大质粒主链相关联的哺乳动物细胞中的DNA沉默(Lu等人，Mol.Ther.，第20卷第11期，第2111-2119页，2012年)。

[0162] 实施例4:形成具有萤火虫荧光素酶表达盒的β-半乳糖苷酶α互补载体

[0163] 在上文提供的实施例中，构建了使用β-半乳糖苷酶突变的α互补作为选择标记而不是抗生素抗性基因的质粒。为了确定当质粒大小增加时DNA复制是否仍然有效，采用标准克隆技术，使用上文定义的最小β-半乳糖苷酶表达盒/复制起点序列(SEQ ID NO:18)替代现有质粒的抗生素选择标记和复制起点。

[0164] 使用标准克隆技术将来自GWIZ-荧光素酶质粒(SEQ ID NO:16)的CMV启动子-荧光素酶-polyA表达盒克隆到P219中。转化到One Shot TOP10细胞中，接种到M9+乳糖平板上，并在37℃下温育2天，得到大菌落。将菌落再次划线分离到LB-IPTG-XGAL平板上并在37℃下温育过夜。

[0165] 使用引物CNFOR(SEQ ID NO:14)和P455R2(SEQ ID NO:15)，针对插入序列筛选转化反应的蓝色菌落。选取两个PCR阳性菌落并用于接种在37℃下生长的6mL LB培养物。从培养物中分离DNA，并且通过用分光光度计测量其OD_{260}来估计DNA收率(表1)。

[0166] 表1:选定克隆的DNA收率

[0167]

样本名称	A260 浓度 (ng/ul)
P469-1	132.69
P469-2	506.91

[0168] 对于克隆P469-2,将500mL烧瓶中的200mL LB用单个蓝色菌落接种,并在37℃下在振荡培养箱中生长18小时。使用Qiagen HiSpeed MaxiPrep试剂盒从该培养物中纯化DNA,并回收440μg的DNA。

[0169] 在GeneWiz测序确认质粒P469-2(SEQ ID NO:17)。

[0170] 在该实施例中,卡那霉素抗性基因和GWIZ-荧光素酶的复制起点被上文定义的最小β-半乳糖苷酶/复制起点成功替代。当该克隆在LB培养基中在不存在选择性压力的情况下生长时,实现了可接受的质粒收率。

[0171] 实施例5:在多种大肠杆菌菌株中测试β-半乳糖苷酶α互补载体功能

[0172] 为了鉴定其中β-半乳糖苷酶α肽可用作选择性标记而不是抗生素抗性基因的另外的大肠杆菌菌株,通过DNA转染到8种不同的菌株中来测试上文所构建的质粒中的一个质粒。

[0173] 表2:细菌菌株

[0174]

菌株	供应商	基因型
Top10	Thermo-Fisher 公司	F- *mcrA* Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lacZ*ΔM15 Δ *lac*X74 *rec*A1 *ara*D139 Δ(*araleu*)7697 *gal*U *gal*K *rps*L (StrR) *end*A1 *nup*G
NEB 5-α	New England Biolabs 公司	*fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*
GT115	InVivogen 公司	F- mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 Δ lacX74 recA1 rspL (StrA) endA1 Δdcm uidA(Δ MluI)::pir-116 ΔsbcC-sbcD
NEB Stable	New England Biolabs 公司	F' *proA+B+ lacI*^{q} *Δ(lacZ)M15 zzf::Tn10* (Tet^{R}) *Δ (ara-leu) 7697 araD139 fhuA ΔlacX74 galK16 galE15 e14- Φ80dlacZΔM15 recA1 relA1 endA1 nupG rpsL* (Str^{R}) *rph spoT1 Δ(mrr-hsdRMS-mcrBC)*
Stellar	Takara Bio USA 公司	F-, endA1, supE44, thi-1, recA1, relA1, gyrA96, phoA, Φ80d lacZΔ M15, Δ(lacZYA-argF) U169, Δ(mrr-hsdRMS-mcrBC), ΔmcrA, λ-
DH10B	Thermo-Fisher 公司	F- mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 recA1 endA1 araD139 Δ (ara, leu)7697 galU galK λ- rpsL nupG /pMON14272 / pMON7124
Stbl3	Thermo-Fisher 公司	F^{-}*mcr*B *mrrhsd*S20(r$_{B}^{-}$, m$_{B}^{-}$) *rec*A13 *sup*E44 *ara*-14 *gal*K2 *lac*Y1 *pro*A2 *rps*L20(Str^{R}) *xyl*-5 λ^{-}*leumtl*-1
XL1-blue	Agilent 公司	*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F′ *proAB lacI*^{q}Z*ΔM15* Tn*10* (Tetr)].

[0175] 结果

[0176] 将50μL表2中的大肠杆菌菌株与1ng的质粒P469-2在无菌微量离心管中的冰上温育30分钟。将细胞在42℃下热休克30秒，并在冰上温育1分钟。将450μL SOC培养基加入除NEB-Stable细胞之外的所有细胞中。将450μL的NEB-Stable扩增培养基(由制造商提供)加入转化的NEB-Stable细胞中。将细胞在37℃并振荡下温育1小时。将细胞制粒并用1mL的D-PBS洗涤3次。将细胞接种到M9-乳糖平板上并在37℃下温育三天。

[0177] 可以预知，在平板上未检测到来自Stb13转化的细胞的菌落，该菌落被包括作为阴性对照。菌株中的五种菌株(Top10、GT115、NEB-Stable、Stellar和DH10B)具有正常大小的菌落。两个菌株(NEB-a和XL1-Blue)具有小菌落。这是期望的，因为与NEB-α(DH5α)和XL1-Blue相似的菌株在purB基因中包含导致在基本培养基上生长缓慢的突变(Jung等人，Appl Environ.Micro.，第76卷，第6307-6309页，2010年)。

[0178] 将XL1-blue和NEB-α平板在37℃下再温育一天。

[0179] 通过将来自M9-乳糖平板的菌落划线分离到LB-IPTG-XGAL平板上并在37℃下温育来获得纯菌落。将蓝色菌落(含有质粒的细胞)再次划线分离到LB-IPTG-XGAL平板上并在37℃下温育，这得到大部分蓝色细胞。

[0180] 包含Φ80dlacZΔM15标记的全部所测试菌株均可通过β-半乳糖苷酶α肽表达质粒P469-2来转化，并在以乳糖作为唯一碳源的M9基本培养基上进行选择。在F附加体上包含标记$lacI^{q}ZΔM15$的菌株XL1-blue的质粒P469-2转染子在M9-乳糖平板上也具有选择性。因此，已证明七种可商购获得的大肠杆菌菌株与β-半乳糖苷酶选择性标记相容。

[0181] 本领域的技术人员应当理解，在不脱离本发明的广义的发明构思的情况下，可对上述实施方案作出修改。因此，应当理解，本发明不局限于所公开的特定实施方案，但本发明旨在涵盖在本发明实质和范围内的修改，如本说明书所定义。

[0182] 序列表

<110> Janssen Biotech, Inc.

<120> 作为非抗生素选择标记的β-半乳糖苷酶α肽及其用途

<130> 688097-553US

<160> 20

<170> PatentIn 版本 3.5

<210> 1
<211> 60
<212> PRT
<213> 人工序列

<220>
<223> 截短的 LazCα 肽

<400> 1

Met Thr Met Ile Thr Asp Ser Leu Ala Val Val Leu Gln Arg Arg Asp
1 5 10 15

Trp Glu Asn Pro Gly Val Thr Gln Leu Asn Arg Leu Ala Ala His Pro
20 25 30

[0183]

Pro Phe Ala Ser Trp Arg Asn Ser Glu Glu Ala Arg Thr Asp Arg Pro
35 40 45

Ser Gln Gln Leu Arg Ser Leu Asn Gly Glu Trp Arg
50 55 60

<210> 2
<211> 419
<212> DNA
<213> 人工序列

<220>
<223> LacZα 盒 1

<400> 2
agcgggcagt gagcgcaacg caattaatgt gagttagctc actcattagg caccccaggc 60

tttacacttt atgcttccgg ctcgtatgtt gtgtggaatt gtgagcggat aacaatttca 120

cacaggaaac agctatgacc atgattacgg attcactggc cgtcgtttta caacgtcgtg 180

actgggaaaa ccctggcgtt acccaactta atcgccttgc agcacatccc cctttcgcca 240

gctggcgtaa tagcgaagag gcccgcaccg atcgcccttc ccaacagttg cgcagcctga 300

atggcgaatg gcgctgaggc ccggagggtg gcgggcagga cgcccgccat aaactgccag 360

gcatcaaatt aagcagaagg ccatcctgac ggatggcctt tttgcgtttc tacaaactc 419

<210> 3
<211> 540
<212> DNA
<213> 人工序列

<220>
<223> LacZα盒2

<400> 3
cacgtctcta tggaaatatg acggtgttca caaagttcct taaattttac ttttggttac 60

atatttttc tttttgaaac caaatcttta tctttgtagc actttcacgg tagcgaaacg 120

ttagtttgaa tggaaagatg cctgcagaca cataaagaca ccaaactctc atcaatagtt 180

ccgtaaattt ttattgacag aacttattga cggcagtggc aggtgtcata aaaaaaacca 240

tgagggtaat aaataatgac catgattacg gattcactgg ccgtcgtttt acaacgtcgt 300

gactgggaaa accctggcgt tacccaactt aatcgccttg cagcacatcc ccctttcgcc 360

[0184] agctggcgta atagcgaaga ggcccgcacc gatcgccctt cccaacagtt gcgcagcctg 420

aatggcgaat ggcgctgagg cccggagggt ggcgggcagg acgcccgcca taaactgcca 480

ggcatcaaat taagcagaag gccatcctga cggatggcct ttttgcgttt ctacaaactc 540

<210> 4
<211> 96
<212> DNA
<213> 人工序列

<220>
<223> LacZYA启动子

<400> 4
agcgggcagt gagcgcaacg caattaatgt gagttagctc actcattagg caccccaggc 60

tttacacttt atgcttccgg ctcgtatgtt gtgtgg 96

<210> 5
<211> 38
<212> DNA
<213> 人工序列

<220>
<223> 乳糖操纵子

<400> 5
aattgtgagc ggataacaat ttcacacagg aaacagct 38

<210> 6
<211> 183
<212> DNA
<213> 人工序列

<220>
<223> 截短的 LacZα 肽核苷酸序列

<400> 6
atgaccatga ttacggattc actggccgtc gttttacaac gtcgtgactg ggaaaaccct 60

ggcgttaccc aacttaatcg ccttgcagca catcccccctt tcgccagctg gcgtaatagc 120

gaagaggccc gcaccgatcg cccttcccaa cagttgcgca gcctgaatgg cgaatggcgc 180

tga 183

<210> 7
<211> 102
<212> DNA
<213> 人工序列

[0185]

<220>
<223> rrnBT2 转录终止子

<400> 7
ggcccggagg gtggcgggca ggacgcccgc cataaactgc caggcatcaa attaagcaga 60

aggccatcct gacggatggc ctttttgcgt ttctacaaac tc 102

<210> 8
<211> 255
<212> DNA
<213> 人工序列

<220>
<223> OmpF 启动子

<400> 8
cacgtctcta tggaaatatg acggtgttca caaagttcct taaattttac ttttggttac 60

atatttttc tttttgaaac caaatcttta tctttgtagc actttcacgg tagcgaaacg 120

ttagtttgaa tggaaagatg cctgcagaca cataaagaca ccaaactctc atcaatagtt 180

ccgtaaattt ttattgacag aacttattga cggcagtggc aggtgtcata aaaaaaacca 240

tgagggtaat aaata 255

<210> 9
<211> 7222
<212> DNA
<213> 人工序列

<220>
<223> P215

<400> 9

[0186]

```
taactataac ggtcctaagg tagcgaagct cttcagatgg acagtcagac tgaagagcct   60
ctcttaaggt agctcgagga gcttggccca ttgcatacgt tgtatccata tcataatatg  120
tacatttata ttggctcatg tccaacatta ccgccatgtt gacattgatt attgactagt  180
tattaatagt aatcaattac ggggtcatta gttcatagcc catatatgga gttccgcgtt  240
acataactta cggtaaatgg cccgcctggc tgaccgccca acgacccccg cccattgacg  300
tcaataatga cgtatgttcc catagtaacg ccaataggga ctttccattg acgtcaatgg  360
gtggagtatt tacggtaaac tgcccacttg gcagtacatc aagtgtatca tatgccaagt  420
acgcccccta ttgacgtcaa tgacggtaaa tggcccgcct ggcattatgc ccagtacatg  480
accttatggg actttcctac ttggcagtac atctacgtat tagtcatcgc tattaccatg  540
gtgatgcggt tttggcagta catcaatggg cgtggatagc ggtttgactc acggggattt  600
ccaagtctcc accccattga cgtcaatggg agtttgtttt ggcaccaaaa tcaacgggac  660
tttccaaaat gtcgtaacaa ctccgcccca ttgacgcaaa tgggcggtag gcgtgtacgg  720
tgggaggtct atataagcag agctcgttta gtgaaccgtc ggcgcgccgc caccatggtg  780
agcaagggcg aggaggataa catggccatc atcaaggagt tcatgcgctt caaggtgcac  840
atggagggct ccgtgaacgg ccacgagttc gagatcgagg gcgagggcga gggccgcccc  900
tacgagggca cccagaccgc caagctgaag gtgaccaagg gtggcccccc gcccttcgcc  960
tgggacatcc tgtcccctca gttcatgtac ggctccaagg cctacgtgaa gcaccccgcc 1020
gacatccccg actacttgaa gctgtccttc cccgagggct tcaagtggga gcgcgtgatg 1080
aacttcgagg acggcggcgt ggtgaccgtg acccaggact cctccctgca ggacggcgag 1140
ttcatctaca aggtgaagct gcgcggcacc aacttccccc ccgacggccc cgtaatgcag 1200
aagaagacca tgggctggga ggcctcctcc gagcggatgt accccgagga cggcgccctg 1260
aagggcgaga tcaagcagag gctgaagctg aaggacggcg gccactacga cgctgaggtc 1320
aagaccacct acaaggccaa gaagcccgtg cagctgcccg gcgcctacaa cgtcaacatc 1380
aagttggaca tcacctccca caacgaggac tacaccatcg tggaacagta cgaacgcgcc 1440
```

gagggccgcc actccaccgg cggcatggac gagctgtaca agtagtctag agatacattg 1500

atgagtttgg acaaaccaca actagaatgc agtgaaaaaa atgctttatt tgtgaaattt 1560

gtgatgctat tgctttattt gtaaccatta taagctgcaa taaacaagtt aacaacaaca 1620

attgcattca ttttatgttt caggttcagg gggaggtgtg ggaggttttt taaagcaagt 1680

aaaacctcta caaatgtggt atggctgatt atgatcgcgg ccgcgttcca tgtccttata 1740

tggactcatc tttgcctatt gcgacacaca ctcagtgaac acctactacg cgctgcaaag 1800

agccccgcag gcctgaggtg cccccacctc accactcttc ctatttttgt gtaaaaatcc 1860

agcttcttgt caccacctcc aaggaggggg aggaggagga aggcaggttc ctctaggctg 1920

agccgaatgc ccctctgtgg tcccacgcca ctgatcgctg catgcccacc acctgggtac 1980

acacagtctg tgattcccgg agcagaacgg accctgccca cccggtcttg tgtgctactc 2040

agtggacaga cccaaggcaa gaaagggtga caaggacagg gtcttcccag gctggctttg 2100

agttcctagc accgccccgc ccccaatcct ctgtggcaca tggagtcttg gtccccagag 2160

tcccccagcg gcctccagat ggtctgggag ggcagttcag ctgtggctgc gcatagcaga 2220

catacaacgg acggtgggcc cagacccagg ctgtgtagac ccagcccccc cgccccgcag 2280

[0187] tgcctaggtc acccactaac gccccaggcc ttgtcttggc tgggcgtgac tgttaccctc 2340

aaaagcaggc agctccaggg taaaaggtgc cctgccctgt agagcccacc ttccttccca 2400

gggctgcggc tgggtaggtt tgtagccttc atcacgggcc acctccagcc actggaccgc 2460

tggcccctgc cctgtcctgg ggagtgtggt cctgcgactt ctaagtggcc gcaagccacc 2520

tgactccccc aacaccacac tctacctctc aagcccaggt ctctccctag tgacccaccc 2580

agcacattta gctagctgag ccccacagcc agaggtcctc aggccctgct ttcagggcag 2640

ttgctctgaa gtcggcaagg gggagtgact gcctggccac tccatgccct ccaagagctt 2700

cttctgcagg agcgtacaga acccagggcc ctggcacccg tgcagaccct ggcccacccc 2760

acctgggcgc tcagtgccca agagatgtcc acacctagga tgtcccgcgg tgggtggggg 2820

gcccgagaga cgggcaggcc gggggcaggc ctggccatgc ggggccgaac cgggcactgc 2880

ccagcgtggg gcgcgggggc cacggcgcgc gcccccagcc cccgggccca gcaccccaag 2940

gcggccaacg ccaaaactct ccctcctcct cttcctcaat ctcgctctcg ctctttttt 3000

ttttcgcaaa aggaggggag agggggtaaa aaaatgctgc actgtgcggc gaagccggtg 3060

agtgagcggc gcggggccaa tcagcgtgcg ccgttccgaa agttgccttt tatggctcga 3120

gtggccgcgg cggcgcccta taaaacccag cggcgcgacg cgccaccacc gccgagaccg 3180

[0188]

```
cgtccgcccc gcgagcacag agcctcgcct ttgccgatcc gccgcccgtc cacacccgcc 3240
gccaggtaag cccggccagc cgaccggggc aggcggctca cggcccggcc gcaggaggcc 3300
gcggcccctt cgcccgtgca gagccgccgt ctgggccgca gcgggggggcg catgggggg 3360
gaaccggacc gccgtggggg gcgcgggaga agcccctggg cctccggaga tgggggacac 3420
cccacgccag ttcggaggcg cgaggccgcg ctcgggaggc gcgctccggg ggtgccgctc 3480
tcggggcggg ggcaaccggc ggggtctttg tctgagccgg gctcttgcca atggggatcg 3540
cagggtgggc gcggcggagc ccccgccagg cccggtgggg gctggggcgc cattgcgcgt 3600
gcgcgctggt cctttgggcg ctaactgcgt gcgcgctggg aattggcgct aattgcgcgt 3660
gcgcgctggg actcaaggcg ctaactgcgc gtgcgttctg gggcccgggg tgccgcggcc 3720
tgggctgggg cgaaggcggg ctcggccgga aggggtgggg tcgccgcggc tcccgggcgc 3780
ttgcgcgcac ttcctgcccg agccgctggc cgcccgaggg tgtggccgct gcgtgcgcgc 3840
gcgccgaccc ggcgctgttt gaaccgggcg gaggcggggc tggcgcccgg ttgggagggg 3900
gttggggcct ggcttcctgc cgcgcgccgc ggggacgcct ccgaccagtg tttgcctttt 3960
atggtaataa cgcggccggc ccggcttcct ttgtccccaa tctgggcgcg cgccggcgcc 4020
ccctggcggc ctaaggactc ggctcgccgg aagtggccag ggcgggggcg acctcggctc 4080
acagcgcgcc cggctattct cgcagctcgc caccatgacc gagtacaagc ccacggtgcg 4140
cctcgccacc cgcgacgacg tcccccgggc cgtacgcacc ctcgccgccg cgttcgccga 4200
ctaccccgcc acgcgccaca ccgttgaccc ggaccgccac atcgagcggg tcaccgagct 4260
gcaagaactc ttcctcacgc gcgtcgggct cgacatcggc aaggtgtggg tcgcggacga 4320
cggcgccgcg gtggcggtct ggaccacgcc ggagagcgtc gaagcggggg cggtgttcgc 4380
cgagatcggc ccgcgcatgg ccgagttgag cggttcccgg ctggccgcgc agcaacagat 4440
ggaaggcctc ctggcgccgc accggcccaa ggagcccgcg tggttcctgg ccaccgtcgg 4500
cgtctcgccc gaccaccagg gcaagggtct gggcagcgcc gtcgtgctcc ccggagtgga 4560
ggcggccgag cgcgccgggg tgcccgcctt cctggagacc tccgcgcccc gcaacctccc 4620
cttctacgag cggctcggct tcaccgtcac cgccgacgtc gaggtgcccg aaggaccgcg 4680
cacctggtgc atgacccgca agcccggtgc ctgatgtgcc ttctagttgc cagccatctg 4740
ttgtttgccc ctccccgtg ccttccttga ccctggaagg tgccactccc actgtccttt 4800
cctaataaaa tgaggaaatt gcatcgcatt gtctgagtag gtgtcattct attctggggg 4860
```

gtggggtggg gcaggacagc aaggggggagg attgggaaga caatagcagg catgctgggg 4920

atgcggtggg ctctatggta gggataacag ggtaatagcg ggcagtgagc gcaacgcaat 4980

taatgtgagt tagctcactc attaggcacc ccaggcttta cactttatgc ttccggctcg 5040

tatgttgtgt ggaattgtga gcggataaca atttcacaca ggaaacagct atgaccatga 5100

ttacggattc actggccgtc gttttacaac gtcgtgactg ggaaaaccct ggcgttaccc 5160

aacttaatcg ccttgcagca catccccctt tcgccagctg gcgtaatagc gaagaggccc 5220

gcaccgatcg cccttcccaa cagttgcgca gcctgaatgg cgaatggcgc tgaggcccgg 5280

agggtggcgg gcaggacgcc cgccataaac tgccaggcat caaattaagc agaaggccat 5340

cctgacggat ggccttttg cgtttctaca aactctggca aacagctatt atgggtatta 5400

tgggtgacgt caggtggcac ttttcgggga aatgtgcgcg gaacccctat ttgtttattt 5460

ttctaaatac attcaaatat gtatccgctc atgagacaat aaccctgata aatgcttcaa 5520

taatattgaa aaaggaagag tatgagtatt caacatttcc gtgtcgccct tattcccttt 5580

tttgcggcat tttgccttcc tgtttttgct cacccagaaa cgctggtgaa agtaaaagat 5640

gctgaagatc agttgggtgc acgagtgggt tacatcgaac tggatctcaa cagcggtaag 5700

[0189] atccttgaga gttttcgccc cgaagaacgt tttccaatga tgagcacttt taaagttctg 5760

ctatgtggcg cggtattatc ccgtattgac gccgggcaag agcaactcgg tcgccgcata 5820

cactattctc agaatgactt ggttgagtac tcaccagtca cagaaaagca tcttacggat 5880

ggcatgacag taagagaatt atgcagtgct gccataacca tgagtgataa cactgcggcc 5940

aacttacttc tgacaacgat cggaggaccg aaggagctaa ccgcttttt gcacaacatg 6000

ggggatcatg taactcgcct tgatcgttgg gaaccggagc tgaatgaagc cataccaaac 6060

gacgagcgtg acaccacgat gcctgtagca atggcaacaa cgttgcgcaa actattaact 6120

ggcgaactac ttactctagc ttcccggcaa caattaatag actggatgga ggcggataaa 6180

gttgcaggac cacttctgcg ctcggccctt ccggctggct ggtttattgc tgataaatct 6240

ggagccggtg agcgtgggtc tcgcggtatc attgcagcac tggggccaga tggtaagccc 6300

tcccgtatcg tagttatcta cacgacgggg agtcaggcaa ctatggatga acgaaataga 6360

cagatcgctg agataggtgc ctcactgatt aagcattggt aactgtcaga ccaagtttac 6420

tcatatatac tttagattga tttaaaactt catttttaat ttaaaaggat ctaggtgaag 6480

atccttttg ataatctcat gaccaaaatc ccttaacgtg agttttcgtt ccactgagcg 6540

tcagaccccg tagaaaagat caaaggatct tcttgagatc ctttttttct gcgcgtaatc 6600

tgctgcttgc aaacaaaaaa accaccgcta ccagcggtgg tttgtttgcc ggatcaagag 6660

ctaccaactc tttttccgaa ggtaactggc ttcagcagag cgcagatacc aaatactgtt 6720

cttctagtgt agccgtagtt aggccaccac ttcaagaact ctgtagcacc gcctacatac 6780

ctcgctctgc taatcctgtt accagtggct gctgccagtg gcgataagtc gtgtcttacc 6840

gggttggact caagacgata gttaccggat aaggcgcagc ggtcgggctg aacgggggt 6900

tcgtgcacac agcccagctt ggagcgaacg acctacaccg aactgagata cctacagcgt 6960

gagctatgag aaagcgccac gcttcccgaa gggagaaagg cggacaggta tccggtaagc 7020

ggcagggtcg gaacaggaga gcgcacgagg gagcttccag ggggaaacgc ctggtatctt 7080

tatagtcctg tcgggtttcg ccacctctga cttgagcgtc gatttttgtg atgctcgtca 7140

gggggggcgga gcctatggaa aaacgccagc aacgcggcct ttttacggtt cctggccttt 7200

tgctggcctt ttgctcacat gt 7222

<210> 10
<211> 7343
<212> DNA
<213> 人工序列

[0190]

<220>
<223> P216

<400> 10
taactataac ggtcctaagg tagcgaagct cttcagatgg acagtcagac tgaagagcct 60

ctcttaaggt agctcgagga gcttggccca ttgcatacgt tgtatccata tcataatatg 120

tacatttata ttggctcatg tccaacatta ccgccatgtt gacattgatt attgactagt 180

tattaatagt aatcaattac ggggtcatta gttcatagcc catatatgga gttccgcgtt 240

acataactta cggtaaatgg cccgcctggc tgaccgccca acgacccccg cccattgacg 300

tcaataatga cgtatgttcc catagtaacg ccaataggga ctttccattg acgtcaatgg 360

gtggagtatt tacggtaaac tgcccacttg gcagtacatc aagtgtatca tatgccaagt 420

acgcccccta ttgacgtcaa tgacggtaaa tggcccgcct ggcattatgc ccagtacatg 480

accttatggg actttcctac ttggcagtac atctacgtat tagtcatcgc tattaccatg 540

gtgatgcggt tttggcagta catcaatggg cgtggatagc ggtttgactc acggggattt 600

ccaagtctcc accccattga cgtcaatggg agtttgtttt ggcaccaaaa tcaacgggac 660

tttccaaaat gtcgtaacaa ctccgcccca ttgacgcaaa tgggcggtag gcgtgtacgg 720

tgggaggtct atataagcag agctcgttta gtgaaccgtc ggcgcgccgc caccatggtg 780

agcaagggcg aggaggataa catggccatc atcaaggagt tcatgcgctt caaggtgcac 840

atggagggct ccgtgaacgg ccacgagttc gagatcgagg gcgagggcga gggccgcccc 900

tacgagggca cccagaccgc caagctgaag gtgaccaagg gtggcccccc gcccttcgcc 960

tgggacatcc tgtcccctca gttcatgtac ggctccaagg cctacgtgaa gcaccccgcc 1020

gacatccccg actacttgaa gctgtccttc cccgagggct tcaagtggga gcgcgtgatg 1080

aacttcgagg acggcggcgt ggtgaccgtg acccaggact cctccctgca ggacggcgag 1140

ttcatctaca aggtgaagct gcgcggcacc aacttccccc ccgacggccc cgtaatgcag 1200

aagaagacca tgggctggga ggcctcctcc gagcggatgt accccgagga cggcgccctg 1260

aagggcgaga tcaagcagag gctgaagctg aaggacggcg gccactacga cgctgaggtc 1320

aagaccacct acaaggccaa gaagcccgtg cagctgcccg gcgcctacaa cgtcaacatc 1380

aagttggaca tcacctccca caacgaggac tacaccatcg tggaacagta cgaacgcgcc 1440

gagggccgcc actccaccgg cggcatggac gagctgtaca agtagtctag agatacattg 1500

atgagtttgg acaaaccaca actagaatgc agtgaaaaaa atgctttatt tgtgaaattt 1560

[0191] gtgatgctat tgctttattt gtaaccatta taagctgcaa taaacaagtt aacaacaaca 1620

attgcattca ttttatgttt caggttcagg gggaggtgtg ggaggttttt taaagcaagt 1680

aaaacctcta caaatgtggt atggctgatt atgatcgcgg ccgcgttcca tgtccttata 1740

tggactcatc tttgcctatt gcgacacaca ctcagtgaac acctactacg cgctgcaaag 1800

agccccgcag gcctgaggtg cccccacctc accactcttc ctatttttgt gtaaaaatcc 1860

agcttcttgt caccacctcc aaggaggggg aggaggagga aggcaggttc ctctaggctg 1920

agccgaatgc ccctctgtgg tcccacgcca ctgatcgctg catgcccacc acctgggtac 1980

acacagtctg tgattcccgg agcagaacgg accctgccca cccggtcttg tgtgctactc 2040

agtggacaga cccaaggcaa gaaagggtga caaggacagg gtcttcccag gctggctttg 2100

agttcctagc accgccccgc ccccaatcct ctgtggcaca tggagtcttg gtccccagag 2160

tcccccagcg gcctccagat ggtctgggag ggcagttcag ctgtggctgc gcatagcaga 2220

catacaacgg acggtgggcc cagacccagg ctgtgtagac ccagcccccc cgccccgcag 2280

tgcctaggtc acccactaac gccccaggcc ttgtcttggc tgggcgtgac tgttaccctc 2340

aaaagcaggc agctccaggg taaaaggtgc cctgccctgt agagcccacc ttccttccca 2400

gggctgcggc tgggtaggtt tgtagccttc atcacgggcc acctccagcc actggaccgc 2460

[0192]

```
tggcccctgc cctgtcctgg ggagtgtggt cctgcgactt ctaagtggcc gcaagccacc 2520
tgactccccc aacaccacac tctacctctc aagcccaggt ctctccctag tgacccaccc 2580
agcacattta gctagctgag ccccacagcc agaggtcctc aggccctgct ttcagggcag 2640
ttgctctgaa gtcggcaagg gggagtgact gcctggccac tccatgccct ccaagagctt 2700
cttctgcagg agcgtacaga acccagggcc ctggcacccg tgcagaccct ggcccacccc 2760
acctgggcgc tcagtgccca agagatgtcc acacctagga tgtcccgcgg tgggtggggg 2820
gcccgagaga cgggcaggcc gggggcaggc ctggccatgc ggggccgaac cgggcactgc 2880
ccagcgtggg gcgcgggggc cacggcgcgc gcccccagcc cccgggccca gcaccccaag 2940
gcggccaacg ccaaaactct ccctcctcct cttcctcaat ctcgctctcg ctctttttt 3000
ttttcgcaaa aggaggggag agggggtaaa aaaatgctgc actgtgcggc gaagccggtg 3060
agtgagcggc gcggggccaa tcagcgtgcg ccgttccgaa agttgccttt tatggctcga 3120
gtggccgcgg cggcgcccta taaaacccag cggcgcgacg cgccaccacc gccgagaccg 3180
cgtccgcccc gcgagcacag agcctcgcct ttgccgatcc gccgcccgtc cacacccgcc 3240
gccaggtaag cccggccagc cgaccggggc aggcggctca cggcccggcc gcaggaggcc 3300
gcggcccctt cgcccgtgca gagccgccgt ctgggccgca gcgggggcg catgggggg 3360
gaaccggacc gccgtggggg gcgcgggaga agcccctggg cctccggaga tgggggacac 3420
cccacgccag ttcggaggcg cgaggccgcg ctcgggaggc gcgctccggg ggtgccgctc 3480
tcggggcggg ggcaaccggc ggggtctttg tctgagccgg gctcttgcca atggggatcg 3540
cagggtgggc gcggcggagc ccccgccagg cccggtgggg gctggggcgc cattgcgcgt 3600
gcgcgctggt cctttgggcg ctaactgcgt gcgcgctggg aattggcgct aattgcgcgt 3660
gcgcgctggg actcaaggcg ctaactgcgc gtgcgttctg gggcccgggg tgccgcggcc 3720
tgggctgggg cgaaggcggg ctcggccgga aggggtgggg tcgccgcggc tcccgggcgc 3780
ttgcgcgcac ttcctgcccg agccgctggc cgcccgaggg tgtggccgct gcgtgcgcgc 3840
gcgccgaccc ggcgctgttt gaaccgggcg gaggcggggc tggcgcccgg ttgggagggg 3900
gttggggcct ggcttcctgc cgcgcgccgc ggggacgcct ccgaccagtg tttgcctttt 3960
atggtaataa cgcggccggc ccggcttcct ttgtccccaa tctgggcgcg cgccggcgcc 4020
ccctggcggc ctaaggactc ggctcgccgg aagtggccag ggcgggggcg acctcggctc 4080
acagcgcgcc cggctattct cgcagctcgc caccatgacc gagtacaagc ccacggtgcg 4140
```

cctcgccacc cgcgacgacg tcccccgggc cgtacgcacc ctcgccgccg cgttcgccga 4200

ctaccccgcc acgcgccaca ccgttgaccc ggaccgccac atcgagcggg tcaccgagct 4260

gcaagaactc ttcctcacgc gcgtcgggct cgacatcggc aaggtgtggg tcgcggacga 4320

cggcgccgcg gtggcggtct ggaccacgcc ggagagcgtc gaagcggggg cggtgttcgc 4380

cgagatcggc ccgcgcatgg ccgagttgag cggttcccgg ctggccgcgc agcaacagat 4440

ggaaggcctc ctggcgccgc accggcccaa ggagcccgcg tggttcctgg ccaccgtcgg 4500

cgtctcgccc gaccaccagg gcaagggtct gggcagcgcc gtcgtgctcc ccggagtgga 4560

ggcggccgag cgcgccgggg tgcccgcctt cctggagacc tccgcgcccc gcaacctccc 4620

cttctacgag cggctcggct tcaccgtcac cgccgacgtc gaggtgcccg aaggaccgcg 4680

cacctggtgc atgacccgca agcccggtgc ctgatgtgcc ttctagttgc cagccatctg 4740

ttgtttgccc ctccccgtg ccttccttga ccctggaagg tgccactccc actgtccttt 4800

cctaataaaa tgaggaaatt gcatcgcatt gtctgagtag gtgtcattct attctggggg 4860

gtggggtggg gcaggacagc aaggggggagg attgggaaga caatagcagg catgctgggg 4920

atgcggtggg ctctatggta gggataacag ggtaatcacg tctctatgga aatatgacgg 4980

[0193] tgttcacaaa gttccttaaa ttttactttt ggttacatat tttttctttt tgaaaccaaa 5040

tctttatctt tgtagcactt tcacggtagc gaaacgttag tttgaatgga aagatgcctg 5100

cagacacata aagacaccaa actctcatca atagttccgt aaatttttat tgacagaact 5160

tattgacggc agtggcaggt gtcataaaaa aaaccatgag ggtaataaat aatgaccatg 5220

attacggatt cactggccgt cgttttacaa cgtcgtgact gggaaaaccc tggcgttacc 5280

caacttaatc gccttgcagc acatcccccc ttcgccagct ggcgtaatag cgaagaggcc 5340

cgcaccgatc gcccttccca acagttgcgc agcctgaatg gcgaatggcg ctgaggcccg 5400

gagggtggcg ggcaggacgc ccgccataaa ctgccaggca tcaaattaag cagaaggcca 5460

tcctgacgga tggccttttt gcgtttctac aaactctggc aaacagctat tatgggtatt 5520

atgggtgacg tcaggtggca cttttcgggg aaatgtgcgc ggaaccccta tttgtttatt 5580

tttctaaata cattcaaata tgtatccgct catgagacaa taaccctgat aaatgcttca 5640

ataatattga aaaaggaaga gtatgagtat tcaacatttc cgtgtcgccc ttattccctt 5700

ttttgcggca ttttgccttc ctgtttttgc tcacccagaa acgctggtga aagtaaaaga 5760

tgctgaagat cagttgggtg cacgagtggg ttacatcgaa ctggatctca acagcggtaa 5820

gatccttgag agttttcgcc ccgaagaacg ttttccaatg atgagcactt ttaaagttct 5880

[0194]

gctatgtggc gcggtattat cccgtattga cgccgggcaa gagcaactcg gtcgccgcat 5940

acactattct cagaatgact tggttgagta ctcaccagtc acagaaaagc atcttacgga 6000

tggcatgaca gtaagagaat tatgcagtgc tgccataacc atgagtgata acactgcggc 6060

caacttactt ctgacaacga tcggaggacc gaaggagcta accgcttttt tgcacaacat 6120

gggggatcat gtaactcgcc ttgatcgttg ggaaccggag ctgaatgaag ccataccaaa 6180

cgacgagcgt gacaccacga tgcctgtagc aatggcaaca acgttgcgca aactattaac 6240

tggcgaacta cttactctag cttcccggca acaattaata gactggatgg aggcggataa 6300

agttgcagga ccacttctgc gctcggccct tccggctggc tggtttattg ctgataaatc 6360

tggagccggt gagcgtgggt ctcgcggtat cattgcagca ctggggccag atggtaagcc 6420

ctcccgtatc gtagttatct acacgacggg gagtcaggca actatggatg aacgaaatag 6480

acagatcgct gagataggtg cctcactgat taagcattgg taactgtcag accaagttta 6540

ctcatatata ctttagattg atttaaaact tcatttttaa tttaaaagga tctaggtgaa 6600

gatccttttt gataatctca tgaccaaaat cccttaacgt gagttttcgt tccactgagc 6660

gtcagacccc gtagaaaaga tcaaaggatc ttcttgagat cctttttttc tgcgcgtaat 6720

ctgctgcttg caaacaaaaa aaccaccgct accagcggtg gtttgtttgc cggatcaaga 6780

gctaccaact ctttttccga aggtaactgg cttcagcaga gcgcagatac caaatactgt 6840

tcttctagtg tagccgtagt taggccacca cttcaagaac tctgtagcac cgcctacata 6900

cctcgctctg ctaatcctgt taccagtggc tgctgccagt ggcgataagt cgtgtcttac 6960

cgggttggac tcaagacgat agttaccgga taaggcgcag cggtcgggct gaacgggggg 7020

ttcgtgcaca gcccagct tggagcgaac gacctacacc gaactgagat acctacagcg 7080

tgagctatga gaaagcgcca cgcttcccga agggagaaag gcggacaggt atccggtaag 7140

cggcagggtc ggaacaggag agcgcacgag ggagcttcca gggggaaacg cctggtatct 7200

ttatagtcct gtcgggtttc gccacctctg acttgagcgt cgatttttgt gatgctcgtc 7260

aggggggcgg agcctatgga aaaacgccag caacgcggcc tttttacggt tcctggcctt 7320

ttgctggcct tttgctcaca tgt 7343

<210> 11
<211> 2329
<212> DNA
<213> 人工序列

<220>
<223> P217

<400> 11
agcgggcagt gagcgcaacg caattaatgt gagttagctc actcattagg caccccaggc 60

tttacacttt atgcttccgg ctcgtatgtt gtgtggaatt gtgagcggat aacaatttca 120

cacaggaaac agctatgacc atgattacgg attcactggc cgtcgtttta caacgtcgtg 180

actgggaaaa ccctggcgtt acccaactta atcgccttgc agcacatccc cctttcgcca 240

gctggcgtaa tagcgaagag gcccgcaccg atcgcccttc ccaacagttg cgcagcctga 300

atggcgaatg gcgctgaggc ccggagggtg gcgggcagga cgcccgccat aaactgccag 360

gcatcaaatt aagcagaagg ccatcctgac ggatggcctt tttgcgtttc tacaaactct 420

ggcaaacagc tattatgggt attatgggtg acgtcaggtg gcacttttcg gggaaatgtg 480

cgcggaaccc ctatttgttt atttttctaa atacattcaa atatgtatcc gctcatgaga 540

caataaccct gataaatgct tcaataatat tgaaaaagga agagtatgag tattcaacat 600

ttccgtgtcg cccttattcc ctttttttgcg gcattttgcc ttcctgtttt tgctcaccca 660

gaaacgctgg tgaaagtaaa agatgctgaa gatcagttgg gtgcacgagt gggttacatc 720

[0195] gaactggatc tcaacagcgg taagatcctt gagagttttc gccccgaaga acgttttcca 780

atgatgagca cttttaaagt tctgctatgt ggcgcggtat tatcccgtat tgacgccggg 840

caagagcaac tcggtcgccg catacactat tctcagaatg acttggttga gtactcacca 900

gtcacagaaa agcatcttac ggatggcatg acagtaagag aattatgcag tgctgccata 960

accatgagtg ataacactgc ggccaactta cttctgacaa cgatcggagg accgaaggag 1020

ctaaccgctt ttttgcacaa catgggggat catgtaactc gccttgatcg ttgggaaccg 1080

gagctgaatg aagccatacc aaacgacgag cgtgacacca cgatgcctgt agcaatggca 1140

acaacgttgc gcaaactatt aactggcgaa ctacttactc tagcttcccg gcaacaatta 1200

atagactgga tggaggcgga taaagttgca ggaccacttc tgcgctcggc ccttccggct 1260

ggctggttta ttgctgataa atctggagcc ggtgagcgtg ggtctcgcgg tatcattgca 1320

gcactggggc cagatggtaa gccctcccgt atcgtagtta tctacacgac ggggagtcag 1380

gcaactatgg atgaacgaaa tagacagatc gctgagatag gtgcctcact gattaagcat 1440

tggtaactgt cagaccaagt ttactcatat atactttaga ttgatttaaa acttcatttt 1500

taatttaaaa ggatctaggt gaagatcctt tttgataatc tcatgaccaa aatcccttaa 1560

cgtgagtttt cgttccactg agcgtcagac cccgtagaaa agatcaaagg atcttcttga 1620

gatccttttt ttctgcgcgt aatctgctgc ttgcaaacaa aaaaaccacc gctaccagcg 1680

gtggtttgtt tgccggatca agagctacca actcttttc cgaaggtaac tggcttcagc 1740

agagcgcaga taccaaatac tgttcttcta gtgtagccgt agttaggcca ccacttcaag 1800

aactctgtag caccgcctac atacctcgct ctgctaatcc tgttaccagt ggctgctgcc 1860

agtggcgata agtcgtgtct taccgggttg gactcaagac gatagttacc ggataaggcg 1920

cagcggtcgg gctgaacggg gggttcgtgc acacagccca gcttggagcg aacgacctac 1980

accgaactga gatacctaca gcgtgagcta tgagaaagcg ccacgcttcc cgaagggaga 2040

aaggcggaca ggtatccggt aagcggcagg gtcggaacag gagagcgcac gagggagctt 2100

ccagggggaa acgcctggta tctttatagt cctgtcgggt ttcgccacct ctgacttgag 2160

cgtcgatttt tgtgatgctc gtcaggggggg cggagcctat ggaaaaacgc cagcaacgcg 2220

gccttttac ggttcctggc cttttgctgg ccttttgctc acatgttaac tataacggtc 2280

ctaaggtagc gaagctcggt gggctctatg gtagggataa cagggtaat 2329

[0196]

<210> 12
<211> 1143
<212> DNA
<213> 人工序列

<220>
<223> P218

<400> 12
agcgggcagt gagcgcaacg caattaatgt gagttagctc actcattagg caccccaggc 60

tttacacttt atgcttccgg ctcgtatgtt gtgtggaatt gtgagcggat aacaatttca 120

cacaggaaac agctatgacc atgattacgg attcactggc cgtcgtttta caacgtcgtg 180

actgggaaaa ccctggcgtt acccaactta atcgccttgc agcacatccc cctttcgcca 240

gctggcgtaa tagcgaagag gcccgcaccg atcgcccttc ccaacagttg cgcagcctga 300

atggcgaatg gcgctgaggc ccggagggtg gcgggcagga cgcccgccat aaactgccag 360

gcatcaaatt aagcagaagg ccatcctgac ggatggcctt tttgcgtttc tacaaactca 420

aaggatcttc ttgagatcct ttttttctgc gcgtaatctg ctgcttgcaa acaaaaaaac 480

caccgctacc agcggtggtt tgtttgccgg atcaagagct accaactctt tttccgaagg 540

taactggctt cagcagagcg cagataccaa atactgttct tctagtgtag ccgtagttag 600

gccaccactt caagaactct gtagcaccgc ctacatacct cgctctgcta atcctgttac 660

cagtggctgc tgccagtggc gataagtcgt gtcttaccgg gttggactca agacgatagt 720

taccggataa ggcgcagcgg tcgggctgaa cgggggggttc gtgcacacag cccagcttgg 780

agcgaacgac ctacaccgaa ctgagatacc tacagcgtga gctatgagaa agcgccacgc 840

ttcccgaagg gagaaaggcg gacaggtatc cggtaagcgg cagggtcgga acaggagagc 900

gcacgaggga gcttccaggg ggaaacgcct ggtatcttta tagtcctgtc gggtttcgcc 960

acctctgact tgagcgtcga tttttgtgat gctcgtcagg ggggcggagc ctatggaaaa 1020

acgccagcaa cgcggccttt ttacggttcc tggccttttg ctggcctttt gctcacatgt 1080

taactataac ggtcctaagg tagcgaagct cggtgggctc tatggtaggg ataacagggt 1140

aat 1143

<210> 13
<211> 1047
<212> DNA
<213> 人工序列

<220>
<223> P219

<400> 13

[0197] agcgggcagt gagcgcaacg caattaatgt gagttagctc actcattagg caccccaggc 60

tttacacttt atgcttccgg ctcgtatgtt gtgtggaatt gtgagcggat aacaatttca 120

cacaggaaac agctatgacc atgattacgg attcactggc cgtcgtttta caacgtcgtg 180

actgggaaaa ccctggcgtt acccaactta atcgccttgc agcacatccc cctttcgcca 240

gctggcgtaa tagcgaagag gcccgcaccg atcgcccttc ccaacagttg cgcagcctga 300

atggcgaatg gcgctgaaag cttaaaggat cttcttgaga tccttttttt ctgcgcgtaa 360

tctgctgctt gcaaacaaaa aaaccaccgc taccagcggt ggtttgtttg ccggatcaag 420

agctaccaac tctttttccg aaggtaactg gcttcagcag agcgcagata ccaaatactg 480

ttcttctagt gtagccgtag ttaggccacc acttcaagaa ctctgtagca ccgcctacat 540

acctcgctct gctaatcctg ttaccagtgg ctgctgccag tggcgataag tcgtgtctta 600

ccgggttgga ctcaagacga tagttaccgg ataaggcgca gcggtcgggc tgaacggggg 660

gttcgtgcac acagcccagc ttggagcgaa cgacctacac cgaactgaga tacctacagc 720

gtgagctatg agaaagcgcc acgcttcccg aagggagaaa ggcggacagg tatccggtaa 780

gcggcagggt cggaacagga gagcgcacga gggagcttcc agggggaaac gcctggtatc 840

tttatagtcc tgtcgggttt cgccacctct gacttgagcg tcgatttttg tgatgctcgt 900

caggggggcg gagcctatgg aaaaacgcca gcaacgcggc ctttttacgg ttcctggcct 960

tttgctggcc ttttgctcac atgttaacta taacggtcct aaggtagcga agctcggtgg 1020

gctctatggt agggataaca gggtaat 1047

<210> 14
<211> 25
<212> DNA
<213> 人工序列

<220>
<223> CNFOR

<400> 14
tgtgtggaat tgtgagcgga taaca 25

<210> 15
<211> 27
<212> DNA
<213> 人工序列

<220>
<223> P455R2

[0198] <400> 15
tggcgttact atgggaacat acgtcat 27

<210> 16
<211> 6732
<212> DNA
<213> 人工序列

<220>
<223> GWIZ 荧光素酶

<400> 16
tcgcgcgttt cggtgatgac ggtgaaaacc tctgacacat gcagctcccg gagacggtca 60

cagcttgtct gtaagcggat gccgggagca gacaagcccg tcagggcgcg tcagcgggtg 120

ttggcgggtg tcggggctgg cttaactatg cggcatcaga gcagattgta ctgagagtgc 180

accatatgcg gtgtgaaata ccgcacagat gcgtaaggag aaaataccgc atcagattgg 240

ctattggcca ttgcatacgt tgtatccata tcataatatg tacatttata ttggctcatg 300

tccaacatta ccgccatgtt gacattgatt attgactagt tattaatagt aatcaattac 360

ggggtcatta gttcatagcc catatatgga gttccgcgtt acataactta cggtaaatgg 420

cccgcctggc tgaccgccca acgacccccg cccattgacg tcaataatga cgtatgttcc 480

[0199]

```
catagtaacg ccaataggga ctttccattg acgtcaatgg gtggagtatt tacggtaaac  540
tgcccacttg gcagtacatc aagtgtatca tatgccaagt acgccccta ttgacgtcaa  600
tgacggtaaa tggcccgcct ggcattatgc ccagtacatg accttatggg actttcctac  660
ttggcagtac atctacgtat tagtcatcgc tattaccatg gtgatgcggt tttggcagta  720
catcaatggg cgtggatagc ggtttgactc acggggattt ccaagtctcc accccattga  780
cgtcaatggg agtttgtttt ggcaccaaaa tcaacgggac tttccaaaat gtcgtaacaa  840
ctccgcccca ttgacgcaaa tgggcggtag gcgtgtacgg tgggaggtct atataagcag  900
agctcgttta gtgaaccgtc agatcgcctg gagacgccat ccacgctgtt ttgacctcca  960
tagaagacac cgggaccgat ccagcctccg cggccgggaa cggtgcattg gaacgcggat 1020
tccccgtgcc aagagtgacg taagtaccgc ctatagactc tataggcaca ccccttggc 1080
tcttatgcat gctatactgt ttttggcttg gggcctatac acccccgctt ccttatgcta 1140
taggtgatgg tatagcttag cctataggtg tgggttattg accattattg accactcccc 1200
tattggtgac gatactttcc attactaatc cataacatgg ctctttgcca caactatctc 1260
tattggctat atgccaatac tctgtccttc agagactgac acggactctg tatttttaca 1320
ggatggggtc ccatttatta tttacaaatt cacatataca acaacgccgt cccccgtgcc 1380
cgcagttttt attaaacata gcgtgggatc tccacgcgaa tctcgggtac gtgttccgga 1440
catgggctct tctccggtag cggcggagct tccacatccg agccctggtc ccatgcctcc 1500
agcggctcat ggtcgctcgg cagctccttg ctcctaacag tggaggccag acttaggcac 1560
agcacaatgc ccaccaccac cagtgtgccg cacaaggccg tggcggtagg gtatgtgtct 1620
gaaaatgagc gtggagattg ggctcgcacg gctgacgcag atggaagact taaggcagcg 1680
gcagaagaag atgcaggcag ctgagttgtt gtattctgat aagagtcaga ggtaactccc 1740
gttgcggtgc tgttaacggt ggagggcagt gtagtctgag cagtactcgt tgctgccgcg 1800
cgcgccacca gacataatag ctgacagact aacagactgt tcctttccat gggtcttttc 1860
tgcagtcacc gtcgtcgaca cgtgtgatca gatatcgcgg ccgctctagg aagctttcca 1920
tggaagacgc caaaaacata aagaaaggcc cggcgccatt ctatccgctg gaagatggaa 1980
ccgctggaga gcaactgcat aaggctatga agagatacgc cctggttcct ggaacaattg 2040
cttttacaga tgcacatatc gaggtggaca tcacttacgc tgagtacttc gaaatgtccg 2100
ttcggttggc agaagctatg aaacgatatg ggctgaatac aaatcacaga atcgtcgtat 2160
```

gcagtgaaaa ctctcttcaa ttctttatgc cggtgttggg cgcgttattt atcggagttg 2220

cagttgcgcc cgcgaacgac atttataatg aacgtgaatt gctcaacagt atgggcattt 2280

cgcagcctac cgtggtgttc gtttccaaaa aggggttgca aaaaattttg aacgtgcaaa 2340

aaaagctccc aatcatccaa aaaattatta tcatggattc taaaacggat taccagggat 2400

ttcagtcgat gtacacgttc gtcacatctc atctacctcc cggttttaat gaatacgatt 2460

ttgtgccaga gtccttcgat agggacaaga caattgcact gatcatgaac tcctctggat 2520

ctactggtct gcctaaaggt gtcgctctgc ctcatagaac tgcctgcgtg agattctcgc 2580

atgccagaga tcctattttt ggcaatcaaa tcattccgga tactgcgatt ttaagtgttg 2640

ttccattcca tcacggtttt ggaatgttta ctacactcgg atatttgata tgtggatttc 2700

gagtcgtctt aatgtataga tttgaagaag agctgtttct gaggagcctt caggattaca 2760

agattcaaag tgcgctgctg gtgccaaccc tattctcctt cttcgccaaa agcactctga 2820

ttgacaaata cgatttatct aatttacacg aaattgcttc tggtggcgct cccctctcta 2880

aggaagtcgg ggaagcggtt gccaagaggt tccatctgcc aggtatcagg caaggatatg 2940

ggctcactga gactacatca gctattctga ttacacccga gggggatgat aaaccgggcg 3000

[0200] cggtcggtaa agttgttcca ttttttgaag cgaaggttgt ggatctggat accgggaaaa 3060

cgctgggcgt taatcaaaga ggcgaactgt gtgtgagagg tcctatgatt atgtccggtt 3120

atgtaaacaa tccggaagcg accaacgcct tgattgacaa ggatggatgg ctacattctg 3180

gagacatagc ttactgggac gaagacgaac acttcttcat cgttgaccgc ctgaagtctc 3240

tgattaagta caaaggctat caggtggctc ccgctgaatt ggaatccatc ttgctccaac 3300

accccaacat cttcgacgca ggtgtcgcag gtcttcccga cgatgacgcc ggtgaacttc 3360

ccgccgccgt tgttgttttg gagcacggaa agacgatgac ggaaaaagag atcgtggatt 3420

acgtcgccag tcaagtaaca accgcgaaaa agttgcgcgg aggagttgtg tttgtggacg 3480

aagtaccgaa aggtcttacc ggaaaactcg acgcaagaaa aatcagagag atcctcataa 3540

aggccaagaa gggcggaaag atcgccgtgt aattctagac caggcgcctg gatccagatc 3600

acttctggct aataaaagat cagagctcta gagatctgtg tgttggtttt ttgtggatct 3660

gctgtgcctt ctagttgcca gccatctgtt gtttgccccc cccccgtgcc ttccttgacc 3720

ctggaaggtg ccactcccac tgtcctttcc taataaaatg aggaaattgc atcgcattgt 3780

ctgagtaggt gtcattctat tctgggggggt ggggtggggc aggacagcaa gggggaggat 3840

tgggaagaca atagcaggca tgctggggat gcggtgggct ctatgggtac ctctctctct 3900

[0201]

ctctctctct ctctctctct ctctctctct cggtacctct ctctctctct ctctctctct 3960

ctctctctct ctctctcggt accaggtgct gaagaattga cccggttcct cctgggccag 4020

aaagaagcag gcacatcccc ttctctgtga cacaccctgt ccacgcccct ggttcttagt 4080

tccagcccca ctcataggac actcatagct caggagggct ccgccttcaa tcccacccgc 4140

taaagtactt ggagcggtct ctccctccct catcagccca ccaaaccaaa cctagcctcc 4200

aagagtggga agaaattaaa gcaagatagg ctattaagtg cagagggaga gaaaatgcct 4260

ccaacatgtg aggaagtaat gagagaaatc atagaatttc ttccgcttcc tcgctcactg 4320

actcgctgcg ctcggtcgtt cggctgcggc gagcggtatc agctcactca aaggcggtaa 4380

tacggttatc cacagaatca ggggataacg caggaaagaa catgtgagca aaaggccagc 4440

aaaaggccag gaaccgtaaa aaggccgcgt tgctggcgtt tttccatagg ctccgccccc 4500

ctgacgagca tcacaaaaat cgacgctcaa gtcagaggtg gcgaaacccg acaggactat 4560

aaagatacca ggcgtttccc cctggaagct ccctcgtgcg ctctcctgtt ccgaccctgc 4620

cgcttaccgg atacctgtcc gcctttctcc cttcgggaag cgtggcgctt tctcaatgct 4680

cacgctgtag gtatctcagt tcggtgtagg tcgttcgctc caagctgggc tgtgtgcacg 4740

aacccccccgt tcagcccgac cgctgcgcct tatccggtaa ctatcgtctt gagtccaacc 4800

cggtaagaca cgacttatcg ccactggcag cagccactgg taacaggatt agcagagcga 4860

ggtatgtagg cggtgctaca gagttcttga agtggtggcc taactacggc tacactagaa 4920

ggacagtatt tggtatctgc gctctgctga agccagttac cttcggaaaa agagttggta 4980

gctcttgatc cggcaaacaa accaccgctg gtagcggtgg ttttttgtt tgcaagcagc 5040

agattacgcg cagaaaaaaa ggatctcaag aagatccttt gatcttttct acggggtctg 5100

acgctcagtg gaacgaaaac tcacgttaag ggattttggt catgagatta tcaaaaagga 5160

tcttcaccta gatcctttta aattaaaaat gaagttttaa atcaatctaa agtatatatg 5220

agtaaacttg gtctgacagt taccaatgct taatcagtga ggcacctatc tcagcgatct 5280

gtctatttcg ttcatccata gttgcctgac tccgggggggg gggggcgctg aggtctgcct 5340

cgtgaagaag gtgttgctga ctcataccag gcctgaatcg ccccatcatc cagccagaaa 5400

gtgagggagc cacggttgat gagagctttg ttgtaggtgg accagttggt gattttgaac 5460

ttttgctttg ccacggaacg gtctgcgttg tcgggaagat gcgtgatctg atccttcaac 5520

tcagcaaaag ttcgatttat tcaacaaagc cgccgtcccg tcaagtcagc gtaatgctct 5580

gccagtgtta caaccaatta accaattctg attagaaaaa ctcatcgagc atcaaatgaa 5640

actgcaattt attcatatca ggattatcaa taccatattt ttgaaaaagc cgtttctgta 5700

atgaaggaga aaactcaccg aggcagttcc ataggatggc aagatcctgg tatcggtctg 5760

cgattccgac tcgtccaaca tcaatacaac ctattaattt cccctcgtca aaaataaggt 5820

tatcaagtga gaaatcacca tgagtgacga ctgaatccgg tgagaatggc aaaagcttat 5880

gcatttcttt ccagacttgt tcaacaggcc agccattacg ctcgtcatca aaatcactcg 5940

catcaaccaa accgttattc attcgtgatt gcgcctgagc gagacgaaat acgcgatcgc 6000

tgttaaaagg acaattacaa acaggaatcg aatgcaaccg gcgcaggaac actgccagcg 6060

catcaacaat attttcacct gaatcaggat attcttctaa tacctggaat gctgttttcc 6120

cggggatcgc agtggtgagt aaccatgcat catcaggagt acggataaaa tgcttgatgg 6180

tcggaagagg cataaattcc gtcagccagt ttagtctgac catctcatct gtaacatcat 6240

tggcaacgct acctttgcca tgtttcagaa acaactctgg cgcatcgggc ttcccataca 6300

atcgatagat tgtcgcacct gattgcccga cattatcgcg agcccattta tacccatata 6360

aatcagcatc catgttggaa tttaatcgcg gcctcgagca agacgtttcc cgttgaatat 6420

[0202] ggctcataac acccccttgta ttactgttta tgtaagcaga cagttttatt gttcatgatg 6480

atatattttt atcttgtgca atgtaacatc agagattttg agacacaacg tggctttccc 6540

ccccccccca ttattgaagc atttatcagg gttattgtct catgagcgga tacatatttg 6600

aatgtattta gaaaaataaa caaatagggg ttccgcgcac atttccccga aaagtgccac 6660

ctgacgtcta agaaaccatt attatcatga cattaaccta taaaaatagg cgtatcacga 6720

ggcccctttcg tc 6732

<210> 17
<211> 5070
<212> DNA
<213> 人工序列

<220>
<223> P469-2

<400> 17
tagggataac agggtaatag cgggcagtga gcgcaacgca attaatgtga gttagctcac 60

tcattaggca ccccaggctt tacactttat gcttccggct cgtatgttgt gtggaattgt 120

gagcggataa caatttcaca caggaaacag ctatgaccat gattacggat tcactggccg 180

tcgttttaca acgtcgtgac tgggaaaacc ctggcgttac ccaacttaat cgccttgcag 240

[0203]

```
cacatccccc tttcgccagc tggcgtaata gcgaagaggc ccgcaccgat cgcccttccc  300
aacagttgcg cagcctgaat ggcgaatggc gctgaaagct taaaggatct tcttgagatc  360
ctttttttct gcgcgtaatc tgctgcttgc aaacaaaaaa accaccgcta ccagcggtgg  420
tttgtttgcc ggatcaagag ctaccaactc tttttccgaa ggtaactggc ttcagcagag  480
cgcagatacc aaatactgtt cttctagtgt agccgtagtt aggccaccac ttcaagaact  540
ctgtagcacc gcctacatac ctcgctctgc taatcctgtt accagtggct gctgccagtg  600
gcgataagtc gtgtcttacc gggttggact caagacgata gttaccggat aaggcgcagc  660
ggtcgggctg aacgggggt tcgtgcacac agcccagctt ggagcgaacg acctacaccg  720
aactgagata cctacagcgt gagctatgag aaagcgccac gcttcccgaa gggagaaagg  780
cggacaggta tccggtaagc ggcagggtcg gaacaggaga gcgcacgagg gagcttccag  840
ggggaaacgc ctggtatctt tatagtcctg tcgggtttcg ccacctctga cttgagcgtc  900
gatttttgtg atgctcgtca ggggggcgga gcctatggaa aaacgccagc aacgcgggtg  960
cgcataatgt atattatgtt aaattaacta taacggtcct aaggtagcga atggccattg 1020
catacgttgt atccatatca taatatgtac atttatattg gctcatgtcc aacattaccg 1080
ccatgttgac attgattatt gactagttat taatagtaat caattacggg gtcattagtt 1140
catagcccat atatggagtt ccgcgttaca taacttacgg taaatggccc gcctggctga 1200
ccgcccaacg acccccgccc attgacgtca ataatgacgt atgttcccat agtaacgcca 1260
atagggactt tccattgacg tcaatgggtg gagtatttac ggtaaactgc ccacttggca 1320
gtacatcaag tgtatcatat gccaagtacg ccccctattg acgtcaatga cggtaaatgg 1380
cccgcctggc attatgccca gtacatgacc ttatgggact ttcctacttg gcagtacatc 1440
tacgtattag tcatcgctat taccatggtg atgcggtttt ggcagtacat caatgggcgt 1500
ggatagcggt ttgactcacg gggatttcca agtctccacc ccattgacgt caatgggagt 1560
ttgttttggc accaaaatca acgggacttt ccaaaatgtc gtaacaactc cgccccattg 1620
acgcaaatgg gcggtaggcg tgtacggtgg gaggtctata taagcagagc tcgtttagtg 1680
aaccgtcaga tcgcctggag acgccatcca cgctgttttg acctccatag aagacaccgg 1740
gaccgatcca gcctccgcgg ccgggaacgg tgcattggaa cgcggattcc ccgtgccaag 1800
agtgacgtaa gtaccgccta tagactctat aggcacaccc ctttggctct tatgcatgct 1860
atactgtttt tggcttgggg cctatacacc cccgcttcct tatgctatag gtgatggtat 1920
```

agcttagcct ataggtgtgg gttattgacc attattgacc actcccctat tggtgacgat 1980

actttccatt actaatccat aacatggctc tttgccacaa ctatctctat tggctatatg 2040

ccaatactct gtccttcaga gactgacacg gactctgtat ttttacagga tggggtccca 2100

tttattattt acaaattcac atatacaaca acgccgtccc ccgtgcccgc agtttttatt 2160

aaacatagcg tgggatctcc acgcgaatct cgggtacgtg ttccggacat gggctcttct 2220

ccggtagcgg cggagcttcc acatccgagc cctggtccca tgcctccagc ggctcatggt 2280

cgctcggcag ctccttgctc ctaacagtgg aggccagact taggcacagc acaatgccca 2340

ccaccaccag tgtgccgcac aaggccgtgg cggtagggta tgtgtctgaa aatgagcgtg 2400

gagattgggc tcgcacggct gacgcagatg gaagacttaa ggcagcggca gaagaagatg 2460

caggcagctg agttgttgta ttctgataag agtcagaggt aactcccgtt gcggtgctgt 2520

taacggtgga gggcagtgta gtctgagcag tactcgttgc tgccgcgcgc gccaccagac 2580

ataatagctg acagactaac agactgttcc tttccatggg tcttttctgc agtcaccgtc 2640

gtcgacacgt gtgatcagat atcgcggccg ctctaggaag ctttccatgg aagacgccaa 2700

aaacataaag aaaggcccgg cgccattcta tccgctggaa gatggaaccg ctggagagca 2760

[0204] actgcataag gctatgaaga gatacgccct ggttcctgga acaattgctt ttacagatgc 2820

acatatcgag gtggacatca cttacgctga gtacttcgaa atgtccgttc ggttggcaga 2880

agctatgaaa cgatatgggc tgaatacaaa tcacagaatc gtcgtatgca gtgaaaactc 2940

tcttcaattc tttatgccgg tgttgggcgc gttatttatc ggagttgcag ttgcgcccgc 3000

gaacgacatt tataatgaac gtgaattgct caacagtatg ggcatttcgc agcctaccgt 3060

ggtgttcgtt tccaaaaagg ggttgcaaaa aattttgaac gtgcaaaaaa agctcccaat 3120

catccaaaaa attattatca tggattctaa aacggattac cagggatttc agtcgatgta 3180

cacgttcgtc acatctcatc tacctcccgg ttttaatgaa tacgattttg tgccagagtc 3240

cttcgatagg gacaagacaa ttgcactgat catgaactcc tctggatcta ctggtctgcc 3300

taaaggtgtc gctctgcctc atagaactgc ctgcgtgaga ttctcgcatg ccagagatcc 3360

tatttttggc aatcaaatca ttccggatac tgcgatttta agtgttgttc cattccatca 3420

cggttttgga atgtttacta cactcggata tttgatatgt ggatttcgag tcgtcttaat 3480

gtatagattt gaagaagagc tgtttctgag gagccttcag gattacaaga ttcaaagtgc 3540

gctgctggtg ccaaccctat tctccttctt cgccaaaagc actctgattg acaaatacga 3600

tttatctaat ttacacgaaa ttgcttctgg tggcgctccc ctctctaagg aagtcgggga 3660

agcggttgcc aagaggttcc atctgccagg tatcaggcaa ggatatgggc tcactgagac 3720

tacatcagct attctgatta cacccgaggg ggatgataaa ccgggcgcgg tcggtaaagt 3780

tgttccattt tttgaagcga aggttgtgga tctggatacc gggaaaacgc tgggcgttaa 3840

tcaaagaggc gaactgtgtg tgagaggtcc tatgattatg tccggttatg taaacaatcc 3900

ggaagcgacc aacgccttga ttgacaagga tggatggcta cattctggag acatagctta 3960

ctgggacgaa gacgaacact tcttcatcgt tgaccgcctg aagtctctga ttaagtacaa 4020

aggctatcag gtggctcccg ctgaattgga atccatcttg ctccaacacc ccaacatctt 4080

cgacgcaggt gtcgcaggtc ttcccgacga tgacgccggt gaacttcccg ccgccgttgt 4140

tgttttggag cacggaaaga cgatgacgga aaaagagatc gtggattacg tcgccagtca 4200

agtaacaacc gcgaaaaagt tgcgcggagg agttgtgttt gtggacgaag taccgaaagg 4260

tcttaccgga aaactcgacg caagaaaaat cagagagatc ctcataaagg ccaagaaggg 4320

cggaaagatc gccgtgtaat tctagaccag gccctggatc cagatcactt ctggctaata 4380

aaagatcaga gctctagaga tctgtgtgtt ggtttttttgt ggatctgctg tgccttctag 4440

ttgccagcca tctgttgttt gcccctcccc cgtgccttcc ttgaccctgg aaggtgccac 4500

[0205] tcccactgtc ctttcctaat aaaatgagga aattgcatcg cattgtctga gtaggtgtca 4560

ttctattctg gggggtgggg tggggcagga cagcaagggg gaggattggg aagacaatag 4620

caggcatgct ggggatgcgg tgggctctat gggtacctct ctctctctct ctctctctct 4680

ctctctctct ctctctctgg tacctctctc tctctctctc tctctctctc tctctctctc 4740

tctggtaccc aggtgctgaa gaattgaccc ggttcctcct gggccagaaa gaagcaggca 4800

catccccttc tctgtgacac accctgtcca cgcccctggt tcttagttcc agccccactc 4860

ataggacact catagctcag gagggctccg ccttcaatcc cacccgctaa agtacttgga 4920

gcggtctctc cctccctcat cagcccacca aaccaaacct agcctccaag agtgggaaga 4980

aattaaagca agataggcta ttaagtgcag agggagagaa aatgcctcca acatgtgagg 5040

aagtaatgag agaaatcata gaatttcttc 5070

<210> 18
<211> 938
<212> DNA
<213> 人工序列

<220>
<223> β-半乳糖苷酶表达盒/pUC57 复制起点

<400> 18
agcgggcagt gagcgcaacg caattaatgt gagttagctc actcattagg caccccaggc 60

tttacacttt atgcttccgg ctcgtatgtt gtgtggaatt gtgagcggat aacaatttca 120

cacaggaaac agctatgacc atgattacgg attcactggc cgtcgtttta caacgtcgtg 180

actgggaaaa ccctggcgtt acccaactta atcgccttgc agcacatccc cctttcgcca 240

gctggcgtaa tagcgaagag gcccgcaccg atcgcccttc ccaacagttg cgcagcctga 300

atggcgaatg gcgctgaaag cttaaaggat cttcttgaga tccttttttt ctgcgcgtaa 360

tctgctgctt gcaaacaaaa aaaccaccgc taccagcggt ggtttgtttg ccggatcaag 420

agctaccaac tctttttccg aaggtaactg gcttcagcag agcgcagata ccaaatactg 480

ttcttctagt gtagccgtag ttaggccacc acttcaagaa ctctgtagca ccgcctacat 540

acctcgctct gctaatcctg ttaccagtgg ctgctgccag tggcgataag tcgtgtctta 600

ccgggttgga ctcaagacga tagttaccgg ataaggcgca gcggtcgggc tgaacggggg 660

gttcgtgcac acagcccagc ttggagcgaa cgacctacac cgaactgaga tacctacagc 720

gtgagctatg agaaagcgcc acgcttcccg aagggagaaa ggcggacagg tatccggtaa 780

[0206] gcggcagggt cggaacagga gagcgcacga gggagcttcc agggggaaac gcctggtatc 840

tttatagtcc tgtcgggttt cgccacctct gacttgagcg tcgatttttg tgatgctcgt 900

caggggggcg gagcctatgg aaaaacgcca gcaacgcg 938

<210> 19
<211> 615
<212> DNA
<213> 人工序列

<220>
<223> pUC57 复制起点

<400> 19
aaaggatctt cttgagatcc ttttttttctg cgcgtaatct gctgcttgca aacaaaaaaa 60

ccaccgctac cagcggtggt ttgtttgccg gatcaagagc taccaactct ttttccgaag 120

gtaactggct tcagcagagc gcagatacca aatactgttc ttctagtgta gccgtagtta 180

ggccaccact tcaagaactc tgtagcaccg cctacatacc tcgctctgct aatcctgtta 240

ccagtggctg ctgccagtgg cgataagtcg tgtcttaccg ggttggactc aagacgatag 300

ttaccggata aggcgcagcg gtcgggctga acgggggggtt cgtgcacaca gcccagcttg 360

gagcgaacga cctacaccga actgagatac ctacagcgtg agctatgaga aagcgccacg 420

cttcccgaag ggagaaaggc ggacaggtat ccggtaagcg gcagggtcgg aacaggagag 480

cgcacgaggg agcttccagg gggaaacgcc tggtatcttt atagtcctgt cgggtttcgc 540

cacctctgac ttgagcgtcg atttttgtga tgctcgtcag gggggcggag cctatggaaa 600

aacgccagca acgcg 615

<210> 20
<211> 237
<212> DNA
[0207] <213> 人工序列

<220>
<223> ColE1 二聚体解离元件

<400> 20
gaaaccatga aaaatggcag cttcagtgga ttaagtgggg gtaatgtggc ctgtaccctc 60

tggttgcata ggtattcata cggttaaaat ttatcaggcg cgatcgcgca gtttttaggg 120

tggtttgttg ccatttttac ctgtctgctg ccgtgatcgc gctgaacgcg ttttagcggt 180

gcgtacaatt aagggattat ggtaaatcca cttactgtct gccctcgtag ccatcga 237

序列表
<110> Janssen Biotech, Inc.

<120> 作为非抗生素选择标记的β-半乳糖苷酶α肽及其用途
<130> 688097-553US
<160> 20
<170> PatentIn 版本 3.5
<210> 1
<211> 60
<212> PRT
<213> 人工序列
<220>
<223> 截短的 LazCα肽
<400> 1

```
Met Thr Met Ile Thr Asp Ser Leu Ala Val Val Leu Gln Arg Arg Asp
1               5                  10                  15
Trp Glu Asn Pro Gly Val Thr Gln Leu Asn Arg Leu Ala Ala His Pro
            20                  25                  30
Pro Phe Ala Ser Trp Arg Asn Ser Glu Glu Ala Arg Thr Asp Arg Pro
        35                  40                  45
Ser Gln Gln Leu Arg Ser Leu Asn Gly Glu Trp Arg
    50                  55                  60
```

<210> 2
<211> 419
<212> DNA
<213> 人工序列
<220>
[0001] <223> LacZα盒 1
<400> 2

```
agcgggcagt gagcgcaacg caattaatgt gagttagctc actcattagg caccccaggc   60
tttacacttt atgcttccgg ctcgtatgtt gtgtggaatt gtgagcggat aacaatttca  120
cacaggaaac agctatgacc atgattacgg attcactggc cgtcgtttta caacgtcgtg  180
actgggaaaa ccctggcgtt acccaactta atcgccttgc agcacatccc cctttcgcca  240
gctggcgtaa tagcgaagag gcccgcaccg atcgcccttc ccaacagttg cgcagcctga  300
atggcgaatg gcgctgaggc ccggagggtg gcgggcagga cgcccgccat aaactgccag  360
gcatcaaatt aagcagaagg ccatcctgac ggatggcctt tttgcgtttc tacaaactc   419
```

<210> 3
<211> 540
<212> DNA
<213> 人工序列
<220>
<223> LacZα盒 2
<400> 3

```
cacgtctcta tggaaatatg acggtgttca caaagttcct taaattttac ttttggttac   60
atatttttc ttttgaaac caaatcttta tctttgtagc actttcacgg tagcgaaacg  120
ttagtttgaa tggaaagatg cctgcagaca cataaagaca ccaaactctc atcaatagtt  180
ccgtaaattt ttattgacag aacttattga cggcagtggc aggtgtcata aaaaaaacca  240
tgagggtaat aaataatgac catgattacg gattcactgg ccgtcgtttt acaacgtcgt  300
gactgggaaa accctggcgt tacccaactt aatcgccttg cagcacatcc cctttcgcc   360
agctggcgta atagcgaaga ggcccgcacc gatcgccctt cccaacagtt gcgcagcctg  420
aatggcgaat ggcgctgagg cccggagggt ggcgggcagg acgcccgcca taaactgcca  480
ggcatcaaat taagcagaag gccatcctga cggatggcct ttttgcgttt ctacaaactc  540
```

<210> 4
<211> 96
<212> DNA
<213> 人工序列

<220>
<223> LacZYA 启动子
<400> 4
agcgggcagt gagcgcaacg caattaatgt gagttagctc actcattagg caccccaggc 60
tttacacttt atgcttccgg ctcgtatgtt gtgtgg 96
<210> 5
<211> 38
<212> DNA
<213> 人工序列
<220>
<223> 乳糖操纵子
<400> 5
aattgtgagc ggataacaat ttcacacagg aaacagct 38
<210> 6
<211> 183
<212> DNA
<213> 人工序列
<220>
<223> 截短的 LacZα 肽核苷酸序列
<400> 6
atgaccatga ttacggattc actggccgtc gttttacaac gtcgtgactg ggaaaaccct 60
ggcgttaccc aacttaatcg ccttgcagca catcccccctt tcgccagctg gcgtaatagc 120
gaagaggccc gcaccgatcg cccttcccaa cagttgcgca gcctgaatgg cgaatggcgc 180
tga 183
<210> 7
<211> 102
<212> DNA
[0002] <213> 人工序列
<220>
<223> rrnBT2 转录终止子
<400> 7
ggcccggagg gtggcgggca ggacgcccgc cataaactgc caggcatcaa attaagcaga 60
aggccatcct gacggatggc ctttttgcgt ttctacaaac tc 102
<210> 8
<211> 255
<212> DNA
<213> 人工序列
<220>
<223> OmpF 启动子
<400> 8
cacgtctcta tggaaatatg acggtgttca caaagttcct taaattttac ttttggttac 60
atatttttc tttttgaaac caaatcttta tctttgtagc actttcacgg tagcgaaacg 120
ttagtttgaa tggaaagatg cctgcagaca cataaagaca ccaaactctc atcaatagtt 180
ccgtaaattt ttattgacag aacttattga cggcagtggc aggtgtcata aaaaaaacca 240
tgagggtaat aaata 255
<210> 9
<211> 7222
<212> DNA
<213> 人工序列
<220>
<223> P215
<400> 9
taactataac ggtcctaagg tagcgaagct cttcagatgg acagtcagac tgaagagcct 60
ctcttaaggt agctcgagga gcttggccca ttgcatacgt tgtatccata tcataatatg 120
tacatttata ttggctcatg tccaacatta ccgccatgtt gacattgatt attgactagt 180
tattaatagt aatcaattac ggggtcatta gttcatagcc catatatgga gttccgcgtt 240

acataactta cggtaaatgg cccgcctggc tgaccgccca acgacccccg cccattgacg 300
tcaataatga cgtatgttcc catagtaacg ccaataggga ctttccattg acgtcaatgg 360
gtggagtatt tacggtaaac tgcccacttg gcagtacatc aagtgtatca tatgccaagt 420
acgcccccta ttgacgtcaa tgacggtaaa tggcccgcct ggcattatgc ccagtacatg 480
accttatggg actttcctac ttggcagtac atctacgtat tagtcatcgc tattaccatg 540
gtgatgcggt tttggcagta catcaatggg cgtggatagc ggtttgactc acggggattt 600
ccaagtctcc accccattga cgtcaatggg agtttgtttt ggcaccaaaa tcaacgggac 660
tttccaaaat gtcgtaacaa ctccgcccca ttgacgcaaa tgggcggtag gcgtgtacgg 720
tgggaggtct atataagcag agctcgttta gtgaaccgtc ggcgcgccgc caccatggtg 780
agcaagggcg aggaggataa catggccatc atcaaggagt tcatgcgctt caaggtgcac 840
atggagggct ccgtgaacgg ccacgagttc gagatcgagg gcgagggcga gggccgcccc 900
tacgagggca cccagaccgc caagctgaag gtgaccaagg gtggcccccc gcccttcgcc 960
tgggacatcc tgtcccctca gttcatgtac ggctccaagg cctacgtgaa gcaccccgcc 1020
gacatccccg actacttgaa gctgtccttc cccgagggct tcaagtggga gcgcgtgatg 1080
aacttcgagg acggcggcgt ggtgaccgtg acccaggact cctccctgca ggacggcgag 1140
ttcatctaca aggtgaagct gcgcggcacc aacttccctc ccgacggccc cgtaatgcag 1200
aagaagacca tgggctggga ggcctcctcc gagcggatgt accccgagga cggcgccctg 1260
aagggcgaga tcaagcagag gctgaagctg aaggacggcg gccactacga cgctgaggtc 1320
aagaccacct acaaggccaa gaagcccgtg cagctgcccg gcgcctacaa cgtcaacatc 1380
aagttggaca tcacctccca caacgaggac tacaccatcg tggaacagta cgaacgcgcc 1440
gagggccgcc actccaccgg cggcatggac gagctgtaca agtagtctag agatacattg 1500
atgagtttgg acaaaccaca actagaatgc agtgaaaaaa atgctttatt tgtgaaattt 1560
gtgatgctat tgctttattt gtaaccatta taagctgcaa taaacaagtt aacaacaaca 1620
attgcattca ttttatgttt caggttcagg gggaggtgtg ggaggttttt taaagcaagt 1680
aaaacctcta caaatgtggt atggctgatt atgatcgcgg ccgcgttcca tgtccttata 1740
tggactcatc tttgcctatt gcgacacaca ctcagtgaac acctactacg cgctgcaaag 1800
agccccgcag gcctgaggtg cccccacctc accactcttc ctatttttgt gtaaaaatcc 1860
agcttcttgt caccacctcc aaggaggggg aggaggagga aggcaggttc ctctaggctg 1920
[0003] agccgaatgc ccctctgtgg tcccacgcca ctgatcgctg catgcccacc acctgggtac 1980
acacagtctg tgattcccgg agcagaacgg accctgccca cccggtcttg tgtgctactc 2040
agtggacaga cccaaggcaa gaaagggtga caaggacagg gtcttcccag gctggctttg 2100
agttcctagc accgccccgc ccccaatcct ctgtggcaca tggagtcttg gtccccagag 2160
tcccccagcg gcctccagat ggtctgggag ggcagttcag ctgtggctgc gcatagcaga 2220
catacaacgg acggtgggcc cagacccagg ctgtgtagac ccagcccccc cgccccgcag 2280
tgcctaggtc acccactaac gccccaggcc ttgtcttggc tgggcgtgac tgttaccctc 2340
aaaagcaggc agctccaggg taaaaggtgc cctgccctgt agagcccacc ttccttccca 2400
gggctgcggc tgggtaggtt tgtagccttc atcacgggcc acctccagcc actggaccgc 2460
tggcccctgc cctgtcctgg ggagtgtggt cctgcgactt ctaagtggcc gcaagccacc 2520
tgactccccc aacaccacac tctacctctc aagcccaggt ctctccctag tgacccaccc 2580
agcacattta gctagctgag ccccacagcc agaggtcctc aggccctgct ttcagggcag 2640
ttgctctgaa gtcggcaagg gggagtgact gcctggccac tccatgccct ccaagagctt 2700
cttctgcagg agcgtacaga acccagggcc ctggcacccg tgcagaccct ggcccacccc 2760
acctgggcgc tcagtgccca agagatgtcc acacctagga tgtcccgcgg tgggtggggg 2820
gcccgagaga cgggcaggcc gggggcaggc ctggccatgc ggggccgaac cgggcactgc 2880
ccagcgtggg gcgcgggggc cacggcgcgc gcccccagcc cccgggccca gcaccccaag 2940
gcggccaacg ccaaaactct ccctcctcct cttcctcaat ctcgctctcg ctctttttt 3000
ttttcgcaaa aggaggggag agggggtaaa aaaatgctgc actgtgcggc gaagccggtg 3060
agtgagcggc gcggggccaa tcagcgtgcg ccgttccgaa agttgccttt tatggctcga 3120
gtggccgcgg cggcgcccta taaaacccag cggcgcgacg cgccaccacc gccgagaccg 3180
cgtccgcccc gcgagcacag agcctcgcct ttgccgatcc gccgcccgtc cacacccgcc 3240
gccaggtaag cccggccagc cgaccggggc aggcggctca cggcccggcc gcaggaggcc 3300
gcggcccctt cgcccgtgca gagccgccgt ctgggccgca gcgggggggcg catgggggg 3360
gaaccggacc gccgtggggg gcgcgggaga agcccctggg cctccggaga tgggggacac 3420
cccacgccag ttcggaggcg cgaggccgcg ctcgggaggc gcgctccggg ggtgccgctc 3480
tcggggcggg ggcaaccggc ggggtctttg tctgagccgg gctcttgcca atggggatcg 3540
cagggtgggc gcggcggagc ccccgccagg cccggtgggg gctggggcgc cattgcgcgt 3600
gcgcgctggt cctttgggcg ctaactgcgt gcgcgctggg aattggcgct aattgcgcgt 3660

gcgcgctggg actcaaggcg ctaactgcgc gtgcgttctg gggcccgggg tgccgcggcc 3720
tgggctgggg cgaaggcggg ctcggccgga aggggtgggg tcgccgcggc tcccgggcgc 3780
ttgcgcgcac ttcctgcccg agccgctggc cgcccgaggg tgtggccgct gcgtgcgcgc 3840
gcgccgaccc ggcgctgttt gaaccgggcg gaggcggggc tggcgcccgg ttgggagggg 3900
gttggggcct ggcttcctgc cgcgcgccgc ggggacgcct ccgaccagtg tttgcctttt 3960
atggtaataa cgcggccggc ccggcttcct ttgtccccaa tctgggcgcg cgccggcgcc 4020
ccctggcggc ctaaggactc ggctcgccgg aagtggccag ggcgggggcg acctcggctc 4080
acagcgcgcc cggctattct cgcagctcgc caccatgacc gagtacaagc ccacggtgcg 4140
cctcgccacc cgcgacgacg tcccccgggc cgtacgcacc ctcgccgccg cgttcgccga 4200
ctaccccgcc acgcgccaca ccgttgaccc ggaccgccac atcgagcggg tcaccgagct 4260
gcaagaactc ttcctcacgc gcgtcgggct cgacatcggc aaggtgtggg tcgcggacga 4320
cggcgccgcg gtggcggtct ggaccacgcc ggagagcgtc gaagcggggg cggtgttcgc 4380
cgagatcggc ccgcgcatgg ccgagttgag cggttcccgg ctggccgcgc agcaacagat 4440
ggaaggcctc ctggcgccgc accggcccaa ggagcccgcg tggttcctgg ccaccgtcgg 4500
cgtctcgccc gaccaccagg gcaagggtct gggcagcgcc gtcgtgctcc ccggagtgga 4560
ggcggccgag cgcgccgggg tgcccgcctt cctggagacc tccgcgcccc gcaacctccc 4620
cttctacgag cggctcggct tcaccgtcac cgccgacgtc gaggtgcccg aaggaccgcg 4680
cacctggtgc atgacccgca agcccggtgc ctgatgtgcc ttctagttgc cagccatctg 4740
ttgtttgccc ctccccgtg ccttccttga ccctggaagg tgccactccc actgtccttt 4800
cctaataaaa tgaggaaatt gcatcgcatt gtctgagtag gtgtcattct attctggggg 4860
gtggggtggg gcaggacagc aagggggagg attgggaaga caatagcagg catgctgggg 4920
atgcggtggg ctctatggta gggataacag ggtaatagcg ggcagtgagc gcaacgcaat 4980
taatgtgagt tagctcactc attaggcacc ccaggcttta cactttatgc ttccggctcg 5040
tatgttgtgt ggaattgtga gcggataaca atttcacaca ggaaacagct atgaccatga 5100
ttacggattc actggccgtc gttttacaac gtcgtgactg ggaaaaccct ggcgttaccc 5160
aacttaatcg ccttgcagca catccccctt tcgccagctg gcgtaatagc gaagaggccc 5220
gcaccgatcg cccttcccaa cagttgcgca gcctgaatgg cgaatggcgc tgaggcccgg 5280
agggtggcgg gcaggacgcc cgccataaac tgccaggcat caaattaagc agaaggccat 5340
[0004] cctgacggat ggccttttg cgtttctaca aactctggca aacagctatt atgggtatta 5400
tgggtgacgt caggtggcac ttttcgggga aatgtgcgcg gaacccctat ttgtttattt 5460
ttctaaatac attcaaatat gtatccgctc atgagacaat aaccctgata aatgcttcaa 5520
taatattgaa aaaggaagag tatgagtatt caacatttcc gtgtcgccct tattcccttt 5580
tttgcggcat tttgccttcc tgtttttgct cacccagaaa cgctggtgaa agtaaaagat 5640
gctgaagatc agttgggtgc acgagtgggt tacatcgaac tggatctcaa cagcggtaag 5700
atccttgaga gttttcgccc cgaagaacgt tttccaatga tgagcacttt taaagttctg 5760
ctatgtggcg cggtattatc ccgtattgac gccgggcaag agcaactcgg tcgccgcata 5820
cactattctc agaatgactt ggttgagtac tcaccagtca cagaaaagca tcttacggat 5880
ggcatgacag taagagaatt atgcagtgct gccataacca tgagtgataa cactgcggcc 5940
aacttacttc tgacaacgat cggaggaccg aaggagctaa ccgcttttt gcacaacatg 6000
ggggatcatg taactcgcct tgatcgttgg gaaccggagc tgaatgaagc cataccaaac 6060
gacgagcgtg acaccacgat gcctgtagca atggcaacaa cgttgcgcaa actattaact 6120
ggcgaactac ttactctagc ttcccggcaa caattaatag actggatgga ggcggataaa 6180
gttgcaggac cacttctgcg ctcggccctt ccggctggct ggtttattgc tgataaatct 6240
ggagccggtg agcgtgggtc tcgcggtatc attgcagcac tggggccaga tggtaagccc 6300
tcccgtatcg tagttatcta cacgacgggg agtcaggcaa ctatggatga acgaaataga 6360
cagatcgctg agataggtgc ctcactgatt aagcattggt aactgtcaga ccaagtttac 6420
tcatatatac tttagattga tttaaaactt catttttaat ttaaaaggat ctaggtgaag 6480
atccttttg ataatctcat gaccaaaatc ccttaacgtg agttttcgtt ccactgagcg 6540
tcagaccccg tagaaaagat caaaggatct tcttgagatc ctttttttct gcgcgtaatc 6600
tgctgcttgc aaacaaaaaa accaccgcta ccagcggtgg tttgtttgcc ggatcaagag 6660
ctaccaactc tttttccgaa ggtaactggc ttcagcagag cgcagatacc aaatactgtt 6720
cttctagtgt agccgtagtt aggccaccac ttcaagaact ctgtagcacc gcctacatac 6780
ctcgctctgc taatcctgtt accagtggct gctgccagtg gcgataagtc gtgtcttacc 6840
gggttggact caagacgata gttaccggat aaggcgcagc ggtcgggctg aacgggggt 6900
tcgtgcacac agcccagctt ggagcgaacg acctacaccg aactgagata cctacagcgt 6960
gagctatgag aaagcgccac gcttcccgaa gggagaaagg cggacaggta tccggtaagc 7020
ggcagggtcg gaacaggaga gcgcacgagg gagcttccag ggggaaacgc ctggtatctt 7080

tatagtcctg tcgggtttcg ccacctctga cttgagcgtc gatttttgtg atgctcgtca 7140
gggggggcgga gcctatggaa aaacgccagc aacgcggcct ttttacggtt cctggccttt 7200
tgctggcctt ttgctcacat gt 7222

<210> 10
<211> 7343
<212> DNA
<213> 人工序列
<220>
<223> P216
<400> 10

taactataac ggtcctaagg tagcgaagct cttcagatgg acagtcagac tgaagagcct 60
ctcttaaggt agctcgagga gcttggccca ttgcatacgt tgtatccata tcataatatg 120
tacatttata ttggctcatg tccaacatta ccgccatgtt gacattgatt attgactagt 180
tattaatagt aatcaattac ggggtcatta gttcatagcc catatatgga gttccgcgtt 240
acataactta cggtaaatgg cccgcctggc tgaccgccca acgacccccg cccattgacg 300
tcaataatga cgtatgttcc catagtaacg ccaatagggа ctttccattg acgtcaatgg 360
gtggagtatt tacggtaaac tgcccacttg gcagtacatc aagtgtatca tatgccaagt 420
acgccccta ttgacgtcaa tgacggtaaa tggcccgcct ggcattatgc ccagtacatg 480
accttatggg actttcctac ttggcagtac atctacgtat tagtcatcgc tattaccatg 540
gtgatgcggt tttggcagta catcaatggg cgtggatagc ggtttgactc acggggattt 600
ccaagtctcc accccattga cgtcaatggg agtttgtttt ggcaccaaaa tcaacgggac 660
tttccaaaat gtcgtaacaa ctccgcccca ttgacgcaaa tgggcggtag gcgtgtacgg 720
tgggaggtct atataagcag agctcgttta gtgaaccgtc ggcgcgccgc caccatggtg 780
agcaagggcg aggaggataa catggccatc atcaaggagt tcatgcgctt caaggtgcac 840
atggagggct ccgtgaacgg ccacgagttc gagatcgagg gcgagggcga gggccgcccc 900
tacgagggca cccagaccgc caagctgaag gtgaccaagg gtggcccccт gcccttcgcc 960
tgggacatcc tgtcccctca gttcatgtac ggctccaagg cctacgtgaa gcaccccgcc 1020
gacatccccg actacttgaa gctgtccttc cccgagggct tcaagtggga gcgcgtgatg 1080
[0005] aacttcgagg acggcggcgt ggtgaccgtg acccaggact cctccctgca ggacggcgag 1140
ttcatctaca aggtgaagct gcgcggcacc aacttccccт ccgacggccc cgtaatgcag 1200
aagaagacca tgggctggga ggcctcctcc gagcggatgt accccgagga cggcgccctg 1260
aagggcgaga tcaagcagag gctgaagctg aaggacggcg gccactacga cgctgaggtc 1320
aagaccacct acaaggccaa gaagcccgtg cagctgcccg gcgcctacaa cgtcaacatc 1380
aagttggaca tcacctccca caacgaggac tacaccatcg tggaacagta cgaacgcgcc 1440
gagggccgcc actccaccgg cggcatggac gagctgtaca agtagtctag agatacattg 1500
atgagtttgg acaaaccaca actagaatgc agtgaaaaaa atgctttatt tgtgaaattt 1560
gtgatgctat tgctttattt gtaaccatta taagctgcaa taaacaagtt aacaacaaca 1620
attgcattca ttttatgttt caggttcagg gggaggtgtg ggaggttttt taaagcaagt 1680
aaaacctcta caaatgtggt atggctgatt atgatcgcgg ccgcgttcca tgtccttata 1740
tggactcatc tttgcctatt gcgacacaca ctcagtgaac acctactacg cgctgcaaag 1800
agccccgcag gcctgaggtg cccccacctc accactcttc ctatttttgt gtaaaaatcc 1860
agcttcttgt caccacctcc aaggaggggg aggaggagga aggcaggttc ctctaggctg 1920
agccgaatgc ccctctgtgg tcccacgcca ctgatcgctg catgcccacc acctgggtac 1980
acacagtctg tgattcccgg agcagaacgg accctgccca cccggtcttg tgtgctactc 2040
agtggacaga cccaaggcaa gaaagggtga caaggacagg gtcttcccag gctggctttg 2100
agttcctagc accgcccccgc ccccaatcct ctgtggcaca tggagtcttg gtccccagag 2160
tcccccagcg gcctccagat ggtctgggag ggcagttcag ctgtggctgc gcatagcaga 2220
catacaacgg acggtgggcc cagacccagg ctgtgtagac ccagcccccc cgccccgcag 2280
tgcctaggtc acccactaac gccccaggcc ttgtcttggc tgggcgtgac tgttaccctc 2340
aaaagcaggc agctccaggg taaaaggtgc cctgccctgt agagcccacc ttccttccca 2400
gggctgcggc tgggtaggtt tgtagccttc atcacgggcc acctccagcc actggaccgc 2460
tggcccctgc cctgtcctgg ggagtgtggt cctgcgactt ctaagtggcc gcaagccacc 2520
tgactccccc aacaccacac tctacctctc aagcccaggt ctctccctag tgacccaccc 2580
agcacattta gctagctgag ccccacagcc agaggtcctc aggccctgct ttcagggcag 2640
ttgctctgaa gtcggcaagg gggagtgact gcctggccac tccatgccct ccaagagctt 2700
cttctgcagg agcgtacaga acccagggcc ctggcacccg tgcagaccct ggcccacccc 2760
acctgggcgc tcagtgccca agagatgtcc acacctagga tgtcccgcgg tgggtggggg 2820

gcccgagaga cgggcaggcc gggggcaggc ctggccatgc ggggccgaac cgggcactgc 2880
ccagcgtggg gcgcgggggc cacggcgcgc gcccccagcc cccgggccca gcaccccaag 2940
gcggccaacg ccaaaactct ccctcctcct cttcctcaat ctcgctctcg ctctttttt 3000
ttttcgcaaa aggaggggag agggggtaaa aaaatgctgc actgtgcggc gaagccggtg 3060
agtgagcggc gcggggccaa tcagcgtgcg ccgttccgaa agttgccttt tatggctcga 3120
gtggccgcgg cggcgcccta taaaacccag cggcgcgacg cgccaccacc gccgagaccg 3180
cgtccgcccc gcgagcacag agcctcgcct ttgccgatcc gccgcccgtc cacacccgcc 3240
gccaggtaag cccggccagc cgaccggggc aggcggctca cggcccggcc gcaggaggcc 3300
gcggcccctt cgcccgtgca gagccgccgt ctgggccgca gcgggggggcg catgggggg 3360
gaaccggacc gccgtggggg gcgcgggaga agcccctggg cctccggaga tgggggacac 3420
cccacgccag ttcggaggcg cgaggccgcg ctcgggaggc gcgctccggg ggtgccgctc 3480
tcggggcggg ggcaaccggc ggggtctttg tctgagccgg gctcttgcca atggggatcg 3540
cagggtgggc gcggcggagc ccccgccagg cccggtgggg gctggggcgc cattgcgcgt 3600
gcgcgctggt cctttgggcg ctaactgcgt gcgcgctggg aattggcgct aattgcgcgt 3660
gcgcgctggg actcaaggcg ctaactgcgc gtgcgttctg gggcccgggg tgccgcggcc 3720
tgggctgggg cgaaggcggg ctcggccgga aggggtgggg tcgccgcggc tcccgggcgc 3780
ttgcgcgcac ttcctgcccg agccgctggc cgcccgaggg tgtggccgct gcgtgcgcgc 3840
gcgccgaccc ggcgctgttt gaaccgggcg gaggcggggc tggcgcccgg ttgggagggg 3900
gttggggcct ggcttcctgc cgcgcgccgc ggggacgcct ccgaccagtg tttgcctttt 3960
atggtaataa cgcggccggc ccggcttcct ttgtccccaa tctgggcgcg cgccggcgcc 4020
ccctggcggc ctaaggactc ggctcgccgg aagtggccag ggcgggggcg acctcggctc 4080
acagcgcgcc cggctattct cgcagctcgc caccatgacc gagtacaagc ccacggtgcg 4140
cctcgccacc cgcgacgacg tcccccgggc cgtacgcacc ctcgccgccg cgttcgccga 4200
ctaccccgcc acgcgccaca ccgttgaccc ggaccgccac atcgagcggg tcaccgagct 4260
gcaagaactc ttcctcacgc gcgtcgggct cgacatcggc aaggtgtggg tcgcggacga 4320
cggcgccgcg gtggcggtct ggaccacgcc ggagagcgtc gaagcggggg cggtgttcgc 4380
cgagatcggc ccgcgcatgg ccgagttgag cggttcccgg ctggccgcgc agcaacagat 4440
ggaaggcctc ctggcgccgc accggcccaa ggagcccgcg tggttcctgg ccaccgtcgg 4500
[0006] cgtctcgccc gaccaccagg gcaagggtct gggcagcgcc gtcgtgctcc ccggagtgga 4560
ggcggccgag cgcgccgggg tgcccgcctt cctggagacc tccgcgcccc gcaacctccc 4620
cttctacgag cggctcggct tcaccgtcac cgccgacgtc gaggtgcccg aaggaccgcg 4680
cacctggtgc atgacccgca agcccggtgc ctgatgtgcc ttctagttgc cagccatctg 4740
ttgtttgccc ctcccccgtg ccttccttga ccctggaagg tgccactccc actgtccttt 4800
cctaataaaa tgaggaaatt gcatcgcatt gtctgagtag gtgtcattct attctggggg 4860
gtggggtggg gcaggacagc aagggggagg attgggaaga caatagcagg catgctgggg 4920
atgcggtggg ctctatggta gggataacag ggtaatcacg tctctatgga aatatgacgg 4980
tgttcacaaa gttccttaaa ttttactttt ggttacatat tttttctttt tgaaaccaaa 5040
tctttatctt tgtagcactt tcacggtagc gaaacgttag tttgaatgga aagatgcctg 5100
cagacacata aagacaccaa actctcatca atagttccgt aaatttttat tgacagaact 5160
tattgacggc agtggcaggt gtcataaaaa aaaccatgag ggtaataaat aatgaccatg 5220
attacggatt cactggccgt cgttttacaa cgtcgtgact gggaaaaccc tggcgttacc 5280
caacttaatc gccttgcagc acatcccccc ttcgccagct ggcgtaatag cgaagaggcc 5340
cgcaccgatc gcccttccca acagttgcgc agcctgaatg gcgaatggcg ctgaggcccg 5400
gagggtggcg ggcaggacgc ccgccataaa ctgccaggca tcaaattaag cagaaggcca 5460
tcctgacgga tggcctttt gcgtttctac aaactctggc aaacagctat tatgggtatt 5520
atgggtgacg tcaggtggca cttttcgggg aaatgtgcgc ggaacccct tttgtttatt 5580
tttctaaata cattcaaata tgtatccgct catgagacaa taaccctgat aaatgcttca 5640
ataatattga aaaaggaaga gtatgagtat tcaacatttc cgtgtcgccc ttattccctt 5700
ttttgcggca ttttgccttc ctgtttttgc tcacccagaa acgctggtga aagtaaaaga 5760
tgctgaagat cagttgggtg cacgagtggg ttacatcgaa ctggatctca acagcggtaa 5820
gatccttgag agttttcgcc ccgaagaacg ttttccaatg atgagcactt ttaaagttct 5880
gctatgtggc gcggtattat cccgtattga cgccgggcaa gagcaactcg gtcgccgcat 5940
acactattct cagaatgact tggttgagta ctcaccagtc acagaaaagc atcttacgga 6000
tggcatgaca gtaagagaat tatgcagtgc tgccataacc atgagtgata acactgcggc 6060
caacttactt ctgacaacga tcggaggacc gaaggagcta accgcttttt tgcacaacat 6120
gggggatcat gtaactcgcc ttgatcgttg ggaaccggag ctgaatgaag ccataccaaa 6180
cgacgagcgt gacaccacga tgcctgtagc aatggcaaca acgttgcgca aactattaac 6240

tggcgaacta cttactctag cttcccggca acaattaata gactggatgg aggcggataa 6300
agttgcagga ccacttctgc gctcggccct tccggctggc tggtttattg ctgataaatc 6360
tggagccggt gagcgtgggt ctcgcggtat cattgcagca ctggggccag atggtaagcc 6420
ctcccgtatc gtagttatct acacgacggg gagtcaggca actatggatg aacgaaatag 6480
acagatcgct gagataggtg cctcactgat taagcattgg taactgtcag accaagttta 6540
ctcatatata ctttagattg atttaaaact tcatttttaa tttaaaagga tctaggtgaa 6600
gatccttttt gataatctca tgaccaaaat cccttaacgt gagttttcgt tccactgagc 6660
gtcagacccc gtagaaaaga tcaaaggatc ttcttgagat cctttttttc tgcgcgtaat 6720
ctgctgcttg caaacaaaaa aaccaccgct accagcggtg gtttgtttgc cggatcaaga 6780
gctaccaact ctttttccga aggtaactgg cttcagcaga gcgcagatac caaatactgt 6840
tcttctagtg tagccgtagt taggccacca cttcaagaac tctgtagcac cgcctacata 6900
cctcgctctg ctaatcctgt taccagtggc tgctgccagt ggcgataagt cgtgtcttac 6960
cgggttggac tcaagacgat agttaccgga taaggcgcag cggtcgggct gaacgggggg 7020
ttcgtgcaca cagcccagct tggagcgaac gacctacacc gaactgagat acctacagcg 7080
tgagctatga gaaagcgcca cgcttcccga agggagaaag gcggacaggt atccggtaag 7140
cggcagggtc ggaacaggag agcgcacgag ggagcttcca gggggaaacg cctggtatct 7200
ttatagtcct gtcgggtttc gccacctctg acttgagcgt cgatttttgt gatgctcgtc 7260
aggggggcgg agcctatgga aaaacgccag caacgcggcc tttttacggt tcctggcctt 7320
ttgctggcct tttgctcaca tgt 7343

<210> 11
<211> 2329
<212> DNA
<213> 人工序列
<220>
<223> P217
<400> 11

agcgggcagt gagcgcaacg caattaatgt gagttagctc actcattagg caccccaggc 60
tttacacttt atgcttccgg ctcgtatgtt gtgtggaatt gtgagcggat aacaatttca 120
[0007] cacaggaaac agctatgacc atgattacgg attcactggc cgtcgtttta caacgtcgtg 180
actgggaaaa ccctggcgtt acccaactta atcgccttgc agcacatccc cctttcgcca 240
gctggcgtaa tagcgaagag gcccgcaccg atcgcccttc ccaacagttg cgcagcctga 300
atggcgaatg gcgctgaggc ccggagggtg gcgggcagga cgcccgccat aaactgccag 360
gcatcaaatt aagcagaagg ccatcctgac ggatggcctt tttgcgtttc tacaaactct 420
ggcaaacagc tattatgggt attatgggtg acgtcaggtg gcacttttcg gggaaatgtg 480
cgcggaaccc ctatttgttt atttttctaa atacattcaa atatgtatcc gctcatgaga 540
caataaccct gataaatgct tcaataatat tgaaaaagga agagtatgag tattcaacat 600
ttccgtgtcg cccttattcc ctttttttgcg gcattttgcc ttcctgtttt tgctcaccca 660
gaaacgctgg tgaaagtaaa agatgctgaa gatcagttgg gtgcacgagt gggttacatc 720
gaactggatc tcaacagcgg taagatcctt gagagttttc gccccgaaga acgttttcca 780
atgatgagca cttttaaagt tctgctatgt ggcgcggtat tatcccgtat tgacgccggg 840
caagagcaac tcggtcgccg catacactat tctcagaatg acttggttga gtactcacca 900
gtcacagaaa agcatcttac ggatggcatg acagtaagag aattatgcag tgctgccata 960
accatgagtg ataacactgc ggccaactta cttctgacaa cgatcggagg accgaaggag 1020
ctaaccgctt ttttgcacaa catgggggat catgtaactc gccttgatcg ttgggaaccg 1080
gagctgaatg aagccatacc aaacgacgag cgtgacacca cgatgcctgt agcaatggca 1140
acaacgttgc gcaaactatt aactggcgaa ctacttactc tagcttcccg gcaacaatta 1200
atagactgga tggaggcgga taaagttgca ggaccacttc tgcgctcggc ccttccggct 1260
ggctggttta ttgctgataa atctggagcc ggtgagcgtg ggtctcgcgg tatcattgca 1320
gcactggggc cagatggtaa gccctcccgt atcgtagtta tctacacgac ggggagtcag 1380
gcaactatgg atgaacgaaa tagacagatc gctgagatag gtgcctcact gattaagcat 1440
tggtaactgt cagaccaagt ttactcatat atactttaga ttgatttaaa acttcatttt 1500
taatttaaaa ggatctaggt gaagatcctt tttgataatc tcatgaccaa aatcccttaa 1560
cgtgagtttt cgttccactg agcgtcagac cccgtagaaa agatcaaagg atcttcttga 1620
gatccttttt ttctgcgcgt aatctgctgc ttgcaaacaa aaaaaccacc gctaccagcg 1680
gtggtttgtt tgccggatca agagctacca actcttttte cgaaggtaac tggcttcagc 1740
agagcgcaga taccaaatac tgttcttcta gtgtagccgt agttaggcca ccacttcaag 1800
aactctgtag caccgcctac atacctcgct ctgctaatcc tgttaccagt ggctgctgcc 1860

agtggcgata agtcgtgtct taccgggttg gactcaagac gatagttacc ggataaggcg 1920
cagcggtcgg gctgaacggg gggttcgtgc acacagccca gcttggagcg aacgacctac 1980
accgaactga gatacctaca gcgtgagcta tgagaaagcg ccacgcttcc cgaagggaga 2040
aaggcggaca ggtatccggt aagcggcagg gtcggaacag gagagcgcac gagggagctt 2100
ccaggggaa acgcctggta tctttatagt cctgtcgggt ttcgccacct ctgacttgag 2160
cgtcgatttt tgtgatgctc gtcaggggg cggagcctat ggaaaaacgc cagcaacgcg 2220
gccttttac ggttcctggc cttttgctgg ccttttgctc acatgttaac tataacggtc 2280
ctaaggtagc gaagctcggt gggctctatg gtagggataa cagggtaat 2329

<210> 12
<211> 1143
<212> DNA
<213> 人工序列
<220>
<223> P218
<400> 12

agcgggcagt gagcgcaacg caattaatgt gagttagctc actcattagg caccccaggc 60
tttacacttt atgcttccgg ctcgtatgtt gtgtggaatt gtgagcggat aacaatttca 120
cacaggaaac agctatgacc atgattacgg attcactggc cgtcgtttta caacgtcgtg 180
actgggaaaa ccctggcgtt acccaactta atcgccttgc agcacatccc cctttcgcca 240
gctggcgtaa tagcgaagag gcccgcaccg atcgcccttc ccaacagttg cgcagcctga 300
atggcgaatg gcgctgaggc ccggagggtg gcgggcagga cgcccgccat aaactgccag 360
gcatcaaatt aagcagaagg ccatcctgac ggatggcctt tttgcgtttc tacaaactca 420
aaggatcttc ttgagatcct ttttttctgc gcgtaatctg ctgcttgcaa acaaaaaaac 480
caccgctacc agcggtggtt tgtttgccgg atcaagagct accaactctt tttccgaagg 540
taactggctt cagcagagcg cagataccaa atactgttct tctagtgtag ccgtagttag 600
gccaccactt caagaactct gtagcaccgc ctacatacct cgctctgcta atcctgttac 660
cagtggctgc tgccagtggc gataagtcgt gtcttaccgg gttggactca agacgatagt 720
taccggataa ggcgcagcgg tcgggctgaa cgggggggttc gtgcacacag cccagcttgg 780
[0008] agcgaacgac ctacaccgaa ctgagatacc tacagcgtga gctatgagaa agcgccacgc 840
ttcccgaagg gagaaaggcg gacaggtatc cggtaagcgg cagggtcgga acaggagagc 900
gcacgaggga gcttccaggg ggaaacgcct ggtatcttta tagtcctgtc gggtttcgcc 960
acctctgact tgagcgtcga tttttgtgat gctcgtcagg ggggcggagc ctatggaaaa 1020
acgccagcaa cgcggccttt ttacggttcc tggccttttg ctggcctttt gctcacatgt 1080
taactataac ggtcctaagg tagcgaagct cggtgggctc tatggtaggg ataacagggt 1140
aat 1143

<210> 13
<211> 1047
<212> DNA
<213> 人工序列
<220>
<223> P219
<400> 13

agcgggcagt gagcgcaacg caattaatgt gagttagctc actcattagg caccccaggc 60
tttacacttt atgcttccgg ctcgtatgtt gtgtggaatt gtgagcggat aacaatttca 120
cacaggaaac agctatgacc atgattacgg attcactggc cgtcgtttta caacgtcgtg 180
actgggaaaa ccctggcgtt acccaactta atcgccttgc agcacatccc cctttcgcca 240
gctggcgtaa tagcgaagag gcccgcaccg atcgcccttc ccaacagttg cgcagcctga 300
atggcgaatg gcgctgaaag cttaaaggat cttcttgaga tccttttttt ctgcgcgtaa 360
tctgctgctt gcaaacaaaa aaaccaccgc taccagcggt ggtttgtttg ccggatcaag 420
agctaccaac tctttttccg aaggtaactg gcttcagcag agcgcagata ccaaatactg 480
ttcttctagt gtagccgtag ttaggccacc acttcaagaa ctctgtagca ccgcctacat 540
acctcgctct gctaatcctg ttaccagtgg ctgctgccag tggcgataag tcgtgtctta 600
ccgggttgga ctcaagacga tagttaccgg ataaggcgca gcggtcgggc tgaacggggg 660
gttcgtgcac acagcccagc ttggagcgaa cgacctacac cgaactgaga tacctacagc 720
gtgagctatg agaaagcgcc acgcttcccg aagggagaaa ggcggacagg tatccggtaa 780
gcggcagggt cggaacagga gagcgcacga gggagcttcc agggggaaac gcctggtatc 840
tttatagtcc tgtcgggttt cgccacctct gacttgagcg tcgatttttg tgatgctcgt 900

cagggggcg gagcctatgg aaaaacgcca gcaacgcggc ctttttacgg ttcctggcct 960
tttgctggcc ttttgctcac atgttaacta taacggtcct aaggtagcga agctcggtgg 1020
gctctatggt agggataaca gggtaat 1047
<210> 14
<211> 25
<212> DNA
<213> 人工序列
<220>
<223> CNFOR
<400> 14
tgtgtggaat tgtgagcgga taaca 25
<210> 15
<211> 27
<212> DNA
<213> 人工序列
<220>
<223> P455R2
<400> 15
tggcgttact atgggaacat acgtcat 27
<210> 16
<211> 6732
<212> DNA
<213> 人工序列
<220>
<223> GWIZ 荧光素酶
<400> 16

[0009]

tcgcgcgttt cggtgatgac ggtgaaaacc tctgacacat gcagctcccg gagacggtca 60
cagcttgtct gtaagcggat gccgggagca gacaagcccg tcagggcgcg tcagcgggtg 120
ttggcgggtg tcggggctgg cttaactatg cggcatcaga gcagattgta ctgagagtgc 180
accatatgcg gtgtgaaata ccgcacagat gcgtaaggag aaaataccgc atcagattgg 240
ctattggcca ttgcatacgt tgtatccata tcataatatg tacatttata ttggctcatg 300
tccaacatta ccgccatgtt gacattgatt attgactagt tattaatagt aatcaattac 360
ggggtcatta gttcatagcc catatatgga gttccgcgtt acataactta cggtaaatgg 420
cccgcctggc tgaccgccca acgacccccg cccattgacg tcaataatga cgtatgttcc 480
catagtaacg ccaataggga ctttccattg acgtcaatgg gtggagtatt tacggtaaac 540
tgcccacttg gcagtacatc aagtgtatca tatgccaagt acgccccta ttgacgtcaa 600
tgacggtaaa tggcccgcct ggcattatgc ccagtacatg accttatggg actttcctac 660
ttggcagtac atctacgtat tagtcatcgc tattaccatg gtgatgcggt tttggcagta 720
catcaatggg cgtggatagc ggtttgactc acggggattt ccaagtctcc accccattga 780
cgtcaatggg agtttgtttt ggcaccaaaa tcaacgggac tttccaaaat gtcgtaacaa 840
ctccgcccca ttgacgcaaa tgggcggtag gcgtgtacgg tgggaggtct atataagcag 900
agctcgttta gtgaaccgtc agatcgcctg gagacgccat ccacgctgtt ttgacctcca 960
tagaagacac cgggaccgat ccagcctccg cggccgggaa cggtgcattg gaacgcggat 1020
tccccgtgcc aagagtgacg taagtaccgc ctatagactc tataggcaca ccccttggc 1080
tcttatgcat gctatactgt ttttggcttg gggcctatac accccgctt ccttatgcta 1140
taggtgatgg tatagcttag cctataggtg tgggttattg accattattg accactcccc 1200
tattggtgac gatactttcc attactaatc cataacatgg ctctttgcca caactatctc 1260
tattggctat atgccaatac tctgtccttc agagactgac acggactctg tatttttaca 1320
ggatggggtc ccatttatta tttacaaatt cacatataca acaacgccgt cccccgtgcc 1380
cgcagttttt attaaacata gcgtgggatc tccacgcgaa tctcgggtac gtgttccgga 1440
catgggctct tctccggtag cggcggagct tccacatccg agccctggtc ccatgcctcc 1500
agcggctcat ggtcgctcgg cagctccttg ctcctaacag tggaggccag acttaggcac 1560
agcacaatgc ccaccaccac cagtgtgccg cacaaggccg tggcggtagg gtatgtgtct 1620
gaaaatgagc gtggagattg ggctcgcacg gctgacgcag atggaagact taaggcagcg 1680
gcagaagaag atgcaggcag ctgagttgtt gtattctgat aagagtcaga ggtaactccc 1740
gttgcggtgc tgttaacggt ggagggcagt gtagtctgag cagtactcgt tgctgccgcg 1800

cgcgccacca gacataatag ctgacagact aacagactgt tcctttccat gggtcttttc 1860
tgcagtcacc gtcgtcgaca cgtgtgatca gatatcgcgg ccgctctagg aagctttcca 1920
tggaagacgc caaaaacata aagaaaggcc cggcgccatt ctatccgctg gaagatggaa 1980
ccgctggaga gcaactgcat aaggctatga agagatacgc cctggttcct ggaacaattg 2040
cttttacaga tgcacatatc gaggtggaca tcacttacgc tgagtacttc gaaatgtccg 2100
ttcggttggc agaagctatg aaacgatatg ggctgaatac aaatcacaga atcgtcgtat 2160
gcagtgaaaa ctctcttcaa ttctttatgc cggtgttggg cgcgttattt atcggagttg 2220
cagttgcgcc cgcgaacgac atttataatg aacgtgaatt gctcaacagt atgggcattt 2280
cgcagcctac cgtggtgttc gtttccaaaa aggggttgca aaaaattttg aacgtgcaaa 2340
aaaagctccc aatcatccaa aaaattatta tcatggattc taaaacggat taccagggat 2400
ttcagtcgat gtacacgttc gtcacatctc atctacctcc cggttttaat gaatacgatt 2460
ttgtgccaga gtccttcgat agggacaaga caattgcact gatcatgaac tcctctggat 2520
ctactggtct gcctaaaggt gtcgctctgc ctcatagaac tgcctgcgtg agattctcgc 2580
atgccagaga tcctattttt ggcaatcaaa tcattccgga tactgcgatt ttaagtgttg 2640
ttccattcca tcacggtttt ggaatgttta ctacactcgg atatttgata tgtggatttc 2700
gagtcgtctt aatgtataga tttgaagaag agctgtttct gaggagcctt caggattaca 2760
agattcaaag tgcgctgctg gtgccaaccc tattctcctt cttcgccaaa agcactctga 2820
ttgacaaata cgatttatct aatttacacg aaattgcttc tggtggcgct cccctctcta 2880
aggaagtcgg ggaagcggtt gccaagaggt tccatctgcc aggtatcagg caaggatatg 2940
ggctcactga gactacatca gctattctga ttacacccga gggggatgat aaaccgggcg 3000
cggtcggtaa agttgttcca ttttttgaag cgaaggttgt ggatctggat accgggaaaa 3060
cgctgggcgt taatcaaaga ggcgaactgt gtgtgagagg tcctatgatt atgtccggtt 3120
atgtaaacaa tccggaagcg accaacgcct tgattgacaa ggatggatgg ctacattctg 3180
gagacatagc ttactgggac gaagacgaac acttcttcat cgttgaccgc ctgaagtctc 3240
tgattaagta caaaggctat caggtggctc ccgctgaatt ggaatccatc ttgctccaac 3300
accccaacat cttcgacgca ggtgtcgcag gtcttcccga cgatgacgcc ggtgaacttc 3360
ccgccgccgt tgttgttttg gagcacggaa agacgatgac ggaaaaagag atcgtggatt 3420
acgtcgccag tcaagtaaca accgcgaaaa agttgcgcgg aggagttgtg tttgtggacg 3480
[0010] aagtaccgaa aggtcttacc ggaaaactcg acgcaagaaa aatcagagag atcctcataa 3540
aggccaagaa gggcggaaag atcgccgtgt aattctagac caggcgcctg gatccagatc 3600
acttctggct aataaaagat cagagctcta gagatctgtg tgttggtttt ttgtggatct 3660
gctgtgcctt ctagttgcca gccatctgtt gtttgcccct cccccgtgcc ttccttgacc 3720
ctggaaggtg ccactcccac tgtcctttcc taataaaatg aggaaattgc atcgcattgt 3780
ctgagtaggt gtcattctat tctgggggggt ggggtggggc aggacagcaa gggggaggat 3840
tgggaagaca atagcaggca tgctggggat gcggtgggct ctatgggtac ctctctctct 3900
ctctctctct ctctctctct ctctctctct cggtacctct ctctctctct ctctctctct 3960
ctctctctct ctctctcggt accaggtgct gaagaattga cccggttcct cctgggccag 4020
aaagaagcag gcacatcccc ttctctgtga cacaccctgt ccacgcccct ggttcttagt 4080
tccagcccca ctcataggac actcatagct caggagggct ccgccttcaa tcccacccgc 4140
taaagtactt ggagcggtct ctccctccct catcagccca ccaaaccaaa cctagcctcc 4200
aagagtggga agaaattaaa gcaagatagg ctattaagtg cagagggaga gaaaatgcct 4260
ccaacatgtg aggaagtaat gagagaaatc atagaatttc ttccgcttcc tcgctcactg 4320
actcgctgcg ctcggtcgtt cggctgcggc gagcggtatc agctcactca aaggcggtaa 4380
tacggttatc cacagaatca ggggataacg caggaaagaa catgtgagca aaaggccagc 4440
aaaaggccag gaaccgtaaa aaggccgcgt tgctggcgtt tttccatagg ctccgccccc 4500
ctgacgagca tcacaaaaat cgacgctcaa gtcagaggtg gcgaaacccg acaggactat 4560
aaagatacca ggcgtttccc cctggaagct ccctcgtgcg ctctcctgtt ccgaccctgc 4620
cgcttaccgg atacctgtcc gcctttctcc cttcgggaag cgtggcgctt tctcaatgct 4680
cacgctgtag gtatctcagt tcggtgtagg tcgttcgctc caagctgggc tgtgtgcacg 4740
aaccccccgt tcagcccgac cgctgcgcct tatccggtaa ctatcgtctt gagtccaacc 4800
cggtaagaca cgacttatcg ccactggcag cagccactgg taacaggatt agcagagcga 4860
ggtatgtagg cggtgctaca gagttcttga agtggtggcc taactacggc tacactagaa 4920
ggacagtatt tggtatctgc gctctgctga agccagttac cttcggaaaa agagttggta 4980
gctcttgatc cggcaaacaa accaccgctg gtagcggtgg ttttttgtt tgcaagcagc 5040
agattacgcg cagaaaaaaa ggatctcaag aagatccttt gatcttttct acggggtctg 5100
acgctcagtg gaacgaaaac tcacgttaag ggattttggt catgagatta tcaaaaagga 5160
tcttcaccta gatcctttta aattaaaaat gaagttttaa atcaatctaa agtatatatg 5220

agtaaacttg gtctgacagt taccaatgct taatcagtga ggcacctatc tcagcgatct 5280
gtctatttcg ttcatccata gttgcctgac tccgggggggg gggggcgctg aggtctgcct 5340
cgtgaagaag gtgttgctga ctcataccag gcctgaatcg ccccatcatc cagccagaaa 5400
gtgagggagc cacggttgat gagagctttg ttgtaggtgg accagttggt gattttgaac 5460
ttttgctttg ccacggaacg gtctgcgttg tcgggaagat gcgtgatctg atccttcaac 5520
tcagcaaaag ttcgatttat tcaacaaagc cgccgtcccg tcaagtcagc gtaatgctct 5580
gccagtgtta caaccaatta accaattctg attagaaaaa ctcatcgagc atcaaatgaa 5640
actgcaattt attcatatca ggattatcaa taccatattt ttgaaaaagc cgtttctgta 5700
atgaaggaga aaactcaccg aggcagttcc ataggatggc aagatcctgg tatcggtctg 5760
cgattccgac tcgtccaaca tcaatacaac ctattaattt cccctcgtca aaaataaggt 5820
tatcaagtga gaaatcacca tgagtgacga ctgaatccgg tgagaatggc aaaagcttat 5880
gcatttcttt ccagacttgt tcaacaggcc agccattacg ctcgtcatca aaatcactcg 5940
catcaaccaa accgttattc attcgtgatt gcgcctgagc gagacgaaat acgcgatcgc 6000
tgttaaaagg acaattacaa acaggaatcg aatgcaaccg gcgcaggaac actgccagcg 6060
catcaacaat attttcacct gaatcaggat attcttctaa tacctggaat gctgttttcc 6120
cggggatcgc agtggtgagt aaccatgcat catcaggagt acggataaaa tgcttgatgg 6180
tcggaagagg cataaattcc gtcagccagt ttagtctgac catctcatct gtaacatcat 6240
tggcaacgct acctttgcca tgtttcagaa acaactctgg cgcatcgggc ttcccataca 6300
atcgatagat tgtcgcacct gattgcccga cattatcgcg agcccattta tacccatata 6360
aatcagcatc catgttggaa tttaatcgcg gcctcgagca agacgtttcc cgttgaatat 6420
ggctcataac accccttgta ttactgttta tgtaagcaga cagttttatt gttcatgatg 6480
atatattttt atcttgtgca atgtaacatc agagattttg agacacaacg tggctttccc 6540
cccccccca ttattgaagc atttatcagg gttattgtct catgagcgga tacatatttg 6600
aatgtattta gaaaaataaa caaatagggg ttccgcgcac atttccccga aaagtgccac 6660
ctgacgtcta agaaaccatt attatcatga cattaaccta taaaaatagg cgtatcacga 6720
ggccctttcg tc 6732

<210> 17
<211> 5070
[0011] <212> DNA
<213> 人工序列
<220>
<223> P469-2
<400> 17

tagggataac agggtaatag cgggcagtga gcgcaacgca attaatgtga gttagctcac 60
tcattaggca ccccaggctt tacactttat gcttccggct cgtatgttgt gtggaattgt 120
gagcggataa caatttcaca caggaaacag ctatgaccat gattacggat tcactggccg 180
tcgttttaca acgtcgtgac tgggaaaacc ctggcgttac ccaacttaat cgccttgcag 240
cacatccccc tttcgccagc tggcgtaata gcgaagaggc ccgcaccgat cgcccttccc 300
aacagttgcg cagcctgaat ggcgaatggc gctgaaagct taaaggatct tcttgagatc 360
ctttttttct gcgcgtaatc tgctgcttgc aaacaaaaaa accaccgcta ccagcggtgg 420
tttgtttgcc ggatcaagag ctaccaactc tttttccgaa ggtaactggc ttcagcagag 480
cgcagatacc aaatactgtt cttctagtgt agccgtagtt aggccaccac ttcaagaact 540
ctgtagcacc gcctacatac ctcgctctgc taatcctgtt accagtggct gctgccagtg 600
gcgataagtc gtgtcttacc gggttggact caagacgata gttaccggat aaggcgcagc 660
ggtcgggctg aacgggggggt tcgtgcacac agcccagctt ggagcgaacg acctacaccg 720
aactgagata cctacagcgt gagctatgag aaagcgccac gcttcccgaa gggagaaagg 780
cggacaggta tccggtaagc ggcagggtcg gaacaggaga gcgcacgagg gagcttccag 840
ggggaaacgc ctggtatctt tatagtcctg tcgggtttcg ccacctctga cttgagcgtc 900
gatttttgtg atgctcgtca ggggggcgga gcctatggaa aaacgccagc aacgcgggtg 960
cgcataatgt atattatgtt aaattaacta taacggtcct aaggtagcga atggccattg 1020
catacgttgt atccatatca taatatgtac atttatattg gctcatgtcc aacattaccg 1080
ccatgttgac attgattatt gactagttat taatagtaat caattacggg gtcattagtt 1140
catagcccat atatggagtt ccgcgttaca taacttacgg taaatggccc gcctggctga 1200
ccgcccaacg acccccgccc attgacgtca ataatgacgt atgttcccat agtaacgcca 1260
atagggactt tccattgacg tcaatgggtg gagtatttac ggtaaactgc ccacttggca 1320
gtacatcaag tgtatcatat gccaagtacg ccccctattg acgtcaatga cggtaaatgg 1380
cccgcctggc attatgccca gtacatgacc ttatgggact ttcctacttg gcagtacatc 1440

tacgtattag tcatcgctat taccatggtg atgcggtttt ggcagtacat caatgggcgt 1500
ggatagcggt ttgactcacg gggatttcca agtctccacc ccattgacgt caatgggagt 1560
ttgttttggc accaaaatca acgggacttt ccaaaatgtc gtaacaactc cgccccattg 1620
acgcaaatgg gcggtaggcg tgtacggtgg gaggtctata taagcagagc tcgtttagtg 1680
aaccgtcaga tcgcctggag acgccatcca cgctgttttg acctccatag aagacaccgg 1740
gaccgatcca gcctccgcgg ccgggaacgg tgcattggaa cgcggattcc ccgtgccaag 1800
agtgacgtaa gtaccgccta tagactctat aggcacaccc ctttggctct tatgcatgct 1860
atactgtttt tggcttgggg cctatacacc cccgcttcct tatgctatag gtgatggtat 1920
agcttagcct ataggtgtgg gttattgacc attattgacc actcccctat tggtgacgat 1980
actttccatt actaatccat aacatggctc tttgccacaa ctatctctat tggctatatg 2040
ccaatactct gtccttcaga gactgacacg gactctgtat ttttacagga tggggtccca 2100
tttattattt acaaattcac atatacaaca acgccgtccc ccgtgcccgc agtttttatt 2160
aaacatagcg tgggatctcc acgcgaatct cgggtacgtg ttccggacat gggctcttct 2220
ccggtagcgg cggagcttcc acatccgagc cctggtccca tgcctccagc ggctcatggt 2280
cgctcggcag ctccttgctc ctaacagtgg aggccagact taggcacagc acaatgccca 2340
ccaccaccag tgtgccgcac aaggccgtgg cggtagggta tgtgtctgaa aatgagcgtg 2400
gagattgggc tcgcacggct gacgcagatg gaagacttaa ggcagcggca gaagaagatg 2460
caggcagctg agttgttgta ttctgataag agtcagaggt aactcccgtt gcggtgctgt 2520
taacggtgga gggcagtgta gtctgagcag tactcgttgc tgccgcgcgc gccaccagac 2580
ataatagctg acagactaac agactgttcc tttccatggg tcttttctgc agtcaccgtc 2640
gtcgacacgt gtgatcagat atcgcggccg ctctaggaag ctttccatgg aagacgccaa 2700
aaacataaag aaaggcccgg cgccattcta tccgctggaa gatggaaccg ctggagagca 2760
actgcataag gctatgaaga gatacgccct ggttcctgga acaattgctt ttacagatgc 2820
acatatcgag gtggacatca cttacgctga gtacttcgaa atgtccgttc ggttggcaga 2880
agctatgaaa cgatatgggc tgaatacaaa tcacagaatc gtcgtatgca gtgaaaactc 2940
tcttcaattc tttatgccgg tgttgggcgc gttatttatc ggagttgcag ttgcgcccgc 3000
gaacgacatt tataatgaac gtgaattgct caacagtatg ggcatttcgc agcctaccgt 3060
ggtgttcgtt tccaaaaagg ggttgcaaaa aattttgaac gtgcaaaaaa agctcccaat 3120
[0012] catccaaaaa attattatca tggattctaa aacggattac cagggatttc agtcgatgta 3180
cacgttcgtc acatctcatc tacctcccgg ttttaatgaa tacgattttg tgccagagtc 3240
cttcgatagg gacaagacaa ttgcactgat catgaactcc tctggatcta ctggtctgcc 3300
taaaggtgtc gctctgcctc atagaactgc ctgcgtgaga ttctcgcatg ccagagatcc 3360
tatttttggc aatcaaatca ttccggatac tgcgatttta agtgttgttc cattccatca 3420
cggttttgga atgtttacta cactcggata tttgatatgt ggatttcgag tcgtcttaat 3480
gtatagattt gaagaagagc tgtttctgag gagccttcag gattacaaga ttcaaagtgc 3540
gctgctggtg ccaaccctat tctccttctt cgccaaaagc actctgattg acaaatacga 3600
tttatctaat ttacacgaaa ttgcttctgg tggcgctccc ctctctaagg aagtcgggga 3660
agcggttgcc aagaggttcc atctgccagg tatcaggcaa ggatatgggc tcactgagac 3720
tacatcagct attctgatta cacccgaggg ggatgataaa ccgggcgcgg tcggtaaagt 3780
tgttccattt tttgaagcga aggttgtgga tctggatacc gggaaaacgc tgggcgttaa 3840
tcaaagaggc gaactgtgtg tgagaggtcc tatgattatg tccggttatg taaacaatcc 3900
ggaagcgacc aacgccttga ttgacaagga tggatggcta cattctggag acatagctta 3960
ctgggacgaa gacgaacact tcttcatcgt tgaccgcctg aagtctctga ttaagtacaa 4020
aggctatcag gtggctcccg ctgaattgga atccatcttg ctccaacacc ccaacatctt 4080
cgacgcaggt gtcgcaggtc ttcccgacga tgacgccggt gaacttcccg ccgccgttgt 4140
tgttttggag cacggaaaga cgatgacgga aaaagagatc gtggattacg tcgccagtca 4200
agtaacaacc gcgaaaaagt tgcgcggagg agttgtgttt gtggacgaag taccgaaagg 4260
tcttaccgga aaactcgacg caagaaaaat cagagagatc ctcataaagg ccaagaaggg 4320
cggaaagatc gccgtgtaat tctagaccag gccctggatc cagatcactt ctggctaata 4380
aaagatcaga gctctagaga tctgtgtgtt ggtttttttgt ggatctgctg tgccttctag 4440
ttgccagcca tctgttgttt gcccctcccc cgtgccttcc ttgaccctgg aaggtgccac 4500
tcccactgtc ctttcctaat aaaatgagga aattgcatcg cattgtctga gtaggtgtca 4560
ttctattctg gggggtgggg tggggcagga cagcaagggg gaggattggg aagacaatag 4620
caggcatgct ggggatgcgg tgggctctat gggtacctct ctctctctct ctctctctct 4680
ctctctctct ctctctctgg tacctctctc tctctctctc tctctctctc tctctctctc 4740
tctggtaccc aggtgctgaa gaattgaccc ggttcctcct gggccagaaa gaagcaggca 4800
catcccccttc tctgtgacac accctgtcca cgcccctggt tcttagttcc agccccactc 4860

ataggacact catagctcag gagggctccg ccttcaatcc cacccgctaa agtacttgga 4920
gcggtctctc cctccctcat cagcccacca aaccaaacct agcctccaag agtgggaaga 4980
aattaaagca agataggcta ttaagtgcag agggagagaa aatgcctcca acatgtgagg 5040
aagtaatgag agaaatcata gaatttcttc 5070

<210> 18
<211> 938
<212> DNA
<213> 人工序列
<220>
<223> β-半乳糖苷酶表达盒/pUC57 复制起点
<400> 18
agcgggcagt gagcgcaacg caattaatgt gagttagctc actcattagg caccccaggc 60
tttacacttt atgcttccgg ctcgtatgtt gtgtggaatt gtgagcggat aacaatttca 120
cacaggaaac agctatgacc atgattacgg attcactggc cgtcgtttta caacgtcgtg 180
actgggaaaa ccctggcgtt acccaactta atcgccttgc agcacatccc cctttcgcca 240
gctggcgtaa tagcgaagag gcccgcaccg atcgcccttc ccaacagttg cgcagcctga 300
atggcgaatg gcgctgaaag cttaaaggat cttcttgaga tccttttttt ctgcgcgtaa 360
tctgctgctt gcaaacaaaa aaaccaccgc taccagcggt ggtttgtttg ccggatcaag 420
agctaccaac tctttttccg aaggtaactg gcttcagcag agcgcagata ccaaatactg 480
ttcttctagt gtagccgtag ttaggccacc acttcaagaa ctctgtagca ccgcctacat 540
acctcgctct gctaatcctg ttaccagtgg ctgctgccag tggcgataag tcgtgtctta 600
ccgggttgga ctcaagacga tagttaccgg ataaggcgca gcggtcgggc tgaacggggg 660
gttcgtgcac acagcccagc ttggagcgaa cgacctacac cgaactgaga tacctacagc 720
gtgagctatg agaaagcgcc acgcttcccg aagggagaaa ggcggacagg tatccggtaa 780
gcggcagggt cggaacagga gagcgcacga gggagcttcc agggggaaac gcctggtatc 840
tttatagtcc tgtcgggttt cgccacctct gacttgagcg tcgatttttg tgatgctcgt 900
caggggggcg gagcctatgg aaaaacgcca gcaacgcg 938

[0013]

<210> 19
<211> 615
<212> DNA
<213> 人工序列
<220>
<223> pUC57 复制起点
<400> 19
aaaggatctt cttgagatcc ttttttctg cgcgtaatct gctgcttgca aacaaaaaaa 60
ccaccgctac cagcggtggt ttgtttgccg gatcaagagc taccaactct ttttccgaag 120
gtaactggct tcagcagagc gcagatacca aatactgttc ttctagtgta gccgtagtta 180
ggccaccact tcaagaactc tgtagcaccg cctacatacc tcgctctgct aatcctgtta 240
ccagtggctg ctgccagtgg cgataagtcg tgtcttaccg ggttggactc aagacgatag 300
ttaccggata aggcgcagcg gtcgggctga acgggggggtt cgtgcacaca gcccagcttg 360
gagcgaacga cctacaccga actgagatac ctacagcgtg agctatgaga aagcgccacg 420
cttcccgaag ggagaaaggc ggacaggtat ccggtaagcg gcagggtcgg aacaggagag 480
cgcacgaggg agcttccagg gggaaacgcc tggtatcttt atagtcctgt cgggtttcgc 540
cacctctgac ttgagcgtcg atttttgtga tgctcgtcag gggggcggag cctatggaaa 600
aacgccagca acgcg 615

<210> 20
<211> 237
<212> DNA
<213> 人工序列
<220>
<223> ColE1 二聚体解离元件
<400> 20
gaaaccatga aaaatggcag cttcagtgga ttaagtgggg gtaatgtggc ctgtaccctc 60
tggttgcata ggtattcata cggttaaaat ttatcaggcg cgatcgcgca gtttttaggg 120
tggtttgttg ccatttttac ctgtctgctg ccgtgatcgc gctgaacgcg ttttagcggt 180
gcgtacaatt aagggattat ggtaaatcca cttactgtct gccctcgtag ccatcga 237

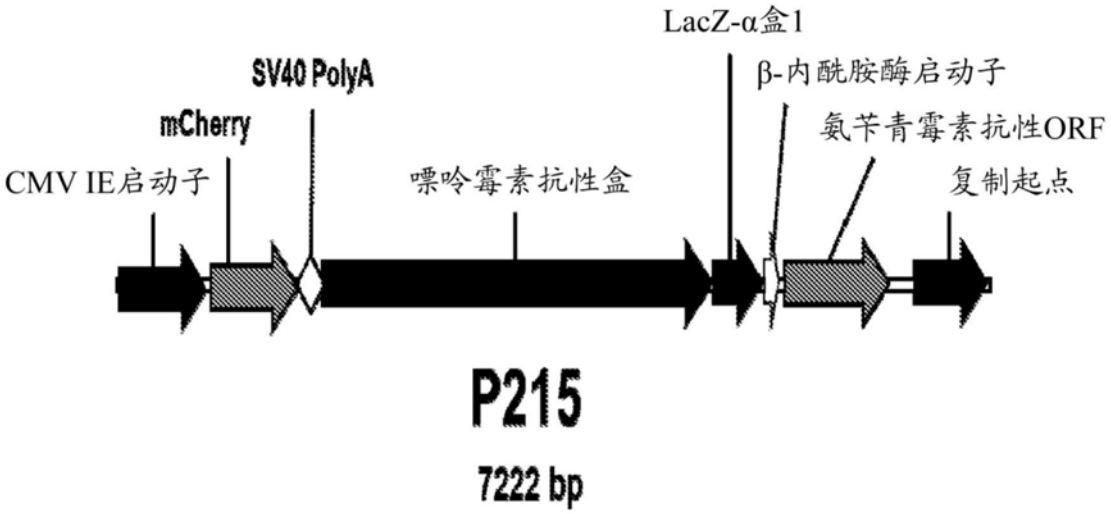

LacZ-α盒1
SV40 PolyA
β-内酰胺酶启动子
mCherry
氨苄青霉素抗性ORF
CMV IE启动子
噻吟霉素抗性盒
复制起点
P215
7222 bp

图1

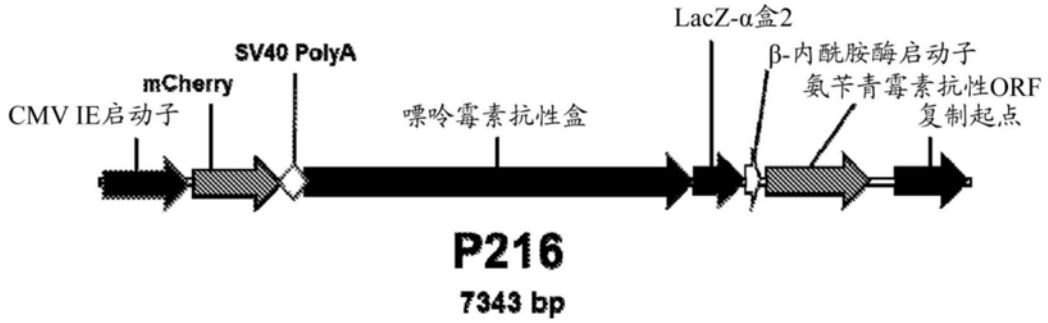

LacZ-α盒2
SV40 PolyA
β-内酰胺酶启动子
mCherry
氨苄青霉素抗性ORF
CMV IE启动子
噻吟霉素抗性盒
复制起点
P216
7343 bp

图2

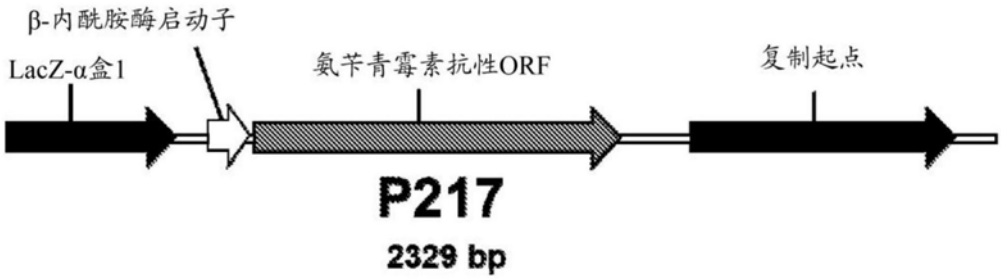

β-内酰胺酶启动子
LacZ-α盒1
氨苄青霉素抗性ORF
复制起点
P217
2329 bp

图3

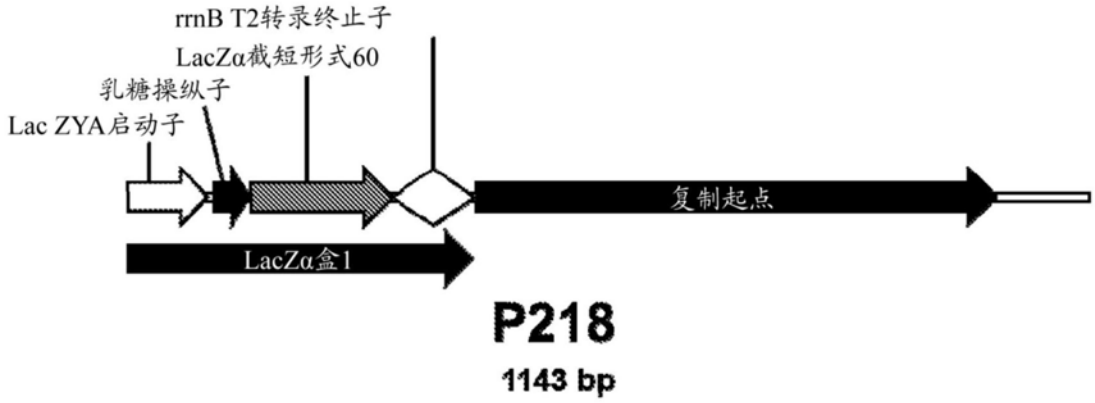

rrnB T2转录终止子
LacZα截短形式60
乳糖操纵子
Lac ZYA启动子
复制起点
LacZα盒1
P218
1143 bp

图4

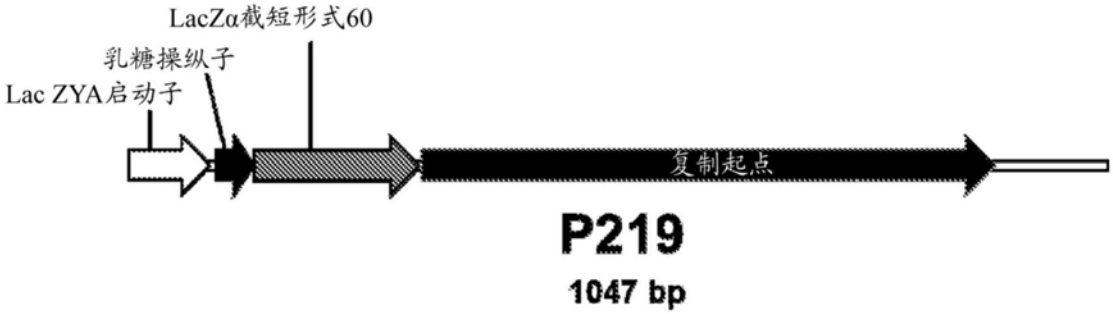

LacZα截短形式60
乳糖操纵子
Lac ZYA启动子
复制起点
P219
1047 bp

图5

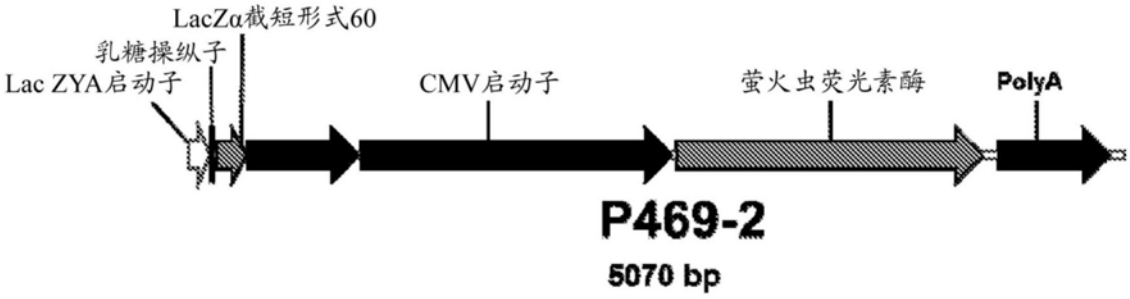

LacZα截短形式60
乳糖操纵子
Lac ZYA启动子
CMV启动子
萤火虫荧光素酶
PolyA
P469-2
5070 bp

图6