

用于非B型DNA序列的克隆/表达载体

相关申请的交叉引用

本申请要求2023年8月29日提交的申请号为202311099203.6的中国专利申请的优先权，其全部内容通过引入并入本文。

发明领域

本申请涉及一种可以用于例如可形成非B型DNA构型的序列的克隆载体或表达载体，以及这些载体的构建和应用。

背景技术

mRNA疫苗是通过转录外源目的基因、或直接合成等工艺而制备的核酸制剂，可通过特定的递送系统导入机体细胞并表达目的蛋白，从而刺激机体产生特异性免疫学反应，使机体获得免疫保护。

成熟mRNA的主体序列是编码区，在其上游5'侧和下游3'侧都有非编码区UTR。真核生物mRNA分子的两端还有5'帽子和3'的poly(A)尾结构，其中poly(A)尾结构对维持mRNA的稳定性、调节mRNA的翻译效率以及维持mRNA的运输发挥了重要作用。针对mRNA的批量合成，目前较为高效的方式为体外转录（IVT），而IVT中对于poly(A)加尾主要有两种方法。第一种为酶法合成，在以线性DNA模板转录出mRNA后，通过添加poly(A)聚合酶，催化由ATP转化成的AMP添加到单链RNA的3'末端，形成Poly(A)尾。该方法的优点是，模板DNA上无需包含Poly(A)序列，操作简单，但存在加尾长度不稳定的局限性。第二种方法为共转录，即由已经存在于模板质粒DNA或者PCR产物上的poly(A)尾编码序列直接转录形成，其优势在于无需使用Poly(A)聚合酶，可减少工艺步骤，节约成本，并且转录加尾长度相对稳定。但是在上游质粒的制备过程中，需要减少培养过程中poly(A)尾编码序列的缺失，这是提供高产优质且均一性好的质粒模板的关键所在。

质粒的规模化生产通常以大肠杆菌发酵的方式进行。在扩增过程中，携带较长poly(A)尾编码序列的质粒本身不太稳定，在复制过程中poly(A)尾编码序列容易缺失。为提高质粒上poly(A)尾编码序列的稳定性，一些科学家采用分段式的poly(A)尾序列，由于每个poly(A)结合蛋白PABP只结合约30个A，两个结合蛋白之间有少量A充当了间隔序列，将间隔序列用其他非A碱基替代，从而降低poly(A)尾编码序列发生重组的概率。此外，可以对间隔序列的长度和碱基进行优化，进一步降低poly(A)尾缺失的风险。但即便如此，A缺失的风险依然存在，并且分段式制备poly(A)尾的质粒会对下游蛋白表达量产生一定的负面影响。此外，另有报道通过低温

(30℃) 发酵的方式, 可改善质粒Poly(A)尾编码序列的完整度。但低温培养细菌生长变慢, 质粒拷贝数降低, 产量明显减少, 周期延长, 对质粒生产造成困难。目前常见的用于mRNA疫苗生产的商业化载体如pVAX1, 在实际生产中, 也存在质粒拷贝数低、产量低和质粒Poly(A)尾序列不纯的现象。

因此, 开发一些更稳定的poly(A)尾克隆载体, 在不影响质粒产量的同时, 减少poly(A)尾序列缺失或不纯的风险, 或将成为新的研究方向。但目前关于载体及其元件对poly(A)尾序列稳定性影响的研究鲜有报道。

类似地, 腺相关病毒(AAV)的反向末端重复序列(ITR)在载体中也存在扩增不稳定的问题。ITR存在于AAV的基因组两端, 各自形成特殊的自我互补T型回文发夹结构, 且GC含量高, 在细菌繁殖过程中容易发生突变或重组丢失。而ITR完整度对于重组AAV(rAAV)高效复制和包装具有很高的重要性, 缺失ITR最好小于20% (Flotte, T. R., and B. J. Carter. (1995) Adeno-associated virus vectors for gene therapy, *Gene therapy* 2(6): 357-362)。单分子实时测序, 将rAAV基因组作为单个完整的分子进行综合分析, 发现HEK293三质粒包装系统中, 缺失的ITR与非载体序列的包装存在相关性。ITR缺失对于rAAV病毒包装及rAAV应用于基因治疗的影响, 情形复杂, 通过载体-宿主相互作用会产生意想不到的生物学后果。如果在临床前阶段使用不完整的ITR序列, 将降低实验内部和跨实验室实验的可重复性, 并可能导致在使用完整ITR序列进行临床试验时的实验重复性, 延迟转化为临床应用。

目前没有较好的方法来避免在rAAV-ITR质粒构建和大肠杆菌扩增过程中发生ITR缺失。一些重组缺陷的商业菌株如JC8111、SURE2、Stabl3、XL10-Gold经常用于含ITR结构的质粒的构建扩增, 但并不总是有效。

对于本申请中任何文件的引用, 并不等同于承认这些文件是本申请的现有技术。

发明内容

本申请的发明人, 经实验发现, 通过i) 延长载体中复制起点与poly(A)尾编码序列、ITR序列等容易在复制中产生突变和缺失的序列的间隔距离; ii) 在poly(A)尾编码序列、ITR等容易在复制中产生突变和缺失的序列, 或其表达盒的上下游添加终止子; 和/或iii) 破坏载体中原有的lac启动子功能区, 可以显著提高这些易缺失或突变序列的扩增和/或表达(转录)稳定性。此外, 上述对于载体的改动还能够提高载体的产量, 特别是提高重组腺相关病毒(rAAV)的产量(如包装滴度)。

因而, 在第一个方面, 本申请提供一种载体, 其可以包含:

i) 复制起点, 以及

ii) a) 目标DNA表达盒, 其包含启动子和目标DNA序列, 或b) 目标DNA序列。

a) 的目标DNA表达盒中的目标DNA序列可以包含能够形成非B型DNA构型的DNA序列。a) 的目标DNA表达盒中的目标DNA序列可以包含能够形成非B型DNA

构型的DNA序列、和克隆位点。该克隆位点可以是多克隆位点。

b) 的目标DNA序列可以包含能够形成非B型DNA构型的DNA序列。b) 的目标DNA序列可以包含能够形成非B型DNA构型的DNA序列、和克隆位点。该克隆位点可以是多克隆位点。

能够形成非B型DNA构型的DNA序列可以为正向重复序列、反向重复序列、或单一嘌呤-单一嘧啶序列，包括但不限于，多聚腺嘌呤（poly(A)）、多聚鸟嘌呤（poly(G)）、多聚胞嘧啶（poly(C)）、或多聚胸腺嘧啶（poly(T)）、shRNA编码序列、和ITR序列。正向重复序列可以为poly(A)尾编码序列。单一嘌呤-单一嘧啶序列可以为poly(A)尾编码序列。反向重复序列可以为ITR或其衍生序列。在一些实施方式中，能够形成非B型DNA构型的DNA序列可以为poly(A)编码序列。

复制起点可以在复制方向上与a) 的目标DNA表达盒中的能够形成非B型DNA构型的DNA序列、或b) 的目标DNA序列中的能够形成非B型DNA构型的DNA序列间隔至少150 bp、200 bp、300 bp、500 bp、或700 bp。在一些实施方式中，复制起点可以在复制方向上与a) 的目标DNA表达盒中的能够形成非B型DNA构型的DNA序列、或b) 的目标DNA序列中的能够形成非B型DNA构型的DNA序列间隔至少300 bp。在一些实施方式中，复制起点可以在复制方向上与a) 的目标DNA表达盒中的能够形成非B型DNA构型的DNA序列、或b) 的目标DNA序列中的能够形成非B型DNA构型的DNA序列间隔至少500 bp。在一些实施方式中，复制起点可以在复制方向上与a) 的目标DNA表达盒中的能够形成非B型DNA构型的DNA序列、或b) 的目标DNA序列中的能够形成非B型DNA构型的DNA序列间隔至少700 bp。

复制起点可以使得载体在合适的宿主细胞中复制。合适的宿主细胞可以是原核细胞如大肠杆菌、或真核细胞。复制起点可以选自pMB1复制起点、pBR322复制起点、ColE1复制起点、R6K复制起点、p15A复制起点、pSC101复制起点、pUC复制起点、pET复制起点和这些复制起点的衍生物。这些复制起点衍生物也可以使得载体在合适的宿主细胞中复制。在一些实施方式中，复制起点可以为pUC复制起点。

载体还可以包含第二复制起点，该第二复制起点可以在复制方向上与a) 的目标DNA表达盒中的能够形成非B型DNA构型的DNA序列、或b) 的目标DNA序列中的能够形成非B型DNA构型的DNA序列间隔至少150 bp、200 bp、300 bp、500 bp、或700 bp。该第二复制起点可以使得载体在合适的宿主细胞中复制，合适的宿主细胞可以是原核细胞如大肠杆菌、或真核细胞。

取决于复制起点的种类，载体还可以相应地包含Rep蛋白表达元件，包括Rep蛋白编码序列和相应的启动子。

载体还可以包含第一终止子和第二终止子，第一终止子和第二终止子分别位于a) 的目标DNA表达盒的上游和下游，并设置成使a) 的目标DNA表达盒之外的转录终止于a) 的目标DNA表达盒的上游和下游；或第一终止子和第二终止子分别位于b) 的目标DNA序列的上游和下游，并设置成使b) 的目标DNA序列之外的转

录终止于b)的目标DNA序列的上游和下游。第一终止子可以为正向终止子、双向终止子、或反向终止子。第二终止子可以为正向终止子、双向终止子、或反向终止子。在一些实施方式中,第一终止子和第二终止子可以均为双向终止子。第一终止子和第二终止子可以独立地选自rrnB T1终止子、rrnB T2终止子、lambda t0终止子、TonB终止子、Amp终止子、T7终止子、和CYC1终止子。在一些实施方式中,第一终止子可以为rrnB T1终止子和rrnB T2终止子。在一些实施方式中,第二终止子可以为lambda t0终止子。

载体还可以包含筛选标记。筛选标记可以为抗生素抗性基因,例如氨苄青霉素抗性基因、四环素抗性基因、氯霉素抗性基因、卡那霉素抗性基因、或链霉素抗性基因等。特别地,筛选标记不存在于载体待转化/转染的宿主细胞的基因组中。在一些实施方式中,筛选标记为卡那霉素抗性基因。该载体还可以在筛选标记的上游包含用于该筛选标记的表达的启动子。

载体可以为环状载体。载体可以为非病毒载体。载体可以为质粒载体。

载体中不包含能够引起转录的lac启动子。

a)的目标DNA表达盒可以包含IIS型限制性内切酶(IIS酶)识别位点。b)的目标DNA序列可以包含IIS酶识别位点。除a)的目标DNA表达盒或b)的目标DNA序列外,载体中不包含与a)的目标DNA表达盒和b)的目标DNA序列中相同的IIS酶识别位点。

a)的目标DNA表达盒可以是适于在真核细胞中表达目标DNA序列的表达盒,其可以包含使得目标DNA序列在真核细胞中表达的组件。a)的目标DNA表达盒中包含的启动子可以是使得目标DNA序列在真核细胞中表达的启动子。a)的目标DNA表达盒还可以包含增强子、转录终止子等,特别是使得目标DNA序列在真核细胞中表达的增强子、和转录终止子等。适用于目标DNA序列在真核细胞中表达的转录终止子可以选自SV40 poly(A)、hGH poly(A)、BGH poly(A)、和rbGlob poly(A)。在一些实施方式中,a)的目标DNA表达盒可以是适于在真核细胞中表达mRNA的表达盒,其可以包含任选的增强子、启动子、5'UTR、开放阅读框(ORF)、3'UTR、Poly(A)尾编码序列、和转录终止子。

a)的目标DNA表达盒还可以包含用于载体线性化的酶切位点。

在一些实施方式中,载体可以包含:

i)复制起点,

ii) a)目标DNA表达盒,其包含启动子、和目标DNA序列,其中目标DNA序列包含能够形成非B型DNA构型的DNA序列,或b)目标DNA序列,其包含能够形成非B型DNA构型的DNA序列,以及

iii)筛选标记和控制该筛选标记表达的启动子,

其中复制起点可以在复制方向上与a)的目标DNA表达盒中的能够形成非B型DNA构型的DNA序列、或b)的目标DNA序列中的能够形成非B型DNA构型的DNA

序列间隔至少500 bp，或特别地间隔至少700 bp。

在一些实施方式中，载体可以包含：

i) 复制起点，

ii) 第一终止子，

iii) a) 目标DNA表达盒，其包含启动子、和目标DNA序列，其中目标DNA序列包含能够形成非B型DNA构型的DNA序列，或b) 目标DNA序列，其包含能够形成非B型DNA构型的DNA序列，

iv) 第二终止子，以及

v) 筛选标记和控制该筛选标记表达的启动子，

其中第一终止子和第二终止子可以分别位于a) 的目标DNA表达盒的上游和下游，并设置成使得载体中a) 的目标DNA表达盒之外的转录终止于a) 的目标DNA表达盒的上游和下游，或第一终止子和第二终止子分别位于b) 的目标DNA序列的上游和下游，并设置成使b) 的目标DNA序列之外的转录终止于b) 的目标DNA序列的上游和下游，

其中复制起点可以在复制方向上与a) 的目标DNA表达盒中的能够形成非B型DNA构型的DNA序列、或b) 的目标DNA序列中的能够形成非B型DNA构型的DNA序列间隔至少500 bp、或700 bp。

在一些实施方式中，载体可以按序包含pUC复制起点、rrnB T1终止子和rrnB T2终止子、目标DNA序列、lambda t0终止子、和卡那霉素抗性元件。

在一些实施方式中，载体可以按序包含pUC复制起点、rrnB T1终止子和rrnB T2终止子、CMV增强子、CMV启动子、目标DNA序列、bGH poly(A)信号、lambda t0终止子、和卡那霉素抗性元件。

在一些实施方式中，载体可以按序包含pUC复制起点、rrnB T1终止子和rrnB T2终止子、CMV增强子、CMV启动子、包含poly(A)尾编码序列的mRNA功能区序列、bGH poly(A)信号、lambda t0终止子、和卡那霉素抗性元件。在一些实施方式中，载体可以按序包含pUC复制起点、rrnB T1终止子和rrnB T2终止子、CMV增强子、CMV启动子、5'UTR、开放阅读框(ORF)、3'UTR、Poly(A)尾编码序列、载体线性化位点、bGH poly(A)信号、lambda t0终止子、和卡那霉素抗性元件。

相应地，本申请提供一种载体骨架，其可以包含：

i) 复制起点，以及

ii) a) 克隆位点，b) 克隆位点和能够形成非B型DNA构型的DNA序列，c) DNA表达盒，其包含启动子和克隆位点，或d) DNA表达盒，其包含启动子、克隆位点和能够形成非B型DNA构型的DNA序列。

能够形成非B型DNA构型的DNA序列可以为正向重复序列、反向重复序列、或单一嘌呤-单一嘧啶序列，包括但不限于，多聚腺嘌呤(poly(A))、多聚鸟嘌呤(poly(G))、多聚胞嘧啶(poly(C))、或多聚胸腺嘧啶(poly(T))、shRNA编码序

列、和ITR序列。正向重复序列可以为poly(A)尾编码序列。单一嘌呤-单一嘧啶序列可以为poly(A)尾编码序列。反向重复序列可以为ITR或其衍生序列。在一些实施方式中，d)的DNA表达盒中的能够形成非B型DNA构型的DNA序列可以为poly(A)尾编码序列。

复制起点可以在复制方向上与a)的克隆位点、b)中的能够形成非B型DNA构型的DNA序列、c)DNA表达盒中的克隆位点、或d)DNA表达盒中的能够形成非B型DNA构型的DNA序列间隔至少150 bp、200 bp、300 bp、500 bp、或700 bp。在一些实施方式中，复制起点可以在复制方向上与a)的克隆位点、b)中的能够形成非B型DNA构型的DNA序列、c)DNA表达盒中的克隆位点、或d)DNA表达盒中的能够形成非B型DNA构型的DNA序列间隔至少500 bp。在一些实施方式中，复制起点可以在复制方向上与a)的克隆位点、b)中的能够形成非B型DNA构型的DNA序列、c)DNA表达盒中的克隆位点、或d)DNA表达盒中的能够形成非B型DNA构型的DNA序列间隔至少700 bp。

复制起点可以使得载体在合适的宿主细胞中复制。合适的宿主细胞可以是原核细胞如大肠杆菌、或真核细胞。复制起点可以选自pMB1复制起点、pBR322复制起点、ColE1复制起点、R6K复制起点、p15A复制起点、pSC101复制起点、pUC复制起点、pET复制起点和这些复制起点的衍生物。这些复制起点衍生物也可以使得载体在合适的宿主细胞中复制。在一些实施方式中，复制起点可以为pUC复制起点。

载体骨架还可以包含第二复制起点，该第二复制起点可以在复制方向上与a)的克隆位点、b)中的能够形成非B型DNA构型的DNA序列、c)DNA表达盒中的克隆位点、或d)DNA表达盒中的能够形成非B型DNA构型的DNA序列间隔至少150 bp、200 bp、300 bp、500 bp、或700 bp。该第二复制起点可以使得载体在合适的宿主细胞中复制，合适的宿主细胞可以是原核细胞如大肠杆菌、或真核细胞。

取决于复制起点的种类，载体还可以相应地包含Rep蛋白表达元件，包括Rep蛋白编码序列和相应的启动子。

载体骨架还可以包含第一终止子和第二终止子，第一终止子和第二终止子分别位于a)的克隆位点的上游和下游，并设置成使得载体中a)的克隆位点之外的转录终止于a)的克隆位点的上游和下游；分别位于b)的克隆位点和能够形成非B型DNA构型的DNA序列的上游和下游，并设置成使得载体中b)的克隆位点和能够形成非B型DNA构型的DNA序列之外的转录终止于b)的克隆位点和能够形成非B型DNA构型的DNA序列的上游和下游；分别位于c)的DNA表达盒上游和下游，并设置成使得载体中c)的DNA表达盒的转录终止于c)的DNA表达盒的上游和下游；或分别位于d)的DNA表达盒的上游和下游，并设置成使得载体中d)的DNA表达盒之外的转录终止于d)的DNA表达盒的上游和下游。第一终止子可以为正向终止子、双向终止子、或反向终止子。第二终止子可以为正向终止子、双向终止子、或反向终止子。在一些实施方式中，第一终止子和第二终止子可以均为双向终止子。

第一终止子和第二终止子可以独立地选自rrnB T1终止子、rrnB T2终止子、lambda t0终止子、TonB终止子、Amp终止子、T7终止子、和CYC1终止子。在一些实施方式中，第一终止子可以为rrnB T1终止子和rrnB T2终止子。在一些实施方式中，第二终止子可以为lambda t0终止子。

载体骨架中的a)的克隆位点可以为多克隆位点。载体骨架中的b)的克隆位点可以为多克隆位点。载体骨架中c) DNA表达盒中的克隆位点可以为多克隆位点。载体d) DNA表达盒中的克隆位点可以为多克隆位点。

在载体骨架的a)的克隆位点处可以插入任何序列，特别地为目标DNA序列，或者启动子和目标DNA序列，其中目标DNA序列可以包含能够形成非B型DNA构型的DNA序列。在载体骨架的b)的克隆位点处可以插入任何序列。在载体骨架的c) DNA表达盒中的克隆位点处可以插入任何序列，特别地为目标DNA序列，其中目标DNA序列可以包含能够形成非B型DNA构型的DNA序列。在载体骨架的d) DNA表达盒中的克隆位点处可以插入任何序列。在一些实施方式中，d) DNA表达盒中的能够形成非B型DNA构型的DNA序列可以为poly(A)编码序列，在d) DNA表达盒中的克隆位点处可以插入mRNA编码序列。

载体骨架还可以包含筛选标记。筛选标记可以为抗生素抗性基因，例如氨苄青霉素抗性基因、四环素抗性基因、氯霉素抗性基因、卡那霉素抗性基因、或链霉素抗性基因等。特别地，筛选标记不存在于载体待转化/转染的宿主细胞的基因组中。在一些实施方式中，筛选标记为卡那霉素抗性基因。该载体还可以在筛选标记的上游包含用于该筛选标记的表达的启动子。

载体骨架可以为环状载体骨架。载体可以为非病毒载体骨架。载体可以为质粒载体骨架。

载体骨架中不包含能够引起转录的lac启动子。

待在a)克隆位点中加入的序列、b)中的能够形成非B型DNA构型的DNA序列、待在c) DNA表达盒的克隆位点加入的序列、或在d) DNA表达盒中的能够形成非B型DNA构型的DNA序列可以包含IIS酶识别位点。除上述序列外，载体骨架中不包含与a)~d)中相同的IIS酶识别位点。

c)和d)的DNA表达盒可以是适于在真核细胞中表达DNA序列的表达盒，其可以包含使得在克隆位点加入的DNA序列、或者使得在克隆位点加入的DNA序列和能够形成非B型DNA构型的DNA序列在真核细胞中得以表达的组件。c)和d)的DNA表达盒中包含的启动子可以是使得在克隆位点加入的DNA序列、或者使得在克隆位点加入的DNA序列和能够形成非B型DNA构型的DNA序列在真核细胞中得以表达的启动子。c)和d)的DNA表达盒还可以包含增强子、转录终止子等，特别是使得在克隆位点加入的DNA序列、或者使得在克隆位点加入的DNA序列和能够形成非B型DNA构型的DNA序列在真核细胞中得以表达的增强子、和转录终止子等。适用于DNA序列在真核细胞中表达的转录终止子可以选自SV40 poly(A)、hGH

poly(A)、BGH poly(A)、和rbGlob poly(A)。在一些实施方式中，d) 的DNA表达盒可以是适用于在真核细胞中表达mRNA的表达盒，其中能够形成非B型DNA构型的DNA序列可以为poly(A)尾编码序列，其中DNA表达盒可以包含任选的增强子、启动子、5'UTR、克隆位点、3'UTR、poly(A)尾编码序列、和转录终止子。c) 或d) 的DNA表达盒还可以包含用于载体线性化的酶切位点。

在一些实施方式中，载体骨架可以包含：

i) 复制起点，

ii) a) 克隆位点，b) 克隆位点和能够形成非B型DNA构型的DNA序列，c) DNA表达盒，其包含启动子和克隆位点，或d) DNA表达盒，其包含启动子、克隆位点和能够形成非B型DNA构型的DNA序列，以及

iii) 筛选标记和控制该筛选标记表达的启动子，

其中复制起点可以在复制方向上与a) 的克隆位点、b) 的能够形成非B型DNA构型的DNA序列、c) DNA表达盒中的克隆位点、或d) DNA表达盒中的能够形成非B型DNA构型的DNA序列间隔至少500 bp或700 bp。

在一些实施方式中，载体骨架可以包含：

i) 复制起点，

ii) 第一终止子，

iii) a) 克隆位点，b) 克隆位点和能够形成非B型DNA构型的DNA序列，c) DNA表达盒，其包含启动子和克隆位点，或d) DNA表达盒，其包含启动子、克隆位点和能够形成非B型DNA构型的DNA序列，

iv) 第二终止子，以及

v) 筛选标记和控制该筛选标记表达的启动子，

其中第一终止子和第二终止子可以分别位于a) 的克隆位点的上游和下游，并设置成使得载体中a) 的克隆位点之外的转录终止于a) 的克隆位点的上游和下游，

第一终止子和第二终止子可以分别位于b) 的克隆位点和能够形成非B型DNA构型的DNA序列的上游和下游，并设置成使得载体中b) 的克隆位点和能够形成非B型DNA构型的DNA序列之外的转录终止于b) 的克隆位点和能够形成非B型DNA构型的DNA序列的上游和下游，

第一终止子和第二终止子可以分别位于c) 的DNA表达盒的上游和下游，并设置成使得载体中c) 的DNA表达盒之外的转录终止于c) 的DNA表达盒的上游和下游，或

第一终止子和第二终止子可以分别位于d) 的DNA表达盒的上游和下游，并设置成使得载体中d) 的DNA表达盒之外的转录终止于d) 的DNA表达盒的上游和下游；

其中复制起点可以在复制方向上与a) 克隆位点、b) 的能够形成非B型DNA构型的DNA序列，c) DNA表达盒中的克隆位点、或d) DNA表达盒中的能够形成非

B型DNA构型的DNA序列间隔至少500 bp、或700 bp。

在一些实施方式中，载体骨架可以按序包含pUC复制起点、rrnB T1终止子和rrnB T2终止子、克隆位点、lambda t0终止子、和卡那霉素抗性元件。

在一些实施方式中，载体骨架可以按序包含pUC复制起点、rrnB T1终止子和rrnB T2终止子、克隆位点、目标DNA序列、lambda t0终止子、和卡那霉素抗性元件。

在一些实施方式中，载体骨架可以按序包含pUC复制起点、rrnB T1终止子和rrnB T2终止子、CMV增强子、CMV启动子、克隆位点、bGH poly(A)信号、lambda t0终止子、和卡那霉素抗性元件。

在一些实施方式中，载体骨架可以按序包含pUC复制起点、rrnB T1终止子和rrnB T2终止子、CMV增强子、CMV启动子、克隆位点、poly(A)尾编码序列、载体线性化位点、bGH poly(A)信号、lambda t0终止子、和卡那霉素抗性元件。在一些实施方式中，载体骨架可以按序包含pUC复制起点、rrnB T1终止子和rrnB T2终止子、CMV增强子、CMV启动子、5'UTR、克隆位点、3'UTR、poly(A)尾编码序列、载体线性化位点、bGH poly(A)信号、lambda t0终止子、和卡那霉素抗性元件。

在第二个方面，本申请提供一种载体，其可以包含：

i) 复制起点，以及

ii) a) 目标DNA表达盒，其按序包含第一ITR或其衍生序列、启动子、目标DNA序列、和第二ITR或其衍生序列，或b) 目标DNA序列组，其按序包含第一ITR或其衍生序列、目标DNA序列、和第二ITR或其衍生序列。

复制起点可以在复制方向上与a) 的目标DNA表达盒或b) 的目标DNA序列组间隔至少150 bp、200 bp、300 bp、500 bp、700 bp、800 bp、900 bp、1000 bp、1100 bp、1200 bp、1300 bp、1400 bp、1500 bp、1600 bp或1700 bp。在一些实施方式中，复制起点可以在复制方向上与a) 的目标DNA表达盒或b) 的目标DNA序列组间隔至少300 bp。在一些实施方式中，复制起点可以在复制方向上与a) 的目标DNA表达盒或b) 的目标DNA序列组间隔至少500 bp。在一些实施方式中，复制起点可以在复制方向上与a) 的目标DNA表达盒或b) 的目标DNA序列组间隔至少700 bp。

复制起点可以使得载体在合适的宿主细胞中复制。合适的宿主细胞可以是原核细胞如大肠杆菌、或真核细胞。复制起点可以选自pMB1复制起点、pBR322复制起点、ColE1复制起点、R6K复制起点、p15A复制起点、pSC101复制起点、pUC复制起点、pET复制起点和这些复制起点的衍生物。这些复制起点衍生物也可以使得载体在合适的宿主细胞中复制。在一些实施方式中，复制起点可以为pUC复制起点。

载体还可以包含第二复制起点，该第二复制起点可以在复制方向上与a) DNA表达盒或b) 目标DNA序列组间隔至少150 bp、200 bp、300 bp、500 bp、700 bp、800 bp、900 bp、1000 bp、1100 bp、1200 bp、1300 bp、1400 bp、1500 bp、1600 bp、或1700 bp。该第二复制起点可以使得载体在合适的宿主细胞中复制，合适的宿主

细胞可以是原核细胞如大肠杆菌、或真核细胞。

取决于复制起点的种类，载体还可以相应地包含Rep蛋白表达元件，包括Rep蛋白编码序列和相应的启动子。

载体还可以包含第一终止子和第二终止子，分别位于a)的目标DNA表达盒的上游和下游，并设置成使得载体中a)的目标DNA表达盒之外的转录终止于a)的目标DNA表达盒的上下游；或分别位于b)的目标DNA序列组的上游和下游，并设置成使得载体中b)的目标DNA序列组之外的转录终止于b)目标DNA序列组的上下游。第一终止子可以为正向终止子、双向终止子、或反向终止子。第二终止子可以为正向终止子、双向终止子、或反向终止子。在一些实施方式中，第一终止子和第二终止子可以均为双向终止子。第一终止子和第二终止子可以独立地选自rrnB T1终止子、rrnB T2终止子、lambda t0终止子、TonB终止子、Amp终止子、T7终止子、和CYC1终止子。在一些实施方式中，第一终止子可以为rrnB T1终止子和rrnB T2终止子。在一些实施方式中，第二终止子可以为lambda t0终止子。

载体还可以包含筛选标记。筛选标记可以为抗生素抗性基因，例如氨苄青霉素抗性基因、四环素抗性基因、氯霉素抗性基因、卡那霉素抗性基因、或链霉素抗性基因等。特别地，筛选标记不存在于载体待转化/转染的宿主细胞的基因组中。在一些实施方式中，筛选标记为卡那霉素抗性基因。该载体还可以在筛选标记的上游包含用于该筛选标记的表达的启动子。

第一ITR和第二ITR可以为野生型的AAV-ITR，包含例如SEQ ID NO: 9所示的核苷酸序列。

第一ITR衍生序列和第二ITR衍生序列可以形成与天然AAV-ITR相似的回文结构，例如包含一个回文臂和一个回文茎、或仅包含一个回文茎，且保持有Rep蛋白结合力。在一些实施方式中，第一ITR衍生序列和第二ITR衍生序列可以包含SEQ ID NO: 10、11或12所示的核苷酸序列。

第一ITR或其衍生序列、与第二ITR或其衍生序列可以包含相同、或不同的核苷酸序列。

载体可以为环状载体。载体可以为非病毒载体。载体可以为质粒载体。

载体中不包含能够引起转录的lac启动子。

a)的目标DNA表达盒可以包含IIS型限制性内切酶(IIS酶)识别位点。b)的目标DNA序列组可以包含IIS酶识别位点。除a)的目标DNA表达盒或b)的目标DNA序列组外，载体中不包含与a)目标DNA表达盒和b)目标DNA序列组相同的IIS酶识别位点。目标DNA序列可以为任何待表达的DNA序列，如mRNA编码序列和poly(A)尾编码序列。目标DNA序列可以包含能够形成非B型DNA构型的DNA序列。目标DNA序列可以包含克隆位点，例如多克隆位点。

a)的目标DNA表达盒可以是适于在真核细胞中表达目标DNA序列的表达盒，其可以包含使得目标DNA序列在真核细胞中表达的组件。a)的目标DNA表达盒中

包含的启动子可以是使得目标DNA序列在真核细胞中表达的启动子。a) 的目标DNA表达盒还可以包含增强子、转录终止子等，特别是使得目标DNA序列在真核细胞中表达的增强子、和转录终止子等。适用于目标DNA序列在真核细胞中表达的转录终止子可以选自SV40 poly(A)、hGH poly(A)、BGH poly(A)、和rbGlob poly(A)。在一些实施方式中，a) 的目标DNA表达盒可以是适于在真核细胞中表达mRNA的表达盒，其可以包含任选的增强子、启动子、5'UTR、开放阅读框(ORF)、3'UTR、Poly(A)尾编码序列、和转录终止子。a) 的目标DNA表达盒还可以包含用于载体线性化的酶切位点。

在一些实施方式中，载体可以包含：

i) 复制起点，

ii) a) 目标DNA表达盒，其按序包含第一ITR或其衍生序列、启动子、目标DNA序列、和第二ITR或其衍生序列，或b) 目标DNA序列组，其按序包含第一ITR或其衍生序列、目标DNA序列、和第二ITR或其衍生序列，以及

iii) 筛选标记和控制该筛选标记表达的启动子，

其中复制起点可以在复制方向上与a) DNA表达盒或b) 目标DNA序列组间隔至少500 bp或700 bp。

在一些实施方式中，载体骨架可以包含：

i) 复制起点，

ii) 第一终止子，

iii) a) 目标DNA表达盒，其按序包含第一ITR或其衍生序列、启动子、目标DNA序列、和第二ITR或其衍生序列，或b) 目标DNA序列组，其按序包含第一ITR或其衍生序列、目标DNA序列、和第二ITR或其衍生序列，

iv) 第二终止子，以及

v) 筛选标记和控制该筛选标记表达的启动子，

其中第一终止子和第二终止子分别位于a) 的目标DNA表达盒的上游和下游，并设置成使得载体中a) 的目标DNA表达盒之外的转录终止于a) 的目标DNA表达盒的上游和下游，或第一终止子和第二终止子分别位于b) 的目标DNA序列组的上游和下游，并设置成使得载体中b) 的目标DNA序列组之外的转录终止于b) 的目标DNA序列组的上游和下游，

其中复制起点可以在复制方向上与a) DNA表达盒或b) 目标DNA序列组间隔至少500 bp或700 bp。

在一些实施方式中，载体可以按序包含pUC复制起点、rrnB T1终止子和rrnB T2终止子、第一ITR或其衍生序列、CMV增强子、鸡β-肌动蛋白启动子、目标DNA序列、bGH poly(A)信号、第二ITR或其衍生序列、lambda t0终止子、和卡那霉素抗性元件。

相应地，本申请提供一种载体骨架，其可以包含：

i) 复制起点, 以及

ii) a) DNA序列组, 其按序包含第一ITR或其衍生序列、克隆位点、和第二ITR或其衍生序列, 或b) DNA表达盒, 其按序包含第一ITR或其衍生序列、启动子、克隆位点、和第二ITR或其衍生序列。

复制起点可以在复制方向上与a) 的DNA序列组、或b) 的DNA表达盒间隔至少150 bp、200 bp、300 bp、500 bp、700 bp、800 bp、900 bp、1000 bp、1100 bp、1200 bp、1300 bp、1400 bp、1500 bp、1600 bp、或1700 bp。在一些实施方式中, 复制起点可以在复制方向上与a) 的DNA序列组、或b) 的DNA表达盒间隔至少500 bp。在一些实施方式中, 复制起点可以在复制方向上与a) 的DNA序列组、或b) 的DNA表达盒间隔至少700 bp。

复制起点可以使得载体在合适的宿主细胞中复制。合适的宿主细胞可以是原核细胞如大肠杆菌、或真核细胞。复制起点可以选自pMB1复制起点、pBR322复制起点、ColE1复制起点、R6K复制起点、p15A复制起点、pSC101复制起点、pUC复制起点、pET复制起点和这些复制起点的衍生物。这些复制起点衍生物也可以使得载体在合适的宿主细胞中复制。在一些实施方式中, 复制起点可以为pUC复制起点。

载体骨架还可以包含第二复制起点, 该第二复制起点可以在复制方向上与a) 的DNA序列组、或b) 的DNA表达盒间隔至少150 bp、200 bp、300 bp、500 bp、700 bp、800 bp、900 bp、1000 bp、1100 bp、1200 bp、1300 bp、1400 bp、1500 bp、1600 bp、或1700 bp。该第二复制起点可以使得载体在合适的宿主细胞中复制, 合适的宿主细胞可以是原核细胞如大肠杆菌、或真核细胞。

取决于复制起点的种类, 载体还可以相应地包含Rep蛋白表达元件, 包括Rep蛋白编码序列和相应的启动子。

载体骨架还可以包含第一终止子和第二终止子, 分别位于a) 的DNA序列组的上游和下游, 并设置成使得载体中a) 的DNA序列组之外的转录终止于a) 的DNA序列组的上下游, 或第一终止子和第二终止子分别位于b) 的DNA表达盒的上游和下游, 并设置成使得载体中 b) 的DNA表达盒之外的转录终止于b) 的DNA表达盒的上游和下游。第一终止子可以为正向终止子、双向终止子、或反向终止子。第二终止子可以为正向终止子、双向终止子、或反向终止子。在一些实施方式中, 第一终止子和第二终止子可以均为双向终止子。第一终止子和第二终止子可以独立地选自rrnB T1终止子、rrnB T2终止子、lambda t0终止子、TonB终止子、Amp终止子、T7终止子、和CYC1终止子。在一些实施方式中, 第一终止子可以为rrnB T1终止子和rrnB T2终止子。在一些实施方式中, 第二终止子可以为lambda t0终止子。

载体骨架中a) 的DNA序列组中的克隆位点可以为多克隆位点。载体骨架中b) 的DNA表达盒中的克隆位点可以为多克隆位点。

在a) 的DNA序列组中, 克隆位点处可以插入任何序列, 特别地为, 目标DNA序列, 或者启动子和目标DNA序列。目标DNA序列可以为任何待表达的DNA序列,

如mRNA编码序列等。

在b)的DNA表达盒中,克隆位点处可以插入任何序列,特别地为目标DNA序列。目标DNA序列可以为任何待表达的DNA序列,如mRNA编码序列等。

载体骨架还可以包含筛选标记。筛选标记可以为抗生素抗性基因,例如氨苄青霉素抗性基因、四环素抗性基因、氯霉素抗性基因、卡那霉素抗性基因、或链霉素抗性基因等。特别地,筛选标记不存在于载体待转化/转染的宿主细胞的基因组中。在一些实施方式中,筛选标记为卡那霉素抗性基因。该载体还可以在筛选标记的上游包含用于该筛选标记的表达的启动子。

第一ITR和第二ITR可以为野生型的AAV-ITR,包含例如SEQ ID NO: 9所示的核苷酸序列。

第一ITR衍生序列和第二ITR衍生序列可以形成与天然AAV-ITR相似的回文结构,例如包含一个回文臂和一个回文茎、或仅包含一个回文茎,且保持有Rep蛋白结合力。在一些实施方式中,第一ITR衍生序列和第二ITR衍生序列可以包含SEQ ID NO: 10、11或12所示的核苷酸序列。

第一ITR或其衍生序列、与第二ITR或其衍生序列可以包含相同、或不同的核苷酸序列。

载体骨架可以为环状载体骨架。载体骨架可以为非病毒载体骨架。载体骨架可以为质粒载体骨架。

载体骨架中不包含能够引起转录的lac启动子。

b)的DNA表达盒可以是适于在真核细胞中DNA序列的表达盒,其可以包含使得DNA序列在真核细胞中表达的组件。b)的DNA表达盒中包含的启动子可以是使得DNA序列在真核细胞中表达的启动子。b)的目标DNA表达盒还可以包含增强子、转录终止子等,特别是使得DNA序列在真核细胞中表达的增强子、和转录终止子等。适用于DNA序列在真核细胞中表达的转录终止子可以选自SV40 poly(A)、hGH poly(A)、BGH poly(A)、和rbGlob poly(A)。在一些实施方式中,b)的目标DNA表达盒可以是适于在真核细胞中表达mRNA的表达盒,其可以包含任选的增强子、启动子、5'UTR、开放阅读框(ORF)、3'UTR、和转录终止子。b)的DNA表达盒还可以包含用于载体线性化的酶切位点。

在一些实施方式中,载体骨架可以包含:

i) 复制起点,

ii) a) DNA序列组,其按序包含第一ITR或其衍生序列、克隆位点、和第二ITR或其衍生序列,或b) DNA表达盒,其按序包含b)第一ITR或其衍生序列、启动子、克隆位点、和第二ITR或其衍生序列,以及

iii) 筛选标记和控制该筛选标记表达的启动子,

其中复制起点可以在复制方向上与a) DNA序列组、或b) DNA表达盒间隔至少500 bp、或700 bp。

在一些实施方式中，载体骨架可以包含：

i) 复制起点，

ii) 第一终止子，

iii) a) DNA序列组，其按序包含a) 第一ITR或其衍生序列、克隆位点、和第二ITR或其衍生序列，或b) DNA表达盒，其按序包含b) 第一ITR或其衍生序列、启动子、克隆位点、和第二ITR或其衍生序列，

iv) 第二终止子，以及

v) 筛选标记和控制该筛选标记表达的启动子，

其中第一终止子和第二终止子分别位于a) 的DNA序列组的上游和下游，并设置成使得由载体中a) 的DNA序列组之外的转录终止于a) 的DNA序列组的上游和下游，或

第一终止子和第二终止子分别位于b) 的DNA表达盒的上游和下游，并设置成使得由载体中b) 的DNA表达盒之外的转录终止于b) 的DNA表达盒的上游和下游，

其中复制起点可以在复制方向上与a) DNA序列组、或b) DNA表达盒间隔至少500 bp、或700 bp。

在一些实施方式中，载体骨架可以按序包含pUC复制起点、rrnB T1终止子和rrnB T2终止子、第一ITR或其衍生序列、CMV增强子、鸡 β -肌动蛋白启动子、克隆位点、bGH poly(A)信号、第二ITR或其衍生序列、lambda t0终止子、和卡那霉素抗性元件。

在第三个方面，本申请提供一种用于制备本申请的载体的方法，包括根据上述本申请载体骨架的设计，从头合成载体。

本申请提供一种用于制备本申请的载体骨架的方法，包括根据上述本申请载体骨架的设计，从头合成载体。

本申请还提供一种用于改造载体或载体骨架的方法，包括i) 提供一种载体或载体骨架，以及ii) 对该载体或载体骨架进行改造，使得改造后的载体或载体骨架包含：

a) 复制起点，

b) 第一终止子，

c) c-i) 克隆位点，

c-ii) DNA表达盒，其包含启动子和克隆位点，

c-iii) DNA表达盒，其包含启动子、克隆位点和能够形成非B型DNA构型的DNA序列，

c-iv) 目标DNA序列，其包含能够形成非B型DNA构型的DNA序列，

c-v) 目标DNA表达盒，其包含启动子和目标DNA序列，目标DNA序列包含能够形成非B型DNA构型的DNA序列，

c-vi) DNA序列组，其按序包含第一ITR或其衍生序列、克隆位点、

和第二ITR或其衍生序列，

c-vii) DNA表达盒，其按序包含第一ITR或其衍生序列、启动子、克隆位点、和第二ITR或其衍生序列，

c-viii) 目标DNA表达盒，其按序包含第一ITR或其衍生序列、启动子、目标DNA序列、和第二ITR或其衍生序列，或

c-ix) 目标DNA序列组，其按序包含第一ITR或其衍生序列、目标DNA序列、和第二ITR或其衍生序列，

d) 第二终止子，以及

e) 筛选标记和控制该筛选标记表达的启动子，

其中第一终止子和第二终止子分别位于c) 的序列的上游和下游，且设置成使得c) 的序列以外的转录终止于c) 的序列的上游和下游，

可选地，复制起点在复制方向上与c-i) 的克隆位点、c-ii) 的DNA表达盒中的克隆位点、c-iii) 的DNA表达盒中的能够形成非B型DNA构型的DNA序列、c-iv) 的目标DNA序列中的能够形成非B型DNA构型的DNA序列、c-v) 目标DNA表达盒中的能够形成非B型DNA构型的DNA序列、c-vi) 的DNA序列组、c-vii) 的DNA表达盒、c-viii) 的目标DNA表达盒、或c-ix) 的目标DNA序列组间隔至少150 bp、200 bp、300 bp、500 bp或700 bp，以及

可选地，载体不含活性lac启动子。

c) 的序列中的克隆位点可以是多克隆位点。

具体地，本申请还提供一种用于改造载体或载体骨架的方法，包括：

i) 提供一种载体或载体骨架，其包含a) 复制起点，以及b) b-i) 克隆位点，b-ii) DNA表达盒，其包含启动子和克隆位点，b-iii) DNA表达盒，其包含启动子、克隆位点和能够形成非B型DNA构型的DNA序列，b-iv) 目标DNA序列，其包含能够形成非B型DNA构型的DNA序列，b-v) 目标DNA表达盒，其包含启动子和目标DNA序列，目标DNA序列包含能够形成非B型DNA构型的DNA序列，b-vi) DNA序列组，其按序包含第一ITR或其衍生序列、克隆位点、和第二ITR或其衍生序列，b-vii) DNA表达盒，其按序包含第一ITR或其衍生序列、启动子、克隆位点、和第二ITR或其衍生序列，b-viii) 目标DNA表达盒，其按序包含第一ITR或其衍生序列、启动子、目标DNA序列、和第二ITR或其衍生序列，或b-ix) 目标DNA序列组，其按序包含第一ITR或其衍生序列、目标DNA序列、和第二ITR或其衍生序列；以及

ii) 在载体或载体骨架中加入第一终止子和第二终止子，使得其分别位于b) 的序列的上游和下游，设置成使b) 序列之外的转录终止于b) 序列的上游和下游。

该方法还可以包括使复制起点在复制方向上与b-i) 的克隆位点、b-ii) 的DNA表达盒中的克隆位点、b-iii) 的DNA表达盒中的能够形成非B型DNA构型的DNA序列、b-iv) 的目标DNA序列中的能够形成非B型DNA构型的DNA序列、b-v) 目标DNA表达盒中的能够形成非B型DNA构型的DNA序列、b-vi) 的DNA序列组、b-vii)

的DNA表达盒、b-viii)的目标DNA表达盒、或b-ix)的目标DNA序列组间隔至少150 bp、200 bp、300 bp、500 bp、或700 bp。

在载体或载体骨架还包含lac启动子的情况下，该方法还包括使得lac启动子失活。

能够形成非B型DNA构型的DNA序列可以为正向重复序列、反向重复序列、或单一嘌呤-单一嘧啶序列，包括但不限于，多聚腺嘌呤（poly(A)）、多聚鸟嘌呤（poly(G)）、多聚胞嘧啶（poly(C)）、或多聚胸腺嘧啶（poly(T)）、shRNA编码序列、和ITR序列。正向重复序列可以为poly(A)尾编码序列。单一嘌呤-单一嘧啶序列可以为poly(A)尾编码序列。反向重复序列可以为ITR或其衍生序列。在一些实施方式中，能够形成非B型DNA构型的DNA序列可以为poly(A)编码序列、ITR或其衍生序列。

复制起点可以使得载体或载体骨架在合适的宿主细胞中复制。合适的宿主细胞可以是原核细胞如大肠杆菌、或真核细胞。复制起点可以选自pMB1复制起点、pBR322复制起点、ColE1复制起点、R6K复制起点、p15A复制起点、pSC101复制起点、pUC复制起点、pET复制起点和这些复制起点的衍生物。这些复制起点衍生物也可以使得载体在合适的宿主细胞中复制。在一些实施方式中，复制起点可以为pUC复制起点。

取决于复制起点的种类，载体还可以相应地包含Rep蛋白表达元件，包括Rep蛋白编码序列和相应的启动子。

第一终止子可以为正向终止子、双向终止子、或反向终止子。第二终止子可以为正向终止子、双向终止子、或反向终止子。在一些实施方式中，第一终止子和第二终止子可以均为双向终止子。第一终止子和第二终止子可以独立地选自rrnB T1终止子、rrnB T2终止子、lambda t0终止子、TonB终止子、Amp终止子、T7终止子、和CYC1终止子。在一些实施方式中，第一终止子可以为rrnB T1终止子和rrnB T2终止子。在一些实施方式中，第二终止子可以为lambda t0终止子。

载体或载体骨架可以包含筛选标记。筛选标记可以为抗生素抗性基因，例如氨苄青霉素抗性基因、四环素抗性基因、氯霉素抗性基因、卡那霉素抗性基因、或链霉素抗性基因等。特别地，筛选标记不存在于载体待转化/转染的宿主细胞的基因组中。在一些实施方式中，筛选标记为卡那霉素抗性基因。该载体还可以在筛选标记的上游包含用于该筛选标记的表达的启动子。

载体或载体骨架可以为环状。载体或载体骨架可以为非病毒来源。载体或载体骨架可以为质粒载体、或质粒载体骨架。

在第四个方面，本申请提供一种组合物，其包含本申请的载体骨架或载体，包括经本申请方法制备得到的载体骨架或载体、以及经本申请方法改造得到的载体骨架或载体。

在第五个方面，本申请提供本申请的载体骨架或载体，包括经本申请方法制备

得到的、以及经本申请方法改造得到的那些，在复制和/或表达目标DNA序列，特别是含有能够形成非B型DNA构型的DNA序列（例如ITR、或poly(A)尾编码序列）的DNA序列中的用途。

具体地，本申请提供一种用于复制目标DNA序列的方法，包括：

i) 提供本申请第一方面中所述的载体，

提供本申请第一方面中所述的载体骨架，在a) 的克隆位点处插入目标DNA序列、或者启动子和目标DNA序列，在b) 的克隆位点处插入目标DNA序列，在c) 的DNA表达盒的克隆位点处插入目标DNA序列，或在d) 的DNA表达盒的克隆位点处插入目标DNA序列，获得新的载体，

提供本申请第二方面中所述的载体，或

提供本申请第二方面中所述的载体骨架，在a) 的DNA序列组的克隆位点处插入目标DNA序列、或启动子和目标DNA序列，或在b) DNA表达盒的克隆位点处插入目标DNA序列，得到新的载体；

ii) 将步骤i) 的载体转化/转染到宿主细胞中；

iii) 在利于载体复制的条件下培养宿主细胞。

本申请还提供一种用于表达目标DNA序列的方法，包括：

i) 提供本申请第一方面中所述的包含a) 目标DNA表达盒的载体，或

提供本申请第一方面中所述的载体骨架，在a) 的克隆位点处插入启动子和目标DNA序列，在b) 的克隆位点处插入目标DNA序列，在c) 的DNA表达盒的克隆位点处插入目标DNA序列，或在d) 的DNA表达盒的克隆位点处插入目标DNA序列，获得新的载体；以及

ii) 将步骤i) 的载体转化/转染到宿主细胞中，在利于目标DNA序列表达的条件下培养宿主细胞，或使步骤i) 的载体线性化、并进行体外转录；或者

i) 提供本申请第二方面所述的包含a) 目标DNA表达盒的载体，

或提供本申请第二方面中所述的载体骨架，在a) 的DNA序列组的克隆位点处插入启动子和目标DNA序列，或在b) DNA表达盒的克隆位点处插入目标DNA序列，得到新的载体，

ii) 将步骤i) 的载体转化/转染到宿主细胞中，并在利于目标DNA序列表达的条件下培养宿主细胞，

使步骤i) 的载体线性化、并进行体外转录；或者

使步骤i) 的载体与包装质粒、辅助质粒共转化/转染宿主细胞，在利于载体包装的条件下培养宿主细胞，回收包装好的载体，转化/转染宿主细胞，并在利于目标DNA序列表达的条件下培养宿主细胞。

该方法还可以在步骤i) 和步骤ii) 之间包含将步骤i) 的载体转化/转染到宿主细胞中，以及在利于载体复制的条件下培养宿主细胞。

在第六个方面，本申请提供本申请的载体骨架或载体，包括经本申请方法制备

得到的、以及经本申请方法改造得到的那些，在制备重组腺相关病毒载体（rAAV）中的用途。

具体地，本申请提供一种用于制备重组腺相关病毒载体的方法，包括：

i) 提供本申请第二方面中所述的载体，或

提供本申请第二方面所述的载体骨架，在a) 的DNA表达盒的克隆位点处插入目标DNA序列、或启动子和目标DNA序列，或在b) DNA表达盒的克隆位点处插入目标DNA序列，得到新的载体；

ii) 可选地，将步骤i) 的载体转化/转染到宿主细胞中，并在利于载体复制的条件下培养宿主细胞；以及

iii) 收获步骤i) 或ii) 的载体，将其与包装质粒、辅助质粒共转化/转染宿主细胞，在利于载体包装的条件下培养宿主细胞。

基于以下具体描述和实施例，当前公开内容的其他特征和优势之处将会更加明晰，具体描述和实施例不应解读为限制性的。在本申请中引用的所有文献、Genbank记录、专利和已公开专利申请的内容通过引用的方式明确地包含在本文中。

附图说明

以下以示例方式给出但不意在将本发明限制于所述具体实施方式的具体描述，可以结合附图更好地理解。

图1示出pUC57-kana测试质粒的示意图。

图2示出poly(A)尾编码序列缺失的失败克隆（A）、poly(A)尾编码序列不纯且测序底峰高于主峰20%的失败克隆（B）、以及poly(A)尾编码序列较纯且测序底峰低于主峰20%的阳性克隆（C）的测序示意图。

图3示出在Ori与poly(A)尾编码序列间不同长度间隔序列的pUC57-kana质粒的单克隆成功率。

图4示出在Ori与poly(A)尾编码序列间的247 bp间隔序列中含有终止子的pUC57-kana质粒的示意图。

图5示出在247 bp间隔序列中不含终止子（A）、或含有终止子（B）的pUC57-kana质粒的测序示意图，以及上述两种质粒的单克隆成功率（C）。

图6示出用465 bp间隔序列破坏lacP功能区的pUC57-kana质粒的示意图。

图7示出lacP功能区对含有poly(A)尾编码序列的质粒单克隆的构建成功率的影响。

图8示出GS-MK质粒的图谱。

图9示出GS-PV质粒的图谱。

图10示出插入有不同poly(A)尾编码序列的GS-MK载体的测序峰形图。

图11示出插入有不同poly(A)尾编码序列的GS-PV载体的测序峰形图。

图12示出插入有不同poly(A)尾编码序列的GS-CMV载体的测序峰形图。

图13示出含有poly(A)尾编码序列的pUC57-kana、GS-MK、PVAX1、GS-PV和GS-CMV的单克隆成功率。

图14示出含有poly(A)尾编码序列的pUC57-kana、GS-MK、PVAX1、GS-PV和GS-CMV的传代稳定性。

图15示出含有poly(A)尾编码序列的pUC57-kana、GS-MK、PVAX1、GS-PV和GS-CMV的产量。

图16示出含有poly(A)尾编码序列的pUC57-kana、GS-MK、PVAX1、GS-PV和GS-CMV的体外转录产物纯度。

图17示出含有poly(A)尾编码序列的pUC57-kana、GS-MK、PVAX1、GS-PV和GS-CMV转录出的mRNA的蛋白表达量。

图18示出4种含有lacP的质粒图谱。

图19示出加上ITR序列的pUC57-kana载体示意图(A)，以及加上ITR序列、调整Ori位置、破坏lacP功能区、以及在5' ITR位置的下游和3' ITR的上游加入终止子序列的载体示意图(B)。

图20示出pAAV-tTA质粒示意图(A)，以及更换启动子、并在5' ITR的下游和3' ITR的上游添加终止子序列的载体示意图(B)。

图21示出pUC57 kana、GenS-pUC57-AAV-NoTer、和GenS-pUC57-AAV-Ter进行基因克隆后的酶切电泳图。泳道1-3: pUC57 kana载体、SmaI酶切产物、和标准大小DNA标记; 泳道4-6: GenS-pUC57-AAV-NoTer载体、SmaI酶切产物、和标准大小DNA标记; 泳道7-9: GenS-pUC57-AAV-Ter载体、SmaI酶切产物、和标准大小DNA标记。

具体实施方式

本文中用到的术语，除非特别指出，均具有字典、教科书、技术工具书中的普通含义，或本领域技术人员通常所理解的意思。以下对于一些术语的描述，仅出于便于理解本申请的目的，而不意在对这些术语进行特别的限定，除非特别指出。

如本文和所附权利要求书中所使用的，单数形式“一个”、“一种”和“该”包括所指对象的复数形式，除非上下文另有明确规定。例如，“一个载体”可以包括多个该种载体。

术语“克隆载体”是指一类具有自我复制能力，含有复制起点、筛选标记和克隆位点等基因元件，可以携带DNA片段进入宿主细胞并大量扩增DNA片段的DNA分子。克隆载体可以包含能够使载体在一种或多种宿主细胞中自主复制的核酸序列，例如复制起点。

术语“表达载体”是指在克隆载体中加入与表达调控有关的元件，例如启动子、核糖体结合位点、终止子等而得到的载体。

载体骨架是指具备基本的复制元件或表达元件、以及其他一些所需的元件的

载体。可以在载体骨架中经克隆位点插入、或经同源重组替换入感兴趣的DNA序列，从而实现感兴趣DNA序列的复制、表达等。

“复制起点”或“Ori”是指DNA复制起始的位点，控制着质粒的复制，并决定了质粒的宿主属性和拷贝数。Ori通常富含腺嘌呤（A）和胸腺嘧啶（T），容易解链，给复制调控元件如Rep蛋白提供结合的空间。常见的复制起点包括pUC（核苷酸序列如SEQ ID NO: 23所示）、ColE1、pMB1、pSC101、R6K等。根据复制起点控制的不同，载体可以分为松弛型和严紧型。其中在严紧型质粒载体中，质粒的复制与宿主细胞染色体的复制表现出受相同因素控制，宿主细胞复制一次，质粒也复制一次，每个宿主细胞中只有1-2个质粒；而在松弛型质粒中，当宿主细胞染色体复制停止后仍然能继续复制，表现出受到与染色体复制不同的因素控制，每个细胞中一般含有几十至上百个左右质粒。此外，根据载体中复制起点调控元件的设计，载体可以分别窄宿主类型、和宽宿主类型。其中在窄宿主类型的载体中，Ori依赖于宿主细胞的复制元件或由宿主细胞提供的Rep蛋白，而在宽宿主类型的载体中，载体自主包含Rep蛋白表达元件。“Rep蛋白”可以与复制起点结合，沿着一定的方向移动，将双链分解开。载体包含一个或多个复制起点，用于在一种或多种宿主细胞中自主复制。

“筛选标记”是指载体中起着特异性标记作用的基因，可以用来检验载体是否已成功转化/转染到宿主细胞中。筛选标记可以是载体固有的，也可以是重组加入的。筛选标记可以是抗性基因，优选为宿主细胞所不具备的抗性基因。这样，当载体成功导入宿主细胞时，可以赋予宿主细胞相应的抗性，从而判断载体已进入宿主细胞。常用的筛选标记包括氨苄青霉素（Ampr）抗性基因、氯霉素（Camr）抗性基因、卡那霉素（Kanr）抗性基因、四环素（Tetr）抗性基因、嘌呤霉素（Puro）抗性基因、G418抗性基因、潮霉素 β （Hygr）抗性基因、博来霉素抗性基因、杀稻瘟菌素（通过干扰核糖体中肽键的形成来特异性抑制原核和真核生物的蛋白质合成）抗性基因等。

“克隆位点”是指载体中含有的一段DNA片段，包含一个或多个限制性内切酶识别位点。克隆位点经过相应的限制性内切酶的处理后，可以使得经过相同限制性内切酶处理的外源DNA插入。根据序列上限制性内切酶识别位点的数量，克隆位点可以分为单克隆位点和多克隆位点（MCS）。克隆位点可以包含NheI、BmtI、AflII、HindIII、Acc65I、KpnI、BamHI、ApoI、RcoRI、PstI、EcoRV、BstXI和/或NotI的识别位点。本领域技术人员可以常规确定克隆位点的序列。克隆位点中可以包含IIS型限制性内切酶的识别位点。IIS型限制性内切酶是一组比较特殊的内切酶，在识别位点下游的特定位置处切割DNA，对切割位点的序列没有要求。在一些实施方式中，待插入或插入载体骨架的目标DNA序列，如mRNA编码序列和poly(A)尾编码序列的端部则含有IIS酶识别位点，而载体骨架不含对应的IIS酶识别位点，例如可以通过同义突变去除BsaI、BspQI、BspMI、BpiI、BsmBI等IIS位点。如此，

含有目标DNA序列的载体在线性化后，IIS型限制性内切酶仅在例如poly(A)尾编码序列之外进行识别，并在poly(A)编码序列内切割。首先，得到的酶切片段不会因为存在多个IIS酶切位点而显得杂乱，其次，酶切片段中的polyT序列不会加入非T序列，相应得到的poly(A)尾也很纯，而纯的poly(A)尾可以维持较高的mRNA的蛋白表达水平（Holtkamp S *et al.*, (2006) Modification of antigen-encoding RNA increases stability, translational efficacy, and T-cell stimulatory capacity of dendritic cells. *Blood*. 108(13):4009-4017）。这种设计也便于载体进行Golden Gate克隆。

“启动子”是指RNA聚合酶识别、结合并起始转录的DNA序列，含有RNA聚合酶特异性结合和转录起始所需的保守序列。多数启动子位于基因转录起始点的上游，本身不会被转录。启动子可以分为组成型启动子、组织/细胞特异型启动子、和诱导型启动子。组成型启动子又称为非特异性表达启动子，在其调控下，基因表达在不同细胞、组织器官或发育阶段没有明显差异，常见的这类启动子包括Ubc、CMV、SV40、EF1a、U6、CAG（由CMV增强子和鸡 β -肌动蛋白启动子组成）等。在组织/细胞特异型启动子的调控下，外源基因仅在特定的细胞中特异表达。诱导表达型启动子是指在某些特定的物理或化学信号的刺激下，启动子开始大幅度提高基因的转录水平，比较常见的是四环素诱导型表达启动子。

“增强子”是指是DNA上可与转录因子结合的区域。在与转录因子结合之后，基因的转录作用将会加强。增强子可能位于基因上游，也可能位于下游，且不一定接近所要作用的基因。常见的增强子包括CMV增强子等。

“终止子”是指能够终止RNA聚合酶转录并起始新合成RNA与转录机构分离的DNA序列。终止子通常位于基因的3'调控序列后。原核终止子可以分为两类，一类依赖蛋白辅因子才能实现终止作用，这种蛋白质辅因子通常称为 ρ 因子（Rho因子）；一类不依赖于蛋白质辅因子就能实现终止作用，这类终止子在序列上有一些共通的特点，即有一段富含GC的反向重复序列，转录而来的RNA易形成发夹结构，破坏RNA-DNA杂合链5'端的正常结构，使RNA聚合酶暂停转录，在GC反向重复序列后紧随一段由4-8个A组成的序列，转录产物为寡U，使RNA-DNA杂合链3'端出现不稳定的U·A区域，释放后形成弱键连接的多聚尿嘧啶环。几乎所有常用的细菌表达载体都使用不依赖 ρ 因子的终止子，包括一些自然存在的终止子如T7和rrnB、和一些工程化的高效率终止子如T0。没有任何终止子可以100%的效率启动终止步骤，但是一些工程化的终止子的终止效率可以接近95%。

在本申请中，适用于在真核细胞中进行DNA序列表达的终止子，在某些情况下可称为“转录终止子”。真核终止子是指位于基因或操纵组末端的DNA片段，可中断真核系统转录作用、结合加尾酶促进聚腺苷酸化。常用的真核终止子包括SV40 poly(A)、hGH poly(A)、bGH poly(A)和rbGlob poly(A)信号。“bGH poly(A)信号”是指牛生长因子（BGH）聚腺苷酸化信号序列，“hGH poly(A)信号”是指人生长因子（hGH）聚腺苷酸化信号序列，“SV40 poly(A)信号”是指猴空泡病毒聚腺苷酸

化信号序列，rbGlob poly(A)信号是指兔β球蛋白聚腺苷酸化信号序列，均可以有效终止转录和促进mRNA聚腺苷酸化。

本申请中的“DNA表达盒”、或“目标DNA表达盒”是指包含启动子、目标DNA序列（或者用于目标DNA序列插入的克隆位点）、以及可选的bGH poly(A)信号等的表达组件。当将该DNA表达盒插入在载体中且含有目标DNA序列时，可以在一定条件下进行目标DNA序列的转录，以及可能的进一步翻译。本申请中的“DNA表达盒”、或“目标DNA表达盒”，在一些实施方式中，是指适于在真核细胞中进行DNA序列表达的表达盒。

本申请中的“目标DNA序列”可指任意的DNA序列片段，特别是感兴趣的DNA序列片段。当该“目标DNA序列”为mRNA编码序列时，可包含5'UTR、开放阅读框（ORF）和3'UTR。“开放阅读框”或“ORF”是指基因序列的一部分，包含一段可以编码蛋白的核苷酸序列，是理论上的氨基酸编码区。“目标DNA序列”还可以是目标蛋白的编码序列，目标蛋白可以是任意感兴趣的蛋白，如抗体、酶等。本申请的“目标DNA序列”可包含能够形成非B型DNA构型的DNA序列。

“体外转录”或“IVT”是指在无细胞环境下，以线性DNA为模板转录生成RNA的过程。体外转录反应通常需要纯化的含有启动子的线性DNA模板、三磷酸核糖核苷酸、含DTT和镁离子的缓冲液体系、以及适当量的RNA聚合酶。

“宿主细胞”涵盖含有载体且优选地支持所述载体的复制或者复制和表达的细胞。宿主细胞可以是原核细胞（例如大肠杆菌）或真核细胞（例如酵母、昆虫、两栖动物或哺乳动物细胞）。在一些实施方式中，宿主细胞为大肠杆菌细胞，如大肠杆菌TOP10细胞，或重组菌株GenITRv1。在一些实施方式中，宿主细胞为真核细胞，如HEK293细胞。

术语“转化”是指含有外源基因的载体（如重组质粒）将外源基因导入原核宿主细胞的技术，而“转染”是指将外源基因或载体（如重组质粒）导入真核细胞（通常为哺乳动物细胞）的技术。转化和转染一般利用感受态细胞进行，也可以应用人工脂质体法、基因枪法、显微注射法、电穿孔法等。“感受态”细胞是指经物理化学方法诱导后，处于适合摄取和容纳外来DNA的状态的细胞。例如，大肠杆菌经二价阳离子（例如CaCl₂）冷处理，可使细胞壁变得可透过质粒DNA。

“腺相关病毒”或“AAV”是细小病毒科的一种小型非致病性病毒，依赖辅助病毒进行复制。AAV的基因组大约5 kb，由一段正极性或负极性的单链DNA组成。基因组的末端是短的回文/反向末端重复序列（ITR），是基因组中唯一的顺式作用元件，可以折叠成发夹结构并作为病毒DNA复制的起点，参与基因组的整合、拯救、复制及包装等。重组腺相关病毒rAAV作为DNA递送载体，因其安全性高、免疫原性低、适用范围广、作用时间长等优点，已成为基因研究和基因治疗最重要的载体。迄今为止，许多不同血清型的AAV已从人类或灵长类动物中分离出来。例如，AAV2的基因组长度为4680个核苷酸，包含两个开放阅读框。

“反向末端重复序列”或“ITR”是指AAV基因组末端的反向重复序列，能够形成回文结构。这是重组AAV与野生AAV中唯一共有的遗传成分，在重组AAV的复制、包装、以及外源基因持续表达等方面发挥着重要作用。ITR可以按序分为7个区段、A、C、C’、B、B’、A’和D区，其中A和A’反向互补形成大的回文茎，B和B’反向互补形成小的回文臂，C和C’反向互补形成小的回文臂。在A-A’区存在一个Rep蛋白结合位点，在B-B’区或C-C’区的顶端存在另一Rep蛋白结合位点，在A-A’与D的交界处存在TRS序列。通过减少结构中的B和B’回文臂、和/或C和C’回文臂，而保留关键的一个或两个Rep蛋白结合位点、以及TRS序列，可以得到结构简化且保持有上述rAAV复制、包装、和外源基因表达功能的ITR。本申请中的“ITR衍生序列”即指能够形成完整野生ITR结构、或部分野生ITR结构，且保持有rAAV复制、包装、和外源基因表达功能的反向回文序列。

乳糖操纵子是参与乳糖分解的基因群，由调节基因（lacI）、启动子（lacP）、操纵基因（lacO）、和结构基因（lacZ、lacY、lacA）组成。本申请中的“lac启动子”或“lacP”即指乳糖操纵子中的启动子，是转录起始时RNA聚合酶的结合位点。lacP存在“泄漏”的情况，即启动子的转录不会完全“关闭”，总存在一些基础转录（F. Jacob, J. Monod (1961) *Genetic regulatory mechanisms in the synthesis of proteins*, *J. Mol. Biol.* 3: 318-356）。lacP常被构建在质粒克隆载体中，用于在细菌中表达重组蛋白，且为引入基因提供了高水平表达的能力。对于表达克隆，通常在中高拷贝数质粒（如pUC、pBluescript和pET载体及其衍生物）中构建结构，以增加重组基因的拷贝数，这在一定程度上增加了lacP的“泄漏”表达（Nielsen, Brent L., Van C. Willis, and Chin-Yo Lin. (2007) Western blot analysis to illustrate relative control levels of the lac and ara promoters in *Escherichia coli*, *Biochemistry and Molecular Biology Education* 35(2): 133-137）。

“非B型DNA构型”是指不同于典型的右螺旋双链（B型）的DNA构型，包括A型DNA构型、Z型DNA构型、H型DNA构型、G-四联体、三链体、发夹或十字形、滑移链等。研究表明，这些DNA构型至少短暂地存在于DNA复制、转录、修复或重组过程中，会阻碍这些进程的进行，增加DNA突变率。poly(A)尾的编码序列由大量腺嘌呤（A）或胸腺嘧啶（T）构成，容易形成H型DNA构型或三链体；反向重复序列如ITR或shRNA编码序列容易形成发夹或十字形结构；正向重复序列容易形成滑移链（Guiblet WM *et al.*, (2021) Non-B DNA: a major contributor to small- and large-scale variation in nucleotide substitution frequencies across the genome. *Nucleic Acids Res.* 49(3):1497-1516）。

“正向重复序列”是指一段DNA序列以相同的方向重复，两段重复之间可以包含间隔序列。正向重复序列，特别是在富含鸟嘌呤（G）时，容易在复制和转录中形成G-四联体等非B型DNA构型。

“反向重复序列”是指反方向互补的序列，中间包含或不包含间隔序列。反向重复序列的一个实例是ITR或shRNA编码序列，易在例如复制和转录中形成发夹或十

字形结构。

“单一嘌呤-单一嘧啶序列”是指一条链基本或完全为嘌呤、而互补链基本或完全为嘧啶的DNA双链序列。特殊的实例包括polyA序列、polyT序列、polyG序列、和polyC序列。这类序列可形成H型DNA构型或三链体。poly(A)尾编码序列也可归入这一类序列，包括但不限于polyT:polyA。

“Poly(A)尾”或“polyA尾”序列可以是连续的腺嘌呤构成，比如连续120个腺嘌呤构成的多聚核苷酸序列；也可以是连续的腺嘌呤中存在一段或多段非腺嘌呤间隔序列构成的分段式多聚核苷酸序列，比如由连续30个腺嘌呤、3到20个非腺嘌呤、30个腺嘌呤、3到20个非腺嘌呤和30个腺嘌呤构成的多聚核苷酸序列。

本申请中的“上游”和“下游”是指载体中在某一转录方向上分别位于目标核酸序列之前和之后（特别是紧邻目标核酸序列）的位置。当载体中存在多个启动子，从而具有多个转录方向时，目标核酸序列可以具有多个配对的“上游”和“下游”。特别地，本申请中第一终止子和第二终止子的设置，使得位于第一终止子和第二终止子之间的目标核酸序列，不会随着载体中目标核酸序列之外发生的转录行为而转录。

成熟mRNA有着较长的poly(A)尾结构，在维持mRNA的稳定性、调节mRNA的翻译效率以及维持mRNA的运输中发挥重要的作用。针对mRNA的批量合成，目前较为高效的方式为体外转录，而IVT中对于poly(A)加尾主要有两种方法。第一种为酶法合成，在以线性DNA模板转录出mRNA后，通过添加poly(A)聚合酶，催化由ATP转化成的AMP添加到单链RNA的3'末端，形成Poly(A)尾。该方法的优点是，转录模板无需含有Poly(A)编码序列，操作简单，但存在加尾长度不稳定的局限性。第二种方法为共转录，即由已经存在于模板质粒DNA或者PCR产物上的poly(A)尾编码序列直接转录形成，其优势在于无需使用Poly(A)聚合酶，可减少工艺步骤，节约成本，并且转录加尾长度相对稳定。但是在上游质粒的制备过程中，需要减少培养过程中poly(A)尾编码序列的缺失，这是提供高产优质且均一性好的质粒模板的关键所在。

本申请的发明人，在实验中发现，通过i) 增加载体中复制起点与poly(A)尾编码序列的间隔距离，ii) 在poly(A)尾编码序列、或含有poly(A)尾编码序列的mRNA表达盒的上下游添加终止子，和/或iii) 破坏载体中原有的lac启动子功能区，可以显著提高poly(A)尾编码序列的扩增和/或表达（转录）稳定性，从而提高质粒体外转录的纯度、和/或产出mRNA的蛋白表达量。此外，通过对某些中低拷贝载体复制起点序列的改造，还能够进一步提高载体的产量。

因而，本申请提供一种载体或载体骨架，其可以包含：

- i) 复制起点，以及
- ii) a) 目标DNA表达盒，其包含启动子和目标DNA序列，b) 目标DNA序列，c) 克隆位点，d) 克隆位点和能够形成非B型DNA构型的DNA序列，e) DNA表达

盒，其包含启动子和克隆位点，或f) DNA表达盒，其包含启动子、克隆位点和能够形成非B型DNA构型的DNA序列。

在ii)中包含a)、e)或f)的载体或载体骨架，可以作为克隆载体使用，也可以作为表达载体使用。a)的目标DNA序列可以为任何序列，可以包含形成B型DNA构型的序列、和/或形成非B型DNA构型的序列。在e)的克隆位点中可以插入目标DNA序列，该目标DNA序列可以包含可以形成非B型DNA构型的DNA序列。在f)的克隆位点中可以插入任何序列，包括形成B型DNA构型的序列、和/或形成非B型DNA构型的序列。

在ii)中包含b)、或d)的载体或载体骨架，可以仅作为克隆载体使用。b)的目标DNA序列可以为任何序列，可以包含形成B型DNA构型的序列、和/或形成非B型DNA构型的序列。在d)的克隆位点中可以插入任何序列，包括形成B型DNA构型的序列、和/或形成非B型DNA构型的序列。

在ii)中包含c)的载体骨架中，取决于克隆位点内待插入的序列，可以用作克隆载体、或克隆/表达载体。在克隆位点内插入启动子和目标DNA序列的情况下，载体可用作克隆/表达载体。

复制起点可以在复制方向上与a)目标DNA表达盒中的目标DNA序列、b)目标DNA序列、c)克隆位点、d)能够形成非B型DNA构型的DNA序列、e)DNA表达盒中的克隆位点、或f)DNA表达盒中的能够形成非B型DNA构型的DNA序列间隔至少150 bp、200 bp、300 bp、500 bp、或700 bp。间隔至少150 bp、200 bp、300 bp、500 bp、或700 bp的目的在于，将复制起点在复制或转录解链时对载体中存在的可形成非B型DNA构型的DNA序列，例如poly(A)尾的编码序列，的影响尽量降至最低。可形成非B型DNA构型的DNA序列，可能会在DNA解链时形成非B型DNA构型，从而阻碍复制或转录的进行，并增加自身的突变，如删除等。随着解链时间的延长，造成非B型DNA突变的几率也越大，而复制起点处的解链时间，由于蛋白结合等，可能会比别处更长。因而，延长复制起点与非B型DNA构型的DNA序列的间隔距离，可减少非B型DNA构型的DNA序列的突变。本申请的发明人经实验发现，将复制起点与可形成非B型DNA构型的DNA序列的间隔序列从100 bp增加至247 bp、500 bp、甚至700 bp时，在复制过程中保持完整poly(A)尾编码序列的载体的比例可以从0.00%提高至3.13%、46.67%、甚至83.3%。复制起点可以使得载体在合适的宿主细胞中复制。合适的宿主细胞可以是原核细胞如大肠杆菌、或真核细胞。在一些实施方式中，复制起点可以为pUC复制起点，其可以包含例如SEQ ID NO: 23的序列。

在例如穿梭质粒载体的情况下，载体或载体骨架还可以包含第二复制起点。同理，该第二复制起点可以在复制方向上与a)目标DNA表达盒中的目标DNA序列、b)目标DNA序列、c)克隆位点、d)能够形成非B型DNA构型的DNA序列、e)DNA表达盒中的克隆位点、或f)DNA表达盒中的能够形成非B型DNA构型的DNA序列

间隔至少150 bp、200 bp、300 bp、500 bp、或700 bp。该第二复制起点可以使得载体在合适的宿主细胞中复制，合适的宿主细胞可以是原核细胞如大肠杆菌、或真核细胞。

取决于载体或载体骨架中复制起点的种类，载体或载体骨架还可以相应地包含Rep蛋白表达元件，包括Rep蛋白编码序列和相应的启动子，使得载体或载体骨架成为宽宿主类型。

载体或载体骨架还可以包含第一终止子和第二终止子，分别位于ii)的序列，即a)目标DNA表达盒、b)目标DNA序列、c)克隆位点、d)克隆位点和能够形成非B型DNA构型的DNA序列、e)DNA表达盒的上游和下游，并设置成使得载体或载体骨架中在上述序列之外的转录终止于这些序列的上下游。这样的设计可以使得大大减少在ii)的区域内的解链，从而保持这些区域中的序列的稳定性，特别是可形成非B型DNA构型的DNA序列如poly(A)尾编码序列的稳定性。本申请的发明人经实验发现，当复制起点与ii)中的特定序列间隔距离为247 bp，而在ii)的序列的两端加入终止子时，复制过程中保持完整poly(A)尾编码序列的载体的比例可以从3.13%提高到37.14%。第一终止子可以为正向终止子、双向终止子、或反向终止子。第二终止子可以为正向终止子、双向终止子、或反向终止子。在一些实施方式中，第一终止子和第二终止子可以均为双向终止子。“正向终止子”是指能够使终止子所在链的转录终止的终止子，“反向终止子”是指能够使终止子所在链的互补链上的转录终止的终止子，“双向终止子”是指能够使得终止子所在链及其互补链上的转录均终止的终止子。可以根据载体或载体骨架的具体结构，例如筛选标记的转录方向，来设置终止子。可以使用任何终止子，只要能够确保由载体或载体骨架中ii)的序列区域以外的启动子引起的转录终止于ii)的序列的上下游，使得在ii)的序列区域以外发生的转录不会引起这些区域的解链。在通用载体骨架中，可以设置双向终止子。在一些实施方式中，第一终止子可以为rrnB T1终止子和rrnB T2终止子。在一些实施方式中，第二终止子可以为lambda t0终止子。rrnB T1终止子、rrnB T2终止子、和lambda t0终止子为本领域所知，可以分别包含SEQ ID NO: 24、25、或26所示的核苷酸序列。

载体或载体骨架中的克隆位点可以为单克隆位点、或多克隆位点。克隆位点可以根据实际需要设置。

载体或载体骨架还可以包含筛选标记。筛选标记可以为抗生素抗性基因，例如氨苄青霉素抗性基因、四环素抗性基因、氯霉素抗性基因、或卡那霉素抗性基因等。特别地，筛选标记不存在于载体或载体骨架待转化/转染的宿主细胞的基因组中。在一些实施方式中，筛选标记为卡那霉素抗性基因。该载体或载体骨架还可以在筛选标记的上游包含用于该筛选标记的表达的启动子。筛选标记除可以用于确认载体是否导入宿主细胞外，还可能在一定情况下促进载体的复制、和/或表达。例如，当载体或载体骨架包含卡那霉素抗性基因，而宿主细胞的培养基中含有卡那霉素

时，载体的复制可能会加速。

载体或载体骨架可以为环状。载体或载体骨架可以为非病毒载体或非病毒载体骨架。载体或载体骨架可以为质粒载体或质粒载体骨架。在一些实施方式中，载体或载体骨架可以为环状质粒载体或环状载体骨架。

鉴于lac启动子的“泄露性”，特别地，载体或载体骨架中不包含功能完整的lac启动子。发明人经实验证明，在去除lacP功能区活性后，载体复制或传代过程中保持完整poly(A)尾编码序列的载体的比例可以从31.25%提高到50%。即便载体或载体骨架中已经包含两个终止子，由于终止子不能百分百保证终止转录，lac启动子区的失活也是有利的。

在一些实施方式中，本申请的上述载体或载体骨架可以构造成复制、和/或表达mRNA及其poly(A)尾。

除在克隆位点插入所需序列外，也可以通过同源重组等方法在本申请的载体或载体骨架中加入感兴趣的DNA序列，如可形成非B型DNA构型的DNA序列、特定的mRNA编码序列、启动子等。通过同源重组等方法得到的载体或载体骨架，也在本申请的保护范围内。

重组腺相关病毒rAAV作为DNA递送载体，因其安全性高、免疫原性低、适用范围广、作用时间长等优点，已成为基因研究和基因治疗最重要的载体。rAAV的两端包含野生AAV的ITR、或其衍生序列，是rAAV中唯一的顺式作用元件，参与rAAV的复制、包装和外源基因表达等。

然而，ITR的回文序列会形成一种高度稳定的T型发夹结构，加上GC含量较高，容易在细菌繁殖过程中发生突变或重组丢失。ITR完整度对rAAV高效复制和包装的重要性。单分子实时测序，将rAAV基因组作为单个完整的分子进行综合分析，发现HEK293三质粒包装系统中，缺失的ITR与非载体序列的包装有关。ITR缺失对于rAAV病毒包装及rAAV应用于基因治疗的影响，情形复杂，通过载体-宿主相互作用会产生意想不到的生物学后果。如果在临床前阶段使用不完整的ITR序列，将降低实验室内部和跨实验室实验的可重复性，并可能导致在使用完整ITR序列进行临床试验时的实验重复性，延迟转化为临床应用。由于ITR在某些情况下在载体包装过程中具有自我修复的能力，因此ITR的部分损失可能未被注意到，这一事实进一步加剧了这一问题（Wilmott, P., Lisowski, L., Alexander, I. E., & Logan, G. J. (2019). A user's guide to the inverted terminal repeats of adeno-associated virus. *Human gene therapy methods*, 30(6): 206-213）。

目前没有较好的方法来避免在rAAV-ITR质粒构建和大肠杆菌扩增过程中发生ITR缺失。一些重组缺陷的商业菌株如JC8111、SURE2、Stab13、XL10-Gold经常用于含ITR结构的质粒的构建扩增，但并不总是有效。

本申请的发明人意外地发现，通过i) 延长载体中复制起点与ITR序列的间隔距离，ii) 在ITR的上下游添加终止子，和/或iii) 破坏载体中原有的lac启动子功能区，

可以显著提高ITR序列的扩增和/或表达（转录）稳定性。此外，上述对于载体的改动还能提高载体产量，和/或重组腺相关病毒（rAAV）的产量例如包装滴度。

因而，本申请提供一种载体或载体骨架，其可以包含：

i) 复制起点，以及

ii) a) 目标DNA表达盒，其按序包含第一ITR或其衍生序列、启动子、目标DNA序列、和第二ITR或其衍生序列，b) 目标DNA序列组，其按序包含第一ITR或其衍生序列、目标DNA序列、和第二ITR或其衍生序列，c) DNA序列组，其按序包含第一ITR或其衍生序列、克隆位点、和第二ITR或其衍生序列，或d) DNA表达盒，其按序包含第一ITR或其衍生序列、启动子、克隆位点、和第二ITR或其衍生序列。

在ii) 中含有a) 和d) 的载体或载体骨架，可以用作复制载体，也可以在AAV包装后进行目标DNA序列的表达。在d) 的克隆位点处可以插入目标DNA序列，例如mRNA编码序列、或mRNA编码序列加poly(A)尾编码序列。目标DNA序列可以是任何序列，包括可形成B型DNA构型的DNA序列、和/或可形成非B型DNA构造的DNA序列。

在ii) 中包含b) 的载体，可以用作复制载体。目标DNA序列可以是任何序列，包括可形成B型DNA构型的DNA序列、和/或可形成非B型DNA构造的DNA序列。

在ii) 中含有c) 的载体骨架，可以在克隆位点处插入目标DNA序列，成为克隆载体。目标DNA序列可以是任何序列，如可形成B型DNA构型的DNA序列、和/或可形成非B型DNA构造的DNA序列等。也可以在克隆位点处插入启动子和目标DNA序列，成为克隆/表达载体。目标DNA序列可以是任何序列，如可形成B型DNA构型的DNA序列、和/或可形成非B型DNA构造的DNA序列等。

复制起点可以在复制方向上与a) 目标DNA表达盒、b) 目标DNA序列组、c) DNA序列组、或d) DNA表达盒间隔至少150 bp、200 bp、300 bp、500 bp、或700 bp。间隔至少150 bp、200 bp、300 bp、500 bp、或700 bp的目的在于，将复制起点在复制或转录解链时对载体中存在的可形成非B型DNA构型的DNA序列，例如ITR，的影响尽量降至最低。ITR，可能会在DNA解链时形成发夹或十字形结构，从而阻碍复制或转录的进行，并增加自身的突变，如删除等。随着解链时间的延长，造成ITR突变的几率也越大，而复制起点处的解链时间，由于蛋白结合等，可能会比别处更长。因而，延长复制起点与ITR的间隔距离，可减少ITR的突变。如a) 目标DNA表达盒、b) 目标DNA序列组、c) DNA序列组、或d) DNA表达盒还包含其他可形成非B型DNA构型的DNA序列，通过拉长与复制起点的距离，也可以减少其突变率。发明人经实验表明，当将载体中的复制起点与筛选标记交换位置，拉长复制起点在复制方向上与ITR的间隔距离时，复制带来的载体酶切杂带比例可以从平均60.63%降低至9.96%。在对载体删除lacP功能区的实验中，也可以看出，lacP与ITR距离越近，由例如转录引起的ITR序列不稳定性就越高，这从侧面印证了复制起点与ITR的间隔距离对ITR序列稳定性的影响。复制起点可以使得载体在合适的宿主细胞中

复制。合适的宿主细胞可以是原核细胞如大肠杆菌、或真核细胞。在一些实施方式中，复制起点可以为pUC复制起点，其可以包含例如SEQ ID NO: 23的序列。

在例如穿梭质粒载体的情况下，载体或载体骨架还可以包含第二复制起点。同理，该第二复制起点可以在复制方向上与a) 目标DNA表达盒、b) 目标DNA序列组、c) DNA序列组、或d) DNA表达盒间隔至少150 bp、200 bp、300 bp、500 bp、或700 bp。该第二复制起点可以使得载体在合适的宿主细胞中复制，合适的宿主细胞可以是原核细胞如大肠杆菌、或真核细胞。

取决于载体或载体骨架中复制起点的种类，载体或载体骨架还可以相应地包含Rep蛋白表达元件，包括Rep蛋白编码序列和相应的启动子，使得载体成为宽宿主类型。

载体或载体骨架还可以包含第一终止子和第二终止子，分别位于a) 目标DNA表达盒、b) 目标DNA序列组、c) DNA序列组、或d) DNA表达盒的上游和下游，并设置成使得载体或载体骨架中a) 目标DNA表达盒、b) 目标DNA序列组、c) DNA序列组、或d) DNA表达盒之外的转录终止于a) 目标DNA表达盒、b) 目标DNA序列组、c) DNA序列组、或d) DNA表达盒的上下游。这样的设计可以使得在a) 目标DNA表达盒、b) 目标DNA序列组、c) DNA序列组、或d) DNA表达盒以外发生的转录不会引起a) 目标DNA表达盒、b) 目标DNA序列组、c) DNA序列组、或d) DNA表达盒区域内的解链，从而保持该区域中的序列的稳定性，特别是ITR的稳定性。发明人经实验发现，当在ITR两端加上终止子后，复制带来的载体酶切杂带比例可以从平均9.96%进一步降低至7.31%，或者从21.90%降低至17.42%。第一终止子可以为正向终止子、双向终止子、或反向终止子。第二终止子可以为正向终止子、双向终止子、或反向终止子。在一些实施方式中，第一终止子和第二终止子可以均为双向终止子。可以根据载体的具体结构，例如筛选标记的转录方向，来设置终止子。可以使用任何终止子，只要能够确保载体中a) 目标DNA表达盒、b) 目标DNA序列组、c) DNA序列组、或d) DNA表达盒之外的转录终止于该区域的上下游，使得在该区域以外发生的转录不会引起该区域内的解链。在通用载体骨架中，可以设置双向终止子。在一些实施方式中，第一终止子可以为rrmB T1终止子和rrmB T2终止子。在一些实施方式中，第二终止子可以为lambda t0终止子。rrmB T1终止子、rrmB T2终止子、和lambda t0终止子为本领域所知，可以分别包含SEQ ID NO: 24、25、或26所示的核苷酸序列。

载体或载体骨架中的克隆位点可以为单克隆位点、或多克隆位点。克隆位点可以根据实际需要设置。

载体或载体骨架还可以包含筛选标记。筛选标记可以为抗生素抗性基因，例如氨苄青霉素抗性基因、四环素抗性基因、氯霉素抗性基因、或卡那霉素抗性基因等。优选地，筛选标记不存在于载体或载体骨架待转化/转染的宿主细胞的基因组中。在一些实施方式中，筛选标记为卡那霉素抗性基因。该载体或载体骨架还可以在筛

选标记的上游包含用于该筛选标记的表达的启动子。筛选标记除可以用于确认载体是否导入宿主细胞外，还可能在一定情况下促进载体的复制。例如，当载体或载体骨架包含卡那霉素抗性基因，而宿主细胞的培养基中含有卡那霉素时，载体的复制可能会加速。

载体或载体骨架可以为环状。载体或载体骨架可以为非病毒载体或非病毒载体骨架。载体或载体骨架可以为质粒载体或质粒载体骨架。在一些实施方式中，载体或载体骨架可以为环状质粒载体或环状质粒载体骨架。

鉴于lac启动子的“泄露性”，优选载体或载体骨架中不包含功能完整的lac启动子。实验证明，在去除lacP功能区活性后，复制、传代或包装后保持完整ITR的载体的比例可以从平均50.00%提高到75.00%。即便载体或载体骨架中已经包含两个终止子，由于终止子不能保证百分百终止转录，lac启动子区的失活也是有利的。

第一ITR或其衍生序列、与第二ITR或其衍生序列可以包含相同、或不同的核苷酸序列。第一ITR和第二ITR可以为野生型的AAV-ITR，包含例如SEQ ID NO: 9所示的核苷酸序列。第一ITR衍生序列和第二ITR衍生序列可以形成与天然AAV-ITR相似的回文结构，例如包含一个回文臂和一个回文茎、或仅包含一个回文茎，且保持有Rep蛋白结合力。在一些实施方式中，第一ITR衍生序列和第二ITR衍生序列可以包含SEQ ID NO: 10、11或12所示的核苷酸序列。

在一些实施方式中，本申请的上述载体或载体骨架可以构造成复制、和/或表达带有ITR的载体，或者用于复制和包装rAAV载体。

可以通过同源重组等方法在本申请的载体或载体骨架中加入感兴趣的DNA序列，如可形成非B型DNA构型的DNA序列，如ITR等。通过同源重组等方法得到的载体或载体骨架，也在本申请的保护范围内。

在本申请中，发明人发现并验证了i) 拉长复制起点与可形成非B型DNA构型的DNA序列的间隔距离，ii) 在可形成非B型DNA构型的DNA序列或其表达盒两端添加终止子，以及iii) 去除载体中的lacP功能区，可以显著地poly(A)尾编码序列以及ITR序列的复制和/或表达稳定性，并相应提升载体的产量、和/或体外转录纯度等。发明人有理由相信，上述这些对于载体的改造，也同样有利于其他正向重复序列、反向重复序列、单一嘌呤-单一嘧啶序列的复制和/或表达稳定性。

根据本申请的载体设计，可以以从头合成的方式等制备载体。

此外，也可以根据本申请的载体设计，对已有的载体做出改造。例如，可以在既有载体或载体骨架中，i) 根据具体情况，使得载体或载体骨架中的复制起点与克隆位点、或可形成非B型DNA构型的DNA序列间隔至少150 bp、200 bp、300 bp、500 bp或700 bp，ii) 在含有lac启动子的情况下，使得lac启动子失活，和/或iii) 根据具体情况，在载体或载体骨架中克隆位点、可形成非B型DNA构型的DNA序列、或可形成非B型DNA构型的DNA序列的表达盒的上下游分别加上终止子，使得载体或载体骨架中在上述区域之外的转录终止于这些区域的上下游。

本申请的载体，包括制备得到的、和改造得到的，可以用于复制和/或表达目标DNA序列，特别是含有正向重复序列、单一嘌呤-单一嘧啶序列、或反向重复序列的目标DNA片段。正向重复序列或单一嘌呤-单一嘧啶序列可以为例如poly(A)尾编码序列、polyA、polyG、polyC、或polyT。反向重复序列可以为例如ITR、或shRNA编码序列。

在使用本申请的载体复制和/或表达目标DNA序列时，涉及载体对宿主细胞的转化或转染、宿主细胞的培养、体外转录等，其相应的条件为本领域技术人员所知，可以根据实际情况而调整。

本申请的载体，包括制备得到的、和改造得到的，可以用于制备重组腺相关病毒载体（rAAV）。具体地，可以i）提供本申请的两端为ITR的目标DNA序列组、或目标DNA表达盒的载体，或在本申请的两端为ITR的目标DNA序列组、或目标DNA表达盒的克隆位点处插入待复制和/或表达的DNA序列，获得新的载体；ii）可选地，将这些载体转化/转染到宿主细胞中，在利于载体复制的条件下培养宿主细胞；以及iii）收获i）或ii）的载体，将其与包装质粒、辅助质粒共转化/转染宿主细胞，在利于载体包装的条件下培养宿主细胞。在不需要扩增的情况下，步骤ii）可以省略，而直接进行三质粒包装。上述步骤中宿主细胞的培养条件，包装质粒和辅助质粒的选择，为本领域技术人员所知，可以根据实际情况调整。腺相关病毒载体的包装也可以采用三质粒包装以外的方法，只要能达到载体包装的目的即可。

除非另有定义，本文中使用的所有技术和科学术语均具有本发明所属领域普通技术人员所通常理解的含义。

如无特殊说明，以下序列均为载体正义链上的序列。

实施例1：调整载体中复制起点（Ori）的位置以提高Poly(A)尾编码序列的稳定性

Ori序列通常富含腺嘌呤（A）和胸腺嘧啶（T），富含A-T的DNA序列更易解链，解链的Ori区域给复制起始提供了蛋白结合的空间，从而起始质粒复制。而且，文献表明，Ori元件可能存在启动活性。当Ori的3'端邻近poly(A)尾编码序列时，在质粒复制过程中，该编码序列极易发生缺失或突变。

取pUC57-kana-T1/T2终止子载体（南京金斯瑞，货号：C240PHC090-2），构建插入有eGFP-120 A-mRNA编码序列（含有120个连续腺嘌呤）、且Ori 3'端与poly(A)尾结构序列之间间隔距离不等的质粒载体，来测试Ori序列的3'端与poly(A)尾结构序列之间的距离对于poly(A)尾结构序列稳定性的影响。构建的载体图谱如图1所示，其中Ori序列的3'端与poly(A)尾结构序列之间为100 bp、247 bp、500 bp、或700 bp的间隔序列（序列分别如SEQ ID NO: 1-4所示，其中两端斜体下划线的序列为用于与pUC57-kana-T1/T2终止子载体进行同源重组的序列，中间为随机序列）。

具体地，取pUC57-kana-T1/T2终止子载体1.5 ng，加入10×cutsmart缓冲液5 μL、XhoI 2.5 μL、和ApaLI 2.5 μL，加水至50 μL，在37℃下孵育1 h，以XhoI和ApaLI酶切pUC57-kana-T1/T2终止子载体。获得的线性化载体产物经AxyPrep-96 DNA凝胶

回收试剂盒（爱思进生物技术有限公司）回收。之后，取回收的线性化载体0.8 ng，加入0.8 ng人工合成的上述各间隔序列（两端包含与载体两端同源的同源臂序列，30-50 bp）、Genbuilder plus酶+缓冲液mix（江苏金斯瑞生物科技有限公司）10 μ L，加水至20 μ L，在50°C下孵育60 min，以利用同源重组的方式将间隔序列片段连入线性化载体片段中。用获得的连接产物转化大肠杆菌TOP10感受态细胞，在LB平板上于37°C培养12 h。

对菌落进行PCR筛选。具体地，菌落PCR反应体系（试剂均来自江苏金斯瑞生物科技有限公司）包含正向引物 1 μ L、反向引物1 μ L、Taq DNA聚合酶 1 μ L、10 $\times$ 缓冲液1 μ L、dNTP 1 μ L，加水至10 μ L，在PCR仪（2720 Thermo Cycler）中执行以下程序，95°C 1 min；95°C 5 s，55°C 15 s，72°C 1 min，30个循环；72°C 1 min；4°C 储存。每种重组质粒载体筛选48个插入有间隔序列的克隆，分别接种到新鲜的4 mL LB液体培养基中，37°C 12 h震荡培养。

使用AxyPrep质粒DNA小量试剂盒（爱思进生物技术有限公司），对培养的菌液进行质粒抽提，之后由Applied Biosystems公司3730XL测序仪对获得的质粒进行Sanger法测序。测序结果经过SnapGene软件比对分析结构，当测序序列与标准序列比对一致，且Poly(A)尾测序底峰不超过主峰的20%，判断为阳性克隆。以此计算每种重组质粒载体构建的单克隆成功率，即阳性克隆数占总挑斑克隆数的比例。图2示出阳性克隆（C，poly(A)尾编码序列较纯且测序底峰低于主峰20%）和失败克隆（A，poly(A)尾编码序列缺失；B，poly(A)尾编码序列不纯且测序底峰高于主峰20%）的测序峰形。结果如图3所示，当Ori 3'距离poly(A)尾编码序列的间隔长度为100 bp时，质粒单克隆成功率为0%，即质粒的poly(A)尾编码序列测序纯度较差，测序底峰较高；随着间隔变长，质粒的克隆成功率上升；当间隔长度为700 bp时，质粒单克隆成功率可达83.33%，即Poly(A)尾编码序列的测序纯度较高，测序底峰低于poly(A)尾主峰的10%。这表明，Ori距离poly(A)尾编码序列位置越远，质粒的poly(A)尾编码序列的稳定性越高。

实施例2：添加终止子元件以提高载体中Poly(A)尾编码序列的稳定性

取实施例1中构建的间隔序列为247 bp的重组质粒载体，用247 bp含有lambda t0终止子的序列（序列如SEQ ID NO: 5所示，其中中部加粗下划线的序列为lambda t0终止子，两端下划线斜体的部分用于同源重组）替换质粒中的247 bp间隔序列，质粒图谱如图4所示。

如实施例1所述，用获得的质粒连接产物转化大肠杆菌TOP10感受态细胞，在LB平板上于37°C培养12 h；对菌落进行PCR筛选；将筛选的插入终止子的克隆接种到新鲜的4 mL LB液体培养基中，37°C 12 h震荡培养；对培养的菌液进行质粒抽提，进行Sanger法测序和分析。

结果如图5（A-C）所示，在Ori与poly(A)尾编码序列之间添加终止子元件，可使单克隆成功率从3.13%提升至37.14%，相比于247 bp的无终止子间隔序列，质粒

测序底峰低，Poly(A)尾编码序列的纯度更高。这表明，终止子元件的添加，可以通过例如阻断载体骨架中泄露的启动活性对于下游功能区的影响，提高poly(A)尾编码序列的稳定性。

实施例3：删除lac启动子功能区以提高载体中Poly(A)尾编码序列的稳定性

如实施例1所述构建插入有eGFP-120 A-mRNA编码序列的载体，其中在未做其他改造的pUC57-kana-T1/T2终止子载体中，Ori与poly(A)尾编码序列之间为总长465 bp的lac启动子功能区，其从5'至3'方向包含CAP结合位点、lac启动子、乳糖操纵子和M13 rev元件。结果发现，质粒的单克隆构建成功率为31.25%。

通过碱基替代，将lac启动子功能区的序列改为SEQ ID NO: 6，使功能区失活，构建出新的载体，图谱如图6所示。之后，筛选lac启动子功能区序列改变的克隆，LB培养基震荡培养，进行质粒抽提，并测序分析。

结果显示在图7中。lac启动子功能区的失活可将单克隆成功率从31.25%提升至50%，提高质粒载体中poly(A)尾编码序列的稳定性。

实施例4：GS-MK、GS-PV和GS-CMV系列载体的构建

根据实施例1、2和3的结果，构建优化的载体。

具体地，在pUC57-kana质粒载体中删除M13 rev、lac启动子、乳糖操纵子、和CAP结合位点等非必要元件，在多克隆位点上下游分别增加T1/T2终止子和lambda t0终止子元件，并拉大Ori的3'端与多克隆位点的间隔，使得在多克隆位点插入序列如mRNA编码序列后，Ori 3'端与poly(A)编码序列间隔大于700 bp。

此外，对载体骨架上的BsaI、BspQI、BspMI、BpiI、BsmBI等IIS位点进行同义突变，去除这些酶切位点，而在待插入的目标DNA序列，如mRNA编码序列和poly(A)尾编码序列的端部加入IIS酶识别位点。如此，插入目标DNA序列的载体在线性化后，IIS型限制性内切酶仅在poly(A)尾编码序列之外进行识别，并在poly(A)编码序列内切割，从而得到较纯的酶切片段，且酶切片段中的polyT序列不会加入非T序列，并从而得到纯的poly(A)尾，纯poly(A)尾可以维持较高的mRNA的蛋白表达水平（Holtkamp S *et al.*, (2006) Modification of antigen-encoding RNA increases stability, translational efficacy, and T-cell stimulatory capacity of dendritic cells. *Blood*. 108(13):4009-4017）。这种设计也便于载体进行Golden Gate克隆。

最终构建得到的载体称为GS-MK载体，自5'至3'方向包含：复制起点、rrnB T1终止子和rrnB T2终止子元件、多克隆位点（MCS）、lambda t0终止子元件、和卡那霉素抗性元件。质粒图谱见图8。

类似地，以PVAX1载体出发，同义突变载体上的BsaI、BspQI、BspMI和SpeI酶切位点，在CMV启动子的上游以及BGH poly(A)尾信号的下流分别插入T1/T2终止子和lambda t0终止子。最终构建的GS-PV载体自5'至3'方向包含：复制起点、rrnB T1终止子和rrnB T2终止子、CMV增强子和启动子、多克隆位点、bGH poly(A)信号、lambda t0终止子、和卡那霉素抗性元件，图谱如图9所示。

此外，基于GS-PV载体，修复原PVAX1载体复制起点中相比pUC复制起点的一处T-C突变。新得到的载体称为GS-CMV。该载体可提高质粒的拷贝数，降低质粒抽提时出现的基因组与产量不足的风险。

实施例5：GS-MK、GS-PV和GS-CMV系列载体的功能测试

在GS-MK、GS-PV和GS-CMV的多克隆位点处插入多种poly(A)尾编码序列，包括100个腺嘌呤（100个A）、120个腺嘌呤（120个A）、分段式的30个腺嘌呤+30个腺嘌呤+43个腺嘌呤（30个A+3个非A核苷酸+30个A+3个非A核苷酸+43个A）、和分段式的31个腺嘌呤+71个腺嘌呤（31个A+8个非A序列+71个A）。

具体地，使用NheI和ApaI酶切GS-MK载体，得到线性化载体。人工合成待插入的基因片段，该片段自5'至3'方向包含：多克隆位点、eGFP序列（该序列的作用是增加Ori 3'与poly(A)尾编码序列的间隔碱基数，使两者距离>700 bp）、上述的各种poly(A)尾编码序列和IIS酶识别位点。使用Gibson的方式得到重组质粒。以GS-MK-100A-BspQI为例，插入基因序列如SEQ ID NO: 27所示。

类似地，使用NheI和ApaI酶切GS-PV载体或GS-CMV载体，得到线性化载体。人工合成待插入的基因片段，该片段自5'至3'方向包含：多克隆位点、上述各种poly(A)尾编码序列和IIS酶识别位点。使用Gibson的方式得到重组质粒。以GS-CMV-100A-BspQI为例，插入基因序列如SEQ ID NO: 28所示。

如上所述，用获得的插入有poly(A)尾编码序列的质粒转化大肠杆菌TOP10感受态细胞，在LB平板上于37°C培养12 h；对菌落进行PCR筛选；将筛选的插入poly(A)尾编码序列的克隆接种到新鲜的4 mL LB液体培养基中，37°C 12 h震荡培养；对培养的菌液进行质粒抽提，进行Sanger法测序和分析。

三种载体的测序峰形图分别显示在图10-图12中。综合来看，质粒测序的底峰较低，Poly(A)尾编码序列的纯度较高。

实施例6：GS-MK、GS-PV和GS-CMV系列载体与商业化载体的对比

在实施例4中构建的GS-MK、GS-PV、GS-CMV载体，以及商业化的Puc57-kana、PVAX1载体中，插入eGFP-100A mRNA编码序列，分析对比不同载体的单克隆成功率、poly(A)尾编码序列的传代稳定性、质粒的产量、体外转录（IVT）的纯度、和mRNA真核蛋白的表达水平。插入上述载体的eGFP-100A mRNA编码序列如SEQ ID NO: 29所示，包含T7启动子、5'UTR、eGFP编码序列、3'UTR、poly(A)尾编码序列、和线性化酶切位点。

根据图谱，Puc57-kana中含有LacZ，不含终止子，其中eGFP-100A mRNA编码序列插入在多克隆位点，插入后poly(A)尾编码序列和复制起点3'的间隔约为400 bp。根据图谱，pVAX1载体中不含终止子，其中eGFP-100A mRNA编码序列插入在多克隆位点，插入后poly(A)和复制起点3'的间隔约为700 bp。

poly(A)尾编码序列传代稳定性的检测方法如下，即将质粒转化感受态细胞后涂布平板，挑取单克隆菌落接种至LB培养基培养，37°C、150 rpm·min⁻¹的条件下培

养12 h后（第一代菌液），转接100微升菌液进入50 mL的LB培养基中传代培养，在37°C、150 rpm·min⁻¹的条件下培养12h后（第二代菌液），转接100微升菌液进入50 mL的LB培养基中传代培养，37°C、150 rpm·min⁻¹的条件下培养12h后（第三代菌液），取2 mL菌液进行质粒抽提并进行Sanger测序验证，获得阳性克隆的概率，传代稳定性即阳性克隆数占总挑斑克隆数的比例。

质粒产量的测量方法如下，即将质粒转化感受态细胞后涂布平板，挑取单克隆菌落接种至400 mL 3×LB培养基，在37°C、150 rpm·min⁻¹的培养箱摇菌13 h，富集菌液，抽提质粒获得的DNA产量。

IVT和纯度检测描述如下。具体地，在包含100 µg质粒、10 µL BspQI酶、30 µL cutsmart缓冲液、补足水至300 µL的300 µL酶切体系中，37°C过夜酶切，使用AxyPrep-96 DNA凝胶回收试剂盒回收线性化产物，按照mMessenger mMachine T7 Ultra体外转录试剂盒的说明书，将质粒线性化产物进行体外转录并纯化，利用Agilent 5200片段分析仪检测IVT产物的纯度。

最后，测试mRNA的真核表达水平。具体地，接种293T细胞，在37°C、5%CO₂培养箱培养至汇合度达70-80%，按照Lipofectamine 2000试剂盒的说明书进行细胞转染操作，配置转染复合液A液包含5 µL Opit-和0.5 µL Lip2000，B液包含5 µL Opit-MEM和0.1 µg mRNA，A液与B液混合后室温孵育5 min，将从上述IVT得到的mRNA产物加入到细胞中，在培养箱中继续培养。转染结束24 h、48 h后，采用倒置荧光显微镜观察GFP荧光转染效果，并利用Image J软件进行GFP荧光读值。

实验数据表明，GS-MK、GS-PV和GS-CMV载体的单克隆成功率显著高于pUC57-kana和PVAX1载体（图13）；GS-MK、GS-PV和GS-CMV载体的重组质粒poly(A)尾编码序列传代稳定性显著高于pUC57-kana（图14）；在相同的摇菌发酵条件下，GS-MK、GS-CMV载体的重组质粒产量显著高于PVAX1和GS-PV载体（图15）；并且由于GS-MK、GS-PV和GS-CMV载体质粒具有更高的Poly(A)尾编码序列稳定性，其IVT纯度和真核表达蛋白效果也显著优于pUC57-Kana（图16和17）。

实施例7：删除lac启动子功能区以提高载体中ITR结构稳定性

选择4个含有lac启动子（lacP）功能区元件（SEQ ID NO: 7）和AAV ITR的质粒，质粒图谱如图18所示，lacP在启动方向上距离最近的ITR的长度记录在表1中。

用SEQ ID NO: 8替代lac启动子序列，删除lacP功能区中的lac操纵子、lac启动子、和CAP结合位点，构建新的质粒。

将构建前后的质粒分别转化GenITRV1菌株（南京金斯瑞），37°C过夜培养，并挑选8个单克隆，LB 37°C培养12 h后，抽提质粒，使用ITR特殊试剂盒（南京金斯瑞）处理质粒，并进行Sanger测序。ITR区域内无缺失或突变的克隆判断为Sanger测序通过，ITR区域Sanger测序出现缺失或突变的，则判定为不通过。统计lacP缺失前后各质粒的Sanger测序通过率，并总结在表1中。

表1. 删除lacP元件功能区前后的质粒克隆Sanger测序通过率

	lacP到ITR的距离	Sanger测序通过率	
		lacP删除前	lacP删除后
质粒1	66 bp	62.50%	87.50%
质粒2	42 bp	25%	50%
质粒3	66 bp	62.50%	87.50%
质粒4	50 bp	50%	75%
平均值	N/A	50.00%	75.00%

从结果可以看出，lacP元件距离ITR越近，Sanger测序的通过率越低，说明lacP元件对ITR稳定性有一定的影响。在缺失lacP元件后，四个质粒的Sanger测序平均通过率，由50%提升到75%，大大提升了ITR结构的稳定性。

实施例8：ITR稳定性载体的设计和构建

pUC57载体的改造

实施例6中用作对照的pUC57质粒是一种常用的分子生物学实验载体，在克隆表达、DNA测序、基因突变等方面都有广泛应用，具有高拷贝数、易于操作、多克隆位点、选择性标记基因等特点。基于pUC57载体的原核载体骨架设计的载有目的基因的AAV-ITR穿梭质粒，广泛应用于AAV基因治疗。但是，由于ITR序列的特殊性，使其在大肠杆菌中复制不稳定，容易发生缺失，需要优化改造。

取pUC57 Kana（GenBank: LT671993.1）的原核载体骨架，进行改造。首先，在得到的载体两端加上ITR序列，即野生型AAV2-ITR序列（SEQ ID NO: 9）或者截短的AAV2-ITR序列（SEQ ID NO: 10、11、或12），如图19（A）所示。其次，将pUC Ori的位置调整到KanR基因上游，将pUC Ori末端与3’ITR序列前端的距离拉长至1747 bp。再次，将质粒中的lac启动子替换为同等长度的无义序列（SEQ ID NO: 13），该替换序列来源于大肠杆菌天然质粒，与大肠杆菌基因组或pUC57质粒载体没有同源性。形成的新载体骨架称为GenS-pUC57-AAV-NoTer。

在GenS-pUC57-AAV-NoTer的基础上，分别在5’ITR位置的下游和3’ITR的上游加入终止子序列，得到GenS-pUC57-AAV-Ter，如图19（B）所示。

之后，在载体GenS-pUC57-AAV-Ter和GenS-pUC57-AAV-NoTer的5’ITR和3’ITR之间插入CMV增强子、鸡β-肌动蛋白启动子、EGFP、bGH poly(A)信号，并将载体骨架上的IIS酶识别位点如BsaI、AarI、BpiI、BsmBI和BspQI进行同义突变，去除这些酶切位点，制备成通用载体GenS-pUC57-AAV-Noter-CMV（SEQ ID NO: 14，含SEQ ID NO: 9的ITR）和GenS-pUC57-AAV-Ter-CMV（SEQ ID NO: 15，含SEQ ID NO: 9的ITR），通过金斯瑞生物的基因合成部门，从头构建组装。

pSub201载体的改造

AAV基因治疗的创始人Samulski于1987年10月在Journal of Virology 61(10): pp.3096-3101首次描述了pSub201（addgene.org/vector-database/4231/）。该载体包含

AAV-2腺相关病毒野生型基因组编码区和顺式末端重复序列（ITR）。这个载体是克隆的理想选择，因为其设计方式是用XbaI进行限制性消化，在质粒骨架中保持完整的AAV-ITR，并去除AAV基因组编码区，替换为感兴趣的基因（GOI）。因此，该载体后续广泛被选做模板进行克隆，以应用于rAAV病毒包装及基因治疗，如pAAV-tTA质粒（GenBank: MZ708023.1）。但该载体的原核骨架部分设计存在一定缺陷，会造成ITR结构的不稳定，因此需要进行一定的改造，以获得高完整性的ITR结构。

取pAAV-tTA质粒（GenBank: MZ708023.1）质粒的原核载体骨架（图20（A）），将源自噬菌体的fl ori替换为额外的一段序列（SEQ ID NO: 16，不含与大肠杆菌基因组同源的部分），将右侧的pUC Ori与氨苄抗性基因交换位置，在5' ITR和3' ITR之间插入CMV增强子、鸡β-肌动蛋白启动子、EGFP、bGH poly(A)信号等元件，并将载体中常用的IIS酶识别位点如BsaI、AarI、BpiI、BsmBI、BspQI和常用于验证ITR结构完整性的酶切位点SmaI、AhdI、BglI、BssHII等进行同义突变，得到的载体称为GenS-pSUB201-AAV-NoTer（SEQ ID NO: 17，含SEQ ID NO: 10和11的ITR）。在此基础上，在5' ITR的下游和3' ITR的上游添加终止子序列，得到载体GenS-pSUB201-AAV-Ter（SEQ ID NO: 18，含SEQ ID NO: 10和11的ITR），如图20（B）所示，其中pUC Ori距离3' ITR的距离为1814 bp。新设计的载体，通过金斯瑞生物的基因合成部门，从头构建组装。

实施例9：pUC57系列载体的ITR结构稳定性测试

取实施例8中的pUC57 Kana（已加入两个ITR，即GenS-pUC57-AAV-NoTer调整Ori与KanR基因互换位置之前）、GenS-pUC57-AAV-NoTer、和GenS-pUC57-AAV-Ter，通过同源重组，在两个ITR之间插入基因片段1、基因片段2、基因片段3或基因片段4（分别如SEQ ID NO: 19-22所示）。

得到的载体在3*LB培养基中37℃条件下摇瓶培养16-18小时，抽提质粒，对质粒进行SmaI酶切，其中SmaI酶切位点位于ITR上，ITR的全部缺失或部分序列缺失会导致酶切失败。对酶切产物进行琼脂糖凝胶电泳，部分结果见图21。除目的条带以外的条带统称为杂带，使用Gel Image System软件（GIS）分析杂带占比，具体数值见表2。

表2. pUC57相关载体的SmaI酶切杂带占比

	pUC57 Kana	GenS-pUC57-AAV-NoTer	GenS-pUC57-AAV-Ter
基因1	100%	12.01%	9.56%
基因2	78.87%	21.65%	19.67%
基因3	36.79%	6.18%	0.00%
基因4	26.84%	0.00%	0.00%
平均值	60.63%	9.96%	7.31%

从结果可以看出，未经改造的pUC57 Kana，克隆的杂带比例较高，甚至出现杂带比例为100%的情况。而优化后的载体，杂带占比显著降低，表明载体在Ori位置、终止子添加和lacP删除方面的优化，明显提升了ITR结构的稳定性。特别是在ITR两端添加终止子的GenS-pUC57-AAV-Ter载体，杂带比显著低于GenS-pUC57-AAV-NoTer载体。

进一步地，使用ITR测序试剂盒（南京金斯瑞），对上述抽提的质粒进行测序。测序结果显示，在未经改造的pUC57 kana载体的测序中，出现缺失、底峰、双峰、测不通等现象，Sanger测序的通过率为0%，挑取另外的克隆仍然易出现以上情况。而优化后载体的测序峰图无明显底峰，ITR完整性高，Sanger测序的通过率皆为100%。

此外，还测试了三种载体的摇瓶抽提产量。具体地，将三种载体中基因片段合成正确的克隆转化GenITRv1感受态细胞，37℃过夜培养后，挑单克隆于装有4 mL LB培养基的试管中，30℃、180 rpm振荡培养16小时，然后将400 μL培养好的种子液转接于装有3×LB培养基的400 mL大瓶中，37℃、200 rpm振荡培养13小时，之后离心收菌，使用碱裂解法裂解细胞，异丙醇沉淀，最后用TE溶解质粒沉淀，用紫外分光光度计测量质粒的浓度。结果总结于表3中，可以看出，优化后的两个载体的产量相比于未改造的pUC57 Kana载体，分别提高了44%和26%。

表3. pUC57相关载体的摇瓶抽提产量（μg）

	pUC57 Kana	GenS-pUC57-AAV-NoTer	GenS-pUC57-AAV-Ter
基因1	469.1	1014.2	731.1
基因2	367.9	546.4	540.3
基因3	500.4	699.1	637.3
基因4	724.9	715.5	698.7
平均产量	515.6	743.8	651.9

进一步地，评价三种载体对于后续AAV包装的影响。具体地，使用三质粒包装系统，以三种载体（含插入基因1）为穿梭质粒，另分别使用AAV2和AAVDJ的包装质粒pRC2-mi342（Takara，货号：6230）/pRC-DJ（Cell Biolabs INC，货号VPK-420-DJ-8）以及辅助质粒pHelper（Takara，货号：6230）。穿梭质粒工业标准为1.5 μg/ml，将穿梭质粒、包装质粒和辅助质粒按照摩尔比2：4：1加入AAV-MAX转染试剂盒（ThermoFisher，A50515/ A50516），按照试剂盒的说明书进行病毒包装。参照AAVpro® Titration试剂盒（用实时PCR）（TaKaRa，货号：6233）第二版试剂盒说明书进行病毒滴度的测定，评估AAV的病毒包装的滴度。结果总结在表4中。

表4. 三种载体的病毒包装滴度

	AAV2-第一次	AAV2第二次	AAVDJ第一次	AAVDJ第二次
pUC57 Kana	1.03E+10	3.54E+09	6.65E+09	4.73E+09
GenS-pUC57-	1.38E+10	4.09E+09	8.31E+09	4.98E+09

AAV-NoTer				
GenS-pUC57-AAV-Ter	1.74E+10	4.31E+09	7.42E+09	6.30E+09

如表4所示，优化后载体获得的病毒包装滴度高于未改造的pUC57 Kana，且GenS-pUC57-AAV-Ter载体高于GenS-pUC57-AAV-NoTer载体。这与前述的SmaI酶切结果一致，即GenS-pUC57-AAV-Ter载体的SmaI酶切杂带比例更低，ITR完整性更高。

实施例10： pSub201系列载体的ITR结构稳定性测试

使用实施例8中的pAAV-tTA、GenS-pSUB201-AAV-NoTer、和GenS-pSUB201-AAV-Ter，分别克隆4个与实施例8相同的基因片段（基因1-4）。具体地，通过同源重组，在两个ITR之间插入基因片段1、基因片段2、基因片段3或基因片段4。

按照实施例9中所述，将载体转化GenITRv1感受态细胞，挑取单克隆，使用包含该质粒的单克隆进行摇瓶发酵且抽提质粒。三种载体的产量见表5。可以看出，经过改造的GenS-pSUB201-AAV-NoTer、和GenS-pSUB201-AAV-Ter载体，在摇瓶发酵后，相比于未改造的pAAV-tTA载体，平均产量分别提升66%和35%。

表5. 三种载体在摇瓶发酵后的产量（μg）

载体	pAAV-tTA	GenS-pSUB201-AAV-NoTer	GenS-pSUB201-AAV-Ter
基因1	200.0	426.3	704.3
基因2	300.0	915.9	740.0
基因3	420.8	542.0	435.1
基因4	714.1	899.6	485.3
平均产量	418.6	696.0	565.2

如实施例9中所述，使用SmaI酶切三种载体，并进行琼脂糖凝胶电泳，除目的条带以外的条带统称为杂带，使用GIS分析SmaI酶切后杂带占比，具体数值见表6。可以看出，未改造的pAAV-tTA载体，杂带比例高，改造后的GenS-pSUB201-AAV-NoTer和GenS-pSUB201-AAV-Ter载体杂带比例明显降低，ITR结构稳定性得到明显提升。

表6. 使用三种载体的SmaI酶切杂带占比

	pAAV-tTA	GenS-pSUB201-AAV-NoTer	GenS-pSUB201-AAV-Ter
基因1	23.50%	31.53%	14.85%
基因2	29.35%	21.32%	15.45%
基因3	30.07%	20.36%	26.11%
基因4	21.50%	14.37%	13.25%
平均值	26.11%	21.90%	17.42%

如实施例9中所述，使用三质粒包装系统，以三种载体（插入有基因1）为穿梭质粒，另使用AAV2包装质粒pRC2-mi342和辅助质粒pHelper（Takara，货号：6230），穿梭质粒工业标准为1.5 μg/ml，将穿梭质粒、包装质粒和辅助质粒按照摩尔比2：4：1加入AAV-MAX转染试剂盒（ThermoFisher，A50515/A50516），按照试剂盒的说明书进行病毒包装。参照AAVpro® Titration试剂盒（用实时PCR）（TaKaRa，货号：6233）第二版试剂盒说明书进行病毒滴度的测定，评估AAV的病毒包装的滴度。结果如表7所示，相对于pAAV-tTA，GenS-pSUB201-AAV-NoTer和GenS-pSUB201-AAV-Ter均将rAAV病毒滴度提升了44%。

表7. 三种载体的病毒包装滴度

	滴度	复孔滴度	平均滴度	平均滴度/D
pAAV-tTA	4.82E+10	3.02E+10	3.92E+10	1
GenS-pSUB201-AAV-NoTer	7.04E+10	4.24E+10	5.64E+10	1.44
GenS-pSUB201-AAV-Ter	6.75E+10	4.51E+10	5.63E+10	1.44

申请中涉及的序列如下。

SEQ ID NO: 1

ggaagagcctagctacctactcgagcttagagggcccggttaaaccctcagatcattgtcacatgtgagcgccaggaaaccgtggccgcgttgctggcgt

SEQ ID NO: 2

ggaagagcctagctacctactcgagcttagagggcccggttaaaccctcagatcagcctcgactctagcttgatggaagacgcaaaaacataaagaaaggcccgccgcttctatccgctggaagatggaacgagcgaactgcataaggctatgaagagcaagctcagatcattgtcacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctggcgtt

SEQ ID NO: 3

ggaagagcctagctacctactcgagcttagaatggaagacgcaaaaacataaagaaaggcccgccgcttctatccgctggaagatggaacgctggagagcaactgcataaggctatgaagagatacgccctggttctggaacaattgctttacagatgcacatacgcaggtggacatcactacgctgagctactcgaaatgtccgttcggttgcagaaagctatgaacgatatgggctgaatacaaatcacagaatcgtcgtatgcagtgaactctctcaattctttatccggtgttggcgcggttattatcgaggtgcagttgcgcccgcgaacgacattataatgaacgtgaattgctcaacagtatgggcatttcgcagcctaccgtggtgttcgtttccaaaaagggttgcaaaaattacatgtgagcaaaaggccagcaaaaggccaggaaccgtataaaaggccgcgttgctggcgtt

SEQ ID NO: 4

ggaagagcctagctacctactcgagcttagaatggaagacgcaaaaacataaagaaaggcccgccgcttctatccgctggaagatggaacgctggagagcaactgcataaggctatgaagagatacgccctggttctggaacaattgctttacagatgcacatacgcaggtggacatcactacgctgagctactcgaaatgtccgttcggttgcagaaagctatgaacgatatgggctgaatacaaatcacagaatcgtcgtatgcagtgaactctctcaattctttatccggtgttggcgcggttattatcgaggtgcagttgcgcccgcgaacgacattataatgaacgtgaattgctcaacagtatgggcatttcgcagcctaccgtggtgttcgtttccaaaaagggttgcaaaaatttgaacgtgcaaaaaagctcccaatcctcaaaaaattattatcatggattctaaaacgattaccagggttctcagtgatgcacgttcgtcacatctcatctcccggttttaataacgattttgtgccagatccttcgataggacaagacaattgcactgatcatgaactccttggaactactggacatgtgagcaaaaggccagcaaaaggccaggaaccgtataaaaggccgcgttgctggcgtt

SEQ ID NO: 5

ggaagagcctagctacctactcgagcttagagggcccggttaaaccctcagatcagcctcgactctagcttgattctcaccataaaaaacgcccggcggaaccgagcgttctgaacaaatccagatggagttctgaggtcattactggatctatcaacaggagtc
caagctcagatcattgtcacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctggcgtt

SEQ ID NO: 6

ggaagagcctagctacctactcagcttagaatggaagacgccccacataaagaaaggcccgccattctatccgctggaagatggaacgctggagagcaactgcataaggctatgaagagatacgccctggttcctggaacaattgctttacagatgcacatatcgaggtggacatcactt acgctgagtacttcgaaatgtccgttcggttggcagaagctatgaacgatatgggctgaatacaaatcacagaatcgtcgtatgcagtgaaaact ctctcaattctttatgccggtgttggcgcggtattataatgaacgtgaattgctcaacagtatgggcatttcgcagcctaccgtggtgttcgttccaa aaaggggttgcaaaaaattacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcttgcctggcggtt

SEQ ID NO: 7

GCGCAACGCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCAGGCTTTACACTTTATG CTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTCACA

SEQ ID NO: 8

TTGTTATCCGCTCACAAATCCACACAACATACGAGCCGGAAGCATAAAGTGTAAGCCTG GGGTGCCTAATGAGTGAGCTAACTCACATTA

SEQ ID NO: 9

Ttggccactccctctctgcgcgtcgtcgtcgtcactgaggccggcgaccaaaggctgcccgcgcccgggctttgccggggcgccctcagt gaggcgcgcgcgcagagggagtgccaactccatcactaggggtcct

SEQ ID NO: 10

AGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAG GCCGGGCGACCAAAGGTCGCCCACGCCCCGGGCTTTGCCCGGGCGGCCTCAGTGAGCG AGCGAGCGCGCAGCTGCCTGCAGG

SEQ ID NO: 11

CCTGCAGGCAGCTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCGTCGGGCGACCTT TGGTCGCCCCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACTCCATC ACTAGGGGTTCT

SEQ ID NO: 12

GCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCCCCACGCCCCGGGCTTTG CCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACTCCATCAC TAGGGGTTCT

SEQ ID NO: 13

TCTGGCGTACCGTTCCTGTCTAAAATCCCTTTAATCGGCCTCCTGTTTAGCTCCCGCTCTG ATTCTAACGAGGAAAGCACGTTATACGTGCTCGTCAAAGCAACgATAGT

SEQ ID NO: 14

GTCGACTAGAGCTCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGT TTGCCCCCTCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCTTTCTTA ATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCACTTCTATTCTGGGGGGTGG GGTGGGGCAGGACAGCAAGGGGGAGGATTGGcAAGACAATAGCAGGAACCCCACTCCCa egcgtAGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACT GAGGCCGCCCGGGCAAAGCCCCGGGCGTCGGGCGACCTTTGGTCGCCCCGGCCTCAGTGA GCGAGCGAGCGCGCAGAGAGGGAGTGGCCAAGGTACCTCTGGCCGTGTTTTACAACG TCGTGACTGGGAAAACCCTGGCGTTACCCAACCTAATCGCCTTGCAGCACATCCCGTCAG ACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTG CTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTGTTGCCGGATCAAGAGCTAC CAACTCTTTTTCCGAAGGTAAGTGGCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTC TAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCTACATACCTCG CTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGCGGATAAGTCGTGTCTTACCGGGT TGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCTGGGCTGAACGGGGGGTTCG TGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGA GCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGC

GGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATC
TTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTGTGATGCTCGTC
AGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACCGCGCCTTTTACGGTTCCTGGCCT
TTTGCTGGCCTTTTGCTCACATGTTCTTTCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGC
CCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCaTGAATGGCGAATGGCGCCTGATGCG
GTATTTTCTCCTTACGCATCTGTGCGGTATTTACACCCGCATATGGTGCACTCTCAGTACA
ATCTGCTCTGATGCCGCATAGTTAAGCCAGCCCCGACACCCGCCAACACCCGCTGACGCG
CCCTGACGGGCTTGTCTGCTCCCGCATCCGCTTACAGACAAGCTGTGACCGagCCGGG
AGCTGCATGTGTGAGAGGTTTTCACCGTCATCACCGAAACGCGCGATGCAGCTCTGGCC
CGTGTCTCAAAATCTCTGATGTTACATTGCACAAGATAAAAAATATATCATCATGAACAATA
AAACTGTCTGCTTACATAAACAGTAATACAAGGGGTGTTATGAGCCATATTCAACGGGAA
ACGTCGAGGCCGCGATTAAATTCCAACATGGATGCTGATTTATATGGGTATAAATGGGCTC
GCGATAATGTCGGGCAATCAGGTGCGACAATCTATCGCTTGTATGGGAAGCCCGATGCGC
CAGAGTTGTTTCTGAAACATGGCAAAGGTAGCGTTGCCAATGATGTTACAGATGAGATGG
TCAGACTAAACTGGCTGACGGAATTTATGCCTCTTCCGACCATCAAGCATTTTATCCGTAC
TCCTGATGATGCATGGTTACTCACCCTGCGATCCCCGGAAAAACAGCATTTCCAGGTATT
AGAAGAATATCCTGATTCAGGTGAAAATATTGTTGATGCGCTGGCAGTGTTCTGCGCCG
GTTGCATTGCTGTTCTGTTGTAATTGTCTTTTAACAGCGATCGCGTATTTTCGcCTCGCTC
AGGCGCAATCACGAATGAATAACGGTTTGGTTGATGCGAGTGATTTTGATGACGAGCGTA
ATGGCTGGCCTGTTGAACAAGTCTGGAAAGAAATGCATAAACTTTTGCCATTCTCACCGG
ATTCAGTCGTCACTCATGGTGATTTCTCACTTGATAACCTTATTTTGTACGAGGGGAAATT
AATAGGTTGTATTGATGTTGGACGAGTCGGAATCGCAGACCGATACCAGGATCTTGCCAT
CCTATGGAACCTGCCTCGGTGAGTTTCTCCTTCATTACAGAAACGGCTTTTTCAAAAATAT
GGTATTGATAATCCTGATATGAATAAATTGCAGTTTCATTTGATGCTCGATGAGTTTCTA
Actgtcagaccaagtttactcatatatacttttagattgatttaaaacttcattttaatttaaaggatctagggaagatccttttgataatctcatgacc
aaaatcccttaacgtgagttttgttccactgagcgtcagacCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATT
ACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTC
AGTGAGCGAGGAAGCGGAAGAaCGCCCAATACGCAAACCGCTCTCCCCGCGCGTTGGC
CGATTCAATATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGATCTG
GCGTACCGTTCTGTCTAAAATCCCTTTAATCGGCCTCCTGTTTAGCTCCCGCTCTGATT
TAACGAGGAAAGCACGTTATACGTGCTCGTCAAAGCAACgATAGTCAGGAAACAGCTATG
ACCATGATTACACCCAATTGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGG
CCGGGCGACCAAAGGTCGCCCGACGCCCCGGGCTTTGCCCCGGGCGGCCTCAGTGAGCGA
GCGAGCGCGCAGAGAGGGAGTGGCCAACTCCATCACTAGGGGTTCTCCTAGGTTAATG
ATTAACCCGCCATGCTACTTATCTACGTAGCCATGCTCTAGGATCTGAATTCTAGTTATTAA
TAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCGCGTTACATAACT
TACGGTAAATGGCCCGCTGGCTGACCGCCCAACGACCCCCGCCATTGACGTCAATAAT
GACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTAT
TTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTA
TTGACGTCAATGACGGTAAATGGCCCGCTGGCATTATGCCCAGTACATGACCTTATGGG
ACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTCGAGGTGAG
CCCCACGTTCTGCTTCACTCTCCCCATCTCCCCCCCCCTCCCCACCCCAATTTTGTATTAT
TTATTTTAAATTATTTTGTGAGCGATGGGGGCGGGGGGGGGGGGGGGGGCGCGCGCCAG
GCGGGGCGGGGCGGGGCGAGGGGCGGGGCGGGGCGAGGCGGAGAGGTGCGGCGGCAG
CCAATCAGAGCGGCGCGCTCCGAAAGTTTCTTTTATGGCGAGGCGGCGGCGGCGGCGG
CCCTATAAAAAGCGAAGCGCGCGGGCGGGGAGTTCGCTGCGaaatgctagcgtttaacttaagcttA
CCGGTgagctcggateccACTAGTccagtgtggtgGCATGCTgcagatatccagcacagtggcgccgcATGGTGAGCA

AGGGCGAGGAGCTGTTACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTA
AACGGCCACAAGTTCAGCGTGTCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGC
TGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCCGTGCCCTGGCCACCCCTCGTG
ACCACCCTGACCTACGGCGTGCAAGTTCAGCCGCTACCCCGACCACATGAAGCAGCA
CGACTTCTTCAAGTCCGCCATGCCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAA
GGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCCTGGTG
AACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACA
AGCTGGAGTACAACACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAAC
GGCATCAAGGTGAACCTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAAGCTCGC
CGACCACTACCAGCAGAACACCCCATCGGCGACGGCCCCGTGCTGCTGCCCCGACAACC
ACTACCTGAGCACCCAGTCCGCCCTGAGCAAAAGACCCCAACGAGAAGCGCGATCATATG
GTCCTGCTGGAGTTCGTGACCGCCGCGGGATCACTCTCGGCATGGACGAGCTGTACAA
GTAACTcgagcttagaggcccggttaaac

SEQ ID NO: 15

GTCGACTAGAGCTCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGT
TTGCCCTCCCCCGTGCTTCTTGACCCTGGAAGGTGCCACTCCCACTGTCTTTTCTTA
ATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGTGG
GGTGGGGCAGGACAGCAAGGGGGAGGATTGGcAAGACAATAGCAGGAACCCCACTCCCa
cggtAGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACT
GAGGCCGCCCCGGGCAAAGCCCCGGGCGTCGGGCGACCTTTGGTCGCCCCGGCCTCAGTGA
GCGAGCGAGCGCGCAGAGAGGGAGTGGCCAAGGTACCTCTGGCCGTGTTTTACAACG
TCGTGACctTAGCTTGGATTCTACCAATAAAAAACGCCCGGCGGCAACCGAGCGTTCT
GAACAAATCCAGATGGAGTTCTGAGGTCACTTGGATCTATCAACAGGAGTCCAAGCTC
AGctactggcggtttTGCTTGACGACATCCCGTCAGACCCCGTAGAAAAGATCAAAGGATCT
TCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAAACAAAAAACACCGCTA
CCAGCGGTGGTTTGGTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACGGC
TTCAGCAGAGCGCAGATACCAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCAC
TTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTG
CTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATA
AGGCGCAGCGGTGCGGCTGAACGGGGGGTTCGTGCACACAGCCAGCTTGGAGCGAAC
GACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCG
AAGGGAGAAAGGCGGACAGGTATCCGTAAGCGGCAGGGTCGGAACAGGAGAGCGCA
CGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCTGTGCGGTTTCGCCACC
TCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAAC
GCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCT
TTCCTTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGT
TGCGCAGCaTGAATGGCGAATGGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGG
TATTTACACCCGCATATGGTGCACCTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGC
CAGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACGGGCTTGCTGCTCCCCGGC
ATCCGCTTACAGACAAGCTGTGACCGacgCCGGGAGCTGCATGTGTGAGAGGTTTTACC
GTCATCACCGAAACGCGCGATGCAGCTCTGGCCCGTGCTCAAAATCTCTGATGTTACAT
TGCACAAGATAAAAAATATCATCATGAACAATAAACTGTCTGCTTACATAAACAGTAAT
ACAAGGGGTGTTATGAGCCATATTCAACGGGAAACGTCGAGGCCGCGATTAAATTCCAAC
ATGGATGCTGATTTATATGGGTATAAATGGGCTCGCGATAATGTCGGGCAATCAGGTGCGA
CAATCTATCGCTTGATGAGGAAGCCCGATGCGCCAGAGTTGTTTCTGAAACATGGCAAAG
GTAGCGTTGCCAATGATGTTACAGATGAGATGGTCAGACTAAACTGGCTGACGGAATTTA
TGCCTCTTCCGACCATCAAGCATTTTATCCGTACTCCTGATGATGCATGGTTACTACCAC

TGCGATCCCCGGAAAAACAGCATTCCAGGTATTAGAAGAATATCCTGATTTCAGGTGAAAA
TATTGTTGATGCGCTGGCAGTGTTCTGCGCCGGTTGCATTTCGATTCTGTGTTGTAATTGT
CCTTTTAACAGCGATCGCGTATTTTCGcCTCGCTCAGGCGCAATCACGAATGAATAACGGTT
TGGTTGATGCGAGTGATTTTGATGACGAGCGTAATGGCTGGCCTGTTGAACAAGTCTGGA
AAGAAATGCATAAACTTTTGCCATTCTCACCGGATTTCAGTCGTCACTCATGGTGATTTCTC
ACTTGATAACCTTATTTTTGACGAGGGGAAATTAATAGGTTGTATTGATGTTGGACGAGTC
GGAATCGCAGACCGATAACCAGGATCTTGCCATCCTATGGAACCTGCCTCGGTGAGTTTTCT
CCTTCATTACAGAAACGGCTTTTTCAAAAATATGGTATTGATAATCCTGATATGAATAAATT
GCAGTTTCATTTGATGCTCGATGAGTTTTTCTAATCAGAATTGGTTAATTGGTTGTAACATT
ATTCAGATTGGGCTTGATTTAAACTTCATTTTAAATTTAGATAACCGTATTACCGCCTTTG
AGTGAGCTGATAACCGCTCGCCGCAGCCGAACGACCgttcttcaataaaacgaaaggctcagtcgaaagactg
ggccttctgtttatctgttgttgcgtgaacgctctcctgagtaggacaaatccggggagcggatttgaacgttgcgaagcaacggcccgga
gggtggcgggcaggacgcccgcataaaactgccaggcatcaaattaagcagaaggccatcctgacggatggccttttgcgtttctacaaactct
gactcttCAGGAAACAGCTATGACCATGATTACACCCAATTGTTGGCCACTCCCTCTCTGCGC
GCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCCCCGACGCCCGGGCTTTGCCCG
GGCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACTCCATCACTAGG
GGTTCCTCCTAGGTTAATGATTAACCCGCCATGCTACTTATCTACGTAGCCATGCTCTAGG
ATCTGAATTCTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGG
AGTTCGCGCTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCC
GCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTG
ACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCA
TATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCC
CAGTACATGACCTTATGGGACTTTCTACTTGGCAGTACATCTACGTATTAGTCATCGCTAT
TACCATGGTCGAGGTGAGCCCCACGTTCTGCTTCACTCTCCCCATCTCCCCCCCCCTCCCC
ACCCCCAATTTTGTATTTATTTATTTTAAATTTTGTGTCAGCGATGGGGGCGGGGGGGG
GGGGGGGGCGCGCGCCAGGCGGGGCGGGGCGGGGCGAGGGGCGGGGCGGGGCGAGGC
GGAGAGGTGCGGCGGCAGCCAATCAGAGCGGCGCGCTCCGAAAGTTTCTTTTATGGCG
AGGCGGGCGGCGGGCGGCCCTATAAAAAGCGAAGCGCGCGGGCGGGGAGTTCGCTG
CGaatgctagcgtttaaacttaagcttACCGGTgagctcggatccACTAGTccagtgtggtgGCATGctgcagatatccagcac
agtggcgccgcATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCATCCTGGTC
GAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCG
ATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCCGTG
CCCTGGCCCCACCCTCGTGACCACCCTGACCTACGGCGTGCACTGCTTCAGCCGCTACCCC
GACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGA
GCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCG
AGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGG
CAACATCCTGGGGCACAAGCTGGAGTACAACACAACAGCCACAACGTCTATATCATGGC
CGACAAGCAGAAGAACGGCATCAAGGTGAACCTCAAGATCCGCCACAACATCGAGGAC
GGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGT
GCTGCTGCCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACG
AGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGC
ATGGACGAGCTGTACAAGTAACtcgagctagaggcccggttaaac

SEQ ID NO: 16

AAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCCCCTGCGTTATCCCCTGATTCTGTGGAT
AACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCG
CAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAAACCGCCTCTCCCC
GCGCGTTGGCCGATTCAATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGG

CAGTGATCTGGCGTACCGTTCTGTCTAAAATCCCTTTAATCGGCCTCCTGTTTAGCTCCC
GCTCTGATTCTAACGAGGAAAAGCACGTTATACGTGCTCGTCAAAGCGTTGATATTCAGTC
AATTACAAACATTAATAACGAAGAGATGACAGAAAAATTTTCATTCTGTGACAGAGAAAA
AGTAGCCGAAGATGACGGTTTGTACATGGAGTTGGCA

SEQ ID NO: 17

GTCGACTAGAGCTCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGT
TTGCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCCTGTCCTTTCCCTA
ATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGG
GGTGGGGCAGGACAGCAAGGGGGAGGATTGGcAAGACAATAGCAGGAACCCCACTCCC
GGTACCAGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCA
CTGAGGCCGGGCGACCAAGGTGCGCCGACGCCGGGCTTTGCCGGGCGGCCTCAGT
GAGCGAGCGAGCGCGCAGCTGCCTGCAGGACTAGTTGATGCGGTATTTTCTCCTTACGCA
TCTGTGCGGTATTTACACCGCATACGTACATGTGTCAGAGGTTTTCACCGTCATCACCG
AACACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTC
TGCGCGTAATCTGCTGCTTGCAAACAAAAAACCACCGCTACCAGCGGTGGTTTGTGTTG
CCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAAGTGGCTTCAGCAGAGCGCAGAT
ACCAAATACTGTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGC
ACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAA
GTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTGCG
GCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACT
GAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCG
GACAGGTATCCGGTAAGCGGCAGGGTCCGAACAGGAGAGCGCACGAGGGAGCTTCCAG
GGGGAACGCCTGGTATCTTTATAGTCCTGTGCGGGTTTCGCCACCTCTGACTTGAGCGTC
GATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCC
TTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTACGCGCGAGTCGAAAGGG
CCTCGTGATACGCCTATTTTATAGGTTAATGTgATGATAATAATGGTTTCTTAGACGTCAG
GTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTCTAAATACATTC
AAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGG
AAGAGTATGAGTATTCAACATTTCCGTGTGCGCCTTATTCCTTTTTTGCGGCATTTTGCT
TCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGG
GTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTTC
GCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTAT
TATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTGCGCGCATACACTATTCTCAGAATG
ACTTGGTTGAGTACTACCCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGA
GAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACA
ACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAAC
TCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATAACCAAACGACGAGCGTGACA
CCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAAGTGGCGAACTACTTA
CTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCAC
TTCTGCGCTCGGCgCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGC
GTGGaTCTCGCGGTATCATTCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGT
TATCTACACcACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGAT
AGGTGCCTCACTGATTAAGCATTGGTAAGTGTGACACCAAGTTTACTCATATATACTTTAG
ATTGATTTAAAACTTCATTTTTAATTTAAAGGATCTAGGTGAAGATCCTTTTTGATAATCT
CATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTcAGTTTATCCCTTAACC
CTGCGTTATCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGC
TCGCCGAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAaCGC

CCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCTTAATGCAGCTGGCACGAC
AGGTTTCCCGACTGGAAAAGCGGGCAGTGATCTGGCGTACCGTTCCTGTCTAAAATCCCTT
TAATCGGCCTCCTGTTTAGCTCCCGCTCTGATTCTAACGAGGAAAGCACGTTATACGTGCT
CGTCAAAGCGTTGATATTCAGTCAATTACAAACATTAATAACGAAGAGATGACAGAAAAA
TTTTCAATTCTGTGACAGAGAAAAAGTAGCCGAAGATGtCGGTTTGTACATGGAGTTGGC
ATATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGCCCCGACAC
CCGCCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTGAAGCAACCATAGTCTCCCGG
CATCCGCTTACAGACAAGCTGTGACCGTCTaCGGGAGCTGAACAATTGCCTGCAGGCAGC
TGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCGTCGGGCGACCTTTGGTCGCCCCGGC
CTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACTCCATCACTAGGGGTTCC
TaccgtTTAATGATTAACCCGCCATGCTACTTATCTACGTAGCCATGCTCTAGGATCTGAATT
CTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTTCATAGCCCATATATGGAGTTCCGC
GTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCATTG
ACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAAT
GGGTGGAGTATTTACGGTAACTGCCCACTTGGCAGTACATCAAGTGATCATATGCCAA
GTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCAGTACAT
GACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGG
TCGAGGTGAGCCCCACGTTCTGCTTCACTCTCCCCATCTCCCCCCCCCTCCCCACCCCCAA
TTTTGTATTTATTTATTTTAAATTATTTTGTGCAGCGATGGGGGCGGGGGGGGGGGGGG
GCGCGCGCCAGGCGGGGCGGGGCGGGGCGAGGGGCGGGGCGGGGCGAGGCGGAGAGG
TGCGGCGGCAGCCAATCAGAGCGGCGCGCTCCGAAAGTTTCCTTTTATGGCGAGGCGGC
GGCGGCGGCGGCCCTATAAAAAGCGAAGCGCGCGGCGGGGCGGGAGTCGCTGCGGaatgcta
gcgtttaaacttaagcttACCGGTgagctcggatccatgcaccagtgtggtgGCATGtgcagatatccagcacagtggcgccgcA
TGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCATCCTGGTCGAGCTGGA
CGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACC
TACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCCGTGCCCTGGCC
CACCTCGTGACCACCCTGACCTACGGCGTGCACTGCTTCAGCCGCTACCCCGACCAT
GAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCA
TCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGA
CACCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCC
TGGGGCACAAGCTGGAGTACAACACAAGCCACAACGTCTATATCATGGCCGACAAG
CAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCG
TGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTG
CCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCCAACGAGAAGC
GCGATCACATGGTCTGCTGGAGTTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGAC
GAGCTGTACAAGTAACTcaggtctagaggccccgtttaaac

SEQ ID NO: 18

GTCGACTAGAGCTCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGT
TTGCCCTCCCCCGTGCCTTCCCTTGACCCTGGAAGGTGCCACTCCCACTGTCTTTCTTA
ATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCAATTCTATTCTGGGGGTGG
GGTGGGGCAGGACAGCAAGGGGGAGGATTGGcAAGACAATAGCAGGAACCCCACTCCC
GGTACCAGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCA
CTGAGGCCGGGCGACCAAGGTGCCCCGACGCCCGGGCTTTGCCCGGGCGGCCTCAGT
GAGCGAGCGAGCGCGCAGCTGCCTGCAGGACTAGTTGATGCGGTATTTCTCCTTACGCA
TCTGTGCGGTgtttttaaataaaacgaaaggctcagtcgaagactgggaccttcgtttatctgtgtgttcggtgaacgctctcctgag
taggacaaatccggcgagcggaattgaacgttgcgaagcaacggccccggagggtggcgggcaggacccccccataaactgccaggcat
caaattaagcagaaggccatcctgacggatggccttttgcgtttctacaactctgactcttGCCAGCCCCGACACCCGCCAA

CACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCT
GTGACC_aTCTCCGGGAGCTGCATGTGTGTCAGAGGTTTTACCGTCATCACCGAACACTGAG
CGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTCTGCGCGTAA
TCTGCTGCTTGCAAACAAAAAACACCGCTACCAGCGGTGGTTTTGTTTGCCGGATCAA
GAGCTACCAACTCTTTTTCCGAAGGTAAGTGGCTTCAGCAGAGCGCAGATACCAAATACT
GTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACA
TACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTA
CCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCCGGCTGAACGGGG
GGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACA
GCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCG
GTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCC
TGGTATCTTTATAGTCTGTGCGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGAT
GCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTC
CTGGCCTTTTTGCTGGCCTTTTGCTCACATGTACGCGGAG_tCGAAAGGGCCTCGTGATAC
GCCTATTTTTATAGGTAAATGT_gATGATAATAATGGTTTCTTAGACGTCAGGTGGCACTTTT
CGGGGAAATGTGCGCGGAACCCCTATTTGTTATTTTTCTAAATACATTCAAATATGTATCC
GCTCATGAGACAATAACCCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAG
TATTCAACATTTCCGTGTGCGCCTTATTCCTTTTTTGCGGCATTTGCCTTCCTGTTTTG
CTCACCCAGAAACGCTGGTGAAAGTAAAGATGCTGAAGATCAGTTGGGTGCACGAGTG
GGTTACATCGAACTGGATCTCAACAGCGGTAAAGATCCTTGAGAGTTTTCGCCCCGAAGA
ACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTG
ACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAG
TACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGT
GCTGCCATAACCATGAGTGATAAACTGCGGCCAACTTACTTCTGACAACGATCGGAGGA
CCGAAGGAGCTAACCGCTTTTTTGACAACATGGGGGATCATGTAACTCGCCTTGATCGT
TGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGT
AGCAATGGCAACAACGTTGCGCAAACTATTAAGTGGCGAACTACTTACTCTAGCTTCCCG
GCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGG
C_gCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGA_tCTCGCGG
TATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACAC_cACG
GGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCCTCACT
GATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAATGG
CACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGATCTGGCGTACCGTTCCCTGTCTAAA
ATCCCTTTAATCGGCCTCCTGTTTAGCTCCCGCTCTGATTCTAACGAGGAAAGCACGTTAT
ACGTGCTCGTCAA_{ct}CTAGCTTGGATTCTCACCAATAAAAAACGCCCGGCGGCAACCGAG
CGTCTGAACAAATCCAGATGGAGTTCTGAGGTCATTACTGGATCTATCAACAGGAGTCC
AAGCTCAG_tctactggcggtttTACAATCTGCTCTGATGCCGCATAGTTAAAACCTCATTTTTAAT
TTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGA
GTTTCAATTGCCTGCAGGCAGCTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCGTC
GGGCGACCTTTGGTCGCCCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGC
CAACTCCATCACTAGGGGTTCC_{Tac}cgfTTAATGATTAACCCGCCATGCTACTTATCTACGTA
GCCATGCTCTAGGATCTGAATCTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCA
TAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCCGCTGGCTGACCG
CCCAACGACCCCCGCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAG
GGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTAC
ATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCCG
CTGGCATTATGCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTAT

CCGTGCCAAGAGTGACGTAAGTACCGCCTATAGAGTCTATAGGCCACAAAAATGCTTT
CTTCTTTTAATACTTTTTTGTATTCTTATTTCTAATACTTTCCCTAATCTCTTTCTTCAG
GGCAATAATGATACAATGTATCATGCCTCTTTGCACCATCTAAAGAATAACAGTGATAATT
TCTGGGTAAAGGCAATAGCAATATTTCTGCATATAAATATTTCTGCATATAAATTGTAACGT
ATGTAAGAGGTTTCATATTGCTAATAGCAGCTACAATCCAGCTACCATTTCTGCTTTTATTTT
ATGGTTGGGATAAGGCTGGATTATTCTGAGTCCAAGCTAGGCCCTTTTGCTAATCATGTTT
ATACCTCTTATCTTCTCCACAGCTCCTGGGCAACGTGCTGGTCTGTGTGCTGGCCATC
ACTTTGGCAAAGAATTGGGATTCGAACATCGATTGAATTCTGAATGGTGGCCGGACATGC
CTCTGGATCTCCAGCCTTTGGAACAGCCAGCCACAGCAACTGCGAGCACGAGGAAATTC
ACCTGGCCGGCTCTATTACGCCTCATGGCGCTCTGCTGGTGGTGTCCGAACACGACCACA
GAGTGATTGAGGCTCTGCCAATGCCGCCGAGTTCTGAATCTGGGCTCTGTGCTGGGAG
TGCCTCTGGCCGAAATTGACGGCGACCTGCTGATCAAGATCCTGCCTCACCTGGATCCTA
CCGCCGAGGGAATGCCTGTGGCCGTGCGTTGTAGAATCGGCAACCCTAGCACCGAGTAC
TGCGGCCTGATGCACAGACCTCCAGAAGGCGGACTGATCATCGAGCTGGAAAGAGCTGG
CCCCAGCATCGATCTGTCTGGAACACTGGCTCCCGCTCTGGAAAGAATCAGAACAGCCG
GAAGCCTGAGAGCCCTGTGCGACGATACAGTGCTGCTGTTCCAGCAGTGCACCGGCTAC
GATAGAGTGATGGTGTACAGATTTCGACGAGCAAGGCCACGGCCTGGTGTCTCTGAATGT
CACGTGCCCCGGCCTGGAAAGCTACTTCGGCAACAGATACCCAGCAGCACCGTGCCTCA
GATGGCCAGACAGCTGTATGTGCGGCAGCGCGTTAGAGTGCTGGTGGACGTCACATACC
AGCCTGTGCCACTGGAACCCAGACTGAGCCCTCTGACAGGCAGAGATCTGGATATGAGC
GGCTGCTTCCTGCGGAGCATGAGCCCTTGCCATCTGCAGTTTCTGAAGGACATGGGCGTC
AGAGCCACACTGGCCGTGTCTCTTGTGTTGGCGGAAAGCTGTGGGGCCTCGTCGTGTG
TCACCACTACCTGCCTCGGTTTCATCAGATTCGAGCTGCGGGCTATCTGCAAGCGGCTGGC
CGAGAGAATCGCCACCAGAATCACAGCCCTGGAATCCTGAGGATCCTCTAGAGTCGACC
TGCAGAAGCTTGCTCGAGCAGCGCTGCTCGAGAGATCTACGGGTGGCATCCCTGTGAC
CCCTCCCCAGTGCTCTCCTGGCCCTGGAAGTTGCCACTCCAGTGCCCACCAGCCTTGTG
CTAATAAAATTAAGTTGCATCATTTTGTCTGACTAGGTGTCTTCTATAATATTATGGGGTG
GAGGGGGGTGGTATGGAGCAAGGGGCAAGTTGGGAAGACAACCTGTAGGGCCTGCGGG
GTCTATTGGGAACCAAGCTGGAGTGCAGTGGCACAATCTTGGCTCACTGCAATCTCCGCC
TCCTGGGTTCAAGCGATTCTCCTGCCTCAGCCTCCCGAGTTGTTGGGATTCCAGGCATGC
ATGACCAGGCTCAGCTAATTTTTGTTTTTTGGTAGAGACGGGGTTTCACCATATTGGCCA
GGCTGGTCTCCAATCCTAATCTCAGGTGATCTACCCACCTTGGCCTCCCAAATTGCTGG
GATTACAGGCGTGAACCACTGCTCCCTTCCCTGTCCTTCTGATTTTGTAGGTAACCGTGA
GTAGTGGAAGGCGGTTCTCTGGCCAACGGGCTGCTTGCTGCTGTTAGCAGTTGGAGAAT
GGATGCCTCTGCCCGGTAAAGGGCACCTGGGGCCGCGCCCGCAGCATTCACTAACTCGT
AACTCTCCTTCCCTTCCTCCAAACACAAAATTCCATGACACTAGACAAGAAAGCTGATGC
TTGGAAAAGAGAATTGGCCTTAAATACCTAGATGGACTGGAGAGACCATCCTTTGTGGTC
CATAGCTTGAGGTCTGAGTACCCTGTCCCCATCCACCTGCAGACCTGCTGAGGCTCCCCT
CTGGCTTTCTCCTGCAGAAACCTCTGCCATGGCCAAGCCTTTGTCTCAAGAAGAATCCAC
CCTCATTGAAAGAGCAACGGCTACAATCAACAGCATCCCCATCTCTGAAGACTACAGCGT
CGCCAGCGCAGCTCTCTAGCGACGGCCGCATCTTCACTGGTGTCAATGTATATCATTTT
ACTGGGGGACCTTGTGCAGAACTCGTGGTGTGGGCACTGCTGCTGCTGCGGCAGCTGG
CAACCTGACTTGTATCGTCGCGATCGGAAATGAGAACAGGGGCATCTTGAGCCCCTGCG
GACGGTGTGACAGGTGCTTCTCGATCTGCATCCTGGGATCAAAGCGATAGTGAAGGAC
AGTGATGGACAGCCGACGGCAGTTGGGATTCTGTAATTGCTGCCCTCTGGTTATGTGTGG
GAGGGCTAAAGAGCAGAAGGAAGGGGTTCTGTTGACTTGTGGAGATGTTGAGGAGA
ATCCAGGACCAGCACCGGGATCCATGGTGAGCGGCTGGCGGCTGTTCAAGAAGATTAGC

GGCGGAGGTGGTTCTGGCGGTGGAGGTTCAATGAACCAGTCCATCCCAGTTGCTCCCAC
CCCACCCCGCAGAGTGCGGCTGAAGCCCTGGCTGGTGGCCCAGGTGAACAGCTGCCAG
TACCCAGGGCTTCAATGGGTCAACGGGGAAAAGAAATTATTCTGCATCCCCTGGAGGCAT
GCCACAAGGCATGGTCCCAGCCAGGACGGAGATAACACCATCTTCAAGGTAAGCCCCGG
GGAGGAGGTTGGCTGGACCTCCAGGGCACCCCTGTCCCCAGAAGAGGAGCGCACATAAC
GCACACAGGCAGCTCCTCGAGGcacctgcGTTT

SEQ ID NO: 22

GTTTgcaggtgcagtcacacgagcagctggttttaagatgctatttccgtataaagcatgagaccgtgacttgccagccccacagagcc
ccgcccctgtccatcactggcatctggactccagcctgggtggggcaagagggaatgagatcatgtcctaacctgatcccttctgccacag
atatcagaacctgacctgctgcccggccaccatggacttcaggtgagatcttcagcttctgctgatcagcgccagcgtgatcagc
cggtatggccacatccagatgaccagagcccttctcttctccgttccgtgggggacagggtaactattacgtccgagcttccagggcat
ccgcaactacctgcagtggtaccagcagaagcctggcaagctcccaagctgctgatctacgcagcttcgacctgcagagcgggtgccctct
cgcttcagtggtatctggtacggactttactctgacctctctccctgcagccggaagatgtggccacttattactgtcagcgttacaccg
cgcccctataccttccgcccagggcaccaggggtgagatcaaggaggtccagccgttctctgctctggggggggtggtccggcggtgg
tggttccaggtgacgtggtggagagcgccggagggtggtgcagccagggcgctggtgcgctgtcatgcgcccgttctggttcacatt
gatgactacgccatgattgggtccgcccggccagggcctggagtggtgagtgccatcacctggaaacctggccacattgactac
gcccagacgctggtgagggcagattcaccatctcccgtgacaacgcgaaaaacagcccttacttgcatgaattccctgcccgtgagggaccc
gcccgtgactactgtgccaaggtgagctacctgtctacggccagttcgttgattatggggccaaaggcactctcgtcaccgtgagcagctcag
cagaagctgatcagcgaggagacctaaccggatccgtgggggtaccgtctcttcagcgcctgagcaactccatcatgtacttcagccactcgt
gcccgttctctccagcgaagcccaccacgacggcagcgccgaccaccaaacaccggcgcccaccatcgcgtcagccccctgtccctg
cgcccagaggcggtgcccggcagcgccggggggcgagtgccacacaggggggctggaccctttgggtttgggtgctggtggtggtggtg
gagtcctggctgtatagctgtgtaaacagtgccctttattattttctgggtgagggagtaaggaggagcaggctctgcacagtgactacatgaac
atgactccccgcccggggcccccgcgaagcattaccagccctatccccaccacgcgacttcgcagcctatcgtccctcgagagagtg
agagtgaagttagcagggagcgagacgccccgcgtaccagcagggccagaaccagctctataacagctcaatctaggacgaagagagg
agtacgatgttttgacaagagacgtggccgggacctgagatggggggaagccgagaagggaagaacctcagggaaggcctgtacaatga
actgcagaaagataagatggcgaggccctacagtgagattgggatgaagggcgagcgccggaggggcaaggggacagatggccctttaccag
ggtctcagtagcaccacgaagacacctacgacgcccctcacatgcagggcctgccccctcgccgttccggagccacgaacttctctgttaaa
gcaagcaggagagctggaagaaaaccccgggtccgggttcggagaggcgaggaagcttgcctacctgcggggatgtgaggagaaatcc
aggctctatgcacagctcagcactgctctgttgcctggtcctcgtactgggtgagggccagcccagggccagggcaccagctctgagaacag
ctgcaccacttccaggaacctgcttaacatgcttcgagatctccgagatgccttcagcagagtgaaagactttcttcaaatgaagatcagctg
gacaactgtgttaaggagctcctgctggaggacttaagggttacctgggttgcgaagccttgcctgagatgatccagttttacctggaggaggt
gatgccccagctgagaaccaagaccagacatcaaggcgcatgtgaactccctgggggagaacctgaagacctcaggtgaggtacgg
cgctgtcatcgatttctccctgtgaaaacaagagcaaggccgtggagcaggtgaagaatgcctttaataagctcaagagaaaggcatctaca
agccatgagtgagtttgacatcttcatcaactacatagaagccatcatgacaatgaagatacgaacgggttccggagccacgaacttctctgtta
aagcaagcaggagacgtggaagaaaaccccgggtccgggttcggagaggcgaggaagcttgcctacctgcggggatgtggaggagaat
ccaggtcctatgcccccctccgggctgcccgtgctgctgctgctaccgctgctggtggtactggtgctgacgcctggccggccggccggcg
gactatccacctgcaagactatcgacatggagctggtgaagcggaagcgcatcgaggccatccggccagatcctgtccaagctgcggctcg
ccagcccccgagccagggggaggtgcccccggcccgtgcccaggccgtgctgcccgtgtacaacagcaccgcgaccgggtggcc
ggggagagtgagaaaccggagcccagcctgaggccgactactacgccaaggaggtcaccgcgtgctaatgttggaaccacacaacgaa
atctatgacaagttaagcagagtagacacagcatatataatgttttaacacatcagagctccgagaagcggtacctgaacctgtgtctctcc
ggcgagagctgctgctgctgaggctcaagttaaaagtggagcagcagctggagctgtaccagaaatacagcaacaattcctggcgatacctcag
caaccggctgctggcaccacgcgactcgccagagtggttatcttttgatgtcaccggaggtgtgcccagtggttgagccgtggaggggaaattg
agggctttaccttagcgeccactgctccagagacagcagggataacacactgcaagtgagacataacgggttactaccggcccgaggtg
acctggccacattcatgcatgaaccggccttctctgcttcatggccaccccgtgagagggcccgacatctgcaagctcccgccagccg
ccgagccctggagaccactattgctcagctccagcggagaagaactgctgctgcccagctgtacattgacttccgaaggacctcggtgg
aagtggtaccacgagcccaaggctaccatgccaacttctgctcgggcccctgcccctacatttgagccctggacacgcagtagcaaggtcc
tgcccctgtacaaccagcataaccggggcgctcggcgccgctgctgctgcccagggcgtggagccgctgcccctgtgtactacgtg

ggccgcaagcccaaggtggagcagctgtccaacatgatcgtgcgctcctgcaagtcagctgacctgactgtgccttctagttgccagccatct
gttgtttgcccctccccgtgccttcttgacctggaaggtgccactcccactgtccttcttaataaaatgaggaaattgcatcgcattgtctgagt
aggtgtcattctattctgggggtgggggtggggcaggacagcaagggggaggattgggaagacaatagcaggcatgtggggatgcgggtgg
gctctatggcttctgagcggaaagaaccagctggggctctaggggtatccccactagtcgtgtaccagctgagagactctaaatccagtgaca
agtcgtctgcctattaccgattttgattctcaacaaatgtgtcacaaagtaaggattctgatgtgtatatcacagacaaaactgtgctagacatgag
gtctatggacttcaagagcaacagtgctgtggcctggagcaacaaatctgactttgcatgtgcaaacgccttcaagatccacctgcGTTT

SEQ ID NO: 23

TTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAACCAC
CGCTACCAGCGGTGGTTTGTGTTGCCGGATCAAGAGCTACCAACTCTTTTTCCG
AAGGTAACCTGGCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGT
AGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCT
CGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTC
TTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCCGG
CTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACAC
CGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAA
GGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGA
GCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTC
GGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGG
GGCGGAGCCTATGGAAA

SEQ ID NO: 24

CAAATAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTT
GTTTGTGCGGTGAACGCTCTCCTGAGTAGGACAAAT

SEQ ID NO: 25

TGGCGGGCAGGACGCCCCGCCATAAACTGCCAGGCATCAAATTAAGCAGAAG
GCCATCCTGACGGATGGCCTTTTTGCGTTTCTACAAA

SEQ ID NO: 26

ATTCTACCAATAAAAAACGCCCCGGCGGCAACCGAGCGTTCTGAACAAATCC
AGATGGAGTTCTGAGGTCATTACTGGATCTATCAACAGGAGTC

SEQ ID NO: 27

gctagcgtttctgcacacttaagcttGAGACGggtaccgagctcggatccgaattctgcagatatccagcacagtggcgg
ccgctcaggtGTGCAGtgtgagcaagggcgaggagctgttaccgggggtgtgcccattctgtcagctggacggc
gacgtaaacggccacaagttcagcgtgtccggcgagggcgagggcgatgccacctacggcaagctgacctgaagttcatc
tgcaccaccggcaagctgcccgtgcccgtggccaccctcgtgaccaccctgacctacggcgtgagtgcttcagccgtacc
ccgaccacatgaagcagcacgacttctcaagtccgccatgccgaaggctacgtccaggagcgcaccatcttctcaagga
cgacggcaactacaagaccgcgccgaggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcac
gacttcaaggaggacggcaacatcctggggcacaagctggagtacaactacaacagccacaacgtctatatcatggccgac
aagcagaagaacggcatcaaggtgaacttaagatccgccacaacatcgaggacggcagcgtgcagctcgccgaccata
ccagcagaacacccccatggcgacggccccgtgctgtgcccgacaaccactacctgagcaccagtcgccctgagca
aagaccccaacgagaagcgcgatcacatggctctgtgagttcgtgaccgcccgggatcactctcgcatggacgagc
tgtacaagaCGTCTCactagtAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AA
AAAAAAAAAAAAAAAAAAAAAGgaagagcctagctactactcagctctagagggcc

SEQ ID NO: 28

gctagcgtttctgcacacttaagcttGAGACGggtaccgagctcggatccgaattctgcagatatccagcacagtggcgg
ccgctcgagtGTGCAGtaCGTCTCactagtAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGGAAGAGCctagctacctactcgagtctagagggccc

SEQ ID NO: 29

TAATACGACTCACTATAagGGGTCTTCTGGTCCCCACAGACTCAGAGAGAACCCGCCACC
ATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCATCCTGGTTCGAGCTGGA
CGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACC
TACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGCAAGCTGCCCCGTGCCCTGGCC
CACCTCTGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACAT
GAAGCAGCACGACTTCTTCAAGTCCGCCATGCCGAAGGCTACGTCCAGGAGCGCACCA
TCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGA
CACCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCC
TGGGGCACAAGCTGGAGTACAACACAAGCCACAACGTCTATATCATGGCCGACAAG
CAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCG
TGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTG
CCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGC
GCGATCACATGGTCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGAC
GAGCTGTACAAGTAAtagtaagctggagcctcgggtgacctgcttcttggcccttgggctccccccagccctcctccccctctg
caccctgaccccggtgctcttgaataaagctgagtgggcgcaactagtAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AA
AAAAAAAAAAAAAAAAAAAAAAAAAAGgaagagc

尽管本申请已经结合一个或多个实施方式进行了描述，应当理解的是，本申请并不受限于这些实施方式。本申请中的描述意在涵盖所有变体形式以及等同物，均包含在所附权利要求的主旨和范围内。所有在本文中引用的文献通过引用的方式全部并入本文。

权 利 要 求 书

1. 一种载体，包含：
 - i) 复制起点，
 - ii) 第一终止子，
 - iii) a) 目标DNA表达盒，其包含启动子和目标DNA序列，其中目标DNA序列包含能够形成非B型DNA构型的DNA序列，或b) 目标DNA序列，其包含能够形成非B型DNA构型的DNA序列，
 - iv) 第二终止子，以及
 - v) 筛选标记和控制该筛选标记表达的启动子，其中，第一终止子和第二终止子分别位于a) 的目标DNA表达盒的上游和下游，并设置成使a) 的目标DNA表达盒之外的转录终止于a) 的目标DNA表达盒的上游和下游；或
第一终止子和第二终止子分别位于b) 的目标DNA序列的上游和下游，并设置成使b) 的目标DNA序列之外的转录终止于b) 的目标DNA序列的上游和下游。
2. 如权利要求1所述的载体，其中复制起点在复制方向上与a) 的目标DNA表达盒中的能够形成非B型DNA构型的DNA序列、或b) 的目标DNA序列中的能够形成非B型DNA构型的DNA序列间隔至少150 bp。
3. 如权利要求2所述的载体，其中复制起点在复制方向上与a) 的目标DNA表达盒中的能够形成非B型DNA构型的DNA序列、或b) 的目标DNA序列中的能够形成非B型DNA构型的DNA序列间隔至少500 bp。
4. 如权利要求1-3中任一项所述的载体，其中能够形成非B型DNA构型的DNA序列为正向重复序列、反向重复序列、或单一嘌呤-单一嘧啶序列。
5. 如权利要求4所述的载体，其中能够形成非B型DNA构型的DNA序列为poly(A)尾编码序列。
6. 如权利要求1-5中任一项所述的载体，其中复制起点为pMB1复制起点、pBR322复制起点、ColE1复制起点、R6K复制起点、p15A复制起点、pSC101复制起点、pUC复制起点、pET复制起点或其衍生物。
7. 如权利要求1-6中任一项所述的载体，其中第一终止子选自rrnB T1终止子、rrnB T2终止子、lambda t0终止子、TonB终止子、Amp终止子、T7终止子和CYC1终

止子中的一种或多种；第二终止子选自rrnB T1终止子、rrnB T2终止子、lambda t0终止子、TonB终止子、Amp终止子、T7终止子和CYC1终止子中的一种或多种。

8. 如权利要求1-7中任一项所述的载体，其中筛选标记为氨苄青霉素抗性基因、四环素抗性基因、氯霉素抗性基因、卡那霉素抗性基因、或链霉素抗性基因。

9. 如权利要求1-8中任一项所述的载体，其中a) 的目标DNA表达盒中的目标DNA序列还包含克隆位点，b) 的目标DNA序列还包含克隆位点。

10. 如权利要求1-9中任一项所述的载体，其中目标DNA表达盒还包含增强子、和/或转录终止子。

11. 如权利要求10所述的载体，其按序包含：

pUC复制起点、rrnB T1终止子和rrnB T2终止子、CMV增强子、CMV启动子、目标DNA序列、bGH poly(A)信号、lambda t0终止子、和卡那霉素抗性元件；或者pUC复制起点、rrnB T1终止子和rrnB T2终止子、目标DNA序列、lambda t0终止子、和卡那霉素抗性元件。

12. 一种载体，包含：

i) 复制起点，

ii) 第一终止子，

iii) a) 目标DNA表达盒，其按序包含第一ITR或其衍生序列、启动子、目标DNA序列、和第二ITR或其衍生序列，或b) 目标DNA序列组，其按序包含第一ITR或其衍生序列、目标DNA序列、和第二ITR或其衍生序列，

iv) 第二终止子，以及

v) 筛选标记和控制该筛选标记表达的启动子，

其中，第一终止子和第二终止子分别位于a) 的目标DNA表达盒的上游和下游，并设置成使a) 的目标DNA表达盒之外的转录终止于a) 的目标DNA表达盒的上游和下游；或

第一终止子和第二终止子分别位于b) 的目标DNA序列组的上游和下游，并设置成使b) 的目标DNA序列组之外的转录终止于b) 的目标DNA序列组的上游和下游。

13. 如权利要求12所述的载体，其中复制起点在复制方向上与a) 的目标DNA表达盒、或b) 的目标DNA序列组间隔至少150 bp，优选地，间隔至少500 bp。

14. 如权利要求12或13所述的载体，其中第一终止子选自rrnB T1终止子、rrnB

T2终止子、lambda t0终止子、TonB终止子、Amp终止子、T7终止子和CYC1终止子中的一种或多种，第二终止子选自rrnB T1终止子、rrnB T2终止子、lambda t0终止子、TonB终止子、Amp终止子、T7终止子和CYC1终止子中的一种或多种。

15. 如权利要求12-14中任一项所述的载体，其中复制起点为pMB1复制起点、pBR322复制起点、ColE1复制起点、R6K复制起点、p15A复制起点、pSC101复制起点、pUC复制起点、pET复制起点或其衍生物。

16. 如权利要求12-15中任一项所述的载体，其中筛选标记为氨苄青霉素抗性基因、四环素抗性基因、氯霉素抗性基因、卡那霉素抗性基因、或链霉素抗性基因。

17. 如权利要求12-16中任一项所述的载体，其中第一ITR或其衍生序列包含SEQ ID NO: 9、10、11或12所示的核苷酸序列，第二ITR或其衍生序列包含SEQ ID NO: 9、10、11或12所示的核苷酸序列。

18. 如权利要求12-17中任一项所述的载体，其中目标DNA表达盒中还包含增强子、和/或转录终止子。

19. 如权利要求18所述的载体，其按序包含：

pUC复制起点、rrnB T1终止子和rrnB T2终止子、第一ITR或其衍生序列、CMV增强子、鸡β-肌动蛋白启动子、目标DNA序列、bGH poly(A)信号、第二ITR或其衍生序列、lambda t0终止子、和卡那霉素抗性元件。

20. 一种用于改造载体的方法，其包括：

i) 提供一种载体，其包含：

a) 复制起点，

b) b-i) 克隆位点，

b-ii) DNA表达盒，其包含启动子和克隆位点，

b-iii) DNA表达盒，其包含启动子、克隆位点和能够形成非B型DNA构型的DNA序列，

b-iv) 目的DNA序列，其包含能够形成非B型DNA构型的DNA序列，

b-v) 目标DNA表达盒，其包含启动子和目标DNA序列，目标DNA序列包含能够形成非B型DNA构型的DNA序列，

b-vi) DNA序列组，其按序包含第一ITR或其衍生序列、克隆位点、和第二ITR或其衍生序列，

b-vii) DNA表达盒，其按序包含第一ITR或其衍生序列、启动子、克隆位点、和第二ITR或其衍生序列，

b-viii) 目标DNA表达盒, 其按序包含第一ITR或其衍生序列、启动子、目标DNA序列、和第二ITR或其衍生序列, 或

b-ix) 目标DNA序列组, 其按序包含第一ITR或其衍生序列、目标DNA序列、和第二ITR或其衍生序列, 和

c) 筛选标记和控制该筛选标记表达的启动子, 以及

ii) 在载体中加入第一终止子和第二终止子, 使其分别位于b) 的序列的上游和下游, 且设置成使得载体中b) 的序列以外的转录终止于b) 的序列的上游和下游。

21. 如权利要求20所述的方法, 还包括使复制起点在复制方向上与b-i) 克隆位点、b-ii) DNA表达盒中的克隆位点、b-iii) DNA表达盒中的能够形成非B型DNA构型的DNA序列、b-iv) 目的DNA序列中的能够形成非B型DNA构型的DNA序列、b-v) 目标DNA表达盒中的能够形成非B型DNA构型的DNA序列、b-vi) DNA序列组、b-vii) DNA表达盒、b-viii) 目标DNA表达盒、或b-ix) 目标DNA序列组间隔至少150 bp, 优选地间隔至少500 bp。

22. 如权利要求20或21所述的方法, 其中载体还包含lac启动子, 所述方法还包括使lac启动子失活。

23. 一种组合物, 包含权利要求1-11中任一项所述的载体、权利要求12-19中任一项所述的载体、或由权利要求20-22中任一项所述的方法得到的载体。

24. 一种用于复制目标DNA序列的方法, 包括:

i) 提供权利要求1-11中任一项所述的载体、或权利要求12-19中任一项所述的载体;

ii) 将步骤i) 的载体转化/转染到宿主细胞中; 以及

iii) 在利于载体复制的条件下培养宿主细胞。

25. 一种用于表达目标DNA序列的方法, 包括:

i) 提供权利要求1-11中任一项所述的包含a) 目标DNA表达盒的载体, 和

ii) 将步骤i) 的载体转化/转染到宿主细胞中, 并在利于目标DNA序列表达的条件下培养宿主细胞, 或者使步骤i) 的载体线性化、并进行体外转录; 或者

i) 提供权利要求12-19中任一项所述的包含a) 目标DNA表达盒的载体,

ii) 将步骤i) 的载体转化/转染到宿主细胞中, 并在利于目标DNA序列表达的条件下培养宿主细胞,

使步骤i) 的载体线性化、并进行体外转录; 或者

使步骤i) 的载体与包装质粒、辅助质粒共转化/转染宿主细胞, 在利于载体包装的条件下培养宿主细胞, 回收包装好的载体, 转化/转染宿主细胞, 并在利于目

标DNA表达的条件下培养宿主细胞。

26. 如权利要求25所述的方法，还在步骤i) 和步骤ii) 之间包括将步骤i) 的载体转化/转染到宿主细胞中，并在利于载体复制的条件下培养宿主细胞。

27. 一种用于制备重组腺相关病毒载体的方法，包括：

i) 提供权利要求12-19中任一项所述的载体；

ii) 将步骤i) 的载体转化/转染到宿主细胞中；

iii) 在利于载体复制的条件下培养宿主细胞；以及

iv) 从步骤iii) 的宿主细胞中收获载体，并将其与包装质粒、辅助质粒共转化/转染宿主细胞，在利于载体包装的条件下培养宿主细胞。