



(12)发明专利申请

(10)申请公布号 CN 109971775 A

(43)申请公布日 2019.07.05

(21)申请号 201810797876.1

(22)申请日 2018.07.19

(66)本国优先权数据

201711438680.5 2017.12.27 CN

(71)申请人 康码(上海)生物科技有限公司

地址 201321 上海市浦东新区芙蓉花路500
弄8号楼3楼

(72)发明人 郭敏 王海鹏 王静 战荣荣

徐开 于雪

(51) Int. Cl.

C12N 15/63(2006.01)

C12N 15/67(2006.01)

C12N 5/10(2006.01)

C12N 1/19(2006.01)

权利要求书2页 说明书22页

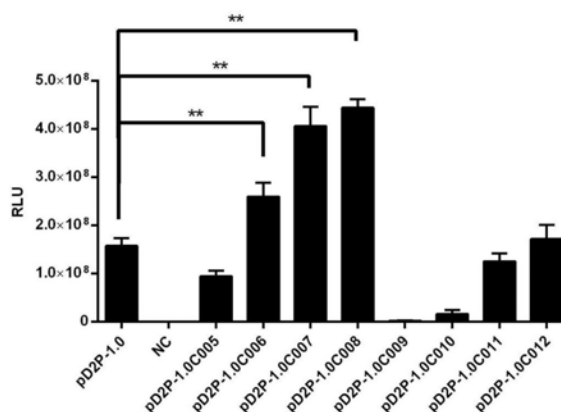
序列表24页 附图3页

(54)发明名称

一种核酸构建物及其调控蛋白合成的方法

(57)摘要

本发明提供了一种核酸构建物及其调控蛋白合成的方法,通过对编码mRNA的5'UTR和3'UTR的DNA核酸序列进行改造,分别插入特异的3' CITE和其对应的5'端序列,使转录生成的mRNA形成环状结构,促使翻译结束的核糖体到达5'UTR快速重新开始新一轮的翻译,利用两端的发夹结构增强mRNA的稳定性,与5'UTR核酸元件以及3'UTR上的poly(A)结构产生协同效应,提高蛋白质合成翻译的效率。



1. 一种核酸构建物,所述核酸构建物具有5'-3'的式A所示的结构:

Z1-Z2-Y1-Z3-Z4-Y2-Z5 (A)

式中,

Z1为启动子元件;

Z2和Z4为环化元件,其中环化元件Z2和Z4通过RNA-RNA直接作用使对应于式A的mRNA转录产物形成一环化结构,并且所述环化元件选自下组:TED、BTE、PTE、I型结构元件、Y型结构元件、T型结构元件、其他能够形成RNA-RNA直接作用的元件、或其组合;

Y1为无或“Z6-Z7-Z8”所示的元件;

Z3为外源蛋白的编码序列;

Y2为无或Z9;

Z5为终止子元件;

Z6为无或增强子元件;

Z7为无或烟草花叶病毒的5'端前导序列(leading sequence)-Ω序列;

Z8为无或Kozak序列;

Z9为腺嘌呤脱氧核糖核酸的多聚链;

各“-”独立地为键或核苷酸连接序列。

2. 一种核酸构建物,其特征在于,所述核酸构建物含有结构如式I所示的核酸序列:

Z1-Z2-Z3-Z4-Z5 (I)

式中,

Z1-Z4分别为用于构成所述构建物的元件;

各“-”独立地为键或核苷酸连接序列;

Z1为启动子元件,较佳地所述启动子元件选自下组:T7启动子、T3启动子、SP6启动子、或其组合;

Z4和Z2为环化元件,并且所述环化元件选自下组:TED、BTE、PTE、I型结构元件、Y型结构元件、T型结构元件、其他能够形成RNA-RNA直接作用的元件、或其组合;

Z3为外源蛋白的编码序列;

Z5为终止子元件,较佳地所述终止子元件选自下组:T7终止子、T3终止子、SP6终止子、或其组合。

3. 一种核酸构建物,其特征在于,所述核酸构建物含有结构如式II所示的核酸序列:

Z1-Z2-Z6-Z7-Z8-Z3-Z4-Z9-Z5 (II)

式中,

Z1-Z9分别为用于构成所述构建物的元件;

各“-”独立地为键或核苷酸连接序列;

Z1为启动子元件,所述启动子元件选自下组:T7启动子、T3启动子、SP6启动子、或其组合;

Z4和Z2为环化元件,所述环化元件选自下组:TED、BTE、PTE、I型结构元件、Y型结构元件、T型结构元件、其他能够形成RNA-RNA直接作用的元件、或其组合;

Z6为增强子元件,所述增强子元件包括IRES元件;

Z7为烟草花叶病毒的5'端前导序列(leading sequence)-Ω序列;

Z8为Kozak序;

Z3为外源蛋白的编码序列;

Z9为腺嘌呤脱氧核糖核酸的多聚链[poly(A)],其特征在于,所述多聚链选自下组:50A、70A、90A、或其组合;

Z5为终止子元件,所述终止子元件选自下组:T7终止子、T3终止子、SP6终止子、或其组合。

4. 一种体外的外源蛋白合成体系,其特征在于,包括:

(a) 权利要求1-3中任一项所述的核酸构建物;和

(b) 酵母细胞提取物。

5. 一种载体或载体组合,其特征在于,所述的载体或载体组合含有权利要求1-3中任一项所述的核酸构建物。

6. 一种基因工程细胞,其特征在于,所述基因工程细胞的基因组的一个或多个位点整合有权利要求1-3任一项所述的核酸构建物,或者所述基因工程细胞中含有权利要求5所述的载体或载体组合。

7. 一种试剂盒,其特征在于,所述试剂盒中包含的试剂选自下组中的一种或多种:

(a) 权利要求1-3任一所述的构建物;

(b) 权利要求5所述的载体或载体组合;

(c) 权利要求6所述的基因工程细胞。

8. 一种如权利要求1-3任一所述的构建物、权利要求5所述的载体或载体组合、权利要求6所述的基因工程细胞或权利要求7所述的试剂盒的用途,其特征在于,用于进行高通量的体外蛋白质合成。

9. 一种体外高通量的外源蛋白的合成方法,其特征在于,包括步骤:

(i) 在真核体外生物合成体系存在下,提供权利要求1-3任一所述的核酸构建物;

(ii) 在适合的条件下,孵育步骤(i)的真核体外生物合成体系一段时间T1,从而合成所述外源蛋白。

10. 如权利要求9所述的方法,其特征在于,所述步骤(ii)中,反应温度为20-37°C,较佳地,22-35°C。

11. 如权利要求9所述的方法,其特征在于,所述步骤(ii)中,反应时间为1-72h,较佳地,2-23h。

12. 如权利要求9所述的方法,其特征在于,还包括(iii)任选地从所述真核体外生物合成体系中,分离或检测所述外源蛋白。

一种核酸构建物及其调控蛋白合成的方法

技术领域

[0001] 本发明涉及生物工程技术领域,具体地,涉及一种核酸构建物及其调控蛋白合成的方法。

背景技术

[0002] 蛋白质是细胞中的重要分子,几乎参与了细胞所有功能的执行。蛋白的序列和结构不同,决定了其功能的不同。在细胞内,蛋白可以作为酶类催化各种生化反应,可以作为信号分子协调生物体的各种活动,可以支持生物形态,储存能量,运输分子,并使生物体运动。在生物医学领域,蛋白质抗体作为靶向药物,是治疗癌症等疾病的重要手段。

[0003] 蛋白质翻译的四个过程包括翻译起始、翻译延伸、翻译终止和核糖体再循环,其中受调控最多和限速的步骤是翻译的起始[1]。在细菌的翻译起始过程中,30S小亚基在3种起始因子的介导下直接在起始密码子的AUG附近形成起始复合体。真核生物的翻译起始则较为复杂,需要11种起始因子,并且核糖体在起始蛋白质合成时与mRNA结合的位点与原核生物不同[2]。真核生物的具有“帽子结构”(cap structure, m^7GpppN)的mRNA翻译起始主要是通过依赖“帽子结构”的扫描机制进行的[1-3]。其中“帽子结构”能够招募翻译起始因子eIF4F,进而招募下游的包括核糖体40S小亚基以及翻译起始因子eIF1、eIF3和eIF5的43S翻译起始前复合物(Pre-initiation complex, PIC),另外一个包含翻译起始因子eIF2和起始tRNA (Met-tRNA_i)的三元复合物也同时被招募到mRNA上。被招募来的复合物沿着5'端到3'端的方向进行扫描,发现第一个起始密码子AUG后开始翻译过程。

[0004] 除了“帽子结构”外,真核生物的大部分mRNA的3'端都有多聚腺嘌呤核苷链[poly (A)],能够被poly (A)结合蛋白(pAB1)结合。除了结合mRNA的3'端polyA, pAB1蛋白还能够结合翻译起始因子eIF4G, pAB1在eIF4G上的结合位点与能够结合mRNA的5'端“帽子结构”eIF4E在eIF4G上的结合位点不同,这样pAB1和eIF4E能够同时与eIF4G结合。在eIF4E, eIF4G和pAB1蛋白的介导下, mRNA的5'端和3'端在空间上十分靠近从而能够形成环状结构,这与电子显微照相观察到的多核糖体是环状的实验结果吻合[4]。真核生物的mRNA环化结构的功能意义目前还没有定论,其可能的解释是帮助核糖体在完成翻译后仍然与mRNA结合在一起,使核糖体从3'末端直接转移到5'末端继续扫描和重新开始翻译过程,减少了核糖体亚基以及翻译起始因子复合物的重新循环组装[2]。生物体内的mRNA翻译受到诸多内在因素(如能荷,氨基酸等)和外在因素(如营养,胁迫压力等)的调控。其中部分调控通过不同信号通路进行信号传导,实现翻译起始因子eIFs的磷酸化水平调控和相关蛋白的活性调节。以蛋白翻译合成调控中关键的信号通路雷帕霉素靶蛋白(target of rapamycin,简称TOR)途径为例,哺乳动物雷帕霉素靶蛋白1(简称mTORC1)的底物之一为翻译抑制因子(4E-Binding Protein,简称4E-BPs)小蛋白家族,是可以与eIF4G竞争性结合eIF4E的小分子蛋白。激活状态下的mTORC1蛋白可以磷酸化4E-BP1上的特定苏氨酸残基,使4E-BP1从eIF4E上解离下来,从而促进eIF4F复合物的形成[9]。通过eIF4F形成的mRNA环化涉及的因素很多,受到的来自不同途径的调控会限制蛋白合成翻译的起始速率。

[0005] 酵母体内一个主要mRNA降解途径是首先对mRNA的去腺苷酸化,然后再进行脱帽反应,最后达到mRNA的5'端到3'端的降解。此外,去腺苷酸化的mRNA的3'端会暴露出来,从而被从3'端到5'端降解[10,11]。在真核生物中的很多mRNA的降解的第一步是去腺苷酸化反应。因此,带有polyA的模板在细胞提取液中的可能比较不稳定,容易被去腺苷酸化而引起降解。

[0006] 体外生物合成系统是指在细菌、真菌、植物细胞或动物细胞的裂解提取物中,通过加入外源编码的核酸DNA、RNA、底物和能量源,完成特定化学分子或生物大分子(DNA、RNA或蛋白质)的体外高效合成。在体外蛋白质合成系统中,利用细胞提取液,通过添加外源mRNA或者DNA模板完成对外源重组蛋白的快速高效表达。目前以原核细胞或真核细胞作为宿主细胞制备的体外蛋白合成表达系统都已实现商业化。由于合成5'帽子的原料7-甲基鸟嘌呤核苷的价格昂贵,而且加帽反应效率较低,真核细胞蛋白合成系统中作为蛋白质翻译模板的mRNA通常设计为不含5'端“帽子结构”,此时通常需要5'非翻译区(即5'UTR)上的特定的RNA结构来起始蛋白质的翻译。

[0007] 目前,常用的此类RNA结构为内部核糖体进入序列(internal ribosome entry site,简称IRES)和一些来自于病毒的翻译增强子,如烟草花叶病毒的Omega序列(简称 Ω 序列)。但因缺少了5'“帽子结构”,在体外生物合成体系中用于蛋白质合成的mRNA很难形成环状结构,这就意味着核糖体在合成结束后需要从mRNA分子上解离下来,需要通过5'UTR重新招募进而起始下一轮的蛋白质翻译,带有polyA的mRNA不稳定,会被细胞裂解液中的酶进行去腺苷酸化反应,容易被降解,这严重影响了蛋白质合成的效率。

[0008] 综上所述,本领域目前的体外生物合成系统的蛋白翻译合成的效率还较低,难以令人满意,因此,本领域亟需开发一种新的核酸构建物,通过所述核酸构建物的转录和翻译,调控外源蛋白的合成,提高所述外源蛋白翻译合成的效率。

发明内容

[0009] 本发明的目的在于提供一种新的可在体外生物合成系统中显著提高蛋白翻译合成效率的核酸构建物。本发明所述核酸构建物通过RNA-RNA相互作用,可产生一种可直接环化的mRNA,即crRNA(circular RNA),从而显著提高体外生物合成体系中蛋白质合成的效率。

[0010] 本发明第一方面提供了一种核酸构建物,所述核酸构建物具有5'-3'的式A所示的结构:

[0011] Z1-Z2-Y1-Z3-Z4-Y2-Z5 (A)

[0012] 式中,

[0013] Z1为启动子元件;

[0014] Z2和Z4为环化元件,其中环化元件Z2和Z4通过RNA-RNA直接作用使对应于式A的mRNA转录产物形成一环化结构,并且所述环化元件选自下组:TED、BTE、PTE、I型结构元件、Y型结构元件、T型结构元件、其他能够形成RNA-RNA直接作用的元件、或其组合;

[0015] Y1为无或“Z6-Z7-Z8”所示的元件;

[0016] Z3为外源蛋白的编码序列;

[0017] Y2为无或Z9;

- [0018] Z5为终止子元件；
- [0019] Z6为无或增强子元件；
- [0020] Z7为无或烟草花叶病毒的5'端前导序列(leading sequence)-Ω序列；
- [0021] Z8为无或Kozak序列；
- [0022] Z9为腺嘌呤脱氧核糖核酸的多聚链；
- [0023] 各“-”独立地为键或核苷酸连接序列。
- [0024] 在另一优选例中,所述RNA-RNA直接作用包括碱基配对相互作用。
- [0025] 在另一优选例中,所述碱基配对相互作用包括完全碱基配对或不完全碱基配对。
- [0026] 在另一优选例中,所述多聚链为碱基长度为10A-200A。
- [0027] 在另一优选例中,所述多聚链选自下组:50A、70A、90A。
- [0028] 在另一优选例中,所述的各核苷酸连接序列的长度为1-20nt,较佳地1-10nt。
- [0029] 在另一优选例中,所述T型结构元件为K1-TSS。
- [0030] 在另一优选例中,所述Y型结构元件为YSS。
- [0031] 在另一优选例中,所述的启动子选自下组:T7启动子、T3启动子、SP6启动子、或其组合。
- [0032] 在另一优选例中,所述终止子元件选自下组:T7终止子、T3终止子、SP6终止子、或其组合。
- [0033] 在另一优选例中,所述Z6、Z8来源于酵母。
- [0034] 在另一优选例中,所述酵母选自下组:酿酒酵母、克鲁维酵母属酵母、或其组合。
- [0035] 在另一优选例中,所述克鲁维酵母属酵母选自下组:乳酸克鲁维酵母(*Kluyveromyces lactis*)、马克斯克鲁维酵母(*Kluyveromyces marxianus*)、多布克鲁维酵母(*Kluyveromyces dobzhanskii*)、或其组合。
- [0036] 在另一优选例中,所述增强子选自下组:IRES元件、RBS元件、非编码序列、或其组合。
- [0037] 在另一优选例中,所述增强子元件包括IRES元件。
- [0038] 在另一优选例中,所述IRES元件来源选自下组的一种或多种细胞:原核细胞、真核细胞。
- [0039] 在另一优选例中,所述真核细胞包括高等真核细胞。
- [0040] 在另一优选例中,所述IRES元件包括内源性IRES元件和外源性IRES元件。
- [0041] 在另一优选例中,所述IRES元件来源选自下组的一种或多种细胞:人(human)、中国仓鼠卵巢细胞(Chinese hamster ovary cell,CHO)、昆虫细胞(insect)、麦胚(Wheat germ cells)、兔网织红细胞(Rabbit reticulocyte)。
- [0042] 在另一优选例中,所述IRES元件选自下组:ScB0I1、ScFL08、ScNCE102、ScMSN1、K1FL08、K1NCE102、K1MSN1、K1B0I1、或其组合。
- [0043] 在另一优选例中,所述Ω序列包括直接重复模块(ACAATTAC)_m和(CAA)_p模块。
- [0044] 在另一优选例中,所述m为1-6,较佳地,2-4。
- [0045] 在另一优选例中,所述p为6-12个,较佳地,8-10个。
- [0046] 在另一优选例中,所述(CAA)_p模块为1-5个,较佳地,1-3个。
- [0047] 在另一优选例中,所述(CAA)_p模块还包括优化的(CAA)_p模块。

- [0048] 在另一优选例中,Z2为3'CITE,而Z4为与所述3'CITE对应的5'端序列。
- [0049] 在另一优选例中,所述的环化元件来自病毒。
- [0050] 在另一优选例中,所述病毒选自下组:PEMV、TBSV、HCRSV、STNV、或其组合。
- [0051] 在另一优选例中,所述的核酸构建物具有式I或式II结构,其中,式I和式II结构如下所述。
- [0052] 在另一优选例中,所述核酸构建物的序列如SEQ ID NO.:2-13、51-53所示。
- [0053] 在另一优选例中,所述核酸构建物的序列如SEQ ID NO.:3、7-9、51-53所示。
- [0054] 在本发明第1a方面,所述核酸构建物含有结构如式I所示的核酸序列:
- [0055] Z1-Z2-Z3-Z4-Z5 (I)
- [0056] 式中,
- [0057] Z1-Z4和Z5分别为用于构成所述构建物的元件;
- [0058] 各“-”独立地为键或核苷酸连接序列;
- [0059] Z1为启动子元件,优选地所述启动子元件选自下组:T7启动子、T3启动子、SP6启动子、或其组合;
- [0060] Z4和Z2为环化元件,优选地所述环化元件选自下组:TED、BTE、PTE、I型结构元件、Y型结构元件、T型结构元件、其他能够形成RNA-RNA直接作用的元件、或其组合;
- [0061] Z3为外源蛋白的编码序列;
- [0062] Z5为终止子元件,优选地所述终止子元件选自下组:T7终止子、T3终止子、SP6终止子、或其组合。
- [0063] 在另一优选例中,所述外源蛋白的编码序列来自原核生物、真核生物。
- [0064] 在另一优选例中,所述外源蛋白的编码序列来自动物、植物、病原体。
- [0065] 在另一优选例中,所述外源蛋白的编码序列来自哺乳动物,较佳地灵长动物,啮齿动物,包括人、小鼠、大鼠。
- [0066] 在另一优选例中,所述的外源蛋白的编码序列编码选自下组的外源蛋白:荧光素蛋白、或荧光素酶(如萤火虫荧光素酶)、绿色荧光蛋白、黄色荧光蛋白、氨酰tRNA合成酶、甘油醛-3-磷酸脱氢酶、过氧化氢酶、肌动蛋白、抗体的可变区域、萤光素酶突变体、 α -淀粉酶、肠道菌素A、丙型肝炎病毒E2糖蛋白、胰岛素前体、干扰素 α A、细胞因子,干扰素 α 2b、白细胞介素-1 β 、溶菌酶素、血清白蛋白、单链抗体段(scFV)、甲状腺素运载蛋白、酪氨酸酶、木聚糖酶、或其组合。
- [0067] 在另一优选例中,所述外源蛋白选自下组:荧光素蛋白、或荧光素酶(如萤火虫荧光素酶)、绿色荧光蛋白、黄色荧光蛋白、氨酰tRNA合成酶、甘油醛-3-磷酸脱氢酶、过氧化氢酶、肌动蛋白、抗体的可变区域、萤光素酶突变、 α -淀粉酶、肠道菌素A、丙型肝炎病毒E2糖蛋白、胰岛素前体、干扰素 α A、细胞因子,干扰素 α 2b,白细胞介素-1 β 、溶菌酶素、血清白蛋白、单链抗体段(scFV)、甲状腺素运载蛋白、酪氨酸酶、木聚糖酶、或其组合。
- [0068] 本发明第1b方面,提供了一种核酸构建物,所述核酸构建物含有结构如式II所示的核酸序列:
- [0069] Z1-Z2-Z6-Z7-Z8-Z3-Z4-Z9-Z5 (II)
- [0070] 式中,
- [0071] Z1-Z9分别为用于构成所述构建物的元件;

- [0072] 各“-”独立地为键或核苷酸连接序列；
- [0073] Z1为启动子元件,优选地,所述启动子元件选自下组:T7启动子、T3启动子、SP6启动子、或其组合；
- [0074] Z4和Z2为环化元件,优选地,所述环化元件选自下组:TED、BTE、PTE、I型结构元件、Y型结构元件、T型结构元件、其他能够形成RNA-RNA直接作用的元件、或其组合；
- [0075] Z6为增强子元件,优选地,所述增强子元件包括IRES元件；
- [0076] Z7为烟草花叶病毒的5'端前导序列(leading sequence)-Ω序列；
- [0077] Z8为Kozak序列,优选地,所述Kozak序列来源于酵母；
- [0078] Z3为外源蛋白的编码序列；
- [0079] Z9为腺嘌呤脱氧核糖核酸的多聚链[poly(A)],优选地,所述多聚链选自下组:50A、70A、90A、或其组合；
- [0080] Z5为终止子元件,优选地,所述终止子元件选自下组:T7终止子、T3终止子、SP6终止子、或其组合。
- [0081] 在本发明第二方面,提供一种体外的外源蛋白合成体系,包括:
- [0082] (a) 本发明第一方面所述的核酸构建物;和
- [0083] (b) 酵母细胞提取物。
- [0084] 在另一优选例中,所述外源蛋白合成体系中,组分(a)的浓度为5-30ng/μL,较佳地,8-20ng/μL,更佳地,10-16ng/μL,以所述外源蛋白合成体系的总体积计。
- [0085] 在另一优选例中,所述外源蛋白合成体系中,组分(b)的浓度(v/v)为20%-70%,较佳地,30-60%,更佳地,40%-50%,以所述外源蛋白合成体系的总体积计。
- [0086] 在另一优选例中,所述酵母或酵母细胞选自下组:酿酒酵母、毕氏酵母、克鲁维酵母、或其组合;较佳地,所述的酵母细胞包括:克鲁维酵母,更佳地为乳酸克鲁维酵母。
- [0087] 在另一优选例中,所述的酵母细胞提取物为对酵母细胞的水性提取物。
- [0088] 在另一优选例中,所述酵母细胞提取物不含酵母内源性的长链核酸分子。
- [0089] 在另一优选例中,所述外源蛋白合成体系还包括选自下组的一种或多种组分:
- [0090] (c1) 镁离子;
- [0091] (c2) 钾离子;
- [0092] (c3) 缓冲剂;
- [0093] (c4) 能量再生系统;
- [0094] (c5) 聚乙二醇;
- [0095] (c6) 用于合成RNA的底物;
- [0096] (c7) 用于合成蛋白的底物;
- [0097] (c8) RNA聚合酶;
- [0098] (c9) 任选的外源蔗糖;
- [0099] (c10) 任选的溶剂,所述溶剂为水或水性溶剂。
- [0100] 在另一优选例中,所述镁离子来源于镁离子源,所述镁离子源选自下组:醋酸镁、谷氨酸镁、或其组合。
- [0101] 在另一优选例中,所述钾离子来源于钾离子源,所述钾离子源选自下组:醋酸钾、谷氨酸钾、或其组合。

[0102] 在另一优选例中,所述缓冲剂选自下组:4-羟乙基哌嗪乙磺酸、三羟甲基氨基甲烷、或其组合。

[0103] 在另一优选例中,所述能量再生系统选自下组:磷酸肌酸/磷酸肌酸酶系统、糖酵解途径及其中间产物能量系统、或其组合。

[0104] 在另一优选例中,所述的合成RNA的底物包括:核苷单磷酸、核苷三磷酸、或其组合。

[0105] 在另一优选例中,所述的合成蛋白的底物包括:1-20种天然氨基酸、以及非天然氨基酸。

[0106] 在另一优选例中,所述核苷三磷酸选自下组:腺嘌呤核苷三磷酸、鸟嘌呤核苷三磷酸、胞嘧啶核苷三磷酸、尿嘧啶核苷三磷酸、或其组合。

[0107] 在另一优选例中,所述氨基酸为选自下组:甘氨酸、丙氨酸、缬氨酸、亮氨酸、异亮氨酸、苯丙氨酸、脯氨酸、色氨酸、丝氨酸、酪氨酸、半胱氨酸、蛋氨酸、天冬酰胺、谷氨酰胺、苏氨酸、天冬氨酸、谷氨酸、赖氨酸、精氨酸、组氨酸、或其组合。

[0108] 在另一优选例中,所述氨基酸包括D型氨基酸和/或L型氨基酸。

[0109] 在另一优选例中,所述聚乙二醇选自下组:PEG3000、PEG8000、PEG6000、PEG3350、或其组合。

[0110] 在另一优选例中,所述聚乙二醇包括分子量(Da)为200-10000的聚乙二醇,较佳地,分子量为3000-10000的聚乙二醇。在另一优选例中,所述RNA聚合酶为T7RNA聚合酶。

[0111] 在另一优选例中,所述的酵母细胞提取物是用包括以下步骤的方法制备:

[0112] (i) 提供酵母细胞;

[0113] (ii) 对酵母细胞进行洗涤处理,获得经洗涤的酵母细胞;

[0114] (iii) 对经洗涤的酵母细胞进行破细胞处理,从而获得酵母粗提物;和

[0115] (iv) 对所述酵母粗提物进行固液分离,获得液体部分,即为酵母细胞提取物。

[0116] 在另一优选例中,所述外源蛋白合成体系包括选自下组的成分:聚乙二醇、4-羟乙基哌嗪乙磺酸、醋酸钾、醋酸镁、核苷三磷酸、氨基酸、磷酸肌酸、二硫苏糖醇(DTT)、磷酸肌酸激酶、蔗糖、本发明第一方面所述的核酸构建物(以萤火虫荧光素酶或绿色荧光蛋白为外源蛋白)、T7RNA聚合酶、酵母细胞提取物、或其组合。

[0117] 本发明第三方面提供了一种载体或载体组合,所述的载体或载体组合含有本发明第一方面所述的核酸构建物。

[0118] 本发明第四方面提供了一种基因工程细胞,所述基因工程细胞的基因组的一个或多个位点整合有本发明第一方面所述的构建物,或者所述基因工程细胞中含有本发明第三方面所述的载体或载体组合。

[0119] 在另一优选例中,所述基因组包括核基因组、细胞质基因组、和/或细胞器基因组。

[0120] 在另一优选例中,所述细胞器包括线粒体、和/或叶绿体。

[0121] 在另一优选例中,所述基因工程细胞包括原核细胞、真核细胞。

[0122] 在另一优选例中,所述真核细胞包括高等真核细胞。

[0123] 在另一优选例中,所述基因工程细胞选自下组:人源细胞(如Hela细胞)、中国仓鼠卵巢细胞、昆虫细胞、麦胚细胞、兔网织红细胞、酵母细胞、或其组合。

[0124] 在另一优选例中,所述基因工程细胞为酵母细胞。

[0125] 在另一优选例中,所述酵母细胞选自下组:酿酒酵母、克鲁维酵母属酵母、或其组合。

[0126] 在另一优选例中,所述克鲁维酵母属酵母选自下组:乳酸克鲁维酵母、马克斯克鲁维酵母、多布克鲁维酵母、或其组合。

[0127] 本发明第五方面提供了一种试剂盒,所述试剂盒中包含的试剂选自下组中的一种或多种:

[0128] (a) 本发明第一方面所述的构建物;

[0129] (b) 本发明第三方面所述的载体或载体组合;

[0130] (c) 本发明第四方面所述的基因工程细胞。

[0131] 在另一优选例中,所述试剂盒还包括 (d) 真核体外生物合成体系(如真核体外蛋白合成体系)。

[0132] 在另一优选例中,所述真核体外生物合成体系选自下组:酵母体外生物合成体系、中国仓鼠卵巢细胞体外生物合成体系、昆虫细胞体外生物合成体系、HeLa细胞体外生物合成体系、或其组合。

[0133] 在另一优选例中,所述真核体外生物合成体系包括真核体外蛋白合成体系。

[0134] 在另一优选例中,所述真核体外蛋白合成体系选自下组:酵母体外蛋白合成体系、中国仓鼠卵巢细胞体外蛋白合成体系、昆虫细胞体外蛋白合成体系、HeLa细胞体外蛋白合成体系、或其组合。

[0135] 在另一优选例中,所述试剂盒还包括 (e) 酵母体外生物合成体系(如酵母体外蛋白合成体系)。

[0136] 在另一优选例中,所述酵母体外生物合成体系(如酵母体外蛋白合成体系)为克鲁维酵母体外生物合成体系(如克鲁维酵母体外蛋白合成体系)(优选乳酸克鲁维酵母体外生物合成体系)(如乳酸克鲁维酵母体外蛋白合成体系)。

[0137] 本发明第六方面提供了一种如本发明第一方面或第二方面所述的构建物、本发明第三方面所述的载体或载体组合、本发明第四方面所述的基因工程细胞或本发明第五方面所述的试剂盒的用途,其特征在于,用于进行高通量的体外蛋白质合成。

[0138] 本发明第七方面提供了一种体外高通量的外源蛋白合成方法,包括步骤:

[0139] (i) 在真核体外生物合成体系存在下,提供本发明第一方面所述的核酸构建物;

[0140] (ii) 在适合的条件下,孵育步骤 (i) 的真核体外生物合成体系一段时间T1,从而合成所述外源蛋白。

[0141] 在另一优选例中,所述方法还包括: (iii) 任选地从所述真核体外生物合成体系中,分离或检测所述外源蛋白。

[0142] 在另一优选例中,所述真核体外生物合成体系为酵母体外生物合成体系(如酵母体外蛋白合成体系)。

[0143] 在另一优选例中,所述酵母体外生物合成体系(如酵母体外蛋白合成体系)为克鲁维酵母体外生物合成体系(如克鲁维酵母体外蛋白合成体系)(优选乳酸克鲁维酵母体外生物合成体系)(如乳酸克鲁维酵母体外蛋白合成体系)。

[0144] 在另一优选例中,所述外源蛋白合成方法包括体外高通量的外源蛋白合成方法。

[0145] 在另一优选例中,所述外源蛋白的编码序列来自原核生物、真核生物。

[0146] 在另一优选例中,所述外源蛋白的编码序列来自动物、植物、病原体。

[0147] 在另一优选例中,所述外源蛋白的编码序列来自哺乳动物,较佳地灵长动物,啮齿动物,包括人、小鼠、大鼠。

[0148] 在另一优选例中,所述的外源蛋白的编码序列编码选自下组的外源蛋白:荧光素蛋白、或荧光素酶(如萤火虫荧光素酶)、绿色荧光蛋白、黄色荧光蛋白、氨酰tRNA合成酶、甘油醛-3-磷酸脱氢酶、过氧化氢酶、肌动蛋白、抗体的可变区域、萤光素酶突变体、 α -淀粉酶、肠道菌素A、丙型肝炎病毒E2糖蛋白、胰岛素前体、干扰素 α A、细胞因子,干扰素 α 2b、白细胞介素-1 β 、溶菌酶素、血清白蛋白、单链抗体段(scFV)、甲状腺素运载蛋白、酪氨酸酶、木聚糖酶、或其组合。

[0149] 在另一优选例中,所述外源蛋白选自下组:荧光素蛋白、或荧光素酶(如萤火虫荧光素酶)、绿色荧光蛋白、黄色荧光蛋白、氨酰tRNA合成酶、甘油醛-3-磷酸脱氢酶、过氧化氢酶、肌动蛋白、抗体的可变区域、萤光素酶突变、 α -淀粉酶、肠道菌素A、丙型肝炎病毒E2糖蛋白、胰岛素前体、干扰素 α A、细胞因子,干扰素 α 2b、白细胞介素-1 β 、溶菌酶素、血清白蛋白、单链抗体段(scFV)、甲状腺素运载蛋白、酪氨酸酶、木聚糖酶、或其组合。

[0150] 在另一优选例中,所述步骤(ii)中,反应温度为20-37 $^{\circ}$ C,较佳地,22-35 $^{\circ}$ C。

[0151] 在另一优选例中,所述步骤(ii)中,反应时间为1-72h,较佳地,2-23h。

[0152] 应理解,在本发明范围内中,本发明的上述各技术特征和在下文(如实施例)中具体描述的各技术特征之间都可以互相组合,从而构成新的或优选的技术方案。限于篇幅,在此不再一一累述。

附图说明

[0153] 图1显示了生物合成中5'-UTR序列对于蛋白质翻译起始的关键性作用。在帽子结构非依赖的翻译起始过程中,5'UTR中的Omega序列在蛋白质因子的协助下招募40S核糖体亚基的示意图。

[0154] 图2显示了不同类型的3'翻译增强子元件CITE(cap-independent translation enhancers,CITE)在体外蛋白质合成体系中影响蛋白质合成效率的差异。将挑选的来自植物病毒的4种3'CITE(k1-TSS、YSS、PTE、TED)插入pD2P-1.0质粒(获自康码(上海)生物科技有限公司,其序列如SEQ ID NO.:1所示)的3'UTR,得到四种质粒模板,分别为pD2P-1.0C001、pD2P-1.0C002、pD2P-1.0C003、pD2P-1.0C004(质粒模板序列SEQ ID NO.:2-5),应用于酵母体外蛋白质合成体系中,测试其合成的荧光素酶的相对光单位值(RLU)。其中,PC为不添加3'CITE的pD2P-1.0质粒合成的荧光素酶的相对光单位值(RLU)。

[0155] 图3显示了将3'CITE与5'发夹结构元件同时应用而影响体外蛋白质合成效率的差异。其中,PC为pD2P-1.0质粒应用于酵母体外蛋白质合成体系中合成的荧光素酶的相对光单位值(RLU)。将4种5'发夹结构序列添加到已经插入了3'CITE的四种模板质粒的 Ω 序列的前端及后端,构建出8种质粒模板pD2P-1.0C005~pD2P-1.0C0012(质粒模板序列SEQ ID NO.:6-13)。

[0156] 图4显示了将3'CITE与5'发夹结构元件同时应用而影响体外蛋白质合成效率的差异。PC为pD2P-1.1质粒(获自康码(上海)生物科技有限公司,其序列如SEQ ID NO.:50所示)应用于酵母体外蛋白质合成体系中合成的增强型绿色荧光蛋白(eGFP)的相对荧光单位值

(RFU)。将pD2P-1.0C006, pD2P-1.0C007和pD2P-1.0C008为克隆载体的模板,将其编码蛋白序列由萤火虫荧光素酶替换为绿色荧光蛋白后,对应得到了pD2P-1.1C001, pD2P-1.1C002和pD2P-1.1C003三个质粒(质粒模板序列SEQ ID NO.:51-53),并应用于酵母体外蛋白质合成体系中合成的增强型绿色荧光蛋白(eGFP)的相对荧光单位值(RFU)。

[0157] 图5显示了本发明核酸构建物提高体外蛋白质翻译起始效率的原理示意图。除了5'UTR元件,3'CITE也能够使得在3'端招募到的真核生物细胞内翻译起始因子复合物eIF4F,使得核糖体小亚基到达RNA的5'端,并且3'CITE与其对应的5'发夹结构发生相互作用,使得mRNA发生环化。

具体实施方式

[0158] 经过广泛而深入的研究,通过大量筛选和摸索,首次意外地发现了一种能够环化mRNA进而增强在体外生物合成体系中合成蛋白质效率的新方法。通过向体外生物合成体系中的mRNA的两端插入环化元件(如3'CITE和其对应的5'发夹结构)来使mRNA形成比较稳定的环状结构,不仅能够使翻译结束的核糖体到达5'UTR快速重新开始新一轮的翻译,还能够利用两端的发夹结构增强mRNA的稳定性,同时还能够与5'UTR上的招募翻译起始因子的RNA元件产生协同效应,更进一步的增加蛋白质翻译的效率。在此基础上,发明人完成了本发明。

[0159] 术语

[0160] 如本文所用,术语“poly(A)”和“poly(A)_n”可互换使用,指腺嘌呤脱氧核糖核酸的多聚链或polyA尾,其中n为 ≥ 10 的正整数。较佳地 $n \geq 30$ 。通常n为10-200,较佳地30-90。

[0161] 如本文所用,术语“本发明的核酸构建物”指具有式A、式I或式II结构的核苷酸序列。本发明的核酸构建物可以是DNA、RNA或其他核酸类似物。

[0162] 如本文所用,对于术语“环化元件Z2和Z4通过RNA-RNA直接作用使得mRNA形成环化结构”,它表示:(i)当本发明的核酸构建物可以是RNA时,由于Z2和Z4为RNA形式的环化元件,因此它们通过RNA-RNA直接作用使得mRNA形成环化结构;和(ii)当本发明的核酸构建物可以是DNA时,由于Z2和Z4为DNA形式的环化元件,因此当DNA形式的核酸构建物被转录后,对应于Z2和Z4的RNA形式的环化元件通过RNA-RNA直接作用使得mRNA形成环化结构。应理解,在本发明中,Z2和Z4元件是一对元件,它们通过RNA-RNA直接作用共同使得转录产物(mRNA)形成环化结构。

[0163] 理论体系的构建:

[0164] 1. 科学背景

[0165] 与真核生物的“帽子结构”依赖的蛋白质翻译相比,在自然界中还存在着一些“帽子结构”非依赖的蛋白质翻译起始方式。细菌mRNA上的核糖体结合位点(ribosome binding site, RBS)在翻译起始因子的辅助下直接招募核糖体。一些病毒和某些真核细胞内的mRNA上的一些特殊的RNA结构能够招募真核细胞蛋白质翻译起始因子,进而利用真核细胞的蛋白质翻译机器,如内部核糖体进入序列(internal ribosome entry site, IRES),和烟草花叶病毒的Omega序列等。这些特殊的RNA元件一般位于mRNA的5'端,通常位于5'UTR内[6,7]。

[0166] 3'UTR在mRNA稳定性和翻译调控中也发挥着重要的调控作用。很多植物的正链RNA病毒的基因组RNA的5'端不含有任何特殊的修饰,简单的为游离的磷酸残基。这类病毒基因

组RNA的一个重要特征为,在RNA的3'UTR中含有帽子结构非依赖的翻译增强子元件(cap-independent translation enhancers,CITE) [8]。3'CITE在功能上可以有效替代5'“帽子结构”功能,使得核糖体由mRNA的5'端进入和扫描。

[0167] 2.基本原理

[0168] 3'CITE的两个重要的特征包括,能够招募真核生物细胞内翻译起始因子复合物eIF4F的亚基,以及能够通过高亲和力的RNA-RNA相互作用(kissing-loop interaction)与位于5'UTR的发夹结构(hairpin loop)相互作用,能够使招募到的翻译起始因子能够到达RNA的5'端,进而起始蛋白质的翻译过程[8]。

[0169] 根据序列和二级结构的不同,目前发现的3'CITE分为以下六种类型:最早在卫星烟草坏死病毒(Satellite tobacco necrosis virus,STNV)中发现的翻译增强结构域(translation enhancer domain,TED)、大麦黄矮病毒样翻译元素(Barley yellow dwarf virus-Like element,BTE)、稷子花叶病毒样翻译元素(Panicum mosaic virus-Like element,PTE)、I型结构元件、Y型结构元件和T型结构元件[8]。除了T型结构元件,一般3'CITE序列的发夹结构可以与相应位于5'UTR中的发夹结构序列直接作用使mRNA形成相对稳定的环状结构。

[0170] 在体外生物合成体系中,将3'CITE和其对应的位于5'UTR的发夹结构整合到不含有“帽子结构”的RNA的两端,能够通过RNA-RNA相互作用直接将RNA环化,新形成的环状RNA(crRNA)能够使翻译结束的核糖体可以不经解离,直接重新到达RNA的5'端。稳定的环状结构还能够阻止RNA酶对RNA的降解,增加稳定性,进一步提高蛋白质合成的效率。同时,某些3'CITE能够通过结合翻译起始因子复合物eIF4F上来招募其他翻译起始因子和复合物,能够与位于5'UTR的其他RNA元件以及3'UTR上的poly(A)元件协同,增加翻译起始的效率[5]。

[0171] 体外外源蛋白合成体系

[0172] 在本发明中,体外外源蛋白合成体系是一种基于真核细胞的转录-翻译偶联的体系,能够以DNA模板为起始合成RNA,或者以DNA或RNA为模板完成蛋白质的体外合成。真核细胞包括酵母细胞、兔网织红细胞、麦胚细胞、昆虫细胞、人源细胞等。真核体外生物合成体系具有能够合成具有复杂结构的RNA或者蛋白质,以及蛋白质翻译后修饰等优点。

[0173] 在本发明中,所述体外外源蛋白合成体系没有特别限制,一种优选的体外外源蛋白合成体系包括酵母体外生物合成体系,作为优选,为酵母体外蛋白合成体系,较佳地为克鲁维酵母表达系统,更佳地为乳酸克鲁维酵母表达系统。

[0174] 酵母(yeast)兼具培养简单、高效蛋白质折叠、和翻译后修饰的优势。其中酿酒酵母(*Saccharomyces cerevisiae*)和毕氏酵母(*Pichia pastoris*)是表达复杂真核蛋白质和膜蛋白的模式生物,酵母也可作为制备体外翻译系统的原料。

[0175] 克鲁维酵母(*Kluyveromyces*)是一种子囊孢子酵母,其中的马克斯克鲁维酵母(*Kluyveromyces marxianus*)和乳酸克鲁维酵母(*Kluyveromyces lactis*)是工业上广泛使用的酵母。与其他酵母相比,乳酸克鲁维酵母具有许多优点,如超强的分泌能力,更好的大规模发酵特性、食品安全的级别、以及同时具有蛋白翻译后修饰的能力等。

[0176] 在本发明中,所述体外外源蛋白合成体系包括:

[0177] (a) 本发明第一方面所述的核酸构建物;和

[0178] (b) 酵母细胞提取物。

[0179] 在一优选实施方式中,所述体外外源蛋白合成体系还包括选自下组的一种或多种组分:

- [0180] (c1) 镁离子;
- [0181] (c2) 钾离子;
- [0182] (c3) 缓冲剂;
- [0183] (c4) 能量再生系统;
- [0184] (c5) 聚乙二醇;
- [0185] (c6) 用于合成RNA的底物;
- [0186] (c7) 用于合成蛋白的底物;
- [0187] (c8) RNA聚合酶;
- [0188] (c9) 任选的外源蔗糖;
- [0189] (c10) 任选的溶剂,所述溶剂为水或水性溶剂。

[0190] 在本发明中,RNA聚合酶没有特别限制,可以选自一种或多种RNA聚合酶,典型的RNA聚合酶为T7RNA聚合酶。

[0191] 在本发明中,所述酵母细胞提取物在体外生物合成体系中的比例不受特别限制,所述酵母细胞提取物在体外生物合成体系中所占体系为20-70%,较佳地,30-60%;更佳地,40-50%。

[0192] 在本发明中,所述的酵母细胞提取物不含完整的细胞,典型的酵母细胞提取物包括用于RNA合成所需的各种类型的RNA聚合酶,和用于蛋白翻译的核糖体、tRNA、氨酰tRNA合成酶、蛋白质合成需要的起始因子和延伸因子以及终止释放因子。此外,酵母细胞提取物中还含有一些源自真核细胞的细胞质中的其他蛋白,尤其是可溶性蛋白。

[0193] 在本发明中,所述的酵母细胞提取物所含蛋白含量为20-100mg/mL,较佳为50-100mg/mL。所述的测定蛋白含量方法为考马斯亮蓝测定方法。

[0194] 在本发明中,所述的酵母细胞提取物的制备方法不受限制,一种优选的制备方法包括以下步骤:

- [0195] (i) 提供酵母细胞;
- [0196] (ii) 对酵母细胞进行洗涤处理,获得经洗涤的酵母细胞;
- [0197] (iii) 对经洗涤的酵母细胞进行破细胞处理,从而获得酵母细胞粗提物;
- [0198] (iv) 对所述酵母细胞粗提物进行固液分离,获得液体部分,即为酵母细胞提取物。

[0199] 本发明中,所述镁离子来源于镁离子源,所述镁离子源选自下组:醋酸镁、谷氨酸镁、或其组合。所述钾离子来源于钾离子源,所述钾离子源选自下组:醋酸钾、谷氨酸钾、或其组合。所述缓冲剂选自下组:4-羟乙基哌嗪乙磺酸、三羟甲基氨基甲烷、或其组合。

[0200] 本发明中,所述能量再生系统选自下组:磷酸肌酸/磷酸肌酸酶系统、糖酵解途径及其中间产物能量系统、或其组合。

[0201] 本发明中,所述的合成RNA的底物包括:核苷单磷酸、核苷三磷酸、或其组合。所述的合成蛋白的底物为氨基酸混合物,包括:1-20种天然氨基酸、以及非天然氨基酸。

[0202] 所述核苷三磷酸选自下组:腺嘌呤核苷三磷酸、鸟嘌呤核苷三磷酸、胞嘧啶核苷三磷酸、尿嘧啶核苷三磷酸、或其组合在本发明中,各种单核苷酸的浓度没有特别限制,每种单核苷酸的浓度为0.5-5mM,较佳地为1.0-2.0mM。

[0203] 所述体外外源蛋白合成体系中的氨基酸混合物可包括天然或非天然氨基酸,可包括D型或L型氨基酸。代表性的氨基酸包括但并不限于20种天然氨基酸:甘氨酸、丙氨酸、缬氨酸、亮氨酸、异亮氨酸、苯丙氨酸、脯氨酸、色氨酸、丝氨酸、酪氨酸、半胱氨酸、蛋氨酸、天冬酰胺、谷氨酰胺、苏氨酸、天冬氨酸、谷氨酸、赖氨酸、精氨酸和组氨酸。每种氨基酸的浓度通常为0.01-0.5mM,较佳地0.02-0.2mM,如0.05mM、0.06mM、0.07mM、0.08mM。

[0204] 在优选例中,所述体外外源蛋白合成体系还含有聚乙二醇或其类似物。聚乙二醇或其类似物的浓度没有特别限制,通常,聚乙二醇或其类似物的浓度(w/v)为0.1-8%,较佳地,0.5-4%,更佳地,1-2%,以所述生物合成体系的总重量计。优选PEG3000、PEG3350、PEG6000或PEG8000。应理解,本发明的体系还包括其他各种分子量的聚乙二醇,如PEG200、400、1500、2000、4000、10000等。

[0205] 在优选例中,所述体外外源蛋白合成体系还含有蔗糖。蔗糖的浓度没有特别限制,通常,蔗糖的浓度为0.03-40wt%,较佳地,0.08-10wt%,更佳地,0.1-5wt%,以所述蛋白合成体系的总重量计。

[0206] 一种特别优选的体外外源蛋白合成体系,含有以下组分:22mM pH为7-8的4-羟乙基哌嗪乙磺酸,30-150mM醋酸钾,1.0-5.0mM醋酸镁,1.5-4mM核苷三磷酸混合物,0.08-0.24mM的氨基酸混合物,25mM磷酸肌酸,1.7mM二硫苏糖醇,0.27mg/mL磷酸肌酸激酶,1%-4%聚乙二醇,0.5%-2%蔗糖,8-20ng/ μ L本发明的核酸构建物(以萤火虫荧光素酶或绿色荧光蛋白为外源蛋白),0.027-0.054mg/mL T7RNA聚合酶,50%体积的酵母细胞提取物。

[0207] Ω 序列

[0208] 如本文所用,术语“ Ω 序列”是烟草花叶病毒基因组的5'端前导序列,是该种病毒的翻译增强子。 Ω 的DNA序列含有68个碱基对,由1-6个(优选2-4个,更优选3个)8碱基对直接重复模块(ACAATTAC)和1-5个(优选1-3个,更优选1个)(CAA)_p模块,其中,p为6-12个,较佳地,8-10个。这两个模块对于 Ω 序列的增强翻译功能是关键。在本发明的酵母体外蛋白质合成体系中, Ω 序列能够起始“帽子结构”非依赖性的蛋白质翻译,这种功能可能是通过招募翻译起始因子eIF4G实现的。但是 Ω 序列起始蛋白质翻译的效率比较低,需要对其构成进行优化,并配合其他DNA元件或者蛋白质来提高蛋白质翻译的效率。

[0209] Kozak序列

[0210] 通过对已知的真核生物的mRNA分子中翻译起始密码子(AUG)上下游序列进行分析,找出的共有序列被称为Kozak序列。Kozak序列被证实能够增强mRNA的翻译起始效率。不同物种的Kozak序列是不相同的,如酿酒酵母细胞(*Saccharomyces cerevisiae*)与哺乳动物细胞的Kozak序列就存在着显著差异。

[0211] 在本发明中,所用的Kozak序列包括6-12个腺嘌呤脱氧核苷寡聚链(较佳地,8-10)、翻译起始密码子(如ATG、ATA、ATT、GTG、TTG等,优选ATG)和丝氨酸密码子(如TCT、TCC、TCA、TCG、AGT、AGC等,优选TCT),来源于克鲁维酵母(优选乳酸克鲁维酵母)。

[0212] 外源编码序列(外源DNA)

[0213] 如本文所用,术语“外源编码序列”与“外源DNA”可互换使用,均指外源的用于指导RNA或者蛋白质合成的DNA分子。通常,所述的DNA分子为线性的或环状的。所述的DNA分子含有编码外源RNA或者外源蛋白的序列。

[0214] 在本发明中,所述的外源编码序列包括但不限于:基因组序列、cDNA序列。所述

的外源编码序列还含有启动子序列、5'非翻译序列、3'非翻译序列。

[0215] 本发明中,所述外源DNA的选择没有特别限制,通常,外源DNA选自下组:小非编码RNA (small non-coding RNA, sncRNA)、长非编码RNA (long non-coding RNA, lncRNA)、转运RNA (transfer RNA, tRNA)、葡萄糖胺-6-磷酸合成酶 (Glucosamine-6-phosphate synthase, glmS) 等核酶、小的核RNA (small nuclear RNA, snRNA)、剪接体 (spliceosome) 等RNA与蛋白质的复合物、其他各种非编码RNA、或其组合。

[0216] 外源DNA还可选自下组:编码荧光素蛋白或荧光素酶(如萤火虫荧光素酶)、绿色荧光蛋白、黄色荧光蛋白、氨酰tRNA合成酶、甘油醛-3-磷酸脱氢酶、过氧化氢酶、肌动蛋白、抗体的可变区域的外源DNA、萤光素酶突变体的DNA、或其组合。

[0217] 外源DNA还可以选自下组:编码 α -淀粉酶、肠道菌素A、丙型肝炎病毒E2糖蛋白、胰岛素前体、干扰素 α A、白细胞介素-1 β 、溶菌酶素、血清白蛋白、单链抗体段 (scFV)、甲状腺素运载蛋白、酪氨酸酶、木聚糖酶的外源DNA、或其组合。

[0218] 在一优选实施方式中,所述外源DNA编码选自下组的蛋白:绿色荧光蛋白 (enhanced GFP, eGFP)、黄色荧光蛋白 (YFP)、大肠杆菌 β -半乳糖苷酶 (β -galactosidase, LacZ)、人赖氨酸-tRNA合成酶 (Lysine-tRNA synthetase)、人亮氨酸-tRNA合成酶 (Leucine-tRNA synthetase)、拟南芥甘油醛3-磷酸脱氢酶 (Glyceraldehyde-3-phosphate dehydrogenase)、鼠过氧化氢酶 (Catalase)、或其组合。

[0219] 核酸构建物

[0220] 本发明提供了一种核酸构建物,所述核酸构建物具有5'-3'的式A所示的结构:

[0221] Z1-Z2-Y1-Z3-Z4-Y2-Z5 (A)

[0222] 式中,

[0223] Z1为启动子元件;

[0224] Z2和Z4为环化元件,其中环化元件Z2和Z4通过RNA-RNA直接作用使对应于式A的mRNA转录产物形成一环化结构,并且所述环化元件选自下组:TED、BTE、PTE、I型结构元件、Y型结构元件、T型结构元件、其他能够形成RNA-RNA直接作用的元件、或其组合;

[0225] Y1为无或“Z6-Z7-Z8”所示的元件;

[0226] Z3为外源蛋白的编码序列;

[0227] Y2为无或Z9;

[0228] Z5为终止子元件;

[0229] Z6为无或增强子元件;

[0230] Z7为无或烟草花叶病毒的5'端前导序列 (leading sequence) - Ω 序列;

[0231] Z8为无或Kozak序列;

[0232] Z9为腺嘌呤脱氧核糖核酸的多聚链;

[0233] 各“-”独立地为键或核苷酸连接序列。

[0234] 在一优选实施方式中,本发明的所述的构建物具有从5'至3'的式I结构:

[0235] Z1-Z2-Z3-Z4-Z5 (I)

[0236] 式中,

[0237] Z1-Z4和Z5分别为用于构成所述构建物的元件;

[0238] 各“-”独立地为键或核苷酸连接序列;

[0239] Z1为启动子元件,优选地所述启动子元件选自下组:T7启动子、T3启动子、SP6启动子、或其组合;

[0240] Z4和Z2为环化元件,优选地所述环化元件选自下组:TED、BTE、PTE、I型结构元件、Y型结构元件、T型结构元件、其他能够形成RNA-RNA直接作用的元件、或其组合;

[0241] Z3为外源蛋白的编码序列;

[0242] Z5为终止子元件,优选地所述终止子元件选自下组:T7终止子、T3终止子、SP6终止子、或其组合。

[0243] 在另一优选实施方式中,本发明的核酸构建物含有结构如式II所示的核酸序列:

[0244] Z1-Z2-Z6-Z7-Z8-Z3-Z4-Z9-Z5 (II)

[0245] 式中,

[0246] Z1-Z9分别为用于构成所述构建物的元件;

[0247] 各“-”独立地为键或核苷酸连接序列;

[0248] Z1为启动子元件,优选地,所述启动子元件选自下组:T7启动子、T3启动子、SP6启动子、或其组合;

[0249] Z4和Z2为环化元件,优选地,所述环化元件选自下组:TED、BTE、PTE、I型结构元件、Y型结构元件、T型结构元件、其他能够形成RNA-RNA直接作用的元件、或其组合;

[0250] Z6为增强子元件,优选地,所述增强子元件包括IRES元件;

[0251] Z7为烟草花叶病毒的5'端前导序列(leading sequence)- Ω 序列;

[0252] Z8为Kozak序列,优选地,所述Kozak序列来源于酵母;

[0253] Z3为外源蛋白的编码序列;

[0254] Z9为腺嘌呤脱氧核糖核酸的多聚链[poly(A)],优选地,所述多聚链选自下组:50A、70A、90A、或其组合;

[0255] Z5为终止子元件,优选地,所述终止子元件选自下组:T7终止子、T3终止子、SP6终止子、或其组合。

[0256] 在本发明中,所述外源蛋白的编码序列的选择没有特别限制,通常,外源蛋白的编码序列选自下组:编码荧光素蛋白、或荧光素酶(如萤火虫荧光素酶)、绿色荧光蛋白、黄色荧光蛋白、氨酰tRNA合成酶、甘油醛-3-磷酸脱氢酶、过氧化氢酶、肌动蛋白、抗体的可变区域的外源DNA、荧光素酶突变体的DNA。

[0257] 外源蛋白的编码序列还可以编码选自下组的蛋白: α -淀粉酶、肠道菌素A、丙型肝炎病毒E2糖蛋白、胰岛素前体、干扰素 α A、白细胞介素-1 β 、溶菌酶素、血清白蛋白、单链抗体段(scFv)、甲状腺素运载蛋白、酪氨酸酶、木聚糖酶。

[0258] 此外,本发明的所述核酸构建物可以是线性的,也可以是环状的。本发明的所述核酸构建物可以是单链的,也可以是双链的。本发明的所述核酸构建物可以是DNA,也可以是RNA,或DNA/RNA杂合。

[0259] 在一优选实施方式中,本发明的核酸构建物的序列如SEQ ID NO.:2-13及SEQ ID NO.:51-53所示。在另一优选例中,所述的构建物还包括选自下组的元件或其组合:poly(A)元件、转运元件、基因靶向元件、筛选标记基因、增强子、抗性基因、转座酶编码基因。

[0260] 多种选择性标志基因均可应用于本发明,包括但不限于:营养缺陷型标记,抗性标记,报告基因标记。选择性标志的应用对于重组细胞(重组子)的筛选起到作用,使得受体细

胞能够与未转化的细胞进行显著区分。营养缺陷型标记是通过转入的标记基因与受体细胞突变基因互补,从而使受体细胞表现野生型生长。抗性标记是指将抗性基因转入受体细胞中,转入的基因使受体细胞在一定的药物浓度下表现抗药性。作为本发明的优选方式,应用抗性标记来实现重组细胞的便捷筛选。

[0261] 在本发明中,在本发明的酵母体外生物合成体系(如酵母蛋白生物合成体系)中应用本发明的核酸构建物,可显著提高外源蛋白翻译的效率,其中,在一优选实施方式中,本发明的核酸构建物(如pD2P-1.0C008)合成的荧光素酶的相对光单位值是pD2P-1.0序列的2.8倍,本发明的pD2P-1.1C003合成的绿色荧光蛋白的RLU是pD2P-1.1序列的1.8倍。

[0262] 载体,基因工程细胞

[0263] 本发明还提供了一种载体或载体组合,所述载体含有本发明的核酸构建物。优选地,所述载体选自:细菌质粒、噬菌体、酵母质粒、或动物细胞载体、穿梭载体;所述的载体为转座子载体。用于制备重组载体的方法是本领域普通技术人员所熟知的。只要其能够在宿主主体内复制和稳定,任何质粒和载体都是可以采用的。

[0264] 本领域普通技术人员可以使用熟知的方法构建含有本发明所述的启动子和/或目的基因序列的表达载体。这些方法包括体外重组DNA技术、DNA合成技术、体内重组技术等。

[0265] 本发明还提供了一种基因工程细胞,所述的基因工程细胞含有所述的核酸构建物或载体或载体组合,或所述的基因工程细胞染色体整合有所述的核酸构建物或载体。

[0266] 在另一优选例中,所述的基因工程细胞还包括含有编码转座酶基因的载体或其染色体上整合有转座酶基因。

[0267] 优选地,所述的基因工程细胞为真核细胞。

[0268] 在另一优选例中,所述真核细胞,包括但不限于:人体细胞、中国仓鼠卵巢细胞、昆虫细胞、麦胚细胞、兔网织红细胞等高等真核生物细胞。

[0269] 在另一优选例中,所述真核细胞,包括但不限于:酵母细胞,优选克鲁维酵母细胞,更优选乳酸克鲁维酵母细胞。

[0270] 本发明的构建物或载体,可以用于转化适当的基因工程细胞。基因工程细胞可以是原核细胞,如大肠杆菌,链霉菌属、农杆菌;或是低等真核细胞,如酵母细胞;或是高等动物细胞,如昆虫细胞。本领域一般技术人员都清楚如何选择适当的载体和基因工程细胞。用重组DNA转化基因工程细胞可用本领域技术人员熟知的常规技术进行。当宿主为原核生物时,如大肠杆菌,可以用CaCl₂法处理,也可用电穿孔法进行。当宿主是真核生物,可选用如下的DNA转染方法:磷酸钙共沉淀法,常规机械方法,包括显微注射、电穿孔、脂质体包装等。转化植物也可使用农杆菌转化或基因枪转化等方法,例如叶盘法、幼胚转化法、花芽浸泡法等。

[0271] 调控体外蛋白合成的方法

[0272] 本发明提供了一种调控体外蛋白合成的方法,包括步骤:

[0273] (i) 在真核体外生物合成体系存在下,提供本发明第一方面所述的核酸构建物;

[0274] (ii) 在适合条件下,孵育步骤(i)的真核体外生物合成体系一段时间T1,从而合成所述外源蛋白。

[0275] 所述的步骤(ii)所述的适合条件为:反应温度20-37℃,较佳地,22-35℃,反应时间为1h-72h,较佳地,2-23h,通过振荡、翻滚、旋转、搅拌、摇匀、翻转或其他方式持续混匀反

应,或者混匀后静置反应。

[0276] 在另一优选例中,所述方法还包括步骤(iii)从所述真核体外生物合成体系中分离或检测所述外源蛋白。

[0277] 试剂盒

[0278] 本发明提供了用于外源蛋白合成的试剂盒,包括:

[0279] (k1) 第一容器,以及位于第一容器内的本发明第一方面所述的核酸构建物;

[0280] (k2) 第二容器,以及位于第二容器内的酵母细胞提取物。在一优选实施方式中,所述的第一容器、第二容器是同一容器或不同容器。

[0281] 一种特别优选的体外外源蛋白合成的试剂盒包含一个体外外源蛋白合成体系,该合成体系包括:酵母细胞提取物,本发明的核酸构建物以及其他合适的试剂。所述合适的试剂包括(但不限于):10-50mM(如22mM),pH为7.4的4-羟乙基哌嗪乙磺酸,30-150mM醋酸钾,1.0-5.0mM醋酸镁,1.5-4mM核苷三磷酸混合物,0.08-0.24mM的氨基酸混合物,25mM磷酸肌酸,1.7mM二硫苏糖醇,0.27mg/mL磷酸肌酸激酶,1%-4%聚乙二醇,0.5%-2%蔗糖,,0.027-0.054mg/mL T7RNA聚合酶。

[0282] 本发明的主要优点包括:

[0283] (1) 本发明首次发现,将任选的启动子、环化元件(如植物病毒的3' CITE和其对应的5'发夹结构)、任选的 Ω 序列、任选的Kozak序列和外源蛋白的编码序列作为核酸构建物,应用于本发明的真核体外生物合成体系(如酵母体外蛋白合成体系)中,可显著提高外源蛋白翻译的效率。

[0284] (2) 本发明对mRNA的5'UTR和3'UTR插入能够相互作用的RNA元件从而使mRNA形成比较稳定的环状结构,还能够利用两端的发夹结构增强mRNA的稳定性,同时还能够与5'UTR上的招募翻译起始因子的RNA元件以及3'UTR上的poly(A)结构产生协同效应,增强真核体外生物合成体系的蛋白质翻译起始效率。含有3' CITE与5'发夹结构的质粒在酵母体外蛋白质合成体系中引起萤火虫荧光素酶(Firefly luciferase,简称Fluc,下同)发出的相对光单位值(Relative Light Units,简称RLU,下同)超过5'单独 Ω 序列的质粒。其中,pD2P-1.0C008在乳酸克鲁维酵母体外生物合成体系中起始合成的Fluc的RLU达到了 4.44×10^8 ,是pD2P-1.0序列的2.8倍(pD2P-1.0序列的RLU为 1.58×10^8)。pD2P-1.1C003在乳酸克鲁维酵母体外生物合成体系中起始合成的绿色荧光蛋白(简称,eGFP,下同)的相对荧光单位值(简称,RFU,下同)达到了472,是pD2P-1.1序列的1.8倍(pD2P-1.1序列的RFU为267)。

[0285] (3) 与酿酒酵母相比,乳酸克鲁维酵母因其安全性和高效性可以被应用于食品和药品领域蛋白质的生产,加上体外生物合成体系的优点,如适应于高通量的蛋白质合成筛选,合成毒性蛋白质和时间短成本低等,所以乳酸克鲁维酵母细胞来源的体外生物合成体系在蛋白质合成相关领域也能够得到广泛的应用。

[0286] (4) 本发明提供的核酸构建物不仅能够提高真核体外生物合成体系蛋白质翻译合成的效率,更主要的是能够增加乳酸克鲁维酵母体外生物合成体系用于不同蛋白质合成的可能性。

[0287] (5) 本发明的核酸构建物不仅提高了蛋白质翻译合成的效率,而且提供了一种设计用于真核体外生物合成体系的DNA元件的新思路与新方法,能够极大的提高相关体系在科研和工业生产领域的应用。

[0288] (6) 本发明首次发现,将强启动子(如T7启动子、T3启动子、SP6启动子)与本发明的核酸构建物组合,也可获得非常高的蛋白合成效率。

[0289] 下面结合具体实施例,进一步阐述本发明。应理解,这些实施例仅用于说明本发明而不适用于限制本发明的范围。下列实施例中未注明具体条件的实验方法,通常按照常规条件,例如Sambrook等人,分子克隆:实验室手册(New York: Cold Spring Harbor Laboratory Press, 1989)中所述的条件,或按照制造厂商所建议的条件。除非另外说明,否则百分比和份数是重量百分比和重量份数。

[0290] 如无特别说明,则本发明实施例中所用的材料和试剂均为市售产品。

[0291] 实施例1植物病毒来源的3'非依赖帽子结构的翻译增强子CITE和相互作用的对应5'UTR的设计

[0292] 1.1构建4个质粒:3'UTR中的非依赖帽子结构的翻译增强子(cap-independent translation enhancers,简称CITE)在功能上可以有效替代5'帽子结构的功能,使得核糖体由mRNA的5'端进入和扫描。不同类型的3'CITE(表1)能够不同程度上促进起始体外蛋白质合成。含有3'CITE编码序列和报告蛋白(优选Fluc)基因的质粒设计:将上述4个3'CITE编码序列插入已有的质粒pD2P-1.0质粒(获自康码(上海)生物科技有限公司,其序列如SEQ ID NO.:1所示)的3'UTR中,形成4个新的质粒,分别为pD2P-1.0C001、pD2P-1.0C002、pD2P-1.0C003、pD2P-1.0C004,其中有1个质粒pD2P-1.0C002起始合成的Fluc的RLU非常显著性地($p < 0.01$)高于模板上仅使用传统的 Ω 序列的原模板PC(positive control),其他3个质粒则与原PC模板没有显著性差异($p > 0.05$)(图2)。4种质粒模板序列如SEQ ID NO.:2-5所示。

[0293] 1.2构建8个质粒:在添加了3'CITE编码序列的4个质粒模板上,进一步在5'端添加其对应的位于5'UTR的发夹结构的编码序列。在改变5'发夹结构(表1)编码序列()与原来5' Ω 序列的上下游顺序后,共有8种质粒模板的设计(pD2P-1.0C005;pD2P-1.0C006、pD2P-1.0C007、pD2P-1.0C008和pD2P-1.0C009、pD2P-1.0C010、pD2P-1.0C011、pD2P-1.0C012)。8种质粒模板序列如SEQ ID NO.:6-13所示,与上述质粒的顺序一一对应。

[0294] 1.3构建3个质粒:在进一步验证Fluc的实验结果的基础上,以加入环化元件并具有增益效果的pD2P-1.0C006,pD2P-1.0C007和pD2P-1.0C008为克隆载体的模板,将其编码蛋白序列由Fluc替换为eGFP后,对应得到了pD2P-1.1C001,pD2P-1.1C002和pD2P-1.1C003三个质粒,序列如SEQ ID NO.:51-53所示。

[0295] 将构建质粒转化感受态细胞过夜培养,挑取6个单克隆进行扩大培养后,进行测序确认正确后,提取质粒保存。

[0296] 表1.3'CITE与相应5'序列信息

病毒来源	3'CITE 类型	3'CITE 作用序列	相应 5'发夹作用序列
PEMV	kl-TSS	UCGCCA	UGGCGA
TBSV	YSS	GGUCG	CGACC

[0298]	HCRSV	PTE	GCCA	UGGC
	STNV	TED	UUCCUG	CAGGAA

[0299] 实施例2:含有3'CITE编码序列和其对应的5'发夹结构编码序列的体外蛋白质合成体系质粒的构建

[0300] 2.1质粒的构建:构建1.1所述4个质粒所具体使用的引物如表2所示,1.2所述8个质粒所具体使用的引物如表3所示,1.3所述3个质粒所具体使用的引物如表4所示。

[0301] 具体构建过程如下:

[0302] 对于待插入的3'CITE片段,以pD2P-1.0 (SEQ ID NO.:1) 或pD2P-1.1 (SEQ ID NO.:50) 为载体质粒。对于待插入的5'发夹结构片段,则分别使用已插入3'CITE的质粒为其各自对应的载体质粒。对于待插入片段和载体分别使用两对引物进行PCR扩增,并各取10 μ L扩增产物进行混合;向20 μ L扩增产物中加入1 μ L Dpn I,37 $^{\circ}$ C孵育6h;将DpnI处理后产物4 μ L加入50 μ L DH5 α 感受态细胞中,冰上放置30min,42 $^{\circ}$ C热激45s后,冰上放置3min,加入200 μ L LB液体培养基37 $^{\circ}$ C振荡培养4h,涂布于含有Amp抗生素的LB固体培养基上过夜培养;挑取6个单克隆进行扩大培养后,进行测序确认正确后,提取质粒保存。

[0303] 表2.3'CITE质粒的PCR扩增引物

[0304]

质粒名称	质粒序列 SEQ ID NO.:	质粒扩增引物名称	SEQ ID NO.:	3'CITE 扩增引物 名称	SEQ ID NO.:
pD2P-1.0C001	2	Kl-TSSPTE-615_2b	14	Kl-TSSPTE-625_1 f	16
		kl-TSS-2f	15	kl-TSS-1b	17
pD2P-1.0C002	3	YSS-62_2b	18	YSS-60_1f	20
		YSS-615_2f	19	YSS-615_1b	21
pD2P-1.0C003	4	PTE-62_2b	22	PTE-62_1f	24
		PTE-615_2f	23	PTE-61_1b	25
pD2P-1.0C004	5	TED-61_2b	26	TED-605_1f	28
		TED-615_2f	27	TED-615_1b	29

[0305] 表3.5'发夹结构质粒的PCR扩增引物

[0306]

质粒名称	质粒序列 SEQ ID NO.:	质粒扩增引物名称	SEQ ID NO.:	5'发夹结构扩增 引物名称	SEQ ID NO.:
pD2P-1.0C005	6	NCEO-605_2b	30	PEMV289U-615_ 1f	32
		NCEO-61_2f	31	PEMV289U-585_ 1b	33
pD2P-1.0C006	7	NCEO-605_2b	30	TBSVU-61_1f	34
		NCEO-61_2f	31	TBSVU-605_1b	35
pD2P-1.0C007	8	NCEO-605_2b	30	HCRSVU-61_1f	36
		NCEO-61_2f	31	HCRSVU-60_1b	37
pD2P-1.0C008	9	NCEO-605_2b	30	SNTVU-61_1f	38
		NCEO-61_2f	31	SNTVU-60_1b	39
pD2P-1.0C009	10	ONCE-60_2b	40	PEMV289-59_1f	42
		NCEFluc-60_2f	41	PEMV289-61_1b	43
pD2P-1.0C010	11	ONCE-60_2b	40	TBSV-59_1f	44
		NCEFluc-60_2f	41	TBSV-59_1b	45
pD2P-1.0C011	12	ONCE-60_2b	40	HCRSV-60_1f	46
		NCEFluc-60_2f	41	HCRSV-62_1b	47
pD2P-1.0C012	13	ONCE-60_2b	40	SNTV-60_1f	48
		NCEFluc-60_2f	41	SNTV-60_1b	49

[0307] 表4.eGFP质粒的PCR扩增引物

[0308]

质粒名称	质粒序列 SEQ ID NO.:	质粒扩增引物名称	SEQ ID NO.:	绿色荧光蛋白序 列扩增引物名称	SEQ ID NO.:
pD2P-1.1C001	51	(9A)O vR	54	O10AGFP iF	56
		(GFP)7YSS vF	55	(LT6)GFP iR	57
pD2P-1.1C002	52	(9A)O vR	54	O10AGFP iF	56
		(GFP)11HPTE vF	58	(LT6)GFP iR	57

[0309]

pD2P-1.1C003	53	(9A)O vR	54	O10AGFP iF	56
		(GFP)15TED vF	59	(LT6)GFP iR	57

[0310] 实施例3:3'CITE和相应5'UTR组件在酵母体外蛋白质合成体系中的应用

[0311] 3.1实验设计:使用不含有3'CITE和/或5'发夹结构编码序列的DNA片段pD2P-1.0(Fluc为外源蛋白)或pD2P-1.1(eGFP为外源蛋白)或作为阳性对照,以不加任何DNA模板的反应组为阴性对照(Negative control,NC),每个样品均设计三组独立实验。

[0312] 3.2 PCR 扩增:利用PCR的方法,并使用引物T7_pET21a_F:CGCGAAATTAATACGACTCACTATAGG(SEQ ID NO.:60)和T7ter_pET21a_R:TCCGGATATAGTTCCTCCTTCAG(SEQ ID NO.:61)将所有质粒中位于T7转录起始序列和终止序列之间包含串联DNA元件与外源蛋白基因的片段进行扩增。

[0313] 3.3纯化:并将扩增得到的DNA片段用乙醇沉淀的方法进行纯化和富集:向PCR产物中加入1/10体积的3M醋酸钠(pH 5.2),然后再加入2.5-3倍体积(该体积为加入醋酸钠之后的体积)的95%的乙醇,置于冰上孵育15min;室温条件下以高于14000g的速度离心30min,弃掉上清;使用70%乙醇进行清洗,然后再以同样速度离心15min,弃掉上清,并用超纯水将沉淀溶解,测定DNA浓度。

[0314] 3.4孵育和检测:按照使用说明,将纯化的DNA片段加入到自制的乳酸克鲁维酵母体外蛋白质合成体系中。并将上述反应体系置于20-30℃的环境中,静置孵育约2-8h。

[0315] 对于以Fluc为外源蛋白的体系,反应结束后,在96孔白色微孔板或者384孔白色微孔板中加入等体积的Fluc底物荧光素(luciferin),立即放置于Envision2120多功能酶标仪(Perkin Elmer),读数,检测Fluc活性,以RLU作为活性参考值,如图2、图3所示。

[0316] 对于以绿色荧光蛋白为外源蛋白的体系,反应结束后,在96孔黑板或384孔黑板中加入10μL反应体系溶液,立即放置于Tecan 2000多功能酶标仪,读数,检测绿色荧光蛋白活性,以RFU作为活性参考值,如图4所示。

[0317] 3.5实验结果:

[0318] (1)含有植物病毒来源的3'CITE与5'发夹结构的质粒的设计

[0319] 将4种不同3'CITE编码序列插入pD2P-1.0质粒的3'UTR,得到四种质粒模板,将4种不同5'发夹结构编码序列插入上述四种质粒模板的Ω前端或者后端,一共形成12个用于体外蛋白质合成的质粒,质粒的设计所采用的方法见实施例1。

[0320] 构建3个质粒:在进一步验证Fluc的实验结果的基础上,以加入环化元件并具有增益效果的pD2P-1.0C006,pD2P-1.0C007和pD2P-1.0C008为克隆载体的模板,将其编码蛋白序列由Fluc替换为eGFP后,对应得到了pD2P-1.1C001,pD2P-1.1C002和pD2P-1.1C003三个质粒,序列如SEQ ID NO.:51-53所示。

[0321] (2)含有植物病毒来源的3'CITE编码序列与5'发夹结构编码序列的质粒的构建

[0322] 经过多次尝试,经基因测序确认,最终构建成功所有15个体外蛋白质合成体系质粒。

[0323] (3)3'CITE与5'发夹结构元件在酵母体外蛋白质合成体系中的应用

[0324] 如图2和图3所示,筛选的12个含有3'CITE编码序列与5'发夹结构编码序列的质粒

在酵母体外蛋白质合成体系中引起Fluc发出的RLU有4个非常显著性 ($p < 0.01$) 地超过改造前pD2P-1.0序列,即pD2P-1.0C002、pD2P-1.0C006、pD2P-1.0C007和pD2P-1.0C008,其RLU分别达到了 2.14×10^8 、 2.60×10^8 、 4.06×10^8 和 4.44×10^8 ,分别是pD2P-1.0序列的1.4倍、1.7倍、2.6倍、和2.8倍 (pD2P-1.0序列的RLU为 1.57×10^8)。另外有3个改造后质粒的RLU非常显著地低于pD2P-1.0序列的RLU,分别为pD2P-1.0C005、pD2P-1.0C009、pD2P-1.0C010。其余两个质粒pD2P-1.0C011和pD2P-1.0C012的RLU与改造前的pD2P-1.0没有显著性差异 ($p > 0.05$)。

[0325] 因为处于检测仪器的RLU与蛋白质浓度关系的线性范围内,活性最高的pD2P-1.0C008的RLU是pD2P-1.0质粒的2.8倍,表明本例中含有3' CITE编码序列与5'发夹结构编码序列的质粒能够提高大约2.8倍的荧光素酶合成。

[0326] 如图4所示,验证5'和3'端环化元件的增益效果的三个质粒在酵母体外蛋白质合成体系中引起绿色荧光蛋白 (eGFP) 发出的RFU有3个非常显著性 ($p < 0.01$) 地超过改造前pD2P-1.1序列,pD2P-1.1C001,pD2P-1.1C002和pD2P-1.1C003的RFU分别达到了398.33,386.33和472,分别是pD2P-1.1序列的1.5倍、1.4倍和1.8倍 (pD2P-1.1序列的RFU为267.33)。

[0327] 因为处于检测仪器相对荧光单位值RFU与蛋白质浓度关系的线性范围内,活性最高的pD2P-1.1C003的RFU是pD2P-1.1质粒的1.8倍,表明本例中含有3' CITE编码序列与5'发夹结构编码序列的质粒能够提高大约1.8倍的绿色荧光蛋白质合成。

[0328] 在本发明中,通过采用特殊设计的式A (或式I或II) 所示的核酸构建物,可以高效地使得mRNA转录产物形成稳定的环状结构。

[0329] 本发明的上述实验结果表明:本发明将3' CITE与5'发夹结构添加到mRNA的两端能够提高蛋白质合成效率,两者能够通过RNA-RNA相互作用直接将RNA环化,新形成的crRNA能够使翻译结束的核糖体可以不经解离,直接重新到达RNA的5'端 (如图5)。

[0330] 此外,稳定的环状结构还能够阻止RNA酶对RNA的降解,提高稳定性,进一步提高蛋白质合成的效率。该设计能够应用于酵母体外蛋白质合成体系中,起始蛋白质合成的效率能够超过常用的 Ω 序列加上90A的尾巴的模板质粒。其中pD2P-1.0C008起始蛋白质合成的量是pD2P-1.0序列的2.8倍,其中pD2P-1.1C003起始蛋白质合成的量是pD2P-1.1序列的1.8倍,提高了酵母体外蛋白质合成体系翻译蛋白质的效率,增加了乳酸克鲁维酵母体外合成体系进行蛋白质合成起始翻译元件的选择性,极大增强了乳酸克鲁维酵母体外蛋白质合成体系的可用性。

[0331] 在本发明提及的所有文献都在本申请中引用作为参考,就如同每一篇文献被单独引用作为参考那样。此外应理解,在阅读了本发明的上述讲授内容之后,本领域技术人员可以对本发明作各种改动或修改,这些等价形式同样落于本申请所附权利要求书所限定的范围。

[0332] 参考文献

[0333] 1.Sonenberg&Hinnebusch.,Regulation of translation initiation in eukaryotes:mechanisms and biological targets.Cell,2009.136 (4) :731-45.

[0334] 2.Dever et al.,Mechanism and Regulation of Protein Synthesis in Saccharomyces cerevisiae.Genetics,2016.203 (1) :65-107.

- [0335] 3.M.,D.J.R.G.,Nucleic Acid.Encyclopedia of Cell Biology.Elsevier, 2015.
- [0336] 4.Wells et al.,Circularization of mRNA by eukaryotic translation initiation factors.Mol.Cell,1998.2:135-140.
- [0337] 5.Ogawa et al.,Identification of short untranslated regions that sufficiently enhance translation in high-quality wheat germ extract.Bioorganic&Medicinal Chemistry Letters,2014.24(16):3724-3727.
- [0338] 6.Komar,A.A.,and Hatzoglou,M.Cellular IRES-mediated translation:the war of ITAFs in pathophysiological states.Cell cycle,2011.10:229-240.
- [0339] 7.Lee,K.M.,et al.Regulation Mechanisms of Viral IRES-Driven Translation.Trends in microbiology,2017.
- [0340] 8.Simon&Miller.,3'Cap-Independent translation enhancers of plant viruses.Annu Rev Microbiol,2013.67:21-42.
- [0341] 9.Roux&Topisirovic.,Regulation of mRNA translation by signaling pathways.Cold Spring Harb Perspect Biol,2012.4(11):843-853.
- [0342] 10.Muhlrad,D.,C.J.Decker,and R.Parker.1995.Turnover mechanisms of the stable yeast PGK1 mRNA.Molecular&Cellular Biology 15(4):2145-2156.
- [0343] 11.Anderson,J.S.J.and R.P.Parker.1998.The 3'to 5'degradation of yeast mRNAs is a general mechanism for mRNA turnover that requires the SKJ2 DEVH box protein and 3'to 5'exonucleases of the exosome complex.EMBO Journal17(5): 1497-1506.

序列表

- <110> 康码(上海)生物科技有限公司
 <120> 一种核酸构建物及其调控蛋白合成的方法
 <130> P2018-1188
 <141> 2018-07-19
 <160> 61
 <170> SIPOSequenceListing 1.0
 <210> 1
 <211> 2043
 <212> DNA
 <213> 人工序列 (artificial sequence)
 <400> 1

```

cgcgaaatta atacgactca ctataggggt atttttaca caattacca caacaacaaa 60
caacaacaaa cattacaatt actatttaca attacaaaa aaaaaaatgt ctgaagacgc 120
caaaaacata aagaaaggcc cggcgccatt ctatcctcta gaggatggaa ccgctggaga 180
gcaactgcat aaggctatga agagatacgc cctggttcct ggaacaattg cttttacaga 240
tgcacatatc gaggtgaaca tcacgtacgc ggaatacttc gaaatgtccg ttcggttggc 300
agaagctatg aaacgatatg ggctgaatac aaatcacaga atcgtcgtat gcagtgaaaa 360
ctctcttcaa ttctttatgc cgggtgttggg cgcgttattt atcggagttg cagttgcgcc 420
cgcgaacgac atttataatg aacgtgaatt gctcaacagt atgaacattt cgcagcctac 480
cgtagtgttt gtttccaaaa aggggttgca aaaaattttg aacgtgcaa aaaaattacc 540
aataatccag aaaattatta tcatggattc taaaacggat taccagggat ttcagtcgat 600
gtacacgttc gtcacatctc atctacctc cggttttaat gaatacgatt ttgtaccaga 660
gtcctttgat cgtgacaaaa caattgcact gataatgaat tcctctggat ctactgggtt 720
acctaagggt gtggcccttc cgcatagaac tgctgcgctc agattctcgc atgccagaga 780
tcctattttt ggcaatcaaa tcattccgga tactgcgatt ttaagtgttg ttccattcca 840
tcacggtttt ggaatgttta ctacactcgg atatttgata tgtggatttc gagtcgtctt 900
aatgtataga tttgaagaag agctgttttt acgatccctt caggattaca aaattcaaag 960
tgcggttgcta gtaccaacc tattttcatt cttcgccaaa agcactctga ttgacaaata 1020
cgatttatct aatttacacg aaattgcttc tgggggcgca cctctttcga aagaagtcgg 1080
ggaagcgggt gcaaaacgct tccatcttc agggatacga caaggatatg ggctcactga 1140
gactacatca gctattctga ttacaccga ggggatgat aaaccgggcg cggctcggtaa 1200
agttgttcca ttttttgaag cgaaggttgt ggatctggat accgggaaaa cgctgggcgt 1260
taatcagaga ggcaattat gtgtcagagg acctatgatt atgtccggtt atgtaacaa 1320
tccggaagcg accaacgcct tgattgacaa ggatggatgg ctacattctg gagacatagc 1380
ttactgggac gaagacgaac acttcttcat agttgaccgc ttgaagtctt taattaaata 1440
caaaggatat caggtggccc ccgctgaatt ggaatcgata ttgttacaac accccaacat 1500
cttcgacgcg ggcggtggcag gtcttcccga cgatgacgcc ggtgaacttc ccgcccgct 1560

```

tgttgttttg gagcacggaa agacgatgac ggaaaaagag atcgtggatt acgtcgccag 1620
 tcaagtaaca accgcgaaaa agttgcgcgg aggagttgtg tttgtggacg aagtaccgaa 1680
 aggtcttacc ggaaaactcg acgcaagaaa aatcagagag atcctcataa aggccaagaa 1740
 gggcggaag tccaaattgg tttaaataag gattaattac ttggatgcca ataaaaaaaa 1800
 aaaagcgaca tagccaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 1860
 aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaactcga gcaccaccac 1920
 caccaccact gagatccggc tgctaacaaa gcccgaaagg aagctgagtt ggctgctgcc 1980
 accgctgagc aataactagc ataaccctt ggggcctcta aacgggtctt gaggggtttt 2040
 ttg 2043

<210> 2

<211> 2040

<212> DNA

<213> 人工序列 (artificial sequence)

<400> 2

cgcgaaatta atacgactca ctataggggt atttttaca caattaccaa caacaacaaa 60
 caacaacaaa cattacaatt actatttaca attacaaaa aaaaaaatgt ctgaagacgc 120
 caaaaacata aagaaaaggcc cggcgccatt ctatcctcta gaggatggaa ccgctggaga 180
 gcaactgcat aaggctatga agagatacgc cctggttctt ggaacaattg cttttacaga 240
 tgcacatatc gaggtgaaca tcacgtacgc ggaatacttc gaaatgtccg ttcggttggc 300
 agaagctatg aaacgatatg ggctgaatac aatcacaga atcgtcgtat gcagtgaaaa 360
 ctctcttcaa ttctttatgc cgggtgttggg cgcgttattt atcggagttg cagttgcgcc 420
 cgcgaacgac atttataatg aacgtgaatt gctcaacagt atgaacattt cgcagcctac 480
 cgtagtgttt gtttccaaaa aggggttgca aaaaattttg aacgtgcaaa aaaaattacc 540
 aataatccag aaaattatta tcatggattc taaaacggat taccaggat ttcagtcgat 600
 gtacacgttc gtcacatctc atctacctc cggttttaat gaatacgatt ttgtaccaga 660
 gtcctttgat cgtgacaaaa caattgcact gataatgaat tcctctggat ctactgggtt 720
 acctaagggt gtggcccttc cgcatagaac tgctgcgctc agattctcgc atgccagaga 780
 tcctattttt ggcaatcaaa tcattccgga tactgcgatt ttaagtgtt ttcattcca 840
 tcacggtttt ggaatgttta ctacactcgg atatttgata tgtggatttc gagtcgtctt 900
 aatgtataga tttgaagaag agctgttttt acgatccctt caggattaca aaattcaaag 960
 tgcgttgcta gtaccaacce tattttcatt cttcgccaaa agcactctga ttgacaaata 1020
 cgatttatct aatttacacg aaattgcttc tgggggcega cctctttcga aagaagtcgg 1080
 ggaagcgggt gcaaaacgct tccatcttc agggatacga caaggatatg ggctcactga 1140
 gactacatca gctattctga ttacaccgga ggggatgat aaaccgggcg cggctcggtaa 1200
 agttgttcca ttttttgaag cgaaggttgt ggatctggat accgggaaa cgctgggcgt 1260
 taatcagaga ggcaattat gtgtcagagg acctatgatt atgtccggtt atgtaacaa 1320
 tccggaagcg accaacgcct tgattgacaa ggatggatgg ctacattctg gagacatagc 1380
 ttactgggac gaagacgaac acttcttcat agttgaccgc ttgaagtctt taattaaata 1440
 caaaggatat caggtggccc ccgctgaatt ggaatcgata ttgttacaac accccaacat 1500

cttcgacgcg ggcgtggcag gtcttcccga cgatgacgcc ggtgaacttc ccgccgccgt 1560
 tgttgttttg gagcacggaa agacgatgac ggaaaaagag atcgtggatt acgtcgccag 1620
 tcaagtaaca accgcgaaaa agttgcgcgg aggagtgtg tttgtggacg aagtaccgaa 1680
 aggtcttacc ggaaaactcg acgcaagaaa aatcagagag atcctcataa aggccaagaa 1740
 gggcggaaaag tccaaattgg tttaaataag gattaattag gttgggacgc aggtgtcggg 1800
 gggatcgcca gccctcagga tagtgagctc ccgcagaggg ataagctatc tccctgcgac 1860
 cttggatgcc aataaaaaaaaa aaaaagcgac atagccaaaa aaaaaaaaaa aaaaaaaaaa 1920
 aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 1980
 aaaaaactcg agctagcata accccttggg gcctctaac ggttcttgag gggttttttg 2040
 <210> 3
 <211> 2132
 <212> DNA
 <213> 人工序列 (artificial sequence)
 <400> 3
 cgcgaaatta atacgactca ctataggggt atttttaca caattacca caacaacaaa 60
 caacaacaaa cattacaatt actatttaca attacaaaa aaaaaaatgt ctgaagacgc 120
 caaaaacata aagaaaaggcc cggcgccatt ctatcctcta gaggatggaa ccgctggaga 180
 gcaactgcat aaggctatga agagatacgc cctggttctt ggaacaattg cttttacaga 240
 tgcacatatc gaggtgaaca tcacgtacgc ggaatacttc gaaatgtccg ttcggttggc 300
 agaagctatg aaacgatatg ggctgaatac aatcacaga atcgtcgtat gcagtgaaaa 360
 ctctcttcaa ttctttatgc cgggtgttggg cgcgttattt atcggagttg cagttgcgcc 420
 cgcgaacgac atttataatg aacgtgaatt gctcaacagt atgaacattt cgcagcctac 480
 cgtagtgttt gtttccaaaa aggggttgca aaaaattttg aacgtgcaaa aaaaattacc 540
 aataatccag aaaattatta tcatggattc taaaacggat taccaggat ttcagtcgat 600
 gtacacgttc gtcacatctc atctacctc cggttttaat gaatacgatt ttgtaccaga 660
 gtcctttgat cgtgacaaaa caattgcact gataatgaat tcctctggat ctactgggtt 720
 acctaagggt gtggcccttc cgcatagaac tgctgcgctc agattctcgc atgccagaga 780
 tcctattttt ggcaatcaaa tcattccgga tactgcgatt ttaagtgtt ttccattcca 840
 tcacggtttt ggaatgttta ctacactcgg atatttgata tgtggatttc gagtcgtctt 900
 aatgtataga tttgaagaag agctgttttt acgatccctt caggattaca aaattcaag 960
 tgcgttgcta gtaccaacc tattttcatt cttcgccaaa agcactctga ttgacaaata 1020
 cgatttatct aatttacacg aaattgcttc tgggggcega cctctttcga aagaagtcgg 1080
 ggaagcgggt gcaaaacgct tccatcttc agggatacga caaggatatg ggctcactga 1140
 gactacatca gctattctga ttacaccgga ggggatgat aaaccgggcg cggctcggtaa 1200
 agttgttcca ttttttgaag cgaaggttgt ggatctggat accgggaaaa cgctgggcgt 1260
 taatcagaga ggcaattat gtgtcagagg acctatgatt atgtccggtt atgtaacaaa 1320
 tccggaagcg accaacgcct tgattgacaa ggatggatgg ctacattctg gagacatagc 1380
 ttactgggac gaagacgaac acttcttcat agttgaccgc ttgaagtctt taattaaata 1440
 caaaggatat caggtggccc ccgctgaatt ggaatcgata ttgttacaac accccaacat 1500

cttcgacgcg ggcgtggcag gtcttcccga cgatgacgcc ggtgaacttc ccgccgccgt 1560
tgttgttttg gagcacgga agacgatgac ggaaaaagag atcgtggatt acgtcgccag 1620
tcaagtaaca accgcgaaaa agttgcgcgg aggagtgtg tttgtggacg aagtaccgaa 1680
aggtcttacc ggaaaactcg acgcaagaaa aatcagagag atcctcataa aggccaagaa 1740
gggcggaaaag tccaaattgg tttaaataag gattaattat cttgttgggt tctggatggt 1800
aggatgacga gtcgactcgg gctccgact aggtttggtc gcctagagga tggagatag 1860
gaaagggctc cgtgtggat cagtcggtc aaagacgcgc ttccaacatg ggcctatggt 1920
cggataagtc ttagcaatac cagccagcat gacttggat ccaataaaaa aaaaaaagcg 1980
acatagccaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 2040
aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaact cgagctagca taacccttg 2100
gggcctctaa acgggtcttg aggggttttt tg 2132

<210> 4

<211> 2087

<212> DNA

<213> 人工序列 (artificial sequence)

<400> 4

cgcgaaatta atacgactca ctataggggt atttttacaa caattaccaa caacaacaaa 60
caacaacaaa cattacaatt actatttaca attacaaaa aaaaaaatgt ctgaagacgc 120
caaaaacata aagaaaggcc cggcgccatt ctatcctcta gaggatggaa ccgctggaga 180
gcaactgcat aaggctatga agagatacgc cctggttcct ggaacaattg cttttacaga 240
tgcacatatc gaggtgaaca tcacgtacgc ggaatacttc gaaatgtccg ttcggttggc 300
agaagctatg aaacgatatg ggctgaatac aatcacaga atcgtcgtat gcagtgaaaa 360
ctctcttcaa ttctttatgc cgggtttggg cgcgttattt atcggagttg cagttgcgcc 420
cgcgaacgac atttataatg aacgtgaatt gctcaacagt atgaacattt cgcagcctac 480
cgtagtgttt gtttccaaaa aggggttgca aaaaattttg aacgtgcaa aaaaattacc 540
aataatccag aaaattatta tcatggattc taaaacggat taccaggat ttcagtcgat 600
gtacacgttc gtcacatctc atctacctc cggttttaat gaatacgatt ttgtaccaga 660
gtcctttgat cgtgacaaaa caattgcaat gataatgaat tcctctggat ctactgggtt 720
acctaagggt gtggccctc cgcatagaac tgctgcgctc agattctcgc atgccagaga 780
tcctattttt ggcaatcaaa tcattccgga tactgcgatt ttaagtgtt ttccattcca 840
tcacggtttt ggaatgttta ctacactcgg atatttgata tgtggatttc gagtcgtctt 900
aatgtataga tttgaagaag agctgttttt acgatecctt caggattaca aaattcaaag 960
tgcgttgcta gtaccaacc tattttcatt cttegccaaa agcactctga ttgacaaata 1020
cgatttatct aatttacacg aaattgcttc tgggggcgca cctctttcga aagaagtcgg 1080
ggaagcgggt gcaaaacgct tccatcttc aggatacga caaggatat ggctcactga 1140
gactacatca gctattctga ttacaccgga ggggatgat aaaccgggcg cggctcgtaa 1200
agttgttcca ttttttgaag cgaaggttgt ggatctggat accgggaaa cgctgggcgt 1260
taatcagaga ggcaattat gtgtcagagg acctatgatt atgtccggtt atgtaaacia 1320
tccggaagcg accaacgcct tgattgacaa ggatggatgg ctacattctg gagacatagc 1380

ttactgggac gaagacgaac acttcttcat agttgaccgc ttgaagtctt taattaaata 1440
 caaaggatat caggtggccc ccgctgaatt ggaatcgata ttgttacaac accccaacat 1500
 cttcgacgcg ggcgtggcag gtcttcccga cgatgacgcc ggtgaacttc ccgccgccgt 1560
 tgtttgtttg gagcacggaa agacgatgac ggaaaaagag atcgtggatt acgtcgccag 1620
 tcaagtaaca accgcgaaaa agttgcgcgg aggagtgtg tttgtggacg aagtaccgaa 1680
 aggtcttacc ggaaaactcg acgcaagaaa aatcagagag atcctcataa aggccaagaa 1740
 gggcggaaaag tccaaattgg tttaaataag gattaattaa ttgtatgggc ctccgtgctg 1800
 ttaggggcag tgaaaacgtc ggtctagcca gccgtcccct gggtagtgtg ctccgtctaa 1860
 gtacaccact actcggtttc cacaaacgat cggactccca aatagtctt ggatgccaat 1920
 aaaaaaaaaa aagcgacata gccaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 1980
 aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaactcgagc 2040
 tagcataacc ccttggggcc tctaaacggg tcttgagggg ttttttg 2087

<210> 5

<211> 2019

<212> DNA

<213> 人工序列 (artificial sequence)

<400> 5

cgcgaaatta atacgactca ctataggggt atttttacaa caattaccaa caacaacaaa 60
 caacaacaaa cattacaatt actatttaca attacaaaa aaaaaaatgt ctgaagacgc 120
 caaaaacata aagaaaggcc cggcgccatt ctatcctcta gaggatggaa ccgctggaga 180
 gcaactgcat aaggctatga agagatacgc cctggttctt ggaacaattg cttttacaga 240
 tgcacatatc gaggtgaaca tcacgtacgc ggaatacttc gaaatgtccg ttcggttggc 300
 agaagctatg aaacgatatg ggctgaatac aatcacaga atcgtcgtat gcagtgaaaa 360
 ctctcttcaa ttctttatgc cgggtgttggg cgcgttattt atcggagttg cagttgcgcc 420
 cgcgaacgac atttataatg aacgtgaatt gctcaacagt atgaacattt cgcagcctac 480
 cgtagtgttt gtttccaaaa aggggttgca aaaaattttg aacgtgcaaa aaaaattacc 540
 aataatccag aaaattatta tcatggattc taaaacggat taccaggat ttcagtcat 600
 gtacacgttc gtcacatctc atctacctc cggttttaat gaatacgatt ttgtaccaga 660
 gtcctttgat cgtgacaaaa caattgcaat gataatgaat tcctctggat ctactgggtt 720
 acctaaagggt gtggcccttc cgcatagaac tgctgcgctc agattctcgc atgccagaga 780
 tcctattttt ggcaatcaaa tcattccgga tactgcgatt ttaagtgttg ttccattcca 840
 tcacggtttt ggaatgttta ctacactcgg atatttgata tgtggatttc gagtcgtctt 900
 aatgtataga tttgaagaag agctgttttt acgatecctt caggattaca aaattcaaa 960
 tgcgttgcta gtaccaacce tattttcatt cttcgccaaa agcactctga ttgacaaata 1020
 cgatttatct aatttacacg aaattgcttc tgggggcgca cctctttcga aagaagtcgg 1080
 ggaagcgggt gcaaaaacgt tccatcttcc aggatacga caaggatat ggctcactga 1140
 gactacatca gctattctga ttacaccgga ggggatgat aaaccgggcg cggtcggtaa 1200
 agttgttcca ttttttgaag cgaaggttgt ggatctggat accgggaaaa cgctgggcgt 1260
 taatcagaga ggcaattat gtgtcagagg acctatgatt atgtccggtt atgtaacaaa 1320

tccggaagcg accaacgcct tgattgacaa ggatggatgg ctacattctg gagacatagc 1380
 ttactgggac gaagacgaac acttcttcat agttgaccgc ttgaagtctt taattaaata 1440
 caaaggatat caggtggccc ccgctgaatt ggaatcgata ttgttacaac accccaacat 1500
 cttcgacgcg ggcgtggcag gtcttcccga cgatgacgcc ggtgaacttc ccgccgccgt 1560
 tgttgttttg gagcacggaa agacgatgac ggaaaaagag atcgtggatt acgtcgccag 1620
 tcaagtaaca accgcgaaaa agttgcgcgg aggagttgtg tttgtggacg aagtaccgaa 1680
 aggtcttacc ggaaaaactcg acgcaagaaa aatcagagag atcctcataa aggccaagaa 1740
 gggcggaaaag tccaaattgg tttaaataag gattaattac gtagctaccc tcctgggtgca 1800
 cttcctgggtg caaagcagaa ccaagggtac ggtggtacgc ttggatgcca ataaaaaaaa 1860
 aaaagcgaca tagccaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 1920
 aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaactcga gctagcataa 1980
 ccccttgggg cctctaaacg ggtcttgagg ggttttttg 2019

<210> 6

<211> 2190

<212> DNA

<213> 人工序列 (artificial sequence)

<400> 6

cgcgaaatta atacgactca ctataggaa aaatcgtta aaaaaacttt tcttctgggt 60
 atttatagag atcagtatga actgtgtcgc taggatcaag cgggtggttca cacctgactt 120
 caccctggc gagggcgtga agtctcaaag catacctaata acaatataa tcccataggt 180
 atttttacaa caattacaa caacaacaaa caacaacaa cattacaatt actatttaca 240
 attacaaaaa aaaaaaatgt ctgaagacgc caaaaacata aagaaaggcc cggcgccatt 300
 ctatcctcta gaggatggaa ccgctggaga gcaactgcat aaggctatga agagatacgc 360
 cctggttcct ggaacaattg cttttacaga tgacatatc gaggtgaaca tcacgtacgc 420
 ggaatacttc gaaatgtccg ttcggttggc agaagctatg aaacgatatg ggctgaatac 480
 aaatcacaga atcgtcgtat gcagtgaaaa ctctcttcaa ttctttatgc cgggtgttggg 540
 cgcgttatatt atcggagttg cagttgcgcc cgcgaaacgc atttataatg aacgtgaatt 600
 gctcaacagt atgaacattt cgcagcctac cgtagtgttt gtttccaaaa aggggttgc 660
 aaaaattttg aacgtgcaaa aaaaattacc aataatccag aaaattatta tcatggattc 720
 taaaacggat taccaggat ttcagtcgat gtacacgttc gtcacatctc atctacctcc 780
 cggttttaat gaatacgatt ttgtaccaga gtctttgat cgtgacaaaa caattgcact 840
 gataatgaat tcctctggat ctactgggtt acctaagggt gtggcccttc cgcatagaac 900
 tgcctgcgtc agattctcgc atgccagaga tectattttt ggcaatcaaa tcattccgga 960
 tactgcgatt ttaagtgtt ttccattcca tcacggtttt ggaatgttta ctacactcgg 1020
 atatttgata tgtggatttc gagtcgtctt aatgtataga tttgaagaag agctgttttt 1080
 acgateccct caggattaca aaattcaaag tgcgttgcta gtaccaacc tattttcatt 1140
 cttcgccaaa agcactctga ttgacaaata cgatttatct aatttacacg aaattgcttc 1200
 tgggggcgca cctctttcga aagaagtcgg ggaagcgggt gcaaaacgct tccatcttcc 1260
 agggatacga caaggatatg ggctcactga gactacatca gctattctga ttacaccgga 1320

ggggatgat aaaccgggcg cgtcggtaa agttgttcca tttttgaag cgaaggttgt 1380
 ggatctggat accgggaaaa cgctgggctg taatcagaga ggcgaattat gtgtcagagg 1440
 acctatgatt atgtccggtt atgtaaaca tccggaagcg accaacgcct tgattgacaa 1500
 ggatggatgg ctacattctg gagacatagc ttactgggac gaagacgaac acttcttcat 1560
 agttgaccgc ttgaagtctt taattaaata caaaggatat caggtggccc ccgctgaatt 1620
 ggaatcgata ttgttacaac accccaacat cttcgacgcg ggcgtggcag gtcttcccga 1680
 cgatgacgcc ggtgaacttc ccgccgccgt tgttgttttg gagcacggaa agacgatgac 1740
 ggaaaaagag atcgtggatt acgtcgccag tcaagtaaca accgcgaaaa agttgcgagg 1800
 aggagtgtg tttgtggacg aagtaccgaa aggtcttacc ggaaaactcg acgcaagaaa 1860
 aatcagagag atcctcataa aggccaagaa gggcgaaag tccaaattgg tttaaataag 1920
 gattaattag gttgggacgc aggtgtcgga gggatcgcca gccctcagga tagtgagctc 1980
 ccgcagaggg ataagctatc tcctgacgac cttggatgcc aataaaaaaa aaaaagcgac 2040
 atagccaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 2100
 aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaactcg agctagcata accccttggg 2160
 gcctctaac ggttcttgag gggttttttg 2190

<210> 7

<211> 2271

<212> DNA

<213> 人工序列 (artificial sequence)

<400> 7

cgcgaaatta atacgactca ctataggaa aaatcggta aaaaaacttt tcttctagaa 60
 attctccagg atttctcgac ctagtctggt tatctggtga cttgcgctac cgttgctttg 120
 cgtagagaat ttctcaaagc atacctaata acaatataat cccataggta tttttacaac 180
 aattaccaac aacaacaaac aacaacaaac attacaatta ctatttaca ttacaaaaaa 240
 aaaaaatgtc tgaagacgcc aaaaacataa agaaaggccc ggcgccattc tctctctag 300
 aggatggaac cgctggagag caactgcata aggctatgaa gagatacgcc ctggttcctg 360
 gaacaattgc ttttacagat gcacatatcg aggtgaacat cacgtacgcg gaatacttgc 420
 aaatgtccgt tcggttgga gaagctatga aacgatatgg gctgaataca aatcacagaa 480
 tcgtcgtatg cagtgaaaac tctcttcaat tctttatgcc ggtgttgggc gcgttattta 540
 tcggagttgc agttgcgcc gcgaacgaca ttataatga acgtgaattg ctcaacagta 600
 tgaacatttc gcagcctacc gtagtgtttg tttccaaaa ggggttgcaa aaaattttga 660
 acgtgcaaaa aaaattacca ataatccaga aaattattat catggattct aaaacggatt 720
 accagggatt tcagtcgatg tacacgttcg tcacatctca tctacctcc ggttttaatg 780
 aatacgattt tgtaccagag tcctttgatc gtgacaaaac aattgcactg ataatagaatt 840
 cctctggatc tactgggtta cctaagggtg tggeccttc gcatagaact gcctgcgtca 900
 gattctcgca tgccagagat cctatTTTTG gcaatcaaat cattccgat actgcgattt 960
 taagtgttgt tccattccat cacggttttg gaatgtttac tacactcgga tatttgatat 1020
 gtggatttcg agtcgtctta atgtatagat ttgaagaaga gctgttttta cgatcccttc 1080
 aggattacaa aattcaaagt gcgttgctag taccaacct attttcatc ttcgcaaaa 1140

gcactctgat tgacaaatac gatttatcta atttacacga aattgcttct gggggcgcac 1200
 ctcttttcgaa agaagtcggg gaagcggttg caaacgctt ccatcttcca gggatacgac 1260
 aaggatatgg gctcactgag actacatcag ctattctgat tacacccgag ggggatgata 1320
 aaccggggcgc ggtcggtaaa gttgttccat ttttgaagc gaaggttgtg gatctggata 1380
 ccgggaaaac gctgggcggt aatcagagag gcgaattatg tgtcagagga cctatgatta 1440
 tgtccggtta tgtaaacat ccggaagcga ccaacgcctt gattgacaag gatggatggc 1500
 tacattctgg agacatagct tactgggacg aagacgaaca cttcttcata gttgaccgct 1560
 tgaagtcttt aattaaatac aaaggatatac aggtggcccc cgctgaattg gaatcgatat 1620
 tgttacaaca cccaacatc ttcgacgagg gcgtggcagg tcttcccgac gatgacgccg 1680
 gtgaacttcc cgccgcccgtt gttgttttgg agcacggaaa gacgatgacg gaaaaagaga 1740
 tcgtggatta cgtcgccagt caagtaacaa ccgcgaaaaa gttgcgcgga ggagttgtgt 1800
 ttgtggacga agtaccgaaa ggtcttaccg gaaaactcga cgcaagaaaa atcagagaga 1860
 tcctcataaa ggccaagaag ggcggaaagt ccaattggt ttaaataagg attaattatc 1920
 ttgttgggtt ctggatgta ggatgacgag tcgactcggg ctccgcacta ggtttggtcg 1980
 cctagaggat ggagatatgg aaagggtctc gtgtggtatc agtcggtcga aagacgcgct 2040
 tccaacatgg gcctatggtc ggataagtct tagcaatacc agccagcatg acttgatgc 2100
 caataaaaaa aaaaaagcga catagccaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 2160
 aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaactc 2220
 gagctagcat aacccttgg ggcctctaaa cgggtcttga ggggtttttt g 2271

<210> 8

<211> 2168

<212> DNA

<213> 人工序列 (artificial sequence)

<400> 8

cgcgaaatta atacgactca ctataggaa aaatcggta aaaaaacttt tcttctggaa 60
 accgtggcaa gtttcccaa gcatacctaa taacaatata atcccatagg tatttttaca 120
 acaattacca acaacaacaa acaacaaca acattacaat tactatttac aattacaaaa 180
 aaaaaaatg tctgaagacg ccaaaaacat aaagaaaggc ccggcgccat tctatcctct 240
 agaggatgga accgctggag agcaactgca taaggctatg aagagatacg ccctggttcc 300
 tggaacaatt gcttttacag atgcacatat cgaggatgac atcacgtacg cggaataactt 360
 cgaaatgtcc gttcggttgg cagaagctat gaaacgatat gggtgaata caaatcacag 420
 aatcgtcgta tgcagtgaaa actctcttca attctttatg ccggtgttgg gcgcgttatt 480
 tatcggagtt gcagttgcgc ccgcgaacga catttataat gaacgtgaat tgctcaacag 540
 tatgaacatt tcgcagccta ccgtagtggt tgtttccaaa aagggttgc aaaaaatttt 600
 gaacgtgcaa aaaaaattac caataatcca gaaaattatt atcatggatt ctaaacgga 660
 ttaccaggga tttcagtcga tgtacacggt cgteacatct catctacctc ccggttttaa 720
 tgaatacgat tttgtaccag agtccttga tcgtgacaaa acaattgcac tgataatgaa 780
 ttctctgga tctactgggt tacctaaggg tgtggcctt ccgcatagaa ctgcctgcgt 840
 cagattctcg catgccagag atcctatfff tggcaatcaa atcattccgg atactgcgat 900

tttaagtgtt gttccattcc atcacggttt tggaatgttt actacactcg gatatttgat 960
 atgtggattt cgagtcgtct taatgtatag atttgaagaa gagctgtttt tacgatccct 1020
 tcaggattac aaaattcaaa gtgcgttgct agtaccaacc ctattttcat tcttcgcaa 1080
 aagcactctg attgacaaat acgatttatc taatttacac gaaattgctt ctgggggctc 1140
 acctctttcg aaagaagtcg gggaagcggg tgcaaaacgc ttccatcttc cagggatacg 1200
 acaaggatat gggctcactg agactacatc agctattctg attacacccg agggggatga 1260
 taaaccgggc gcggtcggta aagttgttcc attttttgaa gcgaaggttg tggatctgga 1320
 taccgggaaa acgctgggct ttaatcagag aggcaatta tgtgtcagag gacctatgat 1380
 tatgtccggg tatgtaaaca atccggaagc gaccaacgc ttgattgaca aggatggatg 1440
 gctacattct ggagacatag cttactggga cgaagacgaa cacttcttca tagttgaccg 1500
 cttgaagtct ttaattaaat acaaaggata tcaggtggcc cccgctgaat tggaatcgat 1560
 attgttaca caccccaaca tcttcgacgc gggcgtggca ggtcttcccg acgatgacgc 1620
 cgggtgaact cccgccgccc ttgttgtttt ggagcacgga aagacgatga cggaaaaaga 1680
 gatcgtggat tacgtcgcca gtcaagtaac aaccgcgaaa aagttgcgcg gaggagtgt 1740
 gtttgtggac gaagtaccga aaggtcttac cggaaaactc gacgcaagaa aatcagaga 1800
 gatcctcata aaggccaaga agggcggaaa gtccaattg gtttaataa ggattaatta 1860
 attgtatggg cctccgtgct gttaggggca gtggaaacgt cggcttagcc agccgtcccc 1920
 tgggtagtgt gctccgtcta agtacaccac tactcgtttt ccacaaacga tcggactccc 1980
 aaaatagtct tggatgcaa taaaaaaaa aaagcgacat agccaaaaa aaaaaaaaa 2040
 aaaaaaaaa aaaaaaaaa aaaaaaaaa aaaaaaaaa aaaaaaaaa aaaaaaaaa 2100
 aaaaaaaaa aaactcgag ctagcataac cccttggggc ctctaacgg gtcttgagg 2160
 gttttttg 2168

<210> 9

<211> 2102

<212> DNA

<213> 人工序列 (artificial sequence)

<400> 9

cgcgaaatta atacgactca ctataggaa aaatcggta aaaaaacttt tcttctagta 60
 aagacaggaa actttactca aagcatacct aataacaata taatcccata ggtattttta 120
 caacaattac caacaacaac aaacaacaaa caacattaca attactattt acaattacaa 180
 aaaaaaaaaa tgtctgaaga cgccaaaaac ataaagaaag gcccgcgcc attctatcct 240
 ctagaggatg gaaccgctgg agagcaactg cataaggcta tgaagagata cgccctgggt 300
 cctggaacaa ttgcttttac agatgcacat atcgaggatg acatcacgta cgcggaatac 360
 ttcgaaatgt ccgttcgggt ggcagaagct atgaaacgat atgggctgaa tacaatcac 420
 agaatcgtcg tatgcagtga aaactctctt caattcttta tgccggtgtt gggcgcgtta 480
 tttatcggag ttgcagttgc gcccgcgaac gacatttata atgaacgtga attgctcaac 540
 agtatgaaca tttcgcagcc taccgtagtg tttgtttcca aaaaggggtt gcaaaaaatt 600
 ttgaacgtgc aaaaaaaaa attccaataatc cagaaaatta ttatcatgga ttctaaaacg 660
 gattaccagg gatttcagtc gatgtacacg ttcgtccat ctcatctacc tcccgtttt 720

aatgaatacg attttgtacc agagtccttt gatcgtgaca aaacaattgc actgataatg 780
aattcctctg gatctactgg gttacctaag ggtgtggccc ttccgcatag aactgcctgc 840
gtcagattct cgcatgccag agatcctatt tttggcaatc aatcattcc ggatactgcg 900
attttaagtg ttgttccatt ccatcacggt tttggaatgt ttactacact cggatatttg 960
atatgtggat ttcgagtcgt cttaatgtat agatttgaag aagagctggt tttacgatcc 1020
cttcaggatt acaaaattca aagtgcgttg ctagtaccaa ccctattttc attcttcgcc 1080
aaaagcactc tgattgacaa atacgattta tctaatttac acgaaattgc ttctgggggc 1140
gcacctcttt cgaaagaagt cggggaagcg gttgcaaac gcttccatct tccagggata 1200
cgacaaggat atgggctcac tgagactaca tcagctattc tgattacacc cgagggggat 1260
gataaaccgg gcgcggtcgg taaagttggt ccattttttg aagcgaaggt tgtggatctg 1320
gataccggga aaacgctggg cgттаатсag agaggcgaat tatgtgtcag aggacctatg 1380
attatgtccg gttatgtaaa caatccggaa gcgaccaacg ccttgattga caaggatgga 1440
tggtacatt ctggagacat agcttactgg gacgaagacg aacacttctt catagttgac 1500
cgcttgaagt ctttaattaa atacaaagga tatcaggtgg cccccgctga attggaatcg 1560
atattgttac aacaccccaa catcttcgac gcgggcgtgg caggtcttcc cgacgatgac 1620
gccggtgaac ttcccgccgc cgttgttgtt ttggagcacg gaaagacgat gacggaaaaa 1680
gagatcgtgg attacgtcgc cagtcaagta acaaccgca aaaagttgcg cggaggagtt 1740
gtgtttgtgg acgaagtacc gaaaggtctt accgaaaaac tcgacgcaag aaaaatcaga 1800
gagatcctca taaaggccaa gaaggcgga aagtccaaat tggtttaaat aaggattaat 1860
tacgtagcta ccctcctggt gcacttctg gtgcaaagca gaaccaaggg tacggtggta 1920
cgcttggatg ccaataaaaa aaaaaaagcg acatagccaa aaaaaaaaaa aaaaaaaaaa 1980
aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 2040
aaaaaaaaact cgagctagca taacccttg gggcctctaa acgggtcttg aggggttttt 2100
tg 2102
<210> 10
<211> 2190
<212> DNA
<213> 人工序列(artificial sequence)
<400> 10
cgcgaaatta atacgactca ctataggggt atttttacaa caattaccaa caacaacaaa 60
caacaaacaa cattacaatt actatttaca attacagaaa aatcggttaa aaaaactttt 120
cttctgggta tttatagaga tcagtatgaa ctgtgtcgtc aggatcaagc ggtggttcac 180
acctgacttc acccctggcg agggcgtgaa gtctcaaagc atacctaata acaatataat 240
cccataaaaa aaaaaaatgt ctgaagacgc caaaaacata aagaaaggcc cggcgccatt 300
ctatcctcta gaggatggaa ccgctggaga gcaactgcat aaggctatga agagatacgc 360
cctggttcct ggaacaattg cttttacaga tgcacatata gaggtgaaca tcacgtacgc 420
ggaatacttc gaaatgtccg ttcggttggc agaagctatg aaacgatatg ggctgaatac 480
aaatcacaga atcgtcgtat gcagtgaaaa ctctcttcaa ttctttatgc cgggtttggg 540
cgcgttattt atcggagttg cagttgcgcc cgcgaaacgc atttataatg aacgtgaatt 600

gctcaacagt atgaacattt cgcagcctac cgtagtgttt gtttccaaaa aggggttgca 660
 aaaaattttg aacgtgcaaa aaaaattacc aataatccag aaaattatta tcatggattc 720
 taaaacggat taccagggat ttcagtcgat gtacacgttc gtcacatctc atctacctcc 780
 cggttttaat gaatacgatt ttgtaccaga gtcctttgat cgtgacaaaa caattgcact 840
 gataatgaat tcctctggat ctactgggtt acctaagggt gtggcccttc cgcatagaac 900
 tgcctgcgtc agattctcgc atgccagaga tcctatTTTT ggcaatcaaa tcattccgga 960
 tactgcgatt ttaagtgtt ttccattcca tcacggtttt ggaatgttta ctacactcgg 1020
 atatttgata tgtggatttc gagtcgtctt aatgtataga tttgaagaag agctgttttt 1080
 acgatccctt caggattaca aaattcaaag tgcgttgcta gtaccaacc tttttcatt 1140
 cttcgccaaa agcactctga ttgacaaata cgatttatct aatttacacg aaattgcttc 1200
 tggggcgca cctctttcga aagaagtcgg ggaagcggtt gcaaaacgct tccatcttc 1260
 agggatacga caaggatatg ggctcactga gactacatca gctattctga ttacaccgga 1320
 ggggatgat aaaccgggcg cggtcggtaa agttgttcca ttttttgaag cgaagttgt 1380
 ggatctggat accgggaaaa cgctgggctg taatcagaga ggcgaattat gtgtcagagg 1440
 acctatgatt atgtccggtt atgtaaaca tccggaagcg accaacgctt tgattgaca 1500
 ggatggatgg ctacattctg gagacatagc ttactgggac gaagacgaac acttcttcat 1560
 agttgaccgc ttgaagtctt taattaaata caaaggatat caggtggccc ccgctgaatt 1620
 ggaatcgata ttgttacaac accccaacat cttcgacgcg ggcgtggcag gtcttcccga 1680
 cgatgacgcc ggtgaacttc ccgccccgt tgttgttttg gagcacgga agacgatgac 1740
 ggaaaaagag atcgtggatt acgtcgccag tcaagtaaca accgcgaaaa agttgcgcgg 1800
 aggagtgtg tttgtggacg aagtaccgaa aggtcttacc ggaaaactcg acgcaagaaa 1860
 aatcagagag atcctcataa aggccaagaa ggcggaag tccaattgg tttaaataag 1920
 gattaattag gttgggacgc aggtgtcgga gggatcgcca gccctcagga tagtgagctc 1980
 ccgcagaggg ataagctatc tcctgacgac cttggatgcc aataaaaaaa aaaaagcgac 2040
 atagccaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 2100
 aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaactcg agctagcata accccttggg 2160
 gcctctaaac ggttcttgag ggttttttg 2190

<210> 11

<211> 2271

<212> DNA

<213> 人工序列 (artificial sequence)

<400> 11

cgcgaaatta atacgactca ctataggggt atttttaca caattacca caacaacaaa 60
 caacaacaaa cattacaatt actatttaca attacagaaa aatcggttaa aaaaactttt 120
 cttctagaaa ttctccagga tttctcgacc tagttcgttt atctggtgac ttgcgctacc 180
 gttgctttgc gtagagaatt tetcaaagca tacctaataa caatataatc ccataaaaa 240
 aaaaaatgtc tgaagacgcc aaaaacataa agaaaggccc ggcgccattc tctctctag 300
 aggatggaac cgctggagag caactgcata aggctatgaa gagatacgcc ctggttctctg 360
 gaacaattgc ttttacagat gcacatatcg aggtgaacat cacgtacgcg gaatacttcg 420

aaatgtccgt tccggttgca gaagctatga aacgatatgg gctgaataca aatcacagaa 480
 tcgtcgtatg cagtgaaaac tctcttcaat tctttatgcc ggtggtgggc gcgttattta 540
 tcggagttgc agttgcgccc gcgaacgaca ttataatga acgtgaattg ctcaacagta 600
 tgaacatttc gcagcctacc gtagtgtttg tttccaaaaa ggggttgcaa aaaattttga 660
 acgtgcaaaa aaaattacca ataatccaga aaattattat catggattct aaaacggatt 720
 accagggatt tcagtcgatg tacacgttcg tcacatctca tctacctccc ggttttaatg 780
 aatacgattt tgtaccagag tcctttgatc gtgacaaaac aattgcactg ataatgaatt 840
 cctctggatc tactgggta cctaagggtg tggcccttc gcatagaact gcctgcgtca 900
 gattctcgca tgccagagat cctatTTTTG gcaatcaaat cattccggat actgcgattt 960
 taagtgttgt tccattccat cacggTTTTG gaatgtttac tacactcgga tatttgatat 1020
 gtggatttcg agtcgtctta atgtatagat ttgaagaaga gctgtTTTTa cgatcccttc 1080
 aggattacaa aattcaaagt gcgttgctag taccaacct attttcatc ttcgcaaaa 1140
 gcaactctgat tgacaaatac gatttatcta atttacacga aattgcttct gggggcgcac 1200
 ctctttcgaa agaagtcggg gaagcggttg caaacgctt ccatcttcca gggatacgac 1260
 aaggatatgg gctcactgag actacatcag ctattctgat tacacccgag ggggatgata 1320
 aaccgggccc ggtcggtaaa gttgttccat tttttgaagc gaaggttgtg gatctggata 1380
 ccgggaaaac gctgggctt aatcagagag gcgaattatg tgtcagagga cctatgatta 1440
 tgtccggta tgtaaacaat ccggaagcga ccaacgctt gattgacaag gatggatggc 1500
 tacattctgg agacatagct tactgggacg aagacgaaca cttcttcata gttgaccgct 1560
 tgaagtcttt aattaaatac aaaggatatac aggtggcccc cgctgaattg gaatcgatat 1620
 tgttacaaca cccaacatc ttcgacgcgg gcgtggcagg tcttcccgac gatgacgccg 1680
 gtgaacttcc cgccgccgtt gttgttttg agcacggaaa gacgatgacg gaaaaagaga 1740
 tcgtggatta cgtcgcagc caagtaaca ccgcgaaaaa gttgcgcgga ggagttgtgt 1800
 ttgtggacga agtaccgaaa ggtcttaccg gaaaactcga cgcaagaaaa atcagagaga 1860
 tcctcataaa ggccaagaag ggcggaagt ccaattggt ttaaataagg attaattatc 1920
 ttgttgggtt ctggatgta ggatgacgag tcgactcggg ctccgcacta gttttggtcg 1980
 cctagaggat ggagatatgg aaagggtctc gtgtggtatc agtcggtcga aagacgcgct 2040
 tccaacatgg gcctatggtc ggataagtct tagcaatacc agccagcatg acttgatgc 2100
 caataaaaaa aaaaaagcga catagccaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 2160
 aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaactc 2220
 gagctagcat aacccttgg ggcctctaaa cgggtcttga ggggtttttt g 2271

<210> 12

<211> 2168

<212> DNA

<213> 人工序列 (artificial sequence)

<400> 12

cgcgaaatta atacgactca ctataggggt atttttacaa caattaccaa caacaacaaa 60
 caacaacaa cattacaatt actatttaca attacagaaa aatcggttaa aaaaactttt 120
 cttctggaaa ccgtggcaag tttcccaaag catacctaata acaatataa tcccataaaa 180

aaaaaaatg tctgaagacg ccaaaaacat aaagaaaggc cggcgccat tctatcctct 240
 agaggatgga accgctggag agcaactgca taaggctatg aagagatacg ccctggttcc 300
 tggaacaatt gcttttacag atgcacatat cgaggtgaac atcacgtacg cggaatactt 360
 cgaaatgtcc gttcggttgg cagaagctat gaaacgatat gggctgaata caaatcacag 420
 aatcgtcgta tgcagtgaaa actctcttca attctttatg ccggtggttg gcgcttatt 480
 tatcggagtt gcagttgcgc ccgcaacga cattataat gaacgtgaat tgctcaacag 540
 tatgaacatt tgcagccta ccgtagtggt tgtttccaaa aagggttgc aaaaaatttt 600
 gaacgtgcaa aaaaaattac caataatcca gaaaattatt atcatggatt ctaaacgga 660
 ttaccaggga tttcagtcga tgtacacggt cgteacatct catctacctc ccggttttaa 720
 tgaatacgat tttgtaccag agtccttga tcgtgacaaa acaattgcac tgataatgaa 780
 ttctctgga tctactgggt tacctaaggg tgtggcctt ccgcatagaa ctgctgcgt 840
 cagattctcg catgccagag atcctatfff tggcaatcaa atcattccgg atactgcgat 900
 tttaaagtgt gttccattcc atcacggttt tggaatgttt actacactcg gatatttgat 960
 atgtggattt cgagtcgtct taatgtatag atttgaagaa gagctgtttt tacgatccct 1020
 tcaggattac aaaattcaaa gtgcgttgct agtaccaccc ctatfffcat tcttcgcaa 1080
 aagcactctg attgacaaa acgatttatc taatttacac gaaattgctt ctgggggcgc 1140
 acctctttcg aaagaagtcg gggaagcggg tgcaaacgc ttccatctc cagggatacg 1200
 acaaggatat gggctcactg agactacatc agctattctg attacaccg aggggatga 1260
 taaaccgggc gcgctcgga aagttgttcc atttttgaa gcaaggttg tggatctgga 1320
 taccgggaaa acgctgggcg ttaatcagag aggcaatta tgtgtcagag gacctatgat 1380
 tatgtccggt tatgtaaca atccggaagc gaccaacgc ttgattgaca aggatggatg 1440
 gctacattct ggagacatag ctactggga cgaagacgaa cacttcttca tagttgaccg 1500
 cttgaagtct ttaattaaat acaaaggata tcaggtggcc cccgctgaat tggaatcgat 1560
 attgttacia cacccaaca tcttcgacgc gggcgtggca ggtcttccc acgatgacgc 1620
 cggatgaact cccgccgcg ttgttgtttt ggagcacgga aagacgatga cggaaaaaga 1680
 gatcgtggat tacgtcgcca gtcaagtaac aaccgcgaaa aagttgcgcg gaggagtgtg 1740
 gtttgtggac gaagtaccga aaggtcttac cggaaaactc gacgcaagaa aatcagaga 1800
 gatcctcata aaggccaaga agggcggaaa gtccaaattg gtttaataa ggattaatta 1860
 attgtatggg cctccgtgct gttaggggca gtggaaactg cggcttagcc agccgtccc 1920
 tggtagtgt gctccgtcta agtacaccac tactcggttt ccacaaacga tcgactccc 1980
 aaaatagtct tggatgcaa taaaaaaaa aaagcgacat agccaaaaa aaaaaaaaa 2040
 aaaaaaaaa aaaaaaaaa aaaaaaaaa aaaaaaaaa aaaaaaaaa aaaaaaaaa 2100
 aaaaaaaaa aaactcgag ctagcataac cccttggggc ctctaacgg gtcttgagg 2160
 gttttttg 2168
 <210> 13
 <211> 2102
 <212> DNA
 <213> 人工序列 (artificial sequence)
 <400> 13

cgcgaaatta atacgactca ctataggggt atttttacaa caattaccaa caacaacaaa 60
 caacaacaaa cattacaatt actatttaca attacagaaa aatcgggtaa aaaaactttt 120
 cttctagtaa agacaggaaa ctttactcaa agcataccta ataacaatat aatcccataa 180
 aaaaaaaaaa tgtctgaaga cgccaaaac ataaagaaag gcccggcgcc attctatcct 240
 ctagaggatg gaaccgctgg agagcaactg cataaggcta tgaagagata cgccctgggt 300
 cctggaacaa ttgcttttac agatgcacat atcgaggatg acatcacgta cgcggaatac 360
 ttcgaaatgt ccgttcgggt ggcagaagct atgaaacgat atgggctgaa tacaatcac 420
 agaatcgtcg tatgcagtga aaactctctt caattcttta tgccgggtgtt gggcgcgtta 480
 tttatcggag ttgcagttgc gcccgcgaac gacatttata atgaacgtga attgctcaac 540
 agtatgaaca tttcgcagcc taccgtagtg tttgtttcca aaaaggggtt gcaaaaaatt 600
 ttgaacgtgc aaaaaaaaaatt accaataatc cagaaaatta ttatcatgga ttctaaaacg 660
 gattaccagg gatttcagtc gatgtacacg ttcgtccat ctcactacc tcccggtttt 720
 aatgaatacg atttgtacc agagtccttt gatcgtgaca aaacaattgc actgataatg 780
 aattcctctg gatctactgg gttacctaac ggtgtggccc ttccgatag aactgcctgc 840
 gtcagattct cgcatgccag agatcctatt tttggcaatc aatcattcc ggatactgcg 900
 attttaagtg ttgttcatt ccatcacggg tttggaatgt ttactacact cggatatttg 960
 atatgtggat ttcgagtcgt cttaatgtat agatttgaag aagagctgtt tttacgatcc 1020
 cttcaggatt acaaaaattca aagtgcgttg ctagtaccaa ccctattttc attcttcgcc 1080
 aaaagcactc tgattgacaa atacgattta tctaatttac acgaaattgc ttctgggggc 1140
 gcacctcttt cgaaagaagt cggggaagcg gttgcaaac gttccatct tccagggata 1200
 cgacaaggat atgggctcac tgagactaca tcagctatc tgattacacc cgagggggat 1260
 gataaacccg gcgcggtcgg taaagttgtt ccattttttg aagcgaaggt tgtggatctg 1320
 gataccggga aaacgctggg cgtaatcag agaggcgaat tatgtgtcag aggacctatg 1380
 attatgtccg gttatgtaaa caatccggaa gcgaccaacg ctttgattga caaggatgga 1440
 tggctacatt ctggagacat agcttactgg gacgaagacg aacacttctt catagttgac 1500
 cgcttgaagt cttaattaa atacaaagga tatcaggtgg cccccgctga attggaatcg 1560
 atattgttac aacaccccaa catcttcgac gcgggcgtgg caggtcttcc cgacgatgac 1620
 gccggtgaac ttcccgccgc cgttgttgtt ttggagcacg gaaagacgat gacggaaaaa 1680
 gagatcgtgg attacgtcgc cagtcaagta acaaccgca aaaagttgcg cggaggagtt 1740
 gtgtttgtgg acgaagtacc gaaaggtctt accggaaaac tcgacgcaag aaaaatcaga 1800
 gagatcctca taaaggccaa gaagggcgga aagtccaaat tggtttaaat aaggattaat 1860
 tacgtagcta ccctcctggt gcacttctg gtgcaaagca gaaccaaggg tacggtggta 1920
 cgcttggatg ccaataaaaa aaaaaaagcg acatagccaa aaaaaaaaaa aaaaaaaaaa 1980
 aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 2040
 aaaaaaaact cgagctagca taacccttg gggcctctaa acgggtcttg aggggttttt 2100
 tg 2102
 <210> 14
 <211> 53
 <212> DNA

<213> 人工序列 (artificial sequence)
<400> 14
cctgcgtccc aacctaatta atccttattt aaaccaattt ggactttccg ccc 53
<210> 15
<211> 36
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 15
cttggatgcc aataaaaaaaaa aaaaagcgac atagcc 36
<210> 16
<211> 47
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 16
ccaaattggt ttaaataagg attaattagg ttgggacgca ggtgtcg 47
<210> 17
<211> 59
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 17
ggctatgtcg cttttttttt tttattggca tccaaggtcg caggagata gcttatccc 59
<210> 18
<211> 58
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 18
catccagaac ccaacaagat aattaatcct tatttaaacc aatttgact ttccgcc 58
<210> 19
<211> 54
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 19
gcaataccag ccagcatgac ttggatgcca ataaaaaaaa aaaagcgaca tagc 54
<210> 20
<211> 57
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 20
ccaaattggt ttaaataagg attaattatc ttgttgggtt ctggatgtta ggatgac 57

<210> 21
<211> 41
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 21
ggcatccaag tcatgctggc tggattgct aagacttatc c 41
<210> 22
<211> 53
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 22
gaggcccata caattaatta atcettatth aaaccaatth ggactttccg ccc 53
<210> 23
<211> 55
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 23
gatcggactc ccaaaatagt cttggatgcc aataaaaaaa aaaaagcgac atagc 55
<210> 24
<211> 51
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 24
gtccaaattg gtttaaataa ggattaatta attgtatggg cctccgtgct g 51
<210> 25
<211> 37
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 25
ggcatccaag actatthttgg gagtccgatc gthttgtg 37
<210> 26
<211> 55
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 26
caggaggta gctacgtaat taatcettat ttaaaccaat ttggactthc cgccc 55
<210> 27
<211> 50
<212> DNA

<213> 人工序列 (artificial sequence)
<400> 27
gggtacggtg gtacgcttgg atgccaataa aaaaaaaaaa gcgacatagc 50
<210> 28
<211> 41
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 28
ggtttaaata aggattaatt acgtagctac cctcctggtg c 41
<210> 29
<211> 35
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 29
attggcatcc aagcgtacca ccgtagcctt ggttc 35
<210> 30
<211> 61
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 30
agaagaaaag ttttttaac cgatttttcc ctatagtgag tcgtattaat ttcgcgagat 60
c 61
<210> 31
<211> 69
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 31
gcatacctaa taacaatata atcccatagg tatttttaca acaattacca acaacaaca 60
acaacaaac 69
<210> 32
<211> 59
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 32
ctcactatag ggaaaaatcg gttaaaaaaa cttttcttct ggtatttat agagatcag 59
<210> 33
<211> 59
<212> DNA
<213> 人工序列 (artificial sequence)

<400> 33
gtaattgttg taaaaatacc tatgggatta tattgttatt aggtatgctt tgagacttc 59
<210> 34
<211> 57
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 34
cgactcacta tagggaaaaa tcggttaaaa aaacttttct tctagaaatt ctccagg 57
<210> 35
<211> 59
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 35
gtaattgttg taaaaatacc tatgggatta tattgttatt aggtatgctt tgagaaatt 59
<210> 36
<211> 51
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 36
cgactcacta tagggaaaaa tcggttaaaa aaacttttct tctggaacc g 51
<210> 37
<211> 59
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 37
gtaattgttg taaaaatacc tatgggatta tattgttatt aggtatgctt tgggaaact 59
<210> 38
<211> 58
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 38
gactcactat agggaaaaat cgggttaaaa aacttttctt ctagtaaaga caggaaac 58
<210> 39
<211> 59
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 39
gtaattgttg taaaaatacc tatgggatta tattgttatt aggtatgctt tgagtaaag 59
<210> 40

<211> 61
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 40
aaagtttttt taaccgattt ttctgtaatt gtaaatagta attgtaatgt tgtttgttgt 60
t 61
<210> 41
<211> 69
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 41
caaagcatac ctaataacaa tataatccca taaaaaaaa aatgtctga agacgccaaa 60
aacataaag 69
<210> 42
<211> 59
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 42
cattacaatt actatttaca attacagaaa aatcggttaa aaaaactttt cttctgggt 59
<210> 43
<211> 59
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 43
cagacatttt tttttttat gggattatat tgttattagg tatgctttga gacttcacg 59
<210> 44
<211> 59
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 44
caacattaca attactattt acaattacag aaaaatcggg taaaaaaact tttcttcta 59
<210> 45
<211> 59
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 45
cagacatttt tttttttat gggattatat tgttattagg tatgctttga gaaattctc 59
<210> 46
<211> 59

<212> DNA
 <213> 人工序列 (artificial sequence)
 <400> 46
 caacattaca attactattht acaattacag aaaaatcggg taaaaaaact tttcttctg 59
 <210> 47
 <211> 59
 <212> DNA
 <213> 人工序列 (artificial sequence)
 <400> 47
 cagacatttt tttttttat gggattatat tggtattagg tatgctttgg gaaacttgc 59
 <210> 48
 <211> 59
 <212> DNA
 <213> 人工序列 (artificial sequence)
 <400> 48
 caacattaca attactattht acaattacag aaaaatcggg taaaaaaact tttcttcta 59
 <210> 49
 <211> 59
 <212> DNA
 <213> 人工序列 (artificial sequence)
 <400> 49
 gacatttttt ttttttatgg gattatattg ttattaggta tgctttgagt aaagtttcc 59
 <210> 50
 <211> 1128
 <212> DNA
 <213> 人工序列 (artificial sequence)
 <400> 50
 cgcgaaatta atacgactca ctataggggt atttttacaa caattaccaa caacaacaaa 60
 caacaacaaa cattacaatt actatthtaca attacaaaa aaaaaaatgg tgagcaaggg 120
 cgaggagctg ttcaccgggg tgggtgccc atctgctcag ctggacggcg acgtaaacgg 180
 ccacaagtgc agcgtgcgcg gcgagggcga gggcgatgcc accaacggca agctgaccct 240
 gaagttcatc tgcaccaccg gcaagctgcc cgtgcccctgg cccaccctcg tgaccaccct 300
 gacctacggc gtgcagtgtc tcagccgcta ccccgaccac atgaagcagc acgacttctt 360
 caagtccgcc atgcccgaag gctacgtcca ggagcgcacc atctcttca aggacgacgg 420
 cacctacaag acccgcgccg aggtgaagtt cgagggcgac accctggtga accgcatcga 480
 gctgaagggc atcgaactca aggaggacgg caacatctc gggcacaagc tggagtacaa 540
 cttcaacagc cacaacgtct atatcacggc cgacaagcag aagaacggca tcaaggcgaa 600
 cttcaagatc cgccacaacg tcgaggacgg cagcgtgcag ctcgccgacc actaccagca 660
 gaacaccccc atcggcgacg gccccgtgct gctgcccgc aaccactacc tgagcaccca 720

gtccaagctg agcaaagacc ccaacgagaa gcgcatcac atggctctgc tggagtctgt 780
gaccgccgcc gggatcactc tcggcatgga cgagctgtac aagtaaataa ggattaatta 840
cttggatgcc aataaaaaaaaa aaaaagcgac atagccaaaa aaaaaaaaaa aaaaaaaaaa 900
aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 960
aaaaaactcg agcaccacca ccaccaccac tgagatccgg ctgctaacia agcccgaag 1020
gaagctgagt tggctgctgc caccgctgag caataactag cataaccctt tggggcctct 1080
aaacgggtct tgaggggttt tttgctgaaa ggaggaacta tatccgga 1128

<210> 51

<211> 1356

<212> DNA

<213> 人工序列 (artificial sequence)

<400> 51

cgcgaaatta atacgactca ctataggaa aatcggta aaaaaacttt tcttctagaa 60
attctccagg atttctcgac ctagtctgtt tatctggtga cttgcgctac cgttgctttg 120
cgtagagaat ttctcaaagc atacctaata acaatataat cccataggta tttttacaac 180
aattaccaac aacaacaaac aacaacaaac attacaatta ctatttacia ttacaaaaaa 240
aaaaaatggt gagcaagggc gaggagctgt tcaccgggtt ggtgcccatc ctggctgagc 300
tggacggcga cgtaaacggc cacaagtcca gcgtgcgagg cgagggcgag ggcgatgcca 360
ccaacggcaa gctgaccctg aagttcatct gcaccaccgg caagctgccc gtgccctggc 420
ccaccctcgt gaccaccctg acctacggcg tgcagtgtt cagccgctac cccgaccaca 480
tgaagcagca cgacttcttc aagtccgcca tgcccgaagg ctacgtccag gagcgacca 540
tctccttcaa ggacgacggc acctacaaga cccgcgccga ggtgaagttc gagggcgaca 600
ccctggtgaa ccgcatcgag ctgaagggca tcgacttcaa ggaggacggc aacatcctgg 660
ggcacaagct ggagtacaac ttcaacagcc acaacgtcta tatcacggcc gacaagcaga 720
agaacggcat caaggcgaac ttcaagatcc gccacaactc cgaggacggc agcgtgcagc 780
tcgccgacca ctaccagcag aacaccccca tcggcgacgg ccccgctgtg ctgcccgaca 840
accactacct gagcaccagc tccaagctga gaaagacc caacgagaag cgcgatcaca 900
tggctctgct ggagtctgtg accgccgccc ggatcactct cggcatggac gagctgtaca 960
agtaaataag gattaattat cttgttgggt tctggatgtt aggatgacga gtcgactcgg 1020
gctccgcaact aggtttggtc gcctagagga tggagatatg gaaagggtct cgtgtggtat 1080
cagtcggctc aaagacgcgc ttccaacatg ggctatggt cggataagtc ttagcaatac 1140
cagccagcat gacttggatg ccaataaaaa aaaaaaagcg acatagccaa aaaaaaaaaa 1200
aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 1260
aaaaaaaaaa aaaaaaaact cgagctagca taacccttg ggcctctaa acgggtcttg 1320
aggggttttt tgctgaaagg aggaactata tccgga 1356

<210> 52

<211> 1253

<212> DNA

<213> 人工序列 (artificial sequence)

<400> 52

cgcgaaatta atacgactca ctatagggaa aaatcggta aaaaaacttt tcttctggaa 60
 accgtggcaa gtttcccaaa gcatacctaa taacaatata atcccatagg tattttttaca 120
 acaattacca acaacaacaa acaacaaaca acattacaat tactattttac aattacaaaa 180
 aaaaaaaaaatg gtgagcaagg gcgaggagct gttcaccggg gtggtgcca tcctggtcga 240
 gctggacggc gacgtaaacc gccacaagtt cagcgtgcgc ggcgagggcg agggcgatgc 300
 caccaacggc aagctgaccc tgaagttcat ctgcaccacc ggcaagctgc ccgtgccctg 360
 gcccaccctc gtgaccacc tgacctacgg cgtgcagtgc ttcagccgct accccgacca 420
 catgaagcag cacgacttct tcaagtccgc catgcccga ggctacgtcc aggagcgcac 480
 catctccttc aaggacgacg gcacctaaa gaccgcgcc gaggtgaagt tcgagggcga 540
 caccctggtg aaccgcatcg agctgaagg catcgacttc aaggaggacg gcaacatcct 600
 ggggcacaag ctggagtaca acttcaacag ccacaacgtc tatatcacgg ccgacaagca 660
 gaagaacggc atcaaggcga acttcaagat ccgccacaac gtcgaggacg gcagcgtgca 720
 gctcgccgac cactaccagc agaacacccc catcggcgac ggccccgtgc tgctgcccga 780
 caaccactac ctgagcacc agtccaagct gagcaaagac cccaacgaga agcgcgatca 840
 catggtcctg ctggagtctg tgaccgccgc cgggatcact ctcggcatgg acgagctgta 900
 caagtaaata aggattaatt aattgtatgg gcctccgtgc tgttaggggc agtggaacg 960
 tcggtctagc cagccgtccc ctgggtagtg tgctccgtct aagtacacca ctactcggtt 1020
 tccacaaacg atcggactcc caaaatagtc ttgatgcca ataaaaaaaa aaaagcgaca 1080
 tagccaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 1140
 aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaactcga gctagcataa ccccttgggg 1200
 cctctaaacg ggtcttgagg ggtttttg c tgaaggagg aactatatcc gga 1253

<210> 53

<211> 1187

<212> DNA

<213> 人工序列 (artificial sequence)

<400> 53

cgcgaaatta atacgactca ctatagggaa aaatcggta aaaaaacttt tcttctagta 60
 aagacaggaa actttactca aagcatacct aataacaata taatcccata ggtattttta 120
 caacaattac caacaacaac aaacaacaaa caacattaca attactattt acaattacaa 180
 aaaaaaaaaa tggtagcaa gggcgaggag ctgttaccg gggtagtgcc catcctggtc 240
 gagctggacg gcgacgtaaa cggccacaag ttcagcgtgc gcggcgagg cgagggcgat 300
 gccaccaacg gcaagctgac cctgaagttc atctgacca ccggcaagct gcccgtgccc 360
 tggcccacc tcgtgaccac cctgacctac ggcgtgcagt gcttcagccg ctaccccgac 420
 cacatgaagc agcacgactt cttcaagtc gccatgccc aaggctacgt ccaggagcgc 480
 accatctcct tcaaggacga cggcacctac aagaccgcg ccgaggtgaa gttcgagggc 540
 gacaccctgg tgaaccgat cgagctgaag ggcacgact tcaaggagga cggcaacatc 600
 ctggggcaca agctggagta caacttcaac agccacaac tctatatcac ggccgacaag 660
 cagaagaacg gcatcaaggc gaacttcaag atccgccaca acgtcgagga cggcagcgtg 720

cagctcgccg accactacca gcagaacacc cccatcgcg acggccccgt gctgctgccc 780
gacaaccact acctgagcac ccagtccaag ctgagcaaag accccaacga gaagcgcgat 840
cacatggctc tgctggagtt cgtgaccgcc gccgggatca ctctcgcat ggacgagctg 900
tacaagtaaa taaggattaa ttacgtagct accctcctgg tgcacttct gtgcaaagc 960
agaaccaagg gtacggtggt acgcttggat gccataaaa aaaaaaaagc gacatagcca 1020
aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 1080
aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa tcgagctagc ataaccctt gggcctcta 1140
aacgggtctt gaggggtttt ttgctgaaag gaggaactat atccgga 1187

<210> 54

<211> 59

<212> DNA

<213> 人工序列 (artificial sequence)

<400> 54

ttttttttt gtaattgtaa atagtaattg taatgttgtt tgttgtttgt tgttgttgg 59

<210> 55

<211> 55

<212> DNA

<213> 人工序列 (artificial sequence)

<400> 55

cggcatggac gagctgtaca agtaaataag gattaattat cttgttgggt tctgg 55

<210> 56

<211> 58

<212> DNA

<213> 人工序列 (artificial sequence)

<400> 56

caacattaca attactattht acaattacaa aaaaaaaaaa tggtagcaa gggcgagg 58

<210> 57

<211> 31

<212> DNA

<213> 人工序列 (artificial sequence)

<400> 57

ccttathtac ttgtacagct cgtccatgcc g 31

<210> 58

<211> 58

<212> DNA

<213> 人工序列 (artificial sequence)

<400> 58

cggcatggac gagctgtaca agtaaataag gattaattaa ttgtatgggc ctccgtgc 58

<210> 59

<211> 59

<212> DNA

<213> 人工序列 (artificial sequence)

<400> 59

cggcatggac gagctgtaca agtaaataag gattaattac gtagctaccc tcctggtgc 59

<210> 60

<211> 27

<212> DNA

<213> 人工序列 (artificial sequence)

<400> 60

cgcgaaatta atacgactca ctatagg 27

<210> 61

<211> 24

<212> DNA

<213> 人工序列 (artificial sequence)

<400> 61

tccgatata gttcctcctt tcag 24

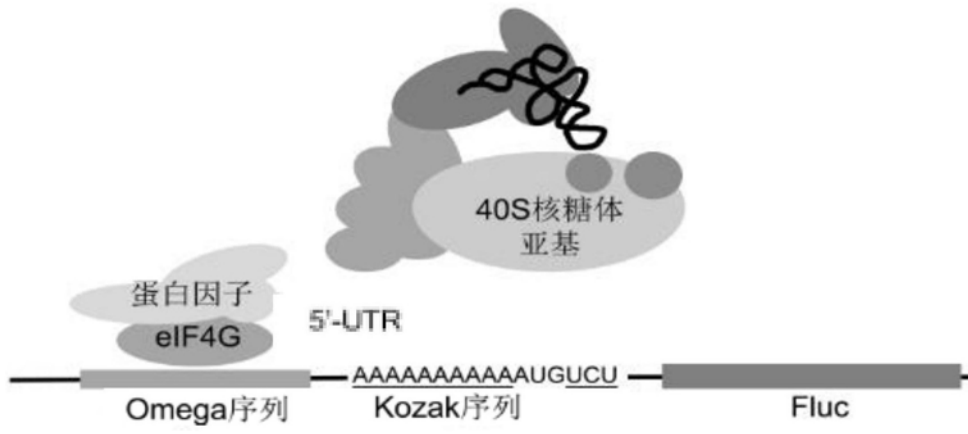


图1

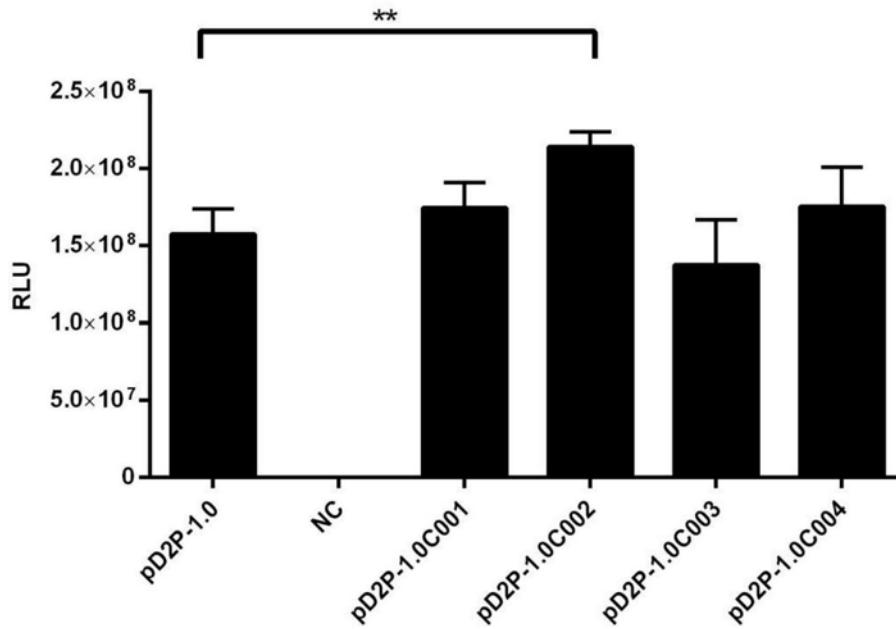


图2

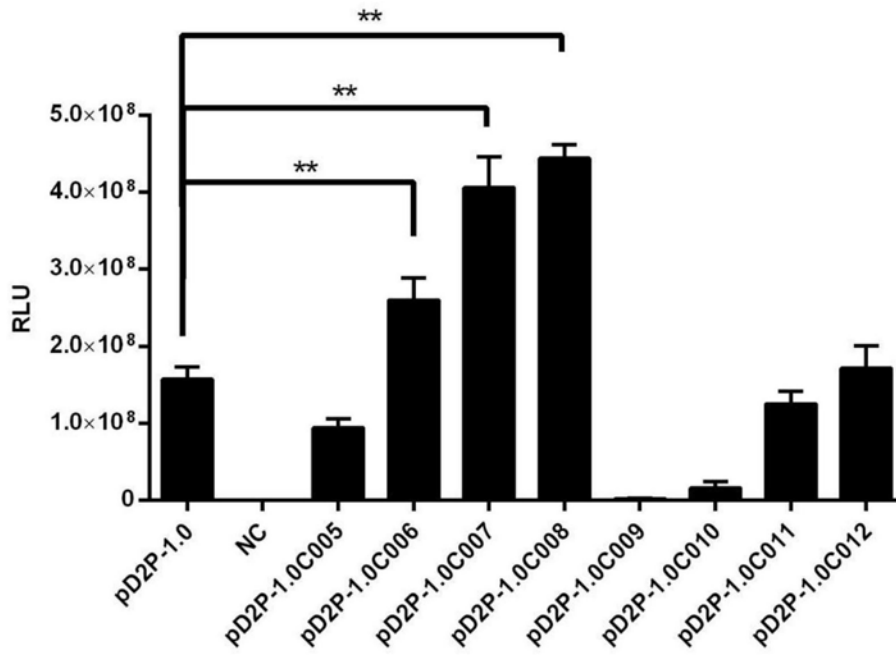


图3

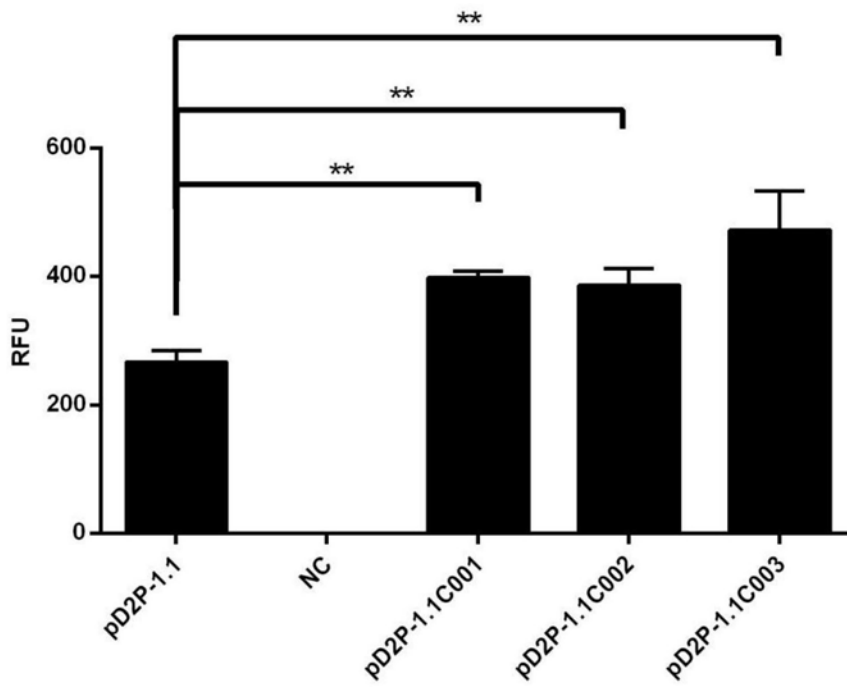


图4

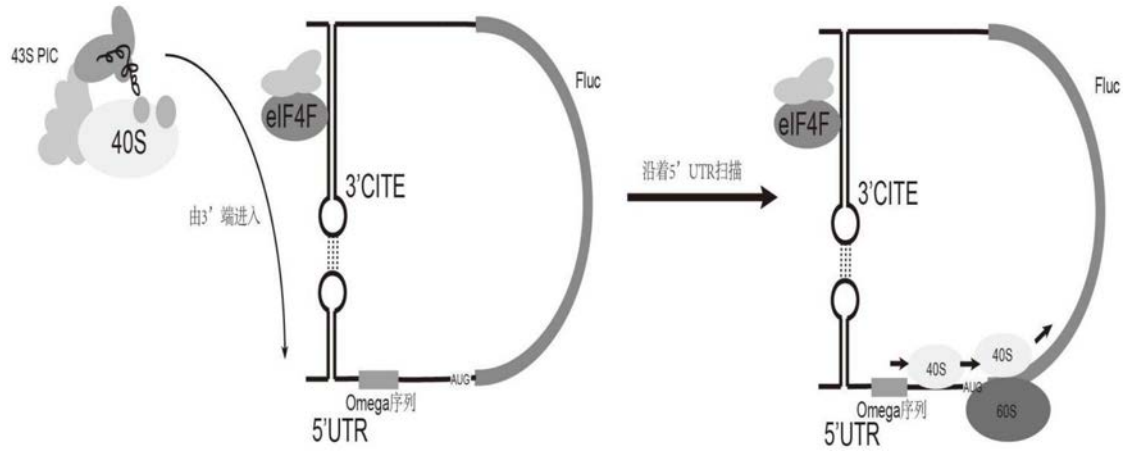


图5