



## (12)发明专利申请

(10)申请公布号 CN 109022478 A

(43)申请公布日 2018.12.18

(21)申请号 201710430876.3

C12N 9/12(2006.01)

(22)申请日 2017.06.09

(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 柴智 刘帅龙  
于雪

(51)Int.Cl.

C12N 15/81(2006.01)

C12P 21/00(2006.01)

C07K 14/00(2006.01)

C07K 14/435(2006.01)

C07K 14/47(2006.01)

C07K 16/00(2006.01)

C12N 9/02(2006.01)

C12N 9/08(2006.01)

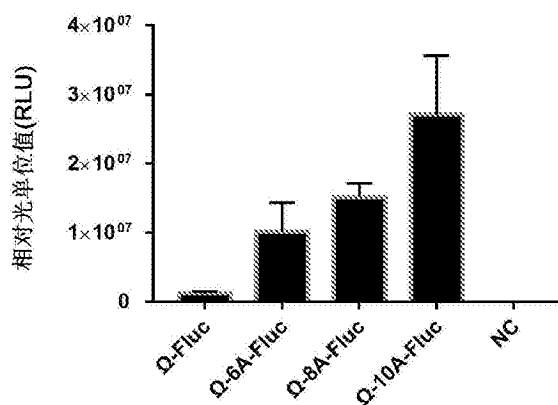
权利要求书2页 说明书12页  
序列表12页 附图1页

### (54)发明名称

一种用于高通量体外蛋白质合成的DNA元件

### (57)摘要

本发明提供了一种用于高通量体外蛋白质合成的DNA元件,具体地,本发明首次意外地发现了一种可大幅度增强蛋白质翻译效率的新型DNA元件,在酵母体外蛋白质合成体系中应用本发明的DNA元件,所合成的荧光素酶活性的相对光单位值比仅含有Ω序列的DNA元件增强约22.3倍。



1. 一种核酸构建物,其特征在于,所述的构建物具有从5'至3'的式I结构:

Z1-Z2-Z3-Z4 (I)

式中,

Z1、Z2、Z3、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为烟草花叶病毒的5'端前导序列(leading sequence)-Ω序列;

Z2为腺嘌呤脱氧核苷酸的寡聚链[oligo(A)]<sub>n</sub>;

Z3为翻译起始密码子;

Z4为丝氨酸密码子;

并且,所述Z2、Z3、Z4共同组成了Kozak序列,所述Kozak序列来源于克鲁维酵母属酵母。

2. 如权利要求1所述的核酸构建物,其特征在于,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。

3. 一种核酸构建物,其特征在于,所述的构建物具有从5'至3'的式II结构:

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

式中,

Z1、Z2、Z3、Z4、Z5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为烟草花叶病毒的5'端前导序列(leading sequence)-Ω序列;

Z2为腺嘌呤脱氧核苷酸的寡聚链[oligo(A)]<sub>n</sub>;

Z3为翻译起始密码子;

Z4为丝氨酸密码子;

Z5为外源蛋白的编码序列;

并且,所述Z2、Z3、Z4共同组成了Kozak序列,所述Kozak序列来源于克鲁维酵母属酵母。

4. 如权利要求3所述的核酸构建物,其特征在于,所述的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、或其组合。

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或3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或3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或3所述的构建物;

(b) 权利要求5所述的载体或载体组合;和

(c) 权利要求6所述的基因工程细胞。

8. 如权利要求1或3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所述试剂盒的用途,其特征在于,用于进行高通量的体外蛋白合成。

9. 一种体外高通量的蛋白合成方法,其特征在于,包括步骤:

(i) 在酵母体外蛋白合成体系存在下,提供权利要求3所述的核酸构建物;

(ii) 在适合的条件下,孵育步骤(i)的酵母体外蛋白合成体系一段时间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,所述方法还包括:(iii) 任选地从所述酵母体外蛋白合成体系中,分离或检测所述外源蛋白。

## 一种用于高通量体外蛋白质合成的DNA元件

### 技术领域

[0001] 本发明涉及生物技术领域,较佳地,涉及一种用于高通量体外蛋白质合成的DNA元件。

### 背景技术

[0002] 在真核细胞内,绝大部分用于编码蛋白质的mRNA分子都含有5'端的“帽子”结构( $m^7GpppN$ )和3'端的多聚腺苷链[poly(A)]结构。这两种结构不仅可以增强 mRNA分子的稳定性,而且是蛋白质翻译所必需的<sup>[1]</sup>。其中,5'“帽子”结构能够被翻译起始因子eIF4E所识别,并招募下游的多种翻译起始因子和核糖体,进而起始蛋白质翻译过程<sup>[1]</sup>。真核细胞内对mRNA进行加“帽子”修饰涉及由3个酶催化的3步反应,而且这个过程是与由RNA聚合酶II催化的转录同时进行的<sup>[2]</sup>。

[0003] 在转录与翻译偶联的体外蛋白质合成体系中,往往使用噬菌体或者病毒的 RNA聚合酶(如T7RNA聚合酶)进行转录,这决定了很难对mRNA进行加“帽子”修饰。在基于真核细胞的体外蛋白质合成体系中,要想利用来源于真核细胞的蛋白质翻译体系(包括翻译起始因子和核糖体等),不具有“帽子”结构的mRNA 必须含有其他元件来招募相关因子,进而起始蛋白质的翻译<sup>[3]</sup>。

[0004] 然而,目前的提高蛋白质合成所用的元件主要有烟草花叶病毒的翻译增强子 $\Omega$ 序列和一些其他病毒的内部核糖体进入序列(Internal Ribosome Entry Sites,IRESs),如EMCV IRES(Encephalomyocarditis virus)。这些元件可以在细胞内和体外蛋白质合成体系中起始“帽子”-非依赖的蛋白质翻译。但是,这些元件起始蛋白质翻译的效率往往比较低,会增加蛋白质合成所需时间,限制最终蛋白质的产量<sup>[4]</sup>。

[0005] 因此,本领域迫切需要开发一种能够增强蛋白质翻译效率的新的DNA元件。

### 发明内容

[0006] 本发明的目的在于提供一种能够增强蛋白质翻译效率的新的DNA元件。

[0007] 本发明第一方面提供了一种核酸构建物,所述的构建物具有从5'至3'的式 I 结构:

[0008] Z1-Z2-Z3-Z4 (I)

[0009] 式中,

[0010] Z1、Z2、Z3、Z4分别为用于构成所述构建物的元件;

[0011] 各“-”独立地为键或核苷酸连接序列;

[0012] Z1为烟草花叶病毒的5'端前导序列(leading sequence)- $\Omega$ 序列;

[0013] Z2为腺嘌呤脱氧核苷酸的寡聚链[oligo(A)]<sub>n</sub>;

[0014] Z3为翻译起始密码子;

[0015] Z4为丝氨酸密码子;

[0016] 并且,所述Z2、Z3、Z4共同组成了Kozak序列,所述Kozak序列来源于克鲁维酵母属

酵母。

[0017] 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(*Kluyveromyces dobzhanskii*)、或其组合。

[0018] 在另一优选例中,所述翻译起始密码子选自下组:ATG、ATA、ATT、GTG、TTG、或其组合。

[0019] 在另一优选例中,所述翻译起始密码子为ATG。

[0020] 在另一优选例中,所述丝氨酸密码子选自下组:TCT、TCC、TCA、TCG、AGT、AGC、或其组合。

[0021] 在另一优选例中,所述丝氨酸密码子为TCT。

[0022] 在另一优选例中,n为6-12,较佳地,8-10。

[0023] 在另一优选例中,所述 $\Omega$ 序列包括直接重复模块(ACAATTAC)<sub>m</sub>和(CAA)<sub>p</sub>模块。

[0024] 在另一优选例中,所述m为1-6,较佳地,2-4。

[0025] 在另一优选例中,所述p为6-12个,较佳地,8-10个。

[0026] 在另一优选例中,所述(CAA)<sub>p</sub>模块为1-5个,较佳地,1-3个。

[0027] 在另一优选例中,所述(CAA)<sub>p</sub>模块还包括优化的(CAA)<sub>p</sub>模块。

[0028] 在另一优选例中,所述核酸构建物的序列如SEQ ID NO.:1-3所示。

[0029] 在另一优选例中,所示Kozak序列如SEQ ID NO.:5所示。

[0030] 本发明第二方面提供了一种核酸构建物,所述的构建物具有从5'至3'的式II结构:

[0031] Z1-Z2-Z3-Z4-Z5 (II)

[0032] 式中,

[0033] Z1、Z2、Z3、Z4、Z5分别为用于构成所述构建物的元件;

[0034] 各“-”独立地为键或核苷酸连接序列;

[0035] Z1为烟草花叶病毒的5'端前导序列(leading sequence)- $\Omega$ 序列;

[0036] Z2为腺嘌呤脱氧核苷酸的寡聚链[oligo(A)]<sub>n</sub>;

[0037] Z3为翻译起始密码子;

[0038] Z4为丝氨酸密码子;

[0039] Z5为外源蛋白的编码序列;

[0040] 并且,所述Z2、Z3、Z4共同组成了Kozak序列,所述Kozak序列来源于克鲁维酵母属酵母。

[0041] 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。

[0042] 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。

[0043] 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。

[0044] 在另一优选例中,所述的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、或其组合。

[0045] 在另一优选例中,所述外源蛋白选自下组: $\alpha$ -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 $\alpha$ A、白细胞介素-1 $\beta$ 、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。

[0046] 在另一优选例中,所述外源蛋白的编码序列的核苷酸序列如SEQ ID NO.:6-12 所示。

[0047] 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或本发明第二方面所述的核酸构建物。

[0048] 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或本发明第二方面所述的构建物,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。

[0049] 在另一优选例中,所述基因工程细胞为酵母细胞。

[0050] 在另一优选例中,所述酵母细胞来自克鲁维酵母属酵母。

[0051] 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(*Kluyveromyces dobzhanskii*)、或其组合。

[0052] 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:

[0053] (a) 本发明第一方面或本发明第二方面所述的构建物;

[0054] (b) 本发明第三方面所述的载体或载体组合;和

[0055] (c) 本发明第四方面所述的基因工程细胞;

[0056] 在另一优选例中,所述试剂盒还包括(d) 酵母体外蛋白合成体系。

[0057] 在另一优选例中,所述酵母体外蛋白合成体系为克鲁维酵母体外蛋白合成体系(优选乳酸克鲁维酵母体外蛋白合成体系)。

[0058] 本发明第六方面提供了如本发明第一方面或本发明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试剂盒的用途,用于进行高通量的体外蛋白合成。

[0059] 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:

[0060] (i) 在酵母体外蛋白合成体系存在下,提供本发明第二方面所述的核酸构建物;

[0061] (ii) 在适合的条件下,孵育步骤(i)的酵母体外蛋白合成体系一段时间T1,从而合成所述外源蛋白。

[0062] 在另一优选例中,所述方法还包括:(iii) 任选地从所述酵母体外蛋白合成体系中,分离或检测所述外源蛋白。

[0063] 在另一优选例中,所述酵母体外蛋白合成体系为克鲁维酵母体外蛋白合成体系(优选乳酸克鲁维酵母体外蛋白合成体系)。

[0064] 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。

[0065] 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。

[0066] 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。

[0067] 在另一优选例中,所述的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、或其组

合。

[0068] 在另一优选例中,所述外源蛋白选自下组: $\alpha$ -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 $\alpha$ A、白细胞介素-1 $\beta$ 、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。

[0069] 在另一优选例中,所述外源蛋白的编码序列的核苷酸序列如SEQ ID NO.:6-12 所示。

[0070] 在另一优选例中,所述步骤(ii)中,反应温度为20-37℃,较佳地,22-35℃。

[0071] 在另一优选例中,所述步骤(ii)中,反应时间为1-10h,较佳地,2-8h。

[0072] 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

[0073] 图1显示了加入了含有乳酸克鲁维酵母特异性Kozak序列的 $\Omega$ -10A/8A/6A-Fluc 3个DNA片段和不含有Kozak序列的 $\Omega$ -Fluc DNA片段在酵母体外蛋白质合成体系中编码合成Fluc蛋白质的相对光单位值(Relative Light Unit,RLU),以及不加入任何DNA片段的酵母体外蛋白质体系的相对光单位值 RLU。

## 具体实施方式

[0074] 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可大幅度增强蛋白质翻译效率的新型DNA元件,本发明的DNA元件包括 $\Omega$ 序列、腺嘌呤脱氧核苷寡聚链、翻译起始密码子和丝氨酸密码子,其中后三个元件构成了克鲁维酵母(如乳酸克鲁维酵母)特异性的Kozak序列,在酵母体外蛋白质合成体系中应用本发明的DNA元件,所合成的荧光素酶活性的相对光单位值比仅含有 $\Omega$ 序列的DNA元件增强约22.3倍。在此基础上,本发明人完成了本发明。

[0075] 酵母体外蛋白质合成体系

[0076] 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。

[0077] 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<sup>[5]</sup>。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<sup>[5]</sup>。

[0078] 在本发明中,酵母体外蛋白质合成体系不受特别限制,一种优选的酵母体外蛋白质合成体系为克鲁维酵母表达系统(更佳地,乳酸克鲁维酵母表达系统)。

[0079] 在本发明中,所述酵母体外蛋白质合成体系包括:

[0080] (a) 酵母细胞提取物;

[0081] (b) 聚乙二醇;

[0082] (c) 任选的外源蔗糖;和

[0083] (d) 任选的溶剂,所述溶剂为水或水性溶剂。

[0084] 在一特别优选的实施方式中,本发明提供的体外蛋白合成体系包括:酵母细胞提取物,4-羟乙基哌嗪乙磺酸,醋酸钾,醋酸镁,腺嘌呤核苷三磷酸(ATP),鸟嘌呤核苷三磷酸(GTP),胞嘧啶核苷三磷酸(CTP),胸腺嘧啶核苷三磷酸(TTP),氨基酸混合物,磷酸肌酸,二硫苏糖醇(DTT),磷酸肌酸激酶,RNA酶抑制剂,荧光素,荧光素酶DNA, RNA聚合酶。

[0085] 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。

[0086] 在本发明中,所述酵母细胞提取物在体外蛋白合成体系中的比例不受特别限制,通常所述酵母细胞提取物在体外蛋白质合成蛋白合成体系中所占体系为 20-70%,较佳地,30-60%,更佳地,40-50%。

[0087] 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蛋白翻译的核糖体、转运RNA、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,酵母提取物中还含有一些源自酵母细胞的细胞质中的其他蛋白,尤其是可溶性蛋白。

[0088] 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。

[0089] 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:

[0090] (i) 提供酵母细胞;

[0091] (ii) 对酵母细胞进行洗涤处理,获得经洗涤的酵母细胞;

[0092] (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;

[0093] (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。

[0094] 在本发明中,所述的固液分离方式不受特别限制,一种优选的方式为离心。

[0095] 在一优选实施方式中,所述离心在液态下进行。

[0096] 在本发明中,所述离心条件不受特别限制,一种优选的离心条件为5000-100000  $\times g$ ,较佳地,8000-30000  $\times g$ 。

[0097] 在本发明中,所述离心时间不受特别限制,一种优选的离心时间为0.5min-2h,较佳地,20min-50min。

[0098] 在本发明中,所述离心的温度不受特别限制,优选的,所述离心在1-10℃下进行,较佳地,在2-6℃下进行。

[0099] 在本发明中,所述的洗涤处理方式不受特别限制,一种优选的洗涤处理方式为采用洗涤液在pH为7-8(较佳地,7.4)下进行处理,所述洗涤液没有特别限制,典型的所述洗涤液选自下组:4-羟乙基哌嗪乙磺酸钾、醋酸钾、醋酸镁、或其组合。

[0100] 在本发明中,所述破细胞处理的方式不受特别限制,一种优选的所述的破细胞处理包括高压破碎、冻融(如液氮低温)破碎。

[0101] 所述体外蛋白质合成体系中的核苷三磷酸混合物为腺嘌呤核苷三磷酸、鸟嘌呤核苷三磷酸、胞嘧啶核苷三磷酸和尿嘧啶核苷三磷酸。在本发明中,各种单核苷酸的浓度没有特别限制,通常每种单核苷酸的浓度为0.5-5mM,较佳地为1.0-2.0mM。

[0102] 所述体外蛋白质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0.05、0.06、0.07、0.08mM。

[0103] 在优选例中,所述体外蛋白质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%,以所述蛋白合成体系的总重量计。代表性的PEG例子包括(但并不限于):PEG3000,PEG8000,PEG6000和PEG3350。应理解,本发明的体系还可包括其他各种分子量的聚乙二醇(如PEG200、400、1500、2000、4000、6000、8000、10000等)。

[0104] 在优选例中,所述体外蛋白质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。

[0105] 一种特别优选的体外蛋白质合成体系,除了酵母提取物之外,还含有以下组分:22mM,pH为7.4的4-羟乙基哌嗪乙磺酸,30-150mM醋酸钾,1.0-5.0 mM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,25 mM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/ $\mu$ l萤火虫荧光素酶的DNA,0.027-0.054mg/mL T7 RNA聚合酶。

[0106] 外源蛋白的编码序列(外源DNA)

[0107] 如本文所用,术语“外源蛋白的编码序列”与“外源DNA”可互换使用,均指外源的用于指导蛋白质合成的DNA分子。通常,所述的DNA分子为线性的或环状的。所述的DNA分子含有编码外源蛋白的序列。

[0108] 在本发明中,所述的编码外源蛋白的序列的例子包括(但并不限于):基因组序列、cDNA序列。所述的编码外源蛋白的序列还含有启动子序列、5'非翻译序列、3'非翻译序列。

[0109] 在本发明中,所述外源DNA的选择没有特别限制,通常,外源DNA选自下组: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、或其组合。

[0110] 外源DNA还可以选自下组:编码 $\alpha$ -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 $\alpha$ A、白细胞介素-1 $\beta$ 、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。

[0111] 一类代表性的外源DNA的序列选自:SEQ ID NO.:6-12。

[0112] 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 $\beta$ -半乳糖苷酶( $\beta$ -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。

[0113] 在一优选实施方式中,所述绿色荧光蛋白的核苷酸序列如SEQ ID NO.:6所示;所述黄色荧光蛋白的核苷酸序列如SEQ ID NO.:7所示;所述大肠杆菌 $\beta$ -半乳糖苷酶的核苷酸序列如SEQ ID NO.:8所示;所述人赖氨酸-tRNA合成酶的核苷酸序列如SEQ ID NO.:9所示;

所述人亮氨酸-tRNA合成酶的核苷酸序列如SEQ ID NO.:10 所示;所述拟南芥甘油醛3-磷酸脱氢酶的核苷酸序列如SEQ ID NO.:11所示;所述鼠过氧化氢酶的核苷酸序列如SEQ ID NO.:12所示。

[0114] Ω 序列

[0115] 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)<sub>p</sub>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元件或者蛋白质来增强蛋白质翻译的效率。

[0116] Kozak序列

[0117] 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序列往往是不同的,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。

[0118] 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)。

[0119] DNA构建物

[0120] 本发明提供了一种DNA构建物,所述DNA构建物含有结构如式I所示的核酸序列:

[0121] Z1-Z2-Z3-Z4 (I)

[0122] 式中,

[0123] Z1、Z2、Z3、Z4分别为用于构成所述构建物的元件;

[0124] 各“-”独立地为键或核苷酸连接序列;

[0125] Z1为烟草花叶病毒的5' 端前导序列(leading sequence)-Ω 序列;

[0126] Z2为腺嘌呤脱氧核苷酸的寡聚链[oligo(A)]<sub>n</sub>;

[0127] Z3为翻译起始密码子;

[0128] Z4为丝氨酸密码子;

[0129] 并且,所述Z2、Z3、Z4共同组成了Kozak序列,所述Kozak序列来源于克鲁维酵母属酵母。

[0130] 本发明还提供了一种DNA构建物,所述的构建物具有从5' 至3' 的式II结构:

[0131] Z1-Z2-Z3-Z4-Z5 (II)

[0132] 式中,

[0133] Z1、Z2、Z3、Z4、Z5分别为用于构成所述构建物的元件;

[0134] 各“-”独立地为键或核苷酸连接序列;

[0135] Z1为烟草花叶病毒的5' 端前导序列(leading sequence)-Ω 序列;

[0136] Z2为腺嘌呤脱氧核苷酸的寡聚链[oligo(A)]<sub>n</sub>;

[0137] Z3为翻译起始密码子;

[0138] Z4为丝氨酸密码子;

[0139] Z5为外源蛋白的编码序列;

[0140] 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、萤光素酶突变体的DNA、或其组合。

[0141] 外源蛋白还可以选自下组: $\alpha$ -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 $\alpha$ A、白细胞介素-1 $\beta$ 、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。

[0142] 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。

[0143] 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:启动子、终止子、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。

[0144] 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。

[0145] 在本发明中,在本发明的酵母体外蛋白质合成体系中应用本发明的DNA构建物,可显著提高蛋白翻译的效率,具体地,应用本发明的DNA构建物所合成的荧光素酶活性的相对光单位值比仅含有 $\Omega$ 序列的DNA元件增强约22.3倍。

[0146] 载体,基因工程细胞

[0147] 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DNA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。只要其能够在宿主体内复制和稳定,任何质粒和载体都是可以采用的。

[0148] 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。

[0149] 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构建物或载体。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。

[0150] 优选地,所述的基因工程细胞为真核细胞。

[0151] 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)。

[0152] 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(如大肠杆菌)时,可以用CaCl<sub>2</sub>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(如显微注射、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。

[0153] 体外高通量的蛋白合成方法

[0154] 本发明提供了一种体外高通量的蛋白合成方法,包括步骤:

[0155] (i) 在酵母体外蛋白合成体系存在下,提供本发明第一方面所述的核酸构建物,并加入外源的用于指导蛋白质合成的DNA分子;或提供本发明第二方面所述的核酸构建物;

[0156] (ii) 在适合的条件下,孵育步骤(i)的酵母体外蛋白合成体系一段时间T<sub>1</sub>,从而合成由所述外源DNA编码的蛋白质或外源蛋白。

[0157] 本发明的主要优点包括:

[0158] (1) 本发明首次发现,将Ω序列和来源于克鲁维酵母(优选乳酸克鲁维酵母)的Kozak序列作为核酸构建物,应用于本发明的酵母体外蛋白质合成体系中,可显著提高蛋白翻译的效率。

[0159] (2) 本发明首次发现,与仅含有Ω序列的DNA元件相比,本发明的DNA构建物所合成的荧光素酶活性的相对光单位值增强约22.3倍。

[0160] (3) 本发明首次发现,在促进蛋白质合成方面,Ω-10A要优于Ω-8A和Ω-6A。

[0161] (4) 本发明的新型DNA元件Ω-10A来源于乳酸克鲁维酵母特异性的Kozak序列,能够显著增强酵母体外合成体系合成蛋白质的效率。与不含有Kozak序列的Ω-Fluc DNA片段相比,Ω-10A-Fluc增加了22.3倍的蛋白质合成量。

[0162] (5) 本发明通过对乳酸克鲁维酵母基因mRNA中翻译起始密码子AUG上下游序列的分析发现,其5'端的序列是A与U富集的,而且含有多腺嘌呤核苷(10A)寡聚链的DNA片段增强蛋白质合成的效率要优于含有8A与6A的DNA片段。以上分析与实验结果表明,本发明所包含的新型DNA元件具有进一步优化的潜力。

[0163] (6) 本发明所包含的新型DNA元件Ω-10A中的Ω序列通常被应用于体外蛋白质合成体系中,来起始蛋白质的合成。本发明通过对Ω序列组成部分的了解,得知具有起始蛋白质翻译能力的可能的组成模块,后续可以针对这些模块的序列、数量以及不同模块之间的排列顺序进行进一步的优化,期望能够更进一步强化其起始蛋白质合成效率的能力。

[0164] 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并不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。

[0165] 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。

[0166] 实施例1:乳酸克鲁维酵母特异性Kozak序列分析

[0167] 1.1 乳酸克鲁维酵母特异性的Kozak序列的确定:根据分析酿酒酵母(*Saccharomyces cerevisiae*)特异性的Kozak序列的方法,检索并收集乳酸克鲁维酵母107个基因的mRNA序列,并将筛选出的乳酸克鲁维酵母mRNA翻译起始密码子AUG 5'端的60个碱基序列和3'端下游的一个密码子序列进行A、U、C和G的含量分析,意外发现在AUG的5'端位置上,腺嘌呤A的比例占优势,而在AUG的3'端确定了一个丝氨酸的密码子UCU占优势,这样就初步确定了乳酸克鲁维酵母特异性的Kozak序列,命名为k1-Kozak序列。

[0168] 1.2构建含有乳酸克鲁维酵母特异性k1-Kozak序列的酵母体外蛋白质合成体系质粒:在原有的酵母体外蛋白质合成体系质粒(修改过的pET21a质粒,获自康码(上海)生物科技有限公司)中的 $\Omega$ 序列与报告基因萤火虫荧光素酶 Fluc的序列之间加入乳酸克鲁维酵母特异性的k1-Kozak序列,在Fluc起始密码子的5'端插入了含有不同数量腺嘌呤脱氧核苷酸的寡聚链,包括10A、8A和6A,并将丝氨酸密码子TCT插入Fluc起始密码子和第二个密码子之间,最终形成了含有乳酸克鲁维酵母特异性k1-Kozak序列的pKL- $\Omega$ -10A-Fluc、pKL- $\Omega$ -8A-Fluc和pKL- $\Omega$ -6A-Fluc的质粒。

[0169] 具体实施过程如下:使用3对引物:

[0170] 06A-605\_f (CAACAATTACCAACAACAACAACAACAACAATTAC AATTACTATTTACAATTACAAAAAATGTCTGAAGACGCCAAAAACAT AAAGAAAGGCC) (SEQ ID NO.:13)  
与OK-60\_b (TTTTTTTGTAAATTGTAAATAGTAATTGTAATGTTGTTGTTGTTGT TG) (SEQ ID NO.:14):

[0171] 08A-605\_f (CAATTACCAACAACAACAACAACAACAACATTACAATTACTATTTACAATTACAAAAAAAAAATGTCTGAAGACGCCAAAAACATA AAGAAAGGCC) (SEQ ID NO.:15)  
与OK-60 b;

[0172] 010A-605\_f (AATTACCAACAACAACAACAACAACATTACAATTACTATTTACAATTACAAAAAAAAAATGTCTGAAGACGCCAAAAACA TAAAGAAAGGCC) (SEQ ID NO.: 16) 与OK-60\_b对质粒pKL- $\Omega$ -Fluc (SEQ ID NO.: 4) 进行PCR扩增,向20 $\mu$ L扩增产物中加入1 $\mu$ L Dpn I, 37 $^{\circ}$ C孵育6h;将Dpn I处理后产物4 $\mu$ L加入50 $\mu$ L DH5 $\alpha$ 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 $\mu$ 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,命名为pKL- $\Omega$ -10A-Fluc (SEQ ID NO.: 1)、pKL- $\Omega$ -8A-Fluc (SEQ ID NO.: 2) 和pKL- $\Omega$ -6A-Fluc (SEQ ID NO.: 3) 质粒。

[0173] 实施例2:k1-Kozak序列在酵母体外蛋白质合成体系中的应用

[0174] 2.1利用PCR的方法,并使用引物T7\_pET21a\_F: CGCGAAATTAATACGACTCACTATAGG (SEQ ID NO.:17) 和 T7ter\_pET21a\_R:TCCGATATAGTTCTCTCTTTCAG (SEQ ID NO.:18) 将质粒pKL- $\Omega$ -10A/8A/6A-Fluc和pKL- $\Omega$ -Fluc中位于T7转录起始序列和终止序列之间包括 $\Omega$ -10A/8A/6A-Fluc的片段和 $\Omega$ -Fluc片段进行扩增。

[0175] 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(pH5.2),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,然后再离心15min,弃掉上清,并用超纯水将沉淀溶解,测定DNA浓度。

[0176] 2.2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25-30℃的环境中,静置孵育约2-6h。反应结束后,在96孔白板或者384孔白板中加入等体积的Fluc底物荧光素(luciferin),立即放置于Envision 2120多功能酶标仪(Perkin Elmer),读数,检测Fluc活性,相对光单位值(Relative Light Unit,RLU)作为活性单位,如图1所示。

[0177] 2.3使用不含有kl-Kozak序列的DNA片段( $\Omega$ -Fluc)作为对照,每个样品均设计三组独立实验,并设计不含有任何DNA片段的实验组作为阴性对照。

[0178] 实验结果

[0179] 1.乳酸克鲁维酵母特异性的kl-Kozak序列

[0180] 乳酸克鲁维酵母的107个基因mRNA的5'端位于起始密码子AUG之前的位置上,腺嘌呤核苷酸(A)的比例占优势,有几个位置上尿嘧啶核苷酸(U)的比例与接近,但略少。所以与酿酒酵母类似,乳酸克鲁维酵母的基因mRNA起始密码子5'端序列是A与U富集的。起始密码子AUG之后的密码子的3个位置上占优势的依次为U、C和U,为丝氨酸的密码子。最终确定的乳酸克鲁维酵母特异性的kl-Kozak序列如SEQ ID NO.:5所示。

[0181] 2.kl-Kozak序列在酵母体外蛋白质合成体系中的应用

[0182] 如图1所示,含有乳酸克鲁维酵母特异性kl-Kozak序列的 $\Omega$ -10A/8A/6A-Fluc 3个DNA片段在酵母体外蛋白质合成体系中编码合成的Fluc蛋白放出的相对光单位值最高达到 $2.72 \times 10^7$ 。而不含有kl-Kozak序列的 $\Omega$ -Fluc DNA片段编码合成的Fluc蛋白的相对光单位值仅有 $1.22 \times 10^6$ 。这表明kl-Kozak序列的插入能够有效增强酵母体外蛋白质合成体系合成蛋白质的效率。因为处于检测仪器相对光单位值RLU与蛋白质浓度关系的线性范围内,活性最高的 $\Omega$ -10A-Fluc片段的相对光单位值是 $\Omega$ -Fluc片段的22.3倍,表明 $\Omega$ -10A能够增强大约22.3倍的蛋白质合成。

[0183] 含有不同长度腺嘌呤核苷寡聚链的DNA片段增强蛋白质合成的效率也是存在差异的,加入了 $\Omega$ -10A-Fluc、 $\Omega$ -8A-Fluc和 $\Omega$ -6A-Fluc片段的酵母体外蛋白质合成体系合成的Fluc蛋白的相对光单位值分别为 $2.72 \times 10^7$ 、 $1.52 \times 10^7$ 和 $1.01 \times 10^7$ 。这些结果表明在促进蛋白质合成方面, $\Omega$ -10A要优于 $\Omega$ -8A和 $\Omega$ -6A。没有加入任何DNA片段的酵母体外蛋白质合成体系作为阴性对照,相对光单位值仅有190.33。

[0184] 本发明结果表明,乳酸克鲁维酵母特异性的kl-Kozak序列能够显著增强酵母体外蛋白质合成体系产生蛋白质的效率, $\Omega$ -10A-Fluc DNA片段的相对光单位值是 $\Omega$ -Fluc的22.3倍,达到了 $2.72 \times 10^7$ ,表明 $\Omega$ -10A能够增强22.3倍的蛋白质合成。

[0185] 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。

[0186] 参考文献:

[0187] 1.Dever,T.E.,et al.(2016)Mechanism and Regulation of Protein Synthesis in *Saccharomyces cerevisiae*.Genetics 203,65-107.

[0188] 2.Kyrieleis,O.J.,et al.(2014)Crystal structure of vaccinia virus mRNA capping enzyme provides insights into the mechanism and evolution of the

capping apparatus. *Structure* 22,452-465.

[0189] 3. Katzen, F., et al. (2005) The past, present and future of cell-free protein synthesis. *Trends in biotechnology* 23,150-156.

[0190] 4. Anastasina, M., et al. (2014) A technique to increase protein yield in a rabbit reticulocyte lysate translation system. *BioTechniques* 56,36-39.

[0191] 5. Spohner, S.C., et al. (2016) *Kluyveromyces lactis*: An emerging tool in biotechnology. *Journal of biotechnology* 222,104-116.

## 序列表

&lt;110&gt; 康码(上海)生物科技有限公司

&lt;120&gt; 一种用于高通量体外蛋白质合成的DNA元件

&lt;130&gt; P2017-1191

&lt;160&gt; 18

&lt;170&gt; PatentIn version 3.5

&lt;210&gt; 1

&lt;211&gt; 1993

&lt;212&gt; DNA

&lt;213&gt; 人工序列

&lt;400&gt; 1

```
cgcgaaatta atacgactca ctataggggt atttttacaa caattaccaa caacaacaaa 60
caacaacaaa cattacaatt actattttaca attacaaaaa 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
cgcgacacgac atttataatg aacgtgaatt gctcaacagt atgaacattt cgcagcctac 480
cgtagtgttt gtttccaaaa aggggttgca aaaaattttg aacgtgcaaa aaaaattacc 540
aataatccag aaaattatta tcatggattc taaaacggat taccagggat ttcagtcgat 600
gtacacgttc gtcacatctc atctacctcc cggttttaat gaatacgatt ttgtaccaga 660
gtcctttgat cgtgacaaaa caattgcact gataatgaat tcctctggat ctactgggtt 720
acctaagggg gtggcccttc cgcatagaac tgcctgcgtc agattctcgc atgccagaga 780
tcctattttt ggcaatcaaa tcattccgga tactgcgatt ttaagtgttg ttccattcca 840
tcacggtttt ggaatgttta ctacactcgg atatttgata tgttgatttc gagtcgtctt 900
aatgtataga tttgaagaag agctgttttt acgatccctt caggattaca aaattcaaag 960
tgcgttgcta gtaccaaccc tattttcatt cttegccaaa agcactctga ttgacaaata 1020
cgattttatc aatttacacg aaattgcttc tgggggcgca cctctttcga aagaagtcgg 1080
ggaagcgggt gcaaaacgct tccatcttcc agggatacga caaggatatg ggctcactga 1140
gactacatca gctattctga ttacaccgga gggggatgat aaaccgggcg cggtcggtaa 1200
agttgttcca ttttttgaag cgaaggttgt ggatctggat accgggaaaa cgctgggcgt 1260
taatcagaga ggcgaattat gtgtcagagg acctatgatt atgtccggtt atgtaaacia 1320
tccggaagcg accaacgcct tgattgacaa ggatggatgg ctacattctg gagacatagc 1380
ttactgggac gaagacgaac acttcttcat agttgaccgc ttgaagtctt taattaaata 1440
caaaggatat cagggtggccc ccgctgaatt ggaatcgata ttgttacaac accccaacat 1500
cttcgacgcg ggcggtggcag gtcttcccga cgatgacgcc ggtgaacttc ccgccgccgt 1560
tgttgttttg gagcacggaa agacgatgac ggaaaaagag atcgtggatt acgtcgccag 1620
```

tcaagtaaca accgcgaaaa agttgcgcgg aggagttgtg tttgtggacg aagtaccgaa 1680  
 aggtcttacc ggaaaactcg acgcaagaaa aatcagagag atcctcataa aggccaagaa 1740  
 gggcggaag tccaaattgg tttaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 1800  
 aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaactcga 1860  
 gcaccaccac caccaccact gagatccggc tgctaacaaa gcccgaagg aagctgagtt 1920  
 ggctgctgcc accgctgagc aataactagc ataaccctt ggggcctcta aacgggtctt 1980  
 gaggggtttt ttg 1993

<210> 2

<211> 1991

<212> DNA

<213> 人工序列

<400> 2

cgcgaaatta atacgactca ctataggggt atttttacaa caattaccaa caacaacaaa 60  
 caacaacaaa cattacaatt actatttaca attacaaaaa aaaaatgtct gaagacgcca 120  
 aaaacataaa gaaaggcccg gcgccattct atcctctaga ggatggaacc gctggagagc 180  
 aactgcataa ggctatgaag agatacgccc tggttcctgg aacaattgct tttacagatg 240  
 cacatatcga ggtgaacatc acgtacgcgg aataactcga aatgtccgtt cggttggcag 300  
 aagctatgaa acgatatggg ctgaatacaa atcacagaat cgtcgtatgc agtgaaaact 360  
 ctcttcaatt ctttatgccg gtgttgggcg cgttatttat cggagttgca gttgcgcccc 420  
 cgaacgacat ttataatgaa cgtgaattgc tcaacagtat gaacatttcg cagcctaccg 480  
 tagtgtttgt ttccaaaaag gggttgcaaa aaattttgaa cgtgcaaaaa aaattaccaa 540  
 taatccagaa aattattatc atggattcta aaacggatta ccagggattt cagtcgatgt 600  
 acacgttcgt cacatctcat ctacctcccg gttttaatga atacgatttt gtaccagagt 660  
 cctttgatcg tgacaaaaca attgcactga taatgaattc ctctggatct actgggttac 720  
 ctaagggtgt ggcccttccg catagaactg cctgcgtcag attctcgcgt gccagagatc 780  
 ctattttttg caatcaaate attccggata ctgcgatttt aagtgttggt ccattccatc 840  
 acggtttttg aatgtttact acactcggat atttgatatg tggatttcga gtcgtcttaa 900  
 tgtatagatt tgaagaagag ctgtttttac gatcccttca ggattacaaa attcaaagtg 960  
 cgttgctagt accaacccta tttcattct tcgcaaaaag cactctgatt gacaaatacg 1020  
 atttatctaa tttacacgaa attgcttctg ggggcgcacc tctttcgaaa gaagtcgggg 1080  
 aagcggttgc aaaacgcttc catcttcag ggatacgaca aggatatggg ctactgaga 1140  
 ctacatcagc tattctgatt acacccgagg gggatgataa accgggcgcg gtcggtaaag 1200  
 ttgttccatt ttttgaagcg aaggttgtgg atctggatac cgggaaaacg ctgggcgtta 1260  
 atcagagagg cgaattatgt gtcagaggac ctatgattat gtccggttat gtaacaatc 1320  
 cggaagcgac caacgccttg attgacaagg atggatggct acattctgga gacatagctt 1380  
 actgggacga agacgaacac ttcttcatag ttgaccgctt gaagtcttta attaaatata 1440  
 aaggatatca ggtggccccc gctgaattgg aatcgatatt gttacaacac cccaacatct 1500  
 tcgacgcggg cgtggcaggt cttcccgcag atgacgccgg tgaacttccc gccgccgttg 1560  
 ttgttttggg gcacggaaaag acgatgacgg aaaaagagat cgtggattac gtcgccagtc 1620

aagtaacaac cgcgaaaaag ttgcgcggag gagttgtgtt tgtggacgaa gtaccgaaag 1680  
 gtcttaccgg aaaactcgac gcaagaaaaa tcagagagat cctcataaag gccagaag 1740  
 gcggaaagtc caaattgggt taaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 1800  
 aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaactcgagc 1860  
 accaccacca ccaccactga gatccggctg ctaacaaagc ccgaaaggaa gctgagttgg 1920  
 ctgctgccac cgctgagcaa taactagcat aaccccttgg ggcctctaaa cgggtcttga 1980  
 ggggtttttt g 1991

<210> 3

<211> 1989

<212> DNA

<213> 人工序列

<400> 3

cgcgaaatta atacgactca ctataggggt atttttacaa caattaccaa caacaacaaa 60  
 caacaacaaa cattacaatt actattttaca attacaaaaa aatgtctga agacgcaaaa 120  
 aacataaaga aaggcccggc gccattctat cctctagagg atggaaccgc tggagagcaa 180  
 ctgcataagg ctatgaagag atacgccttg gttcctggaa caattgcttt tacagatgca 240  
 catatcgagg tgaacatcac gtacgcggaa tacttcgaaa tgtccgttcg gttggcagaa 300  
 gctatgaaac gatatgggct gaatacaaat cacagaatcg tcgtatgcag tgaaaactct 360  
 cttcaattct ttatgccggt gttgggcgcg ttattttatcg gagttgcagt tgcgcccgcg 420  
 aacgacattt ataataaacg tgaattgctc aacagtatga acatttcgca gcctaccgta 480  
 gtgtttgttt ccaaaaaggg gttgcaaaaa attttgaacg tgcaaaaaaa attaccaata 540  
 atccagaaaa ttattatcat ggattctaaa acggattacc agggatttca gtcgatgtac 600  
 acgttcgtca catctcatct acctcccggg tttaatgaat acgattttgt accagagtcc 660  
 tttgatcgtg acaaaacaat tgcactgata atgaattcct ctggatctac tgggttacct 720  
 aagggtgtgg cccttcgca tagaactgcc tgcgtcagat tctcgcatgc cagagatcct 780  
 atttttggca atcaaatcat tccggatact gcgattttaa gtgttgttcc attccatcac 840  
 gggttttgaa tgtttactac actcggatat ttgatattgt gatttcgagt cgtcttaatg 900  
 tatagatttg aagaagagct gtttttacga tcccttcagg attacaaaat tcaaagtgcg 960  
 ttgctagtac caacctatt ttcattcttc gccaaaagca ctctgattga caaatagat 1020  
 ttatctaatt tacacgaaat tgcttctggg ggcgacctc tttcgaaaga agtcggggaa 1080  
 gcggttgcaa aacgcttcca tcttcagggt atacgacaag gatatgggct cactgagact 1140  
 acatcagcta ttctgattac acccgagggg gatgataaac cgggcgcggg cggtaaagtt 1200  
 gttccatttt ttgaagcgaa ggttgtggat ctggataacc ggaaaacgct gggcgttaat 1260  
 cagagaggcg aattatgtgt cagaggacct atgattatgt ccggttatgt aaacaatccg 1320  
 gaagcgacca acgccttgat tgacaaggat ggatggctac attctggaga catagcttac 1380  
 tgggacgaag acgaacactt cttcatagtt gaccgcttga agtctttaat taaatacaaa 1440  
 ggatatcagg tggccccgcg tgaattggaa tcgatattgt tacaacaccc caacatcttc 1500  
 gacgcgggcg tggcaggctc tcccgcgat gacgccggtg aacttcccgc cgccgttggt 1560  
 gttttggagc acggaaagac gatgacggaa aaagagatcg tggattacgt cgccagtcaa 1620

gtaacaaccg cgaaaaagtt gcgcggagga gttgtgtttg tggacgaagt accgaaaggt 1680  
 cttaccggaa aactcgacgc aagaaaaatc agagagatcc tcataaaggc caagaagggc 1740  
 ggaaagtcca aattggttta aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 1800  
 aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa actcgagcac 1860  
 caccaccacc accactgaga tccggctgct aacaaagccc gaaaggaagc tgagttggct 1920  
 gctgccaccg ctgagcaata actagcataa ccccttgggg cctctaaacg ggtcttgagg 1980  
 gggttttttg 1989

<210> 4

<211> 1983

<212> DNA

<213> 人工序列

<400> 4

cgcgaaatta atacgactca ctataggggt atttttaca caattacca caacaacaaa 60  
 caacaacaaa cattacaatt actatttaca attacaatgt ctgaagacgc caaaaacata 120  
 aagaaaggcc cggcgccatt ctatcctcta gaggatggaa ccgctggaga gcaactgcat 180  
 aaggctatga agagatacgc cctggttcct ggaacaattg cttttacaga tgcacatata 240  
 gaggtgaaca tcacgtacgc ggaatacttc gaaatgtccg ttcggttggc agaagctatg 300  
 aaacgatatg ggctgaatac aaatcacaga atcgtcgtat gcagtgaaaa ctctcttcaa 360  
 ttctttatgc cgggtgttggg cgcggttattt atcggagttg cagttgcgcc cgcgacgac 420  
 atttataatg aacgtgaatt gctcaacagt atgaacattt cgagcctac cgtagtgttt 480  
 gtttccaaaa aggggttgca aaaaattttg aacgtgcaaa aaaaattacc aataatccag 540  
 aaaattatta tcatggattc taaaacggat taccagggat ttcagtcgat gtacacgttc 600  
 gtcacatctc atctacctcc cggttttaat gaatacgatt ttgtaccaga gtcctttgat 660  
 cgtgacaaaa caattgcact gataatgaat tcctctggat ctactgggtt acctaagggt 720  
 gtggcccttc cgcatagaac tgcctgcgtc agattctcgc atgccagaga tcctatTTTT 780  
 ggcaatcaaa tcattccgga tactgcgatt ttaagtgttg ttccattcca tcacggtttt 840  
 ggaatgttta ctacactcgg atatttgata tgtggatttc gagtcgtctt aatgtataga 900  
 tttgaagaag agctgttttt acgatccctt caggattaca aaattcaaag tgcgttgcta 960  
 gtaccaaccc tattttcatt cttcgccaaa agcactctga ttgacaaata cgatttatct 1020  
 aatttacacg aaattgcttc tgggggcgca cctctttcga aagaagtcgg ggaagcggtt 1080  
 gcaaacgct tccatcttcc agggatacga caaggatatg ggctcactga gactacatca 1140  
 gctattctga ttacaccgga gggggatgat aaaccgggcg cggtcggtaa agttgttcca 1200  
 ttttttgaag cgaaggttgt ggatctggat accgggaaaa cgctgggcgt taatcagaga 1260  
 ggcgaaattat gtgtcagagg acctatgatt atgtccggtt atgtaaaca tccggaagcg 1320  
 accaacgcct tgattgacaa ggatggatgg ctacattctg gagacatagc ttactgggac 1380  
 gaagacgaac acttcttcat agttgaccgc ttgaagtctt taattaaata caaaggatat 1440  
 cagggtggccc ccgctgaatt ggaatcgata ttgttacaac accccaacat cttcgacgcg 1500  
 ggcggtggcag gtcttccgga cgatgacgcc ggtgaacttc ccgccgccgt tgttgttttg 1560  
 gagcacggaa agacgatgac ggaaaaagag atcgtggatt acgtcgccag tcaagtaaca 1620

accgcgaaaa agttgcgcgg aggagttgtg tttgtggacg aagtaccgaa aggtcttacc 1680  
 ggaaaaactcg acgcaagaaa aatcagagag atcctcataa aggccaagaa gggcggaaag 1740  
 tccaaattgg tttaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 1800  
 aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaactcga gcaccaccac 1860  
 caccaccact gagatccggc tgctaacaaa gcccgaaagg aagctgagtt ggctgctgcc 1920  
 accgctgagc aataactagc ataaccctt ggggcctcta aacgggtctt gaggggtttt 1980  
 ttg 1983

<210> 5

<211> 16

<212> RNA

<213> 人工序列

<400> 5

aaaaaaaaaa augucu 16

<210> 6

<211> 717

<212> DNA

<213> 人工序列

<400> 6

atggtgagca agggcgagga gctgttcacc ggggtggtgc ccatcctggt cgagctggac 60  
 ggcgacgtaa acggccacaa gttcagcgtg tccggcgagg gcgagggcga tgccacctac 120  
 ggcaagctga ccctgaagtt catctgcacc accggcaage tgcccgtgcc ctggcccacc 180  
 ctcgtgacca ccctgacctc cggcgtgcag tgcttcagcc gctaccccga ccacatgaag 240  
 cagcacgact tcttcaagtc cgccatgccc gaaggctacg tccaggagcg caccatcttc 300  
 ttcaaggacg acggcaacta caagaccgc gccgaggtga agttcgaggg cgacaccctg 360  
 gtgaaccgca tcgagctgaa gggcatcgac ttcaaggagg acggcaacat cctggggcac 420  
 aagctggagt acaactacaa cagccacaac gtctatatca tggccgacaa gcagaagaac 480  
 ggcacatcaagg tgaacttcaa gatccgccac aacatcgagg acggcagcgt gcagctcgcc 540  
 gaccactacc agcagaacac ccccatcggc gacggccccg tgctgctgcc cgacaaccac 600  
 tacctgagca ccagtcgcgc cctgagcaaa gaccccaacg agaagcgcca tcacatggtc 660  
 ctgctggagt tcgtgaccgc cgccgggatc actctcgga tggacgagct gtacaag 717

<210> 7

<211> 717

<212> DNA

<213> 人工序列

<400> 7

atggtgagca agggcgagga gctgttcacc ggggtggtgc ccatcctggt cgagctggac 60  
 ggcgacgtaa acggccacaa gttcagcgtg tccggcgagg gcgagggcga tgccacctac 120  
 ggcaagctga ccctgaagct gatctgcacc accggcaage tgcccgtgcc ctggcccacc 180  
 ctcgtgacca ccctgggcta cggcctgcag tgcttcgccc gctaccccga ccacatgaag 240

cagcacgact tcttcaagtc cgccatgccc gaaggctacg tccaggagcg caccatcttc 300  
 ttcaaggacg acggcaacta caagacccgc gccgaggtga agttcgaggg cgacaccctg 360  
 gtgaaccgca tcgagctgaa gggcatcgac ttcaaggagg acggcaacat cctggggcac 420  
 aagctggagt acaactacaa cagccacaac gtctatatca ccgccgacaa gcagaagaac 480  
 ggcatcaagg ccaacttcaa gatccgccac aacatcgagg acggcggcgt gcagctcgcc 540  
 gaccactacc agcagaacac ccccatcggc gacggccccg tgctgctgcc cgacaaccac 600  
 tacctgagct accagtccgc cctgagcaaa gaccccaacg agaagcgcga tcacatggtc 660  
 ctgctggagt tcgtgaccgc cgccgggata actctcggca tggacgagct gtacaag 717

<210> 8

<211> 3048

<212> DNA

<213> 人工序列

<400> 8

atggctgttt tacaacgtcg tgactgggaa aaccctggcg ttacceaaact taatcgccctt 60  
 gcagcacatc cccctttcgc cagctggcgt aatagcgaag aggcccgac cgatcgccct 120  
 tccaacagt tgcgcagcct gaatggcgaa tggcgctttg cctggtttcc ggcaccagaa 180  
 gcggtgccgg aaagctggct ggagtgcgat ctctctgagg ccgatactgt cgtcgteccc 240  
 tcaaactggc agatgcacgg ttacgatgcg cccatctaca ccaacgtgac ctatcccatt 300  
 acggtcaatc cgccgtttgt tcccacggag aatccgacgg gttgttactc gtcacattt 360  
 aatgttgatg aaagctggct acaggaaggc cagacgcgaa ttatTTTTga tggcgTTaac 420  
 tcggcgTTTT atctgtggtg caacgggcgc tgggtcggtt acggccagga cagtcgtttg 480  
 ccgtctgaat ttgacctgag cgcattttta cgcgccggag aaaaccgcct cgcggtgatg 540  
 gtgctgcgct ggagtgcagg cagttatctg gaagatcagg atatgtggcg gatgagcggc 600  
 attttccgtg acgtctcgtt gctgcataaa ccgactacac aaatcagcga tttccatgtt 660  
 gccactcgct ttaatgatga ttccagccgc gctgtactgg aggctgaagt tcagatgtgc 720  
 ggcgagttgc gtgactacct acgggtaaca gtttctttat ggccagggtga aacgcaggtc 780  
 gccagcggca ccgcgccttt cggcggtgaa attatcgatg agcgtgggtg ttatgccgat 840  
 cgcgtcacac tacgtctgaa cgtcgaaaac ccgaaactgt ggagcgccga aatccccgaat 900  
 ctctatcgtg cgggtggtga actgcacacc gccgacggca cgctgattga agcagaagcc 960  
 tgcgatgtcg gtttccgcga ggtgcggatt gaaaatggtc tgctgctgct gaacggcaag 1020  
 ccgttgctga ttccaggcgt taaccgtcac gagcatcatc ctctgcatgg tcaggtcattg 1080  
 gatgagcaga cgatgggtga ggatatcctg ctgatgaagc agaacaactt taacgccgtg 1140  
 cgctgttcgc attatccgaa ccatccgctg tggtaacgc tgtgcgaccg ctacggcctg 1200  
 tatgtggtgg atgaagccaa tattgaaacc cacggcatgg tgccaatgaa tcgtctgacc 1260  
 gatgatccgc gctggctacc ggcatgagc gaacgcgtaa cgcgatggg gcagcgcgat 1320  
 cgtaatcacc cgagtgtgat catctggctg ctgggggaatg aatcaggcca cggcgctaata 1380  
 cacgacgcgc tgtatcgctg gatcaaatct gtcgacett cccgccccgt gcagtatgaa 1440  
 ggccggcgag ccgacaccac ggccaccgat attatttgc cgatgtacgc gcgcgtggat 1500  
 gaagaccage ccttcccggc tgtgccgaaa tggatccatca aaaaatggct ttcgtacct 1560

ggagagacgc gcccgctgat cctttgcgaa tacgcccacg cgatgggtaa cagtcttggc 1620  
 ggtttcgcta aatactggca ggcgtttcgt cagtatcccc gtttacaggg cggcttcgtc 1680  
 tgggactggg tggatcagtc gctgattaaa tatgatgaaa acggcaaccc gtggctcggct 1740  
 tacggcgggtg attttggcga tacgccgaac gatcgccagt tctgtatgaa cggctctggc 1800  
 tttgccgacc gcacgccgca tccagcgcgtg acggaagcaa aacaccagca gcagtttttc 1860  
 cagttccggtt tatccgggca aaccatcgaa gtgaccagcg aatacctgtt ccgtcatagc 1920  
 gataacgagc tcctgcactg gatgggtggcg ctggatggta agccgctggc aagcggtgaa 1980  
 gtgcctctgg atgtcgctcc acaaggtaaa cagttgattg aactgcctga actaccgag 2040  
 ccggagagcg ccgggcaact ctggctcaca gtacgcgtag tgcaaccgaa cgcgaccgca 2100  
 tggtcagaag ccggacacat cagcgcctgg cagcagtggc gtctggctga aaacctcagc 2160  
 gtgacactcc ccgccgcgtc ccacgccatc ccgcactga ccaccagcga aatggatttt 2220  
 tgcacgagc tgggtaataa gcgttggcaa tttaaccgcc agtcaggctt tctttcacag 2280  
 atgtggattg gcgataaaaa acaactgctg acgccgctgc gcgatcagtt caccctgca 2340  
 ccgctggata acgacattgg cgtaagtga gcgaccgca ttgaccctaa cgctggggtc 2400  
 gaacgctgga aggcggcggg ccattaccag gccgaagcag cgttgttgca gtgcacggca 2460  
 gatacacttg ctgatgcggt gctgattacg accgctcacg cgtggcagca tcaggggaaa 2520  
 acctatttta tcagccgga aacctaccgg attgatggta gtggtcaaatt ggcgattacc 2580  
 gttgatgttg aagtggcgag cgataaccg catccggcgc ggattggcct gaactgccag 2640  
 ctggcgcagg tagcagagcg ggtaaaactgg ctcggttag ggccgcaaga aaactatccc 2700  
 gaccgcctta ctgccgcctg ttttgaccgc tgggatctgc cattgtcaga catgtatacc 2760  
 ccgtacgtct tcccagcgca aaacggtctg cgctgcggga cgcgcgaatt gaattatggc 2820  
 ccacaccagt ggcgcgcgca cttccagttc aacatcagcc gctacagtca acagcaactg 2880  
 atggaaacca gccatcgcca tctgctgcac gcggaagaag gcacatggct gaatatcgac 2940  
 ggtttccata tggggattgg tggcgacgac tcctggagcc cgtcagtatc ggcggaattc 3000  
 cagctgagcg ccggtcgcta ccattaccag ttggtctggt gtcaaaaa 3048

<210> 9

<211> 1935

<212> DNA

<213> 人工序列

<400> 9

atggcggccg tgcaggcggc cgaggtgaaa gtggatggca gcgagccgaa actgagcaag 60  
 aatgagctga agagacgcct gaaagctgag aagaaagtag cagagaagga ggccaaacag 120  
 aaagagctca gtgagaaaca gctaagccaa gccactgctg ctgccaccaa ccacaccact 180  
 gataatgggtg tgggtcctga ggaagagagc gtggacccaa atcaatacta caaaatccgc 240  
 agtcaagcaa ttcacagct gaaggtcaat ggggaagacc catacccaca caagttccat 300  
 gtagacatct cactcactga cttcatccaa aaatatagtc acctgcagcc tggggatcac 360  
 ctgactgaca tcaccttaaa ggtggcaggt aggatccatg ccaaaagagc ttctggggga 420  
 aagctcatct tctatgatct tcgaggagag ggggtgaagt tgcaagtcatt ggccaattcc 480  
 agaaattata aatcagaaga agaatttatt catattaata acaaactgcg tcggggagac 540

ataattggag ttcaggggaa tcctggtaaa accaagaagg gtgagctgag catcattccg 600  
 tatgagatca cactgctgtc tccctgtttg catatgttac ctcatcttca ctttggcctc 660  
 aaagacaagg aaacaaggta tcgccagaga tacttggact tgatcctgaa tgactttgtg 720  
 aggcgaaat ttatcatccg ctctaagatc atcacatata taagaagttt cttagatgag 780  
 ctgggattcc tagagattga aactcccatg atgaacatca tcccaggggg agccgtggcc 840  
 aagcctttca tcacttatca caacgagctg gacatgaact tatatatgag aattgctcca 900  
 gaactctatc ataagatgct tgtggttggg ggcacgacc gggtttatga aattggacgc 960  
 cagttccgga atgaggggat tgatttgacg cacaatcctg agttcaccac ctgtgagttc 1020  
 tacatggcct atgcagacta tcacgatctc atggaaatca cggagaagat ggtttcaggg 1080  
 atggtgaagc atattacagg cagttacaag gtcacctacc acccagatgg cccagagggc 1140  
 caagcctacg atgttgactt caccaccacc ttccggcgaa tcaacatggg agaagagctt 1200  
 gagaaagccc tggggatgaa gctgccagaa acgaacctct ttgaaactga agaaactcgc 1260  
 aaaattcttg atgatatctg tgtggcaaaa gctgttgaat gccctccacc tcggaccaca 1320  
 gccaggetcc ttgacaagct tgttggggag ttcttggag tgacttgcat caatcctaca 1380  
 ttcatctgtg atcaccaca gataatgagc cttttggcta aatggcaccg ctctaaagag 1440  
 ggtctgactg agcgctttga gctgtttgtc atgaagaaag agatatgcaa tgcgtatact 1500  
 gagctgaatg atcccatgcg gcagcggcag ctttttgaag aacaggccaa ggccaaggct 1560  
 gcaggtgatg atgaggccat gttcatagat gaaaacttct gtactgccct ggaatatggg 1620  
 ctgcccccca cagctggctg gggcatgggc attgatcgag tcgccatgtt tctcacggac 1680  
 tccaacaaca tcaaggaagt acttctgttt cctgccatga aaccgaaga caagaaggag 1740  
 aatgtagcaa ccactgatac actggaaagc acaacagttg gcatttctgt ctagaaaata 1800  
 ataattgcaa gttgtataac tcaggcgtct ttgcatttct gcgaaagatc aaggtctgca 1860  
 aggggaattct tgtgtgctgc tttccatttg acaccgcagt tctgttcagc catcagaaga 1920  
 gagacaagga attaa 1935

<210> 10

<211> 3531

<212> DNA

<213> 人工序列

<400> 10

atggcggaaa gaaaaggaac agccaaagtg gactttttga agaagattga gaaagaaatc 60  
 caacagaaaat gggatactga gagagtgttt gaggtcaatg catctaattt agagaaacag 120  
 accagcaagg gcaagtatct tgtaaccttc ccatatccat atatgaatgg acgccttcat 180  
 ttgggacaca cgttttcttt atccaaatgt gagtttctg tagggtacca gcgattgaaa 240  
 ggaaaatgtt gtctgtttcc ctttggcctg cactgtactg gaatgcctat taaggcatgt 300  
 gctgataagt tgaaaagaga aatagagctg tatggttgcc cccctgattt tccagatgaa 360  
 gaagaggaag aggaagaaac cagtgttaaa acagaagata taataattaa ggataaagct 420  
 aaaggaaaaa agagtaaagc tgctgctaaa gctggatctt ctaaatacca gtggggcatt 480  
 atgaaatccc ttggcctgtc tgatgaagag atagtaaaat tttctgaagc agaacattgg 540  
 cttgattatt tcccgcact ggctattcag gatttaaaaa gaatgggttt gaaggtagac 600

tggcgtcggt ccttcatcac cactgatgtt aatccttact atgattcatt tgtcagatgg 660  
 caatttttaa cattaagaga aagaaacaaa attaaatttg ggaagcggta tacaatttac 720  
 tctccgaaag atggacagcc ttgcatggat catgatagac aaactggaga ggggtgttga 780  
 cctcaggaat atactttact caaattgaag gtgcttgagc catacccatc taaattaagt 840  
 ggcctgaaag gtaaaaatat tttcttgggt gctgctactc tcagacctga gaccatgttt 900  
 gggcagacaa attgttgggt tcgtcctgat atgaagtaca ttggatttga gacggtgaat 960  
 ggtgatataat tcactctgtac ccaaaaagca gccaggaata tgtcatacca gggctttacc 1020  
 aaagacaatg gcgtgggtgcc tgttgtaag gaattaatgg gggaggaaat tcttgggtga 1080  
 tcactttctg cacctttaac atcatacaag gtgatctatg ttctcccaat gctaactatt 1140  
 aaggaggata aaggcactgg tgtggttaca agtgttccct ccgactcccc tgatgatatt 1200  
 gctgccctca gagacttgaa gaaaaagcaa gccttacgag caaatatgg aattagagat 1260  
 gacatgggtc tgccatttga gccggtgcca gtcattgaaa tcccagggtt tggaatctt 1320  
 tctgctgtaa ccatttgtga tgagtgaag attcagagcc agaatgaccg gaaaaaactt 1380  
 gcagaagcaa aggagaagat atatctaaaa ggattttatg aggttatcat gttggtggat 1440  
 ggattttaaag gacagaaggt tcaagatgta aagaagacta ttcagaaaaa gatgattgac 1500  
 gctggagatg cacttattta catggaacca gagaaacaag tgatgtccag gtcgtcagat 1560  
 gaatgtgttg tggctctgtg tgaccagtgg tacttggatt atggagaaga gaattggaag 1620  
 aaacagacat ctcatgtctt gaagaacctg gaaacattct gtgaggagac caggaggaat 1680  
 tttgaagcca ccttaggttg gctacaagaa catgcttgc caagaactta tggcttaggc 1740  
 actcacctgc cttaggatga gcagtggctg attgaatcac tttctgactc cactatttac 1800  
 atggcatttt acacagtgc acacctattg caggggggta acttgcattg acaggcagag 1860  
 tctccgctgg gcattagacc gcaacagatg accaaggaag tttgggatta tgttttcttc 1920  
 aaggaggctc catttcctaa gactcagatt gcaaaggaaa aattagatca gttaaagcag 1980  
 gagtttgaat tctggtatcc tgttgatctt cgcgtctctg gcaaggatct tgttccaaat 2040  
 catctttcat attaccttta taatcatgtg gctatgtggc cggaacaaag tgacaaatgg 2100  
 cctacagctg tgagagcaaa tggacatctc ctctgaact ctgagaagat gtcaaaatcc 2160  
 acaggcaact tcctcacttt gacccaagct attgacaaat tttcagcaga tggaatgcgt 2220  
 ttggctctgg ctgatgctgg tgacactgta gaagatgcca actttgtgga agccatggca 2280  
 gatgcaggta ttctccgtct gtacacctgg gtagagtggg tgaaagaaat ggttgccaac 2340  
 tgggacagcc taagaagtgg tctgcccagc actttcaatg atagagtttt tgccagtga 2400  
 ttgaatgcag gaattataaa aacagatcaa aactatgaaa agatgatgtt taaagaagct 2460  
 ttgaaaacag ggttttttga gtttcaggcc gcaaaagata agtaccgtga attggctgtg 2520  
 gaagggatgc acagagaact tgtgttccgg tttattgaag ttcagacact tctctcgtct 2580  
 ccattctgtc cacatttgtg tgagcacatc tggacactcc tgggaaagcc tgactcaatt 2640  
 atgaatgctt catggcctgt ggcaggtcct gttaatgaag ttttaataca ctctcacag 2700  
 tatcttatgg aagtaacaca tgacctaga ctacgactca agaactatat gatgccagct 2760  
 aaagggaaga agactgacaa acaaccctg cagaagccct cacattgcac catctatgtg 2820  
 gcaagaact atccacctg gcaacatacc accctgtctg ttctacgtaa acactttgag 2880  
 gccataacg gaaaactgcc tgacaacaaa gtcattgcta gtgaactagg cagtatgcc 2940

gaactgaaga aatacatgaa gaaagtcattg ccatttgttg ccatgattaa ggaaaatctg 3000  
 gagaagatgg ggccctcgat tctggatttg caattagaat ttgatgaaaa ggctgtgctt 3060  
 atggagaata tagtctatct gactaattcg cttgagctag aacacataga agtcaagttt 3120  
 gcctccgaag cagaagataa aatcagggaa gactgctgtc ctgggaaacc acttaatgtt 3180  
 tttagaatag aacctgggtg gtccgtttct ctggtgaatc cccagccatc caatggccac 3240  
 ttctcaacca aaattgaaat caggcaagga gataactgtg attccataat caggcgttta 3300  
 atgaaaaatga atcgaggaat taaagacctt tccaaagtga aactgatgag atttgatgat 3360  
 ccactgttgg ggccctcgacg agttcctgtc ctgggaaagg agtacaccga gaagaccccc 3420  
 atttctgagc atgtctgttt caatgtggac ctcatgagca agaaaattca tctgactgag 3480  
 aatgggataa ggggtggatat tggcgataca ataattctatc tggttcatta a 3531

<210> 11

<211> 1017

<212> DNA

<213> 人工序列

<400> 11

atggctgaca agaagattag gatcggaatc aacggattcg gaagaattgg tcgtttggtt 60  
 gctagagttg ttctccagag ggacgatgtt gagctcgtcg ctgtcaacga ccccttcac 120  
 actactgagt acatgacctt catgttcaag tacgacagtg ttcacggtca atggaaacac 180  
 aatgaactca agatcaagga tgagaagacc cttctcttcg gtgagaagcc agtcaactgtt 240  
 ttccgcatca ggaaccctga ggatatccca tgggccgagg ctggagctga ctacgttggtt 300  
 gagtctactg gtgtcttcac tgacaaagac aaggtgcag ctcacttgaa ggggtggtgcc 360  
 aagaaggttg ttatctctga acccagcaaa gacgtccaa tgtttgttgt tggtgtcaac 420  
 gagcacgaat acaagtccga ccttgacatt gtctccaacg ctagctgcac cactaactgc 480  
 cttgtcctcc ttgccaaggt tatcaatgac agatttggaa ttgttgaggg tcttatgact 540  
 acagtccact caatcactgc tactcagaag actgttgatg ggcccttcaat gaaggactgg 600  
 agagggtggaa gagctgcttc attcaacatt attcccagca gcactggagc tgccaaggct 660  
 gtcggaaagg tgcttccagc tcttaacgga aagttgactg gaatgtcttt ccgtgtccca 720  
 accgttgatg tctcagttgt tgaccttact gtcagactcg agaaagctgc tacctacgaa 780  
 gaaatcaaaa aggctatcaa ggaggaatcc gaaggcaaac tcaagggaat ccttgatac 840  
 accgaggatg atgttgtctc aactgacttc gttggcgaca acaggtcgag catttttgac 900  
 gccaaaggctg gaattgcatt gagcgacaag tttgtgaaat tgggtgcatg gtacgacaac 960  
 gaatgggggtt acagttcccg tgtggtcgac ttgatcgtec acatgtcaaa ggcctaa 1017

<210> 12

<211> 1584

<212> DNA

<213> 人工序列

<400> 12

atgtcggaca gtcgggaccc agccagcgac cagatgaagc agtggaagga gcagcgggcc 60  
 tcgcagagac cagatgtcct gaccaccgga ggcgggaacc caataggaga taaacttaat 120

atcatgaccg cgggggtcccg agggcccctc ctcgttcagg atgtggtttt cactgacgag 180  
 atggcacact ttgacagaga gcggattcct gagagagtgg tacacgcaa aggagcaggt 240  
 gcttttggat actttgaggt caccacagat atcaccagat actccaaggc aaaggtgttt 300  
 gagcatattg gaaagaggac ccctattgcc gttcgattct ccacagtcgc tggagagtca 360  
 ggctcagctg acacagttcg tgaccctcgg gggtttcag tgaaatttta cactgaagat 420  
 ggtaactggg atcttgtggg aaacaacacc cctattttct tcatcaggga tgccatattg 480  
 tttccatcct ttatccatag ccagaagaga aaccacaga ctcacctgaa ggatcctgac 540  
 atgggtctggg acttctggag tcttcgtccc gagtctctcc atcaggtttc tttcttggtc 600  
 agtgaccgag ggattcccga tggtcaccgg cacatgaatg gctatggatc acacaccttc 660  
 aagtttggtta atgcagatgg agaggcagtc tattgcaagt tccattacaa gaccgaccag 720  
 ggcatcaaaa acttgccctgt tggagaggca ggaaggcttg ctcaggaaga tccggattat 780  
 ggccctccgag atcttttcaa tgccatcgcc aatggcaatt acccgtcctg gacgttttac 840  
 atccaggtca tgacttttaa ggaggcagaa actttcccat ttaatccatt tgatctgacc 900  
 aaggtttggc ctcaacaagga ctacctctt ataccagttg gcaaactggg tttaaacaaa 960  
 aatccagtta attactttgc tgaagttgaa cagatggctt ttgacccaag caatatgccc 1020  
 cctggcatcg agcccagccc tgacaaaatg cttcagggcc gcctttttgc ctaccggac 1080  
 actcaccgcc accgcctggg acccaactat ctgcagatac ctgtgaactg tccctaccgc 1140  
 gctcgagtgg ccaactacca gcgtgatggc cccatgtgca tgcatgacaa ccagggtggt 1200  
 gccccaaact attaccccaa cagcttcagc gcaccagagc agcagcgctc agccctggag 1260  
 cacagcgctc agtgcgctgt agatgtgaaa cgcttcaaca gtgctaata agacaatgtc 1320  
 actcaggtgc ggacattcta cacaagggtg ttgaatgagg aggagaggaa acgcctgtgt 1380  
 gagaacattg ctggccacct gaaggacgct cagcttttca ttcagaagaa agcgggtcaag 1440  
 aatttactg acgtccaccc tgactatggg gcccgcaccc aggctcttct ggacaagtac 1500  
 aacgctgaga agcctaagaa cgcaattcac acctacacgc aggccggctc tcacatggct 1560  
 gcgaagggaa aagctaacct gtaa 1584

<210> 13

<211> 99

<212> DNA

<213> 人工序列

<400> 13

caacaattac caacaacaac aaacaacaaa caacattaca attactatattt acaattacaa 60  
 aaaaaatgtc tgaagacgcc aaaaacataa agaaaggcc 99

<210> 14

<211> 53

<212> DNA

<213> 人工序列

<400> 14

ttttttttgt aattgtaaat agtaattgta atgttgtttg ttgtttgttg ttg 53

<210> 15

<211> 98

<212> DNA

<213> 人工序列

<400> 15

caattacca caacaacaaa caacaacaa cattacaatt actatttaca attacaaaa 60  
aaaaatgtct gaagacgcc aaaacataa gaaaggcc 98

<210> 16

<211> 99

<212> DNA

<213> 人工序列

<400> 16

aattaccaac aacaacaaac aacaacaaac attacatta ctatttaca ttacaaaaaa 60  
aaaaaatgtc tgaagagcc aaaaacataa agaaaggcc 99

<210> 17

<211> 27

<212> DNA

<213> 人工序列

<400> 17

cgcgaaatta atacgactca ctatagg 27

<210> 18

<211> 24

<212> DNA

<213> 人工序列

<400> 18

tccgatata gttcctcctt tcag 24

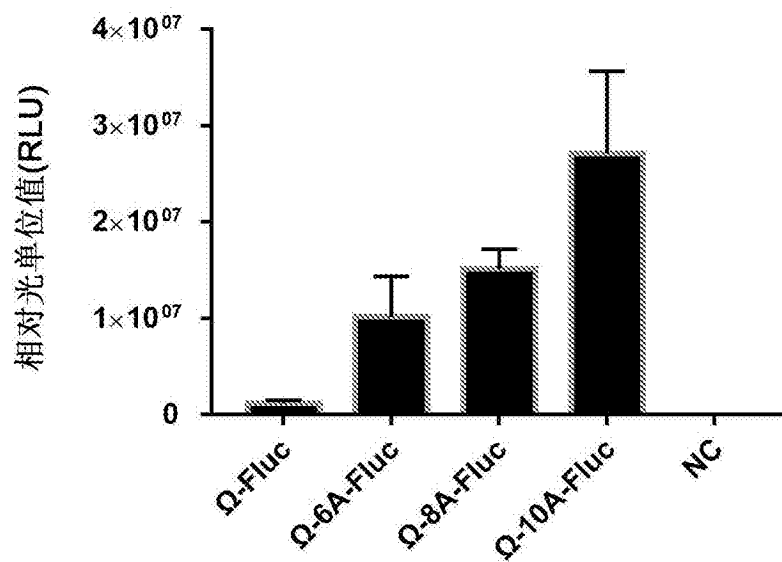


图1