



(12)发明专利申请

(10)申请公布号 CN 110819647 A

(43)申请公布日 2020.02.21

(21)申请号 201810888184.8

(22)申请日 2018.08.07

(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 徐开 于雪

(51)Int.Cl.

C12N 15/62(2006.01)

C12N 15/81(2006.01)

C12N 15/85(2006.01)

C12N 1/19(2006.01)

C12N 5/10(2006.01)

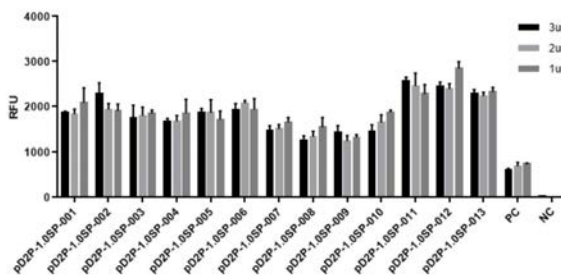
权利要求书1页 说明书17页  
序列表5页 附图1页

(54)发明名称

信号肽相关序列及其在蛋白质合成中的应用

(57)摘要

本发明提供了信号肽相关序列及其在蛋白质合成中的应用,具体地,本发明提供了具有提高蛋白表达效果的信号肽及其编码序列,将信号肽编码序列与外源蛋白的编码序列可操作的相连形成的核酸构建物,可显著提高外源蛋白合成的效率,并简化目标外源蛋白的表达与纯化过程。同时,本发明提供与其相应的载体或载体组合、基因工程细胞、试剂盒,从而能够应用于蛋白质合成中。



1. 一种核酸构建物,该核酸构建物包含编码信号肽的第一核苷酸序列,其与编码外源蛋白的第二核苷酸序列可操作地连接,第一核苷酸序列的3'端在第二核苷酸序列的上游,且第一核苷酸序列选自下组:

- (a) 编码以下任一信号肽的核苷酸序列:氨基酸序列为SEQ ID NO.:14-24的信号肽;
- (b) SEQ ID NO.:1-13所示的任一核苷酸序列。

2. 根据权利要求1所述的核酸构建物,其特征在于:所述的可操作地连接为直接连接或通过连接序列连接。

3. 根据权利要求2所述的核酸构建物,其特征在于:所述的连接序列为SEQ ID NO.:25所示的核苷酸序列。

4. 一种信号肽,其特征在于:其氨基酸序列由权利要求1中的第一核苷酸序列编码产生。

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任一所述的核酸构建物。

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任一所述的核酸构建物;
- (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. 根据权利要求7所述的试剂盒,其特征在于,其还包括(d) 体外生物合成体系。

9. 根据权利要求8所述的试剂盒,其特征在于:所述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。

10. 根据权利要求9所述的试剂盒,其特征在于:所述的酵母体外生物合成体系为克鲁维酵母体外生物合成体系。

11. 一种如权利要求1-3任一所述的核酸构建物、权利要求4所述的信号肽、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7-9任一所述试剂盒在体外蛋白合成体系中的应用。

12. 一种体外蛋白合成方法,其特征在于,其包括以下步骤:

(i) 提供一体外生物合成体系,所述体外生物合成体系含有权利要求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)的体外生物合成体系一段反应时间,从而合成所述外源蛋白。

13. 根据权利要求12所述的体外蛋白合成方法,其特征在于,合适的条件为反应温度为20-37°C,反应时间为1-72h。

14. 根据权利要求12所述的体外蛋白合成方法,其特征在于,所述方法还包括:(iii) 分离或检测所述外源蛋白。

## 信号肽相关序列及其在蛋白质合成中的应用

### 技术领域

[0001] 本发明涉及生物技术领域,具体地,涉及信号肽相关序列及其在蛋白质合成中的应用。

### 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(1,2)。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(2)。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(2)。

[0003] 信号肽是位于蛋白质N端携带蛋白质分泌信息的短肽,它们广泛分布于所有原核生物和真核生物中(3,4)。对于信号肽的研究,着重表现在很多科学和工业领域,包括重组蛋白质的生产,疾病诊断,免疫和很多生物学实验技术(4,5)。很多研究表明信号肽在重组蛋白质生产中有非常重要的作用(6,7)。但是,信号肽在蛋白表达以及跨膜结构形成方面的某些功能仍然很模糊(4,8,9)。

[0004] 除了人们对于细胞内蛋白质合成的了解之外,蛋白质合成也可以在细胞外进行。蛋白质体外合成系统一般是指在细菌、真菌、植物细胞或动物细胞的裂解体系中,加入mRNA或者DNA模板、RNA聚合酶及氨基酸和ATP等组分,完成外源蛋白的快速高效翻译(10,11)。目前,经常实验的商业化体外蛋白表达系统包括大肠杆菌系统(E.coli extract,ECE)、兔网织红细胞(Rabbit reticulocyte lysate,RRL)、麦胚(Wheat germ extract,WGE)、昆虫(Insect cell extract,ICE)和人源系统(11,12)。与传统的体内重组表达系统相比,蛋白质的体外无细胞合成系统具有多种优点,如可表达对细胞有毒害作用或含有非天然氨基酸(如D-氨基酸)的特殊蛋白质,能够直接以PCR产物作为模板同时平行合成多种蛋白质,开展高通量药物筛选和蛋白质组学的研究(10-12)。

[0005] 研究表明部分信号肽序列对蛋白表达有一定的促进作用,而在体外合成所使用的DNA模板通常不具有信号肽相关序列(13)。因此在体外蛋白质合成体系中一般采用在目标蛋白的N端插入长度在30个氨基酸以下的短多肽序列来提升目标蛋白质翻译效率,但有些短肽的插入会显著影响目标蛋白的结构与功能(4,14)。

[0006] 因此,本领域迫切需要提供一种能够应用于体外蛋白质表达系统中,可以显著提高目标蛋白产量,降低蛋白表达的成本,并且提高蛋白质翻译效率的信号肽相关序列。

### 发明内容

[0007] 本发明的目的在于提供一种能够应用于体外蛋白质表达系统中,可以显著提高目标蛋白产量,降低蛋白表达的成本,并且提高蛋白质翻译效率的信号肽相关序列。

[0008] 本发明第一方面,提供一种核酸构建物,该核酸构建物包含编码信号肽的第一核苷酸序列,其与编码外源蛋白的第二核苷酸序列可操作地连接,第一核苷酸序列的3'端在

第二核苷酸序列的上游,且第一核苷酸序列选自下组:

[0009] (a) 编码以下任一信号肽的核苷酸序列:氨基酸序列为SEQ ID NO:14-24的信号肽;

[0010] (b) SEQ ID NO:1-13所示的任一核苷酸序列。

[0011] 在另一优选例中,所述核酸构建物具有从5'至3'的式I结构:

[0012] Z1-Z2-Z3 (I)

[0013] 式中,

[0014] Z1-Z3分别为用于构成所述构建物的元件;

[0015] 各“-”独立地为键或核苷酸连接序列;

[0016] Z1为信号肽的编码序列;

[0017] Z2为无或连接序列;

[0018] Z3为无或外源蛋白的编码序列;

[0019] 其中,所述信号肽的编码序列选自下组:

[0020] (a) 编码如SEQ ID NO.:14-24所示多肽的多核苷酸;

[0021] (b) 序列如SEQ ID NO.:1-13任一所示的多核苷酸;

[0022] (c) 核苷酸序列与SEQ ID NO.:1-13任一所示序列的同源性 $\geq 75\%$  (较佳地 $\geq 85\%$ ,更佳地 $\geq 90\%$ 或 $\geq 95\%$ 或 $\geq 98\%$ 或 $\geq 99\%$ )的多核苷酸;

[0023] (d) 在SEQ ID NO.:1-13任一所示多核苷酸的5'端和/或3'端截短或添加1-60个(较佳地1-30,更佳地1-10个)核苷酸的多核苷酸;

[0024] (e) 与(a)-(d)任一所述的多核苷酸互补的多核苷酸。

[0025] 在另一优选例中,所述核酸构建物具有从5'至3'的式II结构:

[0026] Z1-Z2-Z3 (II)

[0027] 式中,

[0028] Z1-Z3分别为用于构成所述构建物的元件;

[0029] 各“-”独立地为键或核苷酸连接序列;

[0030] Z1为信号肽的编码序列;

[0031] Z2为连接序列;

[0032] Z3为无或外源蛋白的编码序列;

[0033] 其中,所述信号肽的编码序列选自下组:

[0034] (a) 编码如SEQ ID NO.:14-24所示多肽的多核苷酸;

[0035] (b) 序列如SEQ ID NO.:1-13任一所示的多核苷酸;

[0036] (c) 核苷酸序列与SEQ ID NO.:1-13任一所示序列的同源性 $\geq 75\%$  (较佳地 $\geq 85\%$ ,更佳地 $\geq 90\%$ 或 $\geq 95\%$ 或 $\geq 98\%$ 或 $\geq 99\%$ )的多核苷酸;

[0037] (d) 在SEQ ID NO.:1-13任一所示多核苷酸的5'端和/或3'端截短或添加1-60个(较佳地1-30,更佳地1-10个)核苷酸的多核苷酸;

[0038] (e) 与(a)-(d)任一所述的多核苷酸互补的多核苷酸。

[0039] 在另一优选例中,所述核酸构建物具有从5'至3'的式III结构:

[0040] Z1-Z2-Z3 (III)

[0041] 式中,

- [0042] Z1-Z3分别为用于构成所述构建物的元件；
- [0043] 各“-”独立地为键或核苷酸连接序列；
- [0044] Z1为信号肽的编码序列；
- [0045] Z2为连接序列；
- [0046] Z3为外源蛋白的编码序列；
- [0047] 其中,所述信号肽的编码序列选自下组:
- [0048] (a) 编码如SEQ ID NO.:14-24所示多肽的多核苷酸；
- [0049] (b) 序列如SEQ ID NO.:1-13任一所示的多核苷酸；
- [0050] (c) 核苷酸序列与SEQ ID NO.:1-13任一所示序列的同源性 $\geq 75\%$  (较佳地 $\geq 85\%$ ,更佳地 $\geq 90\%$ 或 $\geq 95\%$ 或 $\geq 98\%$ 或 $\geq 99\%$ )的多核苷酸；
- [0051] (d) 在SEQ ID NO.:1-13任一所示多核苷酸的5'端和/或3'端截短或添加1-60个(较佳地1-30,更佳地1-10个)核苷酸的多核苷酸；
- [0052] (e) 与(a)-(d)任一所述的多核苷酸互补的多核苷酸。
- [0053] 在另一优选例中,所述的可操作地连接为直接连接或通过连接序列连接。
- [0054] 在另一优选例中,所述的连接序列为SEQ ID NO.:25所示的核苷酸序列。
- [0055] 在另一优选例中,所述连接序列的氨基酸序列如SEQ ID NO.:26所示。
- [0056] 在另一优选例中,所述信号肽的编码序列为密码子优化的编码序列。
- [0057] 在另一优选例中,所述信号肽的编码序列如SEQ ID NO.:1-13所示。
- [0058] 在另一优选例中,所述信号肽的氨基酸序列具有SEQ ID NO.:14-24所示的序列或其活性片段,或者具有与SEQ ID NO.:14-24所示氨基酸序列 $\geq 85\%$ 的同源性,优选地 $\geq 90\%$ 的同源性;更优选地 $\geq 95\%$ 的同源性;最优选地, $\geq 97\%$ 的同源性,如98%以上,99%以上)且具有与SEQ ID NO.:14-24所示序列相同活性的多肽。
- [0059] 在另一优选例中,所述信号肽的编码序列如SEQ ID NO.:11-13所示。
- [0060] 在另一优选例中,所述信号肽的编码序列如SEQ ID NO.:2-7所示。
- [0061] 在另一优选例中,所述信号肽的编码序列如SEQ ID NO.:1所示。
- [0062] 在另一优选例中,所述信号肽的编码序列如SEQ ID NO.:8-10所示。
- [0063] 在另一优选例中,所述信号肽的氨基酸序列如SEQ ID NO.:22-24所示。
- [0064] 在另一优选例中,所述信号肽的氨基酸序列如SEQ ID NO.:15-20所示。
- [0065] 在另一优选例中,所述信号肽的氨基酸序列如SEQ ID NO.:14所示。
- [0066] 在另一优选例中,所述信号肽的氨基酸序列如SEQ ID NO.:21所示。
- [0067] 在另一优选例中,所述连接序列为密码子优化的连接序列。
- [0068] 在另一优选例中,所述连接序列具有不易形成二级结构的序列(如AT-rich序列,无发卡结构,无G-四链体(G-quadruplex)等),不富含稀有密码子。
- [0069] 在另一优选例中,所述连接序列选自下组:
- [0070] (i) 序列如SEQ ID NO.:25所示的多核苷酸；
- [0071] (ii) 核苷酸序列与SEQ ID NO.:25所示序列的同源性 $\geq 75\%$  (较佳地 $\geq 85\%$ ,更佳地 $\geq 90\%$ 或 $\geq 95\%$ 或 $\geq 98\%$ 或 $\geq 99\%$ )的多核苷酸；
- [0072] (iii) 在SEQ ID NO.:25所示多核苷酸的5'端和/或3'端截短或添加1-60个(较佳地1-30,更佳地1-10个)核苷酸的多核苷酸；

- [0073] (iv) 与 (i)-(iii) 任一所述的多核苷酸互补的多核苷酸。
- [0074] 在另一优选例中,所述外源蛋白来自原核生物、真核生物。
- [0075] 在另一优选例中,所述外源蛋白来自动物、植物、病原体。
- [0076] 在另一优选例中,所述外源蛋白来自哺乳动物,较佳地灵长动物,啮齿动物,包括人、小鼠、大鼠。
- [0077] 在另一优选例中,所述外源蛋白选自下组:荧光素蛋白、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、 $\alpha$ -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。
- [0078] 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、抗体或其可变区域、萤光素酶突变体、或其组合。
- [0079] 在另一优选例中,所述核酸构建物的5'端上游还包括启动子。
- [0080] 在另一优选例中,所述启动子包括组成型或诱导型启动子。
- [0081] 在另一优选例中,所述启动子选自下组:T7启动子、T3启动子、SP6启动子、或其组合。
- [0082] 在另一优选例中,所述核酸构建物还包括增强子元件、RBS核糖体结合序列、间隔序列(Spacer)、其他供RNA转录、翻译的相关序列、或其组合。
- [0083] 在另一优选例中,所述增强子元件包括内部核糖体进入位点元件(IRES)、核糖体结合位点元件、非编码序列、或其组合。
- [0084] 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。
- [0085] 在另一优选例中,所述真核细胞包括高等真核细胞。
- [0086] 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。
- [0087] 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)。
- [0088] 在另一优选例中,所述IRES元件选自下组:ScGPR1、ScFL08、ScNCE102、ScMSN1、K1FL08、K1NCE102、K1MSN1、GAA、Omega、Omega10A、或其组合。
- [0089] 本发明第二方面,提供一种信号肽,其氨基酸序列由第一方面中的第一核苷酸序列编码产生。
- [0090] 在另一优选例中,所述的信号肽的氨基酸序列如SEQ ID No.:14-24中任一所示。
- [0091] 本发明第三方面,提供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的核酸构建物。
- [0092] 本发明第四方面,提供一种基因工程细胞,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中的核酸构建物,或者所述基因工程细胞中含有本发明第三方面中的载体或载体组合。
- [0093] 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。

- [0094] 在另一优选例中,所述真核细胞包括高等真核细胞。
- [0095] 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。
- [0096] 在另一优选例中,所述基因工程细胞为酵母细胞。
- [0097] 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。
- [0098] 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。
- [0099] 本发明第五方面,提供一种试剂盒,所述试剂盒中包含的试剂选自下组中的一种或多种:
- [0100] (a) 本发明第一方面所述的核酸构建物;
- [0101] (b) 本发明第三方面所述的载体或载体组合;和
- [0102] (c) 本发明第四方面所述的基因工程细胞。
- [0103] 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)。
- [0104] 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。
- [0105] 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。
- [0106] 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。
- [0107] 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),优选为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)。
- [0108] 在另一优选例中,所述的酵母体外生物合成体系为克鲁维酵母体外生物合成体系。
- [0109] 本发明第六方面提供第一方面所述的核酸构建物、第二方面所述的信号肽、第三方面所述的载体或载体组合、第四方面所述的基因工程细胞或第五方面所述的试剂盒在体外蛋白合成体系中的应用。
- [0110] 本发明第七方面,提供一种体外蛋白合成方法,其包括以下步骤:
- [0111] (i) 提供一体外生物合成体系,所述体外生物合成体系含有本发明第一方面所述的核酸构建物;
- [0112] (ii) 在合适条件孵育步骤(i)的体外生物合成体系,经过一定反应时间,合成所述外源蛋白。
- [0113] 在另一优选例中,该体外蛋白合成方法还包括(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。
- [0114] 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℃。
- [0115] 在另一优选例中,所述步骤(ii)中,反应时间为1-72h,较佳地,2-24h。

[0116] 在另一优选例中,所述体外生物合成体系可选择酵母体外生物合成体系(如酵母体外蛋白合成体系)。

[0117] 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)。

[0118] 在另一优选例中,所述外源蛋白来自原核生物、真核生物。

[0119] 在另一优选例中,所述外源蛋白来自动物、植物、病原体。

[0120] 在另一优选例中,所述外源蛋白来自哺乳动物,较佳地灵长动物,啮齿动物,包括人、小鼠、大鼠。

[0121] 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、 $\alpha$ -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。

[0122] 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、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、 $\alpha$ -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。

[0123] 本发明第八方面提供了一种分离的多核苷酸,所述多核苷酸选自下组:

[0124] (a) 编码如SEQ ID NO.:14-24所示多肽的多核苷酸;

[0125] (b) 序列如SEQ ID NO.:1-13任一所示的多核苷酸;

[0126] (c) 核苷酸序列与SEQ ID NO.:1-13任一所示序列的同源性 $\geq 75\%$ (较佳地 $\geq 85\%$ ,更佳地 $\geq 90\%$ 或 $\geq 95\%$ 或 $\geq 98\%$ 或 $\geq 99\%$ )的多核苷酸;

[0127] (d) 在SEQ ID NO.:1-13任一所示多核苷酸的5'端和/或3'端截短或添加1-60个(较佳地1-30,更佳地1-10个)核苷酸的多核苷酸;

[0128] (e) 与(a)-(d)任一所述的多核苷酸互补的多核苷酸。

[0129] 在另一优选例中,所述多核苷酸为编码信号肽的核苷酸序列。

[0130] 在另一优选例中,所述多核苷酸包括DNA序列。

[0131] 本发明第九方面提供了一种连接序列,所述连接序列选自下组:

[0132] (i) 序列如SEQ ID NO.:25所示的多核苷酸;

[0133] (ii) 核苷酸序列与SEQ ID NO.:25所示序列的同源性 $\geq 75\%$ (较佳地 $\geq 85\%$ ,更佳地 $\geq 90\%$ 或 $\geq 95\%$ 或 $\geq 98\%$ 或 $\geq 99\%$ )的多核苷酸;

[0134] (iii) 在SEQ ID NO.:25所示多核苷酸的5'端和/或3'端截短或添加1-60个(较佳地1-30,更佳地1-10个)核苷酸的多核苷酸;

[0135] (iv) 与(i)-(iii)任一所述的多核苷酸互补的多核苷酸。

[0136] 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

[0137] 图1显示了从DNA到蛋白质的基本生物过程。

[0138] 图2显示了13种本发明信号肽相关序列在体外蛋白质合成体系中起始合成的增强型绿色荧光蛋白(Enhanced green fluorescent protein,eGFP)的相对荧光单位值(Relative Fluorescence Units,RFUs)。

## 具体实施方式

[0139] 本发明人经过广泛而深入的研究,通过大量筛选和摸索,首次发现了用于提高体外蛋白合成体系的蛋白翻译效率的新型信号肽及包括该信号肽编码序列的核酸构建物,本发明的核酸构建物包括与编码外源蛋白的第二核苷酸序列可操作地相连的编码信号肽的第一核苷酸序列(包括密码子优化的或未优化的信号肽编码序列)。实验表明,在体外蛋白质合成体系(如酵母体外蛋白质合成体系)中应用本发明的核酸构建物或信号肽序列,所合成的外源蛋白的信号强度比对照组显著性的提高( $p < 0.05$ )。本发明还简化了外源蛋白的表达和纯化方式。

[0140] 蛋白质合成体系

[0141] 蛋白质合成是指生物按照从脱氧核糖核酸(DNA)转录得到的信使核糖核酸(mRNA)上的遗传信息合成蛋白质的过程,如图1所示。蛋白质生物合成亦称为翻译(Translation),即把mRNA分子中碱基排列顺序转变为蛋白质或多肽链中的氨基酸排列顺序过程。这是基因表达的第二步,产生基因产物蛋白质的最后阶段。不同的组织细胞具有不同的生理功能,是因为它们表达不同的基因,产生具有特殊功能的蛋白质,参与蛋白质生物合成的成份至少有200种,其主要体是由mRNA、tRNA、核糖核蛋白体以及有关的酶和蛋白质因子共同组成。

[0142] 蛋白质体外合成系统一般是指在细菌、真菌、植物细胞或动物细胞的裂解体系中,加入mRNA或者DNA模板、RNA聚合酶及氨基酸和ATP等组分,完成外源蛋白的快速高效翻译。目前,经常实验的商业化体外蛋白表达系统包括大肠杆菌系统(*E.coli* extract,ECE)、兔网织红细胞(Rabbit reticulocyte lysate,RRL)、麦胚(Wheat germ extract,WGE)、昆虫(Insect cell extract,ICE)和人源系统。

[0143] 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。

[0144] 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。

[0145] 在本发明中,一种优选的蛋白合成体系为体外蛋白合成体系。本发明的体外蛋白质合成体系不受特别限制,一种优选的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)。

[0146] 在本发明中,所述体外蛋白质合成体系包括:酵母细胞提取物和任选的溶剂,所述

溶剂为水或水性溶剂。

[0147] 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还包括:4-羟乙基哌嗪乙磺酸,醋酸钾,醋酸镁,核苷三磷酸混合物,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,萤光素酶DNA, RNA聚合酶。

[0148] 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。

[0149] 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%。

[0150] 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。

[0151] 在本发明中,所述的酵母细胞提取物所含蛋白含量为10-100mg/mL,较佳地为20-80mg/mL。所述的测定蛋白含量方法为考马斯亮蓝测定方法。

[0152] 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:

[0153] (i) 提供酵母细胞;

[0154] (ii) 对酵母细胞进行洗涤处理,获得经洗涤的酵母细胞;

[0155] (iii) 对经洗涤的酵母细胞进行细胞破碎处理,从而获得酵母粗提物;

[0156] (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。

[0157] 在本发明中,所述的固液分离方式不受特别限制,优选为离心。

[0158] 在本发明中,所述离心条件不受特别限制,离心条件为5000-100000g,较佳地,8000-30000g。

[0159] 在本发明中,所述离心时间不受特别限制,离心时间为0.5min-2h,较佳地,20min-50min。

[0160] 在本发明中,所述离心的温度不受特别限制,优选的,离心在1-10°C下进行,较佳地,在2-6°C下进行。

[0161] 在本发明中,所述的洗涤处理方式不受特别限制,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。

[0162] 在本发明中,所述细胞破碎处理的方式不受特别限制,优选的细胞破碎处理包括高压破碎、冻融(如液氮低温)破碎。

[0163] 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。

[0164] 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。

[0165] 在优选例中,所述体外蛋白质合成体系还含有聚乙二醇(PEG)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)。

[0166] 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。

[0167] 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,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/ $\mu$ L萤火虫荧光素酶的DNA,0.027-0.054mg/mL T7RNA 聚合酶。

[0168] 外源蛋白的编码序列(外源DNA)

[0169] 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。

[0170] 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。

[0171] 在本发明中,所述外源DNA的选择没有特别限制,通常,外源DNA编码选自下组的蛋白:荧光素蛋白、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、抗体或其可变区、荧光素酶突变体、或其组合。

[0172] 外源DNA还可以编码选自下组的蛋白: $\alpha$ -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。

[0173] 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。

[0174] 核酸构建物

[0175] 本发明第一方面提供了一种核酸构建物,所述的核酸构建物包含编码信号肽的第一核苷酸序列,其与编码外源蛋白的第二核苷酸序列可操作地连接,第一核苷酸序列的3'端在第二核苷酸序列的上游,且第一核苷酸序列选自下组:

[0176] (a) 编码以下任一信号肽的核苷酸序列:氨基酸序列为SEQ ID NO:14-24的信号肽;

[0177] (b) SEQ ID NO:1-13所示的任一核苷酸序列。

[0178] 术语“可操作地连接”是指两个或多个核苷酸区域或核苷酸序列的功能性的空间排列。例如：编码信号肽核苷酸序列被置于相对于外源蛋白核苷酸序列的特定位置，使得获得提高外源蛋白表达的效果。所述的可操作地连接为直接连接或通过连接序列连接。

[0179] 在一优选实施方式中，本发明的所述核酸构建物具有从5'至3'的式I结构：

[0180] Z1-Z2-Z3 (I)

[0181] 式中，

[0182] Z1-Z3分别为用于构成所述构建物的元件；

[0183] 各“-”独立地为键或核苷酸连接序列；

[0184] Z1为信号肽的编码序列；

[0185] Z2为无或连接序列；

[0186] Z3为无或外源蛋白的编码序列；

[0187] 其中，所述信号肽的编码序列选自下组：

[0188] (a) 编码如SEQ ID NO.:14-24所示多肽的多核苷酸；

[0189] (b) 序列如SEQ ID NO.:1-13任一所示的多核苷酸；

[0190] (c) 核苷酸序列与SEQ ID NO.:1-13任一所示序列的同源性 $\geq 75\%$  (较佳地 $\geq 85\%$ ,更佳地 $\geq 90\%$ 或 $\geq 95\%$ 或 $\geq 98\%$ 或 $\geq 99\%$ )的多核苷酸；

[0191] (d) 在SEQ ID NO.:1-13任一所示多核苷酸的5'端和/或3'端截短或添加1-60个(较佳地1-30,更佳地1-10个)核苷酸的多核苷酸；

[0192] (e) 与(a)-(d)任一所述的多核苷酸互补的多核苷酸。

[0193] 在一优选实施方式中，本发明的所述核酸构建物具有从5'至3'的式II结构：

[0194] Z1-Z2-Z3 (II)

[0195] 式中，

[0196] Z1-Z3分别为用于构成所述构建物的元件；

[0197] 各“-”独立地为键或核苷酸连接序列；

[0198] Z1为信号肽的编码序列；

[0199] Z2为连接序列；

[0200] Z3为无或外源蛋白的编码序列；

[0201] 其中，所述信号肽的编码序列选自下组：

[0202] (a) 编码如SEQ ID NO.:14-24所示多肽的多核苷酸；

[0203] (b) 序列如SEQ ID NO.:1-13任一所示的多核苷酸；

[0204] (c) 核苷酸序列与SEQ ID NO.:1-13任一所示序列的同源性 $\geq 75\%$  (较佳地 $\geq 85\%$ ,更佳地 $\geq 90\%$ 或 $\geq 95\%$ 或 $\geq 98\%$ 或 $\geq 99\%$ )的多核苷酸；

[0205] (d) 在SEQ ID NO.:1-13任一所示多核苷酸的5'端和/或3'端截短或添加1-60个(较佳地1-30,更佳地1-10个)核苷酸的多核苷酸；

[0206] (e) 与(a)-(d)任一所述的多核苷酸互补的多核苷酸。

[0207] 在一优选实施方式中，本发明的所述核酸构建物具有从5'至3'的式III结构：

[0208] Z1-Z2-Z3 (III)

[0209] 式中，

[0210] Z1-Z3分别为用于构成所述构建物的元件；

[0211] 各“-”独立地为键或核苷酸连接序列；

[0212] Z1为信号肽的编码序列；

[0213] Z2为连接序列；

[0214] Z3为外源蛋白的编码序列；

[0215] 其中，所述信号肽的编码序列选自下组：

[0216] (a) 编码如SEQ ID NO.:14-24所示多肽的多核苷酸；

[0217] (b) 序列如SEQ ID NO.:1-13任一所示的多核苷酸；

[0218] (c) 核苷酸序列与SEQ ID NO.:1-13任一所示序列的同源性 $\geq 75\%$  (较佳地 $\geq 85\%$ ,更佳地 $\geq 90\%$ 或 $\geq 95\%$ 或 $\geq 98\%$ 或 $\geq 99\%$ )的多核苷酸；

[0219] (d) 在SEQ ID NO.:1-13任一所示多核苷酸的5'端和/或3'端截短或添加1-60个(较佳地1-30,更佳地1-10个)核苷酸的多核苷酸；

[0220] (e) 与(a)-(d)任一所述的多核苷酸互补的多核苷酸。

[0221] 在一优选实施方式中，本发明所述信号肽的氨基酸序列具有SEQ ID NO.:14-24所示的序列或其活性片段，或者具有与SEQ ID NO.:14-24所示氨基酸序列 $\geq 85\%$ 的同源性，优选地 $\geq 90\%$ 的同源性；更优选地 $\geq 95\%$ 的同源性；最优选地， $\geq 97\%$ 的同源性，如98%以上，99%以上)且具有与SEQ ID NO.:14-24所示序列相同活性的多肽。

[0222] 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编码选自下组的蛋白：荧光素蛋白、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、抗体或其可变区、荧光素酶突变体、或其组合。

[0223] 外源蛋白的编码序列还可以编码选自下组的蛋白： $\alpha$ -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。

[0224] 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。

[0225] 本发明的优选的信号肽序列及其编码信号肽的核苷酸序列如表1所示。

[0226] 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。

[0227] 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。

[0228] 在本发明中，在体外蛋白质合成体系中应用本发明的核酸构建物，可以显著提高

外源蛋白的翻译效率。在一优选实施方式中,在本发明的酵母体外蛋白质合成体系中应用本发明的核酸构建物,可显著提高蛋白翻译的效率。

[0229] 载体

[0230] 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。此外,载体可以为转座子载体。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。

[0231] 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。

[0232] 基因工程细胞

[0233] 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。

[0234] 优选地,所述的基因工程细胞为真核细胞。

[0235] 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)。

[0236] 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。

[0237] 此外,本发明的基因工程细胞可用于生产或提供本发明的核酸构建物。

[0238] 体外高通量的蛋白合成方法

[0239] 本发明的信号肽、含信号肽编码序列的构建物,特别适合在体外生物合成体系中显著提高外源蛋白的合成效率或产量。

[0240] 相应地,本发明还提供了一种体外高通量的蛋白合成方法,包括步骤:

[0241] (i) 在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;

[0242] (i i) 在适合的条件下,孵育步骤(i)的所述体外蛋白合成体系一段时间T1,从而合成所述外源蛋白。

[0243] 在另一优选例中,所述方法还包括:(iii) 任选地从所述体外蛋白合成体系中,分离或检测所述外源蛋白。

[0244] 本发明的主要优点包括:

[0245] (1) 本发明首次发现,将信号肽相关序列和外源蛋白的编码序列作为核酸构建物,应用于本发明的体外蛋白质合成体系中,可用于提高目标蛋白翻译效率并可表达纯化。

[0246] (2) 本发明的信号肽相关序列在翻译起始密码子之后,可以影响mRNA的折叠,从而改变目标蛋白的翻译效率。

[0247] (3) 与其他细胞相比,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,加上体外蛋白质合成体系的优点,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蛋白质合成体系在相关领域也能够得到广泛的应用。

[0248] (4) 本发明提供的信号肽相关序列不仅能够提升目标外源蛋白翻译效率,更主要的是能够增加酵母体外蛋白合成体系(如乳酸克鲁维酵母体外蛋白质合成体系)的针对不同蛋白质合成的可能性。

[0249] (5) 本发明首次开发一种用于提高体外蛋白合成体系的蛋白翻译效率的新型信号肽及包括该信号肽编码序列的核酸构建物,本发明的核酸构建物包括与编码外源蛋白的第二核苷酸序列可操作地相连的编码信号肽的第一核苷酸序列(如密码子优化的信号肽的编码序列),本发明的核酸构建物可显著提高所合成的外源蛋白的表达。

[0250] 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。

[0251] 材料

[0252] 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。

[0253] 实施例中的外源蛋白以eGFP为例。

[0254] 实施例1真核细胞信号肽相关序列的确定

[0255] 1.1信号肽相关序列的来源及确定:随机截取已构建外源蛋白N端相对应的DNA序列,通过实验筛选对于外源蛋白表达有显著提高的序列或元件。

[0256] 具体地,选取并合成碱基长度选择为36,54或者其他长度的30个核苷酸片段,同时采用同义密码子交错替换的方式修改了部分碱基,以提高质粒构建的成功率,其中包括(但并不限于):为了降低信号肽相关序列中的GC含量从而降低了序列的退火温度;或采用优选密码子。构建得到几十个质粒,经分析和筛选,对其中的30个质粒并测试其对外源蛋白表达的作用结果表明,与对照相比,13个信号肽相关序列经过实验验证具有提高蛋白表达的作用,其相应的信号肽的氨基酸序列及编码信号肽的核苷酸序列信息列于表1。其他对于提高蛋白表达没有效果或效果不显著的序列信息未予列出。

[0257] 表1质粒及相关核酸序列

[0258]

名称	信号肽核酸序列	SEQ ID No. :	氨基酸序列	SEQ ID No. :
pD2P-1.0SP-001	AGTGAGCAAAGCCAATTAGATGATTCGACTATAGAC	1	SEQSQLDDSTID	14
pD2P-1.0SP-002	CTGACAACGTGTTCTCCCTAACGTAGCTACATTAAC	2	LTTVLPNVATLN	15
pD2P-1.0SP-003	ATGCTGACAACGTGTTCTCCCTAACGTAGCTACATTAACAG TATGTTTGCCCTG	3	LTTVLPNVATLNSMFAL	16
pD2P-1.0SP-004	AATTGCTCCGCACATTGTATCAAAAAGGCTTTACCT	4	NCSAHCIKKALP	17
pD2P-1.0SP-005	AATTGCTCCGCACATTGTATCAAAAAGGCTTTACCTGCACA GTGGATCCGTTGC	5	NCSAHCIKKALPAQWIRC	18
pD2P-1.0SP-006	AAAACACATATAGTCAGCTCAGTAACAACAACACTA	6	KTHIVSSVTTL	19
pD2P-1.0SP-007	AAAACACATATAGTCAGCTCAGTAACAACAACACTATTGCT AGGTCCATATTA	7	KTHIVSSVTTL LLSIL	20
pD2P-1.0SP-008	TCTGGTGGTCAAATTTTCGTAAAGACGCTGACCGGT	8	SGGQIFVKLTG	21

[0259]

pD2P-1.0SP-009	TCTGGTGGTCAAATTTTCGTCAAACCTAACAGGT	9	SGGQIFVKLTG	21
pD2P-1.0SP-010	TCTGGTGGTCAAATTTTCGTAAAACCTTACTGGT	10	SGGQIFVKLTG	21
pD2P-1.0SP-011	AAGCCTCCAGTATACCCATCGATTTGC	11	KPPVYPSIC	22
pD2P-1.0SP-012	ATGATTACAGAAACATCATCACCGTTCAGATCTATA	12	ITETSSPFRSI	23
pD2P-1.0SP-013	ATGGTCGCTAGAGGTAGAACAGACGAGATATCTACAGATGT TTCAGAGGCTAAT	13	VARGRTDEISTDVSEAN	24
连接序列	GAAAACCTGTATTTCCAAGGAGGTAGTGAGGAAGTGGTGG AAGTGA	25	ENLYFQGGSGGSGGSG	26

[0260] 实施例2含有信号肽相关序列的体外蛋白质合成体系质粒的构建

[0261] 质粒的构建:对于选出的30个信号肽相关序列和连接序列(连接序列上含有TEV酶切位点)使用1对引物扩增,原始含有目标蛋白(以eGFP为例)的质粒骨架使用其对应的反向引物进行扩增。扩增完成后将30个信号肽相关序列+连接序列片段分别插入在目标蛋白的N端。在最终构建的质粒中,30个信号肽相关序列+连接序列核酸序列插入到pD2P-eGFP质粒的ATG起始密码子和eGFP之间。其中13个质粒的名称分别为:pD2P-1.0SP-(001-013)(见表1)。

[0262] 具体构建过程如下:

[0263] 使用2对引物分别进行PCR扩增,并取10 $\mu$ L经鉴定正确的扩增产物进行混合;向10 $\mu$ L扩增产物中加入0.5 $\mu$ L DpnI,37 $^{\circ}$ C孵育6h;将DpnI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。

[0264] 实施例3信号肽相关序列在体外蛋白质合成体系中的应用

[0265] 3.1利用PCR,并使用引物pD2P\_F:CGCGAAATTAATACGACTCACTATAGG (SEQ ID No.: 27) 和pD2P\_R:TCCGGATATAGTTCCTCCTTTCAG (SEQ ID NO.:28),将所有质粒中位于T7转录起始序列和终止序列之间包含信号肽相关序列的片段和pD2P-eGFP片段进行扩增。

[0266] 并将扩增得到的经测序鉴定正确的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。

[0267] 3.2按照使用说明,将纯化的DNA片段加入到自制的体外蛋白质合成体系中。并将上述反应体系置于22-30℃的环境中,静置孵育约2-6h。反应结束后,立即放置于Envision 2120多功能酶标仪(Perkin Elmer),读数,检测eGFP信号强弱,相对荧光单位值(Relative Fluorescence Unit,RFU) 作为活性单位。

[0268] PC(Positive Control)是在增强型绿色荧光蛋白N端仅添加连接序列的实验组,NC(Negative Control)是不加入任何核酸构建物的实验组。1 $\mu$ l、2 $\mu$ l、3 $\mu$ l分别为向体外蛋白质合成体系中加入的DNA模板量,所有反应的总反应体系体积为30 $\mu$ l。

[0269] 实验结果

[0270] 1.体外蛋白质合成体系质粒的构建

[0271] 经过多次尝试,最终构建成功30个含有信号肽相关序列的体外蛋白质合成体系质粒。

[0272] 2.信号肽相关序列在体外蛋白质合成体系中的应用

[0273] 如图2所示,13个包含信号肽相关序列在体外蛋白质合成体系中引起eGFP 发出的RFU值均显著提高(反应3小时后RFU值达到1500以上),最高可达2900。尤其pD2P-1.0SP-012(添加1 $\mu$ lDNA模板量,反应3小时后RFU值达到2900),与未插入信号肽相关序列的对照PC(RFU值为800)相比相对荧光单位值提高了 2.6倍。

[0274] 而对于其他17个未列出的信号肽序列而言,其相对荧光单位值与PC相比无变化或者是变化不明显,大多为800-830,如pD2P-1.0SP-019为823, pD2P-1.0SP-027为816。

[0275] 本发明结果表明,目标蛋白N端的信号肽相关序列可以显著提高目标蛋白的产量,大大提升了目标蛋白质表达纯化效果。提高了目标蛋白的翻译效率并增加了体外合成体系进行蛋白质表达纯化方式的选择性,极大增强了体外蛋白质合成体系的可用性。

[0276] 并且,本发明的研究还发现,将5'-UTR、强启动子(如T7启动子、T3启动子、SP6启动子)、不同的IRES元件(如KLNCE102)与不同的信号肽相关序列,3'-UTR等组合,也可进一步提高目标蛋白的翻译效率。

[0277] 对比例(PC和NC)

[0278] PC(Positive Control)是在增强型绿色荧光蛋白N端仅添加连接序列的实验组,添加1 $\mu$ lDNA模板量的情况下外源蛋白的RFU值为800,反应体系总体积为30 $\mu$ l。

[0279] NC(Negative Control)是不加入任何核酸构建物的实验组,外源蛋白的 RFU值为20,反应体系总体积为30 $\mu$ l。

[0280] 其中,图2中的1 $\mu$ l、2 $\mu$ l、3 $\mu$ l表示分别为向体外蛋白合成体系中加入的 DNA模板量,所有反应的总反应体系体积为30 $\mu$ l,反应时间3小时。

[0281] 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。

[0282] 参考文献:

[0283] 1.Fromm HJ,Hargrove M.Essentials of Biochemistry.2012.

[0284] 2.Garcia RA,Riley MR.Applied biochemistry and biotechnology. Humana Press,;1981.263-264p.

[0285] 3.Martoglio B.Intramembrane proteolysis and post-targeting functions of signal peptides.Biochem Soc Trans.2003;31(6):1243-7.

[0286] 4.Owji H,Nezafat N,Negahdaripour M,Hajiebrahimi A,Ghasemi Y.A Comprehensive Review of Signal Peptides:Structure,Roles,and Applications.Eur J Cell Biol.2018.

[0287] 5.Liu H,Wu R,Yuan L,Tian G,Huang X,Wen Y,et al.Introducing a cleavable signal peptide enhances the packaging efficiency of lentiviral vectors pseudotyped with Japanese encephalitis virus envelope proteins.Virus Res.2017;229:9-16.

[0288] 6.Cui Y,Meng Y,Zhang J,Cheng B,Yin H,Gao C,et al.Efficient secretory expression of recombinant proteins in Escherichia coli with a novel actinomycete signal peptide.Protein Expr Purif. 2017;129:69-74.

[0289] 7.Ling HL,Rahmat Z,Murad AMA,Mahadi NM,Illias RM. Proteome-based identification of signal peptides for improved secretion of recombinant cyclomaltodextrin glucanotransferase in Escherichia coli.Process Biochem.2017;61:47-55.

[0290] 8.Zhang S,Corin K.18-Peptide surfactants in membrane protein purification and stabilization A2-Koutsopoulos,Sotirios BT- Peptide Applications in Biomedicine,Biotechnology and Bioengineering. In Woodhead Publishing;2018.p.485-512.

[0291] 9.Stone TA,Deber CM.Therapeutic design of peptide modulators of protein-protein interactions in membranes.Biochim Biophys Acta-Biomembr.2017;1859(4):577-85.

[0292] 10.Katzen F,Chang G,Kudlicki W.The past,present and future of cell-free protein synthesis.Trends Biotechnol.2005;23(3):150-6.

[0293] 11.Gan R,Jewett MC.A combined cell-free transcription-translation system from Saccharomyces cerevisiae for rapid and robust protein synthesis.Biotechnol J.2014;9(5):641-51.

[0294] 12.Lu Y.Cell-free synthetic biology:Engineering in an open

world.Synth Syst Biotechnol.2017;2(1):23-7.

[0295] 13.Kralicek A V.,Radjainia M,Mohamad Ali NAB,Carraher C, Newcomb RD, Mitra AK.A PCR-directed cell-free approach to optimize protein expression using diverse fusion tags.Protein Expr Purif.2011;80(1):117-24.

[0296] 14.Hansted JG,Pietikäinen L,Hög F,Sperling-Petersen HU, Mortensen KK.Expressivity tag:A novel tool for increased expression in Escherichia coli.J Biotechnol.2011;155(3):275-83.

[0297] 15.Kasi D,Nah HJ,Catherine C,Kim ES,Han K,Ha JC,et al. Enhanced Production of Soluble Recombinant Proteins With an In Situ-Removable Fusion Partner in a Cell-Free Synthesis System. Biotechnol J.2017;12(11):1-6.

[0298] 16.Gräslund S,Nordlund P,Weigelt J,Hal lberg BM,Bray J,Gileadi O,et al.Protein production and purification.Nat Methods. 2008;5(2):135-46。

## 序列表

<110> 康码(上海)生物科技有限公司

<120> 信号肽相关序列及其在蛋白质合成中的应用

<130> P2018-1332

<141> 2018-08-07

<160> 28

<170> SIPOSequenceListing 1.0

<210> 1

<211> 36

<212> DNA

<213> 人工序列 (artificial sequence)

<400> 1

agtgagcaaaa gcccaattaga tgattcgact atagac 36

<210> 2

<211> 36

<212> DNA

<213> 人工序列 (artificial sequence)

<400> 2

ctgacaactg ttctccctaa cgtagctaca ttaaac 36

<210> 3

<211> 54

<212> DNA

<213> 人工序列 (artificial sequence)

<400> 3

atgctgacaaa ctgttctccc taacgtagct acattaaaca gtatgtttgc cctg 54

<210> 4

<211> 36

<212> DNA

<213> 人工序列 (artificial sequence)

<400> 4

aattgctccg cacattgtat caaaaaggct ttacct 36

<210> 5

<211> 54

<212> DNA

<213> 人工序列 (artificial sequence)

<400> 5

aattgctccg cacattgtat caaaaaggct ttacctgcac agtggatccg ttgc 54

<210> 6

- <211> 36  
<212> DNA  
<213> 人工序列 (artificial sequence)  
<400> 6  
aaaacacata tagtcagctc agtaacaaca aacta 36  
<210> 7  
<211> 54  
<212> DNA  
<213> 人工序列 (artificial sequence)  
<400> 7  
aaaacacata tagtcagctc agtaacaaca aactattgc taggttccat atta 54  
<210> 8  
<211> 36  
<212> DNA  
<213> 人工序列 (artificial sequence)  
<400> 8  
tctggtggtc aaattttcgt aaagacgctg accggt 36  
<210> 9  
<211> 36  
<212> DNA  
<213> 人工序列 (artificial sequence)  
<400> 9  
tctggtggtc aaattttcgt caaaactcta acaggt 36  
<210> 10  
<211> 36  
<212> DNA  
<213> 人工序列 (artificial sequence)  
<400> 10  
tctggtggtc aaattttcgt taaaactctt actggt 36  
<210> 11  
<211> 27  
<212> DNA  
<213> 人工序列 (artificial sequence)  
<400> 11  
aagcctccag tatacccatc gatttgc 27  
<210> 12  
<211> 36  
<212> DNA  
<213> 人工序列 (artificial sequence)

<400> 12

atgattacag aaacatcatc accgttcaga tctata 36

<210> 13

<211> 54

<212> DNA

<213> 人工序列 (artificial sequence)

<400> 13

atggctcgcta gaggtagaac agacgagata tctacagatg tttcagaggc taat 54

<210> 14

<211> 12

<212> PRT

<213> 人工序列 (artificial sequence)

<400> 14

Ser Glu Gln Ser Gln Leu Asp Asp Ser Thr Ile Asp

1                    5                    10

<210> 15

<211> 12

<212> PRT

<213> 人工序列 (artificial sequence)

<400> 15

Leu Thr Thr Val Leu Pro Asn Val Ala Thr Leu Asn

1                    5                    10

<210> 16

<211> 17

<212> PRT

<213> 人工序列 (artificial sequence)

<400> 16

Leu Thr Thr Val Leu Pro Asn Val Ala Thr Leu Asn Ser Met Phe Ala

1                    5                    10                    15

Leu

<210> 17

<211> 12

<212> PRT

<213> 人工序列 (artificial sequence)

<400> 17

Asn Cys Ser Ala His Cys Ile Lys Lys Ala Leu Pro

1                    5                    10

<210> 18

<211> 18

<212> PRT

<213> 人工序列 (artificial sequence)

<400> 18

Asn Cys Ser Ala His Cys Ile Lys Lys Ala Leu Pro Ala Gln Trp Ile  
1                   5                   10                   15

Arg Cys

<210> 19

<211> 12

<212> PRT

<213> 人工序列 (artificial sequence)

<400> 19

Lys Thr His Ile Val Ser Ser Val Thr Thr Thr Leu  
1                   5                   10

<210> 20

<211> 18

<212> PRT

<213> 人工序列 (artificial sequence)

<400> 20

Lys Thr His Ile Val Ser Ser Val Thr Thr Thr Leu Leu Leu Gly Ser  
1                   5                   10                   15

Ile Leu

<210> 21

<211> 12

<212> PRT

<213> 人工序列 (artificial sequence)

<400> 21

Ser Gly Gly Gln Ile Phe Val Lys Thr Leu Thr Gly  
1                   5                   10

<210> 22

<211> 9

<212> PRT

<213> 人工序列 (artificial sequence)

<400> 22

Lys Pro Pro Val Tyr Pro Ser Ile Cys  
1                   5

<210> 23

<211> 11

<212> PRT

<213> 人工序列 (artificial sequence)

<400> 23

Ile Thr Glu Thr Ser Ser Pro Phe Arg Ser Ile

1                    5                    10

<210> 24

<211> 17

<212> PRT

<213> 人工序列 (artificial sequence)

<400> 24

Val Ala Arg Gly Arg Thr Asp Glu Ile Ser Thr Asp Val Ser Glu Ala

1                    5                    10                    15

Asn

<210> 25

<211> 48

<212> DNA

<213> 人工序列 (artificial sequence)

<400> 25

gaaaacctgt atttccaagg aggtagtgga ggaagtgggtg gaagtgga 48

<210> 26

<211> 16

<212> PRT

<213> 人工序列 (artificial sequence)

<400> 26

Glu Asn Leu Tyr Phe Gln Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly

1                    5                    10                    15

<210> 27

<211> 27

<212> DNA

<213> 人工序列 (artificial sequence)

<400> 27

cgcgaaatta atacgactca ctatagg 27

<210> 28

<211> 24

<212> DNA

<213> 人工序列 (artificial sequence)

<400> 28

tccgatata gttcctcctt tcag 24

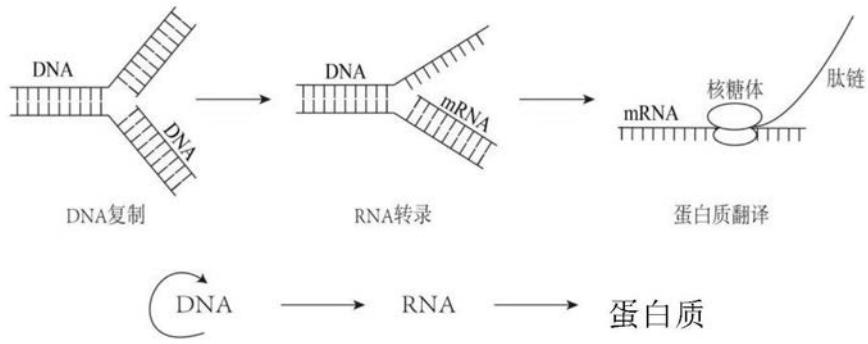


图1

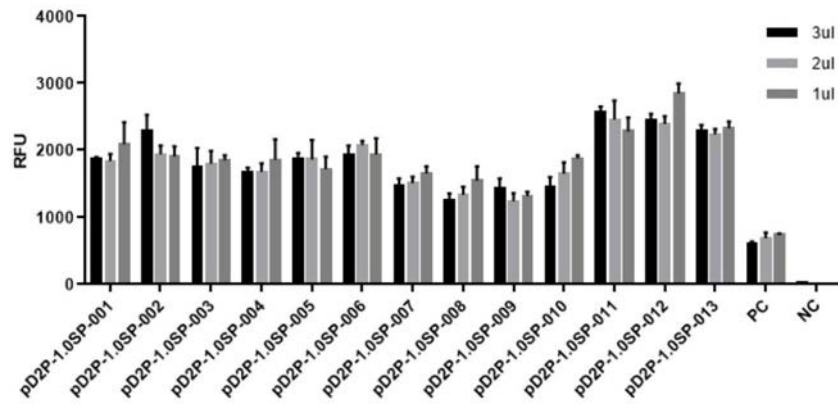


图2