



(12)发明专利申请

(10)申请公布号 CN 110408636 A

(43)申请公布日 2019.11.05

(21)申请号 201810404013.3

(22)申请日 2018.04.28

(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 章小铃 周子鉴
江丹 李海洋 于雪

(51)Int.Cl.

C12N 15/62(2006.01)

C12N 5/10(2006.01)

C07K 19/00(2006.01)

C07K 7/08(2006.01)

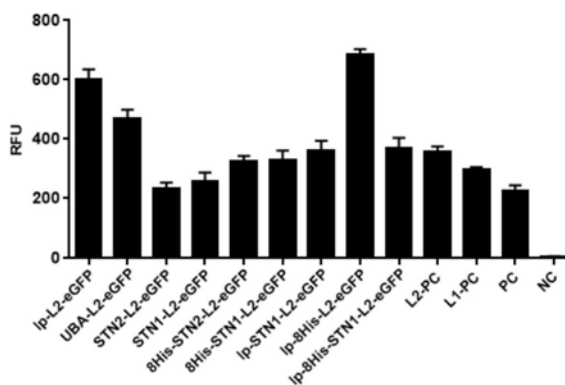
权利要求书3页 说明书17页
序列表3页 附图3页

(54)发明名称

多重标签串联的DNA序列及其在蛋白质表达纯化系统的应用

(57)摘要

本发明提供了一种多重标签串联的DNA序列及其在蛋白质表达纯化系统的应用,具体地,本发明提供了一种核酸构建物,所述核酸构建物具有从5'至3'的式I结构:Z1-Z2-Z3-Z4-Z5(I);式中,Z1-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为前导肽的编码序列;Z2为无或泛素的编码序列;Z3为标签蛋白的编码序列;Z4为连接序列;Z5为无或外源蛋白的编码序列。本发明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。



1. 一种核酸构建物,其特征在于,所述核酸构建物具有从5'至3'的式I结构:

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

式中,

Z1-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为前导肽的编码序列;

Z2为无或泛素的编码序列;

Z3为标签蛋白的编码序列;

Z4为连接序列;

Z5为无或外源蛋白的编码序列;

其中,所述前导肽的编码序列选自下组:

(a) 编码如SEQ ID NO.:2所示多肽的多核苷酸;

(b) 序列如SEQ ID NO.:1所示的多核苷酸;

(c) 核苷酸序列与SEQ ID NO.:1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;

(d) 在SEQ ID NO.:1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;

(e) 与(a)-(d)任一所述的多核苷酸互补的多核苷酸。

2. 如权利要求1所述的核酸构建物,其特征在于,所述泛素的编码序列选自下组:

(a) 编码如SEQ ID NO.:4所示多肽的多核苷酸;

(b) 序列如SEQ ID NO.:3所示的多核苷酸;

(c) 核苷酸序列与SEQ ID NO.:3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;

(d) 在SEQ ID NO.:3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;

(e) 与(a)-(d)任一所述的多核苷酸互补的多核苷酸。

3. 如权利要求1所述的核酸构建物,其特征在于,所述标签蛋白选自下组:链霉亲和素、MBP、GST、Protein、CBP、6His、8His、或其组合。

4. 如权利要求1所述的核酸构建物,其特征在于,所述标签蛋白的编码序列选自下组:

(a) 编码如SEQ ID NO.:6所示多肽的多核苷酸;

(b) 序列如SEQ ID NO.:5或8所示的多核苷酸;

(c) 核苷酸序列与SEQ ID NO.:5或8所示序列的同源性 $\geq 30\%$,较佳地,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;

(d) 在SEQ ID NO.:5或8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;

(e) 与(a)-(d)任一所述的多核苷酸互补的多核苷酸。

5. 如权利要求1所述的核酸构建物,其特征在于,所述标签蛋白具有SEQ ID NO.:6所示的序列或其活性片段,或者具有与SEQ ID NO.:6所示氨基酸序列 $\geq 30\%$ 同源性(优选地, $\geq 85\%$ 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.: 6序列相同活性的多肽。

6. 如权利要求1所述的核酸构建物,其特征在于,所述连接序列选自下组;

(i) 序列如SEQ ID NO.:7或9所示的多核苷酸;

(ii) 核苷酸序列与SEQ ID NO.:7或9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;

(iii) 在SEQ ID NO.:7或9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;

(iv) 与(i)-(iii)任一所述的多核苷酸互补的多核苷酸。

7. 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所述的核酸构建物。

8. 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所述的构建物,或者所述基因工程细胞中含有权利要求7所述的载体或载体组合。

9. 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所述的构建物;

(b) 权利要求7所述的载体或载体组合;和

(c) 权利要求8所述的基因工程细胞。

10. 一种如权利要求1所述的构建物、权利要求7所述的载体或载体组合、权利要求8所述的基因工程细胞或权利要求9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。

11. 一种体外高通量的外源蛋白合成方法,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。

12. 如权利要求11所述的方法,其特征在于,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。

13. 一种分离的多核苷酸,其特征在于,所述多核苷酸选自下组:

(a) 编码如SEQ ID NO.:2所示多肽的多核苷酸;

(b) 序列如SEQ ID NO.:1所示的多核苷酸;

(c) 核苷酸序列与SEQ ID NO.:1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;

(d) 在SEQ ID NO.:1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;

(e) 与(a)-(d)任一所述的多核苷酸互补的多核苷酸。

14. 一种前导肽,其特征在于,所述前导肽的氨基酸序列选自下组:

(A) 具有SEQ ID NO:2所示氨基酸序列的多肽;(B) 具有与SEQ ID NO: 2所示氨基酸序列 $\geq 80\%$ 同源性(优选地, $\geq 90\%$ 的同源性;等优选地 $\geq 95\%$ 的同源性;最优选地, $\geq 97\%$ 的同源性,如98%以上,99%以上)的多肽,且所述多肽具有提高外源蛋白表达效率的功能或活性;

(C) 将SEQ ID NO:2中任一所示氨基酸序列经过1-15(较佳地,2-10,更佳地,3-8)个氨

基酸残基的取代、缺失或添加而形成的,且具有提高外源蛋白表达效率的功能或活性的衍生多肽。

多重标签串联的DNA序列及其在蛋白质表达纯化系统的应用

技术领域

[0001] 本发明涉及生物技术领域,具体地,涉及多重标签串联的DNA序列及其在蛋白质表达纯化系统的应用。

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]。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[1]。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[1, 2]。

[0003] 在从酵母到人类的细胞内都发现了一些翻译增强编码序列和用于蛋白表达纯化的亲和序列,而针对细胞内源的这些元件的研究并不充分,主要是因为细胞内源性元件的作用受别的细胞作用因子等因素的影响,会影响起始蛋白质翻译的效率,而且在纯化过程中容易发生非特异性结合[3, 4]。不同于病毒中表达蛋白较为简单,不同的细胞内源性表达蛋白不具有序列和结构上的共性,通常很难进行高通量具有共性的翻译增强编码序列和用于蛋白表达纯化的亲和序列进行表达纯化[3, 4]。

[0004] 在体外合成所使用的DNA模板通常不具有多标签串联元件,而对不同的蛋白进行设计不同的标签和纯化方法时间和成本都相对较高[5]。而目前使用真核细胞多标签串联元件在体外起始蛋白质的表达与纯化的应用还比较少。

[0005] 因此,本领域迫切需要开发一种可提升目标蛋白的表达效率并可用于蛋白质表达纯化的DNA序列。

发明内容

[0006] 本发明的目的在于提供一种可提升目标蛋白的表达效率并可用于蛋白质表达纯化的DNA序列。

[0007] 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从5'至3'的式I结构:

[0008] Z1-Z2-Z3-Z4-Z5 (I)

[0009] 式中,

[0010] Z1-Z5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为前导肽的编码序列;

[0013] Z2为无或泛素的编码序列;

[0014] Z3为标签蛋白的编码序列;

[0015] Z4为连接序列;

[0016] Z5为无或外源蛋白的编码序列;

[0017] 其中,所述前导肽的编码序列选自下组:

[0018] (a) 编码如SEQ ID NO.:2所示多肽(前导肽的氨基酸序列: ITETSSPFRSIFSHSGK)的多核苷酸;

[0019] (b) 序列如SEQ ID NO.:1所示的多核苷酸;

[0020] (c) 核苷酸序列与SEQ ID NO.:1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;

[0021] (d) 在SEQ ID NO.:1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;

[0022] (e) 与(a)-(d)任一所述的多核苷酸互补的多核苷酸。

[0023] 在另一优选例中,所述前导肽的编码序列为密码子优化的编码序列。

[0024] 在另一优选例中,所述前导肽的编码序列如SEQ ID NO.:1所示。

[0025] 在另一优选例中,所述前导肽的氨基酸序列具有SEQ ID NO.:2所示的序列或其活性片段,或者具有与SEQ ID NO.:2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.:2序列相同活性的多肽。

[0026] 在另一优选例中,所述泛素的编码序列选自下组:

[0027] (a) 编码如SEQ ID NO.:4所示多肽(泛素的氨基酸序列: MQIFVKLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLLIFAGKQLEDGRTLSDYNIQKESTLHLVLRGG)的多核苷酸;

[0028] (b) 序列如SEQ ID NO.:3所示的多核苷酸;

[0029] (c) 核苷酸序列与SEQ ID NO.:3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;

[0030] (d) 在SEQ ID NO.:3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;

[0031] (e) 与(a)-(d)任一所述的多核苷酸互补的多核苷酸。

[0032] 在另一优选例中,所述标签蛋白为野生型或优化的标签蛋白。

[0033] 在另一优选例中,所述标签蛋白选自下组:链霉亲和素、MBP、GST、Protein、CBP、6His、8His、或其组合。

[0034] 在另一优选例中,所述标签蛋白包括优化的链霉亲和素。

[0035] 在另一优选例中,所述标签蛋白带有Biotin标记。

[0036] 在另一优选例中,所述标签蛋白具有(Kd)不大于 10^{-6}mol/L , ($Kd \leq 10^{-6}\text{mol/L}$);或在 10^{-6}mol/L 至 10^{-17}mol/L 之间的编码的蛋白质结合Biotin的结合常数。

[0037] 在另一优选例中,所述标签蛋白的编码序列选自下组:

[0038] (a) 编码如SEQ ID NO.:6所示多肽(链霉亲和素的氨基酸序列: EAGITGTWYNQLGSTFIVTAGADGALTGTYESAVGNAESRYVLTGRYDSAPATDGSALTGWTVAWKNNYR NAHSATTWSGQYVGGAEARINTQWLLTSGTTEANAWKSTLVGHDTFTKVKPSAAS)的多核苷酸;

[0039] (b) 序列如SEQ ID NO.:5或8所示的多核苷酸;

[0040] (c) 核苷酸序列与SEQ ID NO.:5或8所示序列的同源性 $\geq 30\%$,较佳地,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;

[0041] (d) 在SEQ ID NO.:5或8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;

[0042] (e) 与(a)-(d)任一所述的多核苷酸互补的多核苷酸。

[0043] 在另一优选例中,所述标签蛋白具有SEQ ID NO.:6所示的序列或其活性片段,或者具有与SEQ ID NO:6所示氨基酸序列 $\geq 30\%$ 同源性(优选地, $\geq 85\%$ 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.:6序列相同活性的多肽。

[0044] 在另一优选例中,所述连接序列为密码子优化的连接序列。

[0045] 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-quadruplex等等),不富含稀有密码子。

[0046] 在另一优选例中,所述连接序列选自下组;

[0047] (i) 序列如SEQ ID NO.:7或9所示的多核苷酸;

[0048] (ii) 核苷酸序列与SEQ ID NO.:7或9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;

[0049] (iii) 在SEQ ID NO.:7或9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;

[0050] (iv) 与(i)-(iii)任一所述的多核苷酸互补的多核苷酸。

[0051] 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。

[0052] 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。

[0053] 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。

[0054] 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。

[0055] 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、荧光素酶突变、 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、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。

[0056] 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。

[0057] 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。

[0058] 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、Spacer间隔序列、其他供RNA转录、翻译的相关序列、或其组合。

[0059] 在另一优选例中,所述增强子元件包括IRES元件、RBS元件、非编码序列、或其组合。

[0060] 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。

- [0061] 在另一优选例中,所述真核细胞包括高等真核细胞。
- [0062] 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。
- [0063] 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)。
- [0064] 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。
- [0065] 本发明第二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。
- [0066] 本发明第三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一方面所述的载体或载体组合。
- [0067] 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。
- [0068] 在另一优选例中,所述真核细胞包括高等真核细胞。
- [0069] 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。
- [0070] 在另一优选例中,所述基因工程细胞为酵母细胞。
- [0071] 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。
- [0072] 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。
- [0073] 本发明第四方面提供了一种试剂盒,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:
- [0074] (a) 本发明第一方面所述的构建物;
- [0075] (b) 本发明第二方面所述的载体或载体组合;和
- [0076] (c) 本发明第三方面所述的基因工程细胞。
- [0077] 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)。
- [0078] 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。
- [0079] 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。
- [0080] 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。
- [0081] 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)。
- [0082] 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)。

[0083] 本发明第五方面提供了一种如本发明第一方面所述的构建物、本发明第二方面所述的载体或载体组合、本发明第三方面所述的基因工程细胞或本发明第四方面所述试剂盒的用途,用于进行高通量的体外蛋白合成。

[0084] 本发明第六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:

[0085] (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;

[0086] (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。

[0087] 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。

[0088] 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)。

[0089] 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)。

[0090] 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。

[0091] 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。

[0092] 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。

[0093] 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、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。

[0094] 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、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。

[0095] 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。

[0096] 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。

[0097] 本发明第七方面提供了一种分离的多核苷酸,所述多核苷酸选自下组:

[0098] (a) 编码如SEQ ID NO.:2所示多肽的多核苷酸;

[0099] (b) 序列如SEQ ID NO.:1所示的多核苷酸;

[0100] (c) 核苷酸序列与SEQ ID NO.:1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;

[0101] (d) 在SEQ ID NO.:1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;

[0102] (e) 与(a)-(d)任一所述的多核苷酸互补的多核苷酸。

[0103] 在另一优选例中,所述多核苷酸为编码前导肽的核苷酸序列。

[0104] 在另一优选例中,所述多核苷酸包括DNA序列。

[0105] 本发明第八方面提供了一种前导肽,所述前导肽的氨基酸序列选自下组:

[0106] (A) 具有SEQ ID NO:2所示氨基酸序列的多肽;(B) 具有与SEQ ID NO:2 所示氨基酸序列 $\geq 80\%$ 同源性(优选地, $\geq 90\%$ 的同源性;等优选地 $\geq 95\%$ 的同源性;最优选地, $\geq 97\%$ 的同源性,如98%以上,99%以上)的多肽,且所述多肽具有提高外源蛋白表达效率的功能或活性;

[0107] (C) 将SEQ ID NO:2中任一所示氨基酸序列经过1-15(较佳地,2-10,更佳地,3-8)个氨基酸残基的取代、缺失或添加而形成的,且具有提高外源蛋白表达效率的功能或活性的衍生多肽。

[0108] 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:

[0109] (i) 序列如SEQ ID NO.:7所示的多核苷酸;

[0110] (ii) 核苷酸序列与SEQ ID NO.:7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;

[0111] (iii) 在SEQ ID NO.:7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;

[0112] (iv) 与(i)-(iii)任一所述的多核苷酸互补的多核苷酸。

[0113] 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

[0114] 图1显示了mRNA被核糖体复合物识别,并翻译起始的过程。ATG起始密码子之后的20-300个核苷酸序列对于mRNA的翻译效率有决定性作用。

[0115] 图2显示了11种真核细胞一种或多种标签元件串联组合在酵母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),包括前导肽(leading peptide,lp)、泛素(UBA)相关序列、多His标签、Streptavidin序列(STN1/STN2,STN1是直接来自基因库调取的Streptavidin 氨基酸对应的核酸序列,STN2是在氨基酸序列不改变的前提下根据全序列特征采用同义密码子替换的方式修改了部分碱基)及其组合;将一种或多种标签元件与增强型绿色荧光蛋白编码序列之间通过Linker进行连接,形成了lp-L2_eGFP、UBA-L2_eGFP、STN2-L2_eGFP、STN1-L2_eGFP、8His-STN2-L2_eGFP、8His-STN1-L2_eGFP、lp-STN1-L2_eGFP、lp-8His-L2_eGFP、lp-8His-STN1-L2_eGFP、L2_PC、L1_PC和PC;部分含有表达纯化标签的实验组的相对荧光单位值高于300,蛋白表达量可以进行纯化实验。

[0116] 图3显示了在带有STN(STN1、STN2)核酸构建物的反应液中加入biotin beads去结合目标外源蛋白,并进行SDS-PAGE蛋白胶检测结果,考马斯亮蓝染色结果显示条带大小正确并清晰可见。STN1、STN2分别是不连接目标蛋白的实验组,STN1-eGFP、STN2-eGFP分别对应STN(STN1,STN2)后连接目标蛋白(eGFP)的实验组。

[0117] 图4显示了在带有多标签串联序列的核酸构建物(lp-8His-STN1和8His-STN1)以

及部分测试蛋白 (TP1-4) 的反应液中加入 biotin beads 去结合目标外源蛋白, 并进行 SDS-PAGE 蛋白胶检测结果, 考马斯亮蓝染色结果显示条带大小正确并清晰可见。

[0118] 图5显示了带有多标签串联序列的核酸构建物应用于蛋白质表达纯化的基本原理, 多标签串联序列连同目标蛋白编码序列被转录翻译之后, 通过与 biotin beads 特异性结合, 从而在反应液中抓取到目标蛋白。Tag I 用于提升蛋白质的合成效率, Tag II 用于提升蛋白质的纯化效果, Tag III 为连接序列, 部分序列可以同时兼具 Tag I、Tag II 和 Tag III 的功能 (如多 His 标签序列)。

具体实施方式

[0119] 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用于蛋白质表达纯化的 DNA 序列的新型核酸构建物, 本发明的核酸构建物由前导肽的编码序列 (如密码子优化的前导肽的编码序列)、任选的泛素的编码序列、体外蛋白质表达纯化序列 (如链霉亲和素、MBP、GST、Protein、CBP 等)、连接序列 (如密码子优化的连接序列) 以及外源蛋白的编码序列构成, 在任选的 5' UTR 之后使用。

[0120] 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, 所合成的外源蛋白 (如增强型绿色荧光蛋白) 活性的相对荧光单位值 (RFU) 非常高, 高达 705, 与不含多重标签串联的 DNA 序列的 PC 相比, 所合成的外源蛋白 (如增强型绿色荧光蛋白) 活性的相对荧光单位值 (RFU) 提高了 2 倍, 本发明还简化了外源蛋白的表达和纯化方式。在此基础上, 本发明人完成了本发明。

[0121] 蛋白质合成体系

[0122] 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。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。

[0123] 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。与传统的体内重组表达系统相比, 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。

[0124] 其中, 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。

[0125] 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。

[0126] 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)。

[0127] 在本发明中,所述体外蛋白质合成体系包括:

[0128] (a) 酵母细胞提取物;

[0129] (b) 聚乙二醇;

[0130] (c) 任选的外源蔗糖;和

[0131] (d) 任选的溶剂,所述溶剂为水或水性溶剂。

[0132] 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,萤光素酶DNA,RNA聚合酶。

[0133] 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。

[0134] 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%,更佳地,30-60%,更佳地,40-50%。

[0135] 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。

[0136] 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。

[0137] 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:

[0138] (i) 提供酵母细胞;

[0139] (ii) 对酵母细胞进行洗涤处理,获得经洗涤的酵母细胞;

[0140] (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;

[0141] (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。

[0142] 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。

[0143] 在一优选实施方式中,所述离心在液态下进行。

[0144] 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 5000-100000g,更佳地,8000-30000g。

[0145] 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-2 h,更佳地,20min-50min。

[0146] 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,更佳地,在2-6℃下进行。

[0147] 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。

[0148] 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。

[0149] 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1.0-2.0 mM。

[0150] 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0.02-0.2 mM,如0.05、0.06、0.07、0.08mM。

[0151] 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)。

[0152] 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。

[0153] 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/ μ L萤火虫荧光素酶的DNA,0.027-0.054mg/mL T7RNA 聚合酶。

[0154] 外源蛋白的编码序列(外源DNA)

[0155] 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。

[0156] 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。

[0157] 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。

[0158] 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。

[0159] 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。

[0160] 核酸构建物

[0161] 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:

[0162] Z1-Z2-Z3-Z4-Z5 (I)

[0163] 式中,

[0164] Z1、Z2、Z3、Z4、Z5分别为用于构成所述构建物的元件;

[0165] 各“-”独立地为键或核苷酸连接序列;

[0166] Z1为前导肽的编码序列;

[0167] Z2为无或泛素的编码序列;

[0168] Z3为标签蛋白的编码序列;

[0169] Z4为连接序列;

[0170] Z5为无或外源蛋白的编码序列;其中,所述前导肽的编码序列选自下组:

[0171] (a) 编码如SEQ ID NO.:2所示多肽的多核苷酸;

[0172] (b) 序列如SEQ ID NO.:1所示的多核苷酸;

[0173] (c) 核苷酸序列与SEQ ID NO.:1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;

[0174] (d) 在SEQ ID NO.:1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;

[0175] (e) 与(a)-(d)任一所述的多核苷酸互补的多核苷酸。

[0176] 在本发明中,Z1、Z2、Z3、Z4、Z5元件的来源没有特别限制。

[0177] 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、或其组合。

[0178] 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、或其组合。

[0179] 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。

[0180] 在一优选实施方式中,本发明的核酸构建物的序列如表1所示。

[0181] 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。

[0182] 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。

[0183] 在本发明中,在蛋白质合成体系中应用本发明的核酸构建物,并在反应完成后使用biotin-binding技术,可以完成外源蛋白的表达和纯化,具体地,应用本发明的核酸构建物所合成的增强型绿色荧光蛋白量的相对荧光单位值相对较高。

[0184] 载体,基因工程细胞

[0185] 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。只要其能够在宿主体内复制和稳定,任何质粒和载体都是可以采用的。

[0186] 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。

[0187] 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。

[0188] 优选地,所述的基因工程细胞为真核细胞。

[0189] 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)。

[0190] 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(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。

[0191] 体外高通量的蛋白合成方法

[0192] 本发明提供了一种体外高通量的蛋白合成方法,包括步骤:

[0193] (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;

[0194] (ii) 在适合的条件下,孵育步骤(i)的体外蛋白合成体系一段时间T1,从而合成所述外源蛋白。

[0195] 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。

[0196] 本发明的主要优点包括:

[0197] (1) 本发明首次发现,将任选的改变蛋白质起始翻译效率的编码序列、蛋白质表达纯化序列以及外源蛋白的编码序列构成的核酸构建物,应用于本发明的体外蛋白质合成体系中,可显著提高目标蛋白的产量,并可用于外源蛋白的表达与纯化。

[0198] (2) 本发明的蛋白质表达纯化序列能够特异性识别并高效结合beads。这些序列的识别结合效率不仅超过传统的特异性氨基酸biotin序列,而且可以应用于蛋白质合成体系(尤其是体外蛋白合成体系)。

[0199] (3) 本发明应用的可改变蛋白质起始翻译效率的编码序列可增强翻译起始的效率不仅超过传统的翻译增强子Omega序列及部分起始编码序列,而且可以应用于蛋白质合成体系(尤其是体外蛋白合成体系)。

[0200] (4) 与其他系统相比,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。

[0201] (5) 本发明提供的蛋白质表达纯化序列不仅能够提升目标外源蛋白的表达和纯化效果,更主要的是能够增加蛋白质合成体系的针对不同蛋白质合成的可能性。

[0202] (6) 本发明首次开发一种用于蛋白质表达纯化的DNA序列的新型核酸构建物,本发明的核酸构建物由前导肽的编码序列(如密码子优化的前导肽的编码序列)、任选的泛素的编码序列、体外蛋白质表达纯化序列(如链霉亲和素、MBP、GST、Protein、CBP等)、连接序列(如密码子优化的连接序列)以及外源蛋白的编码序列构成,本发明的核酸构建物可显著提高所合成的外源蛋白(如增强型绿色荧光蛋白)的活性,其相对荧光单位值(RFU)非常高,高达705,与不含多重标签串联的DNA序列相比,其相对荧光单位值(RFU)提高了2倍。

[0203] 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。

[0204] 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。

[0205] 外源蛋白以eGFP为例。

[0206] 实施例1:真核细胞改变蛋白质起始翻译效率的编码序列和蛋白质表达纯化序列

[0207] 1.1在基因库中找出相关的核酸序列(表1)。STN1序列是Streptavidin 基因序列,STN2序列是在氨基酸序列不改变前提下根据全序列特征采用同义密码子替换的方式修改了部分碱基。表1中标注Tag I的为改变蛋白质起始翻译效率的编码序列,标注Tag II的为蛋白质表达纯化结合序列,标注为Tag III 的为连接序列。

[0208] 表1 Multi-Tag中相关核酸序列

[0209]

	序列名称	核酸序列
SEQ ID NO. 1	前导肽 (Tag I)	attacagaaacatcatcaccgttcagatctatattctcc cacagtgggaaa
SEQ ID NO. 3	泛素 (Tag I)	CAGATTTTCGTGAAAACCCCTTACGGGGAAGACCATCACC CTCGAGGTTGAACCCTCGGATACGATAGAAAATGTAAAG GCCAAGATCCAGGATAAGGAAGGAATTCCTCCTGATCAG CAGAGACTGATCTTTGCTGGCAAGCAGCTGGAAGATGGA CGTACTTTGTCTGACTACAATATTCAAAGGAGTCTACT CTTCATCTTGTTGAGACTTCGTGGTGGT
SEQ ID NO. 5	STN2 (Tag II) (优化的 STN2)	GAAGCCGGTATCACAGGAACATGGTATAATCAGCTGGGT AGCACTCTGATCGTTACCGCTGGAGCAGATGGTGTCTCTG ACTGGCACCTATGTTACAGCACGTGGTAATGCTGAAGGC TCGTATGTCCTGACAGGTCGCTATGATTCTGCCCCGGCA ACTGATGGATCTGGAACAGCGCTGGGATGGACTGTTGCT TGGAAAAACAACACTATCGTAATGCCATTCCGCCACCACA TGGAGTGGTCAGTATGTTGGAGGCGCCGAAGCTCGTATT AATACACAGTGGCTGCTGACAAGTGGTACAACACTGAAGCC AATGCCTGGAATCAACCCTGGTTGGCCATGACACCTTT ACAAAAGTCAAACCGTCCGCCGCTTCT
SEQ ID NO. 8	STN1 (Tag II) (野生型的 STN1)	GAGGCAGGCATTACAGGCACGTGGTATAACCAGTTGGGG TCAACCTTTATCGTGACAGCCGGGGCCGACGGAGCATTG ACTGGGACCTATGAGAGTGCCGTTGGCAATGCTGAGTCA AGATATGTTCTTACCGGCAGGTACGATTCAGCTCCGGCA ACGGACGGGTCTGGAACGCTTGGGATGGACTGTTGCT TGGAAAAACAACACTATAGGAACGCTCATTCGGCCACAACA TGGTCGGGACAGTACGTCGGAGGTGCCGAGGCTAGGATA AACACGCAGTGGTTGCTAACATCAGGCACAACACTGAAGCC

[0210]

		AATGCCTGGAAGTCTACGCTTGTTGGACACGACACATTT ACCAAGGTGAAACCTAGTGCTGCATCC
SEQ ID NO. 10	6His (Tag I/Tag II/Tag III)	CACCATCACCACCATCAC
SEQ ID NO. 11	8His (Tag I/Tag II/Tag III)	CACCACCACCATCACCACCATCAC
SEQ ID NO. 9	L1 (Tag III) (未优化的 linker)	GGTAGTGGAGGAAGTGGTGGAAAGTGGGA
SEQ ID NO. 7	L2 (Tag III) (优化 的 linker)	GGTTCCTCCTGTTTGCATATGTTACCTCATCTTCACTTT GGCCTCAAAGACAAGGAAACA

[0211] 实施例2:含有真核细胞翻译调控序列和蛋白质表达纯化序列的体外蛋白质合成

体系质粒的构建

[0212] 2.1全基因合成:将STN1,STN2核酸序列、前导肽序列和泛素序列串联起来,并使用全基因合成的方法进行合成。His标签序列直接采用两对引物在模板上扩增。

[0213] 2.2质粒的构建:构建质粒模板为pD2P质粒,将翻译调控序列和蛋白质表达纯化序列插入到翻译起始密码子ATG和eGFP之间。不同的翻译调控序列和蛋白质表达纯化序列可组合,具体实验组合见表2。

[0214] 具体构建过程如下:

[0215] 以pD2P为模板进行质粒构建,使用两对引物分别进行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。

[0216] 表2.真核细胞翻译调控序列和蛋白质表达纯化序列及其组合

[0217]

编号	质粒名称	翻译调控序列和蛋白质表达纯化序列
pD2P-1.0M001	pD2P-STN2_L2_eGFP	STN2 (Tag II)
pD2P-1.0M002	pD2P-STN1_L2_eGFP	STN1 (Tag II)

[0218]

pD2P-1.0M003	pD2P-1p_L2_eGFP	前导肽 (Tag I)
pD2P-1.0M004	pD2P-UBA_L2_eGFP	泛素 (Tag I)
pD2P-1.0M005	pD2P-6H_L2_eGFP	6His (Tag I/Tag II/Tag III)
pD2P-1.0M006	pD2P-8H_L2_eGFP	8His (Tag I/Tag II/Tag III)
pD2P-1.0M007	pD2P-8H_STN2_L2_eGF P	8His (Tag I/Tag II/Tag III) + STN2 (Tag II)
pD2P-1.0M008	pD2P-8H_STN1_L2_eGF P	8His (Tag I/Tag II/Tag III) + STN1 (Tag II)
pD2P-1.0M009	pD2P-1p_STN1_L2_eGF P	前导肽 (Tag I) + STN1 (Tag II)
pD2P-1.0M010	pD2P-1p_8H_L2_eGFP	前导肽 (Tag I) + 8His (Tag I/Tag II/Tag III)
pD2P-1.0M011	pD2P-1p_8H_STN1_L2_ eGFP	前导肽 (Tag I) + 8His (Tag I/Tag II/Tag III) + STN1 (Tag II)

[0219] 实施例3:真核细胞翻译调控序列和蛋白质表达纯化序列在酵母体外蛋白质合成体系中的应用

[0220] 3.1利用PCR的方法,并使用引物pD2P_F: CGCGAAATTAATACGACTCACTATAGG和pD2P_R: TCCGGATATAGTTCCTCCTTTCAG将所有质粒中位于转录起始序列5`UTR和终止序列3`UTR之间包含一种或多种翻译调控序列和蛋白质表达纯化序列标签元件串联组合的片段和pD2P-eGFP片段进行扩增。

[0221] 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。

[0222] 3.2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,如图1所示。

[0223] 3.3将部分在反应液中加入biotin beads去结合目标外源蛋白,并进行 SDS-PAGE

蛋白胶检测,如图2所示。

[0224] 实验结果

[0225] 1. 蛋白质合成体系质粒的构建

[0226] 经过多次尝试,最终构建成功11个蛋白质合成体系质粒,其中包括5个含有翻译调控序列和蛋白质表达纯化序列以多种标签元件进行串联组合的质粒。

[0227] 2. 真核细胞多种标签元件串联组合在酵母体外蛋白质合成体系中的应用

[0228] 如图1所示,筛选的11个真核细胞多种标签元件串联组合在酵母体外蛋白质合成体系中引起增强型绿色荧光蛋白发出的相对荧光单位值(Relative Fluorescence Units, RFU)均达到较高值,部分带有蛋白质表达纯化序列的实验组活性高于200以至于可以进行蛋白纯化实验。

[0229] 本发明结果表明:真核细胞中多种标签元件串联组合序列可以提升目标蛋白的翻译效率并用于检测蛋白质表达纯化效果,并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达到或者超过不含标签的序列。增加了乳酸克鲁维酵母体外合成体系进行蛋白质表达纯化方式的选择性,极大增强了乳酸克鲁维酵母体外蛋白质合成体系的可用性。

[0230] 如图2所示,本发明的核酸构建物中,使用优化的前导肽编码序列,外源蛋白的RFU相比未优化的前导肽的编码序列,提高了1.5倍。本发明的核酸构建物中,使用优化的前导肽编码序列和多His标签串联的编码序列,提高了大约2倍。本发明的核酸构建物中,使用优化的linker,外源蛋白的RFU相比未优化的linker,提高了1.5倍。

[0231] 本发明的核酸构建物中,使用优化的前导肽编码序列和优化的linker构成的核酸构建物,所得到的外源蛋白的RFU相比使用未优化的前导肽编码序列和未优化的linker组合的核酸构建物,提高了2.25倍。

[0232] 如图3所示,利用STN1和STN2的亲亲和纯化序列进行SDS-PAGE跑胶并进行考马斯亮蓝染色检测,所得到的特异性目标条带非常明显,表明了经过优化后的标签的序列非常适合反应后的蛋白纯化实验。

[0233] 如图4所示,利用含有多标签序列的4种测试蛋白进行的纯化实验结果表明,可以看到清晰的目标条带,证明此序列拥有比较广泛的适用性。

[0234] 并且如图5所示,本发明的研究还发现,将结构简单、效率很高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,RBS元件),3`UTR与不同的多种标签元件串联组合序列Multi-tag进行组合,也可提高目标蛋白的翻译效率和获得非常好的蛋白表达纯化效果,其中,目标蛋白(eGFP)的RFU值高达705,与不含多重标签串联的DNA序列相比,其相对荧光单位值(RFU)提高了2倍。

[0235] 对比例

[0236] 在本发明的核酸构建物中,使用未优化的前导肽的编码序列,外源蛋白的RFU为375。

[0237] 在本发明的核酸构建物中,使用链霉亲和素本身的前导肽,外源蛋白的RFU大约为325。

[0238] 在本发明的核酸构建物中,无前导肽,外源蛋白的RFU为300。

[0239] 在本发明的核酸构建物中,无linker,外源蛋白的RFU为225。

[0240] 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。

[0241] 参考文献:

[0242] 1.Garcia RA,Riley MR.Applied biochemistry and biotechnology. Humana Press;1981.263-264.

[0243] 2.Fromm HJ,Hargrove M.Essentials of Biochemistry.2012.

[0244] 3.GräslundS,Nordlund P,Weigelt J,Hallberg BM,Bray J,Gileadi O,et al.Protein production and purification.Nat Methods. 2008;5(2):135-46.

[0245] 4.Korndã¶rferIP,Skerra A.Improved affinity of engineered streptavidin for the Strep-tag II peptide is due to a fixed open conformation of the lid-like loop at the binding site.Protein Sci. 2002;11(4):883-93.

[0246] 5.Hyre DE,Trong I Le,Merritt EA,Eccleston JF,Green NM, Stenkamp RE,et al.Cooperative hydrogen bond interactions in the streptavidin-biotin system.Protein Sci.2006;15(3):459-67。

序列表

<110> 康码(上海)生物科技有限公司

<120> 多重标签串联的DNA序列及其在蛋白质表达纯化系统的应用

<130> P2018-0725

<141> 2018-04-28

<160> 11

<170> SIPOSequenceListing 1.0

<210> 1

<211> 51

<212> DNA

<213> 人工序列 (artificial sequence)

<400> 1

attacagaaa catcatcacc gttcagatct atattctccc acagtgggaa a 51

<210> 2

<211> 17

<212> PRT

<213> 人工序列 (artificial sequence)

<400> 2

Ile Thr Glu Thr Ser Ser Pro Phe Arg Ser Ile Phe Ser His Ser Gly

1 5 10 15

Lys

<210> 3

<211> 225

<212> DNA

<213> 人工序列 (artificial sequence)

<400> 3

cagattttcg tgaaaaccct tacggggaag accatcacc tcgaggttga accctcggat 60
acgatagaaa atgtaaaggc caagatccag gataaggaag gaattcctcc tgatcagcag 120
agactgatct ttgctggcaa gcagctggaa gatggacgta ctttgtctga ctacaatatt 180
caaaaggagt ctactcttca tcttgtgttg agacttcgtg gtggt 225

<210> 4

<211> 76

<212> PRT

<213> 人工序列 (artificial sequence)

<400> 4

Met Gln Ile Phe Val Lys Thr Leu Thr Gly Lys Thr Ile Thr Leu Glu

1 5 10 15

Val Glu Pro Ser Asp Thr Ile Glu Asn Val Lys Ala Lys Ile Gln Asp

	20		25		30														
Lys	Glu	Gly	Ile	Pro	Pro	Asp	Gln	Gln	Arg	Leu	Ile	Phe	Ala	Gly	Lys				
	35		40		45														
Gln	Leu	Glu	Asp	Gly	Arg	Thr	Leu	Ser	Asp	Tyr	Asn	Ile	Gln	Lys	Glu				
	50		55		60														
Ser	Thr	Leu	His	Leu	Val	Leu	Arg	Leu	Arg	Gly	Gly								
65			70		75														

<210> 5
 <211> 378
 <212> DNA
 <213> 人工序列 (artificial sequence)
 <400> 5
 gaagccggta tcacaggaac atggtataat cagctgggta gcactctgat cgttaccgct 60
 ggagcagatg gtgctctgac tggcacctat gttacagcac gtggtaatgc tgaaggctcg 120
 tatgtcctga caggtcgeta tgattctgcc ccggaactg atggatctgg aacagcgctg 180
 ggatggactg ttgcttgaa aaacaactat cgtaatgccc attccgccac cacatggagt 240
 ggtcagtatg ttggaggcgc cgaagctcgt attaatacac agtggctgct gacaagtgg 300
 acaactgaag ccaatgcctg gaaatcaacc ctggttgccc atgacacctt taaaaagtc 360
 aaaccgtccg ccgcttct 378

	210		216		212														
Glu	Ala	Gly	Ile	Thr	Gly	Thr	Trp	Tyr	Asn	Gln	Leu	Gly	Ser	Thr	Phe				
1			5		10									15					
Ile	Val	Thr	Ala	Gly	Ala	Asp	Gly	Ala	Leu	Thr	Gly	Thr	Tyr	Glu	Ser				
			20		25									30					
Ala	Val	Gly	Asn	Ala	Glu	Ser	Arg	Tyr	Val	Leu	Thr	Gly	Arg	Tyr	Asp				
			35		40									45					
Ser	Ala	Pro	Ala	Thr	Asp	Gly	Ser	Gly	Thr	Ala	Leu	Gly	Trp	Thr	Val				
			50		55									60					
Ala	Trp	Lys	Asn	Asn	Tyr	Arg	Asn	Ala	His	Ser	Ala	Thr	Thr	Trp	Ser				
65			70		75									80					
Gly	Gln	Tyr	Val	Gly	Gly	Ala	Glu	Ala	Arg	Ile	Asn	Thr	Gln	Trp	Leu				
			85		90									95					
Leu	Thr	Ser	Gly	Thr	Thr	Glu	Ala	Asn	Ala	Trp	Lys	Ser	Thr	Leu	Val				
			100		105									110					
Gly	His	Asp	Thr	Phe	Thr	Lys	Val	Lys	Pro	Ser	Ala	Ala	Ser						

115	120	125
<210> 7		
<211> 60		
<212> DNA		
<213> 人工序列 (artificial sequence)		
<400> 7		
ggttctccct gtttgcataat gttacctcat cttcactttg gcctcaaaga caaggaaaca		60
<210> 8		
<211> 378		
<212> DNA		
<213> 人工序列 (artificial sequence)		
<400> 8		
gaggcaggca ttacaggcac gtggtataac cagttggggt caacctttat cgtgacagcc		60
ggggccgacg gagcattgac tgggacctat gagagtgccg ttggcaatgc tgagtcaaga		120
tatgttctta ccggcaggta cgattcagct ccggcaacgg acgggtctgg aactgccttg		180
ggatggactg ttgcttgaa aaacaactat aggaacgctc attcggccac aacatggtcg		240
ggacagtacg tcggagggtc cgaggctagg ataaacacgc agtggttgct aacatcaggc		300
acaactgaag ccaatgcctg gaagtctacg cttgttgac acgacacatt taccaaggtg		360
aaacctagtg ctgcatcc		378
<210> 9		
<211> 27		
<212> DNA		
<213> 人工序列 (artificial sequence)		
<400> 9		
ggtagtggag gaagtgggtg aagtgga		27
<210> 10		
<211> 18		
<212> DNA		
<213> 人工序列 (artificial sequence)		
<400> 10		
caccatcacc accatcac		18
<210> 11		
<211> 24		
<212> DNA		
<213> 人工序列 (artificial sequence)		
<400> 11		
caccaccacc atcaccacca tcac		24

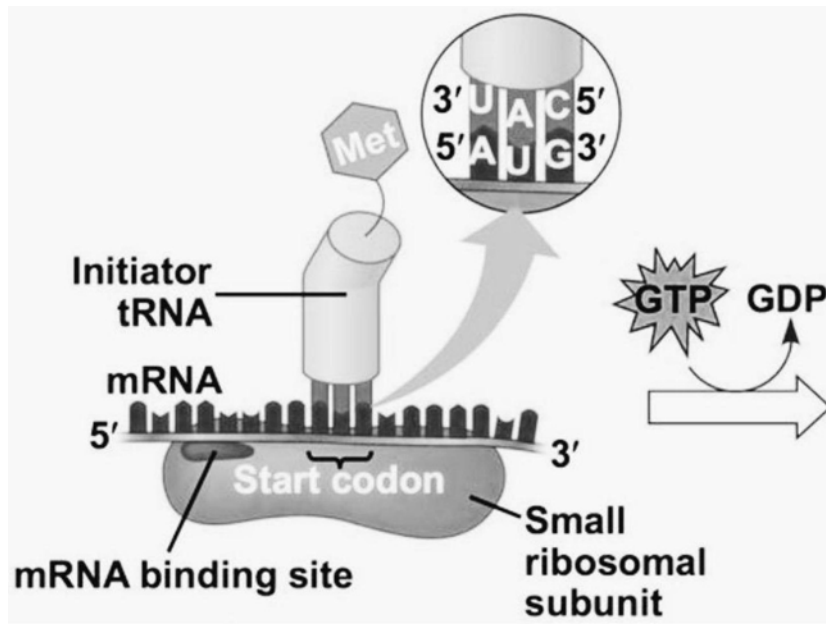


图1

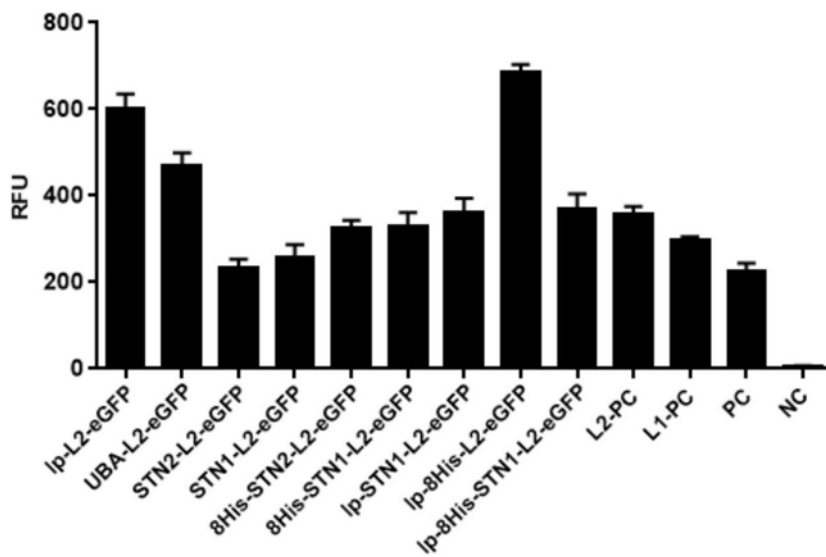


图2

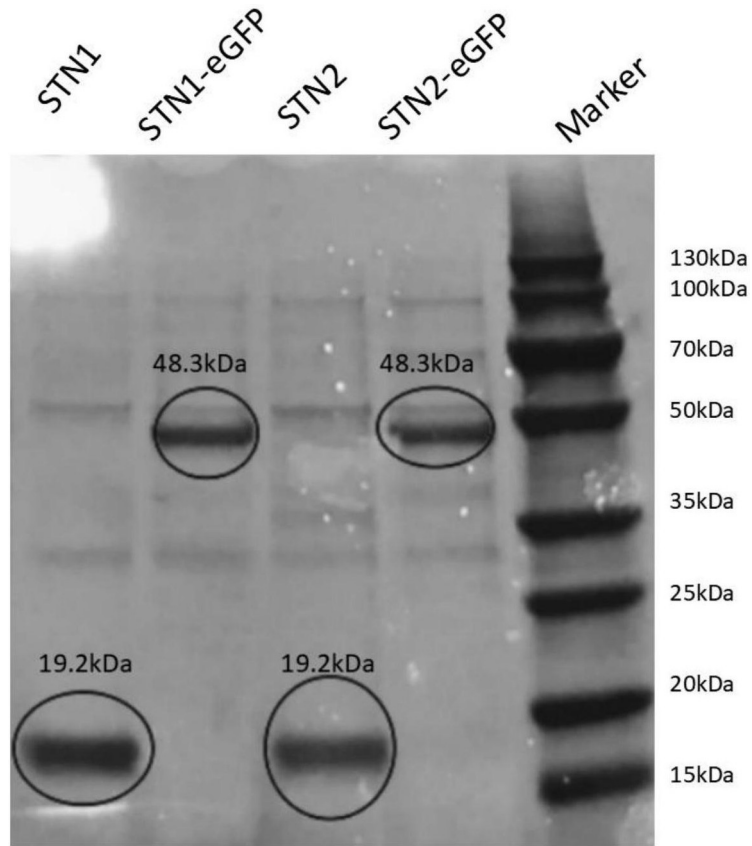


图3

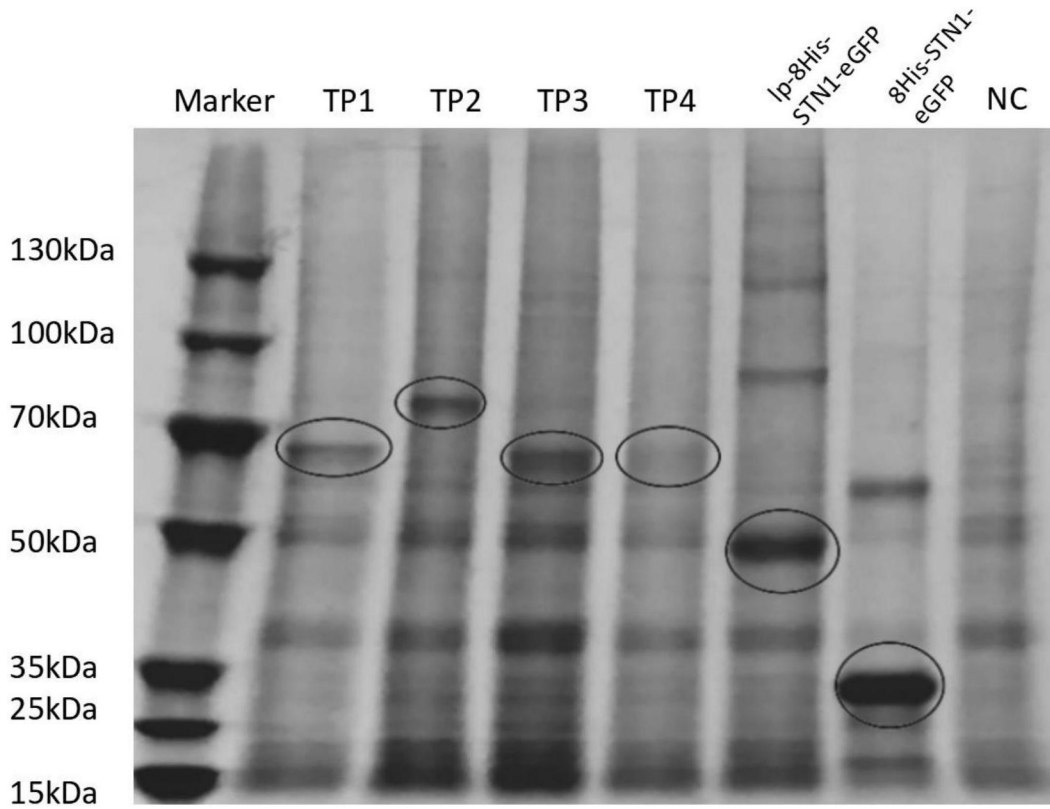
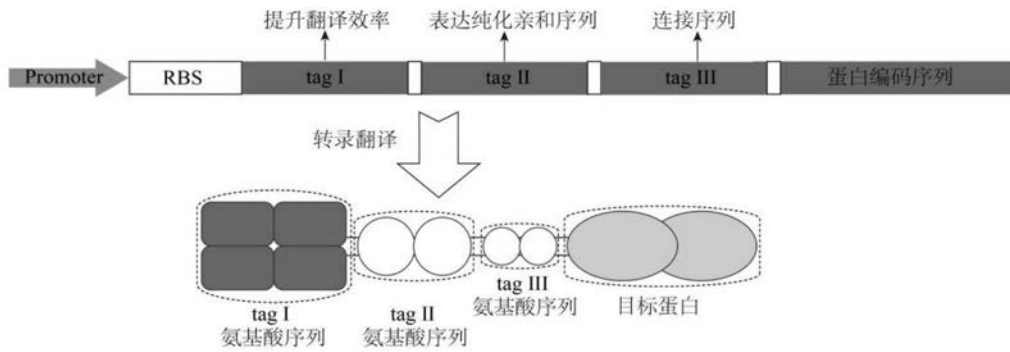


图4



tag I: 前导序列 (Leading Sequences)、泛素相关序列 (Ubiquitin-related Sequences)、多His标签等及其组合

tag II: 多His标签、STN相关序列等及其组合

tag III: 蛋白连接相关序列 (Linker)

图5