



(12)发明专利申请

(10)申请公布号 CN 110408635 A

(43)申请公布日 2019.11.05

(21)申请号 201810402739.3

(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 姜灵轩

王海鹏 杨宁 于雪

(51)Int.Cl.

C12N 15/62(2006.01)

C07K 19/00(2006.01)

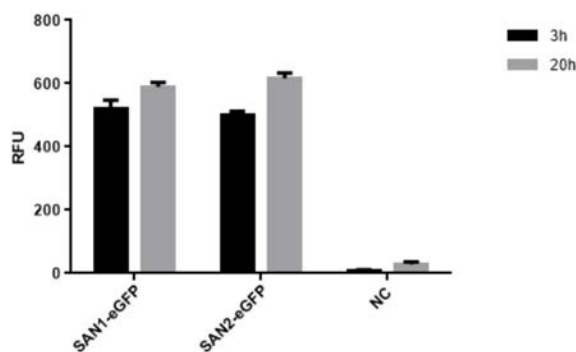
权利要求书2页 说明书18页  
序列表4页 附图2页

(54)发明名称

一种含有链霉亲和素元件的核酸构建物在蛋白质表达、纯化中的应用

(57)摘要

本发明提供了一种含有链霉亲和素元件的核酸构建物在蛋白质表达、纯化中的应用,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(I);式中,Z1、Z2、Z3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为标签蛋白的编码序列;Z2为连接序列;Z3为无或外源蛋白的编码序列,在本发明的蛋白质合成体系(尤其是体外蛋白合成体系)中应用本发明的核酸构建物,可以完成外源蛋白的表达和纯化,并且提高所合成的外源蛋白的RFU值。



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 (I)  
式中,  
Z1、Z2、Z3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为标签蛋白的编码序列;  
Z2为连接序列;  
Z5为无或外源蛋白的编码序列。
2. 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、或其组合。
3. 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.:3或5所示的多核苷酸;
  - (ii) 核苷酸序列与SEQ ID NO.:3或5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.:3或5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。
4. 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。
5. 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4所述的载体或载体组合。
6. 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4所述的载体或载体组合;和
  - (c) 权利要求5所述的基因工程细胞。
7. 一种如权利要求1所述的构建物、权利要求4所述的载体或载体组合、权利要求5所述的基因工程细胞或权利要求6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。
8. 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。
9. 如权利要求8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。
10. 一种连接序列,其特征在于,所述连接序列选自下组:
  - (i) 序列如SEQ ID NO.:3或5所示的多核苷酸;
  - (ii) 核苷酸序列与SEQ ID NO.:3或5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.:3或5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。

## 一种含有链霉亲和素元件的核酸构建物在蛋白质表达、纯化中的应用

### 技术领域

[0001] 本发明涉及生物技术领域,具体地,涉及一种含有链霉亲和素元件的核酸构建物在蛋白质表达、纯化中的应用。

### 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)。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] 酵母表达系统作为一种后起的外源蛋白表达系统,由于兼具原核以及真核表达系统的优点,正在基因工程领域中得到日益广泛的应用,应用此系统可高水平表达蛋白,且具有翻译后修饰功能,故被认可为一种表达大规模蛋白的强有力的工具(3,4)。

[0004] 在从酵母到人类的细胞内都发现了一些链霉亲和素(Streptavidin)、亲和素(Avidin)标签,而针对细胞内源的Streptavidin、Avidin标签的研究并不充分,主要是因为细胞内源性的Streptavidin标签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(4,5)。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的以结合biotin为基础的筛选(5,6)。

[0005] 除了以上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(7,8)。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)和来源系统(7)。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(7,9)。

[0006] 然而,在体外合成所使用的DNA模板通常不具有Streptavidin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(10)。而目前使用真核细胞Streptavidin元件在体外起始蛋白质的表达与纯化的应用还比较少。

[0007] 因此,本领域迫切需要开发一种在真核细胞中能够用于蛋白质表达和纯化的含有链霉亲和素的核酸构建物,将该核酸构建物应用于体外蛋白质表达系统中,可以较为方便的对蛋白进行纯化,提高纯化效率,降低成本。

### 发明内容

[0008] 本发明的目的在于提供一种在真核细胞中能够用于蛋白质表达和纯化的含有链霉亲和素的核酸构建物,将该核酸构建物应用于体外蛋白质表达系统中,可以较为方便的对蛋白进行纯化,提高纯化效率,降低成本。

[0009] 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:

[0010] Z1-Z2-Z3 (I)

[0011] 式中,

[0012] Z1、Z2、Z3分别为用于构成所述构建物的元件;

[0013] 各“-”独立地为键或核苷酸连接序列;

[0014] Z1为标签蛋白的编码序列;

[0015] Z2为连接序列;

[0016] Z5为无或外源蛋白的编码序列。

[0017] 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。

[0018] 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、或其组合。

[0019] 在另一优选例中,所述标签蛋白为优化的链霉亲和素。

[0020] 在另一优选例中,所述标签蛋白带有Biotin标记。

[0021] 在另一优选例中,所述标签蛋白具有(Kd)不大于 $10^{-6}$ mol/L, ( $Kd \leq 10^{-6}$ 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。

[0022] 在另一优选例中,所述标签蛋白的编码序列选自下组;

[0023] (a) 编码如SEQ ID NO.:2所示多肽(链霉亲和素的氨基酸序列:EAGITGTWYNQLGSTFIVTAGADGALTGTYESAVGNAESRYVLTGRYDSAPATDGSALTGWTVAWKNNYRNAHSATTWSGQYVGGAEARINTQWLLTSGTTEANAWKSTLVGHDTFTKVKPSAAS)的多核苷酸;

[0024] (b) 序列如SEQ ID NO.:1或4所示的多核苷酸;

[0025] (c) 核苷酸序列与SEQ ID NO.:1或4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;

[0026] (d) 在SEQ ID NO.:1或4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;

[0027] (e) 与(a)-(d)任一所述的多核苷酸互补的多核苷酸。

[0028] 在另一优选例中,所述标签蛋白具有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 $\geq 30\%$ 同源性,(优选地, $\geq 85\%$ 的同源性,更优选地 $\geq 90\%$ 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。

[0029] 在另一优选例中,所述连接序列为密码子优化的连接序列。

[0030] 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。

[0031] 在另一优选例中,所述连接序列可带有酶切位点。

[0032] 在另一优选例中,所述连接序列选自下组;

[0033] (i) 序列如SEQ ID NO.:3或5所示的多核苷酸;

- [0034] (ii) 核苷酸序列与SEQ ID NO.:3或5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;
- [0035] (iii) 在SEQ ID NO.:3或5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;
- [0036] (iv) 与(i)-(iii)任一所述的多核苷酸互补的多核苷酸。
- [0037] 在另一优选例中,所述连接序列如SEQ ID NO.:3或5所示。
- [0038] 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。
- [0039] 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。
- [0040] 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。
- [0041] 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。
- [0042] 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、 $\alpha$ -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。
- [0043] 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。
- [0044] 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。
- [0045] 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。
- [0046] 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。
- [0047] 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。
- [0048] 在另一优选例中,所述真核细胞包括高等真核细胞。
- [0049] 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。
- [0050] 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)。
- [0051] 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。
- [0052] 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。
- [0053] 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。

[0054] 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。

[0055] 在另一优选例中,所述真核细胞包括高等真核细胞。

[0056] 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。

[0057] 在另一优选例中,所述基因工程细胞为酵母细胞。

[0058] 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。

[0059] 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。

[0060] 本发明第四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:

[0061] (a) 本发明第一方面所述的构建物;

[0062] (b) 本发明第二方面所述的载体或载体组合;和

[0063] (c) 本发明第三方面所述的基因工程细胞。

[0064] 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)。

[0065] 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。

[0066] 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。

[0067] 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。

[0068] 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)。

[0069] 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)。

[0070] 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。

[0071] 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:

[0072] (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;

[0073] (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。

[0074] 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。

[0075] 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)。

[0076] 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)。

[0077] 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。

[0078] 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。

[0079] 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。

[0080] 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、 $\alpha$ -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。

[0081] 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。

[0082] 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。

[0083] 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。

[0084] 本发明第七方面提供了一种连接序列,所述连接序列选自下组:

[0085] (i) 序列如SEQ ID NO.:3或5所示的多核苷酸;

[0086] (ii) 核苷酸序列与SEQ ID NO.:3或5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;

[0087] (iii) 在SEQ ID NO.:3或5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;

[0088] (iv) 与(i)-(iii)任一所述的多核苷酸互补的多核苷酸。

[0089] 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

[0090] 图1显示了从DNA到蛋白质的基本生物过程。

[0091] 图2显示了2种真核细胞Streptavidin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);STN1(直接从基因库调取天然的Streptavidin链霉亲和素基因序列)和STN2(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)序列与增强型绿色荧光蛋白之间通过Linker2进行连接,形成了STN-Linker(命名为SAN);它们的相对荧光单位值均达到了400以上,其中,反应20小时

的相对荧光单位值分别达到了570和620,NC (Negative Control) 是不加入核酸构建物的实验组读值,蛋白表达量可以进行纯化实验。

[0092] 图3显示了在STN1和目标蛋白(eGFP)选用不同的Linker序列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),NC是不加入核酸构建物的实验组读值,可以看出,使用改进的Linker(Linker2和Linker3)的相对荧光单位值有显著提高。

[0093] 图4显示了在反应液中加入biotin beads去结合目标外源蛋白,并进行SDS-PAGE蛋白胶检测结果,结果显示条带大小正确并清晰可见。STN1、STN2分别是指STN(STN1,STN2)后不连接目标蛋白的实验组,SAN1-eGFP、SAN2-eGFP分别对应STN(STN1,STN2)后连接目标蛋白(eGFP)的实验组。

[0094] 图5显示了SAN应用于蛋白质表达纯化的基本原理,SAN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。为分离目标蛋白和SAN片段,可通过在俩片段之间插入酶切位点(如TEV酶切位点)进行酶切反应。

### 具体实施方式

[0095] 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含有可用于目标蛋白表达纯化的Streptavidin-Linker序列的新型核酸构建物,本发明的核酸构建物由任选的5'UTR(例如:启动子、翻译起始位点、酵母来源的IRES增强子,ScGPR1、ScFLO8、ScNCE102、ScMSN1、K1FLO8、K1NCE102、K1MSN1等)、可用于目标蛋白表达纯化的Streptavidin序列、连接Streptavidin序列和外源蛋白的编码序列的Linker(SAN)(优选密码子优化的linker或含有酶切位点的优化的linker)以及外源蛋白的编码序列和3'UTR构成,在本发明的蛋白质合成体系(尤其是体外蛋白合成体系)中应用本发明的核酸构建物,可以完成外源蛋白的表达和纯化,并且所合成的外源蛋白的RFU值可高达600以上,与不带SAN序列的核酸构建物相比,提高了2倍。在此基础上,本发明人完成了本发明。

#### [0096] 蛋白质合成体系

[0097] 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超过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。

[0098] 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。

[0099] 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。

[0100] 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。

[0101] 在本发明中,一种优选的蛋白质合成体系为体外蛋白质合成体系,在本发明中,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)。

[0102] 在本发明中,所述体外蛋白质合成体系包括:

[0103] (a) 酵母细胞提取物;

[0104] (b) 聚乙二醇;

[0105] (c) 任选的外源蔗糖;和

[0106] (d) 任选的溶剂,所述溶剂为水或水性溶剂。

[0107] 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。

[0108] 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。

[0109] 在本发明中,所述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%。

[0110] 在本发明中,所述的细胞提取物不含完整的细胞,典型的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,细胞提取物中还含有一些源自细胞的细胞质中的其他蛋白,尤其是可溶性蛋白。

[0111] 在本发明中,所述的细胞提取物所含蛋白含量为10-100mg/mL,较佳为20-80mg/mL。所述的测定蛋白含量方法为考马斯亮蓝测定方法。

[0112] 在本发明中,所述的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:

[0113] (i) 提供细胞;

[0114] (ii) 对细胞进行洗涤处理,获得经洗涤的细胞;

[0115] (iii) 对经洗涤的细胞进行破细胞处理,从而获得细胞粗提物;

[0116] (iv) 对所述细胞粗提物进行固液分离,获得液体部分,即为细胞提取物。

[0117] 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。

[0118] 在一优选实施方式中,所述离心在液态下进行。

[0119] 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,

较佳地,8000-30000g。

[0120] 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。

[0121] 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。

[0122] 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。

[0123] 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。

[0124] 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。

[0125] 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。

[0126] 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)。

[0127] 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。

[0128] 一种特别优选的体外蛋白质合成体系,除了细胞提取物之外,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,1.0-5.0mM醋酸镁,1.5-4mM核苷三磷酸混合物,0.08-0.24mM的氨基酸混合物,25mM磷酸肌酸,1.7mM二硫苏糖醇,0.27mg/mL磷酸肌酸激酶,1%-4%聚乙二醇,0.5%-2%蔗糖,8-20ng/ $\mu$ l萤火虫荧光素酶的DNA,0.027-0.054mg/mL T7RNA聚合酶。

[0129] 外源蛋白的编码序列(外源DNA)

[0130] 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。

[0131] 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。

[0132] 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、萤光素酶突变体的DNA、或其组合。

[0133] 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。

[0134] 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。

[0135] 核酸构建物

[0136] 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：

[0137] Z1-Z2-Z3 (I)

[0138] 式中，

[0139] Z1、Z2、Z3分别为用于构成所述构建物的元件；

[0140] 各“-”独立地为键或核苷酸连接序列；

[0141] Z1为标签蛋白的编码序列；

[0142] Z2为连接序列；

[0143] Z5为无或外源蛋白的编码序列。

[0144] 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。

[0145] 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。

[0146] 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。

[0147] 在一优选实施方式中，本发明的核酸构建物的序列表1所示。

[0148] 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。

[0149] 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。

[0150] 在本发明中,在本发明的体外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光单位值相对较高。

[0151] 载体,基因工程细胞

[0152] 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。

[0153] 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。

[0154] 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。

[0155] 优选地,所述的基因工程细胞为真核细胞。

[0156] 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)。

[0157] 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,如植物细胞、酵母细胞;或是高等真核细胞,如昆虫细胞、动物细胞等。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。

[0158] 体外高通量的蛋白合成方法

[0159] 本发明提供了一种体外高通量的蛋白合成方法,包括步骤:

[0160] (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;

[0161] (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。

[0162] 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。

[0163] 本发明的主要优点包括:

[0164] (1) 本发明首次发现,将任选的启动子、核糖体结合位点、IRES、用于目标蛋白表达纯化的STN(STN1,STN2)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,可用于外源蛋白的表达与纯化。

[0165] (2) 本发明的Streptavidin-Linker(SAN)氨基酸序列能够特异性识别并高效结合beads。这些氨基酸序列的识别结合效率不仅超过传统的特异性氨基酸序列,而且可以应用于酵母来源(尤其是乳酸克鲁维酵母来源)的体外蛋白质合成体系。

[0166] (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。

[0167] (4) 本发明提供的SAN氨基酸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。

[0168] (5) 本发明首次公开了一种由含有可用于目标蛋白表达纯化的Streptavidin-Linker序列构成的新型核酸构建物,在本发明的蛋白质合成体系(尤其是体外蛋白合成体系)中应用本发明的核酸构建物,可以完成外源蛋白的表达和纯化,并且所合成的外源蛋白的RFU值可高达600以上,与不带SAN序列相比,提高了2倍,与含有未优化的linker的核酸构建物相比,提高了3倍。

[0169] 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。

[0170] 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。

[0171] 外源蛋白以eGFP为例。

[0172] 实施例1真核细胞Streptavidin氨基酸序列确定

[0173] 1.1 Streptavidin氨基酸序列的确定:从基因库中找出相关的核酸序列和对应的氨基酸序列,并作为对应biotin序列,WtStreptavidin序列(STN1)是直接从基因库调取天然链霉亲和素序列,OurStreptavidin序列(STN2)则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。OurStreptavidin和WildtypeStreptavidin,所对应的氨基酸序列相同。Linker的序列确定取决于氨基酸序列尽量和天然蛋白相似,所对应的氨基酸序列的同源性不低于30%,核酸序列不含有强二级结构序列,发卡结构,G-quadruplex等(GC含量不定)。其中,在Linker3中含有TEV酶切位点序列,可用于酶切后得到纯的目标蛋白。

[0174] 表1. 乳酸克鲁维酵母中的相关核酸序列

[0175]

序列名称	核酸序列	SEQ ID NO. :
STN2 (优化的 STN2)	GAAGCCGGTATCACAGG AACATGGTATAATCAGC TGGGTAGCACTCTGATC GTTACCGCTGGAGCAGA TGGTGCTCTGACTGGCA CCTATGTTACAGCACGT GGTAATGCTGAAGGCTC GTATGTCCTGACAGGTC GCTATGATTCTGCCCCG GCAACTGATGGATCTGG AACAGCGCTGGGATGGA	1

[0176]

	CTGTTGCTTGAAAAAC AACTATCGTAATGCCCA TTCCGCCACCACATGGA GTGGTCAGTATGTTGGA GGCGCCGAAGCTCGTAT TAATACACAGTGGCTGC TGACAAGTGGTACAAC GAAGCCAATGCCTGGAA ATCAACCCTGGTTGGCC ATGACACCTTTACAAAA GTCAAACCGTCCGCCGC TTCT	
STN1 (野生型的 STN1)	GAGGCAGGCATTACAGG CACGTGGTATAACCAGT TGGGGTCAACCTTTATC GTGACAGCCGGGGCCGA CGGAGCATTGACTGGGA CCTATGAGAGTGCCGTT GGCAATGCTGAGTCAAG ATATGTTCTTACCGGCA GGTACGATTACAGCTCCG GCAACGGACGGGTCTGG AACTGCCTTGGGATGGA CTGTTGCTTGAAAAAC AACTATAGGAACGCTCA TTCGGCCACAACATGGT CGGGACAGTACGTCGGA GGTGCCGAGGCTAGGAT AAACACGCAGTGGTTGC TAACATCAGGCACAAC GAAGCCAATGCCTGGAA	4

	GTCTACGCTTGTTGGAC ACGACACATTTACCAAG GTGAAACCTAGTGCTGC ATCC		
[0177]	Linker1 (未优化的 linker)	GGTAGTGGAGGAAGTGG TGGAAGTGGA	6
	Linker1.5 (未优化的 linker)	GGGTCGCCCTGTTTGCA TATGTTCCCTCATCTGC ACTTCGGCCTCAACGAC AAGGAGACG	7
	Linker2 (优化的 linker)	GGTTCTCCCTGTTTGCA TATGTTACCTCATCTTC ACTTTGGCCTCAAAGAC AAGGAAACA	3
	Linker3 (优化的 linker, 且含有酶切位点)	GGAAAAATATCAACGGA GAACTTGTATTTCCAGG GAGGTTCTCCCTGTTTG CATATG	5

[0178] 实施例2含有真核细胞Streptavidin核酸序列的体外蛋白质合成体系质粒的构建

[0179] 2.1全基因合成:将人工设计的2个STN核酸序列 (STN1, STN2) 串联起来,并使用全基因合成的方法进行合成。

[0180] 2.2质粒的构建:对于人工设计的2个STN核酸序列 (STN1, STN2) 使用两对引物,后缀为1f和1b的引物从合成的DNA片段上分别扩增相应的STN片段,后缀为2f和2b的引物则用于对pD2P-eGFP质粒的PCR扩增。在最终构建的质粒中,2个STN核酸序列 (STN1, STN2) 插入到pD2P-eGFP质粒的AUG起始密码子和eGFP之间。质粒的名称分别为:pD2P-STN2\_eGFP和pD2P-STN1\_eGFP。

[0181] 通过PCR的方法将Linker2序列插入已有质粒pD2P-STN\_eGFP中,插入到STN (STN1, STN2) 核酸序列和外源蛋白编码序列 (eGFP) 之间,分别构建含有STN-Linker (SAN) 的质粒pD2P-SAN1\_eGFP和pD2P-SAN2\_eGFP,名称及序列号都列在表2中。

[0182] 具体构建过程如下:

[0183] 对于pD2P-STN\_eGFP的质粒构建,使用两对引物分别进行PCR扩增,并各取10 $\mu$ L扩增产物进行混合;向20 $\mu$ L扩增产物中加入1 $\mu$ L Dpn I,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。

[0184] 表2.引物序列

[0185]

质粒名称	引物名称	引物序列	SEQ ID NO. :
pD2P-SAN2_eGFP	STN2-615_1f	CAATTACAAAAAAAAAAAA ATGTCTGGTAGTGAAGCC GGTATCACAGGAACATGG	8
	STN2-eGFP-64_1b	CCTCCCTGGAAATACAAG TTCTCCGTTGATATTTTT CCAGAAGCGGCGGACGGT TTGAC	9
	STN2-eGFP-605_2f	GGAAAAATATCAACGGAG AACTTGTATTTCCAGGGA GGTAGTGGAGGAAGTGGT GGAAG	10
	STN2-59_2b	ATACCGGCTTCACTACCA GACATTTTTTTTTTTTGT AATTGTAATAGTAATTG TAATGTTGTTTGTG	11
pD2P-SAN1_eGFP	STN1-60_1f	CAATTACTATTTACAATT ACAAAAAAAAAAAAATGTC TGGTAGTGAGGCAGGCAT TACAG	12

[0186]	STN1-615_1 b	ATTGGCATCCAAGTAATT AATCCTTATTTACCCCTG GAAATACAAGTTCTCCGT TG	13
	STN1-61_2f	GTATTTCCAGGGGTAAAT AAGGATTAATTACTTGGA TGCCAATAAAAAAAAAAAA AGCG	14
	STN1-59_2b	CACTACCAGACATTTTTT TTTTTTGTAATTGTAAAT AGTAATTGTAATGTTGTT TGTTG	15

[0187] 实施例3:STN核酸序列在体外蛋白质合成体系中的应用

[0188] 3.1利用PCR的方法,并使用引物pD2P\_F:CGCGAAATTAATACGACTCACTATAGG (SEQ ID NO.:16和pD2P\_R:TCCGGATATAGTTCCTCCTTTTCAG (SEQ ID NO.:17) 将所有质粒中位于T7转录起始序列和终止序列之间包含SAN序列的片段和pD2P-eGFP片段进行扩增。

[0189] 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。

[0190] 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,相对光单位值(Relative Fluorescence Unit,RFU)作为活性单位,如图1所示。

[0191] 3.3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。

[0192] 实验结果

[0193] 1.体外蛋白质合成体系质粒的构建

[0194] 经过多次尝试,最终构建成功4个体外蛋白质合成体系质粒,包括所有2个含有eGFP检测蛋白的质粒和2个不带eGFP检测蛋白的质粒(质粒名称是pD2P-SAN1\_eGFP、pD2P-SAN2\_eGFP、pD2P-STN1和pD2P-STN2)。

[0195] 2.SAN序列在体外蛋白质合成体系中的应用

[0196] 如图2所示,筛选的2个SAN序列在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(反应3小时

后RFU值达到400以上),与含有linker3的核酸构建物的RFU值相当,比含有linker1.5的核酸构建物的RFU(含有linker1.5的核酸构建物的RFU仅为含有linker2的核酸构建物的外源蛋白的RFU的60%)提高了3倍,可以进行蛋白纯化实验。

[0197] 本发明结果表明:SAN序列可以用于检测蛋白质表达纯化效果,并能够应用于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常用的tag序列。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。

[0198] 并且,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(如KLNCE102)与不同的STN序列,3'-UTR等组合,也可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600以上,与不带SAN序列相比,其相对荧光单位值(RFU)提高了2倍。

[0199] 对比例

[0200] 本发明的核酸构建物中,无linker,外源蛋白的RFU为85。

[0201] 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。

[0202] 参考文献

[0203] 1.Garcia RA,Riley MR.Applied biochemistry and biotechnology.Humana Press,1981.263-264.

[0204] 2.Fromm HJ,Hargrove M.Essentials of Biochemistry.2012;

[0205] 3.Kenny E,Hinchliffe E.Yeast expression system.EP;1993.

[0206] 4.Gräslund S,Nordlund P,Weigelt J,Hallberg BM,Bray J,Gileadi O,et al.Protein production and purification.Nat Methods.2008;5(2):135-46.

[0207] 5.Korndörfer IP,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.

[0208] 6.Voss S,Skerra A.Mutagenesis of a flexible loop in streptavidin leads to higher affinity for the Strep-tag II peptide and improved performance in recombinant protein purification.Protein Eng.1997;10(8):975-82.

[0209] 7.Katzen F,Chang G,Kudlicki W.The past,present and future of cell-free protein synthesis.Trends Biotechnol.2005;23(3):150-6.

[0210] 8.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.

[0211] 9.Lu Y.Cell-free synthetic biology:Engineering in an open world.Synth Syst Biotechnol[Internet].2017;2(1):23-7.Available from:<http://linkinghub.elsevier.com/retrieve/pii/S2405805X1730008X>

[0212] 10.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> 一种含有链霉亲和素元件的核酸构建物在蛋白质表达、纯化中的应用

<130> P2018-0724

<141> 2018-04-28

<160> 15

<170> SIPOSequenceListing 1.0

<210> 1

<211> 378

<212> DNA

<213> 人工序列 (artificial sequence)

<400> 1

```

gaagccggta tcacaggaac atggtataat cagctgggta gcactctgat cgttaccgct 60
ggagcagatg gtgctctgac tggcacctat gttacagcac gtggtaatgc tgaaggctcg 120
tatgtcctga caggtcgcta tgattctgcc ccggcaactg atggatctgg aacagcgctg 180
ggatggactg ttgcttggaa aaacaactat cgtaatgccc attccgccac cacatggagt 240
ggtcagtatg ttggaggcgc cgaagctcgt attaatacac agtggctgct gacaagtgg 300
acaactgaag ccaatgcctg gaaatcaacc ctggttgccc atgacacctt taaaaagtc 360
aaaccgtccg ccgcttct 378

```

<210> 2

<211> 126

<212> PRT

<213> 人工序列 (artificial sequence)

<400> 2

```

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> 3

<211> 60

<212> DNA

<213> 人工序列 (artificial sequence)

<400> 3

ggttctccct gtttgcataat gttacctcat cttcactttg gcctcaaaga caaggaaaca 60

<210> 4

<211> 378

<212> DNA

<213> 人工序列 (artificial sequence)

<400> 4

gaggcaggca ttacaggcac gtggtataac cagttggggt caacctttat cgtgacagcc 60

ggggccgacg gagcattgac tgggacctat gagagtgccg ttggcaatgc tgagtcaaga 120

tatgttctta ccggcaggta cgattcagct ccggcaacgg acgggtcttg aactgccttg 180

ggatggactg ttgcttgaa aaacaactat aggaacgctc attcggccac aacatggtcg 240

ggacagtacg tcggagggtc cgaggctagg ataaacacgc agtggttgct aacatcaggc 300

acaactgaag ccaatgctg gaagtctacg cttgttgac acgacacatt taccaaggtg 360

aaacctagtg ctgcatcc 378

<210> 5

<211> 57

<212> DNA

<213> 人工序列 (artificial sequence)

<400> 5

ggaaaaatat caacggagaa cttgtatttc cagggaggtt ctccctgttt gcatatg 57

<210> 6

<211> 27

<212> DNA

<213> 人工序列 (artificial sequence)

<400> 6

ggtagtggag gaagtgggtg aagtgga 27

<210> 7

<211> 60

<212> DNA

<213> 人工序列 (artificial sequence)

<400> 7

gggtgcacct gtttgcataat gttccctcat ctgcacttcg gcctcaacga caaggagacg 60

<210> 8

<211> 54  
<212> DNA  
<213> 人工序列 (artificial sequence)  
<400> 8  
caattacaaa aaaaaaaaaat gtctggtagt gaagccgta tcacaggaac atgg 54  
<210> 9  
<211> 59  
<212> DNA  
<213> 人工序列 (artificial sequence)  
<400> 9  
cctccctgga aatacaagtt ctccgttgat atttttccag aagcggcgga cggtttgac 59  
<210> 10  
<211> 59  
<212> DNA  
<213> 人工序列 (artificial sequence)  
<400> 10  
ggaaaaatat caacggagaa ctgtatttc caggagga gtggagga tggtggaag 59  
<210> 11  
<211> 69  
<212> DNA  
<213> 人工序列 (artificial sequence)  
<400> 11  
ataccggctt cactaccaga cattttttt tttgtaatt gtaaatagta attgtaatgt 60  
tgtttggtg 69  
<210> 12  
<211> 59  
<212> DNA  
<213> 人工序列 (artificial sequence)  
<400> 12  
caattactat ttacaattac aaaaaaaaaa atgtctggt agtgaggcag gcattacag 59  
<210> 13  
<211> 56  
<212> DNA  
<213> 人工序列 (artificial sequence)  
<400> 13  
attggcatcc aagtaattaa tccttatta ccctgaaa tacaagttct cgttg 56  
<210> 14  
<211> 58  
<212> DNA

<213> 人工序列 (artificial sequence)

<400> 14

gtatttccag gggtaaataa ggattaatta cttggatgcc aataaaaaaa aaaaagcg 58

<210> 15

<211> 59

<212> DNA

<213> 人工序列 (artificial sequence)

<400> 15

cactaccaga catttttttt ttttgaatt gtaaatagta attgtaatgt tgtttgttg 59

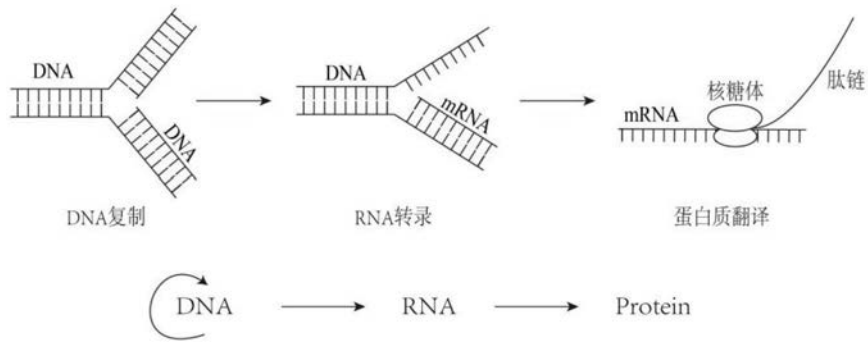


图1

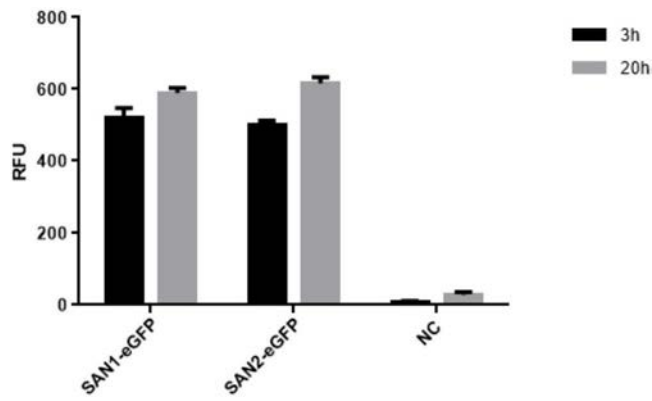


图2

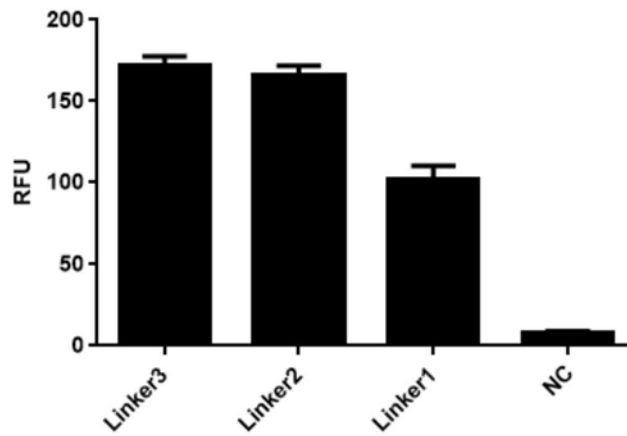


图3

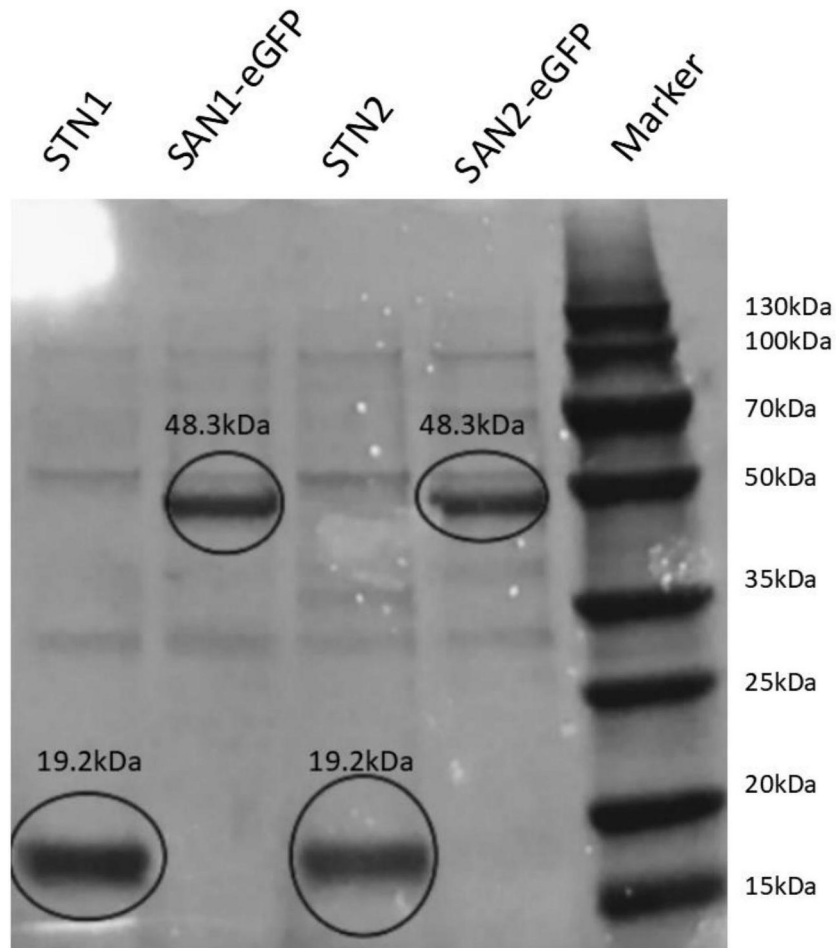


图4

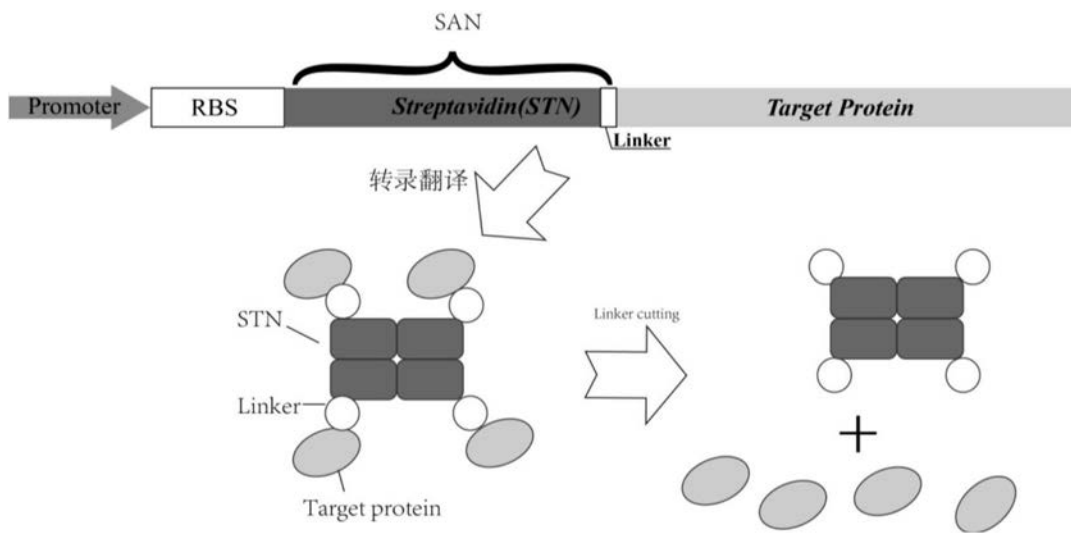


图5