(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 115806976 A
(43)申请公布日 2023.03.17

(21)申请号 202211124976.0
(22)申请日 2019.03.20
(62)分案原申请数据
201910214234.9 2019.03.20
(71)申请人 中国科学院上海营养与健康研究所
地址 200031 上海市徐汇区岳阳路319号
(72)发明人 王泽峰 杨赟 樊晓娟
(74)专利代理机构 上海一平知识产权代理有限公司 31266
专利代理师 王正君 徐迅
(51)Int.Cl.
C12N 15/11(2006.01)
C12N 15/63(2006.01)
C12N 5/10(2006.01)
C12N 1/19(2006.01)
C12R 1/91(2006.01)
C12R 1/865(2006.01)
C12R 1/645(2006.01)

权利要求书3页 说明书12页
序列表（电子公布） 附图1页

(54)发明名称

利用环形RNA进行蛋白翻译及其应用

(57)摘要

本发明提供了一种利用环形RNA进行蛋白翻译及其应用，具体地，本发明提供一种环形RNA构建物，所述环形RNA构建物从5’-3’方向具有式I所示的结构：TI-Z1-Z2(I)，式中，TI为翻译起始元件；Z1为表达外源蛋白的表达盒；Z2为无或其他元件。本发明的环形RNA构建物在体内体外均可显著增强翻译效率。

CN 115806976 A

1.一种环形RNA构建物,其特征在于,所述环形RNA构建物从5’-3’方向具有式I所示的结构:

TI-Z1-Z2 (I)

式中,

TI为翻译起始元件;

Z1为表达外源蛋白的表达盒;

Z2为无或其他元件;

并且,各“-”为键或核苷酸连接序列;

其中,TI元件的长度为6-30nt,较佳地,8-24nt,更佳地,10-20nt;

所述TI元件的编码序列选自下组:

(a)序列如SEQ ID NO.:3－40所示的多核苷酸;

(b)核苷酸序列与SEQ ID NO.:3－40所示序列的同源性≥75%的多核苷酸;

(c)核苷酸序列与SEQ ID NO.:3－40所示序列的同源性≥85%的多核苷酸;

(d)核苷酸序列与SEQ ID NO.:3－40所示序列的同源性≥90%的多核苷酸;

(e)核苷酸序列与SEQ ID NO.:3－40所示序列的同源性≥95%的多核苷酸;

(f)核苷酸序列与SEQ ID NO.:3－40所示序列的同源性≥98%的多核苷酸;

(g)核苷酸序列与SEQ ID NO.:3－40所示序列的同源性≥99%的多核苷酸;

(h)在SEQ ID NO.:3-40所示多核苷酸的5’端和/或3’端截短或添加1-18个核苷酸的多核苷酸;

(i)与(a)-(h)任一所述的多核苷酸互补的多核苷酸。

2.如权利要求1所述的构建物,其特征在于,所述TI元件的编码序列为在SEQ ID NO.:3-40所示多核苷酸的5’端和/或3’端截短或添加1-10个核苷酸的多核苷酸。

3.如权利要求2所述的构建物,其特征在于,所述TI元件的编码序列为在SEQ ID NO.:3-40所示多核苷酸的5’端和/或3’端截短或添加1-6个核苷酸的多核苷酸。

4.一种载体,其特征在于,所述载体含有权利要求1所述的构建物的表达盒。

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。

6.一种反应体系,其特征在于,包括:

(a)权利要求1所述的构建物;

(b)反应所需的其他组分,所述其他组分选自下组:剪接小体、核糖体、翻译起始因子EIF4G2、翻译起始因子EIF4A、翻译起始因子EIF4B、或其组合。

7.如权利要求6所述的反应体系,其特征在于,所述反应体系还包括YTHDF3、PABPC1、和/或hnRNPA1蛋白。

8.一种体外合成蛋白质的方法,其特征在于,包括步骤:

(i)提供权利要求6所述的反应体系;

(ii)在适合的条件下,孵育步骤(i)的合成体系一段时间T1,从而合成所述的蛋白质。

9.一种用于体外蛋白合成的试剂盒,其特征在于,包括:

(k1)第一容器,以及位于第一容器内的权利要求1所述的构建物;

(k2)第二容器,以及位于第二容器内的反应所需的其他组分,所述其他组分选自下组:

剪接小体、核糖体、翻译起始因子EIF4G2、翻译起始因子EIF4A、翻译起始因子EIF4B、或其组合;和

(kt)标签或说明书。

10.如权利要求9所述的试剂盒,其特征在于,所述试剂盒还包括任选的选自下组的容器:

(k3)第三容器,以及位于第三容器内的YTHDF3、PABPC1、和/或hnRNPA1蛋白。

11.一种权利要求1所述的构建物、权利要求4所述的载体、权利要求5所述的基因工程细胞、权利要求6所述的反应体系、权利要求8所述的体外合成蛋白质的方法、或权利要求9所述的试剂盒的用途,其特征在于,用于进行高通量的体外蛋白合成。

12.一种环形RNA构建物,其特征在于,所述环形RNA构建物从5'-3'方向具有式I所示的结构:

TI-Z1-Z2(I)

式中,

TI为翻译起始元件;

Z1为表达外源蛋白的表达盒;

Z2为无或其他元件;

并且,各"-"为键或核苷酸连接序列;

其中,TI元件的长度为6-30nt,较佳地,8-24nt,更佳地,10-20nt;

所述TI元件含有表2所示的选自下组的一种或多种核苷酸序列:

表2

1	TAATAG
2	AAATAG
3	AACAAA。

13.一种载体,其特征在于,所述载体含有权利要求12所述的构建物的表达盒。

14.一种基因工程细胞,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12所述的构建物,或者所述基因工程细胞中含有权利要求13所述的载体。

15.一种反应体系,其特征在于,包括:

(a)权利要求12所述的构建物;

(b)反应所需的其他组分,所述其他组分选自下组:剪接小体、核糖体、翻译起始因子EIF4G2、翻译起始因子EIF4A、翻译起始因子EIF4B、或其组合。

16.如权利要求15所述的反应体系,其特征在于,所述反应体系还包括YTHDF3、PABPC1、和/或hnRNPA1蛋白。

17.一种体外合成蛋白质的方法,其特征在于,包括步骤:

(i)提供权利要求15所述的反应体系;

(ii)在适合的条件下,孵育步骤(i)的合成体系一段时间T1,从而合成所述的蛋白质。

18.一种用于体外蛋白合成的试剂盒,其特征在于,包括:

(k1)第一容器,以及位于第一容器内的权利要求12所述的构建物;

(k2)第二容器,以及位于第二容器内的反应所需的其他组分,所述其他组分选自下组:剪接小体、核糖体、翻译起始因子EIF4G2、翻译起始因子EIF4A、翻译起始因子EIF4B、或其组

合;和

(kt)标签或说明书。

19.如权利要求18所述的试剂盒,其特征在于,所述试剂盒还包括任选的选自下组的容器:

(k3)第三容器,以及位于第三容器内的YTHDF3、PABPC1、和/或hnRNPA1蛋白。

20.一种权利要求12所述的构建物、权利要求13所述的载体、权利要求14所述的基因工程细胞、权利要求15所述的反应体系、权利要求17所述的体外合成蛋白质的方法、或权利要求18所述的试剂盒的用途,其特征在于,用于进行高通量的体外蛋白合成。

利用环形RNA进行蛋白翻译及其应用

[0001] 本申请是申请日为2019年3月20日、申请号为201910214234.9、发明名称为“利用环形RNA进行蛋白翻译及其应用”的发明专利申请的分案申请。

技术领域

[0002] 本发明涉及生物技术领域，具体地，涉及利用环形RNA进行蛋白翻译及其应用。

背景技术

[0003] 蛋白质是生物体内最重要的生物大分子，其突变或者表达异常都会导致疾病发生。因而，通过蛋白质替代或者表达可以治疗相应的疾病。常见的蛋白质替代或者表达疗法包括基于核糖核酸(DNA)载体的递送系统、基于脱氧核糖核酸(RNA)载体的递送系统以及蛋白质递送系统。这些方法都需要通过信使RNA翻译产生蛋白质。在真核生物体内常见的翻译起始方式是帽依赖性翻译，其主要通过翻译起始因子识别信使RNA 5′端的特殊帽结构来起始翻译，该类翻译方式只存在于线性信使RNA中存在。此外，还存在着一类非帽依赖性的翻译起始，其主要通过特定蛋白因子与RNA元件间的互作启动翻译。这类翻译可以在线性或者环形RNA中起始翻译。常见的非帽依赖性翻译起始元件是病毒RNA中一些具有特定二级结构的元件，它们可以借用宿主细胞的翻译系统表达自己所需的蛋白质。例如，脑心肌炎病毒或者丙型肝炎病毒等RNA中含有的内部核糖体进入位点(Internal Ribosome Entry Site：IRES)元件。

[0004] 环形RNA是一类不同于线性RNA的单链闭环RNA形式，因其结构特异性导致其不易被核酸外切酶降解，顾其相较于线性RNA更稳定。所以，通过环形RNA翻译表达蛋白具有更加持续长效的特点，是替代线性RNA翻译的一种重要手段。但因环状RNA的翻译只能利用非帽依赖性翻译的反式，如何设计并选择合适的非帽翻译起始元件是该应用的关键技术。一种常见的手段是利用病毒IRES启动环形RNA的翻译，但由于致病病毒RNA在宿主体内可能存在免疫排斥反应，同时病毒来源的RNA元件基本都含有复杂的RNA二级结构而且较长的序列，限制了基于病毒IRES的表达系统构建及后期的基因治疗应用。

[0005] 因此，本领域迫切需要开发一种非病毒来源的非帽依赖性翻译起始元件。

发明内容

[0006] 本发明的目的在于提供一种非病毒来源的非帽依赖性翻译起始元件。

[0007] 本发明第一方面提供了一种环形RNA构建物，所述环形RNA构建物从5’-3’方向具有式I所示的结构：

[0008] TI-Z1-Z2(I)

[0009] 式中，

[0010] TI为翻译起始元件；

[0011] Z1为表达外源蛋白的表达盒；

[0012] Z2为无或其他元件；

[0013] 并且,各“-”为键或核苷酸连接序列;

[0014] 其中,TI元件的长度为6-30nt,较佳地,8-24nt,更佳地,10-20nt;

[0015] 在TI元件中,A的含量≥35%,较佳地,≥45%,更佳地,≥60%;

[0016] 在TI元件中,T的含量≥20%,较佳地,≥30%,更佳地,≥50%;

[0017] 在TI元件中,A+T的含量≥65%,较佳地,≥80%,更佳地,≥90%;

[0018] 在TI元件中,G的含量≤35%,较佳地,≤25%,更佳地,≤10%。

[0019] 在另一优选例中,所述环形RNA构建物为环形信使RNA构建物。

[0020] 在另一优选例中,所述TI元件中,A的含量为35-100%,较佳地,45-100%,更佳地,60-100%。

[0021] 在另一优选例中,所述TI元件中,T的含量为20-100%,较佳地,30-100%,更佳地,50-100%。

[0022] 在另一优选例中,所述TI元件中,A+T的含量为65-100%,较佳地,80-100%,更佳地,90-100%。

[0023] 在另一优选例中,所述TI元件中,G的含量为0-35%,较佳地,0-25%,更佳地,0-10%。

[0024] 在另一优选例中,所述TI元件含有表1所示的选自下组的一种或多种核苷酸序列:

[0025] 表1

[0026]

1	AATATA	11	ATAAAT	21	AAAATT	31	TAAATA	41	AATAAG	51	TATACT
2	AAAATA	12	ATATTA	22	AAATAC	32	TTATAA	42	AAATTC	52	AGATAT
3	AAATAT	13	AATAAT	23	TAATAT	33	AGAAGA	43	GAGATA	53	TCAAGC
4	AAATAA	14	TATATA	24	CATATA	34	ATTATA	44	CAAAAA	54	AAGAAT
5	AATAAA	15	ATAATA	25	ATTAAT	35	TATAAA	45	AATTTA	55	AAACAT
6	ATATAA	16	AATTAA	26	ACATAT	36	AACATA	46	AAGATA	56	ATTATT
7	AAAAAA	17	ATAAAA	27	AATACA	37	TTAATA	47	AACATT	57	ACAAAA
8	AAATTA	18	TATAAT	28	ATACAA	38	TATATT	48	ATTAGG	58	AAAAGA
9	ATATAT	19	ATAAGA	29	AATTAT	39	AATACT	49	ACATAA	59	AATCAA
10	AAAAAT	20	AATATT	30	ATATAG	40	ATATAC	50	GAAGAA	60	AAAGAC

[0027]

61	TAAGAA	71	GGAGAT	81	TAAACA	91	TATACA
62	ATAAAC	72	TAATCT	82	CGAAAC	92	GACATA
63	TAGATT	73	TAAAAA	83	TATTAA	93	TGAATA
64	ATAAAG	74	AAATCC	84	AATAGA	94	TAAGAC
65	AATATC	75	ATCAAG	85	AATTCA	95	AACTGA
66	TAATAA	76	ATACTG	86	ATAAGT	96	TTATAT
67	ATTCGA	77	CATTAG	87	AAACAA	97	TTTAAA
68	TATTTT	78	TGACAT	88	ATACTA	98	TAATAG
69	TAATTA	79	ATTTAA	89	ATATCT	99	AAATAG
70	TATATG	80	AGATTA	90	AAGAAG	100	AACAAA。

[0028] 在另一优选例中,所述TI元件在表1所示的核苷酸序列的5’端和/或3’端添加1-24个(较佳地1-15,更佳地1-10个,更佳地,1－6个)核苷酸,且具有TI元件的功能。

[0029] 在另一优选例中,所述TI元件的编码序列选自下组;

[0030] (a)序列如SEQ ID NO.:1－40所示的多核苷酸;

[0031] (b)核苷酸序列与SEQ ID NO.:1－40所示序列的同源性≥75%(较佳地≥85%,更佳地≥90%或≥95%或≥98%或≥99%)的多核苷酸;

[0032] (c)在SEQ ID NO.:1-40所示多核苷酸的5'端和/或3'端截短或添加1-18个(较佳地1-10,更佳地1-6个)核苷酸的多核苷酸;

[0033] (d)与(a)-(c)任一所述的多核苷酸互补的多核苷酸。

[0034] 在另一优选例中,所述TI元件具有如SEQ ID NO.:1-40所示的序列。

[0035] 在另一优选例中,所述TI元件的编码序列如SEQ ID NO.:1-40所示。

[0036] 在另一优选例中,所述Z1元件含有终止密码子。

[0037] 在另一优选例中,所述Z1元件不含有终止密码子。

[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、萤光素酶突变体的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、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αA、白细胞介素-1β、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Z2元件选自下组:PolyA、多克隆位点、适配体、miRNA结合位点、翻译增强元件、或其组合。

[0044] 在另一优选例中,所述TI元件的一个或多个腺嘌呤(A)被甲基化。

[0045] 在另一优选例中,所述环形RNA的构建物的序列如SEQ ID NO.:61所示。

[0046] 本发明第二方面提供了一种载体,所述载体含有本发明第一方面所述的构建物的表达盒。

[0047] 在另一优选例中,所述表达盒含有第一内含子和第二内含子。

[0048] 在另一优选例中,所述第一内含子和第二内含子完全互补或不完全互补。

[0049] 在另一优选例中,所述载体具有SEQ ID NO.:62所示的序列。

[0050] 在另一优选例中,所述第一内含子的序列如SEQ ID NO.:63所示。

[0051] 在另一优选例中,所述第二内含子的序列如SEQ ID NO.:64所示。

[0052] 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核酸构建物,或者所述基因工程细胞中含有本发明第二方面所述的载体。

[0053] 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。

[0054] 在另一优选例中，所述真核细胞包括高等真核细胞。

[0055] 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。

[0056] 在另一优选例中，所述基因工程细胞为酵母细胞。

[0057] 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。

[0058] 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。

[0059] 本发明第四方面提供了一种反应体系，包括：

[0060] (a)本发明第一方面所述的构建物；

[0061] (b)反应所需的其他组分，所述其他组分选自下组：剪接小体、核糖体、翻译起始因子EIF4G2、翻译起始因子EIF4A、翻译起始因子EIF4B、或其组合。

[0062] 在另一优选例中，所述反应体系还包括YTHDF3、PABPC1、和/或hnRNPA1蛋白。

[0063] 在另一优选例中，所述反应体系为体外反应体系。

[0064] 本发明第五方面提供了一种体外合成蛋白质的方法，包括步骤：

[0065] (i)提供本发明第四方面所述的反应体系；

[0066] (ii)在适合的条件下，孵育步骤(i)的合成体系一段时间T1，从而合成所述的蛋白质。

[0067] 另一优选例中，所述的方法还包括：(iii)任选地从所述体外反应体系中，分离或检测所述的蛋白质。

[0068] 在另一优选例中，所述步骤(ii)中，反应温度为25－42℃，较佳地，30－40℃，更佳地，35－37℃。

[0069] 在另一优选例中，所述步骤(ii)中，反应时间T1为1小时－20小时，较佳地，2小时－12小时，更佳地，3小时－6小时。

[0070] 本发明第六方面提供了一种用于体外蛋白合成的试剂盒，包括：

[0071] (k1)第一容器，以及位于第一容器内的本发明第一方面所述的构建物；

[0072] (k2)第二容器，以及位于第二容器内的反应所需的其他组分，所述其他组分选自下组：剪接小体、核糖体、翻译起始因子EIF4G2、翻译起始因子EIF4A、翻译起始因子EIF4B、或其组合；和

[0073] (kt)标签或说明书。

[0074] 在另一优选例中，所述的第一容器、第二容器是同一容器或不同容器。

[0075] 在另一优选例中，所述试剂盒还包括任选的选自下组的一个或多个容器：

[0076] (k3)第三容器，以及位于第三容器内的YTHDF3、PABPC1、和/或hnRNPA1蛋白。

[0077] 本发明第七方面提供了一种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、本发明第三方面所述的基因工程细胞、本发明第四方面所述的反应体系、或本发明第六方面所述的试剂盒的用途，用于进行高通量的体外蛋白合成。

[0078] 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

[0079] 图1显示了流式细胞仪筛选获得的不同细胞群体。

[0080] 图2显示了western blot检测翻译起始元件特征序列的活性。

[0081] 图3显示了western blot检测对抗学习产生的翻译起始元件的活性。

具体实施方式

[0082] 经过广泛而深入的研究,通过大量筛选和摸索,首次意外地筛到一种特定的翻译起始元件,该翻译起始元件具有很高的翻译活性,将本发明的该翻译启示元件插入环形RNA表达载体,在体内体外均可显著增强翻译效率。在此基础上,本发明人完成了本发明。

[0083] 第一内含子

[0084] 在本发明中,第一内含子3'端含有一个剪接受体位点,内部包含一段与第二内含子配对的顺式元件(长度50bp-300bp)。

[0085] 序列如下所示:

[0086] GACTGAACATGGAGGAATTGAGGTTGGGTATTTCCCCTGAGGTAGGAAAAAGGCTGGGTCAGTTTCCCGTTAGCCGTCAAGTCCTCATCACATCTTTAAGCCTTCCATGCAGGATAAAGGGCTGCAGAGCTATTTTCAAATTGACATCAAACTGGATTTCTGTTGACTTCGTCTTCCCTTTTTAAGGTCCACAGAAGAAGATGGGAAGGAAAGAAGTCTGAGGGCATCTTATTTGCACTCCGCTGTCATTTCTAAGGAAGGCCTTTAATGCCAAATTCTCATCTTTTATGTCCCCACTAAATCCTAAGGTTCTTGAACTTCTGATCAGACAGCCAAAAAATGAACCATCAACTAGCTTAACCTAACATATGTGAGGATAGAGGACTGGGACAGCTCTCTGGGCCACTGGAGAGTCAGACAGGCCTGCCCTCTGTGTGACTTGACCGCGGTCTCTTTCTTCCAG(SEQ ID NO.:63)

[0087] 第二内含子

[0088] 在本发明中,第二内含子5'端含有一个剪接供体位点,内部包含一段与第一内含子配对的顺式元件(长度50bp-300bp)。

[0089] 序列如下所示:

[0090] GTAAGTCTCGACGGATCCCAAATAAGATGCCCTCAGACTTCTTTCCTTCCCATCTTCTTCTGTGGACCTTAAAAAGGGAAGACGAAGTCAACAGAAATCCAGTTTGATGTCAATTTGAAAATAGCTCTGCAGCCCTTTATCCTGCATGGAAGGCTTAAAGATGTGATGAGGACTTGACGGCTAACGGGAAACTGACCCAGCCTTTTTCCTACCTCAGGGGAAATACCCAACCTCAATTCCTCCATGTTCAG(SEQ ID NO.:64)

[0091] 环形RNA构建物

[0092] 本发明第一方面提供了一种环形RNA构建物,所述环形RNA构建物从5'-3'方向具有式I所示的结构:

[0093] TI-Z1-Z2 (I)

[0094] 式中,

[0095] TI为翻译起始元件;

[0096] Z1为表达外源蛋白的表达盒;

[0097] Z2为无或其他元件;

[0098] 并且,各"-"为键或核苷酸连接序列;

[0099] 其中,TI元件的长度为6-30nt,较佳地,8-24nt,更佳地,10-20nt;

[0100] 在TI元件中,A的含量≥35%,较佳地,≥45%,更佳地,≥60%;

[0101] 在TI元件中,T的含量≥20%,较佳地,≥30%,更佳地,≥50%;

[0102] 在TI元件中,A+T的含量≥65%,较佳地,≥80%,更佳地,≥90%;

[0103] 在TI元件中,G的含量≤35%,较佳地,≤25%,更佳地,≤10%。

[0104] 在一优选实施方式中,所述TI元件中,A的含量为35-100%,较佳地,45-100%,更佳地,60-100%。

[0105] 在一优选实施方式中,所述TI元件中,T的含量为20-100%,较佳地,30-100%,更佳地,50-100%。

[0106] 在一优选实施方式中,所述TI元件中,A+T的含量为65-100%,较佳地,80-100%,更佳地,90-100%。

[0107] 在一优选实施方式中,所述TI元件中,G的含量为0-35%,较佳地,0-25%,更佳地,0-10%。

[0108] 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。

[0109] 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αA、白细胞介素-1β、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。

[0110] 此外,本发明的所述核酸构建物是环状的。本发明的所述核酸构建物是单链的。本发明的所述核酸构建物是RNA。

[0111] 在一优选实施方式中,本发明的环形RNA构建物的序列如SEQ ID NO.:61所示。

[0112] 以GFP为例的环形RNA序列:

[0113] 5'-AAAAAAAAAAATCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAA-3'(SEQ ID NO.:61)

[0114] 在一优选实施方式中,以GFP为例的环形RNA前体(含有第一内含子、第二内含子)的序列如下所示:

[0115] 5'-GACTGAACATGGAGGAATTGAGGTTGGGTATTTCCCCTGAGGTAGGAAAAAGGCTGGGTCAGTTTCCCGTTAGCCGTCAAGTCCTCATCACATCTTTAAGCCTTCCATGCAGGATAAAGGGCTGCAGAGCTATTTTCAAATTGACATCAAACTGGATTTCTGTTGACTTCGTCTTCCCTTTTTAAGGTCCACAGAAGAAGATGGGAAGGAAAGAAGTCTGAGGGCATCTTATTTGCACTCCGCTGTCATTTCTAAGGAAGGCCTTTAATGCCAAATTCTCATCTTTTATGTCC

CCACTAAATCCTAAGGTTCTTGAACTTCTGATCAGACAGCCAAAAAATGAACCATCAACTAGCTTAACCTAACATA
TGTGAGGATAGAGGACTGGGACAGCTCTCTGGGCCACTGGAGAGTCAGACAGGCCTGCCCTCTGTGTGACTTGACC
GCGGTCTCTTTCTTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGT
TCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCA
CAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAAC
TTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCG
ACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCG
CGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAAAA
AAAAAAAATCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGAC
GTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCA
TCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAG
CCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGTAAGTCTC
GACGGATCCCAAATAAGATGCCCTCAGACTTCTTTCCTTCCCATCTTCTTCTGTGGACCTTAAAAAGGGAAGACGA
AGTCAACAGAAATCCAGTTTGATGTCAATTTGAAAATAGCTCTGCAGCCCTTTATCCTGCATGGAAGGCTTAAAGA
TGTGATGAGGACTTGACGGCTAACGGGAAACTGACCCAGCCTTTTTCCTACCTCAGGGGAAATACCCAACCTCAAT
TCCTCCATGTTCAG-3’(SEQ ID NO.:62)

[0116] 在一优选实施方式中,本发明的TI元件含有表1所示的选自下组的核苷酸序列:

[0117] 表1

[0118]

1	AATATA	11	ATAAAT	21	AAAATT	31	TAAATA	41	AATAAG	51	TATACT
2	AAAATA	12	ATATTA	22	AAATAC	32	TTATAA	42	AAATTC	52	AGATAT
3	AAATAT	13	AATAAT	23	TAATAT	33	AGAAGA	43	GAGATA	53	TCAAGC
4	AAATAA	14	TATATA	24	CATATA	34	ATTATA	44	CAAAAA	54	AAGAAT
5	AATAAA	15	ATAATA	25	ATTAAT	35	TATAAA	45	AATTTA	55	AAACAT
6	ATATAA	16	AATTAA	26	ACATAT	36	AACATA	46	AAGATA	56	ATTATT
7	AAAAAA	17	ATAAAA	27	AATACA	37	TTAATA	47	AACATT	57	ACAAAA
8	AAATTA	18	TATAAT	28	ATACAA	38	TATATT	48	ATTAGG	58	AAAAGA
9	ATATAT	19	ATAAGA	29	AATTAT	39	AATACT	49	ACATAA	59	AATCAA
10	AAAAAT	20	AATATT	30	ATATAG	40	ATATAC	50	GAAGAA	60	AAAGAC

[0119]

61	TAAGAA	71	GGAGAT	81	TAAACA	91	TATACA
62	ATAAAC	72	TAATCT	82	CGAAAC	92	GACATA
63	TAGATT	73	TAAAAA	83	TATTAA	93	TGAATA
64	ATAAAG	74	AAATCC	84	AATAGA	94	TAAGAC
65	AATATC	75	ATCAAG	85	AATTCA	95	AACTGA
66	TAATAA	76	ATACTG	86	ATAAGT	96	TTATAT
67	ATTCGA	77	CATTAG	87	AAACAA	97	TTTAAA

[0120]

68	TATTTT	78	TGACAT	88	ATACTA	98	TAATAG
69	TAATTA	79	ATTTAA	89	ATATCT	99	AAATAG
70	TATATG	80	AGATTA	90	AAGAAG	100	AACAAA

[0121] 在一优选实施方式中,本发明的TI元件的编码序列如SEQ ID NO.:1-40所示。

[0122] 在本发明中,应用本发明的环形RNA构建物具有很高的翻译活性,可显著增强体内或体外的翻译效率。

[0123] 反应体系

[0124] 本发明提供了一种反应体系,包括:

[0125] (a)本发明第一方面所述的构建物;

[0126] (b)反应所需的其他组分,所述其他组分选自下组:剪接小体、核糖体、翻译起始因子EIF4G2、翻译起始因子EIF4A、翻译起始因子EIF4B、或其组合。

[0127] 在另一优选例中,所述反应体系还包括YTHDF3、PABPC1、和/或hnRNPA1蛋白。

[0128] 在本发明中,反应体系可以在体外也可以在体内。

[0129] 试剂盒

[0130] 本发明提供了用于体外蛋白合成的试剂盒,包括:

[0131] (k1)第一容器,以及位于第一容器内的权利要求1所述的核酸构建物;

[0132] (k2)第二容器,以及位于第二容器内的反应所需的其他组分,所述其他组分选自下组:剪接小体、核糖体、翻译起始因子EIF4G2、翻译起始因子EIF4A、翻译起始因子EIF4B、或其组合;和

[0133] (kt)标签或说明书。

[0134] 在一优选实施方式中,所述的第一容器、第二容器是同一容器或不同容器。

[0135] 外源蛋白的编码序列(外源DNA)

[0136] 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。

[0137] 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。

[0138] 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。

[0139] 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αA、白细胞介素-1β、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。

[0140] 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。

[0141] 体外蛋白合成方法

[0142] 本发明提供了一种体外蛋白合成方法,包括步骤:

[0143] (i)提供本发明第二方面所述的反应体系;

[0144] (ii)在适合的条件下,孵育步骤(i)的合成体系一段时间T1,从而合成所述的蛋白质。

[0145] 另一优选例中,所述的方法还包括:(iii)任选地从所述反应体系中,分离或检测所述的蛋白质。

[0146] 本发明的主要优点包括:

[0147] (1)本发明首次开发了一套用以设计合成具有高翻译活性同时序列结构和长度可控的非病毒来源的新型人工合成的真核生物翻译起始元件的方法,并以此翻译起始元件驱动环形RNA的翻译。

[0148] (2)本发明首次筛选到一种特定的翻译起始元件,该翻译起始元件非常短,只有6-30nt,但是具有很高的翻译活性,并且将该翻译起始元件插入到环形RNA表达载体,在体内体外均可显著增强翻译效率。

[0149] 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。

[0150] 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。

[0151] 外源蛋白以GFP为例。

[0152] 实验方法

[0153] 1.基于本发明特有的一种环状RNA报告基因系统(该环状RNA报告基因可以在翻译起始元件的驱动下表达产生绿色荧光蛋白)构建一套含有百万个不同序列的文库,通过细胞转染及流式细胞仪筛选不同的细胞群体(阴性:没有绿色荧光,阳性:具有不同强度的绿色荧光)。将收集的不同细胞群体进行扩增子测序并结合计算生物学分析解析阴性和不同阳性细胞内包含的序列信息,根据这些序列信息从中提取不同长度的序列特征;

[0154] 2.以这些阴性和阳性序列特征为基础,通过对抗神经网络学习生成非病毒来源的翻译起始元件模块;

[0155] 3.将生成的翻译起始元件模块及模块组合元件插入环状RNA表达载体,通过细胞转染和western blot检测翻译起始元件的活性。

[0156] 实验结果

[0157] 1.基于环状RNA的高通量筛选系统分离出含有不同绿色荧光强度的细胞群体(阳性)及没有荧光的细胞群体(阴性)。

[0158] 结果如表1和图1所示。结果表明,环形RNA系统可以表达绿色荧光蛋白并用于筛选。同时,该系统可以分离出具有不同荧光强度的细胞群体,说明文库中不同的插入序列对环形RNA的翻译起始具有差异化的效果。

[0159] 2.通过筛选获得的不同细胞群体的扩增子测序数据的计算生物学分析挖掘获得

的高绿色荧光细胞群体包含的序列特征，这些序列通常都含有大量AU碱基较少包含易形成RNA二级结构的GC碱基，表明该类翻译起始元件的核心模块不依赖于结构同时也表明其与病毒IRES具有明显差异，是一类新的可促进环形RNA的非帽依赖翻译的元件，如表1所示。

[0160] 表1具有高绿色荧光信号的细胞群体所含翻译起始元件的序列特征(以6个碱基形式展示的前100个序列特征)

[0161]

1	AATATA	11	ATAAAT	21	AAAATT	31	TAAATA	41	AATAAG	51	TATACT
2	AAAATA	12	ATATTA	22	AAATAC	32	TTATAA	42	AAATTC	52	AGATAT
3	AAATAT	13	AATAAT	23	TAATAT	33	AGAAGA	43	GAGATA	53	TCAAGC
4	AAATAA	14	TATATA	24	CATATA	34	ATTATA	44	CAAAAA	54	AAGAAT
5	AATAAA	15	ATAATA	25	ATTAAT	35	TATAAA	45	AATTTA	55	AAACAT
6	ATATAA	16	AATTAA	26	ACATAT	36	AACATA	46	AAGATA	56	ATTATT
7	AAAAAA	17	ATAAAA	27	AATACA	37	TTAATA	47	AACATT	57	ACAAAA
8	AAATTA	18	TATAAT	28	ATACAA	38	TATATT	48	ATTAGG	58	AAAAGA
9	ATATAT	19	ATAAGA	29	AATTAT	39	AATACT	49	ACATAA	59	AATCAA
10	AAAAAT	20	AATATT	30	ATATAG	40	ATATAC	50	GAAGAA	60	AAAGAC

[0162]

61	TAAGAA	71	GGAGAT	81	TAAACA	91	TATACA
62	ATAAAC	72	TAATCT	82	CGAAAC	92	GACATA
63	TAGATT	73	TAAAAA	83	TATTAA	93	TGAATA
64	ATAAAG	74	AAATCC	84	AATAGA	94	TAAGAC
65	AATATC	75	ATCAAG	85	AATTCA	95	AACTGA
66	TAATAA	76	ATACTG	86	ATAAGT	96	TTATAT
67	ATTCGA	77	CATTAG	87	AAACAA	97	TTTAAA
68	TATTTT	78	TGACAT	88	ATACTA	98	TAATAG
69	TAATTA	79	ATTTAA	89	ATATCT	99	AAATAG
70	TATATG	80	AGATTA	90	AAGAAG	100	AACAAA

[0163] 3.通过细胞转染和western blot检测翻译起始元件基础特征序列的翻译活性(6个碱基的序列特征)，结果如图2所示。结果表明，这些翻译元件均具有翻译起始功能，但不同序列具有的翻译起始能力有差异。这表明基于这些不同的序列特征组合将可能可以得到具有不同活性的翻译起始元件。

[0164] 4.基于阳性和阴性细胞群体特征序列的对抗学习产生的翻译起始元件(以12个碱基为例)。分别列出具有不同翻译活性强度的前20个序列。高活性的翻译起始元件基本为AT富集序列，如表2所示。

[0165] 表2对抗学习产生的翻译起始元件(12个碱基为例)

[0166]

SEQ ID NO.:	强活性元件	SEQ ID NO.:	中等活性元件	SEQ ID NO.:	弱活性元件
1	AAAAAAATATAA	21	GCGTGAGTCAAG	41	CTAACTGGGCGT
2	AAAAAAAATATA	22	ATAATGGGTTAG	42	CTTATGTTTCAA
3	AAAAAAATATAT	23	TGACCCAACTTC	43	TTCCATGGACAT
4	AAAAAAAATAAA	24	GTAGTATTTACG	44	ATCAGACCGAAT
5	AAAATAAATATA	25	AAACATCAGTAT	45	AAGGAGTGACCA
6	AAAAAAAAATAT	26	ATTGGTGATTTG	46	GTCGAGCCGCTG
7	AAAAAAAAATAA	27	AGAATCCTAACT	47	ACATTGTGTGCA
8	AAAAAATATAAT	28	TTTAAGTACGTT	48	ATCACTAGGCGC
9	AAAAATATATAA	29	CCATACCCGGTA	49	GTCCCTCCGCAT
10	AAAAAATATATA	30	GATATTGTAGGC	50	GGGCCTAAGATT
11	AAAAAAATAAAT	31	TAGATGGATACT	51	GTACGCTGTTGC
12	AAAAATATATAT	32	TTTAACTGGGTC	52	AGTCTAACCCGA
13	AAAAAAAAAATA	33	CTGTGGATGAAC	53	AGGGACTTTACG
14	AAAAAAATATAG	34	ATGAGGAGTGGA	54	ACTTGCGCGTGA
15	AAATAAATATAA	35	GAACGGCAAAAG	55	GTGACAAGCGCC
16	AAAATAAATAAA	36	GAAGAAAGGCAC	56	TGACTGTTACTC
17	AAAAAAAATAAT	37	AATCTAACTTTT	57	TGAACTCCAATT
18	AAAAAATATAAA	38	TATCACGAATGG	58	CGTGCGTTTCTT
19	AAAATATATAAT	39	TTACATGGTCTA	59	CGCTACACAAAA
20	AAAATATATATA	40	TCCGCAGAGAAC	60	TCCGCCAAGTGG

[0167] 5.将对抗学习得到的不同强度翻译起始元件插入环形RNA表达载体，利用细胞转染和western blot检测其翻译活性。结果如图3所示。结果显示，表2中的强活性元件能够翻译产生更多的GFP蛋白，中等和弱活性元件也有一定的翻译效率，但是比强活性元件的翻译效率低。

[0168] 结果表明，本发明的对抗学习方法可以有效推测翻译起始元件的活性，可以利用该方法产生不同长度和强度的翻译起始元件，并且本发明的翻译起始元件具有很高的翻译活性，能够显著增强翻译效率。

[0169] 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。

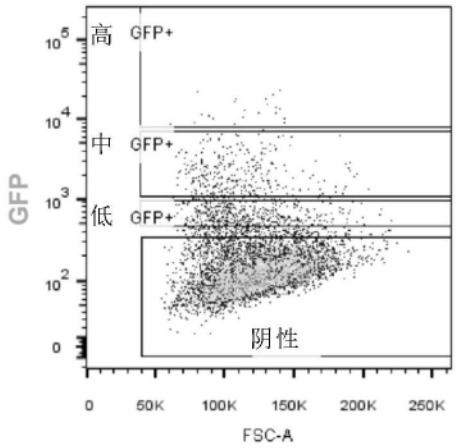
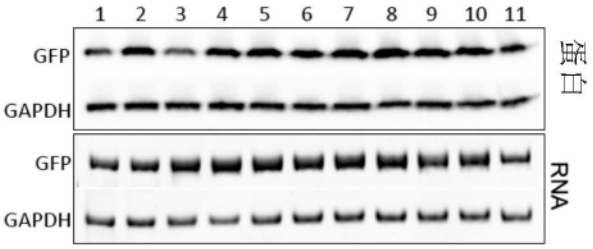
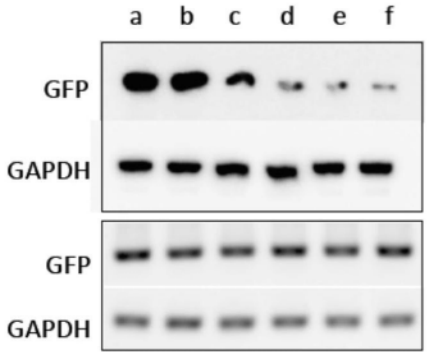


图1



1	ATATAT	7	AATATA
2	ATAAGA	8	AAATAC
3	TTATAT	9	AATTAA
4	AATACT	10	ATACAA
5	TAATAA	11	AAAAAA
6	TTAATA		

图2



a	AAAAAAATATAA
b	AAAAAAAATATA
c	GCGTGAGTCAAG
d	ATAATGGGTTAG
e	CTAACTGGGCGT
f	CTTATGTTTCAA

图3