



(10) 授权公告号 CN 115335526 B

(45) 授权公告日 2025.04.18

(21) 申请号 202180025543.7

(22) 申请日 2021.02.05

(65) 同一申请的已公布的文献号
申请公布号 CN 115335526 A

(43) 申请公布日 2022.11.11

(30) 优先权数据
62/971,356 2020.02.07 US

(85) PCT国际申请进入国家阶段日
2022.09.28

(86) PCT国际申请的申请数据
PCT/US2021/016885 2021.02.05

(87) PCT国际申请的公布数据
W02021/158964 EN 2021.08.12

(73) 专利权人 罗切斯特大学
地址 美国纽约

(72) 发明人 道格拉斯·马修·安德松

(74) 专利代理机构 北京康信知识产权代理有限
责任公司 11240

专利代理师 杜升

(51) Int.Cl.
C12N 15/63 (2006.01)
C12N 15/113 (2006.01)
A61K 48/00 (2006.01)
C12N 15/86 (2006.01)

(56) 对比文件

Anderson DM, Anderson KM, Nelson BR, McAnally JR, Bezprozvannaya S, Shelton JM, Bassel-Duby R, Olson EN. A myocardin-adjacent lncRNA balances SRF-dependent gene transcription in the heart. *Genes Dev.* 2021,835-840.

审查员 杜凌燕

权利要求书3页 说明书71页
序列表57页 附图16页

(54) 发明名称

核酶介导的RNA组装和表达

(57) 摘要

本发明提供了用于使用核酶介导的RNA分子的顺式切割和反式剪接来表达目的蛋白或融合蛋白的组合物、系统和方法。

1. 一种用于产生编码目的蛋白的RNA分子的组合物,包含:
编码第一RNA分子的核酸分子,所述第一RNA分子包含编码目的蛋白的第一部分的编码区和3'核酶;和
编码第二RNA分子的核酸分子,所述第二RNA分子包含编码目的蛋白的第二部分的编码区和5'核酶;
其中所述3'核酶是HDV核酶家族的成员,并且所述5'核酶是HH核酶家族的成员。
2. 如权利要求1所述的组合物,其中所述组合物:
 - a) 进一步包含一种或多种编码一种或多种附加RNA分子的附加核酸分子,每种附加RNA分子均包含编码目的蛋白的结构域的编码区、5'核酶,和3'核酶;或者
 - b) 进一步包含一种或多种编码一种或多种附加RNA分子的附加核酸分子,每种附加RNA分子均包含编码目的蛋白的结构域的编码区、5'核酶,和3'核酶识别序列。
3. 如权利要求2所述的组合物,其中所述组合物进一步包含与所述3'核酶识别序列可以相互作用的核酶,所述核酶诱导所述3'识别序列的去除。
4. 如权利要求2所述的组合物,其中所述3'核酶识别序列包含VS-S并且其中所述核酶是VS-Rz。
5. 编码第一RNA分子的核酸分子和编码第二RNA分子的核酸分子在制备用于产生编码目的蛋白的RNA分子的药剂中的用途,所述第一RNA分子包含编码目的蛋白的第一部分的编码区和3'核酶,所述第二RNA分子包含编码目的蛋白的第二部分的编码区和5'核酶,其中所述产生包括:
向细胞或组织施用所述编码第一RNA分子的核酸分子;和
向所述细胞或组织施用所述编码第二RNA分子的核酸分子;
其中所述3'核酶是HDV核酶家族的成员,并且所述5'核酶是HH核酶家族的成员。
6. 如权利要求5所述的用途,其中所述产生进一步包括:
 - a) 向所述细胞或组织施用一种或多种编码一种或多种附加RNA分子的附加核酸分子,每种附加RNA分子均包含编码目的蛋白的结构域的编码区、5'核酶,和3'核酶;或者
 - b) 向所述细胞或组织施用一种或多种编码一种或多种附加RNA分子的附加核酸分子,每种附加RNA分子均包含编码目的蛋白的结构域的编码区、5'核酶,和3'核酶识别序列;和/或
 - c) 向所述细胞或组织施用连结酶以诱导所述RNA分子的组装。
7. 如权利要求6所述的用途,其中所述产生进一步包括向所述细胞或组织施用与所述3'核酶识别序列相互作用的核酶,所述核酶诱导所述3'识别序列的去除。
8. 如权利要求6所述的用途,其中所述3'核酶识别序列包含VS-S并且其中所述核酶是VS-Rz。
9. 如权利要求6所述的用途,其中所述连结酶是RNA 2',3'-环状磷酸酯和5'-OH (RtcB) 连结酶。
10. 一种产生编码目的蛋白的RNA分子的体外方法,包括:
提供第一RNA分子,所述第一RNA分子包含编码目的蛋白的第一部分的编码区和3'核酶;
提供第二RNA分子,所述第二RNA分子包含编码目的蛋白的第二部分的编码区和5'核

酶;以及

提供连结酶以诱导由所述第一RNA分子的所述编码区和所述第二RNA分子的所述编码区组装所述RNA分子;

其中所述3'核酶是HDV核酶家族的成员,并且所述5'核酶是HH核酶家族的成员。

11. 一种产生编码目的重复结构域蛋白的RNA分子的体外方法,包括:

a) 提供第一RNA分子,所述第一RNA分子包含编码目的蛋白的第一部分的编码区和3'核酶;

b) 提供一种或多种附加RNA分子,所述一种或多种附加RNA分子包含编码目的蛋白的结构域的编码区、5'核酶,和3'核酶识别序列;

c) 提供连结酶以连结所述第一RNA分子的所述编码区和所述一种或多种附加RNA分子的所述编码区;

d) 提供识别所述3'核酶识别序列并催化所述3'核酶识别序列的去除的核酶;

e) 重复步骤b) -d) 一次或多次以产生编码多个重复结构域的RNA分子;

f) 提供最后RNA分子,所述最后RNA分子包含编码目的蛋白的最后部分的编码区和5'核酶;以及

g) 提供连结酶以连结所述一种或多种附加RNA分子的所述编码区和所述最后RNA分子的所述编码区,从而产生编码重复结构域蛋白的完整RNA分子;

其中所述3'核酶是HDV核酶家族的成员,并且所述5'核酶是HH核酶家族的成员。

12. 一种用于产生编码目的蛋白的RNA分子和包含编码以下的核酸的环状RNA分子的组合物:

目的蛋白的第一部分;

包含5'核酶、货物序列和3'核酶的合成内含子;和

目的蛋白的第二部分;

其中所述3'核酶是HDV核酶家族的成员,并且所述5'核酶是HH核酶家族的成员;

其中所述货物序列是选自由以下组成的组的一种或多种:编码目的治疗性蛋白的序列、CRISPR指导RNA序列、小RNA序列和反式切割核酶序列。

13. 如权利要求12所述的组合物,其中所述小RNA序列包括选自由以下组成的组的一种或多种:微RNA (miRNA)、Piwi相互作用RNA (piRNA)、小干扰RNA (siRNA)、小核仁RNA (snoRNA)、小tRNA来源的RNA (tsRNA)、小rDNA来源的RNA (srRNA) 和小核RNA (snRNA)。

14. 如权利要求12或13所述的组合物,其中

a) 所述目的蛋白是选自由以下组成的组的一种或多种:治疗性蛋白、报告蛋白和Cas9蛋白;和/或

b) 进一步包含选自由以下组成的组的一种或多种:RtcB连结酶和编码RtcB连结酶的核酸。

15. 一种非治疗目的的递送编码目的蛋白的RNA分子和环状RNA分子的方法,包括:

向细胞或组织施用编码以下的核酸:目的蛋白的第一部分;包含顺式切割5'核酶、货物序列和顺式切割3'核酶的合成内含子;以及目的蛋白的第二部分;

其中所述3'核酶是HDV核酶家族的成员,并且所述5'核酶是HH核酶家族的成员;

其中所述货物序列是选自由以下组成的组的一种或多种:编码目的治疗性蛋白的序

列、CRISPR指导RNA序列、小RNA序列和反式切割核酶序列。

16. 如权利要求15所述的方法,其中所述小RNA序列包括选自由以下组成的组的一种或多种:微RNA (miRNA)、Piwi相互作用RNA (piRNA)、小干扰RNA (siRNA)、小核仁RNA (snoRNA)、小tRNA来源的RNA (tsRNA)、小rDNA来源的RNA (srRNA)和小核RNA (snRNA)。

17. 如权利要求15或16所述的方法,

a) 其中所述目的蛋白是选自由以下组成的组的一种或多种:治疗性蛋白、报告蛋白和Cas9蛋白;和/或

b) 进一步包括向所述细胞或组织施用选自由以下组成的组的一种或多种:RtcB连结酶和编码RtcB连结酶的核酸。

核酶介导的RNA组装和表达

[0001] 相关申请的引证

[0002] 本申请要求2020年2月7日提交的美国临时专利申请号62/971,356的优先权,该美国临时专利申请的内容通过引用整体并入本文。

背景技术

[0003] 在某些情况下,由于质粒和载体的大小限制,全长蛋白的表达受到限制。例如,在治疗环境中,一些编码全长蛋白的核酸超过了AAV的包装大小,从而限制了它们在基因治疗环境中的适用性。此外,某些生物学和工业相关的蛋白含有大量重复,这可能使表达变得困难。

[0004] 因此,本领域需要改进的用于有效蛋白表达的组合物和方法。本发明满足了这一未被满足的需求。

发明内容

[0005] 在一个实施方案中,本发明包括用于产生编码目的蛋白的RNA分子的系统,该系统包含:编码第一RNA分子的核酸分子,该第一RNA分子包含编码目的蛋白的第一部分的编码区和3'核酶;和编码第二RNA分子的核酸分子,该第二RNA分子包含编码目的蛋白的第二部分的编码区和5'核酶。

[0006] 在一个实施方案中,3'核酶催化自身脱离第一RNA分子,从而产生3'P或2'3'cP末端。在一个实施方案中,5'核酶催化自身脱离第二RNA分子,从而产生5'OH末端。在一个实施方案中,3'P或2'3'cP末端连结到5'OH末端从而形成包含第一RNA分子的编码区和第二RNA分子的编码区的RNA分子。在一个实施方案中,3'核酶是HDV核酶家族的成员。在一个实施方案中,所述5'核酶是HH核酶家族的成员。

[0007] 在一个实施方案中,该系统进一步包含一种或多种编码一种或多种附加RNA分子的附加核酸分子,每种附加RNA分子均包含编码目的蛋白的结构域的编码区、5'核酶,和3'核酶。

[0008] 在一个实施方案中,该系统进一步包含一种或多种编码一种或多种附加RNA分子的附加核酸分子,每种附加RNA分子均包含编码目的蛋白的结构域的编码区、5'核酶,和3'核酶识别序列。在一个实施方案中,该系统进一步包含与3'核酶识别序列相互作用的核酶,该核酶诱导3'识别序列的去除。在一个实施方案中,3'核酶识别序列包含VS-S并且其中该核酶是VS-Rz。

[0009] 在一个实施方案中,本发明涉及用于产生编码目的蛋白的RNA分子的方法,该方法包括:向细胞或组织施用编码第一RNA分子的核酸分子,该第一RNA分子包含编码目的蛋白的第一部分的编码区和3'核酶;以及向细胞或组织施用编码第二RNA分子的核酸分子,所述第二RNA分子包含编码目的蛋白的第二部分的编码区和5'核酶。

[0010] 在一个实施方案中,3'核酶催化自身脱离第一RNA分子,从而产生3'P或2'3'cP末端。在一个实施方案中,5'核酶催化自身脱离第二RNA分子,从而产生5'OH末端。在一个实施

方案中,3'P或2'3'cP末端连结到5'OH末端从而形成包含第一RNA分子的编码区和第二RNA分子的编码区的RNA分子。在一个实施方案中,3'核酶是HDV核酶家族的成员。在一个实施方案中,所述5'核酶是HH核酶家族的成员。

[0011] 在一个实施方案中,该方法进一步包括向细胞或组织施用一种或多种编码一种或多种附加RNA分子的附加核酸分子,每种附加RNA分子均包含编码目的蛋白的结构域的编码区、5'核酶,和3'核酶。

[0012] 在一个实施方案中,该方法进一步包括向细胞或组织施用一种或多种编码一种或多种附加RNA分子的附加核酸分子,每种附加RNA分子均包含编码目的蛋白的结构域的编码区、5'核酶,和3'核酶识别序列。在一个实施方案中,该方法进一步包括向细胞或组织施用与3'核酶识别序列相互作用的核酶,该核酶诱导3'识别序列的去除。在一个实施方案中,3'核酶识别序列包含VS-S并且其中该核酶是VS-Rz。在一个实施方案中,该方法进一步包括向细胞或组织施用连结酶以诱导RNA分子的组装。在一个实施方案中,连结酶是RNA2',3'-环状磷酸酯和5'-OH(RtcB)连结酶。

[0013] 在一个实施方案中,本发明包括产生编码目的蛋白的RNA分子的体外方法,该方法包括:提供包含编码目的蛋白的第一部分的编码区和3'核酶的第一RNA分子;提供包含编码目的蛋白的第二部分的编码区和5'核酶的第二RNA分子;以及提供连结酶以诱导由第一RNA分子的编码区和第二RNA分子的编码区组装RNA分子。

[0014] 在一个实施方案中,本发明包括产生编码目的重复结构域蛋白的RNA分子的体外方法,该方法包括以下步骤:a)提供包含编码目的蛋白的第一部分的编码区和3'核酶的第一RNA分子;b)提供包含编码目的蛋白的结构域的编码区、5'核酶和3'核酶识别序列的一种或多种附加RNA分子;c)提供连结酶以将该第一RNA分子的编码区和该一种或多种附加RNA分子的编码区连结;d)提供识别3'核酶识别序列并催化该3'核酶识别序列的去除的核酶;e)重复步骤b)-d)一次或多次以产生编码多个重复结构域的RNA分子;f)提供包含编码目的蛋白的最后一个部分的编码区和5'核酶的最后一种RNA分子;以及g)提供连结酶以将该一种或多种附加RNA分子的编码区和该最后一种RNA分子的编码区连结,从而产生编码重复结构域蛋白的完整RNA分子。

[0015] 在一个实施方案中,本发明包括治疗受试者中由目的大蛋白中的突变引起的疾病或病症的方法,该方法包括:向所述受试者施用包含编码目的蛋白的第一部分的编码区和3'核酶的第一核酸分子;以及向所述受试者施用包含编码目的蛋白的第二部分的编码区和5'核酶的第二核酸。

[0016] 在一个实施方案中,该疾病或病症是选自由以下组成的组的一种或多种:杜氏肌营养不良症;常染色体隐性多囊肾病;血友病A;Stargardt黄斑变性;肢带型肌营养不良;DFNB9;神经感觉非综合征性隐性耳聋;囊性纤维化;威尔逊病(Wilson Disease);常染色体隐性三好肌营养不良和耳聋9型(Miyoshi Muscular Dystrophy and Deafness;Autosomal Recessive 9);尤塞氏综合症I型和常染色体隐性耳聋2型;常染色体隐性耳聋3型和非综合征性听力损失;尤塞氏综合症I型;常染色体隐性耳聋16(DFNB16);梅尼埃病(Meniere's disease,MD);常染色体显性耳聋12型和常染色体隐性耳聋21型;尤塞氏综合症1F型(USH1F)和DFNB23;常染色体隐性耳聋28型和非综合征性听力损失;常染色体隐性耳聋30型和非综合征性听力损失;常染色体隐性耳脊椎肥大骨骺发育不良和常染色体显性耳椎肥大

骨骺发育不良;常染色体隐性耳聋77型和常染色体隐性非综合征性感觉神经性耳聋Dfnb型;常染色体隐性非综合征性听力障碍DFNB84;常染色体隐性耳聋84B型和罕见性遗传性耳聋;常染色体显性周围神经病变、肌病、声音嘶哑和听力损失和耳聋4A型;先天性血小板减少症;感觉性听力损失;DFNA56;HXB;常染色体显性耳聋56;六臂蛋白(hexabrachion);癫痫性脑病;Timothy综合征和长Qt综合征8型;X-连锁视网膜病症;醛固酮增多症;脊髓小脑性共济失调42型;原发性醛固酮增多症;癫痫发作和神经系统异常和窦房结功能障碍和耳聋;神经发育病症;低钾性周期性麻痹;癫痫;发育性和癫痫性脑病;布罗迪肌病(Brody myopathy);Darier氏病/心脏病;血管性血友病(von Willebrand disease)和齐薇格综合征(Zellweger syndrome)。

[0017] 在一个实施方案中,本发明包括用于产生编码目的蛋白的RNA分子和环状RNA分子的系统,该环状RNA分子包含编码以下的核酸:目的蛋白的第一部分;包含5'核酶、货物序列和3'核酶的合成内含子;和目的蛋白的第二部分。

[0018] 在一个实施方案中,目的蛋白是选自由以下组成的组的一种或多种:治疗性蛋白、报告蛋白和Cas9蛋白。

[0019] 在一个实施方案中,货物序列是选自由以下组成的组的一种或多种:编码目的治疗性蛋白的序列、CRISPR指导RNA序列、小RNA序列和反式切割核酶序列。在一个实施方案中,所述小RNA序列包括选自由以下组成的组的一种或多种:微RNA(miRNA)、Piwi相互作用RNA(piRNA)、小干扰RNA(siRNA)、小核仁RNA(snoRNA)、小tRNA来源的RNA(tsRNA)、小rDNA来源的RNA(srRNA)和小核RNA(snRNA)。

[0020] 在一个实施方案中,合成内含子的3'核酶是HH核酶家族的成员。在一个实施方案中,合成内含子的5'核酶是选自由以下组成的组的一种或多种:HDV核酶家族成员、HDV核酶家族成员和VS-S核酶识别序列。在一个实施方案中,该系统进一步包含选自由以下组成的组的一种或多种:RtcB连结酶和编码RtcB连结酶的核酸。

[0021] 在一个实施方案中,本发明包括递送编码目的蛋白的RNA分子和环状RNA分子的方法,该方法包括向细胞或组织施用编码以下的核酸:目的蛋白的第一部分;包含顺式切割5'核酶、货物序列和顺式切割3'核酶的合成内含子;以及目的蛋白的第二部分。

[0022] 在一个实施方案中,目的蛋白是选自由以下组成的组的一种或多种:治疗性蛋白、报告蛋白和Cas9蛋白。

[0023] 在一个实施方案中,货物序列是选自由以下组成的组的一种或多种:编码目的治疗性蛋白的序列、CRISPR指导RNA序列、小RNA序列和反式切割核酶序列。在一个实施方案中,所述小RNA序列包括选自由以下组成的组的一种或多种:微RNA(miRNA)、Piwi相互作用RNA(piRNA)、小干扰RNA(siRNA)、小核仁RNA(snoRNA)、小tRNA来源的RNA(tsRNA)、小rDNA来源的RNA(srRNA)和小核RNA(snRNA)。

[0024] 在一个实施方案中,该方法进一步包括向细胞或组织施用选自由以下组成的组的一种或多种:RtcB连结酶和编码RtcB连结酶的核酸。

附图说明

[0025] 当结合附图阅读时,能更好地理解本发明实施方案的以下详细描述。应该理解,本发明不限于附图中所示的实施方案的精确布置和手段。

[0026] 图1包括图1A至图1E,描绘了哺乳动物细胞中的核酶介导的反式剪接和表达。图1A示出了描绘编码具有3'HDV核酶的GFP的N-端(Nt)半分子以及具有5'锤头状(HH)核酶的GFP的C-端(Ct)半分子的载体的图。图1B描绘了证明Nt-GFP-HDV和HH-Ct-GFP在COS7和HEK293T细胞中的共表达导致可检测到的GFP荧光,但在单独转染时未能如此的示例性结果。图1C-1D描绘了使用对每个独立RNA(G1和G2)特异的引物进行的RT-PCR扩增(图1C)和桑格序列分析(sanger sequence analysis)(图1D)的示例性结果,该结果示出了核酶的去除了以及GFP编码序列的无疤痕反式剪接和恢复。图1E描绘了使用对GFP特异的抗体的示例性蛋白质印迹结果,该结果示出了针对GFP预测的全长蛋白大小。

[0027] 图2包括图2A至图2E,描绘了用于量化核酶序列对哺乳动物细胞中反式剪接的影响的基于荧光素酶的报告基因的开发。图2A示出了描绘编码具有3'HDV核酶的荧光素酶的N-端(Nt)半分子以及具有5'锤头状(HH)核酶的荧光素酶的C-端(Ct)半分子的载体的图。图2B-2C描绘了使用对每个独立Luc RNA(L1和L2)特异的引物进行的RT-PCR扩增(图2B)和桑格序列分析(图2C)的示例性结果,该示例性结果示出了核酶的去除了和荧光素酶开放阅读框的无疤痕反式剪接。图2D-2E展示了不同HDV(图2D)和HH(图2E)核酶序列对哺乳动物细胞中的反式剪接的影响。此外,核酶催化核苷酸的突变导致荧光素酶活性丧失(图2D,最后一列和图2E,最后一列)。

[0028] 图3包括图3A至图3D,展示了Nt和Ct载体对蛋白质表达的调节。图3A示出了描绘阻止Nt载体编码的蛋白质表达的C-端蛋白降解序列的布置的图。图3B描绘了证明不同蛋白降解序列在阻止由编码全长GFP的Nt载体表达GFP-HDV方面的效率的示例性结果。图3C示出了描绘用于阻止Ct载体中的蛋白序列翻译的N-端翻译控制序列的布置的图。图3D描绘了证明不同的GFP序列修饰或翻译控制序列在阻止哺乳动物细胞中的GFP荧光方面的效率的示例性结果。

[0029] 图4包括图4A至图4D,展示了哺乳动物细胞中单个和多重核酶介导的反式剪接。图4A示出了描绘编码4xMTS和全长GFP(无起始ATG密码子)的载体的图,其中核酶介导靶向线粒体的GFP蛋白的反式剪接和表达。图4B描绘了证明这些载体的共表达导致定位于线粒体的绿色荧光与mitotracker CMXRos的红色荧光重叠的示例性结果。图4C示出了描绘用于阅读框1中靶向线粒体的GFP蛋白(4xMTS-GFP)和阅读框2中靶向肉豆蔻酰化膜的红色荧光蛋白(F2-Myr-RFP)的多重反式剪接和表达的载体的图。图4D描绘了证明哺乳动物Cos7细胞中所有四种载体的共表达导致线粒体中的特异性绿色荧光和膜中的红色荧光的示例性结果。

[0030] 图5包括图5A和图5B,展示了使用优化的核酶序列和顺式剪接剪接受体和剪接供体序列增强的核酶介导的反式剪接。图5A示出了描绘嵌合剪接供体(SD)和剪接受体(SA)序列在通用Nt-GFP-3'Rz和5'Rz-Ct-GFP反式剪接GFP报告基因(其中Rz表示顺式切割核酶)中的布置的图。图5B描绘了在转染后18小时(前三列)或转染后36小时(最后一列)单载体转染(前两列)或共转染(后两列)后Cos7细胞中的GFP荧光的示例性结果。第一行描绘了未优化的HH和HDV核酶的使用,第二行描绘了优化的Twister核酶和RzB核酶的使用,最后一行描绘了Twister核酶和RzB核酶与SD和SA序列的组合。

[0031] 图6包括图6A至图6D,展示了核酶介导的大蛋白编码基因的反式剪接。图6A示出了描绘编码用于使用AAV载体进行递送的分裂的 μ 抗肌萎缩蛋白(μ Dystrophin)-GFP融合蛋白的载体的图。图6B-6C描绘了对用Nt-Dys和Ct-Dys载体转染的细胞进行的RT-PCR(图6B)和

桑格测序(图6C)分析的示例性结果,该结果示出特异性反式剪接。图6D描绘了使用共聚焦显微镜成像的用Nt和Ct抗肌萎缩蛋白载体转染的细胞的GFP荧光的示例性结果,其示出抗肌萎缩蛋白的预测膜定位。

[0032] 图7包括图7A至图7C,展示了用慢病毒递送用于在靶细胞中进行反式剪接的含有核酶的RNA。图7A示出了描绘慢病毒基因转移载体中Nt和Ct分裂的GFP表达盒的负向取向的图。图7B描绘了展示只有与编码Nt-GFP和Ct-GFP基因的慢病毒共转导的细胞显示出GFP荧光的示例性结果。图7C示出了描绘慢病毒基因转移载体中Nt和Ct分裂的Dys表达盒的负向取向的图。

[0033] 图8包括图8A和图8B,展示了核酶介导的毒性DTA基因的反式剪接和表达。图8A示出了描绘编码分裂的Nt和Ct DTA基因的载体的图。图8B描绘了展示用Nt-DTA和Ct-DTA共同转染的细胞导致共同转染的GFP报告基因的表达降低(这与DTA在哺乳动物细胞中的翻译抑制功能一致)的示例性结果。

[0034] 图9描绘了展示外源RNA调节酶的共表达可以增强或抑制哺乳动物细胞中核酶介导的反式剪接的示例性结果。

[0035] 图10包括图10A至图10D,展示了RtcB足以在体外催化核酶介导的反式剪接。图10A示出了描绘分裂的荧光素酶反式剪接报告基因的图,该报告基因含有上游T7 RNA启动子以允许体外RNA转录。图10B示出了展示体外反式剪接的荧光素酶RNA依赖于使用制造商推荐的反应条件添加RtcB蛋白(NEB)的示例性RT-PCR结果。图10C示出了描绘蛛丝蛋白(Spidroin)的保守N-端(N1L)和C-端(N3R)结构域的反式剪接载体的图。图10D描绘了展示来自大肠杆菌的RtcB连结酶足以催化核酶切割的N1L和N3R编码RNA的反式连结的示例性桑格测序结果。

[0036] 图11描绘了使用RtcB、VS-S和VS-Rz在体外定向连结经核酶催化的RNA。

[0037] 图12包括图12A至12D,描绘了使用反式切割核酶进行RNA的反式剪接。图12A描绘了以顺式切割的核酶的二级结构。图12B描绘了能够以反式切割的工程化核酶。图12C和图12D描绘了展示反式切割核酶在删除致病突变(诸如移码或过早终止密码子)以恢复蛋白表达和功能方面的潜在应用的图。

[0038] 图13包括图13A和图13B,描绘了可用于RNA的无疤痕反式剪接的代表性核酶的二级结构。图13A描绘了可用于无疤痕5'切割的代表性核酶。图13B描绘了可用于无疤痕3'切割的代表性核酶。N=任何核苷酸。红色剪刀划定切割位点。红色核苷酸指示催化突变。橙色核苷酸代表待被反式剪接的RNA序列。深蓝色核苷酸指示形成茎所需的核酶序列。浅蓝色指示茎1中的三级稳定基序(TSM),该茎1与茎2环相互作用。HH-锤头状,HDV-丁型肝炎病毒,Rz-核酶。

[0039] 图14包括图14A至图14C,描绘了无疤痕切割和可利用反式活化核酶诱导的RNA反式剪接和表达。图14A描绘了显示VS核酶可以分裂成两个组分(一个小VS-S茎环,其缺乏自催化活性;和一个较大VS-Rz,其在被以反式递送时诱导VS-S切割)的图。该VS-S/Vs-Rz核酶对可用于产生可诱导的无疤痕反式剪接。图14B示出了描绘用于利用VS-S/Vs-Rz反式活化核酶对产生可诱导的RNA反式剪接系统的方法的图。只有在递送或表达VS-Rz后,Nt-GFP-VS-S RNA才会产生合适的RNA端,该端可以参与使用共表达的Ct-GFP RNA的反式剪接。图14C示出了描绘用于产生具有N-端序列、可变或非可变重复区和C-端序列的RNA的方法的

图。“重复”RNA含有5'自催化核酶和3'反式活化核酶(诸如VS-S),其允许根据反式活化VS-Rz和连结酶(诸如RtcB)的选择性添加来控制重复添加。

[0040] 图15包括图15A至图15E,描绘了核酶介导的反式剪接以及稳定的内含子RNA序列的产生。图15A示出了描绘使用顺式切割核酶来介导两种独立RNA的反式剪接的图。图15B示出了描绘使用内部顺式切割核酶来产生合成内含子的图。图15C描绘了展示合成内含子的有效顺式切割和独立RNA的反式剪接产生功能性蛋白(GFP)的示例性结果。图15D和图15E示出了描绘使用内部顺式切割核酶来产生经反式剪接和翻译的报告基因和内含子序列“货物”(其可以是任何有用的RNA序列或基因表达盒)的图。

[0041] 图16包括图16A至图16C,描绘了用于体内核酶介导的反式剪接的优化核酶序列的示例性结果。图16A描绘了使用萤光素酶反式剪接报告基因的相对核酶活性的比较。含有三级稳定基序并且在低镁浓度下具有活性的RzB锤头状核酶变体在哺乳动物细胞中表现出最大的萤光素酶活性。图16B描绘了HDV核酶(HDV68和基因组HDV)与Twister核酶(Twst)的比较。位于Nt-Luc的3'末端上的Twister核酶提供了最大的萤光素酶活性,该萤光素酶活性被催化失活突变(Twst mut)消除。图16C描绘了Twister核酶序列修饰的比较。P1茎的缩短降低了报告基因的活性。第一残基的修饰揭示Twister核酶可以耐受1位处的A核苷酸(U1A)。

具体实施方式

[0042] 定义

[0043] 除非另有定义,否则本文中使用的所有技术和科学术语具有与本发明所属领域的普通技术人员通常理解的含义相同的含义。

[0044] 一般而言,本文使用的命名法以及细胞培养、分子遗传学、有机化学、核酸化学和杂交中的实验室程序是本领域熟知和常用的。

[0045] 标准技术被用于核酸和肽合成。该技术和程序一般根据本领域的常规方法和贯穿本文档提供的各种通用参考文献(例如,Sambrook和Russell,2012,Molecular Cloning,A Laboratory Approach,Cold Spring Harbor Press,Cold Spring Harbor,NY,以及Ausubel等人,2012,Current Protocols in Molecular Biology,John Wiley&Sons,NY)来执行。

[0046] 本文使用的命名法和下文描述的分析化学和有机合成中使用的实验室程序是本领域熟知和常用的那些。标准技术或其修改形式被用于化学合成和化学分析。

[0047] 在本发明的上下文中(尤其是在权利要求书的上下文中)使用的术语“一个/种(a)”、“一个/种(an)”、“该/所述”以及类似术语应被解读为涵盖单数和复数两者,除非在本文另外地指示或明显地与上下文矛盾。

[0048] 当提及可测量的值诸如量、持续时间等时,如本文所用的“约”意在涵盖指定值相差 $\pm 20\%$,或 $\pm 10\%$,或 $\pm 5\%$,或 $\pm 1\%$,或 $\pm 0.1\%$ 的变化,因为此类变化适用于执行公开的方法。

[0049] “反义”特别是指编码蛋白质的双链DNA分子的非编码链的核酸序列,或与该非编码链基本上同源的序列。如本文所定义,反义序列与编码蛋白质的双链DNA分子的序列互补。反义序列不必仅与DNA分子编码链的编码部分互补。反义序列可以与编码蛋白质的DNA分子的编码链上指定的调控序列互补,该调控序列控制编码序列的表达。

[0050] 当提及将分子(例如核酸分子)固定至固体支持物时,如本文所用的术语“附接”旨在涵盖直接或间接、共价或非共价附接,除非另有明确或上下文指示。

[0051] 如本文可互换使用的,“微球体”、“珠粒”或其语法等价物描述了能够充当用于附接生物分子(例如,核酸分子)的固体支持物的小的离散颗粒。

[0052] “疾病”是动物的一种如下健康状态,其中动物不能维持体内平衡,并且其中,如果疾病没有改善,则动物的健康继续恶化。

[0053] 相对来说,动物中的“病症”是一种如下健康状态,其中动物能够维持体内平衡,但其中动物的健康状态不如没有病症时的健康状态。如果不加以治疗,病症不一定会导致动物健康状态进一步下降。

[0054] 如果疾病或病症的体征或症状的严重性、患者经历此种体征或症状的频率,或两者都降低,则疾病或病症“缓和”。

[0055] “编码”是指多核苷酸中的特定核苷酸序列(诸如基因、cDNA或mRNA)的固有特性,该特定核苷酸序列用于在生物过程中充当合成具有确定的核苷酸序列(即rRNA、tRNA和mRNA)或确定的氨基酸序列以及由此产生的生物学特性的其他聚合物和大分子的模板。因此,如果对应于基因的mRNA的转录和翻译在细胞或其他生物系统中产生蛋白质,则该基因编码蛋白。编码链(其核苷酸序列与该mRNA序列相同,并且通常提供在序列表中)和非编码链(用作基因或cDNA转录的模板)都可以称为编码该基因或cDNA的蛋白质或其他产物。

[0056] 术语“患者”、“受试者”、“个体”等在本文中可互换使用,并且是指无论在体外还是在体内都适用于本文所述的方法的任何动物或细胞。在一个实施方案中,受试者包括脊椎动物和无脊椎动物。无脊椎动物包括但不限于黑腹果蝇和秀丽隐杆线虫。脊椎动物包括但不限于灵长类动物、啮齿动物、家养动物或狩猎动物。灵长类动物包括但不限于黑猩猩、食蟹猴、蜘蛛猴和猕猴(例如,恒河猴)。啮齿动物包括但不限于小鼠、大鼠、土拨鼠、雪貂、兔子和仓鼠。家养动物和狩猎动物包括但不限于牛、马、猪、鹿、野牛、水牛、猫科物种(例如家猫)、犬科物种(例如狗、狐狸、狼)、鸟类物种(例如,鸡、鹌鹑、鸵鸟)和鱼(例如斑马鱼、鳟鱼、鲑鱼和鲈鱼)。在一些实施方案中,受试者是哺乳动物,例如灵长类动物,例如人。在某些非限制性实施方案中,患者、受试者或个体是人。

[0057] 如本文所用的关于抗体的术语“特异性地结合”意指识别特定抗原但基本上不识别或结合样品中的其他分子的抗体。例如,与来自一个物种的抗原特异性结合的抗体也可以与来自一个或多个物种的该抗原结合。但是,这种跨物种反应性本身不会改变抗体的特异性分类。在另一个例子中,特异性结合抗原的抗体也可以结合抗原的不同等位基因形式。然而,这种交叉反应性本身不会改变抗体的特异性分类。

[0058] 在一些情况下,术语“特异性结合(specific binding)”或“特异性地结合(specifically binding)”可用于指抗体、蛋白或肽与第二化学物质的相互作用,以表示该相互作用取决于该化学物质上存在的特定结构(例如,抗原决定簇或表位);例如,抗体识别并结合至特定的蛋白结构,而不是一般蛋白。如果抗体对表位“A”具有特异性,则在含有经标记的“A”和抗体的反应中,含有表位A(或游离的未经标记的A)的分子的存在将减少与该抗体结合的经标记的A的数量。

[0059] 基因的“编码区”由基因的编码链的核苷酸残基和基因的非编码链的核苷酸组成,它们分别与由该基因转录产生的mRNA分子的编码区同源或互补。

[0060] mRNA分子的“编码区”也由mRNA分子的核苷酸残基组成,该核苷酸残基在mRNA分子的翻译过程中与转移RNA分子的反密码子区匹配或编码终止密码子。因此,编码区可以包括这样的核苷酸残基,该核苷酸残基包含不存在于由mRNA分子编码的成熟蛋白中的氨基酸残基(例如,蛋白输出信号序列中的氨基酸残基)的密码子。

[0061] 如本文用于指代核酸的“互补”是指两条核酸链的区域之间或同一核酸链的两个区域之间的序列互补性的广义概念。已知如果残基是胸腺嘧啶或尿嘧啶,则第一核酸区域的腺嘌呤残基能够与和第一区域反平行的第二核酸区域的残基形成特定的氢键(“碱基配对”)。类似地,已知第一条核酸链的胞嘧啶残基能够与第二条核酸链的残基配对,如果该残基是鸟嘌呤,则该第二条核酸链与第一条链反平行。核酸的第一区域与相同或不同核酸的第二区域在下面的情况下互补:如果当这两个区域以反平行方式被布置时,第一区域的至少一个核苷酸残基能够与第二区域的残基配对。在一个实施方案中,第一区域包括第一部分并且第二区域包括第二部分,由此,当第一部分和第二部分以反平行方式被布置时,第一部分的核苷酸残基的至少约50%、至少约75%、至少约90%,或至少约95%能够与第二部分中的核苷酸残基进行碱基配对。在一个实施方案中,第一部分的所有核苷酸残基都能够与第二部分中的核苷酸残基进行碱基配对。

[0062] 如本文所用的术语“DNA”被定义为脱氧核糖核酸。

[0063] 如本文所用的术语“表达”被定义为特定核苷酸序列的由其启动子驱动转录和/或翻译。

[0064] 如本文所用的术语“表达载体”是指含有对能够被转录的基因产物的至少一部分编码的核酸序列的载体。在某些情况下, RNA分子然后被翻译成蛋白质、多肽或肽。在其他情况下,这些序列不被翻译,例如,在反义分子、siRNA、核酶等的生产中。表达载体可以含有多种控制序列,该控制序列是指特定宿主生物中可操作地连接的编码序列的转录和可能翻译所必需的核酸序列。除了管控转录和翻译的控制序列外,载体和表达载体还可含有具有其他功能的核酸序列。

[0065] 如本文所用,术语“野生型”是本领域技术人员理解的术语,并且意指与突变体或变体形式有区别的如在自然界中存在的生物体、菌株、基因或特征的典型形式。

[0066] 术语“同源性”是指一定程度的互补性。可能存在部分同源性或完全同源性(即同一性)。同源性常常使用序列分析软件(例如,Genetics Computer Group.(University of Wisconsin Biotechnology Center.1710University Avenue.Madison,Wis.53705)的序列分析软件包)来测量。此类软件通过将同源性程度分配给各种取代、缺失、插入和其他修饰来匹配相似的序列。保守性取代通常包括以下各组内的取代:甘氨酸、丙氨酸;缬氨酸、异亮氨酸、亮氨酸;天冬氨酸、谷氨酸、天冬酰胺、谷氨酰胺;丝氨酸、苏氨酸;赖氨酸、精氨酸;以及苯丙氨酸、酪氨酸。

[0067] “分离的”是指改变或脱离自然状态。例如,在活动物中在正常环境中自然存在的核酸或肽不是“分离的”,但部分或完全从其自然环境的共存物质中分离的同一核酸或肽是“分离的”。分离的核酸或蛋白质可以基本上纯化的形式存在,或者可以存在于非天然环境例如宿主细胞中。

[0068] 术语“分离的”在与核酸相关被使用时,如“分离的寡核苷酸”或“分离的多核苷酸”是指这样的核酸序列,该核酸序列是从至少一种通常与其来源相关的污染物中鉴定和分离

出来的。因此,分离的核酸以不同于其在自然界中被发现的形式或环境存在。相对来说,非分离的核酸(例如DNA和RNA)以它们在自然界中存在的状态被发现。例如,给定DNA序列(例如,基因)被发现在宿主细胞染色体上与相邻基因相邻;RNA序列(例如,编码特定蛋白的特定mRNA序列)被发现在细胞中以与编码众多蛋白的许多其他mRNA的混合物的形式存在。然而,分离的核酸包括,举例来说,通常表达该核酸的细胞中的这样的核酸,该核酸位于与天然细胞的染色体位置不同的染色体位置或者以其他方式侧接有与在自然界中发现的核酸序列不同的核酸序列。分离的核酸或寡核苷酸可以单链或双链形式存在。当使用分离的核酸或寡核苷酸来表达蛋白时,寡核苷酸至少包含有义链或编码链(即寡核苷酸可以是单链的),但可以同时包含有义链和反义链(即寡核苷酸可以是双链的)。

[0069] 术语“分离的”在与多肽相关被使用时,如在“分离的蛋白”或“分离的多肽”中,是指这样的多肽,该多肽是从至少一种通常与其来源相关的污染物中鉴定和分离出来的。因此,分离的多肽以不同于其在自然界中被发现的形式或环境存在。相反,非分离的多肽(例如蛋白和酶)以它们在自然界中存在的状态被发现。

[0070] 所谓“核酸”是指任何核酸,无论是由脱氧核糖核苷或核糖核苷组成,也无论是由磷酸二酯连键或经修饰的连键(诸如磷酸三酯、氨基磷酸酯、硅氧烷、碳酸酯、羧甲基酯、亚氨基乙酸酯(acetamidate)、氨基甲酸酯、硫醚、桥联氨基磷酸酯、桥联亚甲基磷酸酯、硫代磷酸酯、甲基磷酸酯、二硫代磷酸酯、桥联硫代磷酸酯或矾连键以及此类连键的组合)组成。术语核酸还具体包括由除五种生物学上存在的碱基(腺嘌呤、鸟嘌呤、胸腺嘧啶、胞嘧啶和尿嘧啶)以外的碱基组成的核酸。术语“核酸”通常是指大的多核苷酸。

[0071] 本文使用常规符号来描述多核苷酸序列:单链多核苷酸序列的左手端是5'端;双链多核苷酸序列的左手方向称为5'方向。

[0072] 核苷酸从5'到3'被添加到新生RNA转录本中的方向称为转录方向。与mRNA具有相同序列的DNA链称为“编码链”;DNA链上的位于DNA上参考点的5'处的序列称为“上游序列”;DNA链上的位于DNA上参考点的3'处的序列称为“下游序列”。

[0073] “表达盒”意指包含编码序列的核酸分子,该编码序列可操作地连接至该编码序列的转录和任选的翻译所需的启动子/调控序列。

[0074] 如本文所用的术语“可操作地连接”是指核酸序列的这样的连接,该连接使得产生能够指导给定基因的转录和/或所需蛋白分子的合成的核酸分子。该术语还指编码氨基酸的序列的这样的连接,该连接使得产生功能性(例如,具有酶促活性的、能够与结合配偶体结合、能够抑制等)的蛋白或多肽。

[0075] 如本文所用,术语“启动子/调控序列”意指表达可操作地连接至该启动子/调控序列的基因产物所需的核酸序列。在一些情况下,该序列可以是核心启动子序列,而在其他情况下,该序列还可以包括增强子序列和基因产物表达所需的其他调控元件。该启动子/调控序列可以例如是以可诱导方式表达基因产物的序列。

[0076] 如本文所用,用于杂交的“严格条件”是指与靶序列具有互补性的核酸主要与靶序列杂交并且基本上不与非靶序列杂交的条件。严格条件一般是序列依赖性的,并且取决于许多因素。一般而言,序列越长,序列与其靶序列特异性地杂交的温度越高。严格条件的非限制性实例详细描述于Tijssen(1993),Laboratory Techniques In Biochemistry And Molecular Biology-Hybridization With Nucleic Acid Probes第1部分,第二章

“Overview of principles of hybridization and the strategy of nucleic acid probe assay”, Elsevier, N.Y.

[0077] “杂交”是指一种或多种多核苷酸反应形成经由核苷酸残基的碱基之间的氢键键合而稳定的复合物的反应。该氢键键合可以通过Watson Crick碱基配对、Hoogsteen结合,或以任何其他序列特异性方式发生。该复合物可包括形成双链体结构的两条链、形成多链复合物的三条或更多条链、单个自杂交链,或这些的任何组合。杂交反应可以构成更广泛过程中的步骤,例如PCR的启动,或酶对多核苷酸的切割。能够与给定序列杂交的序列称为该给定序列的“互补序列”。

[0078] “诱导型”启动子是这样的核苷酸序列,该核苷酸序列在与编码或指定基因产物的多核苷酸可操作地连接时,导致基因产物基本上仅在对应于启动子的诱导物存在时才产生。

[0079] “组成型”启动子是这样的核苷酸序列,该核苷酸序列在与编码或指定基因产物的多核苷酸可操作地连接时,导致基因产物在细胞中在该细胞的大部分或所有生理条件下产生。

[0080] 如本文所用的术语“多核苷酸”被定义为核苷酸链。此外,核酸是核苷酸的聚合物。因此,如本文所用的核酸和多核苷酸是可互换的。本领域技术人员具有核酸是可以水解成单体“核苷酸”的多核苷酸的常识。单体核苷酸可以水解成核苷。如本文所用,多核苷酸包括但不限于通过本领域可用的任何方式获得的所有核酸序列,所述方式包括但不限于重组方式,即使用普通的克隆技术和PCR等从重组文库或细胞基因组中克隆核酸序列,以及合成方式。

[0081] 在本发明的上下文中,使用常用核酸碱基的以下缩写。“A”是指腺苷,“C”是指胞嘧啶,“G”是指鸟苷,“T”是指胸苷,“U”是指尿苷。

[0082] 如本文所用,术语“肽”、“多肽”和“蛋白质”可互换使用,是指由通过肽键共价连接的氨基酸残基组成的化合物。蛋白质或肽必须含有至少两种氨基酸并且对于可构成蛋白质或肽序列的氨基酸的最大数目没有设限。多肽包括包含通过肽键彼此联接的两种或更多种氨基酸的任何肽或蛋白质。如本文所用,该术语既指短链(其在本领域中也通常称为例如肽、寡肽和寡聚体),也指更长链(其在本领域中一般称为蛋白质),其有很多类型。“多肽”包括例如生物活性片段、基本上同源的多肽、寡肽、同二聚体、异二聚体、多肽的变体、经修饰的多肽、衍生物、类似物、融合蛋白等。多肽包括天然肽、重组肽、合成肽或它们的组合。

[0083] 如本文所用的术语“RNA”被定义为核糖核酸。

[0084] 如本文所用的术语“核酶”是指能够充当酶的RNA分子。例如,一些核酶能够切割RNA分子。RNA切割核酶通常至少由催化结构域和由催化结构域识别的识别序列组成。催化结构域可以是与识别序列相同的RNA分子的一部分,因此介导顺式切割。替代地,催化结构域可以是与包含识别序列的RNA分子分开的RNA分子,并因此介导反式切割。

[0085] “重组多核苷酸”是指具有非天然地联接在一起的序列的多核苷酸。扩增或组装的重组多核苷酸可包含在合适的载体中,并且该载体可用于转化合适的宿主细胞。

[0086] 重组多核苷酸也可以具有非编码功能(例如,启动子、复制起点、核糖体结合位点等)。

[0087] 如本文所用的术语“重组多肽”被定义为通过使用重组DNA方法产生的多肽。

[0088] 如本文所用,术语“固体表面”、“固体支持物”以及它们的其他语法等同物是指适合于或可以被修饰成适合于生物分子(例如,核酸分子)的附接的任何材料。

[0089] 如本文所用,术语“标签”是指生物分子(例如核酸分子)的任何化学修饰,该化学修饰提供附加功能性(例如,与固体支持物的附接、荧光可视化等)。

[0090] 如本文所用的术语“变体”是分别在序列上与参考核酸序列或肽序列不同但保留了参考分子的基本生物学特性的核酸序列或肽序列。核酸变体序列的变化可能不会改变参考核酸编码的肽的氨基酸序列,或可能会导致氨基酸取代、添加、缺失、融合和截短。肽变体序列的变化通常是有限的或保守的,因此参考肽和变体的序列总体上非常相似,并且在许多区域中是相同的。变体和参考肽的氨基酸序列可能因呈任何组合的一个或多个取代、添加、缺失而不同。核酸或肽的变体可以是天然存在的(诸如等位基因变体),或者可以是尚不知道是否天然存在的变体。核酸和肽的非天然存在的变体可以通过诱变技术或通过直接合成来制备。

[0091] “载体”是包含分离的核酸并且可用于将分离的核酸递送至细胞内部的物质组合物。许多载体是本领域已知的,该载体包括但不限于线性多核苷酸、与离子或两亲化合物缔合的多核苷酸、质粒和病毒。因此,术语“载体”包括自主复制的质粒或病毒。该术语还应解读为包括促进核酸转移到细胞中的非质粒和非病毒化合物,诸如,例如聚赖氨酸化合物、脂质体等。病毒载体的实例包括但不限于腺病毒载体、腺相关病毒载体、逆转录病毒载体等。

[0092] 范围:在整个本公开中,本发明的各个方面可以以范围格式呈现。应当理解,以范围格式进行描述只是为了方便和简洁,并且不应被解释为对本发明的范围的僵硬限制。因此,范围的描述应该被认为已明确公开了所有可能的子范围以及该范围内的各个数值。例如,诸如从1至6的范围的描述应该被认为已明确公开了诸如1至3、1至4、1至5、2至4、2至6、3至6等的子范围,以及该范围内的各个数字,例如1、2、2.7、3、4、5、5.3和6。这在任何宽度范围的情况下均适用。

[0093] 详述

[0094] 本发明提供了用于有效且可靠地连结两种或更多种单独的RNA分子以产生编码蛋白和融合蛋白的更大的单种RNA分子的组合物和方法。本发明利用核酶介导的多种RNA分子的反式剪接来组装编码目的蛋白或融合蛋白的单种RNA分子。本发明可用于有效地生产融合蛋白、嵌合蛋白等。进一步地,本发明可用于产生大的全长蛋白,其编码序列可能太大而不能包装到单个载体中。进一步地,本发明的技术还允许快速且容易地组合两种不同的序列,这对于产生新的蛋白组合或文库序列可具有倍增效应(multiplier effect)。这可能例如对于产生合成抗体(如纳米抗体)或对于酶的功能选择特别有用。

[0095] 本发明还提供了用于有效地递送一种或多种具有侧接有核酶的合成内含子的RNA分子的组合物和方法。侧接有核酶的合成内含子可被置于编码目的蛋白的N-端部分的第一RNA部分与编码目的蛋白的C-端部分的第二RNA部分之间。侧接有核酶的合成内含子可以包含货物序列,例如,编码治疗性蛋白或包含功能性RNA的序列。两种核酶的使用允许顺式剪接以产生以下三个RNA片段:1) 编码目的蛋白的N-端部分的第一RNA部分,2) 侧接有核酶的合成内含子,和3) 编码目的蛋白的C-端部分的第二RNA部分。所述顺式剪接产生用于连结的相容末端。经顺式剪接的合成内含子的相容末端的连结产生环状RNA分子,该环状RNA分子比线性RNA分子更耐降解。编码目的蛋白的N-端部分的第一RNA部分和编码目的蛋白的C-端

部分的第二RNA部分的相容末端的连结产生编码全长目的蛋白的RNA分子。全长目的蛋白可以是例如治疗性蛋白、CRISPR-Cas蛋白,或报告蛋白,该报告蛋白用于提供包含侧接有核酶的合成内含子的环状RNA分子中的货物序列的递送和表达的代表性指标(proxy indicator)。

[0096] 在一方面,本发明提供一种或多种编码两种或更多种RNA分子的核酸分子。在某些实施方案中,该RNA分子中的一种或多种包含核酶。在一个实施方案中,该RNA分子中的一种或多种包含编码区和核酶。在某些实施方案中,核酶自切割脱离RNA分子,从而留下编码区。可以在本发明的上下文中使用的示例性核酶包括但不限于锤头状(HH)核酶、丁型肝炎病毒(HDV)核酶、Varkud卫星(VS)核酶、Sister核酶、Twister-sister核酶、发夹状核酶、Hatchet核酶和Pistol核酶家族的成员。

[0097] 例如,在一个实施方案中,所述组合物包含编码第一RNA分子的核酸分子,其中第一RNA分子包含编码区和3'核酶,其中该3'核酶能够催化自身脱离该RNA分子,从而留下具有3'P或2'3'环状磷酸酯(cP)末端的编码区。在一个实施方案中,3'核酶包括HDV核酶。进一步地,在一个实施方案中,所述组合物包含编码第二RNA分子的核酸分子,其中该第二RNA分子包含编码区和5'核酶,其中该5'核酶能够催化自身脱离该RNA分子,从而留下具有5'OH末端的编码区。在一个实施方案中,5'核酶包括HH核酶。在某些情况下,连结酶将第一RNA分子的编码区与第二RNA分子的编码区联接在一起,从而形成编码目的蛋白的较长的RNA分子。

[0098] 例如,在一个实施方案中,所述组合物包含第一RNA分子,其中第一RNA分子包含编码区和3'核酶,其中该3'核酶能够催化自身脱离该RNA分子,从而留下具有3'P或2'3'环状磷酸酯(cP)末端的编码区。在一个实施方案中,3'核酶包括HDV核酶。进一步地,在一个实施方案中,所述组合物包含第二RNA分子,其中该第二RNA分子包含编码区和5'核酶,其中该5'核酶能够催化自身脱离该RNA分子,从而留下具有5'OH末端的编码区。在一个实施方案中,5'核酶包括HH核酶。在某些情况下,连结酶将第一RNA分子的编码区与第二RNA分子的编码区联接在一起,从而形成编码目的蛋白的较长的RNA分子。

[0099] 在某些实施方案中,第一RNA包含编码目的蛋白的第一部分的编码区,并且第二RNA包含编码目的蛋白的第二部分的编码区,因此核酶介导的RNA分子切割和连结酶介导的RNA分子组装导致产生编码具有第一和第二部分的蛋白的RNA分子。本发明可用于由多种RNA产生全长蛋白,每种RNA均包含编码全长蛋白的一部分的编码区。进一步地,本发明可用于产生包含多个结构域的融合蛋白,其中每种RNA分子均包含编码融合蛋白的结构域的编码区。例如,本发明可被用于通过由包含编码前导序列、N-端标签或C-端标签的编码序列的第一RNA和包含编码蛋白质的编码序列的第二RNA分子组装RNA来产生编码具有前导序列、N-端标签、C-端标签等的蛋白质的RNA分子。

[0100] 在某些实施方案中,本发明涉及由三种或更多种单独的RNA分子形成单种RNA分子。例如,在某些方面,所述组合物包含编码第一RNA分子的核酸分子,其中第一RNA分子包含编码蛋白质的N-端区域的编码区;编码第二RNA分子的核酸分子,其中第二RNA分子包含编码蛋白质的C-端区域的编码区;以及一种或多种编码一种或多种附加RNA分子的核酸分子,每种附加RNA分子均包含编码蛋白质的结构域(例如,重复结构域)的编码区。在一个实施方案中,第一RNA分子包含编码N-端区域的编码区和3'核酶,其中3'核酶能够催化自身脱离RNA分子,从而留下具有3'P或2'3'环状磷酸酯(cP)末端的编码区。在一个实施方案中,3'

核酶包括HDV核酶。在一个实施方案中,第二RNA分子包含编码C-端区域的编码区和5'核酶,其中5'核酶能够催化自身脱离RNA分子,从而留下具有5'OH末端的编码区。在一个实施方案中,5'核酶包括HH核酶。在一个实施方案中,该附加RNA分子各自包含编码蛋白结构域的编码区、3'核酶和5'核酶。在一个实施方案中,3'核酶是HDV核酶。在一个实施方案中,5'核酶是HH核酶。在某些方面,3'核酶能够催化自身脱离RNA分子并且5'核酶能够催化自身脱离RNA分子,从而留下具有5'OH和3'P或2'3'cP末端的编码区。在一个实施方案中,该附加RNA分子各自包含编码蛋白的结构域的编码区、5'核酶和3'核酶识别序列。在某些方面,5'核酶能够催化自身脱离RNA分子,从而留下具有5'OH末端的编码区;并且3'核酶识别序列与核酶相互作用从而诱导3'核酶识别序列从RNA分子中剪接出来,从而留下具有3'P或2'3'cP末端的编码区。在一个实施方案中,3'核酶识别序列包含与VS核酶相互作用的Vsv1序列。该技术可被用于产生编码具有多个重复结构域的蛋白的RNA分子,这通过如下方式实现:通过依次提供核酶(例如VS核酶)以与3'核酶识别序列相互作用从而产生3'P或2'3'cP末端来依次添加编码重复结构域的编码区,并且将编码区连结到另一个编码重复结构域的编码区的5'OH末端。在某些方面,重复结构域的顺序添加可以在固体衬底或支持物上执行,在该固体衬底或支持物中,编码N-端区域的第一RNA分子与衬底或支持物结合。

[0101] 在某些方面,多个RNA分子在核酶介导的5'OH和3'P或2'3'cP末端产生后被连结在一起。在一些情况下,该RNA分子通过存在于发生RNA组装的天然细胞或组织中的内源性连接酶被连结在一起。在一些情况下,本发明的方法包括添加外源连接酶以诱导经加工的RNA分子连结在一起的步骤。在一个实施方案中,连接酶是RNA 2',3'-环状磷酸酯和5'-OH (RtcB) 连接酶。

[0102] 组合物

[0103] 在一个实施方案中,本发明涉及一种组合物,包含一种或多种编码一种或多种核酶的核酸分子。在一个实施方案中,本发明包括一种或多种含有一种或多种核酶的RNA分子。在一些实施方案中,该一种或多种RNA分子包含至少第一RNA分子和第二RNA分子。

[0104] 在一些实施方案中,所述组合物的所述一种或多种核酶能够从所述一种或多种RNA分子自发顺式切割。在一些实施方案中,所述一种或多种核酶是3'核酶。在一些实施方案中,所述3'核酶在自发顺式切割后在剩余的一种或多种RNA分子上产生3'P或2'3'cP末端。在一些实施方案中,所述一种或多种核酶是5'核酶。在一些实施方案中,所述5'核酶在自发顺式切割后在剩余的一种或多种RNA分子上产生5'OH末端。在一些实施方案中,所述3'P或2'3'cP末端和所述5'OH末端可以被连结在一起。

[0105] 在一些实施方案中,所述第一RNA分子包含3'核酶。在一些实施方案中,所述3'核酶来自选自以下组成的组的一个或多个家族:锤头状(HH)核酶、丁型肝炎病毒(HDV)核酶、Varkud卫星(VS)核酶、Twister(Twst)核酶、Sister核酶、Twister-sister(TS)核酶、发夹状核酶、Hatchet核酶和Pistol核酶,或它们的保持了顺式切割功能的变体或片段。在一些实施方案中,3'核酶包含一个或多个核苷酸的突出端。在一个实施方案中,该突出端包含与第一RNA分子内的所述3'核酶上游的序列杂交的核苷酸序列。在一些实施方案中,该突出端提高了自发顺式切割的效率。

[0106] 在一些实施方案中,所述第二RNA分子包含5'核酶。在一些实施方案中,所述5'核酶来自选自以下组成的组的一个或多个家族:锤头状(HH)核酶、丁型肝炎病毒(HDV)核

酶、Varkud卫星(VS)核酶、Twister(Twst)核酶、Sister核酶、Twister-sister(TS)核酶、发夹状核酶、Hatchet核酶和Pistol核酶,或它们的保持了顺式切割功能的变体或片段。在一些实施方案中,所述5'核酶包含一个或个种核苷酸的突出端。在一个实施方案中,该突出端包含与第二RNA分子内的所述5'核酶下游的序列杂交的核苷酸序列。在一些实施方案中,该突出端提高了自发顺式切割的效率。

[0107] 在一个实施方案中,所述组合物的HDV核酶包含选自由以下组成的组的一种或多种:HDV、HDV68、HDV67、HDV56、genHDV和抗HDV,或它们的变体或片段。在一个实施方案中,HDV68包含SEQ ID NO:9的核酸序列。在一个实施方案中,HDV67包含SEQ ID NO:10的核酸序列。在一个实施方案中,HDV56包含SEQ ID NO:11的核酸序列。在一个实施方案中,genHDV包含SEQ ID NO:12的核酸序列。在一个实施方案中,抗HDV包含SEQ ID NO:13的核酸序列。

[0108] 在一个实施方案中,HH核酶包含在茎1突出端中的一个或多个核苷酸,该核苷酸与该HH核酶上游或下游序列的核苷酸杂交。在一个实施方案中,该茎1突出端中的核苷酸数目可以是1个或更多个核苷酸、2个或更多个核苷酸、4个或更多个核苷酸、6个或更多个核苷酸、8个或更多个核苷酸、10个或更多个核苷酸、12个或更多个核苷酸、14个或更多个核苷酸、16个或更多个核苷酸、18个或更多个核苷酸,或20个或更多个核苷酸。在一个实施方案中,包含一个或多个核苷酸茎1突出端的HH核酶包含选自由以下组成的组的核酸序列:SEQ ID NO:111、SEQ ID NO:112、SEQ ID NO:113、SEQ ID NO:114、SEQ ID NO:115、SEQ ID NO:116、SEQ ID NO:117和SEQ ID NO:118,其中指定为N的核苷酸对应于与所述HH核酶下游序列的核苷酸杂交的核苷酸。在一个实施方案中,HH核酶具有在茎3突出端中的一个或多个核苷酸。在一个实施方案中,HH核酶具有5个核苷酸的茎3突出端。在一个实施方案中,HH核酶包含SEQ ID NO:105的核酸序列,其中指定为N的核苷酸对应于与所述HH核酶上游序列的核苷酸杂交的核苷酸。在一个实施方案中,HH核酶在茎2环中被修饰。在一个实施方案中,具有经修饰的茎2环的HH核酶包含选自由以下组成的组的核酸序列:SEQ ID NO:119、SEQ ID NO:120、SEQ ID NO:121、SEQ ID NO:122、SEQ ID NO:123和SEQ ID NO:124,其中指定为N的核苷酸对应于与所述HH核酶下游序列的核苷酸杂交的核苷酸。在一个实施方案中,HH核酶在茎1中被修饰成包括三级稳定基序(TSM)。在一个实施方案中,所述HH核酶在茎2环中被修饰,并且在茎1中被修饰成包括三级稳定基序(TSM)。在一个实施方案中,经修饰的HH核酶顺式切割比HH核酶更有效。在一个实施方案中,经修饰的HH核酶是RzB。在一个实施方案中,RzB包含SEQ ID NO:125的核酸序列,其中指定为N的核苷酸对应于与所述HH核酶下游序列的核苷酸杂交的核苷酸。

[0109] 在一个实施方案中,Twister核酶包含SEQ ID NO:32的核酸序列。在一个实施方案中,Twister核酶包含在P1茎突出端中的一个或多个核苷酸。在一个实施方案中,P1茎突出端中的核苷酸数目可以是1个或更多、2个或更多、3个或更多、4个或更多,或5个或更多。

[0110] 在一个实施方案中,包含一个或多个核苷酸P1茎突出端的Twister核酶包含选自由SEQ ID NO:106、SEQ ID NO:107、SEQ ID NO:108、SEQ ID NO:109和SEQ ID NO:110组成的组的核酸序列,其中指定为N的核苷酸对应于与所述Twister核酶下游序列的核苷酸杂交的核苷酸。

[0111] 在一些实施方案中,所述组合物的所述一种或多种核酶由第一部分和第二部分组成。在一些实施方案中,所述第一部分被掺入到所述一种或多种RNA分子中。在一些实施方

案中,所述第一部分是核酶识别序列。在一些实施方案中,所述第二部分是单独引入的。在一些实施方案中,所述第一部分从所述一种或多种RNA分子的顺式切割仅在所述第一部分和所述第二部分彼此接触时发生。在一些实施方案中,所述一种或多种核酶是VS核酶。在一个实施方案中,所述VS核酶包含SEQ ID NO:14的核酸序列。在一个实施方案中,所述第一部分是VS核酶茎环(VS-S)。在一个实施方案中,VS-S包含SEQ ID NO:15的核酸序列。在一个实施方案中,所述第二部分是不含茎环的VS的剩余部分(VS-Rz)。在一个实施方案中,VS-Rz包含SEQ ID NO:16的核酸序列。

[0112] 核酶是以顺式切割以产生如本文所述的独特RNA 3'和5'端的自催化RNA。然而,顺式切割核酶可以被工程化成以反式切割,使得靶RNA可以以核苷酸特异性方式被切割,从而产生相似的RNA端。在一些实施方案中,本发明包括包含编码单种RNA分子的单个核酸分子的组合物,该单种RNA分子包含反式切割工程化核酶。在一个实施方案中,该反式切割工程化核酶能够反式切割单独的RNA分子。在一个实施方案中,该反式切割工程化核酶识别单独的RNA分子中的特定核酸序列。在一些实施方案中,该反式切割工程化核酶靶向致病突变以进行删除。在一些实施方案中,该致病突变位于外显子中。在一些实施方案中,该致病突变位于内含子中。在一些实施方案中,所述组合物包含两种反式切割工程化核酶,该两种反式切割工程化核酶靶向致病突变的上游和下游。在一些实施方案中,在致病突变的上游和下游的反式切割导致该致病突变的去除。在一些实施方案中,在该致病突变的反式切割后,该基因的剩余部分被反式剪接在一起。在一些实施方案中,经反式剪接的基因表达为功能性蛋白。

[0113] 如本文所述,已经经受核酶介导的切割的RNA分子的3'P或2'3'cP末端和5'OH末端可以被连结在一起。因此,编码较大全长蛋白的分离的部分的RNA序列可以以无痕的方式被反式剪接在一起,以实现全长蛋白的表达。在一个实施方案中,本发明涉及一种组合物,该组合物包含一种或多种编码目的蛋白的两个或更多个部分并且编码一种或多种核酶的核酸分子。在一个实施方案中,本发明涉及一种组合物,该组合物包含一种或多种编码目的蛋白的两个或更多个部分并且包含一种或多种核酶的RNA分子。

[0114] 在一个实施方案中,所述一种或多种编码目的蛋白的两个或多个部分的核酸分子包括编码目的蛋白的第一部分的第一核酸分子和编码目的蛋白的第二部分的第二核酸分子。在一个实施方案中,所述第一核酸包含第一RNA分子。在一个实施方案中,所述第二核酸包含第二RNA分子。在一个实施方案中,所述第一RNA分子在3'末端与3'核酶连接。在一个实施方案中,所述第二RNA分子在5'端与5'核酶连接。在一个实施方案中,在3'和5'核酶序列的顺式切割后,第一RNA分子的3'P或2'3'cP末端被连结到第二RNA分子的5'OH末端,从而产生编码全长目的蛋白的单种RNA分子。在一个实施方案中,该全长目的蛋白的功能与相同序列的内源表达的全长蛋白的功能是同一的。

[0115] 在一个实施方案中,该全长目的蛋白包含治疗性蛋白。在一个实施方案中,该治疗性蛋白包含选自由以下(但是不限于以下)组成的组的一种或多种:Utrophin、抗肌萎缩蛋白(Dystrophin)、Dysferlin、Myoferlin、囊性纤维化跨膜电导调节剂(CFTR)、凝血因子VIII、纤维囊蛋白(Fibrocystin)、视网膜特异性磷脂-转运ATP酶(ABCA4)、耳畸蛋白(Otoferlin)、铜转运ATP酶2、MYO7A、MYO15A、CDH23、STRC、OTOG、TECTA、PCDH15、TRIOBP、MYO3A、COL11A2、LOXHD1、PTPRQ、OTOGL、MYH14、MYH9、TNC、CACNA1A、CACNA1C、CACNA1F、

CACNA1H、CACNA1G、CACNA1D、CACNA1B、CACNA1S、CACNA1I、CACNA1E、ATP2A1、ATP2A2、Adcy6、FKBP12-雷帕霉素结合域和Cas9。在一个实施方案中,该全长目的蛋白是重组酶。在一个实施方案中,该重组酶是选自由以下(但是不限于以下)组成的组的一种或多种:CRE重组酶、FLP重组酶。在一个实施方案中,该全长目的蛋白是真核/原核抗生素抗性基因产物。在一个实施方案中,该真核/原核抗生素抗性基因产物是选自由以下(但是不限于以下)组成的组的一种或多种:氨苄青霉素、卡那霉素、杀稻瘟菌素、嘌呤霉素、新霉素和潮霉素。在某些实施方案中,该全长目的蛋白是抗体。在一个实施方案中,该抗体能够结合至目的靶蛋白。在一些实施方案中,该抗体是抗体片段、合成抗体、纳米抗体或它们的保持了结合至靶蛋白的能力的片段或变体。在一个实施方案中,该全长目的蛋白包括合成重复蛋白,包括但不限于构成水凝胶、合成蜘蛛丝和胶原蛋白的那些。在一个实施方案中,该合成重复蛋白包括选自由以下(但不限于以下)组成的组的一种或多种:Spidroin、丝蛋白(Silk)、角蛋白(Keratin)、胶原蛋白(Collagen)、弹性蛋白(Elastin)、节肢弹性蛋白(Resilin)、鱿鱼环齿蛋白(Squid Ring Teeth)、 β 螺旋管蛋白(beta-solenoid protein)、锌指核酸酶(Zinc Finger Nuclease)(ZFN)和Tal效应器核酸酶(TALEN)。在一个实施方案中,该全长目的蛋白包括可抑制哺乳动物包装细胞中慢病毒颗粒的产生的有毒蛋白或抗病毒蛋白。在一个实施方案中,该有毒蛋白是细胞自杀基因。在一个实施方案中,该细胞自杀基因包括选自由以下(但是不限于以下)组成的组的一种或多种:白喉毒素A(DTA)、HSV-tk、蓖麻毒素、霍乱毒素、主要朊病毒蛋白、百日咳毒素、Ectatomin、芋螺肽(Conopeptide)、相思豆毒素(Abrin)、维罗毒素(Verotoxin)、破伤风痉挛毒素(Tetanospasmin)、肉毒杆菌毒素、假单胞菌外毒素A、炭疽、皂草素和商陆抗病毒蛋白(pokeweed antiviral protein,PAP)。在一个实施方案中,抗病毒蛋白包括选自由以下(但是不限于以下)组成的组的一种或多种:干扰素诱导的GTP结合蛋白(MxA)、髓过氧化物酶(MPO)和干扰素。

[0116] 编码目的蛋白的一部分的N-端或C-端RNA分子可能会在核酶介导的切割之前经受翻译,或者当被单独表达时,可能会导致不需要的或截短的蛋白表达。然而,可以利用蛋白降解序列的翻译控制器(translational control)来限制这种不需要的表达。在一个实施方案中,所述组合物的所述一种或多种RNA分子包含编码蛋白降解序列的翻译控制器的核酸序列。在一个实施方案中,所述第一RNA分子包含编码蛋白降解序列的翻译控制器的核酸序列。在一个实施方案中,所述第二RNA分子包含编码蛋白降解序列的翻译控制器的核酸序列。在一些实施方案中,所述蛋白降解序列的翻译控制器在核酶序列的切割和剪接之前阻止蛋白的部分表达。在一些实施方案中,所述蛋白降解序列的翻译控制器包括选自由以下组成的组的一种或多种:hCL1-PEST序列、E1A-PEST序列、核酸的poly(A)序列的去除、通过poly A尾进行的用于产生poly K尾的模拟翻译、ATG终止密码子的删除、N-端NTG密码子内的沉默突变、编码作为翻译抑制剂发挥作用的四个小上游ORF的酵母GCN4序列的5'UTR、酵母GCN4序列的5'UTR的小内部片段。在一些实施方案中,所述蛋白降解序列的翻译控制器包含一种或多种选自由以下组成的组的核酸序列:SEQ ID NO:43、SEQ ID NO:44、SEQ ID NO:45、SEQ ID NO:46、SEQ ID NO:47、SEQ ID NO:48、SEQ ID NO:49、SEQ ID NO:77、SEQ ID NO:79和SEQ ID NO:104。在一些实施方案中,所述蛋白降解序列的翻译控制器包含一种或多种选自由以下组成的组的氨基酸序列:SEQ ID NO:52、SEQ ID NO:53、SEQ ID NO:54、SEQ ID NO:55、SEQ ID NO:56、SEQ ID NO:57、SEQ ID NO:58、SEQ ID NO:59、SEQ ID NO:60、SEQ

ID NO:61、SEQ ID NO:62、SEQ ID NO:63、SEQ ID NO:64、SEQ ID NO:65、SEQ ID NO:66、SEQ ID NO:67、SEQ ID NO:68、SEQ ID NO:69、SEQ ID NO:70、SEQ ID NO:71、SEQ ID NO:72、SEQ ID NO:73、SEQ ID NO:74、SEQ ID NO:76、SEQ ID NO:78和SEQ ID NO:80。

[0117] 在某些方面,为了进一步阻止不需要的或截短的蛋白表达, RNA核定位信号可用于阻止未经剪接的RNA分子的胞浆输出和翻译。在一个实施方案中,所述组合物的所述一种或多种RNA分子包含编码RNA核定位序列的核酸序列。在一个实施方案中,所述第一RNA分子包含编码RNA核定位序列的核酸序列。在一个实施方案中,所述第二RNA分子包含编码RNA核定位序列的核酸序列。在一个实施方案中,所述RNA核定位序列在核酶序列的切割和剪接之前阻止不完全蛋白 (partial protein) 的胞质RNA输出和翻译。在一个实施方案中,所述RNA核定位序列包含一种或多种选自以下组成的组的核酸序列: SEQ ID NO:50和SEQ ID NO:51。

[0118] 在一些实施方案中,所述组合物进一步包含一种或多种附加RNA分子,每种附加RNA分子均包含编码目的蛋白的结构域的编码区;5'核酶;和3'核酶。在一些实施方案中,所述系统进一步包含一种或多种编码一种或多种附加RNA分子的附加核酸分子,每种附加RNA分子均包含编码目的蛋白的结构域的编码区、5'核酶,和3'核酶。

[0119] 在一些实施方案中,所述组合物进一步包含一种或多种附加RNA分子,每种附加RNA分子均包含编码目的蛋白的结构域的编码区;5'核酶;和3'核酶识别序列。在一些实施方案中,所述系统进一步包含一种或多种编码一种或多种附加RNA分子的附加核酸分子,每种附加RNA分子均包含编码目的蛋白的结构域的编码区、5'核酶,和3'核酶识别序列。

[0120] 由剪接体进行的前mRNA剪接已被证明可通过沉积促进首轮翻译的因子或者通过促进RNA加工和向细胞质的输出来增强mRNA翻译。在转基因内添加嵌合顺式剪接内含子也已被证明可促进转基因蛋白表达。因此,在某些实施方案中,被剪接体识别和顺式剪接的剪接供体和剪接受体位点的添加可以增强来自分裂的前体RNA分子的蛋白表达。在一个实施方案中,所述组合物包含一种或多种RNA分子,该RNA分子包含剪接供体或剪接受体序列。在一个实施方案中,所述组合物的所述第一RNA分子包含剪接供体序列。在一个实施方案中,所述剪接供体序列连接至核酶序列之后的第一RNA分子的3'末端。在一个实施方案中,所述组合物的所述第二RNA分子包含剪接受体序列。在一个实施方案中,所述剪接受体序列连接至核酶序列之前的第二RNA分子的5'末端。在一个实施方案中,剪接供体和剪接受体序列的纳入核增强了酶介导的反式剪接后的蛋白表达。

[0121] 由于蛋白在其中被翻译的三个开放阅读框,核酶介导的同时多种不同功能蛋白的反式剪接和表达也可能是可能的。通过利用这一特性,可以使用位于三个不同的不相容开放阅读框中的RNA的反式剪接来产生功能性蛋白。在一个实施方案中,本发明的组合物包含至少四种核酸分子,该核酸分子包括至少两对核酸分子。在一个实施方案中,每对核酸分子编码目的蛋白的至少两个部分并且编码至少两种核酶。在一个实施方案中,所述组合物包含至少四种RNA分子,该至少四种RNA分子包括至少两对RNA分子。在一个实施方案中,每对RNA分子编码目的蛋白的至少两个部分并且包含至少两种核酶。

[0122] 在一个实施方案中,所述至少两对RNA分子包括第一对RNA分子和第二对RNA分子。在一个实施方案中,所述第一对RNA分子包括第一RNA分子和第二RNA分子。在一个实施方案中,所述第二对RNA分子包括第三RNA分子和第四RNA分子。在一些实施方案中,所述第三RNA

分子和所述第四RNA分子与所述第一RNA分子和所述第二RNA分子具有不同的开放阅读框,使得在自发顺式切割后,所述第一RNA分子或所述第二RNA分子与所述第三RNA分子或第四RNA分子的连结不能翻译全长的功能性蛋白产物。

[0123] 在一个实施方案中,所述至少两对RNA分子进一步包括第三对RNA分子。在一个实施方案中,第三对RNA分子包括第五RNA分子和第六RNA分子。在一些实施方案中,所述第五RNA分子和所述第六RNA分子具有与所述第一对RNA分子和所述第二对RNA分子不同的开放阅读框,以使得在自发顺式切割后,仅所述第一对、第二对或第三对RNA分子的连结可以翻译全长的功能性蛋白产物。

[0124] 如本文所述,当一种RNA含有3'核酶而另一种RNA含有5'核酶时,两种独立RNA之间可以发生核酶介导的反式剪接。然而,当在同一种RNA分子中以顺式被转录时,两种核酶可以介导它们自己的无疤痕去除。这种方法类似地产生两个独立的具有3'-P和5'OH端的RNA,该RNA可以在细胞中经受反式剪接和翻译。在所述3'和5'核酶之间纳入货物序列也会产生在连结后产生环化RNA分子的可能性。

[0125] 在一个实施方案中,本发明涉及一种组合物,该组合物包含单一的编码目的蛋白的两个或更多个部分并且编码一种或多种核酶的核酸分子。在一个实施方案中,本发明涉及一种组合物,该组合物包含单一的编码目的蛋白的两个或更多个部分并且包含一种或多种核酶的RNA分子。

[0126] 在一个实施方案中,所述单个核酸分子编码RNA的第一部分、合成内含子,和RNA的第二部分。在一个实施方案中,该合成内含子包含5'核酶和3'核酶。在一个实施方案中,RNA的所述第一部分编码目的蛋白的第一部分。在一个实施方案中,RNA的所述第二部分编码目的蛋白的第二部分。在一个实施方案中,所述单个核酸包含按以下顺序连接的序列:(编码目的蛋白的第一部分的RNA的第一部分)-(合成内含子的5'核酶)-(合成内含子的3'核酶)-(编码目的蛋白的第二部分的RNA的第二部分)。在一个实施方案中,所述目的蛋白的所述第一部分是GFP的N-端部分。在一个实施方案中,所述合成内含子的5'核酶包括HDV。在一个实施方案中,RNA的第一部分和合成内含子的5'核酶包含SEQ ID NO:127的核酸序列,其中小写字母表示5'核酶序列并且大写字母表示编码GFP的N-端部分的序列(参见实施例4,“含有具有和不具有货物的内部合成核酶内含子的GFP”)。在一个实施方案中,所述目的蛋白的所述第二部分是GFP的C-端部分。在一个实施方案中,所述合成内含子的所述3'核酶包含HH。在一个实施方案中,所述RNA的第二部分和所述合成内含子的3'核酶包含SEQ ID NO:128的核酸序列,其中小写字母表示3'核酶序列并且大写字母表示编码GFP的C-端部分的序列。(参见实施例4,“含有具有和不具有货物的内部合成核酶内含子的GFP”)。

[0127] 在一个实施方案中,所述合成内含子包含置于所述5'核酶和所述3'核酶之间的货物序列。在一个实施方案中,所述单个核酸包含按以下顺序连接的序列:(编码目的蛋白的第一部分的RNA的第一部分)-(合成内含子的5'核酶)-(货物序列)-(合成内含子的3'核酶)-(编码目的蛋白的第二部分的RNA的第二部分)。

[0128] 在一个实施方案中,所述合成内含子的5'核酶序列不需要双侧侧接序列来发挥活性。在一个实施方案中,由包含不需要双侧侧接序列来发挥活性的5'核酶序列的合成内含子末端的连结产生的环状RNA可以以环状形式和经再切割的线性形式存在。在一个实施方案中,所述核酶序列是HDV核酶。

[0129] 在一个实施方案中,合成内含子的5'核酶序列确实需要双侧侧接序列来发挥活性。在一个实施方案中,由包含确实需要双侧侧接序列来发挥活性的5'核酶序列的合成内含子末端的连结产生的环状RNA可仅以环状形式存在。在一个实施方案中,所述核酶序列是HH核酶。

[0130] 在一个实施方案中,所述合成内含子的5'核酶序列是核酶识别序列。在一个实施方案中,该核酶识别序列需要添加反式切割核酶以进行诱导型切割。在一个实施方案中,所述核酶识别序列包含VS-S。在一些实施方案中,VS-S由包含SEQ ID NO:15的核酸序列编码。在一个实施方案中,所述反式切割核酶包含VS-Rz。在一些实施方案中,VS-S由包含SEQ ID NO:16的核酸序列编码。

[0131] 在一个实施方案中,5'核酶序列和3'核酶序列的自切割产生以下三个单独的RNA分子:1) 第一片段,包含编码目的蛋白的第一部分的RNA的第一部分;2) 第二片段,包含合成内含子;3) 第三片段,包含编码目的蛋白的第二部分的RNA的第二部分。在一个实施方案中,第二片段的相容末端被连结从而产生包含含有货物序列的合成内含子的环状RNA分子。在实施方案中,第一片段和第三片段被连结在一起从而产生单个全长线性RNA分子。

[0132] 在一个实施方案中,合成内含子的货物序列是选自由以下组成的组的一种或多种:编码目的治疗性蛋白的序列、CRISPR指导RNA序列、小RNA序列和反式切割核酶序列。在一个实施方案中,所述小RNA序列包括选自由以下组成的组的一种或多种:微RNA(miRNA)、Piwi相互作用RNA(piRNA)、小干扰RNA(siRNA)、小核仁RNA(snoRNA)、小tRNA来源的RNA(tsRNA)、小rDNA来源的RNA(srRNA)和小核RNA(snRNA)。

[0133] 在一个实施方案中,所述单个全长线性RNA分子编码全长目的蛋白。在一个实施方案中,该全长目的蛋白是治疗性蛋白。在一个实施方案中,该治疗性蛋白可以是但不限于选自由以下组成的组的一种或多种:Utrophin、抗肌萎缩蛋白、Dysferlin、Myoferlin、囊性纤维化跨膜电导调节剂(CFTR)、凝血因子VIII、纤维囊蛋白、视网膜特异性磷脂-转运ATP酶(ABCA4)、耳畸蛋白、铜转运ATP酶2、MYO7A、MYO15A、CDH23、STRC、OTOGL、TECTA、PCDH15、TRIOBP、MYO3A、COL11A2、LOXHD1、PTPRQ、OTOGL、MYH14、MYH9、TNC、CACNA1A、CACNA1C、CACNA1F、CACNA1H、CACNA1G、CACNA1D、CACNA1B、CACNA1S、CACNA1I、CACNA1E、ATP2A1、ATP2A2、Adcy6、FKBP12-雷帕霉素结合域和Cas9。在一个实施方案中,该全长目的蛋白是重组酶。在一个实施方案中,该重组酶是选自由以下(但是不限于以下)组成的组的一种或多种:CRE重组酶、FLP重组酶。在一个实施方案中,该全长目的蛋白是真核/原核抗生素抗性基因产物。在一个实施方案中,该真核/原核抗生素抗性基因产物是选自由以下(但是不限于以下)组成的组的一种或多种:氨苄青霉素、卡那霉素、杀稻瘟菌素、嘌呤霉素、新霉素和潮霉素。在一个实施方案中,该全长目的蛋白是报告蛋白。在一个实施方案中,该报告蛋白是选自由以下组成的组的一种或多种:绿色荧光蛋白(GFP)、红色荧光蛋白(RFP)和荧光素酶(Luc)。在一个实施方案中,该报告蛋白用作用于评估货物序列的递送和表达的代表性指标。在某些实施方案中,该全长目的蛋白是抗体。在一个实施方案中,该抗体能够结合至目的靶蛋白。在一些实施方案中,该抗体是抗体片段、合成抗体、纳米抗体或它们的保持了结合至靶蛋白的能力的片段或变体。

[0134] 在某些方面,本发明的技术可用于组装全长RNA病毒基因组。在一个实施方案中,所述一种或多种编码本发明的一种或多种核酶的核酸分子编码RNA病毒基因组的一个或多

个部分。在一个实施方案中,包含本发明的一种或多种核酶的所述一种或多种RNA分子包含RNA病毒基因组的一个或多个部分。

[0135] 在一个实施方案中,所述一种或多种核酸分子包含编码RNA病毒基因组的第一部分并且编码3'核酶的第一核酸分子。在一个实施方案中,所述一种或多种核酸分子包含编码RNA病毒基因组的第二部分并且编码5'核酶的第二核酸。在一个实施方案中,所述一种或多种RNA分子包括包含RNA病毒基因组的第一部分和3'核酶的第一RNA分子。在一个实施方案中,所述一种或多种RNA分子包括包含RNA病毒基因组的第二部分和5'核酶的第二RNA分子。在一个实施方案中,所述组合物包含编码连结酶的核酸或连结酶。在一个实施方案中,在3'和5'核酶顺式切割后,RNA病毒基因组的第一部分和RNA病毒基因组的第二部分被连结在一起,从而产生全长RNA病毒基因组。示例性的RNA病毒包括但不限于:冠状病毒、副粘病毒、正粘病毒、逆转录病毒、慢病毒、甲病毒、黄病毒、弹状病毒、麻疹病毒、新城病病毒和小核糖核酸病毒。

[0136] 在一些实施方案中,本发明包括包含编码连结酶的核酸的组合物。在一些实施方案中,该连结酶介导3'P或2'3'cP末端和5'OH末端的连结。在一些实施方案中,该连结酶是RNA 2',3'-环状磷酸酯和5'-OH(RtcB)连结酶。在一些实施方案中,该RtcB连结酶来自选自自由以下组成的组的生物体的一种或多种结构域:真核生物、细菌和古生菌。在一些实施方案中,该生物体选自自由以下组成的组:人、大肠杆菌、耐辐射异常球菌(*Deinococcus radiodurans*)、霍里科什火球菌(*Pyrococcus horikoshii*)、火球菌属种(*Pyrococcus sp.*) ST04,和热球菌属种EP。在一些实施方案中,编码连结酶的核酸序列是选自自由以下组成的一种或多种:SEQ ID NO:82、SEQ ID NO:84、SEQ ID NO:86、SEQ ID NO:88、SEQ ID NO:90、SEQ ID NO:92。在一些实施方案中,编码连结酶的核酸序列编码一种或多种选自自由以下组成的组的氨基酸序列:SEQ ID NO:81、SEQ ID NO:83、SEQ ID NO:85、SEQ ID NO:87、SEQ ID NO:89、SEQ ID NO:91。

[0137] 核酸

[0138] 在一些实施方案中,本发明的一种或多种核酸包含与本文所述的核酸序列基本上同源的核酸序列。例如,在一些实施方案中,该核酸与原始核酸序列的同一性程度为至少60%、至少65%、至少70%、至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%,或至少99.5%。

[0139] 在一些实施方案中,本发明的一种或多种核酸包含作为本文所述核酸序列的一部分的核酸序列。例如,在一些实施方案中,该核酸具有相对于原始核酸序列的至少60%、至少65%、至少70%、至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%或至少99.5%的长度。

[0140] 在一些实施方案中,本发明的一种或多种核酸包含作为本文所述核酸序列的一部分的核酸序列,并且与本文所述核酸序列基本上同源。例如,在一些实施方案中,该核酸与原始核酸序列的同一性程度为至少60%、至少65%、至少70%、至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少

98%、至少99%、或至少99.5%，并且/或者该核酸具有相对于原始核酸序列的至少60%、至少65%、至少70%、至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%、或至少99.5%的长度。

[0141] 本发明的核酸可以包括任何类型的核酸，包括但不限于DNA和RNA。例如，在一个实施方案中，所述组合物包含编码本发明的融合蛋白的分离的DNA分子，包括例如分离的cDNA分子。在一个实施方案中，所述组合物包含编码本发明的融合蛋白或其功能片段的分离的RNA分子。

[0142] 可以修饰本发明的核酸分子以提高在血清中或在用于细胞培养的生长培养基中的稳定性。可以添加修饰以增强稳定性、功能性和/或特异性以及最小化本发明核酸分子的免疫刺激特性。例如，为了增强稳定性，可以稳定3'-残基以防止降解，例如，它们可以被选择成使得它们由嘌呤核苷酸，特别是腺苷或鸟苷核苷酸组成。可替代地，嘧啶核苷酸被经修饰的类似物的取代（例如尿苷被2'-脱氧胸苷的取代）能被耐受并且不影响该分子的功能。

[0143] 在本发明的一个实施方案中，所述核酸分子可以含有至少一种经修饰的核苷酸类似物。例如，可以通过掺入经修饰的核苷酸类似物来稳定末端。

[0144] 核苷酸类似物的非限制性例子包括糖和/或骨架被修饰的核糖核苷酸（即包括对磷酸-糖骨架的修饰）。例如，可以修饰天然RNA的磷酸二酯键以纳入氮或硫杂原子中的至少一者。在示例性的骨架被修饰的核糖核苷酸中，连接到相邻核糖核苷酸的磷酸酯基团被经修饰的基团（例如硫代磷酸酯基团）置换。在示例性糖被修饰的核糖核苷酸中，2'-OH-基团被选自H、OR、R、卤基、SH、SR、NH₂、NHR、NR₂或ON的基团置换，其中R是C₁-C₆烷基、烯基或炔基，并且卤基是F、Cl、Br或I。

[0145] 修饰的其他实例是核碱基被修饰的核糖核苷酸，即，含有至少一种非天然存在的核碱基而不是天然存在的核碱基的核糖核苷酸。碱基可以被修饰成阻断腺苷脱氨酶的活性。示例性经修饰的核碱基包括但不限于在5-位被修饰的尿苷和/或胞苷，例如5-(2-氨基)丙基尿苷、5-溴尿苷；在8位被修饰的腺苷和/或鸟苷，例如8-溴鸟苷；脱氮核苷酸，例如7-脱氮-腺苷；O-烷基化核苷酸和N-烷基化核苷酸，例如N6-甲基腺苷是合适的。应当注意，上述修饰可以组合。

[0146] 在一些情况下，核酸分子包含以下化学修饰中的至少一种：一个或多个核苷酸的2'-H、2'-O-甲基或2'-OH修饰。在某些实施方案中，本发明的核酸分子可以对核酸酶具有增强的抗性。为了增加核酸酶抗性，核酸分子可以包含，例如，2'-经修饰的核糖单元和/或硫代磷酸酯键。例如，2'-羟基(OH)可以被许多不同的“氧基”或“脱氧基”取代基修饰或置换。为了增加核酸酶抗性，本发明的核酸分子可以包含2'-O-甲基、2'-氟、2'-O-甲氧基乙基、2'-O-氨基丙基、2'-氨基和/或硫代磷酸酯键。锁核酸(LNA)、乙烯核酸(ENA)（例如2'-4'-乙烯桥连的核酸），以及某些核碱基修饰（诸如2-氨基-A、2-硫代（例如2-硫代-U）、G-clamp修饰）的纳入也可以增加与靶标的结合亲和力。

[0147] 在一个实施方案中，所述核酸分子包括2'-经修饰的核苷酸，例如2'-脱氧、2'-脱氧-2'-氟、2'-O-甲基、2'-O-甲氧基乙基(2'-O-MOE)、2'-O-氨基丙基(2'-O-AP)、2'-O-二甲基氨基乙基(2'-O-DMAOE)、2'-O-二甲基氨基丙基(2'-O-DMAP)、2'-O-二甲基氨基乙氧基乙基(2'-O-DMAEOE)或2'-O-N-甲基乙酰胺基(2'-O-NMA)。在一个实施方案中，所述核酸分子

包括至少一个2'-O-甲基被修饰的核苷酸,并且在一些实施方案中,所述核酸分子的所有核苷酸都包括2'-O-甲基修饰。

[0148] 在某些实施方案中,本发明的核酸分子具有以下一种或多种性质:

[0149] 本文讨论的核酸剂包括在其他方面未经修饰的RNA和DNA以及已经例如为了提高功效而被修饰的RNA和DNA,以及核苷替代物的聚合物。未经修饰的RNA是指其中核酸的成分(即糖、碱基和磷酸酯部分)与自然界中或人体内天然存在的成分相同或基本相同的分子。本领域将稀有或不常见但天然存在的RNA称为经修饰RNA,参见例如Limbach等人(Nucleic Acids Res., 1994, 22:2183-2196)。此类常常被称为经修饰RNA的稀有或不常见的RNA通常是转录后修饰的结果,并且在如本文所用的术语未经修饰的RNA的范围内。如本文所用的经修饰RNA是指其中核酸的组分(即糖、碱基和磷酸酯部分)中的一种或多种组分不同于自然界中存在的一种或多种组分或不同于人体中存在的一种或多种组分的分子。虽然它们被称为“经修饰的RNA”,但它们当然会因为修饰而包括严格来说不是RNA的分子。核苷替代物是其中核糖磷酸酯骨架被允许碱基以正确的空间关系呈现的非核糖磷酸酯构建体置换,从而使得杂交基本上类似于在核糖磷酸酯骨架情况下所见的杂交的分子(例如核糖磷酸酯骨架的不带电的模拟物)。

[0150] 本发明的核酸的修饰可以存在于磷酸酯基团、糖基团、骨架、N-端、C-端或核碱基中的一者或多者处。

[0151] 载体

[0152] 本发明还包括包含一种或多种插入有一种或多种本发明的核酸分子的载体的组合物。在一个实施方案中,该载体编码至少两种RNA分子。在一个实施方案中,该系统包含至少两种RNA分子。在一些实施方案中,该至少两种RNA分子由同一载体编码。在一些实施方案中,该至少两种RNA分子包含在同一载体内。在一个实施方案中,所述至少两种RNA分子包含第一RNA分子和第二RNA分子。

[0153] 在一些实施方案中,本发明包括编码至少两种RNA分子的至少两种载体。在一些实施方案中,该至少两种载体包含至少两种RNA分子。在一些实施方案中,该至少两种载体编码单独RNA分子。在一个实施方案中,该至少两种载体包含单独RNA分子。在一些实施方案中,该至少两种单独RNA分子包含第一RNA分子和第二RNA分子。在一些实施方案中,第一RNA分子由第一载体编码并且第二RNA分子由第二载体编码。在一些实施方案中,第一RNA分子包含第一载体并且第二RNA分子包含第二载体。

[0154] 在一些实施方案中,本发明进一步包括编码一种或多种附加RNA分子的载体。在一些实施方案中,本发明进一步包括一种或多种包含一种或多种附加RNA分子的载体。在一些实施方案中,每种附加RNA分子均包含编码目的蛋白的结构域的编码区、5'核酶,和3'核酶。在一些实施方案中,每种附加RNA分子均包含编码目的蛋白的结构域的编码区、5'核酶,和3'核酶识别序列。

[0155] 本领域存在很多可用于本发明的合适载体。简而言之,编码本发明融合蛋白的天然或合成核酸的表达通常通过将编码本发明融合蛋白或其部分的核酸可操作地连接至启动子,并且将构建体掺入到表达载体中来实现。待使用的载体适合于在真核细胞中复制和任选地整合。典型的载体含有转录和翻译终止子、起始序列和可用于调节所需核酸序列表达的启动子。

[0156] 本发明的载体也可使用标准基因递送方案用于核酸免疫和基因疗法。用于基因递送的方法是本领域已知的。参见,例如,美国专利号5,399,346、5,580,859、5,589,466,该美国专利通过引用整体并入本文。在另一个实施方案中,本发明提供了一种基因治疗载体。

[0157] 本发明的分离的核酸可以被克隆到许多类型的载体中。例如,所述核酸可被克隆到包括但不限于质粒、噬菌粒、噬菌体衍生物、动物病毒和粘粒的载体中。特别感兴趣的载体包括表达载体、复制载体、探针产生载体和测序载体。

[0158] 进一步地,该载体可以以病毒载体的形式被提供给细胞。病毒载体技术在本领域中是众所周知的并且例如在Sambrook等人(2012,Molecular Cloning:A Laboratory Manual,Cold Spring Harbor Laboratory,New York)中以及在其他病毒学和分子生物学手册中有描述。可用作载体的病毒包括但不限于逆转录病毒、腺病毒、腺相关病毒、疱疹病毒和慢病毒。一般而言,合适的载体含有在至少一种生物体中起作用的复制起点、启动子序列、方便的限制性内切核酸酶位点和一种或多种选择标记(例如,WO 01/96584、WO 01/29058和美国专利号6,326,193)。

[0159] 进一步地,已经开发了许多另外的用于将基因转移到哺乳动物细胞中的基于病毒的系统。例如,逆转录病毒为基因递送系统提供了一个方便的平台。可以使用本领域已知的技术将选定的基因插入到载体中并包装在逆转录病毒颗粒中。然后可以分离出重组病毒并且在体内或离体将其递送至受试者的细胞。许多逆转录病毒系统是本领域已知的。在一些实施方案中,使用腺病毒载体。许多腺病毒载体是本领域已知的。

[0160] 在一个实施方案中,所述组合物包含衍生自腺相关病毒(AAV)的载体。术语“AAV载体”意指衍生自腺相关病毒血清型的载体,包括但不限于AAV-1、AAV-2、AAV-3、AAV-4、AAV-5、AAV-6、AAV-7、AAV-8和AAV-9。AAV载体已成为用于治疗各种病症的强大基因递送工具。AAV载体具有许多使其非常适合于基因疗法的特征,包括缺乏致病性、最小的免疫原性以及以稳定且有效的方式转导有丝分裂后细胞的能力。通过选择AAV血清型、启动子和递送方法的适当组合,可以使AAV载体内包含的特定基因的表达特异性地靶向至一种或多种类型的细胞。

[0161] AAV载体可以具有一个或多个AAV野生型基因(优选rep和/或cap基因)的全部或部分缺失,但保留了功能性侧接ITR序列。尽管同源性程度很高,但不同的血清型对不同的组织具有嗜性。AAV1的受体是未知的;然而,已知AAV1比AAV2更有效地转导骨骼肌和心肌。由于大多数研究都是用假型载体完成的,其中侧接有AAV2 ITR的载体DNA被包装到交替血清型的衣壳中,很明显,生物学差异与衣壳有关,而不是与基因组有关。最近的证据表明,包装在AAV 1衣壳中的DNA表达盒在转导心肌细胞方面的效率比包装在AAV2衣壳中的那些高至少1个log₁₀。在一个实施方案中,病毒递送系统是腺相关病毒递送系统。该腺相关病毒可以是血清型1(AAV 1)、血清型2(AAV2)、血清型3(AAV3)、血清型4(AAV4)、血清型5(AAV5)、血清型6(AAV6)、血清型7(AAV7)、血清型8(AAV8)或血清型9(AAV9)。

[0162] 用于组装成载体的理想AAV片段包括帽蛋白(包括vp1、vp2、vp3和高变区)、rep蛋白(包括rep 78、rep 68、rep 52和rep 40),以及编码这些蛋白的序列。这些片段可以很容易地用于各种载体系统和宿主细胞。此类片段可单独使用,与其他AAV血清型序列或片段组合使用,或与来自其他AAV或非AAV病毒序列的元件组合使用。如本文所用,人工AAV血清型包括但不限于具有非天然存在的衣壳蛋白的AAV。此种人工衣壳可以通过任何合适的技术

使用选择的AAV序列(例如, vp1衣壳蛋白的片段)与异源序列的组合来产生, 该异源序列可以从不同的选择的AAV血清型、同一AAV血清型的非连续部分、非AAV病毒来源, 或非病毒来源获得。人工AAV血清型可以是但不限于嵌合AAV衣壳、重组AAV衣壳或“人源化”AAV衣壳。因此, 适用于表达一种或多种蛋白的示例性AAV或人工AAV包括AAV2/8(参见美国专利号7, 282, 199)、AAV2/5(可从美国国立卫生研究院获得)、AAV2/9(国际专利公布号W02005/033321)、AAV2/6(美国专利号6, 156, 303)和AAVrh8(国际专利公布号W02003/042397), 等等。

[0163] 在一个实施方案中, 所述组合物包含慢病毒载体以递送一种或多种本发明的核酸。在一个实施方案中, 本发明包括一种慢病毒载体, 该慢病毒载体包含一种或多种编码一种或多种目的蛋白的RNA分子。例如, 来源于逆转录病毒(诸如慢病毒)的载体是用于实现长期基因转移的合适工具, 因为它们允许转基因的长期稳定整合以及转基因在子细胞中的繁殖。慢病毒载体与来源于肿瘤逆转录病毒(诸如鼠白血病病毒)的载体相比具有附加优势, 因为它们可以转导非增殖细胞, 如肝细胞。它们还具有低免疫原性的额外优势。

[0164] 在某些实施方案中, 该载体还包括常规控制元件, 该常规控制元件以允许其在用质粒载体转染或感染本发明产生的病毒的细胞中转录、翻译和/或表达的方式可操作地连接到转基因。如本文所用, “可操作地连接的”序列包括与目的基因邻接的表达控制序列和反式或远距离作用以控制目的基因的表达控制序列。表达控制序列包括适当的转录起始、终止、启动子和增强子序列; 高效的RNA处理信号, 诸如剪接和多聚腺苷酸(polyA)信号; 稳定细胞质mRNA的序列; 提高翻译效率的序列(即Kozak共有序列); 增强蛋白稳定性的序列; 以及在需要时, 增强编码产物分泌的序列。包括天然的、组成型的、诱导型的和/或组织特异性的启动子在的大量表达控制序列在本领域中是已知的并且可以被使用。

[0165] 附加启动子元件(例如增强子)调节转录启动的频率。通常, 这些位于起始位点上游30-110bp的区域, 但是最近已证明许多启动子也含有位于起始位点下游的功能元件。启动子元件之间的间距通常是灵活的, 因此当元件相对于彼此倒置或移动时, 启动子功能得以保留。在胸苷激酶(tk)启动子中, 在活性开始下降之前, 启动子元件之间的间距可以增加至50bp。根据启动子, 似乎各个元件可以协同或独立地发挥用来活化转录。

[0166] 合适的启动子的一个例子是立即早期巨细胞病毒(CMV)启动子序列。该启动子序列是强组成型启动子序列, 其能够驱动与其可操作连接的任何多核苷酸序列的高水平表达。合适的启动子的另一个例子是延伸生长因子-1 α (EF-1 α)。然而, 也可以使用其他组成型启动子序列, 包括但不限于猴病毒40(SV40)早期启动子、小鼠乳腺肿瘤病毒(MMTV)、人类免疫缺陷病毒(HIV)长末端重复(LTR)启动子、MoMuLV启动子、禽白血病病毒启动子、爱泼斯坦-巴尔病毒(Epstein-Barrvirus)立即早期启动子、劳斯肉瘤(Roussarcoma)病毒启动子, 以及人基因启动子, 诸如但不限于肌动蛋白启动子、肌球蛋白启动子、血红蛋白启动子和肌酸激酶启动子。进一步地, 本发明不应限于使用组成型启动子。诱导型启动子也被认为是本发明的一部分。诱导型启动子的使用提供了一种分子开关, 该分子开关能够在需要与其可操作连接的多核苷酸序列的表达时开启此种表达, 或者在不需要表达时关断表达。诱导型启动子的实例包括但不限于金属硫氨酸(metallothionine)启动子、糖皮质激素启动子、孕酮启动子和四环素启动子。

[0167] 在载体上发现的增强子序列也调节其中所含基因的表达。通常, 增强子与蛋白因

子结合以增强该基因的转录。增强子可能位于它所调节的基因的上游或下游。增强子也可以是组织特异性的,以增强特定细胞或组织类型中的转录。在一个实施方案中,本发明的载体包含一种或多种增强子以促进存在于载体内的基因的转录。

[0168] 为了评估本发明的融合蛋白的表达,待引入到细胞中的表达载体还可以含有可选择标记基因或报告基因或两者,以促进从寻求通过病毒载体转染或感染的细胞群中鉴定出和选择出表达细胞。在其他方面,可选择标记可被携带在单独的DNA片段上并且被用于共转染程序中。可选择标记和报告基因都可侧接有适当的调节序列,以使其能够在宿主细胞中表达。有用的可选择标记包括例如抗生素抗性基因,诸如neo等。

[0169] 报告基因被用于识别可能转染的细胞和评估调控序列的功能性。一般而言,报告基因是这样的基因,该基因不存在于受体生物体或组织中或者不被受体生物体或组织表达,并且编码这样的多肽,该多肽的表达可通过一些容易检测的特性(例如酶活性)来证明。在将DNA引入到受体细胞中后的合适时间测定报告基因的表达。合适的报告基因可以包括编码荧光素酶、 β -半乳糖苷酶、氯霉素乙酰转移酶、分泌型碱性磷酸酶或绿色荧光蛋白基因的基因(例如,Ui-Tei等人,2000FEBS Letters 479:79-82)。合适的表达系统是众所周知的并且可以使用已知技术制备或商业获得。一般而言,具有显示出最高水平的报告基因表达的最小5'侧接区的构建体被鉴定为启动子。此类启动子区域可与报告基因连接并且可被用于评估剂调节启动子驱动的转录的能力。

[0170] 蛋白质

[0171] 在一些实施方案中,本发明包括包含连结酶的组合物。在一些实施方案中,连结酶介导RNA分子的3'P或2'3'cP末端与RNA分子的5'OH末端的连结。在一些实施方案中,该连结酶是RNA 2',3'-环状磷酸酯和5'-OH(RtcB)连结酶。在一些实施方案中,该RtcB连结酶来自自由以下组成的组的生物体的一种或多种结构域:真核生物、细菌和古生菌。在一些实施方案中,该生物体选自自由以下组成的组:人、大肠杆菌、耐辐射异常球菌、霍里科什火球菌、火球菌属种ST04,和热球菌属种EP。在一些实施方案中,该连结酶包含一种或多种选自自由以下组成的组的氨基酸序列:SEQ ID NO:81、SEQ ID NO:83、SEQ ID NO:85、SEQ ID NO:87、SEQ ID NO:89、SEQ ID NO:91。

[0172] 在一些实施方案中,本发明的一种或多种蛋白包含与本文所述的氨基酸序列基本上同源的氨基酸序列。例如,在一些实施方案中,该蛋白具有相对于原始氨基酸序列为至少60%、至少65%、至少70%、至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%或至少99.5%的同源性程度。

[0173] 在一些实施方案中,本发明的一种或多种蛋白包含作为本文所述氨基酸序列的一部分的氨基酸序列。例如,在一些实施方案中,该蛋白具有相对于原始氨基酸序列为至少60%、至少65%、至少70%、至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%或至少99.5%的长度。

[0174] 在一些实施方案中,本发明的一种或多种蛋白包含作为本文所述氨基酸序列的一部分的氨基酸序列,并且与本文所述氨基酸序列基本上同源。例如,在一些实施方案中,所

述蛋白具有相对于原始氨基酸序列为至少60%、至少65%、至少70%、至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%、或至少99.5%的同一性程度,并且/或者所述蛋白具有相对于原始氨基酸序列为至少60%、至少65%、至少至少70%、至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%或至少99.5%的长度。

[0175] 药物组合物

[0176] 本发明还涵盖本发明的药物组合物或其盐用于实施本发明的方法的用途。此种药物组合物可以由呈适合于施用于受试者的形式的至少一种本发明的核酸或其盐组成,或者该药物组合物可以包含至少一种本发明的核酸或其盐,以及一种或多种药学上可接受的载体、一种或多种附加成分,或这些的一些组合。本发明的核酸可以存在于生理上可接受的盐的形式,例如以与如本领域所熟知的生理上可接受的阳离子或阴离子的组合的药物组合物中。

[0177] 在一个实施方案中,可以施用可用于实施本发明方法的药物组合物以递送介于1ng/kg/天和100mg/kg/天之间的剂量。在另一个实施方案中,可以施用可用于实施本发明的药物组合物以递送介于1ng/kg/天和500mg/kg/天之间的剂量。

[0178] 本发明的药物组合物中活性成分、药学上可接受的载体和任何其他成分的相对量将根据被治疗受试者的身份、大小和状况以及进一步取决于组合物的施用途径而变化。举例来说,该组合物可以包含介于0.1%和100% (w/w) 之间的活性成分。

[0179] 可用于本发明方法的药物组合物可适当地被开发用于口服、直肠、阴道、肠胃外、局部、肺部、鼻内、颊部、眼部施用或另外的施用途径。在本发明的方法内有用的组合物可以直接施用于哺乳动物的皮肤或任何其他组织。其他考虑的制剂包括脂质体配制物、含有活性成分的重新密封的红细胞和基于免疫学的制剂。一种或多种施用途径对于技术人员来说是显而易见的,并且将取决于许多因素,包括所治疗疾病的类型和严重程度、所治疗的兽医或人受试者的类型和年龄等。

[0180] 本文所述的药物组合物的制剂可以通过药理学领域中已知的或药理学领域中以后开发的任何方法来制备。一般而言,此类制备方法包括以下步骤:将活性成分与载体或一种或多种其他辅助成分联合,以及然后,如果有必要或需要,将产品成型或包装成所需的单剂量或多剂量单位。

[0181] 如本文所用,“单位剂量”是包含预定量的活性成分的药物组合物的离散量。活性成分的量一般等于将被施用给受试者的活性成分的剂量或此种剂量的方便分数,诸如,例如此种剂量的二分之一或三分之一。单位剂型可以是单次日剂量,或多次日剂量(例如,每天约1至4次或更多次)的一次剂量。当使用多个日剂量时,每次剂量的单位剂型可以相同或不同。

[0182] 在一个实施方案中,使用一种或多种药学上可接受的赋形剂或载体配制本发明的组合物。在一个实施方案中,本发明的药物组合物包含治疗有效量的本发明的核酸和药学上可接受的载体。有用的药学上可接受的载体包括但不限于甘油、水、生理盐水、乙醇和其

他药学上可接受的盐溶液,诸如磷酸盐和有机酸盐。这些和其他药学上可接受的载体的例子在Remington's Pharmaceutical Sciences(1991,Mack Publication Co.,New Jersey)中有描述。

[0183] 该载体可以是溶剂或分散介质,其含有例如水、乙醇、多元醇(例如甘油、丙二醇和液体聚乙二醇等)、它们的合适混合物,以及植物油。例如,通过使用诸如卵磷脂的包衣、通过在分散体的情况下保持所需的颗粒大小,以及通过使用表面活性剂,可以保持适当的流动性。可以通过各种抗菌剂和抗真菌剂,例如对羟基苯甲酸酯、氯丁醇、苯酚、抗坏血酸、硫柳汞等来实现对微生物作用的预防。在许多情况下,所述组合物中包含等渗剂,例如糖、氯化钠或多元醇,诸如甘露醇和山梨糖醇。可通过在可注射组合物中纳入延迟吸收的剂,例如单硬脂酸铝或明胶来延长该组合物的吸收。在一个实施方案中,药学上可接受的载体不是单独的DMSO。

[0184] 制剂可以与常规赋形剂(即适用于口服、阴道、肠胃外、鼻、静脉内、皮下、肠或本领域已知的任何其他合适的施用方式的药学上可接受的有机或无机载体物质)混合使用。可以对药物配制物进行灭菌,并且如果需要,可以将其与助剂(例如润滑剂、防腐剂、稳定剂、润湿剂、乳化剂、用于影响渗透压的盐、缓冲剂、着色剂、调味剂和/或芳香物质等)混合。还可以在需要时将它们与其他活性剂,例如其他镇痛剂组合。

[0185] 如本文所用,“附加成分”包括但不限于以下中的一种或多种:赋形剂;表面活性剂;分散剂;惰性稀释剂;造粒和崩解剂;粘合剂;润滑剂;甜味剂;调味剂;着色剂;防腐剂;生理可降解组合物,诸如明胶;水性媒介物和溶剂;油性媒介物和溶剂;悬浮剂;分散剂或润湿剂;乳化剂、缓和剂;缓冲剂;盐类;增稠剂;填充物;乳化剂;抗氧化剂;抗生素;抗真菌剂;稳定剂;以及药学上可接受的聚合或疏水材料。可以包含在本发明的药物组合物中的其他“附加成分”在本领域中是已知的并且例如在Genaro编辑(1985,Remington's Pharmaceutical Sciences,Mack Publishing Co.,Easton,PA)中有描述,其通过引用并入本文。

[0186] 本发明组合物可包含按组合物总重量计约0.005%至2.0%的防腐剂。防腐剂用于防止在暴露于环境中的污染物的情况下变质。根据本发明有用的防腐剂的实例包括但不限于选自苯甲醇、山梨酸、对羟基苯甲酸酯、酰亚胺脲以及它们的组合组成的组的那些。示例性防腐剂是约0.5%至2.0%苯甲醇和0.05%至0.5%山梨酸的组合。

[0187] 在一个实施方案中,所述组合物包含抗氧化剂和抑制核酸降解的螯合剂。一些化合物的示例性抗氧化剂是按所述组合物总重量计在约0.01重量%至0.3重量%范围内的BHT、BHA、 α -生育酚和抗坏血酸,以及在0.03重量%至0.1重量%范围内的BHT。在一个实施方案中,螯合剂以按所述组合物总重量计0.01重量%至0.5重量%的量存在。示例性螯合剂包括重量范围为约0.01%至0.20%的依地酸盐(例如依地酸二钠)和柠檬酸。在一些实施方案中,螯合剂在按所述组合物总重量计的0.02重量%至0.10重量%的范围内。螯合剂可用于螯合组合物中可能对制剂的保质期有害的金属离子。虽然BHT和依地酸二钠分别是一些化合物的示例性抗氧化剂和螯合剂,但其他合适的等效的抗氧化剂和螯合剂因此可以被取代,如本领域技术人员已知的。

[0188] 可以使用常规方法制备液体悬浮液以实现活性成分在水性或油性媒介物中的悬浮。水性媒介物包括例如水和等渗生理盐水。油性媒介物包括例如杏仁油、油性酯、乙醇、植

物油(如花生油、橄榄油、芝麻油或椰子油)、分馏植物油和矿物油(如液体石蜡)。液体悬浮液可进一步包含一种或多种附加成分,该附加成分包括但不限于悬浮剂、分散剂或润湿剂、乳化剂、缓和剂、防腐剂、缓冲剂、盐、调味剂、着色剂和甜味剂。油性悬浮液可进一步包含增稠剂。已知的悬浮剂包括但不限于山梨糖醇糖浆、氢化可食用脂肪、海藻酸钠、聚乙烯吡咯烷酮、黄耆胶、阿拉伯树胶和纤维素衍生物(如羧甲基纤维素钠、甲基纤维素、羟丙基甲基纤维素)。已知的分散剂或润湿剂包括但不限于天然存在的磷脂(诸如卵磷脂);烯化氧与脂肪酸、与长链脂族醇、与衍生自脂肪酸和己糖醇的偏酯,或与衍生自脂肪酸和己糖醇酐的偏酯的缩合产物(例如分别为硬脂酸聚氧乙烯酯、十七烯氧基鲸蜡醇(heptadecaethyleneoxy cetanol)、单油酸聚氧乙烯山梨糖醇酯和单油酸聚氧乙烯山梨醇酐酯)。已知的乳化剂包括但不限于卵磷脂和阿拉伯胶。已知的防腐剂包括但不限于对羟基苯甲酸甲酯、对羟基苯甲酸乙酯或对羟基苯甲酸正丙酯、抗坏血酸和山梨酸。已知的甜味剂包括例如甘油、丙二醇、山梨糖醇、蔗糖和糖精。用于油性悬浮液的已知增稠剂包括例如蜂蜡、硬石蜡和鲸蜡醇。

[0189] 活性成分在水性或油性溶剂中的液体溶液可以以与液体悬浮液基本相同的方式制备,主要区别在于活性成分溶解在溶剂中,而不是悬浮在溶剂中。如本文所用,“油性”液体是一种包含含碳液体分子并且表现出比水更小的极性特征的液体。本发明的药物组合物的液体溶液可以包含针对液体悬浮液描述的组分中的每一种组分,应当理解,悬浮剂将不一定有助于活性成分在溶剂中的溶解。水性溶剂包括例如水和等渗生理盐水。油性溶剂包括例如杏仁油、油性酯、乙醇、植物油(如花生油、橄榄油、芝麻油或椰子油)、分馏植物油和矿物油(如液体石蜡)。

[0190] 本发明药物配制物的粉末和颗粒制剂可以使用已知方法制备。此类制剂可被直接施用给受试者,被用来例如形成片剂、填充胶囊或通过向其中添加水性或油性媒介物来制备水性或油性悬浮液或溶液。这些制剂中的每一种制剂均可以进一步包含分散剂或润湿剂、悬浮剂和防腐剂中的一种或多种。这些制剂中还可以包含附加赋形剂,诸如填充剂和甜味剂、调味剂或着色剂。

[0191] 本发明的药物组合物还可以以水包油乳剂或油包水乳剂的形式制备、包装或销售。油相可以为植物油(诸如橄榄油或花生油)、矿物油(诸如液体石蜡),或这些的组合。此类组合物可进一步包含一种或多种乳化剂,诸如天然存在的树胶(诸如阿拉伯树胶或黄耆胶)、天然存在的磷脂(诸如大豆或卵磷脂)、衍生自脂肪酸和己糖醇酐的酯或偏酯(诸如脱水山梨糖醇单油酸酯),以及此类偏酯与环氧乙烷的缩合产物(诸如单油酸聚氧乙烯山梨醇酐酯)。这些乳剂还可以包含附加成分,包括例如甜味剂或调味剂。

[0192] 用化学组合物浸渍或涂覆材料的方法在本领域中是已知的,并且包括但不限于将化学组合物沉积或结合到表面上的方法、在材料的合成期间将化学组合物掺入到该材料的结构中的方法(即,诸如用生理可降解材料),以及将水或油溶液或悬浮液吸收到吸收性材料中(随后进行干燥或不进行干燥)的方法。

[0193] 施用方案可以影响有效量的构成。可在疾病诊断之前或之后将治疗性制剂施用给受试者。进一步地,可以每天或依次施用若干个分开的剂量以及交错的剂量,或者该剂量可以被连续输注,或者可以是团注。进一步地,治疗性制剂的剂量可以根据治疗或预防情况的紧急程度按比例增加或减少。

[0194] 可以使用已知程序,以有效预防或治疗疾病的剂量和时间段向受试者(包括哺乳

动物,例如人)施用本发明组合物。实现治疗效果所需的核酸的有效量可以根据诸如以下的因素而变化:所用特定核酸的活性;施用时间;核酸的排泄率;治疗持续时间;与该核酸组合使用的其他药物、化合物或材料;接受治疗的受试者的疾病或病症的状态、年龄、性别、体重、状况、一般健康状况和既往病史;以及医学领域中众所周知的类似因素。可以调整给药方案以提供最佳的治疗响应。例如,可以每天施用若干分开的剂量或者可以根据治疗情况的紧急程度按比例减小剂量。本发明的核酸化合物的有效剂量范围的非限制性实例为约1至5,000mg/kg体重/每天。本领域普通技术人员将能够在无需过多实验的情况下研究相关因素并确定治疗性核酸的有效量。

[0195] 该核酸可以按每天数次的频率施用于受试者,或者可以按更低的频率施用,例如每天一次、每周一次、每两周一次、每月一次,或者按甚至更低的频率施用,例如每几个月一次或甚至一年一次或更久一次。应当理解,在非限制性实例中,每天给药的核酸量可以每天、隔天、每2天、每3天、每4天或每5天施用。例如,就每隔一天施用而言,可以在星期一启动5mg/天剂量,然后在星期三施用第一后续的5mg/天剂量,然后在星期五施用第二后续的5mg/天剂量,依此类推。给药频率对技术人员来说将是显而易见的,并且取决于许多因素,例如但不限于所治疗疾病的类型和严重程度、动物的类型和年龄等。

[0196] 本发明药物组合物中活性成分的实际剂量水平可以变化,以获得有效实现特定受试者、组合物和施用方式的所需治疗反应,而不对受试者有毒的活性成分的量。

[0197] 具有本领域普通技术的医生,例如医师或兽医可以容易地确定和开出所需组合物有效量。例如,医师或兽医可以使药物组合物中使用的本发明核酸的开始剂量低于实现期望治疗效果所需的水平,并逐渐增加该剂量直至实现所需效果。

[0198] 在特定的实施方案中,将核酸配制成剂量单位形式是特别有利的,以便于施用和剂量的均匀性。如本文所用的剂量单位形式是指适合作为用于待治疗的受试者的单一剂量的物理上离散单位;每个单位均含有经计算可产生期望治疗效果的与所需药物媒介物联合的预定量的治疗性核酸。本发明的剂量单位形式由以下决定由并且直接取决于以下:(a)核酸的独特特性和要达到的特定治疗效果,以及(b)复合/配制此种核酸用于治疗受试者疾病的技术中固有的限制。

[0199] 在一个实施方案中,本发明组合物以每天1至5次或更多次的剂量施用于受试者。在另一个实施方案中,本发明组合物以包括但不限于每天一次、每两天一次、每三天一次至每周一次,和每两周一次的剂量范围施用于受试者。对本领域技术人员而言,很明显,本发明的各种组合组合物的施用频率将因受试者而异,这取决于许多因素,包括但不限于年龄、待治疗的疾病或病症、性别、整体健康状况和其他因素。因此,本发明不应被解读为限于任何特定的剂量方案,并且将由主治医生在考虑与任何受试者有关的所有其他因素的情况下确定给该受试者施用的精确剂量和组合物。

[0200] 用于施用的本发明组合物可以在以下范围内:约1mg至约10,000mg、约20mg至约9,500mg、约40mg至约9,000mg、约75mg至约8,500mg、约150mg至约7,500mg、约200mg至约7,000mg、约3050mg至约6,000mg、约500mg至约5,000mg、约750mg至约4,000mg、约1mg至约3,000mg、约10mg至约2,500mg、约20mg至约2,000mg、约25mg至约1,500mg、约50mg至约1,000mg、约75mg至约900mg、约100mg至约800mg、约250mg至约750mg、约300mg至约600mg、约400mg至约500mg,以及它们之间的任何和所有全部或部分增量。

[0201] 在一些实施方案中,本发明组合物的剂量为约1mg至约2,500mg。在一些实施方案中,用于本文所述组合物中的本发明组合物的剂量小于约10,000mg,或小于约8,000mg,或小于约6,000mg,或小于约5,000mg,或小于约3,000mg,或小于约2,000mg,或小于约1,000mg,或小于约500mg,或小于约200mg,或小于约50mg。类似地,在一些实施方案中,如本文所述的第二组合物(即,用于治疗与通过本发明组合物治疗的疾病相同的疾病或另一种疾病的药物)的剂量小于约1,000mg,或小于约800mg,或小于约600mg,或小于约500mg,或小于约400mg,或小于约300mg,或小于约200mg,或小于约100mg,或小于约50mg,或小于约40mg,或小于约30mg,或小于约25mg,或小于约20mg,或小于约15mg,或小于约10mg,或小于约5mg,或小于约2mg,或小于约1mg,或小于约0.5mg,以及它们的任何和所有全部或部分增量。

[0202] 在一个实施方案中,本发明涉及一种包装的药物组合物,该组合物包括容器,该容器容纳单独或与第二药剂组合的治疗有效量的本发明核酸;以及使用核酸治疗、预防或减轻受试者的一种或多种疾病症状的说明书。

[0203] 术语“容器”包括用于容纳药物组合物的任何容器。例如,在一个实施方案中,容器是含有药物组合物的包装。在其他实施方案中,容器不是含有药物组合物的包装,即容器是含有包装的药物组合物或未包装的药物组合物和药物组合物的使用说明书的器皿,诸如盒子或小瓶。此外,包装技术在本领域中是众所周知的。应当理解,药物组合物的使用说明书可以包含在含有药物组合物的包装上,因此说明书与经包装产品形成了增强的功能关系。然而,应当理解,说明书可以含有与所述核酸执行其预期功能(例如,治疗或预防受试者的疾病,或将成像剂或诊断剂递送至受试者)的能力有关的信息。

[0204] 本发明的任何组合物的施用途径包括口服、鼻、肠胃外、舌下、经皮、经粘膜(例如,舌下、舌、(经)颊部和鼻(内))、膀胱内、十二指肠内、胃内、直肠、腹膜内、皮下、肌肉内、真皮内、动脉内、静脉内或施用。

[0205] 合适的组合物和剂型包括例如片剂、胶囊、囊片、丸剂、凝胶帽、锭剂、分散体、悬浮液、溶液、糖浆剂、颗粒剂、珠粒、透皮贴剂、凝胶、粉末、小丸剂(pellet)、乳浆剂、糖锭(lozenge)、霜剂、糊剂、膏药、洗剂、圆片(disc)、栓剂、用于鼻或口服施用的液体喷雾剂、用于吸入的干粉剂或雾化制剂、用于膀胱内施用的组合物和制剂等。应当理解,可用于本发明的制剂和组合物不限于本文所述的特定制剂和组合物。

[0206] 系统

[0207] 在一些实施方案中,本发明涉及用于独立RNA分子的顺式切割和反式剪接的系统。在一些实施方案中,本发明涉及单种RNA分子的顺式切割和反式剪接系统。在一些实施方案中,独立RNA分子或单个RNA分子片段的顺式切割和反式剪接产生编码如本文所述的全长目的蛋白的单个RNA分子。在一些实施方案中,该系统包含如本文所述的连结酶或编码连结酶(诸如RtcB)的核酸。

[0208] 在一个实施方案中,本发明涉及一种用于从编码全长蛋白的第一部分和第二部分的两个单独RNA分子通过核酶的顺式切割和该两个独立RNA分子的反式剪接来产生编码全长蛋白的单种RNA的诱导型系统。在一些实施方案中,该系统包含如本文所述的核酶识别序列和核酶。在一些实施方案中,该系统包含如本文所述的连结酶或编码连结酶的核酸。

[0209] 在一个实施方案中,本发明涉及一种组装全长RNA病毒基因组的系统。示例性的RNA病毒包括但不限于:冠状病毒、副粘病毒、正粘病毒、逆转录病毒、慢病毒、甲病毒、黄病

毒、弹状病毒、麻疹病毒、新城病病毒和小核糖核酸病毒。在一个实施方案中,该系统包含编码RNA病毒基因组的第一部分并且编码3'核酶的第一核酸。在一个实施方案中,该系统包含编码RNA病毒基因组的第二部分并且编码5'核酶的第二核酸。在一个实施方案中,该系统包含RNA病毒基因组的第一部分,和3'核酶。在一个实施方案中,该系统包含RNA病毒基因组的第二部分,和5'核酶。在一个实施方案中,该系统包含编码连结酶的核酸或连结酶。在一个实施方案中,在3'和5'核酶顺式切割后,RNA病毒基因组的第一部分和RNA病毒基因组的第二部分被连结在一起,从而产生全长RNA病毒基因组。

[0210] 体内

[0211] 在一个实施方案中,本发明涉及一种用于通过编码一种或多种全长蛋白的部分的独立RNA分子的顺式切割和反式剪接来递送和表达该全长蛋白的系统。在一些实施方案中,该系统允许递送和表达超过传统载体包装尺寸的大蛋白(例如,超过AAV载体包装尺寸的抗肌萎缩蛋白)、其中的核酸构建体难以体外合成的合成重复结构域蛋白(例如,合成蜘蛛丝),或有毒/抗病毒蛋白(例如DTA)。在一个实施方案中,本发明包括用于递送和表达一种或多种全长目的蛋白的AAV系统。在一些实施方案中,该系统包含如本文所述的连结酶或编码连结酶的核酸。

[0212] 在一个实施方案中,本发明包括一种慢病毒递送系统用以递送一种或多种编码一种或多种目的蛋白的核酸分子。在一个方面,慢病毒递送系统包含(1)包装质粒,(2)包膜质粒,和(3)转移质粒。在一个实施方案中,该转移质粒编码第一RNA分子和第二RNA分子。

[0213] 在一个实施方案中,本发明包括双慢病毒递送系统,包含第一慢病毒载体和第二慢病毒载体。在一个实施方案中,第一慢病毒载体系统包含(1)包装质粒,(2)包膜质粒,和(3)第一转移质粒。在一个实施方案中,第二慢病毒载体系统包含(1)包装质粒,(2)包膜质粒,和(3)第二转移质粒。在一个实施方案中,第一转移质粒编码第一RNA分子。在一个实施方案中,第二转移质粒编码第二RNA分子。

[0214] 在一个实施方案中,包装质粒包含编码gag-pol多聚蛋白的核酸序列。在一个实施方案中,gag-pol多聚蛋白包含催化死亡的整合酶。在一个实施方案中,gag-pol多聚蛋白包含D116N整合酶突变。

[0215] 在一个实施方案中,包膜质粒包含编码包膜蛋白的核酸序列。在一个实施方案中,包膜质粒包含编码HIV包膜蛋白的核酸序列。在一个实施方案中,包膜质粒包含编码水疱性口炎病毒g-蛋白(VSV-g)包膜蛋白的核酸序列。在一个实施方案中,可以基于所需的细胞类型来选择包膜蛋白。

[0216] 在一个实施方案中,单个转移质粒的第一RNA分子包含编码目的蛋白的第一部分的蛋白编码区和3'核酶。在一个实施方案中,单个转移质粒的第二RNA分子包含编码目的蛋白的第二部分的蛋白编码区和5'核酶。在一个实施方案中,该转移质粒包含5'长末端重复(LTR)序列和3'LTR序列。在一个实施方案中,3'LTR是自失活(SIN)LTR。因此,在一个实施方案中,5'LTR包含U3序列、R序列和U5序列,并且3'LTR包含R序列和U5序列,但不包含U3序列。在一个实施方案中,5'LTR和3'LTR与编码目的蛋白的第一部分和目的蛋白的第二部分的序列侧接。

[0217] 在一个实施方案中,第一转移质粒的第一RNA分子包含编码目的蛋白的第一部分的蛋白编码区和3'核酶。在一个实施方案中,第二转移质粒的第二RNA分子包含编码目的蛋

白的第二部分的蛋白编码区和5'核酶。在一个实施方案中,该第一和第二转移质粒包含5'长末端重复(LTR)序列和3'LTR序列。在一个实施方案中,3'LTR是自失活(SIN)LTR。因此,在一个实施方案中,5'LTR包含U3序列、R序列和U5序列,并且3'LTR包含R序列和U5序列,但不包含U3序列。在一个实施方案中,第一转移质粒的5'LTR和3'LTR与编码目的蛋白的第一部分和3'核酶的序列侧接。在一个实施方案中,第二转移质粒的5'LTR和3'LTR与编码目的蛋白的第二部分和5'核酶的序列侧接。

[0218] 在一个实施方案中,将包装质粒、包膜质粒和转移质粒引入到细胞中。在一个实施方案中,该细胞转录并且翻译编码gag-pol蛋白的核酸序列以产生gag-pol多聚蛋白。在一个实施方案中,该细胞转录并且翻译编码包膜蛋白的核酸序列以产生包膜蛋白。在一个实施方案中,该细胞转录单个转移质粒以提供第一RNA分子和第二RNA分子。在一个实施方案中,该细胞转录第一转移质粒以提供第一RNA分子并且转录第二转移质粒以提供第二RNA分子。在一个实施方案中,gag-pol蛋白、包膜多聚蛋白、第一RNA分子和第二RNA分子被包装到病毒颗粒中。在一个实施方案中,病毒颗粒是从细胞培养基中收集的。在一个实施方案中,该病毒颗粒转导靶细胞,其中3'核酶催化自身脱离第一RNA分子,从而产生3'P或2'3'cP末端,5'核酶催化自身脱离第二RNA分子,从而产生5'OH末端,内源性RNA 2',3'-环状磷酸酯,并且5'-OH(RtcB)连结酶将该3'P或2'3'cP末端连结到该5'OH末端,从而产生编码目的蛋白的完整RNA分子,并且该细胞翻译目的蛋白。

[0219] 在一个实施方案中,包装质粒、包膜质粒和第一转移质粒被引入到细胞中。在一个实施方案中,该细胞转录并且翻译编码gag-pol蛋白的核酸序列以产生gag-pol多聚蛋白。在一个实施方案中,该细胞转录并且翻译编码包膜蛋白的核酸序列以产生包膜蛋白。在一个实施方案中,该细胞转录第一转移质粒以提供第一RNA分子。在一个实施方案中,gag-pol蛋白、包膜多聚蛋白、第一RNA分子被包装到第一病毒颗粒中。在一个实施方案中,第一病毒颗粒是从细胞培养基中收集的。

[0220] 在一个实施方案中,包装质粒、包膜质粒和第二转移质粒被引入到细胞中。在一个实施方案中,该细胞转录并且翻译编码gag-pol蛋白的核酸序列以产生gag-pol多聚蛋白。在一个实施方案中,该细胞转录并且翻译编码包膜蛋白的核酸序列以产生包膜蛋白。在一个实施方案中,该细胞转录第二转移质粒以提供第二RNA分子。在一个实施方案中,gag-pol蛋白、包膜多聚蛋白、第二RNA分子被包装到第二病毒颗粒中。在一个实施方案中,第二病毒颗粒是从细胞培养基中收集的。

[0221] 在一个实施方案中,第一病毒颗粒和第二病毒颗粒转导靶细胞,其中3'核酶催化自身脱离第一RNA分子,从而产生3'P或2'3'cP末端,5'核酶催化自身脱离第二RNA分子,从而产生5'OH末端,内源性RNA 2',3'-环状磷酸酯,并且5'-OH(RtcB)连结酶将该3'P或2'3'cP末端连结到该5'OH末端,从而产生编码目的蛋白的完整RNA分子,并且该细胞翻译目的蛋白。在一个实施方案中,本发明涉及一种阻止来自分裂的前体RNA分子的不想要的蛋白表达的系统。在一个实施方案中,该系统包含在如本文所述的分裂的前体RNA分子中掺入蛋白降解序列的翻译控制器。

[0222] 在一个实施方案中,本发明涉及一种用于从编码目的蛋白的部分的两对或更多对独立RNA分子通过核酶的顺式切割以及该成对独立RNA分子的反式剪接来表达两种或更多种目的蛋白的系统。在一个实施方案中,每个单个对的独立RNA分子具有单独的阅读框,使

得不需要的对的反式剪接不会导致如本文所述的全长功能蛋白的翻译。在一些实施方案中,该系统包含如本文所述的连结酶或编码连结酶的核酸。

[0223] 在一个实施方案中,本发明包括用于递送和表达全长目的蛋白和货物序列的系统。在一个实施方案中,所述系统包含编码目的蛋白的第一部分的RNA的第一部分,该RNA的第一部分以其3'末端连接到合成内含子;以及编码的目的蛋白的第二部分的RNA的第二部分,该RNA的第二部分以其5'末端连接到合成内含子。在一个实施方案中,所述合成内含子在任一侧侧接有5'核酶序列和3'核酶序列。在一个实施方案中,所述合成内含子包含被置于所述5'核酶序列和3'核酶序列之间的货物序列。在一个实施方案中,5'核酶序列和3'核酶序列的自切割产生以下三个单独的RNA分子:1) 第一片段,包含编码目的蛋白的第一部分的RNA的第一部分;2) 第二片段,包含合成内含子;3) 第三片段,包含编码目的蛋白的第二部分的RNA的第二部分。在一个实施方案中,第二片段的相容末端被连结从而产生包含含有货物序列的合成内含子的环状RNA分子。在实施方案中,第一片段和第三片段被连结在一起从而产生单个全长线性RNA分子。在一个实施方案中,全长目的蛋白包括治疗性蛋白、报告蛋白、重组酶、抗生素抗性基因产物、抗体或Cas9蛋白。在一个实施方案中,货物序列包含治疗性核酸序列(例如,miRNA序列或CRISPR指导RNA序列)或编码治疗性蛋白。在一些实施方案中,全长目的蛋白包含Cas9并且货物序列包含指导RNA序列,从而使Cas9靶向特定基因组序列以进行编辑。在一些实施方案中,该系统包含如本文所述的连结酶或编码连结酶的核酸。

[0224] 在一个实施方案中,本发明包括一种用于基因编辑的系统,该系统包含一种或多种反式切割工程化核酶。在一些实施方案中,所述系统包含两种反式切割工程化核酶,该两种反式切割工程化核酶靶向致病突变的上游和下游。在一些实施方案中,在致病突变的上游和下游的反式切割导致该致病突变的去除。在一些实施方案中,在该致病突变的反式切割后,该基因的剩余部分被反式剪接在一起。在一些实施方案中,经反式剪接的基因表达为功能性蛋白。在一些实施方案中,该系统包含如本文所述的连结酶或编码连结酶的核酸。

[0225] 体外

[0226] 在一个实施方案中,本发明包括用于产生编码目的蛋白的RNA分子的体外系统。在一个实施方案中,该系统包含至少两种RNA分子。在一个实施方案中,所述至少两种RNA分子包含第一RNA分子和第二RNA分子。

[0227] 在一个实施方案中,所述第一RNA分子包含编码目的蛋白的第一部分的编码区。在一个实施方案中,所述第一RNA分子包含3'核酶。在一个实施方案中,所述第一RNA分子包含如本文所述的编码目的蛋白的第一部分的编码区和3'核酶。

[0228] 在一个实施方案中,所述第二RNA分子包含编码目的蛋白的第二部分的编码区。在一个实施方案中,所述第二RNA分子包含5'核酶。在一个实施方案中,所述第二RNA分子包含如本文所述的编码目的蛋白的第二部分的编码区和5'核酶。

[0229] 在一个实施方案中,用于产生编码目的蛋白的RNA分子的体外系统进一步包含连结酶。在一个实施方案中,该连结酶诱导由第一RNA分子的编码区和第二RNA分子的编码区组装RNA分子。在一个实施方案中,所述连结酶是如本文所述的RNA 2',3'-环状磷酸酯和5'-OH(RtcB)连结酶。

[0230] 在一个实施方案中,本发明包括用于产生编码目的重复结构域蛋白的RNA分子的体外系统。在一个实施方案中,所述系统包含第一RNA分子、一种或多种附加RNA分子和最后

一种RNA分子。

[0231] 在一个实施方案中,所述第一RNA分子包含编码目的蛋白的第一部分的编码区。在一个实施方案中,所述第一RNA分子包含3'核酶。在一个实施方案中,所述第一RNA分子包含编码目的蛋白的第一部分的编码区和3'核酶。在一个实施方案中,3'核酶催化自身脱离第一RNA分子,从而产生3'P或2'3'cP末端。在一个实施方案中,所述第一RNA分子进一步包含5'标签。在一个实施方案中,所述5'标签介导所述第一RNA分子与固体支持物的附接。

[0232] 在一个实施方案中,所述一种或多种附加RNA分子包含编码目的蛋白的结构域的编码区、5'核酶,和3'核酶识别序列。在一个实施方案中,所述5'核酶切割自身从而产生5'OH末端。在一个实施方案中,所述3'核酶识别序列包含如本文所述的VS-S序列。

[0233] 在一个实施方案中,所述最后一种RNA分子包含编码目的蛋白的最后一个部分的编码区。在一个实施方案中,所述最后一种RNA分子包含5'核酶。在一个实施方案中,所述最后一种RNA分子包含编码目的蛋白的最后一个部分的编码区和5'核酶。在一个实施方案中,所述5'核酶切割自身从而产生5'OH末端。

[0234] 在一个实施方案中,该系统进一步包含核酶。在一个实施方案中,所述核酶包含如本文所述的VS-Rz。在一个实施方案中,如本文所述,所述VS-Rz识别VS-S,并且介导VS-S从一种或多种附加RNA分子上的切割。在一个实施方案中,所述切割产生3'P或2'3'cP末端。

[0235] 在一个实施方案中,该系统包含连结酶。在一些实施方案中,该连结酶将第一RNA分子的3'P或2'3'cP末端连结到该一种或多种附加RNA分子的5'OH末端。在一些实施方案中,该连结酶将该一种或多种附加RNA分子的3'P或2'3'cP末端连结到最后一种RNA分子的5'OH末端。在一些实施方案中,连结酶将第一RNA分子的3'P或2'3'cP末端连结到该一种或多种附加RNA分子的5'OH末端,并且将该一种或多种附加RNA分子的3'P或2'3'cP末端连结到最后一种RNA分子的5'OH末端,从而产生编码N-端结构域、一个或多个附加结构域和C-端结构域的完整RNA分子。在一些实施方案中,该连结酶是如本文所述的RNA 2',3'-环状磷酸酯和5'-OH(RtcB)连结酶。

[0236] 方法

[0237] 在一些实施方案中,本发明涉及独立RNA分子的顺式切割和反式剪接的方法。在一些实施方案中,本发明涉及单种RNA分子的顺式切割和反式剪接方法。在一些实施方案中,独立RNA分子或单个RNA分子片段的顺式切割和反式剪接产生编码如本文所述的全长目的蛋白的单个RNA分子。在一些实施方案中,该方法包括施用如本文所述的连结酶或编码连结酶的核酸。

[0238] 在一个实施方案中,本发明涉及一种用于从编码全长蛋白的第一部分和第二部分的两个单独RNA分子通过核酶的顺式切割和该两个独立RNA分子的反式剪接来产生编码全长蛋白的单种RNA的诱导型方法。在一些实施方案中,该方法包括如本文所述的核酶识别序列和核酶。在一些实施方案中,该方法包括施用如本文所述的连结酶或编码连结酶的核酸。

[0239] 体内

[0240] 在一个实施方案中,本发明包括产生编码目的蛋白的RNA分子的方法。在一些实施方案中,该方法包括向细胞或组织施用至少两种核酸分子。在一个实施方案中,该至少两种核酸分子包括第一RNA分子和第二RNA分子。在一些实施方案中,该至少两种核酸分子编码第一RNA分子和第二RNA分子。

[0241] 在一个实施方案中,所述第一RNA分子包含编码目的蛋白的第一部分的编码区。在一个实施方案中,所述第一RNA分子包含3'核酶。在一个实施方案中,所述第一RNA分子包含编码目的蛋白的第一部分的编码区和3'核酶。在一个实施方案中,所述3'核酶催化自身脱离第一RNA分子,从而产生3'P或2'3'cP末端。在一个实施方案中,所述3'核酶是HDV核酶家族的成员。

[0242] 在一个实施方案中,所述第二RNA分子包含编码目的蛋白的第二部分的编码区。在一个实施方案中,所述第二RNA分子包含5'核酶。在一个实施方案中,所述第二RNA分子包含编码目的蛋白的第二部分的编码区和5'核酶。在一个实施方案中,所述5'核酶催化自身脱离第二RNA分子,从而产生5'OH末端。在一个实施方案中,所述5'核酶是HH核酶家族的成员。

[0243] 在一个实施方案中,所述3'P或2'3'cP末端连结到5'OH末端从而形成包含第一RNA分子的编码区和第二RNA分子的编码区的RNA分子。

[0244] 在一个实施方案中,所述方法包括向细胞或组织施用一种或多种编码一种或多种附加RNA分子的附加核酸分子,每种附加RNA分子均包含编码目的蛋白的结构域的编码区、5'核酶,和3'核酶。

[0245] 在一个实施方案中,所述方法包括向细胞或组织施用一种或多种编码一种或多种附加RNA分子的附加核酸分子,每种附加RNA分子均包含编码目的蛋白的结构域的编码区、5'核酶,和3'核酶识别序列。在一个实施方案中,该3'核酶识别序列包含VS-S。在一个实施方案中,该核酶是VS。

[0246] 在一个实施方案中,该方法包括向细胞或组织施用选自由以下组成的组的一种或多种:编码连结酶的核酸分子和连结酶。在一个实施方案中,该连结酶诱导由第一RNA分子的编码区和第二RNA分子的编码区组装RNA分子。在一个实施方案中,连结酶是RNA 2',3'-环状磷酸酯和5'-OH(RtcB)连结酶。

[0247] 在一些实施方案中,该方法包括向细胞或组织施用至少一种AAV载体,该AAV载体编码第一RNA分子和第二RNA分子,该第一RNA分子包含编码目的蛋白的第一部分的蛋白编码区和3'核酶,该第二RNA分子包含编码目的蛋白的第二部分的蛋白编码区和5'核酶。在一些实施方案中,该方法包括施用如本文所述的连结酶或编码连结酶的核酸。

[0248] 在一些实施方案中,该方法包括施用至少两种AAV载体,包括第一AAV载体和第二AAV载体。在一个实施方案中,第一AAV载体编码第一RNA分子,该第一RNA分子包含编码目的蛋白的第一部分的蛋白编码区和3'核酶。在一个实施方案中,第二AAV载体向细胞或组织编码第二RNA分子,该第二RNA分子包含编码目的蛋白的第二部分的蛋白编码区和5'核酶。在一些实施方案中,该方法包括施用如本文所述的连结酶或编码连结酶的核酸。

[0249] 在一些实施方案中,该方法包括向细胞或组织施用至少一种慢病毒载体,该慢病毒载体编码第一RNA分子和第二RNA分子,该第一RNA分子包含编码目的蛋白的第一部分的蛋白编码区和3'核酶,该第二RNA分子包含编码目的蛋白的第二部分的蛋白编码区和5'核酶。在一些实施方案中,该方法包括施用如本文所述的连结酶或编码连结酶的核酸。

[0250] 在一些实施方案中,该方法包括施用至少两种慢病毒载体,包括第一慢病毒载体和第二慢病毒载体。在一个实施方案中,第一慢病毒载体编码第一RNA分子,该第一RNA分子包含编码目的蛋白的第一部分的蛋白编码区和3'核酶。在一个实施方案中,第二慢病毒载体向细胞或组织编码第二RNA分子,该第二RNA分子包含编码目的蛋白的第二部分的蛋白编

码区和5'核酶。在一些实施方案中,该方法包括施用如本文所述的连结酶或编码连结酶的核酸。

[0251] 在一些实施方案中,该方法包括向细胞或组织施用至少一种慢病毒载体递送系统以提供第一RNA分子和第二RNA分子,该第一RNA分子包含编码目的蛋白的第一部分的蛋白编码区和3'核酶,该第二RNA分子包含编码目的蛋白的第二部分的蛋白编码区和5'核酶。在一些实施方案中,该方法包括施用如本文所述的连结酶或编码连结酶的核酸。

[0252] 在一些实施方案中,该方法包括施用至少两种慢病毒载体递送系统,包括第一慢病毒载体递送系统和第二慢病毒载体递送系统。在一个实施方案中,第一慢病毒载体递送系统提供第一RNA分子,该第一RNA分子包含编码目的蛋白的第一部分的蛋白编码区和3'核酶。在一个实施方案中,第二慢病毒载体递送系统向细胞或组织提供第二RNA分子,该第二RNA分子包含编码目的蛋白的第二部分的蛋白编码区和5'核酶。在一些实施方案中,该方法包括施用如本文所述的连结酶或编码连结酶的核酸。

[0253] 在一些实施方案中,该方法包括施用两种或更多种选自由以下组成的组的递送媒介物:AAV载体、慢病毒载体、慢病毒载体递送系统或它们的组合。在一个实施方案中,所述两种或更多种递送媒介物包括第一递送媒介物和第二递送媒介物。在一个实施方案中,第一递送媒介物提供第一RNA分子,该第一RNA分子包含编码目的蛋白的第一部分的蛋白编码区和3'核酶。在一个实施方案中,第二递送媒介物向细胞或组织提供第二RNA分子,该第二RNA分子包含编码目的蛋白的第二部分的蛋白编码区和5'核酶。在一些实施方案中,该方法包括施用如本文所述的连结酶或编码连结酶的核酸。

[0254] 将基因引入到细胞中以及使基因表达达到细胞中的方法是本领域已知的。在表达载体的情况下,可以容易地通过本领域中的任何方法将该载体引入到宿主细胞,例如哺乳动物、细菌、酵母或昆虫细胞中。例如,表达载体可以通过物理、化学或生物学方式转移到宿主细胞中。

[0255] 用于将多核苷酸引入到宿主细胞中的物理方法包括磷酸钙沉淀、脂质转染、粒子轰击、显微注射、电穿孔等。用于产生包含载体和/或外源核酸的细胞的方法在本领域中是众所周知的。参见,例如,Sambrook等人(2012,Molecular Cloning:A Laboratory Manual,Cold Spring Harbor Laboratory,New York)。用于将多核苷酸引入到宿主细胞中的示例性方法是磷酸钙转染。

[0256] 用于将目的多核苷酸引入到宿主细胞中的生物学方法包括使用DNA和RNA载体。病毒载体,并且尤其是逆转录病毒载体,已成为用于将基因插入到哺乳动物例如人细胞中最广泛使用的方法。其他病毒载体可以来源于慢病毒、痘病毒、单纯疱疹病毒I、腺病毒和腺相关病毒等。参见,例如,美国专利号5,350,674和5,585,362。

[0257] 用于将多核苷酸引入到宿主细胞中的化学方式包括胶体分散系统,诸如大分子复合物、纳米胶囊、微球、珠粒和基于脂质的系统,包括水包油乳剂、胶束、混合胶束和脂质体。用作体外和体内递送媒介物的示例性胶体系统是脂质体(例如,人造膜囊泡)。

[0258] 在使用非病毒递送系统的情况下,示例性递送媒介物是脂质体。考虑了使用脂质制剂将所述核酸引入到宿主细胞(体外、离体或体内)中。在另一个方面,所述核酸可以与脂质缔合。与脂质缔合的核酸可以包封在脂质体的水性内部中、散布在脂质体的脂质双层中、通过与脂质体和寡核苷酸缔合的连接分子附接至脂质体、包埋在脂质体中、与脂质体复合、

分散在含有脂质的溶液中、与脂质混合、与脂质组合、作为悬浮液包含在脂质中、与胶束一起被包含或复合,或以其他方式与脂质缔合。与脂质、脂质/DNA或者脂质/表达载体缔合的组合物不限于溶液中的任何特定结构。例如,它们可能以胶束形式存在于双层结构中,或以“塌陷”结构存在。它们也可以简单地散布在溶液中,从而可能形成大小或形状不均匀的聚集体。脂质是脂肪性物质,其可以是天然存在的脂质或合成脂质。例如,脂质包括天然存在于细胞质中的脂肪滴以及含有长链脂肪烃及其衍生物的一类化合物,诸如脂肪酸、醇、胺、氨基醇和醛。

[0259] 适用的脂质可以从商业来源获得。例如,二肉豆蔻基磷脂酰胆碱(“DMPC”)可以从Sigma, St. Louis, MO获得;磷酸二十六烷基酯(“DCP”)可以从K&K Laboratories (Plainview, NY)获得;胆固醇(“Choi”)可以从Calbiochem-Behring获得;二肉豆蔻基磷脂酰甘油(“DMPG”)和其他脂质可以从Avanti Polar Lipids, Inc. (Birmingham, AL)获得。脂质在氯仿或氯仿/甲醇中的储备溶液可储存在约-20℃。氯仿被用作唯一的溶剂,因为它比甲醇更容易蒸发。“脂质体”是一个通用术语,其涵盖通过产生封闭的脂质双层或聚集体形成的各种单层和多层脂质媒介物。脂质体可以被表征为具有含有磷脂双层膜和内部水性介质的囊泡结构。多层脂质体具有由水性介质隔开的多个脂质层。它们在磷脂悬浮在过量的水溶液中时会自发形成。脂质组分在形成封闭结构之前经历自我重排,并在脂质双层之间夹带水和溶解的溶质(Ghosh等人, 1991 *Glycobiology* 5:505-10)。然而,也涵盖在溶液中具有不同于正常囊泡结构的结构的组合物。例如,脂质可能呈现胶束结构或仅作为脂质分子的不均匀聚集体存在。还考虑了lipofectamine-核酸复合物。

[0260] 不管用于将外源核酸引入到宿主细胞中的方法如何,为了确认宿主细胞中重组DNA序列的存在,可以执行多种多样的测定。此类测定包括例如本领域技术人员熟知的“分子生物学”测定,诸如Southern和Northern印迹、RT-PCR和PCR;“生化”测定,诸如检测特定肽的存在或不存在,例如通过免疫学方式(ELISA和蛋白质印迹)或通过本文所述的测定来鉴定落入本发明范围内的剂。

[0261] 在一个实施方案中,本发明涉及一种从编码目的蛋白的部分的两对或更多对独立RNA分子通过核酶的顺式切割以及该成对独立RNA分子的反式剪接来表达两种或更多种目的蛋白的方法。在一个实施方案中,该方法包括施用一对、两对或三对编码或包含RNA分子的核酸分子,其中每个单个对的独立RNA分子具有单独的阅读框,使得不需要的对的反式剪接不会导致全长功能蛋白的翻译。在一个实施方案中,该方法进一步包括向细胞或组织施用选自以下组成的组的一种或多种:编码连结酶的核酸分子和连结酶。在一个实施方案中,所述连结酶是如本文所述的RNA 2', 3'-环状磷酸酯和5'-OH(RtcB)连结酶。

[0262] 在一个实施方案中,本发明包括递送和表达全长目的蛋白和货物序列的方法。在一个实施方案中,所述方法包括向细胞或组织施用编码目的蛋白的第一部分的RNA的第一部分,该RNA的第一部分以其3'末端连接到合成内含子;以及编码目的蛋白的第二部分的RNA的第二部分,该RNA的第二部分以其5'末端连接到合成内含子。在一个实施方案中,所述合成内含子在任一侧侧接有5'核酶序列和3'核酶序列。在一个实施方案中,所述合成内含子包含被置于所述5'核酶序列和3'核酶序列之间的货物序列。在一个实施方案中,5'核酶序列和3'核酶序列的自切割产生以下三个单独的RNA分子:1) 第一片段,包含编码目的蛋白的第一部分的RNA的第一部分;2) 第二片段,包含合成内含子;3) 第三片段,包含编码目的蛋

白的第二部分的RNA的第二部分。在一个实施方案中,第二片段的相容末端被连结从而产生包含含有货物序列的合成内含子的环状RNA分子。在实施方案中,第一片段和第三片段被连结在一起从而产生单个全长线性RNA分子。在一个实施方案中,全长目的蛋白包括治疗性蛋白、报告蛋白、重组酶、抗生素抗性基因产物、抗体或Cas9蛋白。在一个实施方案中,货物序列包含治疗性核酸序列(例如,miRNA序列或CRISPR指导RNA序列)或者编码治疗性蛋白。在一些实施方案中,全长目的蛋白包含Cas9并且货物序列包含指导RNA序列,从而使Cas9靶向特定基因组序列以进行编辑。在一些实施方案中,所述方法包括向细胞或组织施用如本文所述的连结酶或编码连结酶的核酸。

[0263] 在一个实施方案中,本发明包括一种基因编辑方法,该方法包括一种或多种反式切割工程化核酶。在一些实施方案中,所述方法包括施用第一反式切割工程化核酶和第二反式切割工程化核酶,其中该第一反式切割工程化核酶靶向致病突变的上游并且该第二反式切割工程化核酶靶向致病突变的下游。在一些实施方案中,在致病突变的上游和下游的反式切割导致该致病突变的去除。在一些实施方案中,在该致病突变的反式切割后,该基因的剩余部分被反式剪接在一起。在一些实施方案中,经反式剪接的基因表达为功能性蛋白。

[0264] 在一个实施方案中,本发明涉及组装全长RNA病毒基因组的体内方法。示例性的RNA病毒包括但不限于:冠状病毒、副粘病毒、正粘病毒、逆转录病毒、慢病毒、甲病毒、黄病毒、弹状病毒、麻疹病毒、新城病病毒和小核糖核酸病毒。在一个实施方案中,该方法发包括向细胞或组织施用编码RNA病毒基因组的第一部分并且编码3'核酶的第一核酸。在一个实施方案中,所述方法包括向细胞或组织施用编码RNA病毒基因组的第二部分并且编码5'核酶的第二核酸。在一个实施方案中,所述方法包括向细胞或组织施用包含RNA病毒基因组的第一部分和3'核酶的第一RNA分子。在一个实施方案中,所述方法包括向细胞或组织施用包含RNA病毒基因组的第二部分和5'核酶的第二RNA分子。在一个实施方案中,所述方法包括向细胞或组织施用编码如本文所述的连结酶的核酸或连结酶。在一个实施方案中,在3'和5'核酶顺式切割后,RNA病毒基因组的第一部分和RNA病毒基因组的第二部分被连结在一起,从而产生全长RNA病毒基因组。

[0265] 体外

[0266] 在一个实施方案中,本发明包括一种产生编码目的蛋白的RNA分子的体外方法。在一个实施方案中,所述方法包括提供至少两种RNA分子的步骤。在一个实施方案中,所述步骤包括提供第一RNA分子和第二RNA分子。

[0267] 在一个实施方案中,所述第一RNA分子包含编码目的蛋白的第一部分的编码区。在一个实施方案中,所述第一RNA分子包含3'核酶。在一个实施方案中,所述第一RNA分子包含编码目的蛋白的第一部分的编码区和3'核酶。

[0268] 在一个实施方案中,所述第二RNA分子包含编码目的蛋白的第二部分的编码区。在一个实施方案中,所述第二RNA分子包含5'核酶。在一个实施方案中,所述第二RNA分子包含编码目的蛋白的第二部分的编码区和5'核酶。

[0269] 在一个实施方案中,产生编码目的蛋白的RNA分子的体外方法进一步包括提供连结酶。在一个实施方案中,该连结酶诱导由第一RNA分子的编码区和第二RNA分子的编码区组装RNA分子。在一个实施方案中,所述连结酶是如本文所述的RNA 2',3'-环状磷酸酯和5'-OH(RtcB)连结酶。

[0270] 在一个实施方案中,本发明包括一种产生编码目的多结构域蛋白的RNA分子的体外方法。在一个实施方案中,所述方法包括以下步骤:a)提供第一RNA分子,b)提供一种或多种附加RNA分子,c)提供核酶,和d)提供最后一种RNA分子。

[0271] 在一个实施方案中,步骤a)的所述第一RNA分子包含编码目的蛋白的第一部分的编码区。在一个实施方案中,所述第一RNA分子包含3'核酶。在一个实施方案中,所述第一RNA分子包含编码目的蛋白的第一部分的编码区和3'核酶。在一个实施方案中,3'核酶催化自身脱离第一RNA分子,从而产生3'P或2'3'cP末端。在一个实施方案中,所述第一RNA分子进一步包含5'标签。在一个实施方案中,所述5'标签介导所述第一RNA分子与固体支持物的附接。

[0272] 在一个实施方案中,步骤b)的所述一种或多种附加RNA分子包含编码目的蛋白的结构域的编码区、5'核酶,和3'核酶识别序列。在一个实施方案中,所述5'核酶切割自身从而产生5'OH末端。在一个实施方案中,提供连结酶以催化所述第一RNA分子与所述一种或多种附加RNA分子的连结。在一个实施方案中,所述连结酶是如本文所述的RNA 2',3'-环状磷酸酯和5'-OH (RtcB) 连结酶。在一个实施方案中,所述3'核酶识别序列包含如本文所述的VS-S序列。

[0273] 在一个实施方案中,步骤c)的所述核酶包含如本文所述的VS-Rz。在一个实施方案中,所述VS-Rz识别VS-S,并且介导VS-S从所述一种或多种附加RNA分子上的切割。在一个实施方案中,所述切割产生3'P或2'3'cP末端。在一个实施方案中,步骤b)至c)被重复至少一次以产生编码多个结构域的RNA分子。在一个实施方案中,在重复步骤b)之前去除所述VS-Rz。

[0274] 在一个实施方案中,步骤d)的所述最后一种RNA分子包含编码目的蛋白的最后一个部分的编码区。在一个实施方案中,所述最后一种RNA分子包含5'核酶。在一个实施方案中,所述最后一种RNA分子包含编码目的蛋白的最后一个部分的编码区和5'核酶。在一个实施方案中,所述5'核酶催化自身脱离最后一种RNA分子,从而产生5'OH末端。在一个实施方案中,提供连结酶以催化所述一种或多种附加RNA分子与所述最后一种RNA分子的连结,从而产生编码N-端结构域、一个或多个附加结构域和C-端结构域的完整RNA分子。在一个实施方案中,所述连结酶是如本文所述的RNA 2',3'-环状磷酸酯和5'-OH (RtcB) 连结酶。

[0275] 本公开的任何RNA分子都可以从模板DNA(称为“体外转录模板”)体外转录。该DNA的来源可以是例如基因组DNA、质粒DNA、噬菌体DNA、cDNA、合成DNA序列或任何其他合适的DNA来源。在一些实施方案中,体外转录模板编码5'非翻译(UTR)区,含有开放阅读框,并且编码3'UTR和polyA尾。体外转录模板的特定核酸序列组成和长度将取决于由模板编码的mRNA。

[0276] 在一个实施方案中,5'UTR的长度在0至3000个核苷酸之间。可通过不同方法改变要添加到编码区的5'和3'UTR序列的长度,该方法包括但不限于设计与UTR不同区域退火的PCR引物。使用这种方法,本领域的普通技术人员可以在转染转录的RNA后修饰实现最佳翻译效率所需的5'和3'UTR长度。

[0277] 该5'和3'UTR可以是目的基因的天然存在的内源性5'和3'UTR。可替代地,可以通过将UTR序列掺入到正向和反向引物中或通过模板的任何其他修饰来添加对于目的基因不是内源的UTR序列。使用对于目的基因不是内源的UTR序列可用于修饰该RNA的稳定性和/或

翻译效率。例如,已知3'UTR序列中富含AU的元素会降低mRNA的稳定性。因此,可以根据本领域众所周知的UTR的特性来选择或设计3'UTR以增加转录的RNA的稳定性。

[0278] 在一个实施方案中,该5'UTR可以含有内源性基因的Kozak序列。可替代地,当如上所述通过PCR添加对于目的基因不是内源性的5'UTR时,可以通过添加5'UTR序列来重新设计共有Kozak序列。Kozak序列可以提高某些RNA转录本的翻译效率,但似乎并非所有RNA都需要实现高效翻译。许多mRNA对Kozak序列的要求是本领域已知的。在其他实施方案中,5'UTR可以来源于其中的RNA基因组在细胞中是稳定的RNA病毒。在其他实施方案中,可以在3'或5'UTR中使用各种核苷酸类似物以阻止mRNA的外切核酸酶降解。

[0279] 为了能够从DNA模板合成RNA,转录启动子应附接到待转录序列上游的DNA模板。当起RNA聚合酶的启动子作用的序列被添加到正向引物的5'末端时,RNA聚合酶启动子在要转录的开放阅读框上游被掺入到PCR产物中。在一个实施方案中,该启动子是如本文别处所述的T7 RNA聚合酶启动子。其他有用的启动子包括但不限于T3和SP6 RNA聚合酶启动子。T7、T3和SP6启动子的共有核苷酸序列是本领域已知的。

[0280] 在一个实施方案中,该mRNA具有位于5'末端的帽和3'poly(A)尾,其决定细胞中的核糖体结合、翻译启动和mRNA稳定性。在环状DNA模板(例如质粒DNA)上,RNA聚合酶产生长的串联产物,其不适合在真核细胞中表达。在3'UTR末端处被线性化的质粒DNA的转录产生正常大小的mRNA,当它在转录后被多腺苷酸化时,它在真核转染中是有效的。

[0281] 在线性DNA模板上,噬菌体T7 RNA聚合酶可以将转录本的3'端延伸到该模板的最后一个碱基之外(Schenborn和Mierendorf,Nuc Acids Res.,13:6223-36(1985);Nacheva和Berzal-Herranz,Eur.J.Biochem.,270:1485-65(2003))。

[0282] 将polyA/T序列段(stretch)整合到DNA模板中的常规方法是分子克隆。然而,整合到质粒DNA中的polyA/T序列会导致质粒不稳定性,这可以通过使用重组无能细菌细胞进行质粒繁殖来改善。

[0283] 在体外转录后,可以使用poly(A)聚合酶(例如大肠杆菌polyA聚合酶(E-PAP)或酵母polyA聚合酶)进一步延长RNA的poly(A)尾。在一个实施方案中,将poly(A)尾的长度从100个核苷酸增加到300至400个核苷酸导致RNA的翻译效率增加约两倍。此外,将不同化学基团附接到3'末端可以增加mRNA稳定性。此种附接物(attachment)可以含有经修饰的核苷酸/人工核苷酸、适体和其他化合物。例如,可以使用poly(A)聚合酶将ATP类似物掺入到poly(A)尾中。ATP类似物可以进一步增加RNA的稳定性。

[0284] 5'帽还为mRNA分子提供稳定性。在一个实施方案中,通过该方法产生的RNA包括5'帽1结构。此种帽1结构可以使用牛痘加帽酶和2'-O-甲基转移酶(CellScript, Madison, WI)来产生。可替代地,使用本领域已知并且本文中描述的技术提供5'帽(Cougot等人,Trends in Biochem.Sci.,29:436-444(2001);Stepinski等人,RNA,7:1468-95(2001);Elango等人,Biochim.Biophys.Res.Comm.,330:958-966(2005))。

[0285] 本发明的某些实施方案可以使用由惰性衬底或基质(例如载玻片、聚合物珠粒等)构成的固体支持物,该惰性衬底或基质已例如通过施加允许共价附接到生物分子(诸如多核苷酸)的包含反应性基团的中间材料的层或涂层来官能化。此类支持物的实例包括但不限于支撑在惰性衬底(诸如玻璃)上的聚丙烯酰胺水凝胶,特别是如WO 2005/065814和US 2008/0280773(其内容通过引用整体并入本文)中所描述的聚丙烯酰胺水凝胶。在此类实施

方案中,生物分子(例如多核苷酸)可以直接共价附接到中间材料(例如水凝胶),但是中间材料本身可以非共价地附接到衬底或基质(例如玻璃衬底)上。术语“共价附接到固体支持物”应相应解释为包括这种类型的布置。

[0286] 如本领域技术人员将理解的,可能的衬底的数量非常大。可能的衬底包括但不限于玻璃和经修饰或功能化的玻璃、塑料(包括丙烯酸树脂、聚苯乙烯和苯乙烯与其他材料的共聚物、聚丙烯、聚乙烯、聚丁烯、聚氨酯、Teflon™等)、多糖、尼龙或硝酸纤维素、陶瓷、树脂、二氧化硅或二氧化硅基材料(包括硅和经修饰硅)、碳、金属、无机玻璃、塑料、光纤束和多种多样的其他聚合物。

[0287] 在一些实施方案中,固体支持物包括微球或珠粒。合适的珠粒组合物包括但不限于塑料、陶瓷、玻璃、聚苯乙烯、甲基苯乙烯、丙烯酸聚合物、顺磁性材料、氧化钍溶胶(thoria sol)、碳石墨、二氧化钛、胶乳或交联葡聚糖(如琼脂糖)、纤维素、尼龙、交联胶束和聚四氟乙烯,以及本文概述的用于固体支持物的任何其他材料都可以使用。来自Bangs Laboratories, Fishers Ind.的“微球检测指南(Microsphere Detection Guide)”是一个有用的指南。在某些实施方案中,微球是磁性微球或珠粒。

[0288] 珠粒不必是球形的;可以使用不规则颗粒。可替代地或另外地,珠粒可以是多孔的。珠粒尺寸范围从纳米,即100nm到毫米,即1mm,优选约0.2微米至约200微米的珠粒,特别优选约0.5微米至约5微米的珠粒,但是在一些实施方案中可使用更小或更大的珠粒。

[0289] 在一个实施方案中,本发明涉及组装全长RNA病毒基因组的体外方法。示例性的RNA病毒包括但不限于:冠状病毒、副粘病毒、正粘病毒、逆转录病毒、慢病毒、甲病毒、黄病毒、弹状病毒、麻疹病毒、新城病病毒和小核糖核酸病毒。在一个实施方案中,所述方法包括提供包含RNA病毒基因组的第一部分和3'核酶的第一RNA分子。在一个实施方案中,所述方法包括提供包含RNA病毒基因组的第二部分和5'核酶的第二RNA分子。在一个实施方案中,在3'和5'核酶的顺式切割后,如本文所述,RNA病毒基因组的第一部分和RNA病毒基因组的第二部分具有用于连结的相容端。在一个实施方案中,所述方法包括使第一RNA分子和第二RNA分子与如本文所述的连结酶接触,从而产生全长RNA病毒基因组。

[0290] 治疗和用途

[0291] 本发明提供了治疗受试者的疾病或病症、减轻受试者的疾病或病症的症状,以及/或者降低受试者发展疾病或病症的风险的方法。例如,在一个实施方案中,本发明的方法治疗哺乳动物的疾病或病症、减轻哺乳动物的疾病或病症的症状,以及/或者降低哺乳动物发展疾病或病症的风险。在一个实施方案中,本发明的方法治疗植物的疾病或病症、减轻植物的疾病或病症的症状,以及/或者降低植物发展疾病或病症的风险。在一个实施方案中,本发明的方法治疗酵母生物的疾病或病症、减轻酵母生物的疾病或病症的症状,以及/或者降低酵母生物发展疾病或病症的风险。

[0292] 在一个实施方案中,受试者是细胞。在一些实施方案中,该细胞为原核细胞或真核细胞。在一个实施方案中,该细胞是真核细胞。在一个实施方案中,该细胞是植物、动物或真菌细胞。在一些实施方案中,该细胞是植物细胞。在一个实施方案中,该细胞是动物细胞。在一些实施方案中,该细胞是酵母细胞。

[0293] 在一个实施方案中,该受试者是哺乳动物。例如,在一个实施方案中,该受试者是人、非人灵长类动物、狗、猫、马、牛、山羊、绵羊、兔、猪、大鼠或小鼠。在一个实施方案中,该

受试者是非哺乳动物类受试者。例如,在一个实施方案中,该受试者是斑马鱼、果蝇或蛔虫。

[0294] 在一个实施方案中,该疾病或病症由缺失或有缺陷的蛋白引起,该蛋白的核酸序列超过病毒载体的包装尺寸。因此,在一个实施方案中,可以使用本发明的组合物、系统和方法来治疗、减轻该疾病或病症,或者降低该疾病或病症的风险。因此,在一个实施方案中,所述方法包括向受试者施用一种或多种本发明的组合物。进一步地,在一个实施方案中,所述方法包括利用一种或多种本发明系统来治疗受试者的疾病或病症、减轻受试者的疾病或病症的症状,以及/或者降低受试者发展疾病或病症的风险。

[0295] 在一个实施方案中,该疾病或病症是选自由以下组成的组的一种或多种:杜氏肌营养不良症;常染色体隐性多囊肾病;血友病A;Stargardt黄斑变性;肢带型肌营养不良;DFNB9;神经感觉非综合征性隐性耳聋;囊性纤维化;威尔逊病;常染色体隐性三好肌营养不良和耳聋9型;尤塞氏综合征I型和常染色体隐性耳聋2型;常染色体隐性耳聋3型和非综合征性听力损失;尤塞氏综合征I型;常染色体隐性耳聋16 (DFNB16);梅尼埃病(MD);常染色体显性耳聋12型和常染色体隐性耳聋21型;尤塞氏综合征1F型(USH1F)和DFNB23;常染色体隐性耳聋28型和非综合征性听力损失;常染色体隐性耳聋30型和非综合征性听力损失;常染色体隐性耳脊椎肥大骨骺发育不良和常染色体显性耳椎肥大骨骺发育不良;常染色体隐性耳聋77型和常染色体隐性非综合征性感觉神经性耳聋Dfnb型;常染色体隐性非综合征性听力障碍DFNB84;常染色体隐性耳聋84B型和罕见性遗传性耳聋;常染色体显性周围神经病变、肌病、声音嘶哑和听力损失和耳聋4A型;先天性血小板减少症;感觉性听力损失;DFNA56;HXB;常染色体显性耳聋56;六臂蛋白;癫痫性脑病;Timothy综合征和长Qt综合征8型;X-连锁视网膜病症;醛固酮增多症;脊髓小脑性共济失调42型;原发性醛固酮增多症;癫痫发作和神经系统异常和窦房结功能障碍和耳聋;神经发育病症;低钾性周期性麻痹;癫痫;发育性和癫痫性脑病;布罗迪肌病;Darier氏病/心脏病;血管性血友病和齐薇格综合征。在一个实施方案中,该疾病或病症是由适合CRISPR-Cas9介导的编辑的基因突变引起的任何疾病或病症。

[0296] 在一个实施方案中,本发明的方法包括向患有杜氏肌营养不良症的受试者施用组合物,该组合物包含第一核酸和第二核酸,该第一核酸包含编码抗肌萎缩蛋白的第一部分的编码区和3'核酶,该第二核酸包含编码抗肌萎缩蛋白的第二部分的编码区和5'核酶,其中第一核酸转录第一RNA分子并且第二核酸转录第二RNA分子,并且其中3'和5'核酶的顺式切割以及编码抗肌萎缩蛋白的第一部分的编码区和编码抗肌萎缩蛋白的第二部分的编码区的反式剪接产生编码全长抗肌萎缩蛋白的单种RNA分子。

[0297] 在一个实施方案中,本发明的方法包括向患有杜氏肌营养不良症的受试者施用施用组合物,该组合物包含第一核酸和第二核酸,该第一核酸编码SEQ ID NO:129的核酸序列,该第二核酸编码SEQ ID NO:130的核酸序列,其中该第一核酸转录第一RNA分子并且该第二核酸转录第二RNA分子,并且其中3'和5'核酶的顺式切割以及第一RNA分子和第二RNA的反式剪接产生编码全长肌抗肌萎缩蛋白的单个RNA分子。

[0298] 在一个实施方案中,本发明的方法包括向患有杜氏肌营养不良症的受试者施用施用组合物,该组合物包含第一核酸和第二核酸,该第一核酸编码SEQ ID NO:22的核酸序列,该第二核酸编码SEQ ID NO:23的核酸序列,其中该第一核酸转录第一RNA分子并且该第二核酸转录第二RNA分子,并且其中3'和5'核酶的顺式切割以及第一RNA分子和第二RNA的反

式剪接产生编码具有C-端GFP报告蛋白的全长肌抗肌萎缩蛋白的单个RNA分子。在一个实施方案中,第二核酸编码SEQ ID NO:23的片段,其中该片段不包括C-端GFP报告分子的编码序列。

[0299] 在一个实施方案中,所述方法包括向患有杜氏肌营养不良症的受试者施用组合物,该组合物包含第一RNA分子和第二RNA分子,该第一RNA分子编码肌抗肌萎缩蛋白的第一部分并且包含3'核酶,该第二RNA分子编码抗肌萎缩蛋白的第二部分并且包含5'核酶,其中所述3'和5'核酶的顺式切割以及所述第一和第二RNA分子的反式剪接产生编码全长肌抗肌萎缩蛋白的单种RNA分子。

[0300] 在一个实施方案中,所述方法包括向患有杜氏肌营养不良症的受试者施用组合物,该组合物包含第一RNA分子和第二RNA分子,该第一RNA分子包含SEQ ID NO:129的核酸序列,该第二RNA分子包含SEQ ID NO:130的核酸序列,其中3'和5'核酶的顺式切割以及第一和第二RNA分子的反式剪接产生编码全长抗肌萎缩蛋白的单种RNA分子。

[0301] 在一个实施方案中,所述方法包括向患有杜氏肌营养不良症的受试者施用组合物,该组合物包含第一RNA分子和第二RNA分子,该第一RNA分子包含SEQ ID NO:22的核酸序列,该第二RNA分子包含SEQ ID NO:23的核酸序列,其中3'和5'核酶的顺式切割以及第一和第二RNA分子的反式剪接产生编码具有C-端GFP报告蛋白的全长抗肌萎缩蛋白的单种RNA分子。在一个实施方案中,第二核酸编码SEQ ID NO:23的片段,其中该片段不包括C-端GFP报告分子的编码序列。

[0302] 在一个实施方案中,本发明的方法包括向患有选自表1的一种或多种疾病的受试者施用组合物,该组合物包含第一核酸和第二核酸,该第一核酸包含编码对应于表1中相关疾病的治疗性蛋白的第一部分的编码区和3'核酶,该第二核酸包含编码对应于表1中相关疾病的治疗性蛋白的第二部分的编码区和5'核酶,其中第一核酸转录第一RNA分子并且第二核酸转录第二RNA分子,并且其中3'和5'核酶的顺式切割以及编码治疗性蛋白的第一部分的编码区和编码肌治疗性蛋白的第二部分的编码区的反式剪接产生编码全长治疗性蛋白的单种RNA分子。

[0303] 在一个实施方案中,所述方法包括向患有一种或多种选自表1的疾病的受试者施用组合物,该组合物包含第一RNA分子和第二RNA分子,该第一RNA分子编码对应于表1中相关疾病的治疗性蛋白的第一部分并且包含3'核酶,该第二RNA分子编码对应于表1中相关疾病的治疗性蛋白的第二部分并且包含5'核酶,其中3'和5'核酶的顺式切割以及第一和第二RNA分子的反式剪接产生编码全长治疗性蛋白的单种RNA分子。

[0304] 表1.由大基因突变引起的单基因疾病列表(包括蛋白大小(氨基酸数目)、基因符号、蛋白名称和疾病名称)。

	蛋白大小	基因	治疗性蛋白	疾病
[0305]	3,685	DMD	抗肌萎缩蛋白	杜氏肌营养不良症
	4,074	PKHD1	纤维囊蛋白	常染色体隐性多囊肾病
	2,351	F8	凝血因子 VIII	血友病 A

	2,273	ABCA4	视网膜特异性磷脂转运 ATP 酶	Stargardt 黄斑变性
	2,080	DYSF	Dysferlin	肢带型肌营养不良
	1,997	OTOF	耳畸蛋白	DFNB9, 神经感觉非综合征性隐性耳聋
	1480	CFTR	囊性纤维化跨膜电导调节剂	囊性纤维化
	1,465	ATP7B	铜转运 ATP 酶 2	威尔逊病
	2,061	MYOF	Myoferlin	常染色体隐性三好肌营养不良和耳聋 9 型
	2,215	MYO7A	非常规肌球蛋白-VIIa	尤塞氏综合症 I 型和常染色体隐性耳聋 2 型
	3,530	MYO15A	非常规肌球蛋白-XV	常染色体隐性耳聋 3 型和非综合征性听力损失
	3,354	CDH23	钙粘蛋白-23	尤塞氏综合症 I 型
[0306]	1,809	STRC	Stereocilin	常染色体隐性耳聋 16 型(DFNB16)
	2,925	OTOG	Otogelin	梅尼埃病(MD)
	2,155	TECTA	α -覆膜蛋白 (Alpha-tectorin)	常染色体显性耳聋 12 型和常染色体隐性耳聋 21 型
	1,955	PCDH15	原钙粘蛋白 -15(Protocadherin-15)	尤塞氏综合症 1F 型 (USH1F)和 DFNB23
	2,365	TRIOBP	TRIO 和 F-肌动蛋白结合蛋白	常染色体隐性耳聋 28 型和非综合征性听力损失
	1,616	MYO3A	肌球蛋白-IIIa	常染色体隐性耳聋 30 型和非综合征性听力损失
	1,736	COL11A2	胶原蛋白 α -2(XI)链	常染色体隐性耳脊椎肥大骨骺发育不良和常染色体显性耳脊椎肥大骨骺发育不良
	2,067	LOXHD1	含脂氧合酶同源结构	常染色体隐性耳聋

		域的蛋白 1	77 型和常染色体隐性非综合征性感觉神经性耳聋 Dfnb 型
	2,332	PTPRQ 磷脂酰肌醇磷酸酶 PTPRQ	常染色体隐性非综合征性听力障碍 DFNB84
	2,332	OTOGL Otogelin 样蛋白	常染色体隐性耳聋 84B 型和罕见性遗传性耳聋
	1,995	MYH14 肌球蛋白-14	常染色体显性周围神经病变、肌病、声音嘶哑和听力损失和耳聋 4A 型
	1,960	MYH9 肌球蛋白-9	先天性血小板减少症、感觉性听力损失
	2,201	TNC 肌腱蛋白(Tenascin)	DFNA56、HXB、常染色体显性遗传耳聋 56 型、六臂蛋白
[0307]	2,506	CACNA1A 电压依赖性 P/Q-型钙通道亚基 α -1A	癫痫性脑病
	2,221	CACNA1C 电压依赖性 L-型钙通道亚基 α -1C	Timothy 综合征和长 Qt 综合征 8 型
	1,977	CACNA1F 电压依赖性 L-型钙通道亚基 α -1F	X-连锁视网膜病症
	2,353	CACNA1H 电压依赖性 T-型钙通道亚基 α -1H	醛固酮增多症
	2,377	CACNA1G 电压依赖性 T-型钙通道亚基 α -1G	脊髓小脑性共济失调 42 型
	2,161	CACNA1D 电压依赖性 L-型钙通道亚基 α -1D	原发性醛固酮增多症、癫痫发作和神经系统异常以及窦房结功能障碍和耳聋
	2,339	CACNA1B 电压依赖性 N-型钙通道亚基 α -1B	神经发育病症
	1,873	CACNA1S 电压依赖性 L-型钙通道亚基 α -1S	低钾性周期性麻痹
	2,223	CACNA1I 电压依赖性 T-型钙通道亚基 α -1I	癫痫

	2,313	CACNA1E	电压依赖性 R-型钙通道亚基 α -1E	发育性和癫痫性脑病
	1,001	ATP2A1	肌浆网钙/内质网钙 ATP 酶 1	布罗迪肌病
	1,042	ATP2A2	肌浆网钙/内质网钙 ATP 酶 2	Darier 氏病/心脏病
[0308]	2,813	VWF	血管性血友病因子 (von Willebrand factor)	血管性血友病
	1,283	PEX1	过氧化物酶体生物发生因子 1	齐薇格综合征
	4,069	CMYA5	心肌病相关蛋白 5 心肌病	

[0309] 实验实施例

[0310] 参照以下实验实施例进一步详细描述本发明。这些实施例仅仅是为了说明的目的而提供的,除非另有说明,否则这些实施例无意是限制性的。因此,本发明决不应被解释为限于以下实施例,而是应被解释为涵盖由于本文提供的教导而变得明显的任何和所有变化。

[0311] 无需进一步描述,据信本领域普通技术人员可以使用前面的描述和以下说明性实施例来实现和利用本发明并实践所要求保护的方法。因此,以下工作实施例不应被解读为以任何方式限制本公开的其余部分。

[0312] 实施例1:核酶介导的哺乳动物细胞中的RNA组装和表达

[0313] 核酶(Rz)是能够进行核苷酸特异性自切割的小型催化RNA序列(Doherty和Doudna 2000)。核酶介导的RNA切割产生独特的3'磷酸酯和5'-羟基端,它们类似于所有三个生命王国中普遍存在的RNA修复途径的底物。如本文所示,核酶介导的顺式切割可被用于哺乳动物细胞中独立RNA转录本的反式剪接,这种方法称为stitchR(stitch RNA)。值得注意的是,通过stitchR重构信使RNA允许在哺乳动物细胞中有效翻译和表达全长蛋白。正如所证明的, stitchR可被用于组合蛋白编码功能结构域或者被用于通过病毒载体递送和表达大蛋白编码序列。进一步地, RNA 2',3'-环状磷酸酯和5'-OH(RtcB)连结酶的过表达增强了哺乳动物细胞中的stitchR活性,并足以在体外催化stitchR活性。这些数据表征了一种利用核酶对细胞中功能性RNA进行无疤痕反式剪接的新方法,该方法可用于无数研究和治疗应用。

[0314] 自催化RNA序列在自然界中广泛存在并且催化多种多样的生物过程,包括内含子剪接、滚环病毒基因组复制和肽键形成(Weinberg等人,2019年)。已鉴定出具有不同序列和结构特征的至少七个主要核酶家族,包括锤头状(HH)核酶、丁型肝炎病毒(HDV)核酶、Varkud卫星(VS)核酶、Sister核酶、Twister-sister核酶、发夹状核酶、Hatchet核酶和Pistol核酶。研究最广泛的是HH、HDV和Twister核酶家族成员,它们由于小尺寸和切割特性,已在体外和体内用于产生没有核酶序列的具有精确端的RNA(图13)(Ferre-D'Amare和Doudna 1996;Avis等人2012;Zhang等人2017)。

[0315] 在原核生物和真核生物中,包括信使和长非编码RNA在内的大多数细胞RNA是用5'-磷酸酯(P)和3'-羟基(OH)端合成和剪接的。相比之下,许多tRNA和编码ER应激反应蛋白

XBP1的mRNA的非常规顺式剪接由酶途径催化,这产生独特的5'-OH以及3'-P或2'3'环状磷酸酯(cP)末端。最近的研究结果表明,哺乳动物中普遍存在的RNA2',3'-环状磷酸酯和5'-OH(RtcB)连结酶催化了非常规的RNA顺式剪接。此外,RtcB和其他几个酶家族可能起到修复已被应激或外源性核毒素损伤的宿主细胞RNA的作用。由于核酶介导的切割导致相似的终末端,因此核酶切割的RNA可能会经受通过内源性RNA修复途径进行的反式剪接。

[0316] 核酶切割的mRNA在哺乳动物细胞中被反式剪接和翻译

[0317] 为了确定核酶是否可用于哺乳动物细胞中RNA的无疤痕反式剪接,设计了两种表达质粒,该质粒含有荧光报告基因GFP的非重叠N-端(Nt)和C-端(Ct)片段(分别为Nt-GFP和Ct-GFP)。核酶被设计成催化它们自己从GFP片段的相邻核苷酸中去除,包括Nt-GFP上的3' HDV核酶和Ct-GFP上的5' HH核酶(图1A)。当被转染到哺乳动物COS-7或HEK293T细胞中时,单独的编码GFP核酶RNA的Nt或Ct的表达导致无可检测的GFP荧光(图1B)。值得注意的是,Nt和Ct-GFP编码的RNA的共表达在48小时后会产生产生绿色荧光(图1B)。RT-PCR分析和桑格测序显示,在预测的核酶催化切割位点之间发生了单独的Nt-和Ct-GFP RNA的反式剪接(图1C和图1D)。进一步地,通过蛋白质印迹在共转染细胞中检测到全长GFP蛋白(图1E)。这些数据表明,内源性哺乳动物细胞RNA修复途径足以催化独立的经核酶加工的RNA的反式剪接,这些RNA被有效地翻译成全长蛋白。这种RNA反式剪接方法被命名为stitchR。

[0318] 核酶序列和类型对核酶介导的反式剪接的影响

[0319] 为了精确量化细胞中由核酶介导的反式剪接产生的功能性全长蛋白的相对量,使用萤火虫荧光素酶的两个不重叠的半部产生报告基因(图2A)。与我们之前的研究结果一致,只有编码RNA的Nt-荧光素酶核酶和Ct-荧光素酶核酶的共转染才能导致细胞中产生反式剪接和荧光素酶活性(图2B和图2C)。使用该测定法,进一步表征了不同HH和HDV核酶序列对哺乳动物细胞中的反式剪接活性的影响。茎1HH核酶中的6个碱基对(bp)重叠提供了最大的荧光素酶活性和HH催化残基突变消除活性,这与先前关于体外表征的HH核酶活性的报道一致(图2D)。此外,基因组和反基因组HDV核酶序列的荧光素酶活性相当,除了最小的56核苷酸HDV核酶(HDV56),其活性显著降低(图2E)。同样与之前的报道一致,HDV催化所需的核苷酸的C到U突变导致荧光素酶活性完全丧失(图2E)。这些发现表明核酶介导的反式剪接活性依赖于哺乳动物细胞中的核酶切割。

[0320] 使用翻译控制器和/或蛋白降解序列阻止来自Nt或Ct载体的不需要或截短的蛋白表达

[0321] Nt或Ct RNA可能会在核酶介导的切割之前经受翻译,或者当被单独表达时,可能导致不需要的或截断的蛋白表达。为了限制未经剪接的Nt或Ct载体的表达,测试了先前表征的蛋白降解序列的翻译控制器对编码全长GFP的载体的稳定性的功效。在GFP的3'末端添加HDV核酶似乎不会改变GFP荧光(图3A和3B)。为了选择性地阻止GFP的表达,测试了蛋白降解序列hCL1-PEST、E1A-PEST,载体的poly(A)序列的去除,或为产生poly K尾而通过poly A尾进行的模拟翻译的影响(图3A和3B)。所有降解序列都与GFP开放阅读框内克隆,从而通过HDV核酶序列进行翻译。hCL1-PEST的纳入显示GFP荧光显著降低,而EF1a PEST则没有。从表达载体中删除载体poly(A)序列导致GFP表达完全丧失,而为产生poly K尾通过多poly A序列进行的翻译也导致荧光降低。

[0322] 对于Ct编码的GFP报告基因,5' HH核酶的纳入和GFP起始密码子(ATG)的删除仍然

导致弱但可检测的GFP表达,尽管缺乏预测的上游替代ATG(图3C和D)。GFP(GFPcdn)中N-端NTG密码子内的进一步沉默突变进一步降低了GFP检测,然而,弱荧光仍然明显。编码起到翻译抑制剂作用的4个小上游ORF的酵母GCN4基因的5'UTR的纳入消除了可检测的GFP荧光。仅编码该4个uORF的GCN45'UTR的较小内部片段在阻止GFP表达方面同样有效。这些数据表明,蛋白降解序列的翻译控制器可被用于阻止来自单个Nt或Ct载体的不需要的蛋白表达。

[0323] 这些翻译控制器或蛋白降解序列可被用于需要限制不需要或截短的蛋白表达的其他双载体应用,诸如依赖同源重组来产生大蛋白编码开放阅读框的双AAV载体策略。

[0324] 功能蛋白编码RNA的单一和多重反式剪接

[0325] 为了确定核酶介导的反式剪接是否可用于细胞中蛋白编码功能结构域的组合,产生了编码4个线粒体靶向序列(Nt-4xMTS)拷贝的RNA和编码缺乏ATG起始密码子的全长GFP(Ct-GFP)的开放阅读框(图4A)。这两种独立RNA的共表达导致定位于线粒体的GFP的强烈表达,其与红色荧光线粒体标记MitoTracker Red CMXRos重叠(图4B)。这些发现表明,核酶介导的反式剪接可用于快速地组合两种独立RNA,以在细胞中表达特定功能融合蛋白。

[0326] 由于蛋白在其中被翻译的三个开放阅读框,核酶介导的同时多种不同功能蛋白的反式剪接和表达也可能是可能的。通过利用这一特性,可以使用位于相容的三个不同开放阅读框中的RNA的反式剪接来产生功能性蛋白。为了证明这一功能性,在阅读框2(F2)中设计了附加核酶对,其编码肉豆蔻酰化膜靶向序列(Nt-F2-Myr)和红色荧光蛋白(Ct-F2-RFP)(图4C)。这些Nt和Ct载体对还包括hCL1-PEST蛋白降解序列和GCN4翻译抑制序列,以分别限制来自单个Nt和Ct载体的截短蛋白表达。在共转染细胞中,GFP荧光对线粒体具有高度特异性,而RFP荧光对膜具有高度特异性(图4D),证明了这种RNA反式剪接方法在细胞中产生不同功能蛋白的能力。

[0327] 优化的核酶增强核酶介导的反式剪接中的蛋白表达

[0328] 小的序列修饰可以通过改变二级结构、稳定性或与金属离子辅因子的结合来深刻影响核酶的催化活性。使用我们的反式剪接荧光素酶报告基因测定,我们鉴定出了改进的核酶类型和序列修饰,它们增强了哺乳动物细胞内的反式剪接荧光素酶报告基因活性(图16)。含有三级稳定基序(TSM)的RzB锤头状变体核酶显示出比没有TSM的核酶更高的活性(图16A)。进一步地,当被克隆到Nt-Luc的3'处时,Twister(twst)核酶显示出比HDV核酶更大的活性。twister核酶内的催化突变可以类似地消除荧光素酶活性(图16B)并且依赖于P1茎的形成(图16C)。由于Twister核酶在1位需要一个U,这一要求可能会将无疤痕反式剪接的设计限制为以U结束的序列。因此,我们测试了1位处的核苷酸取代是否可以被耐受,并且发现U1A没有显示出显著不同的活性,而U1C或U1G取代保留了活性,但是有所降低(图16C)。

[0329] 优化的剪接供体和受体序列增强了核酶介导的反式剪接中的蛋白表达

[0330] 由剪接体进行的前mRNA剪接已被证明可通过沉积促进首轮翻译的因子或者通过促进RNA加工和向细胞质的输出来增强mRNA翻译。在转基因内添加嵌合顺式剪接内含子也已被证明可促进转基因蛋白表达。然后研究了经反式剪接的RNA是否可以经受通过剪接体进行的顺式剪接,以及这是否会影响经反式剪接的mRNA的翻译和表达。为了测试这一点,将剪接供体(SD)和剪接受体(SA)序列掺入到反式剪接GFP报告基因中,使得经反式剪接的RNA将重构嵌合内含子(图5A)。值得注意的是,与不含SD或SA序列的反式剪接GFP报告基因相比,添加SD和SA序列可稳健地增强GFP荧光(图5B)。RT-PCR和桑格测序显示含有SD和SA序列

的Nt-GFP和Ct-GFP RNA都被反式剪接和顺式剪接,导致正常GFP开放阅读框的恢复(数据未显示)。这些数据表明反式剪接可能发生在细胞核中,并且随后的顺式剪接是增强由经反式剪接的RNA的表达的有用策略。

[0331] 核酶介导的用于使用病毒治疗载体进行递送的大基因序列的反式剪接和表达

[0332] 核酶介导的反式剪接可被用于递送和表达超过治疗性病毒基因治疗载体(如AAV)的包装尺寸限制的大蛋白编码mRNA(图6A)。这可能有助于恢复在许多人单基因疾病中发生突变的大基因(诸如杜氏肌营养不良症(DMD)中的抗肌萎缩蛋白(Dys)、囊性纤维化(CF)中的CFTR、血友病A中的因子VIII(F8)等)的表达。在基于细胞的转染测定中,编码Nt和Ct分裂的 μ 抗肌萎缩蛋白(μ Dystrophin)的载体与C-端GFP标签的共表达物在哺乳动物细胞中被反式剪接(图6B和图6C)并且被定位到膜上(图6D)。这些数据证明了使用核酶介导的反式剪接重构和表达大蛋白编码基因的可行性。

[0333] 用于在细胞中进行反式剪接的经核酶激活(ribozyme-enabled)的RNA的慢病毒递送

[0334] 核酶的自催化自切割可能会阻碍正义RNA病毒(诸如常用的 γ 逆转录病毒和慢病毒载体)对编码核酶的RNA的包装。为了规避这一潜在问题,使Nt和Ct分裂的GFP表达盒编码在第三代慢病毒载体骨架中的负义链上(图7A)。分别为Nt和Ct载体产生慢病毒颗粒,然后将其用于转导HEK293T细胞。用Nt-GFP和Ct-GFP二者转导的细胞显示出绿色荧光表达,而用单独的Nt-GFP或Ct-GFP转导的细胞未显示出可检测到的荧光(图7B)。这些数据表明,慢病毒载体能够递送和表达编码RNA的核酶以进行反式剪接。

[0335] 这种方法也可用于递送超过这些病毒载体的包装大小的大基因序列,诸如Dys(图7C)。核酶介导的反式剪接还可以允许安全操作或重构病毒基因组,诸如慢病毒或大型冠状病毒RNA基因组。

[0336] 使用病毒载体安全操作、递送和表达毒性基因或抗病毒基因

[0337] 核酶介导的反式剪接还可以允许安全操作或重构毒性或抗病毒蛋白,这些蛋白可能会抑制哺乳动物包装细胞中慢病毒颗粒的产生。这些包括许多细胞自杀基因,诸如翻译抑制性白喉毒素A(DTA)(图8A)。我们证明了编码分裂的DTA序列的载体,在反式剪接和表达后,抑制CS2GFP报告基因构建体的共表达,这与DTA在哺乳动物细胞中的翻译抑制作用一致(图8B)。

[0338] 用于增强或抑制核酶介导的反式剪接的酶

[0339] 许多酶家族被建议连结5'-OH和3'-P或2',3'环状磷酸酯(cP)末端,最值得注意的是RtcB,它在所有三个生命领域中都被发现是保守的。克隆并共表达来自真核生物(智人(*H.sapiens*))、细菌(大肠杆菌)和古生菌(霍里科什火球菌(*P.horikoshii*))物种的人密码子优化的RtcB直系同源物,以测量它们对反式剪接荧光素酶报告基因活性的影响。有趣的是,来自霍里科什火球菌的RtcB的共表达导致荧光素酶活性增强(4.5倍),而人和细菌直系同源物分别显示出适度的增强或没有显示出增强(图9)。

[0340] 其他酶家族已被证明可以调节这些RNA端。有趣的是,起到5'-羟基激酶和3'-磷酸酶和2',3'-环状磷酸二酯酶的作用的T4多核苷酸激酶(T4 PNK)的表达显著抑制荧光素酶活性(图9)。这些数据表明,外源酶的共表达可以增强或抑制哺乳动物细胞中核酶介导的反式剪接。

[0341] RtcB足以在体外催化核酶介导的RNA反式剪接

[0342] 由于它们的核苷酸特异性切割,核酶已在体外广泛用于产生精确的RNA末端。接下来试图确定核酶是否可用于体外独立合成的RNA的定向反式剪接。使用T7 RNA聚合酶对Nt-和Ct-萤光素酶-核酶报告基因构建体进行体外RNA转录,发现重组大肠杆菌RtcB的添加对于催化使用RT-PCR检测的反式剪接既必要又充分(图10A和图10B)。类似地,设计了编码蜘蛛丝蛋白Spidroin的结构域的RNA(图10C)。Spidroin是蜘蛛牵引丝的主要组分,这种材料因其拉伸特性而备受推崇,但由于该蛋白的高度重复性,很难在异源系统中合成。Spidroin天然地由多个侧接有保守N-端(N1L)和C-端(N3R)结构域的A和Q重复序列组成。在用T7聚合酶体外合成Spidroin RNA后,发现来自大肠杆菌的重组RtcB连结酶的添加足以催化编码经核酶切割的N1L和N3R的RNA的反式连结,如由RT-PCR和桑格测序所检测的(图10D)。

[0343] 编码多结构域蛋白的RNA的受控串联反式剪接

[0344] 接下来检查添加编码具有侧接核酶的A-Q融合结构域的第三RNA是否会导致串联重复组装,尽管不受控制(图11A)。虽然能够检测到每种单独RNA之间的定向反式剪接,但无法检测到三种或更多种独立RNA片段的组装(数据未示出)。这可能是由于含有与RtcB连结相容的端的RNA的快速环化。作为一种替代方法,反式活化的VS核酶的使用有可能允许体外RNA序列的顺序和受控组装(图11B和图11C)。在这种方法中,3'端RNA核酶仅适用于RtcB在添加和VS-Rz反式切割后的连结。由于VS-Rz反式活化核酶RNA不是共价附接的,因此逐步添加stitchR相容的RNA、VS-Rz和RtcB连结酶可以实现RNA序列的受控串联组装,这可能有助于组装编码生物学或工业上重要的蛋白(如合成蜘蛛丝、弹性蛋白、胶原蛋白等)的重复RNA。

[0345] 使用反式切割核酶对内源性RNA进行反式剪接——纠正致病突变的治疗应用

[0346] 核酶是自催化RNA,其以顺式切割以产生独特的RNA端,我们已经证明这些端被反式剪接,并且随后在哺乳动物细胞中表达(图12A)。值得注意的是,顺式切割核酶可以被工程化成以反式切割,使得靶RNA可以以核苷酸特异性方式被切割,从而产生相似的RNA端(图12B)(Carbonell等人2011;Webb和Luptak 2018)。因此,反式切割核酶可被用于在细胞内或体外催化RNA的无痕反式剪接。这种方法可用于无数应用,其中一个主要应用是通过靶向外显子或内含子序列中的突变侧接序列来删除基因转录本中的致病突变(图12C和图12D)。

[0347] 总之,本文证明了核酶介导的对细胞中表达的独立RNA的切割被有效组装并且能够在哺乳动物细胞中翻译。这种在本文中称为stitchR的方法能够用作一种对用于基础和临床应用的功能性RNA和蛋白进行组合组装的新方法。由于核酶的自催化性质和细胞中存在的内源性RNA修复途径,stitchR只需要表达单独的RNA就可以在细胞中发生反式剪接和翻译。在体外,已证明RtcB连结酶足以进行反式剪接,并且由于RtcB在所有三个生命界中普遍存在且广泛表达,stitchR有可能成为在许多不同生物体中有用的方法。

[0348] 这种系统的稳健性依赖于核酶介导的RNA切割的高效和精确性质,该核酶介导的RNA切割产生可靠且精确的对于恢复蛋白编码开放阅读框架至关重要的核苷酸特异性末端。进一步地,使用完全催化自身的去除的核酶产生RNA的能力允许无疤痕组装,从而产生与其天然对应物基本上无法区分的RNA。

[0349] 虽然核酶切割已在体外被广泛研究,但体内核酶切割被了解得较少,并且被认为受通过与RNA结合蛋白的相互作用进行的折叠和催化所需金属离子的可获得性的影响。

StitchR用作核酶介导的切割的间接读出,有趣的是,本文发现其受到核酶序列和结构变化的显著影响。这表明核酶裂解的优化可能是用于增强体内stitchR活性的有用方法。进一步分析诸如RtcB、RtcA和Archease的RNA修复途径组分的影响也可用作调节stitchR活性的重要因素。

[0350] 核酶已自然进化为以顺式发挥作用以促进其自我切割,然而,许多核酶家族(特别是HDV和HH)已被工程化为以反式切割靶RNA。本文值得注意的是,将反式切割核酶与stitchR组合可以进一步允许在细胞中或体外执行强大的RNA切割和修复方法。这种方法可以用作RNA的核苷酸特异性“剪切和粘贴”方法,该方法可能有助于产生RNA多样性或者有助于去除致病RNA中的某些有害突变。

[0351] 实施例2:可使用经反式活化的核酶诱导的RNA的反式剪接和表达

[0352] 大多数核酶是自催化的,并且只需要金属离子作为辅助因子,在生物环境中很容易找到,这有助于折叠和化学催化。如果供体RNA以G核苷酸结束,则Varkud卫星(VS)核酶可用于无痕反式剪接。有趣的是,VS核酶可以被修饰成允许该核酶的反式活化以诱导催化作用(Guo和Collins 1995;Ouellet等人2009)。当分裂成两个组分时,小VS茎环(VS-S)不足以单独诱导顺式切割,然而,剩余序列VS-Rz的添加促进了VS-S的有效切割(图14A)。这种反式活化特征可以允许诱导型核酶介导的反式切割,其中需要添加VS-Rz序列才能在Nt供体RNA上进行VS-S切割,该Nt供体RNA然后可适合于与含有5'-OH端的Ct受体RNA进行反式剪接(图14B)。含有典型5'-P-和3'-OH RNA端的VS-Rz序列不能参与反式剪接,因此可以作为反应的多周转催化剂。

[0353] 诸如通过所需的反式活化序列(诸如VS-Rz)添加来控制核酶介导的切割的能力可以允许控制可变或非可变RNA序列的添加以产生合成重复RNA(图14C)。一种方法是产生具有独特N-端结构域、独特C-端结构域和内部可变或非可变“重复”结构域的RNA。这种方法需要N-端和C-端RNA分别在3'末端和5'末端含有单种核酶。内部重复RNA在5'和3'末端都需要核酶,以使其在反式剪接过程中同时作为受体和供体发挥作用。然而,在RNA的两个端或同时具有3'-P和5'-OH的RNA上添加核酶会导致连结酶(诸如RtcB)环化(Desai等人,2015),从而阻止参与不断增长的线性链。然而,利用诱导型反式活化核酶可以通过添加和去除VS-Rz和RtcB连结酶来逐步连结5'和3'末端,从而控制RNA结构域合成(图14C)。这种方法可用于产生高度重复的RNA序列,该高度重复的RNA序列随后可被翻译从而产生合成重复蛋白(诸如构成水凝胶、合成蜘蛛丝或胶原蛋白等的蛋白),由于重组,这些蛋白很难生成并编码为DNA。这些方法可能对药物递送、生物材料或工业材料的产生有用(Chambre等人,2020)。

[0354] 实施例3:使用核酶产生稳定的合成内含子序列

[0355] 当一种RNA含有3'核酶而另一种RNA含有5'核酶时,两种独立RNA之间可以发生核酶介导的反式剪接(图15A)。然而,已证明,当在同一RNA中以顺式被转录时,两种核酶可以介导它们自身的无痕去除(图15B)。这种方法同样会产生两个独立的具有3'-P和5'OH端的RNA,它们可以在细胞中经受反式剪接和翻译(图15B)。这也可以通过添加连结酶(诸如RtcB)在体外实现。

[0356] 由核酶产生的内含子序列也含有相容5'-OH和3'-P末端,可以被顺式剪接或环化,是体外RtcB连结酶活性的常见读出。与外显子剪接期间由剪接体产生的快速降解的套索RNA不同,RNA环被认为是高度稳定的,因为它们不再包含5'或3'末端,因此不能被RNA外切

核酸酶降解。可以包含任意数量的功能性或有用的RNA (诸如微RNA、CRISPR指导RNA等) 或基因表达序列的货物序列可以作为“货物”被插入在两种核酶之间(图15C)。这种方法可用于在核酶介导的反式剪接和表达过程中共同递送和表达有用的RNA序列。如果其中一种内部核酶不需要双侧侧接序列来发挥活性, 诸如对于5' HDV核酶, RNA环可以以环状和经再切割的线性形式存在(图15C)。当使用VS-S代替HDV时, 所述系统可以被制成可诱导的, 需要VS-Rz的递送或表达。使用需要双侧侧接序列进行切割的核酶(诸如HH核酶), 可以将切割设计成使得货物RNA的RNA环化是单向的(图15D)。

[0357] 实施例4: 序列

[0358] 编码反式剪接蛋白的核酸序列

[0359] Nt-GFP (SEQ ID NO:1)

[0360] AUGGUGAGCAAGGGCGAGGAGCUGUUCACCGGGUGGUGCCCAUCCUGGUCGAGCUGGACGGCGACGUA
AACGGCCACAAGUUCAGCGUGUCCGGCGAGGGCGAGGGCGAUGCCACCUACGGCAAGCUGACCCUGAAGUUAUCUG
CACCACCGGCAAGCUGCCCCUGGCCACCCUCGUGACCACCCUGACCUACGGCGUGCAGUGCUUCAGCCGCU
ACCCCGACCACAUGAAGCAGCAGACUUCUUAAGUCCGCAUGCCCGAAGGCUACGUCCAGGAGCGCACCAUCUUC
UU

[0361] Ct-GFP (SEQ ID NO:2)

[0362] CAAGGACGACGGCAACUACAAGACCCGCGCCGAGGUGAAGUUCGAGGGCGACACCCUGGUGAACCGCAU
CGAGCUGAAGGGCAUCGACUUAAGGAGGACGGCAACAUCUGGGGCACAAGCUGGAGUACAACUACAACAGCCACA
ACGUCUAUAUCAUGGCCGACAAGCAGAAGAACGGCAUCAAGGUGAACUUAAGAUCGCCACAACAUCGAGGACGGC
AGCGUGCAGCUCGCCACCACUACCAGCAGAACACCCCAUCGGCGACGGCCCGUGCUGCUGCCCGACAACCACUA
CCUGAGCACCCAGUCCGCCUGAGCAAAGACCCCAACGAGAAGCGCGAUCACAUGGUCCUGCUGGAGUUCGUGACCG
CCGCCGGGAUCACUCUGGCAUGGACGAGCUGUACAAGUAGUAA

[0363] Nt-萤光素酶 (SEQ ID NO:3)

[0364] AUGGAAGACGCCAAAAACAUAAGAAAGGCCCGGGCGCAUUCUAUCCGUCUGGAAGAUGGAACCGCUGGA
GAGCAACUGCAUAAGGCUAUGAAGAGAUACGCCUGGUUCCUGGAACAAUUGCUUUACAGAUGCACAUUUCGAGGU
GGACAUCACUUAACGUCGAGUACUUCGAAAUGUCCGUUCGGUUGGCAGAAGCUAUGAAACGAUAUGGGCUGAAUACA
AUCACAGAAUCGUCGUAUGCAGUGAAAACUCUCUUAUUCUUAUGCCGGUGUUGGGCGCGUUAUUUAUCGGAGUU
GCAGUUGCGCCCGGAACGACAUUUUAUGAACGUGAAUUGCUAACAGUAUGGGCAUUUCGCAGCCUACCGUGGU
GUUCGUUUCAAAAAGGGGUUGCAAAAAUUUUGAACGUGCAAAAAAGCUCCCAUCAUCCAAAAUUUAUUAUCA
UGGAUUCUAAAACGGAUUACCAGGGAUUUCAGUCGAUGUACACGUUCGUCACAUCUCAUCUACCUCCCGUUUUAU
GAAUACGAUUUUGUGCCAGAGUCCUUCGAUAGGGACAAGACAUAUUGCACUGAUCAUGAACUCCUCUGGAUCUACUGG
UCUGCCUAAAAGGUGUCGCUUGCCUCAUAGAACUGCCUGCGUGAGAUUCUCGCAUGCCAGAGAUCCAUUUUUGGCA
AUCAAAUCAUCCGGAUACUGCGAUUUUAAGUGUUGUCCAUCACGGUUUUGGAAUGUUUACUACACUCGGA
UAUUUGAUUAUGUGGAUUUCGAGUCGUCUUAUGUAUAGAUUUGAAGAAGAGCUGUUUCUGAGGAGCCUU

[0365] Ct-萤光素酶 (SEQ ID NO:4)

[0366] CAGGAUUACAAGAUUCAAGUGCGCUGCUGGUGCCAACCCUAUUCUCCUUCUUCGCCAAAAGCACUCUG
AUUGACAAUACGAUUUAUCUAAUUUACACGAAAUUGCUUCUGGUGGCGUCCCCUCUCUAAGGAAGUCGGGGAAGC
GGUUGCCAAGAGGUUCAUCUGCCAGGUAUCAGGCAAGGAUAUGGGCUCACUGAGACUACAUCAGCUAUUCUGAUUA
CACCCGAGGGGAUGAUAAACCGGGCGCGGUCGGUAAAGUUGUCCAUUUUUGAAGCGAAGGUUGUGGAUCUGGAU

ACCGGGAAAACGCUGGGCGUAAUCAAAAGAGGCGAACUGUGUGUGAGAGGUCCUAUGAUUAUGUCCGGUUAUGUAAA
CAAUCCGGAAGCGACCAACGCCUUGAUUGACAAGGAUGGAUGGCUACAUUCUGGAGACAUAGCUUACUGGGACGAAG
ACGAACACUUCUUCUUCGUGACCGCCUGAAGUCUCUGAUUAAGUACAAAGGCUAUCAGGUGGCUCGCGUGAAUUG
GAAUCCAUCUUGCUCCAACACCCCAACAUCUUCGACGCGAGGUGUCGAGGUCUUCGCGACGAUGACGCCGGUGAACU
UCCCGCCGCGUUGUUGUUUGGAGCACGGAAAAGACGAUGACGGAAAAAGAGAUUCGUGGAUUACGUCGCCAGUCAAG
UAACAACCGCGAAAAAGUUGCGCGGAGGAGUUGUGUUUGUGGACGAAGUACCGAAAGGUCUUAACCGGAAAACUCGAC
GCAAGAAAAAUCAGAGAGAUCCUCAUAAAAGGCCAAGAAGGGCGGAAAAGAUUCGCCGUGUAGUAA

[0367] N1L (SEQ ID NO:5)

[0368] ATGGGTCAGGCCAATACGCCCTGGAGCAGTAAGGCAAACGCGGATGCCTTTATAAATTCATTCATCAGT
GCAGCATCCAATACTGGTTCCTTCTCTCAAGACCAAATGGAGGACATGTCACTCATCGGCAATACTCTGATGGCTGC
CATGGACAATATGGGAGGCCGCATAACACCATCTAAGTTGCAGGCGTTGGATATGGCCTTCGCATCATCAGTGGCCG
AGATCGCGGCTAGTGAGGGCGGCGACTTGGGAGTCACTACCAACGCGATCGCGGATGCCCTCACTTCTGCTTTTTAT
CAAACGACCGGGGTTGTCAATTCACGATTCATATCTGAGATCAGGAGCCTCATAGGAATGTTGCGCAGGCTTCCGC
AAATGACGTTTTATGCATCTGCTGGCTCTGGCAGCGGGGTGGTGGGTATGGAGCCAGCTCAGCATCTGCGGCTTCTG
CAAGTGCTGCTGCCCCGAGTGGCGTAGCTTATCAGGCTCCTGCTCAGGCTCAAATCAGTTTTACGTTGCGAGGGCAA
CAACCTGTTTCC

[0369] AQ (SEQ ID NO:6)

[0370] GGTCTTATGGACCCGGTGCTAGCGCTGCGGCAGCAGCCGCTGGCGTTATGGCCCAGGTTTCAGGGCAA
CAGGGGCTGGGCAACAAGGACCTGGCCAACAAGGTCCTGGTCAGCAGGGTCCAGGGCAGCAG

[0371] NR3 (SEQ ID NO:7)

[0372] GGCCTGCTTCCGCTGCAGTATCAGTAGGTGGCTATGGACCTCAATCTAGTAGCGCCCCTGTTGCCTCT
GCCGCCGCATCTCGACTTTCAAGTCCC GCCGCTAGTTCCAGGGTCAGTTCCGCGGTATCTAGCTTGTAAGTAGCGG
ACCCACTAATCAAGCGGCACTTTCAAACACAATATCCTCAGTAGTCAGTCAAGTAAGCGCATCAAACCCTGGCTTGT
CAGGGTGTGACGTTCTGGTTCAGGCACTTCTGGAAGTTGTCTCAGCGTTGGTAAGCATCCTGGGTAGCTCCTCCATA
GGTCAAATTAATTATGGCGGAGCGCCAATACACACAAATGGTGGGTCAGAGTGTGGCGCAGGCACTCGCAGGCGA
CTACAAGGATCATGACGGAGACTATAAGGATCATGATATAGATTACAAGGACGATGATGACAAGGCCTAGTAA

[0373] Nt-4xMTS (SEQ ID NO:8)

[0374] AUGAGUGUGUAGACGCCGUUCUUCUGCGAGGGCUUACCGGGUCUGCUAGAAGACUUCGGUCCCCAGG
GCCAAGAUACAUAGCCUCGGAGACCCGAUGUCUGUGCUCACUCCUCUGCUUUUGCGAGGACUGACUGGGUCCGCCAG
ACGACUCCCGGUGCCGAGAGCUAAAAUCCAUAAGCCUGGGAAAAUUGGCAACUAUGUCAGUCCUGACGCCGCUUCUUC
UCCGGGGUCUACAGGGUCUGCAAGAAGGCUGCCUGUACCUCGGGCGAAAAUUCAUAGCUUGGGCGACCCGAUGAGU
GUAUUGACGCCCCUGUUGCUGAGAGGAUUGACUGGGUCAGCGCGCCGGCUCCUGUCCCCCGAGCUAAGAUUCACUC
CCUUGGUAAGCUGAGAAUCCUCCAAUCAACGGUUCGAGAGCAAGAGAUCCGCCGGUCGCCACGAGGCCUCUCGAG

[0375] Nt-DTA (SEQ ID NO:17)

[0376] AUGGACCCCGACGACGUGGUGGACAGCAGCAAGAGCUUCGUGAUGGAGAACUUCAGCAGCUACCACGGC
ACCAAGCCCGGCUACGUGGACAGCAUCCAGAAGGGCAUCCAGAAGCCCAAGAGCGGCACCCAGGGCAACUACGACGA
CGACUGGAAGGGCUUCUACAGCACCGACAACAAGUACGACGUCGCCGGCUACAGCGUGGACAACGAGAACCCCCUGA
GCGGCAAGGCCGGCGGUGGUGAAGGUGACCUACCCCGGCCUGACCAAGGUGCUGGCCCUAAGGUG

[0377] Ct-DTA (SEQ ID NO:18)

[0378] GACAAUGCCGAGACCAUCAAGAAGGAGCUGGGCCUGAGCCUGACCGAGCCCCUGAUGGAGCAGGUGGGC
ACCGAGGAGUUCAUCAAGAGAUUCGGCGACGGCGCCAGCAGAGUGGUGCUGAGCCUGCCCUUCGCCGAGGGCAGCAG
CAGCGUGGAGUACAUCAACAACUGGGAGCAGGCCAAGGCCUGAGCGUGGAGCUGGAGAUAACUUCGAGACCAGAG
GCAAGAGAGGCCAGGACGCCAUGUACGAGUACAUGGCCAGGCUUGCGCCGCAACAGAGUGAGAAGAUAGUAA

[0379] GFP_cdn(无起始ATG密码子) (SEQ ID NO:19)

[0380] GUUAGCAAGGGCGAGGAGCUCUUCACCGGGGUCGUCCCCAUCCUCGUCGAGCUCGACGGCGACGUAAAC
GGCCACAAGUUCAGCGUCUCCGGCGAGGGCGAGGGCGAUGCCACCUACGGCAAGCUCACCCUGAAGUUCAUCUGCAC
CACCGGCAAGCUGCCCGUGCCCUUGGCCACCCUCGUGACCACCCUGACCUACGGCGUGCAGUGCUUCAGCCGCUACC
CCGACCACAUGAAGCAGCAGCAGACUUCUUCAGUCCGCCAUGCCCCAAGGCUACGUCCAGGAGCGCACCAUCUUCUUC
AAGGACGACGGCAACUACAAGACCCGCGCCGAGGUGAAGUUCGAGGGCGACACCCUGGUGAACCGCAUCGAGCUGAA
GGGCAUCGACUUAAGGAGGACGGCAACAUCUGGGGCACAAGCUGGAGUACAACUACAACAGCCACAACGUCUUAUA
UCAUGGCCGACAAGCAGAAGAACGGCAUCAAGGUGAACUUAAGAUCGGCCACAACUUCGAGGACGGCAGCGUGCAG
CUCGCGGACCACUACCAGCAGAACACCCCAUCGGCGACGGCCCCGUGCUGCUGCCCGACAACCACUACCUGAGCAC
CCAGUCCGCCUGAGCAAAGACCCCAACGAGAAGCGCGAUCACAUGGUCCUGCUGGAGUUCGUGACCGCCGCCGGGA
UCACUCUCGGCAUGGACGAGCUGUACAAGUAG

[0381] F2-Myr (SEQ ID NO:20)

[0382] AUGGGUUGUUGUUUCAGCAAGACAGCGGCGAAAGGUGAAGCAGCAGCAGAAAGACCAGGCGAGGCUGCG
GUAGCAUCAAGUCCCUCCAAGGCUAAUGGGCAGGAAAACGGACACGUCAAAGUUGGAAGCGU

[0383] F2-RFP (SEQ ID NO:21)

[0384] AGCCAUCAUCAAGGAGUUCAUGCGCUUCAAGGUGCACAUGGAGGGCUCGUGAACGGCCACGAGUUCGA
GAUCGAGGGCGAGGGCGAGGGCCGCCCUACGAGGGCACCCAGACCGCCAAGCUGAAGGUGACCAAGGGUGGCCCCC
UGCCCUUCGCCUGGGACAUCUCCUGUCCCUUCAGUUCAUGUACGGCUCCAAGGCCUACGUGAAGCACCCCGCCGACAUC
CCCACUACUUGAAGCUGUCCUUCUCCCGAGGGCUUCAAGUGGGAGCGCGUGAUGAACUUCGAGGACGGCGGGCUGGU
GACCGUGACCCAGGACUCCUCCUGCAGGACGGCGAGUUAUCUACAAGGUGAAGCUGCGCGGCACCAACUUCUCCCU
CCGACGGCCCCGUAUUGCAGAAGAAGACCAUGGGCUGGGAGGCCUCCUCCGAGCGGAUGUACCCCGAGGACGGCGCC
CUGAAGGGCGAGAUCAAGCAGAGGCUGAAGCUGAAGGACGGCGGCCACUACGACGCUGAGGUCUAGACCACCUACAA
GGCCAAGAAGCCCGUGCAGCUGCCCGGGCCUACAACGUCAACAUCAAGUUGGACUACCCUCCACAACGAGGACU
ACACCAUCGUGGAACAGUACGAACGCGCCGAGGGCCGCCACUCCACCGCGGCAUGGACGAGCUGUACAAGUAGUAA

[0385] Nt-uDys (SEQ ID NO:22)

[0386] AUGCUUUGGUGGAAGAAGUAGAGGACUGUUAUGAAAGAGAAGAUUCAAAGAAAACAUUCACAAAA
UGGGUAAAUGCACAAUUUCUAAGUUUGGGAAGCAGCAUAUUGAGAACCUCUUCAGUGACCUACAGGAUGGGAGGCG
CCUCCUAGACCUCUUGAAGGCCUGACAGGGCAAAAACUGCCAAAAGAAAAGGAUCCACAAGAGUUAUGCCCUGA
ACAAUGUCAACAAGGCACUGCGGGUUUUGCAGAACAUAUUGUUGAUUUAGUGAAUUAUUGGAAGUACUGACAUCGUA
GAUGGAAAUCAUAAACUGACUCUUGGUUUGAAUUAUAUCCUCCACUGGCAGGUCAAAAUGUAAUGAAAA
UAUCAUGGCUGGAUUGCAACAACCAACAGUGAAAAGAUUCUCCUGAGCUGGGUCCGACAUAACUCGUAAUUAUC
CACAGGUUAAUGUAAUCAACUUCACCACCAGCUGGUCUGAUGGCCUGGCUUUGAAUGCUCUCAUCCAUAUGUCAUAGG
CCAGACCUAUUUGACUGGAAUAGUGUGGUUUGCCAGCAGUCAGCCACACAACGACUGGAACAUGCAUUAACAUCGC
CAGAUUAUAAUAGGCAUAGAGAAAACUACUCGAUCCUGAAGAUGUUGAUUACCACCUAUCAGAUAAAGAUCCAUCU
UAAUGUACAUCACAUCACUCUCCAAGUUUUGCCUCAACAAGUGAGCAUUGAAGCAUCCAGGAAGUGGAAAUGUUG

CCAAGGCCACCUAAAGUGACUAAAGAAGAACAUUUUCAGUUACAUCAUCAAAUUGCACUAUUCUCAACAGAUCACGGU
CAGUCUAGCACAGGGUAUAGAGAGAACUUCUCCCUAAGCCUCGAUUCAAGAGCUAUGCCUACACACAGGCUGCUU
AUGUCACCACCUCUGACCCUACACGGAGCCCAUUUCCUUCACAGCAUUUGGAAGCUCCUGAAGACAAGUCAUUUGGC
AGUUCAUUGAUGGAGAGUGAAGUAAAACCGGACCGUUUAUCAAACAGCUUUAGAAGAAGUAUUUUCGUGGCUUCUUUC
UGCUGAGGACACAUUGCAAGCACAAGGAGAGAUUUCUAAUGAUGUGGAAGUGGUGAAAGACCAGUUUCAUACUCAUG
AGGGGUACAUGAUGGAUUUGACAGCCCAUCAGGGCCGGGUUGGUAUUUAUUCUACAAUUGGGAAGUAAGCUGAUUGGA
ACAGGAAAAUUUAUCAGAAGAUAGAAGAAACUGAAGUACAAGAGCAGAUGAAUCUCCUAAAUAAGAUGGGAAUGCCU
CAGGGUAGCUAGCAUGGAAAAACAAAGCAAUUUACAUAAGAGUUUUAAUGGAUCUCCAGAAUCAGAAACUGAAAGAGU
UGAAUGACUGGCUAACAAAAACAGAAGAAACAAGGAAAAUGGAGGAAGAGCCUCUUGGACCUGAUCUUGAAGAC
CUAAAACGCCAAGUACAACAACAUAAGGUGCUUCAAGAAGAUCAAGAACAAGAACAAGUCAGGGUCAAUUCUCUCAC
UCACAUGGUGGUGGUAGUUGAUGAAUCUAGUGGAGAUACGCAACUGCUGCUUUGGAAGAACAACUUAAGGUUUUGG
GAGAUCGAUGGGCAAACAUCUGUAGAUGGACAGAAGACCGCUGGGUUCUUUACAAGACAUCUUCUCAAUUGGCAA
CGUCUUACUGAAGAACAGUGCCUUUUUAGUGCAUGGCUUCAGAAAAAGAAGUAGUGAACAAGAUUCACACAAC
UGGCUUUAAAGAUCAAAAUGAAAUGUUAUCAAGUCUUCAAAAACUGGCCGUUUUAAAAGCGGAUCUAGAAAAGAAAA
AGCAAUCCAUGGGCAAACUGUAUUCACUCAAAACAAGAUUCUUCUUAACACUGAAGAAUAAGUCAGUGACCCAGAAG
ACGGAAGCAUGGCUGGAUAACUUUGCCCGGUGUUGGGAUAAUUUAGUCCAAAAACUUGAAAAGAGUACAGCACAGAU
UUCACAGGCUGUCACCACCACUCAGCCAUCACUAAACACAGACAACUGUAAUGGAAACAGUAACUACGGUGACCACAA
GGGAACAGAUCCUGGUAAGCAUGCUCUAAAGAGGAACUCCACCACCACCUCUCCAAAAGAAGAGGCAGAUUACUGUG
GAUCUUGAAAGACUCCAGGAACUUAAGAGGCCACGGAUGAGCUGGACCUCUAGCUGCGCCAAGCUGAGGUGAUCAA
GGGAUCCUGGCAGCCCGUGGGCGAUCUCCUCAUUGACUCUCUCCAAGAUCACCUCGAGAAAGUCAAGGCACUUCGAG
GAGAAAUUGCGCCUCUGAAAGAGAACGUGAGCCAC

[0387] Ct-uDys-GFP (SEQ ID NO:23)

[0388] GUCAAUGACCUUGCUCGCCAGCUUACCACUUUGGGCAUUCAGCUCUCACCGUAUAACCUCAGCACUCUG
GAAGACCUGAACACCAGAUGGAAGCUUCUCGAGGUGGCCGUCGAGGACCGAGUCAGGCAGCUGCAUGAAGCCACAG
GGACUUUGGUCCAGCAUCUCAGCACUUUCUUCACGUCUGUCCAGGGUCCCUGGGAGAGAGCCAUCUCGCCAAACA
AAGUGCCCUACUUAUUAUCAAACCACGAGACUCAAAACAACUUGCUGGGACCAUCCAAAUAUGACAGAGCUCUACCAGUCU
UUAGCUGACCUGAAUAAUUGUCAGAUUCUCAGCUUAUAGGACUGCCAUGAAACUCCGAAGACUGCAGAAGGCCCUUG
CUUGGAUCUCUUGAGCCUGUCAGCUGCAUGUGAUGCCUUGGACCAGCACAAACCUCUAGCAAAAUGACCAGCCUUGG
AUAUCCUGCAGAUUAAUUAUUGUUUGACCACUUAUUUAGACCGCCUGGAGCAAGAGCACAACAUAUUGGUCAACGUC
CCUCUCUGCGUGGAUAUGUGUCUGAACUGGCUGCUGAAUGUUUAUGAUACGGGACGAACAGGGAGGAUCCGUGUCCU
GUCUUUUAAAACUGGCAUCAUUUCCUUGUGUAAAAGCACAUUUGGAAGACAAGUACAGAUACCUUUUCAAGCAAGUGG
CAAGUUAACAGGAUUUUGUGACCAGCGCAGGCUGGGCCUCCUUCUGCAUGAUUCUUAUCCAAAUAUCCAAGACAGUUG
GGUGAAGUUGCAUCCUUUGGGGGCAGUAAACAUUGAGCCAAGUGUCCGGAGCUGCUUCCAUAUUGCUAAUAAUAAAGCC
AGAGAUCCAAGCGGCCUUCUCCUAGACUGGAUGAGACUGGAACCCAGUCCAUGGUGUGGCUGCCCGUCCUGCACA
GAGUGGCUGCUGCAGAAACUGCCAAGCAUCAGGCCAAAUGUAACAUCUGCAAAGAGUGUCCAUAUCAUUGGAUUCAGG
UACAGGAGUCUAAAGCACUUUAAUUAUGACAUCUGCCAAAGCUGCUUUUUUCUGGUCGAGUUGCAAAGGCCAUAA
AAUGCACUAUCCAUUGGUGGAUAUUGCACUCCGACUACAUCAGGAGAAGAUGUUCGAGACUUUGCCAAGGUACUAA
AAAACAAUUCGAACCAAAAAGGUUUUUGCGAAGCAUCCCGAAUGGGCUACCUGCCAGUCGAGACUGUCUUAGAG
GGGGACAACAUGGAAACUGACACAUAUCUAGAGGUGAGCAAGGGCGAGGAGCUGUUCACCGGGGUGGUGCCCAUCCU

GGUCGAGCUGGACGGCGACGUAAACGGCCACAAGUUCAGCGUGUCCGGCGAGGGCGAUGCCACCUACGGCA
AGCUGACCCUGAAGUUAUCUGCACCACCGGCAAGCUGCCCGUGCCCGACCCUCGUGACCACCCUGACCUAC
GGCGUGCAGUGCUUCAGCCGUACCCCGACCACAUGAAGCAGCAGACUUCUUAAGUCCGCCAUGCCCGAAGGCUA
CGUCCAGGAGCGCACCAUCUUCUUAAGGACGACGGCAACUACAAGACCCGCGCCGAGGUGAAGUUCGAGGGCGACA
CCCUGGUGAACCGCAUCGAGCUGAAGGGCAUCGACUUAAGGAGGACGGCAACAUCUGGGGGACAAGCUGGAGUAC
AACUACAACAGCCACAACGUCUAUAUCAUGGCCGACAAGCAGAAGAACGGCAUCAAGGUGAACUUAAGAUCGCCA
CAACAUCGAGGACGGCAGCGUGCAGCUCGCCGACCACUACCAGCAGAACACCCCAUCGGCGACGGCCCCGUGCUGC
UGCCCGACAACCACUACCUGAGCACCCAGUCCGCCUGAGCAAAGACCCCAACGAGAAGCGCGAUCACAUGGUCCUG
CUGGAGUUCGUGACCGCCGCCGGGAUCACUCUGGCAUGGACGAGCUGUACAAGUAA

[0389] Nt-miniDys(Δ H2-R15) (SEQ ID NO:129)

[0390] AUGCUUUGGUGGAAGAAGUAGAGGACUGUUAUGAAAGAGAAGUUCAAAAGAAAACAUUCACAAAA
UGGGUAAAUGCACAAUUUCUAAGUUUGGGAAGCAGCAUAUUGAGAACCUCUUCAGUGACCUACAGGAUGGGAGGCG
CCUCCUAGACCUCUCGAAGGCCUGACAGGGCAAAAACUGCCAAAAGAAAAGGAUCCACAAGAGUUAUGCCCUGA
ACAAUGUCAACAAGGCACUGCGGGUUUUGCAGAAACAUAUUGUUAUUAGUGAAUUAUUGGAAGUACUGACAUCGUA
GAUGGAAAUCAUAAACUGACUCUUGGUUUGAUUUGGAAUUAUUAUCCUCCACUGGCAGGUCAAAAUGUAAUGAAAA
UAUCAUGGCUGGAUUGCAACAAACCAACAGUGAAAAGAUUCUCCUGAGCUGGGUCCGACAUAACUCGUAAUUUUC
CACAGGUUAAUGUAAUCAACUUCACCACCAGCUGGUCUGAUGGCCUGGCUUUGAAUGCUCUCAUCCAUAUGUCAUAGG
CCAGACCUAUUUGACUGGAAUAGUGUGGUUUGCCAGCAGUCAGCCACACAACGACUGGAACAUGCAUUAACAUCGC
CAGAUUAUCAAUAGGCAUAGAGAAACUACUCGAUCCUGAAGAUGUUAUACCACCUAUCAGAUAAAGUCCAUUCU
UAAUGUACAUCACAUCACUCUCCAAGUUUUGCCUCAACAAGUGAGCAUUGAAGCCAUCAGGAAGUGGAAAUGUUG
CCAAGGCCACCUAAAGUGACUAAAGAAGAACAUUUUCAGUUACAUCUAAAUGCACUAUUCUCAACAGAUACAGGU
CAGUCUAGCACAGGGAUUGAGAGAACUUCUCCCUAAGCCUCGAUUAAGAGCUAUGCCUACACACAGGCUGCUU
AUGUCACCACCUCUGACCCUACACGGAGCCCAUUUCCUUCACAGCAUUGGAAGCUCUGAAGACAAGUCAUUUGGC
AGUUCAUUGAUGGAGAGUGAAGUAAAACCGGACCGUUUAUCAAACAGCUUUAAGAAGUAUUAUCGUGGCUCUUUC
UGCUGAGGACACAUUGCAAGCACAAGGAGAGAUUUCUAAUGAUGUGGAAGUGGUGAAAGACCAGUUUCAUACUCAUG
AGGGGUACAUGAUGGAUUUGACAGCCCAUCAGGGCCGGGUUGGUAUUAUUCUACAUAUUGGGAAGUAAGCUGAUUGGA
ACAGGAAAAUUAUCAGAAGAUGAAGAAACUGAAGUACAAGAGCAGAUGAAUCUCCUAAAUAAGAUGGGAAUGCCU
CAGGGUAGCUAGCAUGGAAAAACAAGCAAUUUACAUAAGAGUUUUAUUGGAUCUCCAGAAUCAGAAACUGAAAGAGU
UGAAUGACUGGCUAACAAAAACAGAAGAAAGAAACAAGGAAAAUGGAGGAAGAGCCUCUUGGACCUGAUCUUGAAGAC
CUAAAACGCCAAGUACAACAACAUAAGGUGCUUCAAGAAGAUUAGAACAAGAACAAGUCAGGGUCAAUUCUCUCAC
UCACAUGGUGGUGGUAGUUGAUGAAUCUAGUGGAGAUACGCAACUGCUGCUUUGGAAGAACAACUUAAGGUUAUUGG
GAGAUCGAUGGGCAAACAUCUGUAGAUGGACAGAAGACCGCUGGGUUCUUUAACAAGACAUCUCAAUUGGCAA
CGUCUUACUGAAGAACAGUGCCUUUUUAGUGCAUGGCUUUCAGAAAAAGAAGUAGCAGUGAACAAGAUUCACACAAC
UGGCUUUAAAAGAUCAAAUUGAAAUGUUAUCAAGUCUUAACAAACUGGCCGUUUUAAAAGCGGAUCUAGAAAAGAAAA
AGCAAUCCAUGGGCAAACUGUAUUCACUCAAAACAAGAUUCUUCUUAACACUGAAGAAUUAAGUCAGUGACCCAGAAG
ACGGAAGCAUGGCUGGAUAACUUUGCCCGGUGUUGGGAUAAUUUAGUCCAAAAACUUGAAAAGAGUACAGCACAGAU
UUCACAGGAAAUUCUUAUGUGCCUUCUACUUAUUUGACUGAAAUCACUCAUGUCUCAAGCCCUAUUAGAAGUGG
AAACAUCUCAAUGCUCUCCUGACCUCUGUCUAAGGACUUAUGAAGACCUCUUUAAGCAAGAGGAGUCUCUGAAGAAU
AUAAAAGAUAGUCUACAACAAGCUCAGGUCGGAUUGACAUAUUAUAGCAAGAAGACAGCAGCAUUGCAAAGUGC

AACGCCUGUGGAAAGGGUGAAGCUACAGGAAGCUCUCUCCCAGCUUGAUUUCCAUGGGAAAAAGUUAACAAAUGU
 ACAAGGACCGACAAGGGCGAUUUGACAGAUCCGUUGAGAAAUGGCGGCGUUUCAUUAUGAUUAAAAGAUUUUAAU
 CAGUGGCUAACAGAAGCUGAACAGUUUCUCAGAAAGACACAAAUCCUGAGAAUUGGGAACAUGCIAAAUACAAAUG
 GUAUCUUAAGGAACUCCAGGAUGGCAUUGGGCAGCGGCAAACUGUUGUCAGAACAUUGAAUGCAACUGGGGAAGAAA
 UAAUUCAGCAAUCCUCAAACAGAUCCAGUAUUCUACAGGAAAAAUUGGGAAGCCUGAAUCUGCGGUGGCAGGAG
 GUCUGCAAACAGCUGUCAGACAGAAAAAGAGGCUAGAAGAACAAAAGAAUUCUUGUCAGAAUUCAAAGAGAUUU
 AAAUGAAUUUGUUUUAUGGUUGGAGGAAGCAGUAACAUAUGCUAGUAUCCACUUGAACCCUGGAAAAGAGCAGCAAC
 UAAAAGAAAAGCUUGAGCAAGUCAAGUUAACUGGUGGAAGAGUUGCCCCUGCGCCAGGGAAUCCUCAAACAAUAAAAU
 GAAACUGGAGGACCCGUGCUUGAAGUGCUCUCAUAGCCCAGAAGAGCAAGAUAAACUUGAAAAUAAAGCUCAAGCA
 GACAAAUCUCCAGUGGAUAAAGGUUCCAGAGCUUUAACUGAGAAACAAGGAGAAUUGAAGCUCAAAUAAAAGACC
 UUGGGCAGCUUGAAAAAAAGCUUGAAGACCUUGAAGAGCAGUUAAAUCAUCUGCUGCUGUGGUUAUCUCCUAAUAGG
 AAUCAGUUGGAAAUUUAUAAACCAACCAAAACCAAGAAGGACCAUUUGACGUUAAGGAAACUGAAUAGCAGUUAAGC
 UAAACAACCGGAUGUGGAAGAGAUUUUUGUCUAAAAGGGCAGCAUUUGUACAAGGAAAAACCAGCCACUCAGCCAGUGA
 AGAGGAAGUUAGAAGACCUGUCCUCUGAGUGGAAGGCGUAAACCGUUUACUUCAAGAGCUGAGGGCAAAGCAGCCU
 GACCUAGCUCUCCUGGACUGACCACUAUUGGAGCCUCUCCUACUCAGACUGUACUCUGGUGACACAACCUGUGGUUAC
 UAAGGAAACUGCCAUCUCCAAACUAGAAAUGCCAUCUCCUUGAUGUUGGAGGUACCUGCUCUGGCAGAUUUAACC
 GGGCUUGGACAGAACUUAACCGACUGGCUUUCUCUGCUUGAUCAAGUUUAAAAUCACAACGCGUGAUGGUGGGCGAC
 CUUGAGGAUAUCAACGAGAUGAUCAUCAAGCAGAAGGCAACAAUGCAGGAUUUGAACAGAGGGCUGCCCAGUUGGA
 AGAACUCAUUAACCGCUGCCAAAAAUUGAAAAACAAGACCAGCAUCAAGAGGCUAGAACAUAUACGGAUCGAA
 UUGAAAGAAUUCAGAAUCAGUGGGAUGAAGUACAAG

[0391] Ct-miniDys(ΔH2-R15) (SEQ ID NO:130)

[0392] AACACCUUCAGAACCGGAGGCAACAGUUGAAUGAAAUGUUAAAGGAUUCACACAAUGGCUGGAAGCUA
 AGGAAGAAGCUGAGCAGGUCUUAAGGACAGGCCAGAGCCAAGCUGGAGUCAUGGAAGGAGGGUCCCUAUACAGUAGAU
 GCAAUCCAAAAGAAAAUCACAGAAACCAAGCAGUUGGCCAAAGACCUCGCGCAGUGGCAGACAAAUGUAGAUGUGGC
 AAAUGACUUGGCCUGAAAACUUCUCCGGGAUUUAUUCUGCAGAUUACCAGAAAAGUCCACAUGAUAAACAGAGAAUA
 UCAUUGCCUCUUGGAGAAGCAUUCAUAAAAGGGUGAGUGAGCGAGAGGCUGCUUUGGAAGAAACUCAUAGAUUACUG
 CAACAGUUCUCCUGGACCUGGAAAAGUUUCUUGCCUGGCUUACAGAAGCUGAAACAACUGCCAAUGUCCUACAGGA
 UGCUACCCGUAAGGAAAGGCUCUAGAAGACUCCAAGGGAGUAAAAGAGCUGAUGAAACAUGGCAAGACCUCUCCAAAG
 GUGAAAUUGAAGCUCACACAGAUUUUAUCACAACCUGGAUGAAAACAGCCAAAAAAUCCUGAGAUCUCCUGGAAGGU
 UCCGAUGAUGCAGUCCUGUUACAAGACGUUUGGAUACAUGAACUUAAGUGGAGUGAACUUCGGAAAAAGUCUCU
 CAACAUUAGGUCCAUUUGGAAGCCAGUUCUGACCAGUGGAAGCGUCUGCACCUCUUCUGCAGGAACUUCUGGUGU
 GGCUCACAGCUGAAAGAUGAUGAAUUAAGCCGGCAGGCACCUAUUGGAGGCGACUUCAGCAGUUCAGAAGCAGAAC
 GAUGUGCAUAGGGCCUUAAGAGGGAAUUGAAAAACUAAAAGAACCUGUAAUCAUGAGUACUCUUGAGACUGUACGAAU
 AUUUCUGACAGAGCAGCCUUUGGAAGGACUAGAGAAACUCUACCAGGAGCCCAGAGAGCUGCCUCCUGAGGAGAGAG
 CCCAGAAUGUCACUCGGCUUCUACGAAAAGCAGGCUGAGGAGGUCAAUACUGAGUGGGAAAAAUUGAACCGCACUCC
 GCUGACUGGCAGAGAAAAAUAGAUGAGACCCUUGAAAAGACUCCGGGAACUUAAGAGGCCACGGAUGAGCUGGACCU
 CAAGCUGCGCAAGCUGAGGUGAUCAAGGGAUCCUGGCAGCCCUGGGCGAUCUCCUCAUUGACUCUCUCCAAGAUC
 ACCUGGAGAAAGUCAAGGCACUUCGAGGAGAAAUUGCGCCUCUGAAAGAGAACGUGAGCCACGUCUUAUGACCUUGCU
 CGCCAGCUUACCACUUUGGGCAUUCAGCUCUCACCGUAUAACCUCAGCACUCUGGAAGACCUGAACACCAGAUGGAA

GCUUCUGCAGGUGCCGUCGAGGACCGAGUCAGGCAGCUGCAUGAAGCCCACAGGGACUUUGGUCCAGCAUCUCAGC
 ACUUUCUUUCCACGUCUGUCCAGGGUCCUGGGAGAGGCCAUCUCGCCAAACAAAGUGCCCUACUUAUAUCAACCAC
 GAGACUCAAAACAACUUGCUGGGACCAUCCAAAAUGACAGAGCUCUACCAGUCUUUAGCUGACCUGAAUAUUGUCAG
 AUUCUCAGCUUAUAGGACUGCCAUGAAACUCCGAAGACUGCAGAAGGCCUUUGCUUGGAUCUCUUGAGCCUGUCAG
 CUGCAUGUGAUGCCUUGGACCAGCACAACCUCUAAAGCAAAAUGACCAGCCCAUGGAUAUCCUGCAGAUUAUUAUUGU
 UUGACCACUAUUUAUGACCGCCUGGAGCAAGAGCACAACAAUUUGGUCAACGUCCCUUCUCUGCGUGGAUAUGUGUCU
 GAACUGGCUGCUGAAUGUUUAUGAUACGGGACGAACAGGGAGGAUCCGUGUCCUGUCUUUUAAAACUGGCAUCAUUU
 CCCUGUGUAAAAGCACAUUUGGAAGACAAGUACAGAUACCUUUUCAAGCAAGUGGCAAGUUAACAGGAUUUUGUGAC
 CAGCGCAGGCUGGGCCUCCUUCUGCAUGAUUCUAUCCAAAUCCAAGACAGUUGGGUGAAGUUGCAUCCUUUGGGGG
 CAGUAACAUGAGCCAAGUGUCCGGAGCUGCUUCCAAUUUGCUAAUAAUAAGCCAGAGAUCGAAGCGGCCCUUCUCC
 UAGACUGGAUGAGACUGGAACCCAGUCCAUGGUGUGGCUGCCCGUCCUGCACAGAGUGGCUGCUGCAGAAACUGCC
 AAGCAUCAGGCCAAAUGUAACAUCUGCAAAGAGUGUCCAUAUUGGAUUCAGGUACAGGAGUCUAAAGCACUUUAA
 UUAUGACAUCUGCCAAAGCUGCUUUUUUCUGGUCGAGUUGCAAAAGGCCAUAAAUGCACUAUCCCAUGGUGGAU
 AUUGCACUCCGACUACAUCAGGAGAAGAUGUUCGAGACUUUGCCAAGGUACUAAAAACAAAUUUCGAACCAAAGG
 UAUUUUGCGAAGCAUCCCCGAAUUGGCUACCUCCAGUGCAGACUGUCUUAAGAGGGGACAACAUGGAAACUCCCGU
 UACUCUGAUAACUUCUGGCCAGUAGAUCUGCGCCUGCCUCGUCCCUACAGCUUUCACACGAUGAUACUCAUUCAC
 GCAUUGAACAUAUAGCUAGCAGGCUAGCAGAAAUGGAAAACAGCAAUGGAUCUUAUCUAAAUGAUAGCAUCUCUCCU
 AAUGAGAGCAUAGAUGAUGAACAUAUUGUAAUCCAGCAUACUGCCAAAGUUUGAACCAGGACUCCCCCUGAGCCA
 GCCUCGUAGUCCUGCCCAGAUUUUAUUCUAGAGAGUGAGGAAAGAGGGGAGCUAGAGAGAAUCCUAGCAGAUC
 UUGAGGAAGAAAACAGGAAUCUGCAAGCAGAAUAUGACCGUCUAAAGCAGCAGCACGAACAUAAGGCCUGUCCCA
 CUGCCGUCCCUCCUGAAAUGAUGCCCACCUCUCCCCAGAGUCCCGGGAUGCUGAGCUCUUAUGCUGAGGCCAAGCU
 ACUGCGUCAACACAAAGGCCGCCUGGAAGCCAGGAUGCAAUCCUGGAAGACCACAUAACAGCUGGAGUCACAGU
 UACACAGGCUAAGGCAGCUGCUGGAGCAACCCAGGCAGAGGCCAAAGUGAAUGGCACAACGGUGUCCUCUCCUUCU
 ACCUCUCUACAGAGGUCGACAGCAGCAGCCUAUGCUGCUCGAGUGGUUGGCAGUCAAAUCUUGGACUCCAUGGG
 UGAGGAAGAUCUUCUAGUCCUCCCCAGGACACAAGCACAGGGUUAGAGGAGGUGAUGGAGCAACUCAACAACUCCU
 UCCUAGUUAAGAGGAAGAAUACCCUGGAAAGCCAAUGAGAGAGGACACAUAUGUAA

[0393] 用于无疤痕3'RNA切割的核酶核酸序列

[0394] HDV68 (SEQ ID NO:9)

[0395] GGCCGGCAUGGUCCAGCCUCCUCGUGGCGCCGGCUGGGCAACAUGCUUCGGCAUGGCGAAUGGGAC

[0396] HDV68催化突变体 (SEQ ID NO:24)

[0397] 5'-GGCCGGCAUGGUCCAGCCUCCUCGUGGCGCCGGCUGGGCAACAUGCUUCGGCAUGGUGAAUGG
 GAC-3'

[0398] HDV67 (SEQ ID NO:10)

[0399] GGGUCGGCAUGGCAUCUCCACCUCUCCGCGGUCCGACCUGGGCUACUUCGGUAGGCUAAGGGAGAAG

[0400] HDV56 (SEQ ID NO:11)

[0401] GAGGGAUAGUACAGAGCCUCCCCGUGGCUCCCUUGGAUAACCAACUGAUACUGUAC

[0402] 基因组HDV (genHDV) (SEQ ID NO:12)

[0403] GGCCGGCAUGGUCCAGCCUCCUCGUGGCGCCGGCUGGGCAACAUCCGAGGGGACCGUCCCUCCGGU
 AAUGGCGAAUGGGACCCA

- [0404] 抗基因组HDV(抗HDV) (SEQ ID NO:13)
- [0405] GGGUCGGCAUGGCAUCUCCACCUCUCGCGGUCCGACCUGGGCAUCCGAAGGAGGACGCACGUCCACUC
GGAUGGCUAAGGGAGAGCCACU
- [0406] VS核酶(SEQ ID NO:14)
- [0407] GCGGUAGUAAGCAGGGAACUCACCUCCAAUUUCAGUACUGAAAUUGUCGUAGCAGUUGACUACUGUUAU
GUGAUUGGUAGAGGCUAAGUGACGGUAUUGGCGUAAGUCAGUAUUGCAGCACAGCACAAGCCCCGCUUGCGAGAAU
- [0408] VS-S(SEQ ID NO:15)
- [0409] GAAGGGCGUCGUCGCCCCGAG
- [0410] VS-Rz(SEQ ID NO:16)
- [0411] GCGGUAGUAAGCAGGGAACUCACCUCCAAUUUCAGUACUGAAAUUGUCGUAGCAGUUGACUACUGUUAU
GUGAUUGGUAGAGGCUAAGUGACGGUAUUGGCGUAAGUCAGUAUUGCAGCACAGCACAAGCCCCGCUUGCGAGAAU
- [0412] 具有对Nt-Luc特异的茎3突出端的锤头状核酶(SEQ ID NO:25)
- [0413] 5'-GAGCCUUACCGAUGUGUUUCCGGUCUGAUGAGUCCGGUAGCGGACGAAAGGCUC-3'
- [0414] 用于Ct-Luc的具有5nt P1茎的Twister核酶(SEQ ID NO:26)
- [0415] 5'-AGCCUUAACACUGCCAAUGCCGGUCCCAAGCCCCGGAUAAAAGUGGAGGGAGGCU-3'
- [0416] 用于Ct-Luc和T6A突变的具有5nt P1茎的Twister核酶(SEQ ID NO:27)
- [0417] 5'-AGCCUUAACACUGCCAAUGCCGGUCCCAAGCCCCGGAUAAAAGUGGAGGGAGGCU-3'
- [0418] 用于Ct-Luc的具有5nt P1茎的Twister核酶突变体(SEQ ID NO:28)
- [0419] 5'-AGCCUUAACUCUCCAAUGCCGGUCCCAAGCCCCGGAUAAAAGUGGAGGGAGGCU-3'
- [0420] 用于Ct-Luc的具有5nt P1茎的Twister核酶(SEQ ID NO:29) 5'-AGCCUUAACACUGC
CAAUGCCGGUCCCAAGCCCCGGAUAAAAGUGGAGGGAGGCU-3'
- [0421] 用于Ct-Luc的具有2nt P1茎的Twister核酶(SEQ ID NO:30)
- [0422] 5'-AGCCUUAACACUGCCAAUGCCGGUCCCAAGCCCCGGAUAAAAGUGGAGGGAG-3'
- [0423] 用于Ct-Luc的具有1nt P1茎的Twister核酶(SEQ ID NO:31) 5'-AGCCUUAACACUGC
CAAUGCCGGUCCCAAGCCCCGGAUAAAAGUGGAGGG-3'
- [0424] 用于Ct-Luc的不具有P1茎的Twister核酶(SEQ ID NO:32) 5'-AGCCUUAACACUGCCA
AUGCCGGUCCCAAGCCCCGGAUAAAAGUGGAGGG-3'
- [0425] 用于3'的锤头状(HH)核酶(SEQ ID NO:105)
- [0426] 5' NNNNDWHACCGGAUGUGUUUCCGGUCUGAUGAGUCCGGUAGCGGACGAAWHNNNN 3'
- [0427] 具有5nt P1茎的Twister核酶WT(SEQ ID NO:106)
- [0428] 5' NNNNNUAACACUGCCAAUGCCGGUCCCAAGCCCCGGAUAAAAGUGGAGGGNNNNN 3'
- [0429] 具有5nt P1茎的Twister核酶突变体(SEQ ID NO:107)
- [0430] 5' NNNNNUAACUCUCCAAUGCCGGUCCCAAGCCCCGGAUAAAAGUGGAGGGNNNNN 3'
- [0431] 具有有U1A突变的5nt P1茎的Twister核酶(SEQ ID NO:108)
- [0432] 5' NNNNNAACACUGCCAAUGCCGGUCCCAAGCCCCGGAUAAAAGUGGAGGGNNNNN 3'
- [0433] 具有有U1C突变的5nt P1茎的Twister核酶(SEQ ID NO:109)
- [0434] 5' NNNNCAACACUGCCAAUGCCGGUCCCAAGCCCCGGAUAAAAGUGGAGGGNNNNN 3'
- [0435] 具有有U1G突变的5nt P1茎的Twister核酶(SEQ ID NO:110)
- [0436] 5' NNNNGAACACUGCCAAUGCCGGUCCCAAGCCCCGGAUAAAAGUGGAGGGNNNNN 3'

- [0437] 用于无疤痕5'RNA切割的核酶核酸序列
- [0438] 具有对Ct-Luc特异的茎1突出端的锤头状(HH)核酶
- [0439] 16HH(SEQ ID NO:33)
- [0440] 5' -GAAUCUUGUAAUCCUGCUGAUGAGUCCGUGAGGACGAAACGAGUAAGCUCGUC-3'
- [0441] 14HH(SEQ ID NO:34)
- [0442] 5' -AUCUUGUAAUCCUGCUGAUGAGUCCGUGAGGACGAAACGAGUAAGCUCGUC-3'
- [0443] 12HH(SEQ ID NO:35)
- [0444] 5' -CUUGUAAUCCUGCUGAUGAGUCCGUGAGGACGAAACGAGUAAGCUCGUC-3'
- [0445] 8HH(SEQ ID NO:36)
- [0446] 5' -UAAUCCUGCUGAUGAGUCCGUGAGGACGAAACGAGUAAGCUCGUC-3'
- [0447] 6HH(SEQ ID NO:37)
- [0448] 5' -AUCCUGCUGAUGAGUCCGUGAGGACGAAACGAGUAAGCUCGUC-3'
- [0449] 6HH突变体(SEQ ID NO:38)
- [0450] 5' -AUCCUGCUGAUGAGUCCGUGAGGACGAGACGAGUAAGCUCGUC-3'
- [0451] 4HH(SEQ ID NO:39)
- [0452] 5' -CCUGCUGAUGAGUCCGUGAGGACGAAACGAGUAAGCUCGUC-3'
- [0453] 用于5'的锤头状核酶4nt突出端(SEQ ID NO:111)
- [0454] 5' NNNNCUGAUGAGUCCGUGAGGACGAAACGAGUAAGCUCGUC3'
- [0455] 用于5'的锤头状核酶6nt突出端(SEQ ID NO:112)
- [0456] 5' NNNNNNCUGAUGAGUCCGUGAGGACGAAACGAGUAAGCUCGUC 3'
- [0457] 用于5'的锤头状核酶8nt突出端(SEQ ID NO:113)
- [0458] 5' NNNNNNNNCUGAUGAGUCCGUGAGGACGAAACGAGUAAGCUCGUC 3'
- [0459] 用于5'的锤头状核酶10nt突出端(SEQ ID NO:114)
- [0460] 5' NNNNNNNNNNCUGAUGAGUCCGUGAGGACGAAACGAGUAAGCUCGUC 3'
- [0461] 用于5'的锤头状核酶12nt突出端(SEQ ID NO:115)
- [0462] 5' NNNNNNNNNNNNCUGAUGAGUCCGUGAGGACGAAACGAGUAAGCUCGUC 3'
- [0463] 用于5'的锤头状核酶14nt突出端(SEQ ID NO:116)
- [0464] 5' NNNNNNNNNNNNNNCUGAUGAGUCCGUGAGGACGAAACGAGUAAGCUCGUC 3'
- [0465] 用于5'的锤头状核酶16nt突出端(SEQ ID NO:117)
- [0466] 5' NNNNNNNNNNNNNNNNCUGAUGAGUCCGUGAGGACGAAACGAGUAAGCUCGUC 3'
- [0467] 用于5'的TX2锤头状核酶4nt突出端(Huang等人2019)(SEQ ID NO:118)
- [0468] 5' NNNNCUGAUGAGUCCGGUAGCGGACGAAACGCGCUUCGGUGCGUC 3'
- [0469] 用于5'的TX2锤头状核酶6nt突出端(Huang等人,2019)(SEQ ID NO:119)
- [0470] 5' NNNNNNCUGAUGAGUCCGGUAGCGGACGAAACGCGCUUCGGUGCGUC 3'
- [0471] 用于5'的TX2锤头状核酶8nt突出端(Huang等人2019)(SEQ ID NO:120)
- [0472] 5' NNNNNNNNCUGAUGAGUCCGGUAGCGGACGAAACGCGCUUCGGUGCGUC 3'
- [0473] 用于5'的TX2锤头状核酶10nt突出端(Huang等人2019)(SEQ ID NO:121)
- [0474] 5' NNNNNNNNNNCUGAUGAGUCCGGUAGCGGACGAAACGCGCUUCGGUGCGUC 3'
- [0475] 用于5'的TX2锤头状核酶12nt突出端(Huang等人2019)(SEQ ID NO:122)

- [0476] 5' NNNNNNNNNNNNCUGAUGAGUCCGGUAGCGGACGAAACGCGCUUCGGUGCGUC 3'
- [0477] 用于5'的TX2锤头状核酶14nt突出端(Huang等人2019)(SEQ ID NO:123)
- [0478] 5' NNNNNNNNNNNNNNCUGAUGAGUCCGGUAGCGGACGAAACGCGCUUCGGUGCGUC 3'
- [0479] 用于5'的TX2锤头状核酶16nt突出端(Huang等人2019)(SEQ ID NO:124)
- [0480] 5' NNNNNNNNNNNNNNNNCUGAUGAGUCCGGUAGCGGACGAAACGCGCUUCGGUGCGUC 3'
- [0481] 用于5'的RzB锤头状核酶(Saksmerprome等人,2004)(SEQ ID NO:125)
- [0482] 5' NNNNNNUAANNNNNCUGAUGAGUCGUGGGAUGCGACGAAACGCCUUCGGGCGUC 3'
- [0483] RzB(Saksmerprome等人,2004),具有对Ct-Luc特异的茎1突出端(SEQ ID NO:40)
- [0484] 5' -UUGUAAUUAUCCUGCUGAUGAGUCGUGGGAUGCGACGAAACGCCUUCGGGCGUC-3'
- [0485] Nt载体的剪接供体序列(SEQ ID NO:41)
- [0486] 5' -GUAAGUAUCAAGGUACAAGACAGGUUUAAGGAGACCAAUAGAAACUGGGCU-3'
- [0487] Ct载体的剪接受体序列(SEQ ID NO:42)
- [0488] 5' -UGUCGAGACAGAGAAGACUCUUGCGUUUCUGAUAGGCACCUAUUGGUCUACUGACAUCCACUUU
GCCUUUCUCUCCACAG-3'
- [0489] Ct载体的翻译调控序列
- [0490] GCN4 5'UTR uORF(Zhang和Hinnebusch 2011)(SEQ ID NO:43)
- [0491] 5' -AAACAAAACUCACAACACAGGUUACUCUCCCCCUAAAUUCAAAUUUUUUUGCCCAUCAGUUU
CACUAGCGAAUUAACAACUCACCAGCCACACAGCUCACUCAUCUACUUCGCAAUCAAACAAAUAUUUUUUUU
AGUUCAGUUUAUUAAGUUUAUUAUCAGUAUCGUAAUUAUUUUUUUUUUAAGAUAUUGAAAAUUGGCUUGCUAAACCGAU
UAUAUUUUUGUUUUUAAAGUAGAUUAUUUUAGAAAAUUAUUAAGAGAAUUAUGUGUUAAAUUUAUUGAAAGAGAAA
AUUUUUUUUCCCUUAUUAAUUAAGUCCUUUACUUUUUUUUGAAAACUGUCAGUUUUUUGAAGAGUUUUUUGUUUUU
UUACCAAUUGCUAUCAGUACCCGUAGAAUUUUUAUCAAGAUGUUUCCGUAACGGUUACCUUUCUGUCAAAUUAUC
CAGGUUUACUCGCCAAUAAAAUUUCCCUAUACUAUCAUUAAUUAUUUUCAUUAUUUUUACUAAAGUUUUUGUUUACC
AAUUUGUCUGCUCAAGAAAAUAAAUAUUUUUACAAAUAAA-3'
- [0492] sGCN4 5'UTR uORF(SEQ ID NO:104)
- [0493] UUAAGAUAUUGAAAAUUGGCUUGCUAAACCGAUUAUUAUUUUUGUUUUUAAAGUAGAUUAUUUUUAGAA
AAUUUAUUAAGAGAAUUAUGUGUUAAAUUUAUUGAAAGAGAAAAUUUAUUUUUCCCUUAUUAAUUAAGUCCUUUACU
UUUUUGAAAACUGUCAGUUUUUUGAAGAGUUUUUUGUUUACCAAUUGCUAUCAGUACCCGUAGAAUUUUUAU
CAAGAUGUUUCCGUAACGGUUACCU
- [0494] SRY 5'UTR uORF(Calvo等人,2009)(SEQ ID NO:44)
- [0495] 5' -GUUGAGGGGGUGUUGAGGGCGGAGAAUGCAAGUUCAUUACAAAAGUUAACGUAACAAAGAAUC
UGGUAGAAAUGAGUUUUGGAUAGUAAAUAAGUUUCGAACUCUGGCACCUUCAAUUUUUGUCGCACUCUCCUUGUU
UUUGACA-3'
- [0496] Hoxa9 TIE(Leppek等人,2020)(SEQ ID NO:45)
- [0497] 5' -GAAAAACAGAAGAGGGAAGGAUACCAGAGCGGUUCAUACAGGGCCCAGAAACUAGGCGAGGUGA
CCCCUCAGCAAGACAAAACACCUCUUGAUGUUGACUGGCGAUUUUCCCAUCUCCAGUCUGGGGAGCGGGACUAGGC
AUACAGAUGAUGGAGCUUAGAACCCGUGGCUAGGGAUAAAAUUCGUGGGCAGUUUGUCUCAAGAAGUGGGC
CAGGGCGCUUGUGACACAAUCAGGGCGUUUGUGACACAAACCCUUGAGGGUUGGAGUUCUCUCCUUGGCGGUUGC
UCUGGUUGCUCUGUGGGCCUUCCUGUGGAGCAAGGGUGAUCUGGCCGA-3'

- [0498] Hoxa3 TIE (Leppek等人,2020) (SEQ ID NO:46)
- [0499] 5'-AGGACAAUUCGUCUCUUGGGCUGCCGAAGCGACAGCUGUCAGAGAGGCAGAAGCUUCUGGGAGCCGCGGUCUGAAGGCUACGUGUGCUGCCUGGUCAUUCAAGUGUCAUUUUAGGUCCAGAAGUGUCCAAACCACAAGUUCUCAAAAACUCUGAAAAAUGGCUCCCUCC-3'
- [0500] NRAS 5'UTR G-四链体 (Kumari等人,2007) (SEQ ID NO:47)
- [0501] 5'-CGUCCCGUGUGGGAGGGGCGGGUCUGGGUGCGGCCUGC-3'
- [0502] 人IFNG 5'UTR假结 (Kaempfer 2006) (SEQ ID NO:48)
- [0503] CACAUUGUUCUGAUCAUCUGAAGAUCAGCUAUUAGAAGAGAAAGAUCAGUUAAGUCCUUUGGACCUGAUCAGCUUGAUACAAGAACUACUGAUUUAACUUCUUUGGCUUAAUUCUCUCGAAACG
- [0504] 大鼠ODC 5'UTR (Manzella和Blackshear 1990) (SEQ ID NO:49)
- [0505] 5'-UGUCAGUCCUGCAGCCGCCGCCGCCGCCUUCAGUCAGCAGCUCGGCGCCACCUCGGGUCGCGACUGCGGGGUCGACGAGGCGGCUGACGGGGCGGGCGGGGAAGACGGCCGGGUGCGCCUUG-3'
- [0506] RNA核定位信号
- [0507] SIRLOIN RNA核定位信号 (Lubelsky和Ulitsky 2018) (SEQ ID NO:50)
- [0508] 5'-CGCCUCCCGGUUCAAGCGAUUCUCCUGCCUCAGCCUCCCGAGUAGCUG-3'
- [0509] BORG lncRNA NLS (Zhang等人,2014) (SEQ ID NO:51)
- [0510] 5'-ACCUCAGAAUCUACAAGUCAGCCCCAAUAAAUGUUGUUUA-3'
- [0511] 蛋白降解氨基酸序列
- [0512] Nt或Ct载体的N-端和C-端蛋白降解序列
- [0513] FKBP DD (Banaszynski等人,2006) (SEQ ID NO:52)
- [0514] MGVQVETISPGDGRTPFKRGQTTCVVHYTGMLEDGKKVDSSRDRNKPFPKMLGKQEVIRGWEEGVAQMSV
GQRAKLTISPDYAYGATGHPGIIPPHATLVFDVELLKPE
- [0515] C-端蛋白降解序列
- [0516] PEST (增强型ODC PEST) (Li等人,1998) (SEQ ID NO:53)
- [0517] SHGFPPEVEEQAAGTLPMSCAQESGMDRHPAACASARINV*
- [0518] ODC PEST (酵母) (Rogers等人,1986) (SEQ ID NO:54)
- [0519] SHGFPPEVEEQDDGTLPMSCAQESGMDRHPAACASARINV*
- [0520] ODC PEST (人) (SEQ ID NO:55)
- [0521] NPDPPEVEEQDASTLPVSCAWESGMRHRAACASASINV*
- [0522] CL1 (Gilon等人,1998) (SEQ ID NO:56)
- [0523] ACKNWFSSLSHFVIHLNSHGFPPEVEEQAAGTLPMSCAQESGMDRHPAACASARINV*
- [0524] CL1-PEST (SEQ ID NO:57)
- [0525] ACKNWFSSLSHFVIHLNSHGFPPEVEEQAAGTLPMSCAQESGMDRHPAACASARINV*
- [0526] E1A PEST (Rogers等人,1986) (SEQ ID NO:58)
- [0527] SRECNSSTDSCDGSNTPEIHPVPLCPKPVAVRVGGRRQAVECIEDLLNEPGQPLDLSCRRPRP*
- [0528] C-myc PEST (Rogers等人,1986) (SEQ ID NO:59)
- [0529] LHEETPPTTSSDSEEEQEDEEIDVVSVEKR
- [0530] c-Fos PEST (Rogers等人,1986) (SEQ ID NO:60)
- [0531] AAHRKGSSSNPSSDSLSSPTLLAL

- [0532] v-Myb PEST(Rogers等人,1986) (SEQ ID NO:61)
- [0533] PSPPVDHGCLPEESASPARCMIVHQS
- [0534] NPDC1 PEST(SEQ ID NO:62)
- [0535] PPKELDTASSDEENEDGDFTVYECPLAPTGEMEVNPLFDHAALSAPLPAPSSPPALP
- [0536] IkBa PEST(Shumway等人,1999) (SEQ ID NO:63)
- [0537] PESEDEESYDTESEFTEFTEDELPHYDDCVFQQRLTL
- [0538] m.m.AZIN2 PEST(Lambertos和Penafiel 2019) (SEQ ID NO:64)
- [0539] GQLLPAEEDQDAEGVCKPLSCGWEITDTLCVGPVFTPASIM*
- [0540] x.1.AZIN2 PEST(Lambertos和Penafiel 2019) (SEQ ID NO:65)
- [0541] VQLLQRGLQQTEEKENVCTPMSGWEISDSLCTRTFAATSII*
- [0542] 由CRL2泛素连接酶指导的C-末端Degrons(Lin等人,2018)
- [0543] NS1(SEQ ID NO:66)
- [0544] TSLYKKVGMGRK*
- [0545] NS6(SEQ ID NO:67)
- [0546] SLYKKVGTMAAG*
- [0547] NS7(SEQ ID NO:68)
- [0548] YKKVGTMRGRGL*
- [0549] NS12(SEQ ID NO:69)
- [0550] ERAPTGRWGRRG*
- [0551] NS15(SEQ ID NO:70)
- [0552] EGPLWHPRICGS*
- [0553] SELK(SEQ ID NO:71)
- [0554] LRGPSPPMAGG*
- [0555] SELS(SEQ ID NO:72)
- [0556] WRPGRRGPSSGG*
- [0557] 由E3泛素连接酶指导的C-末端Degrons(Koren等人,2018)
- [0558] EMID1(SEQ ID NO:73)
- [0559] RDERG*
- [0560] IRX6(SEQ ID NO:74)
- [0561] GAEAG*
- [0562] 泛素Degrons(Chassin等人,2019)
- [0563] UbVR(SEQ ID NO:75)
- [0564] QIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDYNIQKESTLHLV
LRLRGVRASAS
- [0565] 2xUbVR(SEQ ID NO:76)
- [0566] TSQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDYNIQKESTLH
LVLRLRGVRASASQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDYNIQKES
TLHLVLRRLRGVRASAS
- [0567] 通过polyA尾模拟翻译的序列

[0568] 12x poly K编码尾序列(SEQ ID NO:77)

[0569] AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAA

[0570] 翻译产物12x poly K(SEQ ID NO:78)

[0571] KKKKKKKKKKKK*

[0572] 16x poly K编码尾序列(SEQ ID NO:79)

[0573] AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAA

[0574] 翻译产物16x poly K(SEQ ID NO:80)

[0575] KKKKKKKKKKKKKK*

[0576] 用于增强或阻抑核酶介导的反式剪接的酶

[0577] 人RtcB蛋白序列(SEQ ID NO:81)

[0578] MSRSYNDELQFLEKINKNCWRIKKGFPNMQVEGVFVNDALDKLMFEELRNACRGGVGGFLPAMKQ
IGNVAALPGIVHRSIGLPDVHSGYGFAIGNMAAFDMNDPEAVVSPGGVGFDFINCGVRLLRNLDES DVQPVKQLA
QAMFDHIPVGVGSKGVIPMNAKDLEEALEMVDWSLREGYAWAEDKEHCEEYGRMLQADPNKVSARAKKRGLPQLG
TLGAGNHYAEIQVVDEIFNEYAAKMGIDHKGQVCVMIHSGSRGLGHQVATDALVAMEKAMKRDKIIVNDRQLACA
RIASPEGQDYLKGMAAAGNYAWVNRSSMTFLTRQAFKVFNTTPDDLHLVYDVSJNIKVEQHVVDGKERTLLV
HRKGSTRAFPPHPLIAVDYQLTGQPVLIGGTMGTCSYVLTGTEQGMTETFGTTCHGAGRALSRAKSRRNLDFQDV
LDKLDMGIAIRVASPKLVMEAPESYKNVTDVNTCHDAGISKKAIKLRPIAVIKG*

[0579] 人RtcB人密码子优化的核酸序列(SEQ ID NO:82)

[0580] ATGTCCCGGTCATATAATGACGAGCTGCAATTCCTTGAGAAGATAAATAAGAATTGCTGGCGCATCAA
gAAAGGCTTCGTTCTAATATGCAAGTTGAAGGTGTATTTATGTAAATGACGCTTTGAAAAGTTGATGTTTCGAG
GAACTGAGGAACGCATGTCGCGGTGGaGgtGTCGGGGGTTTTCTTCCCGCTATGAAGCAGATTGGCAATGTGGCGG
CTCTGCCCCGAATTGTGCACCGCTCTATAGGATTGCCTGACGTACACAGCGGCTACGGATTGCCATTGGGAATAT
GGCGGCGTTCGATATGAACGACCCTGAGGCGGTTGTTAGCCCTGGAGGTGTCGGCTTCGATATAAATTGCGGAGTC
AGATTGCTTCGGACAAATTTGGATGAATCTGACGTACAACAGTGAAGAGCAACTTGCACAAGCGATGTTTCGATC
ATATTCCTGTTGGTGTGGGGTCAAAGGGAGTAATCCAATGAACGCGAAAAGACCTGGAAGAAGCATTGGAGATGGG
TGTAGACTGGTCACTGCGAGAAGGTTATGCCTGGGCTGAAGACAAAAGACTGCGAGGAGTACGGTTCGCATGTTG
CAAGCAGACCCAAATAAAGTATCCGCGAGGGCCAAGAAAAGAGGTTTGCCGACGCTGGGGACATTGGGGCCGGTA
ACCACTATGCAGAAATACAAGTAGTGGATGAGATTTTCAATGAGTACGCTGCGAAGAAAATGGGGATCGACCATAA
AGGTCAAGTGTGCGTAATGATACATTCTGGGAGtCGCGGACTCGGGCACCAGTTGCAACGGACGCCCTTGTCGCC
ATGGAAAAAGCGATGAAGCGGGATAAAATCATCGTAAATGATAGGCAATTGGCTTGCCTCGCATTGCGAGTCCGG
AAGGGCAAGACTACTTGAAAGGGATGGCTGCTGCCGGAATTATGCATGGGTCAACCGGAGCAGTATGACATTCTT
GACGCGGCAGGCTTTTGCAAAAGTGTTAATACGACTCCGGACGACCTCGATCTCCATGTTATATATGATGTATCA
CACAATATCGCAAAGGTTGAGCAACACGTTGTGGATGGTAAGGAAAGGACTCTGCTGGTACACCGGAAAGGCAGTA
CACGGGCAATTCGCGCTCATCACCCATTGATCGCAGTCGATTATCAATTGACAGGTCAGCCAGTTCTGATCGGAGG
AACAAATGGGCACATGTAGCTACGTATTGACCGGGACTGAACAGGGGATGACCGAAACTTTTGGCACAACATGCCAT
GGCGCGGGGAGGGCACTCTCCCGAGCTAAAAGTAGGAGGAATCTTGACTTCCAGGATGACTGGATAAGCTgGCCG
ATATGGGGATAGCCATCCGGGTAGCGTCAACCAAATGGTAATGGAGGAAGCTCCTGAAAGCTATAAAAATGTCCT
GACGTTGTCAACACATGCCATGACGCGGGTATATCCAAGAAAGCTATTAAGCTGCGCCCAATAGCTGTAATTAAGG
ATAG

[0581] 大肠杆菌RtcB蛋白序列 (SEQ ID NO:83)

[0582] MNYELLTTENAPVKMWTKGVPVEADARQQLINTAKMPFIFKHIAVMPDVHLGKGSTIGSVIPTKGAIIP
AAVGVDIGCGMNLRTALTAEDLPENLAELRQAIETAVPHGRTTGRCKRDKGAWENPPVNVDAKWAELEAGYQWLTQ
KYPRFLNTNYYKHLGLTGTGNHFIEICLDESQVWIMLHSGSRGIGNAIGTYFIDLAQKEMQETLETLP SRDLAYFM
EGTEYFDDYLKAWAWAQLFASLNRDAMMENVV TALSITQKTVRQPQLAMEEINCHHNYVQKEQHFGEIYVTRKG
AVSARAGQYGIIPGSMGAKSFIVRGLGNEESFCSCSHGAGRVM SRTKAKKLFSVEDQIRATAHVECRKDAEVIDEIP
MAYKDIDAVMAAQSDLVEVIYTLRQVVCVKG

[0583] 大肠杆菌RtcB人密码子优化的核酸序列 (SEQ ID NO:84)

[0584] ATGAATTACGAGCTTCTTACCACTGAGAATGCACCTGTGAAAATGTGGACTAAGGGAGTGCCCGTGGGA
AGCGGACGCAAGGCAGCAGCTCATAAATACAGCTAAGATGCCTTTCATCTTCAAACACATCGCGGTTATGCCCGAC
GTGCACCTCGGAAAAGGCTCTACTATTGGAAGTGTGATTCCGACAAAGGGTGCATCATACTGCTGCCGTGCGGG
TGGACATAGGCTGTGGAATGAATGCCCTGCGAACGGCTCTTACCGCAGAAGATCTTCTGAGAATCTGGCCGAGCT
GCGACAGGCCATTGAAACAGCGTTCCGCATGGTCGGACTACCGGACGGTGCAAAAAGGACAAAAGGTGCGTGGGAA
AACCCtCCCCTAACGTGGATGCGAAATGGGCTGAGTTGGAAGCAGGCTATCAATGGCTTACCCAGAAATATCCAC
GGTTCTTGAACACTAATAACTACAAACACCTGGGGACCTTGGGGACGGGAATCATTTCATCGAAATCTGTCTTGA
TGAGTCTGACCAAGTGTGGATTATGCTTCATAGCGGTAGCCGCGCATTGGTAACGCAATTGGGACATATTTTATT
GACCTCGCGCAgAAAGAGATGCAGGAAACGCTTGAGACGCTGCCGTCCCGAGATCTTGCATTTTTATGGAAGGGA
CGGAATACTTTGACGATTATCTGAAGCGGTAGCATGGGCTCAACTGTTTGCTAGTCTCAACCGAGACGCGATGAT
GGAAAATGTGGTAACAGCACTTCAATCAATCACCCAAAAGACAGTGCAGACAGCCCCAACTCTCGCTATGGAAGAA
ATCAATTGCCACCACAATTACGTTCAgAAAGAGCAACATTTCCGAGAAGAAATTTACGTGACAAGAAAAGGAGCTG
TTAGCGCGAGGGCCGGACAGTACGGCATCATTCTGGGTCAATGGGTGCGAAATCTTTATAGTACGCGGGCTTGGT
AATGAAGAATCCTTCTGCAGCTGTTCTCATGGAGCCGGAAGGGTAATGTCCAGGACTAAGGCCAAGAACTCTTCTC
TGTGGAAGATCAAATTAGAGCTACAGCACATGTTGAATGTAGAAAGGATGCCGAAGTCATAGACGAGATCCCTATGG
CTTACAAAGATATAGATGCTGTAATGGCTGCACAGTCAGACCTCGTAGAGTTATCTACACACTCCGGCAAGTCGTA
TGCGTAAAAGGATAG

[0585] 耐辐射异常球菌RtcB蛋白序列 (SEQ ID NO:85)

[0586] MNGKHITKLGFEQKAVGLALSAAGLREDAGVSRGDILDELRVQNYPEQYQGGGVYADLATHLIEQQAA
QQTRQSACLRAAPLPYRTWGEDLIEPGAHRQMDVAMQLPISRAGALMPDAHVG YGLPIGGVLATENAVIPYGVGVDI
GCSMMLSVFPVAATGLSVDEARSLLLKHTRFGAGVGF EKRDRLDHPVLA EATWDEQPLL RHLFDKAAGQIGSSSGSN
HFVEFGTFTLAQADPQLEGLDPGEYLAVLSHSGSRGFGA QVAGHFTNLAQRLWPALDKEAQKLAWLPLDSEAGQAYW
QAMNLAGRYALANHEQIHARLARALGEKPLLRAQNSHNLA WKQQVNGQELIVHRKGATPAEAGQLGLIPGSMADPGY
LVRGRGNPEALASASHGAGRQLGRKAAERSLAKKDVQAYLKDRGVTLIGGGIDEAPQAYKRIEDVIARQRDLVDVLG
EFRPRVVRMDTGSEDV

[0587] 耐辐射异常球菌RtcB人密码子优化的核酸序列 (SEQ ID NO:86)

[0588] ATGAACGGAAAGCACATCACGAAGTTGGGTTTCCGAGGAGATATTCTCGATGAACTTAGGTCTGTCCAGAATTATCCGGAG
TGTTCTCAGGGAAGACGCAGGCGTTTCCCGAGGAGATATTCTCGATGAACTTAGGTCTGTCCAGAATTATCCGGAG
CAATATCAAGGGGGAGGGGTCTATGCCGACTTGGCGACACACCTTATTGAGCAACAAGCTGCTCAGCAGACTAGGC
AATCCGCCAAGCTGCGAGCAGCACCCTTCCGTACCGAACGTGGGGTGAAGACCTGATCGAGCCAGGCGCACACAG
ACAGATGGATGTAGCAATGCAGCTCCCGATCTCCGGGCGGGAGCGCTGATGCCAGATGCCACGTAGGATACGGA

CTTCCCATTGGAGGCGTGCTCGCTACCGAAAACGCCGTAATCCCCTATGGAGTGGGCGTTGACATCGGTTGCTCAA
TGATGTTGAGTGTTCCTCCCGTGGCTGCAACAGGTCTGTCACTGAGGCGCGTCACTGCTTCTCAAACACAC
GCGCTTCGGTGCAGGGGTCGGATTCGAGAAAACGCGACAGGCTCGACCATCCTGTCTTGGCGGAGGCTACGTGGGAC
GAGCAGCCTTTGCTGAGACACTTGTGATAAAGCTGCTGGCCAGATTGGGTCTCCGGATCAGGGAACCACTTCG
TCGAATTTGGAACCTTCACCCCTCGCACAGGCCGATCCGCGAGTTGGAAGGTTTGGAcCCTGGGGAATACTTGGCTGT
TCTTTCACACTCAGGGAGTAGAGGATTTGGAGCCCAGGTGGCTGGGCATTTTACCAACTTGGCGCAGCGCTTGTGG
CCCGCACTTGATAAGGAAGCTCAAAAACCTCGCATGGCTGCCACTGGATTCTGAGGCTGGGCAAGCcTACTGGCAAG
CCATGAACTTGGCGGGACGATATGCGTTGGCTAACCATGAGCAAATTCACGCCCCACTGGCCCGCGCACTTGGTGAG
AAGCCTCTTCTGCGCGCCAGAACTCCACAATCTGGCCTGAAAACAGCAGGTGAATGGGCAGGAATTGATAGTCCA
CCGCAAAGGGGCTACTCCTGCGGAAGCCGGGCAACTTGGTCTCATCCCTGGCTCCATGGCCGACCCGGGATATTTGG
TCAGGGGAAGGGGAAATCCGGAAGCATTGGCCTCTGCGTCACACGGAGCAGGTAGACAGCTCGGCCGGAAGGCAGCG
GAAAGGTCCCTGGCGAAGAAAGATGTGCAGGCTTACCTTAAAGATAGAGGAGTAACCCCTATCGGGGGCGGGATTGA
CGAGGCTCCCCAGGCGTATAAAAGGATCGAAGACGTCATAGCACGCCAGCGGGACCTTGTGGATGTGTTGGGAGAAT
TTAGGCCACGAGTAGTGCGGATGGATACAGGGTCTGAAGATGTTTAG

[0589] 霍里科什火球菌RtcB蛋白序列(SEQ ID NO:87)

[0590] MVVPLKRIDKIRWEIPKFDKMRVPGRVYADEVLLEKMKNDRTLEQATNVAMLPGIYKYSIVMPDGHQ
GYGFPIGGVAAFDVKEGVISPGGIGYDINCGVRLIRTNLTEKEVRPRIKQLVDTLFKNVPSGVGSQRIKLHWTQI
DDVLVDGAKWAVDNGYGWERDLERLEEGRMGADPEAVSQRAKQRGAPQLGSLGSGNHFLVQVVDKIFDPEVAK
AYGLFEGQVVVMVHTGSRGLGHQVASYDLRIMERAIRKYRIPWPDRELVSVPFQSEEGQRYFSAMKAAANFAWANR
QMI THWVRESFQEVFKQDPEGDLGMDIVYDVAHNIGKVEEHEVDGKRVKIVHRKGATRAFPFGHEAVPRLYRDVG
QPVLIPGSMGTASYILAGTEGAMKETFSTCHGAGRVL SRKAATRQYRGRDRIRQELLNRGIYVRAASMRVVAEEAP
GAYKNVDNVVKVSEAGIAKLVARMRPIGVAKG*

[0591] 霍里科什火球菌RtcB人密码子优化的核酸序列(SEQ ID NO:88)

[0592] ATGGTGGTCCCCTGAAGAGAATAGATAAAAATTCGCTGGGAGATCCCTAAGTTCGACAAAAGGATGAGA
GTACCAGGACGGGTGTATGCAGATGAGGTCTTGCTCGAAAAAATGAAAAATGACCGCACGCTTGAACAGGCAACGAA
CGTCGCAATGCTGCCAGGCATTTATAAATACAGTATTGTGATGCCCGATGGCCACCAGGGGTACGGATTTCCAATTG
GAGGGGTAGCCGCTTTCGATGTTAAAGAGGGCGTAATCAGTCCTGGTGGGATCGGGTACGACATCAATTGTGGAGTC
CGACTGATCAGAACCAATCTCACTGAGAAAAGTAAGGCCAGAATCAAGCAACTGGTTGATACTCTGTTTAAAAA
CGTCCCTTCTGGAGTGGGACAGTCAAGGGCGGATTAACTGCATTGGACTCAAATAGACGATGTACTCGTAGACGGGG
CAAAATGGGCTGTGGACAACGGATATGGATGGGAGCGCGACCTCGAACGGTTGGAAGAAGGTGGTCCGATGGAGGGG
GCCGATCCAGAGGCGGTCTCCCAACGGGCAAAGCAGAGGGGAGCACCCAGCTCGGGTCCCTGGGGTCTGGCAACCA
TTTCTCGAAGTACAGGTGCTAGATAAGATCTTTGATCCTGAAGTAGCGAAAGCGTATGGCCTCTTCGAGGGGCAAG
TGGTTGTGATGGTTCACACTGGTAGCAGAGGTCTTGGGCACCAAGTTGCATCCGACTACTTGCGAATCATGGAGCGC
GCAATTAGGAAGTATAGAATCCCCTGGCCGATAGAGAGCTTGTCTCAGTCCCTTTTCAAAGCGAGGAAGGACAAAG
ATACTTCAGCGCCATGAAAGCCGCGGCAAACTTTGCATGGGCAAATCGGCAGATGATAACTCATTGGGTACGAGAAT
CATTCCAAGAGGTCTTCAAACAAGATCCGGAAGGCGACCTCGGCATGGACATTGTGTACGATGTCGCCACAATATA
GGCAAAGTGGAGGAGCACGAGGTGATGGCAAACGGGTGAAAGTTATAGTCCATCGAAAGGGAGCAACTCGCGCTTT
TCCACCAGGTACGAGGCTGTACCTAGGCTGTATCGGGATGTCGGTCAACCTGTACTCATAACCCGGATCTATGGGCA
CAGCTTCTATATTCTGGCTGGCACTGAAGGAGCAATGAAAGAGACGTTTGGATCTACCTGTCACGGAGCTGGTAGG

GTACTCTCCCGGAAGGCCGCGACACGACAATATCGCGGGGACAGGATCAGACAAGAAGCTTTTGAATAGAGGCATCTA
CGTGCGCGCCGCTAGTATGCGCGTCGTGGCCGAAGAGGCACCTGGGGCTTACAAGAACGTGGATAACGTAGTTAAAG
TAGTAAGTGAAGCCGGCATCGCCAAGCTGGTGGCCCGGATGCGCCCGATTGGCGTGGCAAAGGGTTAG

[0593] 火球菌属种ST04 RtcB蛋白序列 (SEQ ID NO:89)

[0594] MTVPLKRIDRIRWEIPKFDKMRVPGRVYADEVLIIEKMRSDRTLEQAANVAMPLPGIYKYSIVMPDGHQ
GYGFPIGGVAAFDVKEGVISPGGIGYDINCGVRLIRTNLTEKEVRPKIKQLVDTLFKNVPSGVGSQGRIRLHWTQI
DDVLVDGAKWAVDNGYGWERDLERLEEGRMGADPDVAVSQRKQKRGAPQLGSLGSGNHFLVQVVDKIYDEEVAK
AYGLFEGQVVVMVHTGSRGLGHQVASYDLRIMERAIRKYRIPWPDRELVSVPFQSEEGQRYFSAMKAAANFAWANR
QMITHWVRESFQEVFRQDPEGDLGMDIVYDVAHNIGKVEEHEVDGKKVTVIVHRKGATRAFPAGHEAIPRIYRDVG
QPVLIPGSMGTASYVLAGTEGAMKETFSTCHGAGRVLRSKAATRQYRGDRIRNELLRGIYVRAASMRVVAEEAP
GAYKNVDNVVVKVSEAGIAKLVARMRPIGVAKG*

[0595] 火球菌属种ST04 RtcB人密码子优化的核酸序列 (SEQ ID NO:90)

[0596] ATGACCGTCCCTGAAGAGAATAGATAGGATTCGCTGGGAGATCCCTAAGTTCGACAAAAGGATGAGA
GTACCAGGACGGGTGTATGCAGATGAGGTCTTGATCGAGAAAATGAGAAGCGACCGCACGCTTGAACAGGCAGCCAA
CGTCGCAATGCTGCCAGGCATTTATAAATACAGTATTGTGATGCCCGATGGCCACCAGGGGTACGGATTTCCAATTG
GAGGGGTAGCCGCTTTTCGATGTTAAAGAGGGCGTAATCAGTCTGGTGGGATCGGGTACGACATCAATTGTGGAGTC
CGACTGATCAGAACCAATCTCACTGAGAAAAGAAGTAAGGCCAAAATCAAGCAACTGGTTGATACTCTGTTTAAAAA
CGTCCCTTCTGGAGTGGGCAGTCAAGGGCGGATTAGACTGCATTGACTCAAATAGACGATGTACTCGTAGACGGGG
CAAATGGGCTGTGGACAACGGATATGGATGGGAGCGCGACCTCGAACGGTTGGAAGAAGGTGGTGGATGGAGGGG
GCCGATCCAGACGCGGTCTCCCAACGGGCAAAGCAGAGGGGAGCACCCCAGCTCGGGTCCCTGGGGTCTGGCAACCA
TTTCTCGAAGTACAGGTTCGTAGATAAGATCTACGATGAGGAAGTAGCGAAAGCGTATGGCCTCTTCGAGGGGCAAG
TGGTTGTGATGGTTCACACTGGTAGCAGAGGTCTTGGGCACCAAGTTGCATCCGACTACTTGCGAATCATGGAGCGC
GCAATTAGGAAGTATAGAATCCCCTGGCCGATAGAGAGCTTGTCTCAGTCCCTTTCAAAGCGAGGAAGGACAAAG
ATACTTCAGCGCCATGAAAGCCGCGGCAAACCTTGCATGGGCAAATCGGCAGATGATAACTCATTGGGTACGAGAAT
CATTCCAAGAGGTCTTCAGACAAGATCCGGAAGGGCACCTCGGCATGGACATTGTGTACGATGTCGCCCACAATATA
GGCAAAGTGGAGGAGCACGAGGTTCGATGGCAAGAAAGTGACCGTTATAGTCCATCGAAAGGGAGCAACTCGCGCTTT
TCCACCAGGTACGAGGCTATCCCTAGGATCTATCGGGATGTCGGTCAACCTGTACTCATACCCGGATCTATGGGCA
CAGCTTCTATGTGCTGGCTGGCACTGAAAGGAGCAATGAAAGAGACGTTTGGATCTACCTGTACGGAGCTGGTAGG
GTACTCTCCCGGAAGGCCGCGACACGACAATATCGCGGGGACAGGATCAGAAATGAAGCTTTTCAAAGAGGCATCTA
CGTGCGCGCCGCTAGTATGCGCGTCGTGGCCGAAGAGGCACCTGGGGCTTACAAGAACGTGGATAACGTAGTTAAAG
TAGTAAGTGAAGCCGGCATCGCCAAGCTGGTGGCCCGGATGCGCCCGATTGGCGTGGCAAAGGGTTAG

[0597] 热球菌属种EP1 RtcB蛋白序列 (SEQ ID NO:91)

[0598] MEIPLKRLDKIRWEIPKFNRRMRVPGRVYADDTLLQKMRQDKTLEQATNVAMPLPGIYKYSIVMPDGHQ
GYGFPIGGVAAFDVKEGVISPGGVGYDINCGVRLIRTNLVEKEVRPKIKQLIDTLFKNVPSGLGSKGRIRLHWTQL
DDVLADGAKWAVDNGYGWKDDLEHLEEGRMGEGANPNAVSVKAKQKRGAPQLGSLGSGNHFLVQVVDKVFNEEIAK
AYGLFEGQIVVMVHTGSRGLGHQVASYDLRIMEKANRKYNVPWPDRELVSVPFQTEEGQRYFSAMKAAANFAWANR
QMITHWVRESFEEVFKQKAEDLGMHIVYDVAHNIKVEEHEVNVRKIKVVVHRKGATRAFPAGHEAIPKAYRDVGQ
PVLIPGSMGTASYVLAGAEGSMRETFSTCHGAGRVLRSRHAATRQFRGDRLRNELMQRGIYIRAASMRVVAEEAPG
AYKNVDNVVRRVHEAGIANLVARMRPIGVAKG*

[0599] 热球菌属种EP1 RtcB人密码子优化的核酸序列 (SEQ ID NO:92)

[0600] ATGGAGATACCACTCAAACGACTTGACAAGATCCGATGGGAGATTCCCAAATTTAACAGACGAATGAG
AGTTCCGGGAAGAGTTTACGCAGATGATACATTGCTCCAAAAGATGCGACAAGATAAGACGCTCGAaCAAGCCACC
AACGTGGCCATGCTCCCAGGCATTTATAAGTATAGTATAGTCATGCCTGACGGACACCAGGGTTATGGATTCCCGA
TTGGCGGTGTAGCAGCCTTCGACGTAAAAGAGGGAGTAATTAGTCTGGcGGTGTGGTTATGATATTAAGTGTGG
CGTGAGGCTTATCAGGACGAATCTTGTAGAGAAGGAAGTGCACCAAAAAATCAAACAACCTTATAGATACTTTGTTC
AAAAATGTCCCGTCTGGGCTCGGATCAAAGGGTCGATAAAGGCTCCACTGGACTCAACTGGATGATGTTCTGGCTG
ATGGGGCAAAAATGGGCTGTTGACAATGGGTACGGGTGGAAGGATGATCTCGAACATTTGGAGGAGGGcGGACGGAT
GGAGGGCGCAAACCCCAATGCCGTTTCACAGAAAGCGAAGCAAAGGGGAGCGCCACAGCTTGGGTCCCTTGCTCA
GGCAATCATTTCCTCGAAAATCAGGTCGTCGATAAGGTTTTTAACGAAGAGATAGCAAAGGCTTACGGACTCTTG
AAGGTCAGATAGTGGTAATGGTCCATACGGGCTCTCGGGGACTGGGACATCAAGTCGAAGTACTACCTGAGGAT
CATGGAGAAAGCCAATCGCAAGTACAATGTGCCCTGGCCTGACCGGGAGCTTGTTAGCGTGCCTTCCAGACGGAA
GAGGGTCAACGATACTTTAGCGCTATGAAGCGGCAGCTAATTTTCGCTTGGGCAAACAGACAGATGATAACACATT
GGGTTAGAGAGTCTTCGAGGAGGTCTTTAAACAAAAAGCTGAGGACCTTGGAAATGCATATTGTCTATGATGTTGC
CCATAACATAGCAAAAAGTAGAGGAACATGAGGTGAACGGGCGGAAAATTAAGGTCGTAGTACACAGAAAAGGCGCT
ACCAGAGCATTCCCCGAGGACACGAGGCCATACCCAAAGCATATAGAGATGTCGGCCAGCCAGTgCTCATAACGG
GATCTATGGGTACGGCGTCTATGTCTTGGCGGGTGTGAAGGATCAATGAGGGAGACGTTCCGGCTCAACCTGTCA
TGGGGCAGGTCCGGTCTTGTCTCGGCATGCTGCAACTCGGCAGTTCGcGGGGATCGACTCAGGAATGAACTCATG
CAGAGAGGCATTTACATACGCGCTGCCTCCATGCGCGTTGTGCGCGAGGAAGtCCCGGCGCCTATAAGAACGTAG
ACAATGTCGTCAGGGTGGTGCATGAAGCGGGAATTGCGAACTGGTAGCCAGGATGCGCCAATAGGGGTTGCCAAG
GGATAGTAA

[0601] 人Archease蛋白序列 (SEQ ID NO:93)

[0602] MAQEEEDVRDYNLTEEQKAIKAKYPPVNRKYEYLDHTADVQLHAWGDTLEEA FEQCAMAMFGYMTDTG
TVEPLQTVEVETQGDDLQSLFLHFLDEWLYKFSADFFIPREVKVLSIDQRNFKLRSIGWGEEFSLSKHPQGTEVK
AITYSAMQVYNEENPEVFVIIDI*

[0603] 人Archease人密码子优化的核酸序列 (SEQ ID NO:94)

[0604] AGGAACAAAAGGCCATCAAAGCGAAATATCCGCCTGTAAACCGAAAAGTATGAGTACCTGGATCACACT
GCGGACGTCCAGTTGCATGCCTGGGGCGACTCTGGAGGAGGCATTGGAACAATGTGCAATGGCAATGTTTGGCT
ACATGACTGATACAGGCACAGTGGAGCCCTTCAAACGGTAGAGGTAGAAACTCAGGGAGAtGATCTTCAGAGCTT
GCTCTTCCATTTTCTCGACGAATGGTTGTATAAGTTCAGTGCCGACGAGTtTTCATTCCACGCGAAGTGAAGTG
CTGAGTATTGATCAGAGAACTTTAAACTTAGGTCTATTGGGTGGGGTGAAGAGTTCTCTTTGTCTAAACACCCTCA
AGGAACTGAGGTAAAGGCGATAACTTACTCAGCCATGCAGGTATATAACGAGGAGAATCTGAGGTTTTCGTAATCA
TTGATATATAG

[0605] 霍里科什火球菌Archease蛋白序列 (SEQ ID NO:95)

[0606] MKKWEHYEHTADIGIRGYGDSLEEA FEVAIAI LFDVMVNVNKVEKKEVRE IEVEAEDLEALLYSFLEE
LLVIHDIEGLVFRDFEVKIERVNGKYRLRAKAYGEKLDLKKHEPKKEVKAITYHDMKIERLPNGKWMALVPDI*

[0607] 霍里科什火球菌Archease人密码子优化的核酸序列 (SEQ ID NO:96)

[0608] ATGAAGAAATGGGAGCACTATGAGCATACTGCCGACATTGGTATTCCGGGATATGGGGATAGCCTTGAG
GAGGCATTGGAAGCAGTAGCCATCGCGCTCTTTGATGTAATGGTGAACGTGAATAAAGTCGAGAAGAAGGAAGTCCG

AGAAATTGAAGTGGAGGCAGAAGATTTGGAGGCCCTCCTTTATTCATTCCTGGAAGAAGTGTGGTTATTCATGATA
TAGAGGGACTGGTTTTTCAGGGACTTTGAAGTTAAGATAGAGAGAGTAAATGGCAAATACCGACTTCGAGCGAAAGCC
TACGGTGAGAAGCTCGACCTCAAGAAGCACGAACCGAAAGAGGAAGTAAAGGCGATAACCTACCATGATATGAAAAT
TGAACGGTTGCCCAATGGAAAAGTGGATGGCTCAACTCGTTCCAGATATTTAG

[0609] T4多核苷酸激酶(T4 PNK)蛋白序列(SEQ ID NO:97)

[0610] MKKIILTIGCPGSGKSTWAREFIAKNPGFYINRDDYRQSIMAHEERDEYKYTKKKEGIVTGMQFDTA
KSILYGGDSVKGVIISDTNLPERRLAWETFAKEYGWKVEHKVFDVPWTELVKRNSKRGTKAVPIDVLRSMYKSMR
EYLGLPVYNGTPGPKAVIFDVGTLAKMNGRGPYDLEKCDTDVINPMVVVLSKMYALMGYQIVVVSRESGTEKED
PTKYRMRTRKWVEDIAGVPLVMQCQREQGDTRKDDVVKEEIFWKHIAPHFDVKLAIDDRITQVEMWRRIGVECWQV
ASGDF*

[0611] T4 PNK人密码子优化的核酸序列(SEQ ID NO:98)

[0612] ATGAAGAAAATTATACTTACAATCGGATGCCCTGGTAGTGGTAAGAGCACTTGGGCGAGGGAATTTAT
TGCGAAgAACCCtGGATTTTATAATATCAATCGAGACGACTACCGGCAGTCTATTATGGCCACGAGGAACGAGAC
GAATACAAGTATACCAAGAAGAAAGAAGGGATTGTCACGGGTATGCAATTTGACACCGCCAAATCAATACTGTACG
GAGGTGATTCAGTCAAAGGCGTTATCATATCAGACACTAACCTCAATCCTGAACGCCGATTGGCATGGGAAACATT
TGCGAAGGAATACGGTTGGAAGTTGAACACAAGGTGTTTCGATGTCCCGTGGACCGAACTGGTAAAACGCAATTCT
AAACGAGGCACTAAAGCTGTGCCATTGACGTACTTCGAAGTATGTACAAGTCCATGAGAGAGTACCTGGGGCTTC
CCGTCTATAACGGTACGCCGGGCAAACCGAAGGCGGTGATCTTTGACGTAGATGGGACTCTGGCGAAGATGAATGG
TCGCGGACCATAACGATTTGAAAAAATGTGACACAGATGTAATCAACCCAATGGTAGTAGAGCTTAGCAAGATGTAC
GCATTGATGGGcTACCAAATGTCTGGTGTCCGGGCGGGAGTCAGGCACAAAAGAAGATCCGACGAAGTATTATC
GCATGACACGGAAATGGGTGCAAGATATAGCCGGGTgCCTCTCGTTATGCAATGTCAACGAGAACAGGGCGACAC
ACGGAAGGATGACGTAGTGAAGGAGGAAATTTCTGGAAGCATATAGCGCCACACTTTGACGTTAAGCTCGCCATCG
ACGACCGAACTCAGGTGGTCGAGATGTGGCGACGAATTGGCGTAGAGTGTGGCAAGTTGCATCTGGAGATTTTTAG

[0613] 大肠杆菌thpR蛋白序列(SEQ ID NO:99)

[0614] MSEPQRLFFAIDLPAEIREQI IHWRATHFPPEAGRPVAADNLHLTLAFLGEVSAEKEKALSLLAGRIR
QPGFTLTLDDAGQWLRSRVWLGMRQPPRGLIQLANMLRSQAARSGCFQSNRPFPHITLLRDASEAVTI PPPGFN
WSYAVTEFTLYASSFARGRTRYTPLKRWALTQ*

[0615] 大肠杆菌thpR人密码子优化的核酸序列(SEQ ID NO:100)

[0616] ATGAGTGAGCCTCAACGATTGTTCTTTGCCATAGATTTGCCTGCTGAAATTAGAGAGCAAATTATCCAT
TGGAGAGCCACCCATTTCCCCCAGAAGCTGGACGACCAGTCGCAGCGGACAACCTCCACCTTACACTGGCGTTCTT
GGGTGAAGTGAGCGCCGAGAAAGAGAAAGCTCTCTCACTTCTGGCTGGGAGGATTCCGGCAGCCGGGCTTTACCCTTA
CTCTGGATGATGCCGGCCAGTGGCTGAGGTCCAGGGTTGTCTGGCTCGGAATGAGGCAACCACCTAGGGGGCTCATC
CAGCTCGCCAATATGCTGAGATCCCAGGCCGAAGGTCTGGCTGCTTCCAATCAAACAGGCCATTCCACCCGCATAT
TACCTTGCTCAGAGATGCCTCCGAGGCAGTAACTATTCCACCTCCCGGCTTTAACTGGAGTTACGCCGTCACAGAAT
TACTCTGTACGCCTCCAGCTTCGCCCCAGGGAGAACCAGGTACACGCCTTTGAAGCGGTGGGCCTTGACCCAGTAG

[0617] 人PNKP蛋白序列(SEQ ID NO:101)

[0618] MGEVEAPGRLWLESPPGGAPP IFLPSDGQALVLRGPLTQVTDKCSRTQVELVADPETRTVAVKQLG
VNPSTTGTQELKPGLESLGVDLTYLVNGLHPLTLRWEETRTPEQPDPPTPPGTPLVSQDEKRDAELPKRMRKSN
PGWENLEKLLVFTAAGVKPQGVAGFDLDTLITRSGKVFPTGPSDWRI LYPEIPRKLRELEAEGYKLVIFTNQM

SIGRGKLPAAEFKAKVEAVVEKLGVPFQVLVATHAGLYRKPVTGMWDHLQEQANDGTPISIGDSIFVGDAAGRPAN
WAPGRKKKDFSCADRLFALNLGLPFATPEEFFLKWPAAGFELPAFDPRTVSRSGPLCLPESRALLSASPEVVAVG
FPGAGKSTFLKKHLVSAGYVHVNRDTLGSWQRCVTTTETALKQGKRVAIDNTNPDAASRARYVQCARAAGVPCRCF
LFTATLEQARHNNRFREMTDSSHIPVSDMVMYGYRKQFEAPTLAEGFSAILEIPFRLWVEPRLGRLYCQFSEG*

[0619] 人PNKP人密码子优化的核酸序列(SEQ ID NO:102)

[0620] ATGGGCGAGGTGGAGGCCCCGGGCCGCTTGTGGCTCGAGAGCCCCCTGGGGGAGCGCCCCCATCTTC
CTGCCCTCGACGGGCAAGCCTGGTCCTGGGCAGGGGACCCCTGACCCAGGTTACGGACCGGAAGTGCTCCAGAAC
TCAAGTGGAGCTGGTCGCAGATCCTGAGACCCGGACAGTGGCAGTGAAACAGCTGGGAGTTAACCCCTCAACTACCG
GGACCCAGGAGTTGAAGCCGGGTTGGAGGGCTCTCTGGGGGTGGGGGACACACTGTATTTGGTCAATGGCCTCCAC
CCACTGACCCTGCGCTGGGAAGAGACCCGCACACCAGAATCCCAGCCAGATACTCCGCCTGGCACCCCTCTGGTGTC
CCAAGATGAGAAGAGAGATGCTGAGCTGCCGAAGAAGCGTATGCGGAAGTCAAACCCCGGCTGGGAGAACTTGGAGA
AGTTGCTAGTGTTACCCGAGCTGGGGTGAAACCCAGGGCAAGGTGGCTGGCTTTGATCTGGACGGGACGCTCATC
ACCACACGCTCTGGGAAGGTCTTTCCACTGGCCCCAGTGACTGGAGGATCTTGTACCCAGAGATTCCCCGTAAGCT
CCGAGAGCTGGAAGCCGAGGGCTACAAGCTGGTGATCTTACCAACCAGATGAGCATCGGGCGCGGAAGCTGCCAG
CCGAGGAGTTCAAGGCCAAGGTGGAGGCTGTGGTGGAAGCTGGGGGTCCCCTTCCAGGTGCTGGTGCCACGCAC
GCAGGCTTGTACCGGAAGCCGGTGACGGGCATGTGGGACCATCTGCAGGAGCAGGCCAACGACGGCAGGCCATATC
CATCGGGACAGCATCTTTGTGGGAGACGCAGCCGGACGCCCGCCAACCTGGGCCCCGGGGCGGAAGAAGAAAGACT
TCTCCTGCGCCGATCGCCTGTTTGCCTCAACCTTGGCCTGCCCTTCGCCACGCCTGAGGAGTTCTTTCTCAAGTGG
CCAGCAGCCGGCTTCGAGCTCCCAGCCTTTGATCCGAGGACTGTCTCCCGCTCAGGGCCTCTCTGCCTCCCCGAGTC
CAGGGCCCTCCTGAGCGCCAGCCGGAGGTGGTTGTGCAGTGGGATCCCTGGGGCCGGGAAGTCCACCTTTCTCA
AGAAGCACCTCGTGTGCGCCGGATATGTCCACGTGAACAGGGACACGCTAGGCTCCTGGCAGCGCTGTGTGACCAGC
TGTGAGACAGCCCTGAAGCAAGGGAAACGGGTGCCATCGACAACACAAACCCAGACGCCGAGCCGCGCCAGGTA
CGTCCAGTGTGCCGAGCCGCGGGCGTCCCCTGCCGCTGCTTCTCTTACCCGCACTCTGGAGCAGGCGCGCCACA
ACAACCGGTTTCGAGAGATGACGGACTCCTCTCATATCCCCGTGTCAGACATGGTCATGTATGGCTACAGGAAGCAG
TTCGAGGCCCAACGCTGGCTGAAGGCTTCTCTGCCATCCTGGAGATCCCCTTCCGGCTATGGGTGGAGCCGAGGCT
GGGGCGGCTGTACTGCCAGTTCTCCGAGGGCTAG

[0621] 具有含货物和不含货物的内部合成核酶内含子的GFP

[0622] NtGFP-HDV-HH-CtGFP (SEQ ID NO:103)

[0623] AUGGUGAGCAAGGGCGAGGAGCUGUUCACCGGGUGGUGCCCAUCCUGGUCGAGCUGGACGGCGACGU
AAACGGCCACAAGUUCAGCGUUGCCGGCGAGGGCGAUGCCACCUACGGCAAGCUGACCCUGAAGUUAUC
UGCACCACCGCAAGCUGCCCGUGCCUGGCCACCCUCGUGACCACCCUGACCUACGGCGUGCAGUGCUUCAGCC
GCUACCCCGACCACAUGAAGCAGCACGACUUCUUAAGUCCGCAUGCCCGAAGGCUACGUCCAGGAGCGCACCAU
CUUCUUGggccggcauggucccagccuccucgcuggcgccggcugggcaacaugcuucggcauggcgaaugggaccc
cgggacauaacuaguuaaaccaaaucuuugcugaugaguccgugaggacgaaacgaguaagcucgucCAAGGACGA
CGGCAACUACAAGACCCGCGCCGAGGUGAAGUUCGAGGGCGACACCCUGGUGAACCGCAUCGAGCUGAAGGGCAUCG
ACUUAAGGAGGACGGCAACAUCUGGGGCACAAGCUGGAGUACAACUACAACAGCCACAACGUCUAUAUCAUGGCC
GACAAGCAGAAGAACGGCAUCAAGGUGAACUUAAGAUCGCCACAACAUCGAGGACGGCAGCGUGCAGCUCGCCGA
CCACUACCAGCAGAACACCCCAUCGGCGACGGCCCCGUGCUGCUGCCCGACAACCACUACCUGAGCACCCAGUCCG
CCCUGAGCAAAGACCCCAACGAGAAGCGGAUCACAUGGUCCUGCUGGAGUUCGUGACCGCCGCGGGAUCACUCUC

GGCAUGGACGAGCUGUACAAGUAG

[0624] NtGFP-HDV-CARGO-HH-CtGFP (SEQ ID NO:126)

[0625] AUGGUGAGCAAGGGCGAGGAGCUGUUCACCGGGGUGGUGCCCAUCCUGGUCGAGCUGGACGGCGACGU
AAACGGCCACAAGUUCAGCGUGUCCGGCGAGGGCGAGGGCGAUGCCACCUACGGCAAGCUGACCCUGAAGUUAUC
UGCACCACCGGCAAGCUGCCCGUGCCCGGGCCACCCUCGUGACCACCCUGACCUACGGCGUGCAGUGCUUCAGCC
GCUACCCCGACCACAUGAAGCAGCACGACUUCUUAAGUCCGCCAUGCCCGAAGGCUACGUCCAGGAGCGCACCAU
CUUCUUGgcccggcauggucccagccuccucgcuggcgccggcugggcaacaugcuucggcauggcgaaugggacNu
ccuugcugaugaguccgugaggacgaaacgaguaagcucgucCAAGGACGACGGCAACUACAAGACCCGCGCCGAG
GUGAAGUUCGAGGGCGACACCCUGGUGAACCGCAUCGAGCUGAAGGGCAUCGACUUAAGGAGGACGGCAACAUCU
GGGGCACAAGCUGGAGUACAACUACAACAGCCACAACGUCUAUAUCAUGGCCGACAAGCAGAAGAACGGCAUCAAGG
UGAACUUAAGAUCGCCACAACAUCGAGGACGGCAGCGUGCAGCUCGCCACCACUACCAGCAGAACACCCCAUC
GGCGACGGCCCCGUGCUGCUGCCCGACAACCACUACCUGAGCACCCAGUCCGCCUGAGCAAAGACCCCAACGAGAA
GCGCGAUCACAUGGUCCUGCUGGAGUUCGUGACCGCCGCCGGGAUCACUCUCGGCAUGGACGAGCUGUACAAGUAG

[0626] NtGFP-HDV (SEQ ID NO:127)

[0627] AUGGUGAGCAAGGGCGAGGAGCUGUUCACCGGGGUGGUGCCCAUCCUGGUCGAGCUGGACGGCGACGU
AAACGGCCACAAGUUCAGCGUGUCCGGCGAGGGCGAGGGCGAUGCCACCUACGGCAAGCUGACCCUGAAGUUAUC
UGCACCACCGGCAAGCUGCCCGUGCCCGGGCCACCCUCGUGACCACCCUGACCUACGGCGUGCAGUGCUUCAGCC
GCUACCCCGACCACAUGAAGCAGCACGACUUCUUAAGUCCGCCAUGCCCGAAGGCUACGUCCAGGAGCGCACCAU
CUUCUUGgcccggcauggucccagccuccucgcuggcgccggcugggcaacaugcuucggcauggcgaaugggac

[0628] HH-CtGFP (SEQ ID NO:128)

[0629] uccuugcugaugaguccgugaggacgaaacgaguaagcucgucCAAGGACGACGGCAACUACAAGACC
CGCGCCGAGGUGAAGUUCGAGGGCGACACCCUGGUGAACCGCAUCGAGCUGAAGGGCAUCGACUUAAGGAGGACGG
CAACAUCUGGGGCACAAGCUGGAGUACAACUACAACAGCCACAACGUCUAUAUCAUGGCCGACAAGCAGAAGAACG
GCAUCAAGGUGAACUUAAGAUCGCCACAACAUCGAGGACGGCAGCGUGCAGCUCGCCGACCACUACCAGCAGAAC
ACCCCAUCGGCGACGGCCCCGUGCUGCUGCCCGACAACCACUACCUGAGCACCCAGUCCGCCUGAGCAAAGACCC
CAACGAGAAGCGCGAUCACAUGGUCCUGCUGGAGUUCGUGACCGCCGCCGGGAUCACUCUCGGCAUGGACGAGCUGU
ACAAGUAG

[0630] 本文引用的每一个专利、专利申请和出版物的公开内容在此通过引用整体并入本文。虽然已参考特定实施方案公开了本发明,但很明显,在不脱离本发明的真实精神和范围的情况下,本领域的其他技术人员可以设计出本发明的其他实施方案和变型。所附权利要求书旨在被解释为包括所有此类实施方案和等同变化。

- [0001] 序列表
- [0002] <110> 罗切斯特大学
- [0003] 道格拉斯·马修·安德松
- [0004] <120> 核酶介导的RNA组装和表达
- [0005] <130> 204606-0127-00W0
- [0006] <150> US 62/971,356
- [0007] <151> 2020-02-07
- [0008] <160> 130
- [0009] <170> PatentIn version 3.5
- [0010] <210> 1
- [0011] <211> 302
- [0012] <212> RNA
- [0013] <213> 人工序列
- [0014] <220>
- [0015] <223> Nt-GFP
- [0016] <400> 1
- [0017] auggugagca agggcgagga gcuguucacc gggguggugc ccauccuggu cgagcuggac 60
- [0018] ggcgacguaa acggccacaa guucagcgug uccggcgagg gcgagggcga ugccaccuac 120
- [0019] ggcaagcuga cccugaaguu caucugcacc accggcaagc ugcccugucc cuggcccacc 180
- [0020] cucgugacca cccugaccua cggcgugcag ugcuucagcc gcuaccccga ccacaugaag 240
- [0021] cagcacgacu ucuucaaguc cgccaugccc gaaggcuacg uccaggagcg caccaucuuc 300
- [0022] uu 302
- [0023] <210> 2
- [0024] <211> 421
- [0025] <212> RNA
- [0026] <213> 人工序列
- [0027] <220>
- [0028] <223> Ct-GFP
- [0029] <400> 2
- [0030] caaggacgac ggcaacuaca agaccgcgc cgaggugaag uucgagggcg acaccuggu 60
- [0031] gaaccgcauc gacgugaagg gcaucgacuu caaggaggac ggcaacauc uggggcacia 120
- [0032] gcuggaguac aacuacaaca gccacaacgu cuauaucaug gccgacaagc agaagaacgg 180
- [0033] caucaaggug aacuuaaga uccgccacaa caucgaggac ggcagcgugc agcucgccga 240
- [0034] ccacuaccag cagaacaccc ccaucggcga cggccccgug cugcugcccc acaaccacua 300
- [0035] ccugagcacc caguccgcc ugagcaaaga cccaacgag aagcgcgauc acaugguccu 360
- [0036] gcuggaguuc gugaccgccg ccgggaucac ucucggcaug gacgagcugu acaaguagua 420
- [0037] a 421
- [0038] <210> 3
- [0039] <211> 831
- [0040] <212> RNA
- [0041] <213> 人工序列

- [0042] <220>
- [0043] <223> Nt-荧光素酶
- [0044] <400> 3
- [0045] auggaagacg ccaaaaacau aaagaaagc cgggcgccau ucuaucgcu ggaagaugga 60
- [0046] accgcuggag agcaacugca uaaggcuau aagagauacg ccugguucc uggaacaauu 120
- [0047] gcuuuuacag augcacauau cgagguggac aucacuuacg cugaguacuu cgaaaugucc 180
- [0048] guucgguugg cagaagcuau gaaacgauau gggcugaaua caaauacag aaucgucgua 240
- [0049] ugcagugaaa acucucuua auucuuuau cggguuugg gcgcguuuuu uaucggaguu 300
- [0050] gcaguugcgc ccgcaacga cauuuuaau gaacgugaau ugcucaacag uaugggcauu 360
- [0051] ugcagccua ccgugguuu cguuuccaaa aagggguugc aaaaauuuu gaacgugcaa 420
- [0052] aaaaagcucc caaucaucca aaaaauuuu aucauggauu cuaaaacgga uuaccagga 480
- [0053] uuucagucga uguacacguu cgucacauca caucuaccuc ccgguuuuaa ugaauacgau 540
- [0054] uuuguccag aguccuucga uaggacaag acaauugcac ugaucaugaa cuccucugga 600
- [0055] ucuacugguc ugccuaaagg ugucgcucug ccucauagaa cugccugcgu gagauucug 660
- [0056] caugccagag auccuuuuu uggcaucaa aucauuccg auacugcga uuuuaguguu 720
- [0057] guccaaucc aucacgguu uggaauguu acuacacug gauuuugau auguggauuu 780
- [0058] cgagucgucu uaauguauag auuugaaga gagcuguuuc ugaggagccu u 831
- [0059] <210> 4
- [0060] <211> 825
- [0061] <212> RNA
- [0062] <213> 人工序列
- [0063] <220>
- [0064] <223> Ct-荧光素酶
- [0065] <400> 4
- [0066] caggauuaca agaucaaaag ugcgcugcug gugccaacc uauucuccu cuucgcaaaa 60
- [0067] agcacucuga uugacaaaau cgauuuauu auuuuacacg aaauugcuuc ugguggcgc 120
- [0068] cccucucua aggaagucgg ggaagcguu gccaaaggu uccaucugcc agguaucagg 180
- [0069] caaggauaug ggcucacuga gacuacauca gcuauucuga uuacaccga gggggaugau 240
- [0070] aaaccggcg cggucgguu aguuguucca uuuuuugaag cgaagguugu ggaucuggau 300
- [0071] accgggaaa cgcuggcgu uaaucuaaga ggcgaacgu gugugagagg uccuaugau 360
- [0072] auguccguu auguaaaca uccggaagc accaacgcc ugaugacaa ggaugaugg 420
- [0073] cuacauucg gagacauagc uuacugggac gaagacgaac acuucuucg cguugaccg 480
- [0074] cugaagucuc ugauuaagua caaaggcuau cagguggcuc ccgcugaau ggaaucauc 540
- [0075] ugcuccaac acccaacau cuucgacga ggugucgag gucuuccga cgaugacgc 600
- [0076] ggugaacuuc ccgccccgu uguuuuuug gagcacgaa agacgaugac ggaaaaagag 660
- [0077] aucuggauu acgucgccag ucaaguaaca accgcgaaa aguugcgcgg aggaguug 720
- [0078] uuugggacg aaguaccgaa aggucuuacc ggaaaacuc acgcaagaaa aaucagagag 780
- [0079] auccucauaa aggccaagaa gggcgaaaag aucgccguu aguaa 825
- [0080] <210> 5
- [0081] <211> 543
- [0082] <212> DNA
- [0083] <213> 人工序列

[0084] <220>
 [0085] <223> NIL
 [0086] <400> 5
 [0087] atgggtcagg ccaatacgcc ctggagcagt aaggcaaagc cggatgcctt tataaattca 60
 [0088] ttcatcagtg cagcatccaa tactggttcc ttctctcaag accaaatgga ggacatgtca 120
 [0089] ctcatcgga atactctgat ggctgccatg gacaatatgg gaggccgcat aacaccatct 180
 [0090] aagttgcagg cgttggatat ggcttcgca tcatcagtg cagagatgc ggctagttag 240
 [0091] ggcggcgact tgggagtcac taccaacgag atcgcggatg ccctcacttc tgctttttat 300
 [0092] caaacgaccg gggttgtcaa ttcacgattc atatctgaga tcaggagcct cataggaatg 360
 [0093] ttccgagcagg cttccgaaa tgacgtttat gcatctgctg gctctggcag cgggggtggt 420
 [0094] gggatggag ccagctcagc atctgcggct tctgcaagt ctgctgccc gagtggcgta 480
 [0095] gcttatcagg ctctgctca ggctcaaac agttttacgt tgcgagggca acaacctgtt 540
 [0096] tcc 543
 [0097] <210> 6
 [0098] <211> 132
 [0099] <212> DNA
 [0100] <213> 人工序列
 [0101] <220>
 [0102] <223> AQ
 [0103] <400> 6
 [0104] ggtccttatg gaccgggtgc tagcgtcgc gcagcagccg ctggcggtta tggcccaggt 60
 [0105] tcagggaac aggggcctgg gcaacaagga cctggccaac aaggtcctgg tcagcagggt 120
 [0106] ccaggcagc ag 132
 [0107] <210> 7
 [0108] <211> 450
 [0109] <212> DNA
 [0110] <213> 人工序列
 [0111] <220>
 [0112] <223> NR3
 [0113] <400> 7
 [0114] ggcgctgctt ccgctgcagt atcagtaggt ggctatggac ctcaatctag tagcgccect 60
 [0115] gttgcctctg ccgcccatc tcgactttca agtcccgcg ctagttccag ggtcagttcc 120
 [0116] gcggtatcta gcttgtaag tagcggacc actaatcaag cggcacttc aaacacaata 180
 [0117] tcctcagtag tcagtcaagt aagcgcacaa aaccctggct tgtcagggtg tgacgttctg 240
 [0118] gttcaggcac ttctggaagt tgtctcagc ttgtaagca tcctgggtag ctctccata 300
 [0119] ggtcaaatta attatggcgc gagcgcccaa tacacacaaa tggtgggtca gactgtggcg 360
 [0120] caggcactcg caggcacta caaggatcat gacggagact ataaggatca tgatatagat 420
 [0121] tacaaggacg atgatgacaa ggcctagtaa 450
 [0122] <210> 8
 [0123] <211> 453
 [0124] <212> RNA
 [0125] <213> 人工序列

- [0126] <220>
[0127] <223> Nt-4xMTS
[0128] <400> 8
[0129] augagugugu ugacgccguu gcuucugcga gggcuuaccg ggucugcuag aagacuuccg 60
[0130] guccccaggg ccaagauaca uagccucgga gacccgaugu cugugcucac uccucugcuu 120
[0131] uugcgaggac ugacuggguc cgccagacga cucccggugc cgagagcuaa aauccauagc 180
[0132] cugggaaaau uggcaacuau gucaguccug acgccgcuuu uucuccgggg ucuuacaggg 240
[0133] ucugcaagaa ggcugccugu accucgggcg aaaauucaua gcuugggcga cccgaugagu 300
[0134] guauugacgc ccuguuugcu gagaggauug acugggucag cgcgccggcu ccugucuccc 360
[0135] cgagcuaaga uucacucccu ugguaagcug agaauccucc aaucacgggu uccgagagca 420
[0136] agagaucgcg cggucgccac gaggccucuc gag 453
[0137] <210> 9
[0138] <211> 68
[0139] <212> RNA
[0140] <213> 人工序列
[0141] <220>
[0142] <223> HDV68
[0143] <400> 9
[0144] ggccggcaug gucccagccu ccucgcuggc gccggcuggg caacaugcuu cggcauggcg 60
[0145] aaugggac 68
[0146] <210> 10
[0147] <211> 67
[0148] <212> RNA
[0149] <213> 人工序列
[0150] <220>
[0151] <223> HDV67
[0152] <400> 10
[0153] gggucggcau ggcaucucca ccuccucgcg guccgaccug ggcuacuucg guaggcuaag 60
[0154] ggagaag 67
[0155] <210> 11
[0156] <211> 56
[0157] <212> RNA
[0158] <213> 人工序列
[0159] <220>
[0160] <223> HDV56
[0161] <400> 11
[0162] gagggauagu acagagccuc cccguggcuc ccuuggauaa ccaacugaua cuguac 56
[0163] <210> 12
[0164] <211> 87
[0165] <212> RNA
[0166] <213> 人工序列
[0167] <220>

- [0168] <223> 基因组HDV
[0169] <400> 12
[0170] ggccggcaug gucccagccu ccucgcuggc gccggcuggg caacauuccg aggggaccgu 60
[0171] ccccucggua auggcgaaug ggaccca 87
[0172] <210> 13
[0173] <211> 91
[0174] <212> RNA
[0175] <213> 人工序列
[0176] <220>
[0177] <223> 抗基因组HDV
[0178] <400> 13
[0179] gggucggcau ggcaucucca ccuccucgcg guccgaccug ggcauccgaa ggaggacgca 60
[0180] cguccacucg gauggcuaag ggagagccac u 91
[0181] <210> 14
[0182] <211> 144
[0183] <212> RNA
[0184] <213> 人工序列
[0185] <220>
[0186] <223> VS核酶
[0187] <400> 14
[0188] gcgguaguaa gcaggaacu caccuccaau uucaguacug aaauugucgu agcaguugac 60
[0189] uacuguuaug ugauugguag aggcuagug acgguauugg cguaagucag uauugcagca 120
[0190] cagcacaagc ccgcuugcga gaau 144
[0191] <210> 15
[0192] <211> 21
[0193] <212> RNA
[0194] <213> 人工序列
[0195] <220>
[0196] <223> VS-S
[0197] <400> 15
[0198] gaaggcguc gucgccccga g 21
[0199] <210> 16
[0200] <211> 144
[0201] <212> RNA
[0202] <213> 人工序列
[0203] <220>
[0204] <223> VS-Rz
[0205] <400> 16
[0206] gcgguaguaa gcaggaacu caccuccaau uucaguacug aaauugucgu agcaguugac 60
[0207] uacuguuaug ugauugguag aggcuagug acgguauugg cguaagucag uauugcagca 120
[0208] cagcacaagc ccgcuugcga gaau 144
[0209] <210> 17

- [0210] <211> 291
- [0211] <212> RNA
- [0212] <213> 人工序列
- [0213] <220>
- [0214] <223> Nt-DTA
- [0215] <400> 17
- [0216] auggaccccg acgacguggu ggacagcagc aagagcuucg ugauggagaa cuucagcagc 60
- [0217] uaccacggca ccaagcccgg cuacguggac agcauccaga agggcaucca gaagcccaag 120
- [0218] agcggcacc caggcaacua cgacgacgac uggaagggu ucuacagcac cgacaacaag 180
- [0219] uacgacgcug ccggcuacag cguggacaac gagaaccccc ugagcggcaa ggccggcggc 240
- [0220] guggugaagg ugaccuaccc cggccugacc aaggugcugg ccugaaggu g 291
- [0221] <210> 18
- [0222] <211> 297
- [0223] <212> RNA
- [0224] <213> 人工序列
- [0225] <220>
- [0226] <223> Ct-DTA
- [0227] <400> 18
- [0228] gacaaugccg agaccaucaa gaaggagcug ggccugagcc ugaccgagcc ccugauggag 60
- [0229] caggugggca ccgaggaguu caucaagaga uucggcgacg gcgccagcag aguggugcug 120
- [0230] agccugcccu ucgccgaggg cagcagcagc guggaguaca ucaacaacug ggagcaggcc 180
- [0231] aaggcccuga gcguggagcu ggagaucaac uucgagacca gaggcaagag aggccaggac 240
- [0232] gccauguacg aguacauggc ccaggcuucg gccggcaaca gagugagaag auaguaa 297
- [0233] <210> 19
- [0234] <211> 717
- [0235] <212> RNA
- [0236] <213> 人工序列
- [0237] <220>
- [0238] <223> GFPcdn (无起始ATG密码子)
- [0239] <400> 19
- [0240] guuagcaagg gcgaggagcu cuucaccggg gucguccca uccucgucga gcucgacggc 60
- [0241] gacguaaacg gccacaaguu cagcgucucc ggcgaggcg agggcgaugc caccuacggc 120
- [0242] aagcucaccc ugaaguucan cuccaccacc ggcaagcugc ccgugcccug gcccaccuc 180
- [0243] gugaccaccc ugaccuacgg cgugcaguc uucagccgc accccgacca caugaagcag 240
- [0244] cagcacuuc ucaaguccgc caugcccga ggcuacguc aggagcgcac caucuucuc 300
- [0245] aaggacgacg gcaacuacaa gaccgcgcc gaggugaagu ucgaggcgca caccucggug 360
- [0246] aaccgcaucg agcugaagg caucgacuuc aaggaggacg gcaacaucuu ggggcacaag 420
- [0247] cuggaguaca acuaaacag ccacaacguc uauaucaugg ccgacaagca gaagaacggc 480
- [0248] aucaagguga acuucaagau ccgccacaac aucgaggacg gcagcugca gcucgccgac 540
- [0249] cacuaccagc agaacacccc caucggcgac ggccccgugc ugcugcccga caaccacuac 600
- [0250] cugagcacc aguccgccu gagcaaagac cccaacgaga agcgcgauca cauggucug 660
- [0251] cuggaguucg ugaccgccgc cgggaucacu cucggcaugg acgagcugua caaguag 717

- [0252] <210> 20
- [0253] <211> 131
- [0254] <212> RNA
- [0255] <213> 人工序列
- [0256] <220>
- [0257] <223> F2-Myr
- [0258] <400> 20
- [0259] auggguuguu guuucagcaa gacagcggcg aaaggugaag cagcagcaga aagaccaggc 60
- [0260] gaggcugcgg uagcaucaag ucccuccaag gcuaaugggc aggaaaacgg acacgucaaa 120
- [0261] guuggaagcg u 131
- [0262] <210> 21
- [0263] <211> 685
- [0264] <212> RNA
- [0265] <213> 人工序列
- [0266] <220>
- [0267] <223> F2-RFP
- [0268] <400> 21
- [0269] agccaucauc aaggaguuca ugcgcuucaa ggugcacaug gagggcuccg ugaacggcca 60
- [0270] cgaguucgag aucgagggcg agggcgaggg cgccccuac gagggcacc agaccgcaa 120
- [0271] gcugaaggug accaaggug gccccugcc cuucgccugg gacauccugu cccucaguu 180
- [0272] cauguacggc uccaaggccu acgugaagca ccccgccgac auccccgacu acuugaagcu 240
- [0273] guccuucucc gagggcuuca agugggagcg cgugaugaac uucgaggacg gcggcguggu 300
- [0274] gaccgugacc caggacuccu cccugcagga cggcgaguuc aucuacaagg ugaagcugcg 360
- [0275] cggcaccaac uccccuccg acggccccgu aaugcagaag aagaccaugg gcugggaggc 420
- [0276] cuccuccgag cggauguacc cggaggacgg cgcccugaag ggcgagauca agcagaggcu 480
- [0277] gaagcugaag gacggcgcc acuacgacg ugaggucaag accaccuaca aggccaagaa 540
- [0278] gcccugcag cugcccggcg ccuacaacgu caacaucaag uggacauca ccuccacaa 600
- [0279] cgaggacuac accaucgugg aacaguacga acgcgccgag ggccgccacu ccaccggcg 660
- [0280] cauggacgag cuguacaagu aguaa 685
- [0281] <210> 22
- [0282] <211> 2337
- [0283] <212> RNA
- [0284] <213> 人工序列
- [0285] <220>
- [0286] <223> Nt-uDys
- [0287] <400> 22
- [0288] augcuuuggu gggaagaagu agaggacugu uaugaaagag aagauguuca aaagaaaaca 60
- [0289] uucacaaaau ggguaaaugc acauuuuucu aaguuuggga agcagcauau ugagaaccuc 120
- [0290] uucagugacc uacaggauug gaggcgccuc cuagaccucc ucgaaggccu gacagggcaa 180
- [0291] aaacugccaa aagaaaaagg auccacaaga guucaugccc ugaacaaugu caacaaggca 240
- [0292] cugcggguuu ugcagaacaa uaauguugau uuagugaaua uuggaaguac ugacaucgua 300
- [0293] gauggaaauc auaaacugac ucuugguuug auuuggaaua uauuccucca cuggcagguc 360

[0294]	aaaaauguaa ugaaaaauau cauggcugga uugcaacaaa ccaacaguga aaagauucuc	420
[0295]	cugagcuggg uccgacaauc aacucguauu uauccacagg uuaauguaau caacuucacc	480
[0296]	accagcuggu cugauggccu ggcuuugaau gcucucaucc auagucouag gccagaccua	540
[0297]	uuugacugga auaguguggu uugccagcag ucagccacac aacgacugga acaugcauuc	600
[0298]	aacaucgcca gauaucaauu aggcauagag aaacuacucg auccugaaga uguugauacc	660
[0299]	accuauccag auaagaaguc caucuuaaug uacaucacau cacucuucca aguuuugccu	720
[0300]	caacaaguga gcauugaagc cauccaggaa guggaaaugu ugccaaggcc accuaaagug	780
[0301]	acuaaagaag acauuuuca guuacaucou caaaugcacu auucucaaca gaucacgguc	840
[0302]	agucuagcac agggauauga gagaacuucu uccccuaagc cucgauucaa gagcuaugcc	900
[0303]	uacacacagg cugcuuaugu caccaccucu gaccuacac ggagcccau uccuucacag	960
[0304]	cauuuggaag cuccugaaga caagucuuu ggcaguucou ugauggagag ugaaguaaac	1020
[0305]	cuggaccguu aucaaacagc uuuagaagaa guauuaucgu ggcuucuuuc ugcugaggac	1080
[0306]	acauugcaag cacaaggaga gauuucuaau gauguggaag uggugaaaga ccaguuucou	1140
[0307]	acucaugagg gguacaugau ggauuugaca gcccaucagg gccggguugg uaauauucou	1200
[0308]	caauugggaa gaaagcugau uggaacagga aaauuaucag aagaugaaga aacugaagua	1260
[0309]	caagagcaga ugaauccucu aaaucaaga ugggaaugcc ucaggguagc uagcauggaa	1320
[0310]	aaacaagca auuuacouag aguuuuuauug gaucuccaga aucagaaacu gaaagaguug	1380
[0311]	aaugacuggc uaacaaaaac agaagaaaga acaaggaaaa uggaggaaga gccucuugga	1440
[0312]	ccugaucuug aagaccuaaa acgccaagua caacaacoua agguucuua agaagauca	1500
[0313]	gaacaagaac agucagggg caauucucuc acucacaugg uggugguagu ugaugaaucu	1560
[0314]	aguggagau acgcaacugc ugcuuuggaa gaacaacua agguauuggg agaucgaugg	1620
[0315]	gcaaacoucu guagauggac agaagaccgc uggguucuuu uacaagacou ccuucuaaa	1680
[0316]	uggcaacguc uuacugaaga acagugccu uuuagucou ggcuucaga aaaagaagau	1740
[0317]	gcagugaaca agauucacac aacuggcuuu aaagaucaaa augaauguu aucaagucou	1800
[0318]	caaaaacugg ccguuuuaaa agcggaucua gaaaagaaaa agcaauccou gggcaaacug	1860
[0319]	uaucacuca aacaagaucu ucuuucaca cugaagaaua agucagugac ccagaagacg	1920
[0320]	gaagcauggc uggauaacuu ugcccggugu ugggaaauu uaguccaaa acuugaaaag	1980
[0321]	aguacagcac agauuucaca ggcugucacc accacucagc caucacuaac acagacaacu	2040
[0322]	guaauuggaaa caguaacuac ggugaccaca agggaacaga uccugguaaa gcaugcuca	2100
[0323]	gaggaacuuc caccaccacc uccccaaaag aagaggcaga uuacugugga ucuugaaaga	2160
[0324]	cuccaggaac uucaagaggc cacggaugag cuggaccuca agcugcgcca agcugaggug	2220
[0325]	aucaagggau ccuggcagcc cgugggcgau cuccucouug acucucucca agaucaccuc	2280
[0326]	gagaaaguc aggcacuucg aggagaaau gcgccucuga aagagaacgu gagccac	2337
[0327]	<210>	23
[0328]	<211>	1974
[0329]	<212>	RNA
[0330]	<213>	人工序列
[0331]	<220>	
[0332]	<223>	Ct-uDys-GFP
[0333]	<400>	23
[0334]	gucaaugacc uugcucgcca gcuuaccacu uugggcouuc agcucucacc guauaacuc	60
[0335]	agcacucugg aagaccugaa caccagaugg aagcuucugc agguaggccgu cgaggaccga	120

[0336]	gucaggcagc ugcaugaagc ccacagggac uuugguccag caucucagca cuuucuuucc	180
[0337]	acgucugucc agggucccug ggagagagcc aucucgcaa acaaagugcc cuacuauauc	240
[0338]	aaccacgaga cucaaacaac uugcugggac caucccaaaa ugacagagcu cuaccagucu	300
[0339]	uuagcugacc ugaauaaugu cagauucuca gcuuauagga cugccaugaa acuccgaaga	360
[0340]	cugcagaagg ccuuuugcuu ggauucuuug agccugucag cugcauguga ugccuuggac	420
[0341]	cagcacaacc ucaagcaaaa ugaccagccc auggauaucc ugcagauuau uaauuguuug	480
[0342]	accacuauuu augaccgcu ggagcaagag cacaacaauu uggucaacgu ccucucugc	540
[0343]	guggauaugu gucugaacug gcugcugaau guuuauauga cgggacgaac agggaggau	600
[0344]	cguguccugu cuuuuaaac uggcaucau ucccugugua aagcacauuu ggaagacaag	660
[0345]	uacagauacc uuucuaagca aguggcaagu ucaacaggau uuugugacca gcgcaggcug	720
[0346]	ggccuccuuc ugcaugauuc uauccaaaau ccaagacagu ugggugaagu ugcauccuuu	780
[0347]	ggggcagua acauugagcc aaguguccgg agcugcuucc aauuugcuaa uaauagcca	840
[0348]	gagaucgaag cggcccucu ccuagacugg augagacugg aaccaguc cauggugug	900
[0349]	cugcccugucc ugcacagagu ggucugcga gaaacugcca agcaucaggc caaanguaac	960
[0350]	aucugcaaag aguguccau cauuggauuc agguacagga gucuaaagca cuuuauuuu	1020
[0351]	gacaucugcc aaagcugcu uuuuucuggu cgaguugcaa aaggccaua aaugcacuau	1080
[0352]	ccauggugg aauauugcac uccgacuaca ucaggagaag auguucgaga cuuugccaag	1140
[0353]	guacuaaaaa acaauuucg aaccaaagg uauuuugcga agcauccccg aaugggcuac	1200
[0354]	cugccaguc agacugucu agagggggac acauggaaa cugacacaau ucuagaggug	1260
[0355]	agcaagggcg aggagcugu caccggggug gugccaucc uggucgagcu ggacggcgac	1320
[0356]	guaaacggcc acaaguucag cguguccggc gagggcgagg gcgaugccac cuacggcaag	1380
[0357]	cugaccuga aguucaucug caccaccggc aagcugccc ugcccugcc caccucug	1440
[0358]	accaccuga ccuacggcg gcagugcuuc agccgcuacc ccgaccacu gaagcagcac	1500
[0359]	gacuucuca aguccccau gccgaaggc uacguccag agcgcaccu cuucuuaag	1560
[0360]	gacgacggca acuaaagac ccgcgccgag gugaaguuc agggcgacac ccuggugaac	1620
[0361]	cgcaucgagc ugaagggcau cgacuuaag gaggacggca acauccugg gcacaagcug	1680
[0362]	gaguacaacu acaacagcca caacgucuau aucauggccg acaagcagaa gaacggcauc	1740
[0363]	aaggugaacu ucaagaucg ccacaaucc gaggacggca gcgucgagcu cgccgaccac	1800
[0364]	uaccagcaga acacccccau cggcgacggc cccgucugc ugcccgaca ccacuaccug	1860
[0365]	agcaccagu ccgcccugag caaagacccc aacgagaagc gcgaucacau gguccugcug	1920
[0366]	gaguucguga ccgcccggc gaucacucuc ggcauggac agcuguacaa guaa	1974
[0367]	<210>	24
[0368]	<211>	68
[0369]	<212>	RNA
[0370]	<213>	人工序列
[0371]	<220>	
[0372]	<223>	HdV68催化突变体
[0373]	<400>	24
[0374]	ggccggcaug gucccagcu ccucgcuggc gccggcugg caacaugcuu cggcauggug	60
[0375]	aaugggac	68
[0376]	<210>	25
[0377]	<211>	56

- [0378] <212> RNA
- [0379] <213> 人工序列
- [0380] <220>
- [0381] <223> 具有对Nt-Luc特异的茎3突出端的锤头状核酶
- [0382] <400> 25
- [0383] gagccuuacc ggauguguuu uccggucuga ugaguccggu agcggacgaa aggcuc 56
- [0384] <210> 26
- [0385] <211> 54
- [0386] <212> RNA
- [0387] <213> 人工序列
- [0388] <220>
- [0389] <223> 用于Ct-Luc的具有5 nt P1茎的扭状核酶
- [0390] <400> 26
- [0391] agccuuuaca cugccaaugc cgguccaag cccggauaaa aguggagga ggcu 54
- [0392] <210> 27
- [0393] <211> 54
- [0394] <212> RNA
- [0395] <213> 人工序列
- [0396] <220>
- [0397] <223> 用于Ct-Luc和T6A突变的具有5 nt P1茎的扭状核酶
- [0398] <400> 27
- [0399] agccuaaaca cugccaaugc cgguccaag cccggauaaa aguggagga ggcu 54
- [0400] <210> 28
- [0401] <211> 54
- [0402] <212> RNA
- [0403] <213> 人工序列
- [0404] <220>
- [0405] <223> 用于Ct-Luc的具有5 nt P1茎的扭状核酶突变体
- [0406] <400> 28
- [0407] agccuuuacu cuuccaaugc cgguccaag cccggauaaa aguggagga ggcu 54
- [0408] <210> 29
- [0409] <211> 54
- [0410] <212> RNA
- [0411] <213> 人工序列
- [0412] <220>
- [0413] <223> 用于Ct-Luc的具有5 nt P1茎的扭状核酶
- [0414] <400> 29
- [0415] agccuuuaca cugccaaugc cgguccaag cccggauaaa aguggagga ggcu 54
- [0416] <210> 30
- [0417] <211> 51
- [0418] <212> RNA
- [0419] <213> 人工序列

- [0420] <220>
[0421] <223> 用于Ct-Luc的具有2 nt P1茎的扭状核酶
[0422] <400> 30
[0423] agccuuaaca cugccaaugc cggucccaag cccggauaaa aguggagga g 51
[0424] <210> 31
[0425] <211> 49
[0426] <212> RNA
[0427] <213> 人工序列
[0428] <220>
[0429] <223> 用于Ct-Luc的具有1 nt P1茎的扭状核酶
[0430] <400> 31
[0431] agccuuaaca cugccaaugc cggucccaag cccggauaaa aguggaggg 49
[0432] <210> 32
[0433] <211> 49
[0434] <212> RNA
[0435] <213> 人工序列
[0436] <220>
[0437] <223> 用于Ct-Luc的不具有P1茎的扭状核酶
[0438] <400> 32
[0439] agccuuaaca cugccaaugc cggucccaag cccggauaaa aguggaggg 49
[0440] <210> 33
[0441] <211> 53
[0442] <212> RNA
[0443] <213> 人工序列
[0444] <220>
[0445] <223> 对Ct-Luc特异的16 HH茎1突出端
[0446] <400> 33
[0447] gaauuugua auccugcuga ugaguccgug aggacgaaac gaguaagcuc guc 53
[0448] <210> 34
[0449] <211> 51
[0450] <212> RNA
[0451] <213> 人工序列
[0452] <220>
[0453] <223> 对Ct-Luc特异的14 HH茎1突出端
[0454] <400> 34
[0455] aucuuguaau ccugcugaug aguccgugag gacgaaacga gaaagcucgu c 51
[0456] <210> 35
[0457] <211> 49
[0458] <212> RNA
[0459] <213> 人工序列
[0460] <220>
[0461] <223> 对Ct-Luc特异的12 HH茎1突出端

- [0462] <400> 35
[0463] cuuguaaucc ugcugaugag uccgugagga cgaaacgagu aagcucguc 49
[0464] <210> 36
[0465] <211> 45
[0466] <212> RNA
[0467] <213> 人工序列
[0468] <220>
[0469] <223> 对Ct-Luc特异的8 HH茎1突出端
[0470] <400> 36
[0471] uaauccugcu gaugaguccg ugaggacgaa acgaguaagc ucguc 45
[0472] <210> 37
[0473] <211> 43
[0474] <212> RNA
[0475] <213> 人工序列
[0476] <220>
[0477] <223> 对Ct-Luc特异的6 HH茎1突出端
[0478] <400> 37
[0479] auccugcuga ugaguccgug aggacgaaac gaguaagcuc guc 43
[0480] <210> 38
[0481] <211> 43
[0482] <212> RNA
[0483] <213> 人工序列
[0484] <220>
[0485] <223> 对Ct-Lu特异的6 HH突变体茎1突出端
[0486] <400> 38
[0487] auccugcuga ugaguccgug aggacgagac gaguaagcuc guc 43
[0488] <210> 39
[0489] <211> 41
[0490] <212> RNA
[0491] <213> 人工序列
[0492] <220>
[0493] <223> 对Ct-Luc特异的4 HH茎1突出端
[0494] <400> 39
[0495] ccugcugaug aguccgugag gacgaaacga gaaagcucgu c 41
[0496] <210> 40
[0497] <211> 55
[0498] <212> RNA
[0499] <213> 人工序列
[0500] <220>
[0501] <223> 对Ct-Luc特异的RzB茎1突出端
[0502] <400> 40
[0503] uuguauuuu ccugcugaug agucgcuggg augcgacgaa acgccuucgg gcguc 55

- [0504] <210> 41
- [0505] <211> 52
- [0506] <212> RNA
- [0507] <213> 人工序列
- [0508] <220>
- [0509] <223> Nt载体的剪接供体序列
- [0510] <400> 41
- [0511] guaagauca agguuacaag acagguuuuaa ggagaccaau agaaacuggg cu 52
- [0512] <210> 42
- [0513] <211> 81
- [0514] <212> RNA
- [0515] <213> 人工序列
- [0516] <220>
- [0517] <223> Ct载体的剪接受体序列
- [0518] <400> 42
- [0519] ugucgagaca gagaagacuc uugcguuuuc gauaggcacc uauuggucuu acugacaucc 60
- [0520] acuuugccuu ucucuccaca g 81
- [0521] <210> 43
- [0522] <211> 560
- [0523] <212> RNA
- [0524] <213> 人工序列
- [0525] <220>
- [0526] <223> GCN4 5' UTR uORF
- [0527] <400> 43
- [0528] aaacaaaaac ucacaacaca gguuacucuc cccccuaaa ucaaauuuuu uuugcccauc 60
- [0529] aguuucacua gcgaauuua caacucacca gccacacagc ucacucaucu acuucgcaau 120
- [0530] caaaacaaaa uauuuuauuu uaguucaguu uauuaaguua uuaucaguau cguauuaaaa 180
- [0531] aauuaaagau cauugaaaaa uggcuugcua aaccgauuau auuuuguuuu uaaaguagau 240
- [0532] uauuauuaga aaauuauuaa gagaauuauug uguuaaaauuu auugaaagag aaaaauuuuu 300
- [0533] uucccuuauu aaauaaaguc cuuuacuuuu uuugaaaacu gucaguuuuu ugaagaguua 360
- [0534] uuuguuuuugu uaccaauugc uaucauguac ccguagaauu uuauucaaga uguuuccgua 420
- [0535] acgguuaccu uucugucaaa uuauccaggu uuacucgcca auaaaaauuu ccuauacua 480
- [0536] ucauuuuuuu aucauuuuu auuacuaaag uuuuguuuac cauuugucu gcucaagaaa 540
- [0537] auaaaauaaa uacaaauaaa 560
- [0538] <210> 44
- [0539] <211> 148
- [0540] <212> RNA
- [0541] <213> 人工序列
- [0542] <220>
- [0543] <223> SRY 5' UTR uORF
- [0544] <400> 44
- [0545] guugaggggg uguugagggc ggagaaaugc aaguuucauu acaaaaguua acguaacaaa 60

- [0546] gaaucuggua gaaaugaguu uuggauagua aaauaaguuu cgaacucugg cacuuucaa 120
- [0547] uuuugucgca cucuccuugu uuuugaca 148
- [0548] <210> 45
- [0549] <211> 343
- [0550] <212> RNA
- [0551] <213> 人工序列
- [0552] <220>
- [0553] <223> Hoxa9 TIE
- [0554] <400> 45
- [0555] gaaaaaacag aagaggaag gauaccagag cgguucauac agggcccaga aacuaggcga 60
- [0556] ggugaccccu cagcaagaca aacaccucu gauguugacu ggcgauuuuc ccaucucca 120
- [0557] gucuggggag cgggacuagg cauacagaug auggagcuua gaaccgcug gcuagggaau 180
- [0558] aaaaucgcu gggcaguug ugcucaaaga agugggccag ggcgcuugug acacaucag 240
- [0559] ggcuuugug acacaaccc uugagguug gcaguucucu ccuuggcggu ugcucugguu 300
- [0560] gcucuguggg gccuuccug uggagcaagg gugaucuggc cga 343
- [0561] <210> 46
- [0562] <211> 170
- [0563] <212> RNA
- [0564] <213> 人工序列
- [0565] <220>
- [0566] <223> Hoxa3 TIE
- [0567] <400> 46
- [0568] aggacaauuc gucucuuggg cugccgaagc gacagcuguc agagaggcag aagcuucugg 60
- [0569] gagccgcggu cugaaggcua cgugugcugc cuggucauuc aaagugucua uuuuaggucc 120
- [0570] agaagugucc aaaccacaag uucucaaaac ucugaaaaau ggcuccucc 170
- [0571] <210> 47
- [0572] <211> 38
- [0573] <212> RNA
- [0574] <213> 人工序列
- [0575] <220>
- [0576] <223> NRAS 5'UTR G-假结
- [0577] <400> 47
- [0578] cgucccgugu gggagggcg ggucugggug cggccugc 38
- [0579] <210> 48
- [0580] <211> 126
- [0581] <212> RNA
- [0582] <213> 人工序列
- [0583] <220>
- [0584] <223> 人IFNG 5' UTR假结
- [0585] <400> 48
- [0586] cacauuguuc ugaucaucug aagaucagcu auuagaagag aaagaucagu uaaguccuuu 60
- [0587] ggaccugauc agcuugauac aagaacuacu gauuucacu ucuuuggcuu aaucucugc 120

- [0588] gaaacg 126
- [0589] <210> 49
- [0590] <211> 132
- [0591] <212> RNA
- [0592] <213> 人工序列
- [0593] <220>
- [0594] <223> Rat ODC 5'UTR
- [0595] <400> 49
- [0596] ugucaguccc ugcagccgcc gccgccggcc gccuucaguc agcagcucgg cgccaccucc 60
- [0597] ggucggcgac ugcggcgggc ucgacgaggc ggcugacggg gcggcggcgg gaagacggcc 120
- [0598] gggugcgccu ug 132
- [0599] <210> 50
- [0600] <211> 49
- [0601] <212> RNA
- [0602] <213> 人工序列
- [0603] <220>
- [0604] <223> SIRLOIN RNA核定位信号
- [0605] <400> 50
- [0606] cgccucccgg guucaagcga uucuccugcc ucagccucc gaguagcug 49
- [0607] <210> 51
- [0608] <211> 42
- [0609] <212> RNA
- [0610] <213> 人工序列
- [0611] <220>
- [0612] <223> BORG lncRNA NLS
- [0613] <400> 51
- [0614] accucagaau cuacaaguca gcccgaauua aauguuguuu ua 42
- [0615] <210> 52
- [0616] <211> 108
- [0617] <212> PRT
- [0618] <213> 人工序列
- [0619] <220>
- [0620] <223> FKBP DD
- [0621] <400> 52
- [0622] Met Gly Val Gln Val Glu Thr Ile Ser Pro Gly Asp Gly Arg Thr Phe
- [0623] 1 5 10 15
- [0624] Pro Lys Arg Gly Gln Thr Cys Val Val His Tyr Thr Gly Met Leu Glu
- [0625] 20 25 30
- [0626] Asp Gly Lys Lys Val Asp Ser Ser Arg Asp Arg Asn Lys Pro Phe Lys
- [0627] 35 40 45
- [0628] Phe Met Leu Gly Lys Gln Glu Val Ile Arg Gly Trp Glu Glu Gly Val
- [0629] 50 55 60

[0672]	20	25	30
[0673]	Cys Ala Ser Ala Ser Ile Asn Val		
[0674]	35	40	
[0675]	<210> 56		
[0676]	<211> 57		
[0677]	<212> PRT		
[0678]	<213> 人工序列		
[0679]	<220>		
[0680]	<223> CL1		
[0681]	<400> 56		
[0682]	Ala Cys Lys Asn Trp Phe Ser Ser Leu Ser His Phe Val Ile His Leu		
[0683]	1	5	10 15
[0684]	Asn Ser His Gly Phe Pro Pro Glu Val Glu Glu Gln Ala Ala Gly Thr		
[0685]	20	25	30
[0686]	Leu Pro Met Ser Cys Ala Gln Glu Ser Gly Met Asp Arg His Pro Ala		
[0687]	35	40	45
[0688]	Ala Cys Ala Ser Ala Arg Ile Asn Val		
[0689]	50	55	
[0690]	<210> 57		
[0691]	<211> 57		
[0692]	<212> PRT		
[0693]	<213> 人工序列		
[0694]	<220>		
[0695]	<223> CL1-PEST		
[0696]	<400> 57		
[0697]	Ala Cys Lys Asn Trp Phe Ser Ser Leu Ser His Phe Val Ile His Leu		
[0698]	1	5	10 15
[0699]	Asn Ser His Gly Phe Pro Pro Glu Val Glu Glu Gln Ala Ala Gly Thr		
[0700]	20	25	30
[0701]	Leu Pro Met Ser Cys Ala Gln Glu Ser Gly Met Asp Arg His Pro Ala		
[0702]	35	40	45
[0703]	Ala Cys Ala Ser Ala Arg Ile Asn Val		
[0704]	50	55	
[0705]	<210> 58		
[0706]	<211> 68		
[0707]	<212> PRT		
[0708]	<213> 人工序列		
[0709]	<220>		
[0710]	<223> E1A PEST		
[0711]	<400> 58		
[0712]	Ser Arg Glu Cys Asn Ser Ser Thr Asp Ser Cys Asp Ser Gly Pro Ser		
[0713]	1	5	10 15

[0714]	Asn Thr Pro Pro Glu Ile His Pro Val Val Pro Leu Cys Pro Ile Lys
[0715]	20 25 30
[0716]	Pro Val Ala Val Arg Val Gly Gly Arg Arg Gln Ala Val Glu Cys Ile
[0717]	35 40 45
[0718]	Glu Asp Leu Leu Asn Glu Pro Gly Gln Pro Leu Asp Leu Ser Cys Lys
[0719]	50 55 60
[0720]	Arg Pro Arg Pro
[0721]	65
[0722]	<210> 59
[0723]	<211> 31
[0724]	<212> PRT
[0725]	<213> 人工序列
[0726]	<220>
[0727]	<223> C-myc PEST
[0728]	<400> 59
[0729]	Leu His Glu Glu Thr Pro Pro Thr Thr Ser Ser Asp Ser Glu Glu Glu
[0730]	1 5 10 15
[0731]	Gln Glu Asp Glu Glu Glu Ile Asp Val Val Ser Val Glu Lys Arg
[0732]	20 25 30
[0733]	<210> 60
[0734]	<211> 25
[0735]	<212> PRT
[0736]	<213> 人工序列
[0737]	<220>
[0738]	<223> c-Fos PEST
[0739]	<400> 60
[0740]	Ala Ala His Arg Lys Gly Ser Ser Ser Asn Glu Pro Ser Ser Asp Ser
[0741]	1 5 10 15
[0742]	Leu Ser Ser Pro Thr Leu Leu Ala Leu
[0743]	20 25
[0744]	<210> 61
[0745]	<211> 26
[0746]	<212> PRT
[0747]	<213> 人工序列
[0748]	<220>
[0749]	<223> v-Myb PEST
[0750]	<400> 61
[0751]	Pro Ser Pro Pro Val Asp His Gly Cys Leu Pro Glu Glu Ser Ala Ser
[0752]	1 5 10 15
[0753]	Pro Ala Arg Cys Met Ile Val His Gln Ser
[0754]	20 25
[0755]	<210> 62

[0756] <211> 59
 [0757] <212> PRT
 [0758] <213> 人工序列
 [0759] <220>
 [0760] <223> NPDC1 PEST
 [0761] <400> 62
 [0762] Pro Pro Lys Glu Leu Asp Thr Ala Ser Ser Asp Glu Glu Asn Glu Asp
 [0763] 1 5 10 15
 [0764] Gly Asp Phe Thr Val Tyr Glu Cys Pro Gly Leu Ala Pro Thr Gly Glu
 [0765] 20 25 30
 [0766] Met Glu Val Arg Asn Pro Leu Phe Asp His Ala Ala Leu Ser Ala Pro
 [0767] 35 40 45
 [0768] Leu Pro Ala Pro Ser Ser Pro Pro Ala Leu Pro
 [0769] 50 55
 [0770] <210> 63
 [0771] <211> 37
 [0772] <212> PRT
 [0773] <213> 人工序列
 [0774] <220>
 [0775] <223> IkbA PEST
 [0776] <400> 63
 [0777] Pro Glu Ser Glu Asp Glu Glu Ser Tyr Asp Thr Glu Ser Glu Phe Thr
 [0778] 1 5 10 15
 [0779] Glu Phe Thr Glu Asp Glu Leu Pro Tyr Asp Asp Cys Val Phe Gly Gly
 [0780] 20 25 30
 [0781] Gln Arg Leu Thr Leu
 [0782] 35
 [0783] <210> 64
 [0784] <211> 41
 [0785] <212> PRT
 [0786] <213> 人工序列
 [0787] <220>
 [0788] <223> m.m. AZIN2 PEST
 [0789] <400> 64
 [0790] Gly Gln Leu Leu Pro Ala Glu Glu Asp Gln Asp Ala Glu Gly Val Cys
 [0791] 1 5 10 15
 [0792] Lys Pro Leu Ser Cys Gly Trp Glu Ile Thr Asp Thr Leu Cys Val Gly
 [0793] 20 25 30
 [0794] Pro Val Phe Thr Pro Ala Ser Ile Met
 [0795] 35 40
 [0796] <210> 65
 [0797] <211> 43

[0798] <212> PRT
 [0799] <213> 人工序列
 [0800] <220>
 [0801] <223> x.1. AZIN2 PEST
 [0802] <400> 65
 [0803] Val Gln Leu Leu Gln Arg Gly Leu Gln Gln Thr Glu Glu Lys Glu Asn
 [0804] 1 5 10 15
 [0805] Val Cys Thr Pro Met Ser Cys Gly Trp Glu Ile Ser Asp Ser Leu Cys
 [0806] 20 25 30
 [0807] Phe Thr Arg Thr Phe Ala Ala Thr Ser Ile Ile
 [0808] 35 40
 [0809] <210> 66
 [0810] <211> 12
 [0811] <212> PRT
 [0812] <213> 人工序列
 [0813] <220>
 [0814] <223> NS1
 [0815] <400> 66
 [0816] Thr Ser Leu Tyr Lys Lys Val Gly Met Gly Arg Lys
 [0817] 1 5 10
 [0818] <210> 67
 [0819] <211> 12
 [0820] <212> PRT
 [0821] <213> 人工序列
 [0822] <220>
 [0823] <223> NS6
 [0824] <400> 67
 [0825] Ser Leu Tyr Lys Lys Val Gly Thr Met Ala Ala Gly
 [0826] 1 5 10
 [0827] <210> 68
 [0828] <211> 12
 [0829] <212> PRT
 [0830] <213> 人工序列
 [0831] <220>
 [0832] <223> NS7
 [0833] <400> 68
 [0834] Tyr Lys Lys Val Gly Thr Met Arg Gly Arg Gly Leu
 [0835] 1 5 10
 [0836] <210> 69
 [0837] <211> 12
 [0838] <212> PRT
 [0839] <213> 人工序列

[0840] <220>
 [0841] <223> NS12
 [0842] <400> 69
 [0843] Glu Arg Ala Pro Thr Gly Arg Trp Gly Arg Arg Gly
 [0844] 1 5 10
 [0845] <210> 70
 [0846] <211> 12
 [0847] <212> PRT
 [0848] <213> 人工序列
 [0849] <220>
 [0850] <223> NS15
 [0851] <400> 70
 [0852] Glu Gly Pro Leu Trp His Pro Arg Ile Cys Gly Ser
 [0853] 1 5 10
 [0854] <210> 71
 [0855] <211> 12
 [0856] <212> PRT
 [0857] <213> 人工序列
 [0858] <220>
 [0859] <223> SELK
 [0860] <400> 71
 [0861] Leu Arg Gly Pro Ser Pro Pro Pro Met Ala Gly Gly
 [0862] 1 5 10
 [0863] <210> 72
 [0864] <211> 12
 [0865] <212> PRT
 [0866] <213> 人工序列
 [0867] <220>
 [0868] <223> SELS
 [0869] <400> 72
 [0870] Trp Arg Pro Gly Arg Arg Gly Pro Ser Ser Gly Gly
 [0871] 1 5 10
 [0872] <210> 73
 [0873] <211> 5
 [0874] <212> PRT
 [0875] <213> 人工序列
 [0876] <220>
 [0877] <223> EMID1
 [0878] <400> 73
 [0879] Arg Asp Glu Arg Gly
 [0880] 1 5
 [0881] <210> 74

[0882] <211> 5
 [0883] <212> PRT
 [0884] <213> 人工序列
 [0885] <220>
 [0886] <223> IRX6
 [0887] <400> 74
 [0888] Gly Ala Glu Ala Gly
 [0889] 1 5
 [0890] <210> 75
 [0891] <211> 80
 [0892] <212> PRT
 [0893] <213> 人工序列
 [0894] <220>
 [0895] <223> UbVR
 [0896] <400> 75
 [0897] Gln Ile Phe Val Lys Thr Leu Thr Gly Lys Thr Ile Thr Leu Glu Val
 [0898] 1 5 10 15
 [0899] Glu Pro Ser Asp Thr Ile Glu Asn Val Lys Ala Lys Ile Gln Asp Lys
 [0900] 20 25 30
 [0901] Glu Gly Ile Pro Pro Asp Gln Gln Arg Leu Ile Phe Ala Gly Lys Gln
 [0902] 35 40 45
 [0903] Leu Glu Asp Gly Arg Thr Leu Ser Asp Tyr Asn Ile Gln Lys Glu Ser
 [0904] 50 55 60
 [0905] Thr Leu His Leu Val Leu Arg Leu Arg Gly Val Arg Ala Ser Ala Ser
 [0906] 65 70 75 80
 [0907] <210> 76
 [0908] <211> 162
 [0909] <212> PRT
 [0910] <213> 人工序列
 [0911] <220>
 [0912] <223> 2xUbVR
 [0913] <400> 76
 [0914] Thr Ser Gln Ile Phe Val Lys Thr Leu Thr Gly Lys Thr Ile Thr Leu
 [0915] 1 5 10 15
 [0916] Glu Val Glu Pro Ser Asp Thr Ile Glu Asn Val Lys Ala Lys Ile Gln
 [0917] 20 25 30
 [0918] Asp Lys Glu Gly Ile Pro Pro Asp Gln Gln Arg Leu Ile Phe Ala Gly
 [0919] 35 40 45
 [0920] Lys Gln Leu Glu Asp Gly Arg Thr Leu Ser Asp Tyr Asn Ile Gln Lys
 [0921] 50 55 60
 [0922] Glu Ser Thr Leu His Leu Val Leu Arg Leu Arg Gly Val Arg Ala Ser
 [0923] 65 70 75 80

[0924]	Ala Ser Gln Ile Phe Val Lys Thr Leu Thr Gly Lys Thr Ile Thr Leu
[0925]	85 90 95
[0926]	Glu Val Glu Pro Ser Asp Thr Ile Glu Asn Val Lys Ala Lys Ile Gln
[0927]	100 105 110
[0928]	Asp Lys Glu Gly Ile Pro Pro Asp Gln Gln Arg Leu Ile Phe Ala Gly
[0929]	115 120 125
[0930]	Lys Gln Leu Glu Asp Gly Arg Thr Leu Ser Asp Tyr Asn Ile Gln Lys
[0931]	130 135 140
[0932]	Glu Ser Thr Leu His Leu Val Leu Arg Leu Arg Gly Val Arg Ala Ser
[0933]	145 150 155 160
[0934]	Ala Ser
[0935]	<210> 77
[0936]	<211> 39
[0937]	<212> DNA
[0938]	<213> 人工序列
[0939]	<220>
[0940]	<223> 12x多聚K编码尾序列
[0941]	<400> 77
[0942]	aaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaataa 39
[0943]	<210> 78
[0944]	<211> 12
[0945]	<212> PRT
[0946]	<213> 人工序列
[0947]	<220>
[0948]	<223> 翻译产物12x多聚K
[0949]	<400> 78
[0950]	Lys Lys Lys Lys Lys Lys Lys Lys Lys Lys Lys Lys Lys
[0951]	1 5 10
[0952]	<210> 79
[0953]	<211> 51
[0954]	<212> DNA
[0955]	<213> 人工序列
[0956]	<220>
[0957]	<223> 16x多聚K编码尾序列
[0958]	<400> 79
[0959]	aaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaata a 51
[0960]	<210> 80
[0961]	<211> 16
[0962]	<212> PRT
[0963]	<213> 人工序列
[0964]	<220>
[0965]	<223> 翻译产物16x多聚K

[1008]	Met Ile His Ser Gly Ser Arg Gly Leu Gly His Gln Val Ala Thr Asp
[1009]	260 265 270
[1010]	Ala Leu Val Ala Met Glu Lys Ala Met Lys Arg Asp Lys Ile Ile Val
[1011]	275 280 285
[1012]	Asn Asp Arg Gln Leu Ala Cys Ala Arg Ile Ala Ser Pro Glu Gly Gln
[1013]	290 295 300
[1014]	Asp Tyr Leu Lys Gly Met Ala Ala Ala Gly Asn Tyr Ala Trp Val Asn
[1015]	305 310 315 320
[1016]	Arg Ser Ser Met Thr Phe Leu Thr Arg Gln Ala Phe Ala Lys Val Phe
[1017]	325 330 335
[1018]	Asn Thr Thr Pro Asp Asp Leu Asp Leu His Val Ile Tyr Asp Val Ser
[1019]	340 345 350
[1020]	His Asn Ile Ala Lys Val Glu Gln His Val Val Asp Gly Lys Glu Arg
[1021]	355 360 365
[1022]	Thr Leu Leu Val His Arg Lys Gly Ser Thr Arg Ala Phe Pro Pro His
[1023]	370 375 380
[1024]	His Pro Leu Ile Ala Val Asp Tyr Gln Leu Thr Gly Gln Pro Val Leu
[1025]	385 390 395 400
[1026]	Ile Gly Gly Thr Met Gly Thr Cys Ser Tyr Val Leu Thr Gly Thr Glu
[1027]	405 410 415
[1028]	Gln Gly Met Thr Glu Thr Phe Gly Thr Thr Cys His Gly Ala Gly Arg
[1029]	420 425 430
[1030]	Ala Leu Ser Arg Ala Lys Ser Arg Arg Asn Leu Asp Phe Gln Asp Val
[1031]	435 440 445
[1032]	Leu Asp Lys Leu Ala Asp Met Gly Ile Ala Ile Arg Val Ala Ser Pro
[1033]	450 455 460
[1034]	Lys Leu Val Met Glu Glu Ala Pro Glu Ser Tyr Lys Asn Val Thr Asp
[1035]	465 470 475 480
[1036]	Val Val Asn Thr Cys His Asp Ala Gly Ile Ser Lys Lys Ala Ile Lys
[1037]	485 490 495
[1038]	Leu Arg Pro Ile Ala Val Ile Lys Gly
[1039]	500 505
[1040]	<210> 82
[1041]	<211> 1518
[1042]	<212> DNA
[1043]	<213> 人工序列
[1044]	<220>
[1045]	<223> 人RtcB人密码子优化的核酸序列
[1046]	<400> 82
[1047]	atgtcccggt catataatga cgagctgcaa ttccttgaga agataaataa gaattgctgg 60
[1048]	cgcatcaaga aaggcttcgt tcctaataatg caagttgaag gtgtatttta tgtaaataac 120
[1049]	gctttggaaa agttgatggt cgaggaactg aggaacgcat gtcgcggtgg aggtgtcggg 180

[1050]	ggttttcttc ccgctatgaa gcagattggc aatgtggcgg ctctgcccgg aattgtgcac	240
[1051]	cgctctatag gattgcctga cgtaacacagc ggctacggat tcgccattgg gaatatggcg	300
[1052]	gcgttcgata tgaacgaccc tgaggcggtt gttagccctg gaggtgtcgg cttcgatata	360
[1053]	aattgcggag tcagattgct tcggacaaat ttggatgaat ctgacgtaca accagtgaaa	420
[1054]	gagcaacttg cacaagcgat gttcgatcat attcccgtgg gtgtggggtc aaagggagta	480
[1055]	atcccaatga acgcgaaaga cctggaagaa gcattggaga tgggtgtaga ctggtcactg	540
[1056]	cgagaaggtt atgcctgggc tgaagacaaa gagcactcgc aggagtacgg tcgcatgttg	600
[1057]	caagcagacc caaataaagt atccgcgagg gccaaagaaa gaggtttgcc gcagctgggg	660
[1058]	acattggggg ccgtaacca ctatgcagaa atacaagtag tggatgagat tttcaatgag	720
[1059]	tacgtcgcga agaaaatggg gatcgaccat aaaggtcaag tgtgcgtaat gatacattct	780
[1060]	gggagtcgcg gactcgggca ccaagttgca acggacgcc ttgtcgcctat ggaaaaagcg	840
[1061]	atgaagcggg ataaaatcat cgtaaagtat aggcaattgg cttgcgctcg cattgcgagt	900
[1062]	ccggaagggc aagactactt gaaagggatg gctgctgccg ggaattatgc atgggtcaac	960
[1063]	cggagcagta tgacattctt gacgcggcag gcttttgcaa aagtgtttaa tacgactccg	1020
[1064]	gacgacctcg atctccatgt tatatatgat gtatcacaca atatcgcaa gtttgagcaa	1080
[1065]	cacgttgagg atggtaagga aaggactctg ctggtacacc ggaaaggcag tacacgggca	1140
[1066]	ttcccgcctc atcaccatt gatcgcagtc gattatcaat tgacaggtca gccagttctg	1200
[1067]	atcggaggaa caatgggcac atgtagctac gtattgaccg ggactgaaca ggggatgacc	1260
[1068]	gaaacttttg gcacaacatg ccatggcgcg gggagggcac tctcccgagc taaaagtagg	1320
[1069]	aggaatcttg acttccagga tgtactggat aagctggccg atatggggat agccatccgg	1380
[1070]	gtagcgtcac ccaaatggg aatggaggaa gctcctgaaa gctataaaaa tgtcactgac	1440
[1071]	gttgtaaca catgcatga cgcgggtata tccaagaaag ctattaagct gcgccaata	1500
[1072]	gctgtaatta aaggatag 1518	
[1073]	<210> 83	
[1074]	<211> 408	
[1075]	<212> PRT	
[1076]	<213> 人工序列	
[1077]	<220>	
[1078]	<223> 大肠杆菌RtcB蛋白序列	
[1079]	<400> 83	
[1080]	Met Asn Tyr Glu Leu Leu Thr Thr Glu Asn Ala Pro Val Lys Met Trp	
[1081]	1 5 10 15	
[1082]	Thr Lys Gly Val Pro Val Glu Ala Asp Ala Arg Gln Gln Leu Ile Asn	
[1083]	20 25 30	
[1084]	Thr Ala Lys Met Pro Phe Ile Phe Lys His Ile Ala Val Met Pro Asp	
[1085]	35 40 45	
[1086]	Val His Leu Gly Lys Gly Ser Thr Ile Gly Ser Val Ile Pro Thr Lys	
[1087]	50 55 60	
[1088]	Gly Ala Ile Ile Pro Ala Ala Val Gly Val Asp Ile Gly Cys Gly Met	
[1089]	65 70 75 80	
[1090]	Asn Ala Leu Arg Thr Ala Leu Thr Ala Glu Asp Leu Pro Glu Asn Leu	
[1091]	85 90 95	

[1092]	Ala Glu Leu Arg Gln Ala Ile Glu Thr Ala Val Pro His Gly Arg Thr
[1093]	100 105 110
[1094]	Thr Gly Arg Cys Lys Arg Asp Lys Gly Ala Trp Glu Asn Pro Pro Val
[1095]	115 120 125
[1096]	Asn Val Asp Ala Lys Trp Ala Glu Leu Glu Ala Gly Tyr Gln Trp Leu
[1097]	130 135 140
[1098]	Thr Gln Lys Tyr Pro Arg Phe Leu Asn Thr Asn Asn Tyr Lys His Leu
[1099]	145 150 155 160
[1100]	Gly Thr Leu Gly Thr Gly Asn His Phe Ile Glu Ile Cys Leu Asp Glu
[1101]	165 170 175
[1102]	Ser Asp Gln Val Trp Ile Met Leu His Ser Gly Ser Arg Gly Ile Gly
[1103]	180 185 190
[1104]	Asn Ala Ile Gly Thr Tyr Phe Ile Asp Leu Ala Gln Lys Glu Met Gln
[1105]	195 200 205
[1106]	Glu Thr Leu Glu Thr Leu Pro Ser Arg Asp Leu Ala Tyr Phe Met Glu
[1107]	210 215 220
[1108]	Gly Thr Glu Tyr Phe Asp Asp Tyr Leu Lys Ala Val Ala Trp Ala Gln
[1109]	225 230 235 240
[1110]	Leu Phe Ala Ser Leu Asn Arg Asp Ala Met Met Glu Asn Val Val Thr
[1111]	245 250 255
[1112]	Ala Leu Gln Ser Ile Thr Gln Lys Thr Val Arg Gln Pro Gln Thr Leu
[1113]	260 265 270
[1114]	Ala Met Glu Glu Ile Asn Cys His His Asn Tyr Val Gln Lys Glu Gln
[1115]	275 280 285
[1116]	His Phe Gly Glu Glu Ile Tyr Val Thr Arg Lys Gly Ala Val Ser Ala
[1117]	290 295 300
[1118]	Arg Ala Gly Gln Tyr Gly Ile Ile Pro Gly Ser Met Gly Ala Lys Ser
[1119]	305 310 315 320
[1120]	Phe Ile Val Arg Gly Leu Gly Asn Glu Glu Ser Phe Cys Ser Cys Ser
[1121]	325 330 335
[1122]	His Gly Ala Gly Arg Val Met Ser Arg Thr Lys Ala Lys Lys Leu Phe
[1123]	340 345 350
[1124]	Ser Val Glu Asp Gln Ile Arg Ala Thr Ala His Val Glu Cys Arg Lys
[1125]	355 360 365
[1126]	Asp Ala Glu Val Ile Asp Glu Ile Pro Met Ala Tyr Lys Asp Ile Asp
[1127]	370 375 380
[1128]	Ala Val Met Ala Ala Gln Ser Asp Leu Val Glu Val Ile Tyr Thr Leu
[1129]	385 390 395 400
[1130]	Arg Gln Val Val Cys Val Lys Gly
[1131]	405
[1132]	<210> 84
[1133]	<211> 1227

- [1134] <212> DNA
- [1135] <213> 人工序列
- [1136] <220>
- [1137] <223> 大肠杆菌RtcB人密码子优化的核酸序列
- [1138] <400> 84
- [1139] atgaattacg agcttcttac cactgagaat gcacctgtga aaatgtggac taagggagtg 60
- [1140] cccgtggaag cggacgcaag gcagcagctc ataaatacag ctaagatgcc tttcatcttc 120
- [1141] aaacacatcg cggttatgcc cgacgtgcac ctcgaaaag gctctactat tggaaagtgtg 180
- [1142] attccgacaa aggggtcgat catacctgct gccgtcgggg tggacatagg ctgtggaatg 240
- [1143] aatgccctgc gaacggctct taccgcagaa gatcttctctg agaatctggc cgagctgcga 300
- [1144] caggccattg aaacagcggg tccgcattgt cggactaccg gacggtgcaa aagggacaaa 360
- [1145] ggtgcgtggg aaaaccctcc cgtaacgtg gatgcgaaat gggctgagtt ggaagcaggc 420
- [1146] tatcaatggc ttaccagaa atatccacgg ttcttgaaca ctaataacta caaacacctg 480
- [1147] gggaccttgg ggacggggaa tcatttcacg gaaatctgtc ttgatgagtc tgaccaagtg 540
- [1148] tggattatgc ttcatagcgg tagccgcggc attggtaacg caattgggac atattttatt 600
- [1149] gacctcgcgc agaaagagat gcaggaaaac cttgagacgc tgccgtcccg agatcttgcg 660
- [1150] tattttatgg aagggacgga atactttgac gattatctga aggcggtagc atgggctcaa 720
- [1151] ctgtttgcta gtctcaaccg agacgcgatg atggaaaatg tggtaacagc acttcaatca 780
- [1152] atcacccaaa agacagtgcg acagcccaaa actctcgcta tggaaagaaat caattgccac 840
- [1153] cacaattacg ttcagaaaga gcaacatttc ggagaagaaa tttacgtgac aagaaaagga 900
- [1154] gctgttagcg cgaggccgg acagtacggc atcattctctg ggtcaatggg tgcgaaatct 960
- [1155] tttatagtac gcgggcttgg taatgaagaa tccttctgca gctgttctca tggagccgga 1020
- [1156] agggtaatgt ccaggactaa ggccaagaaa ctcttctctg tggaaagatca aattagagct 1080
- [1157] acagcacatg ttgaatgtag aaaggatgcc gaagtcatag acgagatccc tatggcttac 1140
- [1158] aaagatatag atgctgtaat ggctgcacag tcagacctcg tagaggttat ctacacactc 1200
- [1159] cggcaagtcg tatgcgtaaa aggatag 1227
- [1160] <210> 85
- [1161] <211> 470
- [1162] <212> PRT
- [1163] <213> 人工序列
- [1164] <220>
- [1165] <223> 耐辐射奇异球菌RtcB蛋白序列
- [1166] <400> 85
- [1167] Met Asn Gly Lys His Ile Thr Lys Leu Gly Phe Glu Gly Lys Ala Val
- [1168] 1 5 10 15
- [1169] Gly Leu Ala Leu Ser Ala Ala Gly Leu Arg Glu Asp Ala Gly Val Ser
- [1170] 20 25 30
- [1171] Arg Gly Asp Ile Leu Asp Glu Leu Arg Ser Val Gln Asn Tyr Pro Glu
- [1172] 35 40 45
- [1173] Gln Tyr Gln Gly Gly Gly Val Tyr Ala Asp Leu Ala Thr His Leu Ile
- [1174] 50 55 60
- [1175] Glu Gln Gln Ala Ala Gln Gln Thr Arg Gln Ser Ala Lys Leu Arg Ala

[1176]	65	70	75	80
[1177]	Ala Pro Leu Pro Tyr Arg Thr Trp Gly Glu Asp Leu Ile Glu Pro Gly			
[1178]		85	90	95
[1179]	Ala His Arg Gln Met Asp Val Ala Met Gln Leu Pro Ile Ser Arg Ala			
[1180]		100	105	110
[1181]	Gly Ala Leu Met Pro Asp Ala His Val Gly Tyr Gly Leu Pro Ile Gly			
[1182]		115	120	125
[1183]	Gly Val Leu Ala Thr Glu Asn Ala Val Ile Pro Tyr Gly Val Gly Val			
[1184]		130	135	140
[1185]	Asp Ile Gly Cys Ser Met Met Leu Ser Val Phe Pro Val Ala Ala Thr			
[1186]		145	150	155
[1187]	Gly Leu Ser Val Asp Glu Ala Arg Ser Leu Leu Leu Lys His Thr Arg			
[1188]		165	170	175
[1189]	Phe Gly Ala Gly Val Gly Phe Glu Lys Arg Asp Arg Leu Asp His Pro			
[1190]		180	185	190
[1191]	Val Leu Ala Glu Ala Thr Trp Asp Glu Gln Pro Leu Leu Arg His Leu			
[1192]		195	200	205
[1193]	Phe Asp Lys Ala Ala Gly Gln Ile Gly Ser Ser Gly Ser Gly Asn His			
[1194]		210	215	220
[1195]	Phe Val Glu Phe Gly Thr Phe Thr Leu Ala Gln Ala Asp Pro Gln Leu			
[1196]		225	230	235
[1197]	Glu Gly Leu Asp Pro Gly Glu Tyr Leu Ala Val Leu Ser His Ser Gly			
[1198]		245	250	255
[1199]	Ser Arg Gly Phe Gly Ala Gln Val Ala Gly His Phe Thr Asn Leu Ala			
[1200]		260	265	270
[1201]	Gln Arg Leu Trp Pro Ala Leu Asp Lys Glu Ala Gln Lys Leu Ala Trp			
[1202]		275	280	285
[1203]	Leu Pro Leu Asp Ser Glu Ala Gly Gln Ala Tyr Trp Gln Ala Met Asn			
[1204]		290	295	300
[1205]	Leu Ala Gly Arg Tyr Ala Leu Ala Asn His Glu Gln Ile His Ala Arg			
[1206]		305	310	315
[1207]	Leu Ala Arg Ala Leu Gly Glu Lys Pro Leu Leu Arg Ala Gln Asn Ser			
[1208]		325	330	335
[1209]	His Asn Leu Ala Trp Lys Gln Gln Val Asn Gly Gln Glu Leu Ile Val			
[1210]		340	345	350
[1211]	His Arg Lys Gly Ala Thr Pro Ala Glu Ala Gly Gln Leu Gly Leu Ile			
[1212]		355	360	365
[1213]	Pro Gly Ser Met Ala Asp Pro Gly Tyr Leu Val Arg Gly Arg Gly Asn			
[1214]		370	375	380
[1215]	Pro Glu Ala Leu Ala Ser Ala Ser His Gly Ala Gly Arg Gln Leu Gly			
[1216]		385	390	395
[1217]	Arg Lys Ala Ala Glu Arg Ser Leu Ala Lys Lys Asp Val Gln Ala Tyr			

[1218]	405	410	415
[1219]	Leu Lys Asp Arg Gly Val Thr Leu Ile Gly Gly Gly Ile Asp Glu Ala		
[1220]	420	425	430
[1221]	Pro Gln Ala Tyr Lys Arg Ile Glu Asp Val Ile Ala Arg Gln Arg Asp		
[1222]	435	440	445
[1223]	Leu Val Asp Val Leu Gly Glu Phe Arg Pro Arg Val Val Arg Met Asp		
[1224]	450	455	460
[1225]	Thr Gly Ser Glu Asp Val		
[1226]	465	470	
[1227]	<210> 86		
[1228]	<211> 1413		
[1229]	<212> DNA		
[1230]	<213> 人工序列		
[1231]	<220>		
[1232]	<223> 耐辐射异常球菌RtcB人密码子优化的核酸序列		
[1233]	<400> 86		
[1234]	atgaacggaa agcacatcac gaagttgggt ttcgaaggga aggctggttg cctggcattg	60	
[1235]	tctgcggctg gtctcagga agacgcaggc gtttcccag gagatattct cgatgaactt	120	
[1236]	aggtctgtcc agaattatcc ggagcaatat caagggggag gggctctatgc cgacttggcg	180	
[1237]	acacacctta ttgagcaaca agctgtcag cagactaggc aatccgcaa gctgcgagca	240	
[1238]	gcaccacttc cgtaccgaac gtggggtgaa gacctgatcg agccaggcgc acacagacag	300	
[1239]	atggatgtag caatgcagct cccgatctcc cgggcgggag cgctgatgcc agatgcccac	360	
[1240]	gtaggatacg gacttcccat tggaggcgtg ctctgctaccg aaaacgccgt aatcccctat	420	
[1241]	ggagtgggcg ttgacatcgg ttgctcaatg atgttgatg ttttcccggg ggctgcaaca	480	
[1242]	ggtctgtcag tggatgaggc gcggtcactg cttctcaaac acacgcgctt cgggtgcgggg	540	
[1243]	gtcggattcg agaaacgca caggctcgac catcctgtct tggcggagge tacgtgggac	600	
[1244]	gagcagcctt tgctgagaca cttgtttgat aaagctgctg gccagattgg gtcttccgga	660	
[1245]	tcagggaacc acttcgtcga atttggaact ttcaccctcg cacaggccga tccgcagttg	720	
[1246]	gaaggtttgg acctgggga atacttggtt gttctttcac actcaggag tagaggattt	780	
[1247]	ggagcccagg tggctgggca ttttaccac ttggcgcagc gcttgtggcc cgcacttgat	840	
[1248]	aaggaagctc aaaaactcgc atggctgcca ctggattctg aggctgggca agcctactgg	900	
[1249]	caagccatga acttggcggg acgatatgag ttggctaacc atgagcaaat tcacgcccga	960	
[1250]	ctggcccgcg cacttgggtga gaagcctctt ctgcgcgccc agaactceca caatctggcc	1020	
[1251]	tggaacagc aggtgaatgg gcaggaattg atagtccacc gcaaaggggc tactcctgag	1080	
[1252]	gaagccgggc aacttgggtc catccctggc tccatggccg acccgggata tttggctcag	1140	
[1253]	ggaaggggaa atccggaagc attggcctct gcgtcacacg gagcaggtag acagctcggc	1200	
[1254]	cggaaggcag cggaaggctc cctggcgaag aaagatgtgc aggcttacct taaagataga	1260	
[1255]	ggagtaacct ttatcggggg cgggattgac gaggctcccc aggcgtataa aaggatcgaa	1320	
[1256]	gacgtcatag cacgccagcg ggacctgtg gatgtgttgg gagaatttag gccacagta	1380	
[1257]	gtcggatg atacagggtc tgaagatgt tag 1413		
[1258]	<210> 87		
[1259]	<211> 481		

[1260] <212> PRT
 [1261] <213> 人工序列
 [1262] <220>
 [1263] <223> 霍里科什火球菌RtcB蛋白序列
 [1264] <400> 87
 [1265] Met Val Val Pro Leu Lys Arg Ile Asp Lys Ile Arg Trp Glu Ile Pro
 [1266] 1 5 10 15
 [1267] Lys Phe Asp Lys Arg Met Arg Val Pro Gly Arg Val Tyr Ala Asp Glu
 [1268] 20 25 30
 [1269] Val Leu Leu Glu Lys Met Lys Asn Asp Arg Thr Leu Glu Gln Ala Thr
 [1270] 35 40 45
 [1271] Asn Val Ala Met Leu Pro Gly Ile Tyr Lys Tyr Ser Ile Val Met Pro
 [1272] 50 55 60
 [1273] Asp Gly His Gln Gly Tyr Gly Phe Pro Ile Gly Gly Val Ala Ala Phe
 [1274] 65 70 75 80
 [1275] Asp Val Lys Glu Gly Val Ile Ser Pro Gly Gly Ile Gly Tyr Asp Ile
 [1276] 85 90 95
 [1277] Asn Cys Gly Val Arg Leu Ile Arg Thr Asn Leu Thr Glu Lys Glu Val
 [1278] 100 105 110
 [1279] Arg Pro Arg Ile Lys Gln Leu Val Asp Thr Leu Phe Lys Asn Val Pro
 [1280] 115 120 125
 [1281] Ser Gly Val Gly Ser Gln Gly Arg Ile Lys Leu His Trp Thr Gln Ile
 [1282] 130 135 140
 [1283] Asp Asp Val Leu Val Asp Gly Ala Lys Trp Ala Val Asp Asn Gly Tyr
 [1284] 145 150 155 160
 [1285] Gly Trp Glu Arg Asp Leu Glu Arg Leu Glu Glu Gly Gly Arg Met Glu
 [1286] 165 170 175
 [1287] Gly Ala Asp Pro Glu Ala Val Ser Gln Arg Ala Lys Gln Arg Gly Ala
 [1288] 180 185 190
 [1289] Pro Gln Leu Gly Ser Leu Gly Ser Gly Asn His Phe Leu Glu Val Gln
 [1290] 195 200 205
 [1291] Val Val Asp Lys Ile Phe Asp Pro Glu Val Ala Lys Ala Tyr Gly Leu
 [1292] 210 215 220
 [1293] Phe Glu Gly Gln Val Val Val Met Val His Thr Gly Ser Arg Gly Leu
 [1294] 225 230 235 240
 [1295] Gly His Gln Val Ala Ser Asp Tyr Leu Arg Ile Met Glu Arg Ala Ile
 [1296] 245 250 255
 [1297] Arg Lys Tyr Arg Ile Pro Trp Pro Asp Arg Glu Leu Val Ser Val Pro
 [1298] 260 265 270
 [1299] Phe Gln Ser Glu Glu Gly Gln Arg Tyr Phe Ser Ala Met Lys Ala Ala
 [1300] 275 280 285
 [1301] Ala Asn Phe Ala Trp Ala Asn Arg Gln Met Ile Thr His Trp Val Arg

[1302]	290	295	300
[1303]	Glu Ser Phe Gln Glu Val Phe Lys Gln Asp Pro Glu Gly Asp Leu Gly		
[1304]	305	310	315 320
[1305]	Met Asp Ile Val Tyr Asp Val Ala His Asn Ile Gly Lys Val Glu Glu		
[1306]		325	330 335
[1307]	His Glu Val Asp Gly Lys Arg Val Lys Val Ile Val His Arg Lys Gly		
[1308]		340	345 350
[1309]	Ala Thr Arg Ala Phe Pro Pro Gly His Glu Ala Val Pro Arg Leu Tyr		
[1310]		355	360 365
[1311]	Arg Asp Val Gly Gln Pro Val Leu Ile Pro Gly Ser Met Gly Thr Ala		
[1312]		370	375 380
[1313]	Ser Tyr Ile Leu Ala Gly Thr Glu Gly Ala Met Lys Glu Thr Phe Gly		
[1314]		385	390 395 400
[1315]	Ser Thr Cys His Gly Ala Gly Arg Val Leu Ser Arg Lys Ala Ala Thr		
[1316]		405	410 415
[1317]	Arg Gln Tyr Arg Gly Asp Arg Ile Arg Gln Glu Leu Leu Asn Arg Gly		
[1318]		420	425 430
[1319]	Ile Tyr Val Arg Ala Ala Ser Met Arg Val Val Ala Glu Glu Ala Pro		
[1320]		435	440 445
[1321]	Gly Ala Tyr Lys Asn Val Asp Asn Val Val Lys Val Val Ser Glu Ala		
[1322]		450	455 460
[1323]	Gly Ile Ala Lys Leu Val Ala Arg Met Arg Pro Ile Gly Val Ala Lys		
[1324]		465	470 475 480
[1325]	Gly		
[1326]	<210> 88		
[1327]	<211> 1446		
[1328]	<212> DNA		
[1329]	<213> 人工序列		
[1330]	<220>		
[1331]	<223> 霍里科什火球菌RtcB人密码子优化的核酸序列		
[1332]	<400> 88		
[1333]	atggtggttc ccctgaagag aatagataaa attcgctggg agatccctaa gttcgacaaa 60		
[1334]	aggatgagag taccaggacg ggtgtatgca gatgaggctt tgctcgaaaa aatgaaaaat 120		
[1335]	gaccgcacgc ttgaacaggc aacgaacgtc gcaatgctgc caggcattta taaatacagt 180		
[1336]	atttgtatgc ccgatggcca ccaggggtac ggatttcaa ttggaggggt agccgctttc 240		
[1337]	gatgttaaag agggcgtaat cagtctggt gggatcgggt acgacatcaa ttgtggagtc 300		
[1338]	cgactgatca gaaccaatct cactgagaaa gaagtaagc ccagaatcaa gcaactggtt 360		
[1339]	gatactctgt ttaaaaacgt cccttctgga gtgggcagtc aaggcggat taaactgcat 420		
[1340]	tggactcaaa tagacgatgt actcgtagac ggggcaaaat gggctgtgga caacggatat 480		
[1341]	ggatgggagc gcgacctga acggttgaa gaagtggtc ggatggaggg ggccgatcca 540		
[1342]	gaggcggctc cccaacgggc aaagcagagg ggagcacccc agctcgggtc cctggggtct 600		
[1343]	ggcaaccatt tctcgaagt acaggtcgta gataagatct ttgatcctga agtagcgaaa 660		

[1344] gcgtatggcc tcttcgaggg gcaagtgggtt gtgatggttc aacttggttag cagaggtcctt 720
 [1345] gggcaccaag ttgcatccga ctacttgcca atcatggagc gcgcaattag gaagtataga 780
 [1346] atccccctggc cggatagaga gcttgtctca gtcccttttc aaagcgagga aggacaaaga 840
 [1347] tacttcagcg ccatgaaagc cgcggcaaac tttgcatggg caaatcggca gatgataact 900
 [1348] cattgggtac gagaatcatt ccaagaggtc ttcaacaag atccggaagg cgacctcggc 960
 [1349] atggacattg tgtacgatgt cgcccacaat ataggcaaag tggaggagca cgaggtcgat 1020
 [1350] ggcaaacggg tgaaagtat agtccatcga aaggagcaa ctgcgcttt tccaccaggt 1080
 [1351] cagagagctg tacctagct gtatcgggat gtcggtaac ctgtactcat acccgatct 1140
 [1352] atgggcacag cttcctatat tctggctggc actgaaggag caatgaaaga gacgtttgga 1200
 [1353] tctacctgtc acggagctgg taggtactc tcccggagg ccgcgacacg acaatatcgc 1260
 [1354] ggggacagga tcagacaaga acttttgaat agaggeatct acgtgcgcgc cgctagtatg 1320
 [1355] cgcgtcgtgg ccgaagagc acctggggct tacaagaac tggataacgt agttaaagta 1380
 [1356] gtaagtgaag ccggcatcgc caagctggtg gcccgatgc gcccgattgg cgtggcaaag 1440
 [1357] ggtag 1446
 [1358] <210> 89
 [1359] <211> 481
 [1360] <212> PRT
 [1361] <213> 人工序列
 [1362] <220>
 [1363] <223> 火球菌属种ST04 RtcB蛋白序列
 [1364] <400> 89
 [1365] Met Thr Val Pro Leu Lys Arg Ile Asp Arg Ile Arg Trp Glu Ile Pro
 [1366] 1 5 10 15
 [1367] Lys Phe Asp Lys Arg Met Arg Val Pro Gly Arg Val Tyr Ala Asp Glu
 [1368] 20 25 30
 [1369] Val Leu Ile Glu Lys Met Arg Ser Asp Arg Thr Leu Glu Gln Ala Ala
 [1370] 35 40 45
 [1371] Asn Val Ala Met Leu Pro Gly Ile Tyr Lys Tyr Ser Ile Val Met Pro
 [1372] 50 55 60
 [1373] Asp Gly His Gln Gly Tyr Gly Phe Pro Ile Gly Gly Val Ala Ala Phe
 [1374] 65 70 75 80
 [1375] Asp Val Lys Glu Gly Val Ile Ser Pro Gly Gly Ile Gly Tyr Asp Ile
 [1376] 85 90 95
 [1377] Asn Cys Gly Val Arg Leu Ile Arg Thr Asn Leu Thr Glu Lys Glu Val
 [1378] 100 105 110
 [1379] Arg Pro Lys Ile Lys Gln Leu Val Asp Thr Leu Phe Lys Asn Val Pro
 [1380] 115 120 125
 [1381] Ser Gly Val Gly Ser Gln Gly Arg Ile Arg Leu His Trp Thr Gln Ile
 [1382] 130 135 140
 [1383] Asp Asp Val Leu Val Asp Gly Ala Lys Trp Ala Val Asp Asn Gly Tyr
 [1384] 145 150 155 160
 [1385] Gly Trp Glu Arg Asp Leu Glu Arg Leu Glu Glu Gly Gly Arg Met Glu

[1386]		165		170		175
[1387]	Gly Ala Asp Pro Asp Ala Val Ser Gln Arg Ala Lys Gln Arg Gly Ala					
[1388]		180		185		190
[1389]	Pro Gln Leu Gly Ser Leu Gly Ser Gly Asn His Phe Leu Glu Val Gln					
[1390]		195		200		205
[1391]	Val Val Asp Lys Ile Tyr Asp Glu Glu Val Ala Lys Ala Tyr Gly Leu					
[1392]		210		215		220
[1393]	Phe Glu Gly Gln Val Val Val Met Val His Thr Gly Ser Arg Gly Leu					
[1394]		225		230		235
[1395]	Gly His Gln Val Ala Ser Asp Tyr Leu Arg Ile Met Glu Arg Ala Ile					
[1396]		245		250		255
[1397]	Arg Lys Tyr Arg Ile Pro Trp Pro Asp Arg Glu Leu Val Ser Val Pro					
[1398]		260		265		270
[1399]	Phe Gln Ser Glu Glu Gly Gln Arg Tyr Phe Ser Ala Met Lys Ala Ala					
[1400]		275		280		285
[1401]	Ala Asn Phe Ala Trp Ala Asn Arg Gln Met Ile Thr His Trp Val Arg					
[1402]		290		295		300
[1403]	Glu Ser Phe Gln Glu Val Phe Arg Gln Asp Pro Glu Gly Asp Leu Gly					
[1404]		305		310		315
[1405]	Met Asp Ile Val Tyr Asp Val Ala His Asn Ile Gly Lys Val Glu Glu					
[1406]		325		330		335
[1407]	His Glu Val Asp Gly Lys Lys Val Thr Val Ile Val His Arg Lys Gly					
[1408]		340		345		350
[1409]	Ala Thr Arg Ala Phe Pro Pro Gly His Glu Ala Ile Pro Arg Ile Tyr					
[1410]		355		360		365
[1411]	Arg Asp Val Gly Gln Pro Val Leu Ile Pro Gly Ser Met Gly Thr Ala					
[1412]		370		375		380
[1413]	Ser Tyr Val Leu Ala Gly Thr Glu Gly Ala Met Lys Glu Thr Phe Gly					
[1414]		385		390		395
[1415]	Ser Thr Cys His Gly Ala Gly Arg Val Leu Ser Arg Lys Ala Ala Thr					
[1416]		405		410		415
[1417]	Arg Gln Tyr Arg Gly Asp Arg Ile Arg Asn Glu Leu Leu Gln Arg Gly					
[1418]		420		425		430
[1419]	Ile Tyr Val Arg Ala Ala Ser Met Arg Val Val Ala Glu Glu Ala Pro					
[1420]		435		440		445
[1421]	Gly Ala Tyr Lys Asn Val Asp Asn Val Val Lys Val Val Ser Glu Ala					
[1422]		450		455		460
[1423]	Gly Ile Ala Lys Leu Val Ala Arg Met Arg Pro Ile Gly Val Ala Lys					
[1424]		465		470		475
[1425]	Gly					
[1426]	<210> 90					
[1427]	<211> 1446					

- [1428] <212> DNA
- [1429] <213> 人工序列
- [1430] <220>
- [1431] <223> 火球菌属种ST04 RtcB人密码子优化的核酸序列
- [1432] <400> 90
- [1433] atgaccgttc ccctgaagag aatagatagg attcgtggg agatccctaa gttcgacaaa 60
- [1434] aggatgagag taccaggacg ggtgtatgca gatgaggtct tgatcgagaa aatgagaagc 120
- [1435] gaccgcacgc ttgaacaggc agccaacgtc gcaatgctgc caggcattta taaatacagt 180
- [1436] attgtgatgc ccgatggcca ccaggggtac ggatttcaa ttggaggggt agccgctttc 240
- [1437] gatgttaaag agggcgtaat cagtctggt gggatcgggt acgacatcaa ttgtggagtc 300
- [1438] cgactgatca gaaccaatct cactgagaaa gaagtaaggc ccaaaatcaa gcaactggtt 360
- [1439] gatactctgt ttaaaaacgt cccttctgga gtgggcagtc aagggcggat tagactgcat 420
- [1440] tggactcaaa tagacgatgt actcgtagac ggggcaaaat gggctgtgga caacggatat 480
- [1441] ggatgggagc gcgacctcga acggttgaa gaaggtggtc ggatggaggg ggccgatcca 540
- [1442] gacgcggtct cccaacgggc aaagcagagg ggagcacccc agctcgggtc cctggggtct 600
- [1443] ggcaaccatt tcctcgaagt acaggtcgtg gataagatct acgatgagga agtagcgaag 660
- [1444] gcgatggcc tcttcgaggg gcaagtggtt gtgatggtc aactggtag cagaggtctt 720
- [1445] gggaccaag ttgcatccga ctacttcgca atcatggagc gcgcaattag gaagtataga 780
- [1446] atcccctggc cggatagaga gcttgtctca gtcccttttc aaagcgagga aggacaaaaga 840
- [1447] tacttcagcg ccatgaaagc cgcggcaaac tttgatggg caaatcggca gatgataact 900
- [1448] cattgggtac gagaatcatt ccaagagtc ttcagacaag atccggaagg cgacctggc 960
- [1449] atggacattg tgtacgatgt cgcccacaat ataggcaaag tggaggagca cgaggtcgat 1020
- [1450] ggcaagaaag tgaccgttat agtccatcga aaggagcaa ctgcgcgtt tccaccaggt 1080
- [1451] cagaggcta tccttaggat ctatcgggat gtcggtcaac ctgtactcat acccgatct 1140
- [1452] atggcacag cttcctatgt gctggctggc actgaaggag caatgaaaga gacgtttgga 1200
- [1453] tctacctgtc acggagctgg taggtactc tcccgggaag ccgcgacacg acaatatcgc 1260
- [1454] ggggacagga tcagaaatga acttttgcaa agaggcatct acgtgcgcgc cgctagtatg 1320
- [1455] cgcgtcgtgg ccgaagaggc acctggggct tacaagaacg tggataacgt agttaaagta 1380
- [1456] gtaagtgaag ccggcatcgc caagctggtg gcccggatgc gcccgattgg cgtggcaaag 1440
- [1457] ggtag 1446
- [1458] <210> 91
- [1459] <211> 480
- [1460] <212> PRT
- [1461] <213> 人工序列
- [1462] <220>
- [1463] <223> 热球菌属种EP1 RtcB蛋白序列
- [1464] <400> 91
- [1465] Met Glu Ile Pro Leu Lys Arg Leu Asp Lys Ile Arg Trp Glu Ile Pro
- [1466] 1 5 10 15
- [1467] Lys Phe Asn Arg Arg Met Arg Val Pro Gly Arg Val Tyr Ala Asp Asp
- [1468] 20 25 30
- [1469] Thr Leu Leu Gln Lys Met Arg Gln Asp Lys Thr Leu Glu Gln Ala Thr

[1470]	35	40	45
[1471]	Asn Val Ala Met Leu Pro Gly Ile Tyr Lys Tyr Ser Ile Val Met Pro		
[1472]	50	55	60
[1473]	Asp Gly His Gln Gly Tyr Gly Phe Pro Ile Gly Gly Val Ala Ala Phe		
[1474]	65	70	75
[1475]	Asp Val Lys Glu Gly Val Ile Ser Pro Gly Gly Val Gly Tyr Asp Ile		
[1476]	85	90	95
[1477]	Asn Cys Gly Val Arg Leu Ile Arg Thr Asn Leu Val Glu Lys Glu Val		
[1478]	100	105	110
[1479]	Arg Pro Lys Ile Lys Gln Leu Ile Asp Thr Leu Phe Lys Asn Val Pro		
[1480]	115	120	125
[1481]	Ser Gly Leu Gly Ser Lys Gly Arg Ile Arg Leu His Trp Thr Gln Leu		
[1482]	130	135	140
[1483]	Asp Asp Val Leu Ala Asp Gly Ala Lys Trp Ala Val Asp Asn Gly Tyr		
[1484]	145	150	155
[1485]	Gly Trp Lys Asp Asp Leu Glu His Leu Glu Glu Gly Gly Arg Met Glu		
[1486]	165	170	175
[1487]	Gly Ala Asn Pro Asn Ala Val Ser Gln Lys Ala Lys Gln Arg Gly Ala		
[1488]	180	185	190
[1489]	Pro Gln Leu Gly Ser Leu Gly Ser Gly Asn His Phe Leu Glu Ile Gln		
[1490]	195	200	205
[1491]	Val Val Asp Lys Val Phe Asn Glu Glu Ile Ala Lys Ala Tyr Gly Leu		
[1492]	210	215	220
[1493]	Phe Glu Gly Gln Ile Val Val Met Val His Thr Gly Ser Arg Gly Leu		
[1494]	225	230	235
[1495]	Gly His Gln Val Ala Ser Asp Tyr Leu Arg Ile Met Glu Lys Ala Asn		
[1496]	245	250	255
[1497]	Arg Lys Tyr Asn Val Pro Trp Pro Asp Arg Glu Leu Val Ser Val Pro		
[1498]	260	265	270
[1499]	Phe Gln Thr Glu Glu Gly Gln Arg Tyr Phe Ser Ala Met Lys Ala Ala		
[1500]	275	280	285
[1501]	Ala Asn Phe Ala Trp Ala Asn Arg Gln Met Ile Thr His Trp Val Arg		
[1502]	290	295	300
[1503]	Glu Ser Phe Glu Glu Val Phe Lys Gln Lys Ala Glu Asp Leu Gly Met		
[1504]	305	310	315
[1505]	His Ile Val Tyr Asp Val Ala His Asn Ile Ala Lys Val Glu Glu His		
[1506]	325	330	335
[1507]	Glu Val Asn Gly Arg Lys Ile Lys Val Val Val His Arg Lys Gly Ala		
[1508]	340	345	350
[1509]	Thr Arg Ala Phe Pro Ala Gly His Glu Ala Ile Pro Lys Ala Tyr Arg		
[1510]	355	360	365
[1511]	Asp Val Gly Gln Pro Val Leu Ile Pro Gly Ser Met Gly Thr Ala Ser		

[1512]	370	375	380
[1513]	Tyr Val Leu Ala Gly Ala Glu Gly Ser Met Arg Glu Thr Phe Gly Ser		
[1514]	385	390	395
[1515]	Thr Cys His Gly Ala Gly Arg Val Leu Ser Arg His Ala Ala Thr Arg		
[1516]	405	410	415
[1517]	Gln Phe Arg Gly Asp Arg Leu Arg Asn Glu Leu Met Gln Arg Gly Ile		
[1518]	420	425	430
[1519]	Tyr Ile Arg Ala Ala Ser Met Arg Val Val Ala Glu Glu Ala Pro Gly		
[1520]	435	440	445
[1521]	Ala Tyr Lys Asn Val Asp Asn Val Val Arg Val Val His Glu Ala Gly		
[1522]	450	455	460
[1523]	Ile Ala Asn Leu Val Ala Arg Met Arg Pro Ile Gly Val Ala Lys Gly		
[1524]	465	470	475
[1525]	<210> 92		
[1526]	<211> 1446		
[1527]	<212> DNA		
[1528]	<213> 人工序列		
[1529]	<220>		
[1530]	<223> 热球菌属种EP1 RtcB人密码子优化的核酸序列		
[1531]	<400> 92		
[1532]	atggagatac cactcaaacg acttgacaag atccgatggg agattcccaa atttaacaga 60		
[1533]	cgaatgagag ttccggaag agtttacga gatgatacat tgctccaaaa gatgcgacaa 120		
[1534]	gataagacgc tcgaacaagc caccaactgt gccatgctcc caggcattta taagtatagt 180		
[1535]	atagtcatgc ctgacggaca ccagggttat ggattcccga ttggcgggtg agcagccttc 240		
[1536]	gacgtaaaag agggagtaat tagtcttggc ggtgttggtt atgatattaa ctgtggcgtg 300		
[1537]	aggcttatca ggacgaatct tgtagagaag gaagtgcgac caaaaatcaa acaacttata 360		
[1538]	gatactttgt tcaaaaatgt cccgtctggg ctcgatcaa agggtcggat aaggctccac 420		
[1539]	tggactcaac tggatgatgt tctggctgat ggggcaaaat gggctgttga caatgggtac 480		
[1540]	gggtggaagg atgatctcga acatttgag gagggcggac ggatggaggg cgcaaacccc 540		
[1541]	aatgccgttt cacagaaagc gaagcaaagg ggagcggcac agcttgggtc ctttggtca 600		
[1542]	ggcaatcatt tcctcgaat tcaggtcgtc gataaggtt ttaacgaaga gatagcaaag 660		
[1543]	gcttacggac tctttgaagg tcagatagt gtaatgtcc atacgggctc tcggggactg 720		
[1544]	ggacatcaag tcgcaagtga ctacctgag atcatggaga aagccaatcg caagtacaat 780		
[1545]	gtgccctggc ctgaccgga gcttgttagc gtgccctcc agacggaaga gggatcaacga 840		
[1546]	tactttagcg ctatgaaggc ggcagctaat ttcgcttggg caaacagaca gatgataaca 900		
[1547]	cattgggtta gagagtcctt cgaggaggtc tttaaacaaa aagctgagga ctttggaatg 960		
[1548]	catattgtct atgatgttc ccataacata gcaaaagtag aggaacatga ggtgaacggg 1020		
[1549]	cggaaaatta aggtcgtagt acacagaaaa ggcgctacca gagcattccc cgcaggacac 1080		
[1550]	gaggccatac ccaaagcata tagagatgtc ggccagccag tgctcatacc gggatctatg 1140		
[1551]	ggtacggcgt cctatgtctt ggcgggtgct gaaggatcaa tgaggagac gttcggctca 1200		
[1552]	acctgcatg ggcaggctcg ggtcttctcg cggcatgtg caactcggca gttccgctca 1260		
[1553]	gatcgactca ggaatgaact catgcagaga ggcatttaca tacgcgctgc ctccatgcgc 1320		

[1554] gttgtcgccg aggaagctcc cggcgcctat aagaacgtag acaatgtcgt cagggtggtg 1380
 [1555] catgaagcgg gaattgcgaa cttggtagcc aggatgcgcc caataggggt tgccaagga 1440
 [1556] tagtaa 1446
 [1557] <210> 93
 [1558] <211> 167
 [1559] <212> PRT
 [1560] <213> 人工序列
 [1561] <220>
 [1562] <223> 人Archease蛋白序列
 [1563] <400> 93
 [1564] Met Ala Gln Glu Glu Glu Asp Val Arg Asp Tyr Asn Leu Thr Glu Glu
 [1565] 1 5 10 15
 [1566] Gln Lys Ala Ile Lys Ala Lys Tyr Pro Pro Val Asn Arg Lys Tyr Glu
 [1567] 20 25 30
 [1568] Tyr Leu Asp His Thr Ala Asp Val Gln Leu His Ala Trp Gly Asp Thr
 [1569] 35 40 45
 [1570] Leu Glu Glu Ala Phe Glu Gln Cys Ala Met Ala Met Phe Gly Tyr Met
 [1571] 50 55 60
 [1572] Thr Asp Thr Gly Thr Val Glu Pro Leu Gln Thr Val Glu Val Glu Thr
 [1573] 65 70 75 80
 [1574] Gln Gly Asp Asp Leu Gln Ser Leu Leu Phe His Phe Leu Asp Glu Trp
 [1575] 85 90 95
 [1576] Leu Tyr Lys Phe Ser Ala Asp Glu Phe Phe Ile Pro Arg Glu Val Lys
 [1577] 100 105 110
 [1578] Val Leu Ser Ile Asp Gln Arg Asn Phe Lys Leu Arg Ser Ile Gly Trp
 [1579] 115 120 125
 [1580] Gly Glu Glu Phe Ser Leu Ser Lys His Pro Gln Gly Thr Glu Val Lys
 [1581] 130 135 140
 [1582] Ala Ile Thr Tyr Ser Ala Met Gln Val Tyr Asn Glu Glu Asn Pro Glu
 [1583] 145 150 155 160
 [1584] Val Phe Val Ile Ile Asp Ile
 [1585] 165
 [1586] <210> 94
 [1587] <211> 461
 [1588] <212> DNA
 [1589] <213> 人工序列
 [1590] <220>
 [1591] <223> 人Archease人密码子优化的核酸序列
 [1592] <400> 94
 [1593] aggaacaaaa ggcatcaaa gcgaaatc cgcctgtaa ccgaaagtat gactacctgg 60
 [1594] atcacactgc ggacgtccag ttgcatgcct gggcgacac tctggaggag gcattcgaac 120
 [1595] aatgtgcaat ggcaatgttt ggctacatga ctgatacagg cacagtggag ccccttcaaa 180

[1596] cggtagaggt agaaactcag ggagatgac ttcagagctt gctcttccat tttctcgacg 240
 [1597] aatggttga taagttcagt gccgacgagt tcttcattcc acgcgaagtg aaagtgctga 300
 [1598] gtattgatca gagaaacttt aaacttaggt ctattgggtg gggatgaagag ttctctttgt 360
 [1599] ctaaacaccc tcaaggaact gaggtaaagg cgataactta ctcagccatg caggtatata 420
 [1600] acgaggagaa tcctgagggt ttcgtaatca ttgatata g 461
 [1601] <210> 95
 [1602] <211> 142
 [1603] <212> PRT
 [1604] <213> 人工序列
 [1605] <220>
 [1606] <223> 霍里科什火球菌Archease蛋白序列
 [1607] <400> 95
 [1608] Met Lys Lys Trp Glu His Tyr Glu His Thr Ala Asp Ile Gly Ile Arg
 [1609] 1 5 10 15
 [1610] Gly Tyr Gly Asp Ser Leu Glu Glu Ala Phe Glu Ala Val Ala Ile Ala
 [1611] 20 25 30
 [1612] Leu Phe Asp Val Met Val Asn Val Asn Lys Val Glu Lys Lys Glu Val
 [1613] 35 40 45
 [1614] Arg Glu Ile Glu Val Glu Ala Glu Asp Leu Glu Ala Leu Leu Tyr Ser
 [1615] 50 55 60
 [1616] Phe Leu Glu Glu Leu Leu Val Ile His Asp Ile Glu Gly Leu Val Phe
 [1617] 65 70 75 80
 [1618] Arg Asp Phe Glu Val Lys Ile Glu Arg Val Asn Gly Lys Tyr Arg Leu
 [1619] 85 90 95
 [1620] Arg Ala Lys Ala Tyr Gly Glu Lys Leu Asp Leu Lys Lys His Glu Pro
 [1621] 100 105 110
 [1622] Lys Glu Glu Val Lys Ala Ile Thr Tyr His Asp Met Lys Ile Glu Arg
 [1623] 115 120 125
 [1624] Leu Pro Asn Gly Lys Trp Met Ala Gln Leu Val Pro Asp Ile
 [1625] 130 135 140
 [1626] <210> 96
 [1627] <211> 429
 [1628] <212> DNA
 [1629] <213> 人工序列
 [1630] <220>
 [1631] <223> 霍里科什火球菌Archease人密码子优化的核酸序列
 [1632] <400> 96
 [1633] atgaagaaat gggagcacta tgagcatact gccgacattg gtattcgggg atatggggat 60
 [1634] agccttgagg aggcattcga agcagtagcc atcgcgctct ttgatgtaat ggtgaacgtg 120
 [1635] aataaagtcg agaagaagga agtccgagaa attgaagtgg aggcagaaga tttggaggcc 180
 [1636] ctcttttatt cattctgga agaactgttg gttattcatg atatagaggg actggttttc 240
 [1637] aggactttg aagttaagat agagagagta aatggcaaat accgacttcg agcgaaagcc 300

[1638] tacggtgaga agctcgacct caagaagcac gaaccgaaag aggaagtaaa ggcgataacc 360
 [1639] taccatgata tgaaaattga acggttgccc aatggaaagt ggatggctca actcgttcca 420
 [1640] gatatttag 429
 [1641] <210> 97
 [1642] <211> 301
 [1643] <212> PRT
 [1644] <213> 人工序列
 [1645] <220>
 [1646] <223> T4多核苷酸激酶(T4 PNK) 蛋白序列
 [1647] <400> 97
 [1648] Met Lys Lys Ile Ile Leu Thr Ile Gly Cys Pro Gly Ser Gly Lys Ser
 [1649] 1 5 10 15
 [1650] Thr Trp Ala Arg Glu Phe Ile Ala Lys Asn Pro Gly Phe Tyr Asn Ile
 [1651] 20 25 30
 [1652] Asn Arg Asp Asp Tyr Arg Gln Ser Ile Met Ala His Glu Glu Arg Asp
 [1653] 35 40 45
 [1654] Glu Tyr Lys Tyr Thr Lys Lys Lys Glu Gly Ile Val Thr Gly Met Gln
 [1655] 50 55 60
 [1656] Phe Asp Thr Ala Lys Ser Ile Leu Tyr Gly Gly Asp Ser Val Lys Gly
 [1657] 65 70 75 80
 [1658] Val Ile Ile Ser Asp Thr Asn Leu Asn Pro Glu Arg Arg Leu Ala Trp
 [1659] 85 90 95
 [1660] Glu Thr Phe Ala Lys Glu Tyr Gly Trp Lys Val Glu His Lys Val Phe
 [1661] 100 105 110
 [1662] Asp Val Pro Trp Thr Glu Leu Val Lys Arg Asn Ser Lys Arg Gly Thr
 [1663] 115 120 125
 [1664] Lys Ala Val Pro Ile Asp Val Leu Arg Ser Met Tyr Lys Ser Met Arg
 [1665] 130 135 140
 [1666] Glu Tyr Leu Gly Leu Pro Val Tyr Asn Gly Thr Pro Gly Lys Pro Lys
 [1667] 145 150 155 160
 [1668] Ala Val Ile Phe Asp Val Asp Gly Thr Leu Ala Lys Met Asn Gly Arg
 [1669] 165 170 175
 [1670] Gly Pro Tyr Asp Leu Glu Lys Cys Asp Thr Asp Val Ile Asn Pro Met
 [1671] 180 185 190
 [1672] Val Val Glu Leu Ser Lys Met Tyr Ala Leu Met Gly Tyr Gln Ile Val
 [1673] 195 200 205
 [1674] Val Val Ser Gly Arg Glu Ser Gly Thr Lys Glu Asp Pro Thr Lys Tyr
 [1675] 210 215 220
 [1676] Tyr Arg Met Thr Arg Lys Trp Val Glu Asp Ile Ala Gly Val Pro Leu
 [1677] 225 230 235 240
 [1678] Val Met Gln Cys Gln Arg Glu Gln Gly Asp Thr Arg Lys Asp Asp Val
 [1679] 245 250 255

[1680]	Val Lys Glu Glu Ile Phe Trp Lys His Ile Ala Pro His Phe Asp Val
[1681]	260 265 270
[1682]	Lys Leu Ala Ile Asp Asp Arg Thr Gln Val Val Glu Met Trp Arg Arg
[1683]	275 280 285
[1684]	Ile Gly Val Glu Cys Trp Gln Val Ala Ser Gly Asp Phe
[1685]	290 295 300
[1686]	<210> 98
[1687]	<211> 906
[1688]	<212> DNA
[1689]	<213> 人工序列
[1690]	<220>
[1691]	<223> T4 PNK人密码子优化的核酸序列
[1692]	<400> 98
[1693]	atgaagaaaa ttatacttac aatcggatgc cctggtagtg gtaagagcac ttgggcgagg 60
[1694]	gaatttattg cgaagaaccc tggattttat aatatcaatc gagacgacta ccggcagtct 120
[1695]	attatggccc acgaggaacg agacgaatac aagtatacca agaagaaaga agggattgtc 180
[1696]	acgggtatgc aatttgacac cgccaaatca atactgtacg gaggtgattc agtcaaaggc 240
[1697]	gttatcatat cagacactaa cctcaatcct gaacgccgat tggcatggga aacatttgcg 300
[1698]	aaggaatacg gttggaaggt tgaacacaag gtgttcgatg tcccgtggac cgaactggta 360
[1699]	aaacgaatt ctaaacgagg cactaaagct gtgccattg acgtacttcg aagtatgtac 420
[1700]	aagtccatga gagagtacct ggggcttccc gtctataacg gtacgccggg caaacccaag 480
[1701]	gcggtgatct ttgacgtaga tgggactctg gcgaagatga atggtcgcgg accatacgat 540
[1702]	ttggaaaaat gtgacacaga tgtaatcaac ccaatggtag tagagcttag caagatgtac 600
[1703]	gcattgatgg gctaccaaat tgtcgtggtg tccgggcggg agtcaggcac aaaagaagat 660
[1704]	ccgacgaagt attatcgcat gacacggaaa tgggtcgaag atatagccgg ggtgcctctc 720
[1705]	gttatgcaat gtcaacgaga acagggcgac acacggaagg atgacgtagt gaaggaggaa 780
[1706]	atcttctgga agcatatagc gccacacttt gacgttaagc tcgccatcga cgaccgaact 840
[1707]	caggtggctg agatgtggcg acgaattggc gtagagtgtt ggcaagttgc atctggagat 900
[1708]	ttttag 906
[1709]	<210> 99
[1710]	<211> 176
[1711]	<212> PRT
[1712]	<213> 人工序列
[1713]	<220>
[1714]	<223> 大肠杆菌thpR蛋白序列
[1715]	<400> 99
[1716]	Met Ser Glu Pro Gln Arg Leu Phe Phe Ala Ile Asp Leu Pro Ala Glu
[1717]	1 5 10 15
[1718]	Ile Arg Glu Gln Ile Ile His Trp Arg Ala Thr His Phe Pro Pro Glu
[1719]	20 25 30
[1720]	Ala Gly Arg Pro Val Ala Ala Asp Asn Leu His Leu Thr Leu Ala Phe
[1721]	35 40 45

[1722]	Leu Gly Glu Val Ser Ala Glu Lys Glu Lys Ala Leu Ser Leu Leu Ala
[1723]	50 55 60
[1724]	Gly Arg Ile Arg Gln Pro Gly Phe Thr Leu Thr Leu Asp Asp Ala Gly
[1725]	65 70 75 80
[1726]	Gln Trp Leu Arg Ser Arg Val Val Trp Leu Gly Met Arg Gln Pro Pro
[1727]	85 90 95
[1728]	Arg Gly Leu Ile Gln Leu Ala Asn Met Leu Arg Ser Gln Ala Ala Arg
[1729]	100 105 110
[1730]	Ser Gly Cys Phe Gln Ser Asn Arg Pro Phe His Pro His Ile Thr Leu
[1731]	115 120 125
[1732]	Leu Arg Asp Ala Ser Glu Ala Val Thr Ile Pro Pro Pro Gly Phe Asn
[1733]	130 135 140
[1734]	Trp Ser Tyr Ala Val Thr Glu Phe Thr Leu Tyr Ala Ser Ser Phe Ala
[1735]	145 150 155 160
[1736]	Arg Gly Arg Thr Arg Tyr Thr Pro Leu Lys Arg Trp Ala Leu Thr Gln
[1737]	165 170 175
[1738]	<210> 100
[1739]	<211> 531
[1740]	<212> DNA
[1741]	<213> 人工序列
[1742]	<220>
[1743]	<223> 大肠杆菌thpR人密码子优化的核酸序列
[1744]	<400> 100
[1745]	atgagtgagc ctcaacgatt gttctttgcc atagatttgc ctgctgaaat tagagagcaa 60
[1746]	attatccatt ggagagccac ccatttcccc ccagaagctg gacgaccagt cgcagcggac 120
[1747]	aacctccacc ttacactggc gttcttgggt gaagtgagcg ccgagaaaga gaaagctctc 180
[1748]	tcacttctgg ctgggaggat tcggcagccg ggctttacc ttactctgga tgatgccggc 240
[1749]	cagtggctga ggtccagggt tgtctggctc ggaatgagc aaccacctag ggggctcatc 300
[1750]	cagctcgcca atatgetgag atcccaggcc gcaaggtctg gctgcttcca atcaaacagg 360
[1751]	ccattccacc cgcatattac cttgctcaga gatgcctccg aggcagtaac tattccacct 420
[1752]	cccggcttta actggagtta cgccgtcaca gaatttactc tgtacgcctc cagcttcgcc 480
[1753]	cgaggagaaa ccaggtacac gcctttgaag cgggtggcct tgaccagta g 531
[1754]	<210> 101
[1755]	<211> 521
[1756]	<212> PRT
[1757]	<213> 人工序列
[1758]	<220>
[1759]	<223> 人PNKP蛋白序列
[1760]	<400> 101
[1761]	Met Gly Glu Val Glu Ala Pro Gly Arg Leu Trp Leu Glu Ser Pro Pro
[1762]	1 5 10 15
[1763]	Gly Gly Ala Pro Pro Ile Phe Leu Pro Ser Asp Gly Gln Ala Leu Val

[1764]	20	25	30
[1765]	Leu Gly Arg Gly Pro Leu Thr Gln Val Thr Asp Arg Lys Cys Ser Arg		
[1766]	35	40	45
[1767]	Thr Gln Val Glu Leu Val Ala Asp Pro Glu Thr Arg Thr Val Ala Val		
[1768]	50	55	60
[1769]	Lys Gln Leu Gly Val Asn Pro Ser Thr Thr Gly Thr Gln Glu Leu Lys		
[1770]	65	70	75
[1771]	Pro Gly Leu Glu Gly Ser Leu Gly Val Gly Asp Thr Leu Tyr Leu Val		
[1772]	85	90	95
[1773]	Asn Gly Leu His Pro Leu Thr Leu Arg Trp Glu Glu Thr Arg Thr Pro		
[1774]	100	105	110
[1775]	Glu Ser Gln Pro Asp Thr Pro Pro Gly Thr Pro Leu Val Ser Gln Asp		
[1776]	115	120	125
[1777]	Glu Lys Arg Asp Ala Glu Leu Pro Lys Lys Arg Met Arg Lys Ser Asn		
[1778]	130	135	140
[1779]	Pro Gly Trp Glu Asn Leu Glu Lys Leu Leu Val Phe Thr Ala Ala Gly		
[1780]	145	150	155
[1781]	Val Lys Pro Gln Gly Lys Val Ala Gly Phe Asp Leu Asp Gly Thr Leu		
[1782]	165	170	175
[1783]	Ile Thr Thr Arg Ser Gly Lys Val Phe Pro Thr Gly Pro Ser Asp Trp		
[1784]	180	185	190
[1785]	Arg Ile Leu Tyr Pro Glu Ile Pro Arg Lys Leu Arg Glu Leu Glu Ala		
[1786]	195	200	205
[1787]	Glu Gly Tyr Lys Leu Val Ile Phe Thr Asn Gln Met Ser Ile Gly Arg		
[1788]	210	215	220
[1789]	Gly Lys Leu Pro Ala Glu Glu Phe Lys Ala Lys Val Glu Ala Val Val		
[1790]	225	230	235
[1791]	Glu Lys Leu Gly Val Pro Phe Gln Val Leu Val Ala Thr His Ala Gly		
[1792]	245	250	255
[1793]	Leu Tyr Arg Lys Pro Val Thr Gly Met Trp Asp His Leu Gln Glu Gln		
[1794]	260	265	270
[1795]	Ala Asn Asp Gly Thr Pro Ile Ser Ile Gly Asp Ser Ile Phe Val Gly		
[1796]	275	280	285
[1797]	Asp Ala Ala Gly Arg Pro Ala Asn Trp Ala Pro Gly Arg Lys Lys Lys		
[1798]	290	295	300
[1799]	Asp Phe Ser Cys Ala Asp Arg Leu Phe Ala Leu Asn Leu Gly Leu Pro		
[1800]	305	310	315
[1801]	Phe Ala Thr Pro Glu Glu Phe Phe Leu Lys Trp Pro Ala Ala Gly Phe		
[1802]	325	330	335
[1803]	Glu Leu Pro Ala Phe Asp Pro Arg Thr Val Ser Arg Ser Gly Pro Leu		
[1804]	340	345	350
[1805]	Cys Leu Pro Glu Ser Arg Ala Leu Leu Ser Ala Ser Pro Glu Val Val		

[1806]	355	360	365
[1807]	Val Ala Val Gly Phe Pro Gly Ala Gly Lys Ser Thr Phe Leu Lys Lys		
[1808]	370	375	380
[1809]	His Leu Val Ser Ala Gly Tyr Val His Val Asn Arg Asp Thr Leu Gly		
[1810]	385	390	395
[1811]	Ser Trp Gln Arg Cys Val Thr Thr Cys Glu Thr Ala Leu Lys Gln Gly		
[1812]	405	410	415
[1813]	Lys Arg Val Ala Ile Asp Asn Thr Asn Pro Asp Ala Ala Ser Arg Ala		
[1814]	420	425	430
[1815]	Arg Tyr Val Gln Cys Ala Arg Ala Ala Gly Val Pro Cys Arg Cys Phe		
[1816]	435	440	445
[1817]	Leu Phe Thr Ala Thr Leu Glu Gln Ala Arg His Asn Asn Arg Phe Arg		
[1818]	450	455	460
[1819]	Glu Met Thr Asp Ser Ser His Ile Pro Val Ser Asp Met Val Met Tyr		
[1820]	465	470	475
[1821]	Gly Tyr Arg Lys Gln Phe Glu Ala Pro Thr Leu Ala Glu Gly Phe Ser		
[1822]	485	490	495
[1823]	Ala Ile Leu Glu Ile Pro Phe Arg Leu Trp Val Glu Pro Arg Leu Gly		
[1824]	500	505	510
[1825]	Arg Leu Tyr Cys Gln Phe Ser Glu Gly		
[1826]	515	520	
[1827]	<210> 102		
[1828]	<211> 1566		
[1829]	<212> DNA		
[1830]	<213> 人工序列		
[1831]	<220>		
[1832]	<223> 人PNKP人密码子优化的核酸序列		
[1833]	<400> 102		
[1834]	atgggcgagg tggagccccc gggccgcttg tggctcgaga gccccctgg gggagcgccc 60		
[1835]	cccatcttcc tgcctcgga cgggcaagcc ctggtcctgg gcaggggacc cctgaccag 120		
[1836]	gttacggacc ggaagtgtc cagaactcaa gtggagctgg tcgcagatcc tgagaccg 180		
[1837]	acagtggcag tgaacagct gggagttaac ccctcaacta ccgggaccga ggagttgaag 240		
[1838]	ccggggttg agggctctct gggggtgggg gacacactgt atttggtcaa tggcctccac 300		
[1839]	ccactgacc tgcgtggga agagaccgc acaccagaat cccagccaga tactccgct 360		
[1840]	ggcaccctc tgggtgccca agatgagaag agagatgctg agctgccgaa gaagcgtatg 420		
[1841]	cggaagtcaa accccggctg ggagaacttg gagaagtgc tagtggtcac cgcagctggg 480		
[1842]	gtgaaacccc agggcaaggt ggctggcttt gatctggacg ggacgctcat caccacacgc 540		
[1843]	tctgggaagg ttttccac tggccccagt gactggagga tttgtacc agagattccc 600		
[1844]	cgtaagctcc gagagctgga agccgagggc tacaagctgg tgatcttcac caaccagatg 660		
[1845]	agcatcgggc gcgggaagct gccagccgag gatttcaagg ccaaggtgga ggctgtgggt 720		
[1846]	gagaagctgg ggtcccctt ccaggtgtg gtggccacgc acgaggctt gtaccggaag 780		
[1847]	ccggtgacgg gcatgtggga ccatctgcag gagcaggcca acgacggcac gcccatatcc 840		

[1848]	atcggggaca gcatctttgt gggagacgca gccggacgcc cggccaactg ggccccgggg	900
[1849]	cggaagaaga aagacttctc ctgcgccgat cgcctgtttg ccctcaacct tggcctgcc	960
[1850]	ttgccacgc ctgaggagt ctttctcaag tggccagcag ccggcttcga gctcccagcc	1020
[1851]	ttgatccga ggactgtctc ccgctcagg cctctctgcc tccccgagtc cagggcctc	1080
[1852]	ctgagcgcca gcccggaggt ggttgtcga gtgggattcc ctggggccgg gaagtccacc	1140
[1853]	tttctcaaga agcacctcgt gtcggccgga tatgtccacg tgaacaggga cacgctaggc	1200
[1854]	tcttggcagc gctgtgtgac cacgtgtgag acagccctga agcaagggaa acgggtgcc	1260
[1855]	atcgacaaca caaaccaga cgccgcgagc cgcgccaggt acgtccagtg tgcccagacc	1320
[1856]	gcggcgctcc cctgccctg cttctcttc accgccactc tggagcagc gcgccacaac	1380
[1857]	aaccggtttc gagagatgac ggactcctct catatccccg tgtcagacat ggtcatgtat	1440
[1858]	ggctacagga agcagttcga ggccccaac ctggctgaag gtttctctgc catcctggag	1500
[1859]	atcccgttcc ggctatgggt ggagccgagg ctggggcggc tgtactgcca gttctccag	1560
[1860]	ggctag	1566
[1861]	<210>	103
[1862]	<211>	857
[1863]	<212>	RNA
[1864]	<213>	人工序列
[1865]	<220>	
[1866]	<223>	NtGFP-HDV-HH-CtGFP合成内含子
[1867]	<400>	103
[1868]	auggugagca agggcgagga gcuguucacc gggguggugc ccauccuggu cgagcuggac	60
[1869]	ggcgacguaa acggccacaa guucagcgug uccggcgagg gcgagggcga ugccaccuac	120
[1870]	ggcaagcuga ccugaaguu caucugcacc accggcaagc ugcccugucc cuggcccacc	180
[1871]	cucgugacca ccugaccua cggcgugcag ugcuuagcc gcuaccccga ccacaugaag	240
[1872]	cagcacgacu ucuucaaguc cgccaugccc gaaggcuacg uccaggagcg caccaucuuc	300
[1873]	uuggccgca ugguccagc cuccucgug gcgccggcug ggcaacaugc uucggcaugg	360
[1874]	cgaauaggac cccgggacau aacuaguua accaaaucuu ugcugaugag uccgugagga	420
[1875]	cgaaacgagu aagcucgucc aaggacgacg gcaacuacaa gaccgcgcc gaggugaagu	480
[1876]	ucgagggcga caccucggug aaccgcaucg agcugaagg caucgacuuc aaggaggacg	540
[1877]	gcaacaucuu ggggcacaag cuggaguaca acuacaacag ccacaacguc uauaucaugg	600
[1878]	ccgacaagca gaagaacggc aucaagguga acuucaagau ccgccacaac aucgaggacg	660
[1879]	gcagcgugca gcucgccgac cacuaccagc agaacacccc caucggcgac ggccccgugc	720
[1880]	ugcugcccga caaccacuac cugagcacc aguccgccu gagcaaagac cccaacgaga	780
[1881]	agcgcgauca caugguccug cuggaguucg ugaccgccgc cgggaucacu cucggcaugg	840
[1882]	acgagcugua caaguag	857
[1883]	<210>	104
[1884]	<211>	248
[1885]	<212>	RNA
[1886]	<213>	人工序列
[1887]	<220>	
[1888]	<223>	sGCN4 5' UTR uORF
[1889]	<400>	104

- [1890] uuaaagauca uugaaaaaug gcuugcuaaa ccgauuauau uuuguuuua aaguagauua 60
- [1891] uuauuagaaa auuauuaaga gaauuaugug uuaauuuau ugaaagagaa aauuuuuuu 120
- [1892] cccuuuuuaa uuaaaguccu uuacuuuuuu ugaaaacugu caguuuuuug aagaguuuuu 180
- [1893] uguuuuguua ccaauugcua ucauguaccc guagaauuuu auucaagaug uuuccguaac 240
- [1894] gguuaccu 248
- [1895] <210> 105
- [1896] <211> 56
- [1897] <212> RNA
- [1898] <213> 人工序列
- [1899] <220>
- [1900] <223> 用于3'的锤头状(HH)核酶
- [1901] <220>
- [1902] <221> 杂项特征
- [1903] <222> (1) .. (4)
- [1904] <223> n 是 a、c、g 或 u
- [1905] <220>
- [1906] <221> 杂项特征
- [1907] <222> (53) .. (56)
- [1908] <223> n 是 a、c、g 或 u
- [1909] <400> 105
- [1910] nnnndwhacc ggauguguuu uccggucuga ugaguccggu agcggacgaa whnnnn 56
- [1911] <210> 106
- [1912] <211> 54
- [1913] <212> RNA
- [1914] <213> 人工序列
- [1915] <220>
- [1916] <223> 具有5 nt P1茎的Twister WT
- [1917] <220>
- [1918] <221> 杂项特征
- [1919] <222> (1) .. (5)
- [1920] <223> n 是 a、c、g 或 u
- [1921] <220>
- [1922] <221> 杂项特征
- [1923] <222> (50) .. (54)
- [1924] <223> n 是 a、c、g 或 u
- [1925] <400> 106
- [1926] nnnnuaaca cugccaaugc cgguccaag cccggauaaa aguggagggn nnnn 54
- [1927] <210> 107
- [1928] <211> 54
- [1929] <212> RNA
- [1930] <213> 人工序列
- [1931] <220>

- [1932] <223> 具有5 nt P1茎的Twister突变体
- [1933] <220>
- [1934] <221> 杂项特征
- [1935] <222> (1) .. (5)
- [1936] <223> n 是 a、c、g 或 u
- [1937] <220>
- [1938] <221> 杂项特征
- [1939] <222> (50) .. (54)
- [1940] <223> n 是 a、c、g 或 u
- [1941] <400> 107
- [1942] nnnnuaacu cuuccaaugc cggucccaag cccggauaaa aguggaggg nnnn 54
- [1943] <210> 108
- [1944] <211> 54
- [1945] <212> RNA
- [1946] <213> 人工序列
- [1947] <220>
- [1948] <223> 具有有U1A突变的5 nt P1茎的Twister
- [1949] <220>
- [1950] <221> 杂项特征
- [1951] <222> (1) .. (5)
- [1952] <223> n 是 a、c、g 或 u
- [1953] <220>
- [1954] <221> 杂项特征
- [1955] <222> (50) .. (54)
- [1956] <223> n 是 a、c、g 或 u
- [1957] <400> 108
- [1958] nnnnnaaca cugccaaugc cggucccaag cccggauaaa aguggaggg nnnn 54
- [1959] <210> 109
- [1960] <211> 54
- [1961] <212> RNA
- [1962] <213> 人工序列
- [1963] <220>
- [1964] <223> 具有有U1C突变的5 nt P1茎的Twister
- [1965] <220>
- [1966] <221> 杂项特征
- [1967] <222> (1) .. (5)
- [1968] <223> n 是 a、c、g 或 u
- [1969] <220>
- [1970] <221> 杂项特征
- [1971] <222> (50) .. (54)
- [1972] <223> n 是 a、c、g 或 u
- [1973] <400> 109

- [1974] nnnncaaca cugccaaugc cggucccaag cccggauaaa aguggagggn nnnn 54
- [1975] <210> 110
- [1976] <211> 54
- [1977] <212> RNA
- [1978] <213> 人工序列
- [1979] <220>
- [1980] <223> 具有有U1G突变的5 nt P1茎的Twister
- [1981] <220>
- [1982] <221> 杂项特征
- [1983] <222> (1) .. (5)
- [1984] <223> n 是 a、c、g 或 u
- [1985] <220>
- [1986] <221> 杂项特征
- [1987] <222> (50) .. (54)
- [1988] <223> n 是 a、c、g 或 u
- [1989] <400> 110
- [1990] nnnnngaaca cugccaaugc cggucccaag cccggauaaa aguggagggn nnnn 54
- [1991] <210> 111
- [1992] <211> 41
- [1993] <212> RNA
- [1994] <213> 人工序列
- [1995] <220>
- [1996] <223> 用于5'的锤头状核酶4 nt突出端
- [1997] <220>
- [1998] <221> 杂项特征
- [1999] <222> (1) .. (4)
- [2000] <223> n 是 a、c、g 或 u
- [2001] <400> 111
- [2002] nnnnucgaug aguccgugag gacgaaacga gaaagcucgu c 41
- [2003] <210> 112
- [2004] <211> 43
- [2005] <212> RNA
- [2006] <213> 人工序列
- [2007] <220>
- [2008] <223> 用于5'的锤头状核酶6 nt突出端
- [2009] <220>
- [2010] <221> 杂项特征
- [2011] <222> (1) .. (6)
- [2012] <223> n 是 a、c、g 或 u
- [2013] <400> 112
- [2014] nnnnnncuga ugaguccgug aggacgaaac gaguaagcuc guc 43
- [2015] <210> 113

- [2016] <211> 45
[2017] <212> RNA
[2018] <213> 人工序列
[2019] <220>
[2020] <223> 用于5'的锤头状核酶8 nt突出端
[2021] <220>
[2022] <221> 杂项特征
[2023] <222> (1) .. (8)
[2024] <223> n 是 a、c、g 或 u
[2025] <400> 113
[2026] nnnnnnnncu gaugaguccg ugaggacgaa acgaguaagc ucguc 45
[2027] <210> 114
[2028] <211> 47
[2029] <212> RNA
[2030] <213> 人工序列
[2031] <220>
[2032] <223> 用于5'的锤头状核酶10 nt突出端
[2033] <220>
[2034] <221> 杂项特征
[2035] <222> (1) .. (10)
[2036] <223> n 是 a、c、g 或 u
[2037] <400> 114
[2038] nnnnnnnnnn cugaugaguc cgugaggacg aaacgaguaa gcucguc 47
[2039] <210> 115
[2040] <211> 49
[2041] <212> RNA
[2042] <213> 人工序列
[2043] <220>
[2044] <223> 用于5'的锤头状核酶12 nt突出端
[2045] <220>
[2046] <221> 杂项特征
[2047] <222> (1) .. (12)
[2048] <223> n 是 a、c、g 或 u
[2049] <400> 115
[2050] nnnnnnnnnn nncugaugag uccgugagga cgaaacgagu aagcucguc 49
[2051] <210> 116
[2052] <211> 51
[2053] <212> RNA
[2054] <213> 人工序列
[2055] <220>
[2056] <223> 用于5'的锤头状核酶14 nt突出端
[2057] <220>

- [2058] <221> 杂项特征
[2059] <222> (1) .. (14)
[2060] <223> n 是 a、c、g 或 u
[2061] <400> 116
[2062] nnnnnnnnnn nnnncugaug aguccgugag gacgaaacga gaaagcucgu c 51
[2063] <210> 117
[2064] <211> 53
[2065] <212> RNA
[2066] <213> 人工序列
[2067] <220>
[2068] <223> 用于5'的锤头状核酶16 nt突出端
[2069] <220>
[2070] <221> 杂项特征
[2071] <222> (1) .. (16)
[2072] <223> n 是 a、c、g 或 u
[2073] <400> 117
[2074] nnnnnnnnnn nnnnnncuga ugaguccgug aggacgaaac gaguaagcuc guc 53
[2075] <210> 118
[2076] <211> 45
[2077] <212> RNA
[2078] <213> 人工序列
[2079] <220>
[2080] <223> 用于5'的TX2锤头状核酶4 nt突出端
[2081] <220>
[2082] <221> 杂项特征
[2083] <222> (1) .. (4)
[2084] <223> n 是 a、c、g 或 u
[2085] <400> 118
[2086] nnnncugaug aguccgguag cggacgaaac gcgcuucggu gcguc 45
[2087] <210> 119
[2088] <211> 47
[2089] <212> RNA
[2090] <213> 人工序列
[2091] <220>
[2092] <223> 用于5'的TX2锤头状核酶6 nt突出端
[2093] <220>
[2094] <221> 杂项特征
[2095] <222> (1) .. (6)
[2096] <223> n 是 a、c、g 或 u
[2097] <400> 119
[2098] nnnnnncuga ugaguccggu agcggacgaa acgcuucg gucguc 47
[2099] <210> 120

- [2100] <211> 49
[2101] <212> RNA
[2102] <213> 人工序列
[2103] <220>
[2104] <223> 用于5'的TX2锤头状核酶8 nt突出端
[2105] <220>
[2106] <221> 杂项特征
[2107] <222> (1) .. (8)
[2108] <223> n 是 a、c、g 或 u
[2109] <400> 120
[2110] nnnnnnnncu gaugaguccg guagcggacg aaacgcgcuu cggugcguc 49
[2111] <210> 121
[2112] <211> 51
[2113] <212> RNA
[2114] <213> 人工序列
[2115] <220>
[2116] <223> 用于5'的TX2锤头状核酶10 nt突出端
[2117] <220>
[2118] <221> 杂项特征
[2119] <222> (1) .. (10)
[2120] <223> n 是 a、c、g 或 u
[2121] <400> 121
[2122] nnnnnnnnnn cugaugaguc cgguagcggga cgaaacgcgc uucggugcgu c 51
[2123] <210> 122
[2124] <211> 53
[2125] <212> RNA
[2126] <213> 人工序列
[2127] <220>
[2128] <223> 用于5'的TX2锤头状核酶12 nt突出端
[2129] <220>
[2130] <221> 杂项特征
[2131] <222> (1) .. (12)
[2132] <223> n 是 a、c、g 或 u
[2133] <400> 122
[2134] nnnnnnnnnn nncugaugag uccgguagcg gacgaaacgc gcuucggugc guc 53
[2135] <210> 123
[2136] <211> 55
[2137] <212> RNA
[2138] <213> 人工序列
[2139] <220>
[2140] <223> 用于5'的TX2锤头状核酶14 nt突出端
[2141] <220>

- [2142] <221> 杂项特征
[2143] <222> (1) .. (14)
[2144] <223> n 是 a、c、g 或 u
[2145] <400> 123
[2146] nnnnnnnnnn nnnncugaug aguccgguag cggacgaaac gcgcuucggu gcguc 55
[2147] <210> 124
[2148] <211> 57
[2149] <212> RNA
[2150] <213> 人工序列
[2151] <220>
[2152] <223> 用于5'的TX2锤头状核酶16 nt突出端
[2153] <220>
[2154] <221> 杂项特征
[2155] <222> (1) .. (16)
[2156] <223> n 是 a、c、g 或 u
[2157] <400> 124
[2158] nnnnnnnnnn nnnnnncuga ugaguccggu agcggacgaa acgcuucg gugcuc 57
[2159] <210> 125
[2160] <211> 55
[2161] <212> RNA
[2162] <213> 人工序列
[2163] <220>
[2164] <223> 用于5'的RzB锤头状核酶
[2165] <220>
[2166] <221> 杂项特征
[2167] <222> (1) .. (6)
[2168] <223> n 是 a、c、g 或 u
[2169] <220>
[2170] <221> 杂项特征
[2171] <222> (10) .. (14)
[2172] <223> n 是 a、c、g 或 u
[2173] <400> 125
[2174] nnnnnnuaan nnnncugaug agucgcuggg augcgacgaa acgccuucgg gcguc 55
[2175] <210> 126
[2176] <211> 832
[2177] <212> RNA
[2178] <213> 人工序列
[2179] <220>
[2180] <223> NtGFP-HDV-CARGO-HH-CtGFP
[2181] <220>
[2182] <221> 杂项特征
[2183] <222> (371) .. (371)

- [2184] <223> n 是 a、c、g 或 u
- [2185] <400> 126
- [2186] auggugagca agggcgagga gcuguucacc gggguggugc ccauccuggu cgagcuggac 60
- [2187] ggcgacguaa acggccacaa guucagcgug uccggcgagg gcgagggcga ugccaccuac 120
- [2188] ggcaagcuga ccugaaguu caucugcacc accggcaagc ugcccugucc cuggcccacc 180
- [2189] cucgugacca ccugaccua cggcgugcag ugcuuagcc gcuaccccga ccacaugaag 240
- [2190] cagcacgacu ucuucaaguc cgccaugccc gaaggcuacg uccaggagcg caccaucuuc 300
- [2191] uuggccggca uggucccagc cuccucgcug gcgccggcug ggcaacaugc uucggcaugg 360
- [2192] cgaauaggac nuccuugcug augaguccgu gaggacgaaa cgaguaagcu cguccaagga 420
- [2193] cgacggcaac uacaagacc gcgccgaggu gaaguucgag ggcgacacc uggugaaccg 480
- [2194] caucgagcug aagggaucg acuucaagga ggacggcaac auccuggggc acaagcugga 540
- [2195] guacaacuac aacagccaca acgucuauau cauggccgac aagcagaaga acggcaucaa 600
- [2196] ggugaacuuc aagaucgcc acaacaucga ggacggcagc gugcagcug ccgaccacua 660
- [2197] ccagcagaac acccccacg gcgacggccc cgugcugcug cccgacaacc acuaccugag 720
- [2198] caccgagucc gccugagca aagacccaa cgagaagcgc gaucacaugg uccugcugga 780
- [2199] guucgugacc gccgccgga ucacucucgg cauggacgag cuguacaagu ag 832
- [2200] <210> 127
- [2201] <211> 370
- [2202] <212> RNA
- [2203] <213> 人工序列
- [2204] <220>
- [2205] <223> NtGFP-HDV
- [2206] <400> 127
- [2207] auggugagca agggcgagga gcuguucacc gggguggugc ccauccuggu cgagcuggac 60
- [2208] ggcgacguaa acggccacaa guucagcgug uccggcgagg gcgagggcga ugccaccuac 120
- [2209] ggcaagcuga ccugaaguu caucugcacc accggcaagc ugcccugucc cuggcccacc 180
- [2210] cucgugacca ccugaccua cggcgugcag ugcuuagcc gcuaccccga ccacaugaag 240
- [2211] cagcacgacu ucuucaaguc cgccaugccc gaaggcuacg uccaggagcg caccaucuuc 300
- [2212] uuggccggca uggucccagc cuccucgcug gcgccggcug ggcaacaugc uucggcaugg 360
- [2213] cgaauaggac 370
- [2214] <210> 128
- [2215] <211> 461
- [2216] <212> RNA
- [2217] <213> 人工序列
- [2218] <220>
- [2219] <223> HH-CtGFP
- [2220] <400> 128
- [2221] uccuugcuga ugaguccgug aggacgaaac gaguaagcuc guccaaggac gacggcaacu 60
- [2222] acaagaccg cgccgaggug aaguucgagg gcgacaccu ggugaaccgc aucgagcuga 120
- [2223] agggcaucga cuucaaggag gacggcaaca uccuggggca caagcuggag uacaacuaca 180
- [2224] acagccacaa cgucuauauc auggccgaca agcagaagaa cggcaucaag gugaacuua 240
- [2225] agauccgcca caaucgag gacggcagcg ugcagcugc cgaccacuac cagcagaaca 300

[2226]	ccccaucgg cgacggcccc gugcugcugc cgcacaacca cuaccugagc acccaguccg	360
[2227]	ccugagcaa agacccaac gagaagcgcg aucacauggu ccugcuggag uucgugaccg	420
[2228]	ccgccgggau cacucucggc auggacgagc uguacaagua g	461
[2229]	<210>	129
[2230]	<211>	3724
[2231]	<212>	RNA
[2232]	<213>	人工序列
[2233]	<220>	
[2234]	<223>	Nt-miniDys (Δ H2-R15)
[2235]	<400>	129
[2236]	augcuuuggu gggaagaagu agaggacugu uaugaaagag aagauguuca aaagaaaaca	60
[2237]	uucacaaaau ggguaaaugc acauuuuucu aaguuuugga agcagcauau ugagaaccuc	120
[2238]	uucagugacc uacaggauug gaggcgccuc cuagaccucc ucgaaggccu gacagggcaa	180
[2239]	aaacugccaa aagaaaaagg auccacaaga guucaugccc ugaacaaugu caacaaggca	240
[2240]	cugcggguuu ugcagaacaa uaauguugau uuagugaaua uuggaaguac ugacaucgua	300
[2241]	gauggaaauc auaaacugac ucuugguuug auuuggaaua uauuccucca cuggcagguc	360
[2242]	aaaaauguaa ugaaaaauau cauggcugga uugcaacaaa ccaacaguga aaagauucuc	420
[2243]	cugagcuggg uccgacauc aacucguaau uauccacagg uuaauguaau caacuucacc	480
[2244]	accagcuggu cugauggccu ggcuuugaau gcucucaucc auagucouag gccagaccua	540
[2245]	uuugacugga auaguguggu uugccagcag ucagccacac aacgacugga acaugcauuc	600
[2246]	aacaucgcca gauaucauu aggcouagag aaacuacucg auccugaaga uguugauacc	660
[2247]	accuauccag auaagaaguc caucuuaaug uacaucacau cacucuucca aguuuugccu	720
[2248]	caacaaguga gcauugaagc cauccaggaa guggaaaugu ugccaaggcc accuaaagug	780
[2249]	acuaaagaag aacauuuuca guuacaucou caaauagcacu auucucaaca gaucacgguc	840
[2250]	agucuagcac agggauauga gagaacuucu uccccuaagc cucgauuca gagcuugcc	900
[2251]	uacacacagg cugcuuauug caccaccucu gaccuacac ggagcccauu uccuucacag	960
[2252]	cauuuggaag cuccugaaga caagucuuu ggcaguucau ugauggagag ugaaguuaac	1020
[2253]	cuggaccguu aucaaacagc uuuagaagaa guuuuauugc ggcuucuuuc ugcugaggac	1080
[2254]	acauugcaag cacaaggaga gauuuuauu gauguggaag uggugaaaga ccaguuucau	1140
[2255]	acucaugagg gguacaugau ggauuuugaca gcccaucagg gccggguugg uaaauuucua	1200
[2256]	caauugggaa guaagcugau uggaacagga aaauuauagc aagaugaaga aacugaagua	1260
[2257]	caagagcaga ugaauccucu aaauucaaga ugggaauugc ucaggguagc uagcauggaa	1320
[2258]	aaacaaagca auuuacauag aguuuuuauug gaucuccaga aucagaaacu gaaagaguug	1380
[2259]	aaugacuggc uaacaaaaac agaagaaaga acaaggaaaa uggaggaaga gccucuugga	1440
[2260]	ccugaucuug aagaccuaaa acgccaagua caacaacaua aggugcuuca agaagaucua	1500
[2261]	gaacaagaac aagucagggu caauucucuc acucacauug uggugguagu ugaugaauuc	1560
[2262]	aguggagau acgcaucugc ugcuuuggaa gaacaacuaa agguauuggg agaucgaug	1620
[2263]	gcaaacuucu guagauggac agaagaccgc uggguucuuu uacaagacau ccuucuaaa	1680
[2264]	uggcaacguc uuacugaaga acagugccuu uuuagugcau ggcuucaga aaaagaagau	1740
[2265]	gcagugaaca agauucacac aacuggcuuu aaagaucuaa augaaaugu aucaagucuu	1800
[2266]	caaaaacugg ccguuuuaaa agcggaucua gaaaagaaaa agcaauccau gggcaaacug	1860
[2267]	uauucacuca aacaagaucu ucuuucuaa cugaagaaua agucagugac ccagaagacg	1920

[2268]	gaagcauggc	uggauaacuu	ugcccggugu	ugggauaaau	uaguccaaaa	acuugaaaag	1980
[2269]	aguacagcac	agauuucaca	ggaaauuucu	uauuguccuu	cuacuauuuu	gacugaaauc	2040
[2270]	acucaugucu	cacaagcccu	auuagaagug	gaacaacuuc	uccaauugucc	ugaccucugu	2100
[2271]	gcuaaggacu	uugaagaccu	cuuuuagcaa	gaggagucuc	ugaagaauau	aaaagauagu	2160
[2272]	cuacaacaaa	gcucaggucg	gauugacauu	auucauagca	agaagacagc	agcauugcaa	2220
[2273]	agugcaacgc	cuguggaaaag	ggugaagcua	caggaagcuc	ucucccagcu	ugauuuccaa	2280
[2274]	ugggaaaaag	uuacaaaaau	guacaaggac	cgacaagggc	gauuugacag	auccguugag	2340
[2275]	aaauggcggc	guuuucauuu	ugauauaaag	auuuuuauuc	aguggcuaac	agaagcugaa	2400
[2276]	caguuuuca	gaaagacaca	aauccugag	aaugggaac	augcuaaaau	caaauugguu	2460
[2277]	cuuaaggaac	uccaggaugg	cauugggcag	cggcaaacug	uugucagaac	auugaauuca	2520
[2278]	acuggggaag	aaauuuuca	gcaauccuca	aaaacagaug	ccaguauucu	acaggaaaaa	2580
[2279]	uugggaagcc	ugaaucugcg	guggcaggag	gucugcaaac	agcugucaga	cagaaaaaag	2640
[2280]	aggcuagaag	aacaaaagaa	uauuuuguca	gaauuucaa	gagauuuuuu	ugaauuuugu	2700
[2281]	uuauugguug	aggaagcaga	uaacauugcu	aguauccac	uugaaccugg	aaaagagcag	2760
[2282]	caacuuaaag	aaaagcuuga	gcaagucaag	uuacuggugg	aagaguugcc	ccugcgccag	2820
[2283]	ggauuccuca	aacaauuuuu	ugaaacugga	ggaccgcugc	uuguaagugc	ucccauaagc	2880
[2284]	ccagaagagc	aagauaaacu	ugaaaauaag	cucaagcaga	caaauccuca	guggauaaag	2940
[2285]	guuuccagag	cuuuccuga	gaaacaagga	gaaauugaag	cucaauuuuu	agaccuuggg	3000
[2286]	cagcuugaaa	aaaagcuuga	agaccuugaa	gagcaguuaa	aucaucugcu	gcugugguuu	3060
[2287]	ucuccuauuu	ggaucaguu	ggaaauuuuu	aaccaacca	accaagaagg	accuuugac	3120
[2288]	guuaaggaaa	cugaaauagc	aguucaagcu	aaacaaccgg	auguggaaga	gauuuugucu	3180
[2289]	aaagggcagc	auuuguacaa	ggaaaaacca	gccacucagc	cagugaagag	gaaguuagaa	3240
[2290]	gaccuguccu	cugaguggaa	ggcgguaaac	cguuuacuuc	aagagcugag	ggcaaagcag	3300
[2291]	ccugaccuag	cuccuggacu	gaccacuauu	ggagccucuc	cuacucagac	uguuacucug	3360
[2292]	gugacacaac	cugugguuac	uaaggaaacu	gccaucucca	aacuagaaau	gccaucuucc	3420
[2293]	uugauguugg	agguaccugc	ucuggcagau	uucaaccggg	cuuggacaga	acuuaccgac	3480
[2294]	uggcuuucuc	ugcuugauca	aguuuuuuuu	ucacaacgcg	ugaugguggg	cgaccuugag	3540
[2295]	gauaucaacg	agaugaucau	caagcagaag	gcaacaauuc	aggauuugga	acagaggcgu	3600
[2296]	ccccaguugg	agaacucau	uaccgcugcc	caaaauuuuga	aaaacaagac	cagcaaucaa	3660
[2297]	gaggcuagaa	caucauuac	ggaucgaaau	gaaagaauuc	agaaucaug	ggaugaagua	3720
[2298]	caag	3724					
[2299]	<210>	130					
[2300]	<211>	3362					
[2301]	<212>	RNA					
[2302]	<213>	人工序列					
[2303]	<220>						
[2304]	<223>	Ct-miniDys (Δ H2-R15)					
[2305]	<400>	130					
[2306]	aacaccuuca	gaaccggagg	caacaguuga	augaaauguu	aaaggauuca	acacauggc	60
[2307]	uggaagcuua	ggaagaagcu	gagcaggucu	uaggacaggc	cagagccaag	cuggagucuu	120
[2308]	ggaaggaggg	ucccuauaca	guagaugcaa	uccaaaagaa	aucaacagaa	accaagcagu	180
[2309]	uggccaaaga	ccuccgccag	uggcagacaa	auguagaugu	ggcaauugac	uuggcccuga	240

[2310]	aacuucuccg	ggauuuuucu	gcagaugaua	ccagaaaagu	ccacaugaua	acagagaaua	300
[2311]	ucaaugccuc	uuggagaagc	auucauaaaa	gggugaguga	gcgagaggcu	gcuuuggaag	360
[2312]	aaacucauag	auuacugcaa	caguuccecc	uggaccugga	aaaguuucuu	gccuggcuua	420
[2313]	cagaagcuga	aacaacugcc	aauguccuac	aggaugcuac	ccguaaggaa	aggcuccuag	480
[2314]	aagacuccaa	gggaguaaaa	gagcugauga	aacaauaggca	agaccuccaa	ggugaaaauug	540
[2315]	aagcucacac	agauguuuau	cacaaccugg	augaaaacag	ccaaaaaauc	cugagaucce	600
[2316]	uggaagguuc	cgaugaugca	guccuguuac	aaagacguuu	ggauaacaug	aacuucaagu	660
[2317]	ggagugaacu	ucggaaaaag	ucucucaaca	uuaggucceca	uuuggaagcc	aguucugacc	720
[2318]	aguggaagcg	ucugcaccuu	ucucugcagg	aacuucuggu	guggcuacag	cugaaagaug	780
[2319]	augaauuaag	ccggcaggca	ccuauuggag	gcgacuuucc	agcaguucag	aagcagaacg	840
[2320]	augugcauag	ggccuucag	agggaauuga	aaacuaaaga	accuguaauc	augaguacuc	900
[2321]	uugagacugu	acgaauuuu	cugacagagc	agccuuugga	aggacuagag	aaacucuacc	960
[2322]	aggagcccag	agagcugccu	ccugaggaga	gagcccagaa	ugucacucgg	cuucucagaa	1020
[2323]	agcaggcuga	ggaggucaau	acugaguggg	aaaaauugaa	ccugcacucc	gcugacuggc	1080
[2324]	agagaaaaau	agaugagacc	cuugaaagac	uccgggaacu	ucaagaggcc	acggaugagc	1140
[2325]	uggaccucaa	gcugcgccaa	gcugagguga	ucaagggauc	cuggcagccc	gugggcgauc	1200
[2326]	uccucauuga	cucucuccaa	gaucaccugg	agaaagucua	ggcacuucga	ggagaaaauug	1260
[2327]	cgccucugaa	agagaacgug	agccacguca	augaccuugc	ucgccagcuu	accacuuugg	1320
[2328]	gcauucagcu	cucaccguau	aaccucagca	cucuggaaga	ccugaacacc	agauggaagc	1380
[2329]	uucugcaggu	ggccgucgag	gaccgaguca	ggcagcugca	ugaagcccac	agggacuuug	1440
[2330]	guccagcauc	ucagcacuuu	cuuuccacgu	cuguccaggg	ucccugggag	agagccaucu	1500
[2331]	cgccaaacaa	agugcccuac	uauaucaacc	acgagacuca	aacaacuugc	ugggaccauc	1560
[2332]	ccaaaugac	agagcucuc	cagucuuuag	cugaccugaa	uaaugucaga	uucucagcuu	1620
[2333]	auaggacugc	caugaaacuc	cgaagacugc	agaaggcccu	uugcuuggau	cucuugagcc	1680
[2334]	ugucagcugc	augugaugcc	uuggaccagc	acaaccucaa	gcaaaaugac	cagcccagg	1740
[2335]	auauccugca	gauuauuaau	uguuugacca	cuauuuuuga	ccgccuggag	caagagcaca	1800
[2336]	acaauuuggu	caacguccu	cucugcgugg	auaugugucu	gaacuggcug	cugaanguuu	1860
[2337]	augauacggg	acgaacaggg	aggauccgug	uccugucuuu	uaaaacuggc	aucuuuccc	1920
[2338]	uguguaaagc	acauuuggaa	gacaaguaca	gauaccuuuu	caagcaagug	gcaaguucua	1980
[2339]	caggauuuug	ugaccagcgc	aggcugggcc	uccuucugca	ugauucuauc	caaaauccaa	2040
[2340]	gacaguuggg	ugaaguugca	uccuuugggg	gcaguaacau	ugagccaagu	guccggagcu	2100
[2341]	gcuuccaauu	ugcuauuaau	aagccagaga	ucgaagcggc	ccucuuccua	gacuggauga	2160
[2342]	gacuggaacc	ccaguccaug	guguggcugc	ccguccugca	cagaguggcu	gcugcagaaa	2220
[2343]	cugccaagca	ucaggccaaa	uguaacaucu	gcaaaagagug	uccaaucauu	ggauucaggu	2280
[2344]	acaggagucu	aaagcacuuu	aauuuagaca	ucugccaaag	cugcuuuuuu	ucuggucgag	2340
[2345]	uugcaaaagg	ccauaaaaug	cacuauccca	ugguggaaua	uugcacuccg	acuacaucag	2400
[2346]	gagaagaugu	ucgagacuuu	gccaaagguac	uaaaaaacaa	auuucgaacc	aaaagguauu	2460
[2347]	uugcgaagca	uccccgaaug	ggcuaccugc	cagugcagac	ugucuuagag	ggggacaaca	2520
[2348]	uggaaacucc	cguuacucug	aucaacuucu	ggccaguaga	uucugcgccu	gccucguccc	2580
[2349]	cucagcuuuc	acacgaugau	acucauucac	gcauugaaca	uuauugcuagc	aggcuagcag	2640
[2350]	aaauggaaaa	cagcaaugga	ucuuuacuaa	augauagcau	cucuccuauu	gagagcauag	2700
[2351]	augaugaaca	uuuguuaauc	cagcauuacu	gccaaaguuu	gaaccaggac	ucccccuga	2760

[2352] gccagccucg uaguccugcc cagaucuuga uuuccuuaga gagugaggaa agaggggagc 2820
[2353] uagagagaau ccuagcagau cuugaggaag aaaacaggaa ucugcaagca gaauaugacc 2880
[2354] gucuaaagca gcagcacgaa cauaaaggcc ugucceccacu gccgucceccu ccugaaauga 2940
[2355] ugcccaccuc uccccagagu ccccgggagc cugagcucac ugcugaggcc aagcuacugc 3000
[2356] gucaacacaa aggccgccug gaagccagga ugcaaaucuu ggaagaccac aaauaacagc 3060
[2357] uggagucaca guuacacagg cuaaggcagc ugcuggagca accccaggca gaggccaaag 3120
[2358] ugaauaggac aacggugucc ucuccuucua ccucucuaca gagguccgac agcagucagc 3180
[2359] cuaugcugcu ccgagugguu ggcagucaaa cuucggacuc caugggugag gaagaucuuc 3240
[2360] ucaguccucc ccaggacaca agcacaggu uagaggaggu gauggagcaa cucaacaacu 3300
[2361] ccuuccuag uucaagagga agaaauaccc cuggaaagcc aaugagagag gacacaaugu 3360
[2362] aa 3362

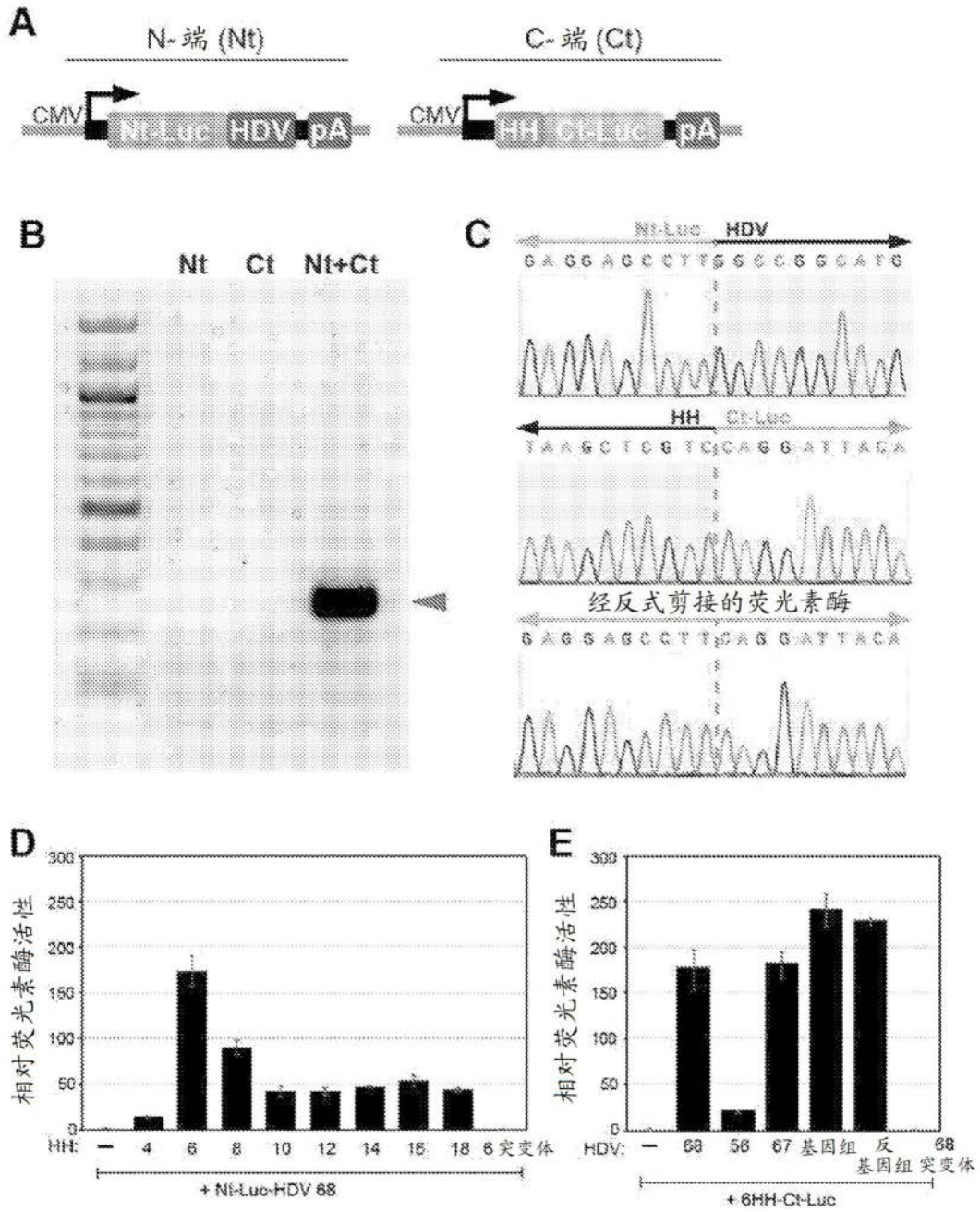


图2

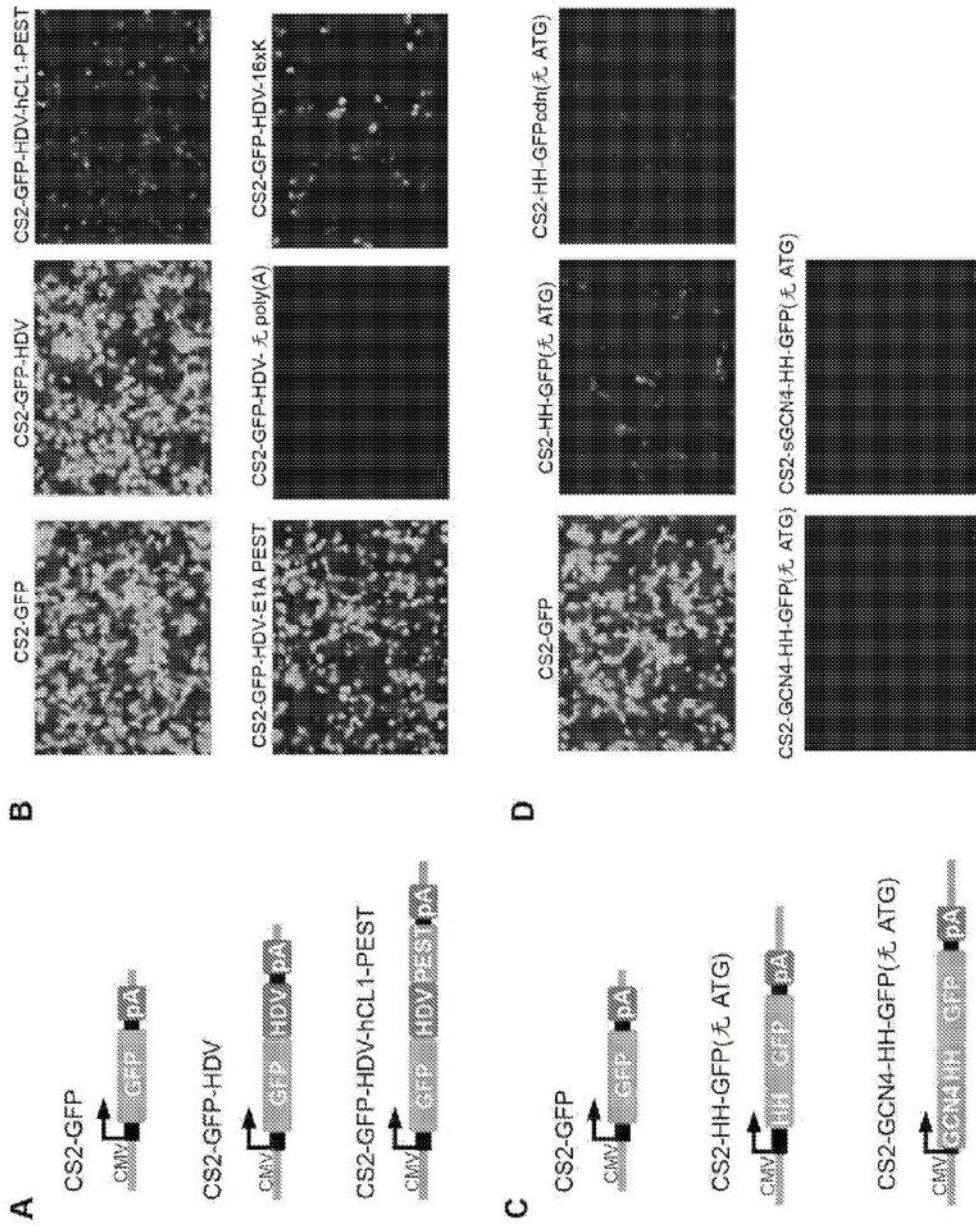


图3

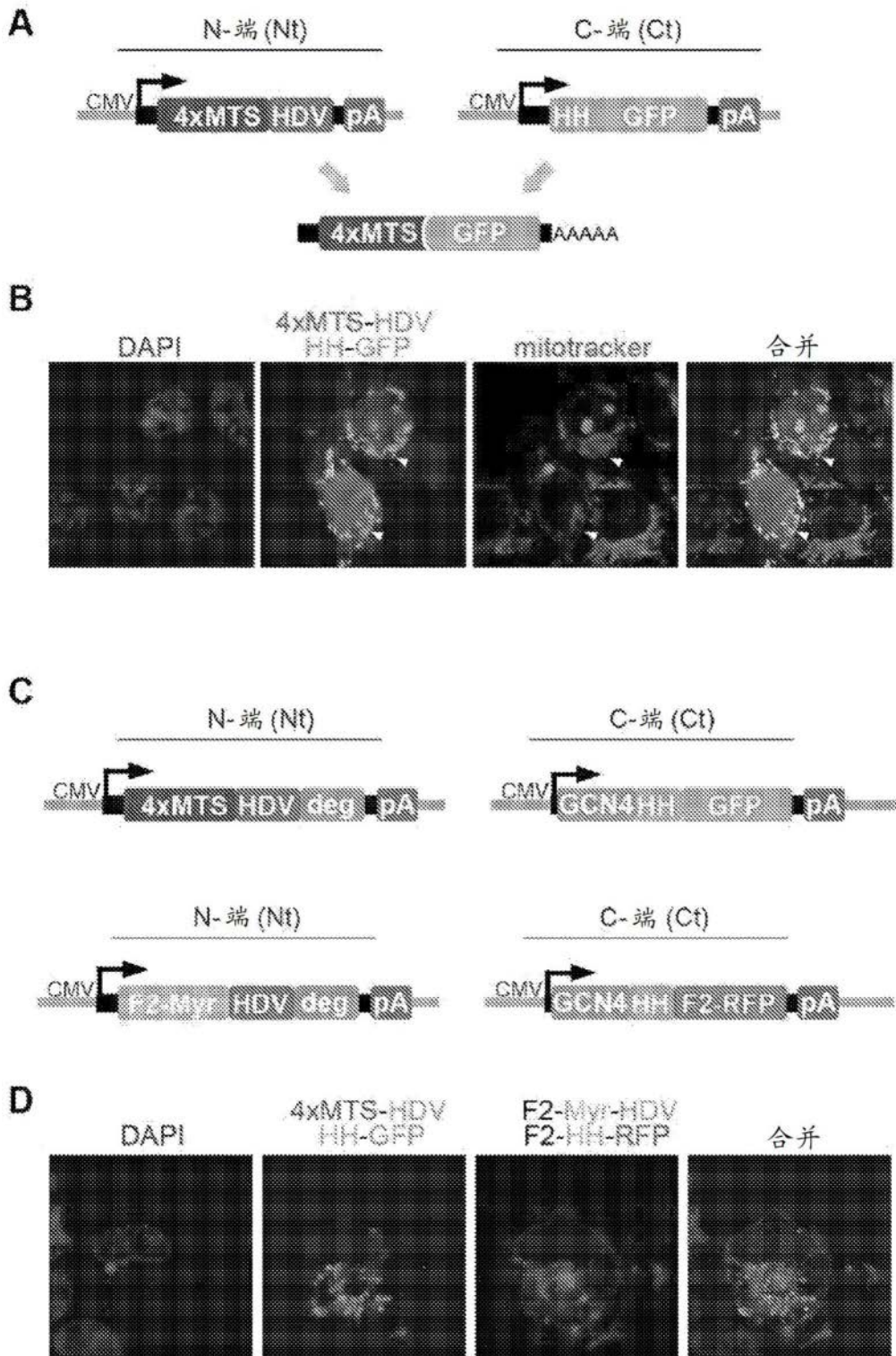


图4

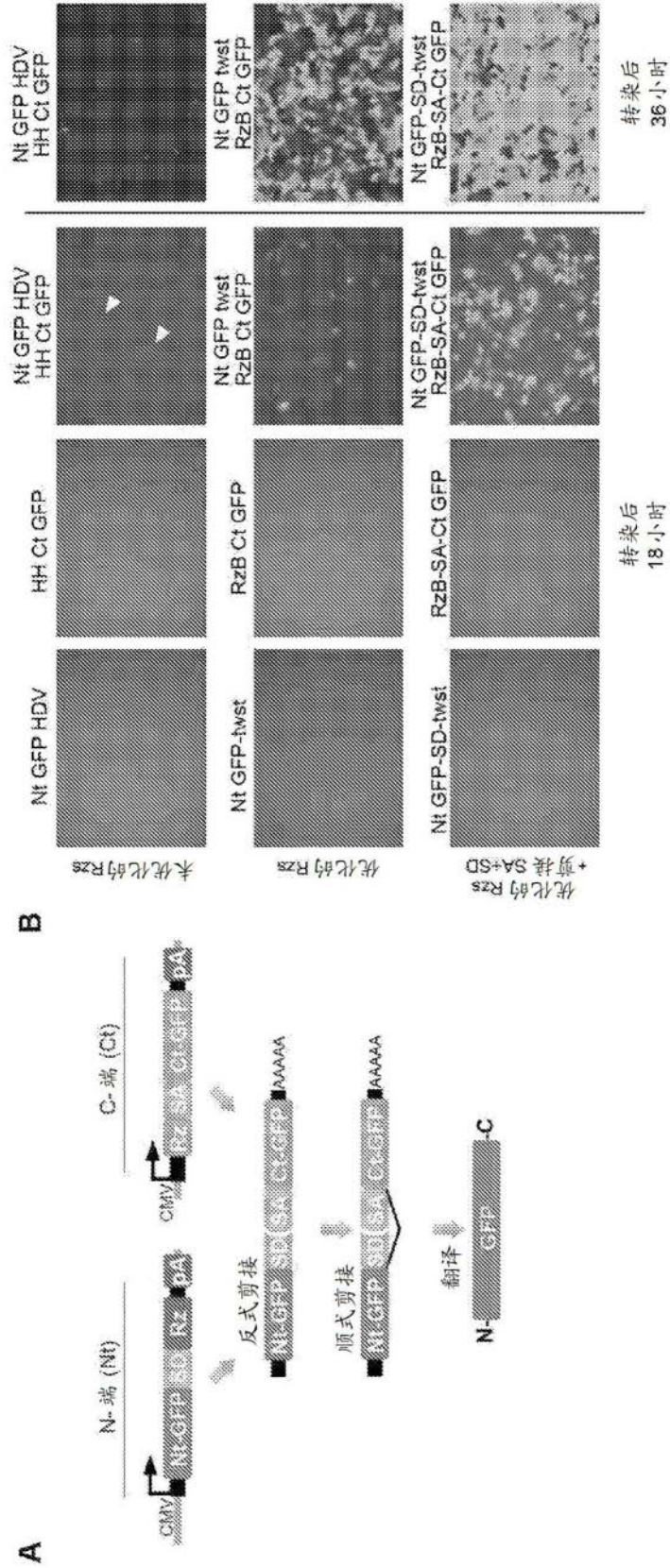


图5

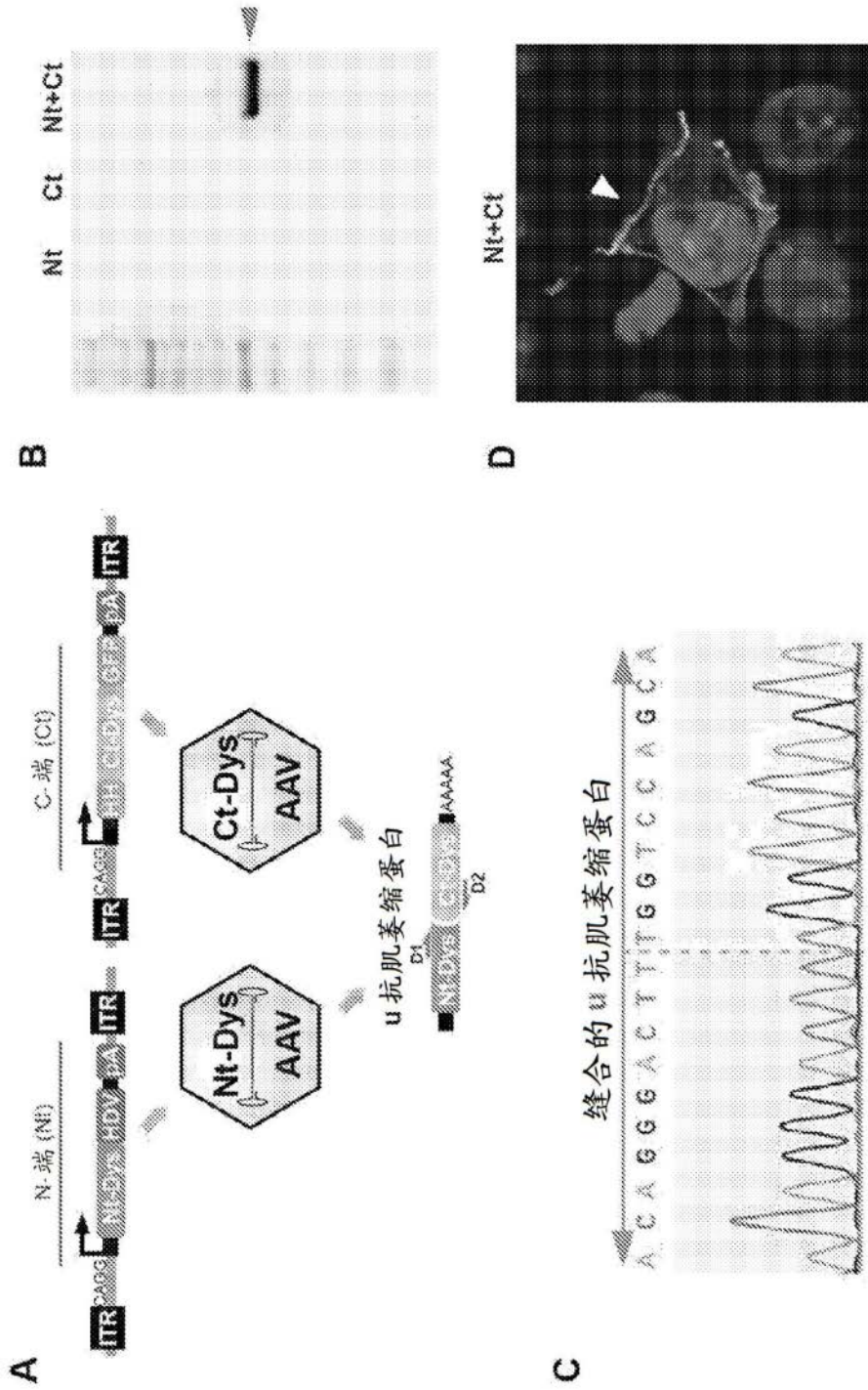


图6

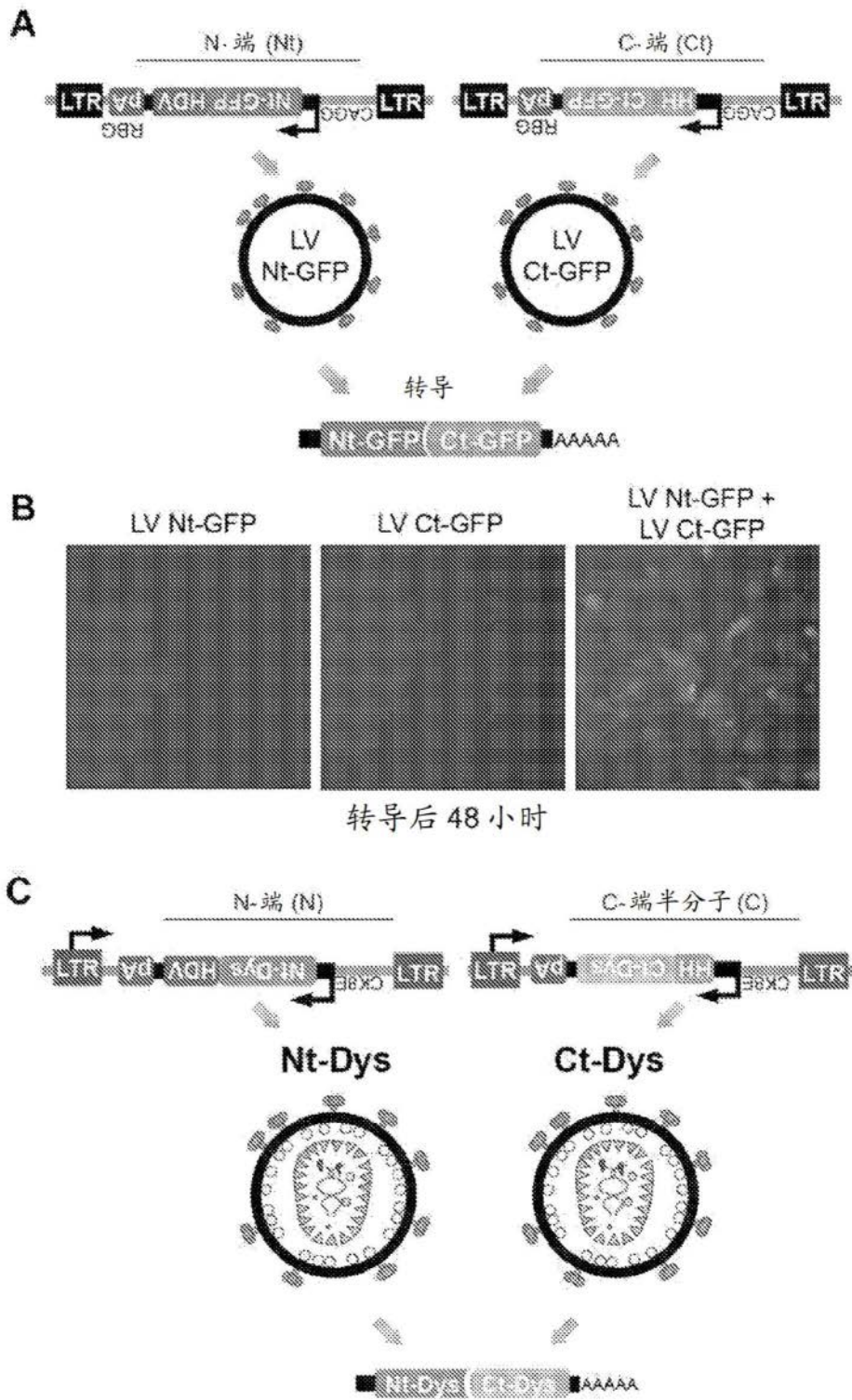


图7

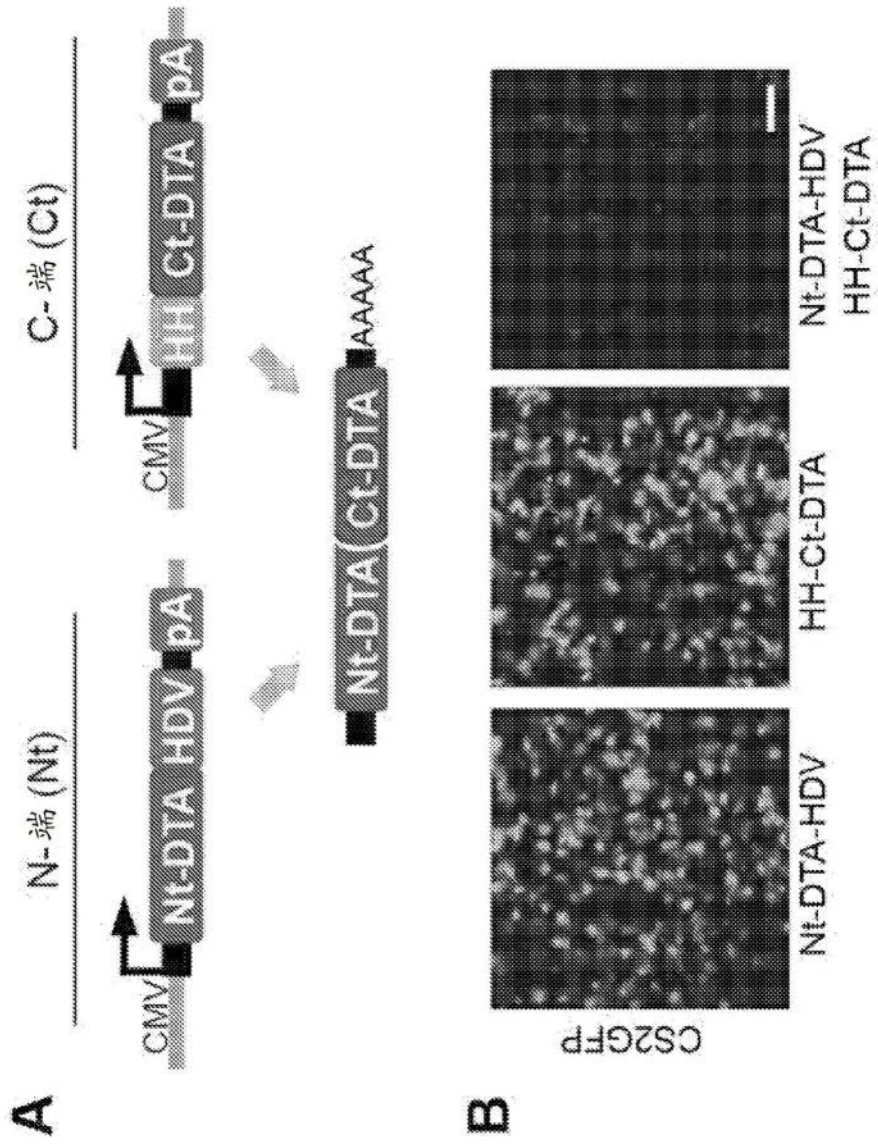


图8

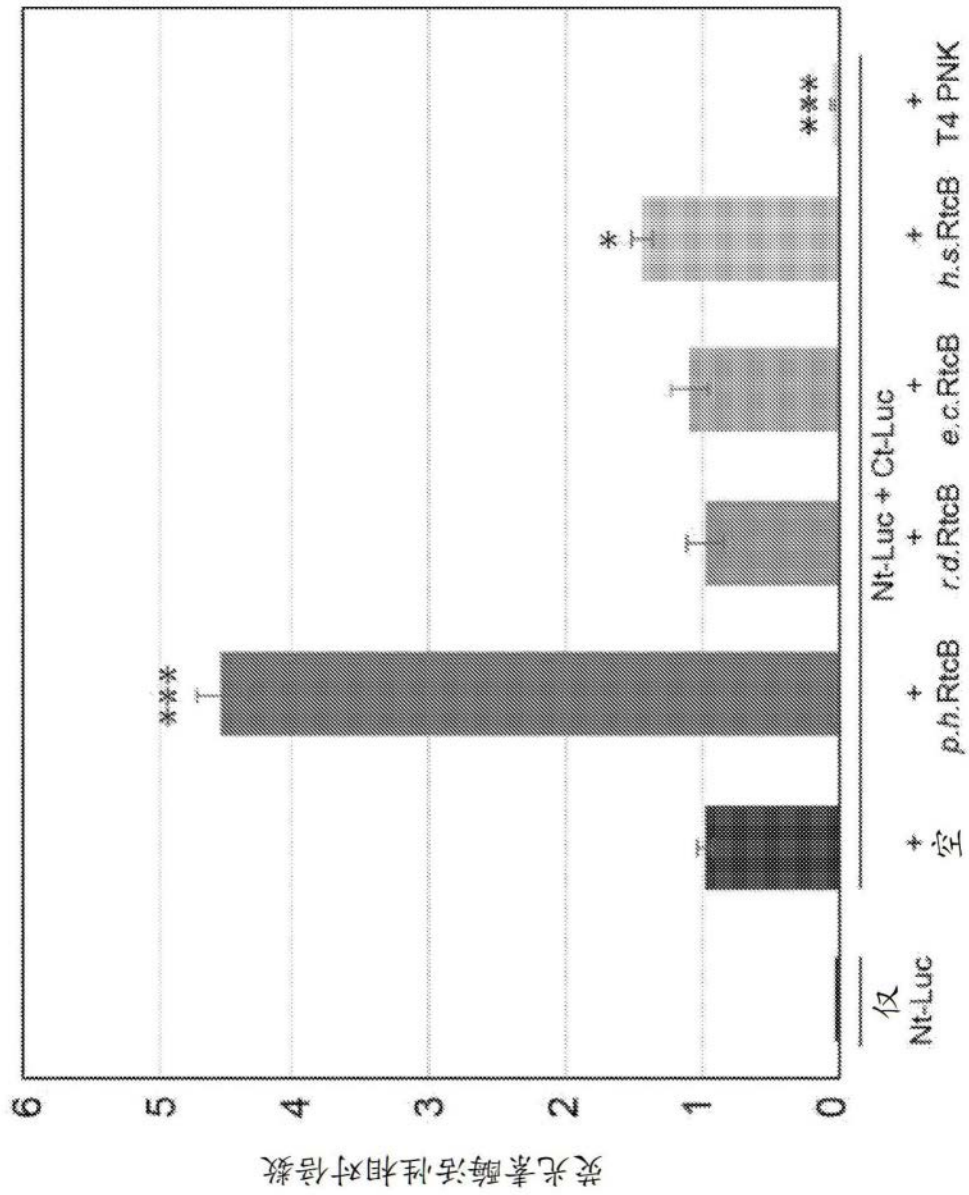


图9

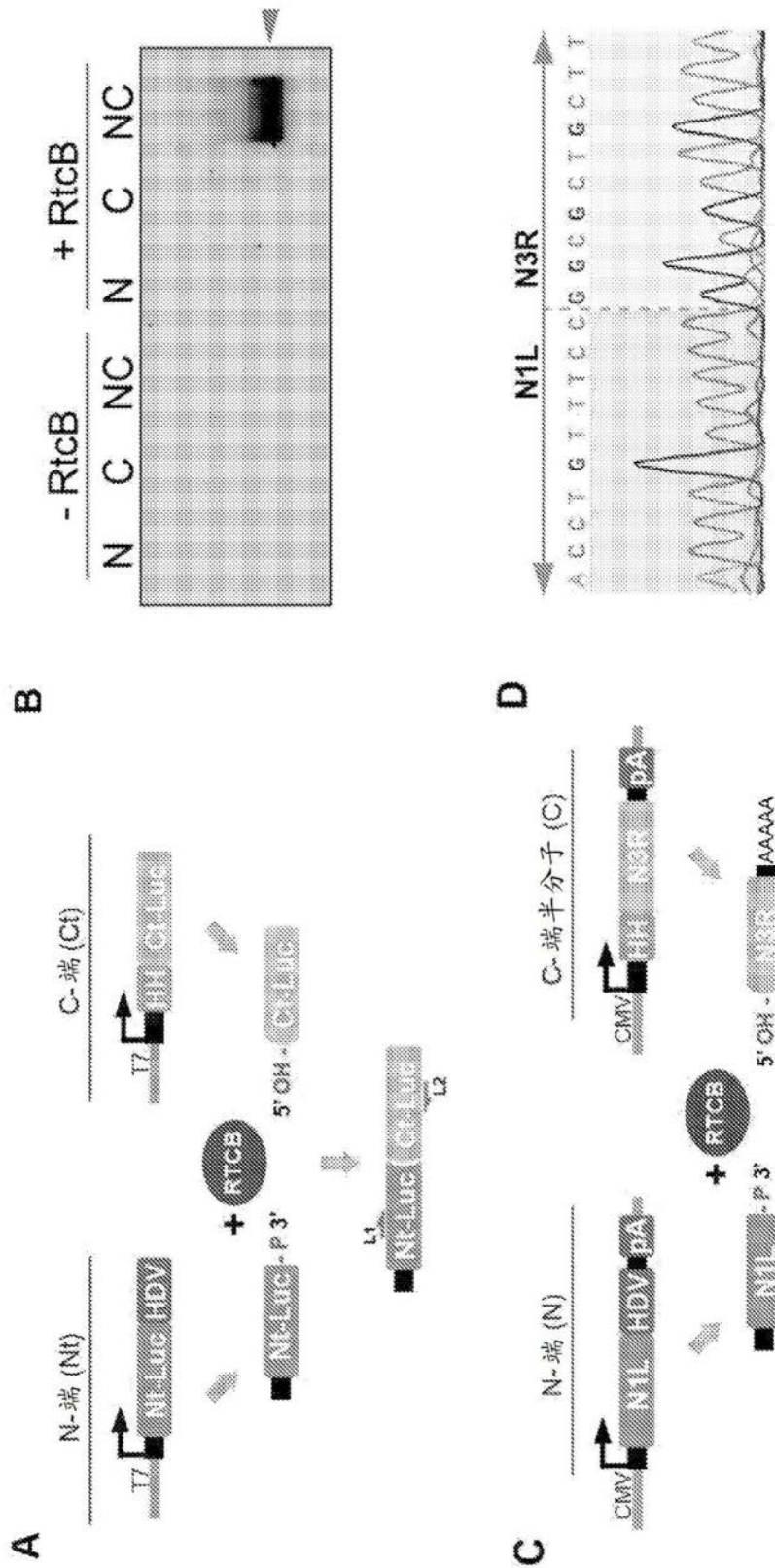


图10

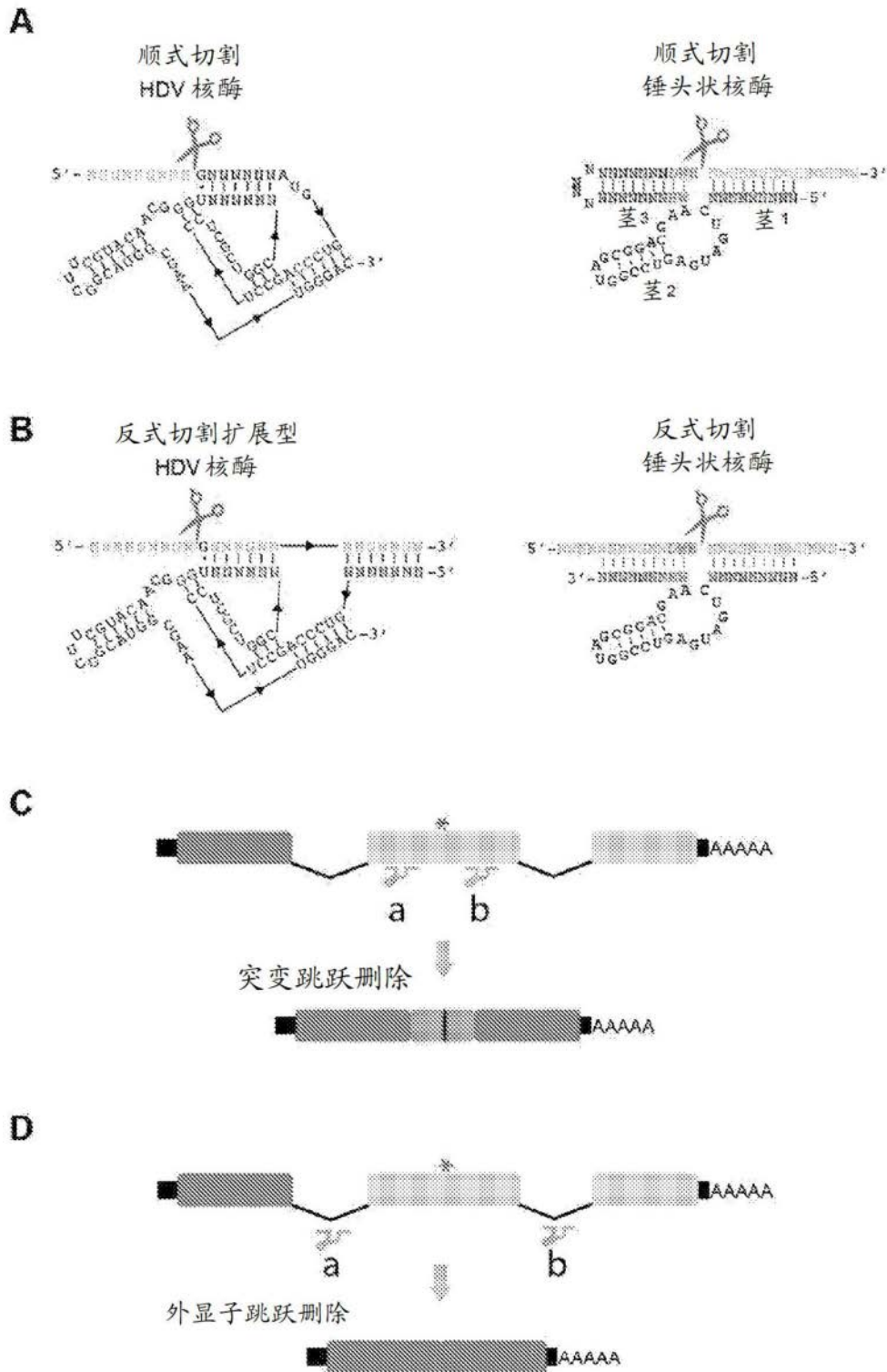


图12

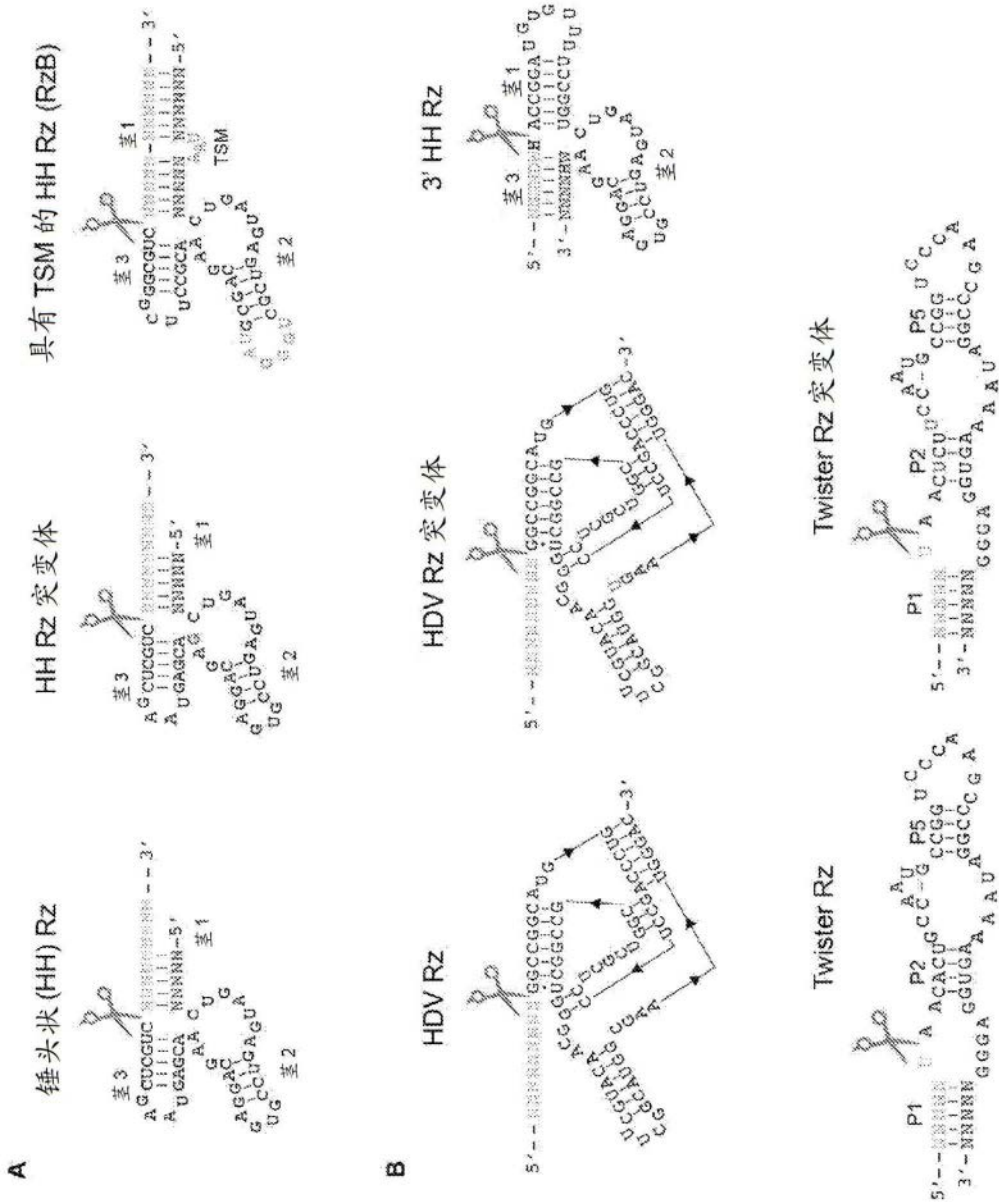


图13

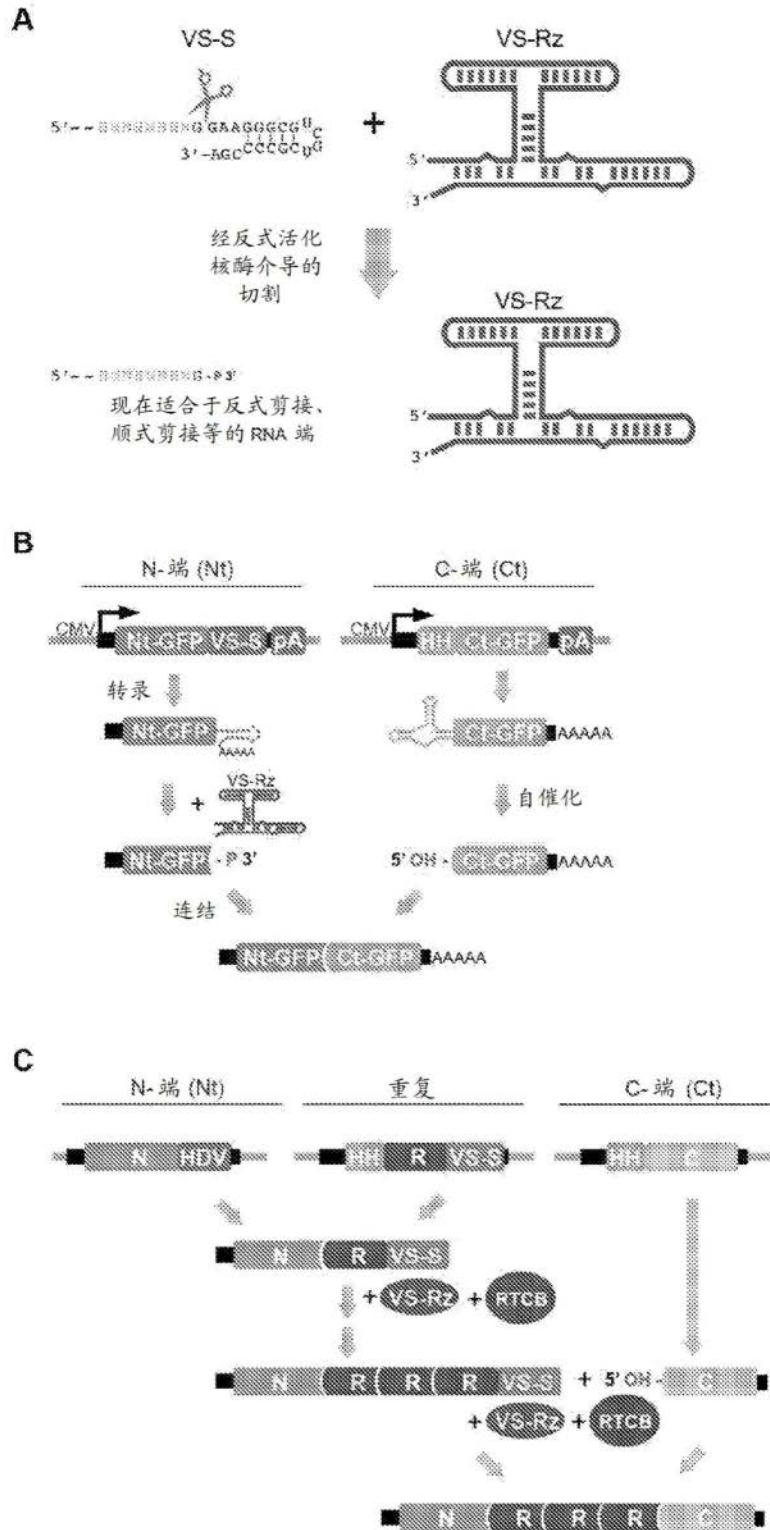


图14

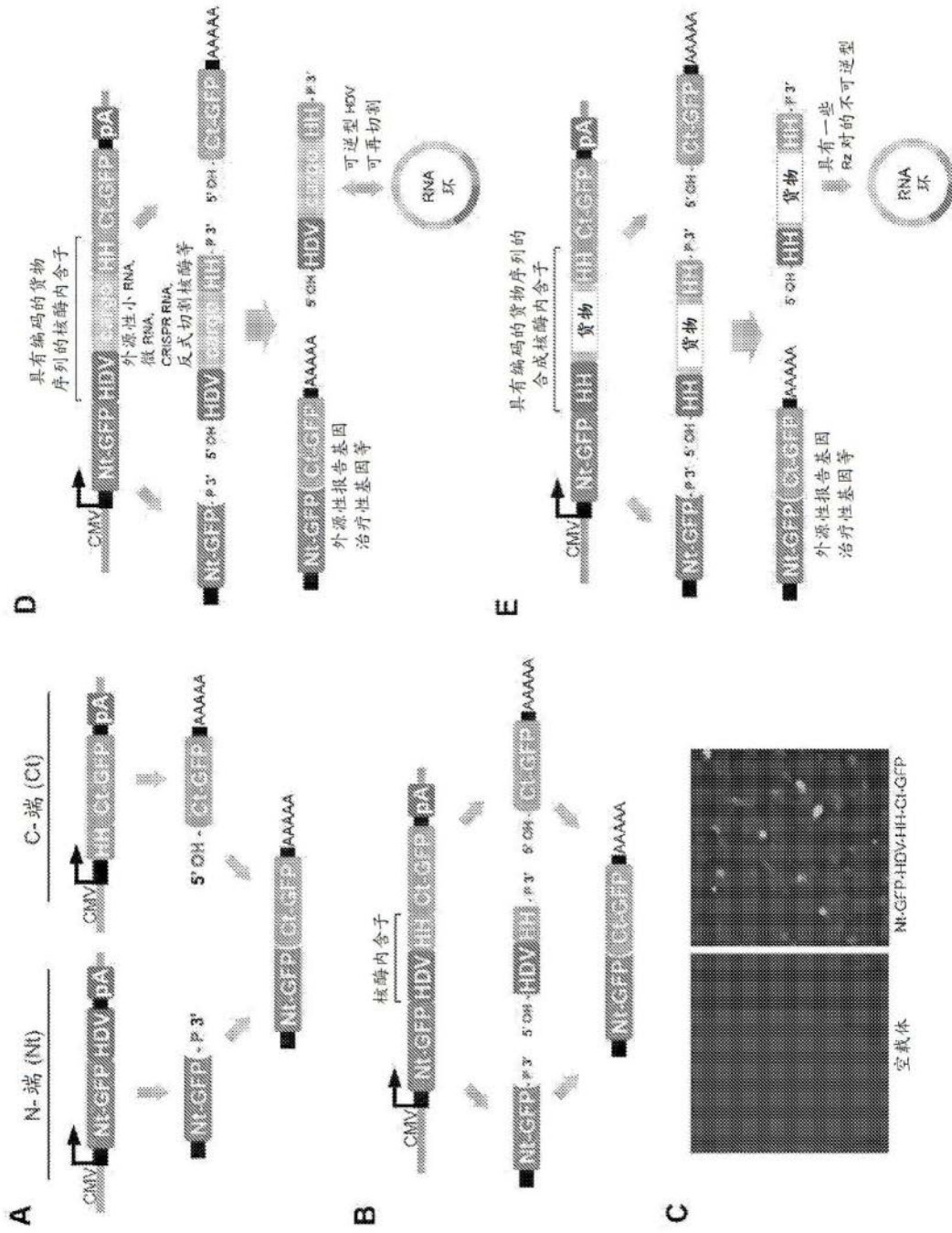


图15

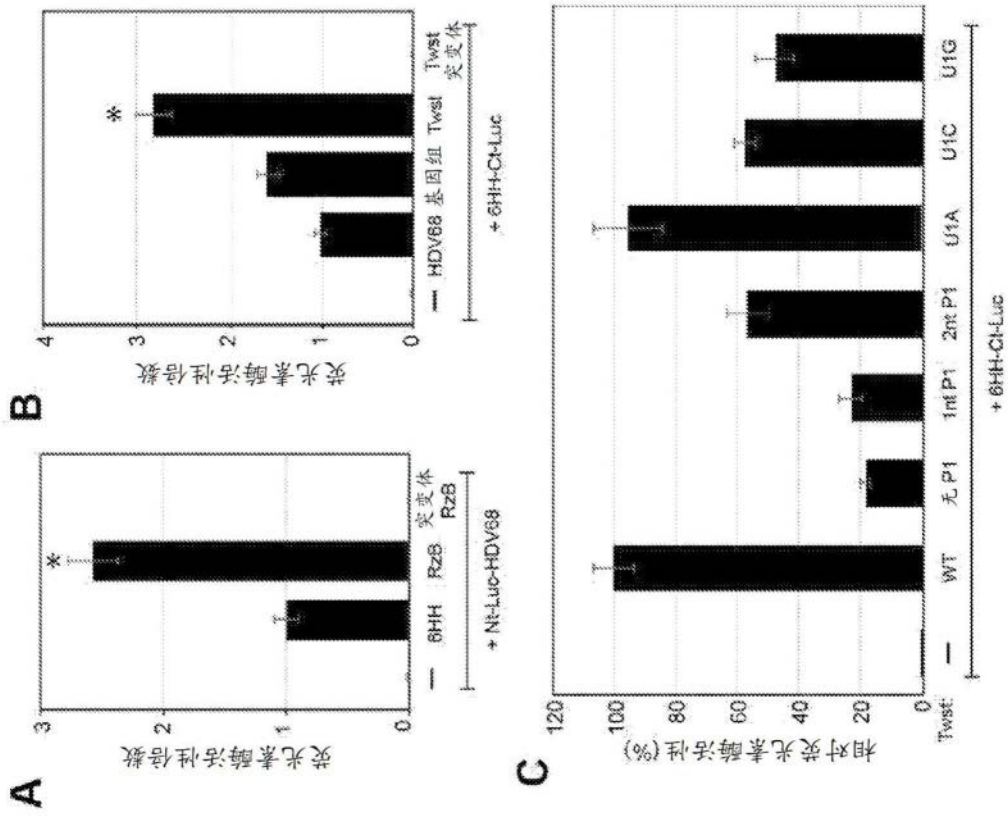


图16