



(12) 发明专利

(10) 授权公告号 CN 111770994 B

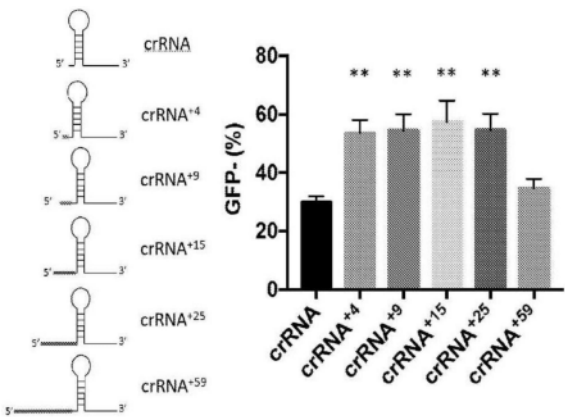
(45) 授权公告日 2025. 03. 18

(21) 申请号 201880077637.7  
(22) 申请日 2018.10.02  
(65) 同一申请的已公布的文献号  
    申请公布号 CN 111770994 A  
(43) 申请公布日 2020.10.13  
(30) 优先权数据  
    62/567,123 2017.10.02 US  
    62/617,138 2018.01.12 US  
    62/697,327 2018.07.12 US  
(85) PCT国际申请进入国家阶段日  
    2020.05.29  
(86) PCT国际申请的申请数据  
    PCT/US2018/054027 2018.10.02  
(87) PCT国际申请的公布数据  
    W02019/070762 EN 2019.04.11  
(73) 专利权人 基因编辑有限公司  
    地址 美国加利福尼亚

(72) 发明人 K·李  
(74) 专利代理机构 永新专利商标代理有限公司  
    72002  
    专利代理师 左路 区斌  
(51) Int.Cl.  
    C12N 15/113 (2006.01)  
    C12N 15/85 (2006.01)  
    C12N 15/90 (2006.01)  
(56) 对比文件  
    CN 106244591 A,2016.12.21  
    Rongfang等.Generation of targeted  
mutant rice using a CRISPR-Cpf1 system.  
《Plant Biotechnology Journal》.2017,图1.  
    Bin等.Engineering CRISPR-Cpf1 crRNAs  
and mRNAs to maximize genome editing  
efficiency.《Nat Biomed Eng》.2017,图1-图3.  
审查员 王小玉  
权利要求书1页 说明书27页  
序列表13页 附图22页

(54) 发明名称  
    修饰的CPF1引导RNA

(57) 摘要  
    本发明提供了包含Cpf1 crRNA、Cpf1 crRNA  
5'的加工序列和加工序列5'的延伸序列的核酸。  
本发明还提供了包含核酸、载体和任选地Cpf1的  
组合物。另外,本发明提供了遗传修饰真核靶细  
胞的方法,其包括使真核靶细胞与核酸或组合物  
接触,以遗传修饰细胞中的靶核酸。



1. 一种核酸,其包含Cpf1 crRNA、位于所述crRNA5'的至少6个核苷酸以及不超过60个核苷酸的延伸序列、以及任选地所述crRNA和所述延伸序列之间的加工序列,其中所述加工序列是被Cpf1自切割的序列,其中所述Cpf1 crRNA包含靶向序列5'的茎环结构域,

且所述延伸不包含所述加工序列、适体或所述Cpf1 crRNA的序列。

2. 权利要求1的核酸,其中所述核酸包含加工序列,并且所述加工序列包含Cpf1阵列的直接重复序列的片段,其中所述直接重复序列包含crRNA序列部分和位于所述crRNA序列部分5'的加工部分,并且所述片段包含所述直接重复序列的加工部分的至少5个邻接核苷酸。

3. 权利要求2的核酸,其中所述加工序列包含所述直接重复序列的加工部分的至少10个核苷酸的片段。

4. 权利要求2的核酸,其中所述加工序列包含所述直接重复序列的整个加工部分。

5. 权利要求1的核酸,其中所述延伸序列包含10至60个核苷酸。

6. 权利要求1的核酸,其中所述核酸仅含有单个Cpf1 crRNA序列。

7. 权利要求1的核酸,其进一步包含所述延伸序列5'的第二加工序列和所述第二加工序列5'的第二延伸序列。

8. 权利要求1的核酸,其进一步包含与所述加工序列的5'或所述延伸序列的5'共价连接的供体核酸。

9. 权利要求1的核酸,其进一步包含与与所述延伸序列和/或加工序列杂交的供体核酸。

10. 权利要求1的核酸,其中所述核酸不包含加工序列。

11. 权利要求1的核酸,其中所述延伸序列包含自杂交序列,或其中所述延伸序列包含重复性三核苷酸基序。

12. 权利要求1的核酸,其中所述延伸序列或其一部分对核酸酶降解是抗性的,或其中所述延伸序列包含一个或多个修饰的核苷酸间键,或其中所述延伸序列包含一种或多种异种核酸(XNA)。

13. 权利要求1的核酸,其中所述核酸进一步包含附着至所述延伸序列的5'末端的生物素和/或抗生物素蛋白或链霉抗生物素蛋白分子。

14. 一种组合物,其包含权利要求1-13中任一项的核酸和载体。

15. 权利要求14的组合物,其中所述核酸在脂质体中,或所述核酸被聚合物纳米颗粒部分或全部包封、或者附着至金属或聚合物纳米颗粒。

16. 权利要求14或15的组合物,进一步包含Cpf1蛋白。

17. 权利要求16的组合物,其中所述Cpf1蛋白与所述核酸复合。

18. 一种体外或离体遗传修饰真核靶细胞的方法,其包括使所述真核靶细胞与Cpf1和权利要求1-13中任一项的核酸或权利要求14-17中任一项的组合物接触,以遗传修饰靶核酸。

19. 权利要求1-13任一项的核酸或权利要求14-17中任一项的组合物在制备用于治疗人类对象的遗传疾病或病症的药物中的用途。

## 修饰的CPF1引导RNA

[0001] 引入作为参考的以电子方式提交的材料

[0002] 相关申请的交叉引用

[0003] 本专利申请要求于2017年10月2日提交的美国临时专利申请号62/567,123;于2018年1月12日提交的美国临时专利申请号62/617,138;以及于2018年7月12日提交的美国临时专利申请号62/697,327的优先权,所述申请的全部公开内容在此引入作为参考。

### 背景技术

[0004] RNA引导的核酸内切酶已证明是用于多重细胞类型和微生物中的基因组改造的有效工具。RNA引导的核酸内切酶在靶核酸内生成位点特异性双链DNA断裂或单链DNA断裂。当靶核酸的切割在细胞内发生时,可以通过非同源末端连接(NHEJ)或同源性定向修复(HDR)来修复核酸中的断裂。

[0005] RNA引导的核酸内切酶及其基因编辑组分(例如,引导RNA)在体外和体内均直接递送到细胞内,具有作为治疗遗传疾病的治疗策略的巨大潜力。然而,目前,以合理效率将这些组分直接递送到细胞内是挑战性的。

[0006] 因此,需要鉴定新的组合物和相关方法,以改善将增强基因组改造的RNA引导的核酸内切酶的细胞递送及其它特性。本发明提供了此类组合物和相关方法。

### 发明内容

[0007] 本发明是包含具有延伸序列的Cpf1 crRNA的核酸。在一个方面,核酸包含Cpf1 crRNA和在Cpf1 crRNA的5'末端上的延伸序列,其中所述延伸序列包含少于约60个核苷酸。在另一个方面,核酸包含Cpf1 crRNA、Cpf1crRNA 5'的加工序列和加工序列5'的延伸序列。还提供的是包含核酸、载体和任选地Cpf1蛋白或编码该蛋白的载体的组合物。

[0008] 本发明还提供了用于遗传修饰真核靶细胞的方法。该方法涉及使真核靶细胞与包含如本文所述的Cpf1 crRNA的核酸接触。

[0009] 在下述节段中更详细地描述了本发明的这些和其它方面。

### 附图说明

[0010] 图1A是比较使用阳离子脂质体(lipofectamine)或电穿孔(核转染),递送与Cpf1复合的未修饰的crRNA的图。

[0011] 图1B和1C是比较Cpf1-crRNA复合物的阳离子脂质体介导的递送和NHEJ生成的图,所述Cpf1-crRNA复合物包含未修饰的crRNA(长41个核苷酸)和延伸的crRNA。

[0012] 图1D是41核苷酸(nt)的未修饰的crRNA和延伸的crRNA的结构示意图。箭头代表Cpf1切割位点。

[0013] 图1E是显示了关于图1D中所示的crRNA构建体的NHEJ效率的图。

[0014] 图1F是描绘了使用阳离子脂质体和用荧光染料标记的具有不同长度的crRNA的Cpf1 RNP的细胞递送的图。

[0015] 图2提供了示出了根据经由电穿孔递送至GFP-HEK细胞的5'延伸crRNA的GFP击倒的基因编辑效率的图(右图),以及crRNA的示意图(左图)。

[0016] 图3A是在AI9小鼠中的体内研究的示意图。

[0017] 图3B是腓肠肌注射部位和成像区段的示意图。

[0018] 图4A是关于使用电穿孔,与供体DNA一起递送的具有不同延伸的crRNA的HDR频率的图。

[0019] 图4B是关于与供体DNA一起递送的具有不同延伸的crRNA的GFP-细胞百分比的图,示出了使用电穿孔的NHEJ效率。

[0020] 图4C是关于与供体DNA一起递送的具有不同延伸的crRNA的BFP-细胞百分比的图,示出了使用电穿孔的NHEJ效率。

[0021] 图4D是关于使用电穿孔,连同和不连同与靶序列不具有同源性的单链DNA(ssDNA)一起递送的延伸crRNA的GFP-细胞百分比的图。

[0022] 图4E是示出了作为关于使用电穿孔的crRNA的GFP-细胞百分比的基因编辑的图,所述crRNA用与靶序列不具有同源性的100nt RNA和9nt RNA进行延伸。

[0023] 图4F是关于使用电穿孔的crRNA的GFP-细胞百分比的图,所述crRNA具有和不具有4nt延伸,并且进一步用化学部分进行修饰。

[0024] 图5A和5B提供了使crRNA和供体DNA缀合的示意图。

[0025] 图5C是凝胶电泳分离的图像,其示出了在用硫醇还原后,从缀合的crRNA/DNA分子中释放供体DNA和crRNA。

[0026] 图6是证实了在用PAsp(DET)(即阳离子聚合物)转染后,与HD-RNA缀合的Cpf1在GFP-HEK细胞中诱导NHEJ的图。

[0027] 图7是证实了在用PAsp(DET)(即阳离子聚合物)转染后,与HD-RNA缀合的Cpf1在GFP-HEK细胞中诱导HDR的图。

[0028] 图8提供了Cpf1蛋白的序列。

[0029] 图9提供了Cpf1加工序列的实例。

[0030] 图10A显示了与供体DNA缀合的crRNA的示意图。

[0031] 图10B示出了在与供体DNA缀合的crRNA中使用的序列。

[0032] 图10C是在原代Ai9成肌细胞中,使用电穿孔,用各种crRNA和Cpf1处理后的RFP+细胞百分比的图。

[0033] 图10D是用100nt DNA或RNA转染到原代Ai9成肌细胞内的RFP细胞百分比的图。

[0034] 图10E是使用电穿孔,用在靶向Serpina1基因的crRNA具有或不具有9nt延伸的Cpf1 RNP转染的HepG2细胞中的NHEJ效率的图。

[0035] 图11A示出了可以用于crRNA延伸中的RNA结构。

[0036] 图11B示出了可以用于提供各种RNA结构的三核苷酸重复。

[0037] 图11C示出了在接吻环(kissing loop)中crRNA的杂交延伸序列的相交,其可以用于形成crRNA多聚体。

[0038] 图11D示出了crRNA的杂交延伸序列的相交,以形成三聚体(小图(i))或八聚体(小图(ii))。

[0039] 图12显示了在表达BFP的HEK293T细胞中,各种Cpf1 crRNA的编辑效率(%BFP-)。

MS是在从5'末端开始的前三个核苷酸上的2'-OMe 3'-硫代磷酸酯修饰,+9du是在从5'末端开始的第9个核苷酸上的2'-脱氧修饰,+9S是在从5'末端开始的前9个核苷酸上的硫代磷酸酯修饰。在电穿孔后7天,用流式细胞术测量BFP敲除效率。平均值±S.E,n=3。通过斯氏t检验,所有延伸的crRNA都显示与未修饰的crRNA在统计上显著的差异,其p值小于0.05。

[0040] 图13A是未修饰的Cas9 sgRNA和Cpf1 crRNA的图解;和

[0041] 图13B和13C是显示了根据GFP击倒,Cas9 sgRNA和Cpf1 crRNA的相对活性的图。

[0042] 图14是用生物素和抗生物素蛋白修饰,并且与包含生物素的靶向分子连接的延伸crRNA的示意图。

[0043] 图15A是对具有延伸的crRNA进行的化学修饰的示意图。

[0044] 图15B是定量在血清中温育后的剩余crRNA量的图。

[0045] 图15C是在使用阳离子脂质体递送具有9nt延伸和化学修饰的crRNA后,GFP-细胞百分比的图。

[0046] 图15D是比较在使用电穿孔递送具有化学修饰的延伸crRNA和未修饰的延伸crRNA连同Cpf1后,GFP-细胞百分比的图。

[0047] 图16是比较Cpf1-crRNA复合物的阳离子聚合物介导的递送和NHEJ生成的图,所述Cpf1-crRNA复合物包含未修饰的crRNA(41nt)、9碱基对的延伸crRNA(总共50nt)、或59碱基对的延伸crRNA(总共100nt)。

## 具体实施方式

[0048] 本发明提供了用于Cpf1的修饰的引导核酸,称为“crRNA”,与常规crRNA分子相比,其具有增强的特性。如本文使用的,crRNA指核酸序列(例如,RNA),其与RNA引导的核酸内切酶Cpf1结合,并且将RNA引导的核酸内切酶靶向待由Cpf1切割的靶核酸内的特定位置。Cpf1是II类CRISPR/Cas系统的RNA引导的核酸内切酶,其涉及V型适应性免疫。Cpf1不像其它CRISPR酶一样需要tracrRNA分子,并且仅需要单个crRNA分子来起作用。Cpf1更喜欢“TTN”PAM基序,其位于其靶的5'上游。另外,关于Cpf1的切割位点错开约3-5个碱基,其产生了“粘性末端”(Kim等人,2016.“Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells”,2016年6月6日在线发表)。具有3-5bp突出端的这些粘性末端被认为促进NHEJ介导的连接,并且改善具有匹配末端的DNA片段的基因编辑。

[0049] 本领域技术人员将了解,Cpf1 crRNA可以来自任何物种或者衍生自或分离自任何来源的任何合成或天然存在的变体或直系同源物。即,Cpf1crRNA可以具有crRNA的所需元件(例如序列或结构),其识别(由其结合)来自任何细菌物种的任何Cpf1多肽或直系同源物、或其合成变体。Cpf1crRNA序列的实例在图9中提供;因此,例如,Cpf1 crRNA的实例包括包含SEQ ID NO:21-39中任一个的那些)。合成变体Cpf1的Cpf1 crRNA序列的实例是对应于通过Inscripta,Inc.(CO,USA)的MAD7Cpf1直系同源物的crRNA。Cpf1变体的另一个实例是经修饰为减少或消除RNA酶活性的Cpf1,例如通过在氨基酸球菌属(Acidaminococcus)Cpf1(AsCpf1)的H800A、K809A、K860A、F864A和R790A,或不同Cpf1直系同源物的相应位置处引入修饰(例如H800A突变或在相应位置处的H→A突变)。

[0050] 通常,crRNA包含靶向结构域和位于靶向结构域5'的茎环结构域。crRNA的总体长度并无特别限制,只要它可以将Cpf1引导至靶核酸内的特定位置。茎-环结构域一般为长度

约19-22个核苷酸(nt),而靶向/引导结构域为约14-25nt(例如,至少约14nt、15nt、16nt、17nt或18nt)的任何地方。在一些实施方案中,Cpf1 crRNA的总体长度可以具有20至100nt、20nt至90nt、20nt至80nt、20nt至70nt、20nt至60nt、20nt至55nt、20nt至50nt、20nt至45nt、20nt至40nt、20nt至35nt、20nt至30nt、或20nt至25nt的长度。

[0051] 本公开内容的一个方面提供了核酸,其包含Cpf1 crRNA、crRNA 5'的延伸序列、以及任选地加工序列,其可以位于crRNA与延伸序列之间、延伸序列内、或延伸序列的5'。

[0052] 延伸序列

[0053] 本发明的核酸包含位于crRNA 5'的延伸序列。延伸序列可以包含核酸的任何组合(即,任何序列)。在一个实施方案中,延伸序列增加了核酸分子的总体负电荷密度,并且改善了核酸包括crRNA的递送。

[0054] 在一些实施方案中,延伸序列可以在细胞中被切割一次。不希望受任何特定的理论或作用机制的束缚,认为Cpf1可以切割延伸序列。然而,在某些应用中,期望使用其中延伸序列不从Cpf1 crRNA中切割的构建体。因此,在一些实施方案中,延伸序列不能被Cpf1 crRNA切割。例如,延伸序列或其一些部分或区域可以包含一个或多个修饰的核苷酸间键合(修饰的“主链”),其对通过Cpf1 crRNA的切割是抗性的(例如,核酸酶抗性的)。修饰的核苷酸间键合的实例包括但不限于:硫代磷酸酯、二硫代磷酸酯、甲基膦酸酯、氨基磷酸酯、2'-O-甲基、2'-O-甲氧基乙基、2'-氟代、桥连核酸(BNA)、或磷酸三酯修饰的键以及其组合。延伸序列或其一些部分还可以包含合成核苷酸,例如对核酸酶抗性的异种核酸(xeno nucleic acid)(XNA)。XNA是这样的核酸,其中DNA或RNA的呋喃核糖环替换为五或六元修饰的核糖分子,例如1,5脱水己糖醇核酸(HNA)、环己烯基核酸(CeNA)和2'4'-C-(N-甲基氨基亚甲基)桥连核酸(BNA)、2'-O,4'-C-亚甲基- $\beta$ -D-核糖核酸或锁核酸(LNA)、ANA(阿拉伯糖核酸)、2'-氟-阿拉伯核糖核酸(FANA)和 $\alpha$ -L-苏呋喃糖基核酸(TNA)。此外,也可以使用其任何组合。

[0055] 延伸序列的长度并无特别限制,只要延伸序列增加总体负电荷密度。例如,延伸序列可以具有至少约2个核苷酸(nt)一直到约1000nt(例如,至少约2、3、4、5、6、7、8、9、10、15、20、25、30、35、40、45、50、55、60、65、70、75、80、85、90、95、100、200、300、400、500、600、700、800或900,且一直到约1000nt)的长度。在一个方面,延伸序列长度不超过约100个核苷酸,例如长度不超过约80个核苷酸、长度不超过约60个核苷酸、或长度不超过约40个核苷酸(例如,长度不超过约30个核苷酸或长度不超过约20个核苷酸)。关于长度的前述下限和上限中的任一个都可以表示为范围。也可以使用较短的序列(例如,不超过约15个核苷酸、或不超过约10个核苷酸)。在一些实施方案中,延伸序列包含至少约2个核苷酸,例如至少约4个核苷酸、至少约6个核苷酸、或甚至至少约9个核苷酸。前述中的任一个都可以表示为范围。因此,例如,延伸序列可以是约2-60个核苷酸(例如约2-40个核苷酸、约2-30个核苷酸、约2-20个核苷酸、约2-15个核苷酸、或约2-10个核苷酸)、约4-60个核苷酸(例如约4-40个核苷酸、约4-30个核苷酸、约4-20个核苷酸、约4-15个核苷酸、或约4-10个核苷酸);约6-60个核苷酸(例如约6-40个核苷酸、约6-30个核苷酸、约6-20个核苷酸、约6-15个核苷酸、或约6-10个核苷酸);或约9-60个核苷酸(例如约9-40个核苷酸、约9-30个核苷酸、约9-20个核苷酸、约9-15个核苷酸、或约9-10个核苷酸)。

[0056] 在一些实施方案中,延伸序列不具有除对核酸构建体赋予更大的总体负电荷密度

外的功能。在这个实施方案中,例如,延伸序列是随机或非编码序列。在一些情况下,例如当使用加工序列时,该序列可以在加工序列切割时被降解,并且从核酸构建体中释放。

[0057] 在其它实施方案中,延伸序列具有与对核酸构建体赋予更大的总体负电荷密度分开且隔开的功能。延伸序列可以具有任何另外的功能。例如,延伸序列可以提供用于另一种核酸,例如供体核酸的杂交位点。另外,在一些实施方案中,延伸序列可以是适体和/或促进细胞结合。然而,有时不期望募集除RNA引导的核酸内切酶外的蛋白质与引导RNA的结合。此外,适体序列通常具有复杂的折叠模式,其可能庞大且不紧凑。因此,在其它实施方案中,延伸序列不是适体序列。

[0058] 在一些实施方案中,延伸序列可以包含编码蛋白质的序列,所述蛋白质的表达在待编辑的靶细胞中是期望的。例如,延伸序列可以包含编码RNA引导的核酸内切酶的序列,例如与核酸构建体中使用的crRNA配对(即,识别并由其引导)的RNA引导的核酸内切酶。延伸序列可以包含例如RNA引导的核酸内切酶的mRNA的序列。


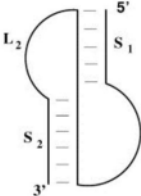
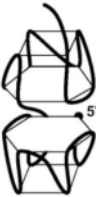

[0059] 在一些实施方案中,延伸部分自折叠(自杂交)以提供结构化的延伸部分。关于所提供结构的类型没有限制。延伸部分可以具有无规卷曲结构;然而,在一些实施方案中,延伸部分具有的结构比相同数目的核苷酸的无规卷曲结构更紧凑,其提供了更大的负电荷密度。通过增加延伸部分的总体长度,分子的负电荷增加。当使用更紧凑的结构时,分子的总体负电荷密度进一步增加。紧性或电荷密度可以根据凝胶电泳中的迁移率来确定。更特别地,如果对于在同一凝胶上一起运行的具有相同数目的核苷酸的两种核酸执行凝胶电泳,则具有较高迁移率(在凝胶中移动最远)的核酸被认为具有更紧凑的结构。

[0060] 在另一个实施方案中,延伸序列包括至少一个半稳定的发夹结构、稳定的发夹结构、假结结构、G-四链体结构、凸环结构、内环结构、分支环结构或其组合。这些类型的核苷酸结构是本领域已知的,并且在图11A中示意性地示出。应理解,图示仅仅出于示出一般结构的目的,而不预期是实际分子结构的详细图示。本领域技术人员认识到,发夹结构例如可以具有散布的非互补区域,其在结构中产生“凸起”或其它变化,并且其它所描绘的结构可以包括类似的变化。给定核苷酸序列的结构可以使用可用的算法来确定(例如,由在Albany的Rensselaer Polytechnic Institute and The RNA Institute, College of Arts and Sciences, State University of New York操作的“The mfold Web Server”;还参见M.Zuker, D.H.Mathews & D.H.Turner. Algorithms and Thermodynamics for RNA Secondary Structure Prediction: A Practical Guide In RNA Biochemistry and Biotechnology, 11-43, J.Barciszewski和B.F.C.Clark, 编辑, NATO ASI Series, Kluwer Academic Publishers, Dordrecht, NL, (1999))。

[0061] 所提供的结构类型可以使用重复性三核苷酸基序(例如,图11B)加以控制。重复性三核苷酸基序是在序列中重复至少两次(例如,重复两次或更多次、三次或更多次、四次或更多次、五次或更多次、六次或更多次、七次或更多次更多次数、八次或更多次、或者十次或更多次)的三个核苷酸的基序。因此,延伸序列可以包含重复性三核苷酸基序。在一个实施方案中,延伸序列包含CAA、UUG、AAG、CUU、CCU、CCA、UAA或其组合的重复性三核苷酸基序,其提供了无规卷曲序列。在另一个实施方案中,延伸序列包含CAU、CUA、UUA、AUG、UAG或其组合的重复性三核苷酸基序,其提供了半稳定的发夹结构。在另一个实施方案中,延伸序列包含重复性CNG三核苷酸基序(例如,CGG、CAG、CUG、CCG)、CGA或CGU的重复性三核苷酸基序或其

组合,其提供了稳定的发夹结构。在另一个实施方案中,延伸序列包含AGG、UGG或其组合的重复性三核苷酸基序,其提供了四链体(或G-四链体)结构。在又一个实施方案中,延伸序列包含前述三核苷酸基序的组合和由此产生的不同结构的组合。例如,延伸序列可以具有包含无规卷曲结构的区域、包含半稳定发夹的区域、包含稳定发夹的区域和/或包含四链体的区域。因此,每个区域可以包含与所示结构相关的重复性三核苷酸基序。下表中呈现了结构的非限制性实例:

[0062] 表1.代表性的RNA延伸序列(35个核苷酸)及其相应结构。

	序列(5'→3')
无规卷曲 	R1: UCCCGAGCUGUGCUUCGUUUCUACACUUGUACAUG R2: CCCUGCGACAGUCAUCUCGGCCGCCAAAGACACAG R3: UUCGUCGCCUUGGGCCGUCGUUUUUCGCUCGUGGG
假结 	S1: UUGGCGCAGUGGGCUAGCGCCACUCAAAAGGCCCA
四链体 	S2: UUAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG G
发夹 	S3: UUCGUCGUCGUCGUCGUCGUCGUCGUCGUCGUCGU S4: UUCUGCUGCUGCUGCUGCUGCUGCUGCUGCUGCUGCUG

[0064] 延伸序列也可以用于产生crRNA多聚体;因此,在另一个实施方案中,提供了crRNA多聚体,其包含两个或更多个crRNA分子(例如3个或更多个、4个或更多个、5个或更多个、6个或更多个、7个或更多个、甚至8个或更多个crRNA分子),其中每个crRNA包含如本文所述的延伸序列,并且多聚体的crRNA分子通过其延伸序列例如经由碱基配对或杂交而连接。因



此,在一个实施方案中,多聚体的每个crRNA包含延伸序列,所述延伸序列包含与多聚体的另一个crRNA的延伸区域充分互补的区域,以促进杂交。互补区域可以具有促进相互作用的任何合适的长度(例如4nt或更多、6nt或更多、8nt或更多、10nt或更多、15nt或更多等)。crRNA多聚体可用于例如同时递送多重crRNA,例如当对于特定治疗策略需要多重crRNA时。此类用途的一个实例是外显子跳跃,其中DNA片段被两个crRNA切割,以恢复功能读码框(例如, Ousterout DG等人(2015), Multiplex CRISPR/Cas9-based genome editing for correction of dystrophin mutations that cause Duchenne Muscular Dystrophy. Nat Commun. 6:6244)。外显子跳跃需要两个crRNA,其各自靶向用于靶向的核中的不同位点(一个在5'位点处,且另一个在3'位点处)。理想地,两个crRNA的比率应该是1:1;然而,难以维持这个比率。通过经由适当的延伸结构将多聚体中的crRNA(例如,各自包含不同的靶向序列)配对,可以促进以所需比率的递送。

[0065] 在一个实施方案中,具有结构化延伸的两个或更多个crRNA参与RNA“接吻”相互作用(又名环-环相互作用),其在一个结构化延伸序列(例如,发夹环)中的未配对核苷酸与第二个crRNA上的另一个结构(例如另一个发夹环)中的未配对核苷酸碱基配对时发生。这种类型的相互作用的实例在图11C中示出。接吻环或其它结构的形成使两个或多个crRNA分子多聚化。这种策略可以用于连接几个crRNA分子。

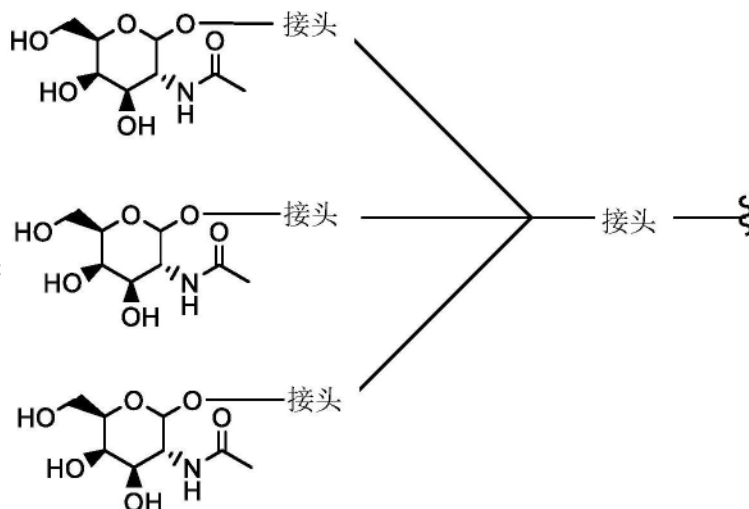
[0066] 每个crRNA上的延伸中的互补序列的杂交也可以用于促进多聚化。例如,可以经由具有自组装能力的延伸区域来构建超分子crRNA结构。例如,三聚体可以由具有适当放置的杂交区域的三个RNA分子形成(例如,图11D,小图(i); Shu D, Shu Y, Haque F, Abdelmawla S, & Guo P (2011) Thermodynamically stable RNA three-way junctions as platform for constructing multi-functional nanoparticles for delivery of therapeutics. Nat Nanotechnol. 6(10):658-667.))。类似地,可以通过组装十六个RNA分子来生成RNA八聚体(图11D,小图(ii); Yu J, Liu Z, Jiang W, Wang G, & Mao C (2014) De novo design of an RNA tile that self-assembles into a homo-octameric nanoprism. Nat Commun. 6:5724.))。

[0067] 前述类型的延伸部分中的任一种都可以连同或不连同加工序列一起使用。在一些实施方案中,核酸可以包括多重加工序列和延伸序列。例如,核酸可以进一步包含第一延伸序列5'的第二加工序列和第二加工序列5'的第二延伸序列。第二加工和延伸序列可以与第一加工和延伸序列(例如,重复)相同,或者第二加工序列和第二延伸序列中的任一或两者可以与第一加工序列和/或延伸序列不同。核酸并不特别限于任何数目的加工和延伸序列,并且可以具有2、3、4、5等个加工和/或延伸序列。

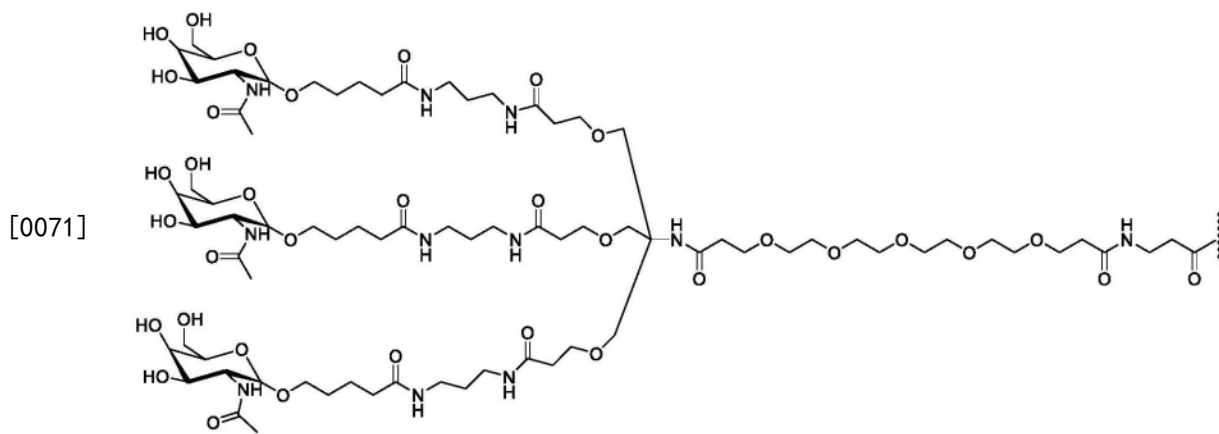
[0068] 核酸构建体的5'末端(即,在5'末端处的加工序列或延伸序列,适用时)可以根据需要进行进一步修饰。例如,5'末端可以用官能团(例如参与生物正交或“点击(click)”化学反应的官能团)进行修饰。例如,核酸的5'末端可以用叠氮化物、四嗪、炔烃、应变烯烃或应变炔烃进行化学修饰。此类修饰可以使用适当地配对的官能团,促进所需化学部分或分子与构建体连接。

[0069] 核酸的5'末端可以任选地经由上述生物正交或“点击”化学修饰为包含生物功能分子。生物功能分子可以是增强RNA引导的核酸内切酶的递送或活性,或提供一些其它所需功能的任何分子,例如将核酸靶向特定的目的地(例如靶向特定蛋白质、细胞受体、组织等

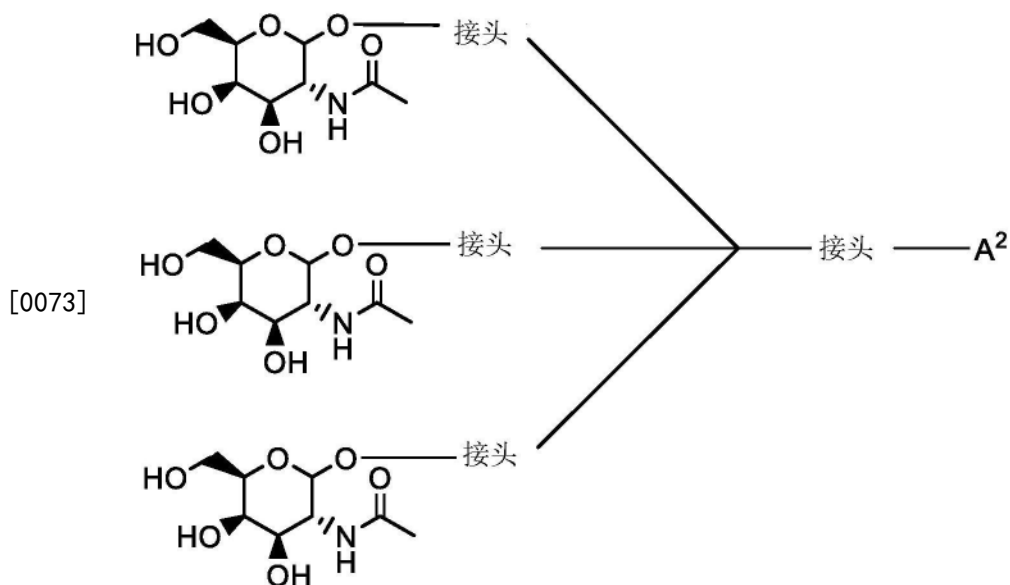
的部分),或促进构建体的追踪(例如,可检测标记,例如荧光标记物、放射性标记等等)。生物功能分子的实例包括例如内溶酶聚合物、供体DNA分子、氨基糖(例如,N-乙酰半乳糖胺(GalNAc)或三-N-GalNAc)、引导和/或示踪RNA(例如,单引导RNA)、以及其它肽、核酸和靶向配体(例如抗体、配体、细胞受体、适体、半乳糖、糖、小分子)。在一个实施方案中,crRNA包含与crRNA延伸部分缀合的生物素或抗生物素蛋白(或链霉抗生物素蛋白)分子,允许适当时,修饰的crRNA结合与抗生物素蛋白/链霉抗生物素蛋白或生物素缀合的另一种分子(例如,靶向分子或肽)(参见例如,图14)。在另一个实施方案中,crRNA延伸部分可以以任何合适的方式例如通过接头与氨基糖共价连接。如本文使用的,“氨基糖”是糖分子,其中羟基已替换为胺基(例如半乳糖胺)和/或其为复杂官能团(例如N-乙酰半乳糖胺(GalNAc);三-N-乙酰半乳糖胺(三触角N-乙酰半乳糖胺))的部分的氮。氨基糖可以进行修饰,以含有任选的间隔基。氨基糖的实例包括N-乙酰半乳糖胺(GalNAc)、三价GalNAc或三触角N-乙酰半乳糖胺。氨基糖基团的一个实例包括下述:



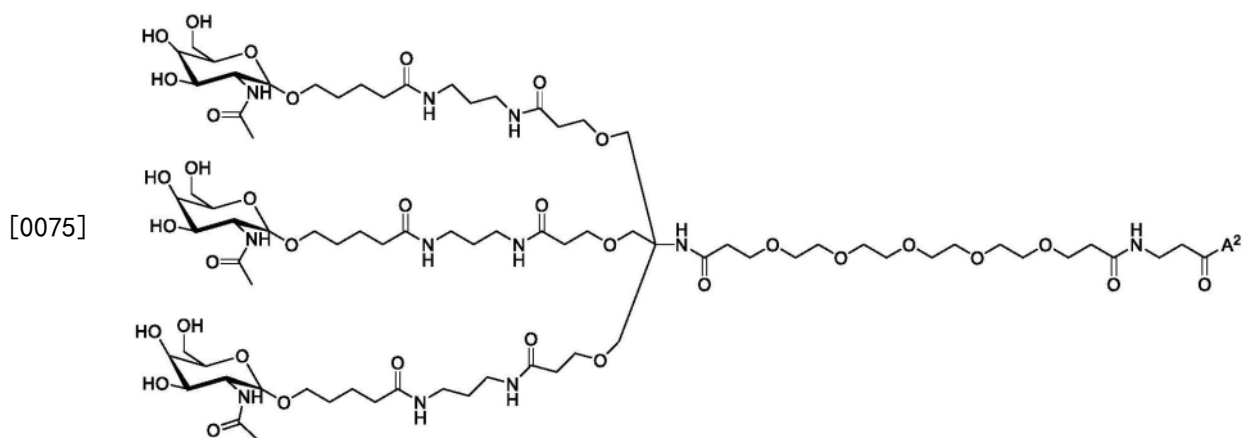
[0070] 其中所述接头可以是本领域通常已知的任何,并且每个接头可以彼此相同或不同。一般地,接头是饱和或不饱和脂肪族或杂脂肪族链。脂肪族或杂脂肪族链通常包含1-30个成员(例如1-30个碳、氮和/或氧原子),并且可以被一个或多个官能团(例如一个或多个酮、醚、酯、酰胺、醇、胺、脲、硫脲、亚砷、砷、磺酰胺和/或二硫化物基)取代。在一些情况下,使用较短的脂肪族或杂脂肪族链(例如,链中的约1-15个成员、约1-10个成员、约1-5个成员、约3-15个成员、约3-10个成员、约5-15个成员、或约5-10个成员)。在其它情况下,使用更长的脂肪族或杂脂肪族链(例如,链中的约5-30个成员、约5-25个成员、约5-20个成员、约10-30个成员、约10-25个成员、约10-20个成员、约15-30个成员、约15-25个成员、或约15-20个成员)。间隔基的实例包括取代的和未取代的烷基、烯基和聚乙二醇(例如,PEG 1-10或PEG 1-5)、或其组合。提供用于说明的更具体实例如下:



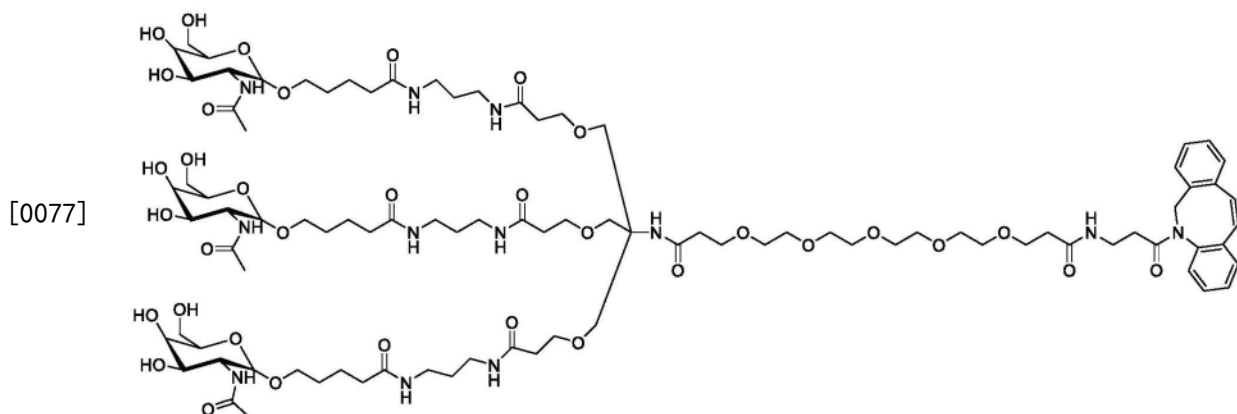
[0072] 在与接头缀合之前,氨基糖可以包含官能团(例如叠氮化物、四嗪、炔烃、应变烯烃或应变炔烃),其允许缀合至与crRNA延伸部分附着的适当配对的官能团(例如在5'末端处)。因此,例如,在与延伸的crRNA缀合之前,氨基糖可以包含:



[0074] 其中A<sup>2</sup>包含叠氮化物、四嗪、炔烃、应变烯烃或应变炔烃,如本文所述。更具体的实例如下:



[0076] 其中A<sup>2</sup>包含叠氮化物、四嗪、炔烃、应变烯烃或应变炔烃,如本文所述,例如:



[0078] 加工序列

[0079] 在一些实施方案中,crRNA包含加工序列。加工序列是无需引导/靶向序列,通过Cpf1在体外或体内自切割的核酸序列。不希望受任何特定的理论或作用机制的束缚,认为当存在时,加工序列在进入细胞后被切割,并且从任何延伸序列中释放crRNA。加工序列可以位于crRNA和延伸序列之间。在这种构型中,在加工序列切割后,crRNA从本文提供的核酸构建体的延伸序列中释放。

[0080] 加工序列也可以位于延伸序列内,位于延伸序列的5',或者可以充当延伸序列。另外,可以使用多重加工序列。例如,第二加工序列可以单独或连同另外的核苷酸序列一起充当延伸序列。然而,延伸序列一般不同于加工序列(如果存在的话)。此外,在一个实施方案中,延伸序列不包含加工序列和/或任何其它完全的(完整的)crRNA序列。

[0081] 在一些实施方案中,加工序列位于crRNA的紧5'(即,直接附着至crRNA序列)。在其它实施方案中,间隔序列可以存在于crRNA和加工序列之间。间隔序列可以具有任何长度(例如1、2、3、4、5、6、7、8、9或10nt),条件是它不阻止加工序列的Cpf1切割或在切割后释放的crRNA的功能。

[0082] 在一个实施方案中,加工序列包含Cpf1阵列的直接重复序列的片段。Cpf1阵列(有时也称为前crRNA)是天然存在的阵列,其包含直接重复序列和每个直接重复之间的间隔序列。阵列的直接重复部分包括两个部分:crRNA序列部分和加工部分。在给定的直接重复内,加工部分位于crRNA序列部分的5',经常为加工部分的紧5'。根据这个实施方案,本文提供的核酸的加工序列包含足以实现Cpf1切割的直接重复的加工部分的至少一个片段。例如,加工序列可以包含直接重复序列的加工部分的至少5个邻接核苷酸的片段,例如直接重复序列的加工部分的至少6、7、8、9、10、11、12、13、14、15、16或17nt(或直接重复序列的整个加工部分),其长度取决于直接重复源于其的种类。在一些实施方案中,加工序列包含直接重复序列的整个加工部分。直接重复可以来自任何微生物的Cpf1阵列。在图9中提供了直接重复序列,以及直接重复序列的加工部分的实例。本发明核酸的加工序列可以包含图9中的任何加工序列的片段或整个序列(例如,SEQ ID NO:2-20)。

[0083] 供体核酸

[0084] 本文提供的核酸构建体可以进一步包含供体核酸(也称为供体多核苷酸)。供体多核苷酸是插入在由RNA引导的核酸内切酶(例如,Cpf1)诱导的切割位点处的核酸。供体多核苷酸的核酸可以是本领域已知的任何类型的核酸。例如,核酸可以是DNA、RNA、DNA/RNA杂交体、人工核酸或其任何组合。在一个实施方案中,供体多核苷酸的核酸是DNA,在本文中也称

为“供体DNA”。

[0085] 供体多核苷酸通常是单链的,并且充当用于产生含有所需序列的双链DNA的模板。供体多核苷酸与侧接切割位点至接近切割位点的基因组序列区域(在约50个碱基或更少内、在约30个碱基或更少内、在约15个碱基或更少内、或在约10个碱基或更少内、在约5个碱基或更少内、或紧邻切割位点)的基因组序列含有足够的同一性(例如85%、90%、95%或100%序列同一性),以支持供体序列与侧接切割位点的基因组序列(供体序列与其具有足够的序列同一性)之间的同源性定向重复。供体多核苷酸序列可以具有任何长度,但必须具有在切割位点的两侧上具有序列同一性的足够数目的核苷酸,以促进HDR。供体多核苷酸的这些区域称为同源臂。同源臂可以具有相同数目的碱基或不同数目的碱基,并且各自一般为长度至少5个核苷酸(例如10个核苷酸或更多、15个核苷酸或更多、20个核苷酸或更多、50个核苷酸或更多、100个核苷酸或更多、150个核苷酸或更多、或甚至200个核苷酸或更多)。供体多核苷酸还含有侧面为同源臂的中心区域,其含有突变或其它目的DNA序列。因此,供体多核苷酸的总体长度通常大于两个同源臂的总长度(例如,约15个核苷酸或更多、约20个核苷酸或更多、50个核苷酸或更多、100个核苷酸或更多、150个核苷酸或更多、或甚至200个核苷酸或更多、250个核苷酸或更多、500个核苷酸或更多、1000个核苷酸或更多、5000个核苷酸或更多)。

[0086] 供体多核苷酸序列通常与靶基因组序列不同。相反,供体多核苷酸序列可以含有关于基因组序列的一个或多个单碱基改变、插入、缺失、倒置或重排,只要同源臂具有足够的序列同一性来支持HDR。供体多核苷酸序列可以进一步包含促进供体多核苷酸的成功插入的检测的序列。

[0087] 可以通过本领域技术人员已知的方法,来保护供体多核苷酸的末端(例如,免于核酸外切降解)。例如,将一个或多个双脱氧核苷酸残基加入线性分子的3'末端,和/或将自互补的寡核苷酸连接至一个或两个末端。用于保护外源多核苷酸免于降解的另外方法包括但不限于,添加末端氨基和使用修饰的核苷酸间键合,例如硫代磷酸酯、氨基磷酸酯和O-甲基核糖或脱氧核糖残基。

[0088] 在一些实施方案中,供体多核苷酸(例如,供体DNA)共价连接至Cpf1 crRNA的5'末端、加工序列的5'末端、或延伸序列的5'末端。在一个优选的实施方案中,供体多核苷酸连接至延伸序列的5'末端。在一些实施方案中,供体DNA和核酸之间的键合是可逆的(例如,二硫键)。

[0089] 在一些实施方案中,供体DNA与核酸构建体共价连接。例如,供体多核苷酸可以与加工序列连接,并且充当位于加工序列5'的延伸序列。在另一个实施方案中,供体多核苷酸可以在延伸序列的5'处连接。

[0090] 核酸和供体DNA可以通过本领域已知的任何方法连接或缀合。在一些实施方案中,修饰供体DNA的3'末端和核酸的5'末端,以促进键合。例如,核酸的5'末端可以被硫代吡啶激活,而供体DNA可以是巯基封端的,从而允许在两个分子之间形成二硫键。在一些实施方案中,与核酸和供体DNA两者互补的桥DNA使两个分子杂交并且靠近,以促进反应。图5A-5C提供了供体DNA与核酸缀合的合成的非限制性实例。

[0091] 在其它实施方案中,核酸和供体DNA可以经由官能团缀合,所述官能团例如参与生物正交或“点击”化学反应的官能团。例如,可以用官能团例如叠氮化物、四嗪、炔烃、应变烯

烃或应变炔烃化学修饰核酸的5'末端,并且可以用适当配对的官能团化学修饰供体DNA的3'末端。例如,如果核酸含有叠氮化物,则叠氮化物将经由叠氮化物-炔烃环加成反应(铜催化的)与供体DNA的炔基反应、或将经由叠氮化物-应变炔烃环加成反应(无需催化剂)与供体DNA的应变炔基反应。同样地,如果核酸包含四嗪,则它将经由四嗪/烯烃环加成反应与应变烯烃反应。类似地,可以使用相反的构型,例如,如果核酸包含的是炔烃、应变炔烃或应变烯烃,则它将通过相同的环加成反应与供体DNA的叠氮化物或四嗪基团反应。

[0092] 在一些实施方案中,核酸和供体DNA经由接头缀合。例如,核酸和供体DNA可以通过自杀式(self-immolative)接头缀合。如本文使用的,“自杀式接头”是在特定条件(例如,特定pH值)下水解的接头,所述条件允许供体DNA从核酸中释放。

[0093] 核酸供体DNA缀合物的接头涵盖本领域已知的任何接头,其能够将供体DNA与核酸共价连接。接头可以在任一末端处附着至供体DNA和核酸。然而,在一些实施方案中,接头附着至核酸的5'末端(例如,crRNA、加工序列或延伸序列的5'末端)和供体DNA的3'末端。可以通过本领域已知的任何方法,例如本文关于供体DNA与核酸的缀合所述的那些,将接头附着至核酸和供体DNA。

[0094] 在另一个实施方案中,供体多核苷酸可以与延伸序列和/或加工序列杂交。因此,例如,延伸序列可以包含与供体多核苷酸充分互补以促进杂交的序列。

[0095] 当供体核酸与延伸序列共价或非共价连接时,有时期望供体核酸与未被Cpf1 crRNA切割的延伸序列或其一部分连接,使得当靶基因由Cpf1编辑时,供体核酸与crRNA紧密结合。认为在某些情况下,通过此类构建体可以达到改善的基因编辑。如上所述,未被Cpf1切割的延伸序列包括例如包含一个或多个修饰的核苷酸间键合或合成核苷酸的延伸序列。

[0096] 组合物和载体

[0097] 本发明还包括包含本文所述的任何核酸分子和载体的组合物。可以使用用于核酸递送的任何合适的载体。在一些实施方案中,载体可以包含能够与本文所述的任何核酸相互作用并促进核酸进入细胞内的分子。

[0098] 在一些实施方案中,载体包含阳离子脂质。阳离子脂质是具有带正电荷的极性首基的两亲分子,所述极性首基经由锚连接至一般包含两个烷基链的非极性疏水性结构域。在一些实施方案中,阳离子脂质形成围绕核酸构建体和任选地Cpf1蛋白的脂质体(例如脂质囊泡)。因此,在一个相关方面,提供了包含核酸构建体和任选地Cpf1蛋白的脂质体。

[0099] 在再一个实施方案中,载体包含阳离子聚合物。本发明组合物的阳离子聚合物的实例包括聚乙烯亚胺(PEI)、聚(精氨酸)、聚(赖氨酸)、聚(组氨酸)、聚-[2-{(2-氨基乙基)氨基}-乙基-天冬酰胺](pAsp(DET))、聚乙二醇(PEG)和聚精氨酸的嵌段共聚物、PEG和聚赖氨酸的嵌段共聚物、PEG和聚{N-[N-(2-氨基乙基)-2-氨基乙基]天冬酰胺}的嵌段共聚物(PEG-pAsp[DET])、({2,2-双[(9Z,12Z)-十八碳-9,12-二烯-1-基]-1,3-二噁烷-5-基}甲基)二甲胺、(3aR,5s,6aS)-N,N-二甲基-2,2-二((9Z,12Z)-十八碳-9,12-二烯-1-基)四氢-3aH-环戊二烯并[d][1,3]二氧杂环戊烯-5-胺、(3aR,5r,6aS)-N,N-二甲基-2,2-二((9Z,12Z)-十八碳-9,12-二烯-1-基)四氢-3aH-环戊二烯并[d][1,3]二氧杂环戊烯-5-胺、(3aR,5R,7aS)-N,N-二甲基-2,2-二((9Z,12Z)-十八碳-9,12-二烯-1-基)六氢苯并[d][1,3]二氧杂环戊烯-5-胺、(3aS,5R,7aR)-N,N-二甲基-2,2-二((9Z,12Z)-十八碳-9,12-二烯-1-基)

六氢苯并[d][1,3]二氧杂环戊烯-5-胺、(2-{2,2-双[(9Z,12Z)-十八碳-9,12-二烯-1-基]-1,3-二噁烷-4-基}乙基)二甲胺、(3aR,6aS)-5-甲基-2-((6Z,9Z)-十八碳-6,9-二烯-1-基)-2-((9Z,12Z)-十八碳-9,12-二烯-1-基)四氢-3aH-[1,3]二氧杂环戊烯并[4,5-c]吡咯、(3aS,7aR)-5-甲基-2,2-二((9Z,12Z)-十八碳-9,12-二烯-1-基)六氢-[1,3]二氧杂环戊烯并[4,5-c]吡啶、(3aR,8aS)-6-甲基-2,2-二((9Z,12Z)-十八碳-9,12-二烯-1-基)六氢-3aH-[1,3]二氧杂环戊烯并[4,5-d]氮杂卓、(6Z,9Z,28Z,31Z)-三十七碳-6,9,28,31-四烯-19-基2-(二甲基氨基)乙酸酯、(6Z,9Z,28Z,31Z)-三十七碳-6,9,28,31-四烯-19-基3-(二甲基氨基)丙酸酯、[6Z,9Z,28Z,31Z)-三十七碳-6,9,28,31-四烯-19-基-4-(二甲基氨基)丁酸酯]、(6Z,9Z,28Z,31Z)-三十七碳-6,9,28,31-四烯-19-基5-(二甲基氨基)戊酸酯、(6Z,9Z,28Z,31Z)-三十七碳-6,9,28,31-四烯-19-基6-(二甲基氨基)己酸酯、(3-{2,2-双[(9Z,12Z)-十八碳-9,12-二烯-1-基]-1,3-二噁烷-4-基}丙基)二甲胺、1-((3aR,5r,6aS)-2,2-二((9Z,12Z)-十八碳-9,12-二烯-1-基)四氢-3aH环戊二烯并[d][1,3]二氧杂环戊烯-5-基)-N,N-二甲基甲胺、1-((3aR,5s,6aS)-2,2-二((9Z,12Z)-十八碳-9,12-二烯-1-基)四氢-3aH环戊二烯并[d][1,3]二氧杂环戊烯-5-基)-N,N-二甲基甲胺、8-甲基-2,2-二((9Z,12Z)-十八碳-9,12-二烯-1-基)-1,3-二氧杂-8-氮杂螺[4.5]癸烷、2-(2,2-二((9Z,12Z)-十八碳-9,12-二烯-1-基)-1,3-二氧戊环-4-基)-N-甲基-N-(吡啶-3-基甲基)乙胺、1,3-双((9Z,12Z)-十八碳-9,12-二烯-1-基)-2-[2-(二甲基氨基)乙基]丙二酸酯、N,N-二甲基-1-((3aR,5R,7aS)-2-((8Z,11Z)-十八碳-8,11-二烯-1-基)-2-((9Z,12Z)-十八碳-9,12-二烯-1-基)六氢苯并[d][1,3]二氧杂环戊烯-5-基)甲胺、N,N-二甲基-1-((3aR,5S,7aS)-2-((8Z,11Z)-十八碳-8,11-二烯-1-基)-2-((9Z,12Z)-十八碳-9,12-二烯-1-基)六氢苯并[d][1,3]二氧杂环戊烯-5-基)甲胺、(1s,3R,4S)-N,N-二甲基-3,4-双((9Z,12Z)-十八碳-9,12-二烯-1-基氧基)环戊烷胺、(1s,3R,4S)-N,N-二甲基-3,4-双((9Z,12Z)-十八碳-9,12-二烯-1-基氧基)环戊烷胺、2-(4,5-二((8Z,11Z)-十七碳-8,11-二烯-1-基)-2-甲基-1,3-二氧戊环-2-基)-N,N-二甲基乙胺、2,3-二((8Z,11Z)-十七碳-8,11-二烯-1-基)-N,N-二甲基-1,4-二氧杂螺[4.5]癸-8-胺、(6Z,9Z,28Z,31Z)-三十七碳-6,9,28,31-四烯-19-基4-(二乙基氨基)丁酸酯、(6Z,9Z,28Z,31Z)-三十七碳-6,9,28,31-四烯-19-基4-[双(丙-2-基)氨基]丁酸酯、N-(4-N,N-二甲基氨基)丁酰基-(6Z,9Z,28Z,31Z)-三十七碳-6,9,28,31-四烯-19-胺、(2-{2,2-双[(9Z,12Z)-十八碳-9,12-二烯-1-基]-1,3-二噁烷-5-基}乙基)二甲胺、(4-{2,2-双[(9Z,12Z)-十八碳-9,12-二烯-1-基]-1,3-二噁烷-5-基}丁基)二甲胺、(6Z,9Z,28Z,31Z)-三十七碳-6,9,28,31-四烯-19-基(2-(二甲基氨基)乙基)氨基甲酸酯、2-(二甲基氨基)乙基(6Z,9Z,28Z,31Z)-三十七碳-6,9,28,31-四烯-19-基氨基甲酸酯、(6Z,9Z,28Z,31Z)-三十七碳-6,9,28,31-四烯-19-基3-(乙基氨基)丙酸酯、(6Z,9Z,28Z,31Z)-三十七碳-6,9,28,31-四烯-19-基4-(丙-2-基氨基)丁酸酯、N1,N1,N2-三甲基-N2-((11Z,14Z)-2-((9Z,12Z)-十八碳-9,12-二烯-1-基)二十碳-11,14-二烯-1-基)乙烷-1,2-二胺、3-(二甲基氨基)-N-((11Z,14Z)-2-((9Z,12Z)-十八碳-9,12-二烯-1-基)二十碳-11,14-二烯-1-基)丙酰胺、(6Z,9Z,28Z,31Z)-三十七碳-6,9,28,31-四烯-19-基4-(甲基氨基)丁酸酯、二甲基({4-[(9Z,12Z)-十八碳-9,12-二烯-1-基氧基]-3-[(9Z,12Z)-十八碳-9,12-二烯-1-基氧基]甲基}丁基))胺、2,3-二((8Z,11Z)-十七碳-8,11-二烯-1-基)-8-甲基-1,4-二氧杂-8-氮杂螺[4.5]癸烷、3-(二甲基氨基)丙基(6Z,9Z,28Z,31Z)-三十七碳-6,9,

28,31-四烯-19-基氨基甲酸酯、2-(二甲基氨基)乙基((11Z,14Z)-2-((9Z,12Z)-十八碳-9,12-二烯-1-基)二十碳-11,14-二烯-1-基)氨基甲酸酯、1-((3aR,4R,6aR)-6-甲氧基-2,2-二((9Z,12Z)-十八碳-9,12-二烯-1-基)四氢呋喃并[3,4-d][1,3]二氧杂环戊烯-4-基)-N,N-二甲基甲胺、(6Z,9Z,28Z,31Z)-三十七碳-6,9,28,31-四烯-19-基4-[乙基(甲基)氨基]丁酸酯、(6Z,9Z,28Z,31Z)-三十七碳-6,9,28,31-四烯-19-基4-氨基丁酸酯、3-(二甲基氨基)丙基((11Z,14Z)-2-((9Z,12Z)-十八碳-9,12-二烯-1-基)二十碳-11,14-二烯-1-基)氨基甲酸酯、1-((3aR,4R,6aS)-2,2-二((9Z,12Z)-十八碳-9,12-二烯-1-基)四氢呋喃并[3,4-d][1,3]二氧杂环戊烯-4-基)-N,N-二甲基甲胺、(3aR,5R,7aR)-N,N-二甲基-2,2-二((9Z,12Z)-十八碳-9,12-二烯-1-基)六氢苯并[d][1,3]二氧杂环戊烯-5-胺、(11Z,14Z)-N,N-二甲基-2-((9Z,12Z)-十八碳-9,12-二烯-1-基)二十碳-11,14-二烯-1-胺、(3aS,4S,5R,7R,7aR)-N,N-二甲基-2-((7Z,10Z)-十八碳-7,10-二烯-1-基)-2-((9Z,12Z)-十八碳-9,12-二烯-1-基)六氢-4,7-亚甲基苯并[d][1,3]二氧杂环戊烯-5-胺、N,N-二甲基-3,4-双((9Z,12Z)-十八碳-9,12-二烯-1-基氧基)丁-1-胺、以及3-(4,5-二((8Z,11Z)-十七碳-8,11-二烯-1-基)-1,3-二氧戊环-2-基)-N,N-二甲基丙-1-胺。也可以使用前述聚合物的任何组合。

[0100] 在其它实施方案中,载体包含聚合物纳米颗粒。例如,本发明组合物可以如国际专利申请号PCT/US2016/052690中所述作为纳米颗粒施用,所述专利申请的全部公开内容明确地引入作为参考。

[0101] Cpf1多肽或编码其的核酸

[0102] 在一些实施方案中,包括上述脂质体实施方案,组合物还包含Cpf1多肽或编码其的核酸。任何Cpf1多肽都可以用于本发明组合物中,尽管所选择的Cpf1应该适当地选择,以便与组合物中的核酸构建体的crRNA组合作用,以在适用时切割靶核酸和/或切割核酸构建体的加工序列。组合物的Cpf1可以是天然存在的Cpf1或变体或突变体Cpf1多肽。在一些实施方案中,Cpf1多肽是酶促活性的,例如,当与引导RNA结合时,Cpf1多肽切割靶核酸。在一些实施方案中,相对于野生型Cpf1多肽(例如,相对于包含图8中所示的氨基酸序列(SEQ ID NO:1)的Cpf1多肽),Cpf1多肽显示出减少的酶促活性,并且保留DNA结合活性。改变Cpf1的酶促活性的突变是本领域已知的。

[0103] 例如,Cpf1可以来自氨基酸球菌属的细菌或来自毛螺菌属(Lachnospiraceae)、或者来自图9中鉴定的任何属或种。图8中提供了Cpf1蛋白序列的实例。在一些实施方案中,Cpf1多肽包含的氨基酸序列与图8中所示的氨基酸序列具有至少30%、至少35%、至少40%、至少45%、至少50%、至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少90%或100%的氨基酸序列同一性。在一些实施方案中,Cpf1多肽包含的氨基酸序列与图8中所示的氨基酸序列的100个氨基酸至200个氨基酸(aa)、200aa至400aa、400aa至600aa、600aa至800aa、800aa至1000aa、1000aa至1100aa、1100aa至1200aa、或1200aa至1300aa的邻接段具有至少30%、至少35%、至少40%、至少45%、至少50%、至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少90%或100%的氨基酸序列同一性。

[0104] 在一些实施方案中,Cpf1多肽包含的氨基酸序列与图8中所示的氨基酸序列的Cpf1多肽的RuvCI结构域具有至少30%、至少35%、至少40%、至少45%、至少50%、至少



55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少90%或100%的氨基酸序列同一性。在一些实施方案中,Cpf1多肽包含的氨基酸序列与图8中所示的氨基酸序列的Cpf1多肽的RuvCII结构域具有至少30%、至少35%、至少40%、至少45%、至少50%、至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少90%或100%的氨基酸序列同一性。在一些实施方案中,Cpf1多肽包含的氨基酸序列与图8中所示的氨基酸序列的Cpf1多肽的RuvCIII结构域具有至少30%、至少35%、至少40%、至少45%、至少50%、至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少90%或100%的氨基酸序列同一性。

[0105] 在一些实施方案中,Cpf1多肽是FnCpf1、Lb3Cpf1、BpCpf1、PeCpf1、SsCpf1、AsCpf1、Lb2Cpf1、CMtCpf1、EeCpf1、MbCpf1、LiCpf1、LbCpf1、PcCpf1、PdCpf1或PmCpf1;或Cpf1多肽,其包含的氨基酸序列与其具有至少30%、至少35%、至少40%、至少45%、至少50%、至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少90%或100%的氨基酸序列同一性。

[0106] 在一些实施方案中,Cpf1多肽包含在对应于图8中所示的氨基酸序列的位置917的氨基酸残基处的氨基酸取代(例如,D→A取代);和/或包含在对应于图8中所示的氨基酸序列的位置1006的氨基酸残基处的氨基酸取代(例如,E→A取代);和/或包含在对应于图8中所示的氨基酸序列的位置1255的氨基酸残基处的氨基酸取代(例如,D→A取代)。

[0107] Cpf1多肽还可以是RNA酶灭活的Cpf1,例如包含在氨基酸球菌属Cpf1(AsCpf1)的H800A、K809A、K860A、F864A或R790A处,或者不同Cpf1直系同源物的相应位置处的修饰的Cpf1。突变型Cpf1蛋白的实例包括在Zetsche等人,“Multiplex Gene Editing by CRISPR-Cpf1 Through Autonomous Processing of a Single crRNA Array,” Nat.Biotechnol.2017,35(1),31-34中公开的蛋白质。Cpf1多肽也可以是dCpf1碱基编辑器(例如Cpf1-胞嘧啶脱氨酶融合蛋白)。实例包括例如在Li等人,Nature Biotechnology,36 324-327 (2018),以及Mahfouz等人,Biochem J.,475(11),1955-1964(2018)中公开的蛋白质。合成变体Cpf1的实例是通过Inscripta,Inc.(CO,USA)的MAD7 Cpf1直系同源物。Cpf1蛋白的另外实例包括国际专利申请号PCT/US2016/052690中公开的那些Cpf1蛋白中的任一种,包括嵌合或突变蛋白,所述专利申请的全部公开内容明确地引入本文作为参考。

[0108] 其它核酸

[0109] 除crRNA之外,组合物还可以进一步包含其它核酸。例如,如本文所述,组合物可以包含供体多核苷酸。可替代地或另外,组合物可以包含并非供体多核苷酸的一种或多种另外的核酸(例如,与待编辑的靶序列、或编辑细胞的任何内源核酸序列没有显著的序列同一性的核酸(例如,不足以允许同源重组的序列同一性水平)。这些另外的核酸可以是RNA或DNA,例如单链RNA或DNA分子(或包含RNA和DNA两者的杂交分子,任选地具有合成核酸残基)。另外的核酸可以是任何长度,例如长度至少5个核苷酸(例如10个核苷酸或更多、15个核苷酸或更多、20个核苷酸或更多、50个核苷酸或更多、100个核苷酸或更多、150个核苷酸或更多、或甚至200个核苷酸或更多)。在一些实施方案中,核酸可以包含500个核苷酸或更多、1000个核苷酸或更多、或甚至5000个核苷酸或更多。然而,在大多数情况下,核酸包含约5000个核苷酸或更少,例如约1000个核苷酸或更少,或甚至500个核苷酸或更少(例如200

个核苷酸或更少)。

[0110] 组合物可以进一步包含编码特定目的蛋白质的核酸,例如,RNA引导的核酸内切酶(例如,Cpf1多肽)。RNA引导的核酸内切酶可以是如本文关于本发明的其它方面所述的任何。

[0111] 二价金属离子

[0112] 在一些实施方案中,组合物基本上或完全不含二价金属离子(例如,镁),其激活所使用的特定Cpf1蛋白,以便减少或防止在递送之前加工序列的过早切割。组合物被视为基本上不含镁,其浓度不允许Cpf1自加工酶促活性。在一些实施方案中,组合物包含约20mM或更少的NaCl,并且基本上或完全不含激活Cpf1蛋白的镁或其它二价离子。

[0113] 遗传修饰真核细胞的方法

[0114] 本发明还提供了遗传修饰真核靶细胞的方法,其包括使真核靶细胞与本文所述的任何核酸或组合物(例如,包含Cpf1 crRNA、crRNA5'的延伸序列、以及任选地crRNA和延伸序列之间的加工序列)接触,以遗传修饰靶核酸。在一些实施方案中,核酸的Cpf1 crRNA包含与靶细胞中的靶序列杂交的靶向序列(例如,茎-环结构域的3')。在一些实施方案中,Cpf1 crRNA包含加工序列,其在进入细胞后被切割,从而从加工序列和延伸序列中释放Cpf1 crRNA。在其它实施方案中,Cpf1 crRNA不包含加工序列。

[0115] 靶核酸是crRNA的靶向序列与之结合并诱导通过Cpf1的切割的多核苷酸(例如, RNA、DNA)。靶核酸包含“靶位点”或“靶序列”,其是crRNA与之杂交的靶核酸中存在的序列,其依次地将核酸内切酶引导至靶核酸。

[0116] “真核靶细胞”可以是本领域已知的任何真核细胞,并且包括体内和体外细胞两者。在一个实施方案中,靶细胞是哺乳动物细胞。

[0117] 任何施用途径都可以用于将组合物递送至哺乳动物。实际上,尽管多于一种途径可以用于施用组合物,但特定的途径可以提供比另一种途径更立即和更有效的反应。当体外或离体施用于细胞时,可以通过任何合适的方法使核酸或组合物与细胞接触。例如,核酸可以在脂质体中、被阳离子聚合物包封、和/或通过电穿孔引入。当施用于对象例如哺乳动物或人时,可以通过各种途径中的任一种来施用组合物。例如,一定剂量的组合物也可以施加到或滴注到体腔内,通过皮肤吸收(例如,经由透皮贴剂),吸入,摄入,局部施加于组织,或者经由例如静脉内、腹膜内、口内、真皮内、皮下或动脉内施用进行肠胃外施用。

[0118] 组合物可以在允许控制或持续释放的装置中或在其上施用,所述装置例如海绵、生物相容性网状物、机械储存器或机械植入物。植入物(参见例如美国专利5,443,505)、装置(参见例如美国专利4,863,457),例如可植入装置,例如机械储存器或植入物或由聚合物组合物组成的装置,对于组合物的施用是特别有用的。组合物也可以以持续释放制剂的形式施用(参见例如,美国专利5,378,475),其包含例如凝胶泡沫、透明质酸、明胶、硫酸软骨素、聚磷酸酯例如双-2-羟乙基-对苯二甲酸乙二醇酯(BHET)和/或聚乳酸乙醇酸。

[0119] 实施例

[0120] 下述实施例进一步说明了本发明,但是,当然,不应以任何方式解释为限制其范围。下表2提供了这些实验中使用的核酸序列。

[0121] 表2

[0122] 补充表1

[0123]

关于 Cas9 的 sgRNA	rGrCrCrGrUrCrCrArGrCrUrCrGrArCrCrArGrGrArGrUrUrUr UrArGrArGrCrUrArGrArArArUrArGrCrArArGrUrUrArArAr ArUrArArGrGrCrUrArGrUrCrCrGrUrUrArUrCrArArCrUrUr GrArArArArArGrUrGrGrCrArCrCrGrArGrUrCrGrGrUrGrCr UrUrUrU
crRNA <sup>+4</sup>	rUrGrGrArUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUr CrGrUrCrGrCrCrGrUrCrCrArGrCrUrCrGrArCrCrA
crRNA <sup>+9</sup>	rGrGrGrArArUrGrGrArUrArArUrUrUrCrUrArCrUrCrUrUrGr UrArGrArUrCrGrUrCrGrCrCrGrUrCrCrArGrCrUrCrGrArCrC rA

[0124]

乱序的 crRNA <sup>+9</sup>	rGrGrUrGrArGrCrArArUrArArUrUrUrCrUrArCrUrCrUrUrGr UrArGrArUrCrGrUrCrGrCrCrGrUrCrCrArGrCrUrCrGrArCrC rA
crRNA <sup>+9</sup> dU	rGrGrUrGrArGrCrAUrUrArArUrUrUrCrUrArCrUrCrUrUrGr UrArGrArUrCrGrUrCrGrCrCrGrUrCrCrArGrCrUrCrGrArCrC rA
crRNA <sup>+9</sup> S	rA*rU*rG*rG*rU*rG*rA*rG*C*rUrArArUrUrUrCrUrArCrUr CrUrUrGrUrArGrArUrCrGrUrCrGrCrCrGrUrCrCrArGrCrUrC rGrArCrCrA
crRNA <sup>+15</sup>	rGrU rCrArA rArGrG rGrArA rUrGrG rArUrA rArUrU rUrCrU rArCrU rCrUrU rGrUrA rGrArU rCrGrU rCrGrC rCrGrU rCrCrA rGrCrU rCrGrA rCrCrA
crRNA <sup>+25</sup>	rArUrG rUrGrU rUrUrU rUrGrU rCrArA rArArG rArCrC rUrUrU rUrUrA rArUrU rUrCrU rArCrU rCrUrU rGrUrA rGrArU rCrGrU rCrGrC rCrGrU rCrCrA rGrCrU rCrGrA rCrCrA
crRNA <sup>+59</sup>	rGrGrCrCrArGrCrUrUrGrCrCrGrGrUrUrUrUrUrUrArGrUrCr GrUrGrCrUrGrCrUrUrCrArUrGrUrGrUrUrUrUrUrGrUrCrAr <u>ArArArGrArCrCrUrUrUrUrUrArArUrUrUrCrUrArCrUrCrUr</u> UrGrUrArGrArUrCrGrUrCrGrCrCrGrUrCrCrArGrCrUrCrGr ArCrCrA
crRNA <sup>+59</sup> _D2	<u>rUrCrArArArArGrArCrCrUrUrUrUrGrUrCrArArArArGrArCr</u> <u>CrUrUrUrUrGrUrCrArArArArGrArCrCrUrUrUrUrGrUrCrAr</u> <u>ArArArGrArCrCrUrUrUrUrUrArArUrUrUrCrUrArCrUrCrUr</u> UrGrUrArGrArUrCrGrUrCrGrCrCrGrUrCrCrArGrCrUrCrGr

[0125]

	ArCrCrA
crRNA <sup>+59</sup> _D3	rGrGrCrGrGrCrUrUrGrCrCrGrGrUrUrUrUrUrArGrUrCrGr UrGrCrUrGrCrUrUrCrArUrGrUrGrUrUrUrUrUrGrUrCrUrAr <u>ArGrArArCrUrUrUrArArAr</u> UrArArUrUrUrCrUrArCrUrCrUr UrGrUrArGrArUrCrGrUrCrGrCrCrGrUrCrCrArGrCrUrCrGr ArCrCrA
Ai9 crRNA	rUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrUrCrCrAr ArArCrUrCrArUrCrArArUrGrUrArUrCrU
Ai9 crRNA <sup>+2</sup>	rUrUrUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrUrCr CrArArArCrUrCrArUrCrArArUrGrUrArUrCrU
Ai9 crRNA <sup>+9</sup>	rArGrArCrCrUrUrUrUrUrArArUrUrUrCrUrArCrUrCrUrUrGr UrArGrArUrUrCrCrArArArCrUrCrArUrCrArArUrGrUrArUr CrU
Ai9 crRNA <sup>+59</sup>	rGrArGrCrArGrCrUrUrGrCrCrGrGrUrUrUrUrUrUrArGrUrCr GrUrGrCrUrGrCrUrUrCrArUrGrUrGrUrUrUrUrUrGrUrCrAr ArArArGrArCrCrUrUrUrUrUrArArUrUrUrCrUrArCrUrCrUr UrGrUrArGrArUrUrCrCrArArArCrUrCrArUrCrArArUrGrUr ArUrCrU
Serpinal crRNA	rUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrUrCrGrUr CrGrArUrGrGrUrCrArGrCrArCrArGrCrC
Serpinal crRNA <sup>+9</sup>	rCrUrCrCrCrCrUrCrCrUrArArUrUrUrCrUrArCrUrCrUrUrGr UrArGrArUrUrCrGrUrCrGrArUrGrGrUrCrArGrCrArCrArGr CrC
具有 ClaI 限制 性酶位点的 ssODN	CTCGCCGGACACGCTGAACTTGTGGCCGCTTACGTCG CCGTCCAGCTCGACCATCGATGGGCACCACCCCGGTG AACAGCTCCTCGCCCTTGCTCACCAT

[0126]

没有任何同源性的 ssDNA	GGGATAGACATGGGTATGGCCTCTAAAAACATGGCGCC AGCAGCTTCAGTCCCTTTCTCATCGATGGTCAGCACAG CCTTATGCACGGCCTGGAGGGGAG
用于 ai9 成肌细胞实验的 ssDNA	CTTGACCTCGGGGGGGATAGACATGGGTATGGCCTCTA AAAACATGGCgCCAGCAGCTTCAGTCCCTTTCTCATCG ATGGTCAGCACAGCCTTATGCACG
用于 ai9 成肌细胞实验的 ssRNA	rArGrArArGrGrGrArCrUrGrArArGrCrUrGrCrUrGrUrU rUrUrA rGrArG rCrUrA rGrArA rArUrA rGrCrA rArGrU rUrArA rArArU rArArG rGrCrU rArGrU rCrCrG rUrUrA rUrCrA rArCrU rUrGrA rArArA rArGrU rGrGrC rArCrC rGrArG rUrCrG rGrUrG rCrUrUrUrU
用于 GFP-HEK 实验的 9nt ssRNA	rGrGrGrArArUrGrGrA
用于 GFP-HEK 实验的 100nt ssRNA	rArArGrUrArArArArCrCrUrCrUrArCrArArArUrGrGrUrUrUr UrArGrArGrCrUrArGrArArArUrArGrCrArArGrUrUrArArAr ArUrArArGrGrCrUrArGrUrCrCrGrUrUrArUrCrArArCrUrUr GrArArArArGrUrGrGrCrArCrCrGrArGrUrCrGrGrUrGrCr UrUrUrU

[0127] \*:硫代磷酸酯

[0128] U:2' 脱氧

[0129] 下划线:前crRNA序列

[0130] 实施例1

[0131] 本实施例说明了与Cas9 sgRNA相比,未修饰的Cpf1 crRNA趋于提供较低的基因编辑效率。

[0132] 使用绿色荧光蛋白(GFP)报告系统,比较了SpCas9和AsCpf1,两者均不含引导RNA的延伸部分。选择了可以被两种核酸酶识别的GFP基因中匹配的前间隔DNA序列,以便直接比较AsCpf1和SpCas9。该系统在图13A中示出。使用电穿孔和阳离子脂质,在多西环素诱导型启动子(GFP-HEK)的控制下,将RNP复合物引入表达GFP基因的HEK293T细胞内。通过测量GFP阴性细胞的群体来确定编辑活性,其中GFP通过NHEJ介导的插入缺失突变被破坏。

[0133] AsCpf1 RNP在电穿孔和阳离子脂质体处理的细胞两者中均显示出比SpCas9更低的基因编辑(图13B和13C)。AsCpf1 RNP电穿孔的细胞为31%GFP阴性,而SpCas9 RNP电穿孔的细胞为41%GFP阴性(图13B)。使用阳离子脂质体和RNAiMax递送系统,AsCpf1处理的细胞

为大约8%GFP阴性,而SpCas9处理的细胞为大约30%GFP阴性(图13C)。

#### [0134] 实施例2

[0135] 本实施例证实了包含Cpf1 crRNA、Cpf1 crRNA5'的加工序列、以及加工序列5'的延伸序列(可逆地超荷电的crRNA)的核酸,增强了在体外通过阳离子脂质的Cpf1递送。

[0136] 阳离子材料如阳离子脂质体和聚阳离子是对于细胞和动物中的核酸最常用的递送媒介物。因此,分析了阳离子脂质阳离子脂质体将Cpf1/crRNA复合物转染到表达绿色荧光蛋白的HEK细胞(GFP-HEK)内的效率。简言之,将设计为经由插入缺失来击倒GFP基因的crRNA与Cpf1复合,并且或电穿孔(核转染)到细胞内或用阳离子脂质体转染。然后经由流式细胞术,通过测量GFP敲除细胞的数目来确定基因编辑效率,其中不再表达GFP的细胞指示该细胞已由Cpf1/crRNA复合物转染。

[0137] 来自这些实验的结果显示,阳离子脂质体不能有效地转染未修饰的Cpf1 crRNA复合物(图1A)。具体地,用阳离子脂质体和Cpf1/crRNA复合物处理的细胞仅具有8%的NHEJ效率,而用Cpf1/crRNA复合物电穿孔的细胞具有40%的NHEJ效率,证实了递送局限性是关于使用阳离子脂质体的低NHEJ效率的主要原因。

[0138] 为了检查序列延伸的效应,长度41个核苷酸的crRNA通过9个核苷酸进行延伸(crRNA<sup>+9</sup>)、或通过59个核苷酸进行延伸(crRNA<sup>+59</sup>)。延伸部分包括自切割成活性crRNA的自加工序列。未修饰的crRNA、crRNA<sup>+9</sup>和crRNA<sup>+59</sup>各自个别地与Cpf1复合,并且使用阳离子脂质体转染到GFP-HEK细胞内。RNP在低盐条件下形成,以防止5'末端延伸的潜在加工。其为转染效率的指示剂的NHEJ水平,通过测量GFP阴性细胞的百分比来确定。

[0139] 如图1B中所示,用自加工序列延伸crRNA显著增强了使用阳离子媒介物的转染效率。经由阳离子脂质体用未延伸的crRNA处理的细胞显示8%GFP阴性,而9碱基延伸的crRNA处理的细胞为18%GFP-,而59碱基延伸的crRNA处理的细胞为37%,其为超过对照未修饰的crRNA的大约4倍增加。另外,用阳离子脂质体测试了crRNA+9、crRNA+15和crRNA+25,并且显示了基因编辑效率的长度依赖性增加(图1C)。

[0140] 为了确定对于这种增强是否存在特定的5'延伸序列要求,比较了三种不同的59核苷酸的5'延伸。上文描述了第一个和原始的59核苷酸延伸的crRNA,并且含有一个AsCpf1前crRNA(crRNA+59),第二个59碱基延伸的crRNA含有串联的四个AsCpf1前crRNA位点(crRNA+59-D2),而第三个59碱基延伸的crRNA含有FnCpf1前crRNA,其后为与人基因组中的任何序列没有同源性的乱序的DNA序列(crRNA+59-D3)(图1D)。这些crRNA使用阳离子脂质体2000进行递送,类似于上述段落。所有三个5'延伸部分都显示等价的编辑活性:crRNA+59细胞为32%GFP阴性,crRNA+59\_D2细胞为30%GFP阴性,而crRNA+59\_D3细胞为27%GFP阴性(图1E)。这提示对于5'延伸增强不存在严格的序列要求,类似于用9核苷酸延伸的crRNA与电穿孔细胞的先前发现。另外,这些结果提供了支持上文假设的证据,即增加crRNA上的负电荷密度,并且从而也增加AsCpf1 RNP复合物上的负电荷密度,可以增强AsCpf1通过阳离子脂质对细胞的递送。

[0141] 在体外DNA切割测定中测试了延伸的crRNA,以确定延伸的crRNA是否增强了Cpf1的固有核酸酶活性。伴随15分钟和60分钟的温育时间,在测试的三种crRNA之间没有观察到活性中的差异:野生型crRNA、crRNA<sup>+9</sup>和crRNA<sup>+59</sup>。当温育时间仅为5分钟时,crRNA<sup>+59</sup>甚至具有比野生型crRNA更缓慢的DNA切割。

[0142] 还研究了5'延伸的crRNA,以确定如果通过质粒而不是作为RNP递送,是否存在Cpf1中增强的基因编辑活性。在crRNA的电穿孔之前24小时转染Cpf1质粒,并且确定基因编辑活性。当从质粒产生Cpf1时,未观察到基因编辑效率中的改善。

[0143] 进一步地,用荧光染料标记crRNA,以确定在经由电穿孔或阳离子脂质体递送后,延伸的crRNA在细胞中是否具有增强的摄取。Cpf1 RNP的电穿孔导致90%以上的细胞对于染料-crRNA呈阳性,并且无论crRNA的长度如何,都显示高度有效的递送。Cpf1 RNP用阳离子脂质体的递送效率取决于crRNA的长度,并且延伸的crRNA比野生型crRNA更有效地递送到HEK 293T细胞内(图1F)。

[0144] 结果显示如本文提供的延伸的crRNA增强了递送和基因编辑效率。

[0145] 实施例3

[0146] 本实施例证实了,Cpf1的活性在使用电穿孔的HEK细胞中,由5'末端延伸得到增强。

[0147] 通过电穿孔,将具有4、9、15、25和59个核苷酸的5'末端延伸的GFP靶向crRNA作为具有AsCpf1的RNP复合物引入GFP-HEK细胞内。关于4至25核苷酸延伸的序列进行乱序,并且59核苷酸延伸由AsCpf1前crRNA,其后为与人基因组序列没有同源性的乱序RNA序列组成。

[0148] 具有4至25核苷酸5'延伸的crRNA全都显示出超过没有延伸部分的crRNA急剧增加的基因编辑。使用未延伸的crRNA电穿孔的细胞为30%GFP阴性(crRNA),使用4至25核苷酸延伸的crRNA为55至60%GFP阴性,而使用59核苷酸延伸的crRNA为37%GFP阴性(图2)。关于4至25核苷酸的5'延伸的crRNA的基因编辑水平与SpCas9 RNP电穿孔细胞的基因编辑水平可比较。

[0149] 结果确认了Cpf1 crRNA的5'延伸增加了基因编辑效率。

[0150] 实施例4

[0151] 本实施例证实了包含Cpf1 crRNA和5'延伸序列的核酸增强了Cpf1在体内通过阳离子脂质的递送以及细胞中的Cpf1活性。

[0152] 对具有延伸的crRNA调查了三种不同的化学修饰:2'-O-甲基修饰、硫代磷酸酯键合和脱氧核苷酸核糖基团(图15A)。4个核苷酸中的前3个用以下进行延伸:2'-O-甲基核苷酸和3'硫代磷酸酯键合(MS)、在9核苷酸的5'-延伸的crRNA的第9个位置处的脱氧核苷酸(9dU),对所有9个核苷酸的3'硫代磷酸酯键合加上在9核苷酸的5'-延伸的crRNA的第9个位置处的脱氧核苷酸(9s)。

[0153] 将含有具有延伸和化学修饰的crRNA的Cpf1 RNP电穿孔到GFP-HEK细胞内,并且通过流式细胞术确定基因编辑活性。具有化学修饰的延伸的crRNA具有与未修饰的延伸的crRNA相似的活性(41%至46%的GFP阴性细胞)(图15D)。

[0154] 另外,使用表达蓝色荧光蛋白(BFP)的HEK293T细胞系(BFP-HEK)检查了这些crRNA。结果呈现于图12中。

[0155] 来自该实验的结果显示了5'延伸增加了AsCpf1的基因编辑效率,以及crRNA的5'末端对于化学修饰的耐受性。进一步地,如果crRNA的5'末端是延伸的,则crRNA的5'化学修饰是可能的,而不损害活性。

[0156] 使用化学修饰的crRNA的关键益处在于它们对通过血清核酸酶的水解更稳定。因此,调查了5'化学修饰的crRNA的血清稳定性。



[0157] 5'化学修饰的crRNA在稀释的胎牛血清中进行温育,并且经由凝胶电泳分析其降解。图15B提供了在血清中的15分钟温育后剩余的crRNA的定量。结果显示未修饰的crRNA在血清中快速降解,而含有硫代磷酸酯主链的crRNA<sup>+9S</sup>对血清中的水解明显更稳定。

[0158] 还研究了5'修饰的crRNA,以确定由于其保护crRNA免于细胞和血清中的核酸酶的能力,它们是否可以增强阳离子脂质体转染Cpf1 RNP的能力。具有crRNA<sup>+9S</sup>的Cpf1在编辑细胞中的基因方面比crRNA<sup>+9</sup>更有效40%,其提示了通过5'crRNA延伸实现的5'crRNA化学修饰在基因编辑中具有众多应用(图15C)。

[0159] Cpf1 RNP的晶体结构最近已得到分辨,并且证实了AsCpf1蛋白与crRNA的磷酸二酯主链形成众多相互作用。因此,未延伸的crRNA的5'-化学修饰具有破坏crRNA与Cpf1之间的重要相互作用的高机率,导致AsCpf1基因编辑活性的破坏。

[0160] 相比之下,具有5'延伸的crRNA看起来耐受化学修饰,因为与AsCpf1蛋白相互作用的核苷酸是未修饰的。这些结果提供了用于在crRNA的5'末端处引入化学修饰的方法,其可以潜在地增强用于离体和体内治疗应用的递送。此类构建体还使得其它分子,例如靶向配体、内体逃逸部分或其它功能性分子,与延伸的crRNA缀合并且与Cpf1分子一起保留。例如,生物素或抗生物素蛋白(或链霉抗生物素蛋白)可以与crRNA延伸缀合,允许修饰的crRNA适当结合与链霉抗生物素蛋白或生物素缀合的另一种分子(例如,靶向分子)(例如图14)。另外,crRNA可以通过延伸与Cpf1 mRNA缀合,允许Cpf1通过mRNA的翻译被递送。一旦Cpf1 mRNA在细胞质中翻译并产生Cpf1蛋白,Cpf1蛋白就识别构建体的crRNA部分,并且任选地加工连接RNA序列,分开Cpf1 mRNA和crRNA。

[0161] 实施例5

[0162] 还执行了实验以确定具有延伸序列的crRNA是否可以增强阳离子聚合物转染Cpf1的能力。特别地,如实施例2中使用的各种长度的crRNA(不含延伸、具有9nt延伸和59nt延伸的crRNA)与Cpf1复合,与阳离子聚合物PAsp(DET)混合,并且加入GFP-HEK细胞中。经由流式细胞术,通过测量GFP阴性细胞的频率,来确定制剂的NHEJ效率。

[0163] 如图16中所示,59碱基的5'延伸将PAsp(DET)介导的AsCpf1 RNP对细胞的递送增强2倍。未延伸的crRNA(crRNA)为8%GFP阴性,9核苷酸延伸的crRNA(crRNA+9)为10%GFP阴性,而59核苷酸延伸的crRNA(crRNA+59)为18%GFP阴性。这些结果证实了,延伸序列可以使用阳离子聚合物改善crRNA对细胞的递送。

[0164] 实施例6

[0165] 本实施例证实了延伸的crRNA增强了在体内的Cpf1递送。

[0166] 执行实验以确定延伸的crRNA(crRNA<sup>+59</sup>)是否可以增强阳离子脂质介导的在体内的Cpf1递送。实验的示意图在图3A中提供。

[0167] 使用先前验证的间隔物在Ai9小鼠中执行研究。Ai9小鼠给予与以下任一组合的阳离子脂质或PAsp(DET)的肌肉注射:AsCpf1-crRNA复合物或AsCpf1-crRNA<sup>+59</sup>复合物。在注射后两周,tdTomato(红色荧光)的表达在腓肠肌的10μm切片中进行成像(图3B中的肌肉图)。对于未延伸和延伸的crRNA收集的图像比较显示了,延伸的crRNA急剧增强了PAsp(DET)在体内递送AsCpf1 RNP的能力。另外,与PAsp(DET)复合的具有延伸的crRNA的RNP(Cpf1 RNP+59),诱导了远离注射部位数毫米以及整个腓肠肌的tdTomato表达。肌肉中的tdTomato表达的高范围很可能是由于肌肉纤维独特的多核性质。这允许TdTomato在肌肉纤维的整个长度

上表达,并且因此可在几毫米的长度上观察到。

[0168] 59个核苷酸延伸的crRNA增强递送,并且通过延伸增强AsCpf1 RNP在体内的编辑效率的能力,支持了Cpf1作为用于动物研究的工具、以及作为用于治疗人疾病,尤其是遗传性肌营养不良的潜在治疗剂的价值。

[0169] 实施例7

[0170] 5.1:延伸的crRNA增加HDR和NHEJ率

[0171] 为了检查5'-延伸是否可以增加除NHEJ水平之外的HDR率,具有含有各种延伸的crRNA的AsCpf1 RNP,连同单链寡核苷酸供体(ssODN)一起引入GFP-HEK细胞内。NHEJ水平通过测量GFP阴性细胞的群体来确定(类似于第一节段),而HDR率使用限制性酶消化测定来定量。对于4和9个核苷酸延伸的crRNA两者,均观察到HDR中的2倍改善(在图4A中,关于crRNA+4的17%HDR频率,以及关于crRNA+9的18%HDR频率,相对于关于未修饰的crRNA的9%)。对于59碱基延伸,观察到HDR中的较小增加(在图4A中,关于crRNA+59的13%HDR率)。

[0172] 有趣的是,用于HDR的ssODN也急剧增加了AsCpf1的NHEJ效率。ssODN将GFP阴性细胞的百分比对于未延伸的crRNA(crRNA)从30%增加到46%,对于4碱基延伸的crRNA(crRNA+4)从55%增加到95%,对于9碱基延伸的crRNA(crRNA+9)从58%增加到93%,并且对于59碱基延伸的crRNA(crRNA+59)从37%增加到58%(图4B)。在BFP报告系统中进一步验证了外源添加的DNA增强了AsCpf1 RNP介导的编辑的发现。与人基因组没有任何同源性的单链DNA(ssDNA)用AsCpf1 RNP电穿孔到BFP-HEK细胞内。类似地,对于未延伸和延伸的crRNA两者,ssDNA的添加使AsCpf1编辑活性增加2倍。BFP阴性群体对于未延伸的crRNA(crRNA)从31%增加到50%,对于4核苷酸延伸的crRNA(crRNA+4)从59%增加到91%,并且对于9核苷酸延伸的crRNA(crRNA+9)从60%增加到95%(图4C)。还执行了另外的实验,以确定外源添加的DNA是否必须与Cpf1 RNP靶位点具有同源性,以便增强基因编辑。AsCpf1 RNP连同与靶序列没有任何同源性单链DNA(ssDNA)一起电穿孔到细胞内,并且测量了基因编辑效率。类似地,对于两种延伸的crRNA,没有同源性的ssDNA的添加也将AsCpf1编辑活性增加到大约90%(图4D)。结果证实了ssDNA可以加强用AsCpf1的编辑。另外,具有5'末端延伸的活性增强与外源ssDNA协同作用,并且它们共同地诱导的基因编辑接近100%。观察到如果阳离子脂质体用作代替电穿孔的递送方法,则ssDNA的添加并不增强Cpf1基因编辑效率。

[0173] 有时优选使用RNA而不是DNA用于基因编辑,因为ssRNA无法整合到基因组内,并且可以更安全地使用。为了测试关于ssRNA而不是DNA的效应,GFP-HEK细胞用Cpf1 RNP和两种不同的ssRNA(9nt和100nt)进行电穿孔,并且测定了所得到的基因编辑水平。具有轻微序列变异的两种100nt ssRNA均急剧增加了Cpf1的基因编辑效率,导致2倍改善,而9nt ssRNA诱导了基因编辑效率中的10%增强(图4E)。

[0174] 这些结果证实了单链核酸可以用于加强细胞中的Cpf1编辑活性。

[0175] 5.2:单分子延伸的crRNA和供体DNA。

[0176] 本实施例的第2部分证实了在单个分子中与供体DNA组合的延伸的crRNA可以增强HDR。

[0177] 绝大多数遗传疾病需要基因校正而不是敲除,并且因此,在开发可以经由HDR校正基因突变的基于Cpf1的治疗剂方面存在极大的兴趣。为了解决生成有效封装供体DNA和Cpf1-crRNA两者的纳米颗粒的问题,执行了实验以确定crRNA和供体DNA是否可以经由可逆

的二硫键组合成单个分子。如先前报道的,一个挑战是crRNA的5'末端对化学修饰相当敏感。因此,测试了在crRNA延伸的自加工序列的5'末端处的化学修饰。

[0178] crRNA在其5'末端处用四个另外的核苷酸进行延伸,并且将化学修饰(即5' DBCO、5' 硫醇或5' 叠氮化物)加入核苷酸的末端。测试了在5'末端处具有和不具有化学修饰的4nt延伸的crRNA的活性。简言之,将5'修饰的超荷电的crRNA(其设计为经由插入缺失形成来击倒GFP基因)与Cpf1复合,并且电穿孔(核转染)到细胞内。然后经由流式细胞术,通过测量GFP敲除细胞的数目来确定基因编辑效率,其中不再表达GFP的细胞指示该细胞已由Cpf1/crRNA复合物转染。

[0179] 图4F显示了化学修饰或核苷酸延伸无一影响Cpf1-crRNA活性,因为所有crRNA复合物都显示约40%的GFP敲除,这是与未修饰的对照crRNA相似的水平。

[0180] 由于可以添加在crRNA的5'末端上的化学而不损失活性,因此crRNA的5'末端被硫代吡啶激活,以与硫醇封端的供体DNA反应(参见图5A-5C)。两个大分子之间的反应具有缓慢的动力学。因此,使用了使用与crRNA和供体DNA两者均互补的桥DNA的方法。所述桥使两个大分子杂交并且靠近,以促进反应,以便增强crRNA与供体DNA之间的缀合产率。缀合产率高达40%,并且产物经由凝胶提取进行纯化。该缀合物被命名为“同源DNA-crRNA”(HD-RNA)。HD-RNA包含二硫键,其应该在细胞质中被还原。硫醇介导的HD-RNA切割通过使其在DTT中温育6小时,并且经由凝胶电泳分析其分子量来确定。图5C中的凝胶比较显示了DTT使HD-RNA还原,并且再生供体DNA和crRNA。

[0181] 将与Cpf1复合的HD-RNA电穿孔到GFP-HEK细胞内,并且将HDR和NHEJ的水平分别与用Cpf1-crRNA和供体DNA电穿孔的细胞进行比较。与用Cpf1-crRNA和供体DNA电穿孔的细胞相比,从用HD-RNA电穿孔的细胞观察到相似的NHEJ和HDR水平,其证实了crRNA与供体DNA经由二硫键的缀合并不影响crRNA或供体DNA的功能性。

[0182] 执行了实验,以确定在用阳离子脂质(即阳离子脂质体)或阳离子聚合物(即PAsp(DET))转染后,HD-RNA是否增强了Cpf1的HDR效率。对于这些实验,使用阳离子脂质体或PAsp(DET),将与Cpf1复合的HD-RNA转染到GFP-HEK细胞内,并且比较了HDR和NHEJ的水平。NHEJ通过测量GFP阴性细胞的频率来确定,并且通过用BFP基因的靶向区域的PCR扩增子执行Surveyor测定来确认。通过分离细胞DNA,并且分析供体DNA中嵌入的限制性酶位点的存在,来确定HDR效率。结果呈现于图6和7中。

[0183] 图6和7证实了在用PAsp(DET)递送后,HD-RNA增强了Cpf1的NHEJ和HDR效率两者。具体地,HDR在高达60%的用PAsp(DET)递送的HD-RNA/Cpf1复合物处理的细胞中检测到,其显著高于用与PAsp(DET)复合的Cpf1/crRNA和供体DNA处理的细胞的HDR率。另外,用与PAsp(DET)一起递送的HD-RNA/Cpf1复合物观察到的60%HDR率甚至高于用Cpf1/crRNA复合物的电穿孔观察到的HDR率,并且提示在Cpf1切割位点的附近具有供体DNA可以有助于HDR。

[0184] 这些结果证实了延伸的crRNA可以增强HDR。

[0185] 实施例9

[0186] 本实施例证实了延伸的crRNA在不同细胞类型中的用途,以及该方法治疗遗传病症的效用。

[0187] HD-RNA具有众多的潜在应用,由于其增强Cpf1在用阳离子脂质递送后在细胞中生成HDR的能力。杜兴氏肌营养不良(DMD)作为关于HD-RNA的最初医学应用进行测试。DMD是由

肌营养不良蛋白基因中的突变引起的早发型致命疾病；它是最常见的先天性肌病，并且大约30%的DMD患者具有单碱基突变或小缺失，其可以潜在地用基于HDR的治疗剂进行治疗。

[0188] 因此，就其经由HDR校正得自mdx小鼠的成肌细胞中的肌营养不良蛋白突变的能力，测试了设计为靶向肌营养不良蛋白基因的HD-RNA。设计了HD-RNA，其可以切割肌营养不良蛋白基因，并且还含有设计为校正存在于其肌营养不良蛋白基因中的C至T突变的供体DNA（参见图10A和10B）。确定了用Cpf1/HD-RNA+阳离子脂质体处理的mdx成肌细胞中的HDR率，并且针对用Cpf1-crRNA、供体DNA和阳离子脂质体处理的mdx成肌细胞进行比较。发现与用Cpf1 RNP和供体DNA处理的细胞相比，与HD-RNA复合的Cpf1在mdx成肌细胞中生成HDR方面更有效。例如，HD-RNA处理的细胞具有5-10%的HDR率，而对照细胞仅具有1%的HDR率。

[0189] 在另一个实验中，用与具有和不具有5'延伸的crRNA复合的AsCpf1RNP，电穿孔从Ai9小鼠中分离的原代成肌细胞，所述Ai9小鼠是在所有3个读码框中都含有终止密码子的转基因小鼠品系，所述读码框与tdTomato报告基因上游的三重聚(A)信号偶联。Ai9小鼠是含有tdTomato报告基因的转基因小鼠品系，所述tdTomato报告基因在与三重聚(A)信号偶联的所有三个读码框中都具有终止密码子。AsCpf1间隔物设计为将多重断裂引入DNA内，其导致通过基因组缺失的终止序列去除。成功的基因编辑通过tdTomato（红色荧光蛋白，RFP）的表达得到指示，所述tdTomato可以通过荧光显微镜检查得到显现，并且使用流式细胞术进行定量。延伸的crRNA使基因编辑增加超过未延伸的crRNA40-50%。用未延伸的crRNA处理的成肌细胞为12%RFP阳性的；用2核苷酸延伸的crRNA处理的成肌细胞为15%RFP阳性的；用9核苷酸延伸的crRNA处理的成肌细胞为18%RFP阳性的，而用59核苷酸延伸的crRNA处理的成肌细胞为16%RFP阳性的（图10C）。另外，使用crRNA连同对靶DNA原代成肌细胞没有序列同源性的ssDNA ssRNA（100nt），测试了基因编辑的效率。ssDNA和ssRNA两者均增强了基因编辑效率（图10D）。

[0190] 总的来说，原代成肌细胞中靶序列的缺失提示，延伸的crRNA增强的基因编辑可广泛地跨越遗传靶和细胞类型应用。这些结果证实了HD-RNA可以用作用于遗传疾病的治疗剂。

#### [0191] 实施例10

[0192] 下述实施例证实了5'crRNA延伸增强的基因编辑效应可广泛地跨越遗传靶和细胞类型应用。使用Serpinal1作为测试平台，测试这些crRNA，以查看它们是否可以增强Cpf1 RNP编辑内源基因的能力。

[0193] 使用电穿孔，将靶向Serpinal1基因、具有crRNA或crRNA<sup>+9</sup>的Cpf1转染到HepG2细胞内。选择Serpinal1进行进一步调查，是因为Serpinal1基因中的突变引起 $\alpha$ 1-抗胰蛋白酶缺乏，这使其成为用于治疗性基因编辑的靶。对来自HepG2细胞的基因组DNA进行液滴数字PCR，以定量NHEJ效率。

[0194] 如图10E中所示，与野生型crRNA相比，具有crRNA<sup>+9</sup>的Cpf1 RNP具有增强的NHEJ效率。这些结果进一步指示，5'crRNA延伸的基因编辑效应可广泛地跨越遗传靶应用。

[0195] 本文引用的所有参考文献，包括出版物、专利申请和专利，都在此引入作为参考，其程度如同每个参考文献被个别且特别地指示引入作为参考，并且在本文中整体阐述一样。

[0196] 在描述本发明的上下文中（尤其是在下述权利要求的上下文中），术语“一个”和

“一种”、以及“该/所述”和“至少一个/种”以及类似指示物的使用,应解释为涵盖单数和复数两者,除非本文另有说明或与上下文明显矛盾。术语“至少一个”随后为一个或多个项目的列表(例如,“A和B中的至少一个”)的使用,应理解为意指选自所列出的项目(A或B)的一个项目、或者两个或更多个所列出的项目(A和B)的任何组合,除非本文另有说明或与上下文明显矛盾。除非另有说明,否则术语“包含”、“具有”、“包括”和“含有”应解释为开放式术语(即,意指“包括但不限于”)。除非本文另有说明,否则本文中数值范围的叙述仅仅预期充当个别地提及落入该范围内的每个分开的值的简写方法,并且每个分开的值都并入说明书内,如同它在本文中被个别地叙述一样。除非本文另有说明或与上下文明显矛盾,否则本文描述的所有方法都可以以任何合适的次序执行。除非另外要求,否则本文提供的任何和所有实例或示例性语言(例如,“如”)的使用仅仅预期更好地阐明本发明,并且不对本发明的范围构成限制。说明书中的任何语言都不应解释为指示任何未请求保护的要素对于实践本发明为必要的。

[0197] 本文描述了本发明的优选实施方案,包括本发明人已知的用于实施本发明的最佳模式。在阅读前述说明书后,那些优选实施方案的变化对于本领域普通技术人员将变得显而易见。本发明人预期熟练的技术人员适当地采用此类变化,并且本发明人希望以不同于如本文具体描述的方式来实践本发明。相应地,本发明包括如通过适用法律所允许的,与其所附的权利要求中所述主题的所有修改和等价物。此外,除非本文另有说明或与上下文明显矛盾,否则上述元件在其所有可能的变化中的任何组合都由本发明涵盖。

[0001] 序列表  
 [0002] <110> 基因编辑有限公司  
 [0003] <120> 修饰的CPF1引导RNA  
 [0004] <130> 512701  
 [0005] <160> 39  
 [0006] <170> PatentIn version 3.5  
 [0007] <210> 1  
 [0008] <211> 1300  
 [0009] <212> PRT  
 [0010] <213> 人工序列  
 [0011] <220>  
 [0012] <223> 合成的  
 [0013] <400> 1  
 [0014] Met Ser Ile Tyr Gln Glu Phe Val Asn Lys Tyr Ser Leu Ser Lys Thr  
 [0015] 1 5 10 15  
 [0016] Leu Arg Phe Glu Leu Ile Pro Gln Gly Lys Thr Leu Glu Asn Ile Lys  
 [0017] 20 25 30  
 [0018] Ala Arg Gly Leu Ile Leu Asp Asp Glu Lys Arg Ala Lys Asp Tyr Lys  
 [0019] 35 40 45  
 [0020] Lys Ala Lys Gln Ile Ile Asp Lys Tyr His Gln Phe Phe Ile Glu Glu  
 [0021] 50 55 60  
 [0022] Ile Leu Ser Ser Val Cys Ile Ser Glu Asp Leu Leu Gln Asn Tyr Ser  
 [0023] 65 70 75 80  
 [0024] Asp Val Tyr Phe Lys Leu Lys Lys Ser Asp Asp Asp Asn Leu Gln Lys  
 [0025] 85 90 95  
 [0026] Asp Phe Lys Ser Ala Lys Asp Thr Ile Lys Lys Gln Ile Ser Glu Tyr  
 [0027] 100 105 110  
 [0028] Ile Lys Asp Ser Glu Lys Phe Lys Asn Leu Phe Asn Gln Asn Leu Ile  
 [0029] 115 120 125  
 [0030] Asp Ala Lys Lys Gly Gln Glu Ser Asp Leu Ile Leu Trp Leu Lys Gln  
 [0031] 130 135 140  
 [0032] Ser Lys Asp Asn Gly Ile Glu Leu Phe Lys Ala Asn Ser Asp Ile Thr  
 [0033] 145 150 155 160  
 [0034] Asp Ile Asp Glu Ala Leu Glu Ile Ile Lys Ser Phe Lys Gly Trp Thr  
 [0035] 165 170 175  
 [0036] Thr Tyr Phe Lys Gly Phe His Glu Asn Arg Lys Asn Val Tyr Ser Ser  
 [0037] 180 185 190  
 [0038] Asn Asp Ile Pro Thr Ser Ile Ile Tyr Arg Ile Val Asp Asp Asn Leu

[0039]	195	200	205
[0040]	Pro Lys Phe Leu Glu Asn Lys Ala Lys Tyr Glu Ser Leu Lys Asp Lys		
[0041]	210	215	220
[0042]	Ala Pro Glu Ala Ile Asn Tyr Glu Gln Ile Lys Lys Asp Leu Ala Glu		
[0043]	225	230	235
[0044]	Glu Leu Thr Phe Asp Ile Asp Tyr Lys Thr Ser Glu Val Asn Gln Arg		
[0045]	245	250	255
[0046]	Val Phe Ser Leu Asp Glu Val Phe Glu Ile Ala Asn Phe Asn Asn Tyr		
[0047]	260	265	270
[0048]	Leu Asn Gln Ser Gly Ile Thr Lys Phe Asn Thr Ile Ile Gly Gly Lys		
[0049]	275	280	285
[0050]	Phe Val Asn Gly Glu Asn Thr Lys Arg Lys Gly Ile Asn Glu Tyr Ile		
[0051]	290	295	300
[0052]	Asn Leu Tyr Ser Gln Gln Ile Asn Asp Lys Thr Leu Lys Lys Tyr Lys		
[0053]	305	310	315
[0054]	Met Ser Val Leu Phe Lys Gln Ile Leu Ser Asp Thr Glu Ser Lys Ser		
[0055]	325	330	335
[0056]	Phe Val Ile Asp Lys Leu Glu Asp Asp Ser Asp Val Val Thr Thr Met		
[0057]	340	345	350
[0058]	Gln Ser Phe Tyr Glu Gln Ile Ala Ala Phe Lys Thr Val Glu Glu Lys		
[0059]	355	360	365
[0060]	Ser Ile Lys Glu Thr Leu Ser Leu Leu Phe Asp Asp Leu Lys Ala Gln		
[0061]	370	375	380
[0062]	Lys Leu Asp Leu Ser Lys Ile Tyr Phe Lys Asn Asp Lys Ser Leu Thr		
[0063]	385	390	395
[0064]	Asp Leu Ser Gln Gln Val Phe Asp Asp Tyr Ser Val Ile Gly Thr Ala		
[0065]	405	410	415
[0066]	Val Leu Glu Tyr Ile Thr Gln Gln Ile Ala Pro Lys Asn Leu Asp Asn		
[0067]	420	425	430
[0068]	Pro Ser Lys Lys Glu Gln Glu Leu Ile Ala Lys Lys Thr Glu Lys Ala		
[0069]	435	440	445
[0070]	Lys Tyr Leu Ser Leu Glu Thr Ile Lys Leu Ala Leu Glu Glu Phe Asn		
[0071]	450	455	460
[0072]	Lys His Arg Asp Ile Asp Lys Gln Cys Arg Phe Glu Glu Ile Leu Ala		
[0073]	465	470	475
[0074]	Asn Phe Ala Ala Ile Pro Met Ile Phe Asp Glu Ile Ala Gln Asn Lys		
[0075]	485	490	495
[0076]	Asp Asn Leu Ala Gln Ile Ser Ile Lys Tyr Gln Asn Gln Gly Lys Lys		
[0077]	500	505	510

[0078]	Asp	Leu	Leu	Gln	Ala	Ser	Ala	Glu	Asp	Asp	Val	Lys	Ala	Ile	Lys	Asp
[0079]				515				520					525			
[0080]	Leu	Leu	Asp	Gln	Thr	Asn	Asn	Leu	Leu	His	Lys	Leu	Lys	Ile	Phe	His
[0081]				530				535					540			
[0082]	Ile	Ser	Gln	Ser	Glu	Asp	Lys	Ala	Asn	Ile	Leu	Asp	Lys	Asp	Glu	His
[0083]	545					550					555				560	
[0084]	Phe	Tyr	Leu	Val	Phe	Glu	Glu	Cys	Tyr	Phe	Glu	Leu	Ala	Asn	Ile	Val
[0085]					565					570					575	
[0086]	Pro	Leu	Tyr	Asn	Lys	Ile	Arg	Asn	Tyr	Ile	Thr	Gln	Lys	Pro	Tyr	Ser
[0087]				580					585					590		
[0088]	Asp	Glu	Lys	Phe	Lys	Leu	Asn	Phe	Glu	Asn	Ser	Thr	Leu	Ala	Asn	Gly
[0089]				595				600					605			
[0090]	Trp	Asp	Lys	Asn	Lys	Glu	Pro	Asp	Asn	Thr	Ala	Ile	Leu	Phe	Ile	Lys
[0091]		610					615					620				
[0092]	Asp	Asp	Lys	Tyr	Tyr	Leu	Gly	Val	Met	Asn	Lys	Lys	Asn	Asn	Lys	Ile
[0093]	625					630					635				640	
[0094]	Phe	Asp	Asp	Lys	Ala	Ile	Lys	Glu	Asn	Lys	Gly	Glu	Gly	Tyr	Lys	Lys
[0095]					645					650					655	
[0096]	Ile	Val	Tyr	Lys	Leu	Leu	Pro	Gly	Ala	Asn	Lys	Met	Leu	Pro	Lys	Val
[0097]				660					665					670		
[0098]	Phe	Phe	Ser	Ala	Lys	Ser	Ile	Lys	Phe	Tyr	Asn	Pro	Ser	Glu	Asp	Ile
[0099]				675				680						685		
[0100]	Leu	Arg	Ile	Arg	Asn	His	Ser	Thr	His	Thr	Lys	Asn	Gly	Ser	Pro	Gln
[0101]		690					695					700				
[0102]	Lys	Gly	Tyr	Glu	Lys	Phe	Glu	Phe	Asn	Ile	Glu	Asp	Cys	Arg	Lys	Phe
[0103]	705					710					715				720	
[0104]	Ile	Asp	Phe	Tyr	Lys	Gln	Ser	Ile	Ser	Lys	His	Pro	Glu	Trp	Lys	Asp
[0105]					725					730				735		
[0106]	Phe	Gly	Phe	Arg	Phe	Ser	Asp	Thr	Gln	Arg	Tyr	Asn	Ser	Ile	Asp	Glu
[0107]				740					745					750		
[0108]	Phe	Tyr	Arg	Glu	Val	Glu	Asn	Gln	Gly	Tyr	Lys	Leu	Thr	Phe	Glu	Asn
[0109]				755				760					765			
[0110]	Ile	Ser	Glu	Ser	Tyr	Ile	Asp	Ser	Val	Val	Asn	Gln	Gly	Lys	Leu	Tyr
[0111]		770					775					780				
[0112]	Leu	Phe	Gln	Ile	Tyr	Asn	Lys	Asp	Phe	Ser	Ala	Tyr	Ser	Lys	Gly	Arg
[0113]	785					790					795				800	
[0114]	Pro	Asn	Leu	His	Thr	Leu	Tyr	Trp	Lys	Ala	Leu	Phe	Asp	Glu	Arg	Asn
[0115]					805					810				815		
[0116]	Leu	Gln	Asp	Val	Val	Tyr	Lys	Leu	Asn	Gly	Glu	Ala	Glu	Leu	Phe	Tyr



[0117]	820						825						830					
[0118]	Arg	Lys	Gln	Ser	Ile	Pro	Lys	Lys	Ile	Thr	His	Pro	Ala	Lys	Glu	Ala		
[0119]	835						840						845					
[0120]	Ile	Ala	Asn	Lys	Asn	Lys	Asp	Asn	Pro	Lys	Lys	Glu	Ser	Val	Phe	Glu		
[0121]	850						855						860					
[0122]	Tyr	Asp	Leu	Ile	Lys	Asp	Lys	Arg	Phe	Thr	Glu	Asp	Lys	Phe	Phe	Phe		
[0123]	865						870						875					
[0124]	His	Cys	Pro	Ile	Thr	Ile	Asn	Phe	Lys	Ser	Ser	Gly	Ala	Asn	Lys	Phe		
[0125]	885						890						895					
[0126]	Asn	Asp	Glu	Ile	Asn	Leu	Leu	Leu	Lys	Glu	Lys	Ala	Asn	Asp	Val	His		
[0127]	900						905						910					
[0128]	Ile	Leu	Ser	Ile	Asp	Arg	Gly	Glu	Arg	His	Leu	Ala	Tyr	Tyr	Thr	Leu		
[0129]	915						920						925					
[0130]	Val	Asp	Gly	Lys	Gly	Asn	Ile	Ile	Lys	Gln	Asp	Thr	Phe	Asn	Ile	Ile		
[0131]	930						935						940					
[0132]	Gly	Asn	Asp	Arg	Met	Lys	Thr	Asn	Tyr	His	Asp	Lys	Leu	Ala	Ala	Ile		
[0133]	945						950						955					
[0134]	Glu	Lys	Asp	Arg	Asp	Ser	Ala	Arg	Lys	Asp	Trp	Lys	Lys	Ile	Asn	Asn		
[0135]	965						970						975					
[0136]	Ile	Lys	Glu	Met	Lys	Glu	Gly	Tyr	Leu	Ser	Gln	Val	Val	His	Glu	Ile		
[0137]	980						985						990					
[0138]	Ala	Lys	Leu	Val	Ile	Glu	Tyr	Asn	Ala	Ile	Val	Val	Phe	Glu	Asp	Leu		
[0139]	995						1000						1005					
[0140]	Asn	Phe	Gly	Phe	Lys	Arg	Gly	Arg	Phe	Lys	Val	Glu	Lys	Gln	Val			
[0141]	1010						1015						1020					
[0142]	Tyr	Gln	Lys	Leu	Glu	Lys	Met	Leu	Ile	Glu	Lys	Leu	Asn	Tyr	Leu			
[0143]	1025						1030						1035					
[0144]	Val	Phe	Lys	Asp	Asn	Glu	Phe	Asp	Lys	Thr	Gly	Gly	Val	Leu	Arg			
[0145]	1040						1045						1050					
[0146]	Ala	Tyr	Gln	Leu	Thr	Ala	Pro	Phe	Glu	Thr	Phe	Lys	Lys	Met	Gly			
[0147]	1055						1060						1065					
[0148]	Lys	Gln	Thr	Gly	Ile	Ile	Tyr	Tyr	Val	Pro	Ala	Gly	Phe	Thr	Ser			
[0149]	1070						1075						1080					
[0150]	Lys	Ile	Cys	Pro	Val	Thr	Gly	Phe	Val	Asn	Gln	Leu	Tyr	Pro	Lys			
[0151]	1085						1090						1095					
[0152]	Tyr	Glu	Ser	Val	Ser	Lys	Ser	Gln	Glu	Phe	Phe	Ser	Lys	Phe	Asp			
[0153]	1100						1105						1110					
[0154]	Lys	Ile	Cys	Tyr	Asn	Leu	Asp	Lys	Gly	Tyr	Phe	Glu	Phe	Ser	Phe			
[0155]	1115						1120						1125					

[0156]	Asp Tyr Lys Asn Phe Gly Asp Lys Ala Ala Lys Gly Lys Trp Thr		
[0157]	1130	1135	1140
[0158]	Ile Ala Ser Phe Gly Ser Arg Leu Ile Asn Phe Arg Asn Ser Asp		
[0159]	1145	1150	1155
[0160]	Lys Asn His Asn Trp Asp Thr Arg Glu Val Tyr Pro Thr Lys Glu		
[0161]	1160	1165	1170
[0162]	Leu Glu Lys Leu Leu Lys Asp Tyr Ser Ile Glu Tyr Gly His Gly		
[0163]	1175	1180	1185
[0164]	Glu Cys Ile Lys Ala Ala Ile Cys Gly Glu Ser Asp Lys Lys Phe		
[0165]	1190	1195	1200
[0166]	Phe Ala Lys Leu Thr Ser Val Leu Asn Thr Ile Leu Gln Met Arg		
[0167]	1205	1210	1215
[0168]	Asn Ser Lys Thr Gly Thr Glu Leu Asp Tyr Leu Ile Ser Pro Val		
[0169]	1220	1225	1230
[0170]	Ala Asp Val Asn Gly Asn Phe Phe Asp Ser Arg Gln Ala Pro Lys		
[0171]	1235	1240	1245
[0172]	Asn Met Pro Gln Asp Ala Asp Ala Asn Gly Ala Tyr His Ile Gly		
[0173]	1250	1255	1260
[0174]	Leu Lys Gly Leu Met Leu Leu Gly Arg Ile Lys Asn Asn Gln Glu		
[0175]	1265	1270	1275
[0176]	Gly Lys Lys Leu Asn Leu Val Ile Lys Asn Glu Glu Tyr Phe Glu		
[0177]	1280	1285	1290
[0178]	Phe Val Gln Asn Arg Asn Asn		
[0179]	1295	1300	
[0180]	<210> 2		
[0181]	<211> 17		
[0182]	<212> RNA		
[0183]	<213> 人工序列		
[0184]	<220>		
[0185]	<223> 合成的		
[0186]	<400> 2		
[0187]	gucuaagaac uuuaau 17		
[0188]	<210> 3		
[0189]	<211> 16		
[0190]	<212> RNA		
[0191]	<213> 人工序列		
[0192]	<220>		
[0193]	<223> 合成的		
[0194]	<400> 3		

[0195] gucaaaagac cuuuuu 16  
[0196] <210> 4  
[0197] <211> 16  
[0198] <212> RNA  
[0199] <213> 人工序列  
[0200] <220>  
[0201] <223> 合成的  
[0202] <400> 4  
[0203] guuucaaaga uuaaaau 16  
[0204] <210> 5  
[0205] <211> 17  
[0206] <212> RNA  
[0207] <213> 人工序列  
[0208] <220>  
[0209] <223> 合成的  
[0210] <400> 5  
[0211] cucuagcagg ccuggca 17  
[0212] <210> 6  
[0213] <211> 17  
[0214] <212> RNA  
[0215] <213> 人工序列  
[0216] <220>  
[0217] <223> 合成的  
[0218] <400> 6  
[0219] auuugaaagc aucuuuu 17  
[0220] <210> 7  
[0221] <211> 17  
[0222] <212> RNA  
[0223] <213> 人工序列  
[0224] <220>  
[0225] <223> 合成的  
[0226] <400> 7  
[0227] ggcuauaaag cuuauuu 17  
[0228] <210> 8  
[0229] <211> 17  
[0230] <212> RNA  
[0231] <213> 人工序列  
[0232] <220>  
[0233] <223> 合成的

---

[0234]	<400> 8
[0235]	gccaaauacc ucuauaa 17
[0236]	<210> 9
[0237]	<211> 15
[0238]	<212> RNA
[0239]	<213> 人工序列
[0240]	<220>
[0241]	<223> 合成的
[0242]	<400> 9
[0243]	gucuaacuac cuuuu 15
[0244]	<210> 10
[0245]	<211> 15
[0246]	<212> RNA
[0247]	<213> 人工序列
[0248]	<220>
[0249]	<223> 合成的
[0250]	<400> 10
[0251]	gucuaacuac cuuuu 15
[0252]	<210> 11
[0253]	<211> 15
[0254]	<212> RNA
[0255]	<213> 人工序列
[0256]	<220>
[0257]	<223> 合成的
[0258]	<400> 11
[0259]	gucuaacuac cuuuu 15
[0260]	<210> 12
[0261]	<211> 17
[0262]	<212> RNA
[0263]	<213> 人工序列
[0264]	<220>
[0265]	<223> 合成的
[0266]	<400> 12
[0267]	cucuaauaag agauaug 17
[0268]	<210> 13
[0269]	<211> 18
[0270]	<212> RNA
[0271]	<213> 人工序列
[0272]	<220>

[0273] <223> 合成的  
[0274] <400> 13  
[0275] ugcuuagaac auuuaaag 18  
[0276] <210> 14  
[0277] <211> 17  
[0278] <212> RNA  
[0279] <213> 人工序列  
[0280] <220>  
[0281] <223> 合成的  
[0282] <400> 14  
[0283] guuuaaaacc acuuuaa 17  
[0284] <210> 15  
[0285] <211> 17  
[0286] <212> RNA  
[0287] <213> 人工序列  
[0288] <220>  
[0289] <223> 合成的  
[0290] <400> 15  
[0291] cucuacaacu gauaaag 17  
[0292] <210> 16  
[0293] <211> 17  
[0294] <212> RNA  
[0295] <213> 人工序列  
[0296] <220>  
[0297] <223> 合成的  
[0298] <400> 16  
[0299] guuuaaaagu ccuauug 17  
[0300] <210> 17  
[0301] <211> 17  
[0302] <212> RNA  
[0303] <213> 人工序列  
[0304] <220>  
[0305] <223> 合成的  
[0306] <400> 17  
[0307] gucuaaaacu cauucag 17  
[0308] <210> 18  
[0309] <211> 18  
[0310] <212> RNA  
[0311] <213> 人工序列

---

[0312]	<220>
[0313]	<223> 合成的
[0314]	<400> 18
[0315]	ugcuuaguac uuauaaag 18
[0316]	<210> 19
[0317]	<211> 17
[0318]	<212> RNA
[0319]	<213> 人工序列
[0320]	<220>
[0321]	<223> 合成的
[0322]	<400> 19
[0323]	gccaagaacc uauagau 17
[0324]	<210> 20
[0325]	<211> 17
[0326]	<212> RNA
[0327]	<213> 人工序列
[0328]	<220>
[0329]	<223> 合成的
[0330]	<400> 20
[0331]	gucuauaaga cauuuau 17
[0332]	<210> 21
[0333]	<211> 19
[0334]	<212> RNA
[0335]	<213> 人工序列
[0336]	<220>
[0337]	<223> 合成的
[0338]	<400> 21
[0339]	aaauucuacu guuguagau 19
[0340]	<210> 22
[0341]	<211> 19
[0342]	<212> RNA
[0343]	<213> 人工序列
[0344]	<220>
[0345]	<223> 合成的
[0346]	<400> 22
[0347]	aaauucuacu cuuguagau 19
[0348]	<210> 23
[0349]	<211> 20
[0350]	<212> RNA

---

[0351]	<213> 人工序列
[0352]	<220>
[0353]	<223> 合成的
[0354]	<400> 23
[0355]	aaauucuacu aaguguagau 20
[0356]	<210> 24
[0357]	<211> 19
[0358]	<212> RNA
[0359]	<213> 人工序列
[0360]	<220>
[0361]	<223> 合成的
[0362]	<400> 24
[0363]	aaauucuacu guuguagau 19
[0364]	<210> 25
[0365]	<211> 19
[0366]	<212> RNA
[0367]	<213> 人工序列
[0368]	<220>
[0369]	<223> 合成的
[0370]	<400> 25
[0371]	aaauucuacu auuguagau 19
[0372]	<210> 26
[0373]	<211> 19
[0374]	<212> RNA
[0375]	<213> 人工序列
[0376]	<220>
[0377]	<223> 合成的
[0378]	<400> 26
[0379]	aaauucuacu auuguagau 19
[0380]	<210> 27
[0381]	<211> 19
[0382]	<212> RNA
[0383]	<213> 人工序列
[0384]	<220>
[0385]	<223> 合成的
[0386]	<400> 27
[0387]	aaauucuacu uuuguagau 19
[0388]	<210> 28
[0389]	<211> 20

[0390] <212> RNA  
[0391] <213> 人工序列  
[0392] <220>  
[0393] <223> 合成的  
[0394] <400> 28  
[0395] aauuucuacu guuuguagau 20  
[0396] <210> 29  
[0397] <211> 20  
[0398] <212> RNA  
[0399] <213> 人工序列  
[0400] <220>  
[0401] <223> 合成的  
[0402] <400> 29  
[0403] aauuucuacu guuuguagau 20  
[0404] <210> 30  
[0405] <211> 20  
[0406] <212> RNA  
[0407] <213> 人工序列  
[0408] <220>  
[0409] <223> 合成的  
[0410] <400> 30  
[0411] aauuucuacu guuuguagau 20  
[0412] <210> 31  
[0413] <211> 19  
[0414] <212> RNA  
[0415] <213> 人工序列  
[0416] <220>  
[0417] <223> 合成的  
[0418] <400> 31  
[0419] aauuucuacu guuguagau 19  
[0420] <210> 32  
[0421] <211> 19  
[0422] <212> RNA  
[0423] <213> 人工序列  
[0424] <220>  
[0425] <223> 合成的  
[0426] <400> 32  
[0427] aauuucuacu auuguagau 19  
[0428] <210> 33



---

[0429]	<211> 19
[0430]	<212> RNA
[0431]	<213> 人工序列
[0432]	<220>
[0433]	<223> 合成的
[0434]	<400> 33
[0435]	aaauucuacu auuguagau 19
[0436]	<210> 34
[0437]	<211> 19
[0438]	<212> RNA
[0439]	<213> 人工序列
[0440]	<220>
[0441]	<223> 合成的
[0442]	<400> 34
[0443]	aaauucuacu uuuguagau 19
[0444]	<210> 35
[0445]	<211> 19
[0446]	<212> RNA
[0447]	<213> 人工序列
[0448]	<220>
[0449]	<223> 合成的
[0450]	<400> 35
[0451]	gaauucuacu uuuguagau 19
[0452]	<210> 36
[0453]	<211> 19
[0454]	<212> RNA
[0455]	<213> 人工序列
[0456]	<220>
[0457]	<223> 合成的
[0458]	<400> 36
[0459]	aaauucuacu aguguagau 19
[0460]	<210> 37
[0461]	<211> 19
[0462]	<212> RNA
[0463]	<213> 人工序列
[0464]	<220>
[0465]	<223> 合成的
[0466]	<400> 37
[0467]	aaauucuacu auuguagau 19

---

[0468]	<210> 38
[0469]	<211> 19
[0470]	<212> RNA
[0471]	<213> 人工序列
[0472]	<220>
[0473]	<223> 合成的
[0474]	<400> 38
[0475]	aaauucuacu guuguagau 19
[0476]	<210> 39
[0477]	<211> 19
[0478]	<212> RNA
[0479]	<213> 人工序列
[0480]	<220>
[0481]	<223> 合成的
[0482]	<400> 39
[0483]	aaauucuacu auuguagau 19

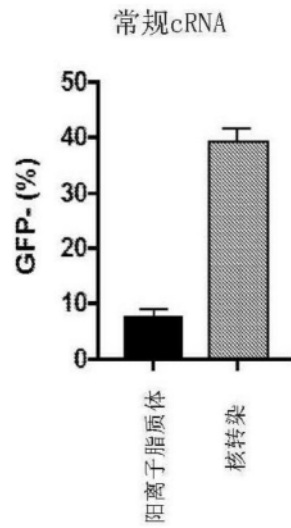


图1A

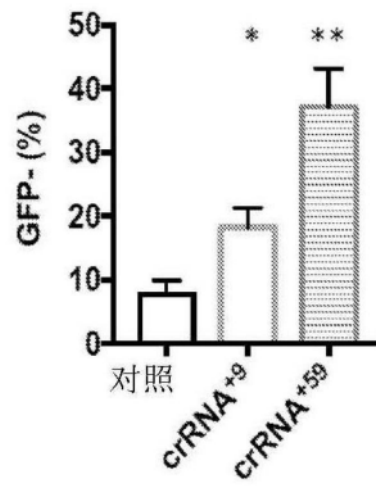


图1B

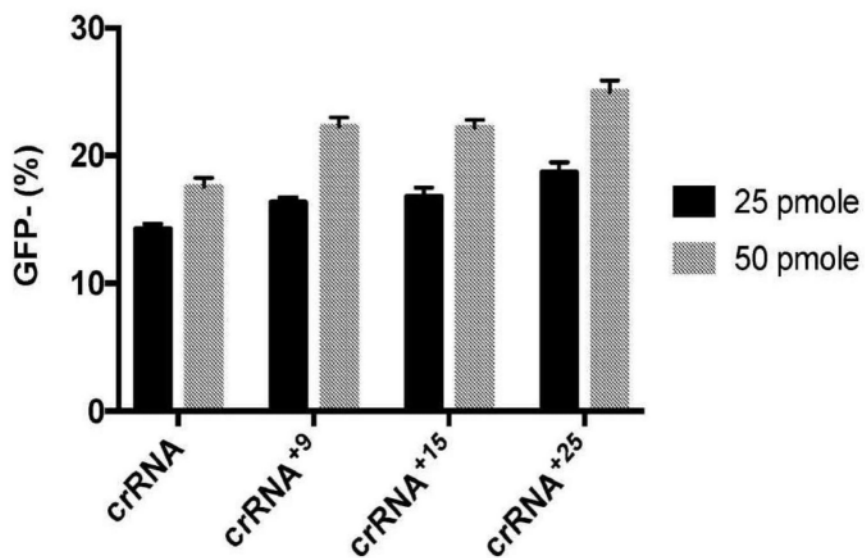


图1C

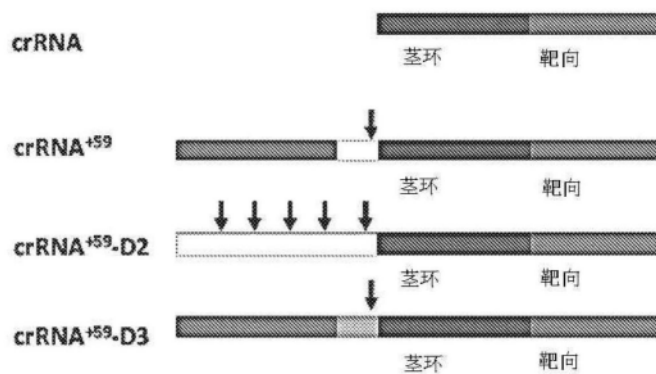


图1D

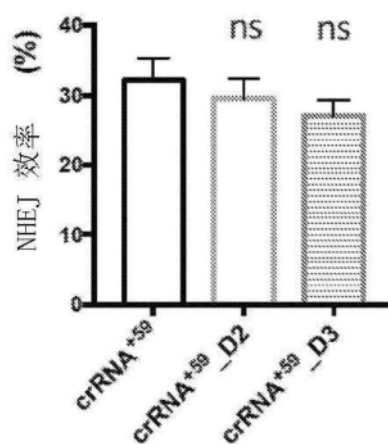


图1E

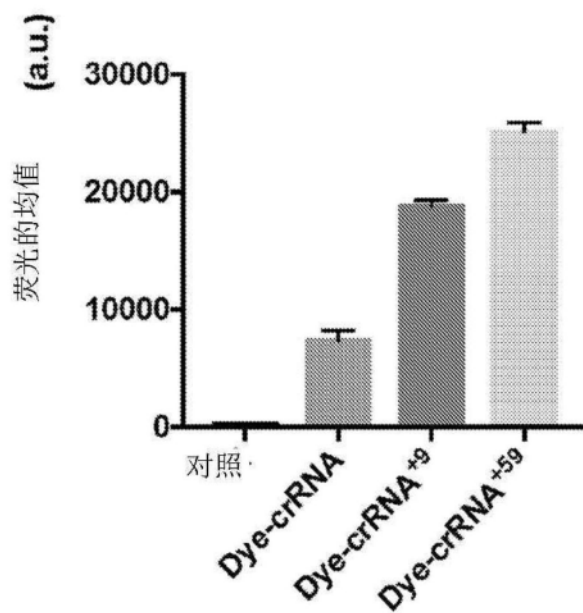


图1F

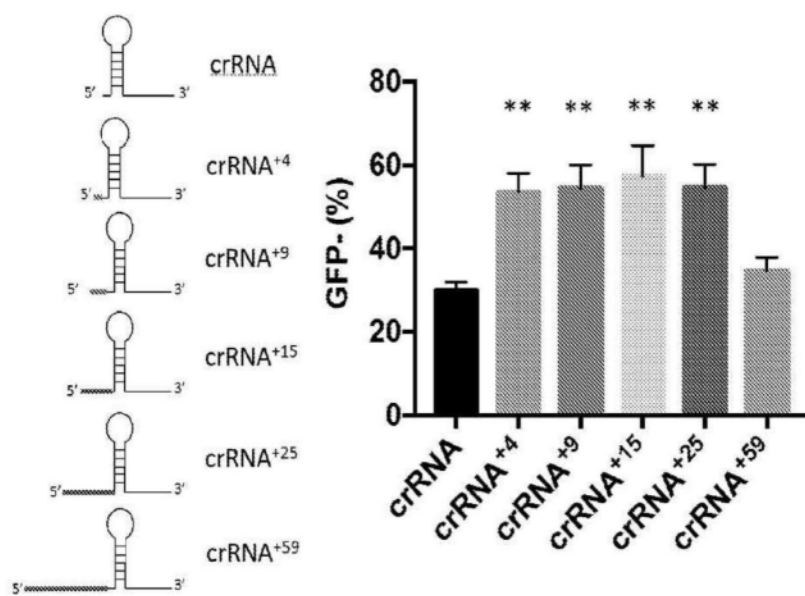


图2

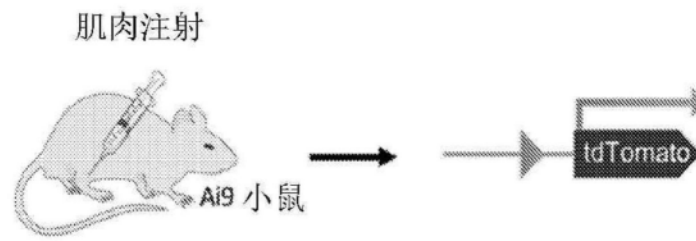


图3A

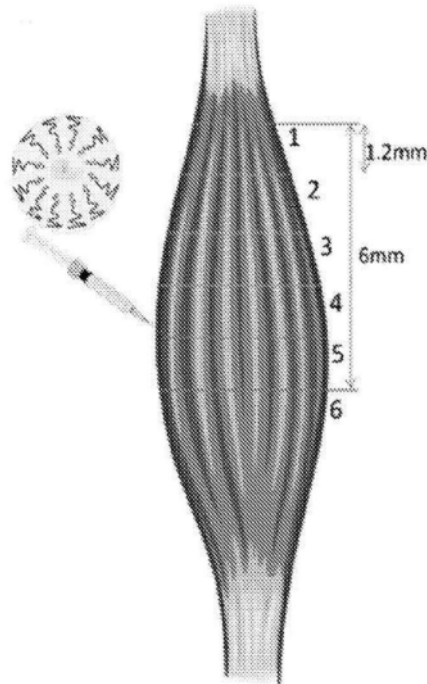


图3B

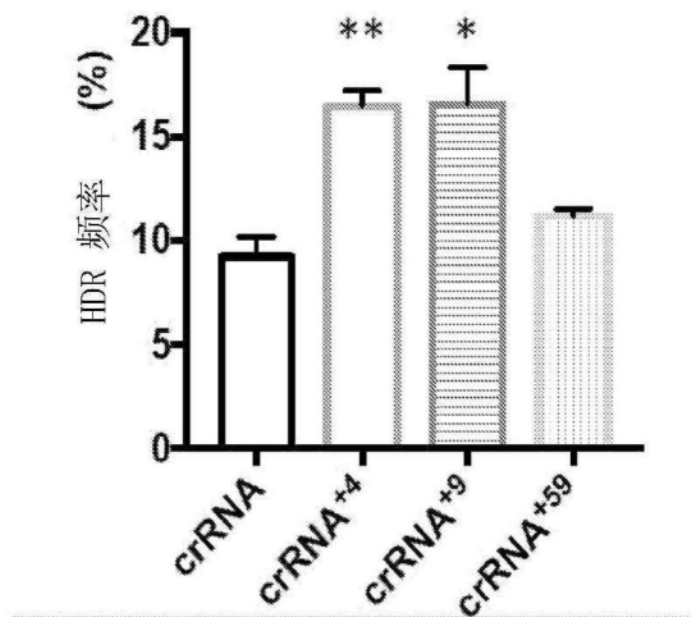


图4A

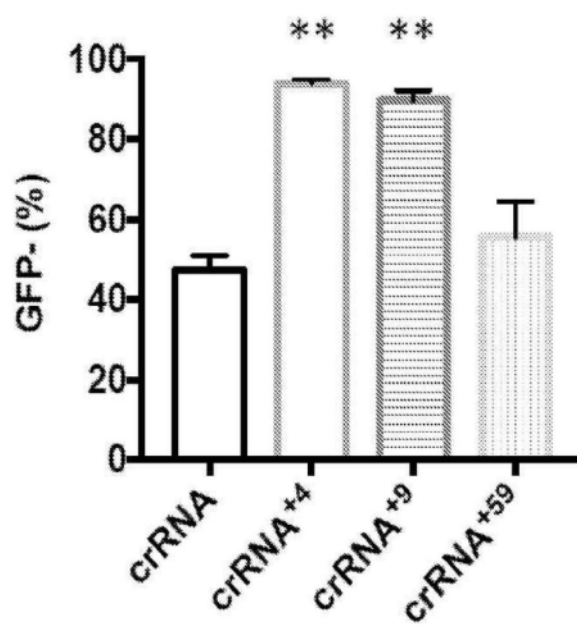


图4B

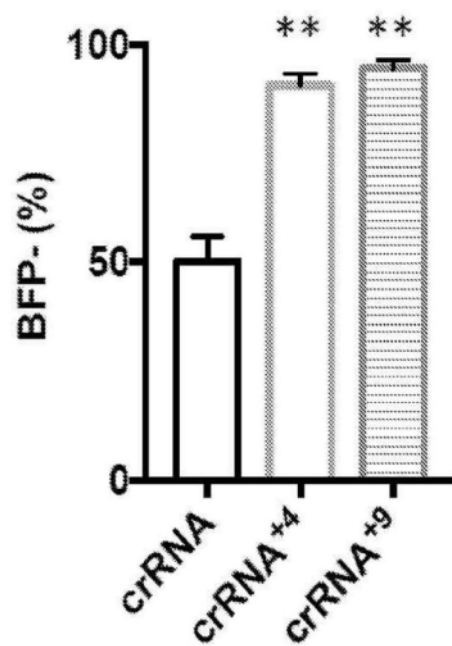


图4C

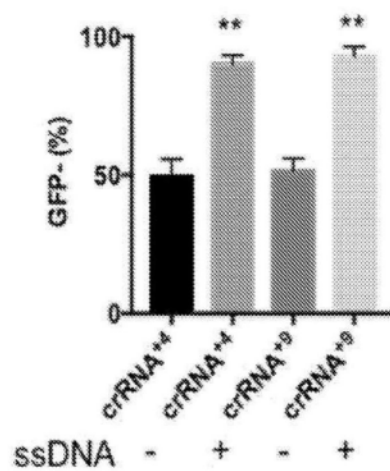


图4D



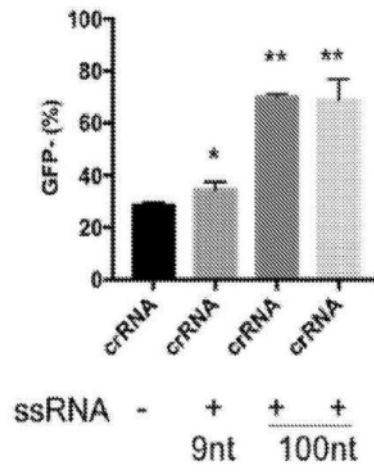


图4E

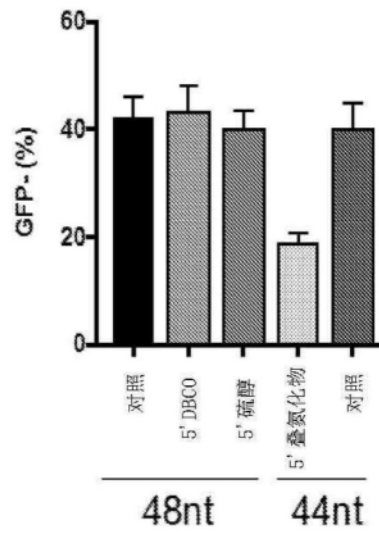


图4F

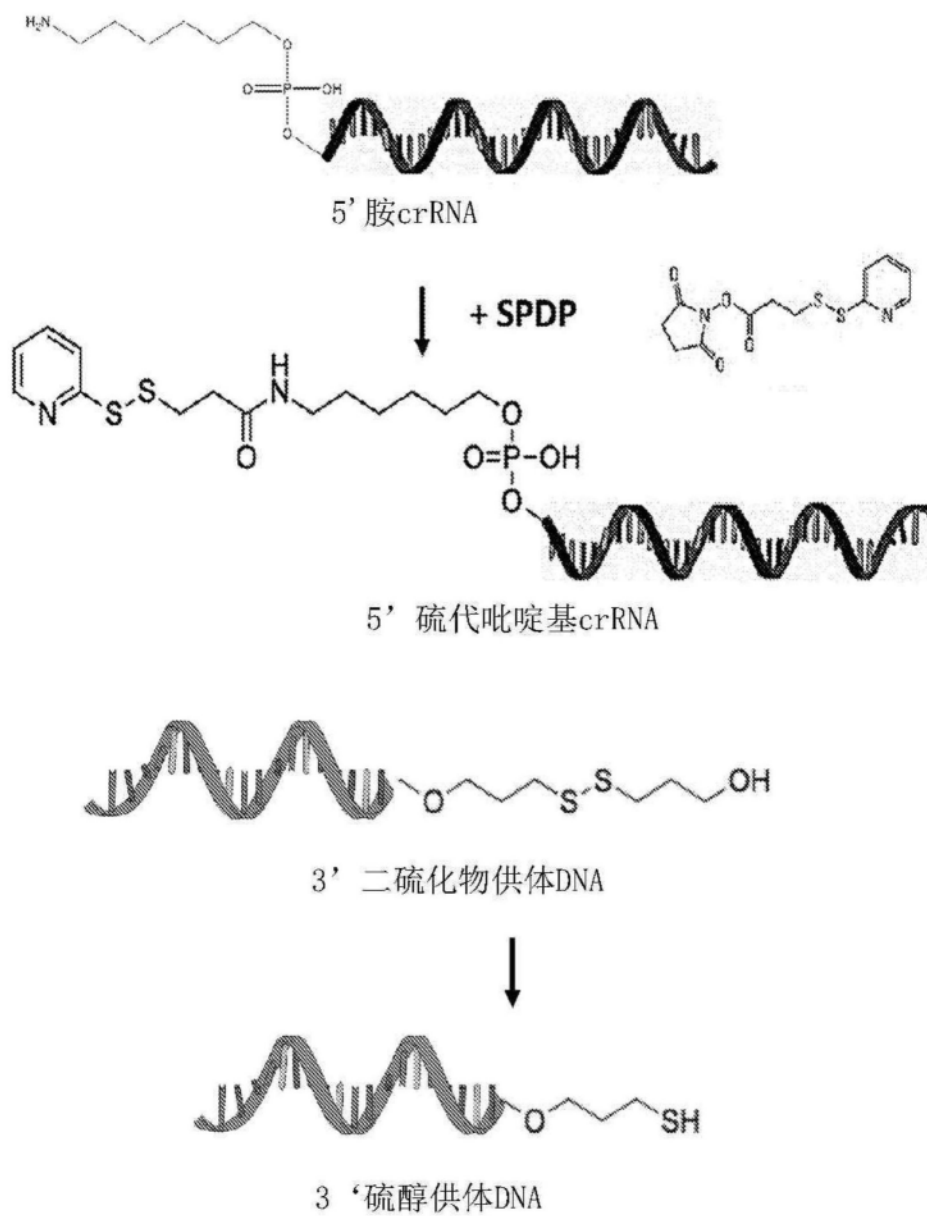


图5A

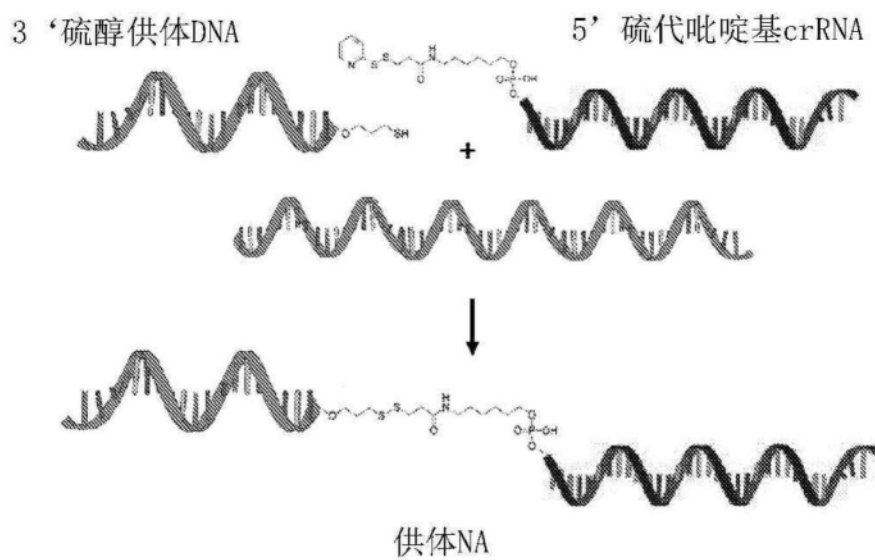


图5B

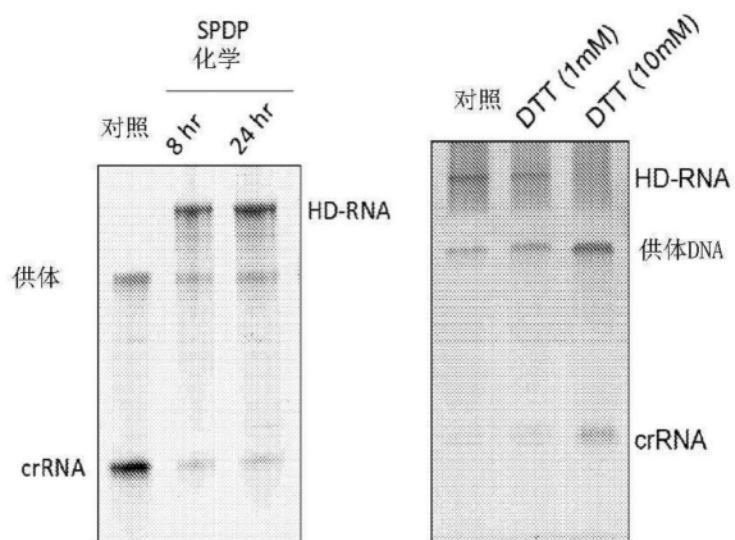


图5C

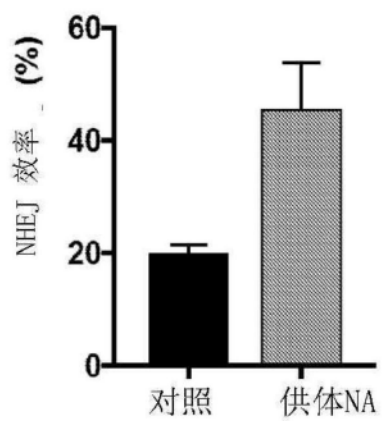


图6

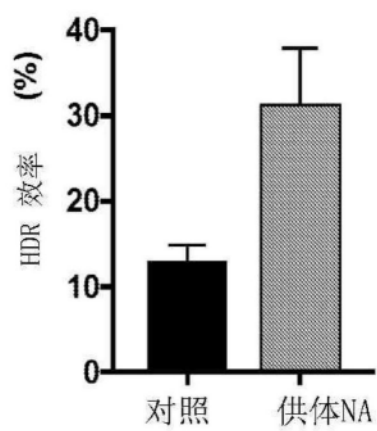


图7

## Francisella tularensis subsp. novicida U112 Cpfl

```

1  mslyqefvnk yslsktlrfe lipqgktlen ikarglildd ekrakdykka kqiidkyhqf
61  fieeilssvc isedllqnys dvyfklkksd ddnlqkdfks akdtikkqis eyikdsekfk
121 nlfngnlida kkggesdlil wlqskdngi elfkansdit didealeiik sfkgwttyfk
181 gfhenrknav ssndiptsii yrivddnlpk flenkakyes lkdkapeain yeqikkdlae
241 eltfdidykt sevnqrvfsl devfeianfn nylngsgitk fntiiggkfv ngentkrkqi
301 neyinlysqq indktlkkyk msvlfkqils dtesksfvid kledsdsvvt tmqsfyeqia
361 afktveeksi ketlsllfdd lkaqkldlsk iyfkndkslt dlsqqvfddy svigtavley
421 itqqiapknl dnpskkeqel iakktekaky lsletiklal eefnkhrdid kqcrfeeila
481 nfaaipmifd eiaqnkdnla qisikyqngg kkdllqasae ddvkaikdli dqtnnllhkl
541 kifhisqsed kanildkdeh fylvfeecyf elanivplyn kirnyitqkp ysdekfklnf
601 enstlangwd knkepdntai lfikddkyyl gvmnkknki fddkaikenk gegykkivyk
661 llpgankmlp kvffsaksik fynpsedilr irnhsthtkn gspqkgyekf efniedcrkf
721 idfyqsisk hpewkdfgr fsdtqrynsi defyrevenq gykltfenis esyidsvvnq
781 gklylfqiyn kdfsayskgr pnhltlywka lfdernlqdv vyklngeael fyrkqsipkk
841 ithpakeaia nknkdnpkke svfeydlikd krftedkfff hcpitinfks sgankfndei
901 nlllkekand vhilsidrg rhlayytlvd gkgniikqdt fniigndrmk tnyhdklaai
961 ekdrdsarkd wkkinnikem kegylsqvvh eiaklvieyn aivvfedlnf gfkrgfrkve
1021 kqvyqklekm lieklnylvf kdnefdktgg vlayqltap fetfkkmqkq tgiiyyvpag
1081 ftskicpvtg fvnqlypkye svksqeffs kfdkicynld kgyfefsfdy knfgdkaakg
1141 kwtiasfgsr linfrnsdkn hnwdtrevyp tkelekllkd ysieyghgec ikaaicgesd
1201 kkffakltsv lntilqmrns ktgteldyli spvadvnqnf fdsrqapknm pqdadangay
1261 higlkgmlml griknqegk klnlviknee yfefvqnrnn (SEQ ID NO: 1)

```

图8

类型/ 物种	Cpf1直系同源物的直接重复序列	
	加工序列	(5'-3') Cpf1 crRNA 序列
FnCpf1	GUCUAAGAACUUUAAAUA (SEQ ID NO: 2)	AAUUUCUACU-GUUGUAGAU (SEQ ID NO: 21)
AsCpf1	GUCAAAAGACCUUUUU (SEQ ID NO: 3)	AAUUUCUACUC-UUGUAGAU (SEQ ID NO: 22)
LbCpf1	GUUUCAAAGAUUAAAUA (SEQ ID NO: 4)	AAUUUCUACUAAGUGUAGAU (SEQ ID NO: 23)
TsCpf1	CUCUAGCAGGCCUGGCA (SEQ ID NO: 5)	AAUUUCUACU-GUUGUAGAU (SEQ ID NO: 24)
SaCpf1	AUUUGAAAGCAUCUUUU (SEQ ID NO: 6)	AAUUUCUACU-AUUGUAGAU (SEQ ID NO: 25)
Pb2Cpf1	GGCUAUAAGCUUAUUU (SEQ ID NO: 7)	AAUUUCUACU-AUUGUAGAU (SEQ ID NO: 26)
PgCpf1	GCCAAUACCUCUAUAAA (SEQ ID NO: 8)	AAUUUCUACU-UUUGUAGAU (SEQ ID NO: 27)
MlCpf1	GUCUAAACUACCUUUU (SEQ ID NO: 9)	AAUUUCUACUGUUUGUAGAU (SEQ ID NO: 28)
Mb2Cpf1	GUCUAAACUACCUUUU (SEQ ID NO: 10)	AAUUUCUACUGUUUGUAGAU (SEQ ID NO: 29)
Mb3Cpf1	GUCUAAACUACCUUUU (SEQ ID NO: 11)	AAUUUCUACUGUUUGUAGAU (SEQ ID NO: 30)
Lb4Cpf1	CUCUAAUAAGAGAUAGA (SEQ ID NO: 12)	AAUUUCUACU-GUUGUAGAU (SEQ ID NO: 31)
Lb5Cpf1	UGC UUAGAACAUUUAAA (SEQ ID NO: 13)	AAUUUCUACU-AUUGUAGAU (SEQ ID NO: 32)
FbCpf1	GUUUAAAACACUUUAAA (SEQ ID NO: 14)	AAUUUCUACU-AUUGUAGAU (SEQ ID NO: 33)
CRbCpf1	CUCUACAACUGAUAAA (SEQ ID NO: 15)	AAUUUCUACU-UUUGUAGAU (SEQ ID NO: 34)
CPbCpf1	GUUUAAAAGUCCUAUUG (SEQ ID NO: 16)	AAUUUCUACU-UUUGUAGAU (SEQ ID NO: 35)
CMaCpf1	GUCUAAAACUCAUUCAG (SEQ ID NO: 17)	AAUUUCUACU-AGUGUAGAU (SEQ ID NO: 36)
BsCpf1	UGC UUAGUACUUAUAAA (SEQ ID NO: 18)	AAUUUCUACU-AUUGUAGAU (SEQ ID NO: 37)
BfCpf1	GCCAAGAACCUAUAGAU (SEQ ID NO: 19)	AAUUUCUACU-GUUGUAGAU (SEQ ID NO: 38)
BoCpf1	GUCUAUAAGACAUUUUA (SEQ ID NO: 20)	AAUUUCUACU-AUUGUAGAU (SEQ ID NO: 39)

图9

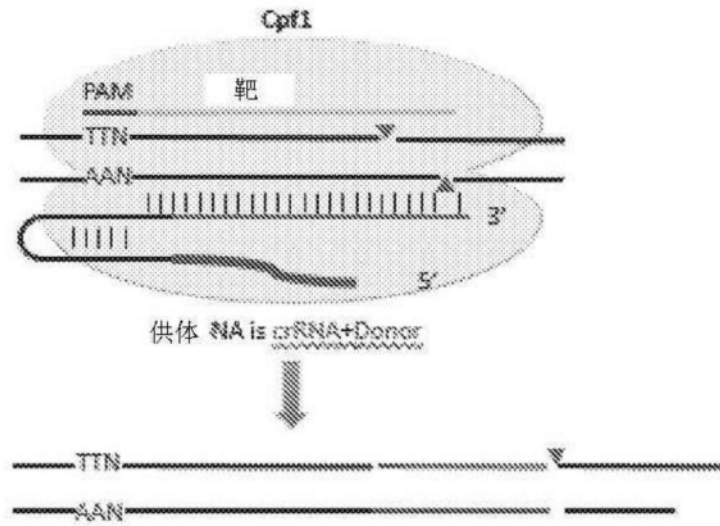


图10A

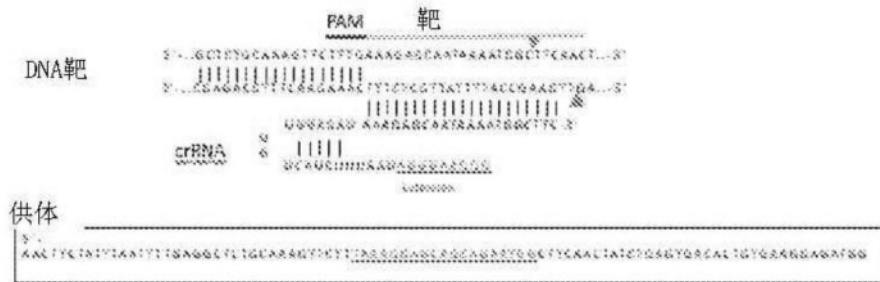


图10B

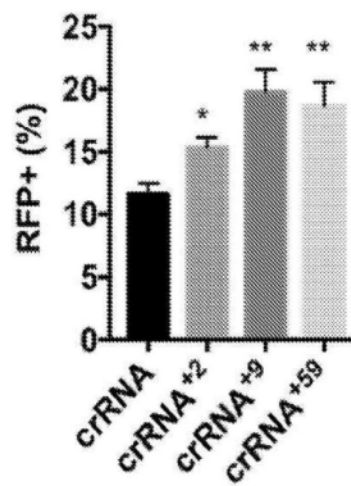


图10C

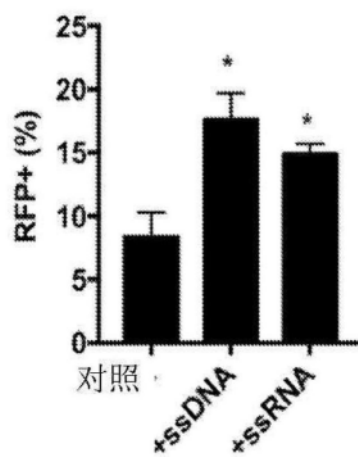


图10D

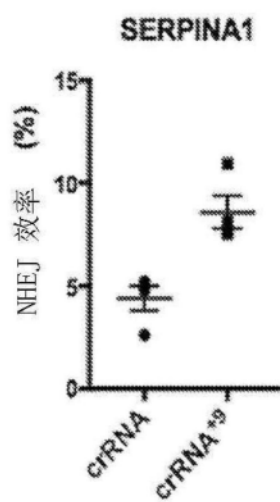


图10E



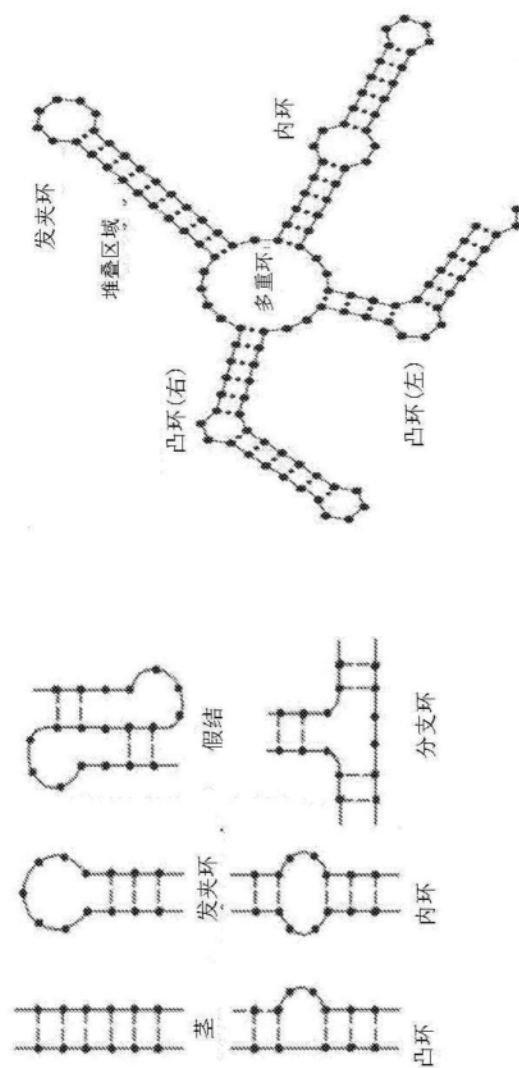


图11A

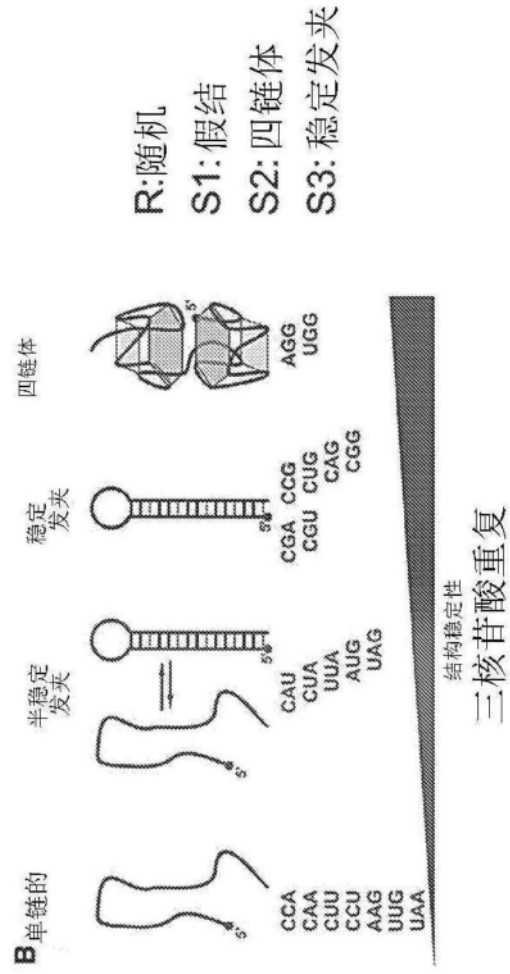


图11B

接吻环复合物

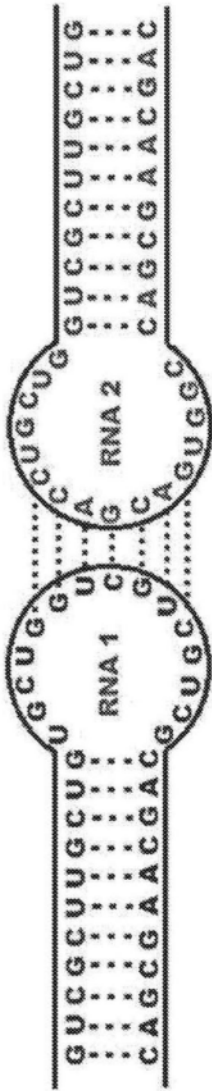


图11C

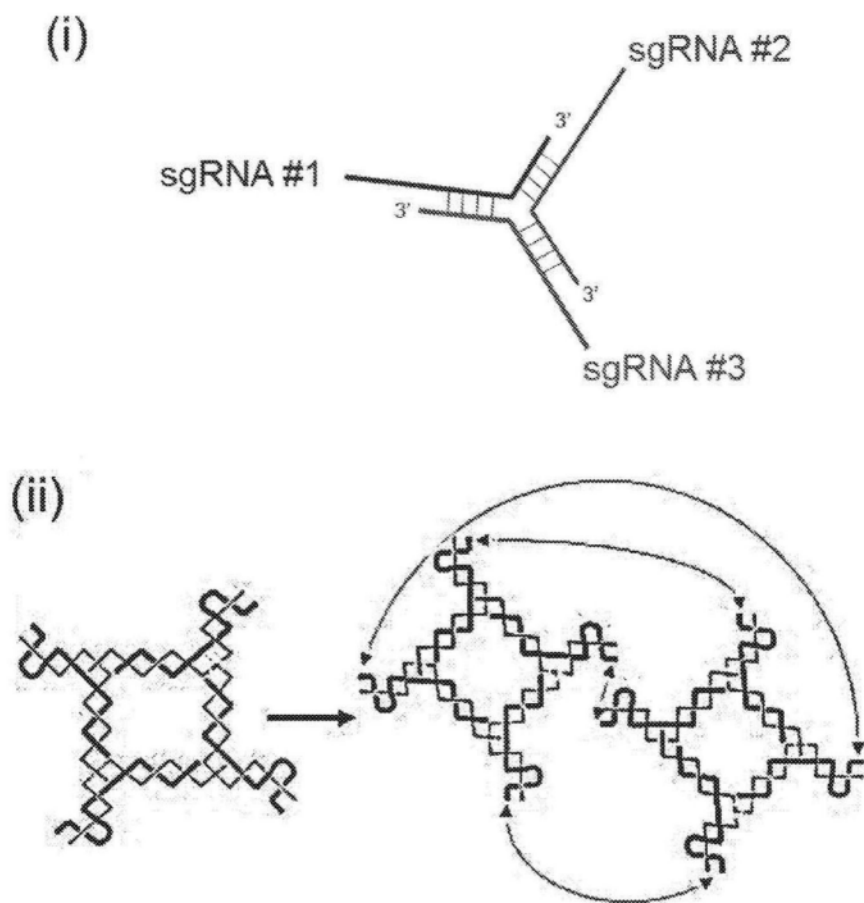


图11D

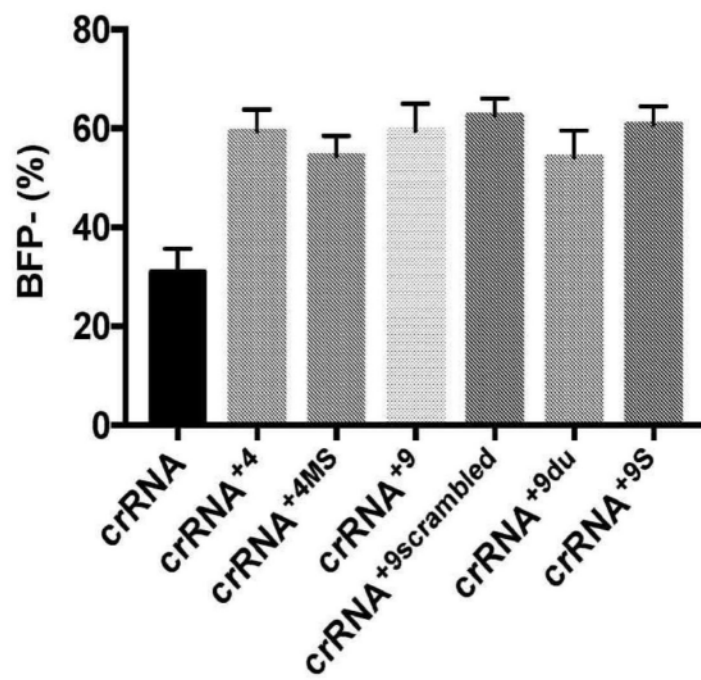


图12

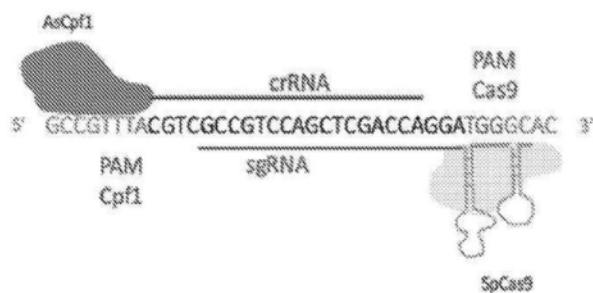


图13A

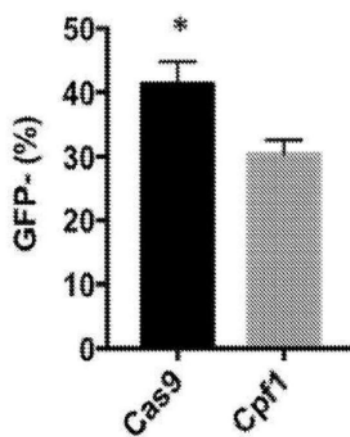


图13B

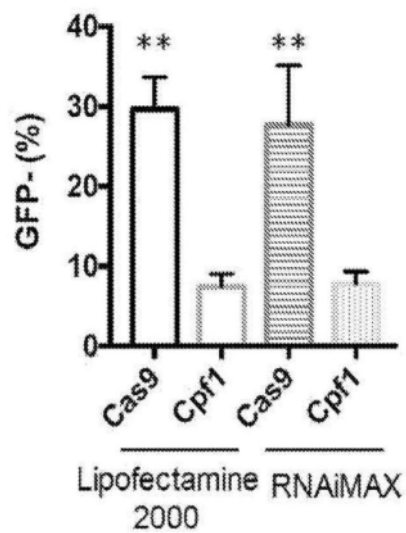


图13C

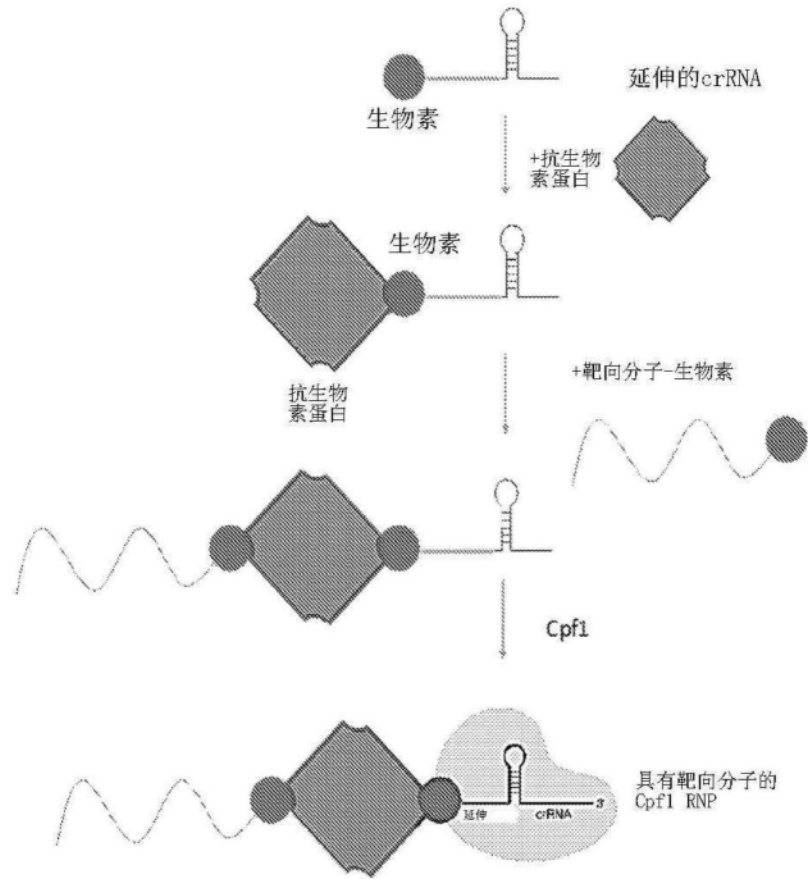


图14

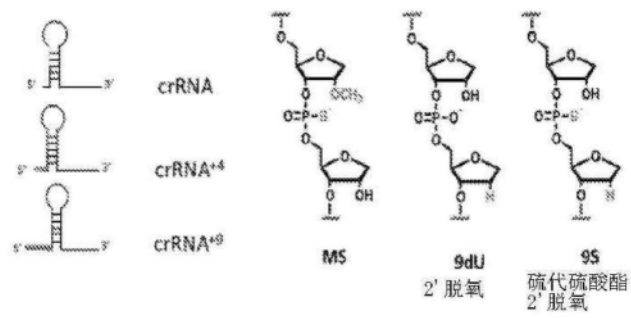


图15A

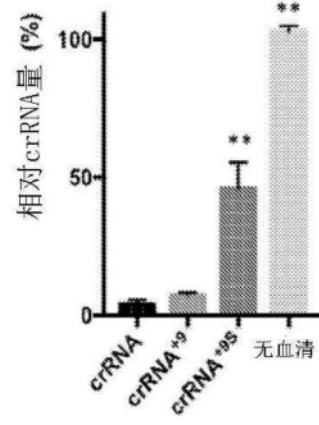


图15B

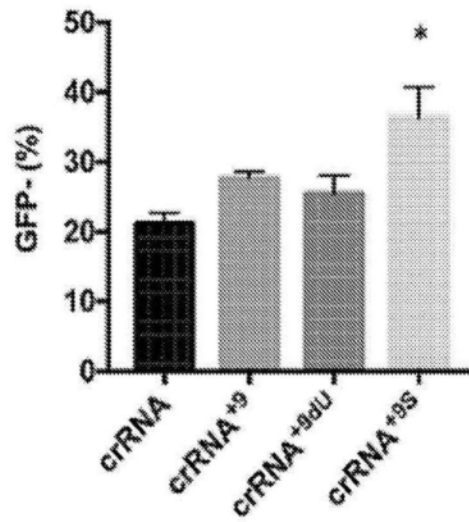


图15C

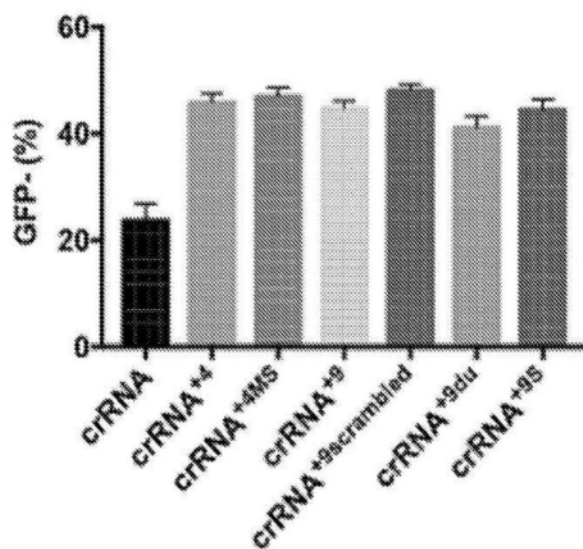


图15D

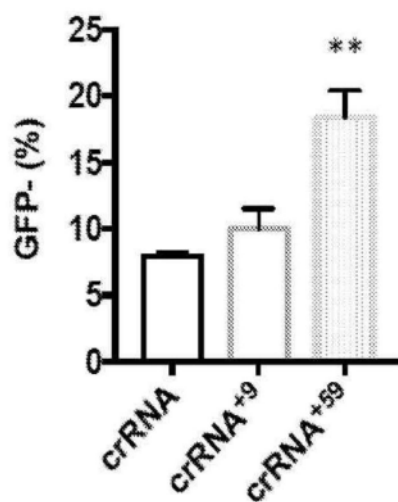


图16