



(12) 发明专利申请

(10) 申请公布号 CN 117242184 A

(43) 申请公布日 2023.12.15

(21) 申请号 202280019354.3

(22) 申请日 2022.01.05

(30) 优先权数据

63/133,945 2021.01.05 US

(85) PCT国际申请进入国家阶段日

2023.09.05

(86) PCT国际申请的申请数据

PCT/US2022/011294 2022.01.05

(87) PCT国际申请的公布数据

WO2022/150372 EN 2022.07.14

(71) 申请人 地平线探索有限公司

地址 英国剑桥

(72) 发明人 H·B·马查多 K·亨普希尔

R·布拉斯伯格 A·史密斯

M·D·拉什顿 P·佩雷斯-杜兰

(74) 专利代理机构 广州嘉权专利商标事务所有
限公司 44205

专利代理师 黄琳娟

(51) Int.Cl.

C12N 15/90 (2006.01)

权利要求书5页 说明书42页

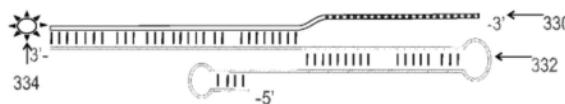
序列表43页 附图21页

(54) 发明名称

向导RNA设计及用于V型Cas系统的复合物

(57) 摘要

本发明提供了一种新型gRNA-配体结合复合物。此复合物可用于将V型Cas蛋白和额外的效应物带到DNA中进行碱基编辑。该系统的设计允许生产在编辑DNA时提供灵活性的高效模块化组分。



1. 一种gRNA-配体结合复合物,其中所述gRNA-配体结合复合物包含:
 - a. gRNA,其中所述gRNA含有60至210个核苷酸,且所述gRNA包含:
 - i. crRNA序列,其中所述RNA序列的长度为35至60个核苷酸,且所述crRNA序列包含Cas结合区和靶向区,其中所述Cas结合区的长度为18至30个核苷酸,其中所述靶向区的长度为18至30个核苷酸,以及
 - ii. tracrRNA序列,其中所述tracrRNA序列的长度为45至120个核苷酸,且其中所述tracrRNA序列包含反重复区和远端区,其中在所述Cas结合区的超过至少18个连续核苷酸上,所述反重复区与所述Cas结合区至少80%互补,且所述Cas结合区与所述反重复区能够杂交形成杂交区,其中所述杂交区能够保持与V型Cas蛋白的RNA结合结构域结合;以及
 - b. 配体结合基团,其中所述配体结合基团(i)直接结合所述gRNA,或者(ii)通过接头结合所述gRNA。
2. 根据权利要求1所述的gRNA-配体结合复合物,其中所述gRNA由核苷酸单链形成,且所述核苷酸单链包括所述crRNA序列和所述tracrRNA序列。
3. 根据权利要求2所述的gRNA-配体结合复合物,其中所述Cas结合区与所述反重复区之间存在环序列。
4. 根据权利要求3所述的gRNA-配体结合复合物,其中所述环序列的长度为4至40个核苷酸。
5. 根据权利要求1所述的gRNA-配体结合复合物,其中所述gRNA由第一核苷酸链和第二核苷酸链组成,其中所述第一核苷酸链包含所述tracrRNA序列,且所述第二核苷酸链包含所述crRNA序列。
6. 根据权利要求1至5中任一项所述的gRNA-配体结合复合物,其中所述gRNA和所述配体结合基团中的至少一个包含至少一个修饰。
7. 根据权利要求3所述的gRNA-配体结合复合物,其中所述环序列包含至少一个修饰。
8. 根据权利要求7所述的gRNA-配体结合复合物,其中所述至少一个修饰是1至30个2'修饰。
9. 根据权利要求8所述的gRNA-配体结合复合物,其中所述1至30个2'修饰是1至10个2'修饰。
10. 根据权利要求9所述的gRNA-配体结合复合物,其中所述1至10个2'修饰位于所述靶向区中。
11. 根据权利要求9所述的gRNA-配体结合复合物,其中所述1至10个2'修饰位于所述Cas结合区中。
12. 根据权利要求9所述的gRNA-配体结合复合物,其中所述1至10个2'修饰位于所述反重复区中。
13. 根据权利要求9所述的RNA-配体结合复合物,其中所述1至10个2'修饰位于所述远端区中。
14. 根据权利要求7所述的gRNA-配体结合复合物,其中所述至少一个修饰是至少一个硫代磷酸酯连接。
15. 根据权利要求6所述的gRNA-配体结合复合物,其中所述至少一个修饰是至少一个硫代磷酸酯连接。

16. 根据权利要求15所述的gRNA-配体结合复合物,其中所述硫代磷酸酯连接位于所述靶向区中。

17. 根据权利要求15所述的gRNA-配体结合复合物,其中所述硫代磷酸酯连接位于所述Cas结合区中。

18. 根据权利要求15所述的gRNA-配体结合复合物,其中所述硫代磷酸酯连接位于所述反重复区中。

19. 根据权利要求15所述的gRNA-配体结合复合物,其中所述硫代磷酸酯连接位于所述远端区中。

20. 根据权利要求1至19中任一项所述的gRNA-配体结合复合物,其中所述配体结合基团包括核苷酸序列,且所述配体结合基团的核苷酸序列包含至少一个2'修饰核苷酸。

21. 根据权利要求20所述的gRNA-配体结合复合物,其中在所述配体结合基团中的所述2'修饰核苷酸是2'-O-甲基修饰核苷酸、2'-氟修饰核苷酸或2-氨基嘌呤修饰核苷酸。

22. 根据权利要求1至21中任一项所述的gRNA-配体结合复合物,其中所述配体结合基团包含硫代磷酸酯连接。

23. 根据权利要求2所述的RNA-配体结合复合物,其中所述gRNA具有5'末端,且所述配体结合基团在gRNA的5'末端直接结合至所述gRNA。

24. 根据权利要求2所述的gRNA-配体结合复合物,其中所述gRNA具有3'末端,且所述配体结合基团在gRNA的3'末端直接结合至所述gRNA。

25. 根据权利要求2所述的gRNA-配体结合复合物,其中所述gRNA具有3'末端和5'末端,且所述配体结合基团直接结合所述gRNA的3'末端和5'末端之外的核苷酸。

26. 根据权利要求25所述的RNA-配体结合复合物,其中所述配体结合基团直接结合所述靶向区中的gRNA。

27. 根据权利要求25所述的gRNA-配体结合复合物,其中所述配体结合基团直接结合所述Cas结合区中的gRNA。

28. 根据权利要求25所述的gRNA-配体结合复合物,其中所述配体结合基团直接结合所述反重复区中的gRNA。

29. 根据权利要求25所述的gRNA-配体结合复合物,其中所述配体结合基团直接结合所述远端区中的gRNA。

30. 根据权利要求3所述的gRNA-配体结合复合物,其中所述配体结合基团直接结合所述环序列处的gRNA。

31. 根据权利要求1所述的gRNA-配体结合复合物,其中所述gRNA包含环序列,且所述配体结合基团直接结合所述环序列处的gRNA。

32. 根据权利要求1至22中任一项所述的gRNA-配体结合复合物,其中所述gRNA-配体结合复合物包含所述接头,且所述配体结合基团通过所述接头结合所述gRNA。

33. 根据权利要求2所述的gRNA-配体结合复合物,其中所述gRNA具有5'末端,且所述接头在所述gRNA的5'末端结合所述gRNA。

34. 根据权利要求32所述的gRNA-配体结合复合物,其中所述gRNA具有3'末端,且所述接头在所述gRNA的3'末端结合所述gRNA。

35. 根据权利要求32所述的gRNA-配体结合复合物,其中所述gRNA具有3'末端和5'末

端,且所述接头在所述3'末端或所述5'末端外的核苷酸处结合所述gRNA。

36. 根据权利要求35所述的gRNA-配体结合复合物,其中所述接头在所述靶向区中结合所述gRNA。

37. 根据权利要求35所述的gRNA-配体结合复合物,其中所述接头在所述Cas结合区中直接结合所述gRNA。

38. 根据权利要求35所述的RNA-配体结合复合物,其中所述接头在所述反重复区中直接结合所述gRNA。

39. 根据权利要求35所述的gRNA-配体结合复合物,其中所述接头在所述远端区中结合所述gRNA。

40. 根据权利要求2所述的gRNA-配体结合复合物,其中在所述Cas结合区与所述反重复区之间存在环序列,且所述gRNA-配体结合复合物包含所述接头,其中所述接头结合所述环序列。

41. 根据权利要求32至40中任一项所述的gRNA-配体结合复合物,其中所述接头选自由经修饰和无修饰的寡核苷酸、经修饰和无修饰的寡肽、无机基团、经修饰和无修饰的多糖、经修饰和无修饰的脂、及其组合所组成的组。

42. 根据权利要求41所述的gRNA-配体结合复合物,其中所述接头包含经修饰的寡核苷酸序列,其中所述经修饰的寡核苷酸序列包含2'修饰和硫代磷酸酯连接中的至少一个。

43. 根据权利要求32至40中任一项所述的gRNA-配体结合复合物,其中所述接头包含乙酰丙酸基团。

44. 根据权利要求32至40中任一项所述的gRNA-配体结合复合物,其中所述接头包含乙二醇基团。

45. 根据权利要求32至40中任一项所述的gRNA-配体结合复合物,其中所述接头选自由18S、9S或C3组成的组。

46. 根据权利要求32至40中任一项所述的gRNA-配体结合复合物,其中所述接头是长度为1至24个核苷酸的核苷酸序列。

47. 根据权利要求1至46中任一项所述的gRNA-配体结合复合物,其中所述配体结合基团包含生物素或链霉亲和素。

48. 根据权利要求1至46中任一项所述的gRNA-配体结合复合物,其中所述配体结合基团能够与选自以下项组成组中的配体结合:MS2、Ku、PP7、SfMu、Sm7、Tat、谷胱甘肽S-转移酶(GST)、CSY4、Q β 、COM、pumilio、抗-His TAG(6H7)、 λ N22+、SNAP-TAG、凝集素和PDGF β -链。

49. 根据权利要求48所述的gRNA-配体结合复合物,其中所述配体是MS2。

50. 根据权利要求1至49中任一项所述的gRNA-配体结合复合物,其中所述配体结合基团包含5'修饰核苷酸,其中所述5'修饰核苷酸包含2'修饰、5'PO₄基团或含氮碱基修饰中的至少一种。

51. 根据权利要求32至50中任一项所述的gRNA-配体结合复合物,其中所述配体结合基团是第一配体结合基团,且所述gRNA-配体结合复合物包含第二配体结合基团,且其中所述接头是第一接头,且所述gRNA-配体结合复合物包含第二接头,其中所述第一配体结合基团附接所述第一接头,且所述第二配体结合基团附接所述第二接头。

52. 根据权利要求51所述的gRNA-配体结合复合物,其中所述第一接头和所述第二接头

各自附接所述crRNA序列。

53. 根据权利要求51所述的RNA-配体结合复合物,其中所述第一接头和所述第二接头各自附接所述tracrRNA序列。

54. 根据权利要求51所述的gRNA-配体结合复合物,其中所述第一接头和所述第二接头中的一个接头附接所述crRNA序列,并且所述第一接头和所述第二接头中的另一个接头附接所述tracrRNA序列。

55. 根据权利要求3所述的gRNA-配体结合复合物,其中所述gRNA-配体结合复合物包含所述接头,且所述配体结合基团是第一配体结合基团,且所述gRNA-配体结合复合物包含第二配体结合基团,且其中所述接头是第一接头,且所述gRNA-配体结合复合物包含第二接头,其中所述第一配体结合基团附接所述第一接头,且所述第二配体结合基团附接所述第二接头,且所述第一接头附接所述环序列。

56. 根据权利要求55所述的gRNA-配体结合复合物,其中所述第二接头附接所述crRNA序列。

57. 根据权利要求56所述的gRNA-配体结合复合物,其中所述第二接头附接所述tracrRNA序列。

58. 根据权利要求1所述的gRNA-配体结合复合物,其中所述配体结合基团是MS2,且所述配体结合基团包含长度为1至12个核苷酸的上游序列及长度为1至12个核苷酸的下游序列,其中所述上游和下游序列紧邻所述MS2的两侧,且所述上游序列与所述下游序列互补。

59. 根据权利要求58所述的gRNA-配体结合复合物,其中每个所述上游序列和所述下游序列的长度均为2个核苷酸。

60. 根据权利要求59所述的gRNA-配体结合复合物,其中每个所述上游序列和所述下游序列均为GC。

61. 根据权利要求1所述的gRNA-配体结合复合物,其中所述gRNA-配体结合复合物包含或编码SEQ ID NO:28。

62. 根据权利要求1所述的gRNA-配体结合复合物,其中所述gRNA-配体结合复合物包含或编码SEQ ID NO:59至SEQ ID NO:65中的任一个。

63. 根据权利要求1所述的gRNA-配体结合复合物,其中所述gRNA-配体结合复合物包含或编码SEQ ID NO:67至SEQ ID NO:71中的任一个。

64. 一种碱基编辑复合物,包含:

a. 权利要求1至63中任一项所述的gRNA-配体结合复合物;以及

b. V型Cas蛋白,其中所述gRNA-配体结合复合物的所述杂交区与所述V型Cas蛋白结合。

65. 根据权利要求64所述的碱基编辑复合物,其中所述V型Cas蛋白是切割酶。

66. 根据权利要求64所述的碱基编辑复合物,其中所述V型Cas蛋白包含活性RuvC结构域。

67. 根据权利要求64所述的碱基编辑复合物,其中所述V型Cas蛋白包含失活的RuvC结构域。

68. 根据权利要求64至67中任一项所述的碱基编辑复合物,其中所述V型Cas蛋白选自由Cas12b、Cas12e和Cas12f组成的组。

69. 根据权利要求64至68中任一项所述的碱基编辑复合物,还包含效应物,其中所述效

应物附接所述配体结合基团配体,且所述配体结合基团配体能够与所述配体结合基团结合。

70.根据权利要求69所述的碱基编辑复合物,其中所述效应物选自自由脱氨酶、逆转录酶转录调节剂和修复酶组成的组。

71.根据权利要求69所述的碱基编辑复合物,其中所述效应物选自自由AID、CDA、APOBEC1、APOBEC3A、APOBEC3B、APOBEC3C、APOBEC3D、APOBEC3F、APOBEC3G、APOBEC3H、ADA、ADAR1、TadA、TADA、TAD3,ADAR2、ADAR3、Dnmt1、Dnmt3a、TET1、TET2和TDG组成的组。

72.根据权利要求69至71中任一项所述的碱基编辑复合物,其中所述配体是MCP。

73.根据权利要求64至72中任一项所述的碱基编辑复合物,还包含半胱氨酸/硒代半胱氨酸标签。

74.根据权利要求64至72中任一项所述的碱基编辑复合物,还包括用于通过点击化学进行环加成的元件。

75.根据权利要求64至72中任一项所述的碱基编辑复合物,其中所述效应物是第一效应物,且所述碱基编辑复合物还包含第二效应物。

76.根据权利要求75所述的碱基编辑复合物,其中所述第二效应物是以共价键或以非共价键方式直接或者通过中间基团结合所述V型Cas蛋白。

77.一种用于碱基编辑的方法,包括将权利要求64至76中任一项所述的碱基编辑复合物暴露于双链DNA。

78.根据权利要求77所述的方法,其中所述方法在体外进行。

79.根据权利要求77所述的方法,其中所述方法在体内进行。

80.根据权利要求77所述的方法,其中所述方法离体进行。

81.一种用于碱基编辑的方法,包括将权利要求65至77中任一项所述的碱基编辑复合物暴露于单链DNA。

82.根据权利要求82所述的方法,其中所述方法在体外进行。

83.根据权利要求82所述的方法,其中所述方法在体内进行。

84.根据权利要求82所述的方法,其中所述方法离体进行。

85.一种在细胞中编辑DNA的方法,所述方法包括将权利要求65至77中任一项所述的碱基编辑复合物暴露于所述细胞。

86.根据权利要求86所述的方法,其中所述细胞是免疫细胞。

87.根据权利要求86所述的方法,其中所述免疫细胞是T细胞。

88.根据权利要求86所述的方法,其中所述免疫细胞是IPSC细胞。

89.一种治疗受试者的方法,所述方法包括权利要求77所述的方法,其中所述暴露发生在受试者体外,并且在所述暴露之后,将所述细胞输注到所述受试者中。

向导RNA设计及用于V型Cas系统的复合物

[0001] 相关申请的交叉引用

[0002] 本申请要求于2021年1月5日提交的美国临时专利申请序列号63/133,945的提交日的权益,如同在此充分陈述一样,该临时专利申请的全部公开内容以引用的方式并入本文中。

技术领域

[0003] 本发明涉及基因编辑领域。

背景技术

[0004] 研究人员正在积极地探索成簇规律间隔短回文重复(CRISPR)系统的使用,以便修饰DNA。到目前为止,在此领域中的绝大部分工作已放在Cas9系统中。在这些系统中,核苷酸的两个单独链或单链核苷酸的区域形式的tracrRNA(反式激活CRISPR RNA)和crRNA(CRISPR RNA)杂交以招募Cas9蛋白,然后将Cas9蛋白引导至与crRNA内部的序列互补的DNA位置。因此,DNA内部的互补序列变成靶位点,并且Cas9蛋白基于其功能结构域可在此靶位点处引发编辑。

[0005] 尽管现在充分认可了Cas9系统的能力,但这些系统并非在所有用途中都是有效的。Cas9系统的局限之一是Cas9系统可对其起作用的功能结构域由所使用的Cas9蛋白的功能结构域所限定。

[0006] 其他的Cas蛋白是已知的。在这些其他的Cas蛋白中,尚未对它们的潜力进行充分发掘,它们是在V型家族中。当人们试图在靶位点或其附近引入多个编辑时,来自V型家族的酶的使用尤其未得到充分发掘。因此,有需求开发改进的gRNA(向导RNA)和复合物以及包含并使用它们的系统。

发明内容

[0007] 本发明提供新型且非显而易见的gRNA-配体结合复合物、碱基编辑复合物,以及用于碱基编辑和基因组修饰的方法。通过使用本发明各种实施方式,技术人员能够高效地且有效地离体、在体外、和在体内引发碱基编辑。此外,本发明的一些实施方式提供了模块化设计,这些设计允许将相同的V型Cas蛋白引导至不同的靶向位点,并且可选地在相同或不同位点与不同的效应蛋白结合。

[0008] 根据第一实施方式,本发明提供一种gRNA-配体结合复合物,其中,该gRNA-配体结合复合物包含:(a)gRNA,其中,该gRNA含有60至210个核苷酸或80至180个核苷酸,并且该gRNA包含(i)crRNA序列,其中,该crRNA序列的长度为36至60个核苷酸或35至60个核苷酸,并且该crRNA序列包含Cas结合区(其中,该Cas结合区的长度为18至30个核苷酸)和靶向区(其中,该靶向区的长度为18至30个核苷酸),以及(ii)tracrRNA序列,其中,该tracrRNA序列的长度为45至120核苷酸,并且其中,该tracrRNA序列包含反重复区和远端区,其中,在Cas结合区的超过至少18个连续核苷酸上,反重复区与Cas结合区至少80%互补,并且Cas结

合区和反重复区能够杂交形成杂交区,其中,该杂交区能够保持与V型Cas蛋白的RNA结合结构域的结合;以及(b)配体结合基团,其中,配体结合基团(i)直接结合gRNA,或者(ii)通过接头结合gRNA。

[0009] 根据第二实施方式,本发明提供一种碱基编辑复合物,包含:本发明的gRNA-配体结合复合物和V型Cas蛋白,其中,gRNA-配体结合复合物的Cas结合区和反重复区与V型Cas蛋白结合。可选地,配体结合基团与配体可逆地结合,该配体附接效应分子到或者是效应分子的一部分。

[0010] 根据第三实施方式,本发明提供一种用于碱基编辑的方法。该方法包括将本发明的碱基编辑复合物暴露于双链DNA(“dsDNA”)或单链DNA(“ssDNA”)。该碱基编辑复合物可在允许碱基编辑的条件下暴露于dsDNA或ssDNA。

[0011] 根据第四实施方式,该gRNA-配体结合复合物包含或编码SEQ ID NO:28。

[0012] 根据第五实施方式,该gRNA-配体结合复合物包含或编码SEQ ID NO:59至SEQ ID NO:65中的任一个。

[0013] 根据第六实施方式,该gRNA-配体结合复合物包含或编码SEQ ID NO:67至SEQ ID NO:71中的任一个。

[0014] 当效应物附接到(或含有)配体时,该系统具有模块化设计。该配体结合基团存在于gRNA-配体结合复合物内部,从而允许该复合物与相应的结合有(或包含于其内部)效应物的配体结合。因此,配体结合基团以允许其能够与配体结合的方式和方位与gRNA结合。类似地,该配体以使得它能够与配体结合基团可逆地结合的方式附接或者与效应物结合。

[0015] 当配体与配体结合基团彼此结合时,与配体结合的效应物将变成含有gRNA-配体结合复合物的任何碱基编辑复合物的一部分。当碱基编辑复合物还含有Cas蛋白时,该Cas蛋白和效应物可以被保留在相同的位置,例如,在目的靶位点或附近。

[0016] 因此,如果技术人员希望使用带Cas蛋白的特定效应物,其只需使该效应物与能够与配体结合基团可逆地结合的配体结合,该配体结合基团是含有该Cas蛋白的碱基编辑复合物的一部分。为了将效应物从一个系统改换至下一个系统,技术人员只需改变效应物-配体。因而,技术人员可以使用相同的gRNA-配体结合复合物以及相关的带有多个不同效应物的Cas蛋白。通过结合和分离它们的配体与配体结合基团,在相同的系统中可以相继地使用多个不同的效应物,或者在不同的系统中同时或相继地使用。

附图说明

[0017] 图1A-图1D是用一个或两个配体结合基团修饰的Cas12b gRNA的图示,其中,tracrRNA和crRNA是分离的寡核苷酸链。

[0018] 图2A-图2D是用一个或两个配体结合基团修饰的Cas12b gRNA的图示,其中,tracrRNA和crRNA是同一寡核苷酸链的一部分。

[0019] 图3A-图3D是在不同位置用配体结合基团修饰的Cas12e gRNA的图示,其中,tracrRNA和crRNA是分离的寡核苷酸链。

[0020] 图4A-图4C是用配体结合基团修饰的Cas12e gRNA的图示,其中,tracrRNA和crRNA是分离的寡核苷酸链。

[0021] 图5A-图5D是在不同位置用配体结合基团修饰的Cas12e gRNA的图示,其中,

tracrRNA和crRNA是同一寡核苷酸链的一部分。

[0022] 图6A-图6C是在不同位置用配体结合基团修饰的Cas12fgRNA的图示,其中, tracrRNA和crRNA是分离的寡核苷酸链。

[0023] 图7A-图7C是用一个或两个配体结合基团修饰的Cas12fgRNA的图示,其中, tracrRNA和crRNA是同一寡核苷酸链的一部分。

[0024] 图8A和图8B是显示C向T转换的百分率的柱状图,这些转换是在HEK-293T细胞系中由Cas12b碱基编辑器与不同效应物所引入的。

[0025] 图9A-图9G是显示C向T转换的百分率的柱状图,这些转换是在HEK-293T细胞系中由Cas12b碱基编辑器在多个位点中所引入的。

[0026] 图10是显示C向T转换的百分率的柱状图,这些转换是在U2OS细胞系中由Cas12b碱基编辑器所引入的。

[0027] 图11A-图11D是显示C向T转换的百分率的柱状图,这些转换是在HEK-293T细胞系中由CasMINI碱基编辑器所引入的。

具体实施方式

[0028] 现在将详细地参考本发明的各种实施方式,其示例如所附的附图所示。在以下描述中,为了提供对本发明的全面理解,列出了许多具体细节。然而,除非另有说明或者从上下文中暗示,否则这些细节旨在是示例性的,而不应被视为以任何方式限制本发明的范围。此外,结合各种或具体实施方式所描述的特征不应被理解未不适用于结合发明中所公开的其他实施方式的使用,除非这种排他性被明确地说明或者从上下文中暗示。

[0029] 本发明中所提供的标题是为了方便读者,并不限制本发明所公开任何实施方式的范围。

[0030] 定义

[0031] 除非另有说明或者从上下文中是显而易见的,以下的术语应具有以下所陈述的含义:

[0032] 短语“2'修饰”是指具有在糖基的2'位置被修饰的糖基的核苷酸单元。2'修饰的一个示例是形成2'-O-烷基修饰核苷酸的2'-O-烷基修饰、或者形成2'卤素修饰核苷酸的2'卤素修饰。

[0033] 短语“2'-O-烷基修饰核苷酸”是指具有糖基的核苷酸单元,例如,脱氧核糖基或核糖基基团,其在2'位置被修饰使得氧原子附接到位于糖的2'位置的碳原子和烷基。在各种实施方式中,烷基基团是由或基本上由碳和氢组成。当O基团和其所附接的烷基被看作是一个基团时,它们可被称为O-烷基基团(例如,-O-甲基、-O-乙基、-O-丙基、-O-异丙基、-O-丁基、-O-异丁基、-O-乙基-O-甲基(-OCH₂CH₂OCH₃)和-O-乙基-OH(-OCH₂CH₂OH)。2'-O-烷基修饰核苷酸可以是经取代的或未经取代的。

[0034] 短语“2'卤素修饰核苷酸”是指具有糖基(例如在2'位置被修饰使得在该位置的碳直接地附接到卤素物质(例如,F1、Cl或Br)的脱氧核糖基团)的核苷酸单元。

[0035] “配体结合基团”是指如适配体的基团,例如,寡核苷酸或肽或者结合到特定配体且可以可逆地或不可逆地与该配体结合的另一种化合物。

[0036] 术语“修饰的核苷酸”是指在碱基、糖和/或磷酸盐的化学结构中具有至少一个修

饰的核苷酸,该修饰包括但不限于:5-位的嘧啶修饰、8-位的嘌呤修饰、在胞嘧啶环外胺处的修饰、和5-溴-尿嘧啶或5-碘尿嘧啶的取代;及2'-修饰,包括但不限于糖修饰核糖核苷酸,其中2'-OH被一个基团(如H、OR、R、卤代、SH、SR、NH₂、NHR、NR₂或CN)所替换。

[0037] 经修饰碱基是指已通过替换或添加一个或多个原子或基团而被修饰的核苷酸碱基,例如,举例来说,腺嘌呤、鸟嘌呤、胞嘧啶、胸腺嘧啶、尿嘧啶、黄嘌呤、肌苷和腺苷。这些类型的修饰的一些示例包括但不限于:烷基化、卤代、硫代、氨基化、酰胺化、或乙酰基化碱基,单独地或以各种组合地。更具体的经修饰碱基包括,例如,5-丙炔基尿苷、5-丙炔基胞苷、6-甲基腺嘌呤、6-甲基鸟嘌呤、N,N-二甲基腺嘌呤、2-丙基腺嘌呤、2-丙基鸟嘌呤、2-氨基腺嘌呤、1-甲基肌苷、3-甲基尿苷、5-甲基胞苷、5-甲基尿苷及具有在5位置的修饰的其他核苷酸、5-(2-氨基)丙基尿苷、5-卤代胞苷、5-卤代尿苷、4-乙酰基胞苷、1-甲基腺苷、2-甲基腺苷、3-甲基胞苷、6-甲基尿苷、2-甲基鸟苷、7-甲基鸟苷、2,2-二甲基鸟苷、5-甲基氨基乙基尿苷、5-甲氧基尿苷,脱氮核苷酸,如7-脱氮-腺苷、6-偶氮基尿苷、6-偶氮基胞苷、6-偶氮基胸苷、5-甲基-2-硫代尿苷、其他硫代碱基,如2-硫代尿苷和4-硫代尿苷和2-硫代胞苷、二氢尿苷、假尿苷、腺苷、古嘌呤、萘基和经取代萘基基团、任意O-和N-烷基化嘌呤和嘧啶,如N6-甲基腺苷、5-甲基羰基甲基尿苷、尿苷5-氧乙酸、吡啶-4-酮、吡啶-2-酮、苯基和经修饰的苯基,如氨基酚或2,4,6-三甲氧基苯、充当G-钳位核苷酸的经修饰的胞嘧啶、8-取代腺嘌呤和鸟嘌呤、5-取代的尿嘧啶和胸腺嘧啶、氮杂嘧啶、羧基羟基烷基核苷酸、羧基烷基氨基烷基核苷酸、及烷基羰基烷基化核苷酸。经修饰的核苷酸还包括那些对于糖基被修饰的经修饰的核苷酸、以及具有糖类或其类似物(不是核糖基)的核苷酸。例如,该糖基可以是或者可基于甘露糖类、阿拉伯糖类、吡喃葡萄糖类、吡喃半乳糖类、4-硫代核糖、和其他糖类、杂环类、或碳环类。

[0038] 短语“用于编码”和术语“编码”表示含有与参考核苷酸序列、等同于参考序列的DNA或RNA相同的序列的序列,或含有与参考序列的DNA或RNA补体的DNA或RNA或序列的序列。因此,当其是指用于编码或编码所列举的DNA序列的序列时,除非另有指定,其是指以下中的任一个序列:相同的DNA序列、该DNA序列的补体、该序列的RNA等效物、或者该序列的RNA补体或者前述中的任一个中,一个或多个核糖核苷酸被其脱氧核糖核苷酸对应物所取代,或者一个或多个脱氧核苷酸被其核糖核苷酸对应物所取代。

[0039] 术语“互补”是指核酸通过传统的Watson-Crick碱基配对或其他非传统类型的碱基对而与另一个核酸序列形成一个或多个氢键的能力。互补百分比表示可以与第二核酸序列形成氢键(例如,Watson-Crick碱基配对)的残基在核酸分子中的百分率(例如,10个中的5、6、7、8、9、10分别为50%、60%、70%、80%、90%和100%互补)。“完全地互补”表示核酸序列的所有相邻残基将与第二核酸序列中相同数量的相邻残基形成氢键。如本发明中使用的“基本互补”是指在一个区域(例如8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、30、35、40、45、50、或更多连续核苷酸)上为至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少97%、至少98%、或至少99%的互补程度,或者是指在严格条件下杂交的两个核酸。

[0040] 术语“杂交”和“产生杂交”是指其中完全地、基本上、或部分地互补的核酸链在特定的杂交条件下聚集到一起以形成双链结构或区域的过程,其中,两条组成链是通过氢键连接。除非另外指出,杂交条件是天然存在的或实验室设计的条件。尽管通常是在腺嘌呤与

胸腺嘧啶或尿嘧啶 (A和T或U) 或者在胞苷与鸟嘌呤之间 (C和G) 形成氢键, 但其他碱基对也可能形成 (参见, 例如, Adams等人, 《核酸的生物化学》第11版, 1992年)。

[0041] 术语“核苷酸”是指核糖核苷酸或脱氧核糖核苷酸或者其修饰形式, 以及其类似物。核苷酸包含以下物质, 其包括嘌呤 (例如, 腺嘌呤、次黄嘌呤、鸟嘌呤、及它们的衍生物和类似物) 以及嘧啶 (例如, 胞嘧啶、尿嘧啶、胸腺嘧啶、及它们的衍生物和类似物)。优选地, 核苷酸包括胞嘧啶、尿嘧啶、胸腺嘧啶、腺嘌呤、或鸟嘌呤基团。此外, 术语“核苷酸”也包括以下物质, 具有可检测的标记, 例如, 举例来说, 放射性或荧光基团、或者附接到该核苷酸的质量标记。术语“核苷酸”也包括在本领域已知的作为通用碱基的核苷酸。例如, 通用碱基包括但不限于3-硝基吡咯、5-硝基吡啶、或水粉蕈素。核苷酸类似物, 例如, 是指包括具有碱基 (如肌苷、腺苷、黄嘌呤)、糖类 (如2'-甲基核糖) 和非天然的磷酸二酯核苷酸间连接 (如甲基磷酸酯、硫代磷酸酯、磷酸乙酸酯) 及肽的核苷酸。

[0042] 术语“受试者”与“患者”在本发明中可互换地使用以指代生物体。例如, 脊椎动物, 优选哺乳动物、更优选人。哺乳动物包括但不限于: 鼠科动物、猿猴、人类、农场动物、运动动物、及宠物 (如犬类和猫类)。在体内获得或者在体外培养的生物体或其他生物实体的组织、细胞和它们的后代也被包含于术语受试者和患者内。此外, 在一些实施方式中, 受试者可以是无脊椎动物, 例如, 昆虫或线虫; 而在其他实施方式中, 受试者可以是植物或真菌。

[0043] 如本发明中所使用的“治疗 (treatment)”、“治疗 (treating)”、“缓和”和“减轻”可互换地使用。这些术语是指用于获得有利或期望结果 (包括但不限于治疗获益和/或预防获益) 的方法。治疗获益表示在治疗下的一个或多个疾病、状况、或症状中的任何治疗学相关的改善或效果。对于预防获益, 可将本发明的复合物施用给受试者、或者受试者的细胞或组织、或者在再次施用之前在体外施用给处于发展为特定疾病、状况、或症状的风险中的另一个受试者, 或者施用给报告有疾病的一个或多个生理学症状的受试者, 即使该疾病、状况、或症状可能尚未得到显现。

[0044] 提供了如本发明中所公开的若干范围的值。可以理解的是, 除非上下文另有明确说明, 在该范围的上限与下限之间每个居中值 (到下限的单位的十分之一) 也被具体地公开。在规定范围内的任何规定值或中间值与该规定范围内的任何其他规定或中间值之间的每个更小范围都包含在本发明中。这些更小范围的上限和下限可独立地被包括或排除在该范围内, 并且更小范围中包含的两者之一、两者均不、或者两者的限值的各范围也被包含于本发明中, 受到该指定范围中的任何特定排除的限值。在该指定范围包括限值中的一个或两个的情况下, 排除这些所包括的限值中的两者之一或两个的范围也被包含于本发明中。

[0045] 术语“约”通常是指标示值的加或减10%。例如, “约10%”可表示9%到11%的范围, 并且“约20”可表示从18到22。“约”的其他含义可从上下文中变得显而易见 (如四舍五入)。例如“约1”也可表示从0.5到1.4。

[0046] 论述

[0047] 根据第一实施方式, 本发明包括一种gRNA-配体结合复合物, 其含有gRNA和配体结合基团两者。此复合物具有保持与V型Cas蛋白的结合的能力。在该gRNA-配体结合复合物内部, gRNA可以直接共价键合至配体结合基团或者通过接头结合配体结合基团。

[0048] gRNA

[0049] 该gRNA-配体结合复合物的gRNA是核苷酸的单链 (该单链具有至少一个是自互补

的区域)或者核苷酸的两条链(其中每条链都具有至少一个是互补于其他链的一个区域的区域)。在该gRNA中,无论它是核苷酸的单链还是核苷酸的两条链,均可存在一个或多个环。该gRNA包含两个部分:tracrRNA和crRNA。

[0050] gRNA中的核苷酸可完全地是RNA或者核糖核苷酸与其他核苷酸(如脱氧核糖核苷酸)的组合。每个核苷酸可以是无修饰的,或者可以修饰一个或多个核苷酸,例如,具有以下修饰中的一个:2'-O-甲基、2'氟或2-氨基嘌呤。在一些实施方式中,在1至40或者2至20或2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、35、或36个核苷酸的一个或多个范围中,在其2'位置存在连续修饰的核苷酸,或者修饰每两个或每三个或每四个核苷酸的修饰模式,而所有其他核苷酸则无修饰。此外或者可替代地,在一对或多对或者每对的连续核苷酸之间,可存在及经修饰或无修饰的核苷酸间连接。

[0051] 在一些实施方式中,crRNA的长度为35至60个核苷酸或者36至60个核苷酸或40至55个核苷酸。crRNA序列中为Cas结合区(该区也可被称为重复区,其长度可为18至30个核苷酸或20至25个核苷酸)和靶向区(该区也可被称为间隔区,其长度可为18至30个核苷酸或20至25个核苷酸)。

[0052] 该靶向区含有靶向序列,其是可变序列,该序列可基于希望Cas蛋白和/或效应物引发碱基编辑的位置来选择。因此,该靶向区可被设计成包含与目的预选靶位点互补且能够杂交的区域。例如,在靶向区与相应靶位点序列之间的互补区域的长度可为约为10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、或多于25个连续核苷酸,或者它可以是与DNA区域上10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、或多于25个连续核苷酸至少80%、至少85%、至少90%、或至少95%互补。该靶向区是不与tracrRNA杂交的区域,并且它可以是Cas结合区的下游。该Cas结合是基于旨在与之结合的Cas蛋白的RNA结合结构域设计的(在Cas结合中,并非所有核苷酸都需要与Cas蛋白直接结合)。

[0053] TracrRNA序列可以,例如长度为30至210个核苷酸或者45至120个核苷酸或者60至100个核苷酸或者70至90个核苷酸。TracrRNA序列包含反重复区和远端区。在一些实施方式中,反重复区的长度为18至60个核苷酸或者25至50个核苷酸或者30至40个核苷酸。远端区的长度可以,例如长度为18至60个核苷酸或者25至50个核苷酸或者30至40个核苷酸。远端区是不与crRNA杂交的区域,并且它可以是反重复区的上游。

[0054] 在Cas结合区的至少18个连续核苷酸上,反重复区与Cas结合区是至少80%、至少85%、至少90%、至少95%、或100%互补,因而Cas结合区与反重复区能够杂交形成杂交区。当tracrRNA与crRNA在杂交区上杂交时,gRNA能够保持与V型Cas蛋白的RNA结合结构域的结合。优选地,这种结合在天然存在的条件和在使用该复合物的实验室条件下均是可能的。

[0055] 如果tracrRNA和crRNA是核苷酸的连续链的部分,那么在tracrRNA与crRNA之间,例如4至20个或8至15个核苷酸,可以存在环区。在5'至3'方向上,该gRNA可包含、基本上由或者由远端区、反重复区、环、Cas结合区和靶向区组成。

[0056] 可结合本发明而使用的tracrRNAs和crRNAs的示例如下所示(全部在5'→3'方向上,并且N是任意核苷酸,例如A、C、G或U)。在下面的每个序列中,显示了特定数量的N核苷酸。然而,在这些序列的任一序列中,N可为16-30,这表示可能存在16、17、18、19、20、21、22、23、24、25、26、27、28、29、或30个核苷酸。

- [0057] 嗜酸耐热菌 (*Alicyclobacillus acidoterrestris*), AacCas12b crRNA
- [0058] SEQ ID NO:1
- [0059] 5'-GUCGGAUCACUGAGCGAGCGAUCUGAGAAGUGGCACNNNNNNNNNNNN NNNNNNNNN-3'
- [0060] SEQ ID NO:2
- [0061] 5'-CGAGCGAUCUGAGAAGUGGCACNNNNNNNNNNNNNNNNNNNN-3'
- [0062] SEQ ID NO:3
- [0063] 5'-CUGAGAAGUGGCACNNNNNNNNNNNNNNNNNNNN-3'
- [0064] AacCas12b tracrRNA
- [0065] SEQ ID NO:4
- [0066] 5'-GUCUAGAGGACAGAAUUUUUCAACGGGUGUGCCAAUGGCCACUUUCC AGGUGGCAAAGCCCGUU
GAGCUUCUCAAAAAG-3'
- [0067] SEQ ID NO:5
- [0068] 5'-GUCUAGAGGACAGAAUUUUUCAACGGGUGUGCCAAUGGCCACUUUCC
AGGUGGCAAAGCCCGUUGAGCUUCUCAAAA-3'
- [0069] SEQ ID NO:6
- [0070] 5'-GUCUAGAGGACAGAAUUUUUCAACGGGUGUGCCAAUGGCCACUUUCC
AGGUGGCAAAGCCCGUUGAG-3'
- [0071] AacCas12b gRNA
- [0072] SEQ ID NO:7
- [0073] 5'-GUCUAGAGGACAGAAUUUUUCAACGGGUGUGCCAAUGGCCACUUUCC AGGUGGCAAAGCCCGUU
GAGCUUCUCAAAUCUGAGAAGUGGCACNNNNNNNNNNNNNNNNNNNN-3'
- [0074] AacCas12b人工嵌合gRNA (artgRNA13)
- [0075] SEQ ID NO:8
- [0076] 5'-GUCGUCUAUAGGACGGCGAGGACAACGGGAGUGCAGUCUCUUUCCA AGAGCAAACACCCCGUU
GGCUUCAAGAGAAGUGGCACNNNNNNNNNNNNNNNNNNNN-3'
- [0077] 嗜热淀粉芽孢杆菌 (*Bacillus thermoamylovorans*), Bth Cas12b tracrRNA
- [0078] SEQ ID NO:9
- [0079] 5'-CGAGGUUCUGUCUUUUGGUCAGGACAACCGUCUAGCUAUAAGUGCUG CAGGGGUGUGAGAAACU
CCUAUUGCUGGACGAUGUCUCUUUUU-3'
- [0080] 嗜热淀粉芽孢杆菌 Bth Cas12b crRNA
- [0081] SEQ ID NO:10
- [0082] 5'-GUCCAAGAAAAAAGAAAUGAUACGAGGCAUAGCACNNNNNNNNNNNN NNNNNNNNN-3'
- [0083] SEQ ID NO:11
- [0084] 5'-AAAUGAUACGAGGCAUAGCACNNNNNNNNNNNNNNNNNNNN-3'
- [0085] SEQ ID NO:12
- [0086] 5'-CGAGGCAUAGCACNNNNNNNNNNNNNNNNNNNN-3'
- [0087] Cas12e, CasX来自 δ 变形杆菌
- [0088] Cas12e, DpbCasX, crRNA
- [0089] SEQ ID NO:13

- [0090] 5'-CCGAUAAGUAAAACGCAUCAAGNNNNNNNNNNNNNNNNNNNN-3'
- [0091] Cas12e, DpbCasX, tracrRNA
- [0092] SEQ ID NO:14
- [0093] 5'-GGCGCGUUUAUCCAUAUACUUUGGAGCCAGUCCCAGCGACUAUGUCGUAUGGACGAAGCGCUUAU
UUAUCGGAGA-3'
- [0094] Cas12e, DpbCasX, gRNA (crRNA+tracrRNA融合, 更短的变体)
- [0095] SEQ ID NO:15
- [0096] 5'-GGCGCGUUUAUCCAUAUACUUUGGAGCCAGUCCCAGCGACUAUGUCGUAUGGACGAAGCGCUUAU
UAUCGGAGAGAAACCGAUAAAGUAAAACGCAUCAANNNNNNNNNNNNNNNNNNNN-3'
- [0097] Cas12e, DpbCasX, gRNA (crRNA+tracrRNA融合, 更长的变体)
- [0098] SEQ ID NO:16
- [0099] 5'-ACAUCUGGCGCGUUUAUCCAUAUACUUUGGAGCCAGUCCCAGCGACUAUGUCGUAUGGACGAAGC
GCUUAUUUAUCGGAGAGAAACCGAUAAAGUAAAACGCAUCAAGNNNNNNNNNNNNNNNNNNNN-3'
- [0100] Cas12f核酸酶
- [0101] Un1Cas12f1 (Cas14a1) gRNA
- [0102] SEQ ID NO:17
- [0103] 5'-GGGCUUCACUGAUAAAAGUGGAGAACCGCUUCACCAAAAGCUGUCCCUUAGGGGAUUAGAACUUGA
GUGAAGGUGGGCUGCUUGCAUCAGCCUAAUGUCGAGAAGUGCUUUCUUCGAAAGUAACCCUCGAAACAAAUUCAU
UUUUCUCUCCAAUUCUGCACAAGAAAGUUGCAGAACCCGAAUAGACGAAUGAAGGAAUGCAACNNNNNNNNNNNN
NNNNNNNN-3'
- [0104] Cas12ftracrRNA
- [0105] SEQ ID NO:18
- [0106] 5'-CUUCACCAAAAAGCUGUCCCUUAGGGGAUUAGAACUUGAGUGAAGGUGGGCUGCUUGCAUCAGCCU
AAUGUCGAGAAGUGCUUUCUUCGAAAGUAACCCUCGAAACAAAUUCAUUU-3'
- [0107] Cas12fcrRNA
- [0108] SEQ ID NO:19
- [0109] 5'-GAAUGAAGGAAUGCAACNNNNNNNNNNNNNNNNNNNN-3'
- [0110] SEQ ID NO:20
- [0111] 5'GUUGCAGAACCCGAAUAGACGAAUGAAGGAAUGCAACNNNNNNNNNN NNNNNNNNN-3'
- [0112] Cas12fgRNA (crRNA+tracrRNA融合)
- [0113] SEQ ID NO:21
- [0114] 5'-CUUCACCAAAAAGCUGUCCCUUAGGGGAUUAGAACUUGAGUGAAGGUG GGCUGCUUGCAUCAGCC
UAAUGUCGAGAAGUGCUUUCUUCGAAAGUAACCCUCGAAACAAAUUCAUUUUGAAAGAAUGAAGGAAUGCAACNN
NNNNNNNNNNNNNNNNNN-3'
- [0115] 配体结合基团
- [0116] 该配体结合基团是能够与配体可逆地结合(通过, 例如形成非共价相互作用)的元
件。在一些实施方式中, 该配体结合基团是适配体。该配体结合基团可直接结合gRNA, 例如
通过共价键或者通过接头。配体结合基团与gRNA的结合(无论直接通过共价键还是通过接
头)均可在多个位置中的任一个位置, 例如, tracrRNA在反重复区中或者远端区中或者在靶

向区中的crRNA中或在Cas结合区中、或者tracrRNA与crRNA之间的环中(如果存在)。配体结合基团直接结合gRNA(如果它结合到在gRNA内部的核苷酸),例如结合到单元的主链磷酸盐或者结合到糖基团或者结合到核苷酸的含氮碱基。

[0117] 通过非限制性的举例,配体结合基团可直接结合到(通过例如共价键)gRNA的3'末端或gRNA的5'末端,如果该gRNA是单链(其对应于tracrRNA的5'末端和crRNA的3'末端)或者结合到crRNA的3'末端、tracrRNA的3'末端、tracrRNA的5'末端或crRNA的5'末端(如果它们是分离的链)。因此,该配体结合基团可结合在gRNA中或者在gRNA的任一链中的第一个或最后一个的核苷酸。或者,该配体结合基团可结合gRNA中的并非第一个或最后一个核苷酸的核苷酸(如果是单链的)或者tracrRNA或crRNA(如果是分离的链)。

[0118] 当配体结合基团是核苷酸序列,并且它直接结合gRNA的5'末端或3'末端时,它可在与gRNA相同的5'至3'方位上。在这些情况下,存在含有配体结合基团和gRNA的核苷酸的连续链,且由此是

[0119] • 5'-[gRNA]-[配体结合基团]-3'或者

[0120] ● 5'-[配体结合基团]-[gRNA]-3'。

[0121] 在其他实施方式中,配体结合基团可在相反的方位上直接地附接gRNA,且由此是

[0122] • 5'-[gRNA]-3'-3'-[配体结合基团]-5'或者

[0123] • 3'-[配体结合基团]-5'-5'-[gRNA]-3'。

[0124] 该配体结合基团也可在不同于5'末端或3'末端的位置附接gRNA。当该配体结合基团是核苷酸序列时,可将它插入至gRNA中,并由此可以是,例如,gRNA的第一段(即配体结合基团5'端)和gRNA的的第二段(即配体结合基团3'端),因而存在一个寡核苷酸序列5'-[gRNA的第一段]-[配体结合基团]-[gRNA的第二段]-3'。相对于不含有该配体结合基团的gRNA,含有gRNA且其中插入有配体结合基团的复合物可以不存在来自Cas结合区或者靶向区的核苷酸的缺失。或者,在插入位置可缺失一个或多个核苷酸(例如,1至10个核苷酸)。

[0125] 在一些实施方式中,配体结合基团形成茎和环复合物。在一些实施方式中,配体结合基团的此茎和环复合物可在位置中处在没有配体结合基团的情况下,存在凸起或另一个茎-环复合物。

[0126] 在一些实施方式中,当配体结合基团不是核苷酸序列并且它作为不同于5'或3'末端的位置而结合gRNA时,它可以例如结合在两个连续核苷酸之间:

[0127] • 5'-[gRNA的第一段]-[配体结合基团]-[gRNA的第二段]-3',

[0128] 或者它可在gRNA的5'或3'末端(在例如2'、3'或5'位置或者含氮碱基)附接磷基团、糖。这些配体结合基团,例如,可在gRNA的凸起或茎环的位置或者在没有凸起或茎-环的位置结合。

[0129] 在其他实施方式中,配体结合基团结合接头,该接头结合gRNA的3'末端或gRNA的5'末端或者在gRNA内的另一个或多个位置。在一些实施方式中,接头和配体结合基团中的每一个都可以独立地包含、基本上由、或者由核苷酸组成。在一些实施方式中,接头和配体结合基团中的每一个都可独立地包含、基本上由、或者由除核苷酸外的基团所组成。

[0130] 当配体结合基团是核苷酸序列并且它通过接头结合gRNA的5'末端或3'末端时,RNA、接头和配体结合基团中的每一个都可处在相同的5'至3'方位。在这些情况下,存在含有配体结合基团和gRNA的核苷酸的连续链

[0131] ●5'-[gRNA]-[接头]-[配体结合基团]-3'或者

[0132] ●5'-[配体结合基团]-[接头]-[gRNA]-3'。

[0133] 在其他实施方式中,配体结合基团和/或接头可以以相反方位直接附接gRNA,并由此为

[0134] ●5'-[gRNA]-3'-3'[接头]-5'-3'-[配体结合基团]-5'或

[0135] ●5'-[gRNA]-3'-5'-[接头]-3'-3'-[配体结合基团]-5'或

[0136] ●3'-[配体结合基团]-5'-3'-[接头]-5'-5'-[gRNA]-3'或

[0137] ●3'-[配体结合基团]-5'-5'-[接头]-3'-5'-[gRNA]-3'。

[0138] 配体结合基团也可在除5'末端或3'外的位置通过接头附接gRNA。当配体结合基团和接头是核苷酸序列时,它们可插入至gRNA中,并由此可以存在例如gRNA的第一段(即,配体结合基团的5')和接头,以及gRNA第二段(即,配体结合基团的3')和接头。在一些实施方式中,存在位于配体结合基团侧面的两个接头。

[0139] 当仅存在一个接头序列时,它可以是配体结合基团的5'或3',使得复合物是

[0140] • 5'[gRNA的第一段]-[接头]-[配体结合基团]-[gRNA的第二段]-3'或者

[0141] • 5'-[gRNA的第一段]-[配体结合基团]-[接头]-gRNA的第二段]-3'。

[0142] 当存在两个接头序列时,第一接头可以是配体结合基团的5',并且第二接头是配体结合基团的3',使得复合物是

[0143] • 5'-[gRNA的第一段]-[第一接头]-[配体结合基团]-[第二接头]-[gRNA的第二段]-3'。

[0144] 在一些实施方式中,gRNA的第一段、接头、配体结合基团、第二接头、gRNA的第二段中的每一个都是在相同的方位上的核苷酸序列。在其他实施方式中,第一接头、配体结合基团和第二接头中的一个或多个是处在与gRNA的第一段和RNA的第二段相反的方位上,gRNA的第一段和RNA的第二段是处在相同的方位上。

[0145] 在一些实施方式中,当配体结合基团是在gRNA的第一段与gRNA的第二段之间时(并且如果存在一个或两个接头,它们也是在gRNA的第一段与gRNA的第二段之间):

[0146] • gRNA的第一段含有远端区的一部分,并且gRNA的第二段含有远端区的剩余部分和整个反重复区;

[0147] • gRNA的第一段含有远端区的一部分,并且gRNA的第二段含有远端区的剩余部分、整个反重复区和crRNA;

[0148] • gRNA的第一段含有整个远端区,并且gRNA的第二段的含有整个反重复区;

[0149] • gRNA的第一段含有整个远端区,并且gRNA的第二段含有整个反重复区和crRNA;

[0150] • gRNA的第一段含有整个远端区和反重复区的一部分,并且gRNA的第二段含有反重复区的剩余部分;

[0151] • gRNA的第一段含有整个远端区和反重复区的一部分,并且gRNA的第二段含有反重复区的剩余部分和crRNA;

[0152] • gRNA的第一段含有tracrRNA,并且gRNA的第二段含有crRNA;

[0153] • gRNA的第一段含有tracrRNA和Cas结合区的一部分,并且RNA的第二段含有Cas结合区的剩余部分和整个靶向区;

[0154] • gRNA的第一段含有Cas结合区的一部分,并且gRNA的第二段含有Cas结合区的

剩余部分和整个靶向区；

[0155] • gRNA的第一段含有tracrRNA和整个Cas结合区,并且gRNA的第二段含有整个靶向区；

[0156] ●gRNA的第一段含有整个Cas结合区,并且gRNA的第二段含有整个靶向区；

[0157] • gRNA的第一段含有tracrRNA、整个Cas结合区和靶向区的一部分,并且gRNA的第二段含有靶向区的剩余部分；以及

[0158] ●gRNA的第一段含有整个Cas结合区和靶向区的一部分,并且gRNA的第二段含有靶向区的剩余部分。

[0159] 相对于不含有配体结合基团(及可选地一个或多个接头)的gRNA,在含有gRNA且其中插入有配体结合基团的复合物中,可以没有来自或在插入区附近的核苷酸的缺失。或者,在插入位置或附近,可以存在一个或多个核苷酸(例如,1至10个核苷酸)的缺失。

[0160] 当存在两个接头时,它们可以是充足互补的,使得它们能够彼此杂交。例如,每个接头的长度都可为1至20个核苷酸,并且接头可以是至少80%、至少85%、至少90%、至少95%、至少98%或100%互补,并且不具有凸起或者具有一个或多个凸起。在一些实施方式中,接头是相同尺寸的;在其他实施方式中,接头是不同尺寸的。

[0161] 当配体结合基团结合(直接或者通过接头)gRNA的环时,结合可以,例如,是在该环中的第一核苷酸、该环中的第二核苷酸、该环中的第三核苷酸、该环中的第四核苷酸、该环中的中心核苷酸,如果该环具有奇数个核苷酸或者在环中的两个最中心核苷酸中的一个,如果该环具有偶数个核苷酸,或者在该环中的最后一个核苷酸。可修饰任何一个或多个前述核苷酸和/或与它们相应的核苷酸5'和/或3'核苷酸间连接。这些修饰可以,例如发生在配体结合基团结合(直接或者通过接头)gRNA处或者仅在在除配体结合基团结合(直接或者通过接头)gRNA处外的位置。例如,该配体结合基团可以附接糖的2'位置或者附接在gRNA寡核苷酸序列中的含氮碱基。

[0162] 在一些实施方式中,配体结合基团包含、基本上由或者由无修饰的或者包含一个或多个经修饰的核苷酸的寡核苷酸序列组成。例如,配体结合基团的长度可以是10至50或者18至50个核苷酸。在一些实施方式中,配体结合基团形成茎-环结构。如果不存在接头,配体结合基团可呈现为紧邻gRNA的5'或3'的gRNA序列的延伸或者tracrRNA或者crRNA的5'或3',或者它可呈现为插入。

[0163] 在一些实施方式中,配体结合基团包含、基本上由、或者由生物素或链霉亲和素组成。

[0164] 在一些实施方式中,配体结合基团是选自由与以下配体结合的基团所组成的组:MS2包被蛋白(MCP)、Ku、PP7包被蛋白(PCP)、Com RNA结合蛋白或其结合结构域、SfMu、Sm7、Tat、谷胱甘肽S-转移酶(GST)、CSY4、Qβ、COM、短小杆菌素、抗-His标签(6H7)、SNAP-标签、λN22、凝集素(在此情况下,配体结合基团可以是碳水化合物或聚糖或寡糖)、以及PDGFβ-链。在一些实施方式中,配体结合基团是包含脱氧核糖核苷酸、核糖核苷酸或两者的组合的适配体。因此,作为非限制性示例,技术人员可使用DNA适配体、RNA适配体、主链中有经修饰核苷酸的DNA适配体、主链中有经修饰核苷酸的RNA适配体及其组合。

[0165] 在一些实施方式中,将天然存在的MS2适配体用作配体结合基团。在其他实施方式中,技术人员使用MS2 C-5突变体或MS2 F-5突变体或经修饰的MS2,例如MS2中在位置10存

在一个或多个(例如,1、2、3、4、5、6、7、8、9、10、11、或12个)经修饰的核苷酸(如氨基嘌呤)的M2,其中,位置10是从适配体的5'末端起的第十核苷酸。2-氨基嘌呤可以,例如是2'脱氧-2-氨基嘌呤或2'核糖-2-氨基嘌呤的2-氨基嘌呤。在任意一个位置的修饰可以是对除另一个位置处的修饰外或者对任何或全部其他位置的修饰的排除。

[0166] 在一些实施方式中,配体结合基团是包含5'经修饰的核苷酸的适配体,其中,5'经修饰的核苷酸包括2'修饰、5'PO₄基团、或含氮碱基修饰中的至少一个。

[0167] 在一些实施方式中,配体结合基团为是或包含适配体-配体对的一部分的适配体,并且如下所述,该效应物连接或者包含适配体-配体对的其他部分。例如,该适配体可包含特异性结合MS2包被蛋白MCP的MS2操作子基序。如本领域技术人员将理解的,可替代地,该适配体可以包含MCP基团(或其他配体),在此情况下,该效应物将包含或者连接到MS2操作子基序(或其他相应的配体结合基团)。

[0168] 接头

[0169] 接头(当存在时)可以是将配体结合基团连接到gRNA的物质。它可以在一个位置处附接配体结合基团和gRNA中的任一个,或者它可在多个位置处附接gRNA和配体结合基团的任一者或两者。在多个位置的附接可允许更好地控制配体结合基团的三维空间,继而允许使用效应物。

[0170] 通过非限制性的举例,该接头可在一个位置处附接gRNA,并且在两个或更多的位置处附接配体结合基团;或者该接头可在一个位置处附接配体结合基团,并且在两个或多个位置处附接gRNA。当该接头在两个或多个位置处附接gRNA时,接头可唯一地在靶向区中、唯一地在Cas结合区中、或者在靶向区和Cas结合区两者中、唯一地在反重复区中、唯一地在远端区中、在反重复区和远端区两者中、在远端区和靶向区两者中、在反重复区和Cas结合区两者中、在反重复区和靶向区两者中、或者在Cas结合区和远端区两者中附接gRNA。

[0171] 在一些实施方式中,接头包含、基本上由、或者由寡核苷酸序列组成,以及可选地,接头包含至少一个或多个的2'修饰,例如接头中的所有核苷酸都是2'修饰核苷酸。该核苷酸序列可随机地或有意地设计为不会在适配体、gRNA或DNA的靶位点中与序列不期望地互补。

[0172] 在一些实施方式中,接头包含、基本上由、或者由至少一个硫代磷酸酯连接组成。

[0173] 在一些实施方式中,接头包含、基本上由、或者由乙酰丙酸基团所组成。

[0174] 在一些实施方式中,接头包含、基本上由、或者由乙二醇基团所组成。

[0175] 在一些实施方式中,该接头包括或者是选自由18S、9S或C3组成的组。

[0176] 在一些实施方式中,接头是长度为1至60或者1至24或者2至20或5至15个核苷酸的核苷酸序列。此外,在一些实施方式中,该接头富含GC,例如具有至少50%、至少60%、至少70%、至少80%或至少90%的GC核苷酸。当接头包含核苷酸时,它可以,例如,是单链或双链的或者部分单链和部分双链的。此外,当接头是寡核苷酸时,该接头可唯一地是RNA,唯一地是DNA或者其组合。

[0177] 在一些实施方式中,接头是配体结合基团的上游或下游的核苷酸序列。当接头是配体结合基团的上游,且gRNA是接头的上游时,可存在与接头互补的另一个序列,其是配体结合基团的下游。类似地,当接头是配体结合基团的下游,且gRNA是接头的下游时,可存在与接头互补的另一个序列,其是配体结合基团的上游。如本领域技术人员将认知的,当寡核

核苷酸自折叠并且各链在5'至3'方向上与各相关部分对齐时,确定互补性。

[0178] 因此,在一些实施方式中,配体结合基团(例如,MS2)具有1至12个核苷酸长的上游序列和1至12个核苷酸长的下游序列,其中,上游和下游序列紧邻配体结合基团的两侧(即,在配体结合基团与上游序列和下游序列中的每一个之间不存在其他核苷酸),并且上游序列与下游序列互补。在一些实施方式中,上游序列和下游序列中的每个序列的长度为1个核苷酸、2个核苷酸、3个核苷酸、4个核苷酸、5个核苷酸、6个核苷酸、7个核苷酸、8个核苷酸、9个核苷酸、10个核苷酸、11个核苷酸、或12个核苷酸。在一个实施方式中,上游序列和下游序列中的每个序列都包含或者是GC。当存在上游和下游序列时,它们也可被称为延伸序列。

[0179] 修饰

[0180] 在一些实施方式中,gRNA或配体结合基团中的至少一个被修饰,或者如果存在接头,gRNA、配体结合基团或接头中的至少一个被修饰。该修饰是指引入不会在天然存在的条件下生成的基团或物质。修饰可用于提高稳定性和特异性中的一者或两者。在一些实施方式中,在引入本发明复合物的系统内,相对于Cas蛋白的活性结构域(例如,RuvC结构域)和一个或多个其他酶的活性结构域中的一个或两者的抗性,改善了稳定性(包括但不限于任何效应物)。当修饰降低脱靶效应的可能性和/或提高本发明的碱基编辑复合物将到达其靶位点的可能性,提高了特异性。可以在核糖、磷酸盐连接、和/或碱基基团处对核苷酸进行修饰。例如,可在gRNA内、反重复区,远端区、靶向区或Cas结合区的一个、多个或所有位置和/或配体结合基团和/或接头处使用硫代磷酸酯主链(如果存在)。

[0181] 在一些实施方式中,该修饰是一个或多个2'经修饰的核苷酸(例如,2'-O-甲基或2'-氟)的存在和/或硫代磷酸酯核苷酸间连接的存在或者gRNA的5'-PO₄基团的引入和/或配体结合基团。

[0182] 在一些实施方式中,选择修饰或一组修饰以使相对于缺少该修饰或该组修饰的gRNA-配体结合复合物被赋予RuvC活性核酸酶结构域抗性。在一些实施方式中,可通过空间位阻导致抗性。在一些实施方式中,修饰是位于其中或其中的一个或多个(如果不是全部核苷酸都在靶向区内)。

[0183] 当存在多于一个修饰时,修饰可以,例如全部是在靶向区中;全部是在Cas结合区中;全部是在反重复区中;全部是在远端区中;全部是在配体结合基团中;全部是在接头中(如果存在);是在靶向区和Cas结合区两者中;是在Cas结合区和配体结合基团两者中;是在Cas联区和接头(如果存在);是在靶向区和配体结合基团两者中;是在靶向区和接头(如果存在)两者中;是在配体结合基团和接头(如果存在)两者中;是在Cas结合区、靶向区和配体结合基团的所有三者中;是在Cas结合区、靶向区和接头(如果存在)中;在Cas结合区、配体结合基团和接头(如果存在)之间;在靶向区、配体结合基团和接头(如果存在)中;是在Cas结合区、靶向区、配体结合基团和接头中的每一个区中(如果存在);是在反重复区和远端区中;是在反重复区和Cas结合区中;是在反重复区和靶向区中;是在远端区和Cas结合区中;是在远端区和靶向区中;是在除远端区外的gRNA的所有区域中;是在除反重复区外的gRNA的所有区域中;是在除Cas结合区外的gRNA的所有区域中;是在除靶向区外的全部gRNA中;是在靶向区、Cas结合区、反重复区和远端区中的每一个区中;是在配体结合基团和gRNA中除反重复区外的所有区域;是在配体结合基团中和除Cas结合区外的gRNA的所有区域中;是在配体结合基团和除靶向区的gRNA外的所有区域中;以及是在配体结合基团及靶向区、

Cas结合区、反重复区、和远端区中的每一个区中。

[0184] 在一些实施方式中,存在1至60个或者1至30个或者1个至10个或者10至20或者20至30个或者30至40个或者40至50个或者50个至60个2'修饰。通过非限制性示例,2'修饰组可位于:靶向区中;反重复区中;远端区中;配体结合基团中,如果该配体结合基团是或者包含寡核苷酸序列;或者在Cas结合区中或者在其组合中。该修饰可以在连续核苷酸上或者可以是在修饰的核苷酸之间以规律或不规律的模式存在一对或多对的无修饰核苷酸。通过进一步的非限制性示例,在gRNA中的位置1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17或18中的任一个或多个包含2'-O-烷基,其中,这些位置是从gRNA或tracrRNA或crRNA的5'末端或3'末端进行测量。

[0185] 在一些实施方式中,除了2'-修饰核苷酸外或者不存在2'-修饰核苷酸的情况下,存在经修饰的核苷酸间连接(如硫代磷酸酯连接)。对于gRNA主链的修饰的示例,配体结合基团(在寡核苷酸中)、和接头(如果存在,且寡核苷酸),包括但不限于具有正常3'-5'连接的硫代磷酸酯、手性硫代磷酸酯、二硫代磷酸酯、磷酸三酯、氨基烷基磷酸三酯、甲基和其他烷基磷酸酯(包括3'-亚烷基磷酸酯、5'-亚烷基磷酸酯和手性磷酸酯)、次磷酸盐、磷酰胺类(包括3'-氨基磷酰胺和氨基烷基磷酰胺、磷酰二胺酯类、硫代磷酰胺类、硫代烷基磷酸酯类、硫代烷基磷酸三酯类、硒代磷酸酯类和硼磷酸盐类、其2'-5'连接的类似物,以及那些具有反向极性的类似物(其中一个或多个的核苷酸间连接是3'至3'、5'至5'或2'至2'连接)。合适的寡核苷酸具有方向极性,其在3'-大部分核苷酸间连接处包含单个3'至3'连接,即,单个倒置核苷残基可以是脱碱基的(该核碱基是缺失的或者在其位置上有羟基),各种盐类(例如,举例来说,钾或钠),混合盐和无酸形式的前述核苷酸间连接也被包含于本发明的范围内。

[0186] 多核苷酸主链的使用也在本发明范围内,这些主链中不包含磷原子,相反它们具有由短链烷基或环烷基核苷间连接、混合杂原子和烷基或环烷基核苷间连接、或者一个或多个短链杂原子或杂环核苷间连接所形成的主链。这些修饰包括具有吗啉代连接(部分地由核苷的糖部分形成);硅氧烷主链;硫化物、亚砷和砷主链;甲酰基和硫代甲酰基主链;亚甲基甲酰基和硫代甲酰基主链;核糖乙酰基主链;含有烯烃的主链;氨基磺酸酯主链;亚甲基亚氨基和亚甲基胍基主链;磺酸酯和磺酰胺主链;酰胺主链;和其他混合有N、O、S和CH₂组成部分的修饰。

[0187] 在一些实施方式中,复合物的一个或多个部分具有1至60个或者1至20个或者1至10个或者10至20个或者20至30个或者30至40个或者40至50或者50至60个硫代磷酸酯连接。这些硫代磷酸酯连接可:全部是在靶向区中;全部是在Cas结合区中;全部是在反重复区中;全部是在远端区中;全部是在配体结合基团中;全部是在接头中(如果存在);是在靶向区和Cas结合区两者中;是在Cas结合区和配体结合基团两者中;是在Cas结合区和接头(如果存在)两者中;是在靶向区和配体结合基团两者中;是在靶向区和接头(如果存在)两者中;是在配体结合基团和接头(如果存在)两者中;是在Cas结合区、靶向区和配体结合基团的所有三者中;是在Cas结合区、靶向区和接头中(如果存在);是在Cas结合区、配体结合基团和接头(如果存在)中;是在靶向区、配体结合基团和接头中(如果存在);是在Cas结合区、靶向区、配体结合基团和接头(如果存在)中的每一个中;是在反重复区和远端区中;是在反重复区和Cas结合区中;是在反重复区和靶向区中;是在远端区和Cas结合区中;是在远端区和靶

向区中；是在除远端区外的gRNA的所有区域中；是在除反重复区外的gRNA的所有区域中；是在除Cas结合区外的gRNA的所有区域中；是在除靶向区外的gRNA的所有区域中；是在靶向区、Cas结合区、反重复区和远端区中的每一个中；是在配体结合基团和除反重复区外的gRNA的所有区域中；是在配体结合基团和除Cas结合区外的gRNA的所有区域中；是在配体结合基团和除靶向区外的gRNA的所有区域；以及是在配体结合基团和靶向区、Cas结合区、反重复区、和远端区中的每一个中。

[0188] 在本发明的复合物内的任何核苷酸可包含一个或多个经取代的糖基团。这些核苷酸可包含选自下列的糖取代基：OH；H；F；O-、S-、或N-烷基；O-、S-、或N-烯基；O-、S-或N-炔基；或O-烷基-Co-烷基，其中，烷基、烯基和炔基可以是经取代的或未经取代的C1至C10烷基或者C2至C10烯基和炔基。 $O((CH_2)_nO)_mCH_3$ 、 $O(CH_2)_nOCH_3$ 、 $O(CH_2)_nNH_2$ 、 $O(CH_2)_nCH_3$ 、 $O(CH_2)_nONH_2$ 、和 $O(CH_2)_nN((CH_2)_nCH_3)_2$ 尤其合适，其中，n和m为1至约10。其他合适的核苷酸包含选自下列的糖取代基：C1至C10低级烷基、经取代的低级烷基、烯基、炔基、烷芳基、芳烷基、O-烷芳基或O-芳烷基、SH、SCH₃、OCN、Cl、Br、CN、CF₃、OCF₃、SOCH₃、SO₂CH₃、ONO₂、NO₂、N₃、NH₂、杂环烷基、杂环烷芳基、氨基烷基氨基、聚烷基氨基、经取代的硅烷基、RNA切割基团、报告基团、嵌入剂、用于改善寡核苷酸的药代动力学性质的基团、或者用于改善寡核苷酸的药效学性质的基团、和具有相似性质的其他取代基。通过非限制性示例，合适的修饰包括2'-甲氧基乙氧基(2'-O-CH₂CH₂OCH₃，也被称为2'-O-(2-甲氧基乙基)或2'-MOE)(Martin等人，*Helv.Chim.Acta*, 1995年, 78, 486-504)或者另一烷氧烷氧基。另一合适的修饰包括2'-二甲氨基氧乙氧基，即， $O(CH_2)_2ON(CH_3)_2$ 基团，也称为2'-DMAOE，和2'-二甲氨基乙氧基乙氧基(在本领域中也称为2'-O-二甲基-氨基-乙氧基-乙基或2'-DMAEOE)，即，2'-O-CH₂-O-CH₂-N(CH₃)₂。

[0189] 其他合适的糖取代基包括甲氧基(-O-CH₃)、氨基丙氧基(-OCH₂CH₂CH₂NH₂)、烯丙基(-CH₂-CH=CH₂)、O-烯丙基(CH₂-CH=CH₂)和氟(F)。2'-糖取代基可在阿拉伯(上)位置或核糖(下)位置。合适的2'-阿拉伯修饰是2'-F。类似的修饰也可在寡聚化合物上的其他位置进行，具体地，在3'端核苷酸的糖的3'位置上、或者在2'-5'连接的寡核苷酸中和5'端核苷酸的5'位置。

[0190] 在本发明复合物中的任何核苷酸也可包含核碱基(在本领域中经常被简称为“碱基”)修饰或取代。经修饰的核碱基包括但不限于：其他合成和天然的核碱基，如5-甲基胞嘧啶(5-me-C)、5-羟基甲基胞嘧啶、黄嘌呤、次黄嘌呤、2-氨基腺嘌呤、腺嘌呤和鸟嘌呤的6-甲基和其他烷基衍生物、腺嘌呤和鸟嘌呤的2-丙基及其他烷基衍生物、2-硫代尿嘧啶、2-硫代胸腺嘧啶和2-硫代胞嘧啶、5-卤代尿嘧啶和胞嘧啶、5-丙炔基(-C≡C-CH₃)尿嘧啶和胞嘧啶及嘧啶碱基的其他炔基衍生物、6-偶氮基尿嘧啶、胞嘧啶和胸腺嘧啶、5-尿嘧啶(假尿嘧啶)、4-硫代尿嘧啶、8-卤代、8-氨基、8-硫代、8-硫代烷基、8-羟基和其他8-经取代的腺嘌呤和鸟嘌呤、5-卤代，具体地，5-溴、5-三氟甲基和其他5-经取代的尿嘧啶和胞嘧啶、7-甲基鸟嘌呤和7-甲基腺嘌呤、2-F-腺嘌呤、2-氨基腺嘌呤、8-氮杂鸟嘌呤和8-氮杂腺嘌呤、7-脱氮鸟嘌呤和7-脱氮腺嘌呤和3-脱氮鸟嘌呤和3-脱氮腺嘌呤。修饰的核碱基还包括但不限于：三环嘧啶，如吩恶嗪胞苷(1H-嘧啶并(5,4-b)(1,4)苯并恶嗪-2(3H)-酮)、吩噻嗪胞苷(1H-嘧啶并(5,4-b)(1,4)苯并噻嗪-2(3H)-酮)、G-钳形物，如经取代的吩恶嗪胞苷(例如9-(2-氨基乙氧基)-H-嘧啶并(5,4-b)(1,4)苯并恶嗪-2(3H)-酮)、咔唑胞苷(2H-嘧啶并(4,5-b)

吡啶-2-酮)、吡啶并吡啶胞苷(H-吡啶并(3',2':4,5)吡咯并(2,3-d)嘧啶-2-酮)和5-甲氧基尿嘧啶。

[0191] 杂环碱基基团也可包括但不限于:其中的嘌呤或嘧啶碱基被其他杂环取代的那些杂环碱基基团,例如7-脱氮-腺嘌呤、7-脱氮鸟苷、2-氨基吡啶和2-吡啶酮。其他核碱基的示例包括在美国专利第3,687,808号中所公开的那些核碱基,在《聚合物科学与工程简明百科全书》中所公开的那些核碱基(第858-859页,Kroschwitz,J.I.编著,John Wiley&Sons,1990年),由Englisch等人在Angewandte Chemie中所公开的那些核碱基(国际版,1991年,30,613),和由Sanghvi,Y.S.,第15章,《反义研究和应用》所公开的那些核碱基(第289-302页,Crooke,S.T.和Lebleu,B.编著,CRC出版社,1993年)。这些核碱基中的某些对于增加寡聚化合物的结合亲和力是有用的:5-经取代的嘧啶、6-氮杂嘧啶和N-2、N-6和O-6经取代的嘌呤,包括2-氨基丙基腺嘌呤、5-丙炔基尿嘧啶和5-丙炔基胞嘧啶。此外,当与2'-O-甲氧基乙基糖修饰组合时,5-甲基胞嘧啶取代可能是有利的。

[0192] 在一些实施方式中,存在与gRNA结合的两个配体结合基团:第一配体结合基团和第二配体结合基团。可选地,可存在两个接头:第一接头和第二接头,其中,第一配体结合基团附接第一接头,并且第二配体结合基团附接第二接头。在这些实施方式中,第一接头和第二接头每个都可附接Cas结合区;或者第一接头和第二接头每个都可附接靶向区;或者第一接头和第二接头中的一个接头可附接结合区,并且第一接头和第二接头中的另一个接头可附接靶向区,或者第一接头和第二接头每个都可附接反重复区;或者第一接头和第二接头每个都可附接远端区;或者第一接头和第二接头中的一个接头可附接远端区,并且第一接头和第二接头中的另一个接头可附接反重复区;或者第一接头和第二接头中的一个接头可附接远端区,并且第一接头和第二接头中的另一个接头可附接Cas结合区;或者第一接头和第二接头中的一个接头可附接远端区,并且第一接头和第二接头中的另一个接头可附接靶向区;或者第一接头和第二接头中的一个接头可附接反重复区,并且第一接头和第二接头中的另一个接头可附接Cas结合区;或者第一接头和第二接头中的一个接头可附接反重复区,并且第一接头和第二接头中的另一个接头可附接靶向区;或者第一接头和第二接头中的一个接头可附接环区(如果存在)并且第一接头和第二接头中的另一个接头可附接靶向区、Cas结合区、反重复区、或远端区中的一个区。

[0193] 图1A是本发明的gRNA-配体结合复合物的图示,其中,配体结合基团114结合Cas12b gRNA中的tracrRNA 112的5'末端,其中,crRNA 110是与tracrRNA的部分杂交的分离的链。图1B是本发明的gRNA-配体结合复合物的图示,其中,配体结合基团124结合tracrRNA 122的3'末端。还示出了crRNA 120。图1C是本发明的gRNA-配体结合复合物的图示,其中,配体结合基团134结合crRNA 130的5'末端。还示出了tracrRNA 132。图1D是本发明的gRNA-配体结合复合物的图示,其中,第一配体结合基团146结合tracrRNA 142的5'末端,并且第二配体结合基团144结合crRNA 140的5'末端。通过非限制性示例,在图1A至1D和在所有其他图中所示出的每个配体结合基团中的每一个都可以,例如,是DNA或RNA适配体、碳水化合物或寡糖基、用于SNAP/CLIP标记的苄基鸟嘌呤或苄基胞嘧啶基团、生物偶联基团(例如,用于点击化学的叠氮炔基)、用于亲和力或共价结合的生物素或其他官能团。在如图1D中所示的双配体结合基团系统中,不同的或相似的官能团可以附接每个分子以进一步提高官能度。

[0194] 图2A是本发明的gRNA-配体结合复合物的图示,其中,配体结合基团214结合Cas12b gRNA中的tracrRNA 212的5'末端,crRNA 210是与tracrRNA相同的核苷酸链的一部分。图2B是本发明的gRNA-配体结合复合物的图示,其中,配体结合基团224结合Cas12b gRNA中的tracrRNA 222与crRNA220之间的环。图2C是本发明的gRNA-配体结合复合物的图示,其中,第一配体结合基团234结合在tracrRNA232与crRNA230之间的环区,并且第二配体结合基团236结合Cas12b gRNA中的tracrRNA的5'末端。图2D是本发明的gRNA-配体结合复合物的图示,其中,配体结合基团244结合tracrRNA区242内的环,并且crRNA 240在Cas12b gRNA中形成gRNA的3'部。

[0195] 图3A是本发明的gRNA-配体结合复合物的图示,其中,配体结合基团314结合Cas12e gRNA中的tracrRNA312的3'末端,Cas12e gRNA具有分离的crRNA链310。图3B是本发明的gRNA-配体结合复合物的图示,其中,配体结合基团324结合Cas12e gRNA中的tracrRNA322的5'末端,Cas12e gRNA具有分离的crRNA链320。图3C是本发明的gRNA-配体结合复合物的图示,其中,配体结合基团334结合Cas12e gRNA中的crRNA330的5'末端,Cas12e gRNA具有分离的tracrRNA332。图3D是本发明的gRNA-配体结合复合物的图示,其中,配体结合基团334结合Cas12e gRNA中的crRNA340的3'末端,Cas12e gRNA中存在形成tracrRNA342的分离的链。

[0196] 图4A是Cas 12e系统的gRNA的图示,其中,第一配体结合基团414附接crRNA410的5'末端,并且第二配体结合基团416附接tracrRNA412的5'末端。图4B是Cas12e系统的gRNA的图示,其中,第一配体结合基团426附接crRNA 420的3'末端,并且第二配体结合基团424附接tracrRNA 422的3'末端。图4C是Cas 12e系统的gRNA的图示,其中,第一配体结合基团434附接tracrRNA432的5'末端,并且第二配体结合基团436附接crRNA430的3'末端。

[0197] 图5A是Cas 12e系统的gRNA的图示,其中,配体结合基团514附接tracrRNA512的5'末端,并且crRNA 510是与crRNA相同的核苷酸链的一部分。图5B是Cas 12e系统的gRNA的图示,其中,配体结合基团524附接tracrRNA 522与crRNA 520之间的环。图5C是Cas 12e系统的gRNA的图示,其中,配体结合基团534附接tracrRNA532的下游的crRNA 530的3'末端,并且与tracrRNA 532形成发夹。图5D是Cas 12e系统的gRNA的图示,其中,配体结合基团544是附接tracrRNA542内的环,并且RNA540与tracrRNA形成发夹。

[0198] 图6A是Cas 12f系统的gRNA的图示,其中,配体结合基团614附接tracrRNA612的5'末端,并且crRNA610是与tracrRNA形成杂交区的分离的链。图6B是Cas 12f系统的gRNA的图示,其中,配体结合基团624附接tracrRNA 622的3'末端,并且crRNA620是与tracrRNA形成杂交区的分离的链。图6C是Cas 12f系统的gRNA的图示,其中,配体结合基团634附接crRNA 630的5'末端,并且示出了tracrRNA 632。

[0199] 图7A示出了为核苷酸单链的gRNA,该核苷酸含有针对Cas 12f系统的tracrRNA和crRNA两者,其中,配体结合基团714结合tracrRNA712的5'末端,其与crRNA 710形成核苷酸连续链。图7B示出了为核苷酸单链的gRNA,其含有针对Cas 12f系统的tracrRNA722和crRNA 720两者,其中,配体结合基团724结合tracrRNA7与crRNA之间的环。图7C示出了Cas 12f系统的gRNA,其中,第一配体结合基团736位于gRNA的tracrRNA 732的5'末端,并且第二配体结合基团734位于在gRNA的tracrRNA与crRNA730之间的环。

[0200] 碱基编辑复合物

[0201] 根据本发明的另一个实施方式,存在一种碱基编辑复合物。该碱基编辑复合物包含、基本上由、或者由本发明的gRNA-配体结合复合物;及V型Cas蛋白组成,其中,gRNA-配体结合复合物的Cas结合区和反重复区与V型Cas蛋白结合。因此,该gRNA能够与Cas蛋白结合。在发生结合后,Cas蛋白将基于靶向序列的一致性而找到靶位点。

[0202] V型Cas蛋白

[0203] 通常,Cas蛋白包含至少一个RNA结合结构域。该RNA结合结构域与在Cas结合区的向导RNA发生相互作用。于本发明中使用的V型Cas蛋白是gRNA配体结合复合物可以在tracrRNA存在下与之结合的V型Cas蛋白,该tracrRNA含有与Cas结合区具有充足互补性的反重复区。在一些实施方式中,V型Cas蛋白是含有RuvC结构域的核酸内切酶。此RuvC结构域可以被突变以使得端核酸酶失活。在一些实施方式中,该蛋白是含有活性或失活RuvC结构域的切割酶。

[0204] 可结合本发明而使用的V型Cas蛋白的示例包括但不限于:活性或失活形式的Cas12b、Cas12e、CasMINI和Cas12f。

[0205] Cas蛋白可以以纯化或分离的形式提供,或者可以是组合物或复合物的一部分。优选地,当在组合物中时,首先将该蛋白纯化到一定程度,更优选地,纯化到高水平的纯度(例如,约80%、90%、95%、或99%或更高)。在其中可存放并运输本发明的复合物和成分的组合物可以是任何类型的所需的组合物,例如,适合用作或者包含于本领域技术人员将会理解的且将会结合本发明而使用用于RNA-向导靶向的组合物水性组合物。

[0206] 在一些实施方式中,V型Cas蛋白包括具有(a)活性、部分失活或失活的V型Cas蛋白和(b)尿嘧啶DNA糖基化酶(UNG)抑制肽(UGI)的融合蛋白。该UGI肽可以直接或者通过由1至10000个氨基酸残基所组成的连接肽与V型Cas蛋白融合。在一些实施方式中,UGI包括来自芽孢杆菌噬菌体PBS2的野生型UGI序列(<https://www.ncbi.nlm.nih.gov/protein/P14739>):MTNLSDIIEKETGKQLVIQESILMLPEEVVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKML(SEQ ID NO:22)。在一些实施方式中,UGI包括SEQ ID NO:22的变体,这些变体包含野生型UGI肽的片段或者SEQ ID NO:22的同源氨基酸序列。在一些实施方式中,同源序列的UGI片段对比野生型UGI肽序列(SEQ ID NO:22)的同源性为至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少97%、至少99%、或至少99.5%。

[0207] 在一些实施方式中,活性或失活的V型Cas蛋白包含两个或更多的UGI肽或变体的融合。UGI肽或UGI肽的变体可以直接连接另一个UGI肽或V型Cas蛋白或者通过1至100个氨基酸残基的接头连接另一个UGI肽或V型Cas蛋白。

[0208] Cas蛋白或Cas蛋白融合可以以纯化或分离的形式提供,或者可以是组合物或复合物的一部分。

[0209] 效应物

[0210] 本发明的碱基编辑复合物可含有附接配体的效应物。该配体能够与配体结合基团可逆地或不可逆地结合。因此,该配体结合基团招募效应物,例如融合或与配体结合的碱基编辑酶,因为该配体结合基团能够保持与配体的结合。这种设计可能是特别有利的,因为它提供了模块化设计,其中,gRNA的核酸序列靶向功能和效应物功能存在于不同的分子中。例如,为了在相同位点连续引入修饰,技术人员可使用与相同配体结合的不同效应物。相反,

为了在不同位点引入相同的修饰,技术人员可使用带有不同gRNAs的相同配体结合基团,同时使用相同的效应物-配体。因此,这种设计允许了技术人员复用系统,而无将效应物融合gRNAs或Cas蛋白的不良负担。

[0211] 可结合本发明使用的效应物的示例为脱氨酶(如具有胞苷脱氨基或腺嘌呤脱氨基活性的脱氨酶)、以及转录调节剂、修复酶、表观遗传修饰物、组蛋白乙酰基酶、去乙酰基酶、甲基化酶(蛋白类和核苷酸的)和去甲基化酶(组蛋白和核苷酸的)。在一些实施方式中,效应物选自由AID、CDA、APOBEC1、APOBEC3A、APOBEC3B、APOBEC3C、APOBEC3D、APOBEC3F、ADA、ADAR和tRNA腺苷脱氨酶所组成的组。表1中提供了效应物的示例和它们所导致的基因变化的类型。

[0212] 表1. 效应蛋白的实例

酶类型	遗传变化	效应蛋白简称	
[0213] 胞苷脱氨酶	C→U/T	AID	
		APOBEC1	
		APOBEC3A	
		APOBEC3B	
		APOBEC3C	
		APOBEC3D	
		APOBEC3F	
[0214]		APOBEC3G	
		APOBEC3H	
	腺苷脱氨酶	A→I/G	ADA
			ADAR1
			TadA
			TADA
			TAD3
	DNA 甲基转移酶	C→Met-C	ADAR2
			ADAR3
			Dnmt1
去甲基化酶	Met-C→C	Dnmt3a	
胞苷去甲基化酶	5mC →5hmC	TET1	
胞苷去甲基化酶	5mC →5hmC	TET2	
	5hmC →5fC/5caC		
糖基化酶	5fc/5caC →C	TDG	

- [0215] 效应蛋白全名:
- [0216] AID:激活诱导性胞苷脱氨酶,又名AICDA
- [0217] APOBEC1:载脂蛋白B mRNA编辑酶,催化性多肽样1。
- [0218] APOBEC3A:载脂蛋白B mRNA编辑酶,催化性肽样3A
- [0219] APOBEC3B:载脂蛋白B mRNA编辑酶,催化性多肽样3B
- [0220] APOBEC3C:载脂蛋白B mRNA编辑酶,催化性多肽样3C
- [0221] APOBEC3D:载脂蛋白B mRNA编辑酶,催化性多肽样3D
- [0222] APOBEC3F:载脂蛋白B mRNA编辑酶,催化性多肽样3F
- [0223] APOBEC3G:载脂蛋白B mRNA编辑酶,催化性多肽样3G
- [0224] APOBEC3H:载脂蛋白B mRNA编辑酶,催化性多肽样3H
- [0225] ADA:腺苷脱氨酶
- [0226] ADAR1:作用于RNA 1的腺苷脱氨酶
- [0227] ADAR2:作用于RNA2的腺苷脱氨酶
- [0228] ADAR3:作用于RNA 3的腺苷脱氨酶
- [0229] Dnmt1:DNA(胞嘧啶-5-)-甲基转移酶1
- [0230] Dnmt3a:DNA(胞嘧啶-5-)-甲基转移酶3- α
- [0231] TadA:tRNA-特异性腺苷脱氨酶
- [0232] TADA:tRNA(腺嘌呤(34))脱氨酶,叶绿体
- [0233] TAD3:tRNA-特异性腺苷脱氨酶TAD3
- [0234] TET1:甲基胞嘧啶双加氧酶TET1
- [0235] TET2:甲基胞嘧啶双加氧酶TET2
- [0236] TDG:G/T错配-特异性胸腺嘧啶DNA糖基化酶
- [0237] 在一些实施方式中,碱基编辑复合物包含两个或更多个效应物。当存在两个效应物时,它们可被称为:第一效应物和第二效应物。每个效应物都可通过不同的配体附接不同的配体结合基团。或者,当存在两个效应物时,一个附接配体并且通过配体结合基团与gRNA结合,而另一个直接附接Cas蛋白。
- [0238] 配体
- [0239] 如上所述,效应物结合配体,例如,通过一个或多个共价键。表2中提供了可用于本发明的各种实施方式中的配体结合基团-配体对的示例的非详尽列表。无修饰和经化学修饰的版本或者配体结合基团和配体均是在本发明的范围内。
- [0240] 表2.

配体结合基团	配体
端粒酶Ku结合结构域	Ku
端粒酶Sm7结合结构域	Sm7
MS2噬菌体操纵子茎-环	MS2包被蛋白(MCP)
PP7噬菌体操纵子茎-环	PP7包被蛋白(PCP)
Q β 噬菌体操纵子茎-环	Q β 包被蛋白[Q65H]
SfMu噬菌体Com茎-环	ComRNA结合蛋白
非天然RNA适配体	相应的适配体配体

生物素	链霉亲和素
寡糖	凝集素
苄基鸟嘌呤或苄基胞嘧啶	SNAP/CLIP标签
6x-His结合基序	6x-His标签
PDGFB β 链结合基序	PDGFB-链
GST结合基序	GST蛋白
Tat结合基序	BIVTat蛋白
Tat结合基序	HIVTat蛋白
Pumilio结合基序	PUM-HD结构域
BoxB结合基序	λ N22+
Csy4结合基序	Csy4[H29A]

[0242] 下面列出了以上结合对的一部分序列。

[0243] 1. 端粒酶Ku结合基序/Ku异源二聚体

[0244] a. Ku结合发夹

[0245] 5' -UUCUUGUCGUACUUAUAGAUCGCUACGUUAUUUCAUUUUGAAAAUCUGAGUCCUGGGAGUGCGA-3' (SEQ ID NO:23)

[0246] b. Ku异源二聚体

[0247] MSGWESYYKTEGDDEEAEQEEENLEASGDYKYSGRDSLIFLVDASKAMFESQSEDELTPFDMSIQCIQSVYISKI I SSDRDL LAVVFGTEKDKNSVNFKNIVLQELDNPGAKRILELDQFKGQQGQKRFQDMMGHGSDYSLS EVLWVCANLFSVDVQFKMSHKRIMLFTNEDNPHGNDSAKASRARTKAGDLRDTGIFLDMHLKPKGGFDISLFYRDI ISIAEDEDLRVHFEESKLEDLLRKVRAKETRKRALSKLKLKNDIVISVGIYNLVQKALKPPPIKLYRETNEPV KTKTRTFNTSTGGLLLPSDTKRSQIYGSRQI ILEKEETEELKRFDDPGLMLMGFKPLVLLKHHYLRPSLFVYPEE SLVIGSSTLFSALLIKCLEKEVAALCRYTPRRNIPPYFVALVPQEEELDDQKIQTTPPGFQLVFLPFADDKRKMPF TEKIMATPEQVGMKAI VEKLRFTYRSDSFENPVLQQHFRNLEALALDLMEPEQAVDLTLPKVEAMNKRGLSLVDE FKELVYPPDYNPEGKVTKRKHDNEGSGSKRPKVEYSEEELKTHISKGTLGKFTVPMLEACRAYGLKSGLKKQELL EALTKHFQD> (SEQ ID NO:24)

[0248] MVRSGNKA AVVLCMDVGF TMSNSIPGIESPFEQAKK VITMFVQRQVFAENKDEIALVLFGT DGTDNPL SGGDQYQ NITVHRHMLPDFDLLEDIESKI QPGSQADFLDALIVSMDVIQHETIGK KFEKRHIEIFTDLSRFSK SQLDII I HSLKKCDISERHS IHWPCLTIGSNLSIRIAAYKSILQERVKKTWTVVDAKTLKKEDIQKETVYCLNDD DETEVLKEDI IQGFRYGS DIVPFSKVDEEQMKYKSEKCF SVLGFCSSQVQRRFFMGNVLVKVAARDDEAAVA LSSLIHALDDLD MVAIVRYAYDKRANPQVGVAFPHIKHNYECLVYVQLPFMEDLRQYMFSSLKNSKKYAPTEAQLN AVDALIDSMSLAKKDEKTDLTLEDLFP TTKIPNPRFQRLFQCLLHRALHPREPLPPIQQHIWNMLNPPAEVTTKSQI PLSKIKTLFPLIEAKKKDQVTAQEIFQDNHEDGPTAK (SEQ ID No:25)

[0249] 2. 端粒酶Sm7结合基序/Sm7同源七聚体

[0250] a. Sm共有位点(单链)

[0251] 5' -AAUUUUUGGA-3' (SEQ ID NO:26)

[0252] b. 单体Sm-样蛋白(古生菌)

[0253] GSVIDVSSQRVNVQRPLDALGNSLNSPVI IKLKGDFRFGVLKSFDLHMNLVLNDAEELEDGEVTRRL GTVLIRGDNIVYISP (SEQ ID NO:27)

- [0254] 3. MS2噬菌体操纵子茎环/MS2包被蛋白
- [0255] a. MS2噬菌体操纵子茎环
- [0256] 5' -GCACAUGAGGAUCACCCAUGUGC-3' (SEQ ID NO:28)
- [0257] b. MS2包被蛋白
- [0258] MASNFTQFVLVDNNGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKG
AWRSYLNMEITPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSIY (SEQ ID NO:29)
- [0259] 4. PP7噬菌体操纵子茎环/PP7包被蛋白
- [0260] a. PP7噬菌体操纵子茎环
- [0261] 5' -AUAAGGAGUUUAUAUGGAAACCCUUA-3' (SEQ ID NO:30)
- [0262] b. PP7包被蛋白 (PCP)
- [0263] MSKTIVLSVGEATRTLTEIQSTADRQIFEEKVGPLVGRRLRTASLRQNGAKTAYRVNLKLDQADVDC
STSVCGELPKVRYTQVWSDVTIVANSTEASRKSLYDLTKSLVATSQVEDLVVNLVPLGR (SEQ ID NO:31)
- [0264] 5. SfMu Com茎环/SfMu Com结合蛋白
- [0265] a. SfMu Com茎环
- [0266] 5' -CUGAAUGCCUGCGAGCAUC-3' (SEQ ID NO:32)
- [0267] b. SfMu Com结合蛋白
- [0268] MKSIRCKNCNKLLFKADSFHDHIEIRCPRCKRHIIMLNACEHPTEKHCGKREKITHSDETVRY (SEQ ID
NO:33)
- [0269] 6. BoxB适配体/ λ N22+
- [0270] a. BoxB适配体
- [0271] 5' -GCCUGAAGAAGGGC-3' (SEQ ID NO:34)
- [0272] b. λ N22+蛋白
- [0273] MNARTRRRERRAEKQAQWKAAN (SEQ ID NO:35)
- [0274] 7. Csy4结合茎环/Csy4 [H29A]
- [0275] a. Csy4结合基序
- [0276] 5' -CUGCCGUUAAGGCAGC-3' (SEQ ID NO:36)
- [0277] b. Csy4 [H29A]
- [0278] MDHYLDIRLRPDPEFPFPAQLMSVLFGLAQAALVAQGGDRIGVSFPDLDESRSRLGERLRIHASADDLR
ALLARPWLEGLRDHLQFGEPVVPHTPYRQVSRVQAKSNPERLRRRLMRRHDLSEEEARKRIPDVARALDLPFV
TLRSQSTGQHFRFLIRHGPLQVTAEEGGFTCYGLSKGGFVVPWF (SEQ ID NO:37)
- [0279] 8. Q β 结合茎环 [Q65H]
- [0280] a. Q β 噬菌体操纵子茎环
- [0281] 5' -ATGCTGTCTAAGACAGCAT-3' (SEQ ID NO:96)
- [0282] b. Q β 包被蛋白 [Q65H]
- [0283] MAKLETVTLGNIGKDGKQTLVLNPRGVNPTNGVASLSQAGAVPALEKRVTVSVSQPSRNRKNYKVHVK
IQNPTACTANGSCDPSVTRQAYADVTFSTQYSTDEERAFVRTELAALLASPLLIDAIDQLNPAY (SEQ ID NO:
97)
- [0284] 在每个前述序列中,技术人员可以,例如,使用在结合对的一个或两个序列中具有一个或多个插入、缺失或取代的相同的一个序列或多个序列。通过非限制性示例,就结合对

的任一个或两个成员而言,可使用与前述序列至少80%、至少85%、至少90%、至少95%、或至少98%相同的序列。

[0285] 其他化学物质

[0286] 在一些实施方式中,将本发明的碱基-编辑复合物与其他化学技术组合。例如,在一些实施方式中,碱基编辑复合物还包含半胱氨酸/硒代半胱氨酸标签。在一些实施方式中,碱基编辑复合物包含用于通过点击化学进行环加成的元件或者与之结合。

[0287] 用于碱基-编辑的方法

[0288] 在另一实施方式中,本发明提供用于碱基编辑的方法。在这些方法中,技术人员将本发明的碱基编辑复合物暴露于双链DNA或者暴露于含有dsDNA的溶液或暴露于含有dsDNA的细胞或者暴露给受试者。该方法可在体外发生或者在体内或离体实施,并且可包括:将碱基编辑复合物作为用于治疗的药物的一部分递送给受试者。

[0289] 这些方法可以,例如,用于修饰选自T细胞(包括原代T细胞)、自然杀伤(NK)细胞、B细胞、或CD34+造血干祖细胞(HSPC)的免疫细胞。该免疫细胞可以是工程化免疫细胞,如包含CAR或TCR的T-细胞。因此,本发明中的方法可以应用于进一步工程化已经过修饰以包含用于治疗 CAR和/或TCR的细胞。通过进一步示例,初级免疫细胞(在宿主动物或患者内自然生成的或源自干细胞或诱导多能干细胞[iPSC]的)可使用本发明中提供的方法和复合物进行基因修饰。合适的干细胞包括但不限于:哺乳动物干细胞(如人干细胞),包括但不限于:造血、神经、胚胎、诱导多能干细胞(iPSC)、间充质、中胚层、肝、胰、肌肉、和视网膜干细胞。其他干细胞包括但不限于:哺乳动物干细胞,如小鼠干细胞,例如,小鼠胚胎干细胞。

[0290] 本发明中也提供了在原核或真核细胞中,体外、体内或离体进行基因组工程(例如,改变或操纵一个或多个基因或一种或多种基因产物的表达)的方法。具体地,本发明中提供的方法可用于在哺乳动物细胞(包括初始T细胞、自然杀伤(NK)细胞、CD34+造血干和祖细胞(HSPC),例如从脐带血或骨髓中分离的HSPC及从它们中分化的细胞)中的靶向碱基编辑中断。

[0291] 本发明中还提供的是由造血干细胞所产生的基因工程细胞,如根据本发明中所描述的方法已被修饰的T细胞。

[0292] 在一些情况下,这些方法被设定为生产由HSCs或iPSCs产生的基因工程T细胞,这些细胞适用作用于治疗用途的“普遍可接受”细胞。造血干细胞(HSC)由成血管干细胞产生,其可以产生HSC、血管平滑肌细胞和成血管细胞,它们分化成血管内皮细胞。HSCs可以产生普通的髓系和普通的淋巴样祖细胞,从中产生T细胞、自然杀伤(NK)细胞、B细胞、成髓细胞、幼红细胞、及其他参与血、骨髓、脾、淋巴结、和胸腺的细胞产生的细胞。这种方法也可以应用于自然杀伤(NK)细胞、CD34+造血干细胞和祖细胞(HSPC),如从脐带血或骨髓和细胞中分离的HSPC和从它们中分化的细胞。

[0293] 在另一个方面,本文中提供的是用于靶向疾病以用于碱基编辑修正的方法。在一些方法中,将碱基编辑复合物递送给受试者用于治疗。该靶序列可以是任何与疾病相关的多核苷酸或基因。根据本发明的内源基因序列的突变或修正的有用用途的示例包括但不限于:疾病相关基因突变的改变、编码剪接位点的序列中的改变、调控序列中的变更、导致功能获得的突变的序列中的改变、和/或导致功能丧失的突变的序列中的改变、及编码蛋白结构特征的序列的改变。

[0294] 将组分递送至细胞中

[0295] 可通过各种方法及各种形式 (DNA、RNA或蛋白) 或这些不同形式的组合将碱基编辑复合物或它们的组分递送给靶细胞和生物体。碱基编辑组分可以作为: (a) 编码用于蛋白效应物或向导RNA的相关序列的DNA多核苷酸; (b) 编码用于蛋白效应物 (信使RNA) 或向导RNA的序列的合成RNA; (c) 用于效应物的纯化蛋白, 被递送。当以蛋白形式递送时, 可以将V型Cas蛋白与向导RNA进行组装, 以形成用于递送进入靶细胞和生物体的核糖核蛋白复合物 (RNP)。

[0296] 例如, 被组装的组分或复合物可通过电穿孔、通过核转染、通过转染、利用纳米粒、利用病毒介导的RNA递送、利用非病毒介导的递送、通过细胞外囊泡 (例如, 外泌体和细胞微泡)、利用真核细胞转移 (例如, 通过重组酵母) 和其他可以包装分子使得它们可在不改变基因组景观的情况下被递送至靶标活细胞的方法一起或单独地递送。

[0297] 其他方法包括但不限于: 包含用于蛋白招募的相关序列的DNA多核苷酸的非整合性瞬时转移, 以便该分子可以转录成所需的RNA分子, 以及含有组分的氨基酸被翻译成蛋白或蛋白片断。这包括但不限于: 仅DNA载体 (例如, 质粒、微环、微载体、微串、前端粒酶产生的DNA分子 (例如Doggybone)、人工染色体 (例如, HAC和黏粒), 利用由纳米粒产生的DNA载体、细胞外囊泡 (例如, 外泌体和细胞微泡)、通过真核细胞转移 (例如, 通过重组酵母)、通过AAV的瞬态病毒转移、非整合病毒颗粒 (例如, 基于慢病毒和逆转录病毒的系统)、细胞穿透肽及其他可以介导DNA引入细胞而不直接整合到基因组景观中的技术。另一种用于引入RNA组分的方法包括使用整合基因转移技术将RNA转录系统稳定地引入至靶细胞的基因组中, 这可以通过控制组成型或启动子诱导型系统来降低RNA表达, 并且这也可以被设计成便于在满足实施后可以被该系统移除 (例如, 引入Cre-Lox重组系统); 这种用于基因转移的技术包括但不限于: 整合病毒颗粒 (例如, 基于慢病毒、腺病毒和逆转录病毒的系统)、转座酶介导的转移 (例如Sleeping Beauty转座子和Piggybac转座子)、开发DNA断裂引入的非同源修复途径 (例如, 应用CRISPR和TALEN) 技术和替代DNA分子、以及支持靶DNA整合进入目的细胞的其他技术。

[0298] 本发明的复合物的各种组分, 如果不是在细胞或溶液中酶促合成的, 则可通过化学方法制造, 或者如果是天然产生的, 则可从天然产生来源中分离和纯化。化学和酶法合成的本发明的各种实施方式的方法对于本领域技术人员是公知的。类似地, 本发明的组分之间的连接引入共价键的方法对于本领域的普通技术人员也是公知的。

[0299] 应用

[0300] 通过非限制性示例, 本发明的复合物可用于招募转录激活因子 (如p65和V64)、以及引入表观遗传修饰或影响HDR的部分。本发明的复合物也可以用于以下用途: 碱基编辑、基因组编辑、基因组筛选、治疗性细胞的生成、基因组标记、表观基因组编辑、核型工程化、染色质成像、转录物组和代谢途径工程化、遗传回路工程化、细胞信号传递感测、细胞事件记录、谱系信息重建、基因驱动、DNA基因分型、miRNA定量、体内克隆、定点突变、基因组多样化和原位蛋白质组学分析。在一些实施方式中, 将细胞或细胞群暴露于本发明的碱基编辑复合物中, 并且通过输注将细胞或细胞群引入受试者。

[0301] 应用也包括人疾病的研究, 如癌症免疫疗法、抗病毒疗法、噬菌体疗法、癌症诊断、病原体筛检、微生物群重构、干细胞重编程、免疫基因组工程化、疫苗开发、和抗体生产。

[0302] 实施例

[0303] 实施例1:用于Cas12e碱基编辑的质粒组分在哺乳动物细胞中的转染(可预见的)

[0304] 载体构建

[0305] Cas12e的编码序列可以是合成获得的,并且在具有红色荧光蛋白-嘌呤霉素融合的T2A多顺反子盒中的小鼠CMV启动子(mCMV)的控制下克隆进载体中。也可获得失活版本的Cas12e以及2xUGI融合失活Cas12e变体,并克隆到前述载体中。可获得用于MS2包被蛋白APOBEC融合(MCP-APOBEC)的编码序列,并且在小鼠CMV启动子的控制下克隆进表达载体中。含有MS2配体结合基团和的gRNA的序列在hU6启动子的控制下被克隆进表达载体中。

[0306] 转染:

[0307] 在转染前一天,可将HEK293T细胞(ATCC,#CRL-11268)以20000细胞/每孔接种于96孔板中。可使用DharmaFECT Duo转染试剂(Horizon Discovery,#T-2010)和200ng Cas12e质粒、或者与UGI融合或未与UGI融合的其失活变体、50ng MCP-APOBEC质粒、和50ng gRNA质粒对细胞进行共转染。该gRNA质粒可由101个核苷酸长度的恒定区、在PP1B或EMX1基因靶内的靶向转录物的不同间隔序列,以及在5'端、3'端、内部的不在5'或3'端的具有MS2配体结合基团或者其组合组成。

[0308] 可在含嘌呤霉素培养基中选取细胞,并在转染后48小时收获细胞用于如下所述的进一步处理,并可使用以下序列。

[0309] Cas12e gRNA序列:

[0310] 5'-GCGCACATGAGGATCACCCATGTGCGGCGCGUUUAUCCAUAUACUUUG GAGCCAGUCCCAGCGA
CUAUGUCGUAUGGACGAAGCGCUUAUUUAUCGGAGAGAAACCGAUAAAGUAAAACGCAUCAAAANNNNNNNNNNNNNN
NNNNNN-3'(SEQ ID NO:38)

[0311] 5'-GGCGCGUUUAUCCAUAUACUUUGGAGCCAGUCCCAGCGACUGCGCAC ATGAGGATCACCCATGT
GCUGUCGUAUGGACGAAGCGCUUAUUUAUCGGAGAGAAACCGAUAAAGUAAAACGCAUCAAAANNNNNNNNNNNNNN
NNNNNN-3'(SEQ ID NO:39)

[0312] 5'-GGCGCGUUUAUCCAUAUACUUUGGAGCCAGUCCCAGCGACUAUGUCG UAUGGACGAAGCGCUUA
UUUAUCGGAGAGCGCACATGAGGATCACCCATGTGCAAACCGAUAAAGUAAAACGCAUCAAAANNNNNNNNNNNNNN
NNNNNN-3'(SEQ ID NO:40)

[0313] 5'-GGCGCGUUUAUCCAUAUACUUUGGAGCCAGUCCCAGCGACUAUGUCG UAUGGACGAAGCGCUUA
UUUAUCGGAGAGAAACCGAUAAAGUAAAACGCAUCAAAANNNNNNNNNNNNNNNNNNNNNGCGCACATGAGGATCACC
CATGTGC-3'(SEQ ID No:41)

[0314] 5'-GCGCACATGAGGATCACCCATGTGCGGCGCGUUUAUCCAUAUACUUUG GAGCCAGUCCCAGCGA
CUAUGUCGUAUGGACGAAGCGCUUAUUUAUCGGAGAGCGCACATGAGGATCACCCATGTGCAAACCGAUAAAGUAAA
ACGCAUCAAAANNNNNNNNNNNNNNNNNNNNNN-3'(SEQ ID NO:42)

[0315] EMX1靶序列:

[0316] 5'-AGAACCGGAGGACAAAGTAC-3'(SEQ ID NO:43)

[0317] 5'-TGGCAATGCGCCACCGTTG-3'(SEQ ID NO:44)

[0318] 5'-TCTTCTGCTCGGACTCAGGC-3'(SEQ ID NO:45)

[0319] 5'-TCTGCTCGGACTCAGGCCCT-3'(SEQ ID NO:46)

[0320] 5'-CCAGCTTCTGCCGTTTGTAC-3'(SEQ ID NO:47)

[0321] PPIB靶序列

[0322] 5'-AAAAACAGTGGATAATTTTG-3' (SEQ ID NO:48)

[0323] 5'-GAAGAGACCAAAGATCACCC-3' (SEQ ID NO:49)

[0324] 5'-CCTCCGCTGTGGATGCTGC-3' (SEQ ID NO:50)

[0325] 5'-TCCTGCTGCTGCCGGGACCT-3' (SEQ ID NO:51)

[0326] 5'-GCGGCCGATGAGAAGAAGAA-3' (SEQ ID NO:52)

[0327] 实施例2:在哺乳动物细胞中的mRNA的电穿孔和用于Cas12e碱基编辑的向导合成(可预示的)

[0328] mRNA制备:

[0329] 可按照mRNA体外转录的标准方案,由携带T7启动子的DNA载体及用于Cas12e、dCas12e-UGI和MCP-APOBEC的编码序列制备信使mRNA。

[0330] RNA合成:

[0331] 可通过Horizon Discovery使用2'-乙酰氧基乙基原酸酯(2'-ACE)或2'-叔丁基二甲基硅基(2'-TBDMS)保护化学品合成crRNA。可通过高效液相色谱(HPLC)或聚丙烯酰胺凝胶电泳(PAGE)对RNA寡核苷酸进行2'-脱保护/脱盐和纯化。在电穿孔前,可使寡核苷酸再悬浮于10mM Tris缓冲液(pH7.5)。

[0332] 可使用Invitrogen™Neon™转染系统,10μL试剂盒,将HEK293T细胞(ATCC,#CRL-11268)进行电穿孔。可将50000细胞、1μg的Cas12e或dCas12e-UGI mRNA和MCP-APOBEC mRNA、和3μM的合成的crRNA和tracrRNA在1150V下进行电穿孔20ms和2个脉冲。化学合成的crRNA可由23个核苷酸长度的恒定区、在PPIB或EMX1基因靶内的靶向转录物的不同间隔序列,以及在5'端、3'端、内部的不在5'或3'端的具有MS2配体结合基团或其组合组成。各序列可在一个或多个连接内含有在一个或多个碱基处的化学修饰。细胞可用全血清培养基接种于96-孔板中并在72小时后收获细胞用于进一步处理。可使用以下序列。

[0333] Cas12e crRNA序列(N,靶序列):

[0334] 5'-GCGCACATGAGGATCACCCATGTGCCCGAUAAGUAAAACGCAUCAAAAGNNNNNNNNNNNNNNNNNNNNNNNN-3' (SEQ ID NO:53)

[0335] 5'-CCGAUAAGUAAAACGCAUCAAAAGNNNNNNNNNNNNNNNNNNNNNNNGCGCACATGAGGATCACCCATGTGC-3' (SEQ ID NO:54)

[0336] Cas12e tracrRNA序列(N,靶序列):

[0337] 5'-GCGCACATGAGGATCACCCATGTGCGGCGGUUUUAUCCAUAUACUUUGGAGCCAGUCCAGCGACUAUGUCGUAUGGACGAAGCGCUUAUUUAUCGGAGA-3' (SEQ ID NO:55)

[0338] 5'-GGCGGUUUUAUCCAUAUACUUUGGAGCCAGUCCAGCGACUGCGCACATGAGGATCACCCATGTGCUGUCGUAUGGACGAAGCGCUUAUUUAUCGGAGA-3' (SEQ ID NO:56)

[0339] 5'-GGCGGUUUUAUCCAUAUACUUUGGAGCCAGUCCAGCGACUAUGUCGUAUGGACGAAGCGCUUAUUUAUCGGAGAGCGCACATGAGGATCACCCATGTGC-3' (SEQ ID NO:57)

[0340] 对于实施例1和2两者:

[0341] 细胞处理

[0342] 细胞可在含有蛋白酶K(Thermo Scientific,#FERE00492)、RNase A(Thermo Scientific,#FEREN0531)和Phusion GC缓冲液(Thermo Scientific,#F-518L)的100μL缓

冲液中于56℃裂解30分钟,接着在95℃下进行5分钟热失活。此细胞裂解液可用于生成跨越含有碱基编辑位点的区域的PCR扩增物。可以Sanger测序对长度在50-100bp之间的未经纯化的PCR扩增物测序。

[0343] 编辑分析

[0344] 可使用Chimera分析工具(开源工具BEAT的改编版)来计算碱基编辑效率(Xu等人2019。BEAT:利用Python程序基于Sanger测序对碱基编辑进行定量。CRISPR杂志2,223-229)。Chimera测定编辑效率,通过首先减去背景噪声以限定样品中的预期变化。这允许了估计编辑效率而无需标准化到对照样品。接下来,Chimera使用绝对中位差(MAD)法从噪声中过滤掉任何异常值,然后评估碱基编辑器在20dp的输入向导序列的跨度上的编辑效率。

[0345] 实施例3-6

[0346] 将以下的材料和方法用于实施例3-6。

[0347] gRNA序列

[0348] 对于实施例3-5(使用Cas12b碱基编辑器),使用表3中列出的序列所编码的gRNA序列。所有gRNA的设计是基于由91nt恒定gRNA序列、靶特异性20nt间隔序列、和7nt poly-TU6终止信号组成的脂环酸芽胞杆菌(*A. acidoterrestris*) Cas12b gRNA。所有修饰都是对gRNA的恒定区进行的并且由RNA适配体发夹的内含物组成。MS2发夹序列(C5变体)的单拷贝被结合至gRNA的5'、3'、茎-环或者内部的polyU(在gRNA内的UUUUU的内部延伸)。在hU6启动子的控制下,将相关gRNA序列克隆至分离的表达载体中。

[0349] 在表3中,N表示20nt靶特异性间隔序列。如前所述的恒定gRNA序列是以粗体突出显示。MS2(C5变体)是以斜体显示,而适配体的延伸是以斜体和下划线显示。在内部polyU延伸引入的突变是以粗体和下划线显示。

[0350] 表3:使用于Cas12b的gRNA序列

gRNA 名称	sgRNA 序列 [5' → 3']	SEQ NO:	ID
脂环酸芽胞杆菌 gRNA (MS2-le ss)	GTCSTGAGGACAGAATTTTCAACGGGTGTGCCAATGGCCACTTTCCAGGTGGCAAAGCCC GTTGAGCTTCTCAAATCTGAGAAGTGGCACNNNNNNNNNNNNNNNNNNNTTTTTTT	58	
5' MS2	GGCACATGAGGATCACCCATGTGCGTCTAGAGGACAGAATTTTCAACGGGTGTGCCAATGGCCACTTTCCAGGTGGCAAAGCCC GTTGAGCTTCTCAAATCTGAGAAGTGGCACNNNNNNNNNNNNNNNNNNNTTTTTTT	59	
5' MS2 U->G	GGCACATGAGGATCACCCATGTGCGTCTAGAGGACAGAATTGTTCAACGGGTGTGCCAATGGCCACTTTCCAGGTGGCAAAGCCC GTTGAGCTTCTCAAATCTGAGAAGTGGCACNNNNNNNNNNNNNNNNNNNTTTTTTT	60	
gRNA 名称	sgRNA 序列 [5' → 3']	SEQ NO:	ID
3' MS2	GGTCTAGAGGACAGAATTTTCAACGGGTGTGCCAATGGCCACTTTCCAGGTGGCAAAGCCC GTTGAGCTTCTCAAATCTGAGAAGTGGCACGCACATGAGGATCACCCATGTGCNNNNNNNNNNNNNNNNNNNTTTTTTT	61	
3' MS2 U->G	GGTCATAGAGGACAGAATTGTTCAACGGGTGTGCCAATGGCCACTTTCCAGGTGGCAAAGCCC GTTGAGCTTCTCAAATCTGAGAAGTGGCACGCACATGAGGATCACCCATGTGCNNNNNNNNNNNNNNNNNNNTTTTTTT	62	
环 MS2	GGTCTAGAGGACAGAATTTTCAACGGGTGTGCCAATGGACATGAGGATCACCCATGTCCAGGTGGCAAAGCCC GTTGAGCTTCTCAAATCTGAGAAGTGGCACNNNNNNNNNNNNNNNNNNNTTTTTTT	63	
环 MS2 U->G	GGTCTAGAGGACAGAATTGTTCAACGGGTGTGCCAATGGACATGAGGATCACCCATGTCCAGGTGGCAAAGCCC GTTGAGCTTCTCAAATCTGAGAAGTGGCACNNNNNNNNNNNNNNNNNNNTTTTTTT	64	
polyU MS2	GGTCTAGAGGACAGAATTGCACATGAGGATCACCCATGTGCTTCAACGGGTGTGCCAATGGCCACTTTCCAGGTGGCAAAGCCC GTTGAGCTTCTCAAATCTGAGAAGTGGCACNNNNNNNNNNNNNNNNNNNTTTTTTT	65	

[0351]

[0352] 对于实施例6(使用CasMINI碱基编辑器),使用由表4中列出的序列所编码的gRNA序列。所有gRNA的设计是基于由162nt恒定gRNA序列、靶特异性23nt间隔序列、和7nt poly-U6终止信号组成的Acidibacillus sulfuroxidans Cas12fgRNA。所有修饰都是对gRNA的恒定组分进行的,并由RNA适配体发夹的内含物和茎-环的截短组成。MS2发夹序列(C5变体)的单拷贝被结合至5'、5'和茎环1的截短,茎环1延伸、茎环2的替换和茎环1的截短、或者重复:反重复的替换。在hU6启动子的控制下,将相关gRNA序列克隆至分离的表达载体中。

[0353] 在表4中,N表示23nt靶特异性间隔序列。恒定gRNA序列是以粗体突出显示。MS2 (C5变体) 是以斜体显示,而适配体的延伸是以斜体和下划线显示。设计1对应于5' MS2,设计2对应于5' MS2和茎环1截短,设计3对应于作为茎环1的延伸定位的MS2,设计4对应于作为茎环2的替换和茎环1的截短定位的MS2,设计5对应于作为重复:反重复区的替换定位的MS2。

[0354] 表4:使用于CasMINI的gRNA序列

gRNA 名称	gRNA 序列 [5' → 3']	SEQ ID NO:
A.Sulfoxidans gRNA (MS2-less)	ATGGGCTTCACTGATAAAGTGGAGAACCGCTTCACCAAAGCTGTCCCTTAGGGGATTAGA ACTTGAGTGAAGGTGGGCTGCTTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGT AACCCTCGAAACAAATTCATTTGAATGAAGGAATGCAACNNNNNNNNNNNNNNNNNNNTTTTTT	66
设计 1	<u>GCACATGAGGATCACCCATGTGCATGGGCTTCACTGATAAAGT</u> GGAGAACCGCTTCACCAAAGCTGTCCCTTAGGGGATTAGA ACTTGAGTGAAGGTGGGCTGCTTGCATCAGCCTAATGTCGA GAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAATTCAT TTGAATGAAGGAATGCAACNNNNNNNNNNNNNNNNNNNTTTTTT	67
设计 2	<u>GCACATGAGGATCACCCATGTGCCGCTTCACCAAAGCTGTCC</u> CTTAGGGGATTAGA ACTTGAGTGAAGGTGGGCTGCTTGCAT CAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCT CGAAACAAATTCATTTGAATGAAGGAATGCAACNNNNNNNN NNNNNNNNNNNNNNNTTTTTT	68
设计 3	GGG <u>GCACATGAGGATCACCCATGTGCAACCGCTTCACCAAAG</u> CTGTCCCTTAGGGGATTAGA ACTTGAGTGAAGGTGGGCTGC TTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGAGTAAC CCTCGAAACAAATTCATTTGAATGAAGGAATGCAACNNNNN NNNNNNNNNNNNNNNNNTTTTTT	69
设计 4	ACCGCTTCACGCACATGAGGATCACCCATGTGCGTGAAGGTGG GCTGCTTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGG AAAGTAACCCTCGAAACAAATTCATTTGAATGAAGGAATGC AACNNNNNNNNNNNNNNNNNNNTTTTTT	70
设计 5	GGGCTTCACTGATAAAGTGGAGAACCGCTTCACCAAAGCT GTCCCTTAGGGGATTAGA ACTTGAGTGAAGGTGGGCTGCTT GCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAA CCCTCGAAACAAAGCACATGAGGATCACCCATGTGCGGAATGC AACNNNNNNNNNNNNNNNNNNNTTTTTT	71

[0357] 质粒设计

[0358] 除了gRNA,所用系统的所有组分都被两个载体上并从CMV启动子表达。第一载体编码增强的人Apobec3A-MCP或安乐蜥 (Anolis carolinensis) Apobec1a-MCP融合蛋白(脱氨酶载体)。第二载体编码通过其C-端与UGI的两个拷贝融合的使dCas12b (D569A、E847A、

D976A) 或dCasMINI (D325A;D509A) (Cas载体)。dCasMINI是dUn1Cas12f1的前述版本。dCas12b-UGI-UGI和dCasMINI-UGI-UGI融合蛋白在Cas序列的N端和UGI序列的C端的两侧有2拷贝的SV40 NLS。此外,载体编码turboRFP的表达以允许监测转染效率。

[0359] 在hU6启动子的控制下,将相关gRNA序列克隆至分离的表达载体中(gRNA表达载体)。

[0360] 如下显示为SEQ ID No:72并用于实施例3、4和5的一个质粒,编码融合两个UGIs的失活Cas12b:dCas12b-UGI-UGI。下面使用以下的字体:

[0361] • 正常:人源化Cas12b

[0362] • 斜体:SV40 NLS

[0363] • 下划线:接头1

[0364] • 斜体和下划线:接头2

[0365] • 双下划线:UGI

[0366] • 粗体和下划线:突变c.1706A>C;c.2540A>C;c.2927A>C

[0367] SEQ ID NO:72:

ATGGCCCCAAAGAAGAAGCGGAAAGTCGCCGTGAAAAGCATTAAAGTGAAACTG
 AGGCTGGACGATATGCCCGAAATCAGAGCCGGCCTCTGGAAGCTGCACAAGGA
 GGTCAACGCCGGCGTTCGTTATTACACAGAGTGGCTGTCTTTACTCAGACAAGA
 AAATTTATATAGAAGGAGCCCCAATGGCGATGGCGAGCAAGAGTGCGACAAAA
 CCGCCGAAGAATGCAAGGCCGAAGTCTCGAAAGACTGAGAGCCAGACAAGT
 TGAGAACGGACACAGAGGACCCGCCGGCTCCGATGATGAACTGCTGCAGCTCG
 CTAGGCAACTGTACGAGCTGCTCGTCCCCAAGCTATCGGAGCTAAAGGAGACG
 CTCAGCAGATCGCTAGGAAGTTTCTGAGCCCCCTCGCCGATAAAGACGCTGTGG
 GCGGTTTAGGAATCGCCAAGGCTGGAAATAAACCTCGTTGGGTGAGGATGAGG
 GAAGCTGGCGAGCCCGGCTGGGAGGAGGAAAAGGAGAAAGCCGAAACTCGTA

[0368]

[0369]

AATCCGCCGACAGAACAGCTGACGTGCTGAGGGCCCTCGCCGACTTCGGTTTA
AAACCCCTCATGAGAGTCTATACCGACTCCGAGATGTCCAGCGTCGAGTGGAAA
CCCTTACGTAAGGGCCAAGCTGTGAGAACATGGGATCGTGATATGTTCCAGCAA
GCTATCGAAAGGATGATGAGCTGGGAGTCTTGGAATCAGAGAGTGGGCCAAGA
ATACGCCAAACTCGTGGAACAGAAGAATCGTTTCGAGCAGAAAACTTTGTGCG
GACAAGAACATTTAGTGCATCTCGTGAACCAACTGCAGCAAGATATGAAGGAG
GCCTCCCCCGGTTTAGAGAGCAAAGAGCAAACCGCTCACTATGTCACCGGTCGT
GCCCTCAGAGGCTCCGACAAAGTGTTGAAAAGTGGGGAAAGCTGGCCCCCGA
CGCTCCTTTTCGATTTATACGACGCCGAGATCAAGAACGTGCAGAGGAGAAATAC
AAGGAGGTTTGGCTCCCACGATTTATTCGCTAAACTCGCCGAGCCGAATACCA
AGCTTTATGGAGAGAGGATGCCAGCTTTCTGACTCGTTACGCCGTGTACAATTCC
ATTCTGAGAAAACCTGAACCACGCCAAAATGTTTGCCACATTCACCTTACCCGAT
GCCACCGCCCACCCTATCTGGACAAGATTCGACAAGCTCGGCGGAAACCTCCAC
CAGTATACCTTTTTATTTAACGAATTCGGAGAGAGGAGGCATGCCATTTCGTTTTTC
ACAAGTTATTAAGGTCGAGAATGGAGTCGCCAGAGAGGTGGACGACGTCACC
GTGCCTATCAGCATGAGCGAACAGCTGGACAATTTACTGCCTCGTGACCCCAAC
GAACCTATCGCCCTCTACTTCAGAGACTACGGAGCTGAGCAGCATTTCACCGGC
GAGTTCGGAGGCGCTAAGATCCAGTGTAGAAGGGATCAACTGGCCCATATGCAT
AGGAGAAGGGGCGCCAGAGATGTCTATCTGAACGTGAGCGTTCGTGTCCAAAG
CCAAAGCGAGGCCAGAGGAGAAAAGGAGACCCCCCTACGCCCGCGTCTTTAGAC
TGGTCGGCGACAATCATCGTGCTTTCGTGCACTTTGATAAGCTGTCCGACTATTT
AGCCGAACACCCCGACGATGGAAAGCTGGGCAGCGAGGGATTATTAAGCGGCC
TCAGAGTCATGAGCGTGGCTCTGGGCCTCAGAACCAGCGCCTCCATCTCCGTCT
TTAGGGTGGCCAGAAAAGACGAGCTGAAGCCCAACAGCAAGGGAAGGGTGCC
CTTTTTCTTCCCTATCAAGGGCAATGACAATTTAGTGGCCGTGCACGAAAGGTCC
CAGTTATTAAGCTGCCCGGCGAGACCGAAAGCAAAGATTTAAGGGCTATCAGA
GAGGAGAGACAGAGAACTTTAAGACAGCTGAGAACCAGCTGGCTTATCTGAG
ATTATTAGTCAGATGTGGCAGCGAGGACGTCCGGTCGTAGAGAGAGGAGCTGGG
CCAAGCTGATTGAACAACCCGTTGATGCCGCTAATCACATGACCCCGATTGGA
GGGAAGCTTTTCGAGAACGAGCTGCAGAACTGAAGTCTTTACACGGCATTTCG
AGCGACAAGGAGTGGATGGACGCCGTGTACGAGTCCGTGAGAAGGGTGTGGA
GGCACATGGGAAAGCAAGTTAGGGATTGGAGGAAAGATGTGAGGTCCGGCGAA
AGACCCAAGATCAGAGGCTACGCCAAGGACGTGGTCCGAGGAAACTCCATCGA
GCAGATCGAGTACCTCGAGAGACAATAAAGTTTTTAAAGTCTTGGTCCTTCTTC
GGCAAGGTCAGCGGCCAAGTCATTCGTGCTGAAAAGGGATCTCGTTTCGCCATC
ACACTGAGAGAGCACATTGACCACGCCAAAGAGGATCGTCTGAAAAAAGTCTGC
CGATCGTATCATTATGGAGGCCCTCGGCTATGTCTATGCTCTGGACGAGAGAGGC
AAGGGAAAATGGGTCGCCAAGTATCCCCCTTGTCAACTGATTTTATTAGCGGAG

CTGTCCGAGTACCAATTTAACAACGATAGGCCTCCCTCCGAGAATAACCGAGCTC
 ATGCAGTGGAGCCATCGTGGCGTGTTC AAGAACTGATCAATCAAGCTCAAGTT
 CACGATTTACTCGTGGGCACCATGTACGCCGCTTTTAGCTCCAGATTCGACGCTA
 GGACCGGCGCCCCGGTATTAGGTGTAGAAGAGTGCCCGCTCGTTGCACCCAAG
 AACATAACCCCGAACCCTTTCCTTGGTGGCTGAATAAGTTCGTCGTCGAGCACA
 CCCTCGACGCTTGCCCTTTACGTGCCGACGACCTCATCCCTACTGGTGAAGGCG
 AGATCTTTGTGAGCCCTTTCAGCGCTGAGGAAGGCGATTTTCACCAGATCCACG
 CCGCTCTGAACGCCGCCAGAATCTGCAACAGAGGCTCTGGAGCGACTTCGATA
 TTTCCCAGATTCGTCTGAGGTGCGATTGGGGAGAGGTGGATGGAGAACTCGTCC
 TCATCCCCAGACTGACTGGTAAGAGGACCGCTGACAGCTATTCCAATAAGGTCT
 TCTATACCAATACTGGTGCACCTACTACGAGAGAGAGAGGGGCAAGAAGAGA
 AGGAAAGTCTTCGCCCAAGAGAAGCTGTCCGAGGAGGAGGCCGAATTATTAGT
 CGAGGCTGATGAGGCTCGTGAAAAGTCCGTGGTTTTAATGAGGGACCCCTCCGG
 CATCATCAACAGAGGCAATTGGACTCGTCAGAAGGAATTCTGGAGCATGGTGAA
 TCAGAGGATCGAGGGCTATCTGGTGAAGCAGATTAGATCTCGTGTGCCTCTGCA
 AGATAGCGCTTGTGAGAATACCGGAGATATTAGCGGCGGGAGCGGCGGGAGCG
GGGGGAGCACTAATCTGAGCGACATCATTGAGAAGGAGACTGGGAAACAGCTG
GTCATTCAGGAGTCCATCCTGATGCTGCCTGAGGAGGTGGAGGAAGTGATCGGC
AACAAGCCAGAGTCTGACATCCTGGTGCACACCGCCTACGACGAGTCCACAGA
TGAGAATGTGATGCTGCTGACCTCTGACGCCCCCGAGTATAAGCCTTGGGCCCT
GGTCATCCAGGATTCTAACGGCGAGAATAAGATCAAGATGCTGAGCGGAGGATC
CGGAGGATCTGGAGGCAGCACCAACCTGTCTGACATCATCGAGAAGGAGACAG
GCAAGCAGCTGGTCATCCAGGAGAGCATCCTAATGCTTCCCGAAGAAGTCGAA
GAAGTGATCGGAAACAAGCCTGAGAGCGATATCCTGGTCCATACTGCGTATGAT
GAAAGTACCGACGAAAACGTAATGCTACTCACATCCGACGCCCCAGAGTATAAG
CCCTGGGCTCTAGTTATAACAAGACTCCAACGGAGAGAACAAAATCAAATGCTG
TCTGGCGGCTCAAAAAGAACC GCCGACGGCAGCGAATTCGAGCCCAAGAAGAAGAG
GAAAGTCTAA

[0370]

[0371] SEQ ID No:72的对应氨基酸序列如下显示为SEQ ID No:73,并在该序列中使用以下的字体:

- [0372] • 正常:Cas12b
- [0373] • 斜体:SV40 NLS
- [0374] • 下划线:接头1
- [0375] • 斜体和下划线:接头2
- [0376] • 双下划线:UGI
- [0377] • 粗体和下划线:突变D569A、E847A、D976A

[0378] SEQ ID NO:73:

[0379] MAPKKKRKVAVKSIKVKLRLDDMPEIRAGLWKLHKEVNAGVRYYTEWLSLLRQE

NLYRRSPNGDGEQECDKTAEECKAELLERLRARQVENGHRGPAGSDDELLQLARQ
 LYELLVPQAIGAKGDAQQIARKFLSPLADKDAVGGGLIAKAGNKPRWVRMREAGE
 PGWEEKEKAETRKSADRTADVLRALADDFGLKPLMRVYTDSEMSSVEWKPLRKG
 QAVRTWDRDMFQQAIERMMSWESWNQRVGQEYAKLVEQKNRFEQKNFVGQEHL
 VHLVNQLQQDMKEASPGLESKEQTAHYVTGRALRGSDKVFKEKWGKLAPDAPFDL
 YDAEIKNVQRRNTRFRGSHDLFAKLAEPEYQALWREDASFLTRYAVYNSILRKLNH
 AKMFATFTLPDATAHPIWTRFDKLGGNLHQYTFLFNEFGERRHAIRFHKLLKVENG
 VISVDDVTVPISMSEQLDNLLPRDPNEPIALYFRDYGAEQHFTGEFGGAKIQCRRDQ
 LAHMHRRRGARDVYLVNSVVRVQSQSEARGERRPPYAAVFRLVGDNHRAVHFVFDK
 LSDYLAEHPDDGKLGSEGLLSGLRVMSVALGLRTSASISVFRVARKDELKPNKSKGR
 VFFFPIKGNNDNLVAVHERSQLLKLPGETESKDLRAIREERQRTLRLQRTQLAYLRL
 VRCGSEDVGRRERSWAKLIEQPVDAANHMTPDWREAFENELQKLKSLHGICSDKE
 [0380] WMDAVYESVRRVWRHMGKQVRDWRKDVRSGERPKIRGYAKDVVGGNSIEQIEY
 LERQYKFLKSWSFFGKVSGQVIRAEKGSRAITLREHIDHAKEDRLKKLADRIIMEA
 LGYVYALDERGKGKWWAKYPPCQLILLAELSEYQFNDRPPSENNQLMQWSHRG
 VFQELINQAQVHDLVGTMYAAFSSRFDARTGAPGIRCRRVPARCTQEHNPPEFPW
 WLNKFVVEHTLDACPLRADDLIPTGEGEIFVSPFSAEEGDFHQIHAALNAAQNLLQQ
 RLWSDFDISQIRLRCDWGEVDGELVLIPRLTGKRTADSYSNKVFYNTGTVYYERE
 RGKKRRKVFAQEKLSEEEAELLVEADEISKSVVLMRDPGIIINRGNWTRQKEFWS
 MVNQRIEGLYLVKQIRSRVPLQDSACENTGDISGGSGGSGGSTNLSDIIEKETGKQLVI
QESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDS
NGENKIKMLSGGSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESD
ILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSKRTADGSEF
EPKKRRKV

[0381] 如下显示为SEQ ID No:74并用于实施例6的第二质粒,编码融合到两个UGI的失活CasMINI:dCasMINI-UGI-UGI。下面使用以下的字体:

- [0382] ●正常:人源化CasMINI
- [0383] • 斜体:SV40 NLS
- [0384] • 下划线:接头1
- [0385] • 斜体和下划线:接头2
- [0386] • 双下划线:UGI
- [0387] • 粗体和下划线:突变c.974A>C;c.975T>A;c.1526A>C;c.1527T>G

[0388] SEQ ID NO:74:
 ATGGCCCCAAGAAAAACGCAAGGTGGCCAAAAACACCATTACCAAACACTG
 AAAGTGCATATTGTGCGTCCGTATAATAGCGCAGAAAGTGGAATAATTGTTGCC
 [0389] GACGAAAAAACAACCGCGAAAAAATCGCACTGGAAAAGAACAAGACAAAG
 TGAAAGAAGCCTGCAGCAACATCTGAAAGTTGCAGCATATTGTACCACACAG
 GTTGAACGTAATGCATGCCTGTTTTGTAAAGCACGTAAACTGGATGACAAATTCT

ACCAAAACTGCGTGGTCAGTTTCCGGATGCAGTTTTTTGGCAAGAAATCAGCG
 AAATTTTTCGCCAGCTGCAGAAACAGGCAGCAGAAATCTATAATCAGAGCCTGA
 TCGAACTGTACTACGAGATTTTTATCAAAGGCAAAGGTATTGCAAATGCCAGCA
 GCGTTGAACATTATCTGAGTAGAGTTTGTATAGACGTGCAGCAGAACTGTTTAA
 AAACGCAGCAATTGCAAGCGGTCTGCGTAGCAAATCAAAGCAATTTTCGTCT
 GAAAGAACTGAAAAACATGAAAAGTGGTCTGCCGACCACCAAAGCGATAATT
 TTCCGATTCCGCTGGTTAAACAGAAAGGTGGTCAGTATACCGGTTTTGAAATTAG
 CAATCATAATAGCGACTTCATCATCAAGATTCCGTTTGGTCGTTGGCAGGTCAAA
 AAAGAGATTGATAAATATCGTCCGTGGGAGAAATTTGACTTTGAACAGGTTTCAG
 AAAAGCCCGAAACCGATTAGCCTGCTGCTGAGCACCCAGCGTCGTAAACGTAAT
 AAAGGTTGGAGCAAAGATGAAGGCACCGAAGCCGAAATCAAAAAGTTATGAA
 TGGCGATTATCAGACCAGCTACATTGAAGTTAAACGTGGCAGCAAATCTGTGA
 AAAAAGCGCATGGATGCTGAATCTGAGCATTGATGTTCCGAAAATTGATAAAGG
 TGTGGATCCGAGCATTATTGGTGGTATTGCAGTTGGTGTAGATCACCGCTGGTT
 TGCGCAATTAACAATGCATTTAGCCGTTATAGCATCAGCGATAACGACCTGTTTC
 ACTTCAACAAGAAAATGTTTGCACGTCGTCGTATCCTGCTGAAAAAAAACCGTC
 ATAAACGTGCAGGTCATGGTGCAAAAAACAACTGAAACCGATCACCATCTGA
 CCGAAAAAAGTGAACGTTTTTCGCAAAAAGCTGATTGAACGTTGGGCATGTGAA
 [0390] ATCGCGGATTTCTTCATTA AAAACAAAGTTGGCACCGTGCAGATGGAAAATCTG
 GAAAGCATGAAACGTAAAGAGGACAGCTATTTTAACATTTCGCTGCGTGGCTTT
 TGGCCGTATGCAGAAATGCAGAACAAAATCGAATTCAACTGAAGCAGTATGGC
 ATCGAAATTCGTAAAGTTGCACCGAATAATACCAGCAAAACCTGTAGCAAATGT
 GGCCATCTGAACAAC TATTTCAACTTCGAGTACCGCAAGAAAAACAAATTCCCG
 CACTTTAAATGCGAAAAATGCAACTTCAAAGAAAACGCCGCGTATAATGCAGCC
 CTGAATATTTCAAACCCGAAACTGAAAAGCACCAAGAGAGACCGAGCGGCGG
GAGCGGCGGGAGCGGGGGGAGCACTAATCTGAGCGACATCATTGAGAAGGAGA
CTGGGAAACAGCTGGTCATT CAGGAGTCCATCCTGATGCTGCCTGAGGAGGTGG
AGGAAGTGATCGGCAACAAGCCAGAGTCTGACATCCTGGTGCACACCGCCTAC
GACGAGTCCACAGATGAGAATGTGATGCTGCTGACCTCTGACGCCCCGAGTAT
AAGCCTTGGGCCCTGGTCATCCAGGATTCTAACGGCGAGAATAAGATCAAGATG
CTGAGCGGAGGATCCGGAGGATCTGGAGGCAGCACCAACCTGTCTGACATCAT
CGAGAAGGAGACAGGCAAGCAGCTGGTCATCCAGGAGAGCATCCTGATGCTGC
CCGAAGAAGTCGAAGAAGTGATCGGAAACAAGCCTGAGAGCGATATCCTGGTC
CATACCGCCTACGACGAGAGTACCGACGAAAATGTGATGCTGCTGACATCCGAC
GCCCCAGAGTATAAGCCCTGGGCTCTGGTCATCCAGGATTCCAACGGAGAGAAC
AAAATCAAATGCTGTCTGGCGGCTCAAAAAGAACC GCCGACGGCAGCGAATTCG
AGCCCAAGAAGAAGAGGAAAGTCTAA

[0391] SEQ ID NO:74的对应氨基酸序列如下显示为SEQ ID No:75,并在该序列中使用以下的字体:

[0392] • 正常:CasMINI

- [0393] • 斜体:SV40 NLS
- [0394] • 下划线:接头1
- [0395] • 斜体和下划线:接头2
- [0396] • 双下划线:UGI
- [0397] • 粗体和下划线:突变D569A、E847A、D976A

[0398] SEQ ID NO:75为:

MAPKKRKRVAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKE
 ACSKHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEIFR
 QLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELFKNAAIAS
 GLRSKIKSNFRLKELKNMKSGLPPTKSDNFPIPLVKQKGGQYTGFEISNHNSDFIIKI
 PFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQRRKRNKGWSKDEGTEA
 EIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSIDVPKIDKGVDPSSIIGGIAVGVR
 SPLVCAINNAFSRYSISDNDLFHFNKKMFARRRILLKKNRHKRAGHGAKNKLKPITI
 LTEKSERFRKKLIERWACEIADFFIKNKVGTVQMENLESMKRKEDSYFNIRLRGFW
 PYAEMQNKIEFKLKQYGIEIRK VAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFK
 CEKCNFKENAAYNAALNISNPKLKSTKERPSGGSGGSGGSTNLSDIIEKETGKQLVI
QESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDS
NGENKIKMLSGGSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESD
ILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSKRTADGSEF
EPKKRKRKV

[0400] 如下显示为SEQ ID No:76的第三质粒,编码融合到MCP、增强的人Apobec3A-MCP的脱氨酶。下面使用以下的字体:

- [0401] • 正常:人Apobec3A
- [0402] • 斜体:SV40 NLS
- [0403] • 下划线:L25-接头
- [0404] • 双下划线:MCP
- [0405] • 粗体和下划线:突变c.307T>G;c.308G>C;c.391T>G

[0406] SEQ ID NO:76:

5'-ATGGCCCCAAGAAGAAGCGGAAAGTGGAAGCCAGCCCAGCATCCGGGCC
 CAGACACTTGATGGATCCACACATATTCACCTTCAACTTTAACAATGGCATTGGA
 AGGCATAAGACCTACCTGTGCTACGAAGTGGAGCGCCTGGACAATGGCACCTCG
 GTCAAGATGGACCAGCACAGGGGCTTTCTACACAACCAGGCTAAGAATCTTCTC
 TGTGGCTTTTACGGCCGCCATGCGGAGCTGCGCTTCTTGGACCTGGTTCCCTTCTT
 TGCAGTTGGACCCGGCCAAATCTACAGGGTCACTTGGTTCATCTCCTGGAGCC
 CCTGCTTCTCCGCGGGCTGTGCCGGGGAAGTGC GTGCGTTCCTTCAGGAGAAC
 ACACACGTGAGACTGCGTATCTTCGCTGCCCGCATCTATGATGACGACCCCCTAT

[0407]

[0408] ATAAGGAGGCACTGCAAATGCTGCGGGATGCTGGGGCCCAAGTCTCCATCATGA
CCTACGATGAATTTAAGCACTGCTGGGACACCTTTGTGGACCACCAGGGATGTC
CCTTCCAGCCCTGGGATGGACTAGATGAGCACAGCCAAGCCCTGAGTGGGAGG
CTGCGGGCCATTCTCCAGAATCAGGGAAACGAGCTGAAGACACCCCTGGGCGA
CACCACACACACCTCTCCACCTTGCCCAGCACCAGAGCTGCTGGGAGGCCCTAT
GGCCAGCAACTTCACACAGTTTGTGCTGGTGGATAATGGAGGAACCGGCGACG
TGACAGTGGCACCATCTAACTTTGCCAATGGCATCGCCGAGTGGATCAGCTCCA
ACTCTCGGAGCCAGGCCTATAAGGTGACCTGTAGCGTGCGGCAGTCTAGCGCCC
AGAATAGAAAGTATAACAATCAAGGTGGAGGTGCCTAAGGGCGCCTGGAGATCCT
ACCTGAACATGGAGCTGACCATCCCAATCTTTGCCACAAATTCTGATTGCGAGC
TGATCGTGAAGGCCATGCAGGGCCTGCTGAAGGACGGCAACCCTATCCCAAGC
GCCATCGCCGCCAATAGCGGAATCTACTGA-3'

[0409] SEQ ID NO:76的对应氨基酸序列如下显示为SEQ ID No:77,并在该序列中使用以下的字体:

- [0410] • 正常:人Apobec3A
- [0411] • 斜体:SV40 NLS
- [0412] • 下划线:L25-接头
- [0413] • 双下划线:MCP
- [0414] • **粗体和下划线**:突变W103A,Y131D
- [0415] SEQ ID NO:77:

[0416] MAPKKKRKVEASPASGPRHLMDPHIFTSNFNNGIGRHKTYLCYEVEERLDNGTS
VKMDQHRGFLHNQAKNLLCGFYGRHAELRFLDLVPSLQLDPAQIYRVTFISWSP
CFSAGCAGEVRAFLQENTHVRLRIFAARIYDDDPLYKEALQMLRDAGAQVSIMTY
DEFKHCWDTFVDHQGCPFQPWDGLDEHSQALSGRLRAILQNQGNELKTPLGDTT
HTSPPCPAPELLGGPMASNFTQFVLVDNNGGTGDVTVAPSNFANGIAEWISSNSRSQA
YKVTCVSRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELVKAMQG
LLKDGNIPIPSAIAANSGIY

[0417] 如下显示为SEQ ID NO:78的第四质粒,编码融合到MCP、安乐蜥Apobec1a-MCP的脱氨酶。下面使用以下的字体:

- [0418] • 正常:安乐蜥属Apobec1a
- [0419] • 斜体:SV40 NLS
- [0420] • 下划线:L25-接头
- [0421] • 双下划线:MCP
- [0422] SEQ ID NO:78:

[0423] 5'-ATGGCCCCAAGAAGAAGCGGAAAGTGGGGTATCAGGCTGCAATTCTATTA
TCAAATTTGTTCTTCAGGTGGCAAATGGAACCAGAGGCGTTTCAGAGGAATTTT
GATCCCAGAGAATTTCCCGAGTGTACTTTACTGCTGTATGAAATCCACTGGGATA
ACAACACCAGTAGGAACTGGTGTACAAACAAACCTGGCCTCCATGCTGAAGAA
AATTTTTTGCAAATTTTTAATGAGAAAATAGATATCAGGCAGGACACACCATGCT
CTATCACTTGGTTTCTGTCTTGGAGTCCTTGTTATCCATGCAGCCAGGCTATAATT

AAGTTCTTGGAAGCACACCCTAACGTGAGCCTGGAGATAAAAGCTGCTCGGCT
 GTACATGCATCAAATCGACTGTAACAAGGAAGGCCTCAGGAATTTAGGTAGAAA
 TAGAGTTTCTATCATGAATCTACCTGATTATCGCCACTGTTGGACAACATTTGTGG
 TTCCTAGAGGGGCAAATGAAGATTATTGGCCACAGGATTTCTTACCAGCCATAAC
 AAATTATTCCAGGGAACCTTGACTCAATTCTTCAGGACGAGCTGAAGACACCCCT
 GGGCGACACCACACACACCTCTCCACCTTGCCCAGCACCAGAGCTGCTGGGAG
 [0424] GCCCTATGGCCAGCAACTTCACACAGTTTGTGCTGGTGGATAATGGAGGAACCG
GCGACGTGACAGTGGCACCATCTAACTTTGCCAATGGCATCGCCGAGTGGATCA
GCTCCAACTCTCGGAGCCAGGCCTATAAGGTGACCTGTAGCGTGCGGCAGTCTA
GCGCCCAGAATAGAAAGTATAACAATCAAGGTGGAGGTGCCTAAGGGCGCCTGG
AGATCCTACCTGAACATGGAGCTGACCATCCCAATCTTTGCCACAAATTCTGATT
GCGAGCTGATCGTGAAGGCCATGCAGGGCCTGCTGAAGGACGGCAACCCTATC
CCAAGCGCCATCGCCGCCAATAGCGGAATCTACTGA-3'

[0425] SEQ ID NO:78的对应氨基酸序列如下显示为SEQ ID No:79,并在该序列中使用以下的字体:

[0426] • 正常:安乐蜥属Apobec1a

[0427] • 斜体:SV40NLS

[0428] • 下划线:L25-接头

[0429] • 双下划线:MCP

[0430] SEQ ID NO:79:

MAPKKKRKVGYYQAAILLSNLFFRWQMEPEAFQRNFDPRFPECTLLLYEIH
 DNNTSRNWCTNKPGLHAEENFLQIFNEKIDIRQDTPCSITWFLSWSPCYPCSQAIKF
 LEAHPNVSLEIKAARLYMHQIDCNKEGLRNLGRNRV SIMNLPDYRHCWTTFFVPR
 [0431] GANEDYWPQDFLPAITNYSRELDSILQDELKTPLGDTTHTSPPCPAPELLGGPMASN
FTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYT
IKVEVPKGAWRSYLNMEITIPFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIY

[0432] 该碱基编辑系统的gRNA组分在分离的载体上表达,表达是由RNA聚合酶III U6启动子(gRNA表达载体)驱动.gRNA以单个单元(包含通过如前所述的人造四环连接的脂环酸芽孢杆菌Cas12b或Acidibacillus sulfuroxidansCas12f的crRNA和tracrRNA组分)表达。每个Cas蛋白的gRNA靶序列列表如表5所示。

靶名称	靶序列[5' → 3']	SEQ ID NO:
Cas12b		
位点 2-sgRNA1	TGTTCCAGTTTCCTTTACAG	80
[0433] 位点 2-sgRNA2	AGGCTGGCCCGCCCCGAGT	93
位点 2-sgRNA4	CAGCCCGCTGGCCCTGTAAA	81
VEGFA-sgRNA1	GCCAGAGCCGGGGTGTGCAG	82
VEGFA-sgRNA6	GGAAGTGTCCAGGGATGCTT	83

[0434]	FANCF-sgRNA3	TCGCGCCCTCCCAGCCGGGC	84
	B2M-sgRNA2	CCGATATTCCTCAGGTACTC	94
	B2M-sgRNA3	AGGTTTACTCACGTCATCCA	95
	CasMINI		
	VEGFA-sgRNA1	CTCCTGGACCCCCTATTTCTGAC	85
	VEGFA-sgRNA2	GCCAGAGCCGGGGTGTGCAGACG	86

[0435] 表5. 用于碱基编辑的gRNA靶位点序列

[0436] 细胞培养和转染

[0437] HEK293细胞在添加10%胎牛血清 (FBS) 和100U mL⁻¹青霉素/链霉素的Dulbecco改良Eagle培养基 (DMEM) 中生长。在转染前24小时, 将10000个细胞接种到96孔板的单孔中, 至实现~70%的转染融合率。24小时后, 用200ng质粒DNA (75ng Cas载体、75ng脱氨酶载体、和50ng gRNA表达载体) 脂质转染细胞, 96孔板每孔0.7μL的DharmaFECT DUO (Horizon discovery)。

[0438] 细胞裂解

[0439] 转染后72小时, 除去培养基, 并用PBS洗涤细胞1x, 将50μL的TrypLE表达酶 (ThermoFisher scientific) 添加到各孔中。在使细胞裂解后, 添加100μL的新鲜DMEM并将20μL的重悬浮细胞转移至96孔板, 并用60μL的引导PCR裂解试剂 (Viagen biotech) 在以下的条件下培养: 55°C 45分钟, 接着95°C 15分钟, 然后在-20°C下保存细胞裂解液。

[0440] 靶向区的PCR扩增

[0441] 使用引导PCR裂解试剂获得每个PCR反应中所用的1μL的细胞裂解液。为了NGS分析, 将Q5高保真2x预混液 (NEB) 用于sgRNA靶位点的扩增, 反应混合体系设置如下:

试剂	体积
Q5 2x 预混液	12.5 μL
正向引物 (10 μM)	12.5 μL
[0442] 反向引物 (10 μM)	12.5 μL
细胞裂解液	1.0 μL
无核酸酶水	9.0 μL
总计	25 μL

[0443] PCR反应在如下确定的热循环条件下进行: :

步骤	温度	时间
[0444] 预变性	98°C	30 秒
30 次循环	98°C	10 秒
	64°C/68°C	30 秒
[0445]	72°C	20 秒
最后延伸	72°C	2 分钟

[0446] 对于Sanger测序分析,使用GoTaq热启动聚合酶(Promega)对目的区进行PCR-扩增。将反应混合体系设置如下:

	试剂	体积
[0447]	5xGoTaq 缓冲液	5 μ L
	MgCl ₂ (25 mM)	2 μ L
	dNTPs (100 mM)	0.1 μ L
	正向引物 (10 μ M)	1.25 μ L
	反向引物(10 μ M)	1.25 μ L
	无核酸酶水	15.4 μ L
	总计	25 μ L

[0448] 表4中详细说明了所使用的引物。

[0449] 表6:引物[5'→3']

[0450]	位点 2-F	AGGACGTCTGCCCAATATGT	SEQ ID NO: 87
	位点 2-R	CCAAGTGAGAAGCCAGTGGA	SEQ ID NO: 88
	VEGFA-F	TCTTCCCTCCCAGTCACTGA	SEQ ID NO: 89
	VEGFA-R	TCACTCTCGAAGACGCTGCT	SEQ ID NO: 90
	FANCF-F	CGATGAGGAGACACTCCAAGAG	SEQ ID NO: 91
	FANCF-R	GCCTTTCTGAAGGTCATAGTGC	SEQ ID No: 92

[0451] 碱基编辑分析

[0452] 提交PCR产物进行Sanger测序(Genewiz)。通过专有的内部软件(Chimera公司)对分析数据。

[0453] 实施例3:使用Cas12b的碱基编辑

[0454] 在此实施例中,位点2扩增区的下一代测序分析显示通过在HEK-293T细胞系中引入Cas12b碱基编辑器而引入的特异性C转换为T。使用了两个不同的效应物,hApobec3A(Apobec3A-MCP)或者AnoApobec(安乐蜥Apobec1a-MCP),它们分别是通过转染的表达在SEQ ID No:76和SEQ ID No:78中突出显示的序列的质粒引入。同时地,dCas12b-UGI-UGI通过转染的表达在SEQ ID No:72中突出显示的序列的质粒引入。gRNAs通过转染的表达对应于SEQ ID No:58至65的gRNA主链的质粒引入。靶位点序列对应于SEQ ID No:80。通过DNA质粒脂转染递送组分,随后在转染后72小时裂解细胞,其后通过PCR扩增靶基因座并通过二代测序进行分析。

[0455] 结果分别报告于图8A和图8B中。这些图显示了在HEK-293细胞中通过DNA质粒转染传递碱基编辑组分所实现的C到T的转换水平。X-轴代表gRNA中MS2适配体的不同位置,并对

应于在表3中突出显示的序列。U->G表示存在Cas12b gRNA支架中的内部的polyU延伸的变化,引入该变化以提高在hU6启动子下转录的效率。Y轴显示了在通过NGS测定前间隔区内在特定C残基处的C向T的编辑百分率。各柱代表靶向的前间隔区内的靶C残基,图示中的数字代表相关C残基的位置,1为PAM近端,20为PAM远端。No-tf=无转染对照。hApobec3A=Apobec3A-MCP;AnoApobec=安乐蜥Apobec1a-MCP。误差棒代表来自2个重复的平均值的标准误差。

[0456] 图8A和图8B表明Cas12b碱基编辑器在HEK-293T细胞系中具有功能,在位点2用hApobec3A和AnoApobec两者都观察到编辑。两个gRNA设计为功能性的,5' MS2设计和PolyU-MS2设计两者都观察到编辑。此外,在不存在MS2适配体(MS2-less)情况没有明显的编辑,表明所观察到的编辑是依赖于通过MCP配体与MS2配体结合基团之间的的相互作用和随后的碱基编辑复合物的组装来募集脱氨酶。与用Cas9碱基编辑器观察到的相反,编辑是在宽窗口上观察到的,在14nt窗口上编辑多达5个C残基。移位后的编辑窗口突出显示了该工具的普遍适用性以及用于碱基编辑的扩大的可靶向位点范围的潜力。

[0457] 实施例4:在多个位点中的碱基编辑

[0458] 本实施例证明了Cas12b碱基编辑器在多个基因组基因座和序列背景中具有功能。碱基编辑分析表明Cas12b碱基编辑器在HEK-293T细胞系的不同的位点中引入C到T的转变。sgRNAs靶向了七个不同的区:VEGFA_sgRNA1、SEQ ID No:82;VEGFA_sgRNA6、SEQ ID No:83;FANCF_sgRNA3、SEQ ID No:84;位点2_sgRNA2、SEQ ID No:93;位点2_sgRNA4、SEQ ID No:81;B2M_sgRNA2、SEQ ID No:94;和B2M_sgRNA3、SEQ ID No:95。使用两种不同的效应物(hApobec3A或者AnoApobec),它们分别是通过转染表达对应于SEQ ID No:76和SEQ ID No:78的序列的质粒引入。Cas12b是通过转染对应于SEQ ID No:72的质粒引入。gRNA(SEQ ID No:59)的5' MS2版本被用于所有条件。靶位点序列对应于SEQ ID No:93-95和81-84。通过DNA质粒脂转染递送组分,随后在转染后72小时裂解细胞,其后通过PCR扩增靶基因座并通过Sanger测序进行分析。间隔序列示于表5,扩增编辑区的引物示于表6。

[0459] 图9A(VEGFA_sgRNA1)、图9B(VEGFA_sgRNA6)、图9C(FANCF_sgRNA3)、图9D(位点2_sgRNA2)、图9E(位点2_sgRNA4)、图9F(B2M_sgRNA2)、和图9G(B2M_sgRNA3)。在HEK-293细胞中通过DNA质粒转染在多个基因组位点和序列背景中传递碱基编辑组分实现的C到T的转换水平。X轴代表所用的试剂(与dCas12b-UGI-UGI结合);hApobec3A=5' MS2 gRNA+Apobec3A-MCP;AnoApobec=5' MS2 gRNA+安乐蜥Apobec1a-MCP;MS2less-hApobec3A=除了gRNA不含有MS2适配体外,如hApobec3A;AnoApobec=除了gRNA不含有MS2适配体外,如AnoApobec;no-tf=无转染对照。Y轴显示了通过Sanger测序和随后的Chimera分析测量的前间隔区内特定C残基处的C向T的编辑百分率。各柱代表靶向的前间隔区内的靶C残基,图示中的数字代表相关C残基的位置,1为PAM近端,20为PAM远端。误差棒代表来自2个重复的标准误差。

[0460] 图9A-9G证明了Cas12b碱基编辑器作为通用碱基编辑工具是可行的,在多个基因组基因座和序列背景下观察到编辑。同样,在前间隔区内的多个Cs残基处观察到编辑,且编辑窗口的尺寸和位置与Cas9碱基编辑系统所观察的尺寸和位置不同,因此进一步增强了此工具的潜在独特功能应用。此外,在不存在MS2适配体(MS2-less)情况下,观察到很少甚至无编辑,表明所观察到的编辑是依赖于通过在MCP配体与MS2配体结合基团之间的相互作用和随后的碱基编辑复合物的组装来募集脱氨酶。

[0461] 实施例5:不同细胞系中的碱基编辑

[0462] 本实施例证明了Cas12b碱基编辑器在不同的细胞系中是有功能的,Sanger测序和Chimera分析显示Cas12b碱基编辑器在U2OS细胞系中引入了C到T的转变。显示了一种sgRNA靶向的一个区域:VEGFA_sgRNA1、SEQ ID No:83。使用两种不同的效应物,hApobec3A或AnoApobec,它们分别是通过质粒的转染引入,这些质粒表达对应于SEQ ID No:76和SEQ ID No:78的序列。通过质粒的转染引入Cas12b,该质粒表达对应于SEQ ID No:72的序列。表达sgRNA 5' MS2版本(SEQ ID No:59)的质粒的转染被用于所有条件。通过DNA质粒脂转染递送组分,随后在转染后72小时裂解细胞,其后通过PCR扩增靶基因座并通过Sanger测序进行分析。

[0463] 图10示出了在U2OS细胞中通过DNA质粒转染递送碱基编辑组分所实现的C到T的转换水平。显示了向导VEGFA_sgRNA1、SEQ ID No:83的数据。X轴代表所用的试剂(与dCas12b-UGI-UGI结合);hApobec3A=5' MS2 gRNA+Apobec3A-MCP;AnoApobec=5' MS2 gRNA+安乐蜥Apobec1a-MCP;MS2less-hApobec3A=除了gRNA不含有MS2适配体外,如hApobec3A;AnoApobec=除了gRNA不含有MS2适配体外,如AnoApobec;no-tf=无转染对照。Y轴显示了通过Sanger测序和随后的Chimera分析测量的前间隔区内特定C残基处的C向T的编辑百分率。各柱代表靶向的前间隔区内的靶C残基,图示中的数字代表相关C残基的位置,1为PAM近端,20为PAM远端。误差棒代表来自2个重复的标准误差。

[0464] 图10表明Cas12b碱基编辑器在U2OS细胞系中对于相同的sgRNAs是有功能的(如图9A中),进一步证明了碱基编辑工具的普遍适用性。编辑是用脱氨酶所观察的和如在HEK-293细胞中所看见的,编辑在宽窗上方是显而易见的。在两种脱氨酶中都观察到编辑,如在HEK-293细胞中观察到的,编辑在宽窗口内是明显的。

[0465] 实施例6:使用CasMINI的碱基编辑

[0466] 本实施例证明了gRNA-配体碱基编辑系统可以应用于不同于Cas12b的其他V型酶。此处展示了CasMINI碱基编辑器在HEK-293T细胞的不同位点中引入C到T的转换。使用两种不同的效应物,hApobec3A或者AnoApobec,它们分别是通过表达对应于SEQ ID No:76和SEQ ID No:78序列的质粒的转染引入。CasMINI通过质粒的转染引入,该质粒表达对应于SEQ ID No:74的序列。RNAs通过表达gRNA主链的质粒的转染引入,该质粒对应于SEQ ID No:66至71。显示了两种不同的sgRNAs靶向的两个不同的区域:VEGFA_sgRNA1、SEQ ID No:85,和VEGFA_sgRNA2,SEQ ID No:86。通过DNA质粒脂转染递送组分,随后在转染后72小时裂解细胞,其后通过PCR扩增靶基因座并通过Sanger测序进行分析。

[0467] 图11A-11D。在HEK-293细胞中通过DNA质粒转染递送CasMINI碱基编辑组分所实现的C到T的转换水平。X轴代表gRNA中MS2适配体的不同位置,并对应于在表4中突出显示的序列:sgRNA-设计1、SEQ ID No:67、对应于5' MS2、sgRNA-设计2,SEQ ID No:68,对应于5' MS2和茎环1的截短、sgRNA-设计3、SEQ ID No:69,对应于作为茎环1的延伸定位的MS2、sgRNA-设计4、SEQ ID No:70,对应于作为茎环2的替换和茎环1的截短定位的MS2、sgRNA-设计5,SEQ ID No:71,对应于作为重复:反重复区的替换定位的MS2。在这些图中的MS-2less对应于SEQ ID No:66。Y轴显示了通过Sanger测序和随后的Chimera分析测量的前间隔区内特定C残基处的C向T的编辑百分率。各柱代表靶向的前间隔区内的靶C残基,图示中的数字代表相关C残基的位置,1为PAM近端,20为PAM远端。误差棒代表来自2个重复的标准误差。

[0468] 图11A、11B、11C、和11D表明CasMINI碱基编辑器在HEK-293T细胞中是有功能的。如Cas12b所证明的,AnoApobec和hApobec3a两者作为MINI碱基编辑器的一部分都是有功能的,并且3种不同的gRNA设计是始终有功能的(设计2、3和4)。有趣的是,配体结合基团的放置对编辑行为具有明显的影响。设计4显示了宽的编辑窗口和广泛提升的编辑水平,而设计2和3显示了对前间隔区内的特定C残基的明显偏好。通过在gRNA内替代置放配体结合基团(位置和它们的尺寸)来改变编辑窗口的这种能力,进一步举例说明了gRNA-配体碱基编辑系统的独特实际应用和灵活性。在不存在MS2适配体(MS2-less)情况下再次观察到很少甚至无编辑,表明所观察到的编辑是依赖于通过MCP配体与MS2配体结合基团之间的相互作用和随后的碱基编辑复合物的组装来募集脱氨酶。

SEQUENCE LISTING

- <110> 地平线探索有限公司
 <120> 向导RNA设计及用于V型Cas系统的复合物
 <130> HORIZON 0109-WO
 <150> US 63/133,945
 <151> 2021-01-05
 <160> 97
 <170> PatentIn version 3.5
 <210> 1
 <211> 56
 <212> RNA
 <213> Alicyclobacillus acidoterrestris
 <220>
 <221> misc_feature
 <222> (37)..(46)
 <223> n is a, c, g, or u
 <220>
 <221> misc_feature
 <222> (47)..(56)
 <223> N is A, C, G, or U
 <400> 1
 gucggaucac ugagcgagcg aucugagaag uggcacnnnn nnnnnnnnnn nnnnnn 56
 <210> 2
 <211> 42
 <212> RNA
 <213> Alicyclobacillus acidoterrestris
 <220>
 <221> misc_feature
 <222> (23)..(42)
 <223> N is A, C, G, or U
 <400> 2
 cgagcgaucu gagaaguggc acnnnnnnnnn nnnnnnnnnn nn 42
 <210> 3
 <211> 34
 <212> RNA
 <213> Alicyclobacillus acidoterrestris
 <220>
 <221> misc_feature

<222>	(14) .. (34)	
<223>	N is A, C, G, or U	
<400>	3	
	cugagaagug gcacnnnnnn nnnnnnnnnn nnnn	34
<210>	4	
<211>	79	
<212>	RNA	
<213>	Alicyclobacillus acidoterrestris	
<400>	4	
	gucuagagga cagaauuuuu caacgggugu gccaauggcc acuuuccagg uggcaaagcc	60
	cguugagcuu cucaaaaag	79
<210>	5	
<211>	78	
<212>	RNA	
<213>	Alicyclobacillus acidoterrestris	
<400>	5	
	gucuagagga cagaauuuuu caacgggugu gccaauggcc acuuuccagg uggcaaagcc	60
	cguugagcuu cucaaaaa	78
<210>	6	
<211>	67	
<212>	RNA	
<213>	Alicyclobacillus acidoterrestris	
<400>	6	
	gucuagagga cagaauuuuu caacgggugu gccaauggcc acuuuccagg uggcaaagcc	60
	cguugag	67
<210>	7	
<211>	111	
<212>	RNA	
<213>	Alicyclobacillus acidoterrestris	
<220>		
<221>	misc_feature	
<222>	(92) .. (111)	
<223>	N is A, C, G, or U	
<400>	7	
	gucuagagga cagaauuuuu caacgggugu gccaauggcc acuuuccagg uggcaaagcc	60
	cguugagcuu cucaaaucug agaaguggca cnnnnnnnnnn nnnnnnnnnn n	111
<210>	8	
<211>	104	
<212>	RNA	

<213>	人工序列	
<220>		
<223>	合成	
<220>		
<221>	misc_feature	
<222>	(85)..(104)	
<223>	N is A, C, G, or U	
<400>	8	
	gucgucuaua ggacggcgag gacaacggga gugcagugcu cuuuccaaga gcaaacaccc	60
	cguuggcuuc aagagaagug gcacnnnnnn nnnnnnnnnn nnnn	104
<210>	9	
<211>	91	
<212>	RNA	
<213>	Bacillus thermoamylovorans	
<400>	9	
	cgagguucug ucuuuugguc aggacaaccg ucuagcuaua agugcugcag gggugugaga	60
	aacuccuauu gcuggacgau gucucuuuuu u	91
<210>	10	
<211>	56	
<212>	RNA	
<213>	Bacillus thermoamylovorans	
<220>		
<221>	misc_feature	
<222>	(37)..(56)	
<223>	N is A, C, G, or U	
<400>	10	
	guccaagaaa aaagaaauga uacgaggcau uagcacnnnn nnnnnnnnnn nnnnnn	56
<210>	11	
<211>	42	
<212>	RNA	
<213>	Bacillus thermoamylovorans	
<220>		
<221>	misc_feature	
<222>	(23)..(42)	
<223>	N is A, C, G, or U	
<400>	11	
	aaaugauacg aggcauuagc acnnnnnnnn nnnnnnnnnn nn	42
<210>	12	
<211>	34	

<212> RNA	
<213> Bacillus thermoamylovorans	
<220>	
<221> misc_feature	
<222> (15) .. (34)	
<223> N is A, C, G, or U	
<400> 12	
cgaggcauua gcacnnnnnn nnnnnnnnnn nnnn	34
<210> 13	
<211> 44	
<212> RNA	
<213> Unknown	
<220>	
<223> Deltaproteobacteria	
<220>	
<221> misc_feature	
<222> (24) .. (44)	
<223> N is A, C, G, or U	
<400> 13	
ccgauaagua aaacgcauca aagnnnnnnn nnnnnnnnnn nnnn	44
<210> 14	
<211> 75	
<212> RNA	
<213> Unknown	
<220>	
<223> Deltaproteobacteria	
<400> 14	
ggcgcguuua uuccauuacu uuggagccag ucccagcgac uaugucguau ggacgaagcg	60
cuaauuuauc ggaga	75
<210> 15	
<211> 122	
<212> RNA	
<213> 人工序列	
<220>	
<223> 合成	
<220>	
<221> misc_feature	
<222> (102) .. (122)	
<223> N is A, C, G, or U	

<400> 15		
ggcgcguua uuccauuacu uuggagccag ucccagcgac uaugucguau ggacgaagcg		60
cuuauuuuauuc ggagagaaac cgauaaguua aacgcaucaa annnnnnnnnn nnnnnnnnnn		120
nn		122
<210> 16		
<211> 128		
<212> RNA		
<213> 人工序列		
<220>		
<223> 合成		
<220>		
<221> misc_feature		
<222> (109) .. (128)		
<223> N is A, C, G, or U		
<400> 16		
acaucuggcg cguuuauucc auuacuugg agccagucc agcgacuaug ucguauggac		60
gaagcgcuaa uuuaucggag agaaaccgau aaguaaaacg caucaaagnn nnnnnnnnnn		120
nnnnnnnnn		128
<210> 17		
<211> 225		
<212> RNA		
<213> 人工序列		
<220>		
<223> 合成		
<220>		
<221> misc_feature		
<222> (206) .. (225)		
<223> N is A, C, G, or U		
<400> 17		
gggcuucacu gauaaaugg agaaccgcuu caccaaaagc ugucccuuag gggauuagaa		60
cuugagugaa ggugggcugc uugcaucagc cuaaugucga gaagugcuuu cuucggaaag		120
uaaccucga aacaaaauca uuuuuuccucu ccaauucugc acaagaaagu ugcagaaccc		180
gaauagacga augaaggaau gcaacnnnnn nnnnnnnnnn nnnnn		225
<210> 18		
<211> 116		
<212> RNA		
<213> 人工序列		
<220>		
<223> 合成		

<400> 18	
cuucaccaaa agcuguccu uaggggauua gaacuugagu gaaggugggc ugcuugcauc	60
agccuaaugu cgagaagugc uuucucgga aaguaaccu cgaacaaa ucauuu	116
<210> 19	
<211> 37	
<212> RNA	
<213> 人工序列	
<220>	
<223> 合成	
<220>	
<221> misc_feature	
<222> (18) .. (37)	
<223> N is A, C, G, or U	
<400> 19	
gaaugaagga augcaacnnn nnnnnnnnnn nnnnnnnn	37
<210> 20	
<211> 57	
<212> RNA	
<213> 人工序列	
<220>	
<223> 合成	
<220>	
<221> misc_feature	
<222> (38) .. (57)	
<223> N is A, C, G, or U	
<400> 20	
guugcagaac ccgaaugac gaaugaagga augcaacnnn nnnnnnnnnn nnnnnnnn	57
<210> 21	
<211> 157	
<212> RNA	
<213> 人工序列	
<220>	
<223> 合成	
<220>	
<221> misc_feature	
<222> (138) .. (157)	
<223> n is a, c, g, or u	
<400> 21	
cuucaccaaa agcuguccu uaggggauua gaacuugagu gaaggugggc ugcuugcauc	60

agccuaaugu cgagaagugc uuucuuucgga aaguaaccu cgaacaaaau ucauuugaaa	120
gaaugaagga augcaacnnn nnnnnnnnnn nnnnnnnn	157
<210> 22	
<211> 84	
<212> PRT	
<213> Unknown	
<220>	
<223> Bacillus phage PBS2	
<400> 22	
Met Thr Asn Leu Ser Asp Ile Ile Glu Lys Glu Thr Gly Lys Gln Leu	
1 5 10 15	
Val Ile Gln Glu Ser Ile Leu Met Leu Pro Glu Glu Val Glu Glu Val	
20 25 30	
Ile Gly Asn Lys Pro Glu Ser Asp Ile Leu Val His Thr Ala Tyr Asp	
35 40 45	
Glu Ser Thr Asp Glu Asn Val Met Leu Leu Thr Ser Asp Ala Pro Glu	
50 55 60	
Tyr Lys Pro Trp Ala Leu Val Ile Gln Asp Ser Asn Gly Glu Asn Lys	
65 70 75 80	
Ile Lys Met Leu	
<210> 23	
<211> 66	
<212> RNA	
<213> 人工序列	
<220>	
<223> 端粒酶Ku结合基序	
<400> 23	
uuuuugucgu acuuauagau cgcuacguua uuucauuuuu gaaaaucuga guccugggag	60
ugcgga	66
<210> 24	
<211> 608	
<212> PRT	
<213> 人工序列	
<220>	
<223> Ku异二聚体	
<400> 24	
Met Ser Gly Trp Glu Ser Tyr Tyr Lys Thr Glu Gly Asp Glu Glu Ala	
1 5 10 15	
Glu Glu Glu Gln Glu Glu Asn Leu Glu Ala Ser Gly Asp Tyr Lys Tyr	

	20		25		30														
Ser	Gly	Arg	Asp	Ser	Leu	Ile	Phe	Leu	Val	Asp	Ala	Ser	Lys	Ala	Met				
	35						40					45							
Phe	Glu	Ser	Gln	Ser	Glu	Asp	Glu	Leu	Thr	Pro	Phe	Asp	Met	Ser	Ile				
	50						55					60							
Gln	Cys	Ile	Gln	Ser	Val	Tyr	Ile	Ser	Lys	Ile	Ile	Ser	Ser	Asp	Arg				
65							70					75			80				
Asp	Leu	Leu	Ala	Val	Val	Phe	Tyr	Gly	Thr	Glu	Lys	Asp	Lys	Asn	Ser				
							85					90			95				
Val	Asn	Phe	Lys	Asn	Ile	Tyr	Val	Leu	Gln	Glu	Leu	Asp	Asn	Pro	Gly				
							100					105			110				
Ala	Lys	Arg	Ile	Leu	Glu	Leu	Asp	Gln	Phe	Lys	Gly	Gln	Gln	Gly	Gln				
							115					120			125				
Lys	Arg	Phe	Gln	Asp	Met	Met	Gly	His	Gly	Ser	Asp	Tyr	Ser	Leu	Ser				
							130					135			140				
Glu	Val	Leu	Trp	Val	Cys	Ala	Asn	Leu	Phe	Ser	Asp	Val	Gln	Phe	Lys				
145							150					155			160				
Met	Ser	His	Lys	Arg	Ile	Met	Leu	Phe	Thr	Asn	Glu	Asp	Asn	Pro	His				
							165					170			175				
Gly	Asn	Asp	Ser	Ala	Lys	Ala	Ser	Arg	Ala	Arg	Thr	Lys	Ala	Gly	Asp				
							180					185			190				
Leu	Arg	Asp	Thr	Gly	Ile	Phe	Leu	Asp	Leu	Met	His	Leu	Lys	Lys	Pro				
							195					200			205				
Gly	Gly	Phe	Asp	Ile	Ser	Leu	Phe	Tyr	Arg	Asp	Ile	Ile	Ser	Ile	Ala				
							210					215			220				
Glu	Asp	Glu	Asp	Leu	Arg	Val	His	Phe	Glu	Glu	Ser	Ser	Lys	Leu	Glu				
225							230					235			240				
Asp	Leu	Leu	Arg	Lys	Val	Arg	Ala	Lys	Glu	Thr	Arg	Lys	Arg	Ala	Leu				
							245					250			255				
Ser	Arg	Leu	Lys	Leu	Lys	Leu	Asn	Lys	Asp	Ile	Val	Ile	Ser	Val	Gly				
							260					265			270				
Ile	Tyr	Asn	Leu	Val	Gln	Lys	Ala	Leu	Lys	Pro	Pro	Pro	Ile	Lys	Leu				
							275					280			285				
Tyr	Arg	Glu	Thr	Asn	Glu	Pro	Val	Lys	Thr	Lys	Thr	Arg	Thr	Phe	Asn				
							290					295			300				
Thr	Ser	Thr	Gly	Gly	Leu	Leu	Leu	Pro	Ser	Asp	Thr	Lys	Arg	Ser	Gln				
305							310					315			320				
Ile	Tyr	Gly	Ser	Arg	Gln	Ile	Ile	Leu	Glu	Lys	Glu	Glu	Thr	Glu	Glu				
							325					330			335				

Leu Lys Arg Phe Asp Asp Pro Gly Leu Met Leu Met Gly Phe Lys Pro		
340	345	350
Leu Val Leu Leu Lys Lys His His Tyr Leu Arg Pro Ser Leu Phe Val		
355	360	365
Tyr Pro Glu Glu Ser Leu Val Ile Gly Ser Ser Thr Leu Phe Ser Ala		
370	375	380
Leu Leu Ile Lys Cys Leu Glu Lys Glu Val Ala Ala Leu Cys Arg Tyr		
385	390	395
Thr Pro Arg Arg Asn Ile Pro Pro Tyr Phe Val Ala Leu Val Pro Gln		
	405	410
		415
Glu Glu Glu Leu Asp Asp Gln Lys Ile Gln Val Thr Pro Pro Gly Phe		
	420	425
		430
Gln Leu Val Phe Leu Pro Phe Ala Asp Asp Lys Arg Lys Met Pro Phe		
	435	440
		445
Thr Glu Lys Ile Met Ala Thr Pro Glu Gln Val Gly Lys Met Lys Ala		
	450	455
		460
Ile Val Glu Lys Leu Arg Phe Thr Tyr Arg Ser Asp Ser Phe Glu Asn		
465	470	475
		480
Pro Val Leu Gln Gln His Phe Arg Asn Leu Glu Ala Leu Ala Leu Asp		
	485	490
		495
Leu Met Glu Pro Glu Gln Ala Val Asp Leu Thr Leu Pro Lys Val Glu		
	500	505
		510
Ala Met Asn Lys Arg Leu Gly Ser Leu Val Asp Glu Phe Lys Glu Leu		
	515	520
		525
Val Tyr Pro Pro Asp Tyr Asn Pro Glu Gly Lys Val Thr Lys Arg Lys		
	530	535
		540
His Asp Asn Glu Gly Ser Gly Ser Lys Arg Pro Lys Val Glu Tyr Ser		
545	550	555
		560
Glu Glu Glu Leu Lys Thr His Ile Ser Lys Gly Thr Leu Gly Lys Phe		
	565	570
		575
Thr Val Pro Met Leu Lys Glu Ala Cys Arg Ala Tyr Gly Leu Lys Ser		
	580	585
		590
Gly Leu Lys Lys Gln Glu Leu Leu Glu Ala Leu Thr Lys His Phe Gln		
	595	600
		605
<210>	25	
<211>	485	
<212>	PRT	
<213>	人工序列	
<220>		

<223> Ku异二聚体

<400> 25

Met	Val	Arg	Ser	Gly	Asn	Lys	Ala	Ala	Val	Val	Leu	Cys	Met	Asp	Val
1				5					10					15	
Gly	Phe	Thr	Met	Ser	Asn	Ser	Ile	Pro	Gly	Ile	Glu	Ser	Pro	Phe	Glu
			20					25					30		
Gln	Ala	Lys	Lys	Val	Ile	Thr	Met	Phe	Val	Gln	Arg	Gln	Val	Phe	Ala
		35					40					45			
Glu	Asn	Lys	Asp	Glu	Ile	Ala	Leu	Val	Leu	Phe	Gly	Thr	Asp	Gly	Thr
	50					55					60				
Asp	Asn	Pro	Leu	Ser	Gly	Gly	Asp	Gln	Tyr	Gln	Asn	Ile	Thr	Val	His
65					70					75				80	
Arg	His	Leu	Met	Leu	Pro	Asp	Phe	Asp	Leu	Leu	Glu	Asp	Ile	Glu	Ser
				85					90					95	
Lys	Ile	Gln	Pro	Gly	Ser	Gln	Gln	Ala	Asp	Phe	Leu	Asp	Ala	Leu	Ile
			100					105					110		
Val	Ser	Met	Asp	Val	Ile	Gln	His	Glu	Thr	Ile	Gly	Lys	Lys	Phe	Glu
		115					120						125		
Lys	Arg	His	Ile	Glu	Ile	Phe	Thr	Asp	Leu	Ser	Ser	Arg	Phe	Ser	Lys
	130					135						140			
Ser	Gln	Leu	Asp	Ile	Ile	Ile	His	Ser	Leu	Lys	Lys	Cys	Asp	Ile	Ser
145					150						155			160	
Glu	Arg	His	Ser	Ile	His	Trp	Pro	Cys	Arg	Leu	Thr	Ile	Gly	Ser	Asn
				165					170				175		
Leu	Ser	Ile	Arg	Ile	Ala	Ala	Tyr	Lys	Ser	Ile	Leu	Gln	Glu	Arg	Val
			180				185						190		
Lys	Lys	Thr	Trp	Thr	Val	Val	Asp	Ala	Lys	Thr	Leu	Lys	Lys	Glu	Asp
		195					200						205		
Ile	Gln	Lys	Glu	Thr	Val	Tyr	Cys	Leu	Asn	Asp	Asp	Asp	Glu	Thr	Glu
	210					215						220			
Val	Leu	Lys	Glu	Asp	Ile	Ile	Gln	Gly	Phe	Arg	Tyr	Gly	Ser	Asp	Ile
225					230					235				240	
Val	Pro	Phe	Ser	Lys	Val	Asp	Glu	Glu	Gln	Met	Lys	Tyr	Lys	Ser	Glu
				245					250					255	
Gly	Lys	Cys	Phe	Ser	Val	Leu	Gly	Phe	Cys	Lys	Ser	Ser	Gln	Val	Gln
			260					265					270		
Arg	Arg	Phe	Phe	Met	Gly	Asn	Gln	Val	Leu	Lys	Val	Phe	Ala	Ala	Arg
		275					280					285			
Asp	Asp	Glu	Ala	Ala	Ala	Val	Ala	Leu	Ser	Ser	Leu	Ile	His	Ala	Leu

290	295	300
Asp Asp Leu Asp Met Val Ala Ile Val Arg Tyr Ala Tyr Asp Lys Arg		
305	310	315
Ala Asn Pro Gln Val Gly Val Ala Phe Pro His Ile Lys His Asn Tyr		320
	325	330
Glu Cys Leu Val Tyr Val Gln Leu Pro Phe Met Glu Asp Leu Arg Gln		335
	340	345
Tyr Met Phe Ser Ser Leu Lys Asn Ser Lys Lys Tyr Ala Pro Thr Glu		350
	355	360
Ala Gln Leu Asn Ala Val Asp Ala Leu Ile Asp Ser Met Ser Leu Ala		365
370	375	380
Lys Lys Asp Glu Lys Thr Asp Thr Leu Glu Asp Leu Phe Pro Thr Thr		
385	390	395
Lys Ile Pro Asn Pro Arg Phe Gln Arg Leu Phe Gln Cys Leu Leu His		
	405	410
Arg Ala Leu His Pro Arg Glu Pro Leu Pro Pro Ile Gln Gln His Ile		415
	420	425
Trp Asn Met Leu Asn Pro Pro Ala Glu Val Thr Thr Lys Ser Gln Ile		430
	435	440
Pro Leu Ser Lys Ile Lys Thr Leu Phe Pro Leu Ile Glu Ala Lys Lys		445
	450	455
Lys Asp Gln Val Thr Ala Gln Glu Ile Phe Gln Asp Asn His Glu Asp		460
465	470	475
Gly Pro Thr Ala Lys		480
	485	

<210> 26

<211> 10

<212> RNA

<213> 人工序列

<220>

<223> 端粒酶Sm7结合基序

<400> 26

aauuuuugga

<210> 27

<211> 83

<212> PRT

<213> 人工序列

<220>

<223> Sm共识位点

<400> 27

Gly Ser Val Ile Asp Val Ser Ser Gln Arg Val Asn Val Gln Arg Pro
 1 5 10 15
 Leu Asp Ala Leu Gly Asn Ser Leu Asn Ser Pro Val Ile Ile Lys Leu
 20 25 30
 Lys Gly Asp Arg Glu Phe Arg Gly Val Leu Lys Ser Phe Asp Leu His
 35 40 45
 Met Asn Leu Val Leu Asn Asp Ala Glu Glu Leu Glu Asp Gly Glu Val
 50 55 60
 Thr Arg Arg Leu Gly Thr Val Leu Ile Arg Gly Asp Asn Ile Val Tyr
 65 70 75 80
 Ile Ser Pro

<210> 28

<211> 23

<212> RNA

<213> 人工序列

<220>

<223> MS2噬菌体操作子茎环

<400> 28

gcacaugagg aucacccaug ugc

23

<210> 29

<211> 117

<212> PRT

<213> 人工序列

<220>

<223> MS2包被蛋白

<400> 29

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

Gln Gly Leu Leu Lys Asp Gly Asn Pro Ile Pro Ser Ala Ile Ala Ala
 100 105 110

Asn Ser Gly Ile Tyr
 115

<210> 30

<211> 26

<212> RNA

<213> 人工序列

<220>

<223> PP7 hage operator stem loop

<400> 30

auaaggaguu uauauggaaa cccuua

26

<210> 31

<211> 128

<212> PRT

<213> 人工序列

<220>

<223> PP7包被蛋白PCP

<400> 31

Met Ser Lys Thr Ile Val Leu Ser Val Gly Glu Ala Thr Arg Thr Leu
 1 5 10 15

Thr Glu Ile Gln Ser Thr Ala Asp Arg Gln Ile Phe Glu Glu Lys Val
 20 25 30

Gly Pro Leu Val Gly Arg Leu Arg Leu Thr Ala Ser Leu Arg Gln Asn
 35 40 45

Gly Ala Lys Thr Ala Tyr Arg Val Asn Leu Lys Leu Asp Gln Ala Asp
 50 55 60

Val Val Asp Cys Ser Thr Ser Val Cys Gly Glu Leu Pro Lys Val Arg
 65 70 75 80

Tyr Thr Gln Val Trp Ser His Asp Val Thr Ile Val Ala Asn Ser Thr
 85 90 95

Glu Ala Ser Arg Lys Ser Leu Tyr Asp Leu Thr Lys Ser Leu Val Ala
 100 105 110

Thr Ser Gln Val Glu Asp Leu Val Val Asn Leu Val Pro Leu Gly Arg
 115 120 125

<210> 32

<211> 19

<212> RNA

<213> 人工序列

<211> 16

<212> RNA

<213> 人工序列

<220>

<223> Csy4结合基序

<400> 36

cugccguaua ggcagc

16

<210> 37

<211> 187

<212> PRT

<213> 人工序列

<220>

<223> Csy4[H29A]

<400> 37

Met Asp His Tyr Leu Asp Ile Arg Leu Arg Pro Asp Pro Glu Phe Pro
 1 5 10 15
 Pro Ala Gln Leu Met Ser Val Leu Phe Gly Lys Leu Ala Gln Ala Leu
 20 25 30
 Val Ala Gln Gly Gly Asp Arg Ile Gly Val Ser Phe Pro Asp Leu Asp
 35 40 45
 Glu Ser Arg Ser Arg Leu Gly Glu Arg Leu Arg Ile His Ala Ser Ala
 50 55 60
 Asp Asp Leu Arg Ala Leu Leu Ala Arg Pro Trp Leu Glu Gly Leu Arg
 65 70 75 80
 Asp His Leu Gln Phe Gly Glu Pro Ala Val Val Pro His Pro Thr Pro
 85 90 95
 Tyr Arg Gln Val Ser Arg Val Gln Ala Lys Ser Asn Pro Glu Arg Leu
 100 105 110
 Arg Arg Arg Leu Met Arg Arg His Asp Leu Ser Glu Glu Glu Ala Arg
 115 120 125
 Lys Arg Ile Pro Asp Thr Val Ala Arg Ala Leu Asp Leu Pro Phe Val
 130 135 140
 Thr Leu Arg Ser Gln Ser Thr Gly Gln His Phe Arg Leu Phe Ile Arg
 145 150 155 160
 His Gly Pro Leu Gln Val Thr Ala Glu Glu Gly Gly Phe Thr Cys Tyr
 165 170 175
 Gly Leu Ser Lys Gly Gly Phe Val Pro Trp Phe
 180 185

<210> 38

<211>	147	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	合成	
<220>		
<221>	misc_feature	
<222>	(127) .. (147)	
<223>	N is A, C, G, T, or U	
<400>	38	
	gcgccatga ggatcaccca tgtgcggcgc guuuauucca uuacuuugga gccagucca	60
	gcgacuaugu cguauggacg aagcgcuuau uuaucggaga gaaaccgaa aguaaaacgc	120
	aucaaannnn nnnnnnnnnn nnnnnnn	147
<210>	39	
<211>	146	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	合成	
<220>		
<221>	misc_feature	
<222>	(126) .. (146)	
<223>	N is A, C, G, T, or U	
<400>	39	
	ggcgcguua uuccauuacu uuggagccag ucccagcgcac ugcgcacatg aggatcaccc	60
	atgtgcuguc guauggacga agcgcuuau uaucggagag aaaccgauaa guaaaacgca	120
	ucaaannnnn nnnnnnnnnn nnnnnn	146
<210>	40	
<211>	146	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	合成	
<220>		
<221>	misc_feature	
<222>	(126) .. (146)	
<223>	N is A, C, G, or U	
<400>	40	
	ggcgcguua uuccauuacu uuggagccag ucccagcgcac uaugucguau ggacgaagcg	60

cuuauuuuau	ggagagcgca	catgaggatc	acccatgtgc	aaaccgauaa	guaaaacgca	120
ucaaannnnn	nnnnnnnnnn	nnnnnn				146
<210>	41					
<211>	147					
<212>	DNA					
<213>	人工序列					
<220>						
<223>	合成					
<220>						
<221>	misc_feature					
<222>	(102)..(122)					
<223>	N is A, C, G, T, or U					
<400>	41					
ggcgcguuuu	uuccauuacu	uuggagccag	ucccagcgac	uaugucguau	ggacgaagcg	60
cuuauuuuau	ggagagaaac	cgauaaguaa	aacgcaucaa	annnnnnnnnn	nnnnnnnnnn	120
nngcgcacat	gaggatcacc	catgtgc				147
<210>	42					
<211>	171					
<212>	DNA					
<213>	人工序列					
<220>						
<223>	合成					
<220>						
<221>	misc_feature					
<222>	(151)..(171)					
<223>	N is A, C., G, T, or U					
<400>	42					
gcgcacatga	ggatcaccca	tgtgcggcgc	guuuauucca	uuacuuugga	gccaguccca	60
gcgacuaugu	cguauggacg	aagcgcuuau	uuaucggaga	gcgcacatga	ggatcaccca	120
tgtgcaaacc	gauaaguaaa	acgcaucaaa	nnnnnnnnnn	nnnnnnnnnn	n	171
<210>	43					
<211>	20					
<212>	DNA					
<213>	Homo sapiens					
<400>	43					
agaaccggag	gacaaagtac					20
<210>	44					
<211>	20					
<212>	DNA					

<213> Homo sapiens	
<400> 44	
tggcaatgcg ccaccggttg	20
<210> 45	
<211> 20	
<212> DNA	
<213> Homo sapiens	
<400> 45	
tcttctgctc ggactcaggc	20
<210> 46	
<211> 20	
<212> DNA	
<213> Homo sapiens	
<400> 46	
tctgctcgga ctcaggccct	20
<210> 47	
<211> 20	
<212> DNA	
<213> Homo sapiens	
<400> 47	
ccagcttctg ccgtttgtac	20
<210> 48	
<211> 20	
<212> DNA	
<213> Homo sapiens	
<400> 48	
aaaaacagtg gataattttg	20
<210> 49	
<211> 20	
<212> DNA	
<213> Homo sapiens	
<400> 49	
gaagagacca aagatcaccc	20
<210> 50	
<211> 20	
<212> DNA	
<213> Homo sapiens	
<400> 50	
cctccgctg tggatgctgc	20

<210> 51	
<211> 20	
<212> DNA	
<213> Homo sapiens	
<400> 51	
tcctgctgct gccgggacct	20
<210> 52	
<211> 20	
<212> DNA	
<213> Homo sapiens	
<400> 52	
gcggccgatg agaagaagaa	20
<210> 53	
<211> 69	
<212> DNA	
<213> 人工序列	
<220>	
<223> 合成	
<220>	
<221> misc_feature	
<222> (49) .. (69)	
<223> N is A, C, G, T, or U	
<400> 53	
gcgcacatga ggatcaccca tgtgcccgau aaguaaacg caucaaagnn nnnnnnnnnn	60
nnnnnnnnnn	69
<210> 54	
<211> 69	
<212> DNA	
<213> 人工序列	
<220>	
<223> 合成	
<220>	
<221> misc_feature	
<222> (24) .. (44)	
<223> N is A, C, G, T, or U	
<400> 54	
ccgauaagua aaacgcauca aagnnnnnnn nnnnnnnnnn nnnngcgac atgaggatca	60
cccatgtgc	69
<210> 55	

<211>	100	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	合成	
<400>	55	
	gcgcacatga ggatcaccca tgtgcggcgc guuuauucca uuacuuugga gccaguccca	60
	gcgacuaugu cguauggacg aagcgcuuau uuaucggaga	100
<210>	56	
<211>	99	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	合成	
<400>	56	
	ggcgcguuuu uuccauuacu uuggagccag ucccagcgac ugcgcacatg aggatcaccc	60
	atgtgcuguc guauggacga agcgcuuau uuaucggaga	99
<210>	57	
<211>	100	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	合成	
<400>	57	
	ggcgcguuuu uuccauuacu uuggagccag ucccagcgac uaugucguau ggacgaagcg	60
	cuuauuuuauc ggagagcgca catgaggatc acccatgtgc	100
<210>	58	
<211>	118	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	合成	
<220>		
<221>	misc_feature	
<222>	(92) .. (111)	
<223>	N is A, C, G, U, or T	
<400>	58	
	gtctagagga cagaatTTTT caacgggtgt gccaatggcc actttccagg tggcaaagcc	60
	cgttgagctt ctcaaactcg agaagtggca cnnnnnnnnn nnnnnnnnnn nttttttt	118

<210>	59	
<211>	142	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	合成	
<220>		
<221>	misc_feature	
<222>	(116)..(135)	
<223>	N is A, C, G, or T	
<400>	59	
	ggcacatgag gatcacccat gtgcgtctag aggacagaat ttttcaacgg gtgtgccaat	60
	ggccactttc caggtggcaa agcccgttga gtttctcaaa tctgagaagt ggcacnnnnn	120
	nnnnnnnnnn nnnnnttttt tt	142
<210>	60	
<211>	142	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	合成	
<220>		
<221>	misc_feature	
<222>	(116)..(135)	
<223>	N is A, C, G, or T	
<400>	60	
	ggcacatgag gatcacccat gtgcgtctag aggacagaat tgttcaacgg gtgtgccaat	60
	ggccactttc caggtggcaa agcccgttga gtttctcaaa tctgagaagt ggcacnnnnn	120
	nnnnnnnnnn nnnnnttttt tt	142
<210>	61	
<211>	142	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	合成	
<220>		
<221>	misc_feature	
<222>	(116)..(135)	
<223>	N is A, C, G, or T	
<400>	61	

ggctctagagg acagaat ttt tcaacgggtg tgccaatggc cactttccag gtggcaaagc	60
ccgttgagct tctcaaatct gagaagtggc acgcacatga ggatcaccca tgtgcnnnnn	120
nnnnnnnnnn nnnnnttttt tt	142
<210> 62	
<211> 142	
<212> DNA	
<213> 人工序列	
<220>	
<223> 合成	
<220>	
<221> misc_feature	
<222> (116) .. (135)	
<223> N is A, C, G, or T	
<400> 62	
ggctctagagg acagaattgt tcaacgggtg tgccaatggc cactttccag gtggcaaagc	60
ccgttgagct tctcaaatct gagaagtggc acgcacatga ggatcaccca tgtgcnnnnn	120
nnnnnnnnnn nnnnnttttt tt	142
<210> 63	
<211> 131	
<212> DNA	
<213> 人工序列	
<220>	
<223> 合成	
<220>	
<221> misc_feature	
<222> (105) .. (124)	
<223> N is A, C, G, or T	
<400> 63	
ggctctagagg acagaat ttt tcaacgggtg tgccaatgga catgaggatc acccatgtcc	60
aggtggcaaaa gcccgttgag cttctcaaat ctgagaagtg gcacnnnnnn nnnnnnnnnn	120
nnnntttttt t	131
<210> 64	
<211> 131	
<212> DNA	
<213> 人工序列	
<220>	
<223> 合成	
<220>	
<221> misc_feature	

<222>	(105) .. (124)	
<223>	N is A, C, G, or T	
<400>	64	
	ggtctagagg acagaattgt tcaacgggtg tgccaatgga catgaggatc acccatgtcc	60
	aggtggcaaaa gcccgttgag cttctcaaat ctgagaagtg gcacnnnnnn nnnnnnnnnn	120
	nnnnnttttt t	131
<210>	65	
<211>	142	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	合成	
<220>		
<221>	misc_feature	
<222>	(116) .. (135)	
<223>	N is A, C, G, or T	
<400>	65	
	ggtctagagg acagaattgc acatgaggat cacccatgtg ctttcaacgg gtgtgccaat	60
	ggccactttc caggtggcaa agcccgttga gttctcaaa tctgagaagt ggcacnnnnn	120
	nnnnnnnnnn nnnnnntttt tt	142
<210>	66	
<211>	192	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	合成	
<220>		
<221>	misc_feature	
<222>	(163) .. (185)	
<223>	N is A, C, G, or T	
<400>	66	
	atgggcttca ctgataaagt ggagaaccgc ttaccaaaa gctgtccctt aggggattag	60
	aacttgagtg aaggtgggct gcttgcata gctaagtgc gagaagtgct ttcttcggaa	120
	agtaaccctc gaaacaaatt catttgaatg aaggaatgca acnnnnnnnn nnnnnnnnnn	180
	nnnnnttttt tt	192
<210>	67	
<211>	215	
<212>	DNA	
<213>	人工序列	

<220>		
<223>	合成	
<220>		
<221>	misc_feature	
<222>	(186)..(208)	
<223>	N is A, C, G, or T	
<400>	67	
gcacatgagg atcacccatg tgcattgggt tcaactgataa agtggagaac cgcttcacca		60
aaagctgtcc cttaggggat tagaacttga gtgaaggtgg gctgcttgca tcagcctaata		120
gtcgagaagt gctttcttcg gaaagtaacc ctcgaaacaa attcatttga atgaaggaat		180
gcaacnnnnn nnnnnnnnnn nnnnnnnntt ttttt		215
<210>	68	
<211>	188	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	合成	
<220>		
<221>	misc_feature	
<222>	(159)..(181)	
<223>	N is A, C, G, or T	
<400>	68	
gcacatgagg atcacccatg tgccgcttca ccaaagctg tcccttaggg gattagaact		60
tgagtgaagg tgggctgctt gcatcagcct aatgtcgaga agtgctttct tcgaaagta		120
accctcgaaa caaattcatt tgaatgaagg aatgcaacnn nnnnnnnnnn nnnnnnnnnn		180
nttttttt		188
<210>	69	
<211>	158	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	合成	
<220>		
<221>	misc_feature	
<222>	(129)..(151)	
<223>	N is A, C, G, or T	
<400>	69	
accgcttcac gcacatgagg atcacccatg tgcgtgaagg tgggctgctt gcatcagcct		60
aatgtcgaga agtgctttct tcgaaagta accctcgaaa caaattcatt tgaatgaagg		120

aatgcaacnn nnnnnnnnnn nnnnnnnnnn nttttttt	158
<210> 70	
<211> 158	
<212> DNA	
<213> 人工序列	
<220>	
<223> 合成	
<220>	
<221> misc_feature	
<222> (129) .. (151)	
<223> N is A, C, G, or T	
<400> 70	
accgcttcac gcacatgagg atcacccatg tgcgtgaagg tgggctgctt gcacagcct	60
aatgtcgaga agtgctttct tcggaaagta accctcgaaa caaatcatt tgaatgaagg	120
aatgcaacnn nnnnnnnnnn nnnnnnnnnn nttttttt	158
<210> 71	
<211> 199	
<212> DNA	
<213> 人工序列	
<220>	
<223> 含有来自Cas12b、SV40 NLS的序列的质粒	
<220>	
<221> misc_feature	
<222> (170) .. (192)	
<223> N is A, C, G, or T	
<400> 71	
gggcttcact gataaagtgg agaaccgctt caccaaaagc tgtcccttag gggattagaa	60
cttgagtgaa ggtgggctgc ttgcatcagc ctaatgtcga gaagtgcttt cttcgaaag	120
taaccctcga aacaaagcac atgaggatca ccatgtgcg gaatgcaacn nnnnnnnnnn	180
nnnnnnnnnn nntttttttt	199
<210> 72	
<211> 4035	
<212> DNA	
<213> 人工序列	
<220>	
<223> 含有来自人源Cas12b和SV40 NLS的序列的质粒	
<400> 72	
atggcccaa agaagaagcg gaaagtcgcc gtgaaaagca ttaaagtgaa actgaggctg	60
gacgatatgc ccgaaatcag agccggcctc tggaagctgc acaaggaggt caacgccggc	120

gttcgttatt	acacagagtg	gctgtcttta	ctcagacaag	aaaatttata	tagaaggagc	180
cccaatggcg	atggcgagca	agagtgcgac	aaaaccgccg	aagaatgcaa	ggccgaactg	240
ctcgaaagac	tgagagccag	acaagttgag	aacggacaca	gaggacccgc	cggtcccgat	300
gatgaactgc	tcgagctcgc	taggcaactg	tacgagctgc	tcgtcccca	agctatcgga	360
gctaaaggag	acgctcagca	gatcgcctag	aagtttctga	gccccctcgc	cgataaagac	420
gctgtgggcg	gttttaggaat	cgccaaggct	ggaaataaac	ctcgttgggt	gaggatgagg	480
gaagctggcg	agccccggctg	ggaggaggaa	aaggagaaag	ccgaaactcg	taaatccgcc	540
gacagaacag	ctgacgtgct	gagggccctc	gccgacttcg	gtttaaaacc	cctcatgaga	600
gtctataccg	actccgagat	gtccagcgtc	gagtggaaac	ccttacgtaa	gggccaagct	660
gtgagaacat	gggatcgtga	tatgttccag	caagctatcg	aaaggatgat	gagctgggag	720
tcttggaatc	agagagtggt	ccaagaatac	gccaaactcg	tggaacagaa	gaatcgtttc	780
gagcagaaaa	actttgtcgg	acaagaacat	ttagtgcate	tcgtgaacca	actgcagcaa	840
gatatgaagg	aggcctcccc	cggtttagag	agcaaagagc	aaaccgctca	ctatgtcacc	900
ggtcgtgcc	tcagaggctc	cgacaaaagt	ttcgaagaag	ggggaagct	ggccccgcac	960
gctcctttcg	atattatacga	cgccgagatc	aagaacgtgc	agaggagaaa	tacaaggagg	1020
tttggtccc	acgatttatt	cgctaaactc	gccgagcccc	aataccaagc	tttatggaga	1080
gaggatgcca	gctttctgac	tcgttacgcc	gtgtacaatt	ccattctgag	aaaactgaac	1140
cacgcaaaaa	tgtttgccac	attcacttta	cccgatgcca	ccgcccacc	tatctggaca	1200
agattcgaca	agctcggcgg	aaacctccac	cagtatacct	ttttatttaa	cgaattcgga	1260
gagaggaggc	atgccattcg	ttttcacaag	ttattaaagg	tcgagaatgg	agtcgccaga	1320
gaggtggacg	acgtcaccgt	gcctatcagc	atgagcgaac	agctggacaa	tttactgcct	1380
cgtgacccca	acgaacctat	cgccctctac	ttcagagact	acggagctga	gcagcatttc	1440
accggcgagt	tcggaggcgc	taagatccag	tgtagaaggg	atcaactggc	ccatattgcat	1500
aggagaaggg	gcgccagaga	tgtctatctg	aacgtgagcg	ttcgtgtcca	aagccaaagc	1560
gaggccagag	gagaaaggag	acccccctac	gccgccgtct	ttagactggt	cgcgacaat	1620
catcgtgctt	tcgtgcactt	tgataagctg	tccgactatt	tagccgaaca	ccccgacgat	1680
ggaaagctgg	gcagcgaggg	attattaagc	ggcctcagag	tcatgagcgt	ggctctgggc	1740
ctcagaacca	gcgcctccat	ctccgtcttt	agggtggcca	gaaaagacga	gctgaagccc	1800
aacagcaagg	gaagggtgcc	ctttttcttc	cctatcaagg	gcaatgacaa	tttagtggcc	1860
gtgcacgaaa	ggccccagtt	attaaagctg	cccggcgaga	ccgaaagcaa	agatttaagg	1920
gctatcagag	aggagagaca	gagaacttta	agacagctga	gaaccagct	ggcttatctg	1980
agattattag	tcagatgtgg	cagcgaggac	gtcggtcgta	gagagaggag	ctgggccaag	2040
ctgattgaac	aacccttga	tgccgctaat	cacatgacc	ccgattggag	ggaagctttc	2100
gagaacgagc	tcagaaaact	gaagtcttta	cacggcattt	gcagcgacaa	ggagtggatg	2160
gacgccgtgt	acgagtccgt	gagaagggtg	tggaggcaca	tgggaaagca	agttagggat	2220
tggaggaaaag	atgtgaggtc	cggcgaaaaga	cccaagatca	gaggctacgc	caaggacgtg	2280
gtcggaggaa	actccatcga	gcagatcgag	tacctcgaga	gacaatacaa	gtttttaaag	2340
tcttggtcct	tcttcggcaa	ggtcagcggc	caagtcattc	gtgctgaaaa	gggatctcgt	2400
ttcgccatca	cactgagaga	gcacattgac	cacgccaag	aggatcgtct	gaaaaaactc	2460

gccgatcgta tcattatgga ggccctcggc tatgtctatg ctctggacga gagaggcaag 2520
 ggaaaatggg tcgccaagta tcccccttgt caactgattt tattagcggg gctgtccgag 2580
 taccaattta acaacgatag gcctccctcc gagaataacc agctcatgca gtggagccat 2640
 cgtggcgtgt ttcaagaact gatcaatcaa gctcaagttc acgatttact cgtgggcacc 2700
 atgtacgccg cttttagctc cagattcgac gctaggaccg gcgcccccg tattaggtgt 2760
 agaagagtgc ccgctcgttg cacccaagaa cataaccccc aaccctttcc ttggtggctg 2820
 aataagttcg tcgctcagca caccctcgac gttgccctt tacgtgccga cgacctcatc 2880
 cctactgggtg aaggcgagat ctttgtgagc ctttcagcg ctgaggaagg cgattttcac 2940
 cagatccacg ccgctctgaa cgccgccag aatctgcaac agaggctctg gagcgacttc 3000
 gatatttccc agattcgtct gaggtgcgat tggggagagg tggatggaga actcgtcctc 3060
 atccccagac tgactggtaa gaggaccgct gacagctatt ccaataaggt cttctatacc 3120
 aatactgggtg tcacctacta cgagagagag aggggcaaga agagaaggaa agtcttcgcc 3180
 caagagaagc tgtccgagga ggaggccgaa ttattagtcg aggctgatga ggctcgtgaa 3240
 aagtccgtgg ttttaatgag ggaccctcc ggcacatca acagaggcaa ttggactcgt 3300
 cagaaggaat tctggagcat ggtgaatcag aggatcgagg gctatctggt gaagcagatt 3360
 agatctcgtg tgccctctga agatagcgtc tgtgagaata ccggagatat tagcggcggg 3420
 agcggcggga gcggggggag cactaatctg agcgacatca ttgagaagga gactgggaaa 3480
 cagctggtea ttcaggagtc catcctgatg ctgcctgagg aggtggagga agtgatcggc 3540
 aacaagccag agtctgacat cctggtgcac accgcctacg acgagtccac agatgagaat 3600
 gtgatgctgc tgacctctga cgccccgag tataagcctt gggcctggt catccaggat 3660
 tctaaccgagc agaataagat caagatgctg agcggaggat ccggaggatc tggaggcagc 3720
 accaacctgt ctgacatcat cgagaaggag acaggcaagc agctggtcat ccaggagagc 3780
 atcctaatac ttcccgaaga agtcgaagaa gtgatcggaa acaagcctga gagcgatata 3840
 ctggtccata ctgcgtatga tgaaagtacc gacgaaaacg taatgctact cacatccgac 3900
 gccccagagt ataagccctg ggctctagtt atacaagact ccaacggaga gaacaaaatc 3960
 aaaatgctgt ctggcggctc aaaaagaacc gccgacggca gcgaattcga gcccaagaag 4020
 aagaggaaaag tctaa 4035

<210> 73

<211> 1344

<212> PRT

<213> 人工序列

<220>

<223> 含有来自Cas12b 和SV40 NLS的序列的蛋白

<400> 73

Met Ala Pro Lys Lys Lys Arg Lys Val Ala Val Lys Ser Ile Lys Val
 1 5 10 15
 Lys Leu Arg Leu Asp Asp Met Pro Glu Ile Arg Ala Gly Leu Trp Lys
 20 25 30
 Leu His Lys Glu Val Asn Ala Gly Val Arg Tyr Tyr Thr Glu Trp Leu

35	40	45
Ser Leu Leu Arg Gln Glu Asn Leu Tyr Arg Arg Ser Pro Asn Gly Asp		
50	55	60
Gly Glu Gln Glu Cys Asp Lys Thr Ala Glu Glu Cys Lys Ala Glu Leu		
65	70	75
Leu Glu Arg Leu Arg Ala Arg Gln Val Glu Asn Gly His Arg Gly Pro		
85	90	95
Ala Gly Ser Asp Asp Glu Leu Leu Gln Leu Ala Arg Gln Leu Tyr Glu		
100	105	110
Leu Leu Val Pro Gln Ala Ile Gly Ala Lys Gly Asp Ala Gln Gln Ile		
115	120	125
Ala Arg Lys Phe Leu Ser Pro Leu Ala Asp Lys Asp Ala Val Gly Gly		
130	135	140
Leu Gly Ile Ala Lys Ala Gly Asn Lys Pro Arg Trp Val Arg Met Arg		
145	150	155
Glu Ala Gly Glu Pro Gly Trp Glu Glu Glu Lys Glu Lys Ala Glu Thr		
165	170	175
Arg Lys Ser Ala Asp Arg Thr Ala Asp Val Leu Arg Ala Leu Ala Asp		
180	185	190
Phe Gly Leu Lys Pro Leu Met Arg Val Tyr Thr Asp Ser Glu Met Ser		
195	200	205
Ser Val Glu Trp Lys Pro Leu Arg Lys Gly Gln Ala Val Arg Thr Trp		
210	215	220
Asp Arg Asp Met Phe Gln Gln Ala Ile Glu Arg Met Met Ser Trp Glu		
225	230	235
Ser Trp Asn Gln Arg Val Gly Gln Glu Tyr Ala Lys Leu Val Glu Gln		
245	250	255
Lys Asn Arg Phe Glu Gln Lys Asn Phe Val Gly Gln Glu His Leu Val		
260	265	270
His Leu Val Asn Gln Leu Gln Gln Asp Met Lys Glu Ala Ser Pro Gly		
275	280	285
Leu Glu Ser Lys Glu Gln Thr Ala His Tyr Val Thr Gly Arg Ala Leu		
290	295	300
Arg Gly Ser Asp Lys Val Phe Glu Lys Trp Gly Lys Leu Ala Pro Asp		
305	310	315
Ala Pro Phe Asp Leu Tyr Asp Ala Glu Ile Lys Asn Val Gln Arg Arg		
325	330	335
Asn Thr Arg Arg Phe Gly Ser His Asp Leu Phe Ala Lys Leu Ala Glu		
340	345	350

Pro Glu Tyr Gln Ala Leu Trp Arg Glu Asp Ala Ser Phe Leu Thr Arg
 355 360 365
 Tyr Ala Val Tyr Asn Ser Ile Leu Arg Lys Leu Asn His Ala Lys Met
 370 375 380
 Phe Ala Thr Phe Thr Leu Pro Asp Ala Thr Ala His Pro Ile Trp Thr
 385 390 395 400
 Arg Phe Asp Lys Leu Gly Gly Asn Leu His Gln Tyr Thr Phe Leu Phe
 405 410 415
 Asn Glu Phe Gly Glu Arg Arg His Ala Ile Arg Phe His Lys Leu Leu
 420 425 430
 Lys Val Glu Asn Gly Val Ala Arg Glu Val Asp Asp Val Thr Val Pro
 435 440 445
 Ile Ser Met Ser Glu Gln Leu Asp Asn Leu Leu Pro Arg Asp Pro Asn
 450 455 460
 Glu Pro Ile Ala Leu Tyr Phe Arg Asp Tyr Gly Ala Glu Gln His Phe
 465 470 475 480
 Thr Gly Glu Phe Gly Gly Ala Lys Ile Gln Cys Arg Arg Asp Gln Leu
 485 490 495
 Ala His Met His Arg Arg Arg Gly Ala Arg Asp Val Tyr Leu Asn Val
 500 505 510
 Ser Val Arg Val Gln Ser Gln Ser Glu Ala Arg Gly Glu Arg Arg Pro
 515 520 525
 Pro Tyr Ala Ala Val Phe Arg Leu Val Gly Asp Asn His Arg Ala Phe
 530 535 540
 Val His Phe Asp Lys Leu Ser Asp Tyr Leu Ala Glu His Pro Asp Asp
 545 550 555 560
 Gly Lys Leu Gly Ser Glu Gly Leu Leu Ser Gly Leu Arg Val Met Ser
 565 570 575
 Val Ala Leu Gly Leu Arg Thr Ser Ala Ser Ile Ser Val Phe Arg Val
 580 585 590
 Ala Arg Lys Asp Glu Leu Lys Pro Asn Ser Lys Gly Arg Val Pro Phe
 595 600 605
 Phe Phe Pro Ile Lys Gly Asn Asp Asn Leu Val Ala Val His Glu Arg
 610 615 620
 Ser Gln Leu Leu Lys Leu Pro Gly Glu Thr Glu Ser Lys Asp Leu Arg
 625 630 635 640
 Ala Ile Arg Glu Glu Arg Gln Arg Thr Leu Arg Gln Leu Arg Thr Gln
 645 650 655
 Leu Ala Tyr Leu Arg Leu Leu Val Arg Cys Gly Ser Glu Asp Val Gly

660	665	670
Arg Arg Glu Arg Ser Trp Ala Lys Leu Ile Glu Gln Pro Val Asp Ala		
675	680	685
Ala Asn His Met Thr Pro Asp Trp Arg Glu Ala Phe Glu Asn Glu Leu		
690	695	700
Gln Lys Leu Lys Ser Leu His Gly Ile Cys Ser Asp Lys Glu Trp Met		
705	710	715
Asp Ala Val Tyr Glu Ser Val Arg Arg Val Trp Arg His Met Gly Lys		
725	730	735
Gln Val Arg Asp Trp Arg Lys Asp Val Arg Ser Gly Glu Arg Pro Lys		
740	745	750
Ile Arg Gly Tyr Ala Lys Asp Val Val Gly Gly Asn Ser Ile Glu Gln		
755	760	765
Ile Glu Tyr Leu Glu Arg Gln Tyr Lys Phe Leu Lys Ser Trp Ser Phe		
770	775	780
Phe Gly Lys Val Ser Gly Gln Val Ile Arg Ala Glu Lys Gly Ser Arg		
785	790	795
Phe Ala Ile Thr Leu Arg Glu His Ile Asp His Ala Lys Glu Asp Arg		
805	810	815
Leu Lys Lys Leu Ala Asp Arg Ile Ile Met Glu Ala Leu Gly Tyr Val		
820	825	830
Tyr Ala Leu Asp Glu Arg Gly Lys Gly Lys Trp Val Ala Lys Tyr Pro		
835	840	845
Pro Cys Gln Leu Ile Leu Leu Ala Glu Leu Ser Glu Tyr Gln Phe Asn		
850	855	860
Asn Asp Arg Pro Pro Ser Glu Asn Asn Gln Leu Met Gln Trp Ser His		
865	870	875
Arg Gly Val Phe Gln Glu Leu Ile Asn Gln Ala Gln Val His Asp Leu		
885	890	895
Leu Val Gly Thr Met Tyr Ala Ala Phe Ser Ser Arg Phe Asp Ala Arg		
900	905	910
Thr Gly Ala Pro Gly Ile Arg Cys Arg Arg Val Pro Ala Arg Cys Thr		
915	920	925
Gln Glu His Asn Pro Glu Pro Phe Pro Trp Trp Leu Asn Lys Phe Val		
930	935	940
Val Glu His Thr Leu Asp Ala Cys Pro Leu Arg Ala Asp Asp Leu Ile		
945	950	955
Pro Thr Gly Glu Gly Glu Ile Phe Val Ser Pro Phe Ser Ala Glu Glu		
965	970	975

Gly Asp Phe His Gln Ile His Ala Ala Leu Asn Ala Ala Gln Asn Leu		
980	985	990
Gln Gln Arg Leu Trp Ser Asp Phe Asp Ile Ser Gln Ile Arg Leu Arg		
995	1000	1005
Cys Asp Trp Gly Glu Val Asp Gly Glu Leu Val Leu Ile Pro Arg		
1010	1015	1020
Leu Thr Gly Lys Arg Thr Ala Asp Ser Tyr Ser Asn Lys Val Phe		
1025	1030	1035
Tyr Thr Asn Thr Gly Val Thr Tyr Tyr Glu Arg Glu Arg Gly Lys		
1040	1045	1050
Lys Arg Arg Lys Val Phe Ala Gln Glu Lys Leu Ser Glu Glu Glu		
1055	1060	1065
Ala Glu Leu Leu Val Glu Ala Asp Glu Ala Arg Glu Lys Ser Val		
1070	1075	1080
Val Leu Met Arg Asp Pro Ser Gly Ile Ile Asn Arg Gly Asn Trp		
1085	1090	1095
Thr Arg Gln Lys Glu Phe Trp Ser Met Val Asn Gln Arg Ile Glu		
1100	1105	1110
Gly Tyr Leu Val Lys Gln Ile Arg Ser Arg Val Pro Leu Gln Asp		
1115	1120	1125
Ser Ala Cys Glu Asn Thr Gly Asp Ile Ser Gly Gly Ser Gly Gly		
1130	1135	1140
Ser Gly Gly Ser Thr Asn Leu Ser Asp Ile Ile Glu Lys Glu Thr		
1145	1150	1155
Gly Lys Gln Leu Val Ile Gln Glu Ser Ile Leu Met Leu Pro Glu		
1160	1165	1170
Glu Val Glu Glu Val Ile Gly Asn Lys Pro Glu Ser Asp Ile Leu		
1175	1180	1185
Val His Thr Ala Tyr Asp Glu Ser Thr Asp Glu Asn Val Met Leu		
1190	1195	1200
Leu Thr Ser Asp Ala Pro Glu Tyr Lys Pro Trp Ala Leu Val Ile		
1205	1210	1215
Gln Asp Ser Asn Gly Glu Asn Lys Ile Lys Met Leu Ser Gly Gly		
1220	1225	1230
Ser Gly Gly Ser Gly Gly Ser Thr Asn Leu Ser Asp Ile Ile Glu		
1235	1240	1245
Lys Glu Thr Gly Lys Gln Leu Val Ile Gln Glu Ser Ile Leu Met		
1250	1255	1260
Leu Pro Glu Glu Val Glu Glu Val Ile Gly Asn Lys Pro Glu Ser		

1265	1270	1275
Asp Ile Leu Val His Thr Ala Tyr Asp Glu Ser Thr Asp Glu Asn		
1280	1285	1290
Val Met Leu Leu Thr Ser Asp Ala Pro Glu Tyr Lys Pro Trp Ala		
1295	1300	1305
Leu Val Ile Gln Asp Ser Asn Gly Glu Asn Lys Ile Lys Met Leu		
1310	1315	1320
Ser Gly Gly Ser Lys Arg Thr Ala Asp Gly Ser Glu Phe Glu Pro		
1325	1330	1335
Lys Lys Lys Arg Lys Val		

1340

<210> 74

<211> 2235

<212> DNA

<213> 人工序列

<220>

<223> 含有来自人源CasMINI和SV40 NLS的序列的质粒

<400> 74

```

atggcccca agaaaaaacg caaggtggcc aaaaacacca ttacaaaac actgaaactg      60
cgtattgtgc gtccgtataa tagcgcagaa gtggaaaaaa ttgttgccga cgaaaaaac      120
aaccgcgaaa aaatcgact ggaaaagaac aaagacaaag tgaagaagc ctgcagcaaa      180
catctgaaag ttgcagcata ttgtaccaca caggttgaac gtaatgatg cctgttttgt      240
aaagcacgta aactggatga caaattctac caaaaactgc gtggtcagtt tccgatgca      300
gttttttggc aagaaatcag cgaaattttt cgccagctgc agaaacaggc agcagaaatc      360
tataatcaga gcctgatcga actgtactac gagattttta tcaaaggcaa aggtattgca      420
aatgccagca gcgttgaaca ttatctgagt agagtttggt atagacgtgc agcagaactg      480
tttaaaaacg cagcaattgc aagcggctcg ctagcaaaa tcaaagcaa ttttcgtctg      540
aaagaactga aaaacatgaa aagtggctcg ccgaccacca aaagcataa ttttccgatt      600
ccgctggtta aacagaaaag tggtcagtat accggttttg aaattagcaa tcataatagc      660
gacttcatca tcaagattcc gtttggtcgt tggcaggtca aaaaagagat tgataaatat      720
cgtccgtggg agaaatttga ctttgaacag gttcagaaaa gcccgaaacc gattagcctg      780
ctgctgagca cccagcgtcg taaacgtaat aaaggttga gcaaagatga aggcaccgaa      840
gccgaaatca aaaaagtat gaatggcgat tatcagacca gctacattga agttaaactg      900
ggcagcaaaa tctgtgaaaa aagcgcattg atgctgaatc tgagcattga tgttccgaaa      960
attgataaag gtgtggatcc gagcattatt ggtggtattg cagttggtgt tagatcaccg     1020
ctggtttgcg caattaacaa tgcatttagc cgttatagca tcagcgataa cgacctgttt     1080
cacttcaaca agaaaatgtt tgcacgtcgt cgtatcctgc tgaaaaaaaa ccgtcataaa     1140
cgtgcaggtc atggtgcaaa aaacaaactg aaaccgatca ccattctgac cgaaaaaagt     1200
gaacgttttc gcaaaaagct gattgaactg tgggcatgtg aaatcgcgga tttcttcatt     1260

```

aaaaacaaag ttggcaccgt gcagatggaa aatctggaaa gcatgaaacg taaagaggac 1320
 agctattttta acattcgcct gcgtggcttt tggccgatg cagaaatgca gaacaaaatc 1380
 gaattcaaac tgaagcagta tggcatcgaa attcgtaaag ttgcaccgaa taataccagc 1440
 aaaacctgta gcaaatgtgg ccatctgaac aactatttca acttcgagta ccgcaagaaa 1500
 aacaaattcc cgcactttaa atgcgaaaaa tgcaacttca aagaaaacgc cgcgtataat 1560
 gcagccctga atatttcaaa cccgaaactg aaaagcacca aagagagacc gagcggcggg 1620
 agcggcggga gcggggggag cactaatctg agcgacatca ttgagaagga gactgggaaa 1680
 cagctgggtca ttcaggagtc catcctgatg ctgcctgagg aggtggagga agtgatcggc 1740
 aacaagccag agtctgacat cctggtgcac accgcctacg acgagtccac agatgagaat 1800
 gtgatgctgc tgacctctga cgccccgag tataagcctt gggccctggt catccaggat 1860
 tctaaccgcg agaataagat caagatgctg agcggaggat ccggaggatc tggaggcagc 1920
 accaacctgt ctgacatcat cgagaaggag acaggcaagc agctggtcat ccaggagagc 1980
 atcctgatgc tgcccgaaga agtcgaagaa gtgatcggaa acaagcctga gagcgatata 2040
 ctggtccata ccgcctacga cgagagtacc gacgaaaatg tgatgctgct gacatccgac 2100
 gccccagagt ataagccctg ggetctggtc atccaggatt ccaacggaga gaacaaaatc 2160
 aaaatgctgt ctggcggctc aaaaagaacc gccgacggca gcgaattcga gcccaagaag 2220
 aagaggaaaag tctaa 2235

<210> 75

<211> 744

<212> PRT

<213> 人工序列

<220>

<223> 含有来自CasMINI和SV40 NLS的序列的蛋白

<400> 75

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

115	120	125
Tyr Tyr Glu Ile Phe Ile Lys Gly Lys Gly Ile Ala Asn Ala Ser Ser		
130	135	140
Val Glu His Tyr Leu Ser Arg Val Cys Tyr Arg Arg Ala Ala Glu Leu		
145	150	155
Phe Lys Asn Ala Ala Ile Ala Ser Gly Leu Arg Ser Lys Ile Lys Ser		
165	170	175
Asn Phe Arg Leu Lys Glu Leu Lys Asn Met Lys Ser Gly Leu Pro Thr		
180	185	190
Thr Lys Ser Asp Asn Phe Pro Ile Pro Leu Val Lys Gln Lys Gly Gly		
195	200	205
Gln Tyr Thr Gly Phe Glu Ile Ser Asn His Asn Ser Asp Phe Ile Ile		
210	215	220
Lys Ile Pro Phe Gly Arg Trp Gln Val Lys Lys Glu Ile Asp Lys Tyr		
225	230	235
Arg Pro Trp Glu Lys Phe Asp Phe Glu Gln Val Gln Lys Ser Pro Lys		
245	250	255
Pro Ile Ser Leu Leu Leu Ser Thr Gln Arg Arg Lys Arg Asn Lys Gly		
260	265	270
Trp Ser Lys Asp Glu Gly Thr Glu Ala Glu Ile Lys Lys Val Met Asn		
275	280	285
Gly Asp Tyr Gln Thr Ser Tyr Ile Glu Val Lys Arg Gly Ser Lys Ile		
290	295	300
Cys Glu Lys Ser Ala Trp Met Leu Asn Leu Ser Ile Asp Val Pro Lys		
305	310	315
Ile Asp Lys Gly Val Asp Pro Ser Ile Ile Gly Gly Ile Ala Val Gly		
325	330	335
Val Arg Ser Pro Leu Val Cys Ala Ile Asn Asn Ala Phe Ser Arg Tyr		
340	345	350
Ser Ile Ser Asp Asn Asp Leu Phe His Phe Asn Lys Lys Met Phe Ala		
355	360	365
Arg Arg Arg Ile Leu Leu Lys Lys Asn Arg His Lys Arg Ala Gly His		
370	375	380
Gly Ala Lys Asn Lys Leu Lys Pro Ile Thr Ile Leu Thr Glu Lys Ser		
385	390	395
Glu Arg Phe Arg Lys Lys Leu Ile Glu Arg Trp Ala Cys Glu Ile Ala		
405	410	415
Asp Phe Phe Ile Lys Asn Lys Val Gly Thr Val Gln Met Glu Asn Leu		
420	425	430

Glu Ser Met Lys Arg Lys Glu Asp Ser Tyr Phe Asn Ile Arg Leu Arg
 435 440 445
 Gly Phe Trp Pro Tyr Ala Glu Met Gln Asn Lys Ile Glu Phe Lys Leu
 450 455 460
 Lys Gln Tyr Gly Ile Glu Ile Arg Lys Val Ala Pro Asn Asn Thr Ser
 465 470 475 480
 Lys Thr Cys Ser Lys Cys Gly His Leu Asn Asn Tyr Phe Asn Phe Glu
 485 490 495
 Tyr Arg Lys Lys Asn Lys Phe Pro His Phe Lys Cys Glu Lys Cys Asn
 500 505 510
 Phe Lys Glu Asn Ala Ala Tyr Asn Ala Ala Leu Asn Ile Ser Asn Pro
 515 520 525
 Lys Leu Lys Ser Thr Lys Glu Arg Pro Ser Gly Gly Ser Gly Gly Ser
 530 535 540
 Gly Gly Ser Thr Asn Leu Ser Asp Ile Ile Glu Lys Glu Thr Gly Lys
 545 550 555 560
 Gln Leu Val Ile Gln Glu Ser Ile Leu Met Leu Pro Glu Glu Val Glu
 565 570 575
 Glu Val Ile Gly Asn Lys Pro Glu Ser Asp Ile Leu Val His Thr Ala
 580 585 590
 Tyr Asp Glu Ser Thr Asp Glu Asn Val Met Leu Leu Thr Ser Asp Ala
 595 600 605
 Pro Glu Tyr Lys Pro Trp Ala Leu Val Ile Gln Asp Ser Asn Gly Glu
 610 615 620
 Asn Lys Ile Lys Met Leu Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser
 625 630 635 640
 Thr Asn Leu Ser Asp Ile Ile Glu Lys Glu Thr Gly Lys Gln Leu Val
 645 650 655
 Ile Gln Glu Ser Ile Leu Met Leu Pro Glu Glu Val Glu Glu Val Ile
 660 665 670
 Gly Asn Lys Pro Glu Ser Asp Ile Leu Val His Thr Ala Tyr Asp Glu
 675 680 685
 Ser Thr Asp Glu Asn Val Met Leu Leu Thr Ser Asp Ala Pro Glu Tyr
 690 695 700
 Lys Pro Trp Ala Leu Val Ile Gln Asp Ser Asn Gly Glu Asn Lys Ile
 705 710 715 720
 Lys Met Leu Ser Gly Gly Ser Lys Arg Thr Ala Asp Gly Ser Glu Phe
 725 730 735
 Glu Pro Lys Lys Lys Arg Lys Val

740

<210> 76
 <211> 1050
 <212> DNA
 <213> 人工序列
 <220>
 <223> 含有来自人Apobec3A和SV40 NLS的序列的质粒
 <400> 76
 atggccccca agaagaagcg gaaagtggaa gccagcccag catccgggcc cagacacttg 60
 atggatccac acatattcac ttccaacttt aacaatggca ttggaaggca taagacctac 120
 ctgtgctacg aagtggagcg cctggacaat ggcacctcgg tcaagatgga ccagcacagg 180
 ggctttctac acaaccaggc taagaatctt ctctgtggct tttacggccg ccatgcggag 240
 ctgcgcttct tggacctggt tccttctttg cagttggacc cggcccaaat ctacagggtc 300
 acttggttca tctcctggag ccctgcttc tccgcggtgt gtgccgggga agtgcgtgcg 360
 ttccctcagg agaacacaca cgtgagactg cgtatcttgc ctgcccgcac ctatgatgac 420
 gaccccctat ataaggaggc actgcaaatg ctgcgggatg ctggggccca agtctccatc 480
 atgacctacg atgaatttaa gcaactgctg gacacctttg tggaccacca gggatgtccc 540
 ttccagccct gggatggact agatgagcac agccaagccc tgagtgggag gctgcggggc 600
 attctccaga atcagggaaa cgagctgaag acaccctggt gcgacaccac acacacctct 660
 ccaccttgcc cagcaccaga gctgctggga ggccctatgg ccagcaactt cacacagttt 720
 gtgctgggtg ataatggagg aaccggcgac gtgacagtgg caccatctaa ctttgccaat 780
 ggcatcgccg agtggatcag ctccaactct cggagccagg cctataaggt gacctgtagc 840
 gtgcggcagt ctagcgccca gaatagaaag tatacaatca aggtggaggt gcctaagggc 900
 gcctggagat cctacctgaa catggagctg accatcccaa tctttgccac aaattctgat 960
 tgcgagctga tcgtgaaggc catgcagggc ctgctgaagg acggcaacc tatccaagc 1020
 gccatcgccg ccaatagcgg aatctactga 1050

<210> 77
 <211> 349
 <212> PRT
 <213> 人工序列
 <220>
 <223> 含有来自人Apobec3A和SV40 NLS的序列的蛋白
 <400> 77
 Met Ala Pro Lys Lys Lys Arg Lys Val Glu Ala Ser Pro Ala Ser Gly
 1 5 10 15
 Pro Arg His Leu Met Asp Pro His Ile Phe Thr Ser Asn Phe Asn Asn
 20 25 30
 Gly Ile Gly Arg His Lys Thr Tyr Leu Cys Tyr Glu Val Glu Arg Leu
 35 40 45

Asp Asn Gly Thr Ser Val Lys Met Asp Gln His Arg Gly Phe Leu His
 50 55 60
 Asn Gln Ala Lys Asn Leu Leu Cys Gly Phe Tyr Gly Arg His Ala Glu
 65 70 75 80
 Leu Arg Phe Leu Asp Leu Val Pro Ser Leu Gln Leu Asp Pro Ala Gln
 85 90 95
 Ile Tyr Arg Val Thr Trp Phe Ile Ser Trp Ser Pro Cys Phe Ser Ala
 100 105 110
 Gly Cys Ala Gly Glu Val Arg Ala Phe Leu Gln Glu Asn Thr His Val
 115 120 125
 Arg Leu Arg Ile Phe Ala Ala Arg Ile Tyr Asp Asp Asp Pro Leu Tyr
 130 135 140
 Lys Glu Ala Leu Gln Met Leu Arg Asp Ala Gly Ala Gln Val Ser Ile
 145 150 155 160
 Met Thr Tyr Asp Glu Phe Lys His Cys Trp Asp Thr Phe Val Asp His
 165 170 175
 Gln Gly Cys Pro Phe Gln Pro Trp Asp Gly Leu Asp Glu His Ser Gln
 180 185 190
 Ala Leu Ser Gly Arg Leu Arg Ala Ile Leu Gln Asn Gln Gly Asn Glu
 195 200 205
 Leu Lys Thr Pro Leu Gly Asp Thr Thr His Thr Ser Pro Pro Cys Pro
 210 215 220
 Ala Pro Glu Leu Leu Gly Gly Pro Met Ala Ser Asn Phe Thr Gln Phe
 225 230 235 240
 Val Leu Val Asp Asn Gly Gly Thr Gly Asp Val Thr Val Ala Pro Ser
 245 250 255
 Asn Phe Ala Asn Gly Ile Ala Glu Trp Ile Ser Ser Asn Ser Arg Ser
 260 265 270
 Gln Ala Tyr Lys Val Thr Cys Ser Val Arg Gln Ser Ser Ala Gln Asn
 275 280 285
 Arg Lys Tyr Thr Ile Lys Val Glu Val Pro Lys Gly Ala Trp Arg Ser
 290 295 300
 Tyr Leu Asn Met Glu Leu Thr Ile Pro Ile Phe Ala Thr Asn Ser Asp
 305 310 315 320
 Cys Glu Leu Ile Val Lys Ala Met Gln Gly Leu Leu Lys Asp Gly Asn
 325 330 335
 Pro Ile Pro Ser Ala Ile Ala Ala Asn Ser Gly Ile Tyr
 340 345

<210> 78

<211>	1008	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	含有来自Anolis Apobec1a和SV40 NLS的序列的质粒	
<400>	78	
atggccccc	agaagaagcg	gaaagtgggg
tatcaggctg	caattctatt	atcaaatttg
60		
ttcttcagg	ggcaaatgga	accagaggcg
ttcagagga	atcttgatcc	cagagaattt
120		
cccagagt	gta ctttactg	ctgtatgaa
atc cactggg	atacaacacc	ag taggaact
180		
tgtagaaca	aacctggcct	ccatgctgaa
gaaaat	tttt tgcaa	at tttt
240		
atagatat	ca ggcaggac	ac accatgct
tet atcact	ttgtgtct	tg gagtcct
300		
tatccatg	ca gccagget	at aattaagt
ttc ttgga	agcac accc	taacgt
360		
ataaaag	ctg ctcggct	gta catgca
atc gactg	ta acaagga	ag cctcag
420		
ttaggtag	aa atagagt	ttc tatcat
gaa ctac	ctgatt atcg	ccactg
480		
tttgtggt	ctc tagaggg	gcaaatga
agat tatt	ggccac	aggattt
540		
acaaatt	at ccaggga	act tgact
caatt cttc	aggac ag	ctgaagac
600		
gacaccac	acacctct	cc accctg
ggc gcacc	agagc	tgctggg
660		
agcaact	tca cacagt	ttgt gctg
gtgat	aatggag	gaa cggc
720		
ccatcta	act ttgcc	aatgg
catcgcc	gag tggat	cgct
780		
tataagg	tga cctgt	agcgt
gcggc	agct agc	gcccaga
840		
gtggagg	tg ctaagg	gcgc
ctggag	atcc tac	ctgaaca
900		
tttgcca	caa attct	gattg
cgagct	gatc gtga	aggcca
960		
ggcaacc	cta tcca	agcgc
catcgcc	gcaaat	agcgaa
1008		
<210>	79	
<211>	335	
<212>	PRT	
<213>	人工序列	
<220>		
<223>	含有来自Anolis Apobec1a和SV40 NLS的序列的蛋白	
<400>	79	
Met Ala Pro Lys Lys Lys Arg Lys Val Gly Tyr Gln Ala Ala Ile Leu		
1 5 10 15		
Leu Ser Asn Leu Phe Phe Arg Trp Gln Met Glu Pro Glu Ala Phe Gln		
20 25 30		
Arg Asn Phe Asp Pro Arg Glu Phe Pro Glu Cys Thr Leu Leu Leu Tyr		
35 40 45		
Glu Ile His Trp Asp Asn Asn Thr Ser Arg Asn Trp Cys Thr Asn Lys		
50 55 60		
Pro Gly Leu His Ala Glu Glu Asn Phe Leu Gln Ile Phe Asn Glu Lys		

65	70	75	80
Ile Asp Ile Arg Gln Asp Thr Pro Cys Ser Ile Thr Trp Phe Leu Ser			
	85	90	95
Trp Ser Pro Cys Tyr Pro Cys Ser Gln Ala Ile Ile Lys Phe Leu Glu			
	100	105	110
Ala His Pro Asn Val Ser Leu Glu Ile Lys Ala Ala Arg Leu Tyr Met			
	115	120	125
His Gln Ile Asp Cys Asn Lys Glu Gly Leu Arg Asn Leu Gly Arg Asn			
	130	135	140
Arg Val Ser Ile Met Asn Leu Pro Asp Tyr Arg His Cys Trp Thr Thr			
	145	150	155
160			
Phe Val Val Pro Arg Gly Ala Asn Glu Asp Tyr Trp Pro Gln Asp Phe			
	165	170	175
Leu Pro Ala Ile Thr Asn Tyr Ser Arg Glu Leu Asp Ser Ile Leu Gln			
	180	185	190
Asp Glu Leu Lys Thr Pro Leu Gly Asp Thr Thr His Thr Ser Pro Pro			
	195	200	205
Cys Pro Ala Pro Glu Leu Leu Gly Gly Pro Met Ala Ser Asn Phe Thr			
	210	215	220
Gln Phe Val Leu Val Asp Asn Gly Gly Thr Gly Asp Val Thr Val Ala			
	225	230	235
240			
Pro Ser Asn Phe Ala Asn Gly Ile Ala Glu Trp Ile Ser Ser Asn Ser			
	245	250	255
Arg Ser Gln Ala Tyr Lys Val Thr Cys Ser Val Arg Gln Ser Ser Ala			
	260	265	270
Gln Asn Arg Lys Tyr Thr Ile Lys Val Glu Val Pro Lys Gly Ala Trp			
	275	280	285
Arg Ser Tyr Leu Asn Met Glu Leu Thr Ile Pro Ile Phe Ala Thr Asn			
	290	295	300
Ser Asp Cys Glu Leu Ile Val Lys Ala Met Gln Gly Leu Leu Lys Asp			
	305	310	315
320			
Gly Asn Pro Ile Pro Ser Ala Ile Ala Ala Asn Ser Gly Ile Tyr			
	325	330	335

<210> 80

<211> 20

<212> DNA

<213> Homo sapiens

<400> 80

tgttccagtt tcctttacag

20

<210> 81	
<211> 20	
<212> DNA	
<213> Homo sapiens	
<400> 81	
cagccccgctg gccctgtaaa	20
<210> 82	
<211> 20	
<212> DNA	
<213> Homo sapiens	
<400> 82	
gccagagccg ggggtgtgcag	20
<210> 83	
<211> 20	
<212> DNA	
<213> Homo sapiens	
<400> 83	
ggaagtgtcc agggatgctt	20
<210> 84	
<211> 20	
<212> DNA	
<213> Homo sapiens	
<400> 84	
tcgcgccctc ccagccgggc	20
<210> 85	
<211> 23	
<212> DNA	
<213> Homo sapiens	
<400> 85	
ctcctggacc ccctatttct gac	23
<210> 86	
<211> 23	
<212> DNA	
<213> Homo sapiens	
<400> 86	
gccagagccg ggggtgtgcag acg	23
<210> 87	
<211> 20	
<212> DNA	

<213> 人工序列	
<220>	
<223> 合成	
<400> 87	
aggacgtctg cccaatatgt	20
<210> 88	
<211> 20	
<212> DNA	
<213> 人工序列	
<220>	
<223> 合成	
<400> 88	
ccaagtgaga agccagtgga	20
<210> 89	
<211> 20	
<212> DNA	
<213> 人工序列	
<220>	
<223> 合成	
<400> 89	
tcttcctcc cagtcactga	20
<210> 90	
<211> 20	
<212> DNA	
<213> 人工序列	
<220>	
<223> 合成	
<400> 90	
tcactctcga agacgctgct	20
<210> 91	
<211> 22	
<212> DNA	
<213> 人工序列	
<220>	
<223> 合成	
<400> 91	
cgatgaggag aactccaag ag	22
<210> 92	
<211> 22	

<212> DNA	
<213> 人工序列	
<220>	
<223> 合成	
<400> 92	
gcctttctga aggtcatagt gc	22
<210> 93	
<211> 20	
<212> DNA	
<213> Homo sapiens	
<400> 93	
aggctggccc gccccgcagt	20
<210> 94	
<211> 20	
<212> DNA	
<213> Homo sapiens	
<400> 94	
ccgatattcc tcaggtactc	20
<210> 95	
<211> 20	
<212> DNA	
<213> Homo sapiens	
<400> 95	
aggtttactc acgtcatcca	20
<210> 96	
<211> 19	
<212> DNA	
<213> Unknown	
<220>	
<223> Qbeta噬菌体操作子茎环	
<400> 96	
atgctgtcta agacagcat	19
<210> 97	
<211> 133	
<212> PRT	
<213> Unknown	
<220>	
<223> Qbeta包被蛋白	
<400> 97	

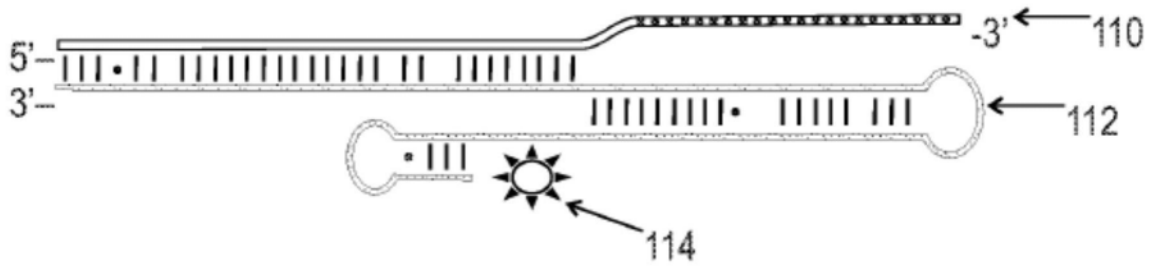


图1A

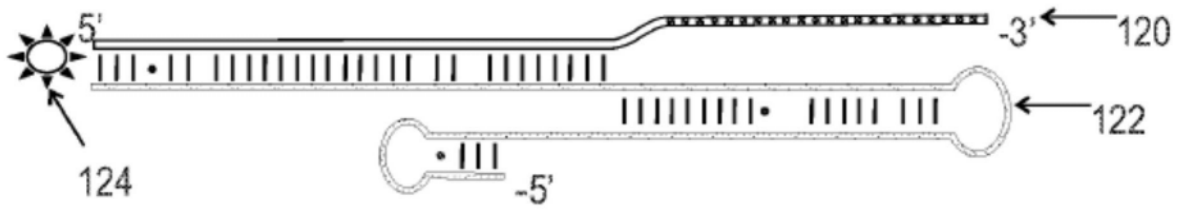


图1B

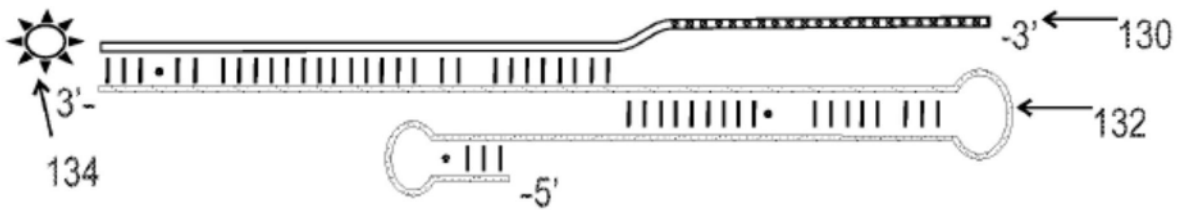


图1C

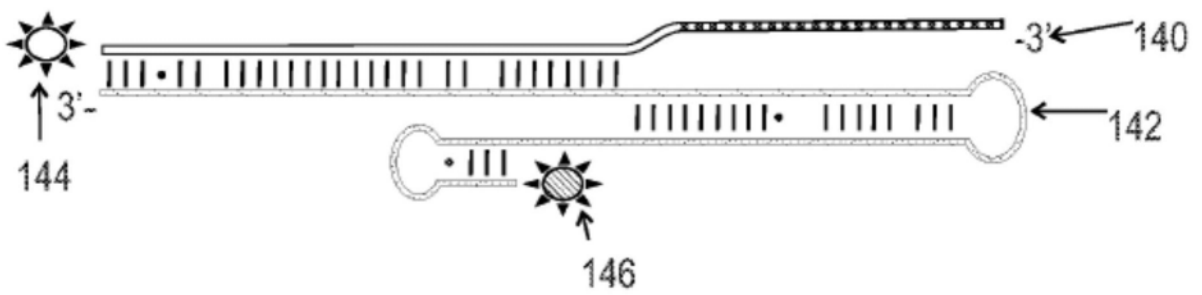


图1D

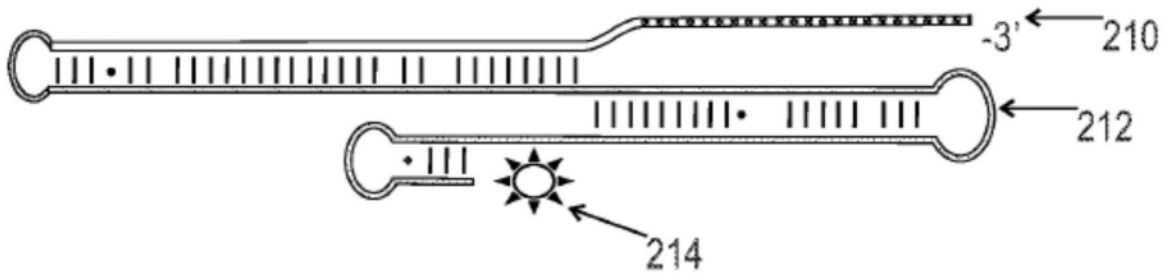


图2A

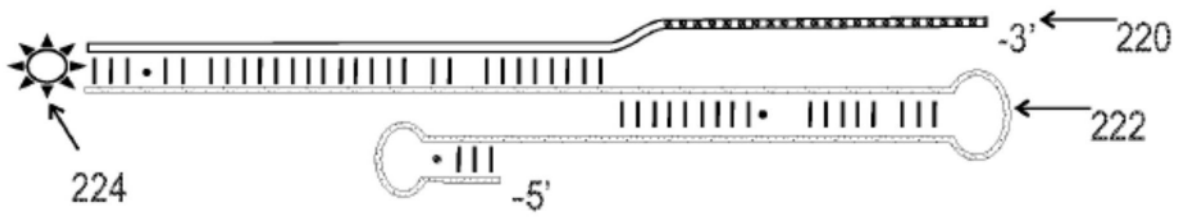


图2B

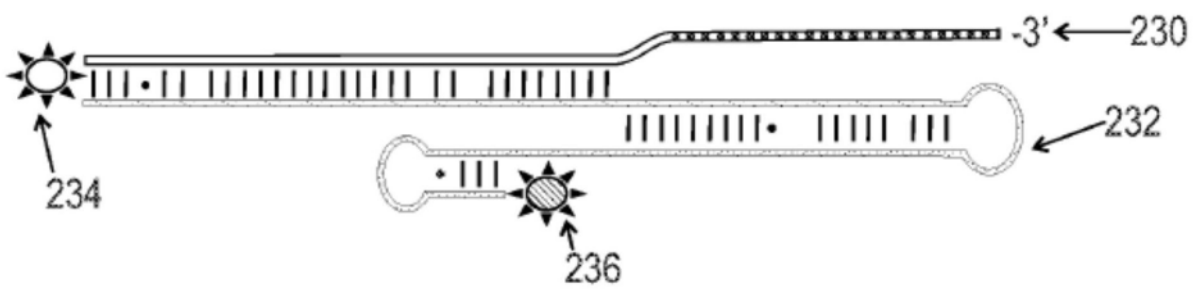


图2C

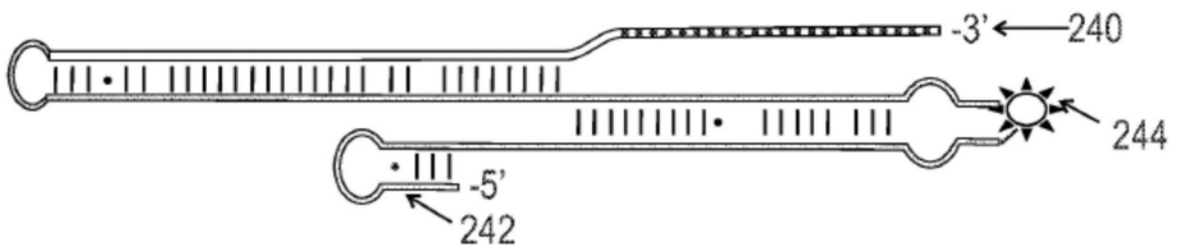


图2D

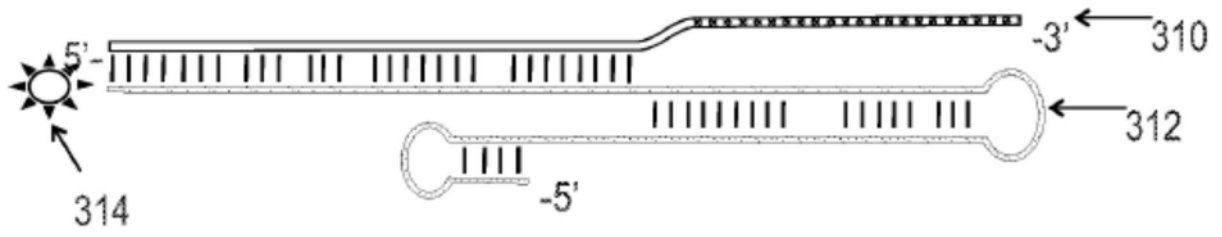


图3A

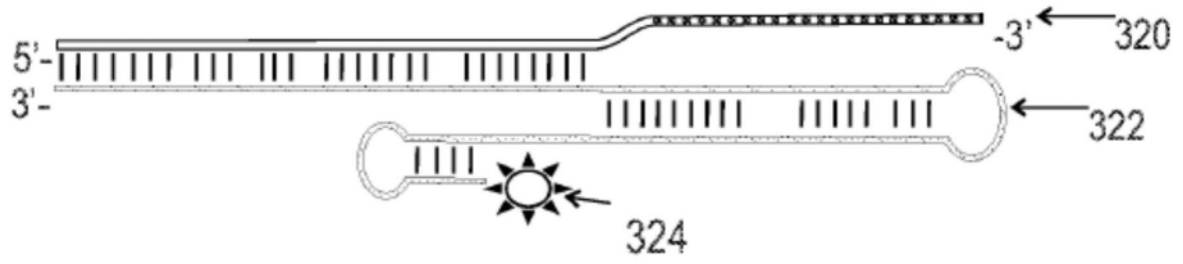


图3B

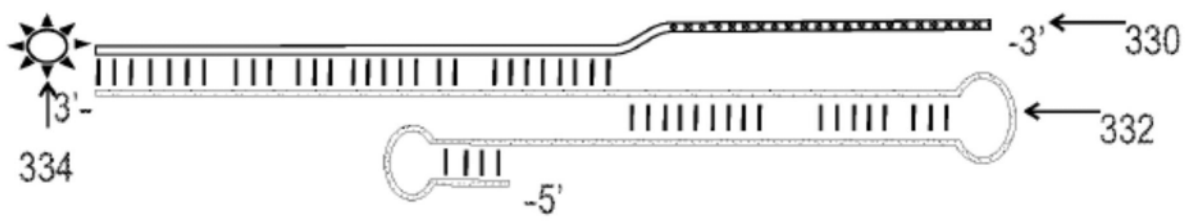


图3C

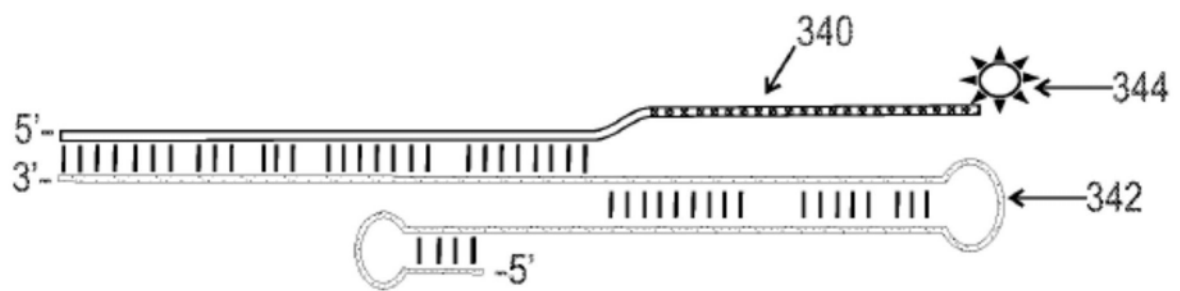


图3D

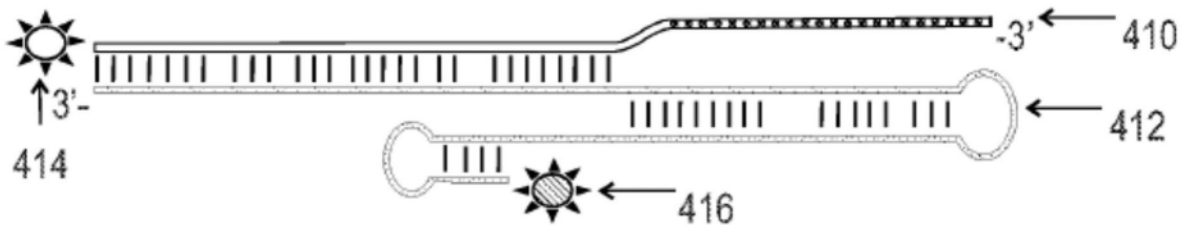


图4A

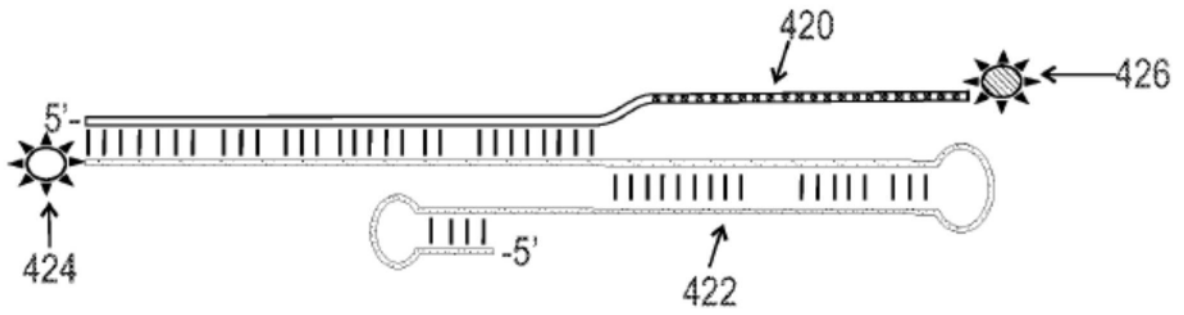


图4B

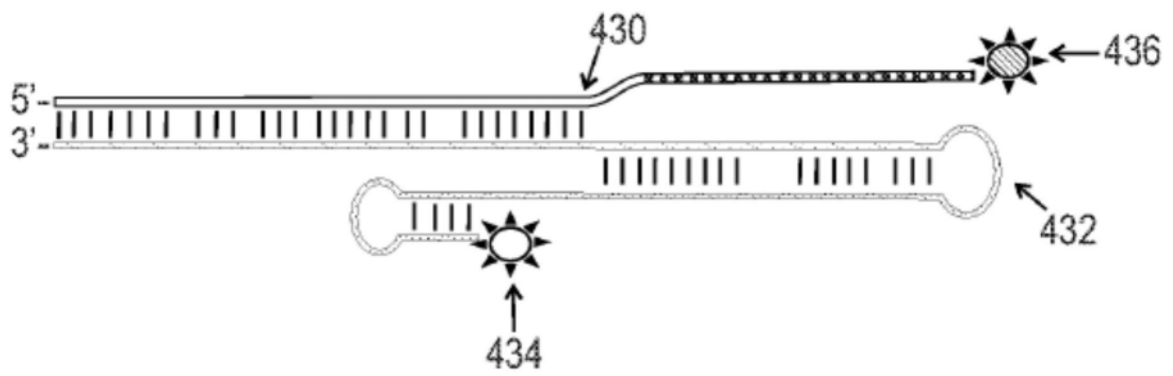


图4C

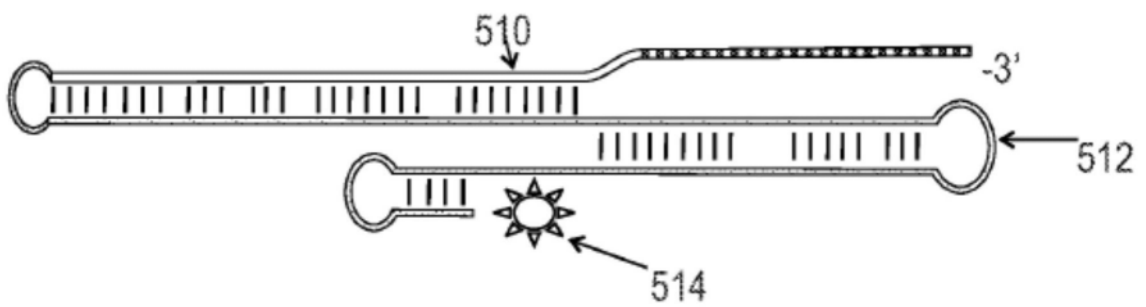


图5A

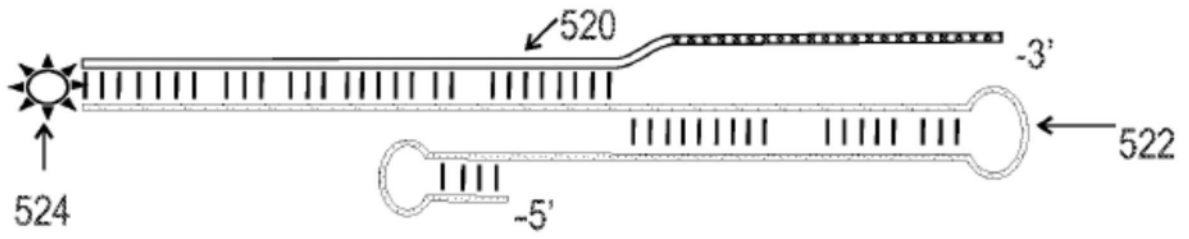


图5B

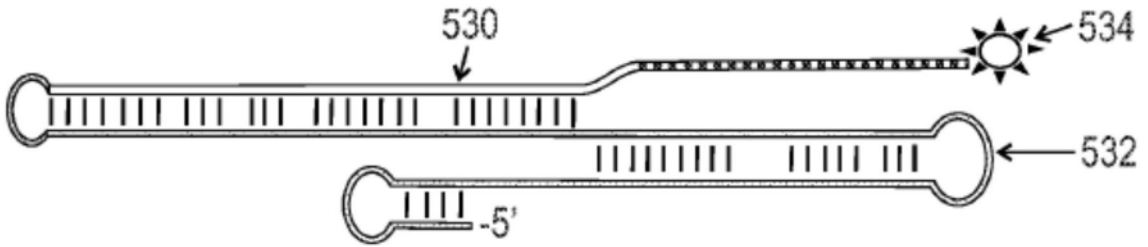


图5C

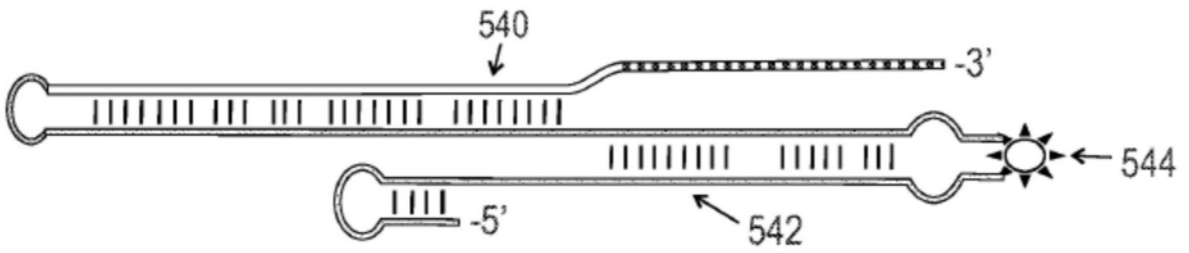


图5D

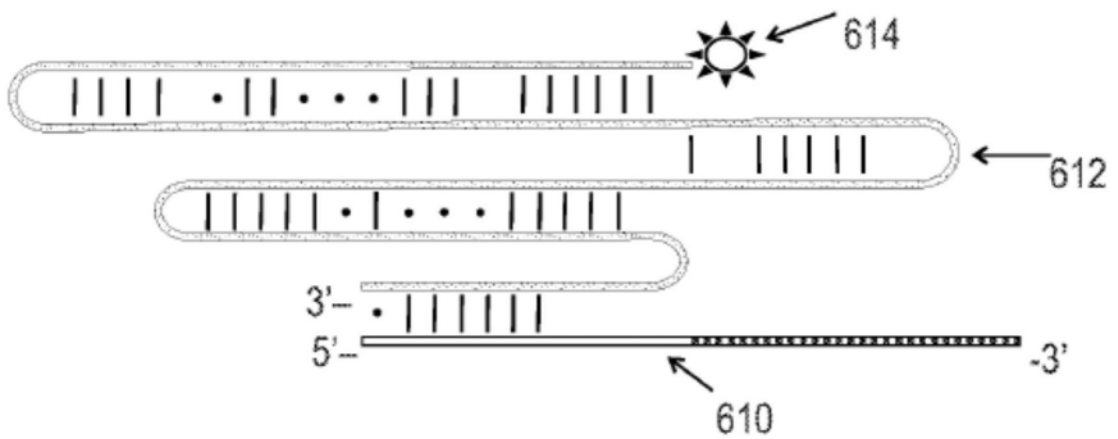


图6A

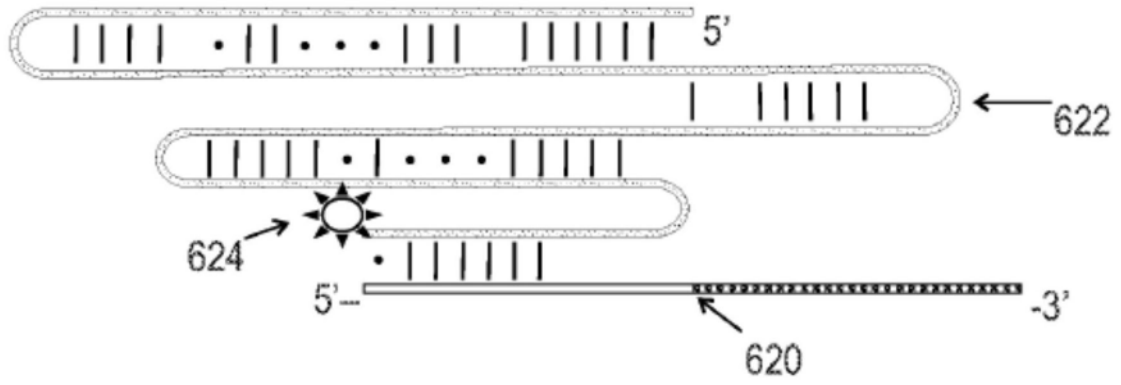


图6B

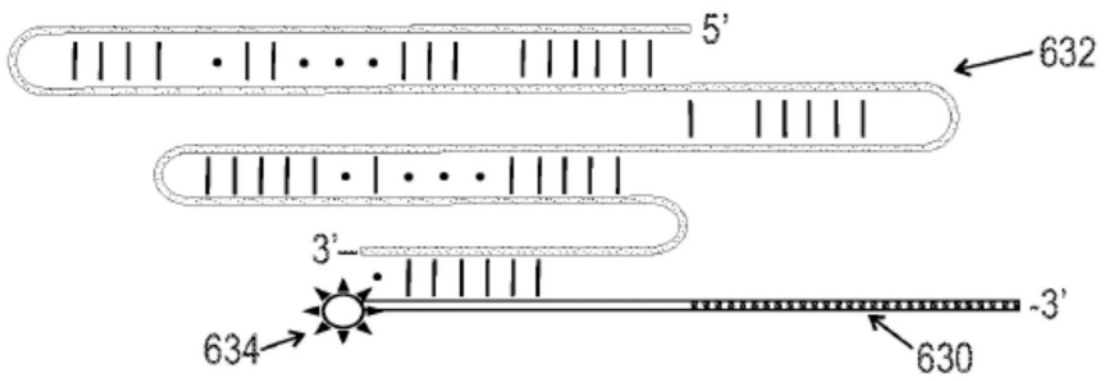


图6C

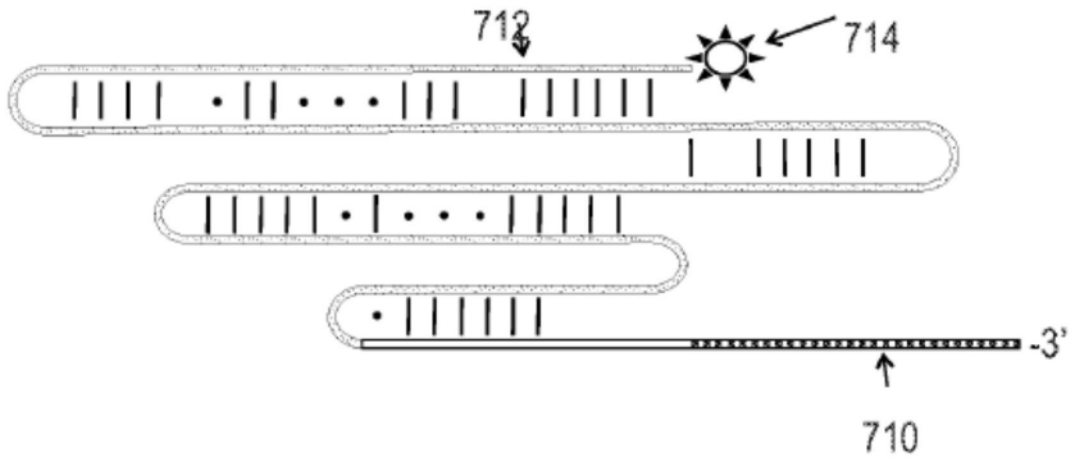


图7A

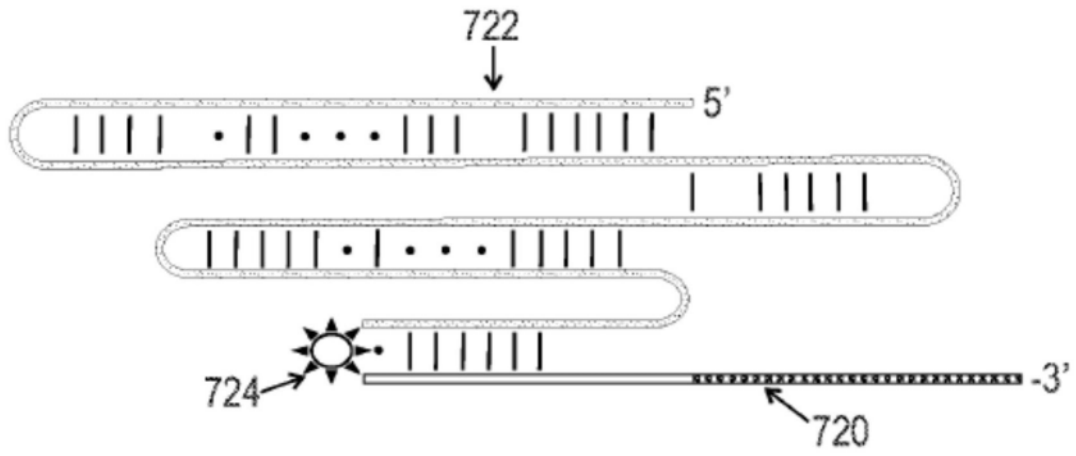


图7B

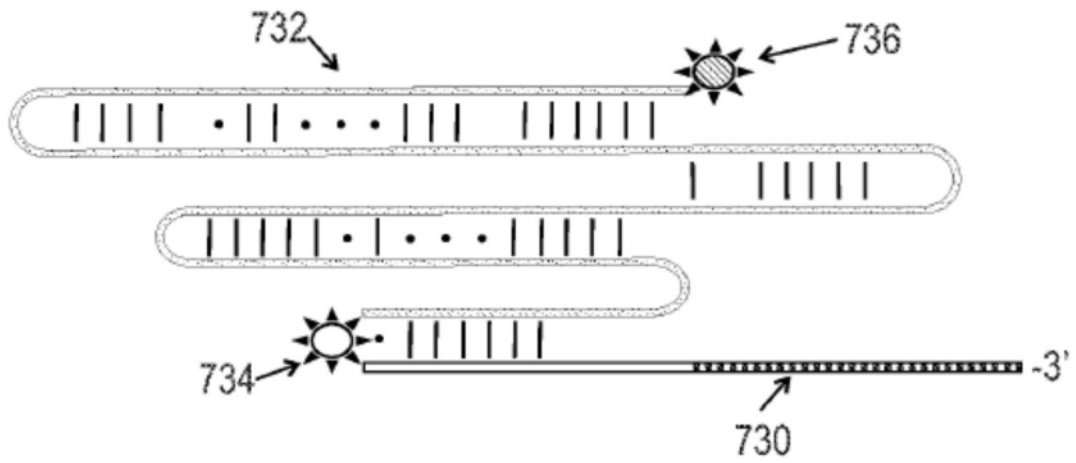


图7C

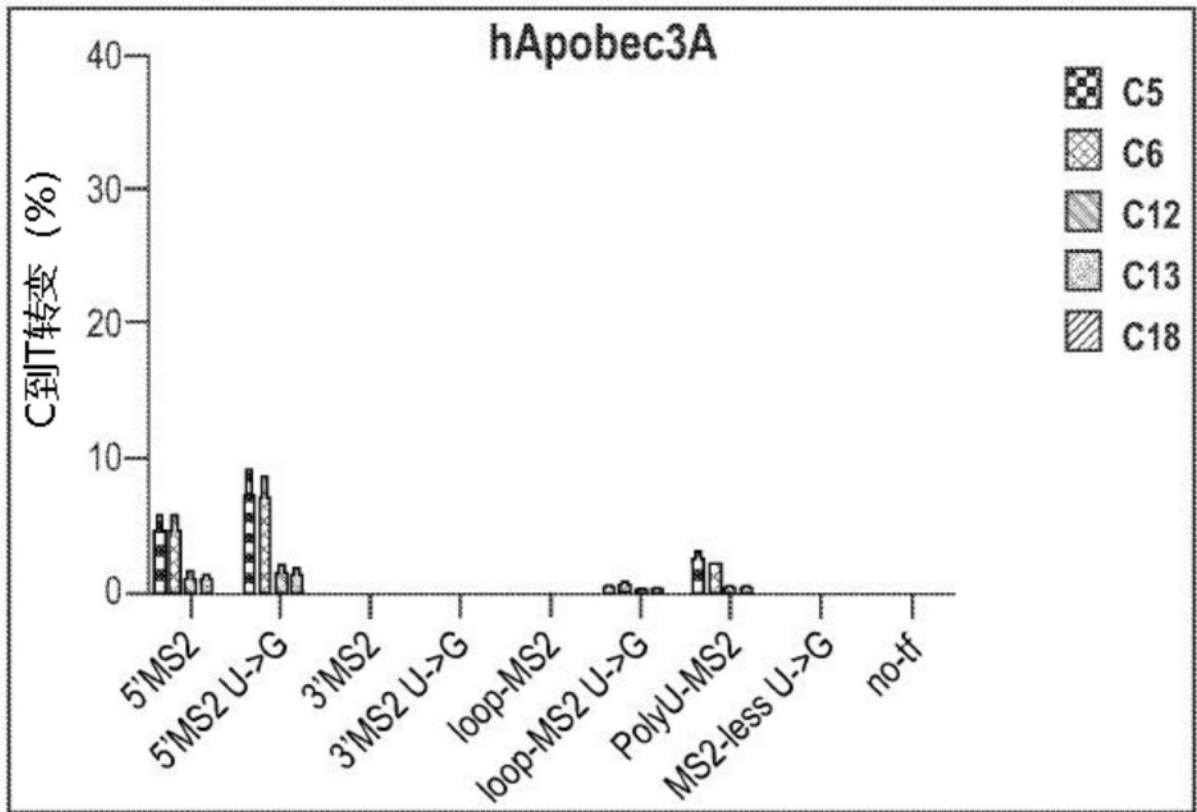


图8A

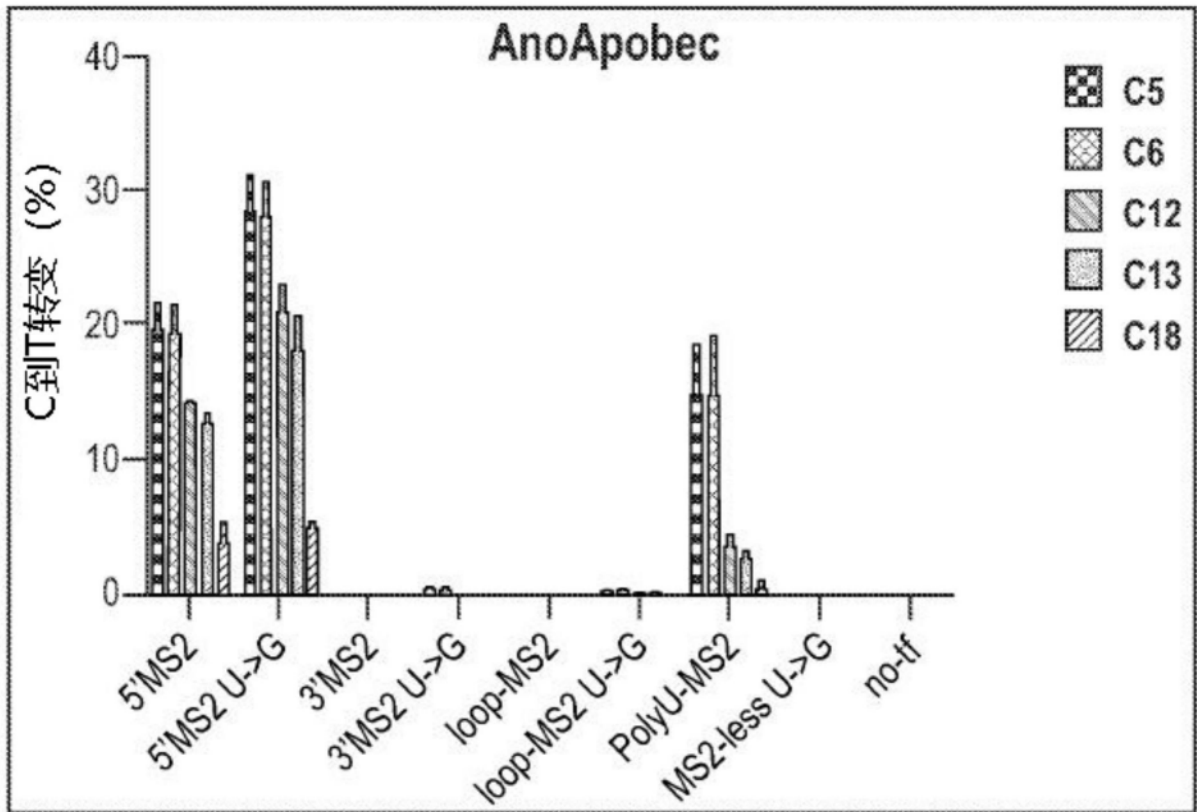


图8B

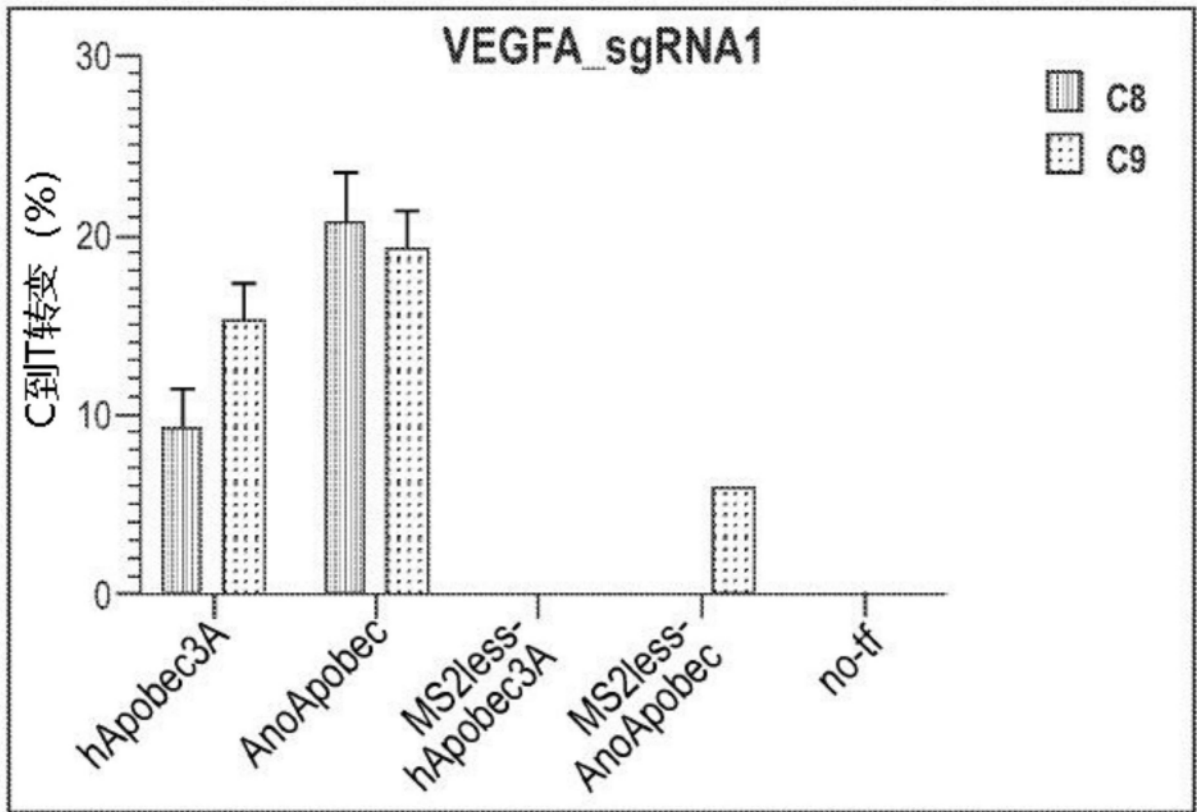


图9A

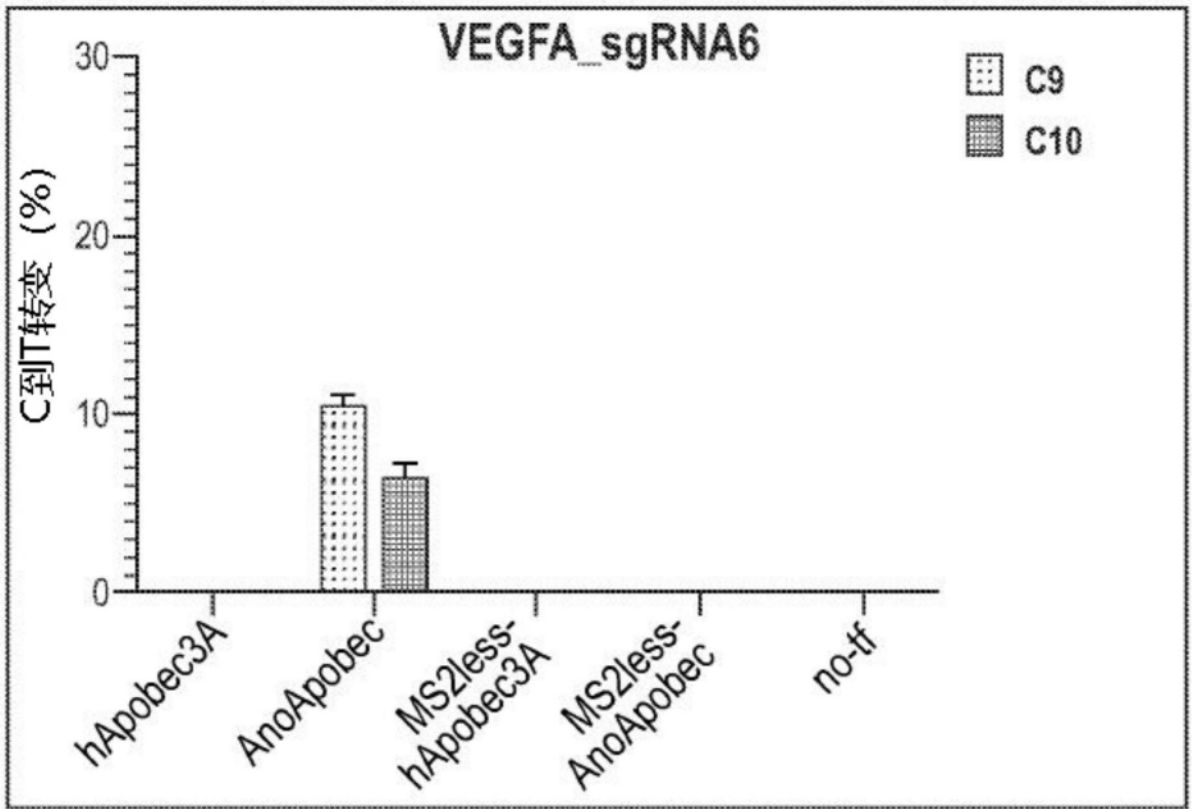


图9B

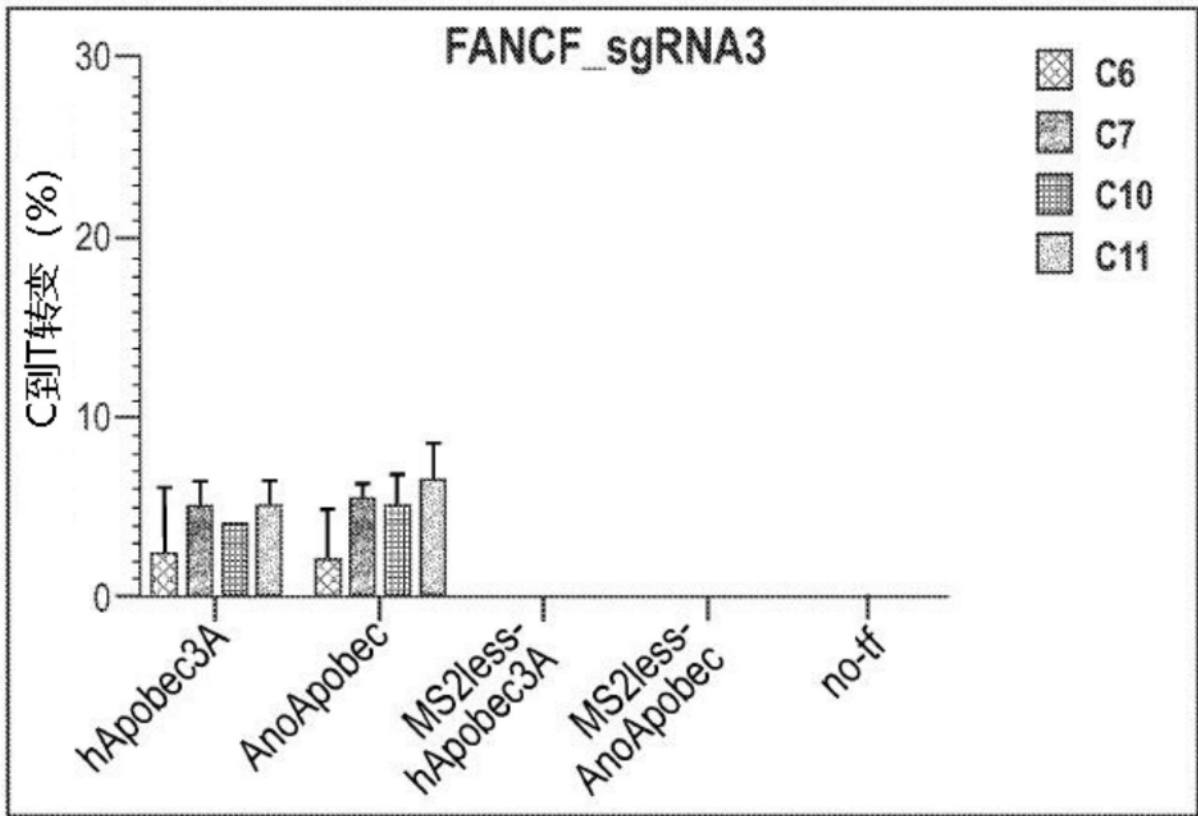


图9C

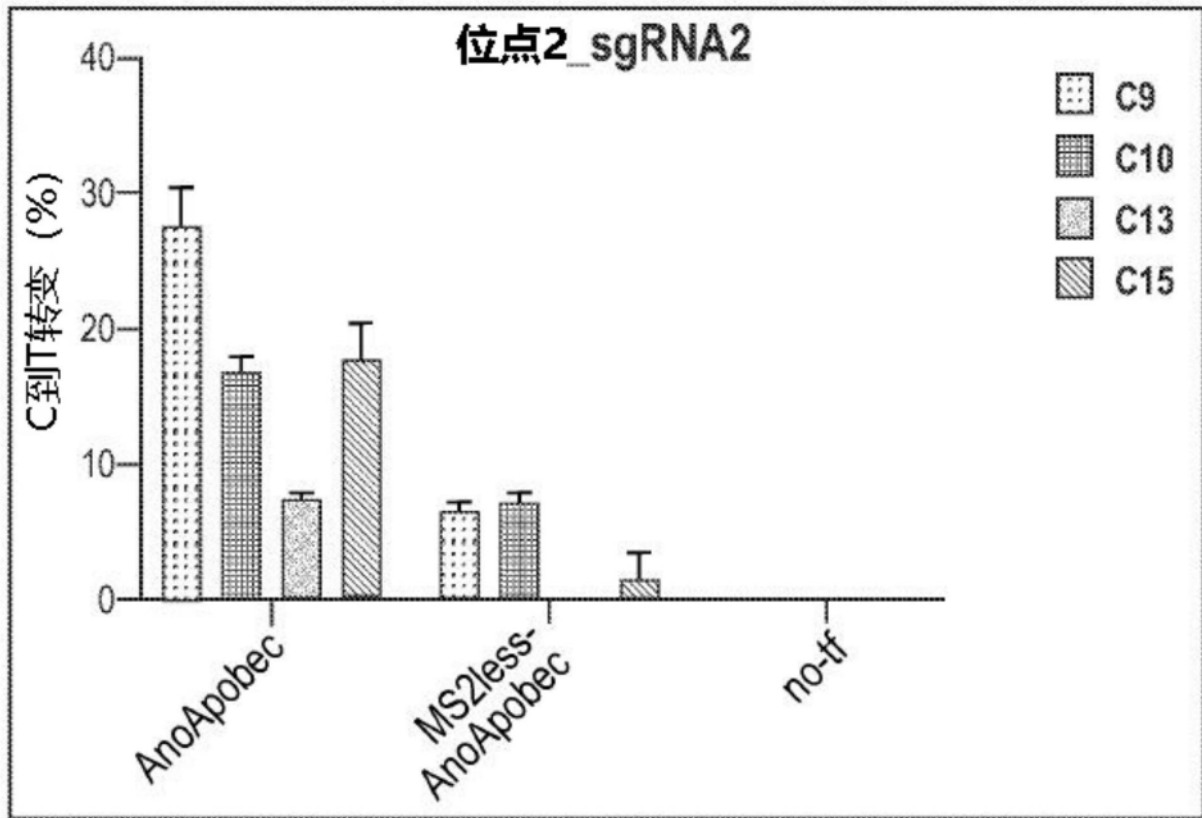


图9D

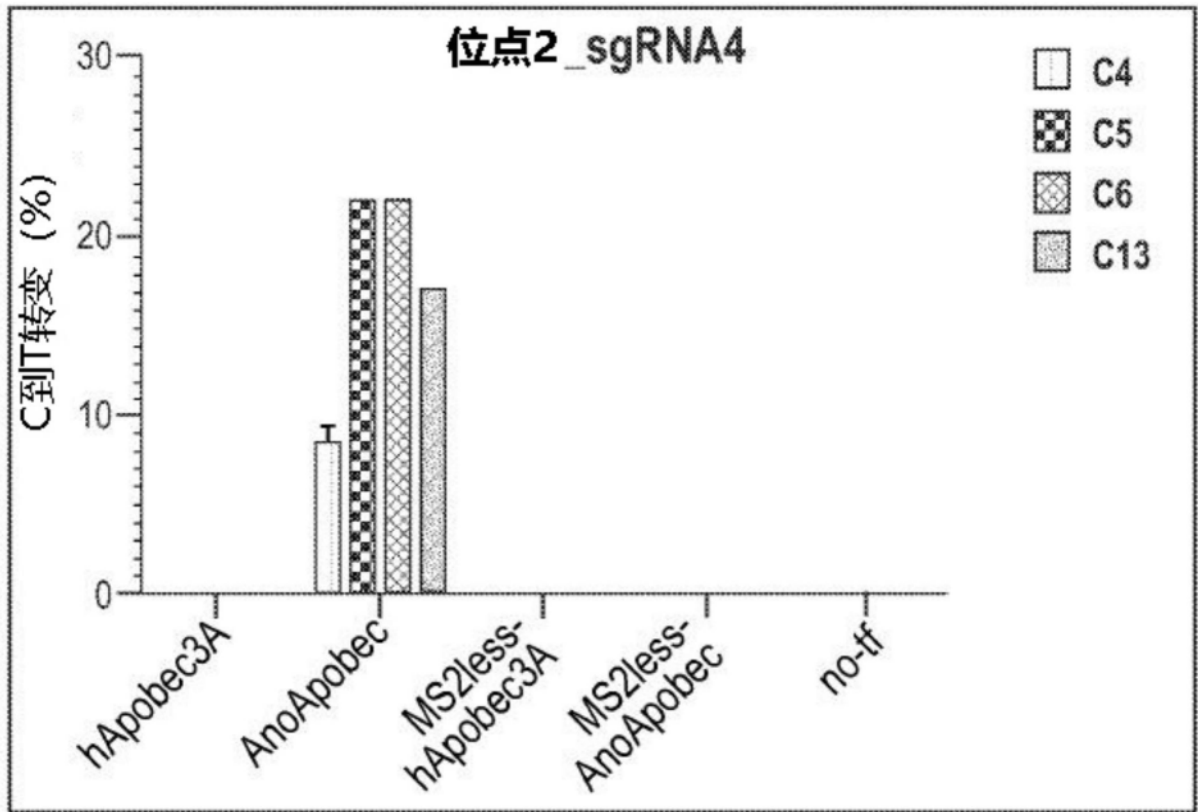


图9E

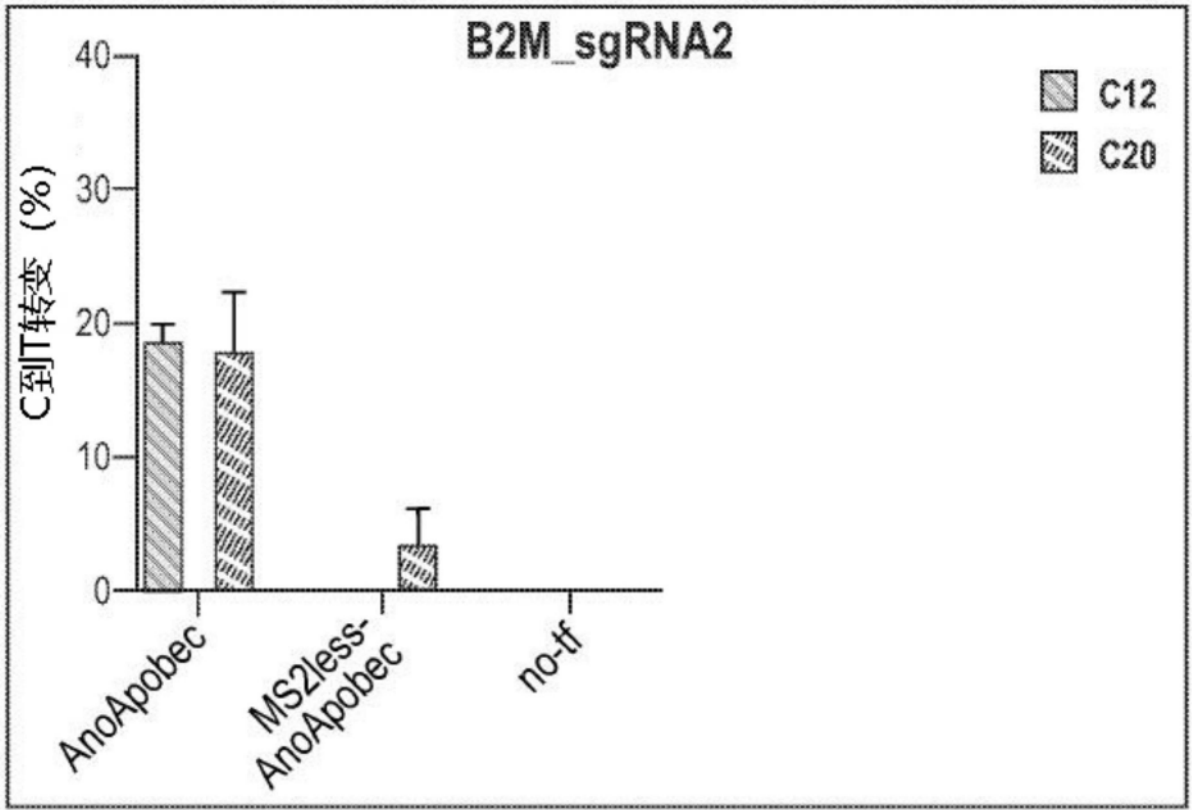


图9F

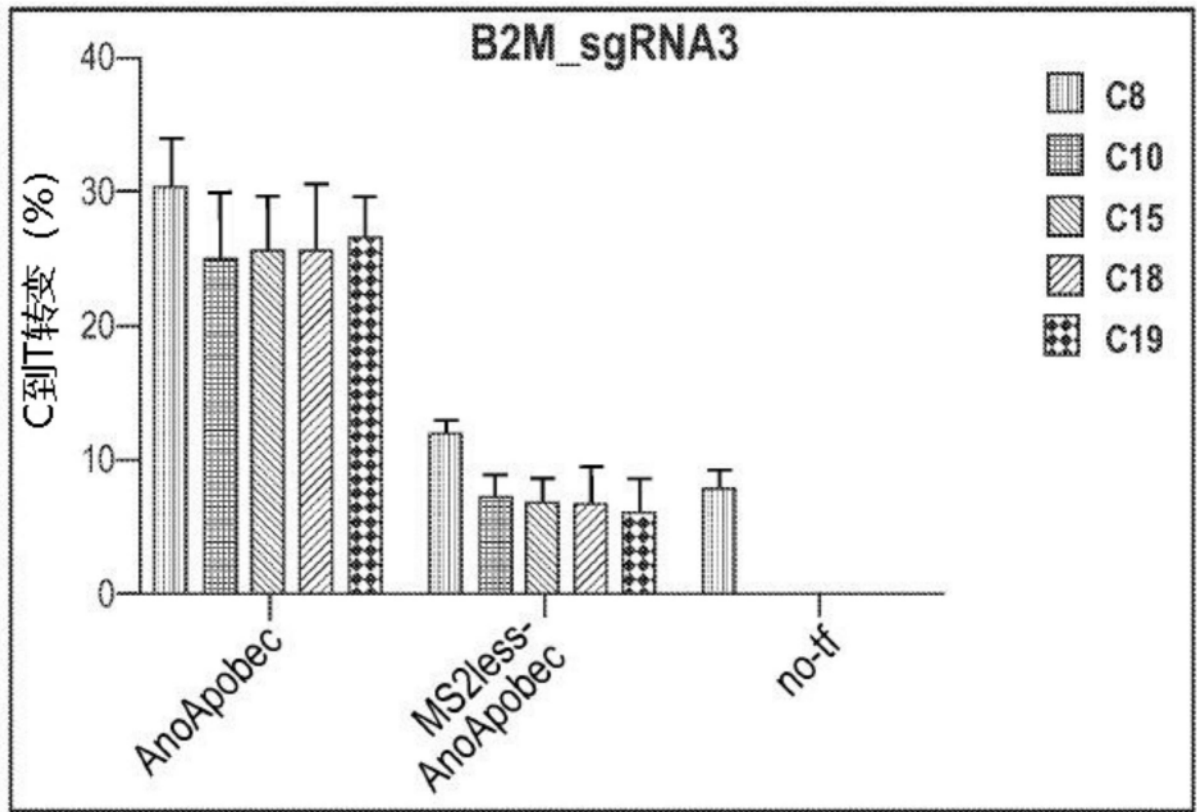


图9G

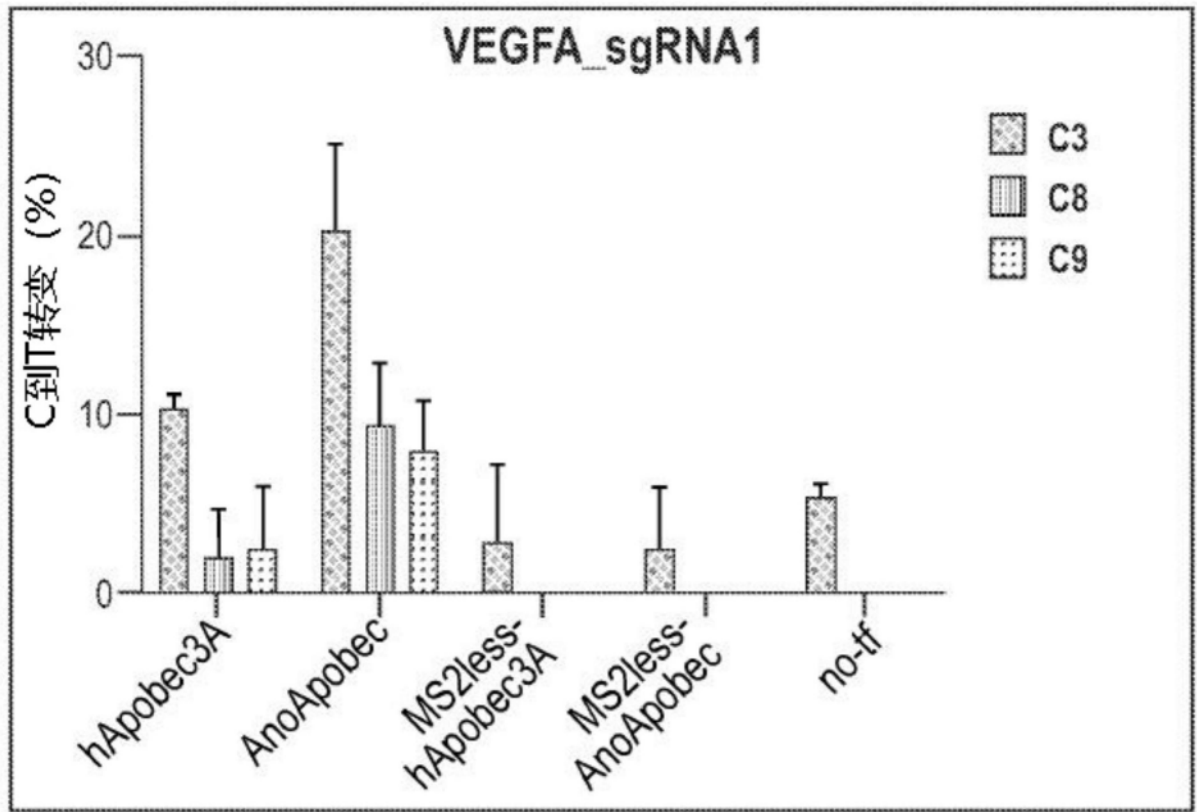


图10

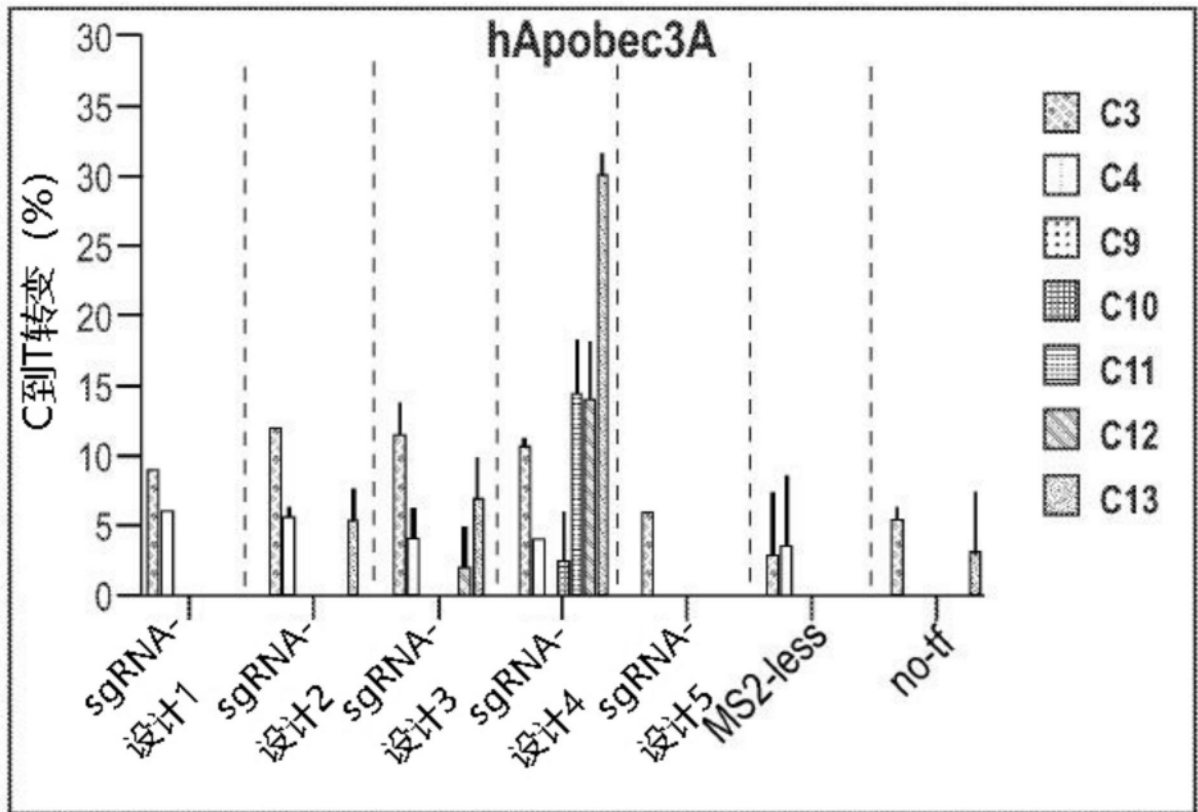


图11A

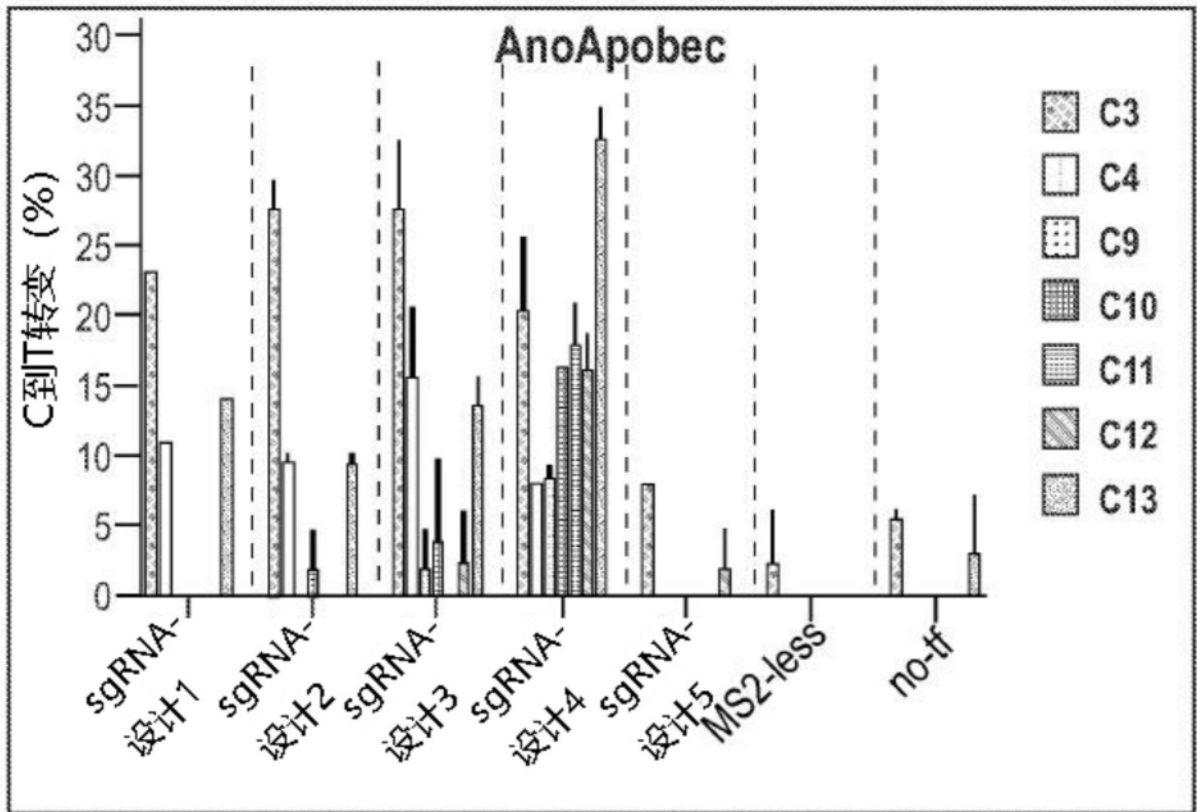


图11B

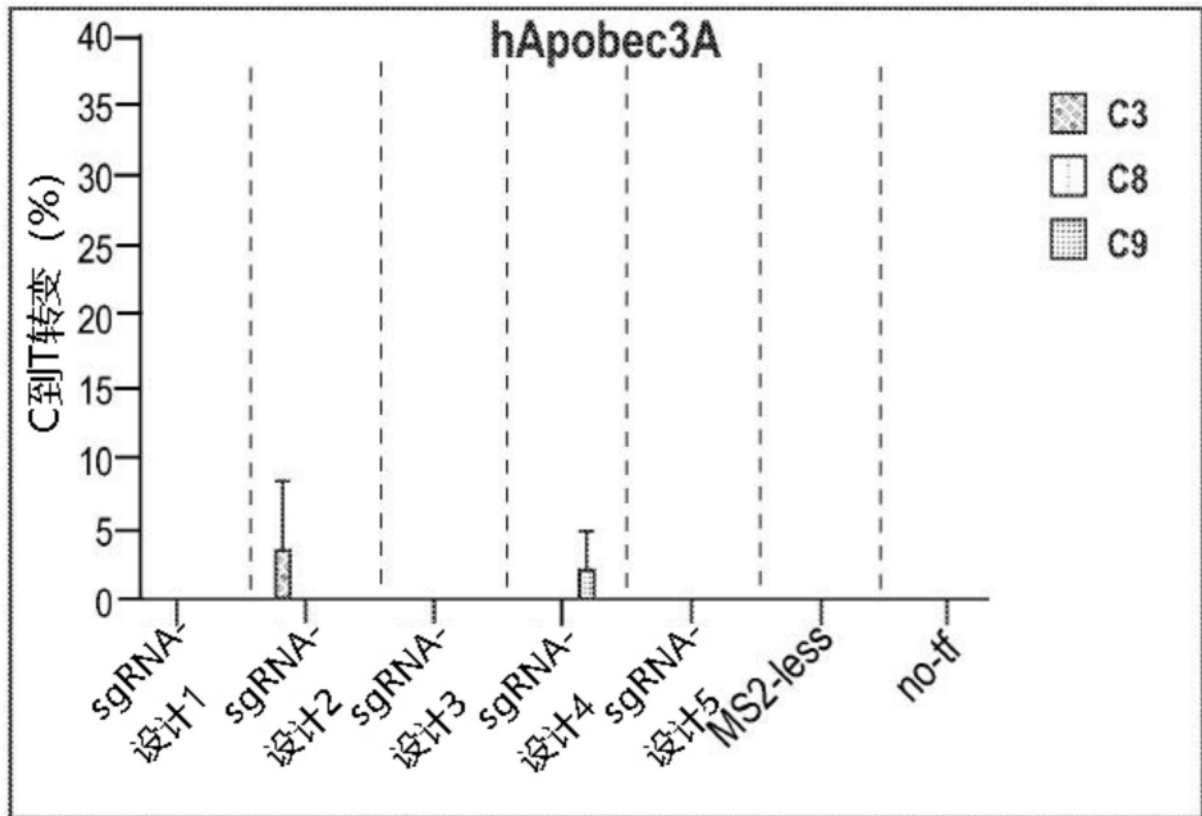


图11C

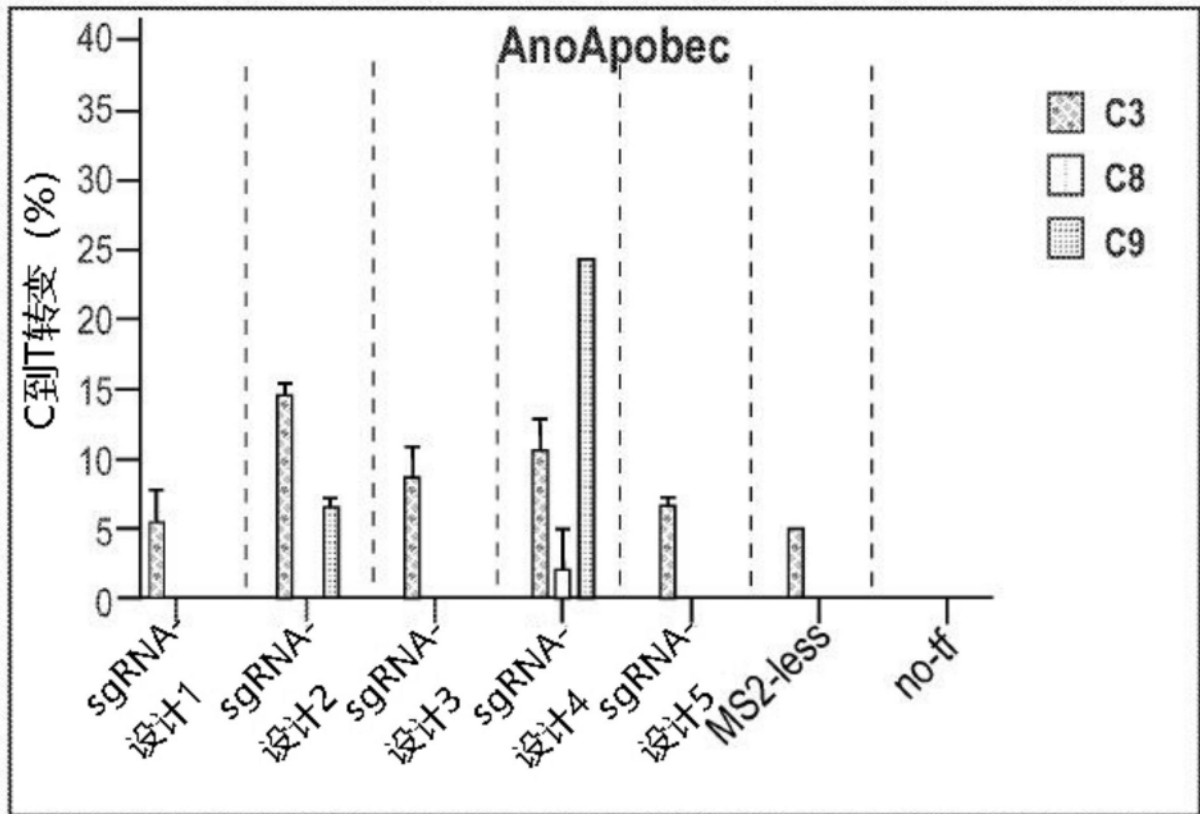


图11D