(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 114040977 B
(45)授权公告日 2024.09.03

(21)申请号 202080026678.0
(22)申请日 2020.01.31
(65)同一申请的已公布的文献号
申请公布号 CN 114040977 A
(43)申请公布日 2022.02.11
(30)优先权数据
62/799,702 2019.01.31 US
(85)PCT国际申请进入国家阶段日
2021.09.30
(86)PCT国际申请的申请数据
PCT/US2020/016285 2020.01.31
(87)PCT国际申请的公布数据
WO2020/160514 EN 2020.08.06
(73)专利权人 比姆医疗股份有限公司
地址 美国马萨诸塞州
(72)发明人 I·斯雷梅克 J·M·格尔克
N·戈代尔利 Y·于
(74)专利代理机构 北京戈程知识产权代理有限公司 11314
专利代理师 程伟
(51)Int.Cl.
C12N 15/62(2006.01)
C12N 15/11(2006.01)
C12N 15/10(2006.01)
(56)对比文件
CN 108290933 A,2018.07.17
CN 108699116 A,2018.10.23
审查员 许文彦

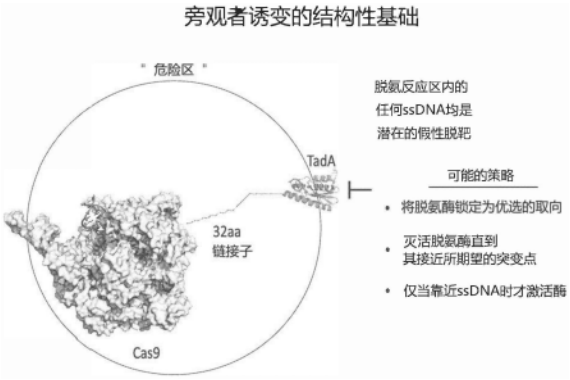
权利要求书2页 说明书235页
序列表（电子公布） 附图52页

(54)发明名称

非靶脱氨反应减低的核碱基编辑器和用于定性核碱基编辑器的测定

(57)摘要

本发明的特征在于具有减低的非靶脱氨反应的碱基编辑器、使用所述碱基编辑器的方法和用于定性具有降低的非靶脱氨反应(例如,与程式化的在靶脱氨反应相比)的碱基编辑器的测定。



CN 114040977 B

1.一种融合蛋白,其包括插入Cas9多肽的柔性环内的脱氨酶,其中该融合蛋白由以下结构组成:

a)NH_2-[Cas9的N末端片段]-[TadA腺苷脱氨酶]-[Cas9的C末端片段]-[核定位信号(nuclear localization signal,NLS)]-COOH,或

NH_2-[Cas9的N末端片段]-[TadA腺苷脱氨酶]-[Cas9的C末端片段]-COOH,

其中该TadA腺苷脱氨酶是插入如SEQ ID NO:1中所编号的氨基酸位置768至769、1015至1016、1022至1023、1026至1027、1029至1030、1040至1041、1054至1055、1068至1069或1247至1248之间;或

b)NH_2-[Cas9的N末端片段]-[APOBEC胞苷脱氨酶]-[Cas9的C末端片段]-[尿嘧啶糖基化酶抑制剂(uracil glycosylase inhibitor,UGI)]-[UGI]-[NLS]-COOH,

或NH_2-[Cas9的N末端片段]-[APOBEC胞苷脱氨酶]-[Cas9的C末端片段]-[UGI]-[UGI]-COOH,

或NH_2-[Cas9的N末端片段]-[APOBEC胞苷脱氨酶]-[Cas9的C末端片段]-[UGI]-[NLS]-COOH,

或NH_2-[Cas9的N末端片段]-[APOBEC胞苷脱氨酶]-[Cas9的C末端片段]-[UGI]-COOH,

其中该APOBEC胞苷脱氨酶是插入如SEQ ID NO:1中所编号的氨基酸位置1016至1017、1023至1024、1029至1030或1069至1070之间;

其中"]-["的每个实例是连接子。

2.如权利要求1所述的融合蛋白,其中该Cas9多肽的N末端片段或C末端片段包括DNA结合结构域。

3.如权利要求1所述的融合蛋白,其中该Cas9多肽的N末端片段或该C末端片段包括RuvC结构域。

4.如权利要求1所述的融合蛋白,其中该Cas9多肽的N末端片段或该C末端片段包括HNH结构域。

5.如权利要求1所述的融合蛋白,其中该APOBEC胞苷脱氨酶是APOBEC1、APOBEC2、APOBEC3A、APOBEC3B、APOBEC3C、APOBEC3D、APOBEC3E、APOBEC3F、APOBEC3G、APOBEC3H或APOBEC4。

6.如权利要求1所述的融合蛋白,其中该APOBEC胞苷脱氨酶是rAPOBEC1。

7.一种融合蛋白,其包括插入Cas9多肽内的脱氨酶,其中该融合蛋白由以下结构组成:

NH_2-[Cas9的N末端片段]-[TadA腺苷脱氨酶]-[Cas9的C末端片段]-[核定位信号(nuclear localization signal,NLS)]-COOH,或

NH_2-[Cas9的N末端片段]-[TadA腺苷脱氨酶]-[Cas9的C末端片段]-COOH,

其中"]-["的每个实例是连接子,

其中该Cas9多肽如SEQ ID NO:1中所编号的氨基酸位置792至872缺失或792到906缺失,且

该TadA腺苷脱氨酶是插入到该缺失的位置处。

8.如权利要求1或7所述的融合蛋白,其中该TadA腺苷脱氨酶是由以下氨基酸序列组成:

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEI MALRQGGLVMQ

NYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDV LHYPGMNHRVEITEGILADECAALLC YFFRMPRQVFNAQKKAQSSTD(SEQ ID NO:19)。

9.如权利要求1或7所述的融合蛋白,其中该连接子为(SGGS)n、(GGGS)n、(GGGGS)n、(G)n、(EAAAK)n、(GGS)n、SGSETPGTSESATPES或(XP)n基序,或其组合,其中n独立地是1和30之间的整数。

10.如权利要求1或7所述的融合蛋白,其中该Cas9多肽的N末端片段在没有连接子的情况下与该TadA腺苷脱氨酶或APOBEC胞苷脱氨酶融合。

11.如权利要求1或7所述的融合蛋白,其中该Cas9多肽的C末端片段在没有连接子的情况下与该TadA腺苷脱氨酶或APOBEC胞苷脱氨酶融合。

12.如权利要求1或7所述的融合蛋白,其中该核定位信号是双分型核定位信号。

13.如权利要求1或7所述的融合蛋白,其中该Cas9多肽是化脓性链球菌Cas9(SpCas9)、金黄色葡萄球菌Cas9(SaCas9)或嗜热链球菌1Cas9(St1Cas9)。

14.如权利要求1或7所述的融合蛋白,其中该Cas9多肽是spCas9并且对由除NGG之外的核苷酸序列组成的原间隔序列临近基序具有特异性,或者其中该Cas9多肽是saCas9并且对由除NNGRRT之外的核苷酸序列组成的原间隔序列临近基序具有特异性,其中“N”是A、C、G或T,“R”是A或G。

15.如权利要求1或7所述的融合蛋白,其中该Cas9多肽是切口酶。

16.如权利要求1或7所述的融合蛋白,其中该Cas9多肽是核酸酶无活性的。

17.如权利要求1或7所述的融合蛋白,其与向导核酸序列复合。

18.如权利要求17所述的融合蛋白,还与靶多核苷酸复合。

19.一种多核苷酸,其编码如权利要求1或7所述的融合蛋白。

20.一种表达载体,其包括如权利要求19所述的多核苷酸。

21.如权利要求20所述的表达载体,其中该表达载体是选自腺-相关病毒(AAV)、逆转录病毒载体、腺病毒载体、慢病毒载体、仙台病毒载体或疱疹病毒载体。

22.如权利要求20或21所述的表达载体,其包括启动子。

23.一种细胞,其包括如权利要求1至18中任一项所述的融合蛋白、如权利要求19所述的多核苷酸或如权利要求20至22中任一项所述的表达载体。

24.如权利要求23所述的细胞,其中该细胞是细菌细胞、植物细胞、昆虫细胞或哺乳动物细胞。

25.一种试剂盒,其包括如权利要求1至18中任一项所述的融合蛋白、如权利要求19所述的多核苷酸或如权利要求20至22中任一项所述的表达载体。

26.一种用于体外碱基编辑的方法,该方法包括使多核苷酸序列与如权利要求1至18中任一项所述的融合蛋白接触,其中该融合蛋白的脱氨酶使多核苷酸中的核碱基脱氨基,从而编辑了该多核苷酸序列。

27.如权利要求26所述的方法,还包括使该多核苷酸序列与向导核酸序列接触以达成核碱基的脱氨反应。

28.如权利要求27所述的方法,其中该接触是在细胞中执行。

29.如权利要求28所述的方法,其中该细胞是哺乳动物细胞。

30.如权利要求29所述的方法,其中该细胞是分化多能的细胞。

非靶脱氨反应减低的核碱基编辑器和用于定性核碱基编辑器的测定

技术领域

[0001] 相关申请的交叉引用

[0002] 本申请要求于2019年1月31日所提交的美国临时专利申请第62/799,702号的权益,其全部内容通过引用并入本文。

背景技术

[0003] 结合了CRISPR-Cas蛋白的精确靶向的脱氨酶,称为核碱基编辑器,能够将特定的点突变引入靶多核苷酸内。核碱基编辑器在不引入双-股DNA断裂的情况下诱导碱基变化,且包含将靶A•T转换为G•C的腺苷碱基编辑器和将靶C•G转换为T•A的胞苷碱基编辑器。然而,在细胞中引入核碱基编辑器有潜力生成不希望得到的碱基编辑器-相关联的编辑,其包含全基因组的假性脱氨反应、旁观者突变和靶近端编辑。假性脱氨反应事件可能发生在整个基因组中,其经由向导RNA对CRISPR-Cas结构域进行编程而被所述碱基编辑器脱氨反应结构域所催化,且独立于靶向的碱基编辑而起作用。不受理论束缚,全基因组的假性脱氨反应事件有潜力发生在单股DNA底物形成之处,例如由于"DNA呼吸"或在DNA复制叉处。靶近端编辑是发生于在靶序列之外的碱基编辑事件,但是在被靶向的区域的上游或下游约200bp以内。旁观者突变是发生于在靶的、Cas9/sgRNA所导向的碱基编辑窗口(其非所期望的靶核碱基)之内的突变。旁观者突变可能导致沉默突变(无氨基酸变化)或非-同义突变(氨基酸变化)二者之一。因此,需要具有减低的非靶脱氨反应的碱基编辑器。

发明内容

[0004] 如下所述,本发明的特征在于核碱基编辑器组合物和方法,以及用于定性具有降低的非靶脱氨反应(例如,与程式化的,在靶脱氨反应相比)的核碱基编辑器的测定。

[0005] 本发明所定义的组合物和制品是结合以下所提供的实施例而被分离或以其他方式被制造。从详细描述和权利要求中,将显见本发明的其他特色和优势。

[0006] 一方面本文提供了融合蛋白,其包括插入Cas9多肽的柔性环内的脱氨酶,其中该融合蛋白包括以下结构:

[0007] NH2-[Cas9的N末端片段]-[脱氨酶]-[Cas9的C末端片段]-COOH,其中"]-["的每个实例是可选的连接子。

[0008] 一方面本文提供了融合蛋白,其包括两侧翼为Cas9多肽的N末端片段和C末端片段的脱氨酶,其中该N末端片段的C端或该C末端片段的N端包括Cas9多肽的柔性环的一部分。

[0009] 在一些实施方式中,该融合蛋白的脱氨酶使靶多核苷酸序列中的靶核碱基脱氨基。在一些实施方式中,当该融合蛋白使靶核碱基脱氨基时,柔性环包括靠近该靶核碱基的氨基酸。在一些实施方式中,该柔性环包括Cas9多肽的α-螺旋结构的一部分。在一些实施方式中,该靶核碱基,与末端融合蛋白(其包括与SEQ ID NO:1的N端或C端融合的脱氨酶)相比,以较低的脱靶脱氨反应被脱氨基。

[0010] 在一些实施方式中,所述靶核碱基与靶多核苷酸序列中的原间隔序列临近基序(PAM)序列相距1至20个核碱基。在一些实施方式中,所述靶核碱基在PAM序列上游2至12个核碱基处。在一些实施方式中,所述柔性环包括一区域,其选自于由位于以下位置的氨基酸残基所组成的群组:530至537,569至579,686至691,768至793,943至947,1002至1040,1052至1077,1232至1248,和1298至1300(如SEQ ID NO:1中所编号的),抑或其相应的区域。在一些实施方式中,所述脱氨酶是插入到以下氨基酸位置之间:768至769,791至792,792至793,1015至1016,1022至1023,1026至1027,1029至1030,1040至1041,1052至1053,1054至1055,1067至1068,1068至1069,1247至1248,或1248至1249(如SEQ ID NO:1中所编号的)抑或其相应的氨基酸位置。

[0011] 在一些实施方式中,所述脱氨酶是插入到以下氨基酸位置之间:768至769,792至793,1022至1023,1026至1027,1040至1041,1068至1069,或1247至1248(如SEQ ID NO:1中所编号的)抑或其相应的氨基酸位置。在一些实施方式中,所述脱氨酶是插入到以下氨基酸位置之间:1016至1017,1023至1024,1029至1030,1040至1041,1069至1070或1247至1248(如SEQ ID NO:1中所编号的)抑或其相应的氨基酸位置。

[0012] 在一些实施方式中,所述N末端片段包括Cas9多肽的氨基酸残基1至529,538至568,580至685,692至942,948至1001,1026至1051,1078至1231,和/或1248至1297(如SEQ ID NO:1中所编号的),抑或其相应的残基。在一些实施方式中,所述C末端片段包括Cas9多肽的氨基酸残基1301至1368,1248至1297,1078至1231,1026至1051,948至1001,692至942,580至685,和/或538至568(如SEQ ID NO:1中所编号的),抑或其相应的残基。在一些实施方式中,所述Cas9多肽的N末端片段或C末端片段结合靶多核苷酸序列。

[0013] 在一些实施方式中,所述Cas9多肽的N末端片段或C末端片段包括DNA结合结构域。在一些实施方式中,该N末端片段或该C末端片段包括RuvC结构域。在一些实施方式中,该N末端片段或该C末端片段包括HNH结构域。在一些实施方式中,该N末端片段和该C末端片段均不包括HNH结构域。在一些实施方式中,该N末端片段和该C末端片段均不包括RuvC结构域。在一些实施方式中,所述Cas9多肽在一个或多个结构性的结构域中包括部分或完全的缺失。在一些实施方式中,所述脱氨酶插入到Cas9多肽的部分或完全缺失的位置处。

[0014] 在一些实施方式中,该缺失是在RuvC结构域内。在一些实施方式中,该缺失是在HNH结构域内。在一些实施方式中,该缺失桥接RuvC结构域和C末端结构域,L-I结构域和HNH结构域,或RuvC结构域和L-I结构域。在一些实施方式中,所述Cas9多肽包括氨基酸的缺失1017至1069(如SEQ ID NO:1中所编号的)抑或其相应的氨基酸。在一些实施方式中,所述Cas9多肽包括氨基酸792至872的缺失(如SEQ ID NO:1中所编号的)抑或其相应的氨基酸。在一些实施方式中,所述Cas9多肽包括氨基酸792至906的缺失(如SEQ ID NO:1中所编号的)抑或其相应的氨基酸。

[0015] 一方面,本文提供了融合蛋白,其包括插入Cas9多肽内的脱氨酶,其中该融合蛋白包括以下结构:

[0016] NH2-[Cas9的N末端片段]-[脱氨酶]-[Cas9的C末端片段]-COOH,其中"]-["的每个实例是可选的连接子,其中该Cas9多肽包括HNH结构域的完全缺失,并且其中脱氨酶是插入到缺失的位置处。

[0017] 在一些实施方式中,所述N末端片段的C末端氨基酸是氨基酸791(如SEQ ID NO:1

中所编号的)。在一些实施方式中,所述C末端片段的N末端氨基酸是氨基酸907(如SEQ ID NO:1中所编号的)。在一些实施方式中,所述C末端片段的N末端氨基酸是氨基酸873(如SEQ ID NO:1中所编号的)。

[0018] 一方面本文提供了融合蛋白,其包括插入Cas9多肽内的脱氨酶,其中该融合蛋白包括以下结构:

[0019] NH2-[Cas9的N末端片段]-[脱氨酶]-[Cas9的C末端片段]-COOH,其中"]-["的每个实例是可选的连接子,而其中该Cas9包括RuvC结构域的完全缺失,并且其中脱氨酶是插入到缺失的位置处。

[0020] 在一些实施方式中,该脱氨酶是胞苷脱氨酶或腺苷脱氨酶。在一些实施方式中,该胞苷脱氨酶是APOBEC胞苷脱氨酶,活化诱导的胞苷脱氨酶(AID)或CDA。在一些实施方式中,该APOBEC脱氨酶是APOBEC1、APOBEC2、APOBEC3A、APOBEC3B、APOBEC3C、APOBEC3D、APOBEC3E、APOBEC3F、APOBEC3G、APOBEC3H或APOBEC4。在一些实施方式中,该APOBEC脱氨酶是rAPOBEC1。在一些实施方式中,上述任一方面的融合蛋白还包括UGI结构域。

[0021] 在一些实施方式中,该腺苷脱氨酶是TadA脱氨酶。在一些实施方式中,该TadA脱氨酶是经修饰的TadA。在一些实施方式中,该TadA脱氨酶是TadA 7.10。在一些实施方式中,该腺苷脱氨酶是TadA二聚体。在一些实施方式中,该TadA二聚体包括TadA 7.10和野生型TadA。在一些实施方式中,所述可选的连接子包括(SGGS)n、(GGGS)n、(GGGGS)n、(G)n、(EAAAK)n、(GGS)n、SGSETPGTSESATPES或(XP)n基序,或其组合,其中n独立地是1和30之间的整数。

[0022] 在一些实施方式中,所述Cas9多肽的N末端片段在没有连接子的情况下与脱氨酶融合。在一些实施方式中,所述Cas9多肽的C末端片段在没有连接子的情况下与脱氨酶融合。在一些实施方式中,上述任一方面的融合蛋白,还包括额外的催化结构域。

[0023] 在一些实施方式中,该额外的催化结构域是第二脱氨酶。在一些实施方式中,该第二脱氨酶与融合蛋白的N端或C端融合。在一些实施方式中,所述脱氨酶是胞苷脱氨酶或腺苷脱氨酶。在一些实施方式中,上述任一方面的融合蛋白还包括核定位信号。在一些实施方式中,该核定位信号是双分型核定位信号。在一些实施方式中,该Cas9多肽是化脓性链球菌Cas9(SpCas9),金黄色葡萄球菌Cas9(SaCas9),嗜热链球菌1Cas9(St1Cas9),或其变体。在一些实施方式中,该Cas9多肽是经修饰的Cas9并且对改变过的PAM有特异性。在一些实施方式中,该Cas9多肽是切口酶。在一些实施方式中,该Cas9多肽是核酸酶无活性的。在一些实施方式中,上述任一方面的融合蛋白与向导核酸序列复合以达成靶核碱基的脱氨反应。在一些实施方式中,该融合蛋白还与所述靶多核苷酸复合。

[0024] 本文提供了编码上述任一方面的融合蛋白的多核苷酸。

[0025] 本文提供了包括上述多核苷酸的表达载体。

[0026] 在一些实施方式中,该表达载体是哺乳动物表达载体。在一些实施方式中,该载体是病毒载体,其选自于由以下所组成的群组:腺-相关病毒(AAV)、逆转录病毒载体、腺病毒载体、慢病毒载体、仙台病毒载体和疱疹病毒载体。在一些实施方式中,该载体包括启动子。

[0027] 本文提供了细胞,其包括上述任一方面的融合蛋白,上述多核苷酸,或上述载体。

[0028] 在一些实施方式中,该细胞是细菌细胞,植物细胞,昆虫细胞,人类细胞,或哺乳动物细胞。

[0029] 本文提供了试剂盒,其包括上述任一方面的融合蛋白,上述多核苷酸,或上述载体。

[0030] 本文提供了一种用于碱基编辑的方法,该方法包括使多核苷酸序列与上述任一方面的融合蛋白接触,其中该融合蛋白的脱氨酶使多核苷酸中的核碱基脱氨基,从而编辑了该多核苷酸序列。

[0031] 在一些实施方式中,该方法还包括使所述靶多核苷酸序列与向导核酸序列接触以达成靶核碱基的脱氨反应。

[0032] 一方面,本文提供了一种用于编辑靶多核苷酸序列中的靶核碱基的方法,该方法包括:使该靶多核苷酸序列与融合蛋白接触,该融合蛋白包括两侧翼为Cas9多肽的N末端片段和C末端片段的脱氨酶,其中该融合蛋白的脱氨酶使靶多核苷酸序列中的靶核碱基脱氨基,并且其中所述N末端片段的C端或所述C末端片段的N端包括Cas9多肽的柔性环的一部分。

[0033] 本文提供了一种用于编辑靶多核苷酸序列中的靶核碱基的方法,该方法包括:使该靶多核苷酸序列与融合蛋白接触,该融合蛋白包括插入Cas9多肽的柔性环内的脱氨酶,其中该融合蛋白包括结构NH2-[Cas9的N末端片段]-[脱氨酶]-[Cas9的C末端片段]-COOH,其中"]-["的每个实例是可选的连接子,其中该融合蛋白的脱氨酶使靶多核苷酸序列中的靶核碱基脱氨基。

[0034] 在一些实施方式中,该方法还包括使所述靶多核苷酸序列与向导核酸序列接触以达成靶核碱基的脱氨反应。在一些实施方式中,该向导核酸序列包括一间隔序列,其序列与靶多核苷酸序列的原间隔序列互补,从而形成了R-环。在一些实施方式中,所述靶核碱基,与末端方法(其包括与SEQ ID NO:1的N端或C端融合的脱氨酶)相比,以较低的脱靶脱氨反应被脱氨基。在一些实施方式中,该融合蛋白的脱氨酶在R-环范围内使不超过两个的核碱基脱氨基。在一些实施方式中,所述靶核碱基与靶多核苷酸序列中的原间隔序列临近基序(PAM)序列相距1至20个核碱基。在一些实施方式中,所述靶核碱基在PAM序列上游2至12个核碱基处。

[0035] 在一些实施方式中,当该融合蛋白的脱氨酶使靶核碱基脱氨基时,柔性环包括靠近该靶核碱基的氨基酸。在一些实施方式中,所述柔性环包括一区域,其选自于由位于以下位置的氨基酸残基所组成的群组:530至537、569至579、686至691、768至793、943至947、1002至1040、1052至1077、1232至1248和1298至1300(如SEQ ID NO:1中所编号的),抑或其相应的区域。在一些实施方式中,所述脱氨酶是插入到以下氨基酸位置之间:768至769、791至792、792至793、1015至1016、1022至1023、1026至1027、1029至1030、1040至1041、1052至1053、1054至1055、1067至1068、1068至1069、1247至1248或1248至1249(如SEQ ID NO:1中所编号的)抑或其相应的氨基酸位置。在一些实施方式中,所述脱氨酶是插入到以下氨基酸位置之间:768至769、792至793、1022至1023、1026至1027、1040至1041、1068至1069或1247至1248(如SEQ ID NO:1中所编号的)抑或其相应的氨基酸位置。在一些实施方式中,所述脱氨酶是插入到以下氨基酸位置之间:1016至1017、1023至1024、1029至1030、1040至1041、1069至1070或1247至1248(如SEQ ID NO:1中所编号的)抑或其相应的氨基酸位置。

[0036] 在一些实施方式中,该N末端片段包括Cas9多肽的氨基酸残基1至529、538至568、580至685、692至942、948至1001、1026至1051、1078至1231,和/或1248至1297(如SEQ ID

NO:1中所编号的),抑或其相应的残基。在一些实施方式中,该C末端片段包括Cas9多肽的氨基酸残基1301至1368、1248至1297、1078至1231、1026至1051、948至1001、692至942、580至685,和/或538至568(如SEQ ID NO:1中所编号的),抑或其相应的残基。在一些实施方式中,所述Cas9多肽的N末端片段或C末端片段结合靶多核苷酸序列。在一些实施方式中,该N末端片段或该C末端片段包括RuvC结构域。在一些实施方式中,该N末端片段或该C末端片段包括HNH结构域。在一些实施方式中,该N末端片段和该C末端片段均不包括HNH结构域。在一些实施方式中,该N末端片段和该C末端片段均不包括RuvC结构域。

[0037] 在一些实施方式中,所述Cas9多肽在一个或多个结构性的结构域中包括部分或完全的缺失。在一些实施方式中,所述脱氨酶插入到Cas9多肽的部分或完全缺失的位置处。在一些实施方式中,该缺失是在RuvC结构域内。在一些实施方式中,该缺失是在HNH结构域内。在一些实施方式中,该缺失桥接RuvC结构域和C末端结构域,L-I结构域和HNH结构域,或RuvC结构域和L-I结构域。在一些实施方式中,所述Cas9多肽包括氨基酸的缺失1017至1069(如SEQ ID NO:1中所编号的)抑或其相应的氨基酸。

[0038] 在一些实施方式中,所述Cas9多肽包括氨基酸的缺失792至872(如SEQ ID NO:1中所编号的)抑或其相应的氨基酸。在一些实施方式中,所述Cas9多肽包括氨基酸的缺失792至906(如SEQ ID NO:1中所编号的)抑或其相应的氨基酸。在一些实施方式中,该脱氨酶是胞苷脱氨酶。在一些实施方式中,该脱氨酶是腺苷脱氨酶。在一些实施方式中,该Cas9多肽是经修饰的Cas9并且对改变过的原间隔序列临近基序(PAM)有特异性。在一些实施方式中,该Cas9多肽是切口酶。在一些实施方式中,该Cas9多肽是核酸酶无活性的。

[0039] 在一些实施方式中,所述接触是在细胞中执行。在一些实施方式中,该细胞是哺乳动物细胞或人类细胞。在一些实施方式中,该细胞是分化多能的细胞。在一些实施方式中,该细胞是在体内的(in vivo)或离体的(ex vivo)。在一些实施方式中,所述接触是在细胞的群体中执行。在一些实施方式中,该细胞的群体是哺乳动物细胞或人类细胞。

[0040] 一方面本文提供了一种用于治疗受试者体内的遗传病况的方法,该方法包括:向该受试者施用融合蛋白,该融合蛋白包括两侧翼为Cas9多肽的N末端片段和C末端片段的脱氨酶或者编码该融合蛋白的多核苷酸,以及向导核酸序列或者编码该向导核酸序列的多核苷酸,其中该向导核酸序列导引该融合蛋白使受试者的靶多核苷酸序列中的靶核碱基脱氨基,从而治疗所述遗传病况。

[0041] 本文提供了一种用于治疗受试者体内的遗传病况的方法,该方法包括:向该受试者施用融合蛋白,该融合蛋白包括插入Cas9多肽的柔性环内的脱氨酶,其中该融合蛋白包括结构NH2-[Cas9的N末端片段]-[脱氨酶]-[Cas9的C末端片段]-COOH,其中"]-["的每个实例是可选的连接子,其中该融合蛋白的脱氨酶使受试者的靶多核苷酸序列中的靶核碱基脱氨基,从而治疗所述遗传病况。

[0042] 在一些实施方式中,该N末端片段的C端或该C末端片段的N端包括所述Cas9多肽的柔性环的一部分。在一些实施方式中,该方法还包括向该受试者施用向导核酸序列以达成靶核碱基的脱氨反应。在一些实施方式中,该靶核碱基包括与该遗传病况相关联的突变。在一些实施方式中,靶核碱基的脱氨反应用野生型核碱基取代了该靶核碱基。在一些实施方式中,靶核碱基的脱氨反应用非-野生型核碱基取代了该靶核碱基,并且其中所述靶核碱基的脱氨反应改善了所述遗传病况的症状。

[0043] 在一些实施方式中,该靶多核苷酸序列在靶核碱基以外的核碱基处包括与所述遗传病况相关联的突变。在一些实施方式中,靶核碱基的脱氨反应改善了所述遗传病况的症状。在一些实施方式中,该靶核碱基与靶多核苷酸序列中的PAM序列相距1至20个核碱基。在一些实施方式中,该靶核碱基在PAM序列上游2至12个核碱基处。在一些实施方式中,当该融合蛋白的脱氨酶使靶核碱基脱氨基时,柔性环包括靠近该靶核碱基的氨基酸。

[0044] 在一些实施方式中,所述柔性环包括一区域,其选自于由位于以下位置的氨基酸残基所组成的群组:530至537、569至579、686至691、768至793、943至947、1002至1040、1052至1077、1232至1248和1298至1300(如SEQ ID NO:1中所编号的),抑或其相应的区域。

[0045] 在一些实施方式中,所述脱氨酶是插入到以下氨基酸位置之间:768至769、791至792、792至793、1015至1016、1022至1023、1026至1027、1029至1030、1040至1041、1052至1053、1054至1055、1067至1068、1068至1069、1247至1248或1248至1249(如SEQ ID NO:1中所编号的)抑或其相应的氨基酸位置。在一些实施方式中,所述脱氨酶是插入到以下氨基酸位置之间:768至769、792至793、1022至1023、1026至1027、1040至1041、1068至1069或1247至1248(如SEQ ID NO:1中所编号的)抑或其相应的氨基酸位置。在一些实施方式中,所述脱氨酶是插入到以下氨基酸位置之间:1016至1017、1023至1024、1029至1030、1040至1041、1069至1070或1247至1248(如SEQ ID NO:1中所编号的)抑或其相应的氨基酸位置。

[0046] 在一些实施方式中,所述N末端片段包括Cas9多肽的氨基酸残基1至529、538至568、580至685、692至942、948至1001、1026至1051、1078至1231和/或1248至1297(如SEQ ID NO:1中所编号的),抑或其相应的残基。在一些实施方式中,所述C末端片段包括Cas9多肽的氨基酸残基1301至1368、1248至1297、1078至1231、1026至1051、948至1001、692至942、580至685和/或538至568(如SEQ ID NO:1中所编号的),抑或其相应的残基。在一些实施方式中,Cas9多肽的N末端片段或C末端片段结合靶多核苷酸序列。在一些实施方式中,该N末端片段或该C末端片段包括RuvC结构域。在一些实施方式中,该N末端片段或该C末端片段包括HNH结构域。

[0047] 在一些实施方式中,该N末端片段和该C末端片段均不包括HNH结构域。在一些实施方式中,该N末端片段和该C末端片段均不包括RuvC结构域。在一些实施方式中,所述Cas9多肽在一个或多个结构性的结构域中包括部分或完全的缺失。在一些实施方式中,所述脱氨酶插入到Cas9多肽的部分或完全缺失的位置处。在一些实施方式中,该缺失是在RuvC结构域内。在一些实施方式中,该缺失是在HNH结构域内。在一些实施方式中,该缺失桥接RuvC结构域和C末端结构域,L-I结构域和HNH结构域,或RuvC结构域和L-I结构域。在一些实施方式中,所述Cas9多肽包括氨基酸的缺失1017至1069(如SEQ ID NO:1中所编号的)抑或其相应的氨基酸。

[0048] 在一些实施方式中,所述Cas9多肽包括氨基酸的缺失792至872(如SEQ ID NO:1中所编号的)抑或其相应的氨基酸。在一些实施方式中,所述Cas9多肽包括氨基酸的缺失792至906(如SEQ ID NO:1中所编号的)抑或其相应的氨基酸。在一些实施方式中,该脱氨酶是胞苷脱氨酶。在一些实施方式中,该脱氨酶是腺苷脱氨酶。在一些实施方式中,该Cas9多肽是经修饰的Cas9并且对改变过的PAM有特异性。在一些实施方式中,该Cas9多肽是切口酶。在一些实施方式中,该Cas9多肽是核酸酶无活性的。在一些实施方式中,该受试者是哺乳动物。在一些实施方式中,该受试者是人。

[0049] 本文提供了用于最优化的碱基编辑的蛋白质文库,其包括复数个融合蛋白,其中该复数个融合蛋白中的每一个都包括两侧翼为Cas9多肽的N末端片段和C末端片段的脱氨酶,其中所述融合蛋白中的每一个的N末端片段都不同于该复数个融合蛋白中其余蛋白的N末端片段或者其中所述融合蛋白中的每一个的C末端片段都不同于该复数个融合蛋白中其余蛋白的C末端片段,其中所述融合蛋白中的每一个的脱氨酶使靶多核苷酸序列中靠近原间隔序列相邻基序(PAM)序列的靶核碱基脱氨基,并且其中该N末端片段或该C末端片段结合该靶多核苷酸序列。

[0050] 在一些实施方式中,对于与PAM序列相距1到20个核碱基的每个核碱基,所述复数个融合蛋白中的至少一个使该核碱基脱氨基。在一些实施方式中,所述复数个融合蛋白中的每一个的Cas9多肽的N末端片段的C端或C末端片段的N端包括Cas9多肽的柔性环的一部分。在一些实施方式中,与末端融合蛋白(其包括与SEQ ID NO:1的N端或C端融合的脱氨酶)相比,所述复数个融合蛋白中的至少一个以较低的脱靶脱氨反应使靶核碱基脱氨基。在一些实施方式中,所述复数个融合蛋白中的至少一个使在PAM序列上游2至12个核碱基处的靶核碱基脱氨基。在一些实施方式中,当该融合蛋白使靶核碱基脱氨基时,所述复数个融合蛋白其中一个的N端片段的C端或C端片段的N端包括靠近靶核碱基的氨基酸。

[0051] 在一些实施方式中,该融合蛋白中的至少一个的脱氨酶是在以下氨基酸位置之间:768至769、791至792、792至793、1015至1016、1022至1023、1026至1027、1029至1030、1040至1041、1052至1053、1054至1055、1067至1068、1068至1069、1247至1248或1248至1249(如SEQ ID NO:1中所编号的)抑或其相应的氨基酸位置。在一些实施方式中,该融合蛋白中的至少一个的脱氨酶是在以下氨基酸位置之间:768至769、792至793、1022至1023、1026至1027、1040至1041、1068至1069或1247至1248(如SEQ ID NO:1中所编号的)抑或其相应的氨基酸位置。在一些实施方式中,该融合蛋白中的至少一个的脱氨酶是在以下氨基酸位置之间:1016至1017、1023至1024、1029至1030、1040至1041、1069至1070或1247至1248(如SEQ ID NO:1中所编号的)抑或其相应的氨基酸位置。在一些实施方式中,该脱氨酶是腺苷脱氨酶。在一些实施方式中,该脱氨酶是胞苷脱氨酶。

[0052] 在一些实施方式中,该Cas9多肽是化脓性链球菌Cas9(SpCas9),金黄色葡萄球菌Cas9(SaCas9),嗜热链球菌1Cas9(St1Cas9),或其变体。在一些实施方式中,该Cas9多肽是经修饰的Cas9并且对改变过的原间隔序列临近基序(PAM)有特异性。在一些实施方式中,该Cas9多肽是切口酶。在一些实施方式中,该Cas9多肽是核酸酶无活性的。

[0053] 定义

[0054] 除非另做界定,否则本文所使用的所有技术和科学术语具有本发明所属领域技术人员一般理解的意义。以下参考文献向技术人员提供本发明中所用的许多术语的通常定义:Singleton等人,Dictionary of Microbiology and Molecular Biology(第二版,1994);The Cambridge Dictionary of Science and Technology(Walker编,1988);The Glossary of Genetics,第五版,R.Rieger等人(编),Springer Verlag(1991);以及Hale&Marham,The Harper Collins Dictionary of Biology(1991)。如本文所用,除非另有说明,否则以下术语具有以下所赋予它们的含义。

[0055] 除非另做界定,否则本文所使用的所有技术和科学术语具有本发明所属领域技术人员一般理解的意义。以下参考文献向技术人员提供本发明中所用的许多术语的通常定

义:Singleton等人,Dictionary of Microbiology and Molecular Biology(第二版,1994);The Cambridge Dictionary of Science and Technology(Walker编,1988);The Glossary of Genetics,第五版,R.Rieger等人(编),Springer Verlag(1991);以及Hale&Marham,The Harper Collins Dictionary of Biology(1991)。如本文所用,除非另有说明,否则以下术语具有以下所赋予它们的含义。

[0056] “腺苷脱氨酶”意为能够催化腺嘌呤或腺苷的水解脱氨反应的多肽或其片段。在一些实施方式中,该脱氨酶或脱氨酶结构域是催化腺苷为肌苷或脱氧腺苷为脱氧肌苷的水解脱氨反应的腺苷脱氨酶。在一些实施方式中,该腺苷脱氨酶催化脱氧核糖核酸(DNA)中的腺嘌呤或腺苷的水解脱氨反应。本文所提供的腺苷脱氨酶(例如,工程化的腺苷脱氨酶,进化版的腺苷脱氨酶)可来自任何生物体,诸如细菌。在一些实施方式中,该脱氨酶或脱氨酶结构域是来自生物体的天然-存在的脱氨酶的变体。在一些实施方式中,该脱氨酶或脱氨酶结构域在自然界中不存在。例如,在一些实施方式中,该脱氨酶或脱氨酶结构域是至少50%,至少55%,至少60%,至少65%,至少70%,至少75%,至少80%,至少85%,至少90%,至少95%,至少96%,至少97%,至少98%,至少99%,或至少99.5%与天然-存在的脱氨酶相同。在一些实施方式中,该腺苷脱氨酶来自细菌,诸如,大肠杆菌(E.coli),金黄色葡萄球菌(S.aureus),伤寒沙门菌(S.typhi),腐烂希瓦氏菌(S.putrefaciens),流感嗜血杆菌(H.influenzae),或新月柄杆菌(C.crescentus)。在一些实施方式中,该腺苷脱氨酶是TadA脱氨酶。在一些实施方式中,该TadA脱氨酶是大肠杆菌TadA(ecTadA)脱氨酶或其片段。

[0057] 例如,相对于全长ecTadA,该截短的ecTadA可能缺少一个或多个N末端氨基酸。在一些实施方式中,相对于全长ecTadA,该截短的ecTadA可能缺少1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,或20个N末端氨基酸残基。在一些实施方式中,相对于全长ecTadA,该截短的ecTadA可能缺少1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,或20个C末端氨基酸残基。在一些实施方式中,该ecTadA脱氨酶不包括N末端的蛋氨酸。在一些实施方式中,该TadA脱氨酶是N末端截短的TadA。在特定的实施方式中,该TadA是PCT/US2017/045381中所述的TadA中的任何一个,其全部内容通过引用并入本文。

[0058] 在一些实施方式中,该腺苷脱氨酶包括氨基酸序列:

[0059] MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTD,其称为“TadA参考序列”。

[0060] 在一些实施方式中,所述TadA脱氨酶是全长的大肠杆菌TadA脱氨酶。例如,在某些实施方式中,该腺苷脱氨酶包括氨基酸序列:

[0061] MRRAFITGVFFLSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTD。

[0062] 然而,应当理解,可用于本申请中的额外的腺苷脱氨酶对熟练技术人员是显而易见的,并且是在本公开揭露的范围内。例如,该腺苷脱氨酶可能是作用在tRNA上的腺苷脱氨酶(AD AT)的同源物。示例性的AD AT同源物包含但不限于:

[0063] 金黄色葡萄球菌(Staphylococcus aureus)TadA:

[0064] MGSHMTNDIYFMTLAIEEAKKAAQLGEVPIGAIITKDDEVIARAHNLRETLQQPTAH AEHIAIERAA

KVLGSWRLEGCTLYVTLEPCVMCAGTIVMSRIPRVVYGADDPKGGCSGS LMNLLQQS NFNHRAIVDKG VLKE AC S TLLTTFFKNLRANKKS TN

[0065] 枯草芽孢杆菌(Bacillus subtilis)TadA:

[0066] MTQDELYMKEAIKEAKKAEEKGEVPIGAVLVINGEIIARAHNLRETEQRSIAHAEML VIDEACKALG TWRLEGATLYVTLEPCPMCAGAVVLSRVEKVVFGAFDPKGGC S GTLMNLLQEERFNHQAEVVSGVLEEECGGML SAFFRELRKKKKAARKNLSE

[0067] 鼠伤寒沙门菌(S.typhimurium)TadA:

[0068] MPPAFITGVTSLSDVELDHEYWMRHALTLAKRAWDEREVPVGAVLVHNHRVIGEGWNRPIGRHDPTAHA EIMALRQGGLVLQNYRLLDTTLYVTLEPCVMCAGAMVHSRIGRVVFGARDAKTGAAGSLIDVLHHPGMNHRVEIIEG VLRDECATLLSDFFRMRRQEIKALKKADRAEGAGPAV

[0069] 腐败希瓦氏菌(S.putrefaciens)TadA:

[0070] MDE YWMQVAMQM AEKAEAAGE VPVGA VLVKDGQQIATGYNLS IS QHDPTAHAEILCLRSAGK KLENYRLLDATLYITLEPCAMCAGAMVHSRIARVVYGARDEKTGAAGT VVNLLQHPAFNHQVEVTSGVLAEACSA QLSRFFKRRRDEKKALKLAQRAQQGIE

[0071] 流感嗜血杆菌F3031(H.influenzae)TadA:

[0072] MDAAKVRSEFDEKMMRYALELADKAEALGEIPVGAVLVDDARNIIGEGWNLSIVQSDPT AH AEIIALRNG AKNIQN YRLLNS TLY VTLEPCTMC AG AILHS RIKRLVFGAS D YKTGAIGSRFHFFDDY KMNHTLEITSGVLAEECSQKLSTFFQKRREEKKIEKALLKSLSD K

[0073] 新月柄杆菌(C.crescentus)TadA:

[0074] MRTDESEDQDHRMMRLALDAARAAAEAGETPVGAVILDPSTGEVIATAGNGPIAAHDPTAHAEIAAMRA AAAKLGNYRLTDLTLVVTLEPCAMCAGAISHARIGRVVFGADDPKGGAVVHGPKFFAQPTCHWRPEVTGGVLADESA DLLRGFFRARRKAKI

[0075] 硫还原地杆菌(G.sulfurreducens)TadA:

[0076] MSSLKKTPIRDDAYWMGKAIREAAKAAARDEVPIGAVIVRDGAVIGRGHNLREGSNDPSAHAEMIAIRQ AARRSANWRLTGATLYVTLEPCLMCMGAIILARLERVVFGCYDPKGGAAGSLYDLSADPRLNHQVRLSPGVCQEECG TMLSDFFRDLRRRKKAKATPALFIDERKVPPEP

[0077] TadA7.10

[0078] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLC YFFRMPRQVFNAQKKAQSSTD

[0079] 含有TadA7.10或TadA7.10变体的示例性序列包含

[0080] GSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIG LHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMN HRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD

[0081] TadA7.10 CP65

[0082] TAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHY PGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGSETPGTSESATPESSGSEVEFSHEYWMRH ALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDP

[0083] TadA7.10 CP83

[0084] YRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADEC AALLCYFFRMPRQVFNAQKKAQSSTDGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGA VLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQN

[0085] TadA7.10 CP136

[0086] MNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGSETPGTSESATPESSGSEVEFSH EYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFE PCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPG

[0087] TadA7.10 C-截短

[0088] GSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIG LHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMN HRVEITEGILADECAALLCYFFRMPRQVFN

[0089] TadA7.10 C-截短2

[0090] GSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIG LHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMN HRVEITEGILADECAALLCYFFRMPRQ

[0091] TadA7.10Δ59-66+C-截短

[0092] GSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAHA EIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEG ILADECAALLCYFFRMPRQVFN

[0093] TadA7.10Δ59-66

[0094] GSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVG

[0095] AVLVLNNRVIGEGWNRAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVM

[0096] CAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILAD

[0097] ECAALLCYFFRMPRQVFNAQKKAQSSTD

[0098] “剂”意为任何小分子化学化合物,抗体,核酸分子,或多肽,或其片段。

[0099] “改变一突变”意为通过标准的本领域已知的分子生物学方法或遗传学方法等,诸如PCR定点诱变或同源重组,对多核苷酸序列中已知的一突变进行操作并使其产生变化。

[0100] “改变”意为通过标准的本领域已知的方法,诸如本文所述的方法,所检测到的基因或多肽的结构、表达水平或活性的变化。如本文所用,改变(例如,增加或减少)包含表达水平的10%变化,25%变化,40%变化和表达水平的50%或更大的变化。

[0101] “类似物”意为不相同但具有类似的功能性或结构性特征的分子。例如,多核苷酸类似物保留了相应的天然-存在的多核苷酸的生物学活性,同时具有某些修饰,该修饰相对于天然-存在的多核苷酸增强了类似物的功能。此种修饰可以增加该多核苷酸对DNA的亲和力,半衰期和/或核酸酶抗性。类似物可能包含非天然的核苷酸或氨基酸。

[0102] 在本公开揭露中,“包括(comprises)”、“包括(comprising)”、“含有(containing)”和“具有(having)”等可以具有美国专利法中赋予它们的含义,并且可以表示“包含(includes)”、“包含(including)”等;“基本上由…组成(consisting essentially of)”或“基本上包括(consists essentially)”同样具有美国专利法中赋予的含义,并且该术语是开放式的,允许存在多于所列举出的内容,只要所列举出的内容的基本的或新颖的

特征不被多于所列举出的内容的存在所变更,但排除现有技术的实施方式。

[0103] “碱基编辑器(BE)”或“核碱基编辑器(NBE)”意为结合多核苷酸并且具有核碱基修饰活性的剂。在一个实施方式中,该剂是融合蛋白,其包括具有碱基编辑活性的结构域,即,能够修饰核酸分子(例如DNA)内的碱基(例如A,T,C,G,或U)的结构域。在一些实施方式中,具有碱基编辑活性的该结构域能够使核酸分子内的碱基脱氨基。在一些实施方式中,该碱基编辑器能够使DNA分子内的碱基脱氨基。在一些实施方式中,该碱基编辑器能够使DNA内的胞嘧啶(C)或腺苷脱氨基。在一些实施方式中,该碱基编辑器是胞苷碱基编辑器(CBE)。在一些实施方式中,该碱基编辑器是腺苷碱基编辑器(ABE)。在一些实施方式中,该碱基编辑器是腺苷碱基编辑器(ABE)和胞苷碱基编辑器(CBE)。在一些实施方式中,该碱基编辑器是与腺苷脱氨酶融合的核酸酶-无活性的Cas9(dCas9)。在一些实施方式中,该Cas9是环状变换排列体Cas9(例如spCas9或saCas9)。环状变换排列体Cas9是本领域已知的,并在例如Oakes等,Cell 176,254-267,2019中进行了描述。在一些实施方式中,该碱基编辑器是融合到碱基切除修复的抑制剂上,例如,UGI结构域。在一些实施方式中,该融合蛋白包括融合到脱氨酶上的Cas9切口酶以及碱基切除修复的抑制剂,诸如UGI结构域。在其他实施方式中,该碱基编辑器是无碱基碱基编辑器。

[0104] 在一些实施方式中,腺苷脱氨酶是由TadA进化而来。在一些实施方式中,该多核苷酸可编程的DNA结合结构域是CRISPR相关联的(例如,Cas或Cpf1)酶。在一些实施方式中,该碱基编辑器是与脱氨酶结构域融合的催化失活的Cas9(dCas9)。在一些实施方式中,该碱基编辑器是与脱氨酶结构域融合的Cas9切口酶(nCas9)。在一些实施方式中,该脱氨酶结构域是多核苷酸可编程的DNA结合结构域的N末端或C末端片段。在一些实施方式中,该脱氨酶的两侧翼为多核苷酸可编程的DNA结合结构域的N末端和C末端片段。在一些实施方式中,该脱氨酶结构域是插入到多核苷酸可编程的DNA结合结构域的位点中。在一些实施方式中,该碱基编辑器与碱基切除修复的抑制剂(BER)融合。在一些实施方式中,该碱基切除修复的抑制剂是尿嘧啶DNA糖基化酶抑制剂(UGI)。在一些实施方式中,该碱基切除修复的抑制剂是肌苷碱基切除修复抑制剂。碱基编辑器的细节在国际PCT申请编号PCT/2017/045381(WO2018/027078)和PCT/US2016/058344(WO2017/070632)中被描述,其每个的全部内容通过引用并入本文。亦请参见Komor,A.C.等,“无需双-股DNA切开的基因组DNA中靶碱基的可编程的编辑”Nature 533,420-424(2016);Gaudelli,N.M.等,“无需DNA切开的基因组DNA中A•T到G•C的可编程的碱基编辑”Nature 551,464-471(2017);Komor,A.C.等,“改良过的碱基切除修复抑制合并噬菌体Mu Gam蛋白所产生的C:G-到-T:A碱基编辑器具有更高的效率和更高的产品纯度”Science Advances 3:eaao4774(2017),和Rees,H.A.等,“碱基编辑:活体细胞的基因组和转录组上的精密化学”Nat Rev Genet.2018Dec;19(12):770-788.doi:10.1038/s41576-018-0059-1,其全部内容在此通过引用并入。

[0105] 在一些实施方式中,所述脱氨酶结构域是插入到多核苷酸可编程的DNA结合结构域的区域中。在一些实施方式中,该插入位点是通过napDNAbp的结构性分析而确定的。在一些实施方式中,该插入位点是柔性环。在一些实施方式中,所述脱氨酶结构域是插入到多核苷酸可编程的DNA结合结构域中的位点中,其中该位点是选自于由以下氨基酸位置所组成的群组中的至少一个:1029,1026,1054,1022,1015,1068,1247,1040,1248,和768。在一些实施方式中,插入的脱氨酶结构域代替了多核苷酸可编程的DNA结合结构域的结构域的位

置。在一些实施方式中,该结构域是选自于由以下所组成的群组:RuvC,Rec1,Rec2,和HNH。在一些实施方式中,插入的脱氨酶结构域代替了多核苷酸可编程的DNA结合结构域中一氨基酸残基范围,其中该氨基酸残基范围是选自于由以下Cas9残基所组成的群组:530至537,569至579,686至691,768至793,943至947,1002至1040,1052至1077,1232至1248,和1298至1300(如SEQ ID NO:1中所编号的)抑或其相应的位置。如何通过比较Cas9氨基酸序列而在不同的多核苷酸可编程的DNA结合结构域中鉴别出同源区域,对熟练技术人员将是显而易见的。在一些实施方式中,该碱基编辑器包括插入到多核苷酸可编程的DNA结合结构域的多于一个位点中的多于一个的脱氨酶结构域,其中该位点如上所述。

[0106] 在一些实施方式中,通过将腺苷脱氨酶变体(例如TadA*7.10)克隆进入支架中生成碱基编辑器,该支架包含环状变换排列体Cas9(例如spCAS9)和双分型核定位序列。环状变换排列体Cas9是本领域已知的,并在例如Oakes等,Cell 176,254-267,2019中进行了描述。示例性的环状变换排列体如下所示,其中粗体序列表示衍生自Cas9的序列,斜体序列表示连接子序列,而底线序列表示双分型核定位序列。

[0107] CP5(具有MSP“NGC=带有突变的Pam变体正常Cas9样NGG”PID=蛋白相互作用结构域和“D10A”切口酶):

[0108] **EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFAT**
VRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFMQPTVAY
SVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK
YSLFELENGRKRMLASAKFLQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFV
EQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNL
GAPRAFKYFDTTIARKEYRSTKEVLDATLIHQSITGLYETRIDLSQLGGD*GGSGGSGGS*
***GGSGGSGGSGGM*DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLI**
GALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLV
EEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRG
HFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENL
IAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGD
QYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQL
PEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQ
RTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSR
FAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTV
YNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVE
ISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKT
[0109] **YAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLI**
HDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKP
ENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYY
LQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE
VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQ
ILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAV
VGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEGADKRTADGSEFESPKKKRKV*

[0110] 碱基编辑器系统的核碱基组分与多核苷酸可编程的核苷酸结合组分可共价地或

非共价地彼此相联。例如,在一些实施方式中,可以通过多核苷酸可编程的核苷酸结合结构域而将所述脱氨酶结构域靶向至靶核苷酸序列。在一些实施方式中,多核苷酸可编程的核苷酸结合结构域可以融合到或链接到脱氨酶结构域上。在一些实施方式中,多核苷酸可编程的核苷酸结合结构域可以通过与脱氨酶结构域的非共价相互作用或与其相联而将该脱氨酶结构域靶向到靶核苷酸序列。例如,在一些实施方式中,该核碱基编辑组分,例如该脱氨酶组分,可以包括能够与身为多核苷酸可编程的核苷酸结合结构域的一部分的额外的异源部分或结构域相互作用、相联或形成复合体的另外的异源部分或结构域。在一些实施方式中,该额外的异源部分可能能够与多肽结合、相互作用、相联或形成复合体。在一些实施方式中,该额外的异源部分可能能够与多核苷酸结合、相互作用、相联或形成复合体。在一些实施方式中,该额外的异源部分可能能够与向导多核苷酸结合。在一些实施方式中,该额外的异源部分可能能够与多肽连接子结合。在一些实施方式中,该额外的异源部分可能能够与多核苷酸连接子结合。该额外的异源部分可能是蛋白结构域。在一些实施方式中,该额外的异源部分可能是K同源(KH)结构域,MS2外壳蛋白结构域,PP7外壳蛋白结构域,SfMu Com外壳蛋白结构域,不育α基序,端粒酶Ku结合基序和Ku蛋白,端粒酶Sm7结合基序和Sm7蛋白,或RNA识别基序。

[0111] 碱基编辑器系统可能还包括向导多核苷酸组分。应当理解,该碱基编辑器系统的各组分可以经由共价键、非共价相互作用或其相联和相互作用的任何组合而彼此相联。在一些实施方式中,可以通过向导多核苷酸将脱氨酶结构域靶向至靶核苷酸序列。例如,在一些实施方式中,该碱基编辑器系统的核碱基编辑组分,例如脱氨酶组分,可以包括额外的异源部分或结构域(例如多核苷酸结合结构域,诸如RNA或DNA结合蛋白),其能够与向导多核苷酸的部分或节段(例如,多核苷酸基序)相互作用、相联或形成复合体。在一些实施方式中,该额外的异源部分或结构域(例如多核苷酸结合结构域,诸如RNA或DNA结合蛋白)可以融合到或链接到该脱氨酶结构域上。在一些实施方式中,该额外的异源部分可能能够与多肽结合、相互作用、相联或形成复合体。在一些实施方式中,该额外的异源部分可能能够与多核苷酸结合、相互作用、相联或形成复合体。在一些实施方式中,该额外的异源部分可能能够与向导多核苷酸结合。在一些实施方式中,该额外的异源部分可能能够与多肽连接子结合。在一些实施方式中,该额外的异源部分可能能够与多核苷酸连接子结合。该额外的异源部分可能是蛋白结构域。在一些实施方式中,该额外的异源部分可能是K同源(KH)结构域,MS2外壳蛋白结构域,PP7外壳蛋白结构域,SfMu Com外壳蛋白结构域,不育α基序,端粒酶Ku结合基序和Ku蛋白,端粒酶Sm7结合基序和Sm7蛋白,或RNA识别基序。

[0112] 在一些实施方式中,碱基编辑器系统可以还包括碱基切除修复(BER)的抑制剂组分。应该理解,该碱基编辑器系统的各组分可经由共价键、非共价相互作用或其相联和相互作用的任何组合而彼此相联。该BER的抑制剂组分可能包括碱基切除修复抑制剂。在一些实施方式中,该碱基切除修复的抑制剂可以是尿嘧啶DNA糖基化酶抑制剂(UGI)。在一些实施方式中,该碱基切除修复的抑制剂可以是肌苷碱基切除修复抑制剂。在一些实施方式中,可以通过所述多核苷酸可编程的核苷酸结合结构域将该碱基切除修复的抑制剂靶向到靶核苷酸序列。在一些实施方式中,多核苷酸可编程的核苷酸结合结构域可以融合到或链接到碱基切除修复的抑制剂上。在一些实施方式中,多核苷酸可编程的核苷酸结合结构域可以融合到或链接到脱氨酶结构域和碱基切除修复的抑制剂上。在一些实施方式中,多核苷酸

可编程的核苷酸结合结构域可以通过与碱基切除修复的抑制剂的非共价相互作用或相联而将该碱基切除修复的抑制剂靶向到靶核苷酸序列。例如,在一些实施方式中,该碱基切除修复的抑制剂组分可以包括额外的异源部分或结构域,其能够与身为多核苷酸可编程的核苷酸结合结构域的一部分的额外的异源部分或结构域相互作用、相联或形成复合体。在一些实施方式中,该碱基切除修复的抑制剂可以被所述向导多核苷酸靶向到靶核苷酸序列。例如,在一些实施方式中,该碱基切除修复的抑制剂可以包括额外的异源部分或结构域(例如多核苷酸结合结构域,诸如RNA或DNA结合蛋白),其能够与向导多核苷酸的部分或节段(例如,多核苷酸基序)相互作用、相联或形成复合体。在一些实施方式中,可以将该向导多核苷酸的额外的异源部分或结构域(例如多核苷酸结合结构域,诸如RNA或DNA结合蛋白)融合到或链接到该碱基切除修复的抑制剂上。在一些实施方式中,该额外的异源部分可能能够与多核苷酸结合、相互作用、相联或形成复合体。在一些实施方式中,该额外的异源部分可能能够结合向导多核苷酸。在一些实施方式中,该额外的异源部分可能能够结合多肽连接子。在一些实施方式中,该额外的异源部分可能能够结合多核苷酸连接子。该额外的异源部分可能是蛋白结构域。在一些实施方式中,该额外的异源部分可能是K同源(KH)结构域,MS2外壳蛋白结构域,PP7外壳蛋白结构域,SfMu Com外壳蛋白结构域,不育α基序,端粒酶Ku结合基序和Ku蛋白,端粒酶Sm7结合基序和Sm7蛋白,或RNA识别基序。

[0113] “碱基编辑活性”意为化学性改变多核苷酸内的碱基的作用。在一个实施方式中,第一碱基被转换为第二碱基。在一个实施方式中,该碱基编辑活性是胞苷脱氨酶活性,例如将靶C·G转换为T·A。在另一个实施方式中,该碱基编辑活性是腺苷脱氨酶活性,例如将A·T转换为G·C。

[0114] 术语“Cas9”或“Cas9结构域”是指包括Cas9蛋白或其片段(例如一蛋白,其包括活性的、无活性的或部分活性的Cas9的DNA切开结构域,和/或Cas9的gRNA结合结构域)的RNA-导向的核酸酶。Cas9核酸酶有时也被称为casn1核酸酶或CRISPR(规律成簇间隔短回文重复序列,clustered regularly interspaced short palindromic repeat)-关联核酸酶。CRISPR是一种适应性免疫系统,其针对可移动的遗传元件(病毒,转座因子和接合型质粒)提供保护。CRISPR簇含有间隔序列,即与先前的可移动的元件互补的序列,并靶向入侵的核酸。CRISPR簇被转录并加工成为CRISPR RNA(crRNA)。在II型CRISPR系统中,pre-crRNA的正确加工需要反式-编码的小RNA(tracrRNA),内源性核糖核酸酶3(rnc)以及Cas9蛋白。该tracrRNA作为核糖核酸酶3-辅助的pre-crRNA的加工的向导。随后,Cas9/crRNA/tracrRNA以核酸内裂解的方式切开与所述间隔序列互补的线性或环状dsDNA靶标。不与crRNA互补的靶标股先以核酸内裂解的方式被切割,然后以核酸外裂解的方式被3′-5′修剪。在自然界中,DNA-结合和切开通常需要蛋白质和两种RNA。然而,可以对单一向导RNAs(“sgRNA”,或简称“gNRA”)进行工程化以便将crRNA和tracrRNA二者的各方面都并入单一RNA种类之中。参见例如Jinek M.,Chylinski K.,Fonfara I.,Hauer M.,Doudna J.A.,Charpentier E.Science 337:816-821(2012),其全部内容在此通过引用并入。Cas9识别CRISPR重复序列中的短基序(PAM或原间隔序列临近基序)以帮助区分自体与非自体。Cas9核酸酶序列和结构是本领域技术人员熟知的(参见例如“Complete genome sequence of an M1 strain of Streptococcus pyogenes.”Ferretti等人,J.J.,McShan W.M.,Ajdic D.J.,Savic D.J.,Savic G.,Lyon K.,Primeaux C.,Sezate S.,Suvorov A.N.,Kenton S.,Lai H.S.,Lin

S.P.,Qian Y.,Jia H.G.,Najar F.Z.,Ren Q.,Zhu H.,Song L.,White J.,Yuan X.,Clifton S.W.,Roe B.A.,McLaughlin R.E.,Proc.Natl.Acad.Sci.美国98:4658-4663(2001);“CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III.”Deltcheva E.,Chylinski K.,Sharma C.M.,Gonzales K.,Chao Y.,Pirzada Z.A.,Eckert M.R.,Vogel J.,Charpentier E.,Nature471:602-607(2011);以及“A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.”Jinek M.,Chylinski K.,Fonfara I.,Hauer M.,Doudna J.A.,Charpentier E.Science 337:816-821(2012),其每个的全部内容通过引用并入本文)。已在好几种物种中描述了Cas9的异种同源物,包含但不限于化脓性链球菌和嗜热链球菌。基于本公开揭露,额外的适合的Cas9核酸酶和序列对于本领域技术人员将是显而易见的,并且此种Cas9核酸酶和序列包含来自Chylinski,Rhun,和Charpentier,“The tracrRNA and Cas9families of type II CRISPR-Cas immunity systems”(2013)RNA Biology 10:5,726-737中所揭露的生物体和基因座的Cas9序列;其全部内容通过引用并入本文。

[0115] 核酸酶-灭活的Cas9蛋白可互换地称为“dCas9”蛋白(核酸酶-“死(dead)”Cas9)或催化无活性的Cas9。生成具有无活性的DNA切开结构域的Cas9蛋白(或其片段)的方法是已知的(参见例如Jinek等,Science.337:816-821(2012);Qi等,“Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression”(2013)Cell.28;152(5):1173-83,其每个的全部内容通过引用并入本文)。例如,已知Cas9的DNA切开结构域包含两个亚结构域,HNH核酸酶亚结构域和RuvC1亚结构域。该HNH亚结构域切开与gRNA互补的股,而该RuvC1亚结构域切开非-互补的股。这些亚结构域内的突变可以沉默化Cas9的核酸酶活性。例如,突变D10A和H840A完全灭活化脓性链球菌Cas9的核酸酶活性(Jinek等,Science.337:816-821(2012);Qi等,Cell.28;152(5):1173-83(2013))。在一些实施方式中,Cas9核酸酶具有无活性的(例如灭活的)DNA切开结构域,即是说,该Cas9为切口酶,被称为“nCas9”蛋白(出于“切口酶(nickase)”Cas9)。在一些实施方式中,提供了包括Cas9的片段的蛋白。例如,在一些实施方式中,蛋白包括两个Cas9结构域之一:(1)Cas9的gRNA结合结构域;或(2)Cas9的DNA切开结构域。在一些实施方式中,包括Cas9或其片段的蛋白被称为“Cas9变体”。Cas9变体与Cas9或其片段共享同源性。例如,Cas9变体与野生型Cas9是至少约70%相同、至少约80%相同、至少约90%相同、至少约95%相同、至少约96%相同、至少约97%相同、至少约98%相同、至少约99%相同、至少约99.5%相同或至少约99.9%相同。在一些实施方式中,与野生型Cas9相比,该Cas9变体可能具有1、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、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50或更多个氨基酸的变化。在一些实施方式中,该Cas9变体包括Cas9的片段(例如gRNA结合结构域或DNA-切开结构域),使得该片段与野生型Cas9的相应片段是至少约70%相同、至少约80%相同、至少约90%相同、至少约95%相同、至少约96%相同、至少约97%相同、至少约98%相同、至少约99%相同、至少约99.5%相同或至少约99.9%相同。在一些实施方式中,该片段是相应的野生型Cas9的氨基酸长度的至少30%、至少35%、至少40%、至少45%、至少50%、至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或至少99.5%。

[0116] 在一些实施方式中,该片段的长度是至少100个氨基酸。在一些实施方式中,该片段的长度是至少100、150、200、250、300、350、400、450、500、550、600、650、700、750、800、850、900、950、1000、1050、1100、1150、1200、1250或至少1300个氨基酸。

[0117] 在一些实施方式中,野生型Cas9对应于来自化脓性链球菌的Cas9(NCBI参考序列:NC_017053.1,核苷酸和氨基酸序列如下)

[0118] ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGCGTCGGATGGGCGGTGATCACTGAT
GATTATAAGGTTCCGTCTAAAAAGTTCAAGGTTCTGGGAAATACAGACCGCCACAGTATCAAAAAAAATCTTATAGG
GGCTCTTTTATTTGGCAGTGGAGAGACAGCGGAAGCGACTCGTCTCAAACGGACAGCTCGTAGAAGGTATACACGTC
GGAAGAATCGTATTTGTTATCTACAGGAGATTTTTTCAAATGAGATGGCGAAAGTAGATGATAGTTTCTTTCATCGA
CTTGAAGAGTCTTTTTTGGTGGAAGAAGACAAGAAGCATGAACGTCATCCTATTTTTGGAAATATAGTAGATGAAGT
TGCTTATCATGAGAAATATCCAACTATCTATCATCTGCGAAAAAAATTGGCAGATTCTACTGATAAAGCGGATTTGC
GCTTAATCTATTTGGCCTTAGCGCATATGATTAAGTTTCGTGGTCATTTTTTGATTGAGGGAGATTTAAATCCTGAT
AATAGTGATGTGGACAAACTATTTATCCAGTTGGTACAAATCTACAATCAATTATTTGAAGAAAACCCTATTAACGC
AAGTAGAGTAGATGCTAAAGCGATTCTTTCTGCACGATTGAGTAAATCAAGACGATTAGAAAATCTCATTGCTCAGC
TCCCCGGTGAGAAGAGAAATGGCTTGTTTGGGAATCTCATTGCTTTGTCATTGGGATTGACCCCTAATTTTAAATCA
AATTTTGATTTGGCAGAAGATGCTAAATTACAGCTTTCAAAAGATACTTACGATGATGATTTAGATAATTTATTGGC
GCAAATTGGAGATCAATATGCTGATTTGTTTTTGGCAGCTAAGAATTTATCAGATGCTATTTTACTTTCAGATATCC
TAAGAGTAAATAGTGAAATAACTAAGGCTCCCCTATCAGCTTCAATGATTAAGCGCTACGATGAACATCATCAAGAC
TTGACTCTTTTAAAAGCTTTAGTTCGACAACAACTTCCAGAAAAGTATAAAGAAATCTTTTTTGATCAATCAAAAAA
CGGATATGCAGGTTATATTGATGGGGGAGCTAGCCAAGAAGAATTTTATAAATTTATCAAACCAATTTTAGAAAAAA
TGGATGGTACTGAGGAATTATTGGTGAAACTAAATCGTGAAGATTTGCTGCGCAAGCAACGGACCTTTGACAACGGC
TCTATTCCCCATCAAATTCACTTGGGTGAGCTGCATGCTATTTTGAGAAGACAAGAAGACTTTTATCCATTTTTAAA
AGACAATCGTGAGAAGATTGAAAAAATCTTGACTTTTCGAATTCCTTATTATGTTGGTCCATTGGCGCGTGGCAATA
GTCGTTTTGCATGGATGACTCGGAAGTCTGAAGAAACAATTACCCCATGGAATTTTGAAGAAGTTGTCGATAAAGGT
GCTTCAGCTCAATCATTTATTGAACGCATGACAAACTTTGATAAAAATCTTCCAAATGAAAAAGTACTACCAAAACA
TAGTTTGCTTTATGAGTATTTTACGGTTTATAACGAATTGACAAAGGTCAAATATGTTACTGAGGGAATGCGAAAAC
CAGCATTTCTTTCAGGTGAACAGAAGAAAGCCATTGTTGATTTACTCTTCAAAACAAATCGAAAAGTAACCGTTAAG
CAATTAAAAGAAGATTATTTCAAAAAAATAGAATGTTTTGATAGTGTTGAAATTTCAGGAGTTGAAGATAGATTTAA
TGCTTCATTAGGCGCCTACCATGATTTGCTAAAAATTATTAAAGATAAAGATTTTTTGGATAATGAAGAAAATGAAG
ATATCTTAGAGGATATTGTTTTAACATTGACCTTATTTGAAGATAGGGGGATGATTGAGGAAAGACTTAAAACATAT
GCTCACCTCTTTGATGATAAGGTGATGAAACAGCTTAAACGTCGCCGTTATACTGGTTGGGGACGTTTGTCTCGAAA
ATTGATTAATGGTATTAGGGATAAGCAATCTGGCAAAACAATATTAGATTTTTTGAAATCAGATGGTTTTGCCAATC
GCAATTTTATGCAGCTGATCCATGATGATAGTTTGACATTTAAAGAAGATATTCAAAAAGCACAGGTGTCTGGACAA
GGCCATAGTTTACATGAACAGATTGCTAACTTAGCTGGCAGTCCTGCTATTAAAAAAGGTATTTTACAGACTGTAAA
AATTGTTGATGAACTGGTCAAAGTAATGGGGCATAAGCCAGAAAATATCGTTATTGAAATGGCACGTGAAAATCAGA
CAACTCAAAAGGGCCAGAAAAATTCGCGAGAGCGTATGAAACGAATCGAAGAAGGTATCAAAGAATTAGGAAGTCAG
ATTCTTAAAGAGCATCCTGTTGAAAATACTCAATTGCAAAATGAAAAGCTCTATCTCTATTATCTACAAAATGGAAG
AGACATGTATGTGGACCAAGAATTAGATATTAATCGTTTAAGTGATTATGATGTCGATCACATTGTTCCACAAAGTT
TCATTAAAGACGATTCAATAGACAATAAGGTACTAACGCGTTCTGATAAAAATCGTGGTAAATCGGATAACGTTCCA

AGTGAAGAAGTAGTCAAAAAGATGAAAAACTATTGGAGACAACTTCTAAACGCCAAGTTAATCACTCAACGTAAGTT
TGATAATTTAACGAAAGCTGAACGTGGAGGTTTGAGTGAACTTGATAAAGCTGGTTTTATCAAACGCCAATTGGTTG
AAACTCGCCAAATCACTAAGCATGTGGCACAAATTTTGGATAGTCGCATGAATACTAAATACGATGAAAATGATAAA
CTTATTCGAGAGGTTAAAGTGATTACCTTAAAATCTAAATTAGTTTCTGACTTCCGAAAAGATTTCCAATTCTATAA
AGTACGTGAGATTAACAATTACCATCATGCCCATGATGCGTATCTAAATGCCGTCGTTGGAACTGCTTTGATTAAGA
AATATCCAAAACTTGAATCGGAGTTTGTCTATGGTGATTATAAAGTTTATGATGTTCGTAAAATGATTGCTAAGTCT
GAGCAAGAAATAGGCAAAGCAACCGCAAAATATTTCTTTTACTCTAATATCATGAACTTCTTCAAAACAGAAATTAC
ACTTGCAAATGGAGAGATTCGCAAACGCCCTCTAATCGAAACTAATGGGGAAACTGGAGAAATTGTCTGGGATAAAG
GGCGAGATTTTGCCACAGTGCGCAAAGTATTGTCCATGCCCCAAGTCAATATTGTCAAGAAAACAGAAGTACAGACA
GGCGGATTCTCCAAGGAGTCAATTTTACCAAAAAGAAATTCGGACAAGCTTATTGCTCGTAAAAAAGACTGGGATCC
AAAAAAATATGGTGGTTTTGATAGTCCAACGGTAGCTTATTCAGTCCTAGTGGTTGCTAAGGTGGAAAAAGGGAAAT
CGAAGAAGTTAAAATCCGTTAAAGAGTTACTAGGGATCACAATTATGGAAAGAAGTTCCTTTGAAAAAAATCCGATT
GACTTTTTAGAAGCTAAAGGATATAAGGAAGTTAAAAAAGACTTAATCATTAAACTACCTAAATATAGTCTTTTTGA
GTTAGAAAACGGTCGTAAACGGATGCTGGCTAGTGCCGGAGAATTACAAAAAGGAAATGAGCTGGCTCTGCCAAGCA
AATATGTGAATTTTTTATATTTAGCTAGTCATTATGAAAAGTTGAAGGGTAGTCCAGAAGATAACGAACAAAAACAA
TTGTTTGTGGAGCAGCATAAGCATTATTTAGATGAGATTATTGAGCAAATCAGTGAATTTTCTAAGCGTGTTATTTT
AGCAGATGCCAATTTAGATAAAGTTCTTAGTGCATATAACAAACATAGAGACAAACCAATACGTGAACAAGCAGAAA
ATATTATTCATTTATTTACGTTGACGAATCTTGGAGCTCCCGCTGCTTTTAAATATTTTGATACAACAATTGATCGT
AAACGATATACGTCTACAAAAGAAGTTTTAGATGCCACTCTTATCCATCAATCCATCACTGGTCTTTATGAAACACG
CATTGATTTGAGTCAGCTAGGAGGTGACTGA

MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNL

[0119] IGALLFGSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSF

FHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLADSTD

KADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQIYNQLFE
ENPINASRVDAKAILSARLSKSRRLENLIAQLPGEKRNGLFGNLIALSLG
LTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLS
DAILLSDILRVNSEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKY
KEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRED
LLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFR
IPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTN
FDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKA
IVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGAYHDL
LKIIKDKDFLDNEENEDILEDIVLTLTLFEDRGMIEERLKTYAHLFDDKV
MKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLI
HDDSLTFKEDIQKAQVSGQGHSLHEQIANLAGSPAIKKGILQTVKIVDEL
[0120] VKVMGHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKE
HPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFIK
DDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFD
NLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDK
LIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALI
KKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKT
EITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKT
EVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAK
VEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKL
PKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGS
PEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHR
DKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLI
HQSITGLYETRIDLSQLGGD

[0121] (单底线:HNH结构域;双底线:RuvC结构域)

[0122] 在一些实施方式中,野生型Cas9对应于,或包括以下的核苷酸和/或氨基酸序列:

[0123] ATGGATAAAAAGTATTCTATTGGTTTAGACATCGGCACTAATTCCGTTGGATGGGCTGTCATAACCGAT
GAATACAAAGTACCTTCAAAGAAATTTAAGGTGTTGGGGAACACAGACCGTCATTCGATTAAAAAGAATCTTATCGG
TGCCCTCCTATTCGATAGTGGCGAAACGGCAGAGGCGACTCGCCTGAAACGAACCGCTCGGAGAAGGTATACACGTC
GCAAGAACCGAATATGTTACTTACAAGAAATTTTTAGCAATGAGATGGCCAAAGTTGACGATTCTTTCTTTCACCGT
TTGGAAGAGTCCTTCCTTGTCGAAGAGGACAAGAAACATGAACGGCACCCCATCTTTGGAAACATAGTAGATGAGGT
GGCATATCATGAAAAGTACCCAACGATTTATCACCTCAGAAAAAAGCTAGTTGACTCAACTGATAAAGCGGACCTGA
GGTTAATCTACTTGGCTCTTGCCCATATGATAAAGTTCCGTGGGCACTTTCTCATTGAGGGTGATCTAAATCCGGAC
AACTCGGATGTCGACAAACTGTTCATCCAGTTAGTACAAACCTATAATCAGTTGTTTGAAGAGAACCCTATAAATGC
AAGTGGCGTGGATGCGAAGGCTATTCTTAGCGCCCGCCTCTCTAAATCCCGACGGCTAGAAAACCTGATCGCACAAT
TACCCGGAGAGAAGAAAAATGGGTTGTTCGGTAACCTTATAGCGCTCTCACTAGGCCTGACACCAAATTTTAAGTCG

AACTTCGACTTAGCTGAAGATGCCAAATTGCAGCTTAGTAAGGACACGTACGATGACGATCTCGACAATCTACTGGC
ACAAATTGGAGATCAGTATGCGGACTTATTTTTGGCTGCCAAAAACCTTAGCGATGCAATCCTCCTATCTGACATAC
TGAGAGTTAATACTGAGATTACCAAGGCGCCGTTATCCGCTTCAATGATCAAAAGGTACGATGAACATCACCAAGAC
TTGACACTTCTCAAGGCCCTAGTCCGTCAGCAACTGCCTGAGAAATATAAGGAAATATTCTTTGATCAGTCGAAAAA
CGGGTACGCAGGTTATATTGACGGCGGAGCGAGTCAAGAGGAATTCTACAAGTTTATCAAACCCATATTAGAGAAGA
TGGATGGGACGGAAGAGTTGCTTGTAAAACTCAATCGCGAAGATCTACTGCGAAAGCAGCGGACTTTCGACAACGGT
AGCATTCCACATCAAATCCACTTAGGCGAATTGCATGCTATACTTAGAAGGCAGGAGGATTTTTATCCGTTCCTCAA
AGACAATCGTGAAAAGATTGAGAAAATCCTAACCTTTCGCATACCTTACTATGTGGGACCCCTGGCCCGAGGGAACT
CTCGGTTCGCATGGATGACAAGAAAGTCCGAAGAAACGATTACTCCATGGAATTTTGAGGAAGTTGTCGATAAAGGT
GCGTCAGCTCAATCGTTCATCGAGAGGATGACCAACTTTGACAAGAATTTACCGAACGAAAAAGTATTGCCTAAGCA
CAGTTTACTTTACGAGTATTTCACAGTGTACAATGAACTCACGAAAGTTAAGTATGTCACTGAGGGCATGCGTAAAC
CCGCCTTTCTAAGCGGAGAACAGAAGAAAGCAATAGTAGATCTGTTATTCAAGACCAACCGCAAAGTGACAGTTAAG
CAATTGAAAGAGGACTACTTTAAGAAAATTGAATGCTTCGATTCTGTCGAGATCTCCGGGGTAGAAGATCGATTTAA
TGCGTCACTTGGTACGTATCATGACCTCCTAAAGATAATTAAAGATAAGGACTTCCTGGATAACGAAGAGAATGAAG
ATATCTTAGAAGATATAGTGTTGACTCTTACCCTCTTTGAAGATCGGGAAATGATTGAGGAAAGACTAAAAACATAC
GCTCACCTGTTCGACGATAAGGTTATGAAACAGTTAAAGAGGCGTCGCTATACGGGCTGGGGACGATTGTCGCGGAA
ACTTATCAACGGGATAAGAGACAAGCAAAGTGGTAAAACTATTCTCGATTTTCTAAAGAGCGACGGCTTCGCCAATA
GGAACTTTATGCAGCTGATCCATGATGACTCTTTAACCTTCAAAGAGGATATACAAAAGGCACAGGTTTCCGGACAA
GGGGACTCATTGCACGAACATATTGCGAATCTTGCTGGTTCGCCAGCCATCAAAAAGGGCATACTCCAGACAGTCAA
AGTAGTGGATGAGCTAGTTAAGGTCATGGGACGTCACAAACCGGAAAACATTGTAATCGAGATGGCACGCGAAAATC
AAACGACTCAGAAGGGGCAAAAAAACAGTCGAGAGCGGATGAAGAGAATAGAAGAGGGTATTAAAGAACTGGGCAGC
CAGATCTTAAAGGAGCATCCTGTGGAAAATACCCAATTGCAGAACGAGAAACTTTACCTCTATTACCTACAAAATGG
AAGGGACATGTATGTTGATCAGGAACTGGACATAAACCGTTTATCTGATTACGACGTCGATCACATTGTACCCCAAT
CCTTTTTGAAGGACGATTCAATCGACAATAAAGTGCTTACACGCTCGGATAAGAACCGAGGGAAAAGTGACAATGTT
CCAAGCGAGGAAGTCGTAAAGAAAATGAAGAACTATTGGCGGCAGCTCCTAAATGCGAAACTGATAACGCAAAGAAA
GTTCGATAACTTAACTAAAGCTGAGAGGGGTGGCTTGTCTGAACTTGACAAGGCCGGATTTATTAAACGTCAGCTCG
TGGAAACCCGCCAAATCACAAAGCATGTTGCACAGATACTAGATTCCCGAATGAATACGAAATACGACGAGAACGAT
AAGCTGATTCGGGAAGTCAAAGTAATCACTTTAAAGTCAAAATTGGTGTCGGACTTCAGAAAGGATTTTCAATTCTA
TAAAGTTAGGGAGATAAATAACTACCACCATGCGCACGACGCTTATCTTAATGCCGTCGTAGGGACCGCACTCATTA
AGAAATACCCGAAGCTAGAAAGTGAGTTTGTGTATGGTGATTACAAAGTTTATGACGTCCGTAAGATGATCGCGAAA
AGCGAACAGGAGATAGGCAAGGCTACAGCCAAATACTTCTTTTATTCTAACATTATGAATTTCTTTAAGACGGAAAT
CACTCTGGCAAACGGAGAGATACGCAAACGACCTTTAATTGAAACCAATGGGGAGACAGGTGAAATCGTATGGGATA
AGGGCCGGGACTTCGCGACGGTGAGAAAAGTTTTGTCCATGCCCCAAGTCAACATAGTAAAGAAAACTGAGGTGCAG
ACCGGAGGGTTTTCAAAGGAATCGATTCTTCCAAAAAGGAATAGTGATAAGCTCATCGCTCGTAAAAAGGACTGGGA
CCCGAAAAAGTACGGTGGCTTCGATAGCCCTACAGTTGCCTATTCTGTCCTAGTAGTGGCAAAAGTTGAGAAGGGAA
AATCCAAGAAACTGAAGTCAGTCAAAGAATTATTGGGGATAACGATTATGGAGCGCTCGTCTTTTGAAAAGAACCCC
ATCGACTTCCTTGAGGCGAAAGGTTACAAGGAAGTAAAAAAGGATCTCATAATTAAACTACCAAAGTATAGTCTGTT
TGAGTTAGAAAATGGCCGAAAACGGATGTTGGCTAGCGCCGGAGAGCTTCAAAAGGGGAACGAACTCGCACTACCGT
CTAAATACGTGAATTTCCTGTATTTAGCGTCCCATTACGAGAAGTTGAAAGGTTCACCTGAAGATAACGAACAGAAG

CAACTTTTTGTTGAGCAGCACAAACATTATCTCGACGAAATCATAGAGCAAATTTCGGAATTCAGTAAGAGAGTCAT
CCTAGCTGATGCCAATCTGGACAAAGTATTAAGCGCATACAACAAGCACAGGGATAAACCCATACGTGAGCAGGCGG
AAAATATTATCCATTTGTTTACTCTTACCAACCTCGGCGCTCCAGCCGCATTCAAGTATTTTGACACAACGATAGAT
CGCAAACGATACACTTCTACCAAGGAGGTGCTAGACGCGACACTGATTCACCAATCCATCACGGGATTATATGAAAC
TCGGATAGATTTGTCACAGCTTGGGGGTGACGGATCCCCCAAGAAGAAGAGGAAAGTCTCGAGCGACTACAAAGACC
ATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGGCTGCAGGA

[0124]
MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNL
IGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSF
FHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTD
KADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFE
ENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLG
LTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLS
DAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKY
KEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRED
LLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFR
IPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTN
FDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKA
IVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL
LKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKV
MKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLI
HDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDEL
VKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILK
EHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFL
KDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF
DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDEND
KLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTAL
IKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFK
TEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKK
TEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIK
LPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKG
SPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH
RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATL
IHQSITGLYETRIDLSQLGGD

[0126] (单底线:HNH结构域;双底线:RuvC结构域)

[0127] 在一些实施方式中,野生型Cas9对应于来自化脓性链球菌的Cas9(NCBI参考序列:

NC_002737.2)(核苷酸序列如下);和Uniprot参考序列:Q99ZW2(氨基酸序列如下)。

[0128] ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGCGTCGGATGGGCGGTGATCACTGAT
GAATATAAGGTTCCGTCTAAAAAGTTCAAGGTTCTGGGAAATACAGACCGCCACAGTATCAAAAAAAATCTTATAGG
GGCTCTTTTATTTGACAGTGGAGAGACAGCGGAAGCGACTCGTCTCAAACGGACAGCTCGTAGAAGGTATACACGTC
GGAAGAATCGTATTTGTTATCTACAGGAGATTTTTTCAAATGAGATGGCGAAAGTAGATGATAGTTTCTTTCATCGA
CTTGAAGAGTCTTTTTTGGTGGAAGAAGACAAGAAGCATGAACGTCATCCTATTTTTGGAAATATAGTAGATGAAGT
TGCTTATCATGAGAAATATCCAACTATCTATCATCTGCGAAAAAAATTGGTAGATTCTACTGATAAAGCGGATTTGC
GCTTAATCTATTTGGCCTTAGCGCATATGATTAAGTTTCGTGGTCATTTTTTGATTGAGGGAGATTTAAATCCTGAT
AATAGTGATGTGGACAAACTATTTATCCAGTTGGTACAAACCTACAATCAATTATTTGAAGAAAACCCTATTAACGC
AAGTGGAGTAGATGCTAAAGCGATTCTTTCTGCACGATTGAGTAAATCAAGACGATTAGAAAATCTCATTGCTCAGC
TCCCCGGTGAGAAGAAAAATGGCTTATTTGGGAATCTCATTGCTTTGTCATTGGGTTTGACCCCTAATTTTAAATCA
AATTTTGATTTGGCAGAAGATGCTAAATTACAGCTTTCAAAAGATACTTACGATGATGATTTAGATAATTTATTGGC
GCAAATTGGAGATCAATATGCTGATTTGTTTTTGGCAGCTAAGAATTTATCAGATGCTATTTTACTTTCAGATATCC
TAAGAGTAAATACTGAAATAACTAAGGCTCCCCTATCAGCTTCAATGATTAAACGCTACGATGAACATCATCAAGAC
TTGACTCTTTTAAAAGCTTTAGTTCGACAACAACTTCCAGAAAAGTATAAAGAAATCTTTTTTGATCAATCAAAAAA
CGGATATGCAGGTTATATTGATGGGGGAGCTAGCCAAGAAGAATTTTATAAATTTATCAAACCAATTTTAGAAAAAA
TGGATGGTACTGAGGAATTATTGGTGAAACTAAATCGTGAAGATTTGCTGCGCAAGCAACGGACCTTTGACAACGGC
TCTATTCCCCATCAAATTCACTTGGGTGAGCTGCATGCTATTTTGAGAAGACAAGAAGACTTTTATCCATTTTTAAA
AGACAATCGTGAGAAGATTGAAAAAATCTTGACTTTTCGAATTCCTTATTATGTTGGTCCATTGGCGCGTGGCAATA
GTCGTTTTGCATGGATGACTCGGAAGTCTGAAGAAACAATTACCCCATGGAATTTTGAAGAAGTTGTCGATAAAGGT
GCTTCAGCTCAATCATTTATTGAACGCATGACAAACTTTGATAAAAATCTTCCAAATGAAAAAGTACTACCAAAACA
TAGTTTGCTTTATGAGTATTTTACGGTTTATAACGAATTGACAAAGGTCAAATATGTTACTGAAGGAATGCGAAAAC
CAGCATTTCTTTCAGGTGAACAGAAGAAAGCCATTGTTGATTTACTCTTCAAAACAAATCGAAAAGTAACCGTTAAG
CAATTAAAAGAAGATTATTTCAAAAAAATAGAATGTTTTGATAGTGTTGAAATTTCAGGAGTTGAAGATAGATTTAA
TGCTTCATTAGGTACCTACCATGATTTGCTAAAAATTATTAAAGATAAAGATTTTTTGGATAATGAAGAAAATGAAG
ATATCTTAGAGGATATTGTTTTAACATTGACCTTATTTGAAGATAGGGAGATGATTGAGGAAAGACTTAAAACATAT
GCTCACCTCTTTGATGATAAGGTGATGAAACAGCTTAAACGTCGCCGTTATACTGGTTGGGGACGTTTGTCTCGAAA
ATTGATTAATGGTATTAGGGATAAGCAATCTGGCAAAACAATATTAGATTTTTTGAAATCAGATGGTTTTGCCAATC
GCAATTTTATGCAGCTGATCCATGATGATAGTTTGACATTTAAAGAAGACATTCAAAAAGCACAAGTGTCTGGACAA
GGCGATAGTTTACATGAACATATTGCAAATTTAGCTGGTAGCCCTGCTATTAAAAAAGGTATTTTACAGACTGTAAA
AGTTGTTGATGAATTGGTCAAAGTAATGGGGCGGCATAAGCCAGAAAATATCGTTATTGAAATGGCACGTGAAAATC
AGACAACTCAAAAGGGCCAGAAAAATTCGCGAGAGCGTATGAAACGAATCGAAGAAGGTATCAAAGAATTAGGAAGT
CAGATTCTTAAAGAGCATCCTGTTGAAAATACTCAATTGCAAAATGAAAAGCTCTATCTCTATTATCTCCAAAATGG
AAGAGACATGTATGTGGACCAAGAATTAGATATTAATCGTTTAAGTGATTATGATGTCGATCACATTGTTCCACAAA
GTTTCCTTAAAGACGATTCAATAGACAATAAGGTCTTAACGCGTTCTGATAAAAATCGTGGTAAATCGGATAACGTT
CCAAGTGAAGAAGTAGTCAAAAAGATGAAAAACTATTGGAGACAACTTCTAAACGCCAAGTTAATCACTCAACGTAA
GTTTGATAATTTAACGAAAGCTGAACGTGGAGGTTTGAGTGAACTTGATAAAGCTGGTTTTATCAAACGCCAATTGG
TTGAAACTCGCCAAATCACTAAGCATGTGGCACAAATTTTGGATAGTCGCATGAATACTAAATACGATGAAAATGAT
AAACTTATTCGAGAGGTTAAAGTGATTACCTTAAAATCTAAATTAGTTTCTGACTTCCGAAAAGATTTCCAATTCTA

TAAAGTACGTGAGATTAACAATTACCATCATGCCCATGATGCGTATCTAAATGCCGTCGTTGGAACTGCTTTGATTA
AGAAATATCCAAAACTTGAATCGGAGTTTGTCTATGGTGATTATAAAGTTTATGATGTTCGTAAAATGATTGCTAAG
TCTGAGCAAGAAATAGGCAAAGCAACCGCAAAATATTTCTTTTACTCTAATATCATGAACTTCTTCAAAACAGAAAT
TACACTTGCAAATGGAGAGATTCGCAAACGCCCTCTAATCGAAACTAATGGGGAAACTGGAGAAATTGTCTGGGATA
AAGGGCGAGATTTTGCCACAGTGCGCAAAGTATTGTCCATGCCCCAAGTCAATATTGTCAAGAAAACAGAAGTACAG
ACAGGCGGATTCTCCAAGGAGTCAATTTTACCAAAAAGAAATTCGGACAAGCTTATTGCTCGTAAAAAAGACTGGGA
TCCAAAAAAATATGGTGGTTTTGATAGTCCAACGGTAGCTTATTCAGTCCTAGTGGTTGCTAAGGTGGAAAAAGGGA
AATCGAAGAAGTTAAAATCCGTTAAAGAGTTACTAGGGATCACAATTATGGAAAGAAGTTCCTTTGAAAAAAATCCG
ATTGACTTTTTAGAAGCTAAAGGATATAAGGAAGTTAAAAAAGACTTAATCATTAAACTACCTAAATATAGTCTTTT
TGAGTTAGAAAACGGTCGTAAACGGATGCTGGCTAGTGCCGGAGAATTACAAAAAGGAAATGAGCTGGCTCTGCCAA
GCAAATATGTGAATTTTTTATATTTAGCTAGTCATTATGAAAAGTTGAAGGGTAGTCCAGAAGATAACGAACAAAAA
CAATTGTTTGTGGAGCAGCATAAGCATTATTTAGATGAGATTATTGAGCAAATCAGTGAATTTTCTAAGCGTGTTAT
TTTAGCAGATGCCAATTTAGATAAAGTTCTTAGTGCATATAACAAACATAGAGACAAACCAATACGTGAACAAGCAG
AAAATATTATTCATTTATTTACGTTGACGAATCTTGGAGCTCCCGCTGCTTTTAAATATTTTGATACAACAATTGAT
CGTAAACGATATACGTCTACAAAAGAAGTTTTAGATGCCACTCTTATCCATCAATCCATCACTGGTCTTTATGAAAC
ACGCATTGATTTGAGTCAGCTAGGAGGTGACTGA

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNL
IGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSF
FHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTD
KADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFE
ENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLG
[0129] LTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLS
DAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKY
KEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRED
LLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFR
IPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTN
FDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKA

[0130] IVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL
LKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKV
MKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLI
HDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDEL
VKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILK
EHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFL
KDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF
DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDEND
KLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTAL
IKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFK
TEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKK
TEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIK
LPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKG
SPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH
RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATL
IHQSITGLYETRIDLSQLGGD(SEQ ID NO:1。单底线:HNH结构域;双底线:RuvC结构域)。

[0131] 在一些实施方式中,Cas9是指来自以下的Cas9:溃疡棒状杆菌(Corynebacterium ulcerans)(NCBI Refs:NC_015683.1,NC_017317.1);白喉棒状杆菌(Corynebacterium diphtheria)(NCBI Refs:NC_016782.1,NC_016786.1);栖蚜蝇螺原体(Spiroplasma syrphidicola)(NCBI Ref:NC_021284.1);中间普雷沃菌(Prevotella intermedia)(NCBI Ref:NC_017861.1);中国台湾螺原体(Spiroplasma taiwanense,China)(NCBI Ref:NC_021846.1);海豚链球菌(Streptococcus iniae)(NCBI Ref:NC_021314.1);波罗的海贝尔氏菌(Belliella baltica)(NCBI Ref:NC_018010.1);扭曲冷弯曲菌I(Psychroflexus torquisI)(NCBI Ref:NC_018721.1);嗜热链球菌(Streptococcus thermophilus)(NCBI Ref:YP_820832.1),无害李斯特菌(Listeria innocua)(NCBI Ref:NP_472073.1),空肠弯曲菌(Campylobacter jejuni)(NCBI Ref:YP_002344900.1)或脑膜炎双球菌(Neisseria.meningitidis)(NCBI Ref:YP_002342100.1)或来自任何其他生物体的Cas9。

[0132] 在一些实施方式中,dCas9对应于,或包括Cas9氨基酸序列的一部分或全部,该序列具有一个或多个灭活Cas9核酸酶活性的突变。例如,在一些实施方式中,dCas9结构域包括D10A和H840A突变,或另一个Cas9中的相应突变。在一些实施方式中,该dCas9包括dCas9(D10A和H840A)的氨基酸序列:

[0133]
MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNL
IGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSF
FHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTD
KADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFE
ENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLG
LTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLS
DAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKY
KEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRED
LLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFR
IPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTN
FDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKA
IVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL
LKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKV
MKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLI
HDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDEL
VKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILK
EHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFL
KDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF
DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDEND
KLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTAL
IKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFK
TEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKK
TEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIK
LPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKG
SPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH
RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATL
IHQSITGLYETRIDLSQLGGD

[0134] (单底线:HNH结构域;双底线:RuvC结构域)。

[0135] 在一些实施方式中,该Cas9结构域包括D10A突变,而在上面所提供的氨基酸序列中位置840的残基,或在任何本文所提供的氨基酸序列的相应位置上的残基,仍然是组氨酸。

[0136] 在其他实施方式中,提供了具有除了D10A和H840A以外突变的dCas9变体,所述突变,例如,导致了核酸酶灭活的Cas9(dCas9)。此种突变,举例来说,包含在D10和H840处其他的氨基酸替换,或在Cas9的核酸酶结构域内的其他替换(例如,在HNH核酸酶亚结构域和/或RuvC1亚结构域内的替换)。在一些实施方式中,提供了dCas9的变体或同源物,其为至少约

70％相同,至少约80％相同,至少约90％相同,至少约95％相同,至少约98％相同,至少约99％相同,至少约99.5％相同,或至少约99.9％相同。在一些实施方式中,提供了dCas9的变体,其所具有的氨基酸序列的较短或较长的程度为约5个氨基酸,约10个氨基酸,约15个氨基酸,约20个氨基酸,约25个氨基酸,约30个氨基酸,约40个氨基酸,约50个氨基酸,约75个氨基酸,约100个氨基酸或更多。

[0137] 在一些实施方式中,如本文所提供的Cas9融合蛋白包括Cas9蛋白的全长氨基酸序列,例如,本文所提供的Cas9序列之一。然而,在其他实施方式中,如本文所提供的融合蛋白不包括全长Cas9序列,而是仅包含其一个或多个片段。本文提供了适合的Cas9结构域和Cas9片段的示例性氨基酸序列,而额外的适合的Cas9结构域和片段的序列对本领域技术人员而言将是显而易见的。

[0138] 在一些实施方式中,Cas9是指来自以下的Cas9:溃疡棒状杆菌(NCBI Refs:NC_015683.1,NC_017317.1);白喉棒状杆菌(NCBI Refs:NC_016782.1,NC_016786.1);栖蚜蝇螺原体(NCBI Ref:NC_021284.1);中间普雷沃菌(NCBI Ref:NC_017861.1);中国台湾螺原体(NCBI Ref:NC_021846.1);海豚链球菌(NCBI Ref:NC_021314.1);波罗的海贝尔氏菌(Belliella baltica)(NCBI Ref:NC_018010.1);扭曲冷弯曲菌I(NCBI Ref:NC_018721.1);嗜热链球菌(NCBI Ref:YP_820832.1);无害李斯特菌(NCBI Ref:NP_472073.1);空肠弯曲菌(NCBI Ref:YP_002344900.1);或脑膜炎双球菌(NCBI Ref:YP_002342100.1)。

[0139] 应当理解,额外的Cas9蛋白(例如核酸酶失活的Cas9(dCas9),Cas9切口酶(nCas9),或核酸酶有活性的Cas9),包含其变体和同源物,是在本公开揭露的范围之内。示例性的Cas9蛋白包含但不限于以下所提供的那些。在一些实施方式中,该Cas9蛋白是核酸酶失活的Cas9(dCas9)。在一些实施方式中,该Cas9蛋白是Cas9切口酶(nCas9)。在一些实施方式中,该Cas9蛋白是核酸酶有活性的Cas9。

[0140] 示例性的催化无活性的Cas9(dCas9):

[0141] DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAK

VEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNE
LALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPI
REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[0142] 示例性的催化的Cas9切口酶(nCas9):

[0143] DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR
RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTD
KADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLEN
LIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAIL
LSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKP
ILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPL
ARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE
GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDN
EENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSD
GFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEM
ARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDH
IVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFI
KRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVG
TALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGE
IVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAK
VEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNE
LALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPI
REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[0144] 示例性的催化有活性的Cas9:

[0145] DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDS
TDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRR
LENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLS
DAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFY
KFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIP
YYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELT
KVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKII
KDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGK
TILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGR
HKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELD
INRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAER
GGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNY
HHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEI
RKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGG
FDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG

RKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADA NLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRID LSQLGGD.

[0146] 在一些实施方式中,Cas9是指来自古菌(例如,纳古菌门)的Cas9,而古菌构成了单-细胞原核微生物的一个域和界。在一些实施方式中,Cas是指CasX或CasY,其已被描述过,例如,在Burstein等,“来自未培育的微生物的新CRISPR-Cas系统”Cell Res.2017Feb 21.doi:10.1038/cr.2017.21中,其全部内容在此通过引用并入。使用基因组-分辨的宏基因组学,识别出了一些CRISPR-Cas系统,包含在生命的古菌域中首次被报道的Cas9。这种相异的Cas9蛋白是在很少-被研究的纳古菌属中作为有活性的CRISPR-Cas的一部分而被发现的。在细菌中,发现了两个之前未知的系统,CRISPR-CasX和CRISPR-CasY,其为迄今为止所发现的最致密紧凑的系统之一。在一些实施方式中,Cas9是指CasX,或CasX的变体。在一些实施方式中,Cas9是指CasY,或CasY的变体。应该理解,其他RNA-导向的DNA结合蛋白可能被用作核酸可编程的DNA结合蛋白(napDNAbp),并且是在本公开揭露的范围内。

[0147] 在一些实施方式中,本文所提供的任何融合蛋白的核酸可编程的DNA结合蛋白(napDNAbp)可能是CasX或CasY蛋白。在一些实施方式中,该napDNAbp是CasX蛋白。在一些实施方式中,该napDNAbp是CasY蛋白。在一些实施方式中,该napDNAbp所包括的氨基酸序列与天然-存在的CasX或CasY蛋白是至少85%,至少90%,至少91%,至少92%,至少93%,至少94%,至少95%,至少96%,至少97%,至少98%,至少99%,或至少99.5%相同。在一些实施方式中,该napDNAbp是天然-存在的CasX或CasY蛋白。在一些实施方式中,该napDNAbp所包括的氨基酸序列与本文所描述的任何CasX或CasY蛋白是至少85%,至少90%,至少91%,至少92%,至少93%,至少94%,至少95%,至少96%,至少97%,至少98%,至少99%,或至少99.5%相同。应该理解,来自其他细菌种类的CasX和CasY也可根据本公开揭露而被使用。

[0148] CasX

[0149] (uniprot.org/uniprot/F0NN87;uniprot.org/uniprot/F0NH53)>tr|F0NN87|F0NN87_SULIH CRISPR-关联的Casx蛋白OS=冰岛硫化叶菌(Sulfolobus islandicus)(菌株HVE10/4)GN=SiH_0402PE=4SV=1

[0150] MEVPLYNIFGDNYIIQVATEAENSTIYNNKVEIDDEELRNVLNLAYKIAKNNEDAAAERRGKAKKKKGE EGETTTSNIILPLSGNDKNPWTETLKCYNFPTTVALSEVFKNFSQVKECEEVSAPSFVKPEFYEFGRSPGMVERTRR VKLEVEPHYLIIAAAGWVLTRLGKAKVSEGDYVGVNVFTPTRGILYSLIQNVNGIVPGIKPETAFGLWIARKVVSSV TNPNVSVVRIYTISDAVGQNPTTINGGFSIDLTKLLEKRYLLSERLEAIARNALSISSNMRERYIVLANYIYEYLTG SKRLEDLLYFANRDLIMNLNSDDGKVRDLKLISAYVNGELIRGEG

[0151] >tr|F0NH53|F0NH53_SULIR CRISPR-关联的蛋白,Casx OS

[0152] =冰岛硫化叶菌(菌株REY15A)GN=SiRe_0771PE=4SV=1

[0153] MEVPLYNIFGDNYIIQVATEAENSTIYNNKVEIDDEELRNVLNLAYKIAKNNEDAAAERRGKAKKKKGE EGETTTSNIILPLSGNDKNPWTETLKCYNFPTTVALSEVFKNFSQVKECEEVSAPSFVKPEFYKFGRSPGMVERTRR VKLEVEPHYLIMAAAGWVLTRLGKAKVSEGDYVGVNVFTPTRGILYSLIQNVNGIVPGIKPETAFGLWIARKVVSSV TNPNVSVVSIYTISDAVGQNPTTINGGFSIDLTKLLEKRDLLSERLEAIARNALSISSNMRERYIVLANYIYEYLTG SKRLEDLLYFANRDLIMNLNSDDGKVRDLKLISAYVNGELIRGEG

[0154] CasY(ncbi.nlm.nih.gov/protein/APG80656.1)

[0155] >APG80656.1 CRISPR-关联的蛋白CasY[未培养的俭菌(Parcubacteria)菌群细菌]

[0156] MSKRHPRISGVKGYRLHAQRLEYTGKSGAMRTIKYPLYSSPSGGRTVPREIVSAINDDYVGLYGLSNFD
DLYNAEKRNEEKVYSVLDFWYDCVQYGAVFSYTAPGLLKNVAEVRGGSYELTKTLKGSHLYDELQIDKVIKFLNKKE
ISRANGSLDKLKKDIIDCFKAEYRERHKDQCNKLADDIKNAKKDAGASLGERQKKLFRDFFGISEQSENDKPSFTNP
LNLTCCLLPFDTVNNNRNRGEVLFNKLKEYAQKLDKNEGSLEMWEYIGIGNSGTAFSNFLGEGFLGRLRENKITELK
KAMMDITDAWRGQEQEEELEKRLRILAALTIKLREPKFDNHWGGYRSDINGKLSSWLQNYINQTVKIKEDLKGHKKD
LKKAKEMINRFGESDTKEEAVVSSLLESIEKIVPDDSADDEKPDIPAIAIYRRFLSDGRLTLNRFVQREDVQEALIK
ERLEAEKKKKPKKRKKKSDAEDEKETIDFKELFPHLAKPLKLVPNFYGDSKRELYKKYKNAAIYTDALWKAVEKIYK
SAFSSSLKNSFFDTDFDKDFFIKRLQKIFSVYRRFNTDKWKPIVKNSFAPYCDIVSLAENEVLYKPKQSRSRKSAAI
DKNRVRLPSTENIAKAGIALARELSVAGFDWKDLLKKEEHEEYIDLIELHKTALALLLAVTETQLDISALDFVENGT
VKDFMKTRDGNLVLEGRFLEMFSQSIVFSELRGLAGLMSRKEFITRSAIQTMNGKQAELLYIPHEFQSAKITTPKEM
SRAFLDLAPAEFATSLEPESLSEKSLLKLKQMRYYPHYFGYELTRTGQGIDGGVAENALRLEKSPVKKREIKCKQYK
TLGRGQNKIVLYVRSSYYQTQFLEWFLHRPKNVQTDVAVSGSFLIDEKKVKTRWNYDALTVALEPVSGSERVFVSQP
FTIFPEKSAEEEGQRYLGIDIGEYGIAYTALEITGDSAKILDQNFISDPQLKTLREEVKGLKLDQRRGTFAMPSTKI
ARIRESLVHSLRNRIHHLALKHKAKIVYELEVSRFEEGKQKIKKVYATLKKADVYSEIDADKNLQTTVWGKLAVASE
ISASYTSQFCGACKKLWRAEMQVDETITTQELIGTVRVIKGGTLIDAIKDFMRPPIFDENDTPFPKYRDFCDKHHIS
KKMRGNSCLFICPFCRANADADIQASQTIALLRYVKEEKKVEDYFERFRKLKNIKVLGQMKKI

[0157] “胞苷脱氨酶”意为能够催化将氨基基团转换为羰基基团的脱氨反应的多肽或其片段。在一个实施方式中,该胞苷脱氨酶将胞嘧啶转换为尿嘧啶,或将5-甲基胞嘧啶转换为胸腺嘧啶。PmCDA1,其衍生自海七鳃鳗(海七鳃鳗胞嘧啶脱氨酶1,“PmCDA1”),AID(活化-诱导的胞苷脱氨酶;AICDA),其衍生自哺乳动物(例如,人类,猪,牛,马,猴等),以及APOBEC是示例性的胞苷脱氨酶。

[0158] PmCDA1的碱基序列和氨基酸序列,以及人类AID的CDS的碱基序列和氨基酸序列如本文下面所示:

[0159] >tr|A5H718|A5H718_PETMA胞嘧啶脱氨酶OS＝海七鳃鳗OX＝7757PE＝2SV＝1

[0160] MTDAEYVRIHEKLDIYTFKKQFFNNKKSVSHRCYVLFELKRRGERRACFWGYAVNKPQSGTERGIHAEI
FSIRKVEEYLRDNPGQFTINWYSSWSPCADCAEKILEWYNQELRGNGHTLKIWACKLYYEKNARNQIGLWNLRDNGV
GLNVMVSEHYQCCRKIFIQSSHNQLNENRWLEKTLKRAEKRRSELSIMIQVKILHTTKSPAV

[0161] >EF094822.1海七鳃鳗分离的PmCDA.21胞嘧啶脱氨酶mRNA,完整的cds

[0162] TGACACGACACAGCCGTGTATATGAGGAAGGGTAGCTGGATGGGGGGGGGGGGAATACGTTCAGAGAGG
ACATTAGCGAGCGTCTTGTTGGTGGCCTTGAGTCTAGACACCTGCAGACATGACCGACGCTGAGTACGTGAGAATCC
ATGAGAAGTTGGACATCTACACGTTTAAGAAACAGTTTTTCAACAACAAAAAATCCGTGTCGCATAGATGCTACGTT
CTCTTTGAATTAAAACGACGGGGTGAACGTAGAGCGTGTTTTTGGGGCTATGCTGTGAATAAACCACAGAGCGGGAC
AGAACGTGGAATTCACGCCGAAATCTTTAGCATTAGAAAAGTCGAAGAATACCTGCGCGACAACCCCGGACAATTCA
CGATAAATTGGTACTCATCCTGGAGTCCTTGTGCAGATTGCGCTGAAAAGATCTTAGAATGGTATAACCAGGAGCTG
CGGGGGAACGGCCACACTTTGAAAATCTGGGCTTGCAAACTCTATTACGAGAAAAATGCGAGGAATCAAATTGGGCT
GTGGAACCTCAGAGATAACGGGGTTGGGTTGAATGTAATGGTAAGTGAACACTACCAATGTTGCAGGAAAATATTCA
TCCAATCGTCGCACAATCAATTGAATGAGAATAGATGGCTTGAGAAGACTTTGAAGCGAGCTGAAAAACGACGGAGC

GAGTTGTCCATTATGATTCAGGTAAAAATACTCCACACCACTAAGAGTCCTGCTGTTTAAGAGGCTATGCGGATGGT
TTTC

[0163] >tr|Q6QJ80|Q6QJ80_HUMAN活化-诱导的胞苷脱氨酶OS＝智人OX＝9606GN＝AICDA
PE＝2SV＝1

[0164] MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGCHVELL

[0165] FLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLYFCEDRK

[0166] AEPEGLRRLHRAGVQIAIMTFKAPV

[0167] >NG_011588.1:5001-15681智人活化诱导的胞苷脱氨酶(AICDA),RefSeqGene
(LRG_17)在染色体12上

[0168] AGAGAACCATCATTAATTGAAGTGAGATTTTTCTGGCCTGAGACTTGCAGGGAGGCAAGAAGACACTCT
GGACACCACTATGGACAGGTAAAGAGGCAGTCTTCTCGTGGGTGATTGCACTGGCCTTCCTCTCAGAGCAAATCTGA
GTAATGAGACTGGTAGCTATCCCTTTCTCTCATGTAACTGTCTGACTGATAAGATCAGCTTGAT

[0169] CAATATGCATATATATTTTTTGATCTGTCTCCTTTTCTTCTATTCAGATCTTATACGCTGTCAGCCCAA
TTCTTTCTGTTTCAGACTTCTCTTGATTTCCCTCTTTTTCATGTGGCAAAAGAAGTAGTGCGTACAATGTACTGATT
CGTCCTGAGATTTGTACCATGGTTGAAACTAATTTATGGTAATAATATTAACATAGCAAATCTT

[0170] TAGAGACTCAAATCATGAAAAGGTAATAGCAGTACTGTACTAAAAACGGTAGTGCTAATTTTCGTAATA
ATTTTGTAAATATTCAACAGTAAAACAACTTGAAGACACACTTTCCTAGGGAGGCGTTACTGAAATAATTTAGCTAT
AGTAAGAAAATTTGTAATTTTAGAAATGCCAAGCATTCTAAATTAATTGCTTGAAAGTCACTAT

[0171] GATTGTGTCCATTATAAGGAGACAAATTCATTCAAGCAAGTTATTTAATGTTAAAGGCCCAATTGTTAG
GCAGTTAATGGCACTTTTACTATTAACTAATCTTTCCATTTGTTCAGACGTAGCTTAACTTACCTCTTAGGTGTGAA
TTTGGTTAAGGTCCTCATAATGTCTTTATGTGCAGTTTTTGATAGGTTATTGTCATAGAACTTATTCTATTCCTACA
TTTATGATTACTATGGATGTATGAGAATAACACCTAATCCTTATACTTTACCTCAATTTAACTCCTTTATAAAGAAC
TTACATTACAGAATAAAGATTTTTTAAAAATATATTTTTTTGTAGAGACAGGGTCTTAGCCCAGCCGAGGCTGGTCT
CTAAGTCCTGGCCCAAGCGATCCTCCTGCCTGGGCCTCCTAAAGTGCTGGAATTATAGACATGAGCCATCACATCCA
ATATACAGAATAAAGATTTTTAATGGAGGATTTAATGTTCTTCAGAAAATTTTCTTGAGGTCAGACAATGTCAAATG
TCTCCTCAGTTTACACTGAGATTTTGAAAACAAGTCTGAGCTATAGGTCCTTGTGAAGGGTCCATTGGAAATACTTG
TTCAAAGTAAAATGGAAAGCAAAGGTAAAATCAGCAGTTGAAATTCAGAGAAAGACAGAAAAGGAGAAAAGATGAAA
TTCAACAGGACAGAAGGGAAATATATTATCATTAAGGAGGACAGTATCTGTAGAGCTCATTAGTGATGGCAAAATGA
CTTGGTCAGGATTATTTTTAACCCGCTTGTTTCTGGTTTGCACGGCTGGGGATGCAGCTAGGGTTCTGCCTCAGGGA
GCACAGCTGTCCAGAGCAGCTGTCAGCCTGCAAGCCTGAAACACTCCCTCGGTAAAGTCCTTCCTACTCAGGACAGA
AATGACGAGAACAGGGAGCTGGAAACAGGCCCCTAACCAGAGAAGGGAAGTAATGGATCAACAAAGTTAACTAGCAG
GTCAGGATCACGCAATTCATTTCACTCTGACTGGTAACATGTGACAGAAACAGTGTAGGCTTATTGTATTTTCATGT
AGAGTAGGACCCAAAAATCCACCCAAAGTCCTTTATCTATGCCACATCCTTCTTATCTATACTTCCAGGACACTTTT
TCTTCCTTATGATAAGGCTCTCTCTCTCTCCACACACACACACACACACACACACACACACACACACACACACAA
ACACACACCCCGCCAACCAAGGTGCATGTAAAAAGATGTAGATTCCTCTGCCTTTCTCATCTACACAGCCCAGGAGG
GTAAGTTAATATAAGAGGGATTTATTGGTAAGAGATGATGCTTAATCTGTTTAACACTGGGCCTCAAAGAGAGAATT
TCTTTTCTTCTGTACTTATTAAGCACCTATTATGTGTTGAGCTTATATATACAAAGGGTTATTATATGCTAATATAG
TAATAGTAATGGTGGTTGGTACTATGGTAATTACCATAAAAATTATTATCCTTTTAAAATAAAGCTAATTATTATTG
GATCTTTTTTAGTATTCATTTTATGTTTTTTATGTTTTTGATTTTTTAAAAGACAATCTCACCCTGTTACCCAGGCT

GGAGTGCAGTGGTGCAATCATAGCTTTCTGCAGTCTTGAACTCCTGGGCTCAAGCAATCCTCCTGCCTTGGCCTCCC
AAAGTGTTGGGATACAGTCATGAGCCACTGCATCTGGCCTAGGATCCATTTAGATTAAAATATGCATTTTAAATTTT
AAAATAATATGGCTAATTTTTACCTTATGTAATGTGTATACTGGCAATAAATCTAGTTTGCTGCCTAAAGTTTAAAG
TGCTTTCCAGTAAGCTTCATGTACGTGAGGGGAGACATTTAAAGTGAAACAGACAGCCAGGTGTGGTGGCTCACGCC
TGTAATCCCAGCACTCTGGGAGGCTGAGGTGGGTGGATCGCTTGAGCCCTGGAGTTCAAGACCAGCCTGAGCAACAT
GGCAAAACGCTGTTTCTATAACAAAAATTAGCCGGGCATGGTGGCATGTGCCTGTGGTCCCAGCTACTAGGGGGCTG
AGGCAGGAGAATCGTTGGAGCCCAGGAGGTCAAGGCTGCACTGAGCAGTGCTTGCGCCACTGCACTCCAGCCTGGGT
GACAGGACCAGACCTTGCCTCAAAAAAATAAGAAGAAAAATTAAAAATAAATGGAAACAACTACAAAGAGCTGTTGT
CCTAGATGAGCTACTTAGTTAGGCTGATATTTTGGTATTTAACTTTTAAAGTCAGGGTCTGTCACCTGCACTACATT
ATTAAAATATCAATTCTCAATGTATATCCACACAAAGACTGGTACGTGAATGTTCATAGTACCTTTATTCACAAAAC
CCCAAAGTAGAGACTATCCAAATATCCATCAACAAGTGAACAAATAAACAAAATGTGCTATATCCATGCAATGGAAT
ACCACCCTGCAGTACAAAGAAGCTACTTGGGGATGAATCCCAAAGTCATGACGCTAAATGAAAGAGTCAGACATGAA
GGAGGAGATAATGTATGCCATACGAAATTCTAGAAAATGAAAGTAACTTATAGTTACAGAAAGCAAATCAGGGCAGG
CATAGAGGCTCACACCTGTAATCCCAGCACTTTGAGAGGCCACGTGGGAAGATTGCTAGAACTCAGGAGTTCAAGAC
CAGCCTGGGCAACACAGTGAAACTCCATTCTCCACAAAAATGGGAAAAAAAGAAAGCAAATCAGTGGTTGTCCTGTG
GGGAGGGGAAGGACTGCAAAGAGGGAAGAAGCTCTGGTGGGGTGAGGGTGGTGATTCAGGTTCTGTATCCTGACTGT
GGTAGCAGTTTGGGGTGTTTACATCCAAAAATATTCGTAGAATTATGCATCTTAAATGGGTGGAGTTTACTGTATGT
AAATTATACCTCAATGTAAGAAAAAATAATGTGTAAGAAAACTTTCAATTCTCTTGCCAGCAAACGTTATTCAAATT
CCTGAGCCCTTTACTTCGCAAATTCTCTGCACTTCTGCCCCGTACCATTAGGTGACAGCACTAGCTCCACAAATTGG
ATAAATGCATTTCTGGAAAAGACTAGGGACAAAATCCAGGCATCACTTGTGCTTTCATATCAACCATGCTGTACAGC
TTGTGTTGCTGTCTGCAGCTGCAATGGGGACTCTTGATTTCTTTAAGGAAACTTGGGTTACCAGAGTATTTCCACAA
ATGCTATTCAAATTAGTGCTTATGATATGCAAGACACTGTGCTAGGAGCCAGAAAACAAAGAGGAGGAGAAATCAGT
CATTATGTGGGAACAACATAGCAAGATATTTAGATCATTTTGACTAGTTAAAAAAGCAGCAGAGTACAAAATCACAC
ATGCAATCAGTATAATCCAAATCATGTAAATATGTGCCTGTAGAAAGACTAGAGGAATAAACACAAGAATCTTAACA
GTCATTGTCATTAGACACTAAGTCTAATTATTATTATTAGACACTATGATATTTGAGATTTAAAAAATCTTTAATAT
TTTAAAATTTAGAGCTCTTCTATTTTTCCATAGTAT

[0172] TCAAGTTTGACAATGATCAAGTATTACTCTTTCTTTTTTTTTTTTTTTTTTTTTTTGAGATGGAGTT
TTGGTCTTGTTGCCCATGCTGGAGTGGAATGGCATGACCATAGCTCACTGCAACCTCCACCTCCTGGGTTCAAGCAA
AGCTGTCGCCTCAGCCTCCCGGGTAGATGGGATTACAGGCGCCCACCACCACACTCGGCTAATGTTTGTATTTTTAG
TAGAGATGGGGTTTCACCATGTTGGCCAGGCTGGTCTCAAACTCCTGACCTCAGAGGATCCACCTGCCTCAGCCTCC
CAAAGTGCTGGGATTACAGATGTAGGCCACTGCGCCCGGCCAAGTATTGCTCTTATACATTAAAAAACAGGTGTGAG
CCACTGCGCCCAGCCAGGTATTGCTCTTATACATTAAAAAATAGGCCGGTGCAGTGGCTCACGCCTGTAATCCCAGC
ACTTTGGGAAGCCAAGGCGGGCAGAACACCCGAGGTCAGGAGTCCAAGGCCAGCCTGGCCAAGATGGTGAAACCCCG
TCTCTATTAAAAATACAAACATTACCTGGGCATGATGGTGGGCGCCTGTAATCCCAGCTACTCAGGAGGCTGAGGCA
GGAGGATCCGCGGAGCCTGGCAGATCTGCCTGAGCCTGGGAGGTTGAGGCTACAGTAAGCCAAGATCATGCCAGTAT
ACTTCAGCCTGGGCGACAAAGTGAGACCGTAACAAAAAAAAAAAAAATTTAAAAAAAGAAATTTAGATCAAGATCCAA
CTGTAAAAAGTGGCCTAAACACCACATTAAAGAGTTTGGAGTTTATTCTGCAGGCAGAAGAGAACCATCAGGGGGTC
TTCAGCATGGGAATGGCATGGTGCACCTGGTTTTTGTGAGATCATGGTGGTGACAGTGTGGGGAATGTTATTTTGGA
GGGACTGGAGGCAGACAGACCGGTTAAAAGGCCAGCACAACAGATAAGGAGGAAGAAGATGAGGGCTTGGACCGAAG

CAGAGAAGAGCAAACAGGGAAGGTACAAATTCAAGAAATATTGGGGGGTTTGAATCAACACATTTAGATGATTAATT
AAATATGAGGACTGAGGAATAAGAAATGAGTCAAGGATGGTTCCAGGCTGCTAGGCTGCTTACCTGAGGTGGCAAAG
TCGGGAGGAGTGGCAGTTTAGGACAGGGGGCAGTTGAGGAATATTGTTTTGATCATTTTGAGTTTGAGGTACAAGTT
GGACACTTAGGTAAAGACTGGAGGGGAAATCTGAATATACAATTATGGGACTGAGGAACAAGTTTATTTTATTTTTT
GTTTCGTTTTCTTGTTGAAGAACAAATTTAATTGTAATCCCAAGTCATCAGCATCTAGAAGACAGTGGCAGGAGGTG
ACTGTCTTGTGGGTAAGGGTTTGGGGTCCTTGATGAGTATCTCTCAATTGGCCTTAAATATAAGCAGGAAAAGGAGT
TTATGATGGATTCCAGGCTCAGCAGGGCTCAGGAGGGCTCAGGCAGCCAGCAGAGGAAGTCAGAGCATCTTCTTTGG
TTTAGCCCAAGTAATGACTTCCTTAAAAAGCTGAAGGAAAATCCAGAGTGACCAGATTATAAACTGTACTCTTGCAT
TTTCTCTCCCTCCTCTCACCCACAGCCTCTTGATGAACCGGAGGAAGTTTCTTTACCAATTCAAAAATGTCCGCTGG
GCTAAGGGTCGGCGTGAGACCTACCTGTGCTACGTAGTGAAGAGGCGTGACAGTGCTACATCCTTTTCACTGGACTT
TGGTTATCTTCGCAATAAGGTATCAATTAAAGTCGGCTTTGCAAGCAGTTTAATGGTCAACTGTGAGTGCTTTTAGA
GCCACCTGCTGATGGTATTACTTCCATCCTTTTTTGGCATTTGTGTCTCTATCACATTCCTCAAATCCTTTTTTTTA
TTTCTTTTTCCATGTCCATGCACCCATATTAGACATGGCCCAAAATATGTGATTTAATTCCTCCCCAGTAATGCTGG
GCACCCTAATACCACTCCTTCCTTCAGTGCCAAGAACAACTGCTCCCAAACTGTTTACCAGCTTTCCTCAGCATCTG
AATTGCCTTTGAGATTAATTAAGCTAAAAGCATTTTTATATGGGAGAATATTATCAGCTTGTCCAAGCAAAAATTTT
AAATGTGAAAAACAAATTGTGTCTTAAGCATTTTTGAAAATTAAGGAAGAAGAATTTGGGAAAAAATTAACGGTGGC
TCAATTCTGTCTTCCAAATGATTTCTTTTCCCTCCTACTCACATGGGTCGTAGGCCAGTGAATACATTCAACATGGT
GATCCCCAGAAAACTCAGAGAAGCCTCGGCTGATGATTAATTAAATTGATCTTTCGGCTACCCGAGAGAATTACATT
TCCAAGAGACTTCTTCACCAAAATCCAGATGGGTTTACATAAACTTCTGCCCACGGGTATCTCCTCTCTCCTAACAC
GCTGTGACGTCTGGGCTTGGTGGAATCTCAGGGAAGCATCCGTGGGGTGGAAGGTCATCGTCTGGCTCGTTGTTTGA
TGGTTATATTACCATGCAATTTTCTTTGCCTACATTTGTATTGAATACATCCCAATCTCCTTCCTATTCGGTGACAT
GACACATTCTATTTCAGAAGGCTTTGATTTTATCAAGCACTTTCATTTACTTCTCATGGCAGTGCCTATTACTTCTC
TTACAATACCCATCTGTCTGCTTTACCAAAATCTATTTCCCCTTTTCAGATCCTCCCAAATGGTCCTCATAAACTGT
CCTGCCTCCACCTAGTGGTCCAGGTATATTTCCACAATGTTACATCAACAGGCACTTCTAGCCATTTTCCTTCTCAA
AAGGTGCAAAAAGCAACTTCATAAACACAAATTAAATCTTCGGTGAGGTAGTGTGATGCTGCTTCCTCCCAACTCAG
CGCACTTCGTCTTCCTCATTCCACAAAAACCCATAGCCTTCCTTCACTCTGCAGGACTAGTGCTGCCAAGGGTTCAG
CTCTACCTACTGGTGTGCTCTTTTGAGCAAGTTGCTTAGCCTCTCTGTAACACAAGGACAATAGCTGCAAGCATCCC
CAAAGATCATTGCAGGAGACAATGACTAAGGCTACCAGAGCCGCAATAAAAGTCAGTGAATTTTAGCGTGGTCCTCT
CTGTCTCTCCAGAACGGCTGCCACGTGGAATTGCTCTTCCTCCGCTACATCTCGGACTGGGACCTAGACCCTGGCCG
CTGCTACCGCGTCACCTGGTTCACCTCCTGGAGCCCCTGCTACGACTGTGCCCGACATGTGGCCGACTTTCTGCGAG
GGAACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCCTCTACTTCTGTGAGGACCGCAAGGCTGAGCCCGAGGGG
CTGCGGCGGCTGCACCGCGCCGGGGTGCAAATAGCCATCATGACCTTCAAAGGTGCGAAAGGGCCTTCCGCGCAGGC
GCAGTGCAGCAGCCCGCATTCGGGATTGCGATGCGGAATGAATGAGTTAGTGGGGAAGCTCGAGGGGAAGAAGTGGG
CGGGGATTCTGGTTCACCTCTGGAGCCGAAATTAAAGATTAGAAGCAGAGAAAAGAGTGAATGGCTCAGAGACAAGG
CCCCGAGGAAATGAGAAAATGGGGCCAGGGTTGCTTCTTTCCCCTCGATTTGGAACCTGAACTGTCTTCTACCCCCA
TATCCCCGCCTTTTTTTCCTTTTTTTTTTTTTGAAGATTATTTTTACTGCTGGAATACTTTTGTAGAAAACCACGAA
AGAACTTTCAAAGCCTGGGAAGGGCTGCATGAAAATTCAGTTCGTCTCTCCAGACAGCTTCGGCGCATCCTTTTGGT
AAGGGGCTTCCTCGCTTTTTAAATTTTCTTTCTTTCTCTACAGTCTTTTTTGGAGTTTCGTATATTTCTTATATTTT
CTTATTGTTCAATCACTCTCAGTTTTCATCTGATGAAAACTTTATTTCTCCTCCACATCAGCTTTTTCTTCTGCTGT

TTCACCATTCAGAGCCCTCTGCTAAGGTTCCTTTTCCCTCCCTTTTCTTTCTTTTGTTGTTTCACATCTTTAAATTT
CTGTCTCTCCCCAGGGTTGCGTTTCCTTCCTGGTCAGAATTCTTTTCTCCTTTTTTTTTTTTTTTTTTTTTTTTTT
AAACAAACAAACAAAAAACCCAAAAAAACTCTTTCCCAATTTACTTTCTTCCAACATGTTACAAAGCCATCCACTCA
GTTTAGAAGACTCTCCGGCCCCACCGACCCCCAACCTCGTTTTGAAGCCATTCACTCAATTTGCTTCTCTCTTTCTC
TACAGCCCCTGTATGAGGTTGATGACTTACGAGACGCATTTCGTACTTTGGGACTTTGATAGCAACTTCCAGGAATG
TCACACACGATGAAATATCTCTGCTGAAGACAGTGGATAAAAAACAGTCCTTCAAGTCTTCTCTGTTTTTATTCTTC
AACTCTCACTTTCTTAGAGTTTACAGAAAAAATATTTATATACGACTCTTTAAAAAGATCTATGTCTTGAAAATAGA
GAAGGAACACAGGTCTGGCCAGGGACGTGCTGCAATTGGTGCAGTTTTGAATGCAACATTGTCCCCTACTGGGAATA
ACAGAACTGCAGGACCTGGGAGCATCCTAAAGTGTCAACGTTTTTCTATGACTTTTAGGTAGGATGAGAGCAGAAGG
TAGATCCTAAAAAGCATGGTGAGAGGATCAAATGTTTTTATATCAACATCCTTTATTATTTGATTCATTTGAGTTAA
CAGTGGTGTTAGTGATAGATTTTTCTATTCTTTTCCCTTGACGTTTACTTTCAAGTAACACAAACTCTTCCATCAGG
CCATGATCTATAGGACCTCCTAATGAGAGTATCTGGGTGATTGTGACCCCAAACCATCTCTCCAAAGCATTAATATC
CAATCATGCGCTGTATGTTTTAATCAGCAGAAGCATGTTTTTATGTTTGTACAAAAGAAGATTGTTATGGGTGGGGA
TGGAGGTATAGACCATGCATGGTCACCTTCAAGCTACTTTAATAAAGGATCTTAAAATGGGCAGGAGGACTGTGAAC
AAGACACCCTAATAATGGGTTGATGTCTGAAGTAGCAAATCTTCTGGAAACGCAAACTCTTTTAAGGAAGTCCCTAA
TTTAGAAACACCCACAAACTTCACATATCATAATTAGCAAACAATTGGAAGGAAGTTGCTTGAATGTTGGGGAGAGG
AAAATCTATTGGCTCTCGTGGGTCTCTTCATCTCAGAAATGCCAATCAGGTCAAGGTTTGCTACATTTTGTATGTGT
GTGATGCTTCTCCCAAAGGTATATTAACTATATAAGAGAGTTGTGACAAAACAGAATGATAAAGCTGCGAACCGTGG
CACACGCTCATAGTTCTAGCTGCTTGGGAGGTTGAGGAGGGAGGATGGCTTGAACACAGGTGTTCAAGGCCAGCCTG
GGCAACATAACAAGATCCTGTCTCTCAAAAAAAAAAAAAAAAAAAAGAAAGAGAGAGGGCCGGGCGTGGTGGCTCAC
GCCTGTAATCCCAGCACTTTGGGAGGCCGAGCCGGGCGGATCACCTGTGGTCAGGAGTTTGAGACCAGCCTGGCCAA
CATGGCAAAACCCCGTCTGTACTCAAAATGCAAAAATTAGCCAGGCGTGGTAGCAGGCACCTGTAATCCCAGCTACT
TGGGAGGCTGAGGCAGGAGAATCGCTTGAACCCAGGAGGTGGAGGTTGCAGTAAGCTGAGATCGTGCCGTTGCACTC
CAGCCTGGGCGACAAGAGCAAGACTCTGTCTCAGAAAAAAAAAAAAAAAAAGAGAGAGAGAGAGAGAAAGAGAACAATAT
TTGGGAGAGAAGGATGGGGAAGCATTGCAAGGAAATTGTGCTTTATCCAACAAAATGTAAGGAGCCAATAAGGGATC
CCTATTTGTCTCTTTTGGTGTCTATTTGTCCCTAACAACTGTCTTTGACAGTGAGAAAAATATTCAGAATAACCATA
TCCCTGTGCCGTTATTACCTAGCAACCCTTGCAATGAAGATGAGCAGATCCACAGGAAAACTTGAATGCACAACTGT
CTTATTTTAATCTTATTGTACATAAGTTTGTAAAAGAGTTAAAAATTGTTACTTCATGTATTCATTTATATTTTATA
TTATTTTGCGTCTAATGATTTTTTATTAACATGATTTCCTTTTCTGATATATTGAAATGGAGTCTCAAAGCTTCATA
AATTTATAACTTTAGAAATGATTCTAATAACAACGTATGTAATTGTAACATTGCAGTAATGGTGCTACGAAGCCATT
TCTCTTGATTTTTAGTAAACTTTTATGACAGCAAATTTGCTTCTGGCTCACTTTCAATCAGTTAAATAAATGATAAA
TAATTTTGGAAGCTGTGAAGATAAAATACCAAATAAAATAATATAAAAGTGATTTATATGAAGTTAAAATAAAAAAT
CAGTATGATGGAATAAACTTG

[0173] 载脂蛋白B mRNA编辑酶,催化多肽-样(APOBEC)是进化上保守的胞苷脱氨酶的家族。这个家族的成员是C-至-U编辑酶。APOBEC样蛋白的N末端结构域是催化结构域,而C末端结构域是假催化结构域。更具体地,该催化结构域是锌依赖性胞苷脱氨酶结构域,且对于胞苷脱氨反应是重要的。APOBEC家族成员包含APOBEC1,APOBEC2,APOBEC3A,APOBEC3B,APOBEC3C,APOBEC3D("APOBEC3E"现在是指这个),APOBEC3F,APOBEC3G,APOBEC3H,APOBEC4,以及活化-诱导的(胞苷)脱氨酶。一些经修饰的胞苷脱氨酶是市售的,包含但不限于:

SaBE3,SaKKH-BE3,VQR-BE3,EQR-BE3,VRER-BE3,YE1-BE3,EE-BE3,YE2-BE3,和YEE-BE3,其可从Addgene获得(质粒85169,85170,85171,85172,85173,85174,85175,85176,85177)。

[0174] 下面提供了根据本公开揭露的各方面可以与Cas9融合的其他示例性脱氨酶。应当清楚,在一些实施方式中,可以使用各个序列的活性结构域,例如,没有定位信号(核定位序列,没有核输出信号,细胞质定位信号)的结构域。

[0175] 人类AID:

[0176] MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLYFCEDRKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHENSVRLSRQLRRILLPLYEVDDLRDAFRTLGL(单底线:核定位序列;双底线:核输出信号)

[0177] 小鼠AID:

[0178] MDSLLMKQKKFLYHFKNVRWAKGRHETYLCYVVKRRDSATSCSLDFGHLRNKSGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVAEFLRWNPNLSLRIFTARLYFCEDRKAEPEGLRRLHRAGVQIGIMTFKDYFYCWNTFVENRERTFKAWEGLHENSVRLTRQLRRILLPLYEVDDLRDAFRMLGF

[0179] (单底线:核定位序列;双底线:核输出信号)

[0180] 犬AID:

[0181] MDSLLMKQRKFLYHFKNVRWAKGRHETYLCYVVKRRDSATSFSLDFGHLRNKSGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGYPNLSLRIFAARLYFCEDRKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENREKTFKAWEGLHENSVRLSRQLRRILLPLYEVDDLRDAFRTLGL(单底线:核定位序列;双底线:核输出信号)

[0182] 牛AID:

[0183] MDSLLKKQRQFLYQFKNVRWAKGRHETYLCYVVKRRDSPTSFSLDFGHLRNKAGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGYPNLSLRIFTARLYFCDKERKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHENSVRLSRQLRRILLPLYEVDDLRDAFRTLGL(单底线:核定位序列;双底线:核输出信号)

[0184] 大鼠AID

[0185] MAVGSKPKAALVGPHWERERIWCFLCSTGLGTQQTGQTSRWLRPAATQDPVSPPRSLLMKQRKFLYHFKNVRWAKGRHETYLCYVVKRRDSATSFSLDFGYLRNKSGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLTGWGALPAGLMSPARPSDYFYCWNTFVENHERTFKAWEGLHENSVRLSRRLRRILLPLYEVD

[0186] DLRDAFRTLGL

[0187] (单底线:核定位序列;双底线:核输出信号)

[0188] 小鼠APOBEC-3

[0189] MGPFCLGCSHRKCYSPIRNLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVSLHHGVFKNKDNIHAEICFLYWFHDKVLKVLSPREEFKITWYMSWSPCFECAEQIVRFLATHHNLSLDIFSSRLYNVQDPETQQNLCRLVQEGAQVAAMDLYEFKKCWKKFVDNGGRRFRPWKRLLTNFRYQDSKLQEILRPCYIPVPSSSSSTLSNICLTKGLPETRFCVEGRRMDPLSEEEFYSQFYNQRVKHLCYYHRMKPYLCYQLEQFNGQAPLKGCLLSEKGKQHAEILFLDKIRSMELSQVTITCYLTWSPCPNCAWQLAAFKRDRPDLILHIYTSRLYFHWKRPFQKGLCSLWQSGILVDVMDLPQF

TDCWTNFVNPKRPFWPWKGLEIISRRTQRRLRRIKESWGLQDLVNDFGNLQLGPPMS(斜体:核酸编辑结构域)

[0190] 大鼠APOBEC-3:

[0191] MGPFCLGCSHRKCYSPIRNLISQETFKFHFKNRLRYAIDRKDTFLCYEVTRKDCDSPVSLHHGVFKNKDNIHAEICFLYWFHDKVLKVLSPREEFKITWYMSWSPCFECAEQVLRFLATHHNLSLDIFSSRLYNIRDPENQQNLCRLVQEGAQVAAMDLYEFKKCWKKFVDNGGRRFRPWKKLLTNFRYQDSKLQEILRPCYIPVPSSSSSTLSNICLTKGLPETRFCVERRRVHLLSEEEFYSQFYNQRVKHLCYYHGVKPYLCYQLEQFNGQAPLKGCLLSEKGKQHAEILFLDKIRSMELSQVIITCYLTWSPCPNCAWQLAAFKRDRPDLILHIYTSRLYFHWKRPFQKGLCSLWQSGILVDVMDLPQFTDCWTNFVNPKRPFWPWKGLEIISRRTQRRLHRIKESWGLQDLVNDFGNLQLGPPMS(斜体:核酸编辑结构域)

[0192] 恒河猴(Rhesus macaque)APOBEC-3G:

[0193] MVEPMDPRTFVSNFNNRPILSGLNTVWLCCEVKTKDPSGPPLDAKIFQGKVYSKAKYHPEMRFLRWFHKWRQLHHDQEYKVTWYVSWSPCTRCANSVATFLAKDPKVTLTIFVARLYYFWKPDYQQALRILCQKRGGPHATMKIMNYNEFQDCWNKFVDGRGKPFKPRNNLPKHYTLLQATLGELLRHLMDPGTFTSNFNNKPWVSGQHETYLCYKVERLHNDTWVPLNQHRGFLRNQAPNIHGFPKGRHAELCFLDLIPFWKLDGQQYRVTCFTSWSPCFSCAQEMAKFISNNEHVSLCIFAARIYDDQGRYQEGLRALHRDGAKIAMMNYSEFEYCWDTFVDRQGRPFQPWDGLDEHSQALSGRLRAI(斜体:核酸编辑结构域;单底线:细胞质定位信号)

[0194] 黑猩猩APOBEC-3G:

[0195] MKPHFRNPVERMYQDTFSDNFYNRPILSHRNTVWLCYEVKTKGPSRPPLDAKIFRGQVYSKLKYHPEMRFFHWFSKWRKLHRDQEYEVTWYISWSPCTKCTRDVATFLAEDPKVTLTIFVARLYYFWDPDYQEALRSLCQKRDGPRATMKIMNYDEFQHCWSKFVYSQRELFEPWNNLPKYYILLHIMLGEILRHSMDPPTFTSNFNNELWVRGRHETYLCYEVERLHNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDVIPFWKLDLHQDYRVTCFTSWSPCFSCAQEMAKFISNNKHVSLCIFAARIYDDQGRCQEGLRTLAKAGAKISIMTYSEFKHCWDTFVDHQGCPFQPWDGLEEHSQALSGRLRAILQNQGN

[0196] (斜体:核酸编辑结构域;单底线:细胞质定位信号)

[0197] 绿猴APOBEC-3G:

[0198] MNPQIRNMVEQMEPDIFVYYFNNRPILSGRNTVWLCYEVKTKDPSGPPLDANIFQGKLYPEAKDHPEMKFLHWFRKWRQLHRDQEYEVTWYVSWSPCTRCANSVATFLAEDPKVTLTIFVARLYYFWKPDYQQALRILCQERGGPHATMKIMNYNEFQHCWNEFVDGQGKPFKPRKNLPKHYTLLHATLGELLRHVMDPGTFTSNFNNKPWVSGQRETYLCYKVERSHNDTWVLLNQHRGFLRNQAPDRHGFPKGRHAELCFLDLIPFWKLDDQQYRVTCFTSWSPCFSCAQKMAKFISNNKHVSLCIFAARIYDDQGRCQEGLRTLHRDGAKIAVMNYSEFEYCWDTFVDRQGRPFQPWDGLDEHSQALSGRLRAI

[0199] (斜体:核酸编辑结构域;单底线:细胞质定位信号)

[0200] 人类APOBEC-3G:

[0201] MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPPLDAKIFRGQVYSELKYHPEMRFFHWFSKWRKLHRDQEYEVTWYISWSPCTKCTRDMATFLAEDPKVTLTIFVARLYYFWDPDYQEALRSLCQKRDGPRATMKIMNYDEFQHCWSKFVYSQRELFEPWNNLPKYYILLHIMLGEILRHSMDPPTFTFNFNNEPWVRGRHETYLCYEVERMHNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDVIPFWKLDLDQDYRVTCFTSWSPCFSCAQEMAKFISKNKHVSLCIFTARIYDDQGRCQEGLRTLAEAGAKISIMTYSEFKHCWDTFVDHQGCPFQPWDGLDEHSQDLSGRLRAILQNQEN

[0202] (斜体:核酸编辑结构域;单底线:细胞质定位信号)

[0203] 人类APOBEC-3F:

[0204] MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPRLDAKIFRGQVYSQPEHHAEMC FLSWFCGNQLPAYKCFQITWFVSWTPCPDCVAKLAEFLAEHPNVTLTISAARLYYYWERDYRRALCRLSQAGARVKI MDDEEFAYCWENFVYSEGQPFMPWYKFDDNYAFLHRTLKEILRNPMEAMYPHIFYFHFKNLRKAYGRNESWLCFTME VVKHHSPVSWKRGVFRNQVDPETHCHAERCFLSWFCDDILSPNTNYEVTWYTSWSPCPECAGEVAEFLARHSNVNLT IFTARLYYFWDTDYQEGLRSLSQEGASVEIMGYKDFKYCWENFVYNDDEPFKPWKGLKYNFLFLDSKLQEILE

[0205] (斜体:核酸编辑结构域)

[0206] 人类APOBEC-3B:

[0207] MNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEVKIKRGRSNLLWDTGVFRGQVYFKPQYHAEM CFLSWFCGNQLPAYKCFQITWFVSWTPCPDCVAKLAEFLSEHPNVTLTISAARLYYYWERDYRRALCRLSQAGARVT IMDYEEFAYCWENFVYNEGQQFMPWYKFDENYAFLHRTLKEILRYLMDPDTFTFNFNN

[0208] DPLVLRRRQTYLCYEVERLDNGTWVLMDQHMGFLCNEAKNLLCGFYGRHAELRFLDLVPSLQLDPAQIY RVTWFISWSPCFSWGCAGEVRAFLQENTHVRLRIFAARIYDYDPLYKEALQMLRDAGAQVSIMTYDEFEYCWDTFVY RQGCPFQPWDGLEEHSQALSGRLRAILQNQGN

[0209] (斜体:核酸编辑结构域)

[0210] 大鼠APOBEC-3B:

[0211] MQPQGLGPNAGMGPVCLGCSHRRPYSPIRNPLKKLYQQTFYFHFKNVRYAWGRKNNFLCYEVNGMDCAL PVPLRQGVFRKQGHIHAELCFIYWFHDKVLRVLSPMEEFKVTWYMSWSPCSKCAEQVARFLAAHRNLSLAIFSSRLY YYLRNPNYQQKLCRLIQEGVHVAAMDLPEFKKCWNKFVDNDGQPFRPWMRLRINFSFYDCKLQEIFSRMNLLREDVF YLQFNNSHRVKPVQNRYYRRKSYLCYQLERANGQEPLKGYLLYKKGEQHVEILFLEKMRSMELSQVRITCYLTWSPC PNCARQLAAFKKDHPDLILRIYTSRLYFWRKKFQKGLCTLWRSGIHVDVMDLPQFADCWTNFVNPQRPFRPWNELEK NSWRIQRRLRRIKESWGL

[0212] 牛APOBEC-3B:

[0213] DGWEVAFRSGTVLKAGVLGVSMTEGWAGSGHPGQGACVWTPGTRNTMNLLREVLFKQQFGNQPRVPAPY YRRKTYLCYQLKQRNDLTLDRGCFRNKKQRHAERFIDKINSLDLNPSQSYKIICYITWSPCPNCANELVNFITRNNH LKLEIFASRLYFHWIKSFKMGLQDLQNAGISVAVMTHTEFEDCWEQFVDNQSRPFQPWDKLEQYSASIRRRLQRILT API

[0214] 黑猩猩APOBEC-3B:

[0215] MNPQIRNPMEWMYQRTFYYNFENEPILYGRSYTWLCYEVKIRRGHSNLLWDTGVFRGQMYSQPEHHAEM CFLSWFCGNQLSAYKCFQITWFVSWTPCPDCVAKLAKFLAEHPNVTLTISAARLYYYWERDYRRALCRLSQAGARVK IMDDEEFAYCWENFVYNEGQPFMPWYKFDDNYAFLHRTLKEIIRHLMDPDTFTFNFNNDPLVLRRHQTYLCYEVERL DNGTWVLMDQHMGFLCNEAKNLLCGFYGRHAELRFLDLVPSLQLDPAQIYRVTWFISWSPCFSWGCAGQVRAFLQEN THVRLRIFAARIYDYDPLYKEALQMLRDAGAQVSIMTYDEFEYCWDTFVYRQGCPFQPWDGLEEHSQALSGRLRAIL QVRASSLCMVPHRPPPPPQSPGPCLPLCSEPPLGSLLPTGRPAPSLPFLLTASFSFPPPASLPPLPSLSLSPGHLPV PSFHSLTSCSIQPPCSSRIRETEGWASVSKEGRDLG

[0216] 人类APOBEC-3C:

[0217] MNPQIRNPMKAMYPGTFYFQFKNLWEANDRNETWLCFTVEGIKRRSVVSWKTGVFRNQVDSETHCHAER CFLSWFCDDILSPNTKYQVTWYTSWSPCPDCAGEVAEFLARHSNVNLTIFTARLYYFQYPCYQEGLRSLSQEGVAVE

IMDYEDFKYCWENFVYNDNEPFKPWKGLKTNFRLLKRRLRESLQ

[0218] 大猩猩APOBEC3C

[0219] MNPQIRNPMKAMYPGTFYFQFKNLWEANDRNETWLCFTVEGIKRRSVVSWKTGVFRNQVDSETHCHAERCFLSWECDDILSPNTNYQVTWYTSWSPCPECAGEVAEFLARHSNVNLTIFTARLYYFQDTDYQEGLRSLSQEGVAVKIMDYKDFKYCWENFVYNDDEPFKPWKGLKYNFRFLKRRLQEILE

[0220] (斜体:核酸编辑结构域)

[0221] 人类APOBEC-3A:

[0222] MEASPASGPRHLMDPHIFTSNFNNGIGRHKTYLCYEVERLDNGTSVKMDQHRGFLHNQAKNLLCGFYGRHAELRFLDLVPSLQLDPAQIYRVTWFISWSPCFSWGCAGEVRAFLQENTHVRLRIFAARIYDYDPLYKEALQMLRDAGAQVSIMTYDEFKHCWDTFVDHQGCPFQPWDGLDEHSQALSGRLRAILQNQGN

[0223] (斜体:核酸编辑结构域)

[0224] 恒河猴APOBEC-3A:

[0225] MDGSPASRPRHLMDPNTFTFNFNNDLSVRGRHQTYLCYEVERLDNGTWVPMDERRGFLCNKAKNVPCGDYGCHVELRFLCEVPSWQLDPAQTYRVTWFISWSPCFRRGCAGQVRVFLQENKHVRLRIFAARIYDYDPLYQEALRTLRDAGAQVSIMTYEEFKHCWDTFVDRQGRPFQPWDGLDEHSQALSGRLRAILQNQGN

[0226] (斜体:核酸编辑结构域)

[0227] 牛APOBEC-3A:

[0228] MDEYTFTENFNNQGWPSKTYLCYEMERLDGDATIPLDEYKGFVRNKGLDQPEKPCHAELYFLGKIHSWNLDRNQHYRLTCFISWSPCYDCAQKLTTFLKENHHISLHILASRIYTHNRFGCHQSGLCELQAAGARITIMTFEDFKHCWETFVDHKGKPFQPWEGLNVKSQALCTELQAILKTQQN

[0229] (斜体:核酸编辑结构域)

[0230] 人类APOBEC-3H:

[0231] MALLTAETFRLQFNNKRRLRRPYYPRKALLCYQLTPQNGSTPTRGYFENKKKCHAEICFINEIKSMGLDETQCYQVTCYLTWSPCSSCAWELVDFIKAHDHLNLGIFASRLYYHWCKPQQKGLRLLCGSQVPVEVMGFPKFADCWENFVDHEKPLSFNPYKMLEELDKNSRAIKRRLERIKIPGVRAQGRYMDILCDAEV

[0232] (斜体:核酸编辑结构域)

[0233] 恒河猴APOBEC-3H:

[0234] MALLTAKTFSLQFNNKRRVNKPYYPRKALLCYQLTPQNGSTPTRGHLKNKKKDHAEIRFINKIKSMGLDETQCYQVTCYLTWSPCPSCAGELVDFIKAHRHLNLRIFASRLYYHWRPNYQEGLLLLCGSQVPVEVMGLPEFTDCWENFVDHKEPPSFNPSEKLEELDKNSQAIKRRLERIKSRSVDVLENGLRSLQLGPVTPSSSIRNSR

[0235] 人类APOBEC-3D:

[0236] MNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEVKIKRGRSNLLWDTGVFRGPVLPKRQSNHRQEVYFRFENHAEMCFLSWFCGNRLPANRRFQITWFVSWNPCLPCVVKVTKFLAEHPNVTLTISAARLYYYRDRDWRWVLLRLHKAGARVKIMDYEDFAYCWENFVCNEGQPFMPWYKFDDNYASLHRTLKEILRNPMEAMYPHIFYFHFKNLLKACGRNESWLCFTMEVTKHHSAVFRKRGVFRNQVDPETHCHAERCFLSWFCDDILSPNTNYEVTWYTSWSPCPECAGEVAEFLARHSNVNLTIFTARLCYFWDTDYQEGLCSLSQEGASVKIMGYKDFVSCWKNFVYSDDEPFKPWKGLQTNFRLLKRRLREILQ

[0237] (斜体:核酸编辑结构域)

[0238] 人类APOBEC-1:

[0239] MTSEKGPSTGDPTLRRRIEPWEFDVFYDPRELRKEACLLYEIKWGMSRKIWRSSGKNTTNHVEVNFIK KFTSERDFHPSMSCSITWFLSWSPCWECSQAIREFLSRHPGVTLVIYVARLFWHMDQQNRQGLRDLVNSGVTIQIM RASEYYHCWRNFVNYPPGDEAHWPQYPPLWMMLYALELHCIILSLPPCLKISRRWQNHLTFFRLHLQNCHYQTIPP HILLATGLIHPSVAWR小鼠APOBEC-1:

[0240] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSVWRHTSQNTSNHVEVNFLEK FTTERYFRPNTRCSITWFLSWSPCGECSRAITEFLSRHPYVTLFIYIARLYHHTDQRNRQGLRDLISSGVTIQIMTE QEYCYCWRNFVNYPPSNEAYWPRYPHLWVKLYVLELYCIILGLPPCLKILRRKQPQLTFFTITLQTCHYQRIPPHLL WATGLK

[0241] 大鼠APOBEC-1:

[0242] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEK FTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTE QESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHIL WATGLK

[0243] 人类APOBEC-2:

[0244] MAQKEEAAVATEAASQNGEDLENLDDPEKLKELIELPPFEIVTGERLPANFFKFQFRNVEYSSGRNKTF LCYVVEAQGKGGQVQASRGYLEDEHAAAHAEEAFFNTILPAFDPALRYNVTWYVSSSPCAACADRIIKTLSKTKNLR LLILVGRLFMWEEPEIQAALKKLKEAGCKLRIMKPQDFEYVWQNFVEQEEGESKAFQPWEDIQENFLYYEEKLADIL K

[0245] 小鼠APOBEC-2:

[0246] MAQKEEAAEAAAPASQNGDDLENLEDPEKLKELIDLPPFEIVTGVRLPVNFFKFQFRNVEYSSGRNKTF LCYVVEVQSKGGQAQATQGYLEDEHAGAHAEEAFFNTILPAFDPALKYNVTWYVSSSPCAACADRILKTLSKTKNLR LLILVSRLFMWEEPEVQAALKKLKEAGCKLRIMKPQDFEYIWQNFVEQEEGESKAFEPWEDIQENFLYYEEKLADIL K

[0247] 大鼠APOBEC-2:

[0248] MAQKEEAAEAAAPASQNGDDLENLEDPEKLKELIDLPPFEIVTGVRLPVNFFKFQFRNVEYSSGRNKTF LCYVVEAQSKGGQVQATQGYLEDEHAGAHAEEAFFNTILPAFDPALKYNVTWYVSSSPCAACADRILKTLSKTKNLR LLILVSRLFMWEEPEVQAALKKLKEAGCKLRIMKPQDFEYLWQNFVEQEEGESKAFEPWEDIQENFLYYEEKLADIL K

[0249] 牛APOBEC-2:

[0250] MAQKEEAAAAAEPASQNGEEVENLEDPEKLKELIELPPFEIVTGERLPAHYFKFQFRNVEYSSGRNKTF LCYVVEAQSKGGQVQASRGYLEDEHATNHAEEAFFNSIMPTFDPALRYMVTWYVSSSPCAACADRIVKTLNKTKNLR LLILVGRLFMWEEPEIQAALRKLKEAGCRLRIMKPQDFEYIWQNFVEQEEGESKAFEPWEDIQENFLYYEEKLADIL K

[0251] 海七鳃鳗CDA1(pmCDA1)

[0252] MTDAEYVRIHEKLDIYTFKKQFFNNKKSVSHRCYVLFELKRRGERRACFWGYAVNKPQSGTERGIHAEI FSIRKVEEYLRDNPGQFTINWYSSWSPCADCAEKILEWYNQELRGNGHTLKIWACKLYYEKNARNQIGLWNLRDNGV GLNVMVSEHYQCCRKIFIQSSHNQLNENRWLEKTLKRAEKRRSELSFMIQVKILHTTKSPAV

[0253] 人类APOBEC3G D316R D317R

[0254] MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPPLDAKIFRGQVYSELKYHPEMR

FFHWFSKWRKLHRDQEYEVTWYISWSPCTKCTRDMATFLAEDPKVTLTIFVARLYYFWDPDYQEALRSLCQKRDGPR ATMKFNYDEFQHCWSKFVYSQRELFEPWNNLPKYYILLHFMLGEILRHSMDPPTFTFNFNNEPWVRGRHETYLCYEV ERMHNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDVIPFWKLDLDQDYRVTCFTSWSPCFSCAQEMAKFISK KHVSLCIFTARIYRRQGRCQEGLRTLAEAGAKISFTYSEFKHCWDTFVDHQGCPFQPWDGLDEHSQDLSGRLRAILQ NQEN

[0255] 人类APOBEC3G链A

[0256] MDPPTFTFNFNNEPWWGRHETYLCYEVERMHNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDVI PFWKLDLDQDYRVTCFTSWSPCFSCAQEMAKFISKNKHVSLCIFTARIYDDQGRCQEGLRTLAEAGAKISFTYSEFK HCWDTFVDHQGCPFQPWDGLDEHSQDLSGRLRAILQ

[0257] 人类APOBEC3G链A D120R D121R

[0258] MDPPTFTFNFNNEPWVRGRHETYLCYEVERMHNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDV IPFWKLDLDQDYRVTCFTSWSPCFSCAQEMAKFISKNKHVSLCIFTARIYRRQGRCQEGLRTLAEAGAKISFMTYSE FKHCWDTFVDHQGCPFQPWDGLDEHSQDLSGRLRAILQ

[0259] 术语“脱氨酶”或“脱氨酶结构域”是指催化脱氨反应的蛋白质或其片段。

[0260] “检测”是指鉴别待检测的分析物的存在、不存在或其数量。在一个实施方式中,多核苷酸或多肽中的序列改变被检测到。在另一个实施方式中,插入/缺失的存在被检测到。

[0261] “可检测标记”意为当链接到目标分子上时,使得该目标分子可经由分光的、光化学的,生物化学的,免疫化学的,或化学的手段而被检测到的组合物。例如,有用的标记包含放射性同位素、磁珠、金属珠、胶体颗粒、荧光染料、电子-致密试剂、酶(例如,像ELISA(酶联免疫吸附测定)中常用的)、生物素、地高辛或半抗原。

[0262] “片段”意为多肽或核酸分子的一部分。这部分含有,该参考核酸分子或多肽的全长的至少约10%,20%,30%,40%,50%,60%,70%,80%,或90%。片段可能含有10,20,30,40,50,60,70,80,90,或100,200,300,400,500,600,700,800,900,或1000个核苷酸或氨基酸。

[0263] “杂交(Hybridization)”意为互补核碱基之间的氢键键结,其可以是沃森-克里克,Hoogsteen或反向Hoogsteen氢键键结。例如,腺嘌呤和胸腺嘧啶是通过氢键的形成而配对的互补核碱基。

[0264] 术语“碱基修复的抑制剂”或“IBR”是指能够抑制核酸修复酶(例如碱基切除修复酶)的活性的蛋白。在一些实施方式中,该IBR是肌苷碱基切除修复的抑制剂。示例性的碱基修复的抑制剂包含APE1,Endo III,Endo IV,Endo V,Endo VIII,Fpg,hOGG1,hNEIL1,T7 Endo1,T4PDG,UDG,hSMUG1,和hAAG的抑制剂。在一些实施方式中,该IBR是Endo V或hAAG的抑制剂。在一些实施方式中,该IBR是催化无活性的EndoV或催化无活性的hAAG。

[0265] 术语“分离的”,“纯化的”,或“生物学纯的”是指材料和天然状态下通常与其相伴的组分的分离的不同程度。“分离”表示与原始来源或周围环境的一种分离程度。“纯化”表示高于分离的一种分离程度。“纯化的”或“生物学纯的”蛋白质充分地不含其他材料,使得任何杂质不会实质地影响蛋白质的生物学特性或造成其他不利后果。也就是说,本发明的核酸或肽,如果以重组DNA技法生产而基本地不含细胞材料、病毒材料或培养介质;或,如果以化学合成而基本地不含化学前体或其他化学品,则本发明的核酸或肽即视为纯化的。纯度和同质性通常用分析化学技法确定,例如,聚丙烯酰胺凝胶电泳或高效液相色谱。术语

“纯化的”可表示核酸或蛋白质在电泳凝胶中基本上产生一条带。对于可以被修饰(例如磷酸化或糖基化)的蛋白质,不同的修饰可能产生不同的分离的蛋白质,其可以被各自纯化。

[0266] “分离的多核苷酸”意为核酸(例如DNA),其不含在本发明的核酸分子所来源生物体的天然-存在的基因组中位于该核酸两侧翼的基因。该术语因此包含,例如,并入载体内的重组DNA;并入自主复制的质粒内或病毒内的重组DNA;或并入原核生物或真核生物的基因组DNA内的重组DNA;或作为独立于其他序列的单独的分子(例如,通过PCR或限制性核酸内切酶消化所产生的cDNA或基因组片段或cDNA片段)存在的重组DNA。另外,该术语包含从DNA分子转录而来的RNA分子,以及编码额外的多肽序列的作为杂合基因一部分的重组DNA。

[0267] “分离的多肽”意为已同与其天然相伴的组分分离的本发明的多肽。通常,以重量计,当多肽至少60%不含与其天然伴随的蛋白质和天然-存在的有机分子时,该多肽即视为分离的。优选地,以重量计,制备为至少75%,更优选至少90%,最优选至少99%,本发明的多肽。可获得分离的本发明的多肽,例如,通过从天然来源提取,通过表达编码此多肽的重组核酸;或通过化学合成该蛋白质。纯度可以通过任何适当的方法测量,例如柱色谱,聚丙烯酰胺凝胶电泳,或通过HPLC(高效液相色谱)分析。

[0268] 术语“连接子”,如本文所用,是指链接两个分子或部分体(例如融合蛋白的两个结构域)的键(例如共价键)、化学基团或分子。在一些实施方式中,连接子连接RNA-可编程的核酸酶的gRNA结合结构域(其包括Cas9核酸酶结构域)和核-酸编辑蛋白(例如,胞苷或腺苷脱氨酶)的催化结构域。在一些实施方式中,连接子连接dCas9和核酸编辑蛋白。通常,连接子是就定位于两个基团、分子或其他部分体的之间或两侧,并经由共价键与两个中的每个都连结,从而将两个连结起来。在一些实施方式中,该连接子是氨基酸或复数个氨基酸(例如,肽或蛋白)。在一些实施方式中,该连接子是有机分子、基团、聚合物或化学部分体。在一些实施方式中,该连接子是5-200个氨基酸的长度,例如,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,25,35,45,50,55,60,65,70,75,80,85,90,95,100,101,102,103,104,105,110,120,130,140,150,160,175,180,190,或200个氨基酸的长度。较长或较短的连接子也是预期中的。在一些实施方式中,连接子包括氨基酸序列SGSETPGTSESATPES,其也称为XTEN连接子。在一些实施方式中,连接子包括氨基酸序列SGGS。在一些实施方式中,连接子包括$(SGGS)_n$,$(GGGS)_n$,$(GGGGS)_n$,$(G)_n$,$(EAAAK)_n$,$(GGS)_n$,SGSETPGTSESATPES,或$(XP)_n$基序,或这些中任何的组合,其中n独立地是1和30之间的整数,且其中X是任何氨基酸。在一些实施方式中,n是1,2,3,4,5,6,7,8,9,10,11,12,13,14,或15。

[0269] 在一些实施方式中,所述核碱基编辑器的各结构域是经由包括以下氨基酸序列的连接子而融合的:SGGSSGSETPGTSESATPESSGGS,SGGSSGGSSGSETPGTSESATPESSGGSSGGS或GGSGGSPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGGSGGS。在一些实施方式中,所述核碱基编辑器的各结构域是经由包括氨基酸序列SGSETPGTSESATPES(其也可被称为XTEN连接子)的连接子而融合的。在一些实施方式中,该连接子是24个氨基酸的长度。在一些实施方式中,该连接子包括氨基酸序列SGGSSGGSSGSETPGTSESATPES。在一些实施方式中,该连接子是40个氨基酸的长度。在一些实施方式中,该连接子包括氨基酸序列SGGSSGGSSGSETPGTSESATPESSGGSSGGSSGGSSGGS。在一些实施方式中,该连接子是64个氨基酸的长度。在一些实施方式中,该连接子包括氨基酸序列SGGSSGGSSGSETPGTSESATPESSGGSSGGSSGGSSGGSSGSETPGTSESATPESSGGSSGGS。在一些实施

方式中,该连接子是92个氨基酸的长度。在一些实施方式中,该连接子包括氨基酸序列PGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATS。

[0270] 术语“突变”,如本文所用,是指序列(例如核酸或氨基酸序列)中的一残基被另一残基替换,或序列中一个或多个残基的缺失或插入。在本文中,突变的描述通常是通过首先指出原始的残基,随后指出该序列中该残基的位置,然后再指出新替换的残基的身份。制造出本文所提供的氨基酸替换(突变)的各种方法在本领域中是众所周知的,例如由Green和Sambrook,Molecular Cloning:A Laboratory Manual(第4版,Cold Spring Harbor Laboratory Press,冷泉港,纽约(2012))所提供。

[0271] 术语“核酸”和“核酸分子”,如本文所用,是指化合物,其包括核碱基和酸性部分体,例如核苷、核苷酸或核苷酸的聚合物。通常,聚合的核酸,例如包括三个或更多个核苷酸的核酸分子,是线性分子,其中相邻的核苷酸经由磷酸二酯键合彼此链接。在一些实施方式中,“核酸”是指单个的核酸残基(例如,核苷酸和/或核苷)。在一些实施方式中,“核酸”是指包括三个或更多个单个核苷酸残基的寡核苷酸链。如本文所用,术语“寡核苷酸”和“多核苷酸”可以互换使用以指代核苷酸的聚合物(例如,至少三个核苷酸的一串)。在一些实施方式中,“核酸”涵盖了RNA以及单股和/或双-股DNA。核酸可以是天然存在的,例如在如下情况中:基因组、转录本、mRNA、tRNA、rRNA、siRNA、snRNA、质粒、黏粒、染色体、染色单体或其他天然存在的核酸分子。另一方面,核酸分子可以是非-天然存在的分子,例如,重组DNA或RNA,人工染色体,工程化的基因组,或其片段,或合成的DNA,RNA,DNA/RNA杂合体,或包含非-天然存在的核苷酸或核苷。此外,术语“核酸”,“DNA”,“RNA”和/或类似术语包含核酸类似物,例如,具有非磷酸二酯主干的类似物。核酸可以从天然来源中被纯化,使用重组表达系统加以生产,并可选地被纯化、被化学合成等。在适当的情况下,例如,在化学合成的分子的情况下,核酸可以包括核苷类似物,诸如具有化学经修饰的碱基或糖,以及主干修饰的类似物。除非另有说明,否则核酸序列是以5′至3′方向呈现。在一些实施方式中,核酸是抑或包括天然核苷(例如,腺苷、胸苷、鸟苷、胞苷、尿苷、脱氧腺苷、脱氧胸苷、脱氧鸟苷和脱氧胞苷);核苷类似物(例如,2-氨基腺苷、2-硫胸苷、肌苷、吡咯-嘧啶、3-甲基腺苷、5-甲基胞苷、2-氨基腺苷、C5-溴尿苷、C5-氟尿苷、C5-碘尿苷、C5-丙炔基-尿苷、C5-丙炔基-胞苷、C5-甲基胞苷、2-氨基腺苷、7-脱氮腺苷、7-脱氮鸟苷、8-氧代腺苷、8-氧代鸟苷、O(6)-甲基鸟嘌呤和2-硫胞苷);化学经修饰的碱基;生物经修饰的碱基(例如,甲基化的碱基);嵌和的碱基;经修饰的糖(例如,2′-氟核糖、核糖、2′-脱氧核糖、阿拉伯糖和己糖);和/或经修饰的磷酸基团(例如,硫代磷酸和5′-N-亚磷酰胺键合)。

[0272] 术语“核定位序列”,“核定位信号”或“NLS”是指促进蛋白质导入细胞核的氨基酸序列。核定位序列在本领域中是已知的,并且描述于例如Plank等,国际PCT申请,PCT/EP2000/011690,其提交于2000年11月23日,以WO/2001/038547公布于2001年5月31日,针对其示例性核定位序列的公开揭露,其内容在此通过引用并入本文。在其他实施方式中,NLS是优化过的NLS,其描述于例如Koblan等,Nature Biotech.2018doi:10.1038/nbt.4172。在一些实施方式中,NLS包括氨基酸序列KRTADGSEFESPKKKRKV、KRPAATKKAGQAKKKK、KKTELQTTNAENKTKKL、KRGINDRNFWRGENGRKTR、RKSGKIAAIVVKRPRK,PKKKRKV或MDSLLMNRRKFLYQFKNVRWAKGRRETYLC。

[0273] 本公开揭露提供了核酸可编程的核酸(例如,DNA或RNA)结合蛋白。该核酸可编程的核酸结合蛋白可以是,例如,“核酸可编程的DNA结合蛋白”或“napDNAbp”。术语“核酸可编程的DNA结合蛋白”或“napDNAbp”是指与核酸(例如DNA或RNA),诸如将napDNAbp导向至特定核酸序列的向导核酸,相联的蛋白质。例如,Cas9蛋白可以与向导RNA相联,该向导RNA将该Cas9蛋白导向至与该向导RNA互补的特定DNA序列。在一些实施方式中,该napDNAbp是Cas9结构域,例如核酸酶有活性的Cas9、Cas9切口酶(nCas9),或核酸酶无活性的Cas9(dCas9)。核酸可编程的DNA结合蛋白的实例包含但不限于:Cas9(例如,dCas9和nCas9)、CasX、CasY、Cpf1、Cas12b/C2c1和Cas12c/C2c3。其他核酸可编程的DNA结合蛋白也在此公开揭露的范围内,尽管它们在本公开揭露中可能未被明确列出。

[0274] 如本文所用,“获得”,正如“获得(一)剂”中那样,包含合成、购买或以其他方式取得该剂。

[0275] 术语“RNA-可编程的核酸酶”和“RNA-导向的核酸酶”与非切开的靶标的一种或多种RNA一起使用(例如结合或相联)。在一些实施方式中,当RNA-可编程的核酸酶与RNA形成复合体时,可称其为核酸酶:RNA复合体。通常,所结合的RNA被称为向导RNA(gRNA)。gRNA可以作为具有两个或多个RNA的复合体存在,也可以作为单一RNA分子存在。以单一RNA分子存在的gRNA可被称为单-向导RNA(sgRNA),尽管“gRNA”可互换使用以指代以单个分子存在或以具有两个或多个分子的复合体存在的向导RNA。通常,以单一RNA种类存在的gRNA包括两个结构域:(1)与靶核酸共享同源性的结构域(例如,并且导引Cas9复合体结合到靶上);(2)结合Cas9蛋白的结构域。在一些实施方式中,结构域(2)对应于被称为tracrRNA的序列,并包括茎-环结构。例如,在一些实施方式中,结构域(2)与Jinek等,Science337:816-821(2012)(其全部内容通过引用并入本文)中所提供的tracrRNA是相同的或同源的。gRNA的其他实例(例如,包括结构域2的那些)可以在2013年9月6日提交的题为“Switchable Cas9Nucleases And Uses Thereof”的美国临时专利申请U.S.S.N.61/874,682中,以及在2013年9月6日提交的题为“Delivery System For Functional Nucleases”的美国临时专利申请U.S.S.N.61/874,746中找到,其每个的全部内容在此通过引用整体并入。在一些实施方式中,gRNA包括两个或更多个的结构域(1)和(2),并可被称为“延伸的gRNA”。如本文所述,延伸的gRNA将结合两个或更多个Cas9蛋白并在两个或更多个独特的区域处结合靶核酸。该gRNA包括与靶位点互补的核苷酸序列,其介导该核酸酶/RNA复合体结合到所述靶位点,并提供了该核酸酶:RNA复合体的序列特异性。在一些实施方式中,该RNA-可编程的核酸酶是(CRISPR-关联系统)Cas9核酸内切酶,例如,来自化脓性链球菌的Cas9(Csn1)(参见例如“Complete genome sequence of an M1 strain of Streptococcus pyogenes.”Ferretti J.J.,McShan W.M.,Ajdic D.J.,Savic D.J.,Savic G.,Lyon K.,Primeaux C,Sezate S.,Suvorov A.N.,Kenton S.,Lai H.S.,Lin S.P.,Qian Y.,Jia H.G.,Najar F.Z.,Ren Q.,Zhu H.,Song L.,White J.,Yuan X.,Clifton S.W.,Roe B.A.,McLaughlin R.E.,Proc.Natl.Acad.Sci.美国98:4658-4663(2001);“CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III.”Deltcheva E.,Chylinski K.,Sharma CM.,Gonzales K.,Chao Y.,Pirzada Z.A.,Eckert M.R.,Vogel J.,Charpentier E.,Nature471:602-607(2011))。

[0276] 术语“重组”,如本文在蛋白质或核酸的情境中使用的,是指自然界中并不存在,而

是人类工程学的产物的蛋白质或核酸。例如,在一些实施方式中,重组蛋白或重组核酸分子包括氨基酸或核苷酸序列,该氨基酸或核苷酸序列与任何天然存在的序列相比包括至少一个,至少两个,至少三个,至少四个,至少五个,至少六个,或至少七个突变。

[0277] “减少”意为至少10%,25%,50%,75%,或100%的负向改变。

[0278] “参考”意为一标准或对照条件。

[0279] “参考序列”是用作序列比对的基准的界定的序列。参考序列可能是指定序列的子集或全部;例如,全长cDNA或基因序列的一节段,或完整的cDNA或基因序列。对多肽来说,参考多肽序列的长度将一般为至少约16个氨基酸,至少约20个氨基酸,至少约25个氨基酸,约35个氨基酸、约50个氨基酸或约100个氨基酸。对核酸来说,参考核酸序列的长度一般为至少约50个核苷酸、至少约60个核苷酸、至少约75个核苷酸、约100个核苷酸或约300个核苷酸或任何约等于它们或在它们之间的整数。

[0280] “特异性结合”意为识别并结合本发明的多肽和/或核酸分子,但基本上不识别和结合样品(例如生物样品)中其他分子的核酸分子、多肽或其复合体(例如,核酸可编程的DNA结合结构域和向导核酸)、化合物或分子。

[0281] 可用于本发明的方法的核酸分子包含任何编码本发明的多肽或该多肽片段的核酸分子。此种核酸分子不需要与内源性核酸序列100%相同,但通常将展现出实质的同一性。与内源性序列具有“实质的同一性”的多核苷酸通常能够与双-股核酸分子的至少一股杂交。可用于本发明的方法的核酸分子包含任何编码本发明的多肽或该多肽片段的核酸分子。此种核酸分子不需要与内源性核酸序列100%相同,但通常将展现出实质的同一性。与内源性序列具有“实质的同一性”的多核苷酸通常能够与双-股核酸分子的至少一股杂交。“杂交”意为在各种严格度的条件下,互补的多核苷酸序列(例如本文所述的基因)或其部分之间配对并且形成双股分子(参见例如,Wahl,G.M.和S.L.Berger(1987),Methods Enzymol.152:399;Kimmel,A.R.(1987),Methods Enzymol.152:507)。

[0282] 例如,严格的盐浓度将一般小于约750mM氯化钠和75mM柠檬酸钠,优选小于约500mM氯化钠和50mM柠檬酸钠,更优选小于约250mM氯化钠和25mM柠檬酸钠。低严格度杂交可在无有机溶剂(例如甲酰胺)的情况下获得,而高严格度杂交可在至少约35%甲酰胺,和更优选至少约50%甲酰胺的存在下获得。严格的温度条件将一般包含至少约30℃,更优选至少约37℃,最优选至少约42℃的温度。变化其他参数,诸如杂交时间、洗涤剂(例如十二烷基硫酸钠(SDS))的浓度、以及载具DNA的包含或排除,是本领域技术人员熟知的。根据需要,通过合并这些不同的条件来实现严格度的各种层级。在一个实施方式中,杂交将发生在30℃下、750mM氯化钠,75mM柠檬酸钠和1%SDS中。在另一个实施方式中,杂交将发生在37℃下、500mM氯化钠,50mM柠檬酸钠,1%SDS,35%甲酰胺和100μg/ml变性的鲑鱼精子DNA(ssDNA)中。在另一个实施方式中,杂交将发生在42℃下、250mM氯化钠,25mM柠檬酸钠,1%SDS,50%甲酰胺和200μg/ml ssDNA中。对于本领域技术人员来说,在这些条件上的有用的变化是显而易见的。

[0283] 对于大多数应用,杂交后的清洗步骤在严格度方面也将变化。清洗的严格度条件可以通过盐浓度和通过温度来定义。如上,可以通过降低盐浓度或通过升高温度而提高清洗的严格度。例如,用于清洗步骤的严格盐浓度将优选是小于约30mM氯化钠和3mM柠檬酸钠,最优选是小于约15mM氯化钠和1.5mM柠檬酸钠。用于清洗步骤的严格温度条件将一般包

含至少约25℃的温度,更优选至少约42℃,甚至更优选至少约68℃。在一实施方式中,清洗步骤将发生在25℃下,30mM氯化钠,3mM柠檬酸钠和0.1%SDS中。在一更优选的实施方式中,清洗步骤将发生在42℃下,15mM氯化钠,1.5mM柠檬酸钠和0.1%SDS中。在一更优选的实施方式中,清洗步骤将发生在68℃下,15mM氯化钠,1.5mM柠檬酸钠和0.1%SDS中。对于本领域技术人员来说,在这些条件上的额外的变化是显而易见的。杂交技法是本领域技术人员熟知的,并且在如下文献中描述,例如:Benton和Davis(Science196:180,1977);Grunstein和Hogness(Proc.Natl.Acad.Sci.,美国72:3961,1975);Ausubel等(Current Protocols in Molecular Biology,Wiley Interscience,纽约,2001);Berger和Kimmel(Guide to Molecular Cloning Techniques,1987,Academic Press,纽约);以及Sambrook等,Molecular Cloning:A Laboratory Manual,Cold Spring Harbor Laboratory Press,纽约。

[0284] “基本上相同”意为与参考氨基酸序列(例如,本文所述的任何一段氨基酸序列)或核酸序列(例如本文所述的任何一段核酸序列)展现出至少50%同一性的多肽或核酸分子。在一个实施方式中,此一序列与用于比较的序列在氨基酸水平或核酸水平上是至少60%,80%或85%,90%,95%或甚至99%相同的。

[0285] 序列同一性通常使用序列分析软件来测量(例如,遗传学计算机小组的序列分析软件包,威斯康星大学生技中心,1710大学大道,麦迪逊,威州53705,BLAST,BESTFIT,GAP,或PILEUP/PRETTYBOX程序)。此种软件通过为各种替换、缺失和/或其他修饰指定同源性程度而匹配相同或相似的序列。保守性替换通常包含以下群组的组内替换:甘氨酸,丙氨酸;缬氨酸,异亮氨酸,亮氨酸;天冬氨酸,谷氨酸,天冬酰胺,谷氨酰胺;丝氨酸,苏氨酸;赖氨酸,精氨酸;和苯丙氨酸,酪氨酸。在确定同一性程度的示例性方法中,可使用BLAST程序,其中e^{-3}和e^{-100}之间的概率分数表示密切相关的序列。

[0286] “受试者”意为哺乳动物,包含但不限于人类或非-人类哺乳动物,诸如牛、马、犬、绵羊或猫。受试者包含家畜,即饲养以产出劳动力和提供商品(诸如食物)的驯养动物,包含但不限于牛、山羊、鸡、马、猪、兔和绵羊。

[0287] 术语“靶位点”是指被核碱基编辑器修饰的核酸分子内的序列。在一个实施方式中,该靶位点被脱氨酶或融合蛋白(该融合蛋白包括脱氨酶(例如,胞苷或腺苷脱氨酶))脱氨基。

[0288] 因为RNA-可编程的核酸酶(例如Cas9)使用RNA:DNA杂交来靶向DNA切开位点,所以这些蛋白质原则上有能力被靶向到由向导RNA所指定的任何序列上。使用RNA-可编程的核酸酶(诸如Cas9)用于位点-特异性切开(例如,以修饰基因组)的方法是本领域已知的(参见例如,Cong,L.等人,Multiplex genome engineering using CRISPR/Cas systems.Science 339,819-823(2013);Mali,P.等人,RNA-guided human genome engineering via Cas9.Science 339,823-826(2013);Hwang,W.Y.等人,Efficient genome editing in zebrafish using a CRISPR-Cas system.Nature biotechnology 31,227-229(2013);Jinek,M.等人,RNA-programmed genome editing in human cells.eLife 2,e00471(2013);Dicarlo,J.E.等人,Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems.Nucleic acids research (2013);Jiang,W.等人,RNA-guided editing of bacterial genomes using CRISPR-Cas

systems.Nature biotechnology 31,233-239(2013);其每个的全部内容通过引用并入本文)。

[0289] 本文所提供的范围应被理解为该范围内所有数值的简略表达。例如,1到50的范围应被理解为包含任何数字、数字的组合或任何子-范围,其来自于由以下所组成的群组:1、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、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49或50。

[0290] 如本文所用,除非特别说明或从上下文中显见,否则术语"或(or)"应被理解为包含在内。如本文所用,除非特别说明或从上下文中显见,否则术语"一(a)","一(an)"和"该(the)"应被理解为单数或复数。

[0291] 如本文所用,除非特别说明或从上下文中显见,否则术语"约(about)"应被理解为在本领域的正常允差范围内,例如在平均值的2个标准差内。"约"可以理解为给定数值的10%、9%、8%、7%、6%、5%、4%、3%、2%、1%、0.5%、0.1%、0.05%或0.01%以内。除非上下文另有明确说明,否则本文所提供的所有数字的数值均由术语"约"修饰。

[0292] 本文对变量的任何定义中的化学基团列表的列举,包含了将该变量定义为所列基团的任何单个基团或所列基团的组合。本文对变量或方面的实施方式的列举,包含了该实施方式作为任何单一实施方式或与任何其他实施方式或该任何其他实施方式的部分的组合。

[0293] 本文所提供的任何组合物和方法可以与本文所提供的任何其他组合物和方法组合。

附图说明

[0294] 图1描绘腺苷核碱基编辑器的模型,并且部分地提供旁观者诱变的结构性基础。

[0295] 图2是描绘用于碱基编辑的靶DNA区位的预测的模型。

[0296] 图3描绘显示图2中靠近靶DNA区位的脱氨酶结构的位置的模型。

[0297] 图4是腺苷核碱基编辑器的模型,其描绘所鉴别出的一个或多个脱氨酶结构域可能插入到Cas9中的区域。靠近脱氨酶结构域所在的环(黄色)可能靶向单股DNA(洋红色)。感兴趣的区域包含标记为A、B、C、D、E、F、G和H的那些。

[0298] 图5是图4中模型的放大视图,其显示区域B、C、D、E和F中的残基。

[0299] 图6是图4中模型的放大视图,其显示区域F、G和H中的残基。

[0300] 图7是图4中模型的放大视图,其显示区域A、B、C、D和E中的残基。

[0301] 图8描绘一种高通量体外脱氨反应测定。通过比较含有核碱基编辑器且有在靶探针(其含有用于该碱基编辑器的底物)的反应和不含该碱基编辑器但有用于检测脱靶脱氨反应的探针的反应,可以将该探针的假性脱氨反应与在靶脱氨反应区分开来。

[0302] 图9是描绘用于脱靶脱氨反应的荧光测定的结果的图表。

[0303] 图10是描绘腺苷碱基编辑器(ABE)与具有反式TadA的ABE系统的比较的图表。

[0304] 图11描绘用于假性脱靶碱基编辑的潜在的底物。

[0305] 图12描绘评估顺式和反式脱氨酶活性的测定。

[0306] 图13是描绘在所述顺式-反式测定中rAPOBEC1的活性的图表。

[0307] 图14是描绘在所述顺式-反式测定中TadA-TadA7.10的活性的图表。

[0308] 图15描绘在碱基编辑器情境(反式ABE)中对于TadA-TadA7.10观察到较低的反式活性。

[0309] 图16是描述GFP表达的剂量-反应的结果的图表。用空载体对pmaxGFP质粒的滴定导致GFP表达水平的降低。

[0310] 图17是描述腺苷核碱基编辑器ABE的顺式和反式活性的剂量-反应的图表。

[0311] 图18是描述胞苷核碱基编辑器BE4的顺式和反式活性的剂量-反应的图表。

[0312] 图19是显示以减低的假性脱氨反应为标准而筛选脱氨酶的结果的图表。脱氨酶ppAPOBEC-2(10),mAPOBEC-2(8),mAPOBEC-3(12),和mfAPOBEC-4(22)显示出高的顺式/反式活性。

[0313] 图20A至Q描绘所检查的编辑器的碱基编辑活性。图20A是ABE7.10的示意图,其TadA经由XTEN连接子与Cas9融合。图20B-Q显示示例性的内部融合碱基编辑器在靶序列GAACACAAAGCATAGACTGC(HEK2)和GGACAGCTTTTCCTAGACAG(T39)的R-环范围内的靶向股上的A变G脱氨反应的百分比,即碱基编辑活性。图20B,ABE7.10的编辑活性。图20C,ISLAY008的编辑活性。图20D,ISLAY003的编辑活性。图20E,ISLAY002的编辑活性。图20F,ISLAY007的编辑活性。图20G,ISLAY001的编辑活性。图20H,ISLAY005的编辑活性。图20I,ISLAY006的编辑活性。图20J,ISLAY004的编辑活性。图20K,ISLAY021的编辑活性。图20L,ISLAY031的编辑活性。图20M,ISLAY020的编辑活性。图20N,ISLAY036的编辑活性。图20O,ISLAY035的编辑活性。图20P,ISLAY028的编辑活性。图20Q,ISLAY009的编辑活性。

[0314] 图21A至B显示示例性的碱基编辑器的示意图。图21A显示示例性的碱基编辑器ABE7.10和示例性的碱基编辑器(IBE002,IBE004、IBE005、IBE006、IBE008、IBE009和IBE020)的示意图。图21B显示上述碱基编辑的空间卡通图表征,其显示脱氨酶插入的区位。

[0315] 图22A-D描绘,与ABE7.10相比,示例性的内部融合碱基编辑器在29个不同基因组靶标处的碱基编辑效率。图22A显示编辑效率被归一化到在最佳位置处的ABE7.10编辑。图22B显示最大编辑效率的IBE,对其总结并将其与ABE7.10比较。图22C显示每个碱基编辑器的峰值编辑位置的高斯平滑法表征。图22D显示来自所述29个被测试的靶标的归一化编辑的热度图。

[0316] 图23描绘通过反式编辑测定所测量的假性脱氨反应,正如29个不同的靶标归一化到在每个位点处的ABE7.10。

[0317] 图24A至F显示A-碱基编辑器在6个基因组基因座处的编辑百分比:HEK4(图24A),FANCF(图24B),HEK-3(图24C),HEK2-YY(图24D),EMX1(图24E),HEK2(图24F)。X-轴:核碱基位置,1为距PAM最远而20为PAM近端(PAM为位置21至23)。Y轴:通过Illumina测序而测量的A变G编辑的百分比。

[0318] 图25A至F显示C-碱基编辑器在6个基因组基因座处的编辑百分比:HEK4(图25A),FANCF(图25B),HEK-3(图25C),HEK2-YY(图25D),EMX1(图25E),HEK2(图25F)。X-轴:核碱基位置,1为距PAM最远而20为PAM近端(PAM为位置21至23)。Y轴:通过Illumina测序而测量的A变G编辑的百分比。

具体实施方式

[0319] 如下所述,本发明的特色为具有减低的非靶脱氨反应的碱基编辑器,使用所述碱

基编辑器的方法,和用于定性具有降低的非靶脱氨反应(例如,与程式化的在靶脱氨反应相比)的碱基编辑器的测定。

[0320] 腺苷脱氨酶

[0321] 在一些实施方式中,本发明的核碱基编辑器包括腺苷脱氨酶结构域。在一些实施方式中,本文所提供的腺苷脱氨酶能够使腺嘌呤脱氨基。在一些实施方式中,本文所提供的腺苷脱氨酶能够使DNA的脱氧腺苷残基中的腺嘌呤脱氨基。该腺苷脱氨酶可衍生自任何适合的生物体(例如,大肠杆菌)。在一些实施方式中,该腺嘌呤脱氨酶是天然-存在的腺苷脱氨酶,其包含一个或多个对应于本文所提供的任何突变(例如,ecTadA中的突变)的突变。本领域技术人员将能够,例如通过序列比对和同源残基的确定,鉴别出任何同源蛋白质中的相对应残基。因此,本领域技术人员将能够在任何天然-存在的腺苷脱氨酶(例如,与ecTadA具有同源性)中生成对应于本文所述的任何突变(例如,在ecTadA中鉴别出的任何突变)。在一些实施方式中,该腺苷脱氨酶来自原核生物。在一些实施方式中,该腺苷脱氨酶来自细菌。在一些实施方式中,该腺苷脱氨酶来自大肠杆菌,金黄色葡萄球菌,伤寒沙门菌,腐败希瓦氏菌,流感嗜血杆菌,新月柄杆菌,或枯草芽孢杆菌。在一些实施方式中,该腺苷脱氨酶来自大肠杆菌。

[0322] 在一个实施方式中,本发明的融合蛋白包括与TadA7.10链接的野生型TadA,而该TadA7.10则链接到Cas9切口酶上。在特定的实施方式中,该融合蛋白包括单一TadA7.10结构域(例如,以单体提供)。在其他实施方式中,该ABE7.10编辑器包括能够形成异二聚体的TadA7.10和TadA(wt)。其相关序列如下:

[0323] TadA(wt):

[0324] SEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTD

[0325] TadA7.10:

[0326] SEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD

[0327] 在一些实施方式中,该腺苷脱氨酶包括的氨基酸序列,与本文所提供的任何腺苷脱氨酶中列出的任一氨基酸序列,是至少60%,至少65%,至少70%,至少75%,至少80%,至少85%,至少90%,至少95%,至少96%,至少97%,至少98%,至少99%,或至少99.5%相同的。应当理解,本文所提供的腺苷脱氨酶可包含一个或多个突变(例如,本文所提供的任何突变)。本公开揭露所提供的任何脱氨酶结构域,均具有一定百分比的同一性,再加上本文所述的任何突变或其组合。在一些实施方式中,该腺苷脱氨酶包括的氨基酸序列,与参考序列或本文所提供的任何腺苷脱氨酶相比,具有1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50个,或更多个突变。在一些实施方式中,该腺苷脱氨酶包括的氨基酸序列,与本领域已知或本文所述的任一氨基酸序列相比,具有至少5,至少10,至少15,至少20,至少25,至少30,至少35,至少40,至少45,至少50,至少60,至少70,至少80,至少90,至少100,至少110,至少120,至少130,至少140,至少150,至少160,或至少170个相同

的接连的氨基酸残基。

[0328] 在一些实施方式中,该腺苷脱氨酶包括在TadA参考序列中的D108X突变,或在另一腺苷脱氨酶中的相应突变,其中X表示在野生型腺苷脱氨酶中相应氨基酸以外的任何氨基酸。在一些实施方式中,该腺苷脱氨酶包括在TadA参考序列中的D108G、D108N、D108V、D108A或D108Y突变或在另一腺苷脱氨酶中的相应突变。然而,应当理解,可类似地比对额外的脱氨酶以鉴别出可以如本文所提供的那样被突变的同源氨基酸残基。

[0329] 在一些实施方式中,该腺苷脱氨酶包括在TadA参考序列中的A106X突变,或在另一腺苷脱氨酶中的相应突变,其中X表示在野生型腺苷脱氨酶中相应氨基酸以外的任何氨基酸。在一些实施方式中,该腺苷脱氨酶包括在TadA参考序列中的A106V突变,或在另一腺苷脱氨酶中的相应突变。

[0330] 在一些实施方式中,该腺苷脱氨酶包括在TadA参考序列中的E155X突变,或在另一腺苷脱氨酶中的相应突变,其中X的存在表示在野生型腺苷脱氨酶中相应氨基酸以外的任何氨基酸。在一些实施方式中,该腺苷脱氨酶包括在TadA参考序列中的E155D、E155G或E155V突变,或在另一腺苷脱氨酶中的相应突变。

[0331] 在一些实施方式中,该腺苷脱氨酶包括在TadA参考序列中的D147X突变,或在另一腺苷脱氨酶中的相应突变,其中X的存在表示在野生型腺苷脱氨酶中相应氨基酸以外的任何氨基酸。在一些实施方式中,该腺苷脱氨酶包括在TadA参考序列中的D147Y突变,或在另一腺苷脱氨酶中的相应突变。

[0332] 应当理解,可将本文所提供的任何突变(例如,基于TadA参考序列的ecTadA氨基酸序列)引入其他腺苷脱氨酶中,诸如金黄色葡萄球菌TadA(saTadA),或其他腺苷脱氨酶(例如细菌的腺苷脱氨酶)。对于熟练技术人员而言显而易见的将是如何通过比对额外的脱氨酶以鉴别出那些与ecTadA中被突变的残基同源的氨基酸残基。因此,可在其他具有同源氨基酸残基的腺苷脱氨酶中制备从ecTadA中所鉴别出的任何突变。还应当理解,本文所提供的任何突变均可在ecTadA或另一腺苷脱氨酶中单独地或以任何组合制备。例如,腺苷脱氨酶可含有在TadA参考序列中的D108N,A106V,E155V和/或D147Y突变,或在另一腺苷脱氨酶中的相应突变。在一些实施方式中,腺苷脱氨酶包括在TadA参考序列中的以下突变群组(各突变群组由“;”分开),或在另一腺苷脱氨酶中的相应突变:D108N和A106V;D108N和E155V;D108N和D147Y;A106V和E155V;A106V和D147Y;E155V和D147Y;D108N、A106V和E55V;D108N、A106V和D147Y;D108N、E55V和D147Y;A106V、E55V和D147Y;和D108N、A106V、E55V和D147Y。然而,应当理解,可以在腺苷脱氨酶(例如ecTadA)中制备本文所提供的相应突变的任何组合。

[0333] 在一些实施方式中,该腺苷脱氨酶包括在TadA参考序列中的一个或多个H8X、T17X、L18X、W23X、L34X、W45X、R51X、A56X、E59X、E85X、M94X、I95X、V102X、F104X、A106X、R107X、D108X、Kl lOX、M118X、N127X、A138X、F149X、M151X、R153X、Q154X、I156X和/或K157X突变,或在另一腺苷脱氨酶中的一个或多个相应突变,其中X的存在表示在野生型腺苷脱氨酶中相应氨基酸以外的任何氨基酸。在一些实施方式中,该腺苷脱氨酶包括在TadA参考序列中的一个或多个H8Y、T17S、L18E、W23L、L34S、W45L、R51H、A56E或A56S,E59G,E85K或E85G、M94L、1951、V102A、F104L、A106V、R107C或R107H或R107P、D108G或D108N或D108V或D108A或D108Y、Kl 101、Ml 18K、N127S、A138V、F149Y、M151V、R153C、Q154L、I156D和/或K157R突变,或在另一腺苷脱氨酶中的一个或多个相应突变。

[0334] 在一些实施方式中,该腺苷脱氨酶包括在TadA参考序列中的一个或多个H8X,D108X和/或N127X突变,或在另一腺苷脱氨酶中的一个或多个相应突变,其中X表示任何氨基酸的存在。在一些实施方式中,该腺苷脱氨酶包括在TadA参考序列中的一个或多个H8Y、D108N和/或N127S突变,或在另一腺苷脱氨酶中的一个或多个相应突变。

[0335] 在一些实施方式中,该腺苷脱氨酶包括在TadA参考序列中的一个或多个H8X、R26X、M61X、L68X、M70X、A106X、D108X、A109X、N127X、D147X、R152X、Q154X,E155X、K161X、Q163X和/或T166X突变,或在另一腺苷脱氨酶中的一个或多个相应突变,其中X表示在野生型腺苷脱氨酶中相应氨基酸以外的任何氨基酸的存在。在一些实施方式中,该腺苷脱氨酶包括在TadA参考序列中的一个或多个H8Y、R26W、M61I、L68Q、M70V、A106T、D108N、A109T、N127S、D147Y、R152C、Q154H或Q154R、E155G或E155V或E155D、K161Q、Q163H和/或T166P突变,或在另一腺苷脱氨酶中的一个或多个相应突变。

[0336] 在一些实施方式中,该腺苷脱氨酶包括一、二、三、四、五或六个突变,其选自于由以下所组成的群组:在TadA参考序列中的H8X、D108X、N127X、D147X、R152X和Q154X,或在另一腺苷脱氨酶中的一个或多个相应突变,其中X表示在野生型腺苷脱氨酶中相应氨基酸以外的任何氨基酸的存在。在一些实施方式中,该腺苷脱氨酶包括一、二、三、四、五、六、七或八个突变,其选自于由以下所组成的群组:在TadA参考序列中的H、155X和Q163X,或在另一腺苷脱氨酶中的一个或多个相应突变,其中X表示在野生型腺苷脱氨酶中相应氨基酸以外的任何氨基酸的存在。在一些实施方式中,该腺苷脱氨酶包括一、二、三、四或五个突变,其选自于由以下所组成的群组:在TadA参考序列中的H8X、D108X、N127X、E155X和T166X,或在另一腺苷脱氨酶中的一个或多个相应突变,其中X表示在野生型腺苷脱氨酶中相应氨基酸以外的任何氨基酸的存在。在一些实施方式中,该腺苷脱氨酶包括一、二、三、四、五或六个突变,其选自于由以下所组成的群组:H8X、A106X、D108X、另一腺苷脱氨酶中的一个或多个突变,其中X表示在野生型腺苷脱氨酶中相应氨基酸以外的任何氨基酸的存在。在一些实施方式中,该腺苷脱氨酶包括一、二、三、四、五、六、七或八个突变,其选自于由以下所组成的群组:在TadA参考序列中的H8X、R126X、L68X、D108X、N127X、D147X和E155X或在另一腺苷脱氨酶中的一个或多个相应突变,其中X表示在野生型腺苷脱氨酶中相应氨基酸以外的任何氨基酸的存在。在一些实施方式中,该腺苷脱氨酶包括一、二、三、四或五个突变,其选自于由以下所组成的群组:在TadA参考序列中的H8X、D108X、A109X、N127X和E155X,或在另一腺苷脱氨酶中的一个或多个相应突变,其中X表示在野生型腺苷脱氨酶中相应氨基酸以外的任何氨基酸的存在。

[0337] 在一些实施方式中,该腺苷脱氨酶包括一、二、三、四、五或六个突变,其选自于由以下所组成的群组:在TadA参考序列中的H8Y、D108N、N127S、D147Y、R152C和Q154H,或在另一腺苷脱氨酶中的一个或多个相应突变。在一些实施方式中,该腺苷脱氨酶包括一、二、三、四、五、六、七或八个突变,其选自于由以下所组成的群组:在TadA参考序列中的H8Y、M61I、M70V、D108N、N127S、Q154R、E155G和Q163H,或在另一腺苷脱氨酶中的一个或多个相应突变。在一些实施方式中,该腺苷脱氨酶包括一、二、三、四或五个突变,其选自于由以下所组成的群组:在TadA参考序列中的H8Y、D108N、N127S、E155V和T166P,或在另一腺苷脱氨酶中的一个或多个相应突变。在一些实施方式中,该腺苷脱氨酶包括一、二、三、四、五或六个突变,其选自于由以下所组成的群组:在TadA参考序列中的H8Y、A106T、D108N、N127S、E155D和

K161Q,或在另一腺苷脱氨酶中的一个或多个相应突变。在一些实施方式中,该腺苷脱氨酶包括一、二、三、四、五、六、七或八个突变,其选自于由以下所组成的群组:在TadA参考序列中的H8Y、R126W、L68Q、D108N、N127S、D147Y和E155V,或在另一腺苷脱氨酶中的一个或多个相应突变。在一些实施方式中,该腺苷脱氨酶包括一、二、三、四或五个突变,其选自于由以下所组成的群组:在TadA参考序列中的H8Y、D108N、A109T、N127S和E155G,或在另一腺苷脱氨酶中的一个或多个相应突变。

[0338] 在一些实施方式中,该腺苷脱氨酶包括在另一腺苷脱氨酶中的一个或多个相应突变。在一些实施方式中,该腺苷脱氨酶包括在TadA参考序列中的D108N,D108G,或D108V突变,或在另一腺苷脱氨酶中的多个相应突变。在一些实施方式中,该腺苷脱氨酶包括在TadA参考序列中的A106V和D108N突变,或在另一腺苷脱氨酶中的多个相应突变。

[0339] 在一些实施方式中,该腺苷脱氨酶包括在TadA参考序列中的R107C和D108N突变,或在另一腺苷脱氨酶中的多个相应突变。在一些实施方式中,该腺苷脱氨酶包括在TadA参考序列中的H8Y、D108N、N127S、D147Y和Q154H突变,或在另一腺苷脱氨酶中的多个相应突变。在一些实施方式中,该腺苷脱氨酶包括在TadA参考序列中的H8Y、R24W、D108N、N127S、D147Y和E155V突变,或在另一腺苷脱氨酶中的多个相应突变。在一些实施方式中,该腺苷脱氨酶包括在TadA参考序列中的D108N、D147Y和E155V突变,或在另一腺苷脱氨酶中的多个相应突变。在一些实施方式中,该腺苷脱氨酶包括在TadA参考序列中的H8Y、D108N和N127S突变,或在另一腺苷脱氨酶中的多个相应突变。在一些实施方式中,该腺苷脱氨酶包括在TadA参考序列中的A106V、D108N、D147Y和E155V突变,或在另一腺苷脱氨酶中的多个相应突变。

[0340] 在一些实施方式中,该腺苷脱氨酶包括在TadA参考序列中的一个或多个S2X、H8X、I49X、L84X、H123X、N127X、I156X和/或K160X突变,或在另一腺苷脱氨酶中的一个或多个相应突变,其中X的存在表示在野生型腺苷脱氨酶中相应氨基酸以外的任何氨基酸。在一些实施方式中,该腺苷脱氨酶包括在TadA参考序列中的一个或多个S2A、H8Y、I49F、L84F、H123Y、N127S、I156F和/或K160S突变,或在另一腺苷脱氨酶中的一个或多个相应突变。

[0341] 在一些实施方式中,该腺苷脱氨酶包括L84X突变腺苷脱氨酶,其中X表示在野生型腺苷脱氨酶中相应氨基酸以外的任何氨基酸。在一些实施方式中,该腺苷脱氨酶包括在TadA参考序列中的L84F突变,或在另一腺苷脱氨酶中的相应突变。

[0342] 在一些实施方式中,该腺苷脱氨酶包括在TadA参考序列中的H123X突变,或在另一腺苷脱氨酶中的相应突变,其中X表示在野生型腺苷脱氨酶中相应氨基酸以外的任何氨基酸。在一些实施方式中,该腺苷脱氨酶包括在TadA参考序列中的H123Y突变,或在另一腺苷脱氨酶中的相应突变。

[0343] 在一些实施方式中,该腺苷脱氨酶包括在TadA参考序列中的I157X突变,或在另一腺苷脱氨酶中的相应突变,其中X表示在野生型腺苷脱氨酶中相应氨基酸以外的任何氨基酸。在一些实施方式中,该腺苷脱氨酶包括在TadA参考序列中的I157F突变,或在另一腺苷脱氨酶中的相应突变。

[0344] 在一些实施方式中,该腺苷脱氨酶包括一、二、三、四、五、六或七个突变,其选自于由以下所组成的群组:在TadA参考序列中的L84X、A106X、D108X、H123X、D147X、E155X和I156X,或在另一腺苷脱氨酶中的一个或多个相应突变,其中X表示在野生型腺苷脱氨酶中相应氨基酸以外的任何氨基酸的存在。在一些实施方式中,该腺苷脱氨酶包括一、二、三、

四、五或六个突变,其选自于由以下所组成的群组:在TadA参考序列中的S2X、I49X、A106X、D108X、D147X和E155X,或在另一腺苷脱氨酶中的一个或多个相应突变,其中X表示在野生型腺苷脱氨酶中相应氨基酸以外的任何氨基酸的存在。在一些实施方式中,该腺苷脱氨酶包括一、二、三、四或五个突变,其选自于由以下所组成的群组:在TadA参考序列中的H8X、A106X、D108X、N127X和K160X,或在另一腺苷脱氨酶中的一个或多个相应突变,其中X表示在该野生型腺苷脱氨酶中相应氨基酸以外的任何氨基酸的存在。

[0345] 在一些实施方式中,该腺苷脱氨酶包括一、二、三、四、五、六或七个突变,其选自于由以下所组成的群组:在TadA参考序列中的L84F、A106V、D108N、H123Y、D147Y、E155V和I156F,或在另一腺苷脱氨酶中的一个或多个相应突变。在一些实施方式中,该腺苷脱氨酶包括一、二、三、四、五或六个突变,其选自于由以下所组成的群组:在TadA参考序列中的S2A、I49F、A106V、D108N、D147Y和E155V。

[0346] 在一些实施方式中,该腺苷脱氨酶包括一、二、三、四或五个突变,其选自于由以下所组成的群组:在TadA参考序列中的H8Y、A106T、D108N、N127S和K160S,或在另一腺苷脱氨酶中的一个或多个相应突变。

[0347] 在一些实施方式中,该腺苷脱氨酶包括在TadA参考序列中的一个或多个E25X、R26X、R107X、A142X和/或A143X突变,或在另一腺苷脱氨酶中的一个或多个相应突变,其中X的存在表示在野生型腺苷脱氨酶中相应氨基酸以外的任何氨基酸。在一些实施方式中,该腺苷脱氨酶包括在TadA参考序列中的一个或多个E25M、E25D、E25A、E25R、E25V、E25S、E25Y、R26G、R26N、R26Q、R26C、R26L、R26K、R107P、R07K、R107A、R107N、R107W、R107H、R107S、A142N、A142D、A142G、A143D、A143G、A143E、A143L、A143W、A143M、A143S、A143Q和/或A143R突变,或在另一腺苷脱氨酶中的一个或多个相应突变。在一些实施方式中,该腺苷脱氨酶包括一个或多个本文所述的相对应于TadA参考序列的突变,或在另一腺苷脱氨酶中的一个或多个相应突变。

[0348] 在一些实施方式中,该腺苷脱氨酶包括在TadA参考序列中的E25X突变,或在另一腺苷脱氨酶中的相应突变,其中X表示在野生型腺苷脱氨酶中相应氨基酸以外的任何氨基酸。在一些实施方式中,该腺苷脱氨酶包括在TadA参考序列中的E25M、E25D、E25A、E25R、E25V、E25S或E25Y突变,或在另一腺苷脱氨酶中的相应突变。

[0349] 在一些实施方式中,该腺苷脱氨酶包括在TadA参考序列中的R26X突变,或在另一腺苷脱氨酶中的相应突变,其中X表示在野生型腺苷脱氨酶中相应氨基酸以外的任何氨基酸。在一些实施方式中,该腺苷脱氨酶包括在TadA参考序列中的R26G、R26N、R26Q、R26C、R26L或R26K突变,或在另一腺苷脱氨酶中的相应突变。

[0350] 在一些实施方式中,该腺苷脱氨酶包括在TadA参考序列中的R107X突变,或在另一腺苷脱氨酶中的相应突变,其中X表示在野生型腺苷脱氨酶中相应氨基酸以外的任何氨基酸。在一些实施方式中,该腺苷脱氨酶包括在TadA参考序列中的R107P、R07K、R107A、R107N、R107W、R107H或R107S突变,或在另一腺苷脱氨酶中的相应突变。

[0351] 在一些实施方式中,该腺苷脱氨酶包括在TadA参考序列中的A142X突变,或在另一腺苷脱氨酶中的相应突变,其中X表示在野生型腺苷脱氨酶中相应氨基酸以外的任何氨基酸。在一些实施方式中,该腺苷脱氨酶包括在TadA参考序列中的A142N、A142D、A142G突变,或在另一腺苷脱氨酶中的相应突变。

[0352] 在一些实施方式中,该腺苷脱氨酶包括在TadA参考序列中的A143X突变,或在另一腺苷脱氨酶中的相应突变,其中X表示在野生型腺苷脱氨酶中相应氨基酸以外的任何氨基酸。在一些实施方式中,该腺苷脱氨酶包括在TadA参考序列中的A143D、A143G、A143E、A143L、A143W、A143M、A143S、A143Q和/或A143R突变,或在另一腺苷脱氨酶中的相应突变。

[0353] 在一些实施方式中,该腺苷脱氨酶包括在TadA参考序列中的一个或多个H36X、N37X、P48X、I49X、R51X、M70X、N72X、D77X、E134X、S146X、Q154X、K157X和/或K161X突变,或在另一腺苷脱氨酶中的一个或多个相应突变,其中X的存在表示在野生型腺苷脱氨酶中相应氨基酸以外的任何氨基酸。在一些实施方式中,该腺苷脱氨酶包括在TadA参考序列中的一个或多个H36L、N37T、N37S、P48T、P48L、I49V、R51H、R51L、M70L、N72S、D77G、E134G、S146R、S146C、Q154H、K157N和/或K161T突变,或在另一腺苷脱氨酶中的一个或多个相应突变。

[0354] 在一些实施方式中,该腺苷脱氨酶包括在TadA参考序列中的H36X突变,或在另一腺苷脱氨酶中的相应突变,其中X表示在野生型腺苷脱氨酶中相应氨基酸以外的任何氨基酸。在一些实施方式中,该腺苷脱氨酶包括在TadA参考序列中的H36L突变,或在另一腺苷脱氨酶中的相应突变。

[0355] 在一些实施方式中,该腺苷脱氨酶包括在TadA参考序列中的N37X突变,或在另一腺苷脱氨酶中的相应突变,其中X表示在野生型腺苷脱氨酶中相应氨基酸以外的任何氨基酸。在一些实施方式中,该腺苷脱氨酶包括在TadA参考序列中的N37T或N37S突变,或在另一腺苷脱氨酶中的相应突变。

[0356] 在一些实施方式中,该腺苷脱氨酶包括在TadA参考序列中的P48X突变,或在另一腺苷脱氨酶中的相应突变,其中X表示在野生型腺苷脱氨酶中相应氨基酸以外的任何氨基酸。在一些实施方式中,该腺苷脱氨酶包括在TadA参考序列中的P48T或P48L突变,或在另一腺苷脱氨酶中的相应突变。

[0357] 在一些实施方式中,该腺苷脱氨酶包括在TadA参考序列中的R51X突变,或在另一腺苷脱氨酶中的相应突变,其中X表示在野生型腺苷脱氨酶中相应氨基酸以外的任何氨基酸。在一些实施方式中,该腺苷脱氨酶包括在TadA参考序列中的R51H,或R51L突变,或在另一腺苷脱氨酶中的相应突变。

[0358] 在一些实施方式中,该腺苷脱氨酶包括在TadA参考序列中的S146X突变,或在另一腺苷脱氨酶中的相应突变,其中X表示在野生型腺苷脱氨酶中相应氨基酸以外的任何氨基酸。在一些实施方式中,该腺苷脱氨酶包括在TadA参考序列中的S146R或S146C突变,或在另一腺苷脱氨酶中的相应突变。

[0359] 在一些实施方式中,该腺苷脱氨酶包括在TadA参考序列中的K157X突变,或在另一腺苷脱氨酶中的相应突变,其中X表示在野生型腺苷脱氨酶中相应氨基酸以外的任何氨基酸。在一些实施方式中,该腺苷脱氨酶包括在TadA参考序列中的K157N突变,或在另一腺苷脱氨酶中的相应突变。

[0360] 在一些实施方式中,该腺苷脱氨酶包括在TadA参考序列中的P48X突变,或在另一腺苷脱氨酶中的相应突变,其中X表示在野生型腺苷脱氨酶中相应氨基酸以外的任何氨基酸。在一些实施方式中,该腺苷脱氨酶包括在TadA参考序列中的P48S、P48T或P48A突变,或在另一腺苷脱氨酶中的相应突变。

[0361] 在一些实施方式中,该腺苷脱氨酶包括在TadA参考序列中的A142X突变,或在另一

腺苷脱氨酶中的相应突变,其中X表示在野生型腺苷脱氨酶中相应氨基酸以外的任何氨基酸。在一些实施方式中,该腺苷脱氨酶包括在TadA参考序列中的A142N突变,或在另一腺苷脱氨酶中的相应突变。

[0362] 在一些实施方式中,该腺苷脱氨酶包括在TadA参考序列中的W23X突变,或在另一腺苷脱氨酶中的相应突变,其中X表示在野生型腺苷脱氨酶中相应氨基酸以外的任何氨基酸。在一些实施方式中,该腺苷脱氨酶包括在TadA参考序列中的W23R或W23L突变,或在另一腺苷脱氨酶中的相应突变。

[0363] 在一些实施方式中,该腺苷脱氨酶包括在TadA参考序列中的R152X突变,或在另一腺苷脱氨酶中的相应突变,其中X表示在野生型腺苷脱氨酶中相应氨基酸以外的任何氨基酸。在一些实施方式中,该腺苷脱氨酶包括在TadA参考序列中的R152P或R52H突变,或在另一腺苷脱氨酶中的相应突变。

[0364] 在一个实施方式中,该腺苷脱氨酶可能包括突变H36L,R51L,L84F,A106V,D108N,H123Y,S 146C,D147Y,E155V,I156F和K157N。在一些实施方式中,该腺苷脱氨酶包括以下相对于TadA参考序列的突变組合,其中组合的每个突变用"_"分开并且每个突变的组合都在括号之间:(A106V_D108N)、(R107C_D108N)、(H8Y_D108N_S 127S_D147Y_Q154H)、(H8Y_R24W_D108N_N127S_D147Y_E155V)、(D108N_D147Y_E155V)、(H8Y_D108N_S 127S)、(H8Y_D108N_N127S_D147Y_Q154H)、(A106V D108N D147Y E155V)(D108QD147Y E155V)(D108M_D147Y_E155V)、(D108L_D147Y_E155V)、(D108K_D147Y_E155V)、(D108I_D147Y_E155V),(D108F_D147Y_E155V)、(A106V_D108N_D147Y)、(A106V_D108M_D147Y_E155V),(E59A_A106V_D108N_D147Y_E155V)、(E59A cat dead_A106V_D108N_D147Y_E155V)、(L84F_A106V_D108N_H123Y_D147Y_E155V_I156Y)、(L84F_A106V_D108N_H123Y_D147Y_E155V_I156F)、(D103A_DO14N)、(G22P_D 103A_D 104N)、(G22P_D 103A_D 104N_S 138A)、(D 103A_D 104N_S138A),(R26G_L84F_A106V_R107H_D108N_H123Y_A142N_A143D_D147Y_E155V_I156F),(E25G_R26G_L84F_A106V_R107H_D108N_H123Y_A142N_A143D_D147Y_E155V_I156F)、(E25D_R26G_L84F_A106V_R107K_D108N_H123Y_A142N_A143G_D147Y_E155V_I156F)、(R26Q_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F)、(E25M_R26G_L84F_A106V_R107P_D108N_H123Y_A142N_A143D_D147Y_E155V_I15 6F)、(R26C_L84F_A106V_R107H_D108N_H123Y_A142N_D147Y_E155V_I156F)、(L84F_A106V_D108N_H123Y_A142N_A143L_D147Y_E155V_I156F)、(R26G_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F)、(E25A_R26G_L84F_A106V_R107N_D108N_H123Y_A142N_A143E_D147Y_E155V_I156F)、(R26G_L84F_A106V_R107H_D108N_H123Y_A142N_A143D_D147Y_E155V_I156F),(A106V_D108N_A142N_D147Y_E155V)、(R26G_A106V_D108N_A142N_D147Y_E155V)、(E25D_R26G_A106V_R107K_D108N_A142N_A143G_D147Y_E155V)、(R26G_A106V_D108N_R107H_A142N_A143D_D147Y_E155V)、(E25D_R26G_A106V_D108N_A142N_D147Y_E155V)、(A106V_R107K_D108N_A142N_D147Y_E155V)、(A106V_D108N_A142N_A143G_D147Y_E155V)、(A106V_D108N_A142N_A143L_D147Y_E155V)、(H36L_R51L_L84F_A106V_D108N_H123Y_S 146C_D147Y_E155V_I156F_K157N),(N37T_P48T_M70L_L84F_A106V_D108N_H123Y_D147Y_I49V_E155V_I156F),(N37S_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_K161T)、(H36L_L84F_A106V_D108N_H123Y_D147Y_Q154H_E155V_I156F)、(N72S_L84F_A106V_D108N_H123Y_S

146R_D147Y_E155V_I156F)、(H36L_P48L_L84F_A106V_D108N_H123Y_E134G_D147Y_E155V_I156F)、57N)、(H36L_L84F_A106V_D108N_H123Y_S 146C_D147Y_E155V_I156F)、(L84F_A106V_D108N_H123Y_S 146R_D147Y_E155V_I156F_K161T)、(N37S_R51H_D77G_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F)、(R51L_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_K157N)、(D24G_Q71R_L84F_H96L_A106V_D108N_H123Y_D147Y_E155V_I156F_K160E),(H36L_G67V_L84F_A106V_D108N_H123Y_S146T_D147Y_E155V_I156F)、(Q71L_L84F_A106V_D108N_H123Y_L137M_A143E_D147Y_E155V_I156F)、(E25G_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_Q159L)、(L84F_A91T_F104I_A106V_D108N_H123Y_D147Y_E155V_I156F)、(N72D_L84F_A106V_D108N_H123Y_G125A_D147Y_E155V_I156F)、(P48S_L84F_S97C_A106V_D108N_H123Y_D147Y_E155V_I156F)、(W23G_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F)、(D24G_P48L_Q71R_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_Q159L),(L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F)、(H36L_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_E155V_I156F_K157N)、(N37S_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F_K161T),(L84F_A106V_D108N_D147Y_E155V_I156F)、(R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N_K161T),(L84F_A106V_D108N_H123Y_S 146C_D147Y_E155V_I156F_K161T)、(L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N_K160E_K161T),(L84F_A106V_D108N_H123Y_S 146C_D147Y_E155V_I156F_K157N_K160E)、(R74QL84F_A106V_D108N_H123Y_D147Y_E155V_I156F)、(R74A_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F)、(L84F_A106V_D108N_H123Y_D147Y_E155V_I156F)、(R74Q_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F)、(L84F_R98Q_A106V_D108N_H123Y_D147Y_E155V_I156F)、(L84F_A106V_D108N_H123Y_R129Q_D147Y_E155V_I156F)、(P48S_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F)、(P48S_A142N),(P48T_I49V_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F_L157N),(P48T_I49V_A142N)、(H36L_P48S_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N)、(H36L_P48S_R51L_L84F_A106V_D108N_H123Y_S146C_A142N_D147Y_E155V_I156F(H36L_P48T_I49V_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N)、(H36L_P48T_I49V_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_E155V_I156F_K157N)、(H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N)、(H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_E155V_I156F_K157N)、(H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_A142N_D147Y_E155V_I156F_K157N)、(W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N)、(W23R_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N)、(W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146R_D147Y_E155V_I156F_K161T)、(H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152H_E155V_I156F_K157N)、(H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152P_E155V_I156F_K157N)、(W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152P_E155V_I156F_K157N)、(W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142A_S146C_D147Y_E155 V_I156F_K157N)、(W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142A_S146C_D147Y_R152P_E155V_I156F_K157N)、(W23L_H36L_P48A_R51L_L84F_A106V_D108N_

H123Y_S146R_D147Y_E155V_I156F_K161T)、(W23R_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152P_E155V_I156F_K157N)、(H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_R152P_E155 V_I156F_K157N)。

[0365] 胞苷脱氨酶

[0366] 在一个实施方式中,本发明的融合蛋白包括胞苷脱氨酶。在一些实施方式中,本文所提供的胞苷脱氨酶能够将胞嘧啶或5-甲基胞嘧啶脱氨基为尿嘧啶或胸腺嘧啶。在一些实施方式中,本文所提供的胞苷脱氨酶能够使DNA中的胞嘧啶脱氨基。该胞苷脱氨酶可衍生自任何适合的生物体。在一些实施方式中,该胞苷脱氨酶是天然-存在的胞苷脱氨酶,其包含一个或多个与本文所提供的任何突变相对应的突变。本领域技术人员将能够,例如通过序列比对和同源残基的确定,鉴别出任何同源蛋白质中的相应残基。因此,本领域技术人员将能够在任何天然-存在的胞苷脱氨酶中生成与本文所述的任何突变相对应的突变。在一些实施方式中,该胞苷脱氨酶来自原核生物。在一些实施方式中,该胞苷脱氨酶来自细菌。在一些实施方式中,该胞苷脱氨酶来自哺乳动物(例如人类)。

[0367] 在一些实施方式中,该胞苷脱氨酶包括的氨基酸序列,与本文所列出的任一胞苷脱氨酶氨基酸序列,是至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或至少99.5%相同的。应当理解,本文所提供的胞苷脱氨酶可包含一个或多个突变(例如,本文所提供的任何突变)。本公开揭露所提供的任何脱氨酶结构域,均具有一定百分比的同一性,再加上本文所述的任何突变或其组合。在一些实施方式中,该胞苷脱氨酶包括的氨基酸序列,与参考序列或本文所提供的任何胞苷脱氨酶相比,具有1、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、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50个或更多个突变。在一些实施方式中,该胞苷脱氨酶包括的氨基酸序列,与本领域已知或本文所述的任一氨基酸序列相比,具有至少5、至少10、至少15、至少20、至少25、至少30、至少35、至少40、至少45、至少50、至少60、至少70、至少80、至少90、至少100、至少110、至少120、至少130、至少140、至少150、至少160或至少170个相同的接连的氨基酸残基。

[0368] 本发明的融合蛋白包括核酸编辑结构域。在一些实施方式中,该核酸编辑结构域可以催化C变U的碱基变化。在一些实施方式中,该核酸编辑结构域是脱氨酶结构域。在一些实施方式中,该脱氨酶是胞苷脱氨酶或腺苷脱氨酶。在一些实施方式中,该脱氨酶是载脂蛋白B mRNA-编辑复合体(APOBEC)家族脱氨酶。在一些实施方式中,该脱氨酶是APOBEC1脱氨酶。在一些实施方式中,该脱氨酶是APOBEC2脱氨酶。在一些实施方式中,该脱氨酶是APOBEC3脱氨酶。在一些实施方式中,该脱氨酶是APOBEC3 A脱氨酶。在一些实施方式中,该脱氨酶是APOBEC3B脱氨酶。在一些实施方式中,该脱氨酶是APOBEC3C脱氨酶。在一些实施方式中,该脱氨酶是APOBEC3D脱氨酶。在一些实施方式中,该脱氨酶是APOBEC3E脱氨酶。在一些实施方式中,该脱氨酶是APOBEC3F脱氨酶。在一些实施方式中,该脱氨酶是APOBEC3G脱氨酶。在一些实施方式中,该脱氨酶是APOBEC3H脱氨酶。在一些实施方式中,该脱氨酶是APOBEC4脱氨酶。在一些实施方式中,该脱氨酶是活化-诱导的脱氨酶(AID)。在一些实施方式中,该脱氨酶是脊椎动物脱氨酶。在一些实施方式中,该脱氨酶是无脊椎动物脱氨酶。在一些实施方式中,该脱氨酶是人类、黑猩猩、大猩猩、猴、牛、狗、大鼠或小鼠脱氨酶。在一些

实施方式中,该脱氨酶是人类脱氨酶。在一些实施方式中,该脱氨酶是大鼠脱氨酶,例如rAPOBEC1。在一些实施方式中,该脱氨酶是海七鳃鳗胞苷脱氨酶1(pmCDA1)。在一些实施方式中,该脱氨酶是人类APOBEC3G。在一些实施方式中,该脱氨酶是人类APOBEC3G的片段。在一些实施方式中,该脱氨酶是包括D316R D317R突变的人类APOBEC3G变体。在一些实施方式中,该脱氨酶是人类APOBEC3G的片段并且包括对应于所述D316R D317R突变的突变。在一些实施方式中,该核酸编辑结构域,与本文所述任何脱氨酶的脱氨酶结构域,是至少80%、至少85%、至少90%、至少92%、至少95%、至少96%、至少97%、至少98%、至少99%或至少99.5%相同的。

[0369] 在某些实施方式中,本文所提供的融合蛋白包括一个或多个改善该融合蛋白的碱基编辑活性的特性。例如,本文所提供的任何融合蛋白可能包括具有减低的核酸酶活性的Cas9结构域。在一些实施方式中,本文所提供的任何融合蛋白可能具有无核酸酶活性的Cas9结构域(dCas9),或切割双股螺旋DNA分子其中一股的Cas9结构域(称为Cas9切口酶(nCas9))。

[0370] 核碱基编辑器的Cas9结构域

[0371] 在某些方面,核酸可编程的DNA结合蛋白(napDNAbp)是Cas9结构域。本文提供了非限制性的示例性Cas9结构域。该Cas9结构域可能是核酸酶活性Cas9结构域、核酸酶无活性的Cas9结构域(dCas9)或Cas9切口酶(nCas9)。在一些实施方式中,该Cas9结构域是核酸酶活性结构域。例如,该Cas9结构域可能是切割双股螺旋核酸(例如,双股螺旋DNA分子的两股)的两股的Cas9结构域。在一些实施方式中,该Cas9结构域包括本文所列出的任一氨基酸序列。在一些实施方式中,该Cas9结构域包括的氨基酸序列,与本文所列出的任一氨基酸序列相比,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96%、至少97%、至少98%、至少99%或至少99.5%相同。在一些实施方式中,该Cas9结构域包括的氨基酸序列,与本文所列出的任一氨基酸序列相比,具有1、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、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50个或更多个突变。在一些实施方式中,该Cas9结构域包括的氨基酸序列,与本文所列出的任一氨基酸序列相比,具有至少10、至少15、至少20、至少30、至少40、至少50、至少60、至少70、至少80、至少90、至少100、至少150、至少200、至少250、至少300、至少350、至少400、至少500、至少600、至少700、至少800、至少900、至少1000、至少1100或至少1200个相同的接连的氨基酸残基。

[0372] 在一些实施方式中,该Cas9结构域是核酸酶-无活性的Cas9结构域(dCas9)。例如,该dCas9结构域可结合到双股螺旋核酸分子上(例如,经由gRNA分子),而不切开该双股螺旋核酸分子的任一股。在一些实施方式中,该核酸酶-无活性的dCas9结构域包括本文所列出的氨基酸序列的D10X突变和H840X突变,或本文所提供的任何氨基酸序列中的相应突变,其中X为任何氨基酸变化。在一些实施方式中,该核酸酶-无活性的dCas9结构域包括本文所列出的氨基酸序列的D10A突变和H840A突变,或本文所提供的任何氨基酸序列中的相应突变。作为一个实例,核酸酶-无活性的Cas9结构域包括克隆载体pPlatTET-gRNA2(登录号BAV54124)中所列出的氨基酸序列。

[0373] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVD

STDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSR
RLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNL
SDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRI
PYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNEL
TKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG
KTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMG
RHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL
DINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAE
RGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINN
YHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGE
IRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYG
GFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELEN
GRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILAD
ANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRI
DLSQLGGD(参见例如,Qi等,“将CRISPR再利用为以供基因表达的序列-特异性控制的RNA-导向的平台”Cell.2013;152(5):1173-83,其全部内容通过引用并入本文)。

[0374] 基于本公开揭露和本领域的知识,额外的适合的核酸酶-无活性的dCas9结构域对本领域技术人员而言将是显而易见的,并且在本公开揭露的范围内。此类额外的示例性的适合的核酸酶-无活性的Cas9结构域包含但不限于,D10A/H840A,D10A/D839A/H840A,以及D10A/D839A/H840A/N863A突变结构域(参见例如,Prashant等,CAS9transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering.,Nature Biotechnology.2013;31(9):833-838,其全部内容通过引用并入本文)。在一些实施方式中,该dCas9结构域包括的氨基酸序列,与本文所提供的任一dCas9结构域相比,是至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或至少99.5%相同。在一些实施方式中,该Cas9结构域包括的氨基酸序列,与本文所列出的任一氨基酸序列相比,具有1、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、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50个或更多个突变。在一些实施方式中,该Cas9结构域包括的氨基酸序列,与本文所列出的任一氨基酸序列相比,具有至少10、至少15、至少20、至少30、至少40、至少50、至少60、至少70、至少80、至少90、至少100、至少150、至少200、至少250、至少300、至少350、至少400、至少500、至少600、至少700、至少800、至少900、至少1000、至少1100或至少1200个相同的接连的氨基酸残基。

[0375] 在一些实施方式中,该Cas9结构域是Cas9切口酶。该Cas9切口酶可能是Cas9蛋白,其能够仅切割双股螺旋核酸分子(例如,双股螺旋DNA分子)的一股。在一些实施方式中,该Cas9切口酶切开双股螺旋核酸分子的靶标股,这意味着该Cas9切口酶切开了与gRNA(例如sgRNA)碱基配对(与其互补)的那一股,而所述gRNA是与该Cas9结合的。在一些实施方式中,

Cas9切口酶包括D10A突变,并在840位置处具有组氨酸。在一些实施方式中,该Cas9切口酶切开双股螺旋核酸分子的非-靶标、非-碱基-被编辑的那一股,这意味着该Cas9切口酶切开了未与gRNA(例如sgRNA)碱基配对的那一股,而该gRNA是与Cas9结合的。在一些实施方式中,Cas9切口酶包括H840A突变并在位置10处具有天冬氨酸残基,或相应的突变。在一些实施方式中,该Cas9切口酶包括的氨基酸序列,与本文所提供的任一Cas9切口酶相比,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96%、至少97%、至少98%、至少99%或至少99.5%相同。基于本公开揭露和本领域的知识,额外的适合的Cas9切口酶对本领域技术人员而言将是显而易见的,并且在本公开揭露的范围内。

[0376] PAM排他性减低的Cas9结构域

[0377] 通常,Cas9蛋白,诸如来自化脓性链球菌的Cas9(spCas9),需要正准的NGG PAM序列以结合特定的核酸区域,其中"NGG"中的"N"是腺苷(A),胸苷(T),或胞嘧啶(C),而所述G为鸟苷。这可能会限制在基因组中编辑目的碱基的能力。在一些实施方式中,可能需要将本文所提供的碱基编辑融合蛋白放置在精确的区位,例如包括PAM上游的靶碱基的区域。参见例如,Komor,A.C.等,"无双-股DNA切开的基因组DNA中的靶碱基的可编程的编辑"Nature 533,420-424(2016),其全部内容在此通过引用并入。因此,在一些实施方式中,本文所提供的任何融合蛋白都可能含有Cas9结构域,该Cas9结构域能够结合不含有正准的(例如NGG)PAM序列的核苷酸序列。结合非-正准PAM序列的Cas9结构域已经在本领域中描述,并且对本领域熟练技术人员而言是显而易见的。例如,结合非-正准PAM序列的Cas9结构域已在如下文献中描述:Kleinstiver,B.P.等,"Engineered CRISPR-Cas9 nucleases with altered PAM specificities"Nature 523,481-485(2015);以及Kleinstiver,B.P.等,"Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition"NatureBiotechnology 33,1293-1298(2015);其每个的全部内容在此通过引用并入。如下的表1中描述了几种PAM变体:

[0378] 表1.Cas9蛋白和相应的PAM序列

[0379]

变体	PAM
spCas9	NGG
spCas9-VRQR	NGA
spCas9-VRER	NGCG
xCas9 (sp)	NGN
saCas9	NNGRRT
saCas9-KKH	NNNRRT
spCas9-MQKSER	NGCG
spCas9-MQKSER	NGCN
spCas9-LRKIQK	NGTN

[0380]

spCas9-LRVSQK	NGTN
spCas9-LRVSQL	NGTN
Cpf1	5’ (TTTV)

[0381] 在一些实施方式中,该Cas9结构域是来自金黄色葡萄球菌的Cas9结构域(SaCas9)。在一些实施方式中,该SaCas9结构域是核酸酶活性的SaCas9,核酸酶无活性的SaCas9(SaCas9d),或SaCas9切口酶(SaCas9n)。在一些实施方式中,该SaCas9包括N579A突变,或本文所提供的任一氨基酸序列中的相应突变。

[0382] 在一些实施方式中,该SaCas9结构域,该SaCas9d结构域,或该SaCas9n结构域可以结合具有非-正准PAM的核酸序列。在一些实施方式中,该SaCas9结构域、该SaCas9d结构域或该SaCas9n结构域可以结合具有NNGRRT PAM序列的核酸序列。在一些实施方式中,该SaCas9结构域包括E781X、N967X和R1014X突变中的一个或多个,或本文所提供的任何氨基酸序列中的相应突变,其中X是任一氨基酸。在一些实施方式中,该SaCas9结构域包括E781K,N967K和R1014H突变中的一个或多个,或者本文所提供的任何氨基酸序列中的一个或多个相应突变。在一些实施方式中,该SaCas9结构域包括E781K,N967K或R1014H突变,或本文所提供的任何氨基酸序列中的相应突变。

[0383] 示例性的SaCas9序列

[0384] KRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLK
RRRRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRG
VHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYV
KEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHC
TYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLK
QIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTI
YQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAI
FNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIE
LAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYS
LEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEE**N**SKKGNRTPFQYLSSSDS
KISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLM
NLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKE
WKKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDK
[0385] KPNRELINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTY
QKLKLIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDY
PNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKIS
NQAEFIASFYNNDLIKINGELYRVIGVNNDLLNRIEVNMIDITYREYLENMNDKRPPRII
KTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG

[0386] 上方的残基N579,其带有底线和以粗体显示,可对其突变(例如,变为A579)以产生SaCas9切口酶。

[0387] 示例性的SaCas9n序列

KRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLK
RRRRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRG
VHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYV
KEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHC
TYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLK
QIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTI
YQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAI
FNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIE

[0388] LAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYS
LEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEE**A**SKKGNRTPFQYLSSSDS
KISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLM
NLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKE
WKKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDK
KPNRELINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTY
QKLKLIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDY
PNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKIS
NQAEFIASFYNNDLIKINGELYRVIGVNNDLLNRIEVNMIDITYREYLENMNDKRPPRII

[0389] KTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG

[0390] 上方的残基A579,其可以由N579突变而来以产生SaCas9切口酶,带有底线和以粗体显示。

[0391] 示例性的SaKKH Cas9

[0392] KRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLK
RRRRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRG
VHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYV
KEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHC
TYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLK
QIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTI
YQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAI
FNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIE
LAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYS
LEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEE**A**SKKGNRTPFQYLSSSDS
KISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLM
NLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKE
WKKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDK
KPNR*A*LINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTY
QKLKLIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDY
PNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKIS
NQAEFIASFY*K*NDLIKINGELYRVIGVNNDLLNRIEVNMIDITYREYLENMNDKRPP*H*II
KTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG.

[0393] 上方的残基A579,其可以由N579突变而来以产生SaCas9切口酶,带有底线和以粗体显示。上方的残基K781,K967,和H1014,其可以由E781,N967,和R1014突变而来以产生SaKKH Cas9,带有底线和以斜体显示。

[0394] 在一些实施方式中,该Cas9结构域是化脓性链球菌的Cas9结构域(SpCas9)。在一些实施方式中,该SpCas9结构域是核酸酶活性的SpCas9,核酸酶无活性的SpCas9(SpCas9d),或SpCas9切口酶(SpCas9n)。在一些实施方式中,该SpCas9包括D9X突变,或本文所提供的任何氨基酸序列中的相应突变,其中X是除了D以外的任何氨基酸。在一些实施方式中,该SpCas9包括D9A突变,或本文所提供的任何氨基酸序列中的相应突变。在一些实施方式中,该SpCas9结构域、该SpCas9d结构域或该SpCas9n结构域可以结合具有非-正准PAM的核酸序列。在一些实施方式中,该SpCas9结构域、该SpCas9d结构域或SpCas9n结构域可以结合具有NGG、NGA或NGCG PAM序列的核酸序列结构域。在一些实施方式中,该SpCas9结构域包括D1134X、R1334X和T1336X突变中的一个或多个,或本文所提供的任何氨基酸序列中的相应突变,其中X是任一氨基酸。在一些实施方式中,该SpCas9结构域包括D1134E、R1334Q和T1336R突变中的一个或多个,或本文所提供的任何氨基酸序列中的相应突变。在一些实施方式中,该SpCas9结构域包括D1134E,R1334Q和T1336R突变,或本文所提供的任何氨基酸序列中的相应的复数个突变。在一些实施方式中,该SpCas9结构域包括D1134X,R1334X和

T1336X突变中的一个或多个,或本文所提供的任何氨基酸序列中的相应突变,其中X是任一氨基酸。在一些实施方式中,该SpCas9结构域包括D1134V、R1334Q和T1336R突变中的一个或多个,或本文所提供的任何氨基酸序列中的相应突变。在一些实施方式中,该SpCas9结构域包括D1134V、R1334Q和T1336R突变,或本文所提供的任何氨基酸序列中的相应的复数个突变。在一些实施方式中,该SpCas9结构域包括D1134X、G1217X、R1334X和T1336X突变中的一个或多个,或本文所提供的任何氨基酸序列中的相应突变,其中X是任一氨基酸。在一些实施方式中,该SpCas9结构域包括D1134V、G1217R、R1334Q和T1336R突变中的一个或多个,或本文所提供的任何氨基酸序列中的相应突变。在一些实施方式中,该SpCas9结构域包括D1134V、G1217R、R1334Q和T1336R突变,或本文所提供的任何氨基酸序列中的相应的复数个突变。

[0395] 在一些实施方式中,本文所提供的任何融合蛋白的Cas9结构域包括的氨基酸序列,与本文所述的Cas9多肽相比,是至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或至少99.5%相同。在一些实施方式中,本文所提供的任何融合蛋白的Cas9结构域包括本文所述的任何Cas9多肽的氨基酸序列。在一些实施方式中,本文所提供的任何融合蛋白的Cas9结构域由本文所述的任何Cas9多肽的氨基酸序列组成。

[0396] 示例性的SpCas9

[0397] DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR
RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTD
KADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLEN
LIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAIL
LSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKP
ILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPL
ARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE
GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDN
EENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSD
GFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEM
ARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDH
IVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFI
KRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVG
TALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGE
IVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAK
VEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNE
LALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPI
REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[0398] 示例性的SpCas9n

[0399] DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR
RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTD
KADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLEN

LIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAIL
LSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKP
ILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPL
ARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE
GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDN
EENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSD
GFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEM
ARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDH
IVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFI
KRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVG
TALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGE
IVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAK
VEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNE
LALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPI
REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[0400] 示例性的SpEQR Cas9

[0401]
DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGE
TAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHP
IFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPD
NSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGL
FGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLS
DAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN
GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLG
ELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWN

FEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKP
AFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL
LKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYT
GWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQG
DSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNS
RERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYD
VDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF

[0402] DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVIT
LKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYD
VRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKG
RDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGF**E**SP
TVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIK
LPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTN
LGAPAAFKYFDTTIDRK**Q**Y**R**STKEVLDATLIHQSITGLYETRIDLSQLGGD

[0403] 上述的残基E1134,Q1334,和R1336,其可以由D1134、R1334和T1336突变而来以产生SpEQR Cas9,带有底线和以粗体显示。

[0404] 示例性的SpVQR Cas9

DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGE
TAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHP
IFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPD
NSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGL
FGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLS

[0405] DAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN
GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLG
ELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWN
FEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKP
AFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL
LKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYT
GWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQG

DSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNS
RERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYD
VDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF
DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVIT
LKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYD

[0406] VRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKG
RDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGF**V**SP
TVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIK
LPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTN
LGAPAAFKYFDTTIDRK**Q**Y**R**STKEVLDATLIHQSITGLYETRIDLSQLGGD

[0407] 上述的残基V1134,Q1334,和R1336,其可以由D1134,R1334,和T1336突变而来以产生SpVQR Cas9,带有底线和以粗体显示。

[0408] 示例性的SpVRER Cas9

DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGE
TAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHP
IFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPD
NSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGL
FGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLS
DAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN
GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLG
ELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWN

[0409] FEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKP
AFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL
LKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYT
GWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQG
DSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNS
RERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYD
VDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF
DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVIT

[0410] LKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYD
VRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKG
RDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGF**V**SP
TVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIK
LPKYSLFELENGRKRMLASA**R**ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTN
LGAPAAFKYFDTTIDRK**E**Y**R**STKEVLDATLIHQSITGLYETRIDLSQLGGD.

[0411] 上述的残基V1134,R1217,E1334,和R1336,其可以由D1134,G1217,R1334,和T1336突变而来以产生SpVRER Cas9,带有底线和以粗体显示。

[0412] 该Cas9核酸酶具有两个功能性核酸内切酶结构域:RuvC和HNH。一经与靶标结合,Cas9经历构象变化,其使核酸酶结构域就定位以切开靶DNA的相对股。Cas9-介导的DNA切开的最终结果是靶DNA内(PAM序列上游约3-4个核苷酸)的双股断裂(DSB)。然后通过两种一般修复途径之一修复所得的DSB:(1)有效率但容易出错的非同源末端连接(NHEJ)途径;或(2)效率较低但高保真的同源定向修复(HDR)途径。

[0413] 非同源末端连接(NHEJ)和/或同源定向修复(HDR)的“效率”可以通过任何方便的方法来计算。例如,在某些情况下,效率可以用成功的HDR的百分比来表示。例如,surveyor核酸酶测定法可用于生成切开产物,而产物与底物的比率可用于计算该百分比。例如,surveyor核酸酶可用于直接切开含有新整合的限制性内切酶识别序列(作为成功的HDR的成果)的DNA。被切开的底物越多,表示HDR百分比越高(HDR的效率越高)。作为说明性的例子,可以使用以下等式[(切开产物)/(底物加上切开产物)]计算HDR的份额(百分比)(例如,(b+c)/(a+b+c),其中“a”是DNA底物的条带强度,而“b”和“c”是切开产物)。

[0414] 在某些情况下,效率可以用成功的NHEJ的百分比来表示。例如,T7核酸内切酶I测定法可用于生成切开产物,而产物与底物的比率可用于计算NHEJ的百分比。T7核酸内切酶I切开错配的异源双股螺旋DNA,其发生于野生和突变DNA股的杂交(NHEJ在原始断裂位点生成小的随机插入或缺失(插入/缺失))。切开越多,表示NHEJ的百分比越高(NHEJ的效率越高)。作为说明性的例子,可以使用以下等式计算NHEJ的份额(百分比):$(1-(1-(b+c)/(a+b+c))^{1/2})\times 100$,其中“a”是DNA底物的条带强度,而“b”和“c”是切开产物(Ran等,Cell.2013Sep.12;154(6):1380-9;以及Ran等,Nat Protoc.2013Nov.;8(11):2281-2308)。

[0415] 所述NHEJ修复途径是最活跃的修复机制,而它经常在DSB位点引起小的核苷酸插入或缺失(插入/缺失)。NHEJ-介导的DSB修复的随机性具有重要的实际意义,因为表达Cas9及gRNA或向导多核苷酸的细胞群体会导致多种多样的突变。在大多数情况下,NHEJ在DNA中引起小的插入/缺失,其导致氨基酸的缺失、插入或移码突变,而这些突变导致被靶向的基因的开放阅读框(ORF)内的提前的终止密码子。理想的最终结果是被靶向的基因内的丧失-功能型突变。

[0416] NHEJ-介导的DSB修复经常中断基因的开放阅读框,而同源定向修复(HDR)可用于生成特定的核苷酸变化,其范围从单核苷酸的变化到大的插入,如添加荧光团或标签。

[0417] 为了将HDR用于基因编辑,可以将含有目标序列的DNA修复模板与(单个或多个)

gRNA和Cas9或Cas9切口酶一起递送到感兴趣的细胞类型中。该修复模板可以含有所期望的编排,以及紧接靶标的上游和下游的额外的同源序列(称为左&右同源臂)。每个同源臂的长度可以取决于将被引入的变化的大小,其中较大的插入需要更长的同源臂。该修复模板可以是单-股寡核苷酸,双-股寡核苷酸或双-股DNA质粒。即使在表达Cas9、gRNA和外源性修复模板的细胞中,HDR的效率通常也是低的(被修饰的等位基因<10%)。由于HDR发生在细胞周期的S和G2阶段,因此可以通过对细胞进行同步化来提高HDR的效率。化学性或遗传性地抑制涉及NHEJ的基因也可以增加HDR频率。

[0418] 在一些实施方式中,Cas9是经修饰的Cas9。给定的gRNA靶向序列可以在整个基因组中具有存在部分同源性的额外的位点。这些位点称为脱靶位点,并在设计gRNA时需要被考虑进去。除了优化gRNA设计外,还可以通过对Cas9进行修饰来提高CRISPR特异性。Cas9通过两个核酸酶结构域RuvC和HNH的合并的活性以生成双-股断裂(DSB)。Cas9切口酶,SpCas9的D10A突变,保留了一个核酸酶结构域并生成DNA切口而非DSB。该切口酶系统也可以与HDR-介导的基因编辑合并以供特定的基因编排。

[0419] 在一些情况下,Cas9是变体Cas9蛋白。当与野生型Cas9蛋白的氨基酸序列相比时,变体Cas9多肽具有的氨基酸序列有一个氨基酸的不同(例如,具有缺失、插入、替换、融合)。在一些实例中,该变体Cas9多肽具有减低该Cas9多肽的核酸酶活性的氨基酸变化(例如,缺失、插入或替换)。例如,在一些实例中,该变体Cas9多肽,具有小于50%、小于40%、小于30%、小于20%、小于10%、小于5%或小于1%的相应的野生型Cas9蛋白的核酸酶活性。在一些情况下,该变体Cas9蛋白没有实质上的核酸酶活性。当一受试Cas9蛋白是不具有实质上的核酸酶活性的变体Cas9蛋白时,其可以被称为“dCas9”。

[0420] 在一些情况下,变体Cas9蛋白具有减低的核酸酶活性。例如,变体Cas9蛋白展现出小于约20%、小于约15%、小于约10%、小于约5%、小于约1%或小于约0.1%的野生型Cas9蛋白(例如野生型Cas9蛋白)的核酸内切酶活性。

[0421] 在一些情况下,变体Cas9蛋白可以切开向导靶序列的互补股,但是其切开双股向导靶序列的非-互补股的能力是减低的。例如,该变体Cas9蛋白可能具有减低RuvC结构域功能的突变(氨基酸替换)。作为非限制性的例子,在一些实施方式中,Cas9变体蛋白具有D10A(在氨基酸位置10的天冬氨酸变为丙氨酸)突变,并因此可以切开双股向导靶序列的互补股,但是其切开双股向导靶序列的非-互补股的能力是减低的(因此,当该变体Cas9蛋白切开双股靶核酸时,会得到单股断裂(SSB)而不是双股断裂(DSB))(参见例如Jinek等,Science.2012Aug.17;337(6096):816-21)。

[0422] 在一些情况下,变体Cas9蛋白可以切开双股向导靶序列的非-互补股,但是其切开向导靶序列的互补股的能力是减低的。例如,该变体Cas9蛋白可能具有减低HNH结构域(RuvC/HNH/RuvC结构域基序)功能的突变(氨基酸替换)。作为非限制性的例子,在一些实施方式中,该变体Cas9蛋白具有H840A(在氨基酸位置840处的组氨酸变为丙氨酸)突变,并因此可以切开向导靶序列的非-互补股,但是其切开向导靶序列的互补股的能力是减低的(因此,当该变体Cas9蛋白切开双股靶核酸时,会得到SSB而不是DSB)。此种Cas9蛋白切开向导靶序列(例如,单股向导靶序列)的能力是减低的,但是保留了结合向导靶序列(例如,单股向导靶序列)的能力。

[0423] 在一些情况下,变体Cas9蛋白切开双股靶DNA的互补股和非-互补股的能力均减

低。作为非限制性的例子,在一些实施方式中,该变体Cas9蛋白带有D10A和H840A两个突变,使得该多肽切开双股靶DNA的互补股和非-互补股的能力均减低。此种Cas9蛋白切开靶DNA(例如,单股靶DNA)的能力是减低的,但是保留了结合靶DNA(例如,单股靶DNA)的能力。

[0424] 作为另一个非限制性的例子,在一些情况下,该变体Cas9蛋白带有W476A和W1126A突变,使得该多肽切开靶DNA的能力是减低的。此种Cas9蛋白切开靶DNA(例如,单股靶DNA)的能力是减低的,但是保留了结合靶DNA(例如,单股靶DNA)的能力。

[0425] 作为另一个非限制性的例子,在一些情况下,该变体Cas9蛋白带有P475A,W476A,N477A,D1125A,W1126A,和D1127A突变,使得该多肽切开靶DNA的能力是减低的。此种Cas9蛋白切开靶DNA(例如,单股靶DNA)的能力是减低的,但是保留了结合靶DNA(例如,单股靶DNA)的能力。

[0426] 作为另一个非限制性的例子,在一些情况下,该变体Cas9蛋白带有H840A,W476A,和W1126A突变,使得该多肽切开靶DNA的能力是减低的。此种Cas9蛋白切开靶DNA(例如,单股靶DNA)的能力是减低的,但是保留了结合靶DNA(例如,单股靶DNA)的能力。作为另一个非限制性的例子,在一些情况下,该变体Cas9蛋白带有H840A,D10A,W476A,和W1126A突变,使得该多肽切开靶DNA的能力是减低的。此种Cas9蛋白切开靶DNA(例如,单股靶DNA)的能力是减低的,但是保留了结合靶DNA(例如,单股靶DNA)的能力。在一些实施方式中,该变体Cas9已在Cas9 HNH结构域的位置840处复原了催化性的His残基(A840H)。

[0427] 作为另一个非限制性的例子,在一些情况下,该变体Cas9蛋白带有H840A,P475A,W476A,N477A,D1125A,W1126A,和D1127A突变,使得该多肽切开靶DNA的能力是减低的。此种Cas9蛋白切开靶DNA(例如,单股靶DNA)的能力是减低的,但是保留了结合靶DNA(例如,单股靶DNA)的能力。作为另一个非限制性的例子,在一些情况下,该变体Cas9蛋白带有D10A,H840A,P475A,W476A,N477A,D1125A,W1126A,和D1127A突变,使得该多肽切开靶DNA的能力是减低的。此种Cas9蛋白切开靶DNA(例如,单股靶DNA)的能力是减低的,但是保留了结合靶DNA(例如,单股靶DNA)的能力。在一些情况下,当变体Cas9蛋白带有W476A和W1126A两个突变时,或当该变体Cas9蛋白带有P475A,W476A,N477A,D1125A,W1126A,和D1127A突变时,该变体Cas9蛋白不能有效地结合到PAM序列上。因此,在一些此类情况下,当在结合的方法中使用此种变体Cas9蛋白时,该方法不需要PAM序列。换句话说,在一些情况下,当在结合的方法中使用此种变体Cas9蛋白时,该方法可以包含向导RNA,但该方法可以在不存在PAM序列的情况下执行(而结合的特异性因此是由向导RNA的靶向节段所提供)。可以使其他残基突变以获得上述效果(即,灭活一个或另外一个核酸酶部分)。作为非限制性的例子,残基D10,G12,G17,E762,H840,N854,N863,H982,H983,A984,D986,和/或A987可以被改变(即,被替换)。同样地,除丙氨酸替换以外的突变也是适合的。

[0428] 在一些实施方式中,具有减低的催化活性的变体Cas9蛋白(例如,当Cas9蛋白具有D10,G12,G17,E762,H840,N854,N863,H982,H983,A984,D986,和/或A987突变时,例如D10A,G12A,G17A,E762A,H840A,N854A,N863A,H982A,H983A,A984A,和/或D986A),只要它保留与向导RNA相互作用的能力,该变体Cas9蛋白仍能以位点-特异性的方式与靶DNA结合(因为它仍可通过向导RNA被导向至靶DNA序列)。

[0429] 在一些实施方式中,该变体Cas蛋白可以是spCas9,spCas9-VRQR,spCas9-VRER,xCas9(sp),saCas9,saCas9-KKH,spCas9-MQKSER,spCas9-LRKIQK,或spCas9-LRVSQL。

[0430] 化脓性链球菌Cas9的替代物可以包含来自Cpf1家族的RNA-导向的核酸内切酶,其在哺乳动物细胞中显示出切开活性。来自普雷沃氏菌(Prevotella)和弗朗西斯菌1(Francisella 1)(CRISPR/Cpf1)的CRISPR是一种类似于CRISPR/Cas9系统的DNA-编辑技术。Cpf1是II类CRISPR/Cas系统的RNA-导向的核酸内切酶。在普雷沃氏菌和弗朗西斯菌细菌中发现了这种获得性免疫机制。Cpf1基因与CRISPR基因座相联,并编码一种核酸内切酶,该核酸内切酶使用向导RNA找到并切开病毒DNA。Cpf1是一种比Cas9小且简单的核酸内切酶,克服了CRISPR/Cas9系统的一些局限性。与Cas9核酸酶不同,Cpf1-介导的DNA切开的结果是带有短3'悬垂的双股断裂。Cpf1的参差的切开模式开辟了定向基因移转(类似于传统的限制性内切酶克隆)的可能性,而定向基因移转可以提高基因编辑的效率。

[0431] 像上述的Cas9变体和异种同源物一样,Cpf1也可以将可被CRISPR靶向的位点(数目)扩展到缺乏SpCas9所偏好的NGG PAM位点的AT-富集区域或AT-富集基因组。该Cpf1基因座含有混合的α/β结构域,RuvC-I,其后为螺旋区域、RuvC-II和锌指-样结构域。该Cpf1蛋白具有RuvC-样核酸内切酶结构域,其与Cas9的RuvC结构域相似。此外,Cpf1不具有HNH核酸内切酶结构域,并且Cpf1的N末端不具有Cas9的α-螺旋识别叶。Cpf1 CRISPR-Cas结构域的结构显示出Cpf1有独特的功能,其被归类为2类V型CRISPR系统。该Cpf1基因座编码的Cas1,Cas2和Cas4蛋白与II型系统相比更类似于I型和III型。功能性的Cpf1不需要反式-激活CRISPR RNA(tracrRNA),因此只需要CRISPR(crRNA)。这有利于基因组编辑,不仅因为Cpf1比Cas9小,而且其具有较小的sgRNA分子(约是Cas9核苷酸数的一半)。与Cas9所靶向的G-富集的PAM相反,该Cpf1-crRNA复合体通过识别原间隔序列临近基序5'-YTN-3'而切开靶DNA或RNA。在识别PAM之后,Cpf1引入一个粘性-末端-样的DNA双-股断裂,其具有4或5个核苷酸的悬垂。

[0432] 包括两个napDNAbp和一个脱氨酶结构域的融合蛋白

[0433] 本公开揭露的一些方面提供了融合蛋白,其包括具有切口酶活性的napDNAbp结构域(例如,nCas结构域)和催化无活性的napDNAbp(例如,dCas结构域)以及核碱基编辑器(例如,腺苷脱氨酶结构域、胞苷脱氨酶结构域),其中至少所述两个napDNAbp结构域是通过一个连接子连接的。应当理解,该Cas结构域可能是本文所提供的任何Cas结构域或Cas蛋白(例如,dCas9和nCas9)。在一些实施方式中,任何Cas结构域、DNA结合蛋白结构域或Cas蛋白包含但不限于Cas9(例如,dCas9和nCas9)、Cas12a/Cpf1、Cas12b/C2c1、Cas12c/C2c3、Cas12d/CasY、Cas12e/CasX、Cas12g、Cas12h和Cas12i。具有与Cas9不同的PAM特异性的可编程的多核苷酸-结合蛋白的一个例子是来自普雷沃氏菌和弗朗西斯菌的规律成簇间隔短回文重复序列(Cpf1)。与Cas9类似,Cpf1也是2类CRISPR效应器。例如,但不限于,在一些实施方式中,该融合蛋白包括以下结构,其中脱氨酶是腺苷脱氨酶或胞苷脱氨酶:

[0434] NH_2-[脱氨酶]-[nCas结构域]-[dCas结构域]-COOH;

[0435] NH_2-[脱氨酶]-[dCas结构域]-[nCas结构域]-COOH;

[0436] NH_2-[nCas结构域]-[dCas结构域]-[脱氨酶]-COOH;

[0437] NH_2-[dCas结构域]-[nCas结构域]-[脱氨酶]-COOH;

[0438] NH_2-[nCas结构域]-[脱氨酶]-[dCas结构域]-COOH;

[0439] NH_2-[dCas结构域]-[脱氨酶]-[nCas结构域]-COOH;

[0440] 在一些实施方式中,上述通用架构中使用的"-"表示可选的连接子的存在。在一些

实施方式中,脱氨酶和napDNAbp(例如,Cas结构域)不通过连接子序列连接,而是直接融合。在一些实施方式中,连接子存在于脱氨酶结构域和napDNAbp之间。在一些实施方式中,脱氨酶或其他核碱基编辑器直接融合到dCas上并且连接子连接dCas和nCas9。在一些实施方式中,脱氨酶和napDNAbp是经由本文所提供的任何连接子融合。例如,在一些实施方式中,脱氨酶和napDNAbp是经由下文标题为"连接子"的部分中所提供的任何连接子融合。在一些实施方式中,dCas结构域和脱氨酶紧邻并且nCas结构域是经由连接子而连接到这些结构域上(5'或3')。

[0441] 具有内部插入的融合蛋白

[0442] 本公开揭露提供了融合蛋白,其包括与核酸可编程的核酸结合蛋白(例如napDNAbp)融合的异源性多肽。异源性多肽可以是在原生或野生型napDNAbp多肽序列中未发现的多肽。该异源性多肽可以在napDNAbp的C末端、N末端与napDNAbp融合,或插入到napDNAbp的一内部区位。在一些实施方式中,该异源性多肽插入到napDNAbp的一内部区位。

[0443] 在一些实施方式中,该异源性多肽是脱氨酶或其功能性片段。例如,融合蛋白可以包括两侧翼为Cas9多肽的N末端片段和C末端片段的脱氨酶。该融合蛋白中的脱氨酶可以是胞苷脱氨酶。该融合蛋白中的脱氨酶可以是腺苷脱氨酶。

[0444] 该脱氨酶可以是环状变换排列体脱氨酶。例如,该脱氨酶可以是环状变换排列体腺苷脱氨酶或环状变换排列体胞苷脱氨酶。在一些实施方式中,该脱氨酶是环状变换排列体TadA,其环状重排是位于氨基酸残基116(如TadA参考序列中所编号的)。在一些实施方式中,该脱氨酶是环状变换排列体TadA,其环状重排是位于氨基酸残基136(如TadA参考序列中所编号的)。在一些实施方式中,该脱氨酶是环状变换排列体TadA,其环状重排是位于氨基酸残基65(如TadA参考序列中所编号的)。

[0445] 该融合蛋白可以包括多于一个脱氨酶。该融合蛋白可以包括,例如,1、2、3、4、5或更多个脱氨酶。在一些实施方式中,该融合蛋白包括一个脱氨酶。在一些实施方式中,该融合蛋白包括两个脱氨酶。融合蛋白中的两个或更多个脱氨酶可以是腺苷脱氨酶、胞苷脱氨酶或其组合。该两个或更多个脱氨酶可以是同二聚体。该两个或更多个脱氨酶可以是异二聚体。该两个或更多个脱氨酶可以串联插入到napDNAbp中。在一些实施方式中,该两个或更多个脱氨酶在napDNAbp中可能不是串联的。

[0446] 在一些实施方式中,该融合蛋白中的napDNAbp是Cas9多肽或其片段。该Cas9多肽可以是变体Cas9多肽。在一些实施方式中,该Cas9多肽是Cas9切口酶(nCas9)多肽或其片段。在一些实施方式中,该Cas9多肽是核酸酶无活性的Cas9(dCas9)多肽或其片段。该融合蛋白中的Cas9多肽可以是全长Cas9多肽。在一些情况下,该融合蛋白中的Cas9多肽可能不是全长Cas9多肽。相对于天然存在的Cas9蛋白,该Cas9多肽可以是被截短的,例如在N末端或C末端。该Cas9多肽可以是经环状重排的Cas9蛋白。

[0447] 该Cas9多肽可以是Cas9多肽的片段、部分或结构域,其仍然能够结合靶多核苷酸和向导核酸序列。

[0448] 在一些实施方式中,该Cas9多肽是化脓性链球菌Cas9(SpCas9),金黄色葡萄球菌Cas9(SaCas9),嗜热链球菌1Cas9(St1Cas9),或其片段或变体。

[0449] 融合蛋白的Cas9多肽可以包括的氨基酸序列,与天然存在的Cas9多肽相比,是至少85%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少

97％、至少98％、至少99％或至少99.5％相同的。

[0450] 融合蛋白的Cas9多肽可以包括的氨基酸序列,与SEQ ID NO:1中所列出的Cas9氨基酸序列相比,是至少85％、至少90％、至少91％、至少92％、至少93％、至少94％、至少95％、至少96％、至少97％、至少98％、至少99％或至少99.5％相同的。

[0451] 所述异源性多肽(例如,脱氨酶)可以在适合的区位插入到napDNAbp(例如,Cas9)中,使得该napDNAbp保留了其结合靶多核苷酸和向导核酸的能力。脱氨酶可以插入到napDNAbp中,而不损害该脱氨酶(例如,碱基编辑活性)或该napDNAbp(例如,结合靶核酸和向导核酸的能力)的功能。脱氨酶可以在例如无序区域或包括高温因子或B-因子的区域处(如晶体学研究所示)插入到napDNAbp中。较不有序、无序或非结构性的蛋白质区域,例如溶剂曝露区域和环,可以用于插入而不损害结构或功能。脱氨酶可以插入到napDNAbp的柔性环区域或溶剂-曝露区域中。在一些实施方式中,该脱氨酶插入到Cas9多肽的柔性环区域中。

[0452] 在一些实施方式中,脱氨酶的插入区位是通过Cas9多肽的晶体结构的B-因子分析来确定的。在一些实施方式中,该脱氨酶插入到包括高于平均B-因子的Cas9多肽区域中(例如,与包括所述无序区域的总蛋白质或蛋白质结构域相比,更高的B-因子)。B因子或温度因子可以表示原子发自其平均位置的波动(例如,由于温度-依赖的原子振动或晶格中的静态无序)。主干原子的高B-因子(例如,高于平均B-因子)可以表示具有相对较高局部移动性的区域。此种区域可用于插入脱氨酶而不损害结构或功能。脱氨酶可以插入到具有Cα原子的残基的区位中,其B-因子与总蛋白质的平均B-因子相比高出50％、60％、70％、80％、90％、100％、110％、120％、130％、140％、150％、160％、170％、180％、190％、200％或大于200％。脱氨酶可以插入到具有Cα原子的残基的区位中,其B-因子与包括该残基的Cas9蛋白结构域的平均B-因子相比高出50％、60％、70％、80％、90％、100％、110％、120％、130％、140％、150％、160％、170％、180％、190％、200％或大于200％。包括高于平均B-因子的Cas9多肽位置可以包含,例如,残基768、792、1052、1015、1022、1026、1029、1067、1040、1054、1068、1246、1247和1248(如SEQ ID NO:1中所编号的)。包括高于平均B-因子的Cas9多肽区域可以包含,例如,残基792-872、792-906和2-791(如SEQ ID NO:1中所编号的)。

[0453] 异源性多肽(例如,脱氨酶)可以插入到napDNAbp中的选自于由以下所组成的群组的氨基酸残基处:768、791、792、1015、1016、1022、1023、1026、1029、1040、1052、1054、1067、1068、1069、1246、1247和1248(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。在一些实施方式中,该异源性多肽插入到以下氨基酸位置之间:768-769、791-792、792-793、1015-1016、1022-1023、1026-1027、1029-1030、1040-1041、1052-1053、1054-1055、1067-1068、1068-1069、1247-1248或1248-1249(如SEQ ID NO:1中所编号的)抑或其相应的氨基酸位置。在一些实施方式中,该异源性多肽插入到以下氨基酸位置之间:769-770、792-793、793-794、1016-1017、1023-1024、1027-1028、1030-1031、1041-1042、1053-1054、1055-1056、1068-1069、1069-1070、1248-1249,或1249-1250(如SEQ ID NO:1中所编号的)抑或其相应的氨基酸位置。在一些实施方式中,该异源性多肽替换了选自于由以下所组成的群组的氨基酸残基:768、791、792、1015、1016、1022、1023、1026、1029、1040、1052、1054、1067、1068、1069、1246、1247和1248(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。应当理解,关于插入位置而对SEQ ID NO:1的参照是为了说明的目

的。如本文所讨论的插入不限于SEQ ID NO:1的Cas9多肽序列,而是包含在变体Cas9多肽(例如Cas9切口酶(nCas9)、核酸酶无活性的Cas9(dCas9)、缺少核酸酶结构域的Cas9变体、截短的Cas9或缺少部分或全部HNH结构域的Cas9结构域)中相应区位的插入。

[0454] 异源性多肽(例如,脱氨酶)可以插入到napDNAbp中的选自于由以下所组成的群组的氨基酸残基处:768、792、1022、1026、1040、1068和1247(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。在一些实施方式中,该异源性多肽插入到以下氨基酸位置之间:768-769、792-793、1022-1023、1026-1027、1029-1030、1040-1041、1068-1069或1247-1248(如SEQ ID NO:1中所编号的)抑或其相应的氨基酸位置。在一些实施方式中,该异源性多肽插入到以下氨基酸位置之间:769-770、793-794、1023-1024、1027-1028、1030-1031、1041-1042、1069-1070或1248-1249(如SEQ ID NO:1中所编号的)抑或其相应的氨基酸位置。在一些实施方式中,该异源性多肽替换了选自于由以下所组成的群组的氨基酸残基:768、792、1022、1026、1040、1068和1247(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。

[0455] 异源性多肽(例如,脱氨酶)可以插入到napDNAbp中的图4,图5,图6,或图7中所示的氨基酸残基处,抑或另一个Cas9多肽中相应的氨基酸残基。异源性多肽(例如,脱氨酶)可以插入到napDNAbp中的选自于由以下所组成的群组的氨基酸残基处:1002、1003、1025、1052-1056、1242-1247、1061-1077、943-947、686-691、569-578、530-539和1060-1077(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。该脱氨酶可以插入到该残基的N端或C端,抑或取代该残基。在一些实施方式中,该脱氨酶插入到该残基的C端。

[0456] 在一些实施方式中,ABE(例如,TadA)插入到选自于由以下所组成的群组的氨基酸处:1015、1022、1029、1040、1068、1247、1054、1026、768、1067、1248、1052和1246(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。在一些实施方式中,ABE(例如,TadA)插入以替代了残基792-872、792-906或2-791(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。在一些实施方式中,该ABE插入到选自于由以下所组成的群组的氨基酸的N端处:1015、1022、1029、1040、1068、1247、1054、1026、768、1067、1248、1052和1246(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。在一些实施方式中,该ABE插入到选自于由以下所组成的群组的氨基酸的C端处:1015、1022、1029、1040、1068、1247、1054、1026、768、1067、1248、1052和1246(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。在一些实施方式中,该ABE插入以替代了选自于由以下所组成的群组的氨基酸:1015、1022、1029、1040、1068、1247、1054、1026、768、1067、1248、1052和1246(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。

[0457] 在一些实施方式中,CBE(例如,APOBEC1)插入到选自于由以下所组成的群组的氨基酸处:1016、1023、1029、1040、1069和1247(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。在一些实施方式中,该ABE插入到选自于由以下所组成的群组的氨基酸的N端处:1016、1023、1029、1040、1069和1247(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。在一些实施方式中,该ABE插入到选自于由以下所组成的群组的氨基酸的C端处:1016、1023、1029、1040、1069和1247(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。在一些实施方式中,该ABE插入以替代了选

自于由以下所组成的群组的氨基酸:1016、1023、1029、1040、1069和1247(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。

[0458] 在一些实施方式中,该脱氨酶插入到氨基酸残基768处(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。在一些实施方式中,该ABE插入到氨基酸768的N端处(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。在一些实施方式中,该ABE插入到氨基酸768的C端处(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。在一些实施方式中,该ABE插入以替代氨基酸768(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。

[0459] 在一些实施方式中,该脱氨酶插入到氨基酸残基791处(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。在一些实施方式中,该ABE插入到氨基酸791的N端处(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。在一些实施方式中,该ABE插入到氨基酸791的C端处(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。在一些实施方式中,该ABE插入以替代氨基酸791(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。

[0460] 在一些实施方式中,该脱氨酶插入到氨基酸残基792处(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。在一些实施方式中,该ABE插入到氨基酸792的N端处(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。在一些实施方式中,该ABE插入到氨基酸792的C端处(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。在一些实施方式中,该ABE插入以替代氨基酸792(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。

[0461] 在一些实施方式中,该脱氨酶插入到氨基酸残基1016处(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。在一些实施方式中,该ABE插入到氨基酸1016的N端处(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。在一些实施方式中,该ABE插入到氨基酸1016的C端处(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。在一些实施方式中,该ABE插入以替代氨基酸1016(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。

[0462] 在一些实施方式中,该脱氨酶插入到氨基酸残基1022处(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。在一些实施方式中,该ABE插入到氨基酸1022的N端处(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。在一些实施方式中,该ABE插入到氨基酸1022的C端处(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。在一些实施方式中,该ABE插入以替代氨基酸1022(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。

[0463] 在一些实施方式中,该脱氨酶插入到氨基酸残基1023处(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。在一些实施方式中,该ABE插入到氨基酸1023的N端处(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。在一些实施方式中,该ABE插入到氨基酸1023的C端处(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。在一些实施方式中,该ABE插入以替代氨基酸1023(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。

[0464] 在一些实施方式中,该脱氨酶插入到氨基酸残基1026处(如SEQ ID NO:1中所编号

的),抑或另一个Cas9多肽中相应的氨基酸残基。在一些实施方式中,该ABE插入到氨基酸1026的N端处(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。在一些实施方式中,该ABE插入到氨基酸1026的C端处(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。在一些实施方式中,该ABE插入以替代氨基酸1026(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。

[0465] 在一些实施方式中,该脱氨酶插入到氨基酸残基1029处(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。在一些实施方式中,该ABE插入到氨基酸1029的N端处(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。在一些实施方式中,该ABE插入到氨基酸1029的C端处(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。在一些实施方式中,该ABE插入以替代氨基酸1029(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。

[0466] 在一些实施方式中,该脱氨酶插入到氨基酸残基1040处(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。在一些实施方式中,该ABE插入到氨基酸1040的N端处(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。在一些实施方式中,该ABE插入到氨基酸1040的C端处(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。在一些实施方式中,该ABE插入以替代氨基酸1040(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。

[0467] 在一些实施方式中,该脱氨酶插入到氨基酸残基1052处(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。在一些实施方式中,该ABE插入到氨基酸1052的N端处(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。在一些实施方式中,该ABE插入到氨基酸1052的C端处(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。在一些实施方式中,该ABE插入以替代氨基酸1052(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。

[0468] 在一些实施方式中,该脱氨酶插入到氨基酸残基1054处(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。在一些实施方式中,该ABE插入到氨基酸1054的N端处(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。在一些实施方式中,该ABE插入到氨基酸1054的C端处(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。在一些实施方式中,该ABE插入以替代氨基酸1054(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。

[0469] 在一些实施方式中,该脱氨酶插入到氨基酸残基1067处(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。在一些实施方式中,该ABE插入到氨基酸1067的N端处(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。在一些实施方式中,该ABE插入到氨基酸1067的C端处(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。在一些实施方式中,该ABE插入以替代氨基酸1067(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。

[0470] 在一些实施方式中,该脱氨酶插入到氨基酸残基1068处(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。在一些实施方式中,该ABE插入到氨基酸1068的N端处(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。在一些实施方式中,该ABE插入到氨基酸1068的C端处(如SEQ ID NO:1中所编号的),抑或另一

个Cas9多肽中相应的氨基酸残基。在一些实施方式中,该ABE插入以替代氨基酸1068(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。

[0471] 在一些实施方式中,该脱氨酶插入到氨基酸残基1069处(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。在一些实施方式中,该ABE插入到氨基酸1069的N端处(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。在一些实施方式中,该ABE插入到氨基酸1069的C端处(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。在一些实施方式中,该ABE插入以替代氨基酸1069(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。

[0472] 在一些实施方式中,该脱氨酶插入到氨基酸残基1246处(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。在一些实施方式中,该ABE插入到氨基酸1246的N端处(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。在一些实施方式中,该ABE插入到氨基酸1246的C端处(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。在一些实施方式中,该ABE插入以替代氨基酸1246(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。

[0473] 在一些实施方式中,该脱氨酶插入到氨基酸残基1247处(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。在一些实施方式中,该ABE插入到氨基酸1247的N端处(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。在一些实施方式中,该ABE插入到氨基酸1247的C端处(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。在一些实施方式中,该ABE插入以替代氨基酸1247(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。

[0474] 在一些实施方式中,该脱氨酶插入到氨基酸残基1248处(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。在一些实施方式中,该ABE插入到氨基酸1248的N端处(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。在一些实施方式中,该ABE插入到氨基酸1248的C端处(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。在一些实施方式中,该ABE插入以替代氨基酸1248(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。

[0475] 在一些实施方式中,异源性多肽(例如,脱氨酶)插入到Cas9多肽的柔性环中。该柔性环部分可以选自于由以下所组成的群组:530-537、569-570、686-691、943-947、1002-1025、1052-1077、1232-1247或1298-1300(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。该柔性环部分可以选自于由以下所组成的群组:1-529、538-568、580-685、692-942、948-1001、1026-1051、1078-1231或1248-1297(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。

[0476] 异源性多肽(例如,脱氨酶)可以插入到对应于如下氨基酸残基的Cas9多肽区域中:1017-1069、1242-1247、1052-1056、1060-1077、1002-1003、943-947、530-537、568-579、686-691,1242-1247、1298-1300、1066-1077、1052-1056或1060-1077(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。

[0477] 异源性多肽(例如,脱氨酶)可以插入以替换Cas9多肽的缺失区域。该缺失区域可以对应于Cas9多肽的N末端或C末端部分。在一些实施方式中,该缺失区域对应于残基792-872(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。在一些实施

方式中,该缺失区域对应于残基792-906(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。在一些实施方式中,该缺失区域对应于残基2-791(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。在一些实施方式中,该缺失区域对应于残基1017-1069(如SEQ ID NO:1中所编号的),抑或其相应的氨基酸残基。

[0478] 异源性多肽(例如,脱氨酶)可以插入到Cas9多肽的结构性或功能性结构域内。异源性多肽(例如,脱氨酶)可以插入到Cas9多肽的两个结构性或功能性结构域之间。异源性多肽(例如,脱氨酶)可以插入以替代Cas9多肽的结构性或功能性结构域,例如,在从该Cas9多肽中删除上述结构域之后。所述Cas9多肽的结构性或功能性结构域可以包含,例如,RuvC I,RuvC II,RuvC III,Rec1,Rec2,PI,或HNH。

[0479] 在一些实施方式中,该Cas9多肽缺少选自于由以下所组成的群组中的一个或多个结构域:RuvC I,RuvC II,RuvC III,Rec1,Rec2,PI,或HNH结构域。在一些实施方式中,该Cas9多肽缺少核酸酶结构域。在一些实施方式中,该Cas9多肽缺少HNH结构域。在一些实施方式中,该Cas9多肽缺少部分的HNH结构域,使得该Cas9多肽具有减低的或消除的HNH活性。

[0480] 在一些实施方式中,该Cas9多肽包括核酸酶结构域的缺失并且脱氨酶被插入以替代该核酸酶结构域。在一些实施方式中,该HNH结构域被删除掉并且所述脱氨酶插入到该HNH结构域的位置中。在一些实施方式中,一个或多个RuvC结构域被删除掉并且所述脱氨酶被插入到该RuvC结构域的位置中。

[0481] 包括异源性多肽的融合蛋白的两侧翼是napDNAbp的N末端和C末端片段。在一些实施方式中,该融合蛋白所包括的脱氨酶的两侧翼是Cas9多肽的N末端片段和C末端片段。该N末端片段或该C末端片段可以结合靶多核苷酸序列。该N末端片段的C端或该C末端片段的N端可以包括Cas9多肽的柔性环的一部分。该N末端片段的C端或该C末端片段的N端可以包括Cas9多肽的α-螺旋结构的一部分。该N末端片段或该C末端片段可以包括DNA结合结构域。该N末端片段或该C末端片段可以包括RuvC结构域。该N末端片段或该C末端片段可以包括HNH结构域。在一些实施方式中,该N末端片段和该C末端片段均不包括HNH结构域。

[0482] 在一些实施方式中,当该融合蛋白使靶核碱基脱氨基时,N末端Cas9片段的C端包括靠近靶核碱基的氨基酸。在一些实施方式中,当该融合蛋白使靶核碱基脱氨基时,C末端Cas9片段的N端包括靠近靶核碱基的氨基酸。不同脱氨酶的插入区位可以不同,以使得靶核碱基与N末端Cas9片段的C端中的或者C末端Cas9片段的N端中的氨基酸是靠近的。例如,ABE的插入位置可以是在选自以下所组成的群组中的氨基酸残基处:1015、1022、1029、1040、1068、1247、1054、1026、768、1067、1248、1052和1246(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。CBE的适合的插入位置可以是选自于由以下所组成的群组中的氨基酸残基:1016、1023、1029、1040、1069和1247(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。在某些实施方式中,ABE的插入可以插入到上面所列出的任一氨基酸残基的N端或C端。在一些实施方式中,ABE的插入可以插入以替代上面所列出的任一氨基酸残基。

[0483] 所述融合蛋白的N末端Cas9片段(即,在融合蛋白中位于脱氨酶侧翼的N末端Cas9片段)可以包括Cas9多肽的N端。所述融合蛋白的N末端Cas9片段可以包括长度为至少约:100、200、300、400、500、600、700、800、900、1000、1100、1200或1300个氨基酸。所述融合蛋白的N末端Cas9片段可以包括对应于以下氨基酸残基的序列:1-56、1-95、1-200、1-300、1-

400、1-500、1-600、1-700、1-718、1-765、1-780、1-906、1-918或1-1100(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。所述N末端Cas9片段可以包括的序列,相对于氨基酸残基1-56、1-95、1-200、1-300、1-400、1-500、1-600、1-700、1-718、1-765、1-780、1-906、1-918或1-1100(如SEQ ID NO:1中所编号的)抑或另一个Cas9多肽中相应的氨基酸残基,包括至少:85%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%或至少99.5%的序列同一性。

[0484] 所述融合蛋白的C末端Cas9片段(即,在融合蛋白中位于脱氨酶侧翼的C末端Cas9片段)可以包括Cas9多肽的C端。所述融合蛋白的C末端Cas9片段可以包括长度为至少约:100、200、300、400、500、600、700、800、900、1000、1100、1200或1300个氨基酸。所述融合蛋白的C末端Cas9片段可以包括对应于以下氨基酸残基的序列:1099-1368、918-1368、906-1368、780-1368、765-1368、718-1368、94-1368或56-1368(如SEQ ID NO:1中所编号的),抑或另一个Cas9多肽中相应的氨基酸残基。所述C末端Cas9片段可以包括的序列,相对于氨基酸残基1099-1368、918-1368、906-1368、780-1368、765-1368、718-1368、94-1368或56-1368(如SEQ ID NO:1中所编号的)抑或另一个Cas9多肽中相应的氨基酸残基,包括:至少85%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%或至少99.5%的序列同一性。

[0485] 所述融合蛋白的N末端Cas9片段和C末端Cas9片段合在一起可能并不对应于全长的天然存在的Cas9多肽序列,例如,如SEQ ID NO:1中所列出的。

[0486] 本文所述的融合蛋白可以凭藉在非靶位点(例如,脱靶的位点)处的减低的脱氨反应而实现靶向的脱氨反应,诸如减低的基因组范围的假性脱氨反应。本文所述的融合蛋白可以凭藉在非靶位点处的减低的旁观者脱氨反应而实现靶向的脱氨反应。所述不需要的脱氨反应或脱靶脱氨反应可以被减低的程度,与例如末端融合蛋白(其包括与Cas9多肽的N端或C端融合的脱氨酶)相比,为至少30%、至少40%、至少50%、至少60%、至少70%、至少80%、至少90%、至少95%或至少99%。所述不需要的脱氨反应或脱靶脱氨反应可以被减低的程度,与例如末端融合蛋白(其包括与Cas9多肽的N端或C端融合的脱氨酶)相比,为至少一倍、至少二倍、至少三倍、至少四倍、至少五倍、至少十倍、至少十五倍、至少二十倍、至少三十倍、至少四十倍、至少五十倍、至少六十倍、至少七十倍、至少八十倍、至少九十倍或至少一百倍。

[0487] 在一些实施方式中,该融合蛋白的脱氨酶在R-环范围内使不超过两个的核碱基脱氨基。在一些实施方式中,该融合蛋白的脱氨酶在R-环范围内使不超过三个的核碱基脱氨基。在一些实施方式中,该融合蛋白的脱氨酶在R-环范围内使不超过2、3、4、5、6、7、8、9或10个的核碱基脱氨基。R-环是三-股核酸结构,其包含DNA:RNA杂合体、DNA:DNA或RNA:RNA互补结构以及与单-股DNA相联的结构。如本文所用,当靶多核苷酸与CRISPR复合体或碱基编辑复合体接触时,可能形成R-环,其中向导多核苷酸(例如向导RNA)的一部分与靶多核苷酸(例如靶DNA)的一部分杂交并置换。在一些实施方式中,R-环包括间隔序列的序列和靶DNA互补序列的杂合区域。R-环区域可能是约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、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49或50个核碱基对的长度。在一些实施方式中,该R-环区域的长度是约20个核碱基对。应当理解,如本文所用,R-环区域不限于与向导多核苷酸杂交的靶DNA股。例如,

R-环区域内的靶核碱基的编辑可能针对包括向导RNA的互补股的DNA股,或者可能针对作为向导RNA的互补股的相对股的DNA股。在一些实施方式中,在R-环区域中的编辑包括了将非-互补股(原间隔序列股)上的核碱基编辑为靶DNA序列中的向导RNA。

[0488] 本文所述的融合蛋白可以在不同于正准的碱基编辑的编辑窗口中实现靶向脱氨反应。在一些实施方式中,靶核碱基是在靶多核苷酸序列中的PAM序列上游约1到约20个碱基处。在一些实施方式中,靶核碱基是在靶多核苷酸序列中的PAM序列上游约2到约12个碱基处。在一些实施方式中,靶核碱基是相距PAM序列或在PAM序列上游:约1到9个碱基对、约2到10个碱基对、约3到11个碱基对、约4到12个碱基对、约5到13个碱基对、约6到14个碱基对、约7到15个碱基对、约8到16个碱基对、约9到17个碱基对、约10到18个碱基对、约11到19个碱基对、约12到20个碱基对、约1到7个碱基对、约2到8个碱基对、约3到9个碱基对、约4到10个碱基对、约5到11个碱基对、约6到12个碱基对、约7到13个碱基对、约8到14个碱基对、约9到15个碱基对、约10到16个碱基对、约11到17个碱基对、约12到18个碱基对、约13到19个碱基对、约14到20个碱基对、约1到5个碱基对、约2到6个碱基对、约3到7个碱基对、约4到8个碱基对、约5到9个碱基对、约6到10个碱基对、约7到11个碱基对、约8到12个碱基对、约9到13个碱基对、约10到14个碱基对、约11到15个碱基对、约12到16个碱基对、约13到17个碱基对、约14到18个碱基对、约15到19个碱基对、约16到20个碱基对、约1到3个碱基对、约2到4个碱基对、约3到5个碱基对、约4到6个碱基对、约5到7个碱基对、约6到8个碱基对、约7到9个碱基对、约8到10个碱基对、约9到11个碱基对、约10到12个碱基对、约11到13个碱基对、约12到14个碱基对、约13到15个碱基对、约14到16个碱基对、约15到17个碱基对、约16到18个碱基对、约17到19个碱基对、约18到20个碱基对。在一些实施方式中,靶核碱基是相距PAM序列或在PAM序列上游大约1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20个或更多个碱基对。在一些实施方式中,靶核碱基是在PAM序列上游约1、2、3、4、5、6、7、8或9个碱基对。在一些实施方式中,靶核碱基是在PAM序列上游约2、3、4或6个碱基对。

[0489] 因此,本文也提供了融合蛋白文库和使用其优化碱基编辑的方法,该经优化的碱基编辑,与正准的碱基编辑器(例如BE4)相比,允许替代的优选碱基编辑窗口。在一些实施方式中,本公开揭露提供了用于优化碱基编辑的蛋白文库,其包括复数个融合蛋白,其中该复数个融合蛋白中的每一个都包括两侧翼为Cas9多肽的N末端片段和C末端片段的脱氨酶,其中每一个融合蛋白的N末端片段都不同于该复数个融合蛋白中其余蛋白的N末端片段,或者,其中每一个融合蛋白的C末端片段都不同于该复数个融合蛋白中其余蛋白的C末端片段,其中每一个融合蛋白的脱氨酶使靠近靶多核苷酸序列中的原间隔序列临近基序(PAM)序列的靶核碱基脱氨基,并且其中所述N末端片段或所述C末端片段结合靶多核苷酸序列。在一些实施方式中,对于CRISPR R-环内的每个核碱基,该复数个融合蛋白中的至少一个使该核碱基脱氨基。在一些实施方式中,对于相距PAM序列1到20个碱基对的靶多核苷酸内的每个核碱基,该复数个融合蛋白中的至少一个使该核碱基脱氨基。在一些实施方式中,本文提供了包括允许最优化的碱基编辑的融合蛋白文库的试剂盒。

[0490] 所述融合蛋白可以包括多于一个的异源性多肽。例如,所述融合蛋白可以额外地包括一个或多个UGI结构域和/或一个或多个核定位信号。所述两个或更多个异源性结构域可以串联插入。所述两个或更多个异源性结构域插入的区位可以使得它们在NapDNAb中不是串联的。

[0491] 融合蛋白可以包括脱氨酶和napDNAbp多肽之间的连接子。该连接子可以是肽或非-肽连接子。例如,该连接子可以是XTEN,(GGGS)n,(GGGGS)n,(G)n,(EAAAK)n,(GGS)n,SGSETPGTSESATPES。在一些实施方式中,所述融合蛋白包括N末端Cas9片段和脱氨酶之间的连接子。在一些实施方式中,所述融合蛋白包括C末端Cas9片段和脱氨酶之间的连接子。在一些实施方式中,napDNAbp的N末端和C末端片段是通过连接子而与脱氨酶连结。在一些实施方式中,该N末端和C末端片段在没有连接子的情况下连接到脱氨酶结构域上。在一些实施方式中,所述融合蛋白包括N末端Cas9片段和脱氨酶之间的连接子,但是不包括C末端Cas9片段和脱氨酶之间的连接子。在一些实施方式中,所述融合蛋白包括C末端Cas9片段和脱氨酶之间的连接子,但是不包括N末端Cas9片段和脱氨酶之间的连接子。

[0492] 示例性的TadA或TadA7.10序列如下所列出:

[0493] SEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD

[0494] GSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD

[0495] TAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDP

[0496] YRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQN

[0497] MNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPG

[0498] GSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFN

[0499] GSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQ

[0500] GSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFN

[0501] GSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD

[0502] 101 Cas9 TadAins 1015

[0503] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR

RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVL
NNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGA
AGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDYDVRKMIAKSEQEIGKATAKYFF
YSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN
SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK
DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD

[0504] 102 Cas9 TadAins 1022

[0505] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVP
VGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGV
RNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDAKSEQEIGKATAKYFF
YSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN
SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK

DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD

[0506] 103 Cas9 TadAins 1029

[0507] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRA
RDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRI
GRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGKATAKYFF
YSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN
SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK
DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD

[0508] 103 Cas9 TadAins 1040

[0509] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV

GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSGSSGSETPGTSESATPESSGSEVEFSHEYW
MRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCV
MCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSS
TDNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN
SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK
DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD

[0510] 105 Cas9 TadAins 1068

[0511] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGEGS
SGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEI
MALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGIL
ADECAALLCYFFRMPRQVFNAQKKAQSSTDTGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN
SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK
DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD

[0512] 106 Cas9 TadAins 1247

[0513] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD

NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLV
LNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTG
AAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD

[0514] 107 Cas9 TadAins 1054

[0515] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGSSGSETPGTSESATP
ESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYR
LIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRM
PRQVFNAQKKAQSSTDGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN
SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK
DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD

[0516] 108 Cas9 TadAins 1026

[0517] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE

NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDE
REVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRV
VFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDQEIGKATAKYFF
YSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN
SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK
DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD

[0518] 109 Cas9 TadAins 768

[0519] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLH
DPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHR
VEITEGILADECAALLCYFFRMPRTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRD
MYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFD
NLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKV
REINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITL
ANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPK
KYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFEL
ENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILA
DANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRI

DLSQLGGD

[0520] 110.1 Cas9 TadAins 1250

[0521] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGSPGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAV
LVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAK
TGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPREDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILA
DANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRI
DLSQLGGD

[0522] 110.2 Cas9 TadAins 1250

[0523] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA

KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGSPGSSGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPV
GAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVR
NAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPREDNEQKQLFVEQHKHYLDEIIEQISEFSKRV
ILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYE
TRIDLSQLGGD

[0524] 110.3 Cas9 TadAins 1250

[0525] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGSPGSSGSSGSETPGTSESATPESGSSSGSEVEFSHEYWMRHALTLAKRARDERE
VPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVF
GVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPREDNEQKQLFVEQHKHYLDEIIEQISEFS
KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITG
LYETRIDLSQLGGD

[0526] 110.4 Cas9 TadAins 1250

[0527] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE

MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGSPGSSGSSGSETPGTSESATPESGSSSGSEVEFSHEYWMRHALTLAKRARDERE
VPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVF
GVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMRREDNEQKQLFVEQHKHYLDEIIEQISEFS
KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITG
LYETRIDLSQLGGD

[0528] 110.5 Cas9 TadAins 1249

[0529] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGSGSSGSSGSETPGTSESATPESGSSSGSEVEFSHEYWMRHALTLAKRARDEREV
PVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFG
VRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMRRPEDNEQKQLFVEQHKHYLDEIIEQISEFS
KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITG
LYETRIDLSQLGGD

[0530] 110.5 Cas9 TadAinsΔ59-66 1250

[0531] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK

PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGSPGSSGSSGSETPGTSESATPESGSSGSEVEFSHEYWMRHALTLAKRARDEREV
PVGAVLVLNNRVIGEGWNRAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGA
AGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDEDNEQKQLFVEQHKHYLDEIIEQ
ISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH
QSITGLYETRIDLSQLGGD

[0532] 110.6 Cas9 TadAins 1251

[0533] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGSPEGSSGSSGSETPGTSESATPESGSSSGSEVEFSHEYWMRHALTLAKRARDER
EVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVV
FGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMRRDNEQKQLFVEQHKHYLDEIIEQISEFS
KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITG
LYETRIDLSQLGGD

[0534] 110.7 Cas9 TadAins 1252

[0535] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGSPEDGSSGSSGSETPGTSESATPESGSSSGSEVEFSHEYWMRHALTLAKRARDE
REVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRV
VFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMRRNEQKQLFVEQHKHYLDEIIEQISEFS
KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITG
LYETRIDLSQLGGD

[0536] 110.8 Cas9 TadAins Δ59-66 C-truncate 1250

[0537] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGSPGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAV

LVLNNRVIGEGWNRAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLM
DVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKV
LSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGG
D

[0538] 111.1 Cas9 TadAins 997

[0539] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDAT
LYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVF
NAQKKAQSSTDGSSGSETPGTSESATPESSGIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIM
NFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLI
ARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIK
LPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQIS
EFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQS
ITGLYETRIDLSQLGGD

[0540] 111.2 Cas9 TadAins 997

[0541] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF

IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDAT
LYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVF
NAQKKAQSSTDGSSGSSGSETPGTSESATPESSGGSSIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYF
FYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR
NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVK
KDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDE
IIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDA
TLIHQSITGLYETRIDLSQLGGD

[0542] 112ΔHNH TadA

[0543] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRA
IGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPG
MNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMN
TKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYD
VRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNI
VKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMER
SSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGS
PEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFK
YFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[0544] 113 N末端单一TadA螺旋截短(trunc)165-端

[0545] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLC
YFFRMPRSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTD
RHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERH
PIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYN
QLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDT
YDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKY
KEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILR
RQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKN

LPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSV
EISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRR
YTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA
IKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNE
KLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQL
LNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLV
SDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYS
NIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSD
KLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDL
IIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIE
QISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLI
HQSITGLYETRIDLSQLGGD

[0546] 114 N末端单一TadA螺旋截短165-端Δ59-65

[0547] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRTAHAEIMALRQGGLVMQNYRLI
DATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPR
SGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKN
LIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIV
DEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDN
LLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQ
SKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYP
FLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVL
PKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED
RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRL
SRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQ
TVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYL
QNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLIT
QRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDF
QFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFK
TEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKK
DWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKY
SLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSK
RVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGL
YETRIDLSQLGGD

[0548] 115.1 Cas9 TadAins1004

[0549] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRA
IGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPG
MNHRVEITEGILADECAALLCYFFRMPRQLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEIT
LANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDP
KKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE
LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVIL
ADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETR
IDLSQLGGD

[0550] 115.2 Cas9 TadAins1005

[0551] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNR
AIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
GMNHRVEITEGILADECAALLCYFFRMPRQESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEIT
LANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDP
KKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE
LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVIL
ADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETR
IDLSQLGGD

[0552] 115.3 Cas9 TadAins1006

[0553] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLEGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWN
RAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHY
PGMNHRVEITEGILADECAALLCYFFRMPRQSEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEIT
LANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDP
KKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE
LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVIL
ADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETR
IDLSQLGGD

[0554] 115.4 Cas9 TadAins1007

[0555] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGW
NRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLH
YPGMNHRVEITEGILADECAALLCYFFRMPRQEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEIT

LANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDP
KKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE
LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVIL
ADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETR
IDLSQLGGD

[0556] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRA
IGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPG
MNHRVEITEGILADECAALLCYFFRMPRQLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEIT
LANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDP
KKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE
LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVIL
ADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETR
IDLSQLGGD

[0557] 115.2 Cas9 TadAins1005

[0558] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF

IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNR
AIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
GMNHRVEITEGILADECAALLCYFFRMPRQESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEIT
LANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDP
KKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE
LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVIL
ADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETR
IDLSQLGGD

[0559] 115.3 Cas9 TadAins1006

[0560] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLEGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWN
RAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHY
PGMNHRVEITEGILADECAALLCYFFRMPRQSEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEIT
LANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDP
KKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE
LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVIL
ADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETR
IDLSQLGGD

[0561] 115.4 Cas9 TadAins1007

[0562] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT

EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGW
NRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLH
YPGMNHRVEITEGILADECAALLCYFFRMPRQEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEIT
LANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDP
KKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE
LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVIL
ADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETR
IDLSQLGGD

[0563] 116.1 Cas9 TadAins C末端截短2 792

[0564] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREV
PVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFG
VRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQSQILKEHPVENTQLQNEKLYLYYLQNGR
DMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF
DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYK
VREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEIT
LANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDP
KKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE
LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVIL
ADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETR
IDLSQLGGD

[0565] 116.2 Cas9 TadAins C-末端 截短2 791

[0566] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST

DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVP
VGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGV
RNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQGSQILKEHPVENTQLQNEKLYLYYLQNGR
DMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF
DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYK
VREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEIT
LANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDP
KKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE
LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVIL
ADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETR
IDLSQLGGD

[0567] 116.3 Cas9 TadAins C-末端 截短2 790

[0568] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKEGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPV
GAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVR
NAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQLGSQILKEHPVENTQLQNEKLYLYYLQNGR
DMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF
DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYK
VREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEIT
LANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDP
KKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE
LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVIL

ADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETR
IDLSQLGGD

[0569] 117 Cas9Δ1017-1069

[0570] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYSSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLH
DPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHR
VEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKE
SILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAK
GYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQH
KHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTST
KEVLDATLIHQSITGLYETRIDLSQLGGD

[0571] 118 Cas9 TadA-CP116ins 1067

[0572] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNMNHR
VEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAK

RARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHS
RIGRVVFGVRNAKTGAAGSLMDVLHYPGGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN
SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK
DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD

[0573] 119 Cas9 TadAins 701

[0574] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIG
EGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMD
VLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDLTFKEDIQKAQVSGQGDSLHEHIANLAGS
PAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ
NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWR
QLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSK
LVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFF
YSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN
SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK
DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD

[0575] 120 Cas9 TadACP136ins 1248

[0576] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE

MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGSMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGSETPGTS
ESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD

[0577] 121 Cas9 TadACP136ins 1052

[0578] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLAMNHRVEITEGILADECA
ALLCYFFRMPRQVFNAQKKAQSSTDGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAV
LVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAK
TGAAGSLMDVLHYPGNGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN
SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK
DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD

[0579] 122 Cas9 TadACP136ins 1041

[0580] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK

PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSMNHRVEITEGILADECAALLCYFFRMPRQV
FNAQKKAQSSTDGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWN
RAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHY
PGNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN
SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK
DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD

[0581] 123 Cas9 TadACP139ins 1299

[0582] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRMNH
RVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLA
KRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIH
SRIGRVVFGVRNAKTGAAGSLMDVLHYPGDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD

[0583] 124 Cas9Δ792-872 TadAins

[0584] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRA
IGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPG
MNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGG
LSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHA
HDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRP
LIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPT
VAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLA
SAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLS
AYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[0585] 125 Cas9Δ792-906 TadAins

[0586] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRA
IGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPG
MNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT
KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDV
RKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIV
KKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERS
SFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSP
EDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKY
FDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[0587] 126 TadA CP65ins 1003

[0588] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVL
HYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGSETPGTSESATPESSGSEVEFSHEYWM
RHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFF
YSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN
SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK
DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD

[0589] 127 TadA CP65ins 1016

[0590] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNA
KTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGSETPGTSESATPESS
GSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPYDVRKMIAKSEQEIGKATAKYFF

YSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN
SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK
DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD

[0591] 128 TadA CP65ins 1022

[0592] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMITAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRV
VFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGSETPGTSE
SATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPAKSEQEIGKATAKYFF
YSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN
SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK
DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD

[0593] 129 TadA CP65ins 1029MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNL
IGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVD
EVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPI
NASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNL
LAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQS
KNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPF
LKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLP
KHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDR
FNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLS
RKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQT
VKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQ
NGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQ

RKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQ
FYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEITAHAEIMALRQGGLVMQNY
RLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFR
MPRQVFNAQKKAQSSTDGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVI
GEGWNRAIGLHDPGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVN
IVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIME
RSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKG
SPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAF
KYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[0594] 130 TadA CP65ins 1041

[0595] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSTAHAEIMALRQGGLVMQNYRLIDATLYVTF
EPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKK
AQSSTDGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLH
DPNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN
SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK
DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD

[0596] 131 TadA CP65ins 1054

[0597] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT

EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANTAHAEIMALRQGGLVM
QNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCY
FFRMPRQVFNAQKKAQSSTDGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNN
RVIGEGWNRAIGLHDPGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN
SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK
DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD

[0598] 132 TadA CP65ins 1246

[0599] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRN
AKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGSETPGTSESATPES
SGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD

[0600] 原间隔序列临近基序

[0601] 术语“原间隔序列临近基序(PAM)”或PAM-样基序是指在CRISPR细菌适应性免疫系统中紧随被Cas9核酸酶所靶向的DNA序列后的2到6个碱基对的DNA序列。在一些实施方式

中,该PAM可以是5’PAM(即,位于原间隔序列5’末端的上游)。在其他实施方式中,该PAM可以是3’PAM(即,位于原间隔序列3’末端的下游)

[0602] 该PAM序列对于结合靶标是必需的,但是确切的序列取决于Cas蛋白的类型。

[0603] 本文所提供的碱基编辑器可以包括CRISPR蛋白-衍生的结构域,该结构域能够结合含有正准的或非-正准的原间隔序列临近基序(PAM)序列的核苷酸序列。PAM位点是靠近靶多核苷酸序列的核苷酸序列。本公开揭露的一些方面提供了碱基编辑器,其包括具有不同PAM特异性的CRISPR蛋白质的全部或部分。例如,Cas9蛋白通常,诸如化脓性链球菌的Cas9(spCas9),需要正准的NGG PAM序列以结合特定的核酸区域,其中“NGG”中的“N”是腺嘌呤(A),胸腺嘧啶(T),鸟嘌呤(G),或胞嘧啶(C),且G为鸟嘌呤。PAM可以是CRISPR蛋白-特异的,并且在不同的碱基编辑器(包括不同的CRISPR蛋白-衍生的结构域)之间可以不同。PAM可以在靶序列的5’或3’。PAM可以在靶序列的上游或下游。PAM的长度可以是1,2,3,4,5,6,7,8,9,10个或更多个核苷酸。通常,PAM的长度在2到6个核苷酸之间。表1中描述了几种PAM变体。

[0604] 在一些实施方式中,该SpCas9对于PAM核酸序列5’-NGC-3’或5’-NGG-3’具有特异性。在上述方面的各种实施方式中,该SpCas9是表1中所列出的Cas9或Cas9变体。在上述方面的各种实施方式中,该经修饰的SpCas9是spCas9-MQKFRAER。在一些实施方式中,该变体Cas蛋白可以是spCas9,spCas9-VRQR,spCas9-VRER,xCas9(sp),saCas9,saCas9-KKH,SpCas9-MQKFRAER,spCas9-MQKSER,spCas9-LRKIQK,或spCas9-LRVSQL。在一个特定的实施方式中,使用了经修饰的SpCas9,其包含氨基酸替换D1135M,S1136Q,G1218K,E1219F,A1322R,D1332A,R1335E,和T1337R(SpCas9-MQKFRAER)并对于改变过的PAM 5’-NGC-3’具有特异性。

[0605] 在一些实施方式中,该PAM是NGT。在一些实施方式中,该NGT PAM是变体。在一些实施方式中,该NGT PAM变体是通过在一个或多个残基1335,1337,1135,1136,1218,和/或1219处的靶向突变而创建的。在一些实施方式中,该NGT PAM变体是通过在一个或多个残基1219,1335,1337,1218处的靶向突变而创建的。在一些实施方式中,该NGT PAM变体是通过在一个或多个残基1135,1136,1218,1219,和1335处的靶向突变而创建的。在一些实施方式中,该NGT PAM变体选自下表2和表3中所提供的靶向突变的集合。

[0606] 表2:在残基1219,1335,1337,1218处的NGT PAM变体突变

[0607]

变体	E1219V	R1335Q	T1337	G1218
1	F	V	T	
2	F	V	R	
3	F	V	Q	
4	F	V	L	
5	F	V	T	R
6	F	V	R	R
7	F	V	Q	R
8	F	V	L	R
9	L	L	T	
10	L	L	R	

11	L	L	Q	
12	L	L	L	
13	F	I	T	
14	F	I	R	
15	F	I	Q	
16	F	I	L	
17	F	G	C	
18	H	L	N	
19	F	G	C	A
20	H	L	N	V
21	L	A	W	
22	L	A	F	
23	L	A	Y	
24	I	A	W	
25	I	A	F	
26	I	A	Y	

[0608] 表3:在残基1135,1136,1218,1219,和1335处的NGT PAM变体突变

[0609]

变体	D1135L	S1136R	G1218S	E1219V	R1335Q
27	G				
28	V				
29	I				
30		A			
31		W			
32		H			
33		K			
34			K		
35			R		
36			Q		
37			T		
38			N		
39				I	
40				A	
41				N	
42				Q	
43				G	
44				L	
45				S	
46				T	
47					L
48					I
49					V
50					N
51					S
52					T
53					F

[0610]

54					Y
55	N1286Q	I1331F			

[0611] 在一些实施方式中,该NGT PAM变体选自表2和3中的变体5,7,28,31,或36。在一些实施方式中,所述这些变体具有改良的NGT PAM识别能力。

[0612] 在一些实施方式中,所述这些NGT PAM变体在残基1219,1335,1337,和/或1218处具有突变。在一些实施方式中,该NGT PAM变体选自下表4中所提供的变体,以得到具有改良的识别能力的突变。

[0613] 表4:在残基1219,1335,1337,和1218处的NGT PAM变体突变

[0614]

变体	E1219V	R1335Q	T1337	G1218
1	F	V	T	
2	F	V	R	
3	F	V	Q	
4	F	V	L	
5	F	V	T	R
6	F	V	R	R
7	F	V	Q	R
8	F	V	L	R

[0615] 在一些实施方式中,该NGT PAM选自下表5中所提供的变体。

[0616] 表5.NGT PAM变体

[0617]

	NGTN 变体	D1135	S1136	G1218	E1219	A1322R	R1335	T1337
变体 1	LRKIQK	L	R	K	I	-	Q	K
变体 2	LRSVQK	L	R	S	V	-	Q	K
变体 3	LRSVQL	L	R	S	V	-	Q	L
变体 4	LRKIRQK	L	R	K	I	R	Q	K
变体 5	LRSVRQK	L	R	S	V	R	Q	K
变体 6	LRSVRQL	L	R	S	V	R	Q	L

[0618] 在一些实施方式中,该Cas9结构域是来自化脓性链球菌的Cas9结构域(SpCas9)。在一些实施方式中,该SpCas9结构域是核酸酶活性的SpCas9,核酸酶无活性的SpCas9(SpCas9d),或SpCas9切口酶(SpCas9n)。在一些实施方式中,该SpCas9包括D9X突变,或本文所提供的任何氨基酸序列中的相应突变,其可能与本文所提供的任何胞苷脱氨酶或腺苷脱氨酶融合。

[0619] 在一些实施方式中,该SpCas9结构域包括突变中的一个或多个D1135X、R1335X和T1336X突变,或本文所提供的任何氨基酸序列中的相应突变,其中X是任一氨基酸。在一些实施方式中,该SpCas9结构域包括突变中的一个或多个D1135E、R1335Q和T1336R突变,或本文所提供的任何氨基酸序列中的相应突变。在一些实施方式中,该SpCas9结构域包括D1135E、R1335Q和T1336R突变,或本文所提供的任何氨基酸序列中的相应突变。在一些实施方式中,该SpCas9结构域包括突变中的一个或多个D1135X、R1335X和T1336X突变,或本文所提供的任何氨基酸序列中的相应突变,其中X是任一氨基酸。在一些实施方式中,该SpCas9结构域包括突变中的一个或多个D1135V、R1335Q和T1336R突变,或本文所提供的任何氨基酸序列中的相应突变。在一些实施方式中,该SpCas9结构域包括D1135V,R1335Q,和T1336R突变,或本文所提供的任何氨基酸序列中的相应突变。在一些实施方式中,该SpCas9结构域包括突变中的一个或多个D1135X、G1217X、R1335X和T1336X突变,或本文所提供的任何氨基酸序列中的相应突变,其中X是任一氨基酸。在一些实施方式中,该SpCas9结构域包括突变中的一个或多个D1135V、G1217R、R1335Q和T1336R突变,或本文所提供的任何氨基酸序列中的相应突变。在一些实施方式中,该SpCas9结构域包括D1135V、G1217R、R1335Q和T1336R突变,或本文所提供的任何氨基酸序列中的相应突变。

[0620] 在一些实施方式中,本文所提供的任何融合蛋白的Cas9结构域包括的氨基酸序列,与本文所述的Cas9多肽相比,是至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或至少99.5%相同的。在一些实施方式中,本文所提供的任何融合蛋白的Cas9结构域包括本文所述的任何Cas9多肽的氨基酸序列。在一些实施方式中,本文所提供的任何融合蛋白的Cas9结构域由本文所述的任何Cas9多肽的氨基酸序列组成。

[0621] 在一些例子中,被本文所揭露的碱基编辑器的CRISPR蛋白-衍生的结构域所识别的PAM可以提供给细胞,且该PAM是在相对于编码该碱基编辑器的嵌入物(例如,AAV嵌入物)而言分开的一寡核苷酸上。在此种实施方式中,在分开的寡核苷酸上提供PAM可以允许靶序列的切开,否则其将不能被切开,因为在与该靶序列相同的多核苷酸上不存在临近的PAM。

[0622] 在一实施方式中,化脓性链球菌Cas9(SpCas9)可用作适于基因组工程化的CRISPR核酸内切酶。然而,可以使用其他蛋白。在一些实施方式中,可以使用不同的核酸内切酶以靶向某些基因组靶标。在一些实施方式中,可以使用具有非-NGG PAM序列的合成的SpCas9-衍生的变体。另外,已经鉴别了来自各种物种的其他Cas9异种同源物,并且这些“非-SpCas9”可以结合多种亦可用于本公开揭露的PAM序列。例如,SpCas9的相对较大的尺寸(大约4千碱基(kb)编码序列)可导致携带所述SpCas9 cDNA的质粒不能在细胞中有效表达。相反地,金黄色葡萄球菌Cas9(SaCas9)的编码序列比SpCas9短约1千碱基,可能允许其在细胞中有效表达。与SpCas9相似,该SaCas9核酸内切酶能够在体外情况下(哺乳动物细胞中)和在体内情况下(小鼠中)对靶基因进行修饰。在一些实施方式中,Cas蛋白可以靶向不同的

PAM序列。在一些实施方式中,例如,靶基因可以与Cas9 PAM(5'-NGG,例如)临近。在其他实施方式中,其他Cas9异种同源物可以具有不同的PAM要求。例如,其他PAM,诸如嗜热链球菌(对于CRISPR1为5'-NNAGAA,对于CRISPR3为5'-NGGNG)和脑膜炎双球菌(5'-NNNNGATT)的那些PAM,也可以在靶基因的临近处被找到。

[0623] 在一些实施方式中,对于化脓性链球菌系统,靶基因序列可以居先于(即,为5'于)5'-NGG PAM,而20-nt的向导RNA序列可以与相对股碱基配对以介导临近PAM的Cas9切开。在一些实施方式中,临近的切割可以是在PAM上游3个碱基对或可以是在PAM上游大约3个碱基对。在一些实施方式中,临近的切割可以是在PAM上游10个碱基对或可以是在PAM上游大约10个碱基对。在一些实施方式中,临近的切割可以是在PAM上游0-20个碱基对或可以是在PAM上游大约0至20个碱基对。例如,临近的切割可以是在紧邻PAM上游的1、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或30个碱基对。临近的切割也可以是在PAM下游的1到30个碱基对。能够结合PAM序列的示例性的SpCas9蛋白的序列如下:

[0624] 示例性的PAM-结合的SpCas9的氨基酸序列如下:

[0625] MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD.

[0626] 示例性的PAM-结合的SpCas9n的氨基酸序列如下:

[0627] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF

YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRI
PYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNEL
TKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG
KTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMG
RHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL
DINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAE
RGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINN
YHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGE
IRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYG
GFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELEN
GRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILAD
ANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRI
DLSQLGGD.

[0628] 示例性的PAM-结合的SpEQR Cas9的氨基酸序列如下：

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNL
IGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSF
FHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTD
KADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFE
ENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLG
LTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLS
DAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKY
KEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRED
LLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFR

[0629] IPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTN
FDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKA
IVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL
LKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKV
MKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLI
HDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDEL
VKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILK
EHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFL
KDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF
DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDEND

[0630] KLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTAL
IKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFK
TEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKK
TEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGF**E**SPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIK
LPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKG
SPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH
RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK**Q**Y**R**STKEVLDATL
IHQSITGLYETRIDLSQLGGD.在这个序列中,残基E1135,Q1335和R1337,其可以由D1135,R1335,和T1337突变而来以产生SpEQR Cas9,带有底线和以粗体显示。

[0631] 示例性的PAM-结合的SpVQR Cas9的氨基酸序列如下:

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNL
IGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSF
FHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTD
KADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFE
ENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLG
LTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLS
DAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKY
KEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRED
LLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFR
IPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTN
FDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKA
[0632] IVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL
LKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKV
MKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLI
HDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDEL
VKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILK
EHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFL
KDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF
DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDEND
KLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTAL
IKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFK
TEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKK
TEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGF**V**SPTVAYSVLVVA

[0633] KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIK
LPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKG
SPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH
RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK**Q**Y**R**STKEVLDATL
IHQSITGLYETRIDLSQLGGD.在这个序列中,残基V1135,Q1335,和R1336,其可以由D1135,R1335,和T1336突变而来以产生SpVQR Cas9,带有底线和以粗体显示。

[0634] 示例性的PAM-结合的SpVRER Cas9的氨基酸序列如下:

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNL
IGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSF
FHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTD
KADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFE
ENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLG
LTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLS
DAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKY
KEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRED
LLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFR
IPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTN
FDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKA
IVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL
LKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKV
[0635] MKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLI
HDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDEL
VKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILK
EHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFL
KDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF
DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDEND
KLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTAL
IKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFK
TEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKK
TEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGF**V**SPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIK
LPKYSLFELENGRKRMLASA**R**ELQKGNELALPSKYVNFLYLASHYEKLKG
SPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH
RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK**E**Y**R**STKEVLDATL
[0636] IHQSITGLYETRIDLSQLGGD.

[0637] 在一些实施方式中,该Cas9结构域是重组Cas9结构域。在一些实施方式中,该重组Cas9结构域是SpyMacCas9结构域。在一些实施方式中,该SpyMacCas9结构域是核酸酶活性的SpyMacCas9,核酸酶无活性的SpyMacCas9(SpyMacCas9d),或SpyMacCas9切口酶(SpyMacCas9n)。在一些实施方式中,该SaCas9结构域,该SaCas9d结构域,或该SaCas9n结构域可以结合具有非-正准的PAM的核酸序列。在一些实施方式中,该SpyMacCas9结构域,该SpCas9d结构域,或该SpCas9n结构域可以结合具有NAA PAM序列的核酸序列。

[0638] 示例性的SpyMacCas9

[0639] MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFGSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLADSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQIYNQLFEENPINASRVDAKAILSARLSKSRRLENLIAQLPGEKRNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNSEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGAYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDRGMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGHSLHEQIANLAGSPAIKKGILQTVKIVDELVKVMGHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFIKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEIQTVGQNGGLFDDNPKSPLEVTPSKLVPLKKELNPKKYGGYQKPTTAYPVLLITDTKQLIPISVMNKKQFEQNPVKFLRDRGYQQVGKNDFIKLPKYTLVDIGDGIKRLWASSKEIHKGNQLVVSKKSQILLYHAHHLDSDLSNDYLQNHNQQFDVLFNEIISFSKKCKLGKEHIQKIENVYSNKKNSASIEELAESFIKLLGFTQLGATSPFNFLGVKLNQKQYKGKKDYILPCTEGTLIRQSITGLYETRVDLSKIGED.

[0640] 在一些情况下,变体Cas9蛋白带有H840A,P475A,W476A,N477A,D1125A,W1126A,和D1218A突变,使得该多肽切开靶DNA或RNA的能力是减低的。此种Cas9蛋白切开靶DNA(例如,单股靶DNA)的能力是减低的,但是保留了结合靶DNA(例如,单股靶DNA)的能力。作为另一个非限制性的例子,在一些情况下,该变体Cas9蛋白带有D10A、H840A、P475A、W476A、N477A、D1125A、W1126A和D1218A突变,使得该多肽切开靶DNA的能力是减低的。

[0641] 此种Cas9蛋白切开靶DNA(例如,单股靶DNA)的能力是减低的,但是保留了结合靶DNA(例如,单股靶DNA)的能力。在一些情况下,当变体Cas9蛋白带有W476A和W1126A突变时,或者当该变体Cas9蛋白带有P475A、W476A、N477A、D1125A、W1126A和D1218A突变时,该变体Cas9蛋白不能有效地结合PAM序列。因此,在一些此类情况下,当在结合的方法中使用此种变体Cas9蛋白时,该方法不需要PAM序列。换句话说,在一些情况下,当在结合的方法中使用此种变体Cas9蛋白时,该方法可以包含向导RNA,但该方法可以在不存在PAM序列的情况下被执行(而结合的特异性因此是由该向导RNA的靶向节段所提供)。可以使其他残基突变以获得上述效果(即,灭活一个或另外一个核酸酶部分)。作为非限制性的例子,残基D10、G12、G17、E762、H840、N854、N863、H982、H983、A984、D986和/或A987可以被改变(即,被替换)。同

样地,除丙氨酸置换以外的突变也是适合的。

[0642] 在一些实施方式中,碱基编辑器的CRISPR蛋白-衍生结构域可以包括具有正准PAM序列(NGG)的Cas9蛋白的全部或部分。在其他实施方式中,碱基编辑器的Cas9-衍生结构域可以采用非-正准的PAM序列。此种序列已经在本领域中描述,并且对本领域熟练技术人员而言是显而易见的。例如,结合非-正准的PAM序列的Cas9结构域已在如下文献中描述:Kleinstiver,B.P.等,"Engineered CRISPR-Cas9nucleases with altered PAM specificities"Nature523,481-485(2015);以及Kleinstiver,B.P.等,"Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition"NatureBiotechnology 33,1293-1298(2015);其每个的全部内容在此通过引用并入。

[0643] 高保真的Cas9结构域

[0644] 本公开揭露的一些方面提供了高保真的Cas9结构域。在一些实施方式中,高保真的Cas9结构域是工程化的Cas9结构域,其包括一个或多个突变,与相应的野生型Cas9结构域相比,该突变降低了该Cas9结构域和DNA的糖-磷酸主干的静电相互作用。不希望被任何特定理论束缚,具有降低的与DNA的糖-磷酸主干的静电相互作用的高保真Cas9结构域可能具有较少的脱靶效应。在一些实施方式中,Cas9结构域(例如,野生型Cas9结构域)包括一个或多个突变,该突变降低了该Cas9结构域与DNA的糖-磷酸主干之间的相联。在一些实施方式中,Cas9结构域包括一个或多个突变,该突变使该Cas9结构域与DNA的糖-磷酸主干之间的相联降低至少1%、至少2%、至少3%、至少4%、至少5%、至少10%、至少15%、至少20%、至少25%、至少30%、至少35%、至少40%、至少45%、至少50%、至少55%、至少60%、至少65%或至少70%。

[0645] 在一些实施方式中,本文所提供的任何Cas9融合蛋白包括N497X、R661X、Q695X和/或Q926X突变中的一个或多个,或本文所提供的任何氨基酸序列中的相应突变,其中X是任一氨基酸。在一些实施方式中,本文所提供的任何Cas9融合蛋白包括N497A、R661A、Q695A和/或Q926A突变中的一个或多个,或本文所提供的任何氨基酸序列中的相应突变。在一些实施方式中,该Cas9结构域包括D10A突变,或本文所提供的任何氨基酸序列中的相应突变。具有高保真度的Cas9结构域是本领域已知的,且对熟练技术人员是显而易见的。例如,具有高保真度的Cas9结构域已在如下文献中被描述:Kleinstiver,B.P.等,"High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects."Nature529,490-495(2016);以及Slaymaker,I.M.等,"Rationally engineered Cas9 nucleases with improved specificity."Science351,84-88(2015);其每个的全部内容通过引用并入本文。

[0646] 高保真Cas9结构域突变(相对于Cas9)以粗体和底线显示

DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGE
TAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHP
IFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPD
NSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGL
FGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLS
DAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN
GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLG
ELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWN
FEEVVDKGASAQSFIERMTAFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKP
AFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL
LKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYT
GWGALSRKLINGIRDKQSGKTILDFLKSDGFANRNFMALIHDDSLTFKEDIQKAQVSGQG
DSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNS
RERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYD
VDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF
DNLTKAERGGLSELDKAGFIKRQLVETRAITKHVAQILDSRMNTKYDENDKLIREVKVIT
LKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYD
VRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKG
RDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSP
TVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIK
LPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTN
LGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[0647] 核酸可编程的DNA结合蛋白

[0648] 本公开揭露的一些方面提供了融合蛋白,该融合蛋白包括充当核酸可编程的DNA结合蛋白的结构域,其可被用于将蛋白,诸如碱基编辑器,导向至特定的核酸(例如DNA或RNA)序列。在特定的实施方式中,融合蛋白包括核酸可编程的DNA结合蛋白结构域和脱氨酶结构域。DNA结合蛋白包含但不限于:Cas9(例如,dCas9和nCas9)、CasX、CasY、Cpf1、Cas12b/C2c1和Cas12c/C2c3。具有与Cas9不同的PAM特异性的核酸可编程的DNA-结合蛋白的一个例子是来自普雷沃氏菌和弗朗西斯菌1的规律成簇间隔短回文重复序列(Cpf1)。与Cas9类似,Cpf1也是2类CRISPR效应器。已经表明,Cpf1介导稳健的DNA干扰,而其特征不同于Cas9。Cpf1是欠缺tracrRNA的单一RNA-导向的内切核酸酶,而它利用T-富集的原间隔序列临近基序(TTN、TTTN或YTN)。此外,Cpf1经由参差的DNA双-股断裂来切开DNA。在16种Cpf1-家族蛋

白中,来自氨基酸球菌属(Acidaminococcus)和毛螺菌科(Lachnospiraceae)的两种酶被证明在人类细胞中具有有效的基因组-编辑活性。Cpf1蛋白是本领域已知的并且之前已经描述过,例如Yamano等,“Crystal structure of Cpf1 in complex with guide RNA and target DNA.”Cell(165)2016,p.949-96;其全部内容在此通过引入并入。

[0649] 亦可用于本发明组合物和方法的核酸酶-无活性的Cpf1(dCpf1)变体可被用作向导核苷酸序列-可编程的DNA-结合蛋白结构域。该Cpf1蛋白具有与Cas9的RuvC结构域相似的RuvC-样核酸内切酶结构域,但不具有HNH核酸内切酶结构域,并且Cpf1的N末端不具有Cas9的α-螺旋识别叶。已经表明,在Zetsche等,Cell,163,759-771,2015(其通过引用并入本文)中,Cpf1的RuvC-样结构域负责切开两条DNA股,并且该RuvC-样结构域的去活化会灭活Cpf1核酸酶的活性。例如,与新凶手弗朗西斯菌(Francisella novicida)Cpf1中的D917A、E1006A或D1255A对应的突变灭活Cpf1核酸酶活性。在一些实施方式中,本公开揭露的dCpf1包括对应于D917A、E1006A、D1255A、D917A/E1006A、D917A/D1255A、E1006A/D1255A或D917A/E1006A/D1255A的突变。应当理解,可根据本公开揭露使用任何灭活Cpf1的RuvC结构域的突变,例如替换突变、缺失或插入。

[0650] 在一些实施方式中,本文所提供的任何融合蛋白的核酸可编程的DNA结合蛋白(napDNAbp)可能是Cpf1蛋白。在一些实施方式中,该Cpf1蛋白是Cpf1切口酶(nCpf1)。在一些实施方式中,该Cpf1蛋白是核酸酶无活性的Cpf1(dCpf1)。在一些实施方式中,该Cpf1,该nCpf1,或该dCpf1包括的氨基酸序列与本文所述的Cpf1序列的同一性是至少85%,至少90%,至少91%,至少92%,至少93%,至少94%,至少95%,至少96%,至少97%,至少98%,至少99%,或至少99.5%。在一些实施方式中,该dCpf1包括的氨基酸序列与本文所述的Cpf1序列的同一性是至少85%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%或至少99.5%,并且包括对应于D917A、E1006A、D1255A、D917A/E1006A、D917A/D1255A、E1006A/D1255A或D917A/E1006A/D1255A的突变。应当理解,可根据本公开揭露使用来自其他细菌种类的Cpf1。

[0651] 野生型新凶手弗朗西斯菌Cpf1(D917,E1006,和D1255是粗体并有底线)

[0652]

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDK
YHQFFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDS
EKFKNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWT
TYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKK
DLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTK
RKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFY
EQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTA
VLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFE
EILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNL
LHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKF
KLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKK
IVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIED
CRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENISESYIDS
VVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQS
IPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKF
NDEINLLLKEKANDVHILSI**D**RGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDK
LAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVF**E**DLNFGFKRGR
FKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYY
VPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDK
AAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAIC
GESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDA**D**A
NGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN

[0653] 新凶手弗朗西斯菌Cpf1 D917A(A917,E1006,和D1255是粗体并有底线)

[0654] MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDK
YHQFFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDS
EKFKNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWT
TYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKK
DLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTK
RKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFY
EQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTA
VLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFE
EILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNL
LHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKF
KLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKK
IVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIED
CRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENISESYIDS
VVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQS
IPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKF
NDEINLLLKEKANDVHILSI**A**RGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDK
LAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVF**E**DLNFGFKRGR
FKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYY
VPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDK
AAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAIC
GESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDA**D**A
NGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN

[0655] 新凶手弗朗西斯菌Cpf1 E1006A(D917,A1006,和D1255是粗体并有底线)

[0656] MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDK
YHQFFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDS
EKFKNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWT

TYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKK
DLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTK
RKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFY
EQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTA
VLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFE
EILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNL
LHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKF
KLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKK
IVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIED

[0657] CRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENISESYIDS
VVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQS
IPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKF
NDEINLLLKEKANDVHILSI**D**RGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDK
LAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVF**A**DLNFGFKRGR
FKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYY
VPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDK
AAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAIC
GESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDA**D**A
NGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN

[0658] 新凶手弗朗西斯菌Cpf1 D1255A(D917,E1006,和A1255是粗体并有底线)
MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDK
YHQFFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDS
EKFKNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWT
TYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKK

[0659] DLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTK
RKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFY
EQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTA
VLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFE
EILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNL

LHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKF
KLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKK
IVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIED
CRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENISESYIDS
VVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQS
IPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKF

[0660] NDEINLLLKEKANDVHILSI**D**RGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDK
LAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVF**E**DLNFGFKRGR
FKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYY
VPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDK
AAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAIC
GESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDA**A**AA
NGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN

[0661] 新凶手弗朗西斯菌Cpf1 D917A/E1006A(A917,A1006,和D1255是粗体并有底线)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDK
YHQFFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDS
EKFKNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWT
TYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKK
DLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTK
RKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFY
EQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTA

[0662] VLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFE
EILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNL
LHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKF
KLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKK
IVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIED
CRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENISESYIDS
VVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQS
IPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKF

NDEINLLLKEKANDVHILSI**A**RGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDK
LAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVF**A**DLNFGFKRGR
FKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYY

[0663] VPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDK
AAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAIC
GESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDA**D**A
NGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN

[0664] 新凶手弗朗西斯菌Cpf1 D917A/D1255A(A917,E1006,和A1255是粗体并有底线)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDK
YHQFFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDS
EKFKNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWT
TYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKK
DLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTK
RKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFY
EQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTA
VLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFE
EILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNL
LHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKF

[0665] KLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKK
IVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIED
CRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENISESYIDS
VVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQS
IPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKF
NDEINLLLKEKANDVHILSI**A**RGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDK
LAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVF**E**DLNFGFKRGR
FKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYY
VPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDK
AAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAIC
GESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDA**A**A

[0666] NGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN

[0667] 新凶手弗朗西斯菌Cpf1 E1006A/D1255A(D917,A1006,和A1255是粗体并有底线)

[0668] MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDK
YHQFFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDS
EKFKNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWT
TYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKK
DLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTK
RKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFY
EQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTA
VLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFE
EILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNL
LHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKF
KLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKK
IVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIED
CRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENISESYIDS
VVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQS
IPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKF
NDEINLLLKEKANDVHILSI**D**RGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDK
LAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVF**A**DLNFGFKRGR
FKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYY
VPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDK
AAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAIC
GESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDA**A**A
NGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN

[0669] 新凶手弗朗西斯菌Cpf1 D917A/E1006A/D1255A(A917,A1006,和A1255是粗体并有底线)

[0670] MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDK
YHQFFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDS

[0671]
EKFKNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWT
TYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKK
DLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTK
RKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFY
EQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTA
VLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFE
EILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNL
LHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKF
KLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKK
IVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIED
CRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENISESYIDS
VVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQS
IPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKF
NDEINLLLKEKANDVHILSIARGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDK
LAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVFADLNFGFKRGR
FKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYY
VPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDK
AAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAIC
GESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDAAA
NGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN

[0672] 在一些实施方式中,融合蛋白中存在的Cas9结构域之一可被对PAM序列没有要求的向导核苷酸序列-可编程的DNA-结合蛋白结构域所取代。

[0673] 在一些实施方式中,该核酸可编程的DNA结合蛋白(napDNAbp)是微生物CRISPR-Cas系统的单一效应器。微生物CRISPR-Cas系统的单一效应器包含但不限于:Cas9,Cpf1,Cas12b/C2c1,和Cas12c/C2c3。通常,微生物CRISPR-Cas系统分为1类和2类系统。1类系统具有多亚基效应器复合体,而2类系统具有单一蛋白效应器。例如,Cas9和Cpf1是2类效应器。除了Cas9和Cpf1,三个相异的2类CRISPR-Cas系统(Cas12b/C2c1,和Cas12c/C2c3)已被如下文献描述:Shmakov等,“多样的2类CRISPR Cas系统的发现与功能性表征”,Mol.Cell,2015Nov.5;60(3):385-397,其全部内容在此通过引用并入。其中两个系统Cas12b/C2c1和Cas12c/C2c3的效应器,含有与Cpf1相关的RuvC-样核酸内切酶结构域。第三个系统含有具有两个预测的HEPN RNase结构域的效应器。成熟CRISPR RNA的生产不依赖于tracrRNA,这不同于通过Cas12b/C2c1生产的CRISPR RNA。Cas12b/C2c1对DNA的切开依赖于CRISPR RNA和tracrRNA两者。

[0674] 酸土脂环酸芽孢杆菌(Alicyclobaccillus acidoterrastris)Cas12b/C2c1

(AacC2c1)与嵌合的单-分子向导RNA(sgRNA)的复合体的晶体结构已被报道,参见例如Liu等,“C2c1-sgRNA Complex Structure Reveals RNA-Guided DNA Cleavage Mechanism”,Mol.Cell,2017Jan.19;65(2):310-322,其全部内容在此通过引用并入。以三元复合体形式与靶DNA结合的酸土脂环酸芽孢杆菌C2c1的晶体结构也被报道。参见例如Yang等,“PAM-dependent Target DNA Recognition and Cleavage by C2C1 CRISPR-Cas endonuclease”,Cell,2016Dec.15;167(7):1814-1828,其全部内容在此通过引用并入。AacC2c1的催化胜任的两个构象(带有靶和非靶DNA股)已被独立地捕捉到,其就定位在单一RuvC催化口袋内,伴随着Cas12b/C2c1-介导的切开导致靶DNA的参差的七-核苷酸断裂。Cas12b/C2c1三元复合体与先前鉴别的Cas9和Cpf1对应物之间的结构比较展示了CRISPR-Cas9系统所用机制的多样性。

[0675] 在一些实施方式中,本文所提供的任何融合蛋白的核酸可编程的DNA结合蛋白(napDNAbp)可能是Cas12b/C2c1,或Cas12c/C2c3蛋白。在一些实施方式中,该napDNAbp是Cas12b/C2c1蛋白。在一些实施方式中,该napDNAbp是Cas12c/C2c3蛋白。在一些实施方式中,该napDNAbp包括的氨基酸序列,与天然-存在的Cas12b/C2c1或Cas12c/C2c3蛋白,是至少85%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%或至少99.5%相同的。在一些实施方式中,该napDNAbp是天然-存在的Cas12b/C2c1或Cas12c/C2c3蛋白。在一些实施方式中,该napDNAbp包括的氨基酸序列,与本文所提供的任一napDNAbp序列,是至少85%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%或至少99.5%相同的。应该理解,来自其他细菌种类的Cas12b/C2c1或Cas12c/C2c3也可根据本公开揭露而被使用。

[0676] Cas12b/C2c1

[0677] (uniprot.org/uniprot/T0D7A2#2)sp|T0D7A2|C2C1_ALIAG CRISPR-关联的核酸内切酶C2c1 OS

[0678] =酸土脂环酸芽孢杆菌(菌株ATCC 49025/DSM 3922/CIP106132/NCIMB 13137/GD3B)GN=c2c1 PE=1SV=1

[0679] MAVKSIKVKLRLDDMPEIRAGLWKLHKEVNAGVRYYTEWLSLLRQENLYRRSPNGDGEQECDKTAEEC
KAELLERLRARQVENGHRGPAGSDDELLQLARQLYELLVPQAIGAKGDAQQIARKFLSPLADKDAVGGLGIAKAGN
KPRWVRMREAGEPGWEEEKEKAETRKSADRTADVLRALADFGLKPLMRVYTDSEMSSVEWKPLRKGQAVRTWDRDM
FQQAIERMMSWESWNQRVGQEYAKLVEQKNRFEQKNFVGQEHLVHLVNQLQQDMKEASPGLESKEQTAHYVTGRAL
RGSDKVFEKWGKLAPDAPFDLYDAEIKNVQRRNTRRFGSHDLFAKLAEPEYQALWREDASFLTRYAVYNSILRKLN
HAKMFATFTLPDATAHPIWTRFDKLGGNLHQYTFLFNEFGERRHAIRFHKLLKVENGVAREVDDVTVPISMSEQLD
NLLPRDPNEPIALYFRDYGAEQHFTGEFGGAKIQCRRDQLAHMHRRRGARDVYLNVSVRVQSQSEARGERRPPYAA
VFRLVGDNHRAFVHFDKLSDYLAEHPDDGKLGSEGLLSGLRVMSVDLGLRTSASISVFRVARKDELKPNSKGRVPF
FFPIKGNDNLVAVHERSQLLKLPGETESKDLRAIREERQRTLRQLRTQLAYLRLLVRCGSEDVGRRERSWAKLIEQ
PVDAANHMTPDWREAFENELQKLKSLHGICSDKEWMDAVYESVRRVWRHMGKQVRDWRKDVRSGERPKIRGYAKDV
VGGNSIEQIEYLERQYKFLKSWSFFGKVSGQVIRAEKGSRFAITLREHIDHAKEDRLKKLADRIIMEALGYVYALD
ERGKGKWVAKYPPCQLILLEELSEYQFNNDRPPSENNQLMQWSHRGVFQELINQAQVHDLLVGTMYAAFSSRFDAR
TGAPGIRCRRVPARCTQEHNPEPFPWWLNKFVVEHTLDACPLRADDLIPTGEGEIFVSPFSAEEGDFHQIHADLNA

AQNLQQRLWSDFDISQIRLRCDWGEVDGELVLIPRLTGKRTADSYSNKVFYTNTGVTYYERERGKKRRKVFAQEKL SEEEAELLVEADEAREKSVVLMRDPSGIINRGNWTRQKEFWSMV NQRIEGYLVKQIRSRVPLQDSACENTGDI

[0680] 包括核定位序列(NLS)的融合蛋白

[0681] 在一些实施方式中,本文所提供的融合蛋白还包括一个或多个(例如,2,3,4,5个)核靶向序列,例如核定位序列(NLS)。在一个实施方式中,使用了双分型NLS。在一些实施方式中,NLS包括氨基酸序列,其促进包括NLS的蛋白质输入细胞核中(例如,通过核转运)。在一些实施方式中,本文所提供的任何融合蛋白还包括核定位序列(NLS)。在一些实施方式中,该NLS融合到该融合蛋白的N端。在一些实施方式中,该NLS融合到该融合蛋白的C端。在一些实施方式中,该NLS融合到Cas9结构域的N端。在一些实施方式中,该NLS融合到nCas9结构域或dCas9结构域的C端。在一些实施方式中,该NLS融合到脱氨酶的N端。在一些实施方式中,该NLS融合到脱氨酶的C端。在一些实施方式中,该NLS经由一个或多个连接子融合到融合蛋白。

[0682] 在一些实施方式中,该NLS在无连接子的情况下融合到融合蛋白。在一些实施方式中,NLS包括本文所提供的或所引用的任一NLS序列的氨基酸序列。额外的核定位序列在本领域中是已知的,并且对本领域熟练技术人员而言是显而易见的。例如,NLS序列在Plank等,PCT/EP2000/011690中描述,其内容通过对其示例性核定位序列的公开揭露的引用而并入本文。在一些实施方式中,NLS包括的氨基酸序列为:PKKKRKVEGADKRTADGSEFES PKKKRKV,KRTADGSEFESPKKKRKV,KRPAATKKAGQAKKKK,KKTELQTTNAENKTKKL,KRGINDRNFWRGENGRKTR,RKSGKIAAIVVKRPRKPKKKRKV,或MDSLLMNRRKFLYQFKNVRWAKGRRETYLC。

[0683] 在一些实施方式中,该NLS存在于连接子中,或连接子位于该NLS的两侧翼,例如,本文所述的连接子。在一些实施方式中,该N端或C端NLS是双分型NLS。双分型NLS包括两个碱性氨基酸簇,它们由相对较短的间隔序列所分开(故,双分型-2个部分,而单分型NLS则不是)。核质蛋白的NLS,KR[PAATKKAGQA]KKKK,是普遍存在的双分型信号的原型:2个碱性氨基酸的簇,其由约10个氨基酸的间隔序列所分开。示例性的双分型NLS的序列如下:

[0684] PKKKRKVEGADKRTADGSEFES PKKKRKV

[0685] 在一些实施方式中,本发明的融合蛋白不包括连接子序列。在一些实施方式中,存在一个或多个所述结构域或蛋白之间的连接子序列。

[0686] 应当理解,本公开揭露的融合蛋白可能包括一个或多个附加的特色。例如,在一些实施方式中,该融合蛋白可能包括抑制剂,细胞质定位序列,输出序列,诸如核输出序列,或其他定位序列,以及可用于该融合蛋白的增溶、纯化或检测的序列标签。本文所提供的适合的蛋白质标签包含但不限于,生物素羧化酶载体蛋白(BCCP)标签,myc-标签,钙调蛋白-标签,FLAG-标签,血凝素(HA)-标签,多组氨酸标签,也称为组氨酸标签或His-标签,麦芽糖结合蛋白(MBP)-标签,nus-标签,谷胱甘肽-硫-转移酶(GST)-标签,绿色荧光蛋白(GFP)-标签,硫氧还蛋白-标签,S-标签,Softags(例如,Softag 1,Softag 3),链霉亲和素-标签,生物素连接酶标签,FlAsH标签,V5标签和SBP-标签。额外的适合的序列对本领域技术人员而言将是显而易见的。在一些实施方式中,该融合蛋白包括一个或多个His标签。

[0687] 连接子

[0688] 在某些实施方式中,连接子可能用于链接本发明的任何肽或肽结构域。该连接子可能是简单如共价键,或它可能是许多个原子长度的聚合的连接子。在某些实施方式中,该

连接子是多肽或基于氨基酸的。在其他实施方式中,该连接子不是肽-样的。在某些实施方式中,该连接子是共价键(例如,碳-碳键,二硫键,碳-杂原子键,等)。在某些实施方式中,该连接子是酰胺键合的碳-氮键。在某些实施方式中,该连接子是环状的或无环的,替换的或未替换的,分枝的或未分枝的,脂肪族的或杂脂肪族的连接子。在某些实施方式中,该连接子是聚合的(例如,聚乙烯,聚乙二醇,聚酰胺,聚酯等)。在某些实施方式中,该连接子包括氨基链烷酸的单体,二聚体或聚合物。在某些实施方式中,该连接子包括氨基链烷酸(例如,甘氨酸,乙酸,丙氨酸,β-丙氨酸,3-氨基丙酸,4-氨基丁酸,5-戊酸等)。在某些实施方式中,该连接子包括氨基己酸(Ahx)的单体,二聚体或聚合物。在某些实施方式中,该连接子是基于碳环部分体(例如,环戊烷,环己烷)。在其他实施方式中,该连接子包括聚乙二醇部分体(PEG)。在其他实施方式中,该连接子包括氨基酸。在某些实施方式中,该连接子包括肽。在某些实施方式中,该连接子包括芳基或杂芳基部分体。在某些实施方式中,该连接子是基于苯环的。该连接子可能包含官能化的部分体,以促进亲核基(例如,硫醇基,氨基)从肽到所述连接子的附接。任何亲电子试剂都可用作该连接子的一部分。示例性的亲电子试剂包含但不限于:活化的酯、活化的酰胺、Michael接受体、烷基卤化物、芳基卤化物、酰基卤化物和异硫氰酸酯。

[0689] 在一些实施方式中,该连接子是一个氨基酸或复数个氨基酸(例如,肽或蛋白质)。在一些实施方式中,该连接子是键(例如共价键)、有机分子、基团、聚合物或化学部分体。在一些实施方式中,该连接子是约3到约104(例如,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、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、55、60、65、70、75、80、85、90、95或100)个氨基酸的长度。

[0690] 具有向导RNA的Cas9复合体

[0691] 本公开揭露的一些方面提供了复合体,该复合体包括本文所提供的任何融合蛋白,以及向导RNA。可以使用任何用于链接所述融合蛋白结构域的方法(例如,范围从(GGGS)n、(GGGGS)n和(G)n形式的非常柔性的连接子到(EAAAK)n,(SGGS)n,SGSETPGTSESATPES(参见例如Guilinger JP,Thompson DB,Liu DR.,Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification.,Nat.Biotechnol.2014;32(6):577-82;其全部内容通过引用并入本文)和(XP)n形式的更刚性的连接子),以取得适于所述核碱基编辑器活性的最佳长度。在一些实施方式中,n是1、2、3、4、5、6、7、8、9、10、11、12、13、14或15。在一些实施方式中,该连接子包括(GGS)n基序,其中n是1、3或7。在一些实施方式中,本文所提供的融合蛋白的Cas9结构域是经由包括氨基酸序列SGSETPGTSESATPES的连接子所融合。

[0692] 在一些实施方式中,该向导核酸(例如向导RNA)的长度是从15到100个核苷酸,并且包括与靶序列互补的至少10个接连的核苷酸的序列。在一些实施方式中,该向导RNA是15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49或50个核苷酸的长度。在一些实施方式中,该向导RNA包括与靶序列互补的15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39或40个接连的核苷酸的序列。在一些实施方式中,该靶序列是DNA序列。在一些实施方式中,该靶序列是细菌、酵母、真菌、昆虫、植物或动物的基因组中的序列。在一些实施方式中,该靶序列是人类基因组中的序列。在一些实施方式中,该靶序列的3'端紧邻

正准的PAM序列(NGG)。在一些实施方式中,该靶序列的3'端紧邻非-正准的PAM序列(例如表1中所列序列)。

[0693] 本公开揭露的一些方面提供了使用本文所提供的融合蛋白或复合体的方法。例如,本公开揭露的一些方面提供了方法,其包括使DNA分子与本文所提供的任何融合蛋白和至少一种向导RNA接触,其中该向导RNA是约15-100个核苷酸的长度并且包括与靶序列互补的至少10个接连的核苷酸的序列。在一些实施方式中,该靶序列的3'端紧邻AGC、GAG、TTT、GTG或CAA序列。在一些实施方式中,该靶序列的3'端紧邻NGA、NGCG、NGN、NNGRRT、NNNRRT、NGCG、NGCN、NGTN、NGTN、NGTN或5'(TTTV)序列。

[0694] 应当理解,各个序列中特定位置或残基的编号取决于所使用的特定蛋白质和编号方案。编号可能不同,例如在成熟蛋白的前体和该成熟蛋白本身中,并且物种与物种之间序列的差异可能影响编号。本领域技术人员将能够通过本领域皆知的方法,例如通过序列比对和同源残基的确定,来鉴别任何同源蛋白以及各自编码核酸中的各自残基。

[0695] 对本领域技术人员而言将是显而易见的是,为了将本文所揭露的任何融合蛋白靶向到靶位点,例如包括待编辑的突变的位点,通常必要与向导RNA一起共-表达该融合蛋白。如本文他处的详解,向导RNA通常包括允许Cas9结合的tracrRNA框架,以及向导序列,其将序列特异性赋予Cas9:核酸编辑酶/结构域融合蛋白。或者,该向导RNA和tracrRNA可作为两个核酸分子而分开提供。在一些实施方式中,该向导RNA包括一种结构,其中该向导序列包括与该靶序列互补的序列。该向导序列通常为20个核苷酸的长度。基于本公开揭露,用于将Cas9:核酸编辑酶/结构域融合蛋白靶向到特定基因组靶位点的适合的向导RNAs的序列对本领域技术人员而言将是显而易见的。此种适合的向导RNA序列通常包括与待编辑的靶核苷酸上游或下游50个核苷酸内的核酸序列互补的向导序列。本文提供了一些适用于将任何所提供的融合蛋白靶向到特定靶序列的示例性的向导RNA序列。

[0696] 使用包括胞苷脱氨酶,腺苷脱氨酶和Cas9结构域的融合蛋白的方法

[0697] 本公开揭露的一些方面提供了使用本文所提供的融合蛋白或复合体的方法。例如,本公开揭露的一些方面提供了方法,其包括使编码突变的DNA分子与本文所提供的任何融合蛋白以及与至少一个向导RNA接触,其中该向导RNA是约15到100个核苷酸的长度,并且包括与靶序列互补的至少10个接连核苷酸的序列。在一些实施方式中,该靶序列的3'末端紧邻正准的PAM序列(NGG)。在一些实施方式中,该靶序列的3'末端并非紧邻正准的PAM序列(NGG)。在一些实施方式中,该靶序列的3'末端紧邻AGC,GAG,TTT,GTG,或CAA序列。在一些实施方式中,该靶序列的3'末端紧邻NGA、NGCG、NGN、NNGRRT、NNNRRT、NGCG、NGCN、NGTN、NGTN、NGTN或5'(TTTV)序列。

[0698] 应当理解,各个序列中特定位置或残基的编号取决于所使用的特定蛋白质和编号方案。编号可能不同,例如在成熟蛋白的前体和成熟蛋白本身中,并且物种与物种之间序列的差异可能影响编号。本领域技术人员将能够通过本领域皆知的方法,例如通过序列比对和同源残基的确定,来鉴别任何同源蛋白以及各自编码核酸中的各自残基。

[0699] 对本领域技术人员而言将是显而易见的是,为了将如本文所揭露的包括Cas9结构域和胞苷脱氨酶或腺苷脱氨酶的任何融合蛋白靶向到靶位点,例如包括待编辑的突变的位点,通常必要与向导RNA(例如,sgRNA)一起共-表达该融合蛋白。如本文他处的详解,向导RNA通常包括允许Cas9结合的tracrRNA框架,以及向导序列,该向导序列将序列特异性赋予

所述Cas9:核酸编辑酶/结构域融合蛋白。或者,该向导RNA和tracrRNA可作为两个核酸分子而分开提供。在一些实施方式中,该向导RNA包括一种结构,其中该向导序列包括与所述靶序列互补的序列。该向导序列通常为20个核苷酸的长度。基于本公开揭露,用于将Cas9:核酸编辑酶/结构域融合蛋白靶向到特定基因组靶位点的适合的向导RNA的序列对本领域技术人员而言将是显而易见的。此种适合的向导RNA序列通常包括与待编辑的靶核苷酸上游或下游50个核苷酸内的核酸序列互补的向导序列。

[0700] 本文提供了一些适合用于将任何所提供的融合蛋白靶向到特定靶序列的示例性的向导RNA序列。

[0701] 碱基编辑器效率

[0702] 本发明的融合蛋白有利地修饰了编码包括突变的特定核苷酸碱基,而不生成显著比例的插入/缺失。“插入/缺失”,如本文所用,是指核酸内的核苷酸碱基的插入或缺失。此种插入或缺失可以导致在基因的编码区域内的移码突变。在一些实施方式中,期望生成有效修饰(例如,使突变)核酸内的特定核苷酸而不在该核酸内生成大量的插入或缺失(即,插入/缺失)的碱基编辑器。在某些实施方式中,本文所提供的任何碱基编辑器能够生成比插入/缺失更大比例的预期修饰(例如,突变)。

[0703] 在一些实施方式中,本文所提供的碱基编辑器能够生成的预期突变与插入/缺失的比率大于1:1。在一些实施方式中,本文所提供的碱基编辑器能够生成的预期突变与插入/缺失的比率是至少1.5:1、至少2:1、至少2.5:1、至少3:1、至少3.5:1、至少4:1、至少4.5:1、至少5:1、至少5.5:1、至少6:1、至少6.5:1、至少7:1、至少7.5:1、至少8:1、至少10:1、至少12:1、至少15:1、至少20:1、至少25:1、至少30:1、至少40:1、至少50:1、至少100:1、至少200:1、至少300:1、至少400:1、至少500:1、至少600:1、至少700:1、至少800:1、至少900:1或至少1000:1,或更高。预期突变和插入/缺失的数目可使用任何适合的方法来确定。

[0704] 在一些实施方式中,本文所提供的碱基编辑器能够限制在核酸区域内的插入/缺失的形成。在一些实施方式中,该区域位于被碱基编辑器所靶向的核苷酸处,或位于被碱基编辑器所靶向的核苷酸的2,3,4,5,6,7,8,9,或10个核苷酸内的区域。在一些实施方式中,本文所提供的任何碱基编辑器能够将在核酸区域的插入/缺失的形成限制为小于1%、小于1.5%、小于2%、小于2.5%、小于3%、小于3.5%、小于4%、小于4.5%、小于5%、小于6%、小于7%、小于8%、小于9%、小于10%、小于12%、小于15%或小于20%。在核酸区域所形成的插入/缺失的数目可能取决于核酸(例如,细胞的基因组内的核酸)暴露于碱基编辑器的时长数。在一些实施方式中,在将核酸(例如,细胞的基因组内的核酸)暴露于碱基编辑器之后至少1小时、至少2小时、至少6小时、至少12小时、至少24小时、至少36小时、至少48小时、至少3天、至少4天、至少5天、至少7天、至少10天或至少14天,再确定插入/缺失的数目或比例。

[0705] 本公开揭露的一些方面基于以下认识:本文所提供的任何碱基编辑器能够有效地在核酸(例如受试者基因组内的核酸)中生成预期突变而不生成显著数目的非预期突变。在一些实施方式中,预期突变是由与gRNA结合的特定碱基编辑器所生成的突变,专门设计用于改变或更正突变。在一些实施方式中,本文所提供的任何碱基编辑器能够生成的预期突变与非预期突变的比率(例如预期突变:非预期突变)大于1:1。在一些实施方式中,本文所提供的任何碱基编辑器能够生成的预期突变与非预期突变的比率是至少1.5:1、至少2:1、

至少2.5:1、至少3:1、至少3.5:1、至少4:1、至少4.5:1、至少5:1、至少5.5:1、至少6:1、至少6.5:1、至少7:1、至少7.5:1、至少8:1、至少10:1、至少12:1、至少15:1、至少20:1、至少25:1、至少30:1、至少40:1、至少50:1、至少100:1、至少150:1、至少200:1、至少250:1、至少500:1或至少1000:1或更高。应当理解,本文中“碱基编辑器”部分中所描述的碱基编辑器的特征可应用于任何融合蛋白,或应用于使用本文所提供的融合蛋白的方法。

[0706] 用于编辑核酸的方法

[0707] 本公开揭露的一些方面提供了用于编辑核酸的方法。在一些实施方式中,该方法是用于编辑编码目标多肽(例如,疾病基因的表达产物)的核酸分子的核碱基的方法。在一些实施方式中,该方法包括以下步骤:a)使核酸(例如,双股DNA序列)的靶区域与包括碱基编辑器和向导核酸(例如,gRNA)的复合体接触,b)诱导所述靶区域的股分离,c)将该靶区域的单股中的所述靶核碱基对的第一核碱基转换为第二核碱基,以及d)使用所述nCas9切割所述靶区域且该切割不超过一股,其中与所述第一核碱基互补的第三核碱基被与所述第二核碱基互补的第四核碱基取代。在一些实施方式中,该方法造成所述核酸中少于20%的插入/缺失的形成。应当理解,在一些实施方式中,步骤b被省略。在一些实施方式中,该方法造成少于19%、18%、16%、14%、12%、10%、8%、6%、4%、2%、1%、0.5%、0.2%或少于0.1%的插入/缺失的形成。在一些实施方式中,该方法还包括用与所述第四核碱基互补的第五核碱基取代所述第二核碱基,从而生成预期的被编辑的碱基对(例如,G·C至A·T)。在一些实施方式中,至少5%的预期的碱基对是被编辑的。在一些实施方式中,至少10%、15%、20%、25%、30%、35%、40%、45%或50%的预期的碱基对是被编辑的。

[0708] 在一些实施方式中,所述靶核苷酸中预期产物与非预期产物的比率是至少2:1、5:1、10:1、20:1、30:1、40:1、50:1、60:1、70:1、80:1、90:1、100:1或200:1或更大。在一些实施方式中,预期突变与所形成的插入/缺失的比率是大于1:1、10:1、50:1、100:1、500:1或1000:1或更大。在一些实施方式中,该被切割的单股(带切口的股)杂交到所述向导核酸上。在一些实施方式中,该被切割的单股与包括所述第一核碱基的股是相对的。在一些实施方式中,该碱基编辑器包括dCas9结构域。在一些实施方式中,该碱基编辑器保护或结合非-被编辑股。在一些实施方式中,该预期的被编辑的碱基对在PAM位点的上游。在一些实施方式中,该预期的被编辑的碱基对在PAM位点的上游的第1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个核苷酸。在一些实施方式中,该预期的被编辑的碱基对在PAM位点的下游。在一些实施方式中,该预期的被编辑的碱基对在所述PAM位点的下游的第1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个核苷酸处。在一些实施方式中,该方法不需要正准的(例如,NGG)PAM位点。在一些实施方式中,该核碱基编辑器包括连接子。在一些实施方式中,该连接子是1-25个氨基酸的长度。在一些实施方式中,该连接子是5-20个氨基酸的长度。在一些实施方式中,连接子是10、11、12、13、14、15、16、17、18,19或20个氨基酸的长度。在一个实施方式中,该连接子是32个氨基酸的长度。在另一个实施方式中,“长连接子”是至少约60个氨基酸的长度。在其他实施方式中,该连接子是约3至100个氨基酸之间的长度。在一些实施方式中,该靶区域包括靶标窗,其中该靶标窗包括所述靶核碱基对。在一些实施方式中,该靶标窗包括1至10个核苷酸。在一些实施方式中,该靶标窗是1至9、1至8、1至7、1至6、1至5、1至4、1至3、1至2或1个核苷酸的长度。在一些实施方式中该靶标窗是1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个核苷酸的长度。在一些实施方式

中,该预期的被编辑的碱基对是在靶标窗内。在一些实施方式中,该靶标窗包括该预期的被编辑的碱基对。在一些实施方式中,使用本文所提供的任何碱基编辑器来执行该方法。

[0709] 在一些实施方式中,本公开揭露提供了用于编辑核苷酸(例如SNP)的方法。在一些实施方式中,本公开揭露提供了用于编辑双-股DNA序列的核碱基对的方法。在一些实施方式中,该方法包括a)使该双-股DNA序列的靶区域与包括碱基编辑器和向导核酸(例如gRNA)的复合体接触,其中该靶区域包括靶核碱基对,b)诱导所述靶区域的股分离,c)将该靶区域的单股中的所述靶核碱基对的第一核碱基转换为第二核碱基,d)切割所述靶区域且该切割不超过一股,其中与所述第一核碱基互补的第三核碱基被与所述第二核碱基互补的第四核碱基取代,并且该第二核碱基被与该第四核碱基互补的第五核碱基取代,从而生成预期的被编辑的碱基对,其中生成该预期的被编辑的碱基对的效率是至少5%。应当理解,在一些实施方式中,步骤b被省略。在一些实施方式中,至少5%的预期的碱基对是被编辑的。在一些实施方式中,至少10%、15%、20%、25%、30%、35%、40%、45%或50%预期的碱基对是被编辑的。在一些实施方式中,该方法造成少于19%、18%、16%、14%、12%、10%、8%、6%、4%、2%、1%、0.5%、0.2%或少于0.1%的插入/缺失的形成。在一些实施方式中,在所述靶核苷酸处的预期产物与非预期产物的比率是至少2:1、5:1、10:1、20:1、30:1、40:1、50:1、60:1、70:1、80:1、90:1、100:1或200:1、或更大。在一些实施方式中,预期突变与所形成的插入/缺失的比率是大于1:1、10:1、50:1、100:1、500:1或1000:1或更大。在一些实施方式中,该被切割的单股杂交到所述向导核酸上。在一些实施方式中,该被切割的单股与包括所述第一核碱基的股是相对的。在一些实施方式中,该预期的被编辑的碱基对在PAM位点的上游。在一些实施方式中,该预期的被编辑的碱基对在PAM位点的上游的第1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个核苷酸。在一些实施方式中,该预期的被编辑的碱基对在PAM位点的下游。在一些实施方式中,该预期的被编辑的碱基对在PAM位点的下游的第1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个核苷酸。在一些实施方式中,该连接子是10、11、12、13、14、15、16、17、18、19或20个氨基酸的长度。在一些实施方式中,该靶区域包括靶标窗,其中该靶标窗包括该靶核碱基对。在一些实施方式中,该靶标窗包括1-10个核苷酸。在一些实施方式中,该靶标窗是1-9、1-8、1-7、1-6、1-5、1-4、1-3、1-2或1个核苷酸的长度。在一些实施方式中,该靶标窗是1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个核苷酸的长度。在一些实施方式中,该预期的被编辑的碱基对出现在靶标窗内。在一些实施方式中,该靶标窗包括该预期的被编辑的碱基对。在一些实施方式中,该核碱基编辑器是本文所提供的任一碱基编辑器。

[0710] 多重编辑

[0711] 在一些实施方式中,本文所提供的碱基编辑器系统能够对一个或多个基因中的复数个核碱基对进行多重编辑。在一些实施方式中,该复数个核碱基对定位于同一基因中。在一些实施方式中,该复数个核碱基对定位于一个或多个基因中,其中至少一个基因定位于不同的基因座。在一些实施方式中,该多重编辑可以包括一个或多个向导多核苷酸。在一些实施方式中,该多重编辑可以包括一个或多个碱基编辑器系统。在一些实施方式中,该多重编辑可以包括一个或多个碱基编辑器系统,其具有单一向导多核苷酸。在一些实施方式中,该多重编辑可以包括一个或多个碱基编辑器系统,其具有复数个向导多核苷酸。在一些实施方式中,该多重编辑可以包括一个或多个向导多核苷酸,其具有单一碱基编辑器系统。在

一些实施方式中,该多重编辑可以包括至少一个向导多核苷酸,其不需要PAM序列以靶向结合到靶多核苷酸序列。在一些实施方式中,该多重编辑可以包括至少一个向导多核苷酸,其需要PAM序列以靶向结合到靶多核苷酸序列。在一些实施方式中,该多重编辑可以包括至少一个不需要PAM序列以靶向结合到靶多核苷酸序列的向导多核苷酸和至少一个需要PAM序列以靶向结合到靶多核苷酸序列的向导多核苷酸的混合物。应当理解,使用如本文所述的任何碱基编辑器的多重编辑的特征可以应用于使用本文所提供的任何碱基编辑器的方法的任何组合。也应当理解,使用如本文所述的任何碱基编辑器的多重编辑可以包括对复数个核碱基对的循序编辑。

[0712] 在一些实施方式中,该复数个核碱基对是在一个或多个基因中。在一些实施方式中,该复数个核碱基对是在同一个基因中。在一些实施方式中,所述一个或多个基因中的至少一个基因是定位于不同的基因座。

[0713] 在一些实施方式中,所述编辑是指,在至少一个蛋白质编码区域中的对所述复数个核碱基对的编辑。在一些实施方式中,所述编辑是指,在至少一个蛋白质非-编码区域中的对所述复数个核碱基对的编辑。在一些实施方式中,所述编辑是指,在至少一个蛋白质编码区域和在至少一个蛋白质非-编码区域中的对所述复数个核碱基对的编辑。

[0714] 在一些实施方式中,该编辑是连同一个或多个向导多核苷酸的。在一些实施方式中,该碱基编辑器系统可以包括一个或多个碱基编辑器系统。在一些实施方式中,该碱基编辑器系统可以包括一个或多个碱基编辑器系统,其与单一向导多核苷酸相连。在一些实施方式中,该碱基编辑器系统可以包括一个或多个碱基编辑器系统,其与复数个向导多核苷酸相连。在一些实施方式中,该编辑是连同一个或多个向导多核苷酸的,所述向导多核苷酸具有单一碱基编辑器系统。在一些实施方式中,该编辑是连同至少一个向导多核苷酸的,所述向导多核苷酸不需要PAM序列以靶向结合到靶多核苷酸序列。在一些实施方式中,该编辑是连同至少一个向导多核苷酸的,所述向导多核苷酸需要PAM序列以靶向结合到靶多核苷酸序列。在一些实施方式中,该编辑是连同至少一个不需要PAM序列以靶向结合到靶多核苷酸序列的向导多核苷酸和至少一个需要PAM序列以靶向结合到靶多核苷酸序列的向导多核苷酸的混合物。应当理解,使用如本文所述的任何碱基编辑器的多重编辑的特征可以应用于使用本文所提供的任何碱基编辑器的方法的任何组合。也应当理解,所述编辑可以包括对复数个核碱基对的循序编辑。

[0715] 宿主细胞中融合蛋白的表达

[0716] 使用本领域熟练技术人员已知的常规方法,本发明的融合蛋白可在几乎任何目标宿主细胞中表达,包含但不限于细菌,酵母,真菌,昆虫,植物和动物细胞。例如,可以用cDNA序列为基础设计适于CDS上游和下游的适合的引物来克隆编码本发明的融合蛋白的DNA。所克隆的DNA可直接或在需要时用限制性酶消化后或在附加适合的连接子和/或核定位信号后,与编码碱基编辑系统的一个或多个其他组分的DNA相连接。该碱基编辑系统在宿主细胞中被翻译成以形成复合体。

[0717] 通过将编码具有核碱基修饰活性(例如,腺苷脱氨酶,胞苷脱氨酶,DNA糖基化酶)的一个或多个结构域的一个或多个多核苷酸可操作地链接至编码napDNAbp的多核苷酸,以制备编码本发明的融合蛋白的多核苷酸而生成融合蛋白。在一些实施方式中,编码napDNAbp的多核苷酸,和编码具有核碱基修饰活性的结构域的DNA可各自与编码结合结构

域的DNA或所述结合结构域结合伙伴的DNA融合,抑或,两种DNA均可与编码分离内含肽的DNA融合,由此所述核酸序列-识别转换模块和所述核酸碱基转换酶在宿主细胞中被翻译以形成复合体。在这些情况下,当需要时,可以将连接子和/或核定位信号链接到所述两种DNA之一或二者的适合位置。

[0718] 编码本文所述的蛋白结构域的DNA可以通过化学合成该DNA而获得,或者通过利用PCR法和Gibson组装法连接合成的部分重叠的寡DNA短链来构建编码其全长的DNA。通过化学合成或通过PCR方法或Gibson组装法的组合来构建全长DNA的优势为,可以根据DNA被导入的宿主设计用于全长CDS的待用密码子。在表达异源DNA时,通过将其蛋白的DNA序列转换为所述宿主生物体中高频使用的密码子,该蛋白质的表达水平预计会增加。既然要使用宿主中密码子使用频率的数据,例如,可以使用在Kazusa DNA研究院的主页上公开的遗传密码使用频率数据库(http://www.kazusa.or.jp/codon/index.html),也可以参考显示每个宿主中密码子使用频率的文件。通过参比已获得的数据和要引入的DNA序列,可将用于该DNA序列的在该宿主中显示低使用频率的密码子转换为编码相同氨基酸并显示出高使用频率的密码子。

[0719] 可以,例如,通过将所述DNA链接到适合的表达载体中启动子的下游,而生产含有编码核酸序列-识别模块和/或核酸碱基转换酶DNA的表达载体。

[0720] 作为表达载体,使用了:大肠杆菌-衍生的质粒(例如、pBR322、pBR325、pUC12、pUC13);枯草芽孢杆菌-衍生的质粒(例如、pUB110、pTP5、pC194);酵母--衍生的质粒(例如pSH19、pSH15);昆虫细胞表达质粒(例如pFast-Bac);动物细胞表达质粒(例如pA1-11、pXT1、pRc/CMV、pRc/RSV、pcDNAI/Neo);噬菌体,诸如λ噬菌体等;昆虫病毒载体,诸如杆状病毒等(例如BmNPV、AcNPV);动物病毒载体,诸如逆转录病毒,牛痘病毒,腺病毒等等。

[0721] 至于启动子,可以使用适合用于基因表达的宿主的任何启动子。在使用DSB的常规方法中,由于宿主细胞的存活率有时会因毒性而显著降低,因此期望通过使用诱导型启动子在诱导开始时增加细胞数目。然而,由于通过表达本发明的核酸-修饰酶复合体也可以提供足够的细胞增殖,因此也可以不受限地使用组成型启动子。

[0722] 例如,当该宿主是动物细胞时,使用了:SRα启动子,SV40启动子,LTR启动子,CMV(巨细胞病毒)启动子,RSV(Rous肉瘤病毒)启动子,MoMuLV(Moloney小鼠白血病病毒)LTR,HSV-TK(单纯疱疹病毒胸苷激酶)启动子等。这些之中,优选CMV启动子、SRα启动子等。在一个实施方式中,该启动子是CMV启动子或SRα启动子。当该宿主细胞是大肠杆菌时,可使用以下任何启动子:trp启动子,lac启动子,recA启动子,λpL启动子,lpp启动子,T7启动子等。当该宿主是芽孢杆菌属时,可使用以下任何启动子:SPO1启动子,SPO2启动子,penP启动子等。当该宿主是酵母时,可使用任何以下启动子:Gal1/10启动子,PHO5启动子,PGK启动子,GAP启动子,ADH启动子等。当该宿主是昆虫细胞时,可以使用任何以下启动子:多角体蛋白启动子,P10启动子等。当该宿主是植物细胞时,可以使用任何以下启动子:CaMV35S启动子,CaMV19S启动子,NOS启动子等。

[0723] 在一些实施方式中,按照需要,该表达载体可能含有增强子,剪接信号,终止子,polyA附加信号,选择标记物(诸如抗药性基因,营养缺陷型互补基因等),复制起点等。

[0724] 可以凭借在本身已知的体外转录系统中通过使用编码DNA(该DNA编码上述核酸序列-识别模块和/或核酸碱基转换酶)的载体作为模板转录为mRNA,来制备编码如本文所述

的蛋白结构域的RNA。

[0725] 可以通过将编码融合蛋白的表达载体引入宿主细胞中并培养该宿主细胞来表达本发明的融合蛋白。可用于本发明的宿主细胞包含细菌细胞,酵母,昆虫细胞,哺乳动物细胞等。

[0726] 埃希氏菌属包含大肠杆菌K12.cndot.DH1[Proc.Natl.Acad.Sci.、美国、60、160(1968)],大肠杆菌JM103[Nucleic Acids Research、9、309(1981)],大肠杆菌JA221[Journal of Molecular Biology、120、517(1978)],大肠杆菌HB101[Journal of Molecular Biology、41、459(1969)],大肠杆菌C600[Genetics、39、440(1954)]等。

[0727] 芽孢杆菌属包含枯草芽孢杆菌M1114[Gene、24、255(1983)],枯草芽孢杆菌207-21[Journal of Biochemistry、95、87(1984)]等。

[0728] 可用于表达本发明的融合蛋白的酵母包含酿酒酵母AH22,AH22R-,NA87-11A,DKD-5D,20B-12,粟酒裂殖酵母NCYC1913,NCYC2036,巴斯德毕赤酵母(Pichia pastoris)KM71等。

[0729] 使用例如病毒载体,诸如AcNPV,在昆虫细胞中表达融合蛋白。昆虫宿主细胞包含以下任何细胞系:甘蓝菜行军虫幼虫-衍生的确立细胞系(草地贪夜蛾细胞;Sf细胞),衍生自粉纹夜蛾(Trichoplusia ni)的中-肠的MG1细胞,衍生自粉纹夜蛾的卵(巢)的“High Five(击掌相庆)TM.”细胞,甘蓝夜蛾(Mamestra brassicae)-衍生的细胞,盐泽灯蛾(Estigmena acrea)-衍生的细胞等。当病毒是BmNPV时,使用了家蚕-衍生的细胞系(家蚕N细胞;BmN细胞)等。Sf细胞包含,例如,Sf9细胞(ATCC CRL1711),Sf21细胞[以上所有、In Vivo、13、213-217(1977)]等。

[0730] 至于昆虫,使用了例如家蚕、果蝇、蟋蟀等的幼虫来表达融合蛋白(Nature、315、592(1985))。

[0731] 哺乳动物细胞系可用于表达融合蛋白。此种细胞系包含猴COS-7细胞,猴Vero细胞,中华仓鼠卵巢(CHO)细胞,dhfr基因-缺陷型CHO细胞,小鼠L细胞,小鼠AtT-20细胞,小鼠骨髓瘤细胞,大鼠GH3细胞,人类FL细胞等,分化多能的干细胞(诸如人类和其他哺乳动物的iPS细胞,ES细胞等),以及从各种组织制备的原代培养细胞。此外,亦可以使用斑马鱼胚胎,爪蟾卵母细胞等。

[0732] 可使用熟练技术人员熟知的方法在培养物中维持植物细胞。植物细胞培养涉及悬浮培养的细胞,愈伤组织,原生质体,叶节段,根节段等,其是从各种植物(例如谷物,诸如水稻,小麦,玉米等,商品作物,诸如番茄,黄瓜,茄子,康乃馨,洋桔梗,烟草,拟南芥)中制备的。

[0733] 所有上述宿主细胞可能是单倍体(单套体),或多倍体(例如,二倍体,三倍体,四倍体等)。在传统的突变引入方法中,原则上仅将突变引入一个同源染色体以产生异种基因类型。因此,除非发生了显性突变,否则所期待的表型不会表达出来,而不幸的是纯合子需要人力和时间。相比之下,根据本发明,由于可以将突变引入基因组中所述同源染色体上的任何等位基因,即使在隐性突变的情况下也可以在仅一代中表达所期待的表型,这是极其有用的,因为其解决了传统方法的难题。

[0734] 使用任何转染方法(例如,溶菌酶方法,感受态方法,PEG方法,CaCl2共沉淀方法,电穿孔方法,显微注射方法,粒子枪方法,脂转染方法,农杆菌方法等)将编码本发明的融合

蛋白的表达载体引入宿主细胞中。基于待转染的宿主细胞而选择转染方法。

[0735] 可以根据例如Proc.Natl.Acad.Sci.美国、69、2110(1972)和Gene、17、107(1982)等中所述方法转化大肠杆菌。可以根据例如Molecular&General Genetics、168、111(1979)等中所述方法将载体引入芽孢杆菌属中。可以根据例如Methods in Enzymology、194、182-187(1991),Proc.Natl.Acad.Sci.美国、75、1929(1978)等中所述方法将载体引入酵母细胞中。可以根据例如Bio/Technology、6、47-55(1988)等中所述方法将载体引入昆虫细胞中。可以根据例如《细胞工程增刊卷8、新细胞工程实验方案》、263-267(1995)(东京秀潤社出版)和Virology、52、456(1973)中所述方法将载体引入哺乳动物细胞中。

[0736] 根据已知方法培养包括本发明的表达载体的细胞,所述方法根据宿主而变化。例如,当培养大肠杆菌或芽孢杆菌属细胞时,优选液体培养基作为用于培养的培养基。该培养基优选地含有碳源,氮源,无机物质和转化株生长所必需的其他组分。碳源的例子包含葡萄糖,糊精,可溶性淀粉,蔗糖等;氮源的例子包含无机或有机物质,诸如铵盐,硝酸盐,玉米浆,蛋白胨,酪蛋白,肉提取物,豆饼,马铃薯提取物等;而无机物质的例子包含氯化钙,磷酸二氢钠,氯化镁等。该培养基亦可含有酵母提取物,维生素,生长促进因子等。该培养基的pH优选地是在约5至约8之间。

[0737] 作为用于培养大肠杆菌的培养基,例如,优选的是含有葡萄糖,酪蛋白氨基酸的M9培养基[Journal of Experiments in Molecular Genetics、431-433、冷泉港实验室、纽约1972]。必要时,例如,剂诸如3β-吲哚丙烯酸可被加入该培养基中以确保启动子的有效功能。大肠杆菌通常在约15至约43℃下培养。必要时,可执行曝气和搅拌。

[0738] 芽孢杆菌属通常在约30至约40℃下培养。必要时,可执行曝气和搅拌。

[0739] 适于培养酵母的培养基的例子包含Burkholder最低限度培养基[Proc.Natl.Acad.Sci.美国、77、4505(1980)],含有0.5%酪蛋白氨基酸的SD培养基[Proc.Natl.Acad.Sci.美国、81、5330(1984)]等。该培养基的pH优选地是约5至约8。该培养通常是在约20℃至约35℃下执行。必要时,可执行曝气和搅拌。

[0740] 作为用于培养昆虫细胞或昆虫的培养基,使用了含有添加剂,诸如去活化的10%牛血清等的Grace's昆虫培养基[Nature、195、788(1962)]。该培养基的pH优选地是约6.2至约6.4。细胞在约27℃下培养。必要时,可执行曝气和搅拌。

[0741] 作为用于培养哺乳动物细胞的培养基,使用了例如,含有约5%至约20%胎牛血清的最低必需培养基(MEM)[Science、122、501(1952)],Dulbecco's改良Eagle培养基(DMEM)[Virology、8、396(1959)],RPMI 1640培养基[The Journal of the American Medical Association、199、519(1967)],199培养基[Proceeding of the Society for the Biological Medicine、73、1(1950)]等。该培养基的pH优选地是约6至约8。培养在约30℃至约40℃下进行。必要时,可执行曝气和搅拌。

[0742] 作为用于培养植物细胞的培养基,例如,使用了MS培养基,LS培养基,B5培养基等。该培养基的pH优选地是约5至约8。培养通常在约20℃至约30℃下进行。必要时,可执行曝气和搅拌。

[0743] 当以高等真核细胞(诸如动物细胞、昆虫细胞、植物细胞等)作为宿主细胞时,将编码本发明的碱基编辑系统的DNA引入宿主细胞中,且所述DNA是在诱导型启动子(例如,金属硫蛋白启动子(由重金属离子诱导),热休克蛋白启动子(由热休克诱导),Tet-ON/Tet-OFF

系统启动子(通过添加或去除四环素或其衍生物),类固醇-应答性启动子(由类固醇激素或其衍生物诱导)等)的调控下,在适当的阶段将诱导物质添加到培养基中(或从培养基中去除)以诱导表达所述核酸-修饰酶复合体,培养一定时间以完成碱基编辑并且,将突变引入靶标基因,可以实现所述碱基编辑系统的瞬时表达。

[0744] 原核细胞,诸如大肠杆菌等,可以利用诱导型启动子。诱导型启动子的例子包含但不限于lac启动子(由IPTG诱导),cspA启动子(由冷休克诱导),araBAD启动子(由阿拉伯糖诱导)等。

[0745] 或者,当以高等真核细胞(诸如动物细胞、昆虫细胞、植物细胞等)作为宿主细胞时,也可以利用上述的诱导型启动子作为载体去除机制。即,给载体安装了在宿主细胞中有功能的复制起点,以及编码复制所需蛋白质的核酸(例如,SV40 ori和大T抗原,oriP和EBNA-1等,其用于动物细胞),而编码所述蛋白的核酸的表达是被上述诱导型启动子所调控。因此,虽然所述载体在诱导物质存在时是自主复制的,但当该诱导物质被去除时,就没有了自主复制,并且所述载体会随着细胞分裂而自然地掉队消失(在Tet-OFF系统载体中,通过添加四环素和多西环素使自主复制变得不可能)。

[0746] 递送系统

[0747] 核碱基编辑器和gRNA的基于-核酸的递送

[0748] 可以通过本领域已知的方法或如本文所述,将根据本公开揭露的编码核碱基编辑器(例如多-效应器核碱基编辑器)的核酸在体外或在体内施用于受试者或递送进入细胞。在一个实施方式中,核碱基编辑器或多-效应器核碱基编辑器的递送是可以通过例如载体(例如病毒或非-病毒载体),基于非-载体的方法(例如使用裸DNA,DNA复合体,脂质纳米颗粒),或其组合。

[0749] 编码核碱基编辑器或多-效应器核碱基编辑器的核酸能以裸DNA或RNA的形式直接被递送进入细胞(例如造血细胞或它们的组细胞,造血干细胞和/或诱导的多能干细胞),例如通过转染或电穿孔的手段,或偶联到可促进所述靶细胞摄取的分子(例如,N-乙酰半乳糖胺)上。也可以使用核酸载体,诸如本文所述的载体。

[0750] 核酸载体可包括一个或多个编码本文所述融合蛋白的结构域的序列。载体亦可包括编码信号肽(例如,用于核定位,核仁定位或线粒体定位)的序列,该序列与编码蛋白质的一序列相关联(例如,插入进或融合到)。作为一个例子,核酸载体可以包含Cas9编码序列,其包含一个或多个核定位序列(例如,来自SV40的核定位序列),以及脱氨酶(例如腺苷脱氨酶和/或胞苷脱氨酶)。

[0751] 所述核酸载体亦可包含任何适合数目的调节/控制元件,例如启动子,增强子,内含子,多腺苷酸化信号,Kozak共有序列或内部核糖体进入位点(IRES)。这些元件在本领域中是众所周知的。对于造血细胞,适合的启动子可以包含IFNbeta或CD45。

[0752] 根据本公开揭露的核酸载体包含重组病毒载体。示例性的病毒载体在本文中列出。也可以使用本领域已知的其他病毒载体。另外,病毒微粒可用于递送以核酸和/或肽形式存在的碱基编辑系统组分。例如,“空”病毒微粒可以被组装以容纳任何适合的货物。病毒载体和病毒微粒也可以被工程化以并入靶向配体以改变靶标的组织特异性。

[0753] 除病毒载体外,非-病毒载体可用于递送根据本公开揭露的编码基因组编辑系统的核酸。非-病毒核酸载体的一重要类别是纳米微粒,其可以是有机的或无机的。纳米微粒

是本领域众所周知的。任何适合的纳米微粒设计均可以用于递送基因组编辑系统组分或编码此类组分的核酸。例如,有机的(例如脂质和/或聚合物)纳米微粒可适用于作为本公开揭露的某些实施方式中的递送载具。用于纳米微粒制剂和/或基因移转的示例性的脂质显示在下面的表6中。

[0754] 表6

[0755]

用于基因移转的脂质		
脂质	缩写	特性
1,2-二油酰基-sn-甘油-3-磷脂酰胆碱	DOPC	辅助者

[0756]

用于基因移转的脂质		
脂质	缩写	特性
1,2-二油酰基-sn-甘油-3-磷脂酰乙醇胺	DOPE	辅助者
胆固醇		辅助者
N-[1-(2,3-二油氧基)丙基]N,N,N-三甲基氯化铵	DOTMA	阳离子
1,2-二油酰基氧基-3-三甲基铵-丙烷	DOTAP	阳离子
二(十八烷基)酰胺基甘氨酰基精胺	DOGS	阳离子
N-(3-氨基丙基)-N,N-二甲基-2,3-双(十二烷基氧基)-1-溴化丙铵	GAP-DLRIE	阳离子
十六烷基三甲基溴化铵	CTAB	阳离子
6-月桂氧基己基鸟氨酸酯	LHON	阳离子
1-(2,3-二油酰基氧基丙基)-2,4,6-三甲基吡啶鎓	20c	阳离子
2,3-二油基氧基-N-[2(精胺羧酰胺基-乙基]-N,N-二甲基-1-三氟乙酸丙铵	DOSPA	阳离子
1,2-二油基-3-三甲基铵-丙烷	DOPA	阳离子
N-(2-羟基乙基)-N,N-二甲基-2,3-双(十四烷氧基)-1-溴化丙铵	MDRIE	阳离子
二肉豆蔻氧基丙基 二甲基 羟基乙基 溴化铵	DMRI	阳离子
3β-[N-(N',N'-二甲基氨基乙烷)-氨基甲酰基]胆固醇	DC-Chol	阳离子
双-胍-三(2-氨基乙基)胺-胆固醇	BGTC	阳离子
1,3-二脱氧-2-(6-羧基-精胺基)-丙基酰胺	DOSPER	阳离子
二(甲基十八烷基)溴化铵	DDAB	阳离子
二(十八烷基)酰胺基甘氨酰基亚精胺	DSL	阳离子
外消旋-[(2,3-二(十六烷基)氧基丙基)(2-羟基乙基)]-二甲基氯化铵	CLIP-1	阳离子
外消旋-[2(2,3-二(十六烷基)氧基丙基-氧基甲基氧基)乙基] 三甲基溴化铵	CLIP-6	阳离子
乙基二肉豆蔻酰基磷脂酰胆碱	EDMPC	阳离子
1,2-二硬脂基氧基-N,N-二甲基-3-氨基丙烷	DSDMA	阳离子

用于基因移转的脂质		
脂质	缩写	特性
1,2-二肉豆蔻酰基-三甲基铵 丙烷	DMTAP	阳离子
O,O'-二肉豆蔻基-N-赖氨酰天冬氨酸	DMKE	阳离子
1,2-二硬脂酰基-sn-甘油-3-乙基磷酸胆碱	DSEPC	阳离子
N-棕榈酰基 D-赤式构型-鞘氨醇基氨基甲酰基-精胺	CCS	阳离子
N-t-叔丁基-NO-十四烷基-3-十四烷基氨基丙脒	diC14-脒	阳离子
十八碳烯酰氧基[乙基-2-十七碳烯基-3 羟基乙基]	DOTIM	阳离子
咪唑鎓氯盐		
N1 -胆固醇羰氧基-3,7-二氮杂壬烷-1,9-二胺	CDAN	阳离子
2-(3-[双(3-氨基-丙基)-氨基]丙基氨基)-N-	RPR209120	阳离子
二(十四烷基)氨基甲酰基甲基乙基乙酰胺		
1,2-二亚油基氧基-3-二甲基氨基丙烷	DLinDMA	阳离子
2,2-二亚油基-4-二甲基氨基乙基-[1,3]-二氧戊环	DLin-KC2-D	阳离子
	MA	
二亚油基-甲基-4-二甲基氨基丁酸酯	DLin-MC3-D	阳离子
	MA	

[0757]

[0758] 表7列出了用于基因移转和/或纳米微粒制剂的示例性的聚合物。

[0759] 表7

用于基因移转的聚合物	
聚合物	缩写
聚（乙烯）二醇，即聚乙二醇	PEG
聚乙烯亚胺	PEI
二硫代（琥珀酸亚胺丙酸酯）	DSP
二甲基-3,3'-二硫代双丙亚氨酸二甲酯	DTBP
聚(乙烯亚胺)双氨基甲酸酯	PEIC
聚(L-赖氨酸)	PLL
组氨酸修饰的 PLL	
聚(N-乙烯基吡咯烷酮)	PVP

[0760]

[0761]

用于基因移转的聚合物	
聚合物	缩写
聚(丙烯亚胺)	PPI
聚(酰胺基胺)	PAMAM
聚(酰胺基乙烯亚胺)	SS-PAEI
三乙烯四胺	TETA
聚(β-氨基酯)	
聚(4-羟基-L-脯氨酸酯)	PHP
聚(烯丙基胺)	
聚(α-[4-氨基丁基]-L-乙醇酸)	PAGA
聚(D,L-乳酸-共聚-乙醇酸)	PLGA
聚(N-乙基-4-乙烯基吡啶鎓溴盐)	
聚(磷腈)	PPZ
聚(磷酸酯)	PPE
聚(氨基磷酸酯)	PPA
聚(N-2-羟丙基甲基丙烯酰胺)	pHPMA
聚 (2-(二甲氨基)甲基丙烯酸乙酯)	pDMAEMA
聚(2-氨基乙基丙烯酸磷脂)	PPE-EA
壳聚糖	
半乳糖基化壳聚糖	
N-十二烷基化壳聚糖	
组蛋白	
胶原	
葡聚糖-精胺	D-SPM

[0762] 表8总结了以供编码本文所述融合蛋白的多核苷酸的递送方法。

[0763] 表8

[0764]

递送	载体/模式	递送进入非-分裂细胞	表达的存续期间	基因组整合	被递送的分子的类型

[0765]

递送	载体/模式	递送进入非-分裂细胞	表达的存续期间	基因组整合	被递送的分子的类型
物理的	例如，电穿孔，粒子枪，磷酸钙转染	是	瞬时的	否	核酸和蛋白质
病毒的	逆转录病毒	否	稳定的	是	RNA
	慢病毒	是	稳定的	是/否（有修饰）	RNA
	腺病毒	是	瞬时的	否	DNA
	腺相关病毒(AAV)	是	稳定的	否	DNA
	牛痘病毒	是	非常瞬时	否	DNA
	单纯疱疹病毒	是	稳定的	否	DNA
非-病毒的	阳离子脂质体	是	瞬时的	取决于被递送的是什么	核酸和蛋白质
	聚合的纳米微粒	是	瞬时的	取决于被递送的是什么	核酸和蛋白质
生物性非-病毒的递送载具	减毒的细菌	是	瞬时的	否	核酸
	工程化的噬菌体	是	瞬时的	否	核酸
	哺乳动物病毒-样粒子	是	瞬时的	否	核酸

[0766]

递送	载体/模式	递送进入非-分裂细胞	表达的存续期间	基因组整合	被递送的分子的类型
	生物性脂质体：红细胞血影与外泌体	是	瞬时的	否	核酸

[0767] 在另一方面，基因组编辑系统各组分或编码该各组分的核酸(例如，核酸结合蛋白(诸如Cas9或其变体)，以及靶向基因组目标核酸序列的gRNA)的递送，可能是通过将核糖核蛋白(RNP)递送至细胞来实现。该RNP包括与所述靶向gRNA形成复合体的所述核酸结合蛋白(例如Cas9)。可使用已知方法将RNP递送至细胞，诸如电穿孔、核转染或阳离子脂质-介导的方法，例如，如Zuris、J.A.等、2015、Nat.Biotechnology、33(1):73-80所报道的。在CRISPR碱基编辑系统中使用RNP是有利的，特别是对于难以转染的细胞，诸如原代细胞。此外，RNP还可以缓和细胞中伴随蛋白质表达可能出现的困境，尤其是当真核启动子，例如CRISPR质粒中可能使用的CMV或EF1A，没有很好地表达时。有利的是，RNP的使用不需要将外源DNA递送到细胞中。此外，由于包括核酸结合蛋白和gRNA复合体的RNP会随时间降解，因此RNP的使用有限制脱靶效应的潜力。以类似于基于质粒的技法的方式，RNP可用于递送结合蛋白(例如Cas9变体)和导引同源定向修复(HDR)。

[0768] 用于驱动编码碱基编辑器的核酸分子的表达的启动子可以包含AAV ITR。这有利于消除对额外启动子元件的需要，该元件会占据所述载体中的空间。所释放的额外空间可用于驱动额外的元件的表达，诸如向导核酸或可筛选标记。ITR活性是相对弱的，所以它可用于减低因所选核酸酶的过度表达而导致的潜在毒性。

[0769] 任何适合的启动子都可以用于驱动所述碱基编辑器的表达，以及适当时所述向导核酸的表达。对于普遍表达，可以使用的启动子包含CMV、CAG、CBh、PGK、SV40、铁蛋白的重链或轻链等。对于脑或其他CNS细胞表达，适合的启动子可以包含：用于所有神经元的SynapsinI，用于兴奋性神经元的CaMKIIα，用于GABA能神经元的GAD67或GAD65或VGAT，等。对于肝脏细胞表达，适合的启动子包含白蛋白启动子。对于肺细胞表达，适合的启动子可以包含SP-B。对于内皮细胞，适合的启动子可以包含ICAM。对于造血细胞，适合的启动子可以包含IFNβ或CD45。对于成骨细胞，适合的启动子可以包含OG-2。

[0770] 在一些实施方式中，本公开揭露的碱基编辑器具有足够小的尺寸以允许分开的多个启动子驱动所述碱基编辑器和相容的向导核酸在同一核酸分子内的表达。例如，载体或病毒载体可包括与编码所述碱基编辑器的核酸可操作地链接的第一启动子和与所述向导核酸可操作地链接的第二启动子。

[0771] 用于驱动向导核酸表达的启动子可以包含Pol III启动子，诸如U6或H1。使用Pol II启动子和内含子框来表达gRNA腺相关病毒(AAV)。

[0772] 病毒载体

[0773] 因此，本文所述的碱基编辑器可以凭借病毒载体递送。在一些实施方式中，本文所述的碱基编辑器可以在核酸上被编码，该核酸是被包含在病毒载体中。在一些实施方式中，

所述碱基编辑器系统的一个或多个组分可以在一个或多个病毒载体上被编码。例如,碱基编辑器和向导核酸可以在单一病毒载体上被编码。在其他实施方式中,该碱基编辑器和向导核酸是在不同的病毒载体上被编码。在任一情形中,该碱基编辑器和向导核酸可以各自可操作地链接到启动子和终止子上。病毒载体上被编码的各个组分的组合可以通过所选病毒载体的基因容量大小限制来确定。

[0774] 使用基于RNA或DNA病毒的系统来递送碱基编辑器,利用了用于将病毒靶向到培养物中或宿主体内的特定细胞的和用于将病毒净载运送到细胞核或宿主细胞基因组的高度进化的过程。病毒载体可以直接施用于培养中的细胞,患者(体内),抑或它们可以用于在体外处理细胞,并且可以任选地将经修饰的细胞施用于患者(离体)。传统的基于病毒的系统可以包含用于基因移转的逆转录病毒、慢病毒、腺病毒、腺相关病毒和单纯疱疹病毒载体。凭借所述逆转录病毒、慢病毒和腺相关病毒基因移转方法,整合到宿主基因组中是有可能的,通常导致插入的转基因的长期表达。此外,在许多不同的细胞类型和靶组织中都观察到了高转导效率。

[0775] 病毒载体可以包含慢病毒(例如基于HIV和FIV的载体)、腺病毒(例如AD100)、逆转录病毒(例如Maloney鼠白血病病毒MML-V)、疱疹病毒载体(例如HSV-2)和腺相关病毒(AAV)或其他质粒或病毒载体类型,特别是使用如下制剂和剂量的质粒或病毒载体类型,例如来自美国专利编号8,454,972(适于腺病毒的制剂、剂量)、美国专利编号8,404,658(适于AAV的制剂、剂量)和美国专利编号5,846,946(适于DNA质粒的制剂、剂量)以及来自临床试验和关于涉及慢病毒、AAV和腺病毒的临床试验的出版物。例如,对于AAV,给药途径、制剂和剂量可以如同美国专利编号8,454,972中的那些和如同涉及AAV的临床试验中的那些。对于腺病毒,给药途径、制剂和剂量可以如同美国专利编号8,404,658中的那些和涉及腺病毒的临床试验中的那些。对于质粒递送,给药途径、制剂和剂量可以如同美国专利编号5,846,946中的那些和如同涉及质粒的临床研究中的那些。剂量可以基于或外推到平均70公斤的个体(例如男性成年人),并且可以针对不同重量和物种的患者、受试者、哺乳动物进行调整。给药频率在医学或兽医从业者(例如医师、兽医)的领域内,取决于通常的因子,包含患者或受试者的年龄、性别、一般健康状况、其他状况以及正在解决的特定病况或症状。该病毒载体可以被注射到目标组织中。对于细胞-类型特异性的碱基编辑,碱基编辑器和可选的向导核酸的表达可由细胞-类型特异性启动子驱动。

[0776] 逆转录病毒的向性可以通过掺入外来包膜蛋白来改变,其扩大了靶细胞的潜在靶群体。慢病毒载体是能够转导或感染非-分裂细胞并且通常产生高病毒滴度的逆转录病毒载体。因此,逆转录病毒基因移转系统的选择将取决于所述靶组织。逆转录病毒载体由顺式-作用的长末端重复序列组成,其包装容量最高为6-10kb的外源序列。最小的顺式-作用LTR足以用于所述载体的复制和包装,然后用于将治疗性基因整合到靶细胞中以提供恒定的转基因表达。广泛使用的逆转录病毒载体包含基于鼠白血病病毒(MuLV)、长臂猿白血病病毒(GaLV)、猿猴免疫缺陷病毒(SIV)、人类免疫缺陷病毒(HIV)及其组合的那些载体(参见例如Buchscher等、J.Virol.66:2731-2739(1992);Johann等、J.Virol.66:1635-1640(1992);Sommnerfelt等、Virol.176:58-59(1990);Wilson等、J.Virol.63:2374-2378(1989);Miller等、J.Virol.65:2220-2224(1991);PCT/US94/05700)。

[0777] 逆转录病毒载体,尤其是慢病毒载体,可能需要小于给定长度的多核苷酸序列以

有效整合到靶细胞中。例如,长度大于9kb的逆转录病毒载体与较小的载体相比可导致低的病毒滴度。在一些实施方式中,本公开揭露的碱基编辑器具有足够的大小以使得其能够经由逆转录病毒载体有效地包装和递送到靶细胞中。在一些实施方式中,碱基编辑器的大小使得其即使在与向导核酸和/或可靶向核酸酶系统的其他组分一起表达时也允许有效包装和递送。

[0778] 在优选瞬时表达的应用中,可以使用基于腺病毒的系统。基于腺病毒的载体能够在许多细胞类型中具有非常高的转导效率并且不需要细胞分裂。使用此种载体,人们已经获得了高滴度和高表达水平。这个载体可以在相对简单的系统中大量生产。腺相关病毒("AAV")载体也可用于用靶核酸转导细胞,例如在核酸和肽的体外生产中,以及用于体内和离体的基因治疗程序(参见例如,West等、Virology160:38-47(1987);美国专利编号4,797,368;WO 93/24641;Kotin、Human GeneTherapy 5:793-801(1994);Muzyczka、J.Clin.Invest.94:1351(1994).重组AAV载体的构建描述于一些出版物中,包含:美国专利编号5,173,414;Tratschin等、Mol.Cell.Biol.5:3251-3260(1985);Tratschin等、Mol.Cell.Biol.4:2072-2081(1984);Hermonat&Muzyczka、PNAS 81:6466-6470(1984);和Samulski等、J.Virol.63:03822-3828(1989)。

[0779] AAV是一种小的单股DNA依赖病毒,其属于细小病毒家族。该4.7kb野生型(wt)AAV基因组是由分别编码四个复制蛋白和三个衣壳蛋白的两个基因组成,并且该基因组的两侧翼都是145-bp的反向末端重复序列(ITR)。所述病毒体由三种衣壳蛋白Vp1、Vp2和Vp3组成,它们是来自同一开放阅读框但出自差异化剪接(Vp1)以及选择性翻译起始位点(分别为Vp2和Vp3)而以1:1:10的比率生产出来。Vp3是病毒体中最丰富的亚单元,并参与定义了该病毒向性的细胞表面的受体识别。在Vp1独特的N末端内已鉴别出一在病毒感染力中起作用的磷脂酶结构域。

[0780] 与野生型AAV相似,重组AAV(rAAV)利用顺式-作用的145bp ITR侧翼包围载体的转基因盒,而提供了最长为4.5kb可用于包装外源DNA。感染之后,rAAV可以表达本发明的融合蛋白,并且不需整合到宿主基因组中仍可持久存在,其通过附加体的形式存在于环状的头尾相连的连环体中。尽管在体外和体内使用此系统而成功的rAAV的例子很多,但是当基因编码序列的长度等于或大于wt AAV基因组时,有限的包装能力限制了AAV-介导的基因递送的使用。

[0781] 可以根据应用选择病毒载体。例如,对于体内基因递送,AAV可能优于其他病毒载体。在一些实施方式中,AAV符合低毒性的要求,这可能是因为纯化方法不需要细胞颗粒的超速-离心,其可以激活免疫反应。在一些实施方式中,AAV符合低概率造成插入诱变的要求,因为它不整合到宿主基因组中。腺病毒通常被用作疫苗,因为它们诱导强烈的免疫原应答。该病毒载体的包装容量可以限制可以被包装到该载体中的碱基编辑器的大小。

[0782] AAV的包装容量为约4.5Kb或4.75Kb(包含两个145碱基反向末端重复序列(ITR)在内)。这意味着所揭露的碱基编辑器以及启动子和转录终止子可以适配到单一病毒载体中。大于4.5或4.75Kb的构建体会导致病毒产量的显著降低。例如,SpCas9是相当大的,其基因本身超过4.1Kb,这使得其不易被装入AAV。因此,本公开揭露的实施方式包含利用所揭露的碱基编辑器,其长度比传统的碱基编辑器短。在一些实例中,该碱基编辑器小于4kb。所揭露的碱基编辑器可小于4.5kb、4.4kb、4.3kb、4.2kb、4.1kb、4kb、3.9kb、3.8kb、3.7kb、3.6kb、

3.5kb、3.4kb、3.3kb、3.2kb、3.1kb、3kb、2.9kb、2.8kb、2.7kb、2.6kb、2.5kb、2kb或1.5kb。在一些实施方式中,所揭露的碱基编辑器的长度是4.5kb或较短。

[0783] AAV可以是AAV1、AAV2、AAV5或其任何组合。可以根据被靶向的细胞选择AAV的类型;例如,对于靶向脑或神经元细胞,可以选择AAV血清型1、2、5或是杂合衣壳AAV1、AAV2、AAV5或其任何组合;对于靶向心脏组织,可以选择AAV4。AAV8可用于对肝脏的递送。关于这些细胞的确定AAV血清型的表列目录可以在Grimm、D.等、J.Virol.82:5887-5911(2008)中找到。

[0784] 慢病毒是复杂的逆转录病毒,其在有丝分裂和有丝分裂后细胞中都具有感染和表达它们基因的能力。最广为人知的慢病毒是人类免疫缺陷病毒(HIV),其使用其他病毒的包膜糖蛋白来靶向广泛的细胞类型。

[0785] 可以如下制备慢病毒。克隆pCasES10(其含有慢病毒移转质粒主干)后,将低传代(p=5)的HEK293FT接种到T-75烧瓶中,在转染前一天在含10%胎牛血清不含抗生素的DMEM中达到50%汇合。20小时后,将培养基更换为OptiMEM(无血清)培养基,4小时后进行转染。细胞用10μg慢病毒移转质粒(pCasES10)和以下包装质粒转染:5μg pMD2.G(VSV-g假型)和7.5μg psPAX2(gag/pol/rev/tat)。转染可在4mL OptiMEM中使用阳离子脂质递送剂(50μl Lipofectamine2000和100ul Plus试剂)进行。6小时后,将该培养基更换为不含抗生素的含10%胎牛血清的DMEM。这些方法在细胞培养过程中使用血清,但无血清的方法是优选的。

[0786] 可以如下纯化慢病毒。48小时后收获病毒上清液。上清液首先清除残渣并经0.45μm低蛋白结合(PVDF)过滤器过滤。然后将它们在超速离心机中以24,000rpm的速度旋转2小时。病毒沉淀物在50μl DMEM中重悬于4℃下过夜。然后将它们等分分装并立即在-80℃下冷冻。

[0787] 在另一个实施方式中,还检视了基于马传染性贫血病毒(EIAV)的最小非-灵长类慢病毒载体。在另一个实施方式中,RetinoStat®TM,一种基于马传染性贫血病毒的慢病毒基因治疗载体,其表达经由视网膜下注射而预期被递送的血管抑制素蛋白质内皮抑素和血管抑素。在另一个实施方式中,检视了使用自我-灭活的慢病毒载体。

[0788] 该系统的任何RNA,例如向导RNA或编码碱基编辑器的mRNA,都能以RNA的形式被递送。可以使用体外转录生成编码碱基编辑器的mRNA。例如,可以使用含有以下元件的PCR框合成核酸酶mRNA:T7启动子、可选的kozak序列(GCCACC)、核酸酶序列和3'UTR,诸如来自β珠蛋白-polyA尾巴的3'UTR。该框可用于经由T7聚合酶的转录。也可以使用体外转录从含有T7启动子的框中转录向导多核苷酸(例如gRNA),紧随该T7启动子的是序列"GG"和向导多核苷酸序列。

[0789] 为了增强表达并减低可能的毒性,可以修饰编码碱基编辑器的序列和/或向导核酸以包含一种或多种经修饰的核苷,例如,使用假-U或5-甲基-C的核苷。

[0790] AAV载体的小的包装能力使得超过该大小的一些基因的递送和/或大的生理调节元件的使用具有挑战性。这些挑战可以通过例如将待递送的一个(或多个)蛋白质分成两个或更多个片段来解决,其中N末端片段与断裂内含肽-N融合,而C末端片段与断裂内含肽-C融合。然后将这些片段包装到两个或多个AAV载体中。在一个实施方式中,内含肽用于连接被嫁接到AAV衣壳蛋白上的多效应器碱基编辑器蛋白的片段或部分。如本文所用,"内含肽"是指将其两侧翼的N末端和C末端的外显肽(例如,待连接的片段)连接起来的自-剪接的蛋

白质内含子(例如,肽)。某些内含肽用于连接异源性蛋白片段是描述于例如Wood等、J.Biol.Chem.289(21);14512-9(2014)中。例如,当被融合到分开的蛋白质片段时,内含肽IntN和内含肽IntC相互识别、剪接自身并同时将侧翼的与它们融合的蛋白质片段的N末端和C末端的外显肽连接起来,从而重构了来自两个蛋白质片段的全长蛋白质。其他适合的内含肽对于本领域技术人员将是显而易见的。

[0791] 本发明的融合蛋白的片段的长度可以变化。在一些实施方式中,蛋白质片段的长度范围为从2个氨基酸至约1000个氨基酸。在一些实施方式中,蛋白质片段的长度范围为从约5个氨基酸至约500个氨基酸。在一些实施方式中,蛋白质片段的长度范围为从约20个氨基酸至约200个氨基酸。在一些实施方式中,蛋白质片段的长度范围为从约10个氨基酸至约100个氨基酸。其他长度的适合的蛋白质片段对本领域技术人员将是显而易见的。

[0792] 在一个实施方式中,通过将大的转基因表达框断裂成两个分开的两半(5′和3′末端,或头和尾)来生成双重AAV载体,其中该框的每一半被包装进单个AAV载体中(其<5kb)。然后,通过两个双重AAV载体共-感染同一细胞,然后进行以下操作即可实现全长转基因表达框的重新-组装:(1)5’和3’基因组(双重AAV重叠载体)之间的同源重组(HR);(2)ITR-介导的5’和3’基因组的尾对头的连环体化(双重AAV反式-剪接载体);或(3)这两种机制的组合(双重AAV杂合载体)。在体内使用双重AAV载体导致全长蛋白质的表达。双重AAV载体平台的使用代表了大于4.7kb大小的转基因的有效且可行的基因移转策略。

[0793] 内含肽

[0794] 在一些实施方式中,核酸酶的一部分或片段(例如Cas9)与内含肽融合。该核酸酶可以被融合到该内含肽的N端或C端。在一些实施方式中,融合蛋白的一部分或片段融合到内含肽并且融合到AAV衣壳蛋白。所述内含肽、核酸酶和衣壳蛋白能以任何排列被融合在一起(例如,核酸酶-内含肽-衣壳,内含肽-核酸酶-衣壳,衣壳-内含肽-核酸酶等)。在一些实施方式中,将内含肽的N端融合到融合蛋白的C端,并将内含肽的C端融合至AAV衣壳蛋白的N端。

[0795] 内含肽(中介蛋白)是在多种生物体中发现的自动-加工结构域,其执行称为蛋白质剪接的过程。蛋白质剪接是一个多-步骤的生化反应,其包括切开和肽键的形成。尽管蛋白质剪接的内源性底物是在含有内含肽的生物体中发现的蛋白质,但内含肽也可被用于化学性操控几乎任何多肽主干。

[0796] 在蛋白质剪接中,所述内含肽通过切开两个肽键将自身从前体多肽中切除出来,从而经由新肽键的形成来连接两侧翼的外显肽(外部蛋白)序列。这种重排发生在翻译之后(或可能与翻译同步)。内含肽-介导的蛋白质剪接自发地发生,其只需要内含肽结构域的折叠。

[0797] 大约5%的内含肽是断裂内含肽,其被转录并翻译为两个单独的多肽,即N-内含肽和C-内含肽,其各自融合到一个外显肽上。一经翻译,所述内含肽片段自发地和非-共价地组装成正准的内含肽结构以进行反式的蛋白质剪接。蛋白质剪接的机制需要一系列酰基-转移反应,其导致在内含肽-外显肽接合处的两个肽键的切开,以及在N-外显肽和C-外显肽之间新肽键的形成。该过程的启动是通过活化连接N-外显肽和该内含肽N端的肽键。几乎所有内含肽在其N端均具有半胱氨酸或丝氨酸,其可攻击N-外显肽C末端残基的羰基碳。保守的苏氨酸和组氨酸(称为TXXH基序)以及常见的天冬氨酸促进了这种N到O/S的酰基-移位,

其导致线性(硫)酯中间体的形成。接下来,该中间体通过对第一个C-外显肽残基(+1)(该残基是半胱氨酸、丝氨酸或苏氨酸)的亲核攻击被反式-(硫)酯化。所得到的分枝(硫)酯中间体可通过独特的转型作用而被变形:内含肽的高度保守的C末端天冬酰胺的环化反应。(发现于高度保守的HNF基序中的)组氨酸和内含肽次末端的组氨酸可促进该过程,并且也可能涉及所述天冬氨酸。此琥珀酰亚胺的形成反应从反应复合体中切除所述内含肽,并留下通过非-肽键合附接的所述外显肽。此结构以独立于内含肽的方式迅速地重排成一稳定的肽键。

[0798] 在一些实施方式中,碱基编辑器(例如ABE、CBE)的N末端片段融合到断裂内含肽-N上,并且C末端片段融合到断裂内含肽-C上。然后将这些片段包装到两个或多个AAV载体中。某些内含肽用于连接异源性蛋白片段是描述于例如Wood等、J.Biol.Chem.289(21);14512-9(2014)。例如,当被融合到分开的蛋白质片段时,内含肽IntN和内含肽IntC相互识别、剪接自身并同时将侧翼的与它们融合的蛋白质片段的N末端和C末端的外显肽连接起来,从而重构了来自两个蛋白质片段的全长蛋白质。其他适合的内含肽对于本领域技术人员将是显而易见的。

[0799] 在一些实施方式中,ABE在SpCas9的选定区域内的Ala、Ser、Thr或Cys残基处被断裂为N末端片段和C末端片段。这些区域对应于通过Cas9晶体结构分析所鉴别出的环区域。每个片段的N端融合到内含肽-N,而每个片段的C端融合到内含肽C,且该融合是在S303、T310、T313、S355、A456、S460、A463、T466、S469、T472、T474、C574、S577、A589和S590氨基酸位置处,其在以下序列中以粗体大写字母示出。

[0800]
```
  1 mdkkysigld igtnsvgwav itdeykvpsk kfkvlgntdr hsikknliga llfdsgetae
 61 atrlkrtarr rytrrknric ylqeifsnem akvddsffhr leesflveed kkherhpifg
121 nivdevayhe kyptiyhlrk klvdstdkad lrliylalah mikfrghfli egdlnpdnsd
181 vdklfiqlvq tynqlfeenp inasgvdaka ilsarlsksr rlenliaqlp gekknglfgn
241 lialslgltp nfksnfdlae daklqlskdt ydddldnlla qigdqyadlf laaknlsdai
301 llSdilrvnT eiTkaplsas mikrydehhq dltllkalvr qqlpekykei ffdqSkngya
361 gyidggasqe efykfikpil ekmdgteell vklnredllr kqrtfdngsi phqihlgelh
```

[0801]

```
 421 ailrrqedfy pflkdnreki ekiltfripy yvgplArgnS rfAwmTrkSe eTiTpwnfee
 481 vvdkgasaqs fiermtnfdk nlpnekvlpk hsllyeyftv yneltkvkyv tegmrkpafl
 541 sgeqkkaivd llfktnrkvt vkqlkedyfk kieCfdSvei sgvedrfnAS lgtyhdllki
 601 ikdkdfldne enedilediv ltltlfedre mieerlktya hlfddkvmkq lkrrrytgwg
 661 rlsrklingi rdkqsgktil dflksdgfan rnfmqlihdd sltfkediqk aqvsgqgdsl
 721 hehianlags paikkgilqt vkvvdelvkv mgrhkpeniv iemarenqtt qkgqknsrer
 781 mkrieegike lgsqilkehp ventqlqnek lylyylqngr dmyvdqeldi nrlsdydvdh
 841 ivpqsflkdd sidnkvltrs dknrgksdnv pseevvkkmk nywrqllnak litqrkfdnl
 901 tkaergglse ldkagfikrq lvetrqitkh vaqildsrmn tkydendkli revkvitlks
 961 klvsdfrkdf qfykvreinn yhhahdayln avvgtalikk ypklesefvy gdykvydvrk
1021 miakseqeig katakyffys nimnffktei tlangeirkr plietngetg eivwdkgrdf
1081 atvrkvlsmp qvnivkktev qtggfskesi lpkrnsdkli arkkdwdpkk yggfdsptva
1141 ysvlvvakve kgkskklksv kellgitime rssfeknpid fleakgykev kkdliiklpk
1201 yslfelengr krmlasagel qkgnelalps kyvnflylas hyeklkgspe dneqkqlfve
1261 qhkhyldeii eqisefskrv iladanldkv lsaynkhrdk pireqaenii hlftltnlga
1321 paafkyfdtt idrkrytstk evldatlihq sitglyetri dlsqlggd
```

[0802] 使用核碱基编辑器以靶向突变

[0803] 如本文所述,评估了靶向一个或多个突变的核碱基编辑器或多-效应器核碱基编辑器的适宜性。在一个实施方式中,用碱基编辑系统与少量的编码报告基因(例如GFP)的载体一起转导单个目标细胞。这些细胞可以是被本领域已知的任何细胞系,包含永生化的人类细胞系,诸如293T,K562或U20S。或者,可以使用原代(例如人类)细胞。此种细胞可能与最终的细胞靶标有关。

[0804] 可以使用病毒载体执行递送。在一个实施方式中,可使用脂质转染(例如Lipofectamine试剂或Fugene试剂)或通过电穿孔执行转染。转染后,可以通过荧光显微镜或流式细胞术确定GFP的表达,以确认一致且高水平的转染。这些初步的转染可以包括不同的核碱基编辑器,以确定哪些编辑器的组合给出最大的活性。

[0805] 如本文所述评估了所述核碱基编辑器的活性,即,通过对细胞的基因组测序以检测靶序列中的改变。为了Sanger测序,将纯化的PCR扩增子克隆到质粒主干中、进行转化、使用miniprep小量制备并使用单个引物测序。测序也可使用下一代测序技法来进行。当使用下一代测序时,扩增子可能为300-500bp,而其预期的切割位点为不对称安插的。PCR之后,可将下一代测序的接头和索引序列条码(例如Illumina多重接头和索引)添加至扩增子的末端,例如用于高通量测序(例如在Illumina MiSeq系统上)。

[0806] 可以选择在初始测试中诱导了最高水平的靶标特异性改变的融合蛋白,以供进一步评估。

[0807] 在特定实施方式中,所述核碱基编辑器或多-效应器碱基编辑器用于靶向目标多核苷酸。在一个实施方式中,将本发明的核碱基编辑器或多-效应器碱基编辑器与向导RNA一起递送到细胞(例如,造血细胞或其祖细胞、造血干细胞和/或诱导的多能干细胞)中,该向导RNA用于靶向细胞基因组内的目标突变,从而改变所述突变。在一些实施方式中,碱基

编辑器被向导RNA靶向以对目标基因的序列引入一个或多个编辑。

[0808] 在一个实施方式中,核碱基编辑器或多-效应器核碱基编辑器用于靶向调节序列,包含但不限于剪接位点、增强子和转录调节元件。然后使用本领域已知的任何方法测定所述改变对受该调节元件控制的基因的表达的影响。

[0809] 在其他实施方式中,本发明的核碱基编辑器或多-效应器核碱基编辑器用于靶向编码互补决定区(CDR)的多核苷酸,从而在表达出的CDR中产生改变。然后,例如通过测量CDR与其抗原的特异性结合,来测定这些改变对CDR功能的影响。

[0810] 在其他实施方式中,本发明的多-效应器核碱基编辑器用于靶向生物体的基因组内的目标多核苷酸。在一个实施方式中,将本发明的多-效应器核碱基编辑器与向导RNA文库一起递送到细胞中,该向导RNA文库用于铺到细胞基因组内的各种序列上,从而系统地改变整个基因组中的序列。

[0811] 该系统可以包括一个或多个不同的载体。在一个方面,该碱基编辑器是密码子优化过的以供在期望的细胞类型中表达,优选真核细胞,优选哺乳动物细胞或人类细胞。

[0812] 一般而言,密码子优化是指为了在目标宿主细胞中增强的表达而修饰核酸序列的过程,其通过用在该宿主细胞的各基因中更频繁或最频繁使用的密码子取代了所述原生序列的至少一个密码子(例如,约或多于约1、2、3、4、5、10、15、20、25、50个或更多个密码子)并且同时维持了原生的氨基酸序列。各种物种对特定氨基酸的某些密码子展现出特定的偏移。密码子偏倚(生物体之间密码子使用的差异)通常与信使RNA(mRNA)的翻译效率相关,而信使RNA的翻译效率又被认为取决于被翻译密码子的特性和特定转移RNA(tRNA)分子的可得性等。细胞中所选定的tRNA的优势通常反映了肽合成中最常用的密码子。因此,可以基于密码子优化来定制基因以在给定生物体中以供最佳的基因表达。密码子使用表很容易获得,例如,在www.kazusa.orjp/codon/(于2002年7月9日访问)上可得的“密码子使用数据库”,并且这些表格可以通过多种方式加以调整。参见Nakamura、Y.等,“源自国际DNA序列数据库中的密码子使用列表:2000年的现况”Nucl.Acids Res.28:292(2000)。用于密码子优化特定序列以供在特定宿主细胞中表达的计算机算法也是可得的,诸如Gene Forge平台(Aptagen公司;Jacobus,宾夕法尼亚州)。在一些实施方式中,编码工程化的核酸酶的序列中的一个或多个密码子(例如1、2、3、4、5、10、15、20、25、50个或更多个,或所有密码子)对应于特定氨基酸的最常用的密码子。

[0813] 包装细胞通常用于形成能够感染宿主细胞的病毒颗粒。此种细胞包含:包装腺病毒的293细胞和包装逆转录病毒的psi.2细胞或PA317细胞。用于基因治疗的病毒载体通常是通过生产将核酸载体包装成病毒颗粒的细胞系来生成的。所述载体通常含有包装和后续整合到宿主中所需的最小限度的病毒序列,其他的病毒序列被以供待表达的一个或多个多核苷酸的表达框所取代。缺失的病毒功能通常由所述包装细胞系反式地提供。例如,用于基因治疗的AAV载体通常仅持有来自AAV基因组的ITR序列,该序列是包装和整合到宿主基因组中所需的。病毒DNA可以在细胞系中被包装,所述细胞系包含辅助者质粒,其编码其他AAV基因,即rep和cap,但欠缺ITR序列。该细胞系也可以用腺病毒作为辅助者而被感染。该辅助者病毒可以促进AAV载体的复制和来自辅助者质粒中的AAV基因的表达。在一些情况下,由于欠缺ITR序列,所述辅助者质粒没有被大量包装。腺病毒的污染可以被减低,例如,通过热处理(腺病毒对其比AAV更敏感)。

[0814] 多-效应器核碱基编辑器的应用

[0815] 所述多-效应器核碱基编辑器可用于靶向目标多核苷酸以产生修饰蛋白质表达的改变。在一个实施方式中,多-效应器核碱基编辑器用于修饰非-编码或调节序列,包含但不限于剪接位点、增强子和转录调节元件。然后使用本领域已知的任何方法测定所述改变对受该调节元件控制的基因的表达的影响。在一个特定的实施方式中,多-效应器核碱基编辑器能够实质上地改变调节序列,从而废除其调节基因表达的能力。有利地,不同于其他RNA-可编程的核酸酶,这可以在基因组靶序列中不生成双-股断裂的情况下完成。

[0816] 所述多-效应器核碱基编辑器可用于靶向目标多核苷酸以产生可修饰蛋白质活性的改变。就诱变的情境而言,例如,与容易出错的PCR和其他基于聚合酶的方法相比,多-效应器核碱基编辑器具有一些优势。因为本发明的多-效应器核碱基编辑器在靶区域中的多个碱基处产生改变,所以相对于由容易出错的PCR引入的突变(不太可能在蛋白质水平表达),此类突变更有可能在蛋白质水平表达,考虑到密码子中的单个核苷酸变化可能仍编码相同的氨基酸(例如,密码子简并性)。不同于容易出错的PCR,其诱导整个多核苷酸的随机改变,本发明的多-效应器核碱基编辑器可用于靶向目标蛋白的小区域内或限定区域内的特定氨基酸。

[0817] 在其他实施方式中,本发明的多-效应器核碱基编辑器用于靶向生物体基因组内的目标多核苷酸。在一个实施方式中,该生物体是微生物组中的细菌(例如,拟杆菌门(Bacteriodetes),疣微菌门(Verrucomicrobia),厚壁菌门(Firmicutes);γ-变形菌纲(Gammaproteobacteria),α-变形菌纲(Alphaproteobacteria),拟杆菌纲(Bacteriodetes),梭菌纲(Clostridia),丹毒丝菌纲(Erysipelotrichia),芽孢杆菌纲(Bacilli);肠杆菌目(Enterobacteriales),拟杆菌目(Bacteriodales),疣微菌目(Verrucomicrobiales),梭菌目(Clostridiales),丹毒丝菌目(Erysiopelotrichales),乳杆菌目(Lactobacillales);肠杆菌科(Enterobacteriaceae),拟杆菌科(Bacteroidaceae),丹毒丝菌科(Erysiopelotrichaceae),普雷沃氏菌科(Prevotellaceae),红蝽菌科(Coriobacteriaceae),以及产碱菌科(Alcaligenaceae);埃希氏菌属(Escherichia),拟杆菌属(Bacteroides),理研菌科的(另枝菌)属(Alistipes),疣微菌科的(阿克曼)属(Akkermansia),梭菌属(Clostridium),乳杆菌属(Lactobacillus))。在另一个实施方式中,该生物体是农业上重要的动物(例如奶牛,绵羊,山羊,马,鸡,火鸡)或植物(例如大豆,小麦,玉米,水稻,烟草,苹果,葡萄,桃子,李子,樱桃)。在一个实施方式中,将本发明的多-效应器核碱基编辑器与向导RNA文库一起递送到细胞中,该向导RNA文库用于铺到细胞的基因组内的各种序列上,从而系统地改变整个基因组中的序列。

[0818] 可在各种蛋白质的任何一个中制造突变,以促进结构功能分析或改变该蛋白质的内源活性。可在,例如酶(例如,激酶,磷酸酶,羧化酶,磷酸二酯酶)中或在酶底物中,在受体或其配体中,以及在抗体及其抗原中制造突变。在一个实施方式中,多-效应器核碱基编辑器靶向:编码该酶的活性位点的核酸分子,编码该受体的配体结合位点的核酸分子,或编码抗体的互补决定区(CDR)的核酸分子。对于酶而言,在活性位点诱导突变可能会增加、降低或废除该酶的活性。突变对酶的影响在酶活性测定中被定性,该测定包含本领域已知和/或对熟练技术人员显而易见的一些测定中的任何测定。对于受体而言,在配体结合位点制造

的突变可能会增加、减少或废除受体对其配体的亲和力。此类突变的影响在受体/配体结合测定法中被测定,该测定法包含本领域已知和/或对熟练技术人员显而易见的一些测定法中的任何测定。对于CDR而言,在CDR内制造的突变可增加、减少或废除与抗原的结合。此外,在CDR内制造的突变可改变抗体对抗原的特异性。然后例如通过测量该CDR与其抗原的特异性结合或在任何其他类型的免疫测定中测定这些改变对CDR功能的影响。

[0819] 药物组合物

[0820] 本公开揭露的其他方面涉及药物组合物,其包括本文所述的任何碱基编辑器、融合蛋白或融合蛋白-向导多核苷酸复合体。在一些实施方式中,该药物组合物还包括药学上可接受的载体。在一些实施方式中,该药物组合物包括另外的剂(例如,用于特异性递送的、增加半衰期的或其他治疗性的化合物)。

[0821] 适合的药学上可接受的载体通常包括惰性物质,其有助于将药物组合物施用于受试者,有助于将药物组合物加工成可递送的制品,或有助于在给药前储存药物组合物。药学上可接受的载体可以包含能够稳定、优化或以其他方式改变制剂的形式、稠度、粘度、pH、药代动力学、溶解度的剂。

[0822] 可以充当药学上-可接受的载体的材料的一些非限制性例子包括:(1)糖,诸如乳糖,葡萄糖和蔗糖;(2)淀粉,诸如玉米淀粉和马铃薯淀粉;(3)纤维素,及其衍生物,诸如羧甲基纤维素钠,甲基纤维素,乙基纤维素,微晶纤维素和乙酸纤维素;(4)黄芪粉;(5)麦芽;(6)明胶;(7)润滑剂,诸如硬脂酸镁,月桂基硫酸钠和滑石;(8)赋形剂,诸如可可脂和栓剂蜡;(9)油,诸如花生油,棉籽油,红花油,芝麻油,橄榄油,玉米油,大豆油等;(10)二醇,诸如丙二醇;(11)多元醇,诸如甘油,山梨糖醇,甘露糖醇和聚乙二醇(PEG);(12)酯,诸如油酸乙酯和月桂酸乙酯;(13)琼脂;(14)缓冲剂,诸如氢氧化镁和氢氧化铝;(15)海藻酸;(16)无热原水;(17)等渗生理盐水;(18)林格式溶液;(19)乙醇;(20)pH缓冲溶液;(21)聚酯,聚碳酸酯和/或聚酐;(22)填充剂,诸如多肽和氨基酸;(23)血清醇,诸如乙醇;以及(23)药物制剂中使用的其他无毒相容性物质。制剂中也可以存在:缓冲剂、润湿剂、乳化剂、稀释剂、包封剂、皮肤渗透促进剂、着色剂、脱模剂、包衣剂、甜味剂、调味剂、加香剂、防腐剂和抗氧化剂。例如,载体可包含但不限于:盐水、缓冲盐水、右旋葡萄糖、精氨酸、蔗糖、水、甘油、乙醇、山梨糖醇、葡聚糖、羧甲基纤维素钠及其组合。

[0823] 药物组合物可以包括一种或多种pH缓冲化合物,以将制剂的pH维持在反映生理pH的预定水平,诸如在约5.0至约8.0的范围内。用于水性液体制剂的pH缓冲化合物可以是氨基酸或氨基酸混合物,诸如组氨酸,或诸如组氨酸和甘氨酸的氨基酸混合物。或者,该pH缓冲化合物优选地是将制剂的pH维持在预定水平,诸如在约5.0至约8.0的范围内,并且不螯合钙离子的剂。此种pH缓冲化合物的说明性实例包含,但不限于,咪唑和乙酸根离子。该pH缓冲化合物能以适合将制剂的pH维持在预定水平的任何数量存在。

[0824] 药物组合物亦可含有一种或多种渗透调节剂,即,将该制剂的渗透特性(例如,张性,渗透度和/或渗透压)调节至接收者个体的血流和血细胞可接受的水平的化合物。该渗透调节剂可以是不螯合钙离子的剂。该渗透调节剂可以是本领域技术人员已知或可获得的可调节制剂的渗透特性的任何化合物。本领域技术人员可凭经验确定给定的渗透调节剂用于本发明制剂的适用性。适合类型的渗透调节剂的说明性实例包含但不限于:盐,诸如氯化钠和乙酸钠;糖,诸如蔗糖,右旋葡萄糖和甘露醇;氨基酸,诸如甘氨酸;以及这些剂和/或剂

类型中的一种或多种的混合物。该一种或多种渗透调节剂能以足以调节制剂的渗透特性的任何浓度存在。

[0825] 在一些实施方式中,将药物组合物配制成用于递送至受试者,例如用于基因编辑。在一些实施方式中,本文所预期的药物组合物的施用可使用传统技法进行,包含但不限于输液、移注或非消化道给药。在一些实施方式中,非消化道给药包含:血管内、静脉内、肌肉内、动脉内、鞘内、肿瘤内、皮内、腹膜内、经气管、皮下、透皮、关节内、囊下、蛛网膜下和胸骨内的输液或注射。在一些实施方式中,施用本文所述药物组合物的适合途径包含但不限于:局部,皮下,穿皮,皮内,病灶内,关节内,腹膜内,膀胱内,经粘膜,牙龈,齿内,耳蜗内,鼓室内,器官内,硬脊膜外,鞘内,肌肉内,静脉内,血管内,骨内,眼周,肿瘤内,脑内和脑室内给药。

[0826] 在一些实施方式中,本文所述的药物组合物是局部施用于疾病位置(例如,肿瘤位置)。在一些实施方式中,本文所述的药物组合物是通过注射,通过导管的方法,通过栓剂的方法,或通过植入物的方法而施用于受试者,该植入物为多孔,无孔,或凝胶状的材料,其包含膜,诸如硅橡胶膜,或纤维。

[0827] 在其他实施方式中,本文所述的药物组合物以控释系统递送。在一个实施方式中,可以使用泵(参见例如,Langer、1990、Science 249:1527-1533;Sefton、1989、CRC Crit.Ref.Biomed.Eng.14:201;Buchwald等、1980、Surgery 88:507;Saudek等、1989、N.Engl.J.Med.321:574)。在另一个实施方式中,可以使用聚合物材料(参见例如《控释的医学应用》(Langer和Wise编、CRC出版社、Boca Raton、佛罗里达州、1974);《可控药物的生物可利用度,药物产品设计和性能》(Smolen和Ball编、Wiley、纽约、1984);Ranger和Peppas、1983、Macromol.Sci.Rev.Macromol.Chem.23:61。亦参见Levy等、1985、Science 228:190;During等、1989、Ann.Neurol.25:351;Howard等、1989、J.Neurosurg.71:105)。讨论了其他的控释系统,例如,在前述Langer的论文中。

[0828] 在一些实施方式中,根据常规程序将该药物组合物配制为适于静脉内或皮下给药给受试者(例如人)的组合物。在一些实施方式中,用于注射给药的药物组合物是在溶液中,该溶液是无菌等渗的并用作增溶剂和局部麻醉剂(诸如利多卡因以减轻注射位置的疼痛)。通常,将各成分要么分开提供,要么以单位剂型混合在一起提供,例如,作为干燥的冻干粉或无水浓缩物,其在标示有活性剂数量的气密密封容器诸如安瓿瓶或小药囊中。如果通过输液方式给药,可以将其分装到含有无菌药物级水或生理盐水的输液瓶中。如果通过注射施用该药物组合物,可以提供一安瓿瓶的无菌注射用水或生理盐水,以便可以在施用之前将各成分混合。

[0829] 用于全身给药的药物组合物可以是液体,例如无菌生理盐水、乳酸林格氏液或Hank's溶液。另外,该药物组合物可以是固体形式,并且在使用前立即重新溶解或重新悬浮。亦检视了冻干形式。该药物组合物可以被容纳在脂质颗粒或囊泡内,诸如脂质体或微晶,其也适合于非消化道给药。该微粒可以为任何适合的结构,诸如单层或多层,只要其中容纳组合物即可。可以将化合物截留到"稳定化的质粒-脂质微粒"(SPLP)中,该微粒含有致融类脂质二油酰基磷脂酰乙醇胺(DOPE),低水平(5-10摩尔百分比)的阳离子脂质,并由聚乙二醇(PEG)涂层使其稳定化(Zhang Y.P.等、Gene Ther.1999、6:1438-47)。带正电荷的脂质,诸如N-[1-(2,3-二油酰基氧基)丙基]-N,N,N-三甲基-甲基硫酸铵,或"DOTAP",对于此

类微粒和囊泡是特别优选的。此种脂质微粒的制备是众所周知的。参见例如,美国专利号4,880,635;4,906,477;4,911,928;4,917,951;4,920,016;和4,921,757;其每个都通过引用并入本文。

[0830] 例如,本文所述的药物组合物可以作为单位剂量予以施用或包装。当用于指代本公开揭露的药物组合物时,术语"单位剂量"是指适合作为受试者的统一单元剂量的物理上离散的单位,而每个单位含有经计算可产生期望的治疗效果的预定数量的活性物质与相关联的所需的稀释剂,即载体或媒介物。

[0831] 此外,可以将该药物组合物作为药物试剂盒提供,其包括(a)含有冻干形式的本发明的化合物的容器,和(b)含有药学上可接受的稀释剂(例如无菌的,用于重构或稀释冻干的本发明的化合物的)的第二容器。可选的伴随此种一个或多个容器的可以是由规范药品或生物产品的制造、使用或销售的政府机构所规定的形式的行政命令,该行政命令反映了该机构对此用于人体给药的试剂盒制造、使用或出售的批准。

[0832] 在另一方面,包含了一种制品,其含有可用于治疗上述疾病的材料。在一些实施方式中,该制品包括容器和标签。适合的容器包含例如瓶子,小药瓶,注射器和试管。该容器可以由各种材料形成,诸如玻璃或塑料。在一些实施方式中,该容器容纳有效治疗本文所述疾病的组合物,并且可以具有无菌的存取口。例如,该容器可以是静脉注射溶液袋或具有可被皮下注射针刺穿的塞子的小药瓶。该组合物中的活性剂是本发明的化合物。在一些实施方式中,容器上的或与容器相关联的标签表明了该组合物用于治疗所选的疾病。该制品可以进一步包括第二容器,其包括药学上-可接受的缓冲液,诸如磷酸盐-缓冲生理盐水,林格氏溶液,或右旋葡萄糖溶液。从商业和使用者的角度来看,它还可以进一步包含其他材料,包含其他缓冲液,稀释剂,过滤器,针头,注射器和带有使用说明的包装插页。

[0833] 在一些实施方式中,本文所述的任何融合蛋白、gRNA和/或复合体是作为药物组合物的一部分被提供。在一些实施方式中,该药物组合物包括本文所提供的任何融合蛋白。在一些实施方式中,该药物组合物包括本文提供的任何复合体。在一些实施方式中,该药物组合物包括核糖核蛋白复合体,其包括与gRNA和阳离子脂质形成复合体的RNA-导向的核酸酶(例如,Cas9)。在一些实施方式中,药物组合物包括gRNA,核酸可编程的DNA结合蛋白,阳离子脂质,和药学上可接受的赋形剂。药物组合物可以任选地包括一种或多种另外的治疗活性物质。

[0834] 在一些实施方式中,将本文所提供的组合物施用给受试者,例如人类受试者,以在受试者体内实现靶向的基因组修饰。在一些实施方式中,细胞是获自于受试者,并与本文所提供的任何药物组合物接触。在一些实施方式中,将从受试者体内取出并与药物组合物离体接触过的细胞重新-引入到受试者体内,其可选地是在细胞中已经实现或检测到所期望的基因组修饰之后。递送包括核酸酶的药物组合物的方法是已知的,并且被描述,例如在美国专利号6,453,242;6,503,717;6,534,261;6,599,692;6,607,882;6,689,558;6,824,978;6,933,113;6,979,539;7,013,219;和7,163,824中,其公开揭露的全部内容通过引用并入本文。尽管本文所提供的药物组合物的描述主要针对适合于施用给人的药物组合物,但是熟练技术人员应理解,此类组合物通常适合于给各种动物或生物体施用,例如,供兽医使用。

[0835] 为了使该组合物适合于向各种动物给药,对适合于向人给药的药物组合物的修饰

是众所周知的,并且一般技术的兽医药理学家可以仅通过普通的实验(如果有的话)来设计和/或执行此种修饰。预期被施用该药物组合物的受试者包含但不限于,人类和/或其他灵长类;哺乳动物,家养动物,宠物,和与商业有关的哺乳动物,诸如牛,猪,马,绵羊,猫,狗,小鼠和/或大鼠;和/或鸟类,包含商业相关的鸟类,诸如鸡,鸭,鹅和/或火鸡。

[0836] 本文所述的药物组合物的制剂可以通过药理学领域中已知的或以后开发的任何方法来制备。通常,此种制备方法包含以下步骤:使(一种或多种)活性成分与赋形剂和/或一种或多种其他辅助成分相结合,然后,如果必要和/或期望,将产品成型和/或包装成为所需的单-剂量或多-剂量单位。药物制剂可另外包括药学上可接受的赋形剂,其如本文所用包含适用于所需特定剂型的任何的和所有的:溶剂,分散介质,稀释剂,或其他液体媒介物,分散或悬浮助剂,表面活性剂,等渗剂,增稠剂或乳化剂,防腐剂,固体粘合剂,润滑剂等。《Remington:药学的科学与实践》、第21版、A.R.Gennaro(Lippincott、Williams&Wilkins出版社、巴尔的摩、马里兰州、2006;其全部内容通过引用并入本文)揭露了用于配制药物组合物的各种赋形剂及其制备的已知技法。关于用于生产包括核酸酶的药物组合物的其他适合的方法、试剂、赋形剂和溶剂,亦请参见PCT申请PCT/US2010/055131(公开号WO2011/053982A8、2010年11月2日提交),其全部内容通过引用并入本文。

[0837] 除非任何传统的赋形剂介质与某种物质或其衍生物不相容,诸如通过产生任何不良的生物学影响或以有害的方式与该药物组合物的任何其他组分相互作用,否则预期其使用范围是在本公开揭露的范围之内。

[0838] 如上所述的组合物可以用有效量给药。该有效量将取决于给药方式、所治疗的特定病况,和所期望的结果。它还可能取决于所述病况的阶段、受试者的年龄和身体状况、并行治疗的性质(如果有的话),以及医师所熟知的类似因素。对于治疗性的应用,它是足以达到医学上所期望结果的那个数量。

[0839] 在一些实施方式中,依据本公开揭露的组合物可用于治疗各种疾病,病症和/或病况中的任何一种。

[0840] 试剂盒,载体,细胞

[0841] 本公开揭露的各个方面提供了包括碱基编辑器系统的试剂盒。在一个实施方式中,该试剂盒包括核酸构建体,该核酸构建体包括编码核碱基编辑器融合蛋白的核苷酸序列。该融合蛋白包括一个或多个脱氨酶结构域(例如胞苷脱氨酶和/或腺嘌呤脱氨酶)和核酸可编程的DNA结合蛋白(napDNAbp)。在一些实施方式中,该试剂盒包括至少一种能够靶向目标核酸分子的向导RNA。在一些实施方式中,该试剂盒包括核酸构建体,该核酸构建体包括编码至少一种向导RNA的核苷酸序列。在一些实施方式中,该试剂盒包括核酸构建体,其包括编码以下的核苷酸序列:(a)如本文所提供的与腺苷脱氨酶和/或胞苷脱氨酶融合的Cas9结构域;以及(b)驱动(a)的序列表达的异源启动子。

[0842] 在一些实施方式中,所述试剂盒提供了使用该试剂盒编辑一个或多个突变的说明书。该说明书通常包含关于使用该试剂盒编辑核酸分子的信息。在其他实施方式中,该说明书包含以下至少一项:注意事项;警告;临床研究;和/或参考文献。该说明书可直接印在容器上(如果有),或者作为贴在容器上的标签,或者作为单独的纸张、小册子、卡片或文件夹提供在容器中或随容器提供。在进一步的实施方式中,试剂盒可以包括标签或单独插页(包装插页)形式的以供适合的操作参数的说明书。在又一个实施方式中,该试剂盒可以包括一

个或多个容器,其具有合适的阳性和阴性对照或对照样品,以用作检测、校准或均一化的标准。该试剂盒可以进一步包括包括药学上-可接受的缓冲液的第二容器,诸如(无菌的)磷酸盐-缓冲盐水、林格氏溶液或右旋葡萄糖溶液。它可以进一步包含从商业和用户的角度来看所期待的其他材料,包含其他缓冲液、稀释剂、过滤器、针头、注射器和带有使用说明书的包装插页。

[0843] 本公开揭露的一些方面提供了包括本文所提供的任何核碱基编辑器或多-效应器核碱基编辑器或融合蛋白的细胞。在一些实施方式中,所述细胞包括本文所提供的任何核苷酸或载体。

[0844] 除非另有说明,本发明的实践使用分子生物学(包含重组技法)、微生物学、细胞生物学、生物化学和免疫学的常规技法,其完全在本领域熟练人员的知识范围内。这些技法在文献中有充分的解释,诸如《分子克隆:实验室手册》第二版(Sambrook,1989);《寡核苷酸合成》(Gait,1984);《动物细胞培养》(Freshney,1987);《酶学方法》《实验免疫学手册》(Weir,1996);《用于哺乳动物细胞的基因移转载体》(Miller和Calos,1987);《当代实验方案-分子生物学》(Ausubel,1987);《PCR:聚合酶链式反应》(Mullis,1994);《当代实验方案-免疫学》(Coligan,1991)。这些技法适用于本发明的多核苷酸和多肽的生产,因此,可以在制备和实践本发明时考虑这些技法。用于特定实施方式的特别有用的技法将在后面部分中讨论。

[0845] 提出以下实施例以向本领域普通技术人员提供如何制备和使用本发明的测定、筛选和治疗方法的完整公开揭露和描述,并非意图限制发明人所视其发明的范围。

[0846] 实施例

[0847] 实施例1:非靶脱氨反应减低的核碱基编辑器的构建

[0848] 核碱基编辑器(例如,通过连接子连接的CRISPR-Cas蛋白和脱氨酶的融合蛋白)可用于将特定的点突变引入到靶多核苷酸中。然而,核碱基编辑器本身具有不需要的全基因组假性脱氨反应、旁观者突变和靶近端编辑的潜力。不受理论的束缚,缩短或去除来自碱基编辑器的连接子将减低潜在的不需要的脱氨反应事件和/或促进所期望的靶脱氨反应(图1)。这可能部分是由于减低了核碱基编辑器的脱氨酶结构域的有效活性半径。尽管通过X-射线晶体学已经确定了与DNA结合的Cas9的结构,但却没有碱基编辑发生之处的DNA部分的结构信息。Cas9的建模预测了碱基编辑发生之处的DNA可能是在靠近Cas9的2个位置(图2)。基于这些预测,将脱氨酶或其片段就定位在这些位置中的一个或多个具有促进在靶碱基编辑的同时也减低不期望的脱氨反应事件的潜力(图3)。在所述腺苷碱基编辑器(例如与TadA融合的Cas9)中鉴定出的几个区域适合TadA脱氨酶或其片段的插入(图4至7)。因此,生成了腺苷脱氨酶碱基编辑器以将TadA或其变体插入到Cas9多肽中所述鉴别出的位置处。

[0849] 实施例2:用于测量在靶和脱靶脱氨反应的高通量体外测定

[0850] 开发了一种体外测定来评估核碱基编辑器和用于定性候选构建体,该测定测量在靶脱氨反应与非靶脱氨反应,包含假性脱氨反应。该测定的基于FRET的版本使用荧光报告基因用于检测,但该测定可改造以用于基于凝胶的结果输出(图8)。用于体外脱氨反应测定的探针包含用于脱氨反应的底物,以及特别以供核碱基编辑器的底物(图8)。除了含有可以被脱氨基的核苷酸外,探针还可能包含PAM序列、靶特异性序列等,甚至随机序列。使用探针组的脱氨反应可以并行执行(例如,高通量的形式)。底物的脱氨反应(C→U或A→I)使得底物可被脱氨反应特异性核酸内切酶(分别为USER/核酸内切酶V)切开(图8)。底物的切开使

得荧光报告蛋白与淬灭剂分子解偶联,从而生成荧光信号(图8)。高的在靶的与脱靶的荧光比率表明碱基编辑器是有效的。可以使用本领域已知的任何相互作用的荧光团和淬灭剂对或FRET供体-受体对。在某些实施方式中,荧光团是FAM、TET、HEX、TAMRA、JOE或ROX中的一种或多种。在各种实施方式中,该淬灭剂是dabcyl、dabsyl、黑洞淬灭剂染料(包含5'Iowa Black® RQ(5IabRQ))中的一种或多种。通常,该淬灭染料是激发匹配的淬灭染料。荧光团-猝灭剂对及其选择在,例如Marras的"分子生物学方法:荧光能量移转核酸探针:设计和实验方案中的《用于荧光核酸杂交探针的荧光团和猝灭剂对的选择》"(编辑:V.V.Didenko © Humana出版公司、托托瓦、新泽西州)中进行了描述。

[0851] 作为该测定的展示,通过将腺苷碱基编辑器的在靶脱氨反应与存在SpCas9(无脱氨反应结构域)或无蛋白质时发生的脱氨反应进行比较,而测定了该腺苷碱基编辑器生成脱靶脱氨反应的潜力(图9)。腺苷碱基编辑器反应所生成的荧光信号高于SpCas9和无蛋白质的反应所生成的,表明ABE的在靶碱基编辑是有效的。在另一个例子中,将腺苷碱基编辑器与反式腺苷碱基编辑(ABE-TadA)(其中SpCas9与TadA反式存在)进行了比较(图10)。与ABE-TadA、SpCas9和无蛋白质的反应相比,ABE生成了更高的荧光,并且其在靶碱基编辑是有效的。可以在这个测定中测试用于假性脱靶碱基编辑的潜在底物,包含单-股结构和分枝的结构,其可能反映基因组中的其他结构(例如,DNA"呼吸",复制叉,转录活性DNA等)(图11)。

[0852] 实施例3:评估顺式和反式的脱氨酶活性的测定

[0853] 开发了一种测定以区分顺式的(脱氨反应结构域与CRISPR-Cas共价结合)和反式的(具有反式提供的脱氨反应结构域的CRISPR-Cas蛋白)脱氨酶活性(图12)。顺式发生的脱氨反应表示通过靶向的碱基编辑的脱氨反应,而反式的脱氨反应表示假性脱氨反应。顺式活性与反式活性的高比率表示具有减低的假性脱氨反应的脱氨酶是是有效的碱基编辑器。

[0854] 在所述顺式-反式测定中测试了大鼠APOBEC1。简言之,用表达所述碱基编辑器BE4(rAPOBEC1-nCas9-UGI-UGI),rAPOBEC1和nCas9,nCas9和向导RNA,或rAPOBEC1和向导RNA的构建体转染了HEK293T细胞。从该细胞中分离了基因组DNA并获得了4个基因组靶位点处的测序。在所有位点处,与反式脱氨酶活性以及缺少至少一个靶向碱基编辑组分的其他对照反应相比,rAPOBEC1显示出更高的顺式脱氨酶活性(图13)。同样地,与反式脱氨酶活性和其他脱氨反应事件相比,TadA7.10也显示出更高的顺式脱氨酶活性(图14)。为了知晓独立于向导的腺苷碱基编辑器的反式影响,将SaCas9-ABE和SaCas9向导,与SpCas9-ABE和SaCas9向导,以及空间位阻的ABE变体和SaCas9向导结合进行了测试(图15)。在这种情境下,SpCas9-ABE对碱基编辑器情境中的TadA-TadA7.10显示出较低的反式活性。使用来自所述SaCas9向导测定的反式测量以及ABE及空间位阻ABE变体的活性而估算了ABE和空间位阻ABE变体的顺式/反式活性比。ABE和空间位阻ABE变体的经估算的比率是相对高的。也进行了顺式和反式的剂量反应研究以确定是否可以通过剂量而调节高的顺式相对于反式的活性(例如,随着剂量增加,顺式活性比反式活性增加得更快)。在所测试的条件下,未观察到顺式相对于反式的活性的剂量反应(图16至18)。

[0855] 所述顺式-反式测定用于评估各种脱氨酶(于下表9中列出)以供减低的假性脱氨反应:

[0856] 表9.使用顺式-反式测定所筛选的脱氨酶

[0857]

1	rAPOBEC-1	9	hAPOBEC-2	17	hAPOBEC-3F	25	btAID
2	mAPOBEC-1	10	ppAPOBEC-2	18	hAPOBEC-3G	26	mAID
3	maAPOBEC-1	11	btAPOBEC-2	19	hAPOBEC-4	27	pmCDA-1
4	hAPOBEC-1	12	mAPOBEC-3	20	mAPOBEC-4	28	pmCDA-2
5	ppAPOBEC-1	13	hAPOBEC-3A	21	rAPOBEC-4	29	pmCDA-5
6	ocAPOBEC1	14	hAPOBEC-3B	22	mfAPOBEC-4	30	yCD
7	mdAPOBEC-1	15	hAPOBEC-3C	23	hAID	31	rAPOBEC-1-Δ177-186
8	mAPOBEC-2	16	hAPOBEC-3D	24	clAID	32	rAPOBEC-1-Δ202-213

[0858] 有趣的是,几种脱氨酶显示出高的顺式/反式活性,包含ppAPOBEC-2、mAPOBEC-2、mAPOBEC-3和mfAPOBEC-4。

[0859] rAPOBEC-1大鼠(Rattus norvegicus)

[0860] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEK FTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTE QESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHIL WATGLK

[0861] mAPOBEC-1小鼠(Mus musculus)

[0862] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSVWRHTSQNTSNHVEVNFLEK FTTERYFRPNTRCSITWFLSWSPCGECSRAITEFLSRHPYVTLFIYIARLYHHTDQRNRQGLRDLISSGVTIQIMTE QEYCYCWRNFVNYPPSNEAYWPRYPHLWVKLYVLELYCIILGLPPCLKILRRKQPQLTFFTITLQTCHYQRIPPHLL WATGLK

[0863] maAPOBEC-1黄金仓鼠(Mesocricetus auratus)

[0864] MSSETGPVVVDPTLRRRIEPHEFDAFFDQGELRKETCLLYEIRWGGRHNIWRHTGQNTSRHVEINFIEK FTSERYFYPSTRCSIVWFLSWSPCGECSKAITEFLSGHPNVTLFIYAARLYHHTDQRNRQGLRDLISRGVTIRIMTE QEYCYCWRNFVNYPPSNEVYWPRYPNLWMRLYALELYCIHLGLPPCLKIKRRHQYPLTFFRLNLQSCHYQRIPPHIL WATGFI

[0865] hAPOBEC-1智人(Homo sapiens)

[0866] MTSEKGPSTGDPTLRRRIEPWEFDVFYDPRELRKEACLLYEIKWGMSRKIWRSSGKNTTNHVEVNFIKK FTSERDFHPSMSCSITWFLSWSPCWECSQAIREFLSRHPGVTLVIYVARLFWHMDQQNRQGLRDLVNSGVTIQIMRA SEYYHCWRNFVNYPPGDEAHWPQYPPLWMMLYALELHCIILSLPPCLKISRRWQNHLTFFRLHLQNCHYQTIPPHIL LATGLIHPSVAWR

[0867] ppAPOBEC-1婆罗洲猩猩(Pongo pygmaeus)

[0868] MTSEKGPSTGDPTLRRRIESWEFDVFYDPRELRKETCLLYEIKWGMSRKIWRSSGKNTTNHVEVNFIKK FTSERRFHSSISCSITWFLSWSPCWECSQAIREFLSQHPGVTLVIYVARLFWHMDQRNRQGLRDLVNSGVTIQIMRA SEYYHCWRNFVNYPPGDEAHWPQYPPLWMMLYALELHCIILSLPPCLKISRRWQNHLAFFRLHLQNCHYQTIPPHIL LATGLIHPSVTWR

[0869] ocAPOBEC1穴兔(Oryctolagus cuniculus)

[0870] MASEKGPSNKDYTLRRRIEPWEFEVFFDPQELRKEACLLYEIKWGASSKTWRSSGKNTTNHVEVNFLEK LTSEGRLGPSTCCSITWFLSWSPCWECSMAIREFLSQHPGVTLIIFVARLFQHMDRRNRQGLKDLVTSGVTVRVMSV SEYCYCWENFVNYPPGKAAQWPRYPPRWMLMYALELYCIILGLPPCLKISRRHQKQLTFFSLTPQYCHYKMIPPYIL LATGLLQPSVPWR

[0871] mdAPOBEC-1灰短尾负鼠(Monodelphis domestica)

[0872] MNSKTGPSVGDATLRRRIKPWEFVAFFNPQELRKETCLLYEIKWGNQNIWRHSNQNTSQHAEINFMEKF TAERHFNSSVRCSITWFLSWSPCWECSKAIRKFLDHYPNVTLAIFISRLYWHMDQQHRQGLKELVHSGVTIQIMSYS EYHYCWRNFVDYPQGEEDYWPKYPYLWIMLYVLELHCIILGLPPCLKISGSHSNQLALFSLDLQDCHYQKIPYNVLV ATGLVQPFVTWR

[0873] mAPOBEC-2小鼠(Mus musculus)

[0874] MAQKEEAAEAAAPASQNGDDLENLEDPEKLKELIDLPPFEIVTGVRLPVNFFKFQFRNVEYSSGRNKTF LCYVVEVQSKGGQAQATQGYLEDEHAGAHAEEAFFNTILPAFDPALKYNVTWYVSSSPCAACADRILKTLSKTKNLR LLILVSRLFMWEEPEVQAALKKLKEAGCKLRIMKPQDFEYIWQNFVEQEEGESKAFEPWEDIQENFLYYEEKLADIL K

[0875] hAPOBEC-2智人(Homo sapiens)

[0876] MAQKEEAAVATEAASQNGEDLENLDDPEKLKELIELPPFEIVTGERLPANFFKFQFRNVEYSSGRNKTF LCYVVEAQGKGGQVQASRGYLEDEHAAAHAEEAFFNTILPAFDPALRYNVTWYVSSSPCAACADRIIKTLSKTKNLR LLILVGRLFMWEEPEIQAALKKLKEAGCKLRIMKPQDFEYVWQNFVEQEEGESKAFQPWEDIQENFLYYEEKLADIL K

[0877] ppAPOBEC-2婆罗洲猩猩(Pongopygmaeus)

[0878] MAQKEEAAAATEAASQNGEDLENLDDPEKLKELIELPPFEIVTGERLPANFFKFQFRNVEYSSGRNKTF LCYVVEAQGKGGQVQASRGYLEDEHAAAHAEEAFFNTILPAFDPALRYNVTWYVSSSPCAACADRIIKTLSKTKNLR LLILVGRLFMWEELEIQDALKKLKEAGCKLRIMKPQDFEYVWQNFVEQEEGESKAFQPWEDIQENFLYYEEKLADIL K

[0879] btAPOBEC-2牛(Bos Taurus)

[0880] MAQKEEAAAAAEPASQNGEEVENLEDPEKLKELIELPPFEIVTGERLPAHYFKFQFRNVEYSSGRNKTF LCYVVEAQSKGGQVQASRGYLEDEHATNHAEEAFFNSIMPTFDPALRYMVTWYVSSSPCAACADRIVKTLNKTKNLR LLILVGRLFMWEEPEIQAALRKLKEAGCRLRIMKPQDFEYIWQNFVEQEEGESKAFEPWEDIQENFLYYEEKLADIL K

[0881] mAPOBEC-3小鼠(Mus musculus)

[0882] MQPQRLGPRAGMGPFCLGCSHRKCYSPIRNLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVS LHHGVFKNKDNIHAEICFLYWFHDKVLKVLSPREEFKITWYMSWSPCFECAEQIVRFLATHHNLSLDIFSSRLYNVQ DPETQQNLCRLVQEGAQVAAMDLYEFKKCWKKFVDNGGRRFRPWKRLLTNFRYQDSKLQEILRPCYISVPSSSSSTL SNICLTKGLPETRFWVEGRRMDPLSEEEFYSQFYNQRVKHLCYYHRMKPYLCYQLEQFNGQAPLKGCLLSEKGKQHA EILFLDKIRSMELSQVTITCYLTWSPCPNCAWQLAAFKRDRPDLILHIYTSRLYFHWKRPFQKGLCSLWQSGILVDV MDLPQFTDCWTNFVNPKRPFWPWKGLEIISRRTQRRLRRIKESWGLQDLVNDFGNLQLGPPMS

[0883] hAPOBEC-3A智人(Homosapiens)

[0884] MEASPASGPRHLMDPHIFTSNFNNGIGRHKTYLCYEVERLDNGTSVKMDQHRGFLHNQAKNLLCGFYGR

HAELRFLDLVPSLQLDPAQIYRVTWFISWSPCFSWGCAGEVRAFLQENTHVRLRIFAARIYDYDPLYKEALQMLRDA
GAQVSIMTYDEFKHCWDTFVDHQGCPFQPWDGLDEHSQALSGRLRAILQNQGN

[0885] hAPOBEC-3B智人(Homosapiens)

[0886] MNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEVKIKRGRSNLLWDTGVFRGQVYFKPQYHAEM
CFLSWFCGNQLPAYKCFQITWFVSWTPCPDCVAKLAEFLSEHPNVTLTISAARLYYYWERDYRRALCRLSQAGARVT
IMDYEEFAYCWENFVYNEGQQFMPWYKFDENYAFLHRTLKEILRYLMDPDTFTFNFNNDPLVLRRRQTYLCYEVERL
DNGTWVLMDQHMGFLCNEAKNLLCGFYGRHAELRFLDLVPSLQLDPAQIYRVTWFISWSPCFSWGCAGEVRAFLQEN
THVRLRIFAARIYDYDPLYKEALQMLRDAGAQVSIMTYDEFEYCWDTFVYRQGCPFQPWDGLEEHSQALSGRLRAIL
QNQGN

[0887] hAPOBEC-3C智人(Homosapiens)

[0888] MNPQIRNPMKAMYPGTFYFQFKNLWEANDRNETWLCFTVEGIKRRSVVSWKTGVFRNQVDSETHCHAER
CFLSWFCDDILSPNTKYQVTWYTSWSPCPDCAGEVAEFLARHSNVNLTIFTARLYYFQYPCYQEGLRSLSQEGVAVE
IMDYEDFKYCWENFVYNDNEPFKPWKGLKTNFRLLKRRLRESLQ

[0889] hAPOBEC-3D智人(Homosapiens)

[0890] MNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEVKIKRGRSNLLWDTGVFRGPVLPKRQSNHRQ
EVYFRFENHAEMCFLSWFCGNRLPANRRFQITWFVSWNPCLPCVVKVTKFLAEHPNVTLTISAARLYYYRDRDWRWV
LLRLHKAGARVKIMDYEDFAYCWENFVCNEGQPFMPWYKFDDNYASLHRTLKEILRNPMEAMYPHIFYFHFKNLLKA
CGRNESWLCFTMEVTKHHSAVFRKRGVFRNQVDPETHCHAERCFLSWFCDDILSPNTNYEVTWYTSWSPCPECAGEV
AEFLARHSNVNLTIFTARLCYFWDTDYQEGLCSLSQEGASVKIMGYKDFVSCWKNFVYSDDEPFKPWKGLQTNFRLL
KRRLREILQ

[0891] hAPOBEC-3F智人(Homosapiens)

[0892] MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPRLDAKIFRGQVYSQPEHHAEMC
FLSWFCGNQLPAYKCFQITWFVSWTPCPDCVAKLAEFLAEHPNVTLTISAARLYYYWERDYRRALCRLSQAGARVKI
MDDEEFAYCWENFVYSEGQPFMPWYKFDDNYAFLHRTLKEILRNPMEAMYPHIFYFHFKNLRKAYGRNESWLCFTME
VVKHHSPVSWKRGVFRNQVDPETHCHAERCFLSWFCDDILSPNTNYEVTWYTSWSPCPECAGEVAEFLARHSNVNLT
IFTARLYYFWDTDYQEGLRSLSQEGASVEIMGYKDFKYCWENFVYNDDEPFKPWKGLKYNFLFLDSKLQEILE

[0893] hAPOBEC-3G智人(Homosapiens)

[0894] MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPPLDAKIFRGQVYSELKYHPEMR
FFHWFSKWRKLHRDQEYEVTWYISWSPCTKCTRDMATFLAEDPKVTLTIFVARLYYFWDPDYQEALRSLCQKRDGPR
ATMKIMNYDEFQHCWSKFVYSQRELFEPWNNLPKYYILLHIMLGEILRHSMDPPTFTFNFNNEPWVRGRHETYLCYE
VERMHNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDVIPFWKLDLDQDYRVTCFTSWSPCFSCAQEMAKFIS
KNKHVSLCIFTARIYDDQGRCQEGLRTLAEAGAKISIMTYSEFKHCWDTFVDHQGCPFQPWDGLDEHSQDLSGRLRA
ILQNQEN

[0895] hAPOBEC-4智人(Homosapiens)

[0896] MEPIYEEYLANHGTIVKPYYWLSFSLDCSNCPYHIRTGEEARVSLTEFCQIFGFPYGTTFPQTKHLTFY
ELKTSSGSLVQKGHASSCTGNYIHPESMLFEMNGYLDSAIYNNDSIRHIILYSNNPCNEANHCCISKMYNFLITYPG
ITLSIYFSQLYHTEMDFPASAWNREALRSLASLWPRVVLSPISGGIWHSVLHSFISGVSGSHVFQPILTGRALADRH
NAYEINAITGVKPYFTDVLLQTKRNPNTKAQEALESYPLNNAFPGQFFQMPSGQLQPNLPPDLRAPVVFVLVPLRDL
PPMHMGQNPNKPRNIVRHLNMPQMSFQETKDLGRLPTGRSVEIVEITEQFASSKEADEKKKKKGKK

[0897] mAPOBEC-4小鼠(Mus musculus)

[0898] MDSLLMKQKKFLYHFKNVRWAKGRHETYLCYVVKRRDSATSCSLDFGHLRNKSGCHVELLFLRYISDWD LDPGRCYRVTWFTSWSPCYDCARHVAEFLRWNPNLSLRIFTARLYFCEDRKAEPEGLRRLHRAGVQIGIMTFKDYFY CWNTFVENRERTFKAWEGLHENSVRLTRQLRRILLPLYEVDDLRDAFRMLGF

[0899] rAPOBEC-4大鼠(Rattus norvegicus)

[0900] MEPLYEEYLTHSGTIVKPYYWLSVSLNCTNCPYHIRTGEEARVPYTEFHQTFGFPWSTYPQTKHLTFYE LRSSSGNLIQKGLASNCTGSHTHPESMLFERDGYLDSLIFHDSNIRHIILYSNNSPCDEANHCCISKMYNFLMNYPE VTLSVFFSQLYHTENQFPTSAWNREALRGLASLWPQVTLSAISGGIWQSILETFVSGISEGLTAVRPFTAGRTLTDR YNAYEINCITEVKPYFTDALHSWQKENQDQKVWAASENQPLHNTTPAQWQPDMSQDCRTPAVFMLVPYRDLPPIHVN PSPQKPRTVVRHLNTLQLSASKVKALRKSPSGRPVKKEEARKGSTRSQEANETNKSKWKKQTLFIKSNICHLLEREQ KKIGILSSWSV

[0901] mfAPOBEC-4食蟹猕猴(Macaca fascicularis)

[0902] MEPTYEEYLANHGTIVKPYYWLSFSLDCSNCPYHIRTGEEARVSLTEFCQIFGFPYGTTYPQTKHLTFY ELKTSSGSLVQKGHASSCTGNYIHPESMLFEMNGYLDSAIYNNDSIRHIILYCNNSPCNEANHCCISKVYNFLITYP GITLSIYFSQLYHTEMDFPASAWNREALRSLASLWPRVVLSPISGGIWHSVLHSFVSGVSGSHVFQPILTGRALTDR YNAYEINAITGVKPFFTDVLLHTKRNPNTKAQMALESYPLNNAFPGQSFQMTSGIPPDLRAPVVFVLLPLRDLPPMH MGQDPNKPRNIIRHLNMPQMSFQETKDLERLPTRRSVETVEITERFASSKQAEEKTKKKKGKK

[0903] hAID智人(Homosapiens)

[0904] MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGCHVELLFLRYISDWD LDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLYFCEDRKAEPEGLRRLHRAGVQIAIMTFKDYFY CWNTFVENHERTFKAWEGLHENSVRLSRQLRRILLPLYEVDDLRDAFRTLGL

[0905] clAID犬(Canis lupus familiaris)

[0906] MDSLLMKQRKFLYHFKNVRWAKGRHETYLCYVVKRRDSATSFSLDFGHLRNKSGCHVELLFLRYISDWD LDPGRCYRVTWFTSWSPCYDCARHVADFLRGYPNLSLRIFAARLYFCEDRKAEPEGLRRLHRAGVQIAIMTFKDYFY CWNTFVENREKTFKAWEGLHENSVRLSRQLRRILLPLYEVDDLRDAFRTLGL

[0907] btAID牛(Bos Taurus)

[0908] MDSLLKKQRQFLYQFKNVRWAKGRHETYLCYVVKRRDSPTSFSLDFGHLRNKAGCHVELLFLRYISDWD LDPGRCYRVTWFTSWSPCYDCARHVADFLRGYPNLSLRIFTARLYFCDKERKAEPEGLRRLHRAGVQIAIMTFKDYF YCWNTFVENHERTFKAWEGLHENSVRLSRQLRRILLPLYEVDDLRDAFRTLGL

[0909] mAID小鼠(Mus musculus)

[0910] MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGCHVELLFLRYISDWD LDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLYFCEDRKAEPEGLRRLHRAGVQIAIMTFKDYFY CWNTFVENHERTFKAWEGLHENSVRLSRQLRRILLPLYEVDDLRDAFRTLGL

[0911] pmCDA-1海七鳃鳗(Petromyzon marinus)

[0912] MAGYECVRVSEKLDFDTFEFQFENLHYATERHRTYVIFDVKPQSAGGRSRRLWGYIINNPNVCHAELIL MSMIDRHLESNPGVYAMTWYMSWSPCANCSSKLNPWLKNLLEEQGHTLTMHFSRIYDRDREGDHRGLRGLKHVSNSF RMGVVGRAEVKECLAEYVEASRRTLTWLDTTESMAAKMRRKLFCILVRCAGMRESGIPLHLFTLQTPLLSGRVVWWR V

[0913] pmCDA-2海七鳃鳗(Petromyzon marinus)

[0914] MELREVVDCALASCVRHEPLSRVAFLRCFAAPSQKPRGTVILFYVEGAGRGVTGGHAVNYNKQGTSIHAEVLLLSAVRAALLRRRRCEDGEEATRGCTLHCYSTYSPCRDCVEYIQEFGASTGVRVVIHCCRLYELDVNRRRSEAEGVLRSLSRLGRDFRLMGPRDAIALLLGGRLANTADGESGASGNAWVTETNVVEPLVDMTGFGDEDLHAQVQRNKQIREAYANYASAVSLMLGELHVDPDKFPFLAEFLAQTSVEPSGTPRETRGRPRGASSRGPEIGRQRPADFERALGAYGLFLHPRIVSREADREEIKRDLIVVMRKHNYQGP

[0915] pmCDA-5海七鳃鳗(Petromyzon marinus)

[0916] MAGDENVRVSEKLDFDTFEFQFENLHYATERHRTYVIFDVKPQSAGGRSRRLWGYIINNPNVCHAELILMSMIDRHLESNPGVYAMTWYMSWSPCANCSSKLNPWLKNLLEEQGHTLMMHFSRIYDRDREGDHRGLRGLKHVSNSFRMGVVGRAEVKECLAEYVEASRRTLTWLDTTESMAAKMRRKLFCILVRCAGMRESGMPLHLFT

[0917] yCD酿酒酵母(Saccharomyces cerevisiae)

[0918] MVTGGMASKWDQKGMDIAYEEAALGYKEGGVPIGGCLINNKDGSVLGRGHNMRFQKGSATLHGEISTLENCGRLEGKVYKDTTLYTTLSPCDMCTGAIIMYGIPRCVVGENVNFKSKGEKYLQTRGHEVVVVDDERCKKIMKQFIDERPQDWFEDIGE

[0919] rAPOBEC-1(Δ177-186)

[0920] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLK

[0921] rAPOBEC-1(Δ202-213)

[0922] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQHYQRLPPHILWATGLK

[0923] 实施例4:CBE和ABE内部融合的构建

[0924] 通过将脱氨酶克隆到SpCas9或具有D10A突变的SpCas9切口酶内的高b-因子位置中而生成CBE和ABE内部融合构建体。在一些情况下,所述Cas9的结构性或功能性结构域被部分地或全部地删除并被TadA结构域所取代(IBE020)。CBE是以相同方式被插入,并且在其C末端上被尿嘧啶DNA糖基化酶抑制剂(UGI)结构域所修饰。

[0925] 示例性的内部融合碱基编辑器在下表10中提供:表10:

[0926]

BE ID	修饰	其他 ID
IBE001	Cas9 TadA ins 1015	ISLAY01
IBE002	Cas9 TadA ins 1022	ISLAY02
IBE003	Cas9 TadA ins 1029	ISLAY03
IBE004	Cas9 TadA ins 1040	ISLAY04
IBE005	Cas9 TadA ins 1068	ISLAY05
IBE006	Cas9 TadA ins 1247	ISLAY06
IBE007	Cas9 TadA ins 1054	ISLAY07
IBE008	Cas9 TadA ins 1026	ISLAY08
IBE009	Cas9 TadA ins 768	ISLAY09
IBE020	ΔHNH TadA 792	ISLAY20
IBE021	N端 融合 单一TadA 螺旋 截短的165-end	ISLAY21
IBE029	TadA- 环状变换排列体116 ins1067	ISLAY29
IBE031	TadA- 环状变换排列体 136 ins1248	ISLAY31
IBE032	TadA- 环状变换排列体 136 ins 1052	ISLAY32
IBE035	Δ 792-872 TadA ins	ISLAY35
IBE036	Δ 792-906 TadA ins	ISLAY36
IBE043	TadA- 环状变换排列体 65 ins1246	ISLAY43
IBE044	TadA ins C端 截短2 791	ISLAY44

[0927]

HR001	GGS-rAPOBEC1-XTEN-ins-位点1_Y1016-D10A-UGIx2	pHRB-043
HR002	GGS-rAPOBEC1-XTEN ins-位点2_A1023-D10A-UGIx2	pHRB-044
HR003	GGS-rAPOBEC1-XTEN ins-位点3_E1029-D10A-UGIx2	pHRB-045
	GGS-rAPOBEC1-XTEN ins-位点4_N1040-D10A-UGIx2	pHRB-046
HR004	GGS-rAPOBEC1-XTEN ins-位点5-T1069-D10A-UGIx2	pHRB-047
HR005	GGS-rAPOBEC1-XTEN ins-位点6-G1247-D10A-UGIx2	pHRB-048

[0928] 所述构建体的序列提供如下。

[0929] Cas9 TadAins 1015

[0930] ATGGACAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCGAC
GAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGG
AGCCCTGCTGTTCGACAGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGAC
GGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGA
CTGGAAGAGTCCTTCCTGGTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGT
GGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCGACCTGC
GGCTGATCTATCTGGCCCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGAC
AACAGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCCATCAACGC
CAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGC
TGCCCGGCGAGAAGAAGAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGC
AACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAACCTGCTGGC
CCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCC
TGAGAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGAC
CTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAA
CGGCTACGCCGGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGA
TGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGC
AGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCCTGAA
GGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACA
GCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGC
GCTTCCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCA
CAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGC
CCGCCTTCCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAG
CAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAA
CGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACGAGG
ACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTAT
GCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAA
GCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACA
GAAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAG
GGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAA
GGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACC
AGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGC
CAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGG
GCGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGA
GCTTTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTG
CCCTCCGAAGAGGTCGTGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAA
GTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGG
TGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGAC

AAGCTGATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTA
CAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGGAACCGCCCTGATCA
AAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGGGTTCTAGCGGCAGCGAGACTCCCGGG
ACCTCAGAGTCCGCCACACCCGAAAGTTCTGGTTCCGAAGTCGAGTTTTCCCATGAGTACTGGATGAGACACGCATT
GACTCTCGCAAAGAGGGCTCGAGATGAACGCGAGGTGCCCGTGGGGGCAGTACTCGTGCTCAACAATCGCGTAATCG
GCGAAGGTTGGAATAGGGCAATCGGACTCCACGACCCCACTGCACATGCGGAAATCATGGCCCTTCGACAGGGAGGG
CTTGTGATGCAGAATTATCGACTTATCGATGCGACGCTGTACGTCACGTTTGAACCTTGCGTAATGTGCGCGGGAGC
TATGATTCACTCCCGCATTGGACGAGTTGTATTCGGTGTTCGCAACGCCAAGACGGGTGCCGCAGGTTCACTGATGG
ACGTGCTGCATTACCCAGGCATGAACCACCGGGTAGAAATCACAGAAGGCATATTGGCGGACGAATGTGCGGCGCTG
TTGTGTTACTTTTTTCGCATGCCCAGGCAGGTCTTTAACGCCCAGAAAAAAGCACAATCCTCTACTGACTACGACGT
GCGGAAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCATGA
ACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGAAACC
GGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGT
GAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACAGCGATAAGCTGATCG
CCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTG
GCCAAAGTGGAAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAG
CAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGC
TGCCTAAGTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGA
AACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCC
CGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCG
AGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGATAAG
CCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTA
CTTTGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCA
TCACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGGGAGGTGAC

[0931] Cas9 TadAins 1015

[0932] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVL

NNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGA
AGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDYDVRKMIAKSEQEIGKATAKYFF
YSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN
SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK
DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD

[0933] Cas9 TadAins 1022

[0934] ATGGACAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCGAC
GAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGG
AGCCCTGCTGTTCGACAGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGAC
GGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGA
CTGGAAGAGTCCTTCCTGGTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGT
GGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCGACCTGC
GGCTGATCTATCTGGCCCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGAC
AACAGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCCATCAACGC
CAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGC
TGCCCGGCGAGAAGAAGAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGC
AACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAACCTGCTGGC
CCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCC
TGAGAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGAC
CTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAA
CGGCTACGCCGGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGA
TGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGC
AGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCCTGAA
GGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACA
GCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGC
GCTTCCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCA
CAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGC
CCGCCTTCCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAG
CAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAA
CGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACGAGG
ACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTAT
GCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAA
GCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACA
GAAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAG
GGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAA
GGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACC
AGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGC

CAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGG
GCGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGA
GCTTTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTG
CCCTCCGAAGAGGTCGTGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAA
GTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGG
TGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGAC
AAGCTGATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTA
CAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGGAACCGCCCTGATCA
AAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGGTTCT
AGCGGCAGCGAGACTCCCGGGACCTCAGAGTCCGCCACACCCGAAAGTTCTGGTTCCGAAGTCGAGTTTTCCCATGA
GTACTGGATGAGACACGCATTGACTCTCGCAAAGAGGGCTCGAGATGAACGCGAGGTGCCCGTGGGGGCAGTACTCG
TGCTCAACAATCGCGTAATCGGCGAAGGTTGGAATAGGGCAATCGGACTCCACGACCCCACTGCACATGCGGAAATC
ATGGCCCTTCGACAGGGAGGGCTTGTGATGCAGAATTATCGACTTATCGATGCGACGCTGTACGTCACGTTTGAACC
TTGCGTAATGTGCGCGGGAGCTATGATTCACTCCCGCATTGGACGAGTTGTATTCGGTGTTCGCAACGCCAAGACGG
GTGCCGCAGGTTCACTGATGGACGTGCTGCATTACCCAGGCATGAACCACCGGGTAGAAATCACAGAAGGCATATTG
GCGGACGAATGTGCGGCGCTGTTGTGTTACTTTTTTCGCATGCCCAGGCAGGTCTTTAACGCCCAGAAAAAAGCACA
ATCCTCTACTGACGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCATGA
ACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGAAACC
GGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGT
GAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACAGCGATAAGCTGATCG
CCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTG
GCCAAAGTGGAAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAG
CAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGC
TGCCTAAGTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGA
AACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCC
CGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCG
AGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGATAAG
CCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTA
CTTTGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCA
TCACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGGGAGGTGAC

[0935] Cas9 TadAins 1022

[0936] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD

NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVP
VGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGV
RNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDAKSEQEIGKATAKYFF
YSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN
SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK
DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD

[0937] IBE003_Cas9:Cas9 TadAins 1029

[0938] ATGGACAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCGAC
GAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGG
AGCCCTGCTGTTCGACAGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGAC
GGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGA
CTGGAAGAGTCCTTCCTGGTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGT
GGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCGACCTGC
GGCTGATCTATCTGGCCCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGAC
AACAGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCCATCAACGC
CAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGC
TGCCCGGCGAGAAGAAGAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGC
AACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAACCTGCTGGC
CCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCC
TGAGAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGAC
CTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAA
CGGCTACGCCGGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGA
TGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGC
AGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCCTGAA
GGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACA
GCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGC
GCTTCCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCA
CAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGC
CCGCCTTCCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAG
CAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAA
CGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACGAGG
ACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTAT

GCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAA
GCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACA
GAAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAG
GGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAA
GGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACC
AGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGC
CAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGG
GCGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGA
GCTTTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTG
CCCTCCGAAGAGGTCGTGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAA
GTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGG
TGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGAC
AAGCTGATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTA
CAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGGAACCGCCCTGATCA
AAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAG
AGCGAGCAGGAAATCGGTTCTAGCGGCAGCGAGACTCCCGGGACCTCAGAGTCCGCCACACCCGAAAGTTCTGGTTC
CGAAGTCGAGTTTTCCCATGAGTACTGGATGAGACACGCATTGACTCTCGCAAAGAGGGCTCGAGATGAACGCGAGG
TGCCCGTGGGGGCAGTACTCGTGCTCAACAATCGCGTAATCGGCGAAGGTTGGAATAGGGCAATCGGACTCCACGAC
CCCACTGCACATGCGGAAATCATGGCCCTTCGACAGGGAGGGCTTGTGATGCAGAATTATCGACTTATCGATGCGAC
GCTGTACGTCACGTTTGAACCTTGCGTAATGTGCGCGGGAGCTATGATTCACTCCCGCATTGGACGAGTTGTATTCG
GTGTTCGCAACGCCAAGACGGGTGCCGCAGGTTCACTGATGGACGTGCTGCATTACCCAGGCATGAACCACCGGGTA
GAAATCACAGAAGGCATATTGGCGGACGAATGTGCGGCGCTGTTGTGTTACTTTTTTCGCATGCCCAGGCAGGTCTT
TAACGCCCAGAAAAAAGCACAATCCTCTACTGACGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCATGA
ACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGAAACC
GGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGT
GAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACAGCGATAAGCTGATCG
CCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTG
GCCAAAGTGGAAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAG
CAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGC
TGCCTAAGTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGA
AACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCC
CGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCG
AGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGATAAG
CCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTA
CTTTGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCA
TCACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGGGAGGTGAC

[0939] Cas9 TadAins 1029

[0940] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST

DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRA
RDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRI
GRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGKATAKYFF
YSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN
SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK
DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD

[0941] IBE004_Cas9:Cas9 TadAins 1040

[0942] ATGGACAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCGAC
GAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGG
AGCCCTGCTGTTCGACAGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGAC
GGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGA
CTGGAAGAGTCCTTCCTGGTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGT
GGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCGACCTGC
GGCTGATCTATCTGGCCCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGAC
AACAGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCCATCAACGC
CAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGC
TGCCCGGCGAGAAGAAGAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGC
AACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAACCTGCTGGC
CCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCC
TGAGAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGAC
CTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAA
CGGCTACGCCGGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGA
TGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGC
AGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCCTGAA
GGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACA
GCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGC

GCTTCCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCA
CAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGC
CCGCCTTCCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAG
CAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAA
CGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACGAGG
ACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTAT
GCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAA
GCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACA
GAAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAG
GGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAA
GGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACC
AGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGC
CAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGG
GCGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGA
GCTTTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTG
CCCTCCGAAGAGGTCGTGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAA
GTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGG
TGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGAC
AAGCTGATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTA
CAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGGAACCGCCCTGATCA
AAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAG
AGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCGGTTCTAGCGGCAGCGAGACTCCCGGGAC
CTCAGAGTCCGCCACACCCGAAAGTTCTGGTTCCGAAGTCGAGTTTTCCCATGAGTACTGGATGAGACACGCATTGA
CTCTCGCAAAGAGGGCTCGAGATGAACGCGAGGTGCCCGTGGGGGCAGTACTCGTGCTCAACAATCGCGTAATCGGC
GAAGGTTGGAATAGGGCAATCGGACTCCACGACCCCACTGCACATGCGGAAATCATGGCCCTTCGACAGGGAGGGCT
TGTGATGCAGAATTATCGACTTATCGATGCGACGCTGTACGTCACGTTTGAACCTTGCGTAATGTGCGCGGGAGCTA
TGATTCACTCCCGCATTGGACGAGTTGTATTCGGTGTTCGCAACGCCAAGACGGGTGCCGCAGGTTCACTGATGGAC
GTGCTGCATTACCCAGGCATGAACCACCGGGTAGAAATCACAGAAGGCATATTGGCGGACGAATGTGCGGCGCTGTT
GTGTTACTTTTTTCGCATGCCCAGGCAGGTCTTTAACGCCCAGAAAAAAGCACAATCCTCTACTGACAACATCATGA
ACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGAAACC
GGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGT
GAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACAGCGATAAGCTGATCG
CCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTG
GCCAAAGTGGAAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAG
CAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGC
TGCCTAAGTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGA
AACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCC
CGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCG
AGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGATAAG

CCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTA
CTTTGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCA
TCACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGGGAGGTGAC

[0943] Cas9 TadAins 1040

[0944] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSGSSGSETPGTSESATPESSGSEVEFSHEYW
MRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCV
MCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSS
TDNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN
SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK
DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD

[0945] IBE005_Cas9:Cas9 TadAins 1068

[0946] ATGGACAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCGAC
GAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGG
AGCCCTGCTGTTCGACAGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGAC
GGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGA
CTGGAAGAGTCCTTCCTGGTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGT
GGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCGACCTGC
GGCTGATCTATCTGGCCCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGAC
AACAGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCCATCAACGC
CAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGC
TGCCCGGCGAGAAGAAGAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGC
AACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAACCTGCTGGC
CCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCC
TGAGAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGAC

CTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAA
CGGCTACGCCGGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGA
TGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGC
AGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCCTGAA
GGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACA
GCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGC
GCTTCCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCA
CAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGC
CCGCCTTCCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAG
CAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAA
CGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACGAGG
ACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTAT
GCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAA
GCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACA
GAAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAG
GGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAA
GGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACC
AGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGC
CAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGG
GCGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGA
GCTTTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTG
CCCTCCGAAGAGGTCGTGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAA
GTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGG
TGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGAC
AAGCTGATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTA
CAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGGAACCGCCCTGATCA
AAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAG
AGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCATGAACTTTTTCAAGACCGAGAT
TACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGAAGGTTCTAGCGGCAGCGAGACTC
CCGGGACCTCAGAGTCCGCCACACCCGAAAGTTCTGGTTCCGAAGTCGAGTTTTCCCATGAGTACTGGATGAGACAC
GCATTGACTCTCGCAAAGAGGGCTCGAGATGAACGCGAGGTGCCCGTGGGGGCAGTACTCGTGCTCAACAATCGCGT
AATCGGCGAAGGTTGGAATAGGGCAATCGGACTCCACGACCCCACTGCACATGCGGAAATCATGGCCCTTCGACAGG
GAGGGCTTGTGATGCAGAATTATCGACTTATCGATGCGACGCTGTACGTCACGTTTGAACCTTGCGTAATGTGCGCG
GGAGCTATGATTCACTCCCGCATTGGACGAGTTGTATTCGGTGTTCGCAACGCCAAGACGGGTGCCGCAGGTTCACT
GATGGACGTGCTGCATTACCCAGGCATGAACCACCGGGTAGAAATCACAGAAGGCATATTGGCGGACGAATGTGCGG
CGCTGTTGTGTTACTTTTTTCGCATGCCCAGGCAGGTCTTTAACGCCCAGAAAAAAGCACAATCCTCTACTGACACC
GGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGT
GAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACAGCGATAAGCTGATCG
CCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTG

GCCAAAGTGGAAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAG
CAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGC
TGCCTAAGTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGA
AACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCC
CGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCG
AGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGATAAG
CCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTA
CTTTGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCA
TCACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGGGAGGTGAC

[0947] Cas9 TadAins 1068

[0948] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGEGS
SGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEI
MALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGIL
ADECAALLCYFFRMPRQVFNAQKKAQSSTDTGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN
SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK
DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD

[0949] IBE006_Cas9:Cas9 TadAins 1247

[0950] ATGGACAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCGAC
GAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGG
AGCCCTGCTGTTCGACAGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGAC
GGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGA
CTGGAAGAGTCCTTCCTGGTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGT
GGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCGACCTGC
GGCTGATCTATCTGGCCCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGAC

AACAGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCCATCAACGC
CAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGC
TGCCCGGCGAGAAGAAGAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGC
AACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAACCTGCTGGC
CCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCC
TGAGAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGAC
CTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAA
CGGCTACGCCGGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGA
TGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGC
AGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCCTGAA
GGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACA
GCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGC
GCTTCCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCA
CAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGC
CCGCCTTCCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAG
CAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAA
CGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACGAGG
ACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTAT
GCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAA
GCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACA
GAAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAG
GGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAA
GGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACC
AGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGC
CAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGG
GCGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGA
GCTTTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTG
CCCTCCGAAGAGGTCGTGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAA
GTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGG
TGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGAC
AAGCTGATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTA
CAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGGAACCGCCCTGATCA
AAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAG
AGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCATGAACTTTTTCAAGACCGAGAT
TACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATA
AGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAG
ACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGA
CCCTAAGAAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCA
AGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCC

ATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTT
CGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAACGAACTGGCCCTGCCCT
CCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCGGTTCTAGCGGCAGCGAGACTCCC
GGGACCTCAGAGTCCGCCACACCCGAAAGTTCTGGTTCCGAAGTCGAGTTTTCCCATGAGTACTGGATGAGACACGC
ATTGACTCTCGCAAAGAGGGCTCGAGATGAACGCGAGGTGCCCGTGGGGGCAGTACTCGTGCTCAACAATCGCGTAA
TCGGCGAAGGTTGGAATAGGGCAATCGGACTCCACGACCCCACTGCACATGCGGAAATCATGGCCCTTCGACAGGGA
GGGCTTGTGATGCAGAATTATCGACTTATCGATGCGACGCTGTACGTCACGTTTGAACCTTGCGTAATGTGCGCGGG
AGCTATGATTCACTCCCGCATTGGACGAGTTGTATTCGGTGTTCGCAACGCCAAGACGGGTGCCGCAGGTTCACTGA
TGGACGTGCTGCATTACCCAGGCATGAACCACCGGGTAGAAATCACAGAAGGCATATTGGCGGACGAATGTGCGGCG
CTGTTGTGTTACTTTTTTCGCATGCCCAGGCAGGTCTTTAACGCCCAGAAAAAAGCACAATCCTCTACTGACTCCCC
CGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCG
AGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGATAAG
CCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTA
CTTTGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCA
TCACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGGGAGGTGAC

[0951] Cas9 TadAins 1247

[0952] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLV
LNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTG
AAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD

[0953] IBE007_Cas9:Cas9 TadAins 1054

[0954] ATGGACAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCGAC

GAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGG
AGCCCTGCTGTTCGACAGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGAC
GGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGA
CTGGAAGAGTCCTTCCTGGTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGT
GGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCGACCTGC
GGCTGATCTATCTGGCCCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGAC
AACAGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCCATCAACGC
CAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGC
TGCCCGGCGAGAAGAAGAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGC
AACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAACCTGCTGGC
CCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCC
TGAGAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGAC
CTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAA
CGGCTACGCCGGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGA
TGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGC
AGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCCTGAA
GGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACA
GCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGC
GCTTCCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCA
CAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGC
CCGCCTTCCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAG
CAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAA
CGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACGAGG
ACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTAT
GCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAA
GCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACA
GAAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAG
GGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAA
GGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACC
AGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGC
CAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGG
GCGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGA
GCTTTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTG
CCCTCCGAAGAGGTCGTGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAA
GTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGG
TGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGAC
AAGCTGATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTA
CAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGGAACCGCCCTGATCA
AAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAG

AGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCATGAACTTTTTCAAGACCGAGAT
TACCCTGGCCAACGGTTCTAGCGGCAGCGAGACTCCCGGGACCTCAGAGTCCGCCACACCCGAAAGTTCTGGTTCCG
AAGTCGAGTTTTCCCATGAGTACTGGATGAGACACGCATTGACTCTCGCAAAGAGGGCTCGAGATGAACGCGAGGTG
CCCGTGGGGGCAGTACTCGTGCTCAACAATCGCGTAATCGGCGAAGGTTGGAATAGGGCAATCGGACTCCACGACCC
CACTGCACATGCGGAAATCATGGCCCTTCGACAGGGAGGGCTTGTGATGCAGAATTATCGACTTATCGATGCGACGC
TGTACGTCACGTTTGAACCTTGCGTAATGTGCGCGGGAGCTATGATTCACTCCCGCATTGGACGAGTTGTATTCGGT
GTTCGCAACGCCAAGACGGGTGCCGCAGGTTCACTGATGGACGTGCTGCATTACCCAGGCATGAACCACCGGGTAGA
AATCACAGAAGGCATATTGGCGGACGAATGTGCGGCGCTGTTGTGTTACTTTTTTCGCATGCCCAGGCAGGTCTTTA
ACGCCCAGAAAAAAGCACAATCCTCTACTGACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGAAACC
GGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGT
GAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACAGCGATAAGCTGATCG
CCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTG
GCCAAAGTGGAAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAG
CAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGC
TGCCTAAGTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGA
AACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCC
CGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCG
AGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGATAAG
CCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTA
CTTTGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCA
TCACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGGGAGGTGAC

[0955] Cas9 TadAins 1054

[0956] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGSSGSETPGTSESATP
ESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYR
LIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRM
PRQVFNAQKKAQSSTDGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN

SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK
DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD

[0957] IBE008_Cas9:Cas9 TadAins 1026

[0958] ATGGACAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCGAC
GAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGG
AGCCCTGCTGTTCGACAGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGAC
GGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGA
CTGGAAGAGTCCTTCCTGGTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGT
GGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCGACCTGC
GGCTGATCTATCTGGCCCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGAC
AACAGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCCATCAACGC
CAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGC
TGCCCGGCGAGAAGAAGAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGC
AACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAACCTGCTGGC
CCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCC
TGAGAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGAC
CTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAA
CGGCTACGCCGGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGA
TGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGC
AGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCCTGAA
GGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACA
GCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGC
GCTTCCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCA
CAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGC
CCGCCTTCCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAG
CAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAA
CGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACGAGG
ACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTAT
GCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAA
GCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACA
GAAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAG
GGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAA
GGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACC
AGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGC
CAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGG
GCGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGA
GCTTTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTG

CCCTCCGAAGAGGTCGTGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAA
GTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGG
TGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGAC
AAGCTGATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTA
CAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGGAACCGCCCTGATCA
AAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAG
AGCGAGGGTTCTAGCGGCAGCGAGACTCCCGGGACCTCAGAGTCCGCCACACCCGAAAGTTCTGGTTCCGAAGTCGA
GTTTTCCCATGAGTACTGGATGAGACACGCATTGACTCTCGCAAAGAGGGCTCGAGATGAACGCGAGGTGCCCGTGG
GGGCAGTACTCGTGCTCAACAATCGCGTAATCGGCGAAGGTTGGAATAGGGCAATCGGACTCCACGACCCCACTGCA
CATGCGGAAATCATGGCCCTTCGACAGGGAGGGCTTGTGATGCAGAATTATCGACTTATCGATGCGACGCTGTACGT
CACGTTTGAACCTTGCGTAATGTGCGCGGGAGCTATGATTCACTCCCGCATTGGACGAGTTGTATTCGGTGTTCGCA
ACGCCAAGACGGGTGCCGCAGGTTCACTGATGGACGTGCTGCATTACCCAGGCATGAACCACCGGGTAGAAATCACA
GAAGGCATATTGGCGGACGAATGTGCGGCGCTGTTGTGTTACTTTTTTCGCATGCCCAGGCAGGTCTTTAACGCCCA
GAAAAAAGCACAATCCTCTACTGACCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCATGA
ACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGAAACC
GGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGT
GAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACAGCGATAAGCTGATCG
CCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTG
GCCAAAGTGGAAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAG
CAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGC
TGCCTAAGTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGA
AACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCC
CGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCG
AGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGATAAG
CCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTA
CTTTGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCA
TCACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGGGAGGTGAC

[0959] Cas9 TadAins 1026

[0960] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTA
RRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVD
STDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSR
RLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNL
SDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRI
PYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNEL
TKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG
KTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMG
RHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL

DINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAE
RGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINN
YHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEGSSGSETPGTSESATPESSGSEVEFSHEYW
MRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPC
VMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQ
SSTDQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKT
EVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSF
EKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPE
DNEQKQLFVEQHKHYLDEI IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFK
YFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[0961] IBE009_Cas9:Cas9 TadAins 768

[0962] ATGGACAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCGAC
GAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGG
AGCCCTGCTGTTCGACAGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGAC
GGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGA
CTGGAAGAGTCCTTCCTGGTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGT
GGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCGACCTGC
GGCTGATCTATCTGGCCCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGAC
AACAGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCCATCAACGC
CAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGC
TGCCCGGCGAGAAGAAGAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGC
AACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAACCTGCTGGC
CCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCC
TGAGAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGAC
CTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAA
CGGCTACGCCGGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGA
TGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGC
AGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCCTGAA
GGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACA
GCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGC
GCTTCCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCA
CAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGC
CCGCCTTCCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAG
CAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAA
CGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACGAGG
ACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTAT
GCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAA
GCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACA
GAAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAG

GGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAA
GGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACC
AGGGTTCTAGCGGCAGCGAGACTCCCGGGACCTCAGAGTCCGCCACACCCGAAAGTTCTGGTTCCGAAGTCGAGTTT
TCCCATGAGTACTGGATGAGACACGCATTGACTCTCGCAAAGAGGGCTCGAGATGAACGCGAGGTGCCCGTGGGGGC
AGTACTCGTGCTCAACAATCGCGTAATCGGCGAAGGTTGGAATAGGGCAATCGGACTCCACGACCCCACTGCACATG
CGGAAATCATGGCCCTTCGACAGGGAGGGCTTGTGATGCAGAATTATCGACTTATCGATGCGACGCTGTACGTCACG
TTTGAACCTTGCGTAATGTGCGCGGGAGCTATGATTCACTCCCGCATTGGACGAGTTGTATTCGGTGTTCGCAACGC
CAAGACGGGTGCCGCAGGTTCACTGATGGACGTGCTGCATTACCCAGGCATGAACCACCGGGTAGAAATCACAGAAG
GCATATTGGCGGACGAATGTGCGGCGCTGTTGTGTTACTTTTTTCGCATGCCCAGGACCACCCAGAAGGGACAGAAG
AACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGT
GGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGG
AACTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCATC
GACAACAAGGTGCTGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCGTGAAGAA
GATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTCGACAATCTGACCAAGGCCG
AGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGCAGATCACAAAG
CACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGACAAGCTGATCCGGGAAGTGAAAGT
GATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTACAAAGTGCGCGAGATCAACAACT
ACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGC
GAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGC
TACCGCCAAGTACTTCTTCTACAGCAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCC
GGAAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTG
CGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTC
TATCCTGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTTCG
ACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGTCCAAGAAACTGAAGAGTGTG
AAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGG
CTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGA
GAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTAC
CTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAA
GCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACA
AAGTGCTGTCCGCCTACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACC
CTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGAGGTACACCAGCACCAA
AGAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGG
GAGGTGAC

[0963] Cas9 TadAins 768

[0964] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK

PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLH
DPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHR
VEITEGILADECAALLCYFFRMPRTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRD
MYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFD
NLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKV
REINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITL
ANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPK
KYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFEL
ENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILA
DANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRI
DLSQLGGD

[0965] IBE020_delt:ΔHNH TadA 792

[0966] ATGGACAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCGAC
GAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGG
AGCCCTGCTGTTCGACAGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGAC
GGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGA
CTGGAAGAGTCCTTCCTGGTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGT
GGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCGACCTGC
GGCTGATCTATCTGGCCCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGAC
AACAGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCCATCAACGC
CAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGC
TGCCCGGCGAGAAGAAGAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGC
AACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAACCTGCTGGC
CCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCC
TGAGAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGAC
CTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAA
CGGCTACGCCGGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGA
TGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGC
AGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCCTGAA
GGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACA
GCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGC
GCTTCCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCA
CAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGC
CCGCCTTCCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAG

CAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAA
CGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACGAGG
ACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTAT
GCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAA
GCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACA
GAAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAG
GGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAA
GGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACC
AGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCTCC
GAAGTCGAGTTTTCCCATGAGTACTGGATGAGACACGCATTGACTCTCGCAAAGAGGGCTCGAGATGAACGCGAGGT
GCCCGTGGGGGCAGTACTCGTGCTCAACAATCGCGTAATCGGCGAAGGTTGGAATAGGGCAATCGGACTCCACGACC
CCACTGCACATGCGGAAATCATGGCCCTTCGACAGGGAGGGCTTGTGATGCAGAATTATCGACTTATCGATGCGACG
CTGTACGTCACGTTTGAACCTTGCGTAATGTGCGCGGGAGCTATGATTCACTCCCGCATTGGACGAGTTGTATTCGG
TGTTCGCAACGCCAAGACGGGTGCCGCAGGTTCACTGATGGACGTGCTGCATTACCCAGGCATGAACCACCGGGTAG
AAATCACAGAAGGCATATTGGCGGACGAATGTGCGGCGCTGTTGTGTTACTTTTTTCGCATGCCCAGGCAGGTCTTT
AACGCCCAGAAAAAAGCACAATCCTCTACTGACGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACA
GCTGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGA
ATGACAAGCTGATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAG
TTTTACAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGGAACCGCCCT
GATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGATCG
CCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCATGAACTTTTTCAAGACC
GAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTG
GGATAAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGG
TGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGAC
TGGGACCCTAAGAAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAA
GGGCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGA
ATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACTCC
CTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAACGAACTGGCCCT
GCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCCGAGGATAATGAGC
AGAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGA
GTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGATAAGCCCATCAGAGAGCA
GGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCA
TCGACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTAC
GAGACACGGATCGACCTGTCTCAGCTGGGAGGTGAC

[0967] ΔHNH TadA 792

[0968] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRA
IGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPG
MNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMN
TKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYD
VRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNI
VKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMER
SSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGS
PEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFK
YFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[0969] IBE021_N-te:N末端融合单一TadA螺旋截短的165-end

[0970] ATGTCCGAAGTCGAGTTTTCCCATGAGTACTGGATGAGACACGCATTGACTCTCGCAAAGAGGGCTCGA
GATGAACGCGAGGTGCCCGTGGGGGCAGTACTCGTGCTCAACAATCGCGTAATCGGCGAAGGTTGGAATAGGGCAAT
CGGACTCCACGACCCCACTGCACATGCGGAAATCATGGCCCTTCGACAGGGAGGGCTTGTGATGCAGAATTATCGAC
TTATCGATGCGACGCTGTACGTCACGTTTGAACCTTGCGTAATGTGCGCGGGAGCTATGATTCACTCCCGCATTGGA
CGAGTTGTATTCGGTGTTCGCAACGCCAAGACGGGTGCCGCAGGTTCACTGATGGACGTGCTGCATTACCCAGGCAT
GAACCACCGGGTAGAAATCACAGAAGGCATATTGGCGGACGAATGTGCGGCGCTGTTGTGTTACTTTTTTCGCATGC
CCAGGTCTGGTGGTTCTTCTGGTGGTTCTAGCGGCAGCGAGACTCCCGGGACCTCAGAGTCCGCCACACCCGAAAGT
TCTGGTGGTTCTTCTGGTGGTTCTGACAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCTGGGC
CGTGATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGCATCAAGA
AGAACCTGATCGGAGCCCTGCTGTTCGACAGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGA
AGATACACCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAG
CTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACA
TCGTGGACGAGGTGGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCGAC
AAGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGA
CCTGAACCCCGACAACAGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAA
ACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAAT
CTGATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGGGCCTGACCCC
CAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGG
ACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTG
CTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGATACGACGA
GCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCG
ACCAGAGCAAGAACGGCTACGCCGGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCC
ATCCTGGAAAAGATGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGAC

CTTCGACAACGGCAGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTT
ACCCATTCCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTG
GCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGT
GGTGGACAAGGGCGCTTCCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACGAGAAGG
TGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAG
GGAATGAGAAAGCCCGCCTTCCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAA
AGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGG
AAGATCGGTTCAACGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAAT
GAGGAAAACGAGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACG
GCTGAAAACCTATGCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCA
GGCTGAGCCGGAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAGTCCGAC
GGCTTCGCCAACAGAAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCA
GGTGTCCGGCCAGGGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCC
TGCAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATG
GCCAGAGAGAACCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAA
AGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACT
ACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTACGATGTGGACCAT
ATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACCGGGGCAA
GAGCGACAACGTGCCCTCCGAAGAGGTCGTGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGA
TTACCCAGAGAAAGTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATC
AAGAGACAGCTGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTA
CGACGAGAATGACAAGCTGATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGG
ATTTCCAGTTTTACAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGGA
ACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAA
GATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCATGAACTTTT
TCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGGGAG
ATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAA
GACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAA
AGAAGGACTGGGACCCTAAGAAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAA
GTGGAAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTT
CGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTA
AGTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAACGAA
CTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCCGAGGA
TAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCT
CCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGATAAGCCCATC
AGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGA
CACCACCATCGACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCG
GCCTGTACGAGACACGGATCGACCTGTCTCAGCTGGGAGGTGAC

[0971] IBE021_N-te:N末端融合单一TadA螺旋截短的165-end

[0972] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLC
YFFRMPRSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTD
RHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERH
PIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYN
QLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDT
YDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKY
KEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILR
RQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKN
LPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSV
EISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRR
YTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA
IKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNE
KLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQL
LNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLV
SDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYS
NIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSD
KLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDL
IIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIE
QISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLI
HQSITGLYETRIDLSQLGGD

[0973] IBE029_ISLA:TadA-CP116ins 1067

[0974] ATGGACAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCGAC
GAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGG
AGCCCTGCTGTTCGACAGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGAC
GGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGA
CTGGAAGAGTCCTTCCTGGTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGT
GGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCGACCTGC
GGCTGATCTATCTGGCCCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGAC
AACAGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCCATCAACGC
CAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGC
TGCCCGGCGAGAAGAAGAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGC
AACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAACCTGCTGGC
CCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCC
TGAGAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGAC
CTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAA
CGGCTACGCCGGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGA
TGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGC
AGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCCTGAA

GGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACA
GCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGC
GCTTCCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCA
CAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGC
CCGCCTTCCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAG
CAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAA
CGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACGAGG
ACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTAT
GCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAA
GCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACA
GAAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAG
GGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAA
GGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACC
AGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGC
CAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGG
GCGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGA
GCTTTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTG
CCCTCCGAAGAGGTCGTGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAA
GTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGG
TGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGAC
AAGCTGATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTA
CAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGGAACCGCCCTGATCA
AAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAG
AGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCATGAACTTTTTCAAGACCGAGAT
TACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACATGAACCACCGGGTAGAAATCACAGAAG
GCATATTGGCGGACGAATGTGCGGCGCTGTTGTGTTACTTTTTTCGCATGCCCAGGCAGGTCTTTAACGCCCAGAAA
AAAGCACAATCCTCTACTGACGGTTCTAGCGGCAGCGAGACTCCCGGGACCTCAGAGTCCGCCACACCCGAAAGTTC
TGGTTCCGAAGTCGAGTTTTCCCATGAGTACTGGATGAGACACGCATTGACTCTCGCAAAGAGGGCTCGAGATGAAC
GCGAGGTGCCCGTGGGGGCAGTACTCGTGCTCAACAATCGCGTAATCGGCGAAGGTTGGAATAGGGCAATCGGACTC
CACGACCCCACTGCACATGCGGAAATCATGGCCCTTCGACAGGGAGGGCTTGTGATGCAGAATTATCGACTTATCGA
TGCGACGCTGTACGTCACGTTTGAACCTTGCGTAATGTGCGCGGGAGCTATGATTCACTCCCGCATTGGACGAGTTG
TATTCGGTGTTCGCAACGCCAAGACGGGTGCCGCAGGTTCACTGATGGACGTGCTGCATTACCCAGGCGGCGAAACC
GGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGT
GAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACAGCGATAAGCTGATCG
CCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTG
GCCAAAGTGGAAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAG
CAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGC
TGCCTAAGTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGA
AACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCC

CGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCG
AGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGATAAG
CCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTA
CTTTGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCA
TCACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGGGAGGTGAC

[0975] IBE029_ISLA:TadA-CP116ins 1067

[0976] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNMNHR
VEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAK
RARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHS
RIGRVVFGVRNAKTGAAGSLMDVLHYPGGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN
SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK
DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD

[0977] IBE031_ISLA:TadACP136ins 1248

[0978] ATGGACAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCGAC
GAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGG
AGCCCTGCTGTTCGACAGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGAC
GGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGA
CTGGAAGAGTCCTTCCTGGTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGT
GGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCGACCTGC
GGCTGATCTATCTGGCCCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGAC
AACAGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCCATCAACGC
CAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGC
TGCCCGGCGAGAAGAAGAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGC
AACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAACCTGCTGGC

CCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCC
TGAGAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGAC
CTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAA
CGGCTACGCCGGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGA
TGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGC
AGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCCTGAA
GGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACA
GCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGC
GCTTCCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCA
CAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGC
CCGCCTTCCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAG
CAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAA
CGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACGAGG
ACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTAT
GCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAA
GCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACA
GAAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAG
GGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAA
GGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACC
AGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGC
CAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGG
GCGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGA
GCTTTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTG
CCCTCCGAAGAGGTCGTGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAA
GTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGG
TGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGAC
AAGCTGATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTA
CAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGGAACCGCCCTGATCA
AAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAG
AGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCATGAACTTTTTCAAGACCGAGAT
TACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATA
AGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAG
ACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGA
CCCTAAGAAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCA
AGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCC
ATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTT
CGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAACGAACTGGCCCTGCCCT
CCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCATGAACCACCGGGTAGAAATC
ACAGAAGGCATATTGGCGGACGAATGTGCGGCGCTGTTGTGTTACTTTTTTCGCATGCCCAGGCAGGTCTTTAACGC

CCAGAAAAAAGCACAATCCTCTACTGACGGTTCTAGCGGCAGCGAGACTCCCGGGACCTCAGAGTCCGCCACACCCG
AAAGTTCTGGTTCCGAAGTCGAGTTTTCCCATGAGTACTGGATGAGACACGCATTGACTCTCGCAAAGAGGGCTCGA
GATGAACGCGAGGTGCCCGTGGGGGCAGTACTCGTGCTCAACAATCGCGTAATCGGCGAAGGTTGGAATAGGGCAAT
CGGACTCCACGACCCCACTGCACATGCGGAAATCATGGCCCTTCGACAGGGAGGGCTTGTGATGCAGAATTATCGAC
TTATCGATGCGACGCTGTACGTCACGTTTGAACCTTGCGTAATGTGCGCGGGAGCTATGATTCACTCCCGCATTGGA
CGAGTTGTATTCGGTGTTCGCAACGCCAAGACGGGTGCCGCAGGTTCACTGATGGACGTGCTGCATTACCCAGGCCC
CGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCG
AGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGATAAG
CCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTA
CTTTGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCA
TCACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGGGAGGTGAC

[0979] IBE031_ISLA:TadACP136ins 1248

[0980] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGSMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGSETPGTS
ESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD

[0981] IBE032_ISLA:TadACP136ins 1052

[0982] ATGGACAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCGAC
GAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGG
AGCCCTGCTGTTCGACAGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGAC
GGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGA
CTGGAAGAGTCCTTCCTGGTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGT

GGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCGACCTGC
GGCTGATCTATCTGGCCCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGAC
AACAGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCCATCAACGC
CAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGC
TGCCCGGCGAGAAGAAGAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGC
AACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAACCTGCTGGC
CCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCC
TGAGAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGAC
CTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAA
CGGCTACGCCGGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGA
TGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGC
AGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCCTGAA
GGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACA
GCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGC
GCTTCCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCA
CAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGC
CCGCCTTCCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAG
CAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAA
CGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACGAGG
ACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTAT
GCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAA
GCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACA
GAAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAG
GGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAA
GGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACC
AGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGC
CAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGG
GCGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGA
GCTTTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTG
CCCTCCGAAGAGGTCGTGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAA
GTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGG
TGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGAC
AAGCTGATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTA
CAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGGAACCGCCCTGATCA
AAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAG
AGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCATGAACTTTTTCAAGACCGAGAT
TACCCTGGCCATGAACCACCGGGTAGAAATCACAGAAGGCATATTGGCGGACGAATGTGCGGCGCTGTTGTGTTACT
TTTTTCGCATGCCCAGGCAGGTCTTTAACGCCCAGAAAAAAGCACAATCCTCTACTGACGGTTCTAGCGGCAGCGAG
ACTCCCGGGACCTCAGAGTCCGCCACACCCGAAAGTTCTGGTTCCGAAGTCGAGTTTTCCCATGAGTACTGGATGAG

ACACGCATTGACTCTCGCAAAGAGGGCTCGAGATGAACGCGAGGTGCCCGTGGGGGCAGTACTCGTGCTCAACAATC
GCGTAATCGGCGAAGGTTGGAATAGGGCAATCGGACTCCACGACCCCACTGCACATGCGGAAATCATGGCCCTTCGA
CAGGGAGGGCTTGTGATGCAGAATTATCGACTTATCGATGCGACGCTGTACGTCACGTTTGAACCTTGCGTAATGTG
CGCGGGAGCTATGATTCACTCCCGCATTGGACGAGTTGTATTCGGTGTTCGCAACGCCAAGACGGGTGCCGCAGGTT
CACTGATGGACGTGCTGCATTACCCAGGCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGAAACC
GGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGT
GAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACAGCGATAAGCTGATCG
CCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTG
GCCAAAGTGGAAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAG
CAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGC
TGCCTAAGTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGA
AACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCC
CGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCG
AGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGATAAG
CCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTA
CTTTGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCA
TCACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGGGAGGTGAC

[0983] IBE032_ISLA:TadACP136ins 1052

[0984] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLAMNHRVEITEGILADECA
ALLCYFFRMPRQVFNAQKKAQSSTDGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAV
LVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAK
TGAAGSLMDVLHYPGNGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN
SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK
DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD

[0985] IBE035_ISLA:Δ792-872 TadAins

[0986] ATGGACAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCGAC
GAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGG
AGCCCTGCTGTTCGACAGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGAC
GGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGA
CTGGAAGAGTCCTTCCTGGTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGT
GGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCGACCTGC
GGCTGATCTATCTGGCCCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGAC
AACAGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCCATCAACGC
CAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGC
TGCCCGGCGAGAAGAAGAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGC
AACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAACCTGCTGGC
CCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCC
TGAGAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGAC
CTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAA
CGGCTACGCCGGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGA
TGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGC
AGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCCTGAA
GGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACA
GCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGC
GCTTCCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCA
CAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGC
CCGCCTTCCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAG
CAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAA
CGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACGAGG
ACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTAT
GCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAA
GCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACA
GAAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAG
GGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAA
GGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACC
AGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCTCC
GAAGTCGAGTTTTCCCATGAGTACTGGATGAGACACGCATTGACTCTCGCAAAGAGGGCTCGAGATGAACGCGAGGT
GCCCGTGGGGGCAGTACTCGTGCTCAACAATCGCGTAATCGGCGAAGGTTGGAATAGGGCAATCGGACTCCACGACC
CCACTGCACATGCGGAAATCATGGCCCTTCGACAGGGAGGGCTTGTGATGCAGAATTATCGACTTATCGATGCGACG
CTGTACGTCACGTTTGAACCTTGCGTAATGTGCGCGGGAGCTATGATTCACTCCCGCATTGGACGAGTTGTATTCGG
TGTTCGCAACGCCAAGACGGGTGCCGCAGGTTCACTGATGGACGTGCTGCATTACCCAGGCATGAACCACCGGGTAG
AAATCACAGAAGGCATATTGGCGGACGAATGTGCGGCGCTGTTGTGTTACTTTTTTCGCATGCCCAGGCAGGTCTTT
AACGCCCAGAAAAAAGCACAATCCTCTACTGACGAAGAGGTCGTGAAGAAGATGAAGAACTACTGGCGGCAGCTGCT

GAACGCCAAGCTGATTACCCAGAGAAAGTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATA
AGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGG
ATGAACACTAAGTACGACGAGAATGACAAGCTGATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTC
CGATTTCCGGAAGGATTTCCAGTTTTACAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGA
ACGCCGTCGTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTG
TACGACGTGCGGAAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAA
CATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACG
GCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTG
AATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACAGCGATAA
GCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGC
TGGTGGTGGCCAAAGTGGAAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATG
GAAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGAT
CATCAAGCTGCCTAAGTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGC
AGAAGGGAAACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAG
GGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCA
GATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACC
GGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCC
TTCAAGTACTTTGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCA
CCAGAGCATCACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGGGAGGTGAC

[0987] IBE035_ISLA:Δ792-872 TadAins

[0988] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRA
IGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPG
MNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGG
LSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHA
HDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRP
LIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPT
VAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLA
SAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLS
AYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[0989] IBE036_ISLA:Δ792-906 TadAins

[0990] ATGGACAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCGAC
GAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGG
AGCCCTGCTGTTCGACAGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGAC
GGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGA
CTGGAAGAGTCCTTCCTGGTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGT
GGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCGACCTGC
GGCTGATCTATCTGGCCCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGAC
AACAGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCCATCAACGC
CAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGC
TGCCCGGCGAGAAGAAGAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGC
AACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAACCTGCTGGC
CCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCC
TGAGAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGAC
CTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAA
CGGCTACGCCGGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGA
TGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGC
AGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCCTGAA
GGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACA
GCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGC
GCTTCCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCA
CAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGC
CCGCCTTCCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAG
CAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAA
CGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACGAGG
ACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTAT
GCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAA
GCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACA
GAAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAG
GGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAA
GGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACC
AGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCTCC
GAAGTCGAGTTTTCCCATGAGTACTGGATGAGACACGCATTGACTCTCGCAAAGAGGGCTCGAGATGAACGCGAGGT
GCCCGTGGGGGCAGTACTCGTGCTCAACAATCGCGTAATCGGCGAAGGTTGGAATAGGGCAATCGGACTCCACGACC
CCACTGCACATGCGGAAATCATGGCCCTTCGACAGGGAGGGCTTGTGATGCAGAATTATCGACTTATCGATGCGACG
CTGTACGTCACGTTTGAACCTTGCGTAATGTGCGCGGGAGCTATGATTCACTCCCGCATTGGACGAGTTGTATTCGG
TGTTCGCAACGCCAAGACGGGTGCCGCAGGTTCACTGATGGACGTGCTGCATTACCCAGGCATGAACCACCGGGTAG
AAATCACAGAAGGCATATTGGCGGACGAATGTGCGGCGCTGTTGTGTTACTTTTTTCGCATGCCCAGGCAGGTCTTT
AACGCCCAGAAAAAAGCACAATCCTCTACTGACGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCT

GGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAATG
ACAAGCTGATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTT
TACAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGGAACCGCCCTGAT
CAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGCCA
AGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCATGAACTTTTTCAAGACCGAG
ATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGA
TAAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGC
AGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGG
GACCCTAAGAAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGG
CAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATC
CCATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTG
TTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAACGAACTGGCCCTGCC
CTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCCGAGGATAATGAGCAGA
AACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTG
ATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGGC
CGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCG
ACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAG
ACACGGATCGACCTGTCTCAGCTGGGAGGTGAC

[0991] IBE036_ISLA:Δ792-906 TadAins

[0992] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRA
IGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPG
MNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT
KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDV
RKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIV
KKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERS
SFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSP
EDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKY
FDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[0993] IBE043_ISLA:TadA CP65ins 1246

[0994] ATGGACAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCGAC
GAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGG
AGCCCTGCTGTTCGACAGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGAC
GGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGA
CTGGAAGAGTCCTTCCTGGTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGT
GGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCGACCTGC
GGCTGATCTATCTGGCCCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGAC
AACAGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCCATCAACGC
CAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGC
TGCCCGGCGAGAAGAAGAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGC
AACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAACCTGCTGGC
CCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCC
TGAGAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGAC
CTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAA
CGGCTACGCCGGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGA
TGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGC
AGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCCTGAA
GGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACA
GCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGC
GCTTCCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCA
CAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGC
CCGCCTTCCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAG
CAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAA
CGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACGAGG
ACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTAT
GCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAA
GCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACA
GAAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAG
GGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAA
GGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACC
AGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGC
CAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGG
GCGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGA
GCTTTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTG
CCCTCCGAAGAGGTCGTGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAA
GTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGG
TGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGAC
AAGCTGATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTA
CAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGGAACCGCCCTGATCA

AAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAG
AGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCATGAACTTTTTCAAGACCGAGAT
TACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATA
AGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAG
ACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGA
CCCTAAGAAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCA
AGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCC
ATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTT
CGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAACGAACTGGCCCTGCCCT
CCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCACTGCACATGCGGAAATCATGGCC
CTTCGACAGGGAGGGCTTGTGATGCAGAATTATCGACTTATCGATGCGACGCTGTACGTCACGTTTGAACCTTGCGT
AATGTGCGCGGGAGCTATGATTCACTCCCGCATTGGACGAGTTGTATTCGGTGTTCGCAACGCCAAGACGGGTGCCG
CAGGTTCACTGATGGACGTGCTGCATTACCCAGGCATGAACCACCGGGTAGAAATCACAGAAGGCATATTGGCGGAC
GAATGTGCGGCGCTGTTGTGTTACTTTTTTCGCATGCCCAGGCAGGTCTTTAACGCCCAGAAAAAAGCACAATCCTC
TACTGACGGTTCTAGCGGCAGCGAGACTCCCGGGACCTCAGAGTCCGCCACACCCGAAAGTTCTGGTTCCGAAGTCG
AGTTTTCCCATGAGTACTGGATGAGACACGCATTGACTCTCGCAAAGAGGGCTCGAGATGAACGCGAGGTGCCCGTG
GGGGCAGTACTCGTGCTCAACAATCGCGTAATCGGCGAAGGTTGGAATAGGGCAATCGGACTCCACGACCCCTCCCC
CGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCG
AGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGATAAG
CCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTA
CTTTGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCA
TCACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGGGAGGTGAC

[0995] IBE043_ISLA:TadA CP65ins 1246

[0996] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN

ELALPSKYVNFLYLASHYEKLKGTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRN
AKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGSETPGTSESATPES
SGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD

[0997] IBE044_ISLA:TadAins C末端截短2 791TGGACAAGAAGTACAGCATCGGCCTGGCCATCGG
CACCAACTCTGTGGGCTGGGCCGTGATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACA
CCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGACAGCGGCGAAACAGCCGAGGCCACCCGG
CTGAAGAGAACCGCCAGAAGAAGATACACCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGA
GATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGAAGAGGATAAGAAGCACGAGC
GGCACCCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAG
AAACTGGTGGACAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCCACATGATCAAGTTCCGGGG
CCACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCT
ACAACCAGCTGTTCGAGGAAAACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGC
AAGAGCAGACGGCTGGAAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTCGGAAACCTGATTGC
CCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGG
ACACCTACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAG
AACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTC
TATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAGA
AGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCGGCTACATTGACGGCGGAGCCAGCCAGGAAGAG
TTCTACAAGTTCATCAAGCCCATCCTGGAAAAGATGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGA
CCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTC
TGCGGCGGCAGGAAGATTTTTACCCATTCCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATC
CCCTACTACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCAC
CCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCTTCCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATA
AGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGCTGACC
AAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGACCT
GCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACT
CCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAG
GACAAGGACTTCCTGGACAATGAGGAAAACGAGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGA
CAGAGAGATGATCGAGGAACGGCTGAAAACCTATGCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGC
GGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATC
CTGGATTTCCTGAAGTCCGACGGCTTCGCCAACAGAAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAA
AGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCC
CCGCCATTAAGAAGGGCATCCTGCAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCC
GAGAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAA
GCGGATCGAAGAGGGCATCAAAGAGCTGGGTTCTAGCGGCAGCGAGACTCCCGGGACCTCAGAGTCCGCCACACCCG
AAAGTTCTGGTTCCGAAGTCGAGTTTTCCCATGAGTACTGGATGAGACACGCATTGACTCTCGCAAAGAGGGCTCGA
GATGAACGCGAGGTGCCCGTGGGGGCAGTACTCGTGCTCAACAATCGCGTAATCGGCGAAGGTTGGAATAGGGCAAT

CGGACTCCACGACCCCACTGCACATGCGGAAATCATGGCCCTTCGACAGGGAGGGCTTGTGATGCAGAATTATCGAC
TTATCGATGCGACGCTGTACGTCACGTTTGAACCTTGCGTAATGTGCGCGGGAGCTATGATTCACTCCCGCATTGGA
CGAGTTGTATTCGGTGTTCGCAACGCCAAGACGGGTGCCGCAGGTTCACTGATGGACGTGCTGCATTACCCAGGCAT
GAACCACCGGGTAGAAATCACAGAAGGCATATTGGCGGACGAATGTGCGGCGCTGTTGTGTTACTTTTTTCGCATGC
CCAGGCAGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTAC
TACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTACGATGTGGACCA
TATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACCGGGGCA
AGAGCGACAACGTGCCCTCCGAAGAGGTCGTGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTG
ATTACCCAGAGAAAGTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCAT
CAAGAGACAGCTGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGT
ACGACGAGAATGACAAGCTGATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAG
GATTTCCAGTTTTACAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGG
AACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGA
AGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCATGAACTTT
TTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGGGA
GATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAAA
AGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGA
AAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAA
AGTGGAAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCT
TCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCT
AAGTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAACGA
ACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCCGAGG
ATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTC
TCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGATAAGCCCAT
CAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTG
ACACCACCATCGACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACC
GGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGGGAGGTGAC

[0998] IBE044_ISLA:TadAins C末端截短2 791

[0999] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVP

VGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGV
RNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQGSQILKEHPVENTQLQNEKLYLYYLQNGR
DMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF
DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYK
VREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEIT
LANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDP
KKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE
LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVIL
ADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETR
IDLSQLGGD

[1000] pHRB-043_GGS-rAP:rAPOBEC1-XTEN-ins-位点

[1001] 1_Y1016-D10A-UGIx2

[1002] ATGGACAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCGAC
GAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGG
AGCCCTGCTGTTCGACAGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGAC
GGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGA
CTGGAAGAGTCCTTCCTGGTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGT
GGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCGACCTGC
GGCTGATCTATCTGGCCCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGAC
AACAGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCCATCAACGC
CAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGC
TGCCCGGCGAGAAGAAGAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGC
AACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAACCTGCTGGC
CCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCC
TGAGAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGAC
CTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAA
CGGCTACGCCGGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGA
TGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGC
AGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCCTGAA
GGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACA
GCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGC
GCTTCCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCA
CAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGC
CCGCCTTCCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAG
CAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAA
CGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACGAGG
ACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTAT
GCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAA
GCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACA

GAAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAG
GGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAA
GGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACC
AGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGC
CAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGG
GCGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGA
GCTTTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTG
CCCTCCGAAGAGGTCGTGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAA
GTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGG
TGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGAC
AAGCTGATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTA
CAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGGAACCGCCCTGATCA
AAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGGAGGCTCTGGAGGAAGCAGCTCT
GAGACAGGACCTGTGGCCGTGGATCCCACACTGCGGAGAAGAATTGAGCCCCACGAGTTCGAGGTGTTCTTCGACCC
CAGAGAGCTGCGGAAAGAGACATGCCTGCTGTACGAGATCAACTGGGGCGGCAGACACTCTATCTGGCGGCACACAA
GCCAGAACACCAACAAGCACGTGGAAGTGAACTTTATCGAGAAGTTTACGACCGAGCGGTACTTCTGCCCCAACACC
AGATGCAGCATCACCTGGTTTCTGAGCTGGTCCCCTTGCGGCGAGTGCAGCAGAGCCATCACCGAGTTTCTGTCCAG
ATATCCCCACGTGACCCTGTTCATCTATATCGCCCGGCTGTACCACCACGCCGATCCTAGAAATAGACAGGGCCTGC
GCGACCTGATCAGCAGCGGAGTGACAATCCAGATCATGACCGAGCAAGAGAGCGGCTACTGCTGGCGGAACTTCGTG
AACTACAGCCCCAGCAACGAAGCCCACTGGCCTAGATATCCTCACCTGTGGGTCCGACTGTACGTGCTGGAACTGTA
CTGCATCATCCTGGGCCTGCCTCCATGCCTGAACATCCTGAGAAGAAAGCAGCCTCAGCTGACCTTCTTCACAATCG
CCCTGCAGAGCTGCCACTACCAGAGACTGCCTCCACACATCCTGTGGGCCACCGGACTTAAGGGCTCTTCTGGATCT
GAAACACCTGGCACAAGTGAGAGCGCCACCCCTGAGAGCTCTGGCGACGTGCGGAAGATGATCGCCAAGAGCGAGCA
GGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCATGAACTTTTTCAAGACCGAGATTACCCTGG
CCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGG
GATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGG
CTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGA
AGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGTCCAAG
AAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCCATCGACTT
TCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTTCGAGCTGG
AAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAACGAACTGGCCCTGCCCTCCAAATAT
GTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTT
TGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCG
ACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATC
ATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGAG
GTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGACACGGATCG
ACCTGTCTCAGCTGGGAGGTGACTCTGGTGGAAGCGGAGGATCTGGCGGCAGCACCAATCTGAGCGACATCATCGAG
AAAGAGACAGGCAAGCAGCTGGTCATCCAAGAGTCCATCCTGATGCTGCCTGAAGAGGTGGAAGAAGTGATCGGCAA
CAAGCCCGAGTCCGACATCCTGGTGCACACCGCCTACGATGAGAGCACCGACGAGAACGTGATGCTGCTGACCTCTG

ACGCCCCTGAGTACAAGCCTTGGGCTCTCGTGATCCAGGACAGCAACGGCGAGAACAAGATCAAGATGCTGAGCGGC
GGCTCTGGTGGCTCTGGCGGATCTACAAACCTGTCCGATATTATTGAGAAAGAAACCGGGAAACAGCTCGTGATTCA
AGAGTCTATTCTCATGCTCCCGGAAGAAGTCGAGGAAGTCATTGGAAACAAGCCTGAGAGCGATATTCTGGTCCATA
CAGCCTACGACGAGTCTACCGATGAGAATGTCATGCTCCTCACCAGCGACGCTCCCGAGTATAAGCCATGGGCACTT
GTCATTCAGGACTCCAATGGGGAAAACAAAATCAAAATGCTCCCAAAGAAAAAACGCAAGGTGGAGGGAGCTGATAA
GCGCACCGCCGATGGTTCCGAGTTCGAAAGCCCCAAGAAGAAGAGGAAAGTCTAACCGGTCATCATCACCATCACCA
TTGAGTTTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGC
CTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGT
AGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGC
TGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCGATACCGTCGACCTCTAGCTAGA
GCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCC
GGAAGCATAAAGTGTAAAGCCTAGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGC
TTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTG
GGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCA
AAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGC
CAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGAC
GCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGC
TCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAG
CTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGC
CCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCA
GCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGG
CTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTT
GATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGA
TCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGT
CATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATAT
ATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCA
TCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAAT
GATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAA
GTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTT
AATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAG
CTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTC
CGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTC
ATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACC
GAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAA
AACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCC
AACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAA
GGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTT
ATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGA
AAAGTGCCACCTGACGTCGACGGATCGGGAGATCGATCTCCCGATCCCCTAGGGTCGACTCTCAGTACAATCTGCTC

TGATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAAAATTT
AAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGC
GATGTACGGGCCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGT
TCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCC
GCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAG
TATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGA
CGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTAT
TAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGA
TTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCG
TAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTGGTTTAG
TGAACCGTCAGATCCGCTAGAGATCCGCGGCCGCTAATACGACTCACTATAGGGAGAGCCGCCACC

[1003] pHRB-043_GGS-rAP:rAPOBEC1-XTEN-ins-位点

[1004] 1_Y1016-D10A-UGIx2

[1005] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTA
RRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVD
STDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSR
RLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNL
SDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRI
PYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNEL
TKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG
KTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMG
RHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL
DINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAE
RGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINN
YHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYGGSGGSSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKE
TCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVT
LFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIIL
GLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKGSSGSETPGTSESATPESSGDVRKMIAKSEQEIG
KATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFS
KESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFL
EAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLF
VEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK
RYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVI
GNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSGGSGGSTNLSDIIEKETGKQL
VIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLPKKKRKVE
GADKRTADGSEFESPKKKRKV*

[1006] pHRB-044_GGS-rAP:rAPOBEC1-XTEN ins-位点

[1007] 2_A1023-D10A-UGIx2

[1008] ATGGACAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCGAC
GAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGG
AGCCCTGCTGTTCGACAGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGAC
GGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGA
CTGGAAGAGTCCTTCCTGGTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGT
GGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCGACCTGC
GGCTGATCTATCTGGCCCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGAC
AACAGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCCATCAACGC
CAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGC
TGCCCGGCGAGAAGAAGAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGC
AACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAACCTGCTGGC
CCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCC
TGAGAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGAC
CTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAA
CGGCTACGCCGGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGA
TGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGC
AGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCCTGAA
GGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACA
GCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGC
GCTTCCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCA
CAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGC
CCGCCTTCCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAG
CAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAA
CGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACGAGG
ACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTAT
GCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAA
GCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACA
GAAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAG
GGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAA
GGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACC
AGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGC
CAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGG
GCGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGA
GCTTTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTG
CCCTCCGAAGAGGTCGTGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAA
GTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGG
TGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGAC
AAGCTGATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTA

CAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGGAACCGCCCTGATCA
AAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGCCGGA
GGCTCTGGAGGAAGCAGCTCTGAGACAGGACCTGTGGCCGTGGATCCCACACTGCGGAGAAGAATTGAGCCCCACGA
GTTCGAGGTGTTCTTCGACCCCAGAGAGCTGCGGAAAGAGACATGCCTGCTGTACGAGATCAACTGGGGCGGCAGAC
ACTCTATCTGGCGGCACACAAGCCAGAACACCAACAAGCACGTGGAAGTGAACTTTATCGAGAAGTTTACGACCGAG
CGGTACTTCTGCCCCAACACCAGATGCAGCATCACCTGGTTTCTGAGCTGGTCCCCTTGCGGCGAGTGCAGCAGAGC
CATCACCGAGTTTCTGTCCAGATATCCCCACGTGACCCTGTTCATCTATATCGCCCGGCTGTACCACCACGCCGATC
CTAGAAATAGACAGGGCCTGCGCGACCTGATCAGCAGCGGAGTGACAATCCAGATCATGACCGAGCAAGAGAGCGGC
TACTGCTGGCGGAACTTCGTGAACTACAGCCCCAGCAACGAAGCCCACTGGCCTAGATATCCTCACCTGTGGGTCCG
ACTGTACGTGCTGGAACTGTACTGCATCATCCTGGGCCTGCCTCCATGCCTGAACATCCTGAGAAGAAAGCAGCCTC
AGCTGACCTTCTTCACAATCGCCCTGCAGAGCTGCCACTACCAGAGACTGCCTCCACACATCCTGTGGGCCACCGGA
CTTAAGGGCTCTTCTGGATCTGAAACACCTGGCACAAGTGAGAGCGCCACCCCTGAGAGCTCTGGCAAGAGCGAGCA
GGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCATGAACTTTTTCAAGACCGAGATTACCCTGG
CCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGG
GATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGG
CTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGA
AGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGTCCAAG
AAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCCATCGACTT
TCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTTCGAGCTGG
AAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAACGAACTGGCCCTGCCCTCCAAATAT
GTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTT
TGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCG
ACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATC
ATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGAG
GTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGACACGGATCG
ACCTGTCTCAGCTGGGAGGTGACTCTGGTGGAAGCGGAGGATCTGGCGGCAGCACCAATCTGAGCGACATCATCGAG
AAAGAGACAGGCAAGCAGCTGGTCATCCAAGAGTCCATCCTGATGCTGCCTGAAGAGGTGGAAGAAGTGATCGGCAA
CAAGCCCGAGTCCGACATCCTGGTGCACACCGCCTACGATGAGAGCACCGACGAGAACGTGATGCTGCTGACCTCTG
ACGCCCCTGAGTACAAGCCTTGGGCTCTCGTGATCCAGGACAGCAACGGCGAGAACAAGATCAAGATGCTGAGCGGC
GGCTCTGGTGGCTCTGGCGGATCTACAAACCTGTCCGATATTATTGAGAAAGAAACCGGGAAACAGCTCGTGATTCA
AGAGTCTATTCTCATGCTCCCGGAAGAAGTCGAGGAAGTCATTGGAAACAAGCCTGAGAGCGATATTCTGGTCCATA
CAGCCTACGACGAGTCTACCGATGAGAATGTCATGCTCCTCACCAGCGACGCTCCCGAGTATAAGCCATGGGCACTT
GTCATTCAGGACTCCAATGGGGAAAACAAAATCAAAATGCTCCCAAAGAAAAAACGCAAGGTGGAGGGAGCTGATAA
GCGCACCGCCGATGGTTCCGAGTTCGAAAGCCCCAAGAAGAAGAGGAAAGTCTAACCGGTCATCATCACCATCACCA
TTGAGTTTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGC
CTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGT
AGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGC
TGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCGATACCGTCGACCTCTAGCTAGA
GCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCC

GGAAGCATAAAGTGTAAAGCCTAGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGC
TTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTG
GGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCA
AAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGC
CAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGAC
GCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGC
TCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAG
CTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGC
CCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCA
GCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGG
CTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTT
GATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGA
TCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGT
CATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATAT
ATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCA
TCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAAT
GATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAA
GTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTT
AATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAG
CTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTC
CGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTC
ATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACC
GAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAA
AACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCC
AACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAA
GGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTT
ATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGA
AAAGTGCCACCTGACGTCGACGGATCGGGAGATCGATCTCCCGATCCCCTAGGGTCGACTCTCAGTACAATCTGCTC
TGATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAAAATTT
AAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGC
GATGTACGGGCCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGT
TCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCC
GCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAG
TATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGA
CGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTAT
TAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGA
TTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCG
TAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTGGTTTAG
TGAACCGTCAGATCCGCTAGAGATCCGCGGCCGCTAATACGACTCACTATAGGGAGAGCCGCCACC

[1009] pHRB-044_GGS-rAP:rAPOBEC1-XTEN ins-位点

[1010] 2_A1023-D10A-UGIx2

[1011] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTA
RRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVD
STDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSR
RLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNL
SDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRI
PYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNEL
TKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG
KTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMG
RHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL
DINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAE
RGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINN
YHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAGGSGGSSSETGPVAVDPTLRRRIEPHEFEVFFD
PRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFL
SRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVL
ELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKGSSGSETPGTSESATPESSGKSEQEIG
KATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFS
KESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFL
EAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLF
VEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK
RYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVI
GNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSGGSGGSTNLSDIIEKETGKQL
VIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLPKKKRKVE
GADKRTADGSEFESPKKKRKV*

[1012] pHRB-045_GGS-rAP:rAPOBEC1-XTEN ins-位点

[1013] 3_E1029-D10A-UGIx2

[1014] ATGGACAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCGAC
GAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGG
AGCCCTGCTGTTCGACAGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGAC
GGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGA
CTGGAAGAGTCCTTCCTGGTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGT
GGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCGACCTGC
GGCTGATCTATCTGGCCCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGAC
AACAGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCCATCAACGC
CAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGC
TGCCCGGCGAGAAGAAGAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGC

AACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAACCTGCTGGC
CCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCC
TGAGAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGAC
CTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAA
CGGCTACGCCGGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGA
TGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGC
AGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCCTGAA
GGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACA
GCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGC
GCTTCCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCA
CAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGC
CCGCCTTCCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAG
CAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAA
CGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACGAGG
ACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTAT
GCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAA
GCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACA
GAAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAG
GGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAA
GGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACC
AGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGC
CAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGG
GCGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGA
GCTTTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTG
CCCTCCGAAGAGGTCGTGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAA
GTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGG
TGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGAC
AAGCTGATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTA
CAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGGAACCGCCCTGATCA
AAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAG
AGCGAGCAGGAAATCGGAGGCTCTGGAGGAAGCAGCTCTGAGACAGGACCTGTGGCCGTGGATCCCACACTGCGGAG
AAGAATTGAGCCCCACGAGTTCGAGGTGTTCTTCGACCCCAGAGAGCTGCGGAAAGAGACATGCCTGCTGTACGAGA
TCAACTGGGGCGGCAGACACTCTATCTGGCGGCACACAAGCCAGAACACCAACAAGCACGTGGAAGTGAACTTTATC
GAGAAGTTTACGACCGAGCGGTACTTCTGCCCCAACACCAGATGCAGCATCACCTGGTTTCTGAGCTGGTCCCCTTG
CGGCGAGTGCAGCAGAGCCATCACCGAGTTTCTGTCCAGATATCCCCACGTGACCCTGTTCATCTATATCGCCCGGC
TGTACCACCACGCCGATCCTAGAAATAGACAGGGCCTGCGCGACCTGATCAGCAGCGGAGTGACAATCCAGATCATG
ACCGAGCAAGAGAGCGGCTACTGCTGGCGGAACTTCGTGAACTACAGCCCCAGCAACGAAGCCCACTGGCCTAGATA
TCCTCACCTGTGGGTCCGACTGTACGTGCTGGAACTGTACTGCATCATCCTGGGCCTGCCTCCATGCCTGAACATCC
TGAGAAGAAAGCAGCCTCAGCTGACCTTCTTCACAATCGCCCTGCAGAGCTGCCACTACCAGAGACTGCCTCCACAC

ATCCTGTGGGCCACCGGACTTAAGGGCTCTTCTGGATCTGAAACACCTGGCACAAGTGAGAGCGCCACCCCTGAGAG
CTCTGGCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCATGAACTTTTTCAAGACCGAGATTACCCTGG
CCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGG
GATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGG
CTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGA
AGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGTCCAAG
AAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCCATCGACTT
TCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTTCGAGCTGG
AAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAACGAACTGGCCCTGCCCTCCAAATAT
GTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTT
TGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCG
ACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATC
ATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGAG
GTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGACACGGATCG
ACCTGTCTCAGCTGGGAGGTGACTCTGGTGGAAGCGGAGGATCTGGCGGCAGCACCAATCTGAGCGACATCATCGAG
AAAGAGACAGGCAAGCAGCTGGTCATCCAAGAGTCCATCCTGATGCTGCCTGAAGAGGTGGAAGAAGTGATCGGCAA
CAAGCCCGAGTCCGACATCCTGGTGCACACCGCCTACGATGAGAGCACCGACGAGAACGTGATGCTGCTGACCTCTG
ACGCCCCTGAGTACAAGCCTTGGGCTCTCGTGATCCAGGACAGCAACGGCGAGAACAAGATCAAGATGCTGAGCGGC
GGCTCTGGTGGCTCTGGCGGATCTACAAACCTGTCCGATATTATTGAGAAAGAAACCGGGAAACAGCTCGTGATTCA
AGAGTCTATTCTCATGCTCCCGGAAGAAGTCGAGGAAGTCATTGGAAACAAGCCTGAGAGCGATATTCTGGTCCATA
CAGCCTACGACGAGTCTACCGATGAGAATGTCATGCTCCTCACCAGCGACGCTCCCGAGTATAAGCCATGGGCACTT
GTCATTCAGGACTCCAATGGGGAAAACAAAATCAAAATGCTCCCAAAGAAAAAACGCAAGGTGGAGGGAGCTGATAA
GCGCACCGCCGATGGTTCCGAGTTCGAAAGCCCCAAGAAGAAGAGGAAAGTCTAACCGGTCATCATCACCATCACCA
TTGAGTTTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGC
CTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGT
AGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGC
TGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCGATACCGTCGACCTCTAGCTAGA
GCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCC
GGAAGCATAAAGTGTAAAGCCTAGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGC
TTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTG
GGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCA
AAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGC
CAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGAC
GCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGC
TCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAG
CTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGC
CCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCA
GCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGG
CTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTT

GATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGA
TCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGT
CATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATAT
ATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCA
TCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAAT
GATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAA
GTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTT
AATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAG
CTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTC
CGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTC
ATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACC
GAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAA
AACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCC
AACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAA
GGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTT
ATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGA
AAAGTGCCACCTGACGTCGACGGATCGGGAGATCGATCTCCCGATCCCCTAGGGTCGACTCTCAGTACAATCTGCTC
TGATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAAAATTT
AAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGC
GATGTACGGGCCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGT
TCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCC
GCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAG
TATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGA
CGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTAT
TAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGA
TTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCG
TAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTGGTTTAG
TGAACCGTCAGATCCGCTAGAGATCCGCGGCCGCTAATACGACTCACTATAGGGAGAGCCGCCACC

[1015] pHRB-045_GGS-rAP:rAPOBEC1-XTEN ins-位点

[1016] 3_E1029-D10A-UGIx2

[1017] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTA
RRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVD
STDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSR
RLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNL
SDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRI
PYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNEL
TKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG

KTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMG
RHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL
DINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAE
RGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINN
YHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGGSGGSSSETGPVAVDPTLRRRIEPHE
FEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSR
AITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLW
VRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKGSSGSETPGTSESATPESSGG
KATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFS
KESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFL
EAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLF
VEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK
RYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVI
GNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSGGSGGSTNLSDIIEKETGKQL
VIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLPKKKRKVE
GADKRTADGSEFESPKKKRKV*

[1018] pHRB-046_GGS-rAP:rAPOBEC1-XTEN ins-位点

[1019] 4_N1040-D10A-UGIx2

[1020] ATGGACAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCGAC
GAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGG
AGCCCTGCTGTTCGACAGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGAC
GGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGA
CTGGAAGAGTCCTTCCTGGTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGT
GGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCGACCTGC
GGCTGATCTATCTGGCCCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGAC
AACAGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCCATCAACGC
CAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGC
TGCCCGGCGAGAAGAAGAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGC
AACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAACCTGCTGGC
CCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCC
TGAGAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGAC
CTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAA
CGGCTACGCCGGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGA
TGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGC
AGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCCTGAA
GGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACA
GCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGC
GCTTCCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCA
CAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGC

CCGCCTTCCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAG
CAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAA
CGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACGAGG
ACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTAT
GCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAA
GCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACA
GAAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAG
GGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAA
GGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACC
AGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGC
CAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGG
GCGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGA
GCTTTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTG
CCCTCCGAAGAGGTCGTGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAA
GTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGG
TGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGAC
AAGCTGATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTA
CAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGGAACCGCCCTGATCA
AAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAG
AGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACGGAGGCTCTGGAGGAAGCAGCTCTGA
GACAGGACCTGTGGCCGTGGATCCCACACTGCGGAGAAGAATTGAGCCCCACGAGTTCGAGGTGTTCTTCGACCCCA
GAGAGCTGCGGAAAGAGACATGCCTGCTGTACGAGATCAACTGGGGCGGCAGACACTCTATCTGGCGGCACACAAGC
CAGAACACCAACAAGCACGTGGAAGTGAACTTTATCGAGAAGTTTACGACCGAGCGGTACTTCTGCCCCAACACCAG
ATGCAGCATCACCTGGTTTCTGAGCTGGTCCCCTTGCGGCGAGTGCAGCAGAGCCATCACCGAGTTTCTGTCCAGAT
ATCCCCACGTGACCCTGTTCATCTATATCGCCCGGCTGTACCACCACGCCGATCCTAGAAATAGACAGGGCCTGCGC
GACCTGATCAGCAGCGGAGTGACAATCCAGATCATGACCGAGCAAGAGAGCGGCTACTGCTGGCGGAACTTCGTGAA
CTACAGCCCCAGCAACGAAGCCCACTGGCCTAGATATCCTCACCTGTGGGTCCGACTGTACGTGCTGGAACTGTACT
GCATCATCCTGGGCCTGCCTCCATGCCTGAACATCCTGAGAAGAAAGCAGCCTCAGCTGACCTTCTTCACAATCGCC
CTGCAGAGCTGCCACTACCAGAGACTGCCTCCACACATCCTGTGGGCCACCGGACTTAAGGGCTCTTCTGGATCTGA
AACACCTGGCACAAGTGAGAGCGCCACCCCTGAGAGCTCTGGCATCATGAACTTTTTCAAGACCGAGATTACCCTGG
CCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGG
GATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGG
CTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGA
AGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGTCCAAG
AAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCCATCGACTT
TCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTTCGAGCTGG
AAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAACGAACTGGCCCTGCCCTCCAAATAT
GTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTT
TGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCG

ACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATC
ATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGAG
GTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGACACGGATCG
ACCTGTCTCAGCTGGGAGGTGACTCTGGTGGAAGCGGAGGATCTGGCGGCAGCACCAATCTGAGCGACATCATCGAG
AAAGAGACAGGCAAGCAGCTGGTCATCCAAGAGTCCATCCTGATGCTGCCTGAAGAGGTGGAAGAAGTGATCGGCAA
CAAGCCCGAGTCCGACATCCTGGTGCACACCGCCTACGATGAGAGCACCGACGAGAACGTGATGCTGCTGACCTCTG
ACGCCCCTGAGTACAAGCCTTGGGCTCTCGTGATCCAGGACAGCAACGGCGAGAACAAGATCAAGATGCTGAGCGGC
GGCTCTGGTGGCTCTGGCGGATCTACAAACCTGTCCGATATTATTGAGAAAGAAACCGGGAAACAGCTCGTGATTCA
AGAGTCTATTCTCATGCTCCCGGAAGAAGTCGAGGAAGTCATTGGAAACAAGCCTGAGAGCGATATTCTGGTCCATA
CAGCCTACGACGAGTCTACCGATGAGAATGTCATGCTCCTCACCAGCGACGCTCCCGAGTATAAGCCATGGGCACTT
GTCATTCAGGACTCCAATGGGGAAAACAAAATCAAAATGCTCCCAAAGAAAAAACGCAAGGTGGAGGGAGCTGATAA
GCGCACCGCCGATGGTTCCGAGTTCGAAAGCCCCAAGAAGAAGAGGAAAGTCTAACCGGTCATCATCACCATCACCA
TTGAGTTTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGC
CTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGT
AGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGC
TGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCGATACCGTCGACCTCTAGCTAGA
GCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCC
GGAAGCATAAAGTGTAAAGCCTAGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGC
TTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTG
GGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCA
AAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGC
CAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGAC
GCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGC
TCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAG
CTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGC
CCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCA
GCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGG
CTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTT
GATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGA
TCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGT
CATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATAT
ATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCA
TCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAAT
GATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAA
GTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTT
AATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAG
CTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTC
CGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTC
ATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACC

GAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAA
AACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCC
AACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAA
GGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTT
ATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGA
AAAGTGCCACCTGACGTCGACGGATCGGGAGATCGATCTCCCGATCCCCTAGGGTCGACTCTCAGTACAATCTGCTC
TGATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAAAATTT
AAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGC
GATGTACGGGCCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGT
TCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCC
GCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAG
TATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGA
CGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTAT
TAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGA
TTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCG
TAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTGGTTTAG
TGAACCGTCAGATCCGCTAGAGATCCGCGGCCGCTAATACGACTCACTATAGGGAGAGCCGCCACC

[1021] pHRB-046_GGS-rAP:rAPOBEC1-XTEN ins-位点

[1022] 4_N1040-D10A-UGIx2

[1023] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTA
RRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVD
STDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSR
RLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNL
SDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRI
PYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNEL
TKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG
KTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMG
RHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL
DINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAE
RGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINN
YHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNGGSGGSSSETGPVAV
DPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKFTTERYFCPNTRCSITW
FLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPS
NEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKGSSGSETPG
TSESATPESSGIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFS
KESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFL
EAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLF

VEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK
RYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVI
GNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSGGSGGSTNLSDIIEKETGKQL
VIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLPKKKRKVE
GADKRTADGSEFESPKKKRKV*

[1024] pHRB-047_GGS-rAP:rAPOBEC1-XTEN ins-位点

[1025] 5-T1069-D10A-UGIx2

[1026] ATGGACAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCGAC
GAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGG
AGCCCTGCTGTTCGACAGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGAC
GGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGA
CTGGAAGAGTCCTTCCTGGTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGT
GGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCGACCTGC
GGCTGATCTATCTGGCCCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGAC
AACAGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCCATCAACGC
CAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGC
TGCCCGGCGAGAAGAAGAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGC
AACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAACCTGCTGGC
CCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCC
TGAGAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGAC
CTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAA
CGGCTACGCCGGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGA
TGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGC
AGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCCTGAA
GGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACA
GCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGC
GCTTCCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCA
CAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGC
CCGCCTTCCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAG
CAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAA
CGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACGAGG
ACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTAT
GCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAA
GCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACA
GAAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAG
GGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAA
GGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACC
AGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGC
CAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGG

GCGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGA
GCTTTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTG
CCCTCCGAAGAGGTCGTGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAA
GTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGG
TGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGAC
AAGCTGATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTA
CAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGGAACCGCCCTGATCA
AAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAG
AGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCATGAACTTTTTCAAGACCGAGAT
TACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGAGGCTCTGGAGGAAGCA
GCTCTGAGACAGGACCTGTGGCCGTGGATCCCACACTGCGGAGAAGAATTGAGCCCCACGAGTTCGAGGTGTTCTTC
GACCCCAGAGAGCTGCGGAAAGAGACATGCCTGCTGTACGAGATCAACTGGGGCGGCAGACACTCTATCTGGCGGCA
CACAAGCCAGAACACCAACAAGCACGTGGAAGTGAACTTTATCGAGAAGTTTACGACCGAGCGGTACTTCTGCCCCA
ACACCAGATGCAGCATCACCTGGTTTCTGAGCTGGTCCCCTTGCGGCGAGTGCAGCAGAGCCATCACCGAGTTTCTG
TCCAGATATCCCCACGTGACCCTGTTCATCTATATCGCCCGGCTGTACCACCACGCCGATCCTAGAAATAGACAGGG
CCTGCGCGACCTGATCAGCAGCGGAGTGACAATCCAGATCATGACCGAGCAAGAGAGCGGCTACTGCTGGCGGAACT
TCGTGAACTACAGCCCCAGCAACGAAGCCCACTGGCCTAGATATCCTCACCTGTGGGTCCGACTGTACGTGCTGGAA
CTGTACTGCATCATCCTGGGCCTGCCTCCATGCCTGAACATCCTGAGAAGAAAGCAGCCTCAGCTGACCTTCTTCAC
AATCGCCCTGCAGAGCTGCCACTACCAGAGACTGCCTCCACACATCCTGTGGGCCACCGGACTTAAGGGCTCTTCTG
GATCTGAAACACCTGGCACAAGTGAGAGCGCCACCCCTGAGAGCTCTGGCGGGGAGATCGTGTGGGATAAGGGCCGG
GATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGG
CTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGA
AGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGTCCAAG
AAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCCATCGACTT
TCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTTCGAGCTGG
AAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAACGAACTGGCCCTGCCCTCCAAATAT
GTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTT
TGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCG
ACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATC
ATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGAG
GTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGACACGGATCG
ACCTGTCTCAGCTGGGAGGTGACTCTGGTGGAAGCGGAGGATCTGGCGGCAGCACCAATCTGAGCGACATCATCGAG
AAAGAGACAGGCAAGCAGCTGGTCATCCAAGAGTCCATCCTGATGCTGCCTGAAGAGGTGGAAGAAGTGATCGGCAA
CAAGCCCGAGTCCGACATCCTGGTGCACACCGCCTACGATGAGAGCACCGACGAGAACGTGATGCTGCTGACCTCTG
ACGCCCCTGAGTACAAGCCTTGGGCTCTCGTGATCCAGGACAGCAACGGCGAGAACAAGATCAAGATGCTGAGCGGC
GGCTCTGGTGGCTCTGGCGGATCTACAAACCTGTCCGATATTATTGAGAAAGAAACCGGGAAACAGCTCGTGATTCA
AGAGTCTATTCTCATGCTCCCGGAAGAAGTCGAGGAAGTCATTGGAAACAAGCCTGAGAGCGATATTCTGGTCCATA
CAGCCTACGACGAGTCTACCGATGAGAATGTCATGCTCCTCACCAGCGACGCTCCCGAGTATAAGCCATGGGCACTT
GTCATTCAGGACTCCAATGGGGAAAACAAAATCAAAATGCTCCCAAAGAAAAAACGCAAGGTGGAGGGAGCTGATAA

GCGCACCGCCGATGGTTCCGAGTTCGAAAGCCCCAAGAAGAAGAGGAAAGTCTAACCGGTCATCATCACCATCACCA
TTGAGTTTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGC
CTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGT
AGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGC
TGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCGATACCGTCGACCTCTAGCTAGA
GCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCC
GGAAGCATAAAGTGTAAAGCCTAGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGC
TTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTG
GGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCA
AAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGC
CAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGAC
GCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGC
TCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAG
CTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGC
CCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCA
GCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGG
CTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTT
GATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGA
TCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGT
CATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATAT
ATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCA
TCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAAT
GATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAA
GTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTT
AATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAG
CTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTC
CGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTC
ATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACC
GAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAA
AACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCC
AACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAA
GGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTT
ATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGA
AAAGTGCCACCTGACGTCGACGGATCGGGAGATCGATCTCCCGATCCCCTAGGGTCGACTCTCAGTACAATCTGCTC
TGATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAAAATTT
AAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGC
GATGTACGGGCCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGT
TCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCC
GCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAG

TATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGA
CGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTAT
TAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGA
TTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCG
TAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTGGTTTAG
TGAACCGTCAGATCCGCTAGAGATCCGCGGCCGCTAATACGACTCACTATAGGGAGAGCCGCCACC

[1027] pHRB-047_GGS-rAP:rAPOBEC1-XTEN ins-位点

[1028] 5-T1069-D10A-UGIx2

[1029] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTA
RRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVD
STDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSR
RLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNL
SDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRI
PYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNEL
TKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG
KTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMG
RHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL
DINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAE
RGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINN
YHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGE
IRKRPLIETNGETGGSGGSSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNT
NKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDL
ISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIAL
QSCHYQRLPPHILWATGLKGSSGSETPGTSESATPESSGGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFS
KESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFL
EAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLF
VEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK
RYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVI
GNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSGGSGGSTNLSDIIEKETGKQL
VIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLPKKKRKVE
GADKRTADGSEFESPKKKRKV*

[1030] pHRB-048_GGS-rAP:rAPOBEC1-XTEN ins-位点

[1031] 6-G1247-D10A-UGIx2

[1032] ATGGACAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCGAC
GAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGG
AGCCCTGCTGTTCGACAGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGAC
GGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGA

CTGGAAGAGTCCTTCCTGGTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGT
GGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCGACCTGC
GGCTGATCTATCTGGCCCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGAC
AACAGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCCATCAACGC
CAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGC
TGCCCGGCGAGAAGAAGAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGC
AACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAACCTGCTGGC
CCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCC
TGAGAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGAC
CTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAA
CGGCTACGCCGGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGA
TGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGC
AGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCCTGAA
GGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACA
GCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGC
GCTTCCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCA
CAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGC
CCGCCTTCCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAG
CAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAA
CGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACGAGG
ACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTAT
GCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAA
GCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACA
GAAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAG
GGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAA
GGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACC
AGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGC
CAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGG
GCGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGA
GCTTTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTG
CCCTCCGAAGAGGTCGTGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAA
GTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGG
TGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGAC
AAGCTGATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTA
CAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGGAACCGCCCTGATCA
AAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAG
AGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCATGAACTTTTTCAAGACCGAGAT
TACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATA
AGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAG

ACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGA
CCCTAAGAAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCA
AGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCC
ATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTT
CGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAACGAACTGGCCCTGCCCT
CCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCGGAGGCTCTGGAGGAAGCAGCTCT
GAGACAGGACCTGTGGCCGTGGATCCCACACTGCGGAGAAGAATTGAGCCCCACGAGTTCGAGGTGTTCTTCGACCC
CAGAGAGCTGCGGAAAGAGACATGCCTGCTGTACGAGATCAACTGGGGCGGCAGACACTCTATCTGGCGGCACACAA
GCCAGAACACCAACAAGCACGTGGAAGTGAACTTTATCGAGAAGTTTACGACCGAGCGGTACTTCTGCCCCAACACC
AGATGCAGCATCACCTGGTTTCTGAGCTGGTCCCCTTGCGGCGAGTGCAGCAGAGCCATCACCGAGTTTCTGTCCAG
ATATCCCCACGTGACCCTGTTCATCTATATCGCCCGGCTGTACCACCACGCCGATCCTAGAAATAGACAGGGCCTGC
GCGACCTGATCAGCAGCGGAGTGACAATCCAGATCATGACCGAGCAAGAGAGCGGCTACTGCTGGCGGAACTTCGTG
AACTACAGCCCCAGCAACGAAGCCCACTGGCCTAGATATCCTCACCTGTGGGTCCGACTGTACGTGCTGGAACTGTA
CTGCATCATCCTGGGCCTGCCTCCATGCCTGAACATCCTGAGAAGAAAGCAGCCTCAGCTGACCTTCTTCACAATCG
CCCTGCAGAGCTGCCACTACCAGAGACTGCCTCCACACATCCTGTGGGCCACCGGACTTAAGGGCTCTTCTGGATCT
GAAACACCTGGCACAAGTGAGAGCGCCACCCCTGAGAGCTCTGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTT
TGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCG
ACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATC
ATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGAG
GTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGACACGGATCG
ACCTGTCTCAGCTGGGAGGTGACTCTGGTGGAAGCGGAGGATCTGGCGGCAGCACCAATCTGAGCGACATCATCGAG
AAAGAGACAGGCAAGCAGCTGGTCATCCAAGAGTCCATCCTGATGCTGCCTGAAGAGGTGGAAGAAGTGATCGGCAA
CAAGCCCGAGTCCGACATCCTGGTGCACACCGCCTACGATGAGAGCACCGACGAGAACGTGATGCTGCTGACCTCTG
ACGCCCCTGAGTACAAGCCTTGGGCTCTCGTGATCCAGGACAGCAACGGCGAGAACAAGATCAAGATGCTGAGCGGC
GGCTCTGGTGGCTCTGGCGGATCTACAAACCTGTCCGATATTATTGAGAAAGAAACCGGGAAACAGCTCGTGATTCA
AGAGTCTATTCTCATGCTCCCGGAAGAAGTCGAGGAAGTCATTGGAAACAAGCCTGAGAGCGATATTCTGGTCCATA
CAGCCTACGACGAGTCTACCGATGAGAATGTCATGCTCCTCACCAGCGACGCTCCCGAGTATAAGCCATGGGCACTT
GTCATTCAGGACTCCAATGGGGAAAACAAAATCAAAATGCTCCCAAAGAAAAAACGCAAGGTGGAGGGAGCTGATAA
GCGCACCGCCGATGGTTCCGAGTTCGAAAGCCCCAAGAAGAAGAGGAAAGTCTAACCGGTCATCATCACCATCACCA
TTGAGTTTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGC
CTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGT
AGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGC
TGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCGATACCGTCGACCTCTAGCTAGA
GCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCC
GGAAGCATAAAGTGTAAAGCCTAGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGC
TTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTG
GGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCA
AAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGC
CAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGAC

GCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGC
TCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAG
CTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGC
CCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCA
GCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGG
CTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTT
GATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGA
TCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGT
CATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATAT
ATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCA
TCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAAT
GATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAA
GTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTT
AATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAG
CTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTC
CGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTC
ATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACC
GAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAA
AACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCC
AACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAA
GGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTT
ATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGA
AAAGTGCCACCTGACGTCGACGGATCGGGAGATCGATCTCCCGATCCCCTAGGGTCGACTCTCAGTACAATCTGCTC
TGATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAAAATTT
AAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGC
GATGTACGGGCCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGT
TCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCC
GCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAG
TATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGA
CGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTAT
TAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGA
TTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCG
TAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTGGTTTAG
TGAACCGTCAGATCCGCTAGAGATCCGCGGCCGCTAATACGACTCACTATAGGGAGAGCCGCCACC

[1033] pHRB-048_GGS-rAP:rAPOBEC1-XTEN ins-位点

[1034] 6-G1247-D10A-UGIx2

[1035] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTA
RRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVD
STDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSR

RLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNL
SDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRI
PYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNEL
TKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG
KTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMG
RHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL
DINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAE
RGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINN
YHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGE
IRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYG
GFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELEN
GRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGGGSGGSSSETGPVAVDPTLRRRIEPHEFEVFFDPREL
RKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYP
HVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYC
IILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKGSSGSETPGTSESATPESSGSPEDNEQKQLF
VEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK
RYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVI
GNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSGGSGGSTNLSDIIEKETGKQL
VIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLPKKKRKVE
GADKRTADGSEFESPKKKRKV*

[1036] 实施例5:ABE内部融合碱基编辑器

[1037] 为了评估细胞内的碱基编辑,使用Lipofectamine 2000(生命科技),用100ng编码sgRNA的质粒和编码碱基编辑器的质粒共-转染HEK293T细胞。4天后,分离基因组DNA,并通过PCR扩增所述被靶向的基因组区域。加上了测序接头以生成PCR产物的文库。含有碱基-被编辑的区域的制备好的PCR文库在Illumina MiSeq上测序。

[1038] IBE003、IBE008、IBE007、IBE002、IBE001、IBE005、IBE006、IBE004、IBE021、IBE031、IBE020、IBE036、IBE035、IBE028和IBE009用于确定在两个不同的靶序列,HEK2(GAACACAAAGCATAGACTGC)和T39(GGACAGCTTTTCCTAGACAG),中的编辑百分比(图20C至Q)。

[1039] 实施例6:内部融合碱基编辑器的效率。

[1040] 使用Lipofectamine 2000(生命科技),用100ng编码sgRNA的质粒和编码碱基编辑器的质粒共-转染HEK293T细胞。4天后,分离基因组DNA,并通过PCR扩增所述被靶向的基因组区域。加上了测序接头以生成PCR产物的文库。含有碱基-被编辑的区域的制备好的PCR文库在Illumina MiSeq上测序。用于这个实验的靶位点显示于下表11中。

[1041] 表11:

[1042]

名称	序列
靶 1	GGCCTCCGTATCACTCTCTGACTGGGGT
靶 2	GTAACTGAACCCCTGCAATCAATGGGAT
靶 3	GCCACAGACTTTTCCATTTGCAGGAGT
靶 4	GCGAAAGGCTCGCGGCGAAGGAAGGAAT
靶 5	GCCTGGCAGATGAGAACCAGGAGGAAT
靶 6	GTATTACTATTATTATCTGAGATGGGGT
靶 7	GCCACAGTGGGAGGGGACATGGGGAAT
靶 8	GCCCTGATCTGCACTGAACAGAGGGGT

[1043]

靶 9	GCCTCAAGTCTGGTTATTTTAGGGGGAT
靶 10	GGTCGACCCTTGGTATCCATGGGGGAT
靶 11	GAAAGAGACAGAGAAGGGGCAGGGGGT
靶 12	GAGTGGGAACTTTCTGATGCCATGGAAT
靶 13	GTGGGACTGATCCCTTAATGTGTGGGGT
靶 14	GCCCAGCTCCAGCCTCTGATGAGGGGT
靶 15	GAAGGCTTTACTGTATTACAGAAGGGGT
靶 16	GGAGCCAGAGACCAGTGGGCAGGGGGT
靶 17	GCTTTCCTTAGCTGTAAAAGAAAGGGAT
靶 18	GAGAAGAAACCAGGGAACAGGTAGGAGT
靶 19	GGCCTCCGTATCACTCTCTGACTGGGGT
靶 20	GATGTGTCTACTGTTACTTACAAGGAAT
靶 21	GACCAGGTCAGCAAACATGTTTGGAAT
靶 22	GCACCCAGGGGTTCTGCAGAGCAGGGAT
靶 23	GCCCAGCAATTCACTGTGAAGAGGGAT
靶 24	GACCAAAACGAGGGACATTTAGGGGAT
靶 25	GCTCCTCTCACCCTTATGACTCAGGGAT
靶 26	GACTCAGCGCCCCTGCCGGGCCTGGGAT
靶 27	GGTCGTAGCCAGTCCGAACCCCGGAGT
靶 28	GCATTCCACTCCGTCCGCCTCCGGAGT
靶 29	GGGTACCTGAGTGGGGTGCATTTGGGGT

[1044] 将测序读数对齐到原始的靶序列,并测量了编辑百分比。参见图22A,其检视了在29个基因组靶标处的相对于ABE7.10的内部融合的效率。以最佳位置(即位置14,在此图中20距PAM最远而1距PAM最近)的ABE7.10编辑对编辑效率进行了归一化。以ABE7.10的最大编辑效率对跨越所有位点的内部碱基编辑器的最大编辑效率进行了归一化(图22B)。基于最大编辑效率(其被归一化到ABE7.10)的有效碱基编辑窗口表明,与ABE7.10相比,内部融合A碱基编辑器中的最大编辑窗口有所改变(图22C、D)。

[1045] 实施例7:内部融合碱基编辑器的假性脱氨反应的评估

[1046] 碱基编辑器(其具有向导)与SaCas9和靶向不同基因座的向导一起被转染。SaCas9所生成的ssDNA易受脱氨反应的影响,而该脱氨反应是通过假性的碱基编辑(假性的=非向导-靶向的)。对所述SaCas9靶基因座测序以测量假性脱氨反应。在每个位点处,反式-编辑被归一化到ABE7.10的反式编辑。通过反式-编辑测定测量了跨越29个不同IBE靶标的假性

脱氨反应,该测定结果在每个位点处被归一化到ABE7.10的反式编辑位点以进行比较。在归一化到某位点处的ABE7.10的反式-编辑之前,对该位点的总的反式-编辑求和。所测试的内部碱基编辑器(IBE002、IBE004、IBE005、IBE006、IBE008、IBE009、IBE020)与ABE7.10相比显示出减低的平均假性脱氨反应。

[1047] 实施例8:内部融合A碱基编辑器的碱基编辑的评估

[1048] 如实施例6(其使用高通量测序)中所述,使用HEK293T细胞评估了使用ABE内部融合的碱基编辑。在这个测定中,设计了向导以靶向6个不同位点:HEK4、GGCACTGCGGCTGGAGGTGG(图24A);FANCF、GTAGGGCCTTCGCGCACCTCA(图24B);HEK-3、LGGCCCAGACTGAGCACGTGA(图24C);HEK2-YY、GGAAACCTTGAATAAGAATGGA(图24D);EMX1、GAGTCCGAGCAGAAGAAGAA(图24E)和HEK2、GAACACAAAGCATAGACTGC(图24F)。

[1049] 每孔接种了10,000-20,000个HEK293T细胞。将75ng的sgRNA和175ng的碱基编辑器或Cas9质粒用1μL Lipofectamine 2000转染。转染四天后,分离了基因组DNA,PCR扩增了所述靶位点并在Illumina MiSeq上对其测序。编辑百分比是通过计算40,000个Illumina测序读数在指定位置具有A到G突变的百分比而得。与ABE7.10相比,内部融合腺苷碱基编辑器展现出不同的最大编辑窗口和减低的脱靶编辑(图24A-F)。

[1050] 实施例9:内部融合C碱基编辑器的碱基编辑的评估

[1051] 如实施例6(其使用高通量测序)中所述,使用HEK293T细胞评估了使用CBE内部融合的碱基编辑。在这个测定中,使用了如下的CBE碱基编辑器:BE4、HR001、HR002、HR003、HR004、HR005。在这个测定中,设计了向导以靶向6个不同位点:HEK4、GGCACTGCGGCTGGAGGTGG(图25A);FANCF、GTAGGGCCTTCGCGCACCTCA(图25B);HEK-3、GGCCCAGACTGAGCACGTGA(图25C);HEK2-YY、GGAAACCTTGAATAAGAATGGA(图25D);EMX1、GAGTCCGAGCAGAAGAAGAA(图25E)和HEK2、GAACACAAAGCATAGACTGC(图25F)。编辑百分比是通过计算40,000个Illumina测序读数在指定位置具有C到T突变的百分比而得。与ABE7.10相比,内部融合胞苷碱基编辑器展现出不同的最大编辑窗口和减低的脱靶编辑(图25A-F)。

[1052] 涵盖所述碱基编辑器系统的方法和组合物及其用途的以下编号的额外的实施方式在本文中展望如下:

[1053] 1.一种融合蛋白,其包括两侧翼为Cas9多肽的N末端片段和C末端片段的脱氨酶,其中该融合蛋白的脱氨酶使靶多核苷酸序列中的靶核碱基脱氨基,其中所述N末端片段或C末端片段结合靶多核苷酸序列,并且其中所述N末端片段的C端或C末端片段的N端包括Cas9多肽的柔性环的一部分。

[1054] 2.如实施方式1所述的融合蛋白,其中该靶核碱基,与包括融合至SEQ ID NO:1的N端或C端的脱氨酶的末端融合蛋白相比,以较低的脱靶脱氨反应被脱氨基。

[1055] 3.如实施方式1或2所述的融合蛋白,其中该靶核碱基与靶多核苷酸序列中的原间隔序列临近基序(PAM)序列相距1至20个核碱基。

[1056] 4.如实施方式3所述的融合蛋白,其中该靶核碱基在PAM序列上游2-12个核碱基处。

[1057] 5.如实施方式1至4中任一项所述的融合蛋白,其中当该融合蛋白使靶核碱基脱氨基时,所述N末端片段的C端或C末端片段的N端包括靠近靶核碱基的氨基酸。

[1058] 6.如实施方式1至4中任一项所述的融合蛋白,其中该N末端片段的C端或该C末端

片段的N端包括Cas9多肽的α-螺旋结构的一部分。

[1059] 7.如实施方式1至4中任一项所述的融合蛋白,其中该N末端片段或该C末端片段包括DNA结合结构域。

[1060] 8.如实施方式1至4中任一项所述的融合蛋白,其中该N末端片段或该C末端片段包括RuvC结构域。

[1061] 9.如实施方式1至-4中任一项所述的融合蛋白,其中该N末端片段和该C末端片段均不包括HNH结构域。

[1062] 10.如实施方式1至4中任一项所述的融合蛋白,其中该Cas9多肽的柔性环包括SEQ ID NO:1中所编号的位置530至537、569至579、686至691、768至793、943至947、1002至1040、1052至1077、1232至1248或1298至1300或其相应的位置之间的氨基酸。

[1063] 11.如实施方式10所述的融合蛋白,其中该N末端片段起始于Cas9多肽的N端并且是一接连的序列,其终止于SEQ ID NO:1中所编号的氨基酸1至529、538至568、580至685、692至942、948至1001、1026至1051、1078至1231或1248至1297或其相应的位置之间的位置。

[1064] 12.如实施方式10所述的融合蛋白,其中该C末端片段起始于SEQ ID NO:1中所编号的位置1301至1368、1248至1297、1078至1231、1026至1051、948至1001、692至942、580至685或538至568并且是一接连的序列,其终止于Cas9多肽的C端。

[1065] 13.如实施方式10所述的融合蛋白,其中该N末端片段的C末端氨基酸是SEQ ID NO:1中所编号的氨基酸1016、1023、1029、1040、1069或1247或其相应的氨基酸。

[1066] 14.如实施方式10所述的融合蛋白,其中该C末端片段的N末端氨基酸是SEQ ID NO:1中所编号的氨基酸1017、1024、1030、1041、1070、1248或其相应的氨基酸。

[1067] 15.如实施方式11至14中任一项所述的融合蛋白,其中该脱氨酶是胞苷脱氨酶。

[1068] 16.如实施方式10所述的融合蛋白,其中该N末端片段起始于Cas9多肽的N端并且是一接连的序列,其终止于SEQ ID NO:1中所编号的位置530至537、569至579、686至691、768至793、943至947、1002至1040、1052至1077、1232至1248或1298至1300或其相应的位置之间。

[1069] 17.如实施方式10所述的融合蛋白,其中该C末端片段起始于SEQ ID NO:1中所编号的位置530至537、569至579、686至691、768至793、943至947、1002至1040、1052至1077、1232至1248或1298至1300之间并且是一接连的序列,其终止于Cas9多肽的C端。

[1070] 18.如实施方式10所述的融合蛋白,其中该N末端片段的C末端氨基酸是SEQ ID NO:1中所编号的氨基酸1022、1029、1040、1068、1069、1247、1054、1026、768、791、792、1248、1052或1246或其相应的氨基酸。

[1071] 19.如实施方式10所述的融合蛋白,其中该C末端片段的N末端氨基酸是SEQ ID NO:1中所编号的氨基酸1023、1030、1041、1069、1070、1248、1055、1026、769、792、793、873、907、1249、1053或1247抑或其相应的氨基酸。

[1072] 20.如实施方式16至19中任一项所述的融合蛋白,其中该脱氨酶是腺苷脱氨酶。

[1073] 21.如实施方式1至20中任一项所述的融合蛋白,还包括额外的催化结构域。

[1074] 22.如实施方式21所述的融合蛋白,其中该额外的催化结构域是胞苷脱氨酶或腺苷脱氨酶。

[1075] 23.如实施方式1至22中任一项所述的融合蛋白还包括该N末端片段和脱氨酶之间

的连接子。

[1076] 24.如实施方式1至22中任一项所述的融合蛋白还包括该C末端片段和脱氨酶之间的连接子。

[1077] 25.如实施方式1至22中任一项所述的融合蛋白还包括核定位信号。

[1078] 26.如实施方式25所述的融合蛋白,其中该核定位信号是双分型核定位信号。

[1079] 27.如实施方式1至26中任一项所述的融合蛋白,其中该Cas9多肽是化脓性链球菌Cas9(SpCas9)、金黄色葡萄球菌Cas9(SaCas9)、嗜热链球菌1Cas9(St1Cas9)或其变体。

[1080] 28.如实施方式1至27中任一项所述的融合蛋白,其中该Cas9多肽是经修饰的Cas9并且对改变过的PAM有特异性。

[1081] 29.如实施方式1至28中任一项所述的融合蛋白,其中该Cas9多肽是切口酶。

[1082] 30.如实施方式1至28中任一项所述的融合蛋白,其中该Cas9多肽是核酸酶无活性的。

[1083] 31.如实施方式中1至30任一项所述的融合蛋白,其与向导核酸序列复合以达成靶核碱基的脱氨反应。

[1084] 32.用于优化碱基编辑的蛋白质文库,其包括复数个融合蛋白,其中该复数个融合蛋白中的每一个都包括两侧翼为Cas9多肽的N末端片段和C末端片段的脱氨酶,其中每一个融合蛋白的N末端片段都不同于该复数个融合蛋白中其余蛋白的N末端片段,或者,其中每一个融合蛋白的C末端片段都不同于该复数个融合蛋白中其余蛋白的C末端片段,其中每一个融合蛋白的脱氨酶使靠近靶多核苷酸序列中的原间隔序列临近基序(PAM)序列的靶核碱基脱氨基,并且其中所述N末端片段或所述C末端片段结合靶多核苷酸序列。

[1085] 33.如实施方式32所述的蛋白质文库,其中对于与PAM序列相距1到20个核碱基的每个核碱基,该复数个融合蛋白中的至少一个使该核碱基脱氨基。

[1086] 34.如实施方式32所述的蛋白质文库,其中该复数个融合蛋白中的每一个的Cas9多肽的N末端片段的C端或C末端片段的N端包括Cas9多肽的柔性环的一部分。

[1087] 35.如实施方式32至34中任一项所述的蛋白质文库,其中与包括融合至SEQ ID NO:1的N端或C端的脱氨酶的末端融合蛋白相比,所述复数个融合蛋白中的至少一个以较低的脱靶脱氨反应使核碱基脱氨基。

[1088] 36.如实施方式32至35中任一项所述的蛋白质文库,其中所述复数个融合蛋白中的至少一个使在PAM序列上游2至12个核碱基处的靶核碱基脱氨基。

[1089] 37.如实施方式32至36中任一项所述的蛋白质文库,其中当该融合蛋白使靶核碱基脱氨基时,所述复数个融合蛋白其中一个的N端片段的C端或C端片段的N端包括靠近靶核碱基的氨基酸。

[1090] 38.如实施方式34至36中任一项所述的蛋白质文库,其中该Cas9多肽的柔性环包括位于SEQ ID NO:1中所编号的位置530至537、569至579、686至691、768至793、943至947、1002至1040、1052至1077、1232至1248或1298至1300(如)抑或其相应的(多个)位置之间的氨基酸。

[1091] 39.如实施方式38所述的蛋白质文库,其中该复数个融合蛋白包括融合蛋白,该融合蛋白包括起始于Cas9多肽的N端的N末端片段并且是终止于SEQ ID NO:1中所编号的位置530至537、569至579、686至691、768至793、943至947、1002至1040、1052至1077、1232至1248

或1298至1300或其相应的(多个)位置之间的接连序列。

[1092] 40.如实施方式38所述的蛋白质文库,其中该复数个融合蛋白包括融合蛋白,该融合蛋白包括C末端片段,其起始于如下氨基酸之间的一SEQ ID NO:1中所编号的位置530至537、569至579、686至691、768至793、943至947、1002至1040、1052至1077、1232至1248或1298至1300或其相应的(多个)位置,且是终止于Cas9多肽的C端的接连序列。

[1093] 41.如实施方式38至40中任一项所述的蛋白质文库,其中该N末端片段的C末端氨基酸是SEQ ID NO:1中所编号的氨基酸1022、1029、1040、1068、1069、1247、1054、1026、768、791、792、1248、1052或1246或其相应的氨基酸。

[1094] 42.如实施方式38至40中任一项所述的蛋白质文库,其中该C末端片段的N末端氨基酸是SEQ ID NO:1中所编号的氨基酸1023、1030、1041、1069、1070、1248、1055、1026、769、792、793、873、907、1249、1053或1247或其相应的氨基酸。

[1095] 43.如实施方式32至42中任一项所述的蛋白质文库,其中该脱氨酶是腺苷脱氨酶。

[1096] 44.如实施方式32至42中任一项所述的蛋白质文库,其中该脱氨酶是胞苷脱氨酶。

[1097] 45.如实施方式32至44中任一项所述的蛋白质文库,其中该Cas9多肽是化脓性链球菌Cas9(SpCas9)、金黄色葡萄球菌Cas9(SaCas9)、嗜热链球菌1Cas9(St1Cas9)或其变体。

[1098] 46.如实施方式32至45中任一项所述的蛋白质文库,其中该Cas9多肽是经修饰的Cas9并且对改变过的原间隔序列临近基序(PAM)有特异性。

[1099] 47.如实施方式32至46中任一项所述的蛋白质文库,其中该Cas9多肽是切口酶。

[1100] 48.如实施方式32至46中任一项所述的蛋白质文库,其中该Cas9多肽是核酸酶无活性的。

[1101] 49.一种细胞,其包括如实施方式1至31中任一项所述的融合蛋白。

[1102] 50.如实施方式49所述的细胞,其中该细胞是哺乳动物细胞或人类细胞。

[1103] 51.一种用于编辑靶多核苷酸序列中的靶核碱基的方法,该方法包括:使该靶多核苷酸与融合蛋白接触,该融合蛋白包括两侧翼为Cas9多肽的N末端片段和C末端片段的脱氨酶,其中该融合蛋白的脱氨酶使靶多核苷酸序列中的靶核碱基脱氨基,其中该N末端片段或C末端片段结合靶多核苷酸序列,并且其中所述N末端片段的C端或所述C末端片段的N端包括Cas9多肽的柔性环的一部分。

[1104] 52.如实施方式51所述的方法,还包括使所述靶多核苷酸序列与向导核酸序列接触以达成靶核碱基的脱氨反应。

[1105] 53.如实施方式52所述的方法,其中该向导核酸序列包括一间隔序列,其序列与靶多核苷酸序列的原间隔序列互补,从而形成了R-环。

[1106] 54.如实施方式53所述的方法,其中该靶核碱基,与包括融合至SEQ ID NO:1的N端或C端的脱氨酶的末端方法相比,以较低的脱靶脱氨反应被脱氨基。

[1107] 55.如实施方式54所述的方法,其中该融合蛋白的脱氨酶在R-环范围内使不超过两个的核碱基脱氨基。

[1108] 56.如实施方式51至55中任一项所述的方法,其中该靶核碱基与靶多核苷酸序列中的PAM序列相距1至20个核碱基。

[1109] 57.如实施方式55所述的方法,其中该靶核碱基在PAM序列上游2-12个核碱基处。

[1110] 58.如实施方式51至57中任一项所述的方法,其中当该融合蛋白的脱氨酶使靶核

碱基脱氨基时,该N端片段的C端或C端片段的N端包括靠近靶核碱基的氨基酸。

[1111] 59.如实施方式51至57中任一项所述的方法,其中该N末端片段或该C末端片段包括RuvC结构域。

[1112] 60.如实施方式51至57中任一项所述的方法,其中该N末端片段和该C末端片段均不包括HNH结构域。

[1113] 61.如实施方式51至57中任一项所述的方法,其中该Cas9多肽的柔性环包括SEQ ID NO:1中所编号的位置530至537、569至579、686至691、768至793、943至947、1002至1040、1052至1077、1232至1248或1298至1300或其相应的位置之间的氨基酸。

[1114] 62.如实施方式61所述的方法,其中该N末端片段起始于Cas9多肽的N端并且是终止于SEQ ID NO:1中所编号的位置530至537、569至579、686至691、768至793、943至947、1002至1040、1052至1077、1232至1248或1298至1300(如)抑或其相应的位置之间的接连序列。

[1115] 63.如实施方式61所述的方法,其中该C末端片段起始于SEQ ID NO:1中所编号的位置530至537、569至579、686至691、768至793、943至947、1002至1040、1052至1077、1232至1248或1298至1300(或其相应的(多个)位置,并且是终止于Cas9多肽的C端的接连序列。

[1116] 64.如实施方式61所述的方法,其中该N末端片段的C末端氨基酸是SEQ ID NO:1中所编号的氨基酸1016、1023、1029、1040、1069或1247或其相应的氨基酸。

[1117] 65.如实施方式61所述的方法,其中该C末端片段的N末端氨基酸是SEQ ID NO:1中所编号的氨基酸1017、1024、1030、1041、1070、1248或其相应的氨基酸。

[1118] 66.如实施方式62至65中任一项所述的方法,其中该脱氨酶是胞苷脱氨酶。

[1119] 67.如实施方式61所述的方法,其中该N末端片段起始于Cas9多肽的N端并且是终止于SEQ ID NO:1中所编号的位置1至529、538至568、580至685、692至942、948至1001、1026至1051、1078至1231或1248至1297或其相应的位置之间的接连序列。

[1120] 68.如实施方式61所述的方法,其中该C末端片段起始于SEQ ID NO:1中所编号的位置1至529、538至568、580至685、692至942、948至1001、1026至1051、1078至1231或1248至1297或其相应的(多个)位置之间并且是终止于Cas9多肽的C端的接连序列。

[1121] 69.如实施方式61所述的方法,其中该N末端片段的C末端氨基酸是SEQ ID NO:1中所编号的氨基酸1022、1029、1040、1068、1069、1247、1054、1026、768、791、792、1248、1052或1246或其相应的氨基酸。

[1122] 70.如实施方式61所述的方法,其中该C末端片段的N末端氨基酸是SEQ ID NO:1中所编号的氨基酸1023、1030、1041、1069、1070、1248、1055、1026、769、792、793、873、907、1249、1053或1247或其相应的氨基酸。

[1123] 71.如实施方式67至70中任一项所述的方法,其中该脱氨酶是腺苷脱氨酶。

[1124] 72.如实施方式51至71中任一项所述的方法,其中该Cas9多肽是经修饰的Cas9并且对改变过的原间隔序列临近基序(PAM)有特异性。

[1125] 73.如实施方式51至72中任一项所述的方法,其中该Cas9多肽是切口酶。

[1126] 74.如实施方式51至72中任一项所述的方法,其中该Cas9多肽是核酸酶无活性的。

[1127] 75.如实施方式51至74中任一项所述的方法,其中该接触是在细胞中执行。

[1128] 76.如实施方式75所述的方法,其中该细胞是哺乳动物细胞或人类细胞。

[1129] 77.如实施方式76所述的方法,其中该细胞是分化多能的细胞。

[1130] 78.如实施方式75至77中任一项所述的方法,其中该细胞是在体内的或离体的。

[1131] 79.如实施方式51至74中任一项所述的方法,其中该接触是在细胞的群体中执行。

[1132] 80.如实施方式79所述的方法,其中该细胞的群体是哺乳动物细胞或人类细胞。

[1133] 81.一种用于治疗受试者体内的遗传病况的方法,该方法包括:向该受试者施用融合蛋白,该融合蛋白包括两侧翼为Cas9多肽的N末端片段和C末端片段的脱氨酶或者编码该融合蛋白的多核苷酸,以及向导核酸序列或者编码该向导核酸序列的多核苷酸,其中该向导核酸序列导引该融合蛋白使受试者的靶多核苷酸序列中的靶核碱基脱氨基,其中该N末端片段或该C末端片段结合靶多核苷酸序列,从而治疗所述遗传病况。

[1134] 82.如实施方式81所述的方法,其中该N末端片段的C端或该C末端片段的N端包括所述Cas9多肽的柔性环的一部分。

[1135] 83.如实施方式81或82所述的方法,还包括向该受试者施用向导核酸序列以达成靶核碱基的脱氨反应。

[1136] 84.如实施方式81至83中任一项所述的方法,其中该靶核碱基包括与所述遗传病况相关联的突变。

[1137] 85.如实施方式84所述的方法,其中该靶核碱基的脱氨反应用野生型核碱基取代了该靶核碱基。

[1138] 86.如实施方式84所述的方法,其中该靶核碱基的脱氨反应用非-野生型核碱基取代了该靶核碱基,并且其中所述靶核碱基的脱氨反应改善了所述遗传病况的症状。

[1139] 87.如实施方式81至83中任一项所述的方法,其中该靶多核苷酸序列在靶核碱基以外的核碱基处包括与所述遗传病况相关联的突变。

[1140] 88.如实施方式87所述的方法,其中该靶核碱基的脱氨反应改善了所述遗传病况的症状。

[1141] 89.如实施方式81至88中任一项所述的方法,其中该靶核碱基与靶多核苷酸序列中的PAM序列相距1至20个核碱基。

[1142] 90.如实施方式89所述的方法,其中该靶核碱基在PAM序列上游2-12个核碱基处。

[1143] 91.如实施方式81至90中任一项所述的方法,其中当该融合蛋白的脱氨酶使靶核碱基脱氨基时,该N端片段的C端或C端片段的N端包括靠近靶核碱基的氨基酸。

[1144] 92.如实施方式81至90中任一项所述的方法,其中该N末端片段或该C末端片段包括RuvC结构域。

[1145] 93.如实施方式81至90中任一项所述的方法,其中该N末端片段和该C末端片段均不包括HNH结构域。

[1146] 94.如实施方式81至90中任一项所述的方法,其中该Cas9多肽的柔性环包括位于SEQ ID NO:1中所编号的位置530至537、569至579、686至691、768至793、943至947、1002至1040、1052至1077、1232至1248或1298至1300或其相应的位置的之间氨基酸。

[1147] 95.如实施方式94所述的方法,其中该N末端片段起始于Cas9多肽的N端并且是终止于如下氨基酸的一SEQ ID NO:1中所编号的位置V530至P537、F569至E579、D686至R691、Y943至D947、L1052至E1077、P1002至S1025、Y1232至G1247或R1298至K1300之间的接连序列。

[1148] 96.如实施方式94所述的方法,其中该C末端片段起始于SEQ ID NO:1中所编号的位置530至537、569至579、686至691、768至793、943至947、1002至1040、1052至1077、1232至1248或1298至1300(如)并且是终止于Cas9多肽的C端的接连序列。

[1149] 97.如实施方式94所述的方法,其中该N末端片段的C末端氨基酸是SEQ ID NO:1中所编号的氨基酸1016、1023、1029、1040、1069、1022、1029、1040、1068、1069、1247、1054、1026、768、791、792、1246、1247、1248或1052或其相应的氨基酸。

[1150] 98.如实施方式94所述的方法,其中该C末端片段的N末端氨基酸是SEQ ID NO:1中所编号的氨基酸1017、1023、1024、1030、1041、1069、1070、1247、1248、1249、1055、1026、769、792、793、873、907或1053(或其相应的氨基酸。

[1151] 99.如实施方式81至98中任一项所述的方法,其中该脱氨酶是胞苷脱氨酶。

[1152] 100.如实施方式81至98中任一项所述的方法,其中该脱氨酶是腺苷脱氨酶。

[1153] 101.如实施方式81至100中任一项所述的方法,其中该Cas9多肽是经修饰的Cas9并且对改变过的PAM有特异性。

[1154] 102.如实施方式81至101中任一项所述的方法,其中该Cas9多肽是切口酶。

[1155] 103.如实施方式81至101中任一项所述的方法,其中该Cas9多肽是核酸酶无活性的。

[1156] 104.如实施方式81至103中任一项所述的方法,其中该受试者是哺乳动物。

[1157] 105.如实施方式81至104中任一项所述的方法,其中该受试者是人类。

[1158] 其他实施方式

[1159] 从前面的描述中将显而易见的是,可对如本文所描述的发明进行变化和修改以将其适用于各种用途和条件。此类实施方式也在以下权利要求的范围内。

[1160] 本文对变量的任何定义中的元件列表的列举,包含了将该变量定义为所列元件的任何单个元件或组合(或子组合)。本文对实施方式的列举,包含了该实施方式作为任何单一实施方式或与任何其他实施方式或其部分的组合。

[1161] 本说明书中提及的所有专利和出版物通过引用并入本文,其范围如同明确地和个别地指出每个独立的专利和出版物通过引用并入。

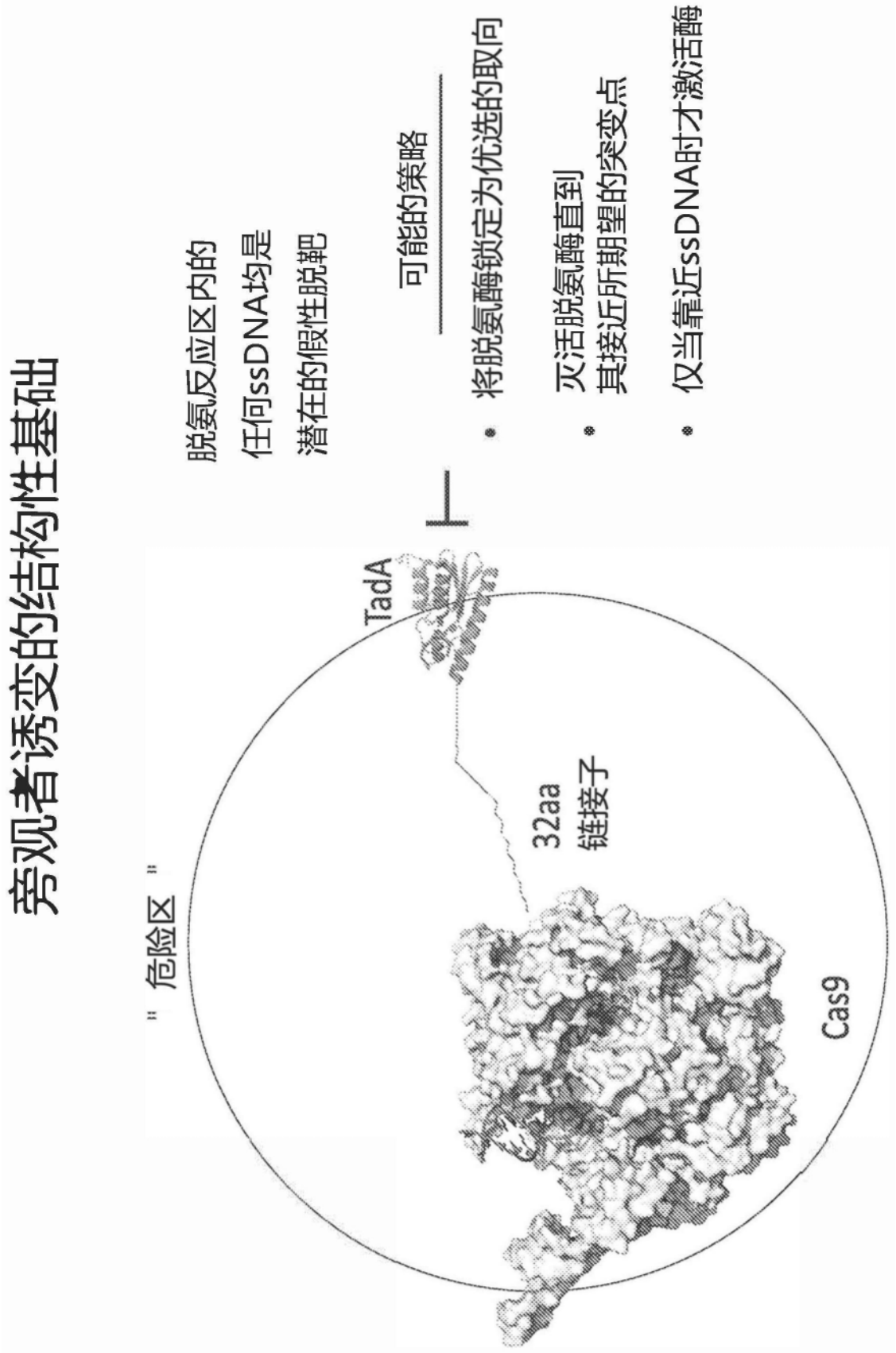

旁观者诱变的结构性基础
" 危险区 "
TadA
32aa
链接子
Cas9
脱氨反应区内的
任何ssDNA均是
潜在的假性脱靶
可能的策略
• 将脱氨酶锁定为优选的取向
• 灭活脱氨酶直到
其接近所期望的突变点
• 仅当靠近ssDNA时才激活酶

图1

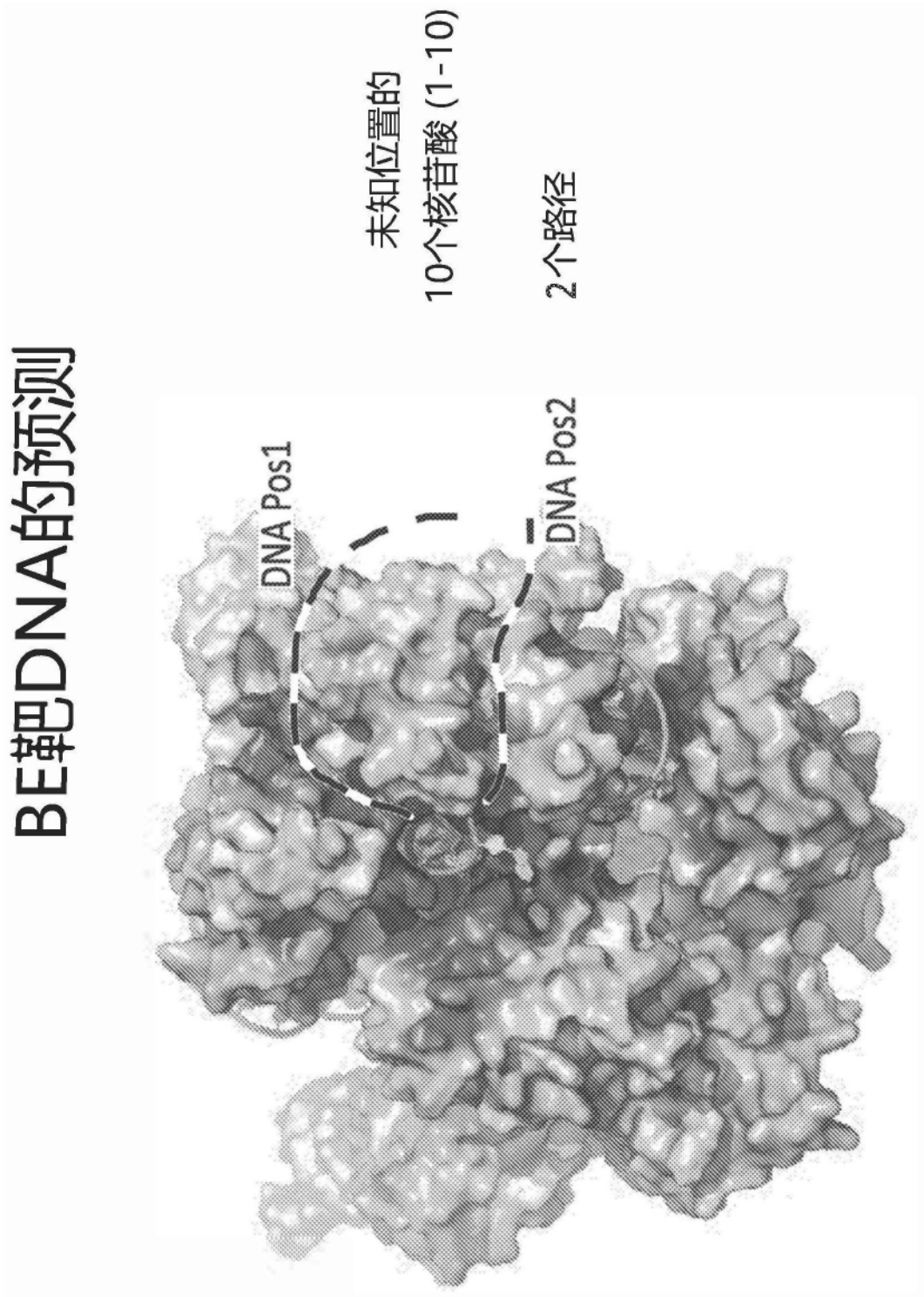


图2

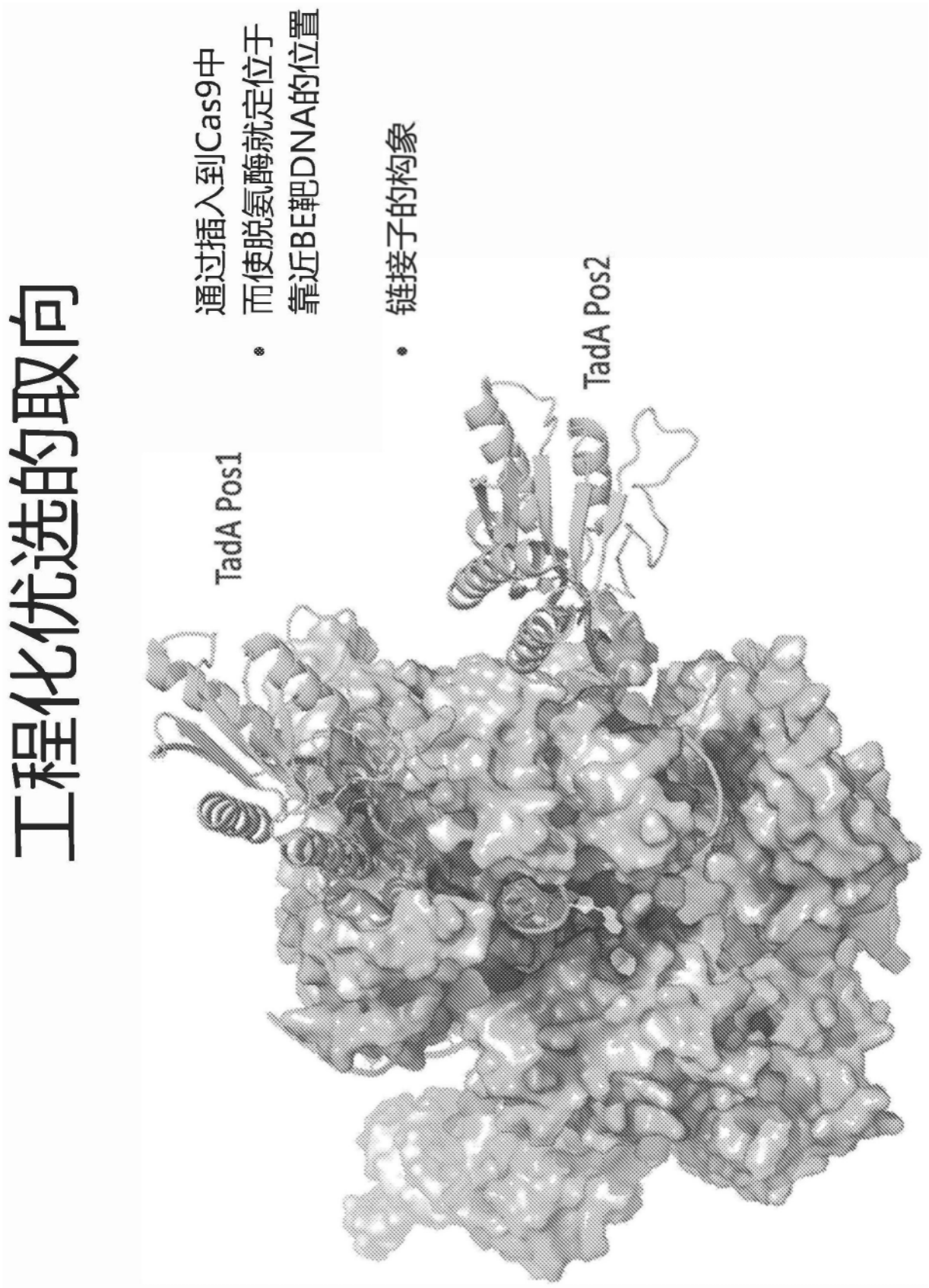

工程化优选的取向
通过插入到Cas9中
而使脱氨酶就定位于
靠近BE靶DNA的位置
链接子的构象
TadA Pos1
TadA Pos2

图3

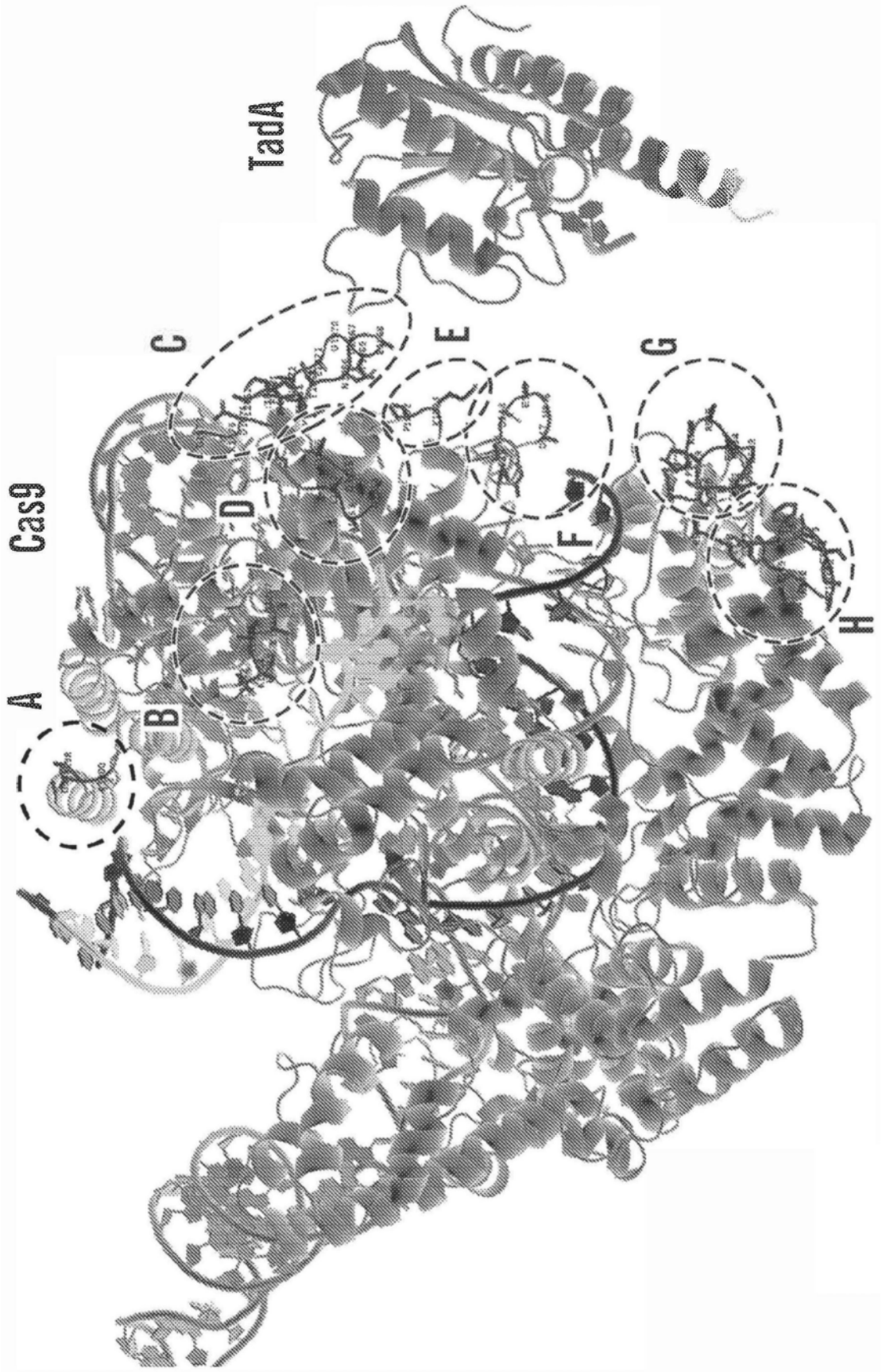


图4

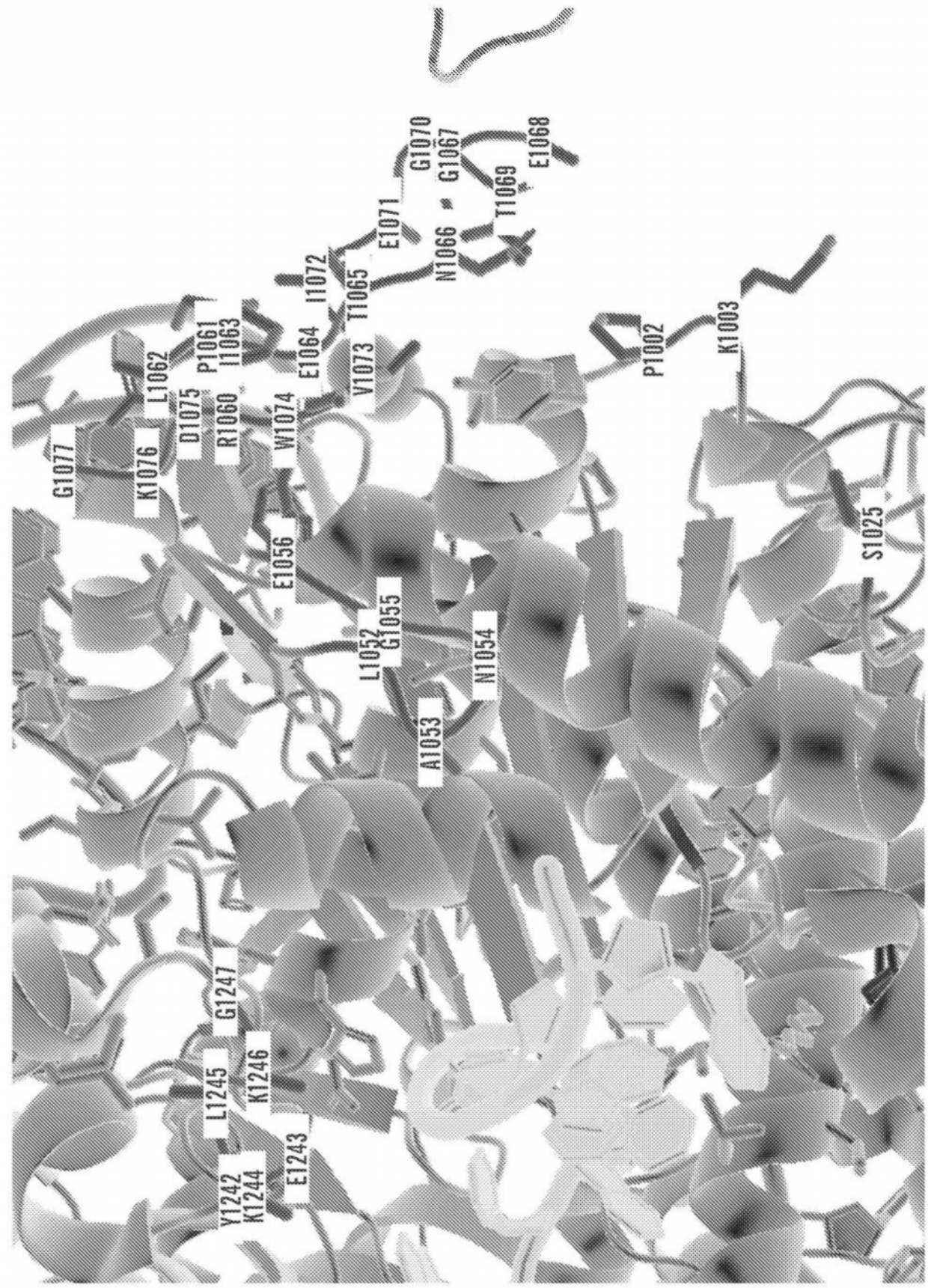

G1077
K1076
L1062
D1075
P1061
I1063
R1060
W1074
E1064
I1072
V1073
T1065
E1071
G1070
G1067
N1066
T1069
E1068
P1002
K1003
S1025
E1056
L1052
G1055
N1054
A1053
G1247
L1245
K1246
E1243
Y1242
K1244

图5

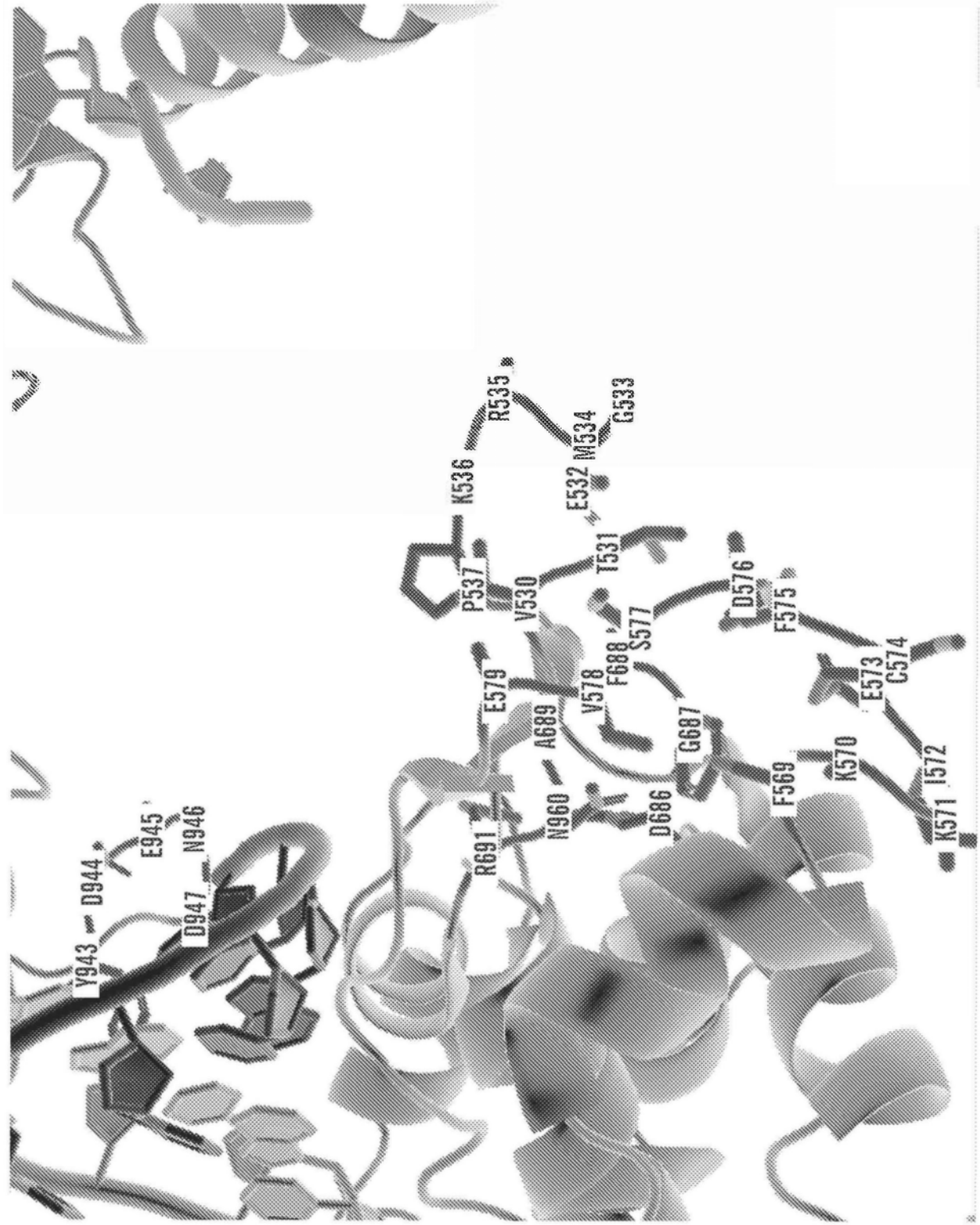

Y943
D944
E945
N946
D947
R691
N960
A689
E579
V578
F688
S577
G687
D686
K536
R535
M534
G533
E532
T531
P537
V530
D576
F575
E573
C574
F569
K570
I572
K571

图6

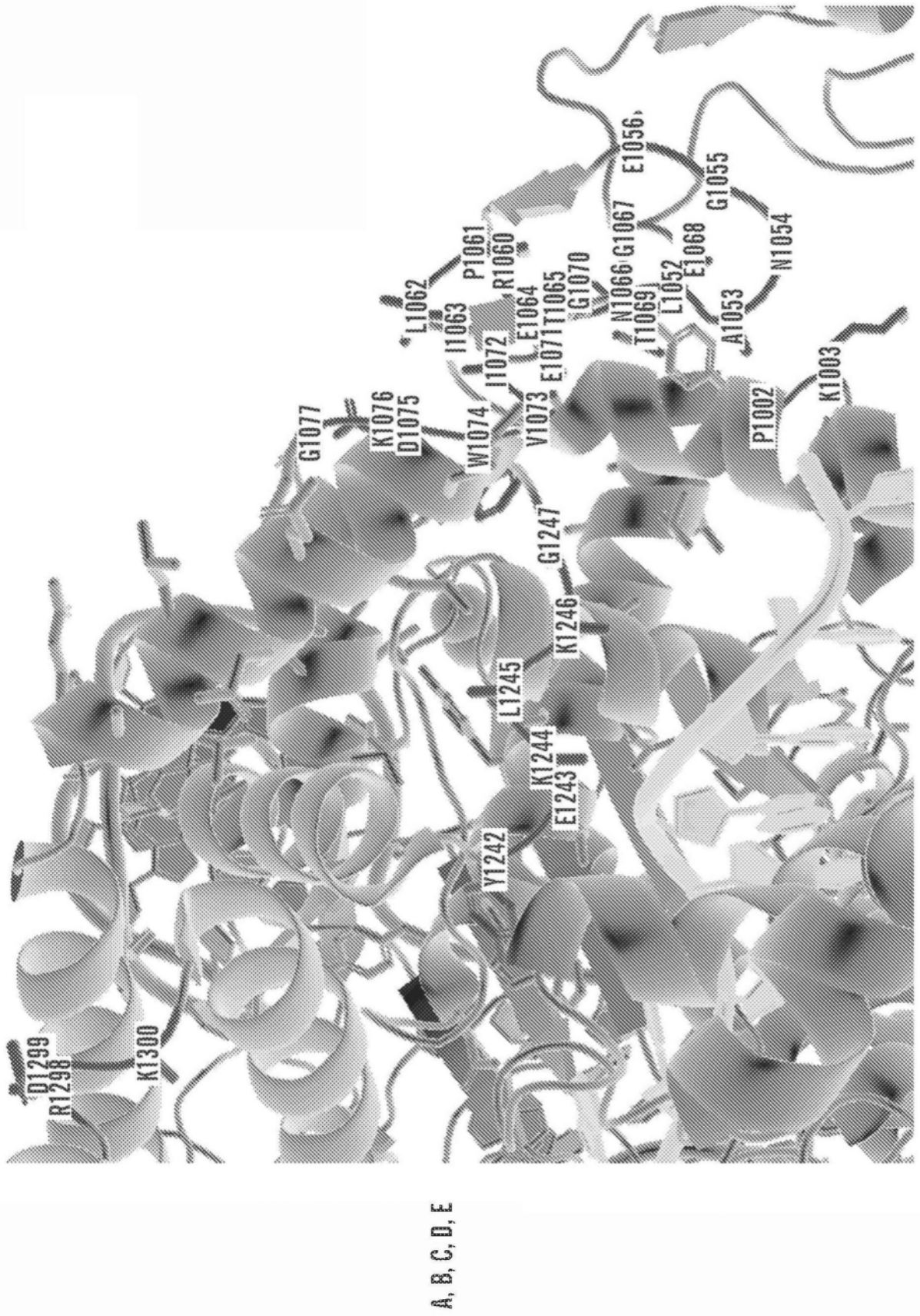

D1299
R1298
K1300
G1077
K1076
D1075
W1074
L1062
P1061
R1060
I1063
I1072
L1064
E1071
1065
V1073
G1070
N1066
G1067
E1056
T1069
L1052
E1068
G1055
A1053
N1054
P1002
K1003
G1247
K1246
L1245
K1244
E1243
V1242
A, B, C, D, E

图7

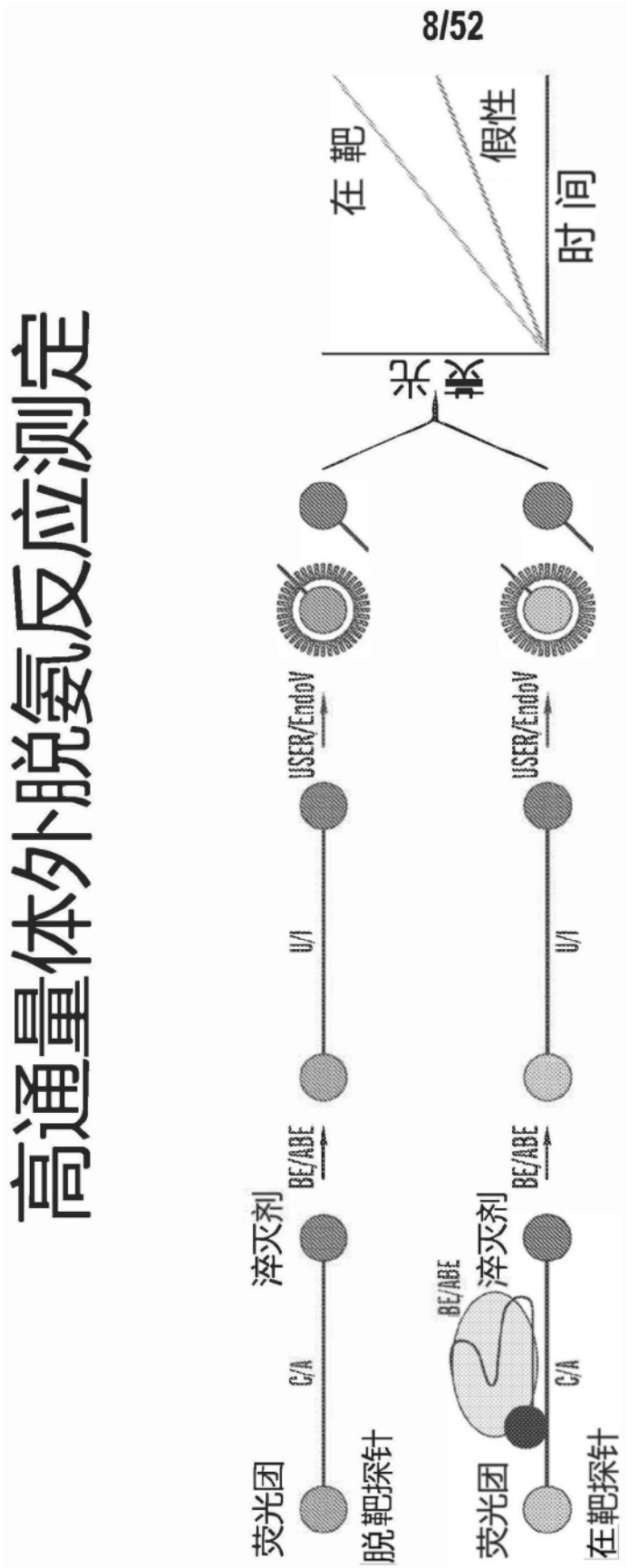

8/52
高通量体外脱氨反应测定
荧光团
C/A
脱靶探针
淬灭剂
BE/ABE
U/I
USER/EndoV
荧光团
C/A
在靶探针
淬灭剂
BE/ABE
U/I
USER/EndoV
采集
在靶
假性
时间

图8

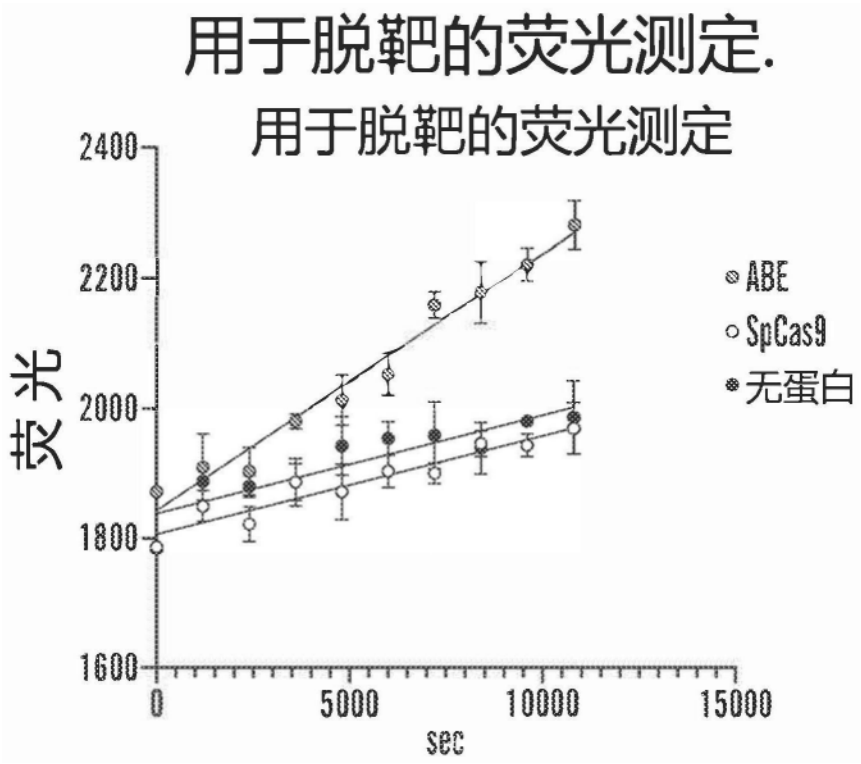

用于脱靶的荧光测定.
用于脱靶的荧光测定
荧光
2400
2200
2000
1800
1600
0
5000
10000
15000
sec
ABE
SpCas9
无蛋白

图9

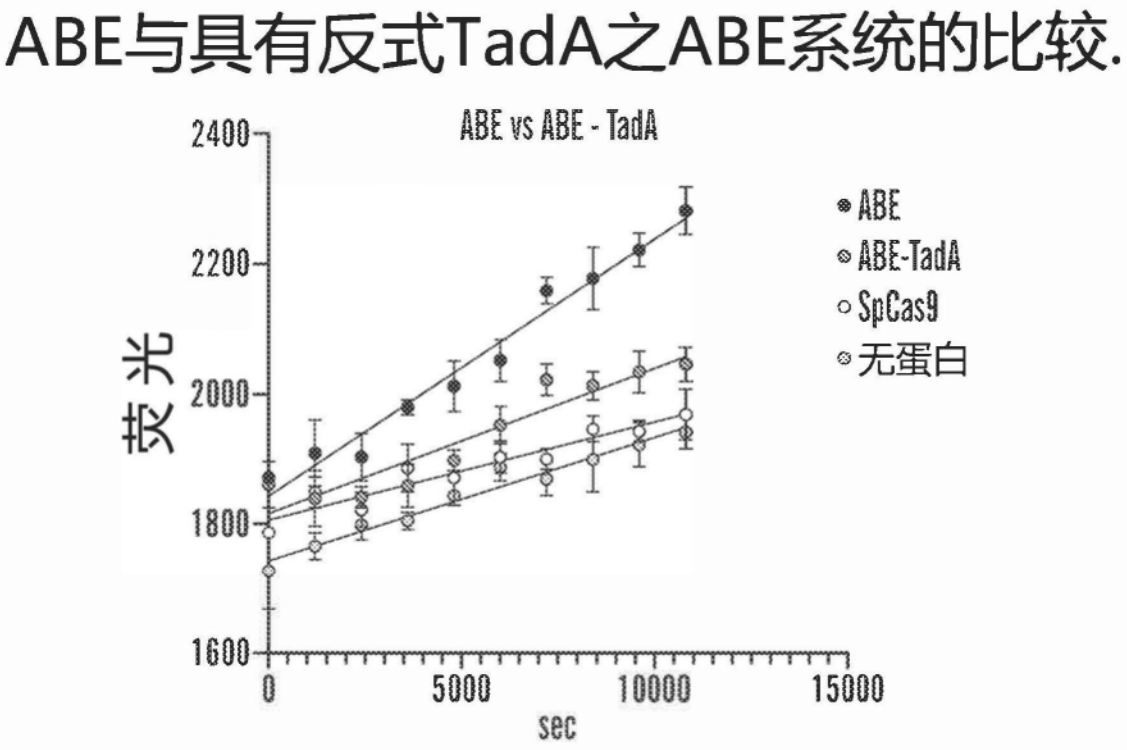

ABE与具有反式TadA之ABE系统的比较.
ABE vs ABE - TadA
荧光
2400
2200
2000
1800
1600
0
5000
10000
15000
sec
ABE
ABE-TadA
SpCas9
无蛋白

图10

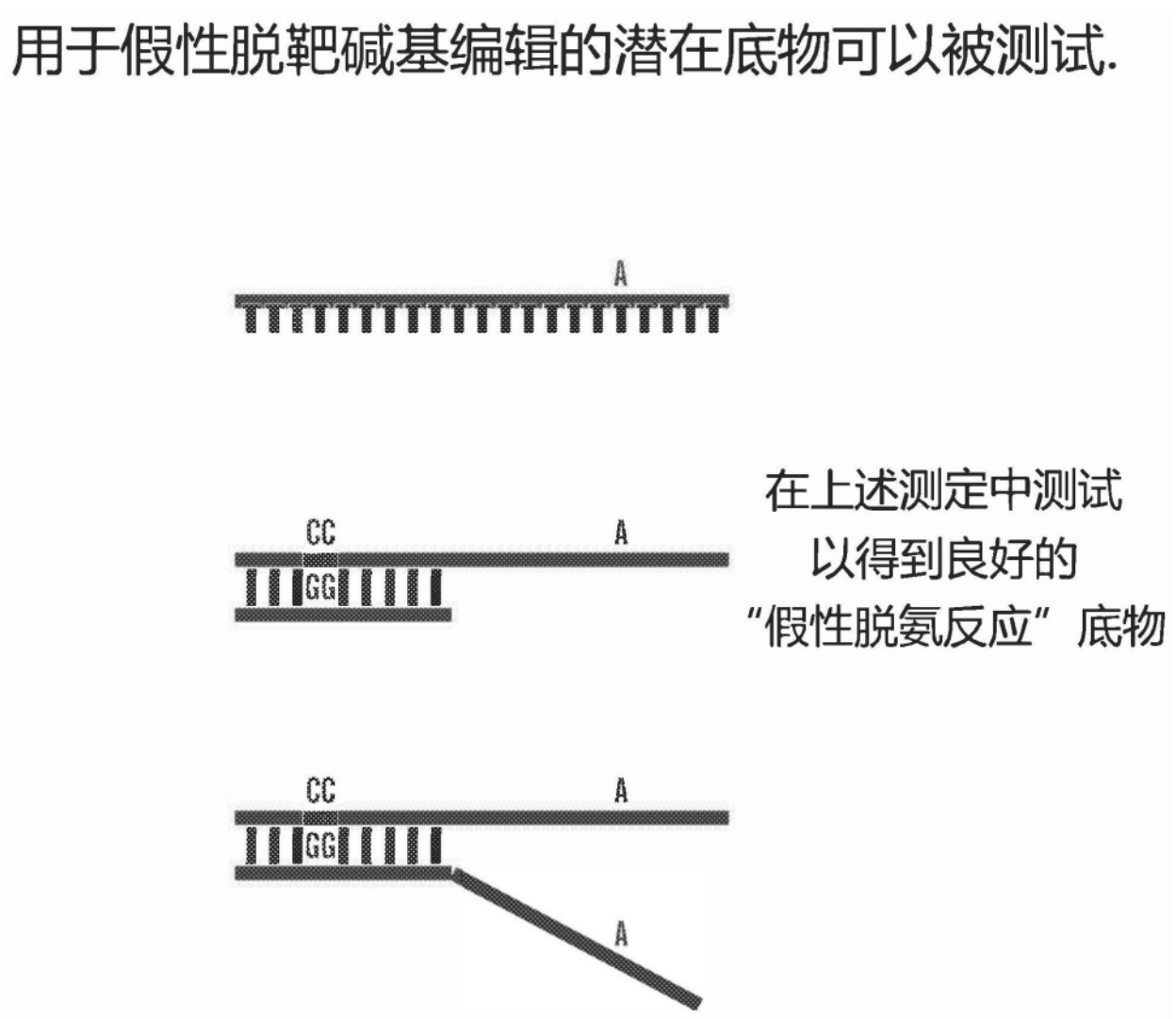

用于假性脱靶碱基编辑的潜在底物可以被测试.
A
CC
A
GG
在上述测定中测试
以得到良好的
“假性脱氨反应”底物
CC
A
GG
A

图11

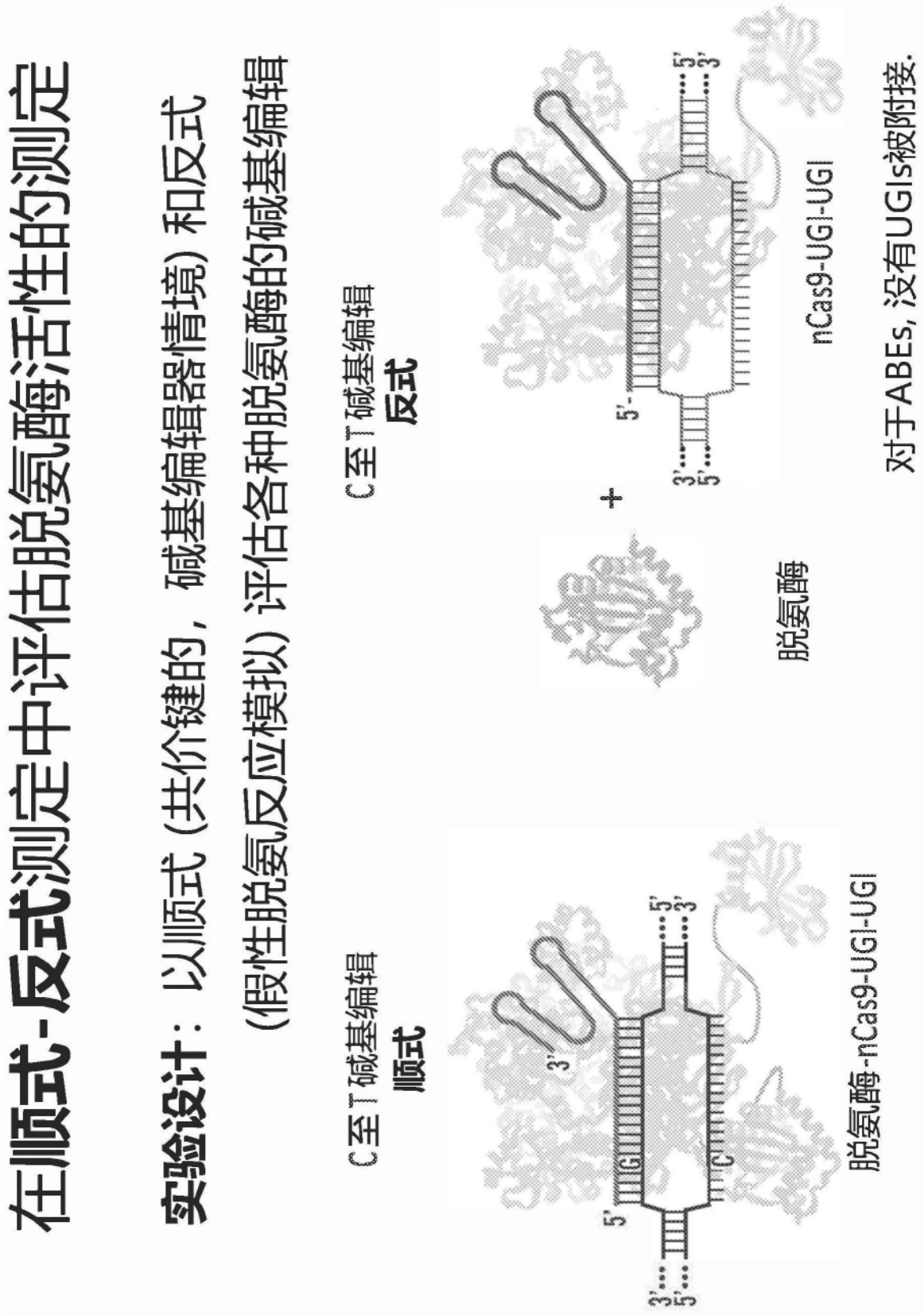

在顺式-反式测定中评估脱氨酶活性的测定
实验设计：以顺式（共价键的，碱基编辑器情境）和反式
（假性脱氨反应模拟）评估各种脱氨酶的碱基编辑
C至T碱基编辑
顺式
C至T碱基编辑
反式
脱氨酶-nCas9-UGI-UGI
脱氨酶
nCas9-UGI-UGI
对于ABEs，没有UGIs被附接.

图12

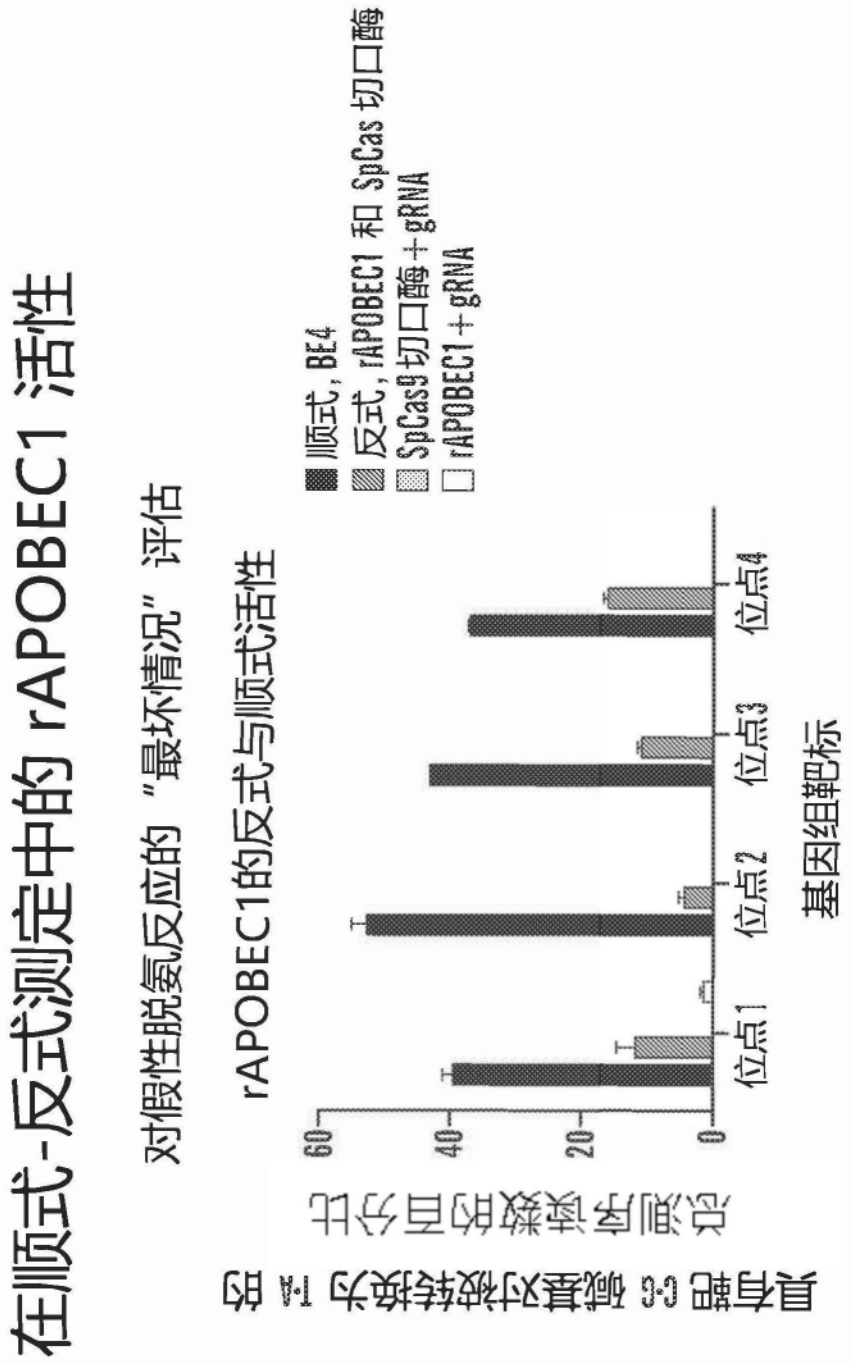

在顺式-反式测定中的 rAPOBEC1 活性
对假性脱氨反应的“最坏情况”评估
rAPOBEC1的反式与顺式活性
顺式，BE4
反式，rAPOBEC1 和 SpCas 切口酶
SpCas9 切口酶 + gRNA
rAPOBEC1 + gRNA
具有靶 C:G 碱基对被转换为 T:A 的
总测序读数的百分比
60
40
20
0
位点1
位点2
位点3
位点4
基因组靶标

图13

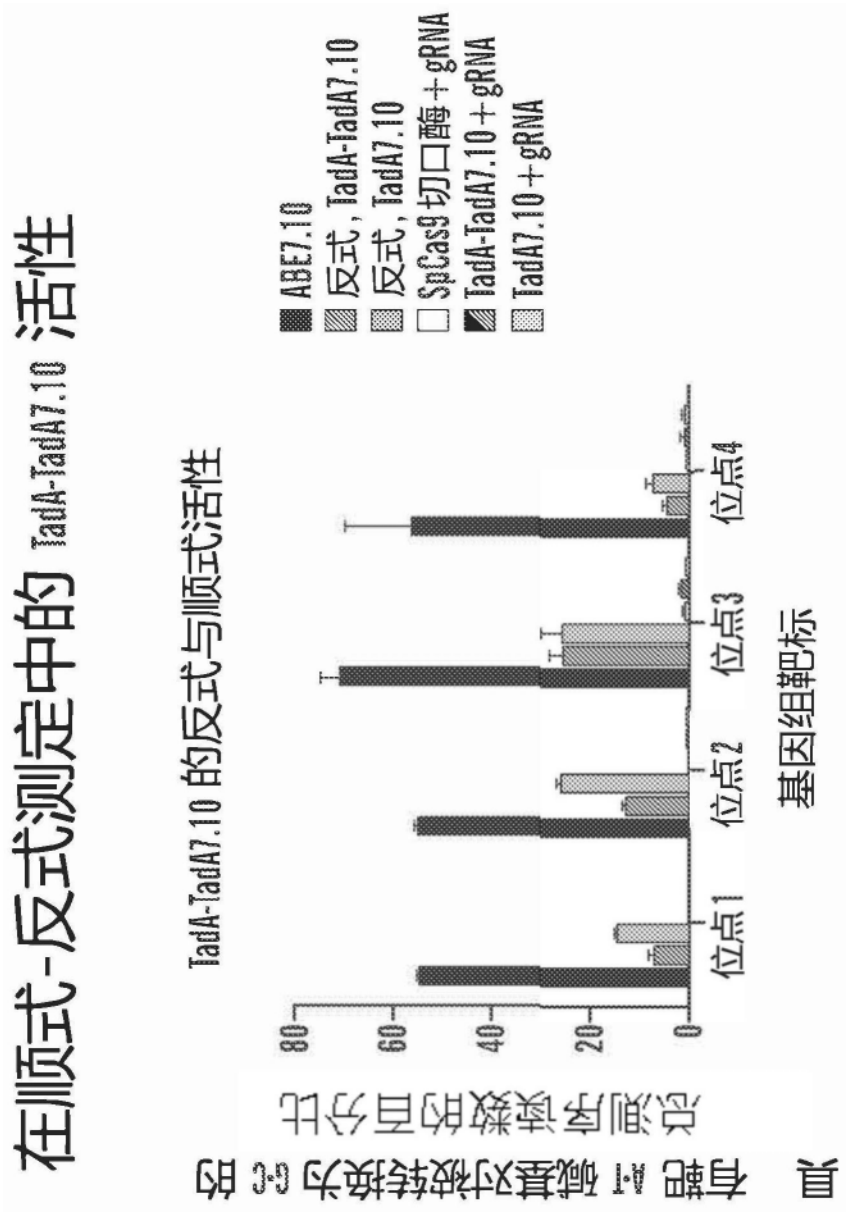

在顺式-反式测定中的 TadA-TadA7.10 活性
TadA-TadA7.10 的反式与顺式活性
ABE7.10
反式，TadA-TadA7.10
反式，TadA7.10
SpCas9 切口酶 + gRNA
TadA-TadA7.10 + gRNA
TadA7.10 + gRNA
0
20
40
60
80
位点1
位点2
位点3
位点4
基因组靶标

图14

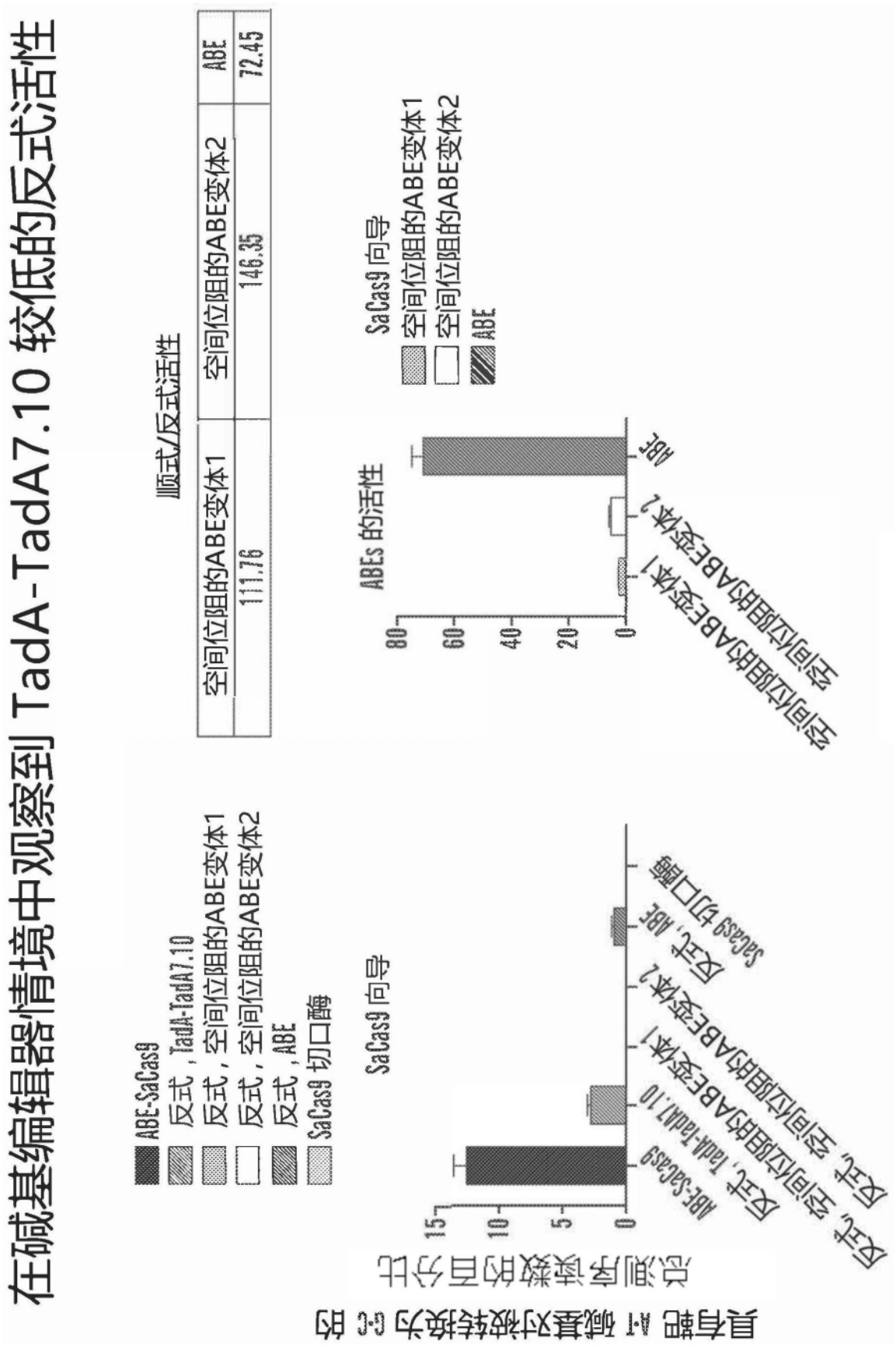

在碱基编辑器情境中观察到 TadA-TadA7.10 较低的反式活性
ABE-SaCas9
反式，TadA-TadA7.10
反式，空间位阻的ABE变体1
反式，空间位阻的ABE变体2
反式，ABE
SaCas9 切口酶
SaCas9 向导
具有靶 A-T 碱基对被转换为 G-C 的
总测序读数的百分比
15
10
5
0
ABE-SaCas9
反式，TadA-TadA7.10
反式，空间位阻的ABE变体1
反式，空间位阻的ABE变体2
反式，ABE
SaCas9 切口酶
顺式/反式活性
空间位阻的ABE变体1
空间位阻的ABE变体2
ABE
111.76
146.35
72.45
ABEs 的活性
SaCas9 向导
空间位阻的ABE变体1
空间位阻的ABE变体2
ABE
80
60
40
20
0
空间位阻的ABE变体1
空间位阻的ABE变体2
ABE

图15

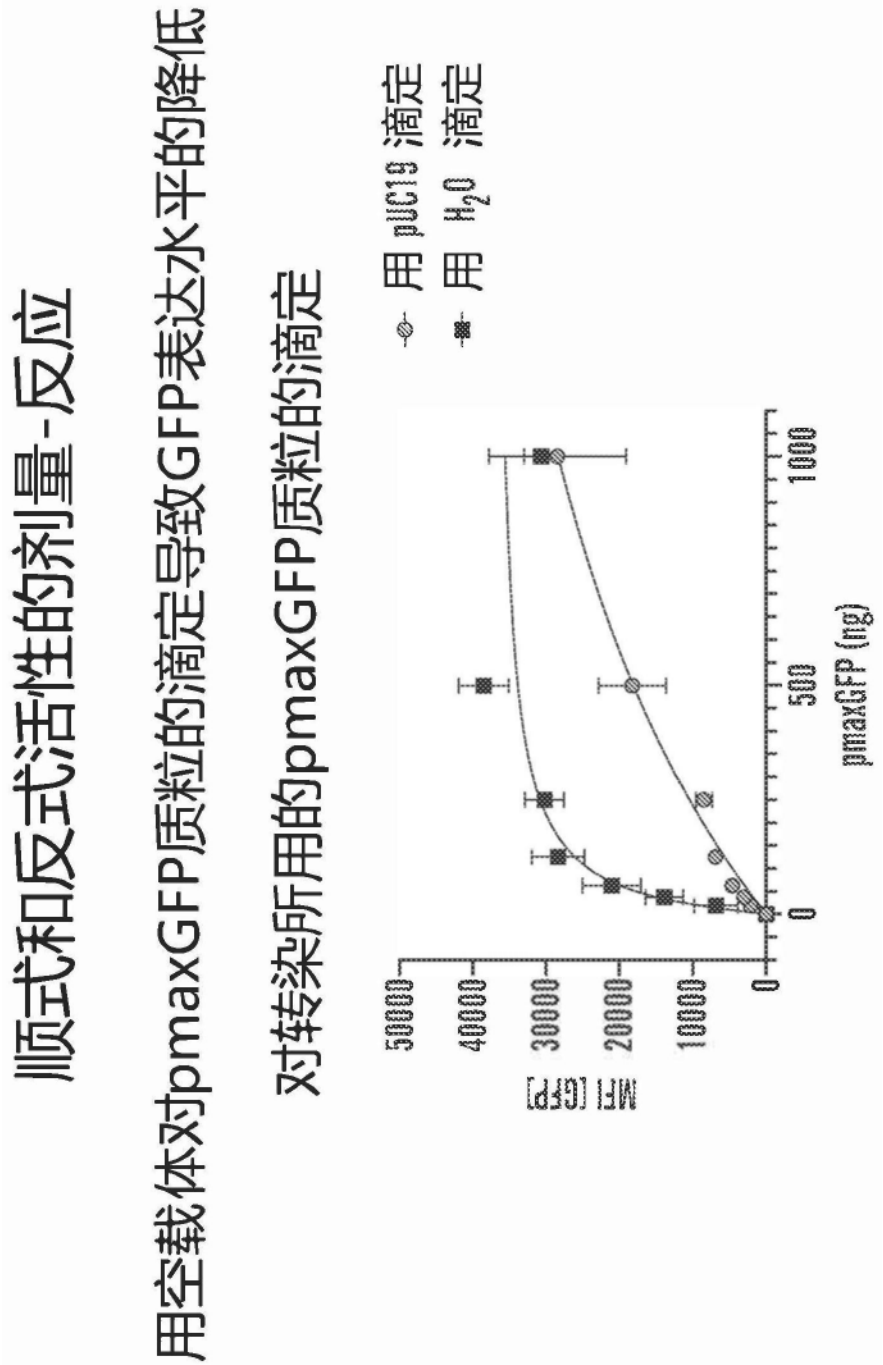

顺式和反式活性的剂量-反应
用空载体对pmaxGFP质粒的滴定导致GFP表达水平的降低
对转染所用的pmaxGFP质粒的滴定
用 pUC19 滴定
用 H2O 滴定
MFI (GFP)
50000
40000
30000
20000
10000
0
0
500
1000
pmaxGFP (ng)

图16

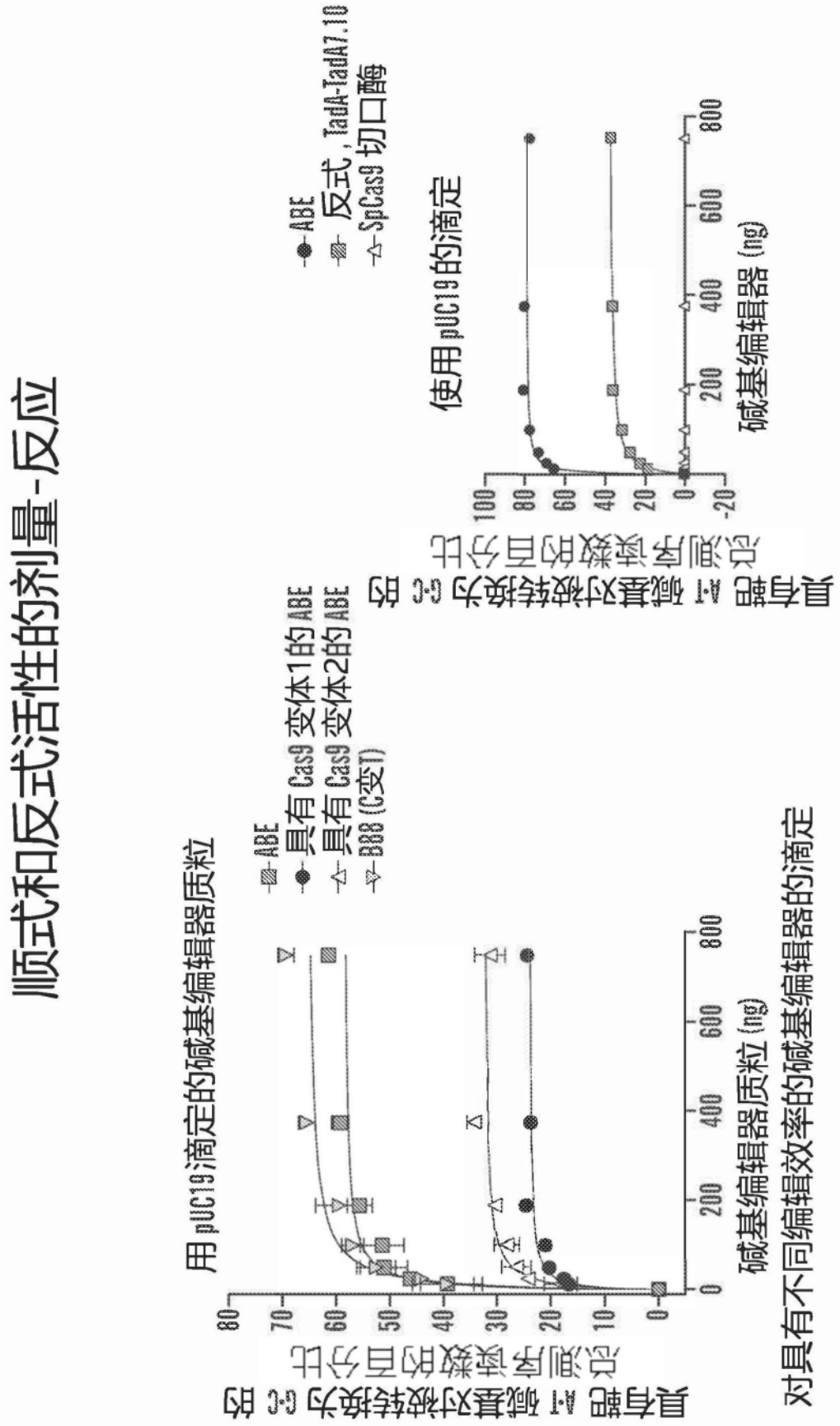

顺式和反式活性的剂量-反应
用 pUC19 滴定的碱基编辑器质粒
ABE
具有 Cas9 变体1的 ABE
具有 Cas9 变体2的 ABE
BE3 (C变T)
具有靶 A·T 碱基对被转换为 G·C 的
总测序读数的百分比
碱基编辑器质粒 (ng)
对具有不同编辑效率的碱基编辑器的滴定
使用 pUC19 的滴定
ABE
反式，TadA-TadA7.10
SpCas9 切口酶
具有靶 A·T 碱基对被转换为 G·C 的
总测序读数的百分比
碱基编辑器 (ng)

图17

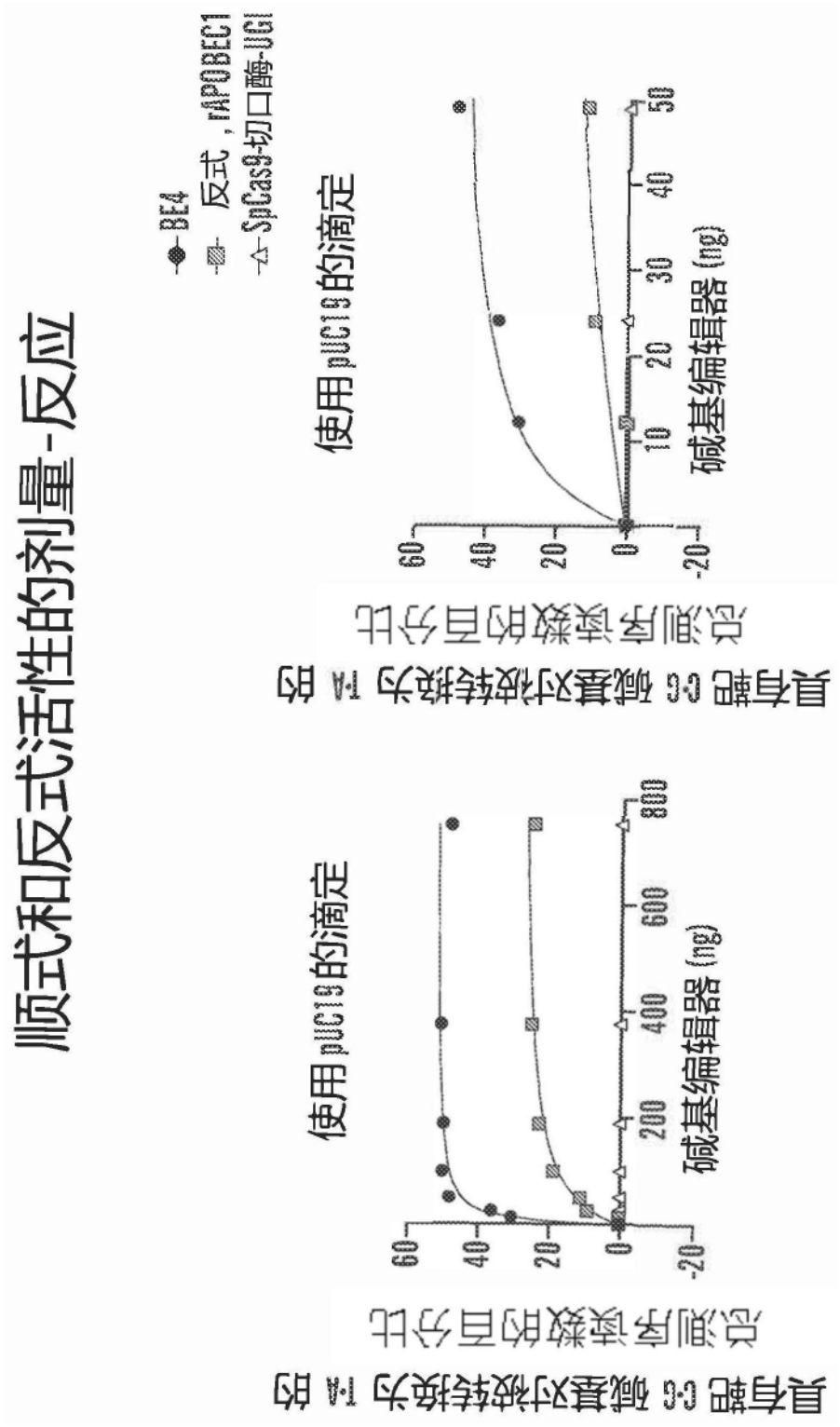

顺式和反式活性的剂量-反应
BE4
反式，rAPOBEC1
SpCas9-切口酶-UGI
使用pUC19的滴定
碱基编辑器 (ng)
总测序读数的百分比
具有靶C·G碱基对被转换为T·A的
使用pUC19的滴定
碱基编辑器 (ng)
总测序读数的百分比
具有靶C·G碱基对被转换为T·A的

图18

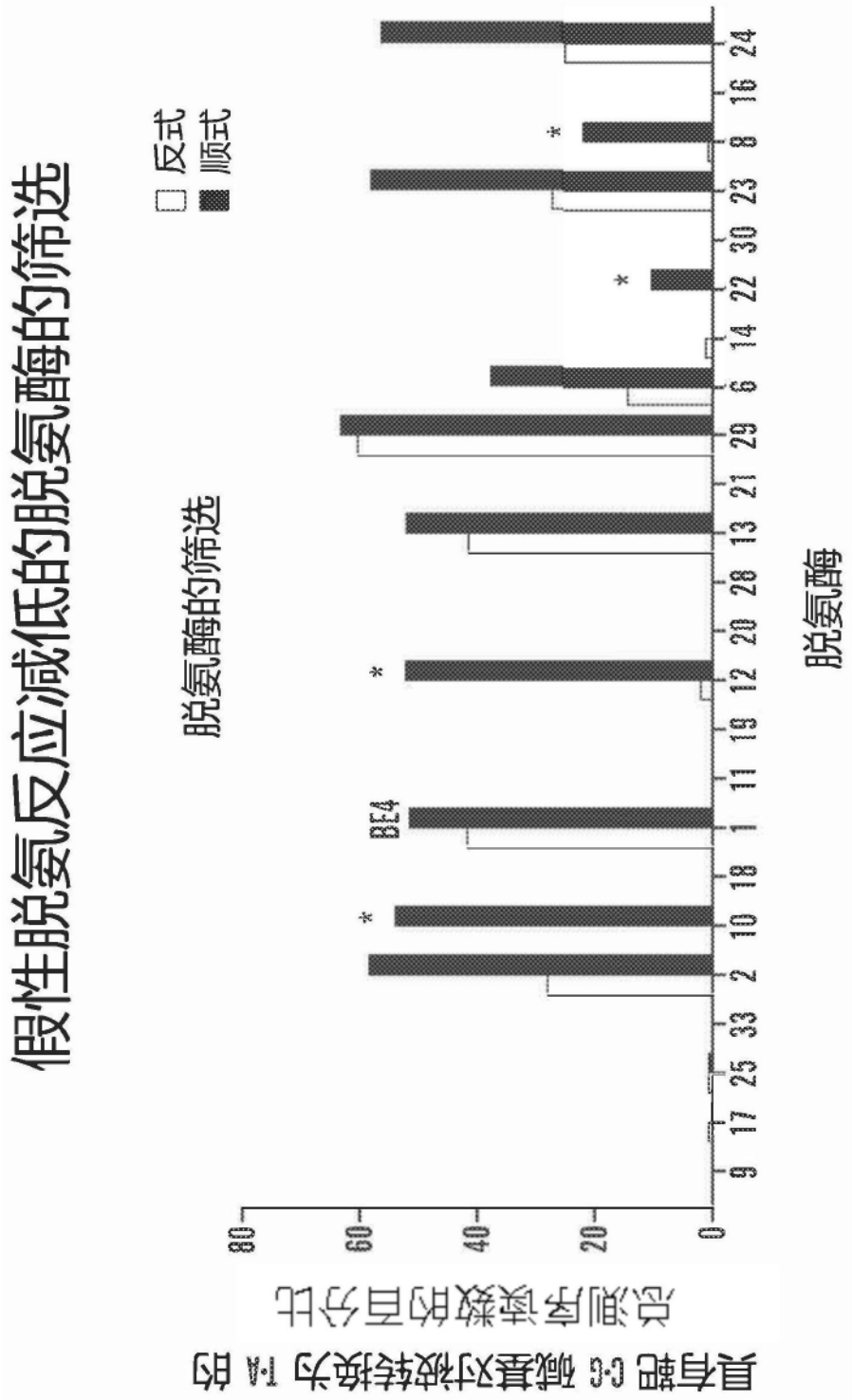

假性脱氨反应减低的脱氨酶的筛选
□ 反式
■ 顺式
脱氨酶的筛选
具有靶 C·C 碱基对被转换为 T·A 的
总测序读数的百分比
80
60
40
20
0
9 17 25 33 2 10 18 BE4 1 11 19 12 20 28 13 21 29 6 14 22 30 23 8 16 24
*
脱氨酶

图19

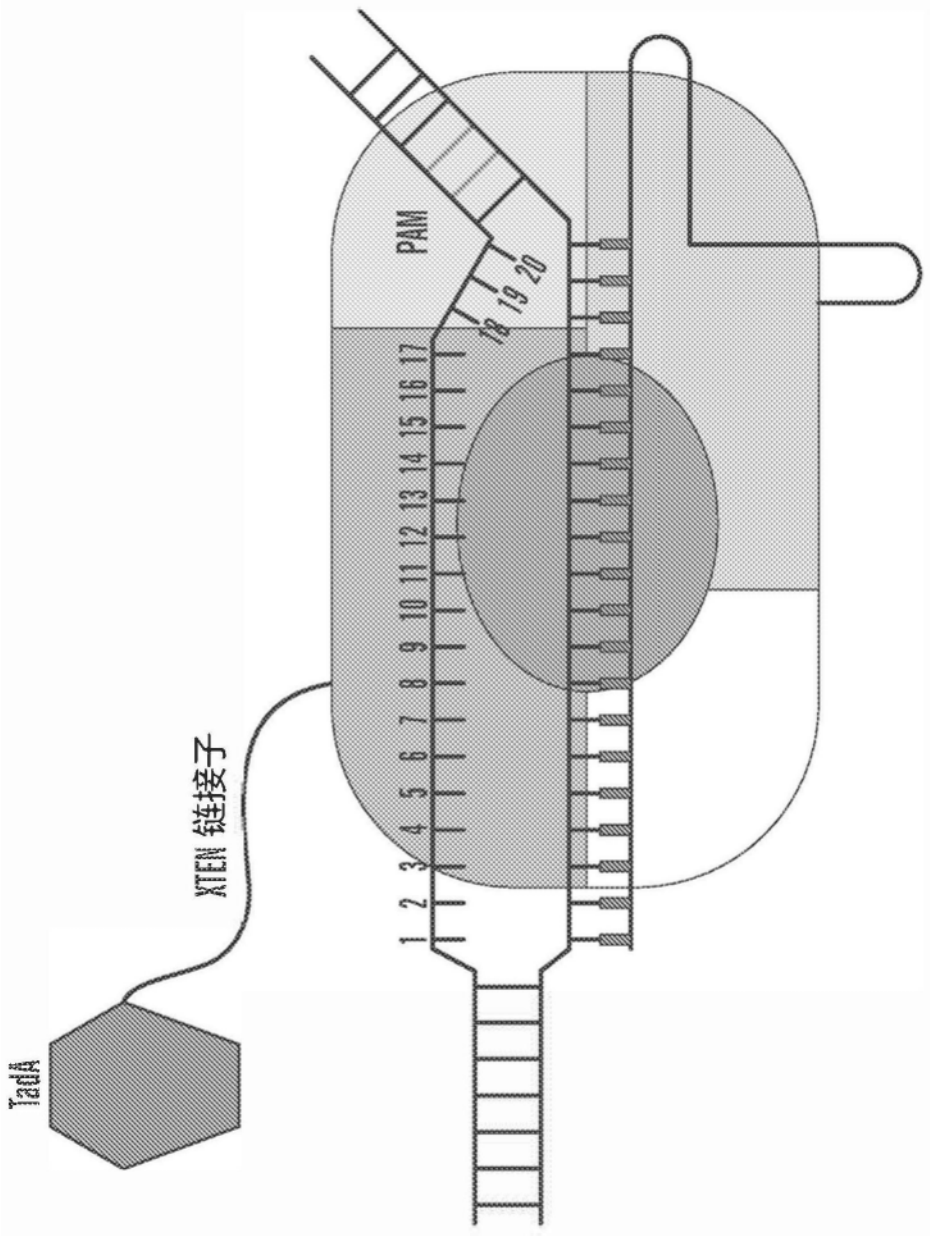


图20A

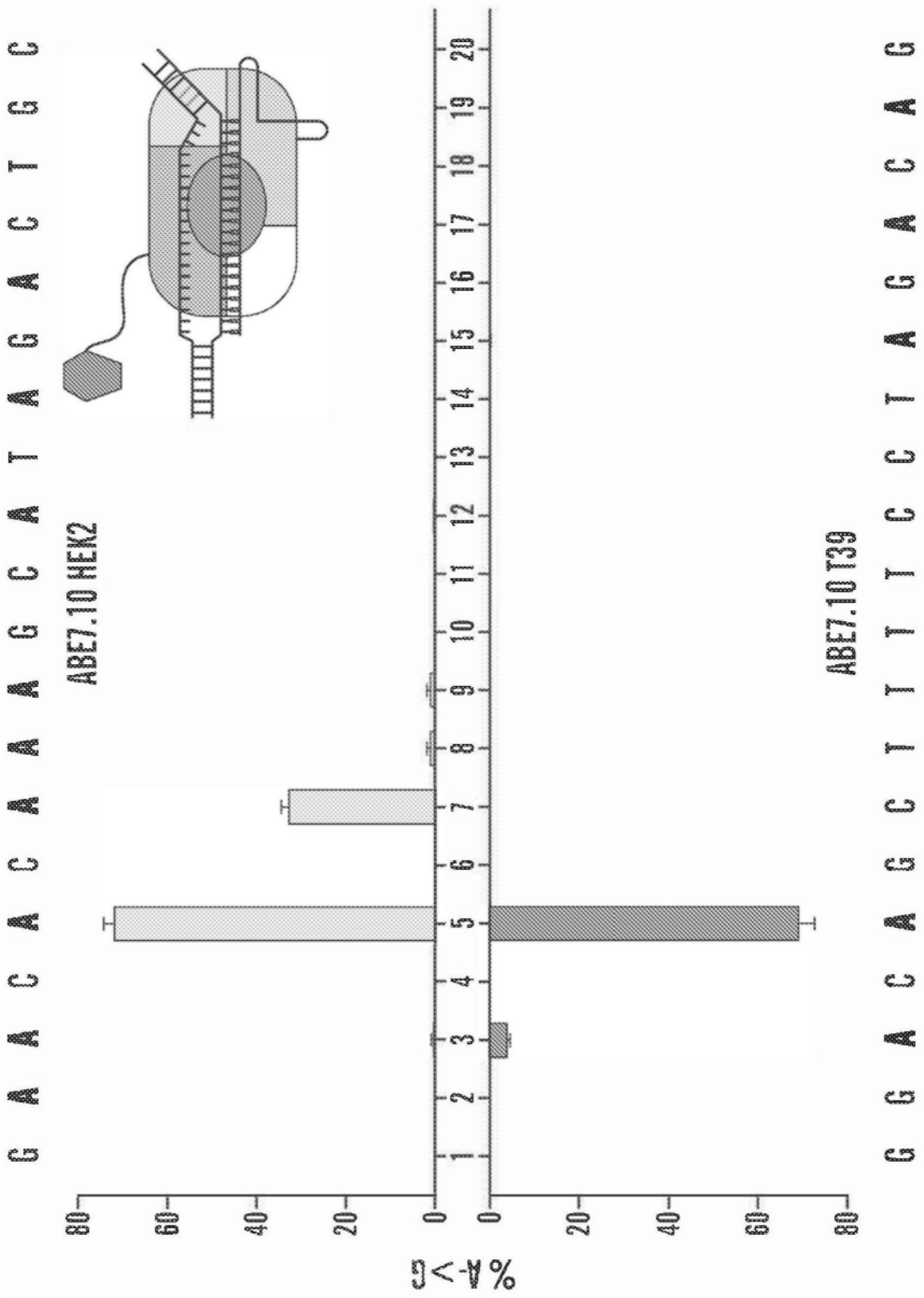

GAACAAAAGCCATAGCACTGCC
GGACGCTTTTTCCTAGCAACAG
ABE7.10 HEK2
ABE7.10 T39
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
80 60 40 20 0
0 20 40 60 80
%

图20B

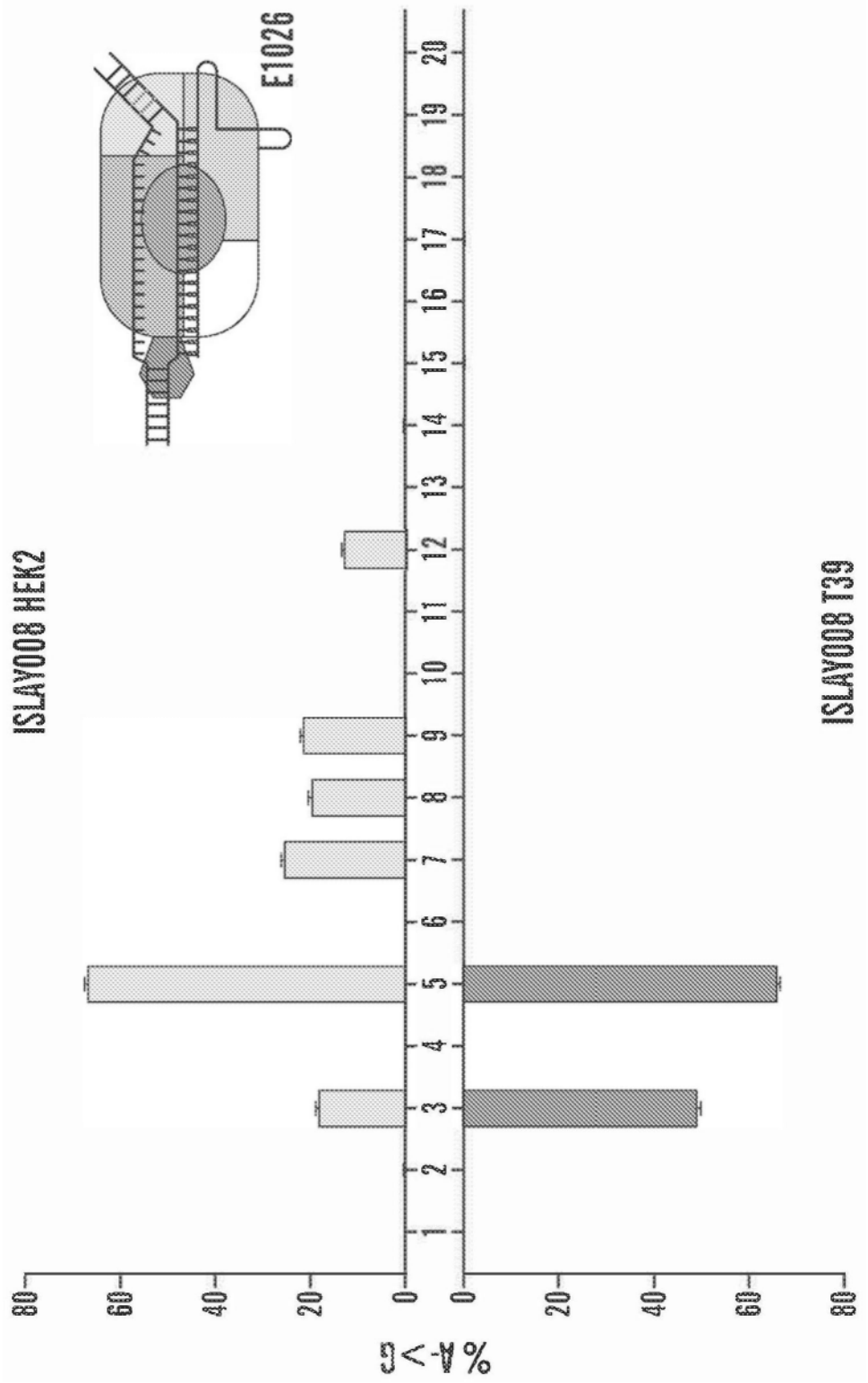

E1026
ISLAY008 HEK2
ISLAY008 T39
80
60
40
20
0
0
20
40
60
80
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
%

图20C

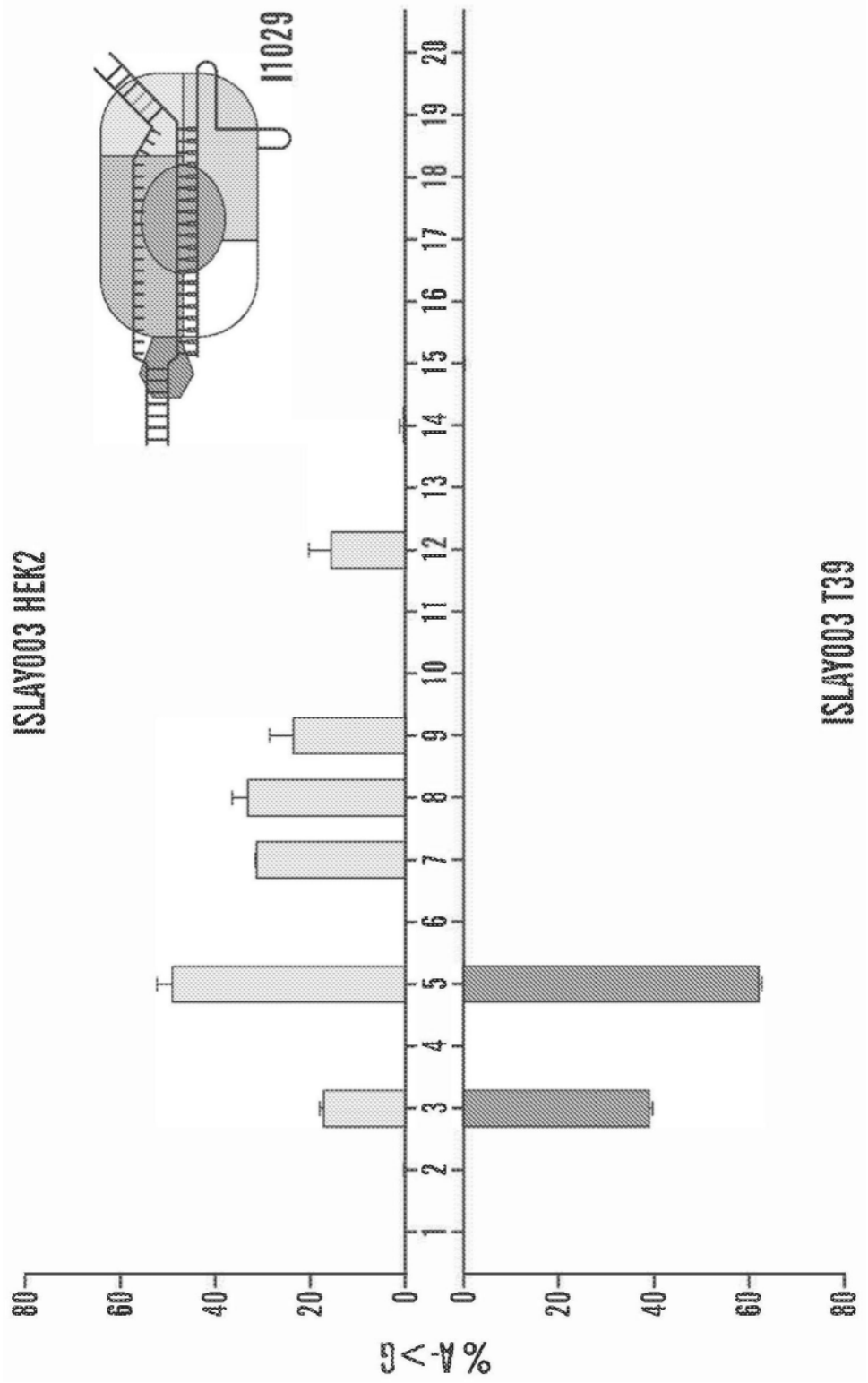

I1029
ISLAY003 HEK2
ISLAY003 T39
80
60
40
20
0
0
20
40
60
80
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20

图20D

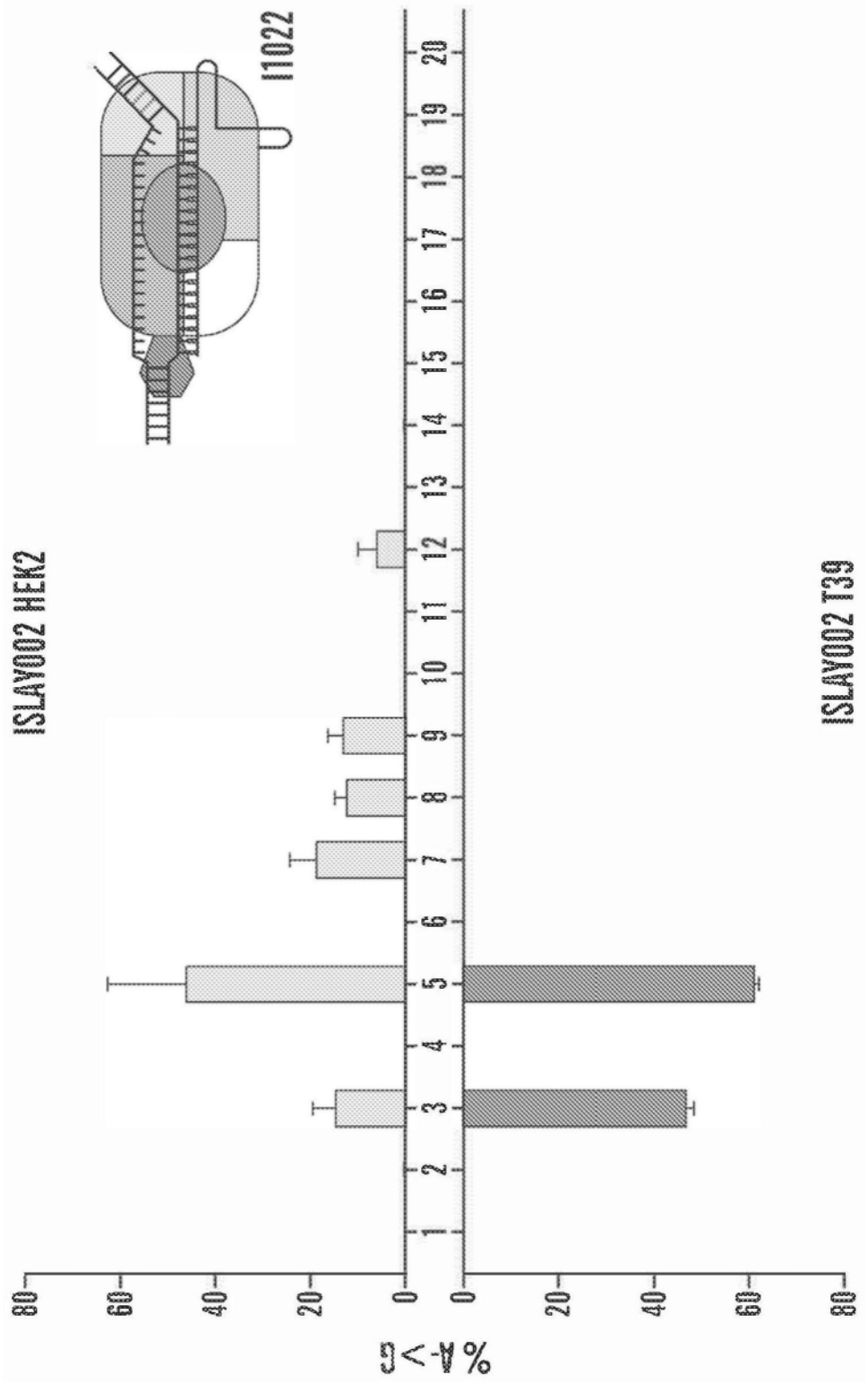

I1022
ISLAY002 HEK2
ISLAY002 T39
80
60
40
20
0
0
20
40
60
80
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20

图20E

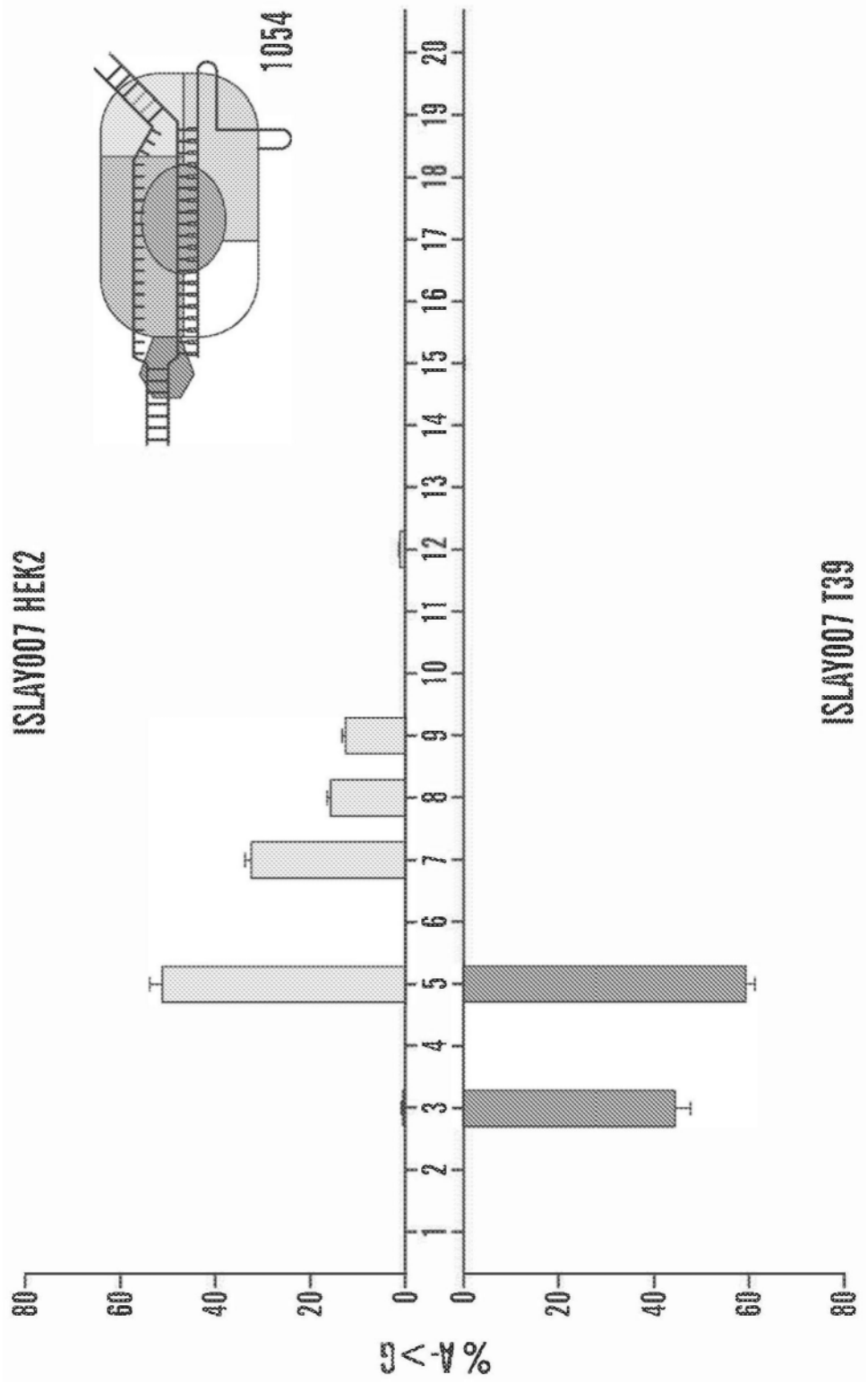

1054
ISLAY007 HEK2
ISLAY007 T39
80
60
40
20
0
0
20
40
60
80
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20

图20F

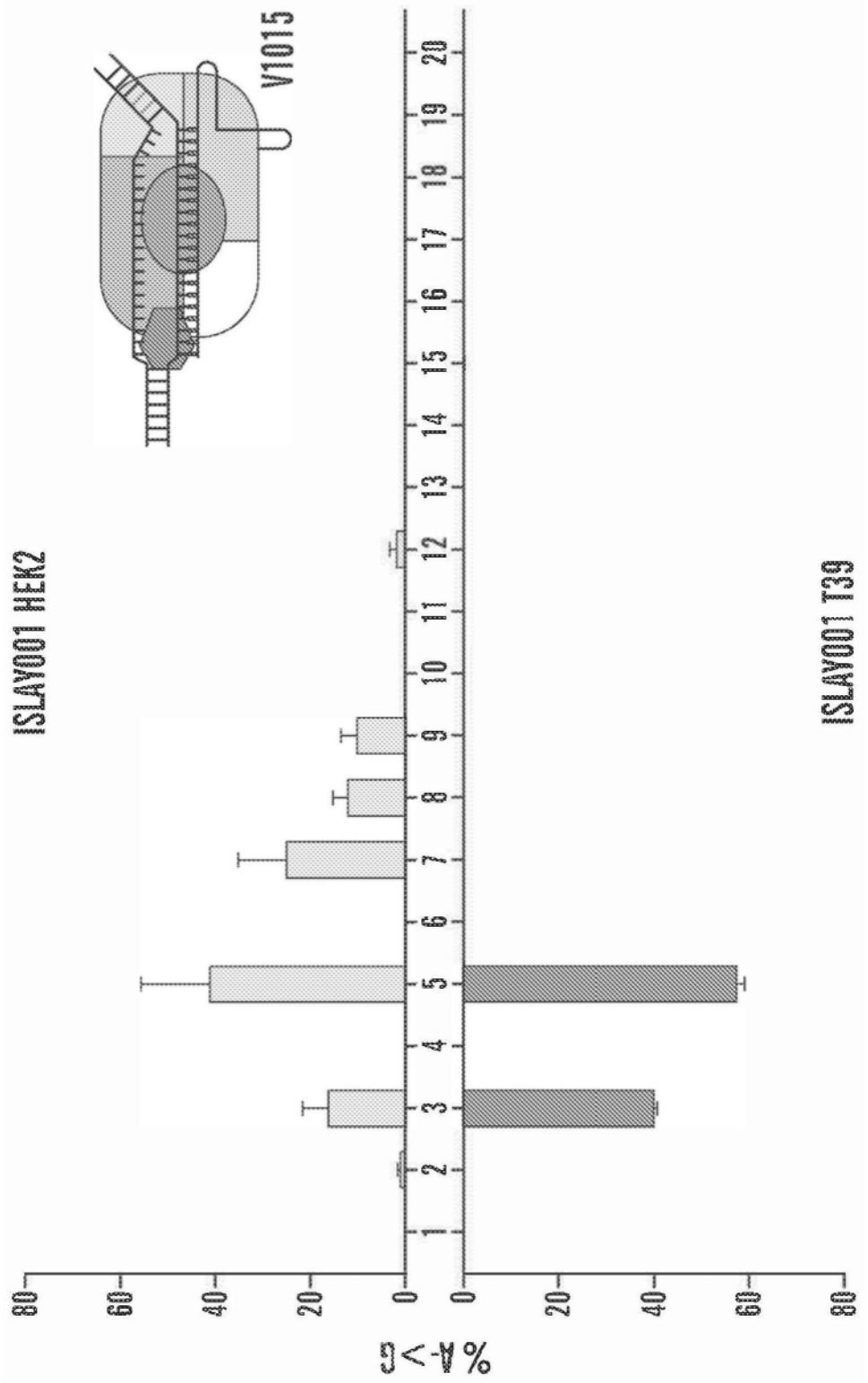

V1015
ISLAY001 HEK2
ISLAY001 T39
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
80
60
40
20
0
0
20
40
60
80
%

图20G

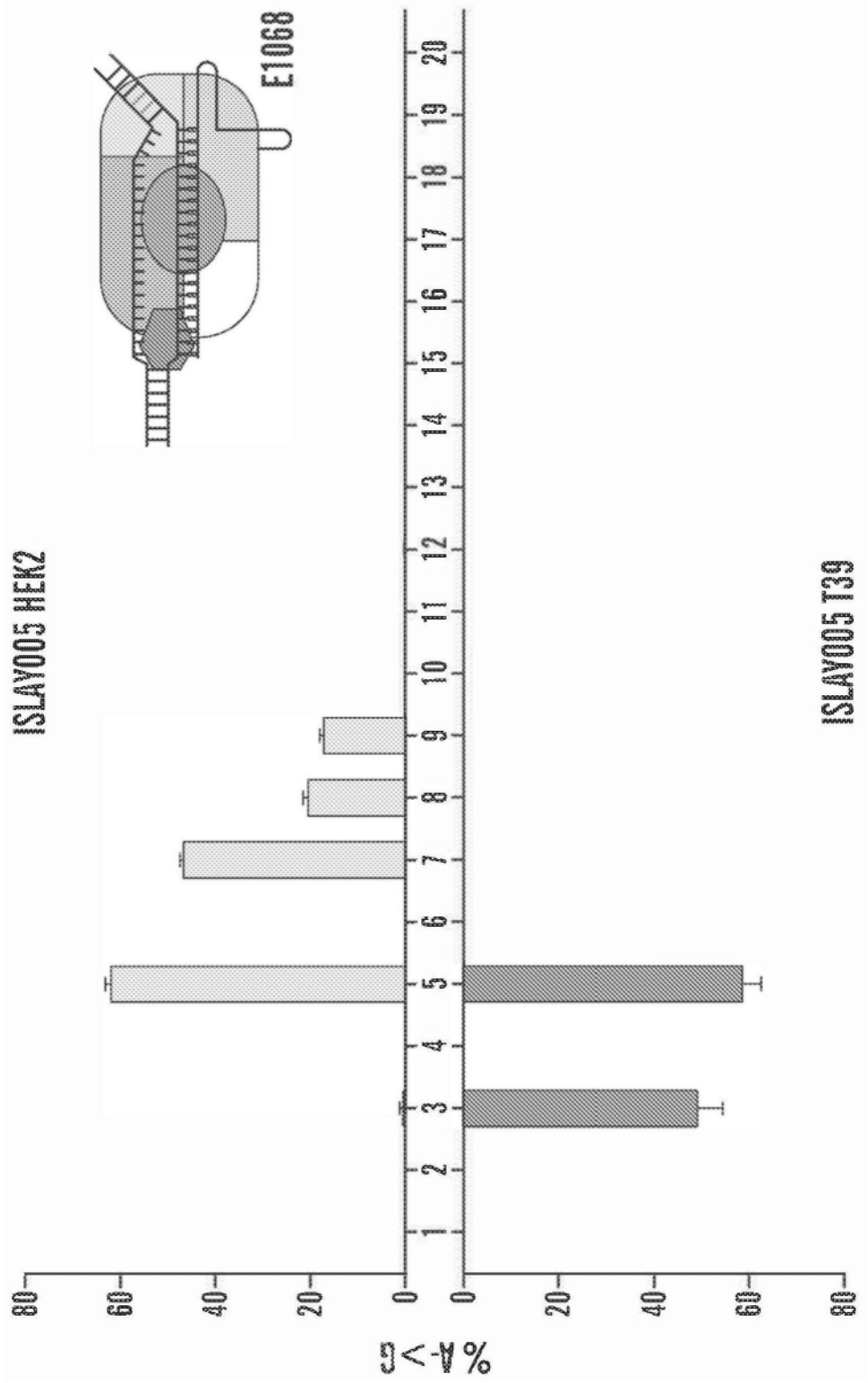

E1068
ISLAY005 HEK2
ISLAY005 T39
80
60
40
20
0
0
20
40
60
80
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20

图20H

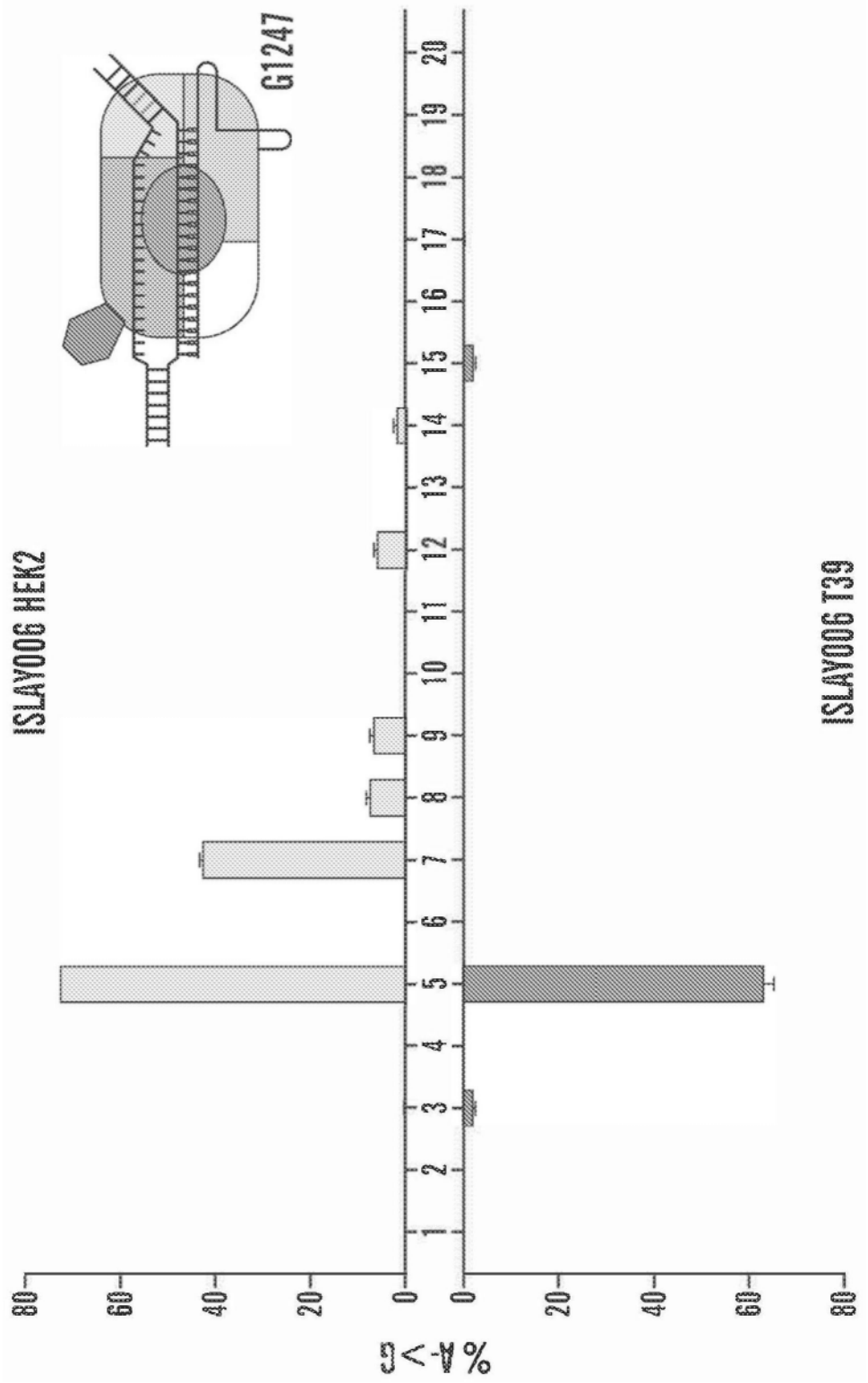

G1247
ISLAY006 HEK2
ISLAY006 T39
80
60
40
20
0
0
20
40
60
80
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
%

图20I

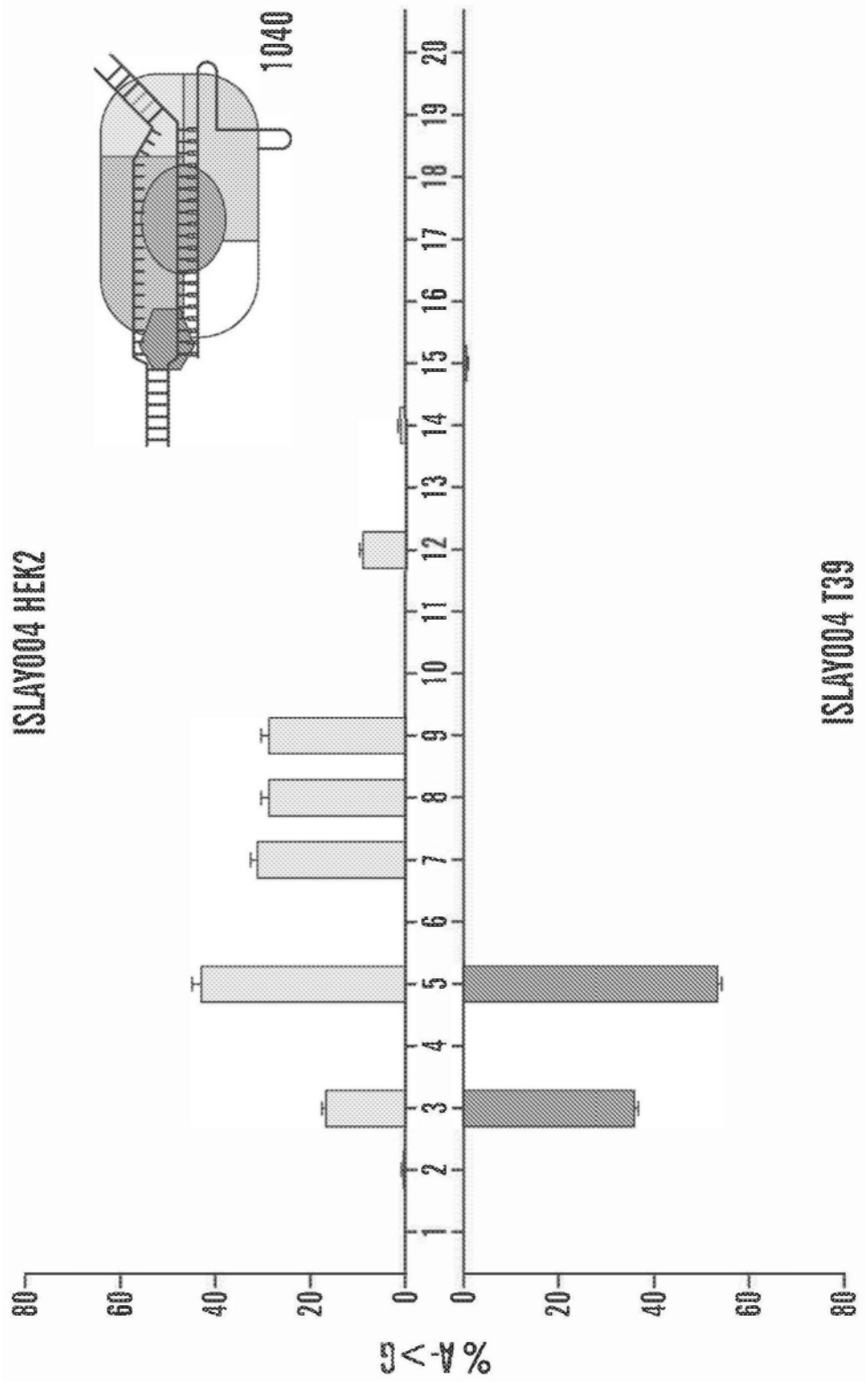

1040
ISLAY004 HEK2
ISLAY004 T39
80
60
40
20
0
0
20
40
60
80
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20

图20J

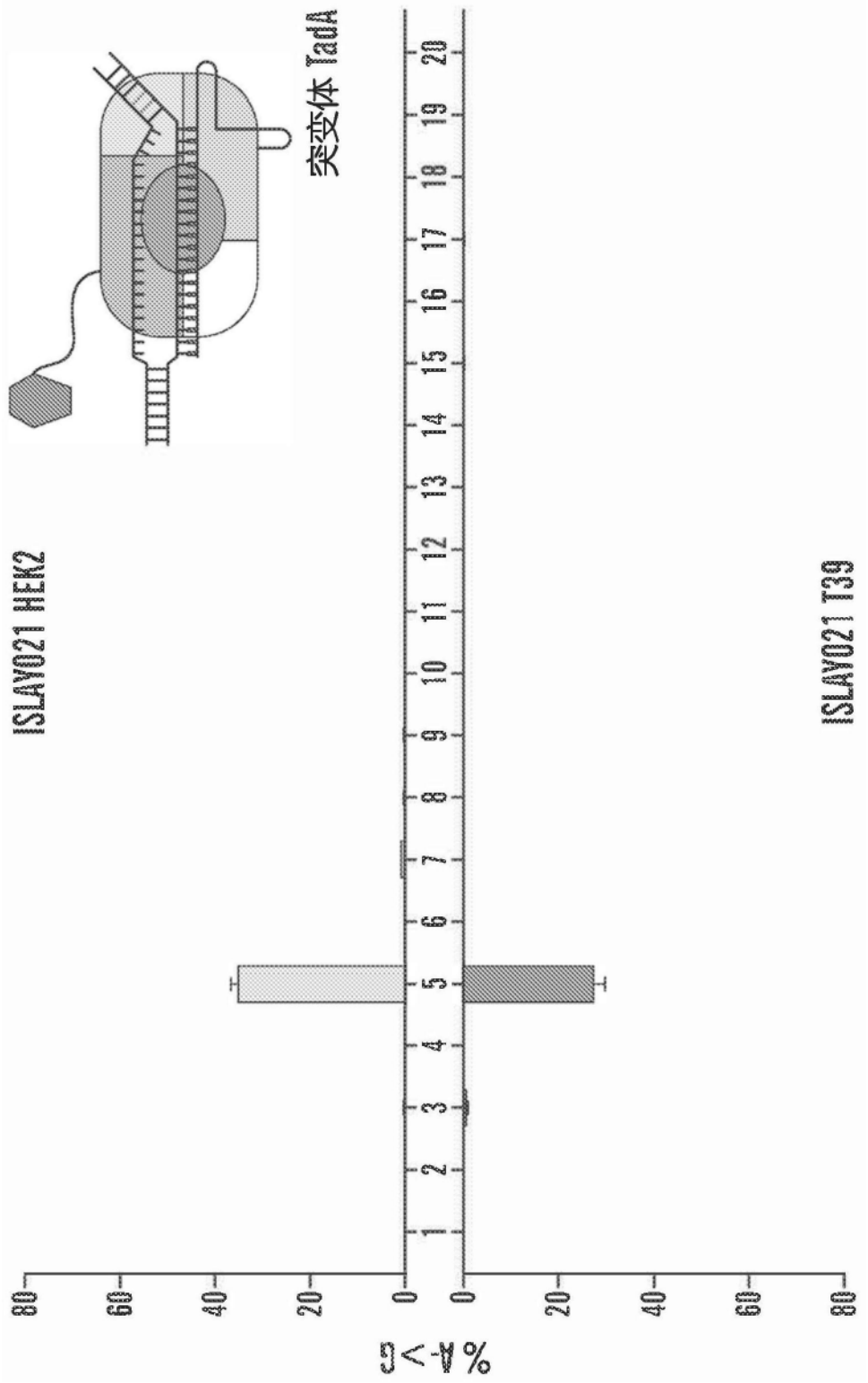

突变体 TadA
ISLAY021 HEK2
ISLAY021 T39
编辑 %
80
60
40
20
0
0
20
40
60
80
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20

图20K

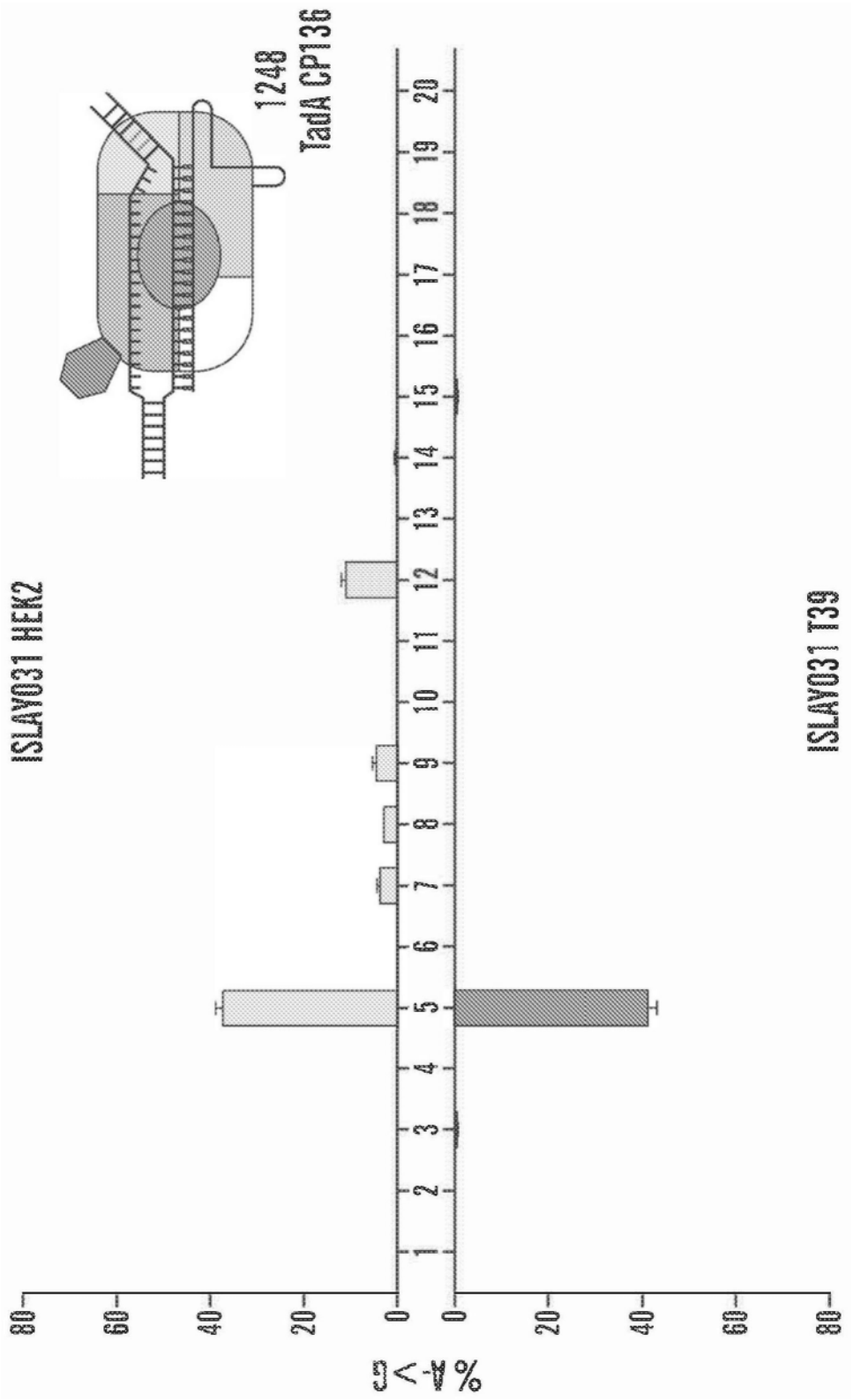

1248
TadA CP136
ISLAY031 HEK2
ISLAY031 T39
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
80
60
40
20
0
0
20
40
60
80
%

图20L

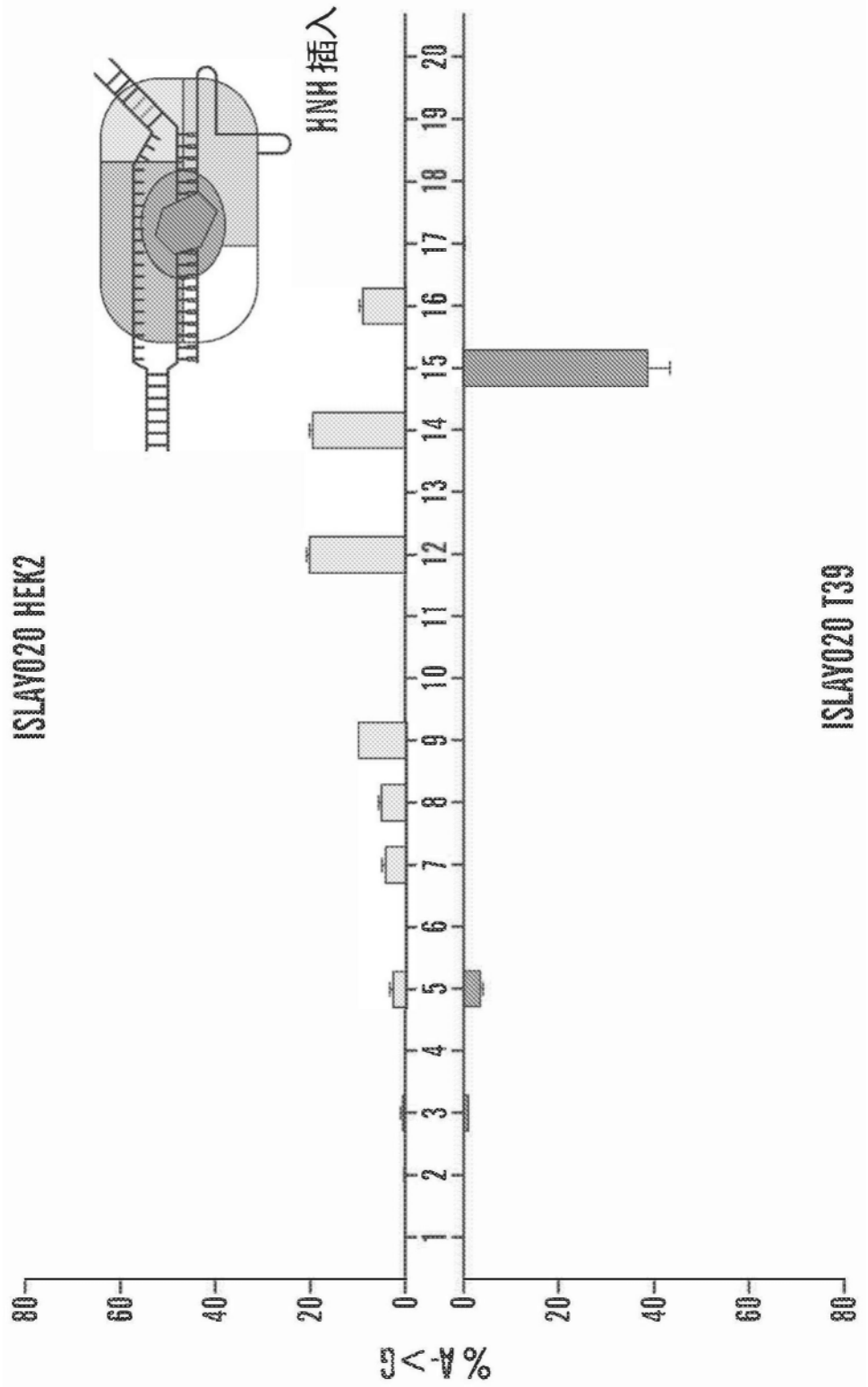

HNH插入
ISLAY020 HEK2
ISLAY020 T39
80
60
40
20
0
0
20
40
60
80
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20

图20M

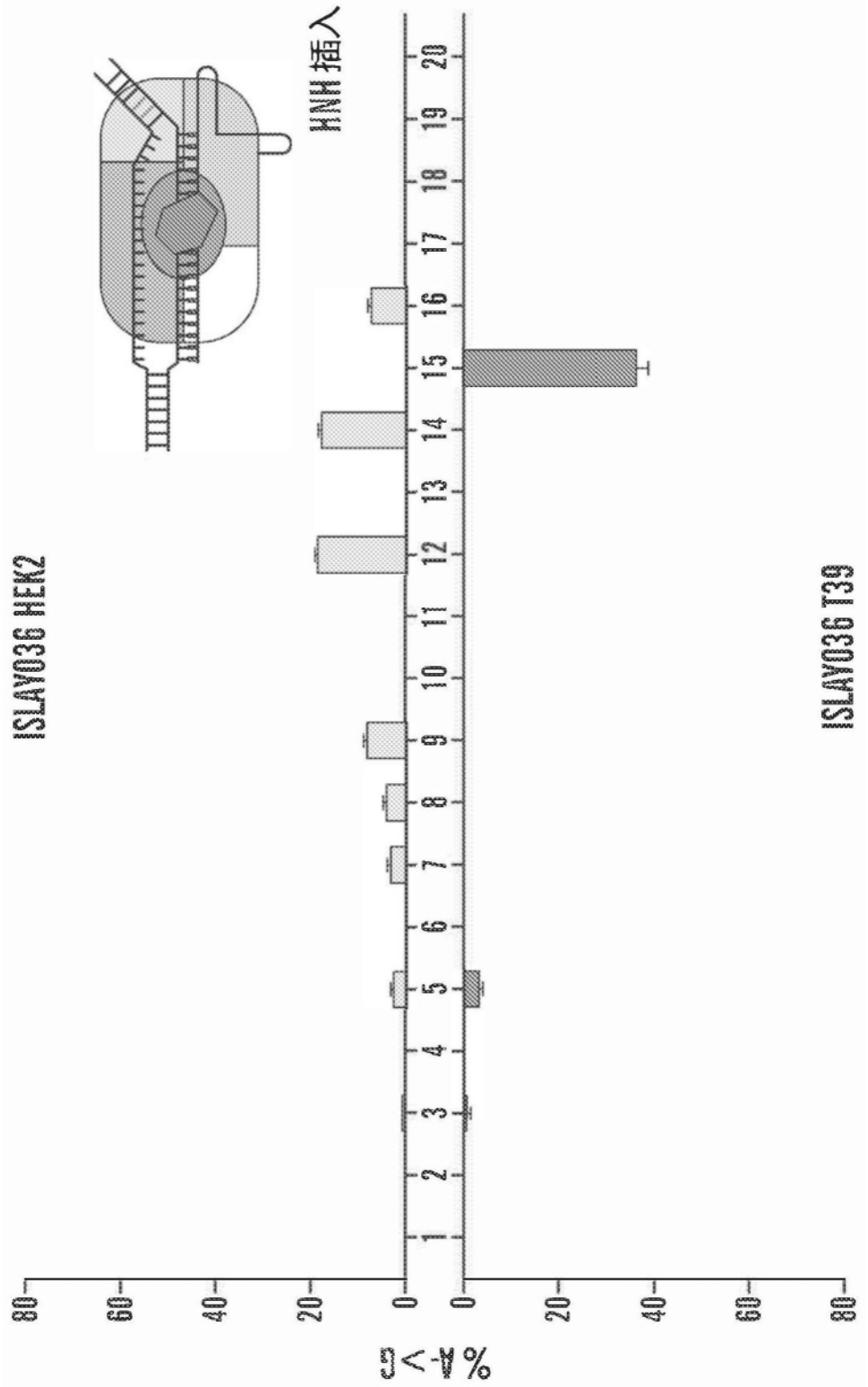

HNH插入
ISLAY036 HEK2
ISLAY036 T39
80
60
40
20
0
0
20
40
60
80
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20

图20N

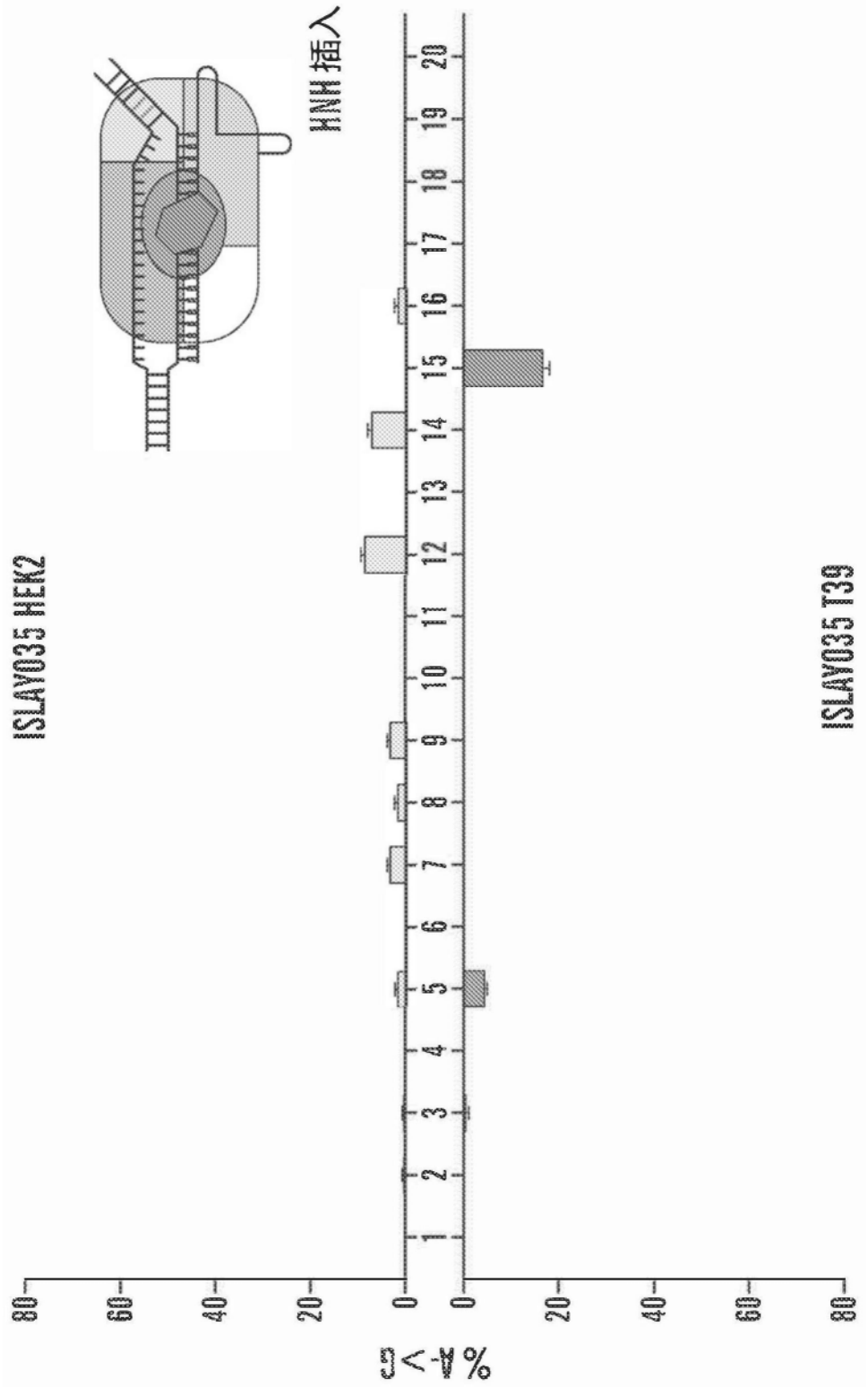

HNH插入
ISLAY035 HEK2
ISLAY035 T39
80
60
40
20
0
0
20
40
60
80
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
%插入

图200

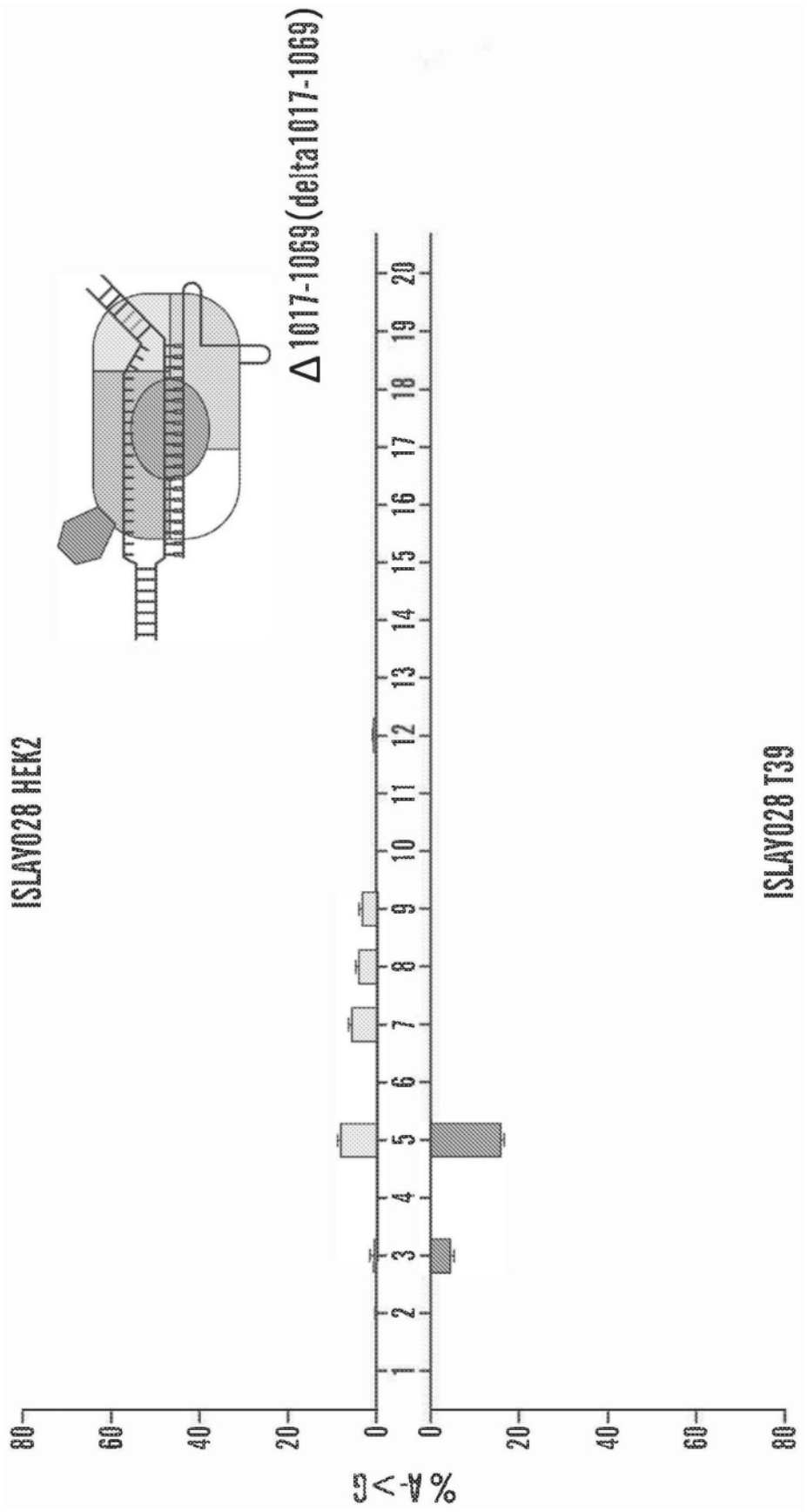

Δ1017-1069(delta1017-1069)
ISLAY028 HEK2
ISLAY028 T39
ヒット%
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
0
20
40
60
80

图20P

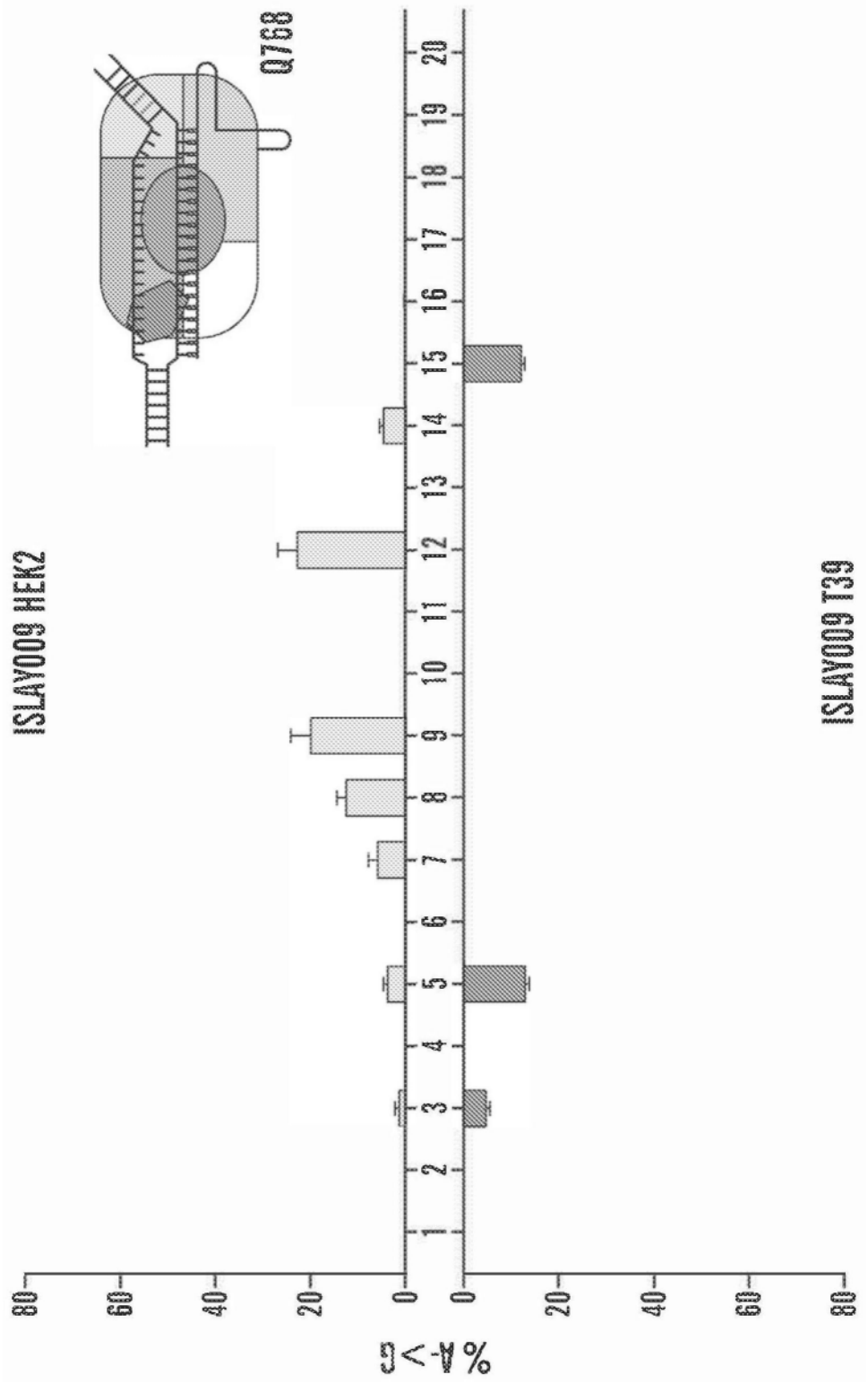

Q768
ISLAY009 HEK2
ISLAY009 T39
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
80
60
40
20
0
0
20
40
60
80
%

图20Q

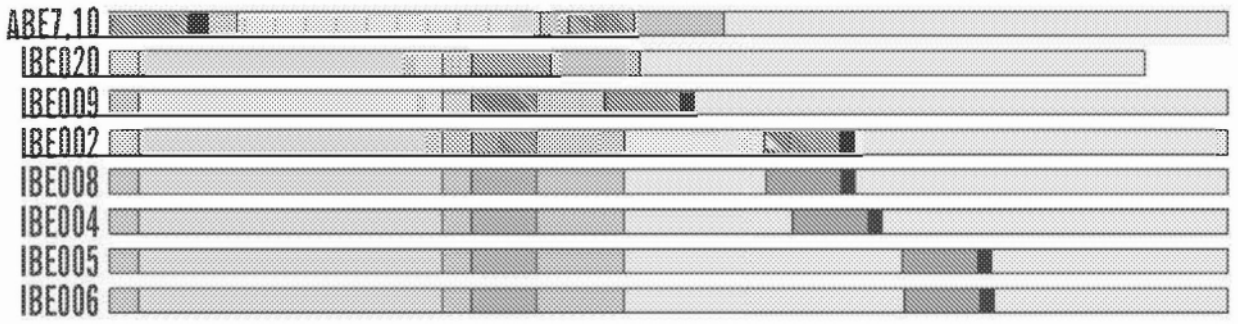

ABE7.10
IBE020
IBE009
IBE002
IBE008
IBE004
IBE005
IBE006

图21A

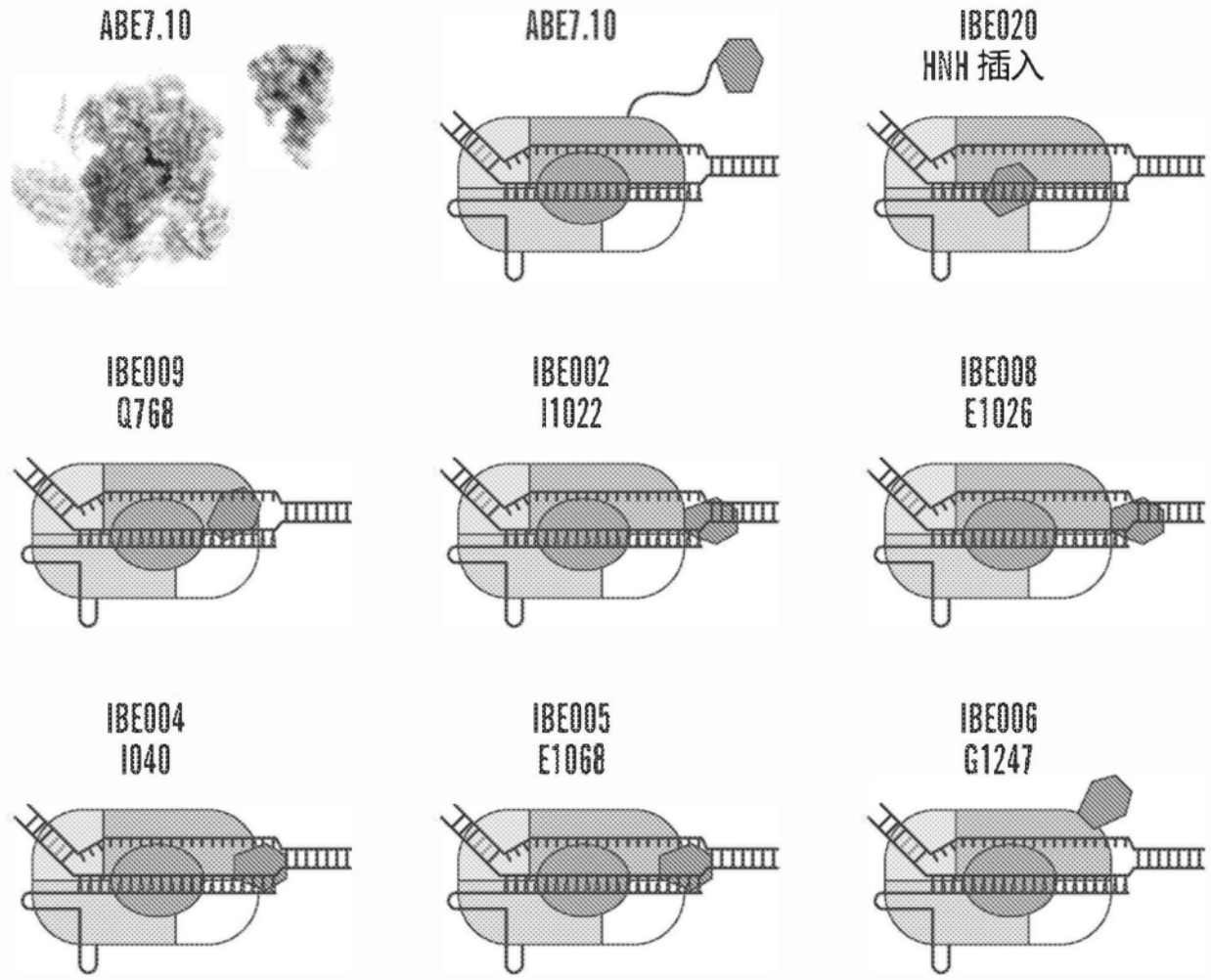

ABE7.10
ABE7.10
IBE020
HNH 插入
IBE009
Q768
IBE002
I1022
IBE008
E1026
IBE004
I040
IBE005
E1068
IBE006
G1247

图21B

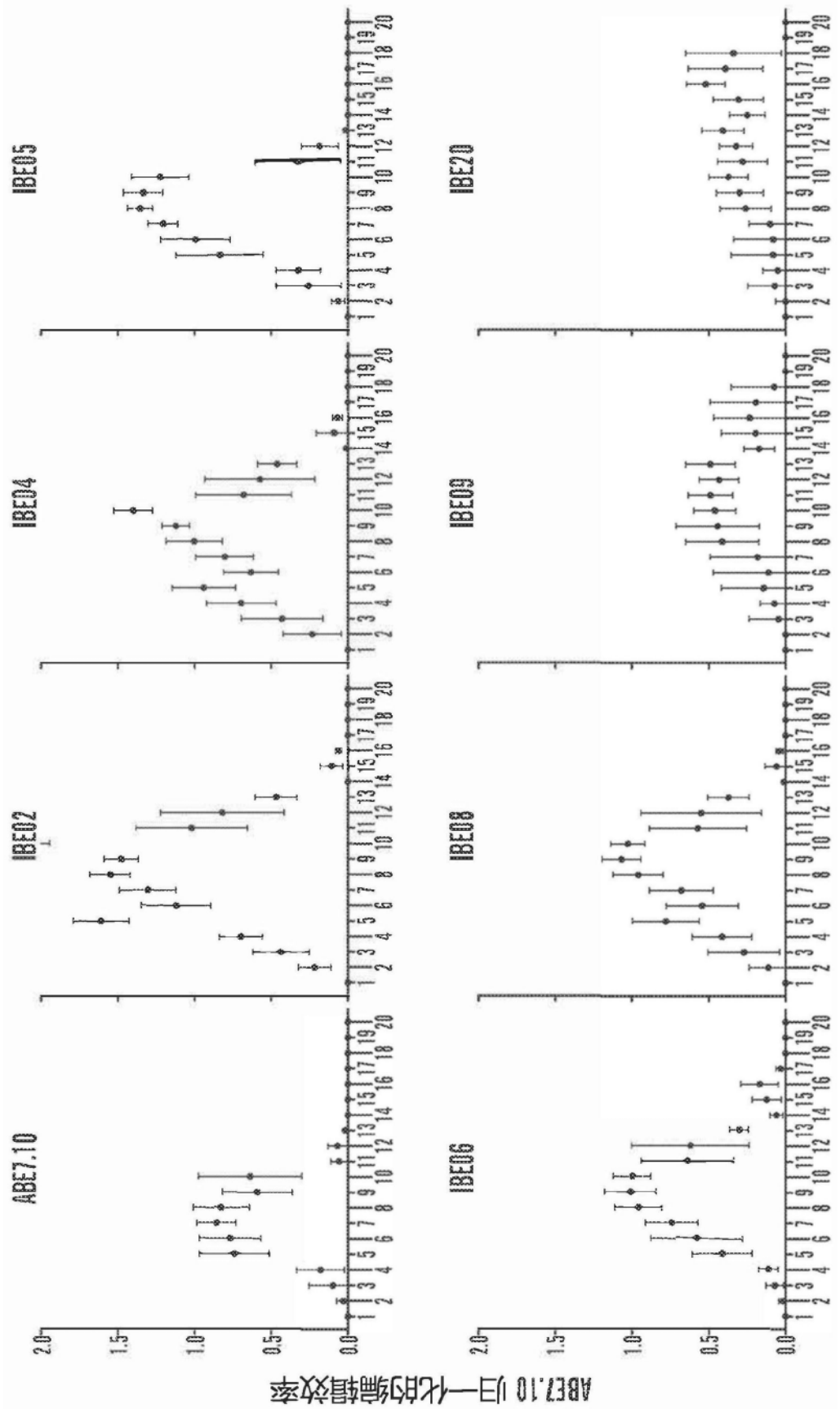

ABE7.10
IBE02
IBE04
IBE05
IBE06
IBE08
IBE09
IBE20
ABE7.10 归一化的编辑效率

图22A

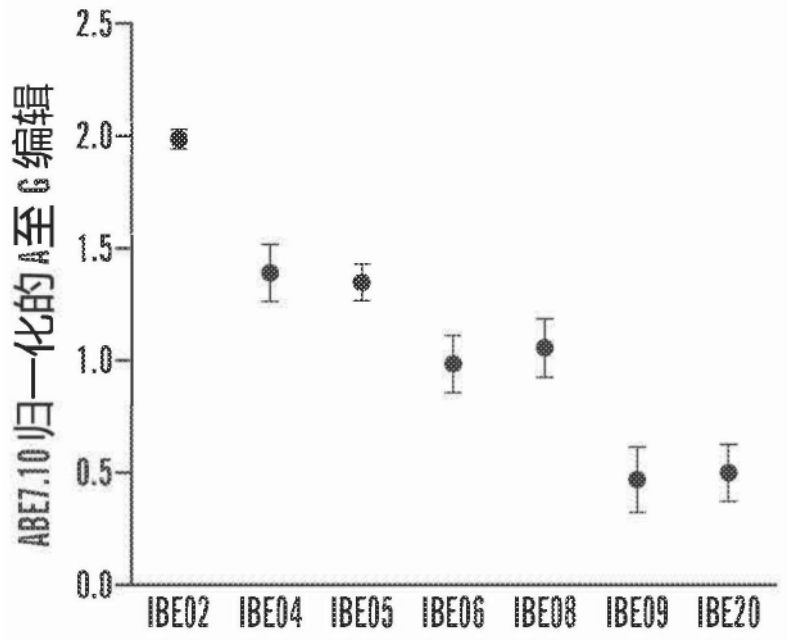

ABE7.10归一化的A至G编辑
2.5
2.0
1.5
1.0
0.5
0.0
IBE02
IBE04
IBE05
IBE06
IBE08
IBE09
IBE20

图22B

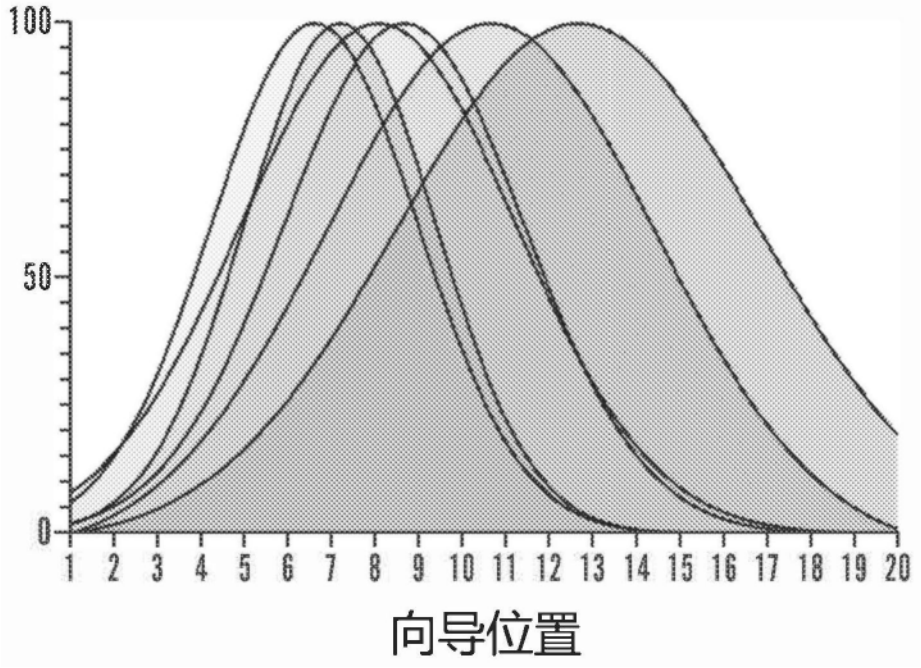

100
50
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
向导位置

图22C

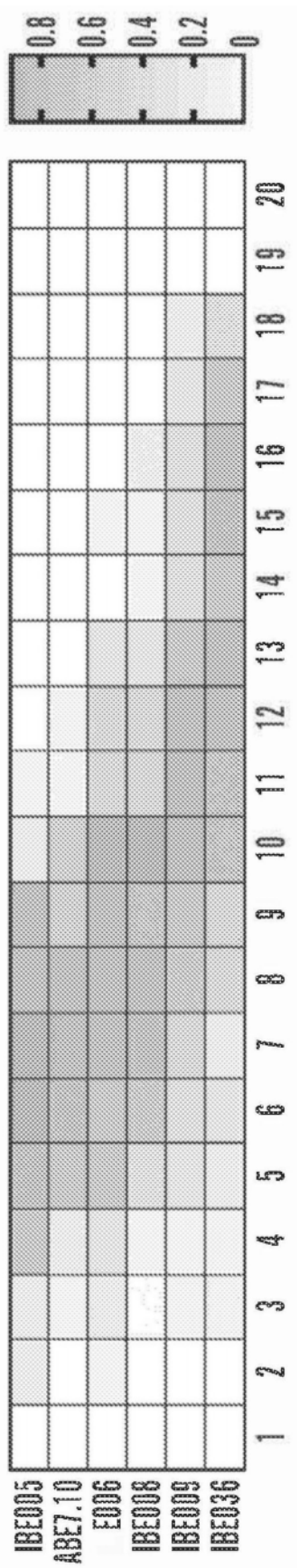

0.8
0.6
0.4
0.2
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
IBE005
ABE7.10
E006
IBE008
IBE009
IBE036

图22D

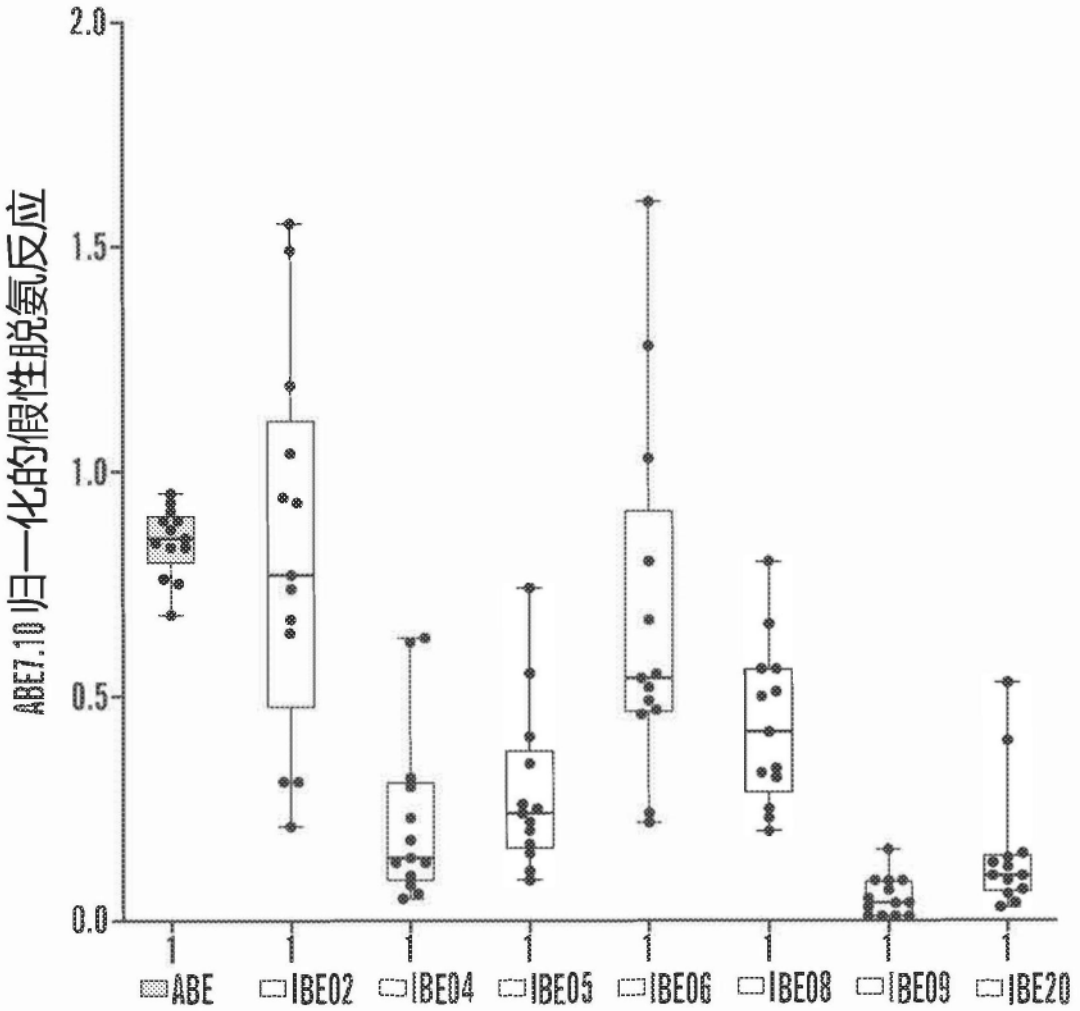

ABE7.10归一化的假性脱氨反应
2.0
1.5
1.0
0.5
0.0
1
ABE
IBE02
IBE04
IBE05
IBE06
IBE08
IBE09
IBE20

图23

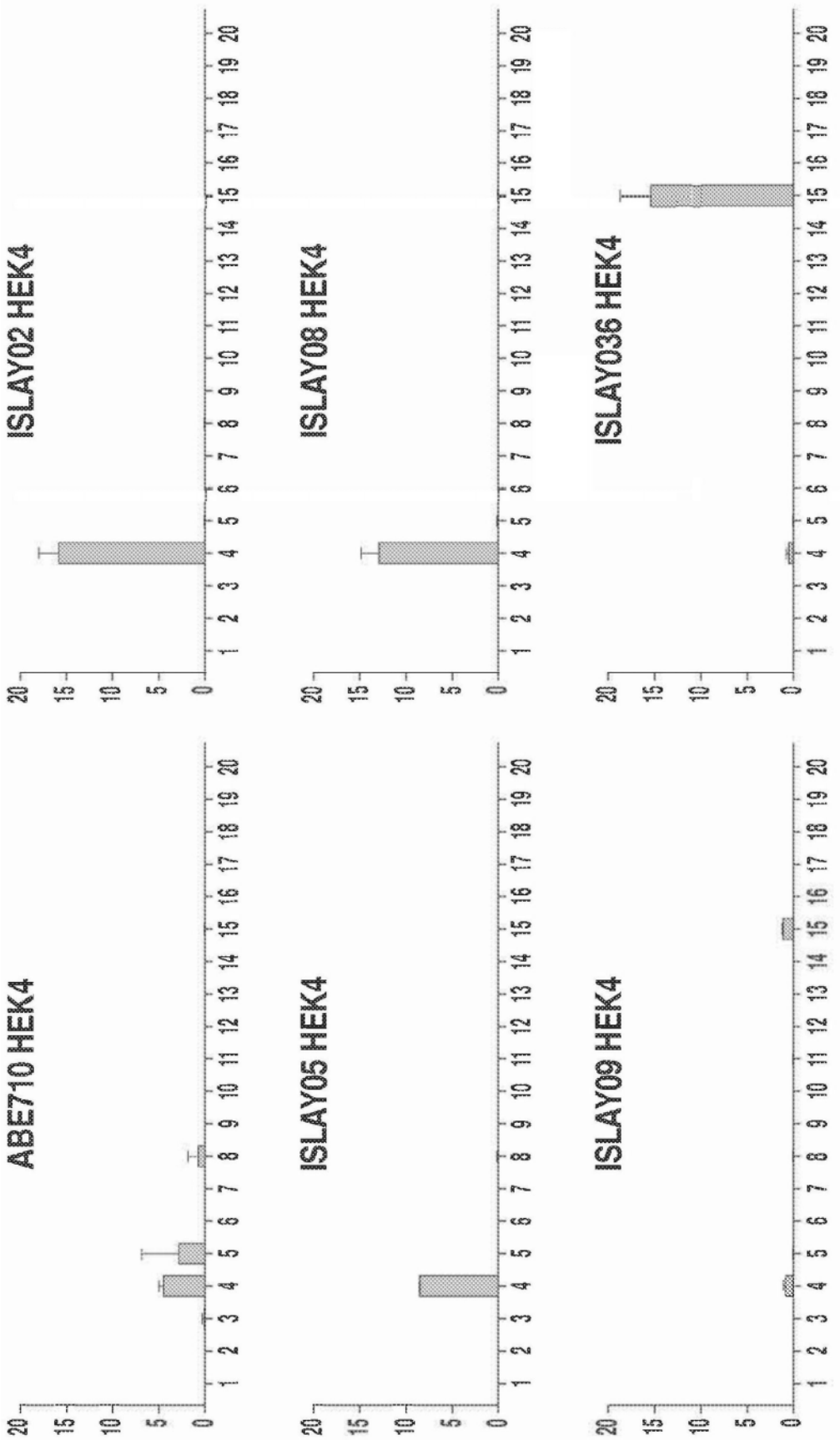

ISLAY02 HEK4
ISLAY08 HEK4
ISLAY036 HEK4
ABE710 HEK4
ISLAY05 HEK4
ISLAY09 HEK4

图24A

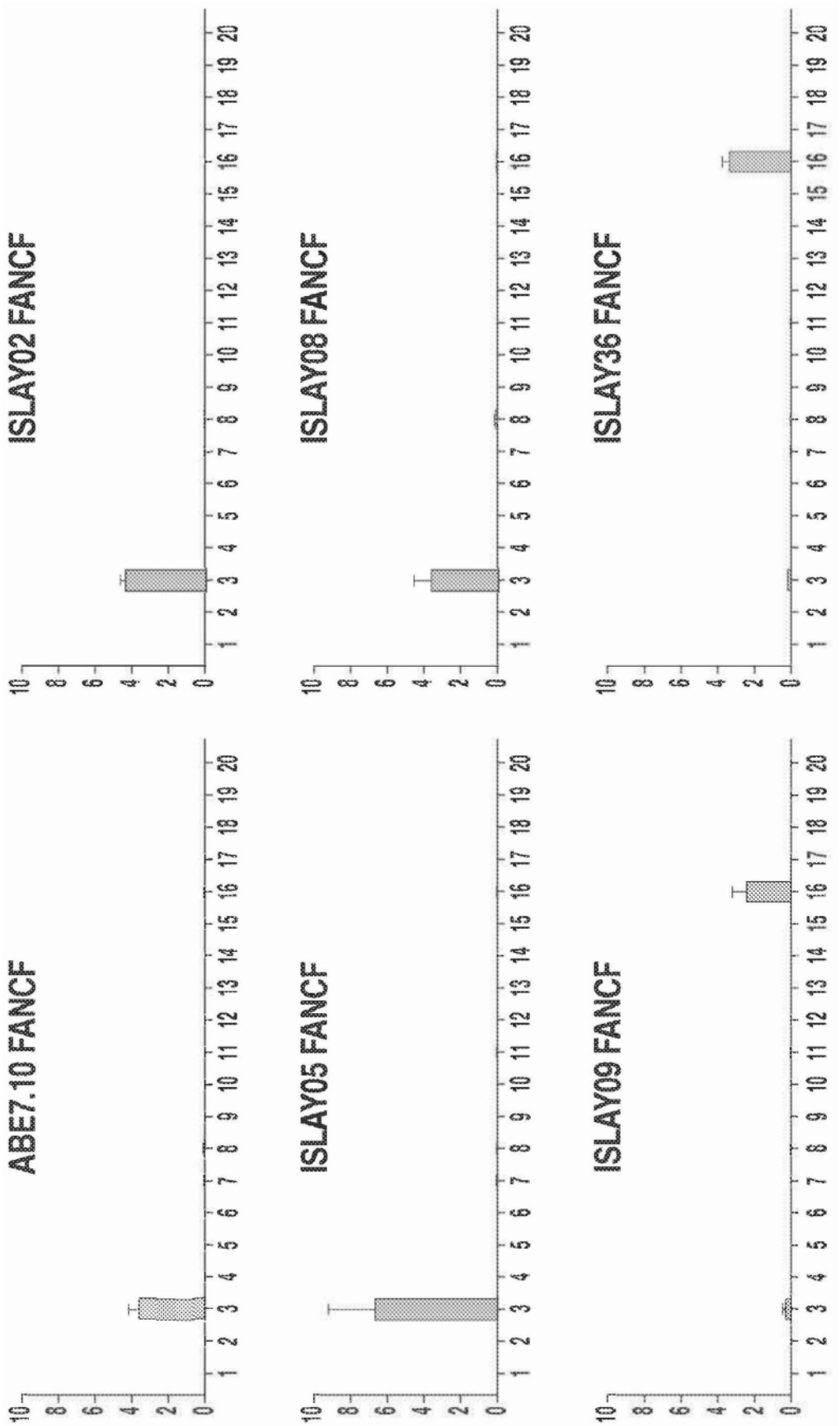

ISLAY02 FANCF
ISLAY08 FANCF
ISLAY36 FANCF
ABE7.10 FANCF
ISLAY05 FANCF
ISLAY09 FANCF

图24B

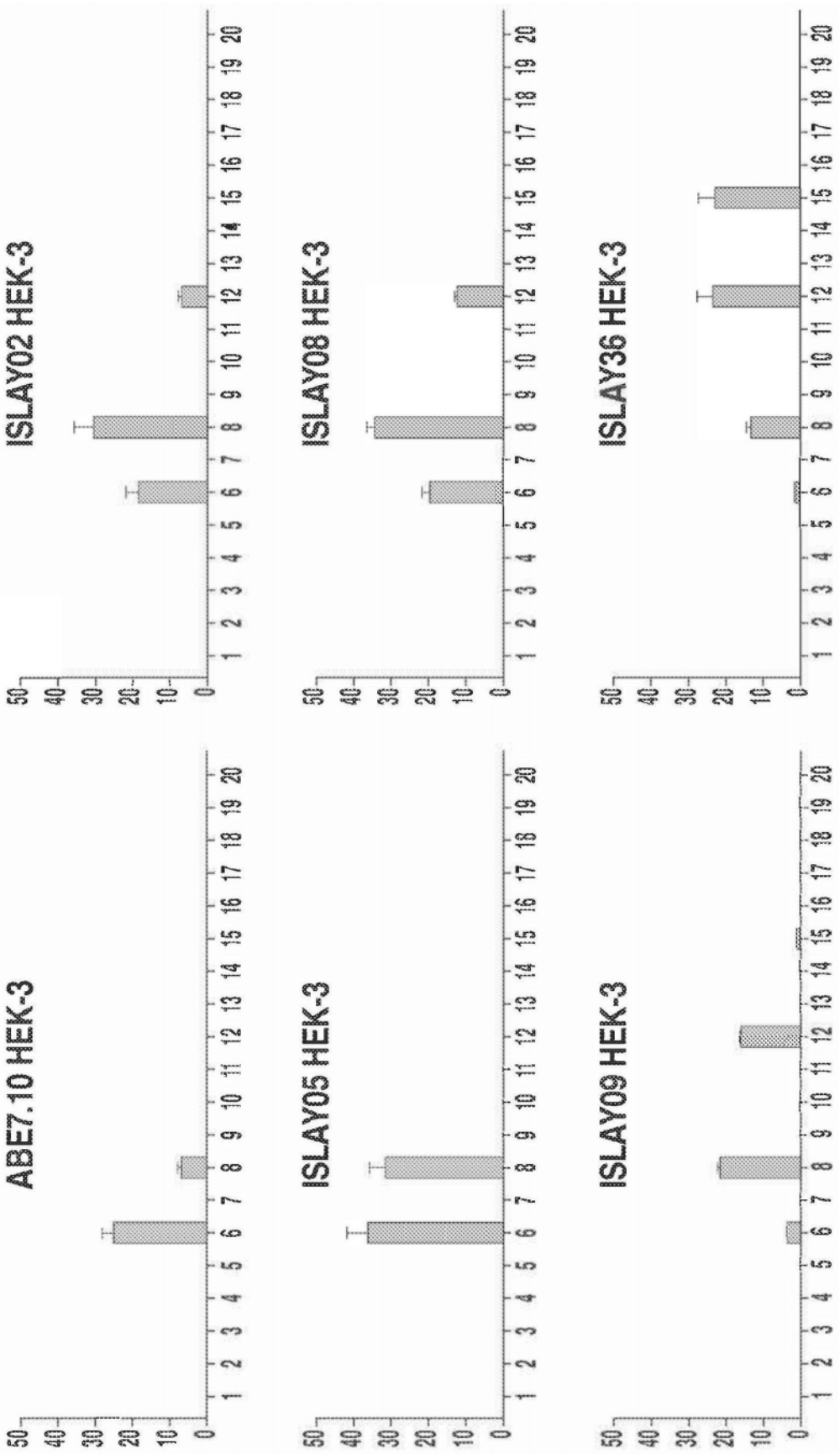

ISLAY02 HEK-3
ISLAY08 HEK-3
ISLAY36 HEK-3
ABE7.10 HEK-3
ISLAY05 HEK-3
ISLAY09 HEK-3

图24C

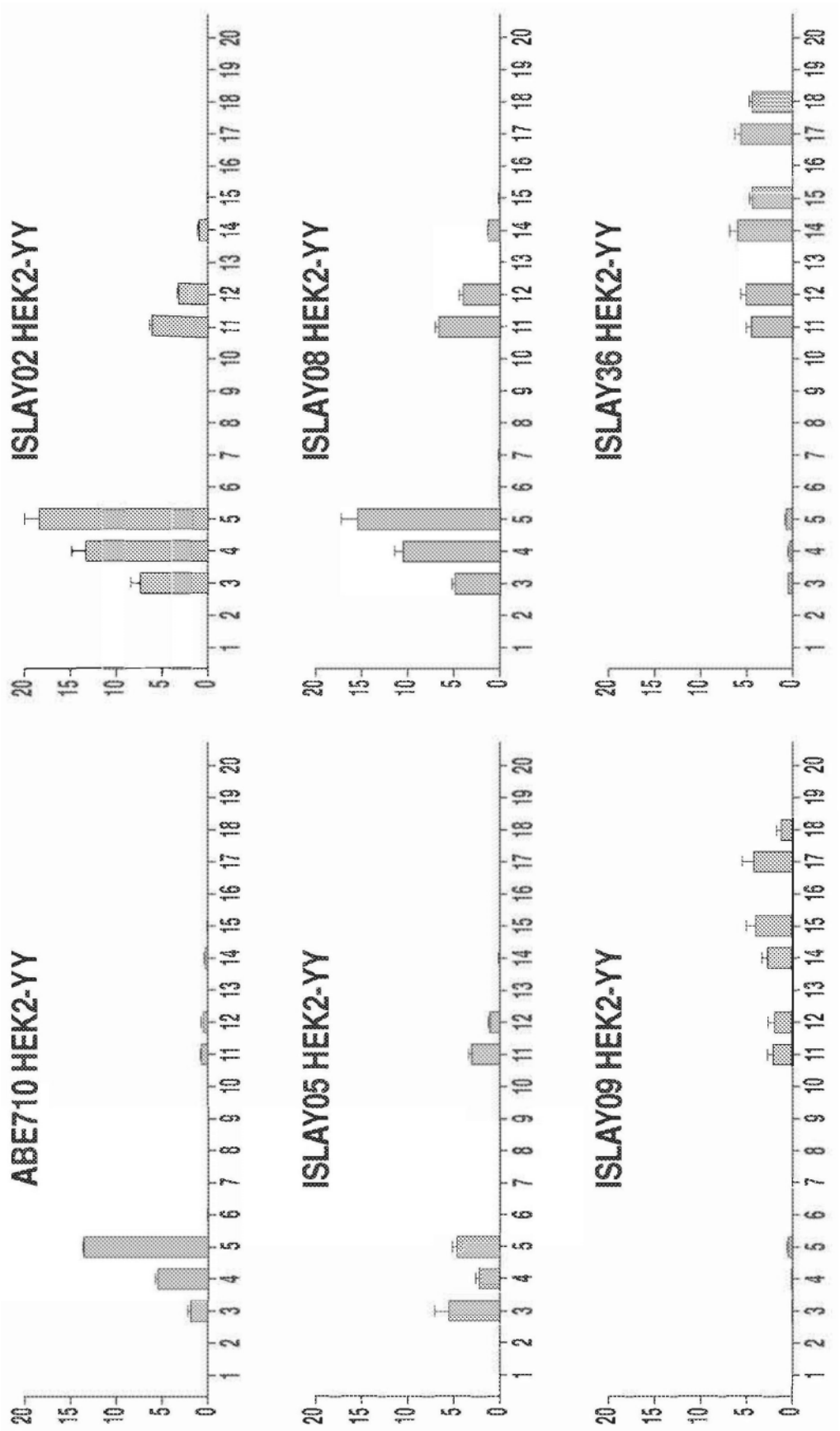

ISLAY02 HEK2-YY
ISLAY08 HEK2-YY
ISLAY36 HEK2-YY
ABE710 HEK2-YY
ISLAY05 HEK2-YY
ISLAY09 HEK2-YY

图24D

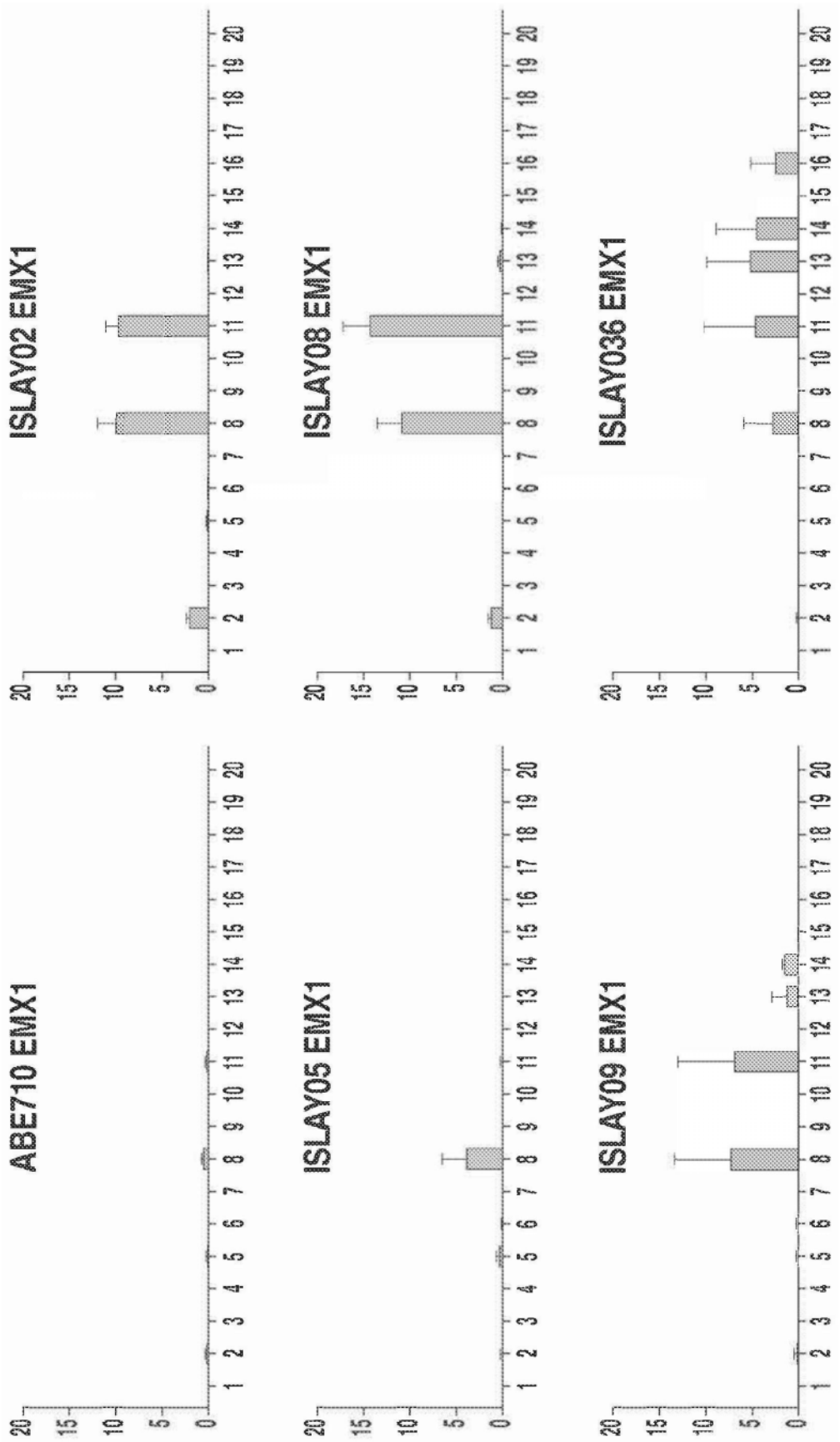

ISLAY02 EMX1
ISLAY08 EMX1
ISLAY036 EMX1
ABE710 EMX1
ISLAY05 EMX1
ISLAY09 EMX1

图24E

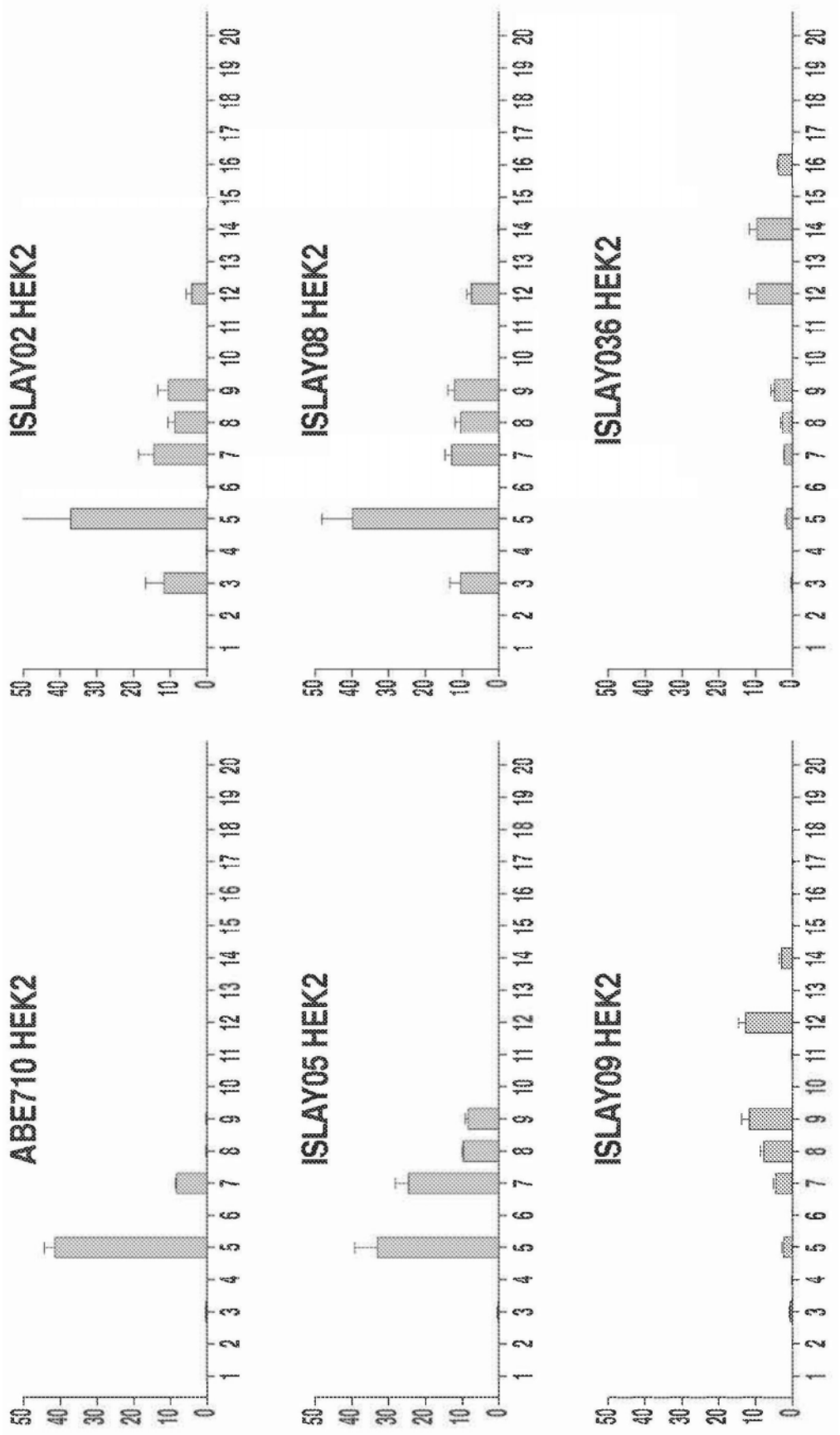

ISLAY02 HEK2
ISLAY08 HEK2
ISLAY036 HEK2
ABE710 HEK2
ISLAY05 HEK2
ISLAY09 HEK2

图24F

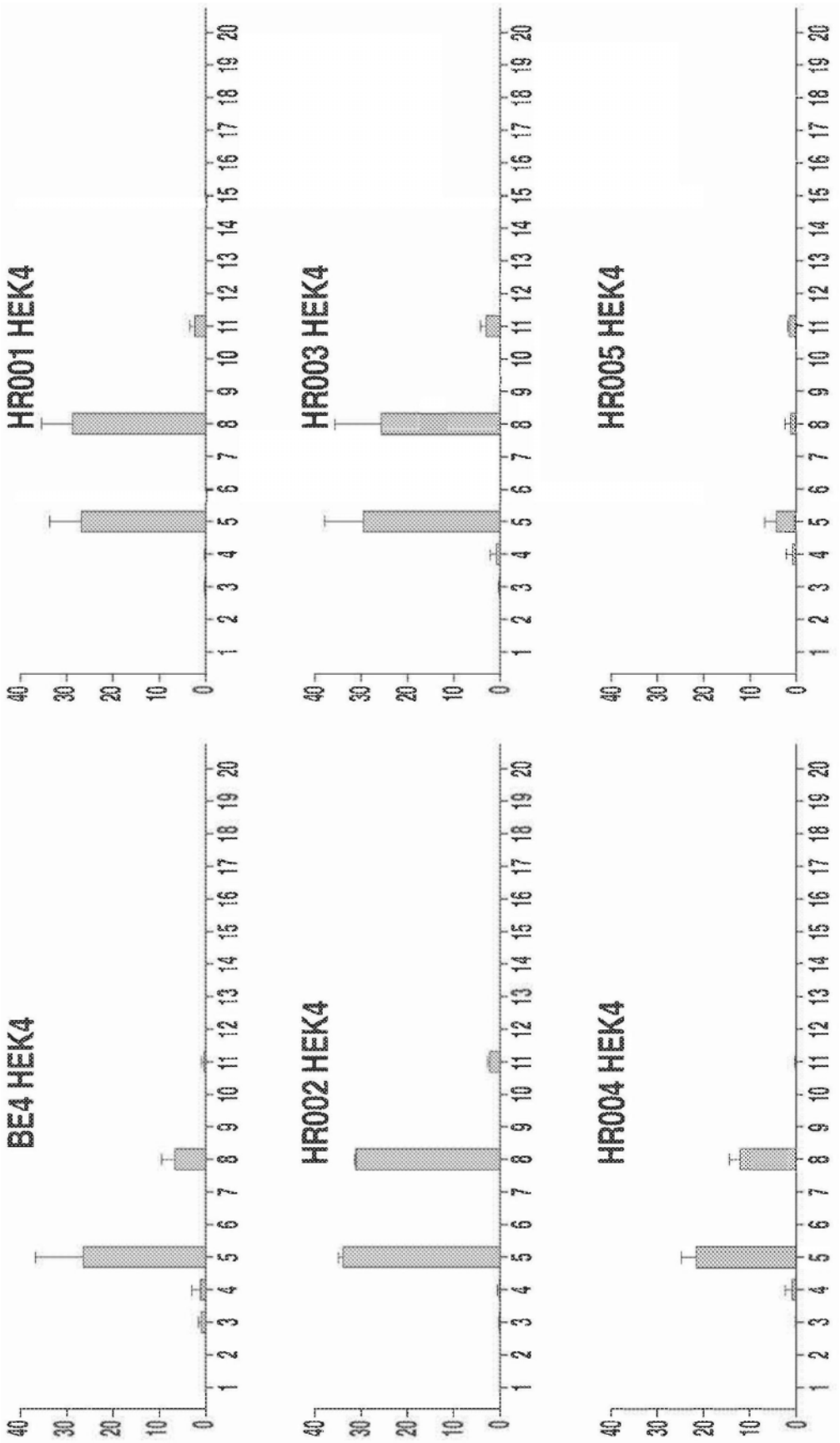

HR001 HEK4
HR003 HEK4
HR005 HEK4
BE4 HEK4
HR002 HEK4
HR004 HEK4

图25A

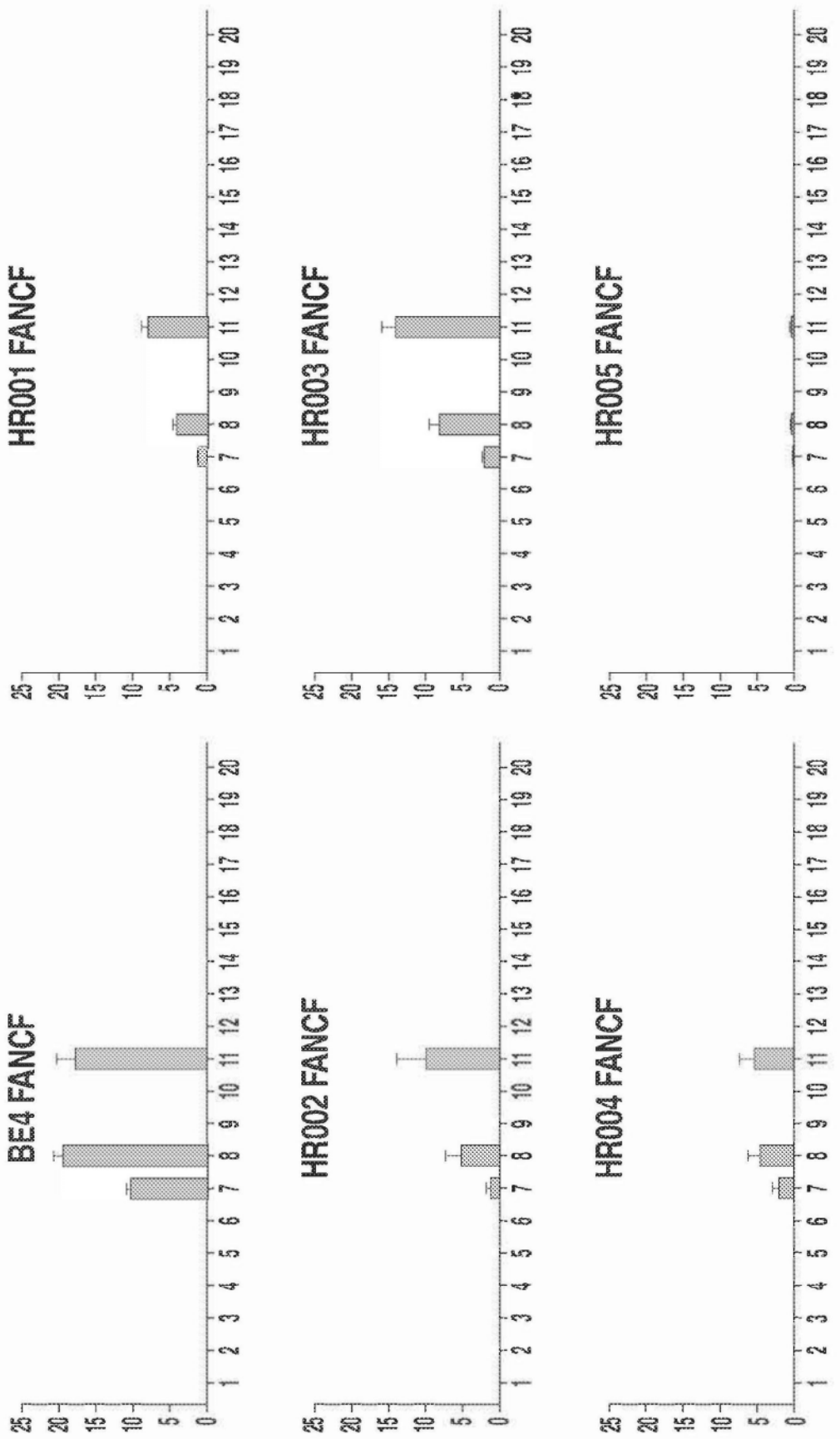

HR001 FANCF
HR003 FANCF
HR005 FANCF
BE4 FANCF
HR002 FANCF
HR004 FANCF

图25B

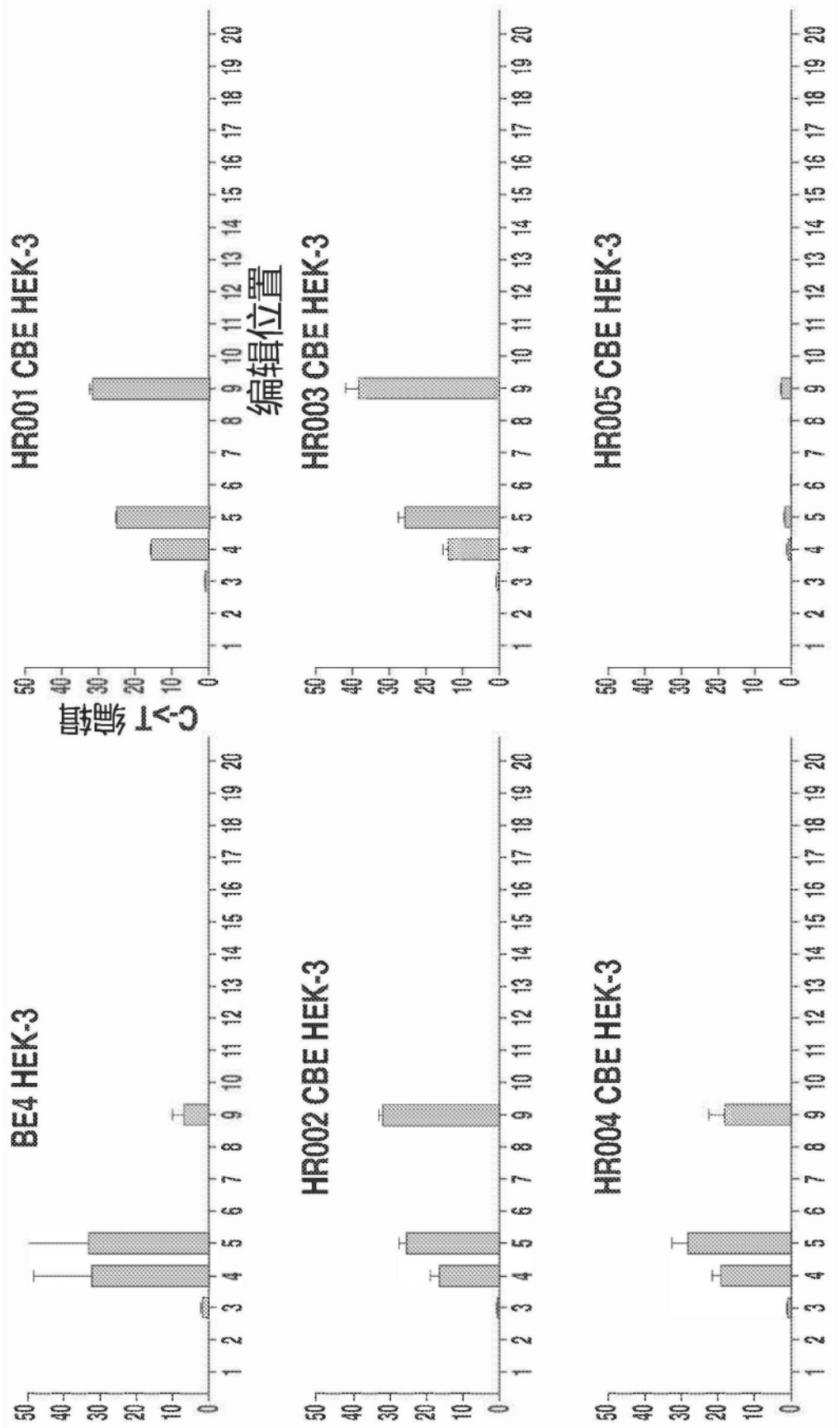

BE4 HEK-3
HR001 CBE HEK-3
HR002 CBE HEK-3
HR003 CBE HEK-3
HR004 CBE HEK-3
HR005 CBE HEK-3
C-T 编辑
编辑位置

图25C

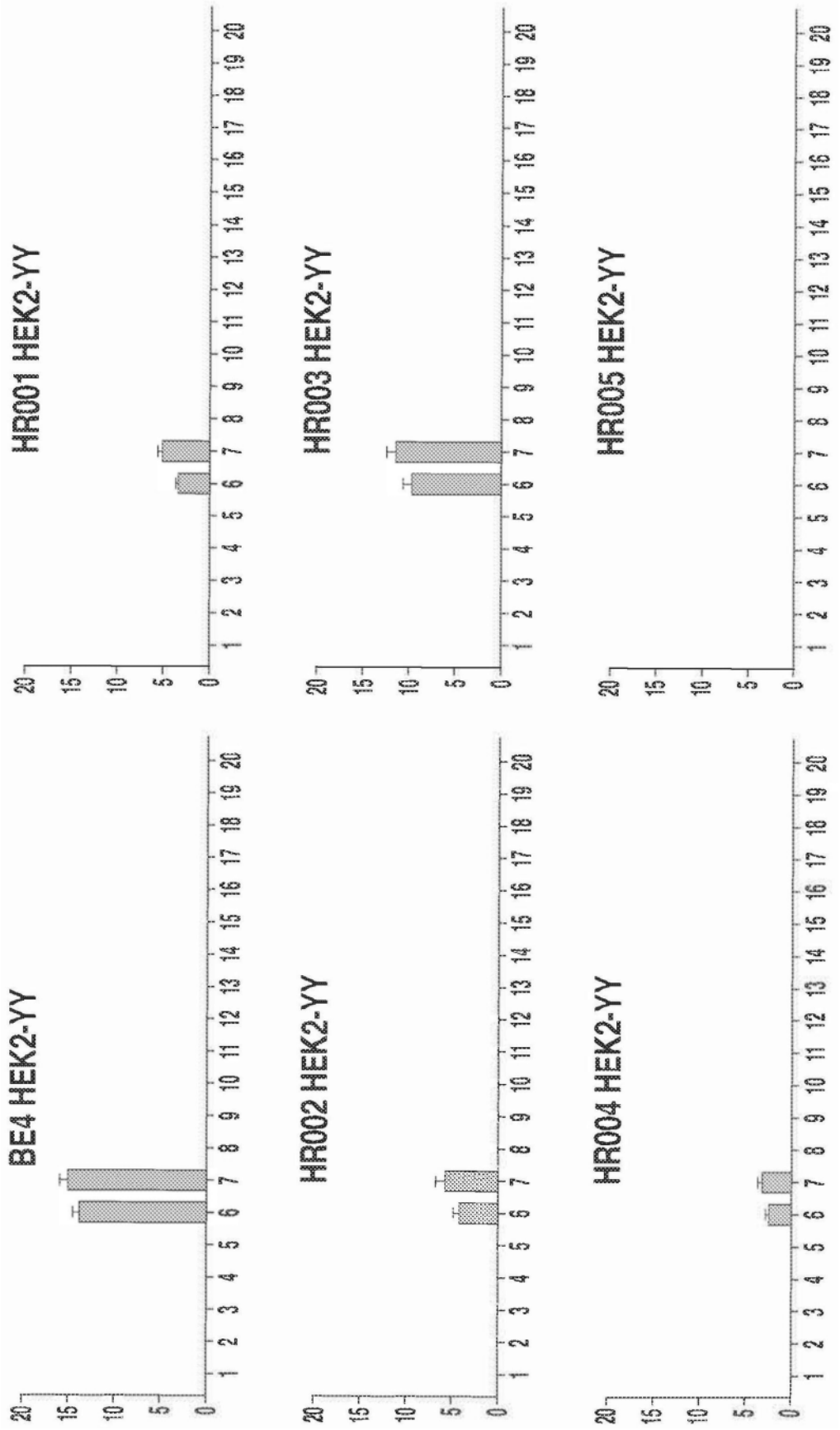

HR001 HEK2-YY
HR003 HEK2-YY
HR005 HEK2-YY
BE4 HEK2-YY
HR002 HEK2-YY
HR004 HEK2-YY

图25D

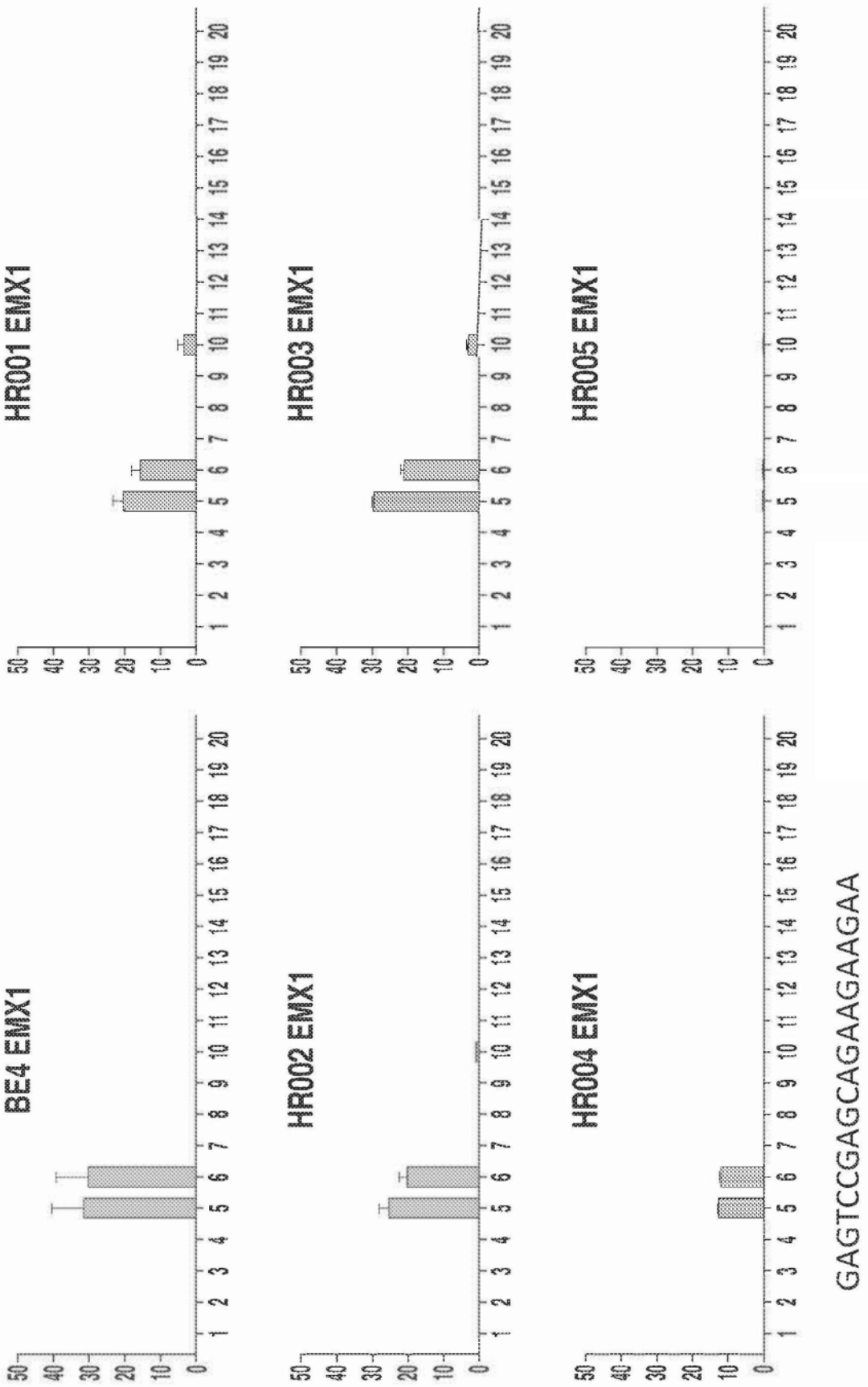

HR001 EMX1
HR003 EMX1
HR005 EMX1
BE4 EMX1
HR002 EMX1
HR004 EMX1
GAGTCCGAGCAGAAGAAGAA

图25E

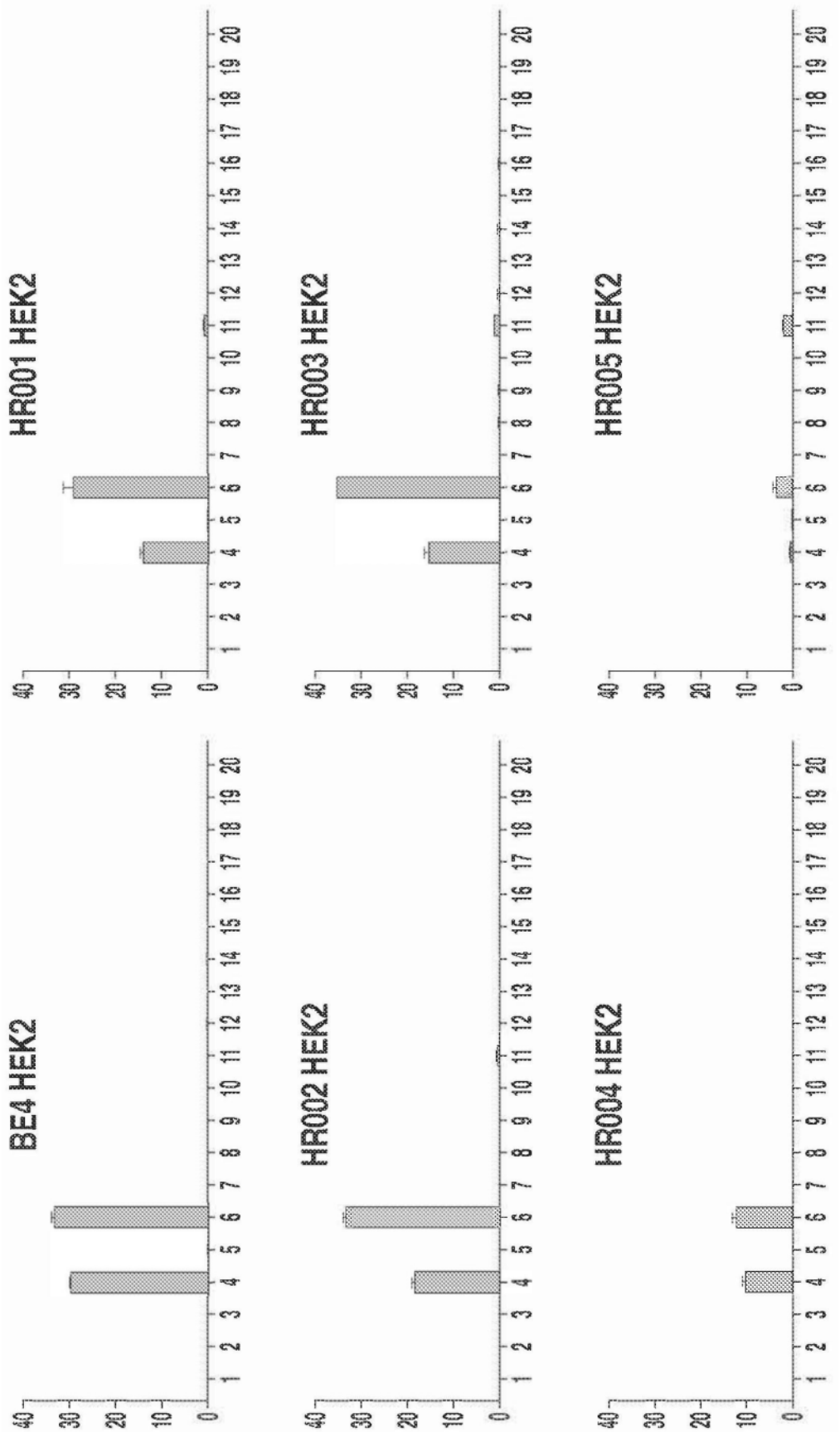

HR001 HEK2
HR003 HEK2
HR005 HEK2
BE4 HEK2
HR002 HEK2
HR004 HEK2

图25F