



(12) 发明专利

(10) 授权公告号 CN 112969790 B

(45) 授权公告日 2024. 07. 05

(21) 申请号 201980073055.6
(22) 申请日 2019.09.07
(65) 同一申请的已公布的文献号
申请公布号 CN 112969790 A
(43) 申请公布日 2021.06.15
(30) 优先权数据
62/728,703 2018.09.07 US
62/779,404 2018.12.13 US
(85) PCT国际申请进入国家阶段日
2021.05.06
(86) PCT国际申请的申请数据
PCT/US2019/050111 2019.09.07
(87) PCT国际申请的公布数据
W02020/051561 EN 2020.03.12
(73) 专利权人 比姆医疗股份有限公司
地址 美国马萨诸塞州

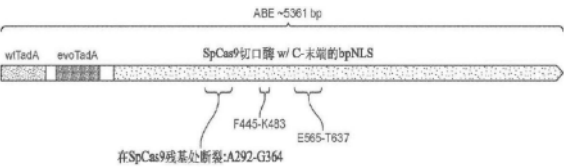
(72) 发明人 D·布赖森
(74) 专利代理机构 北京戈程知识产权代理有限公司 11314
专利代理师 程伟
(51) Int.Cl.
C12N 15/11 (2006.01)
C12N 15/62 (2006.01)
C12N 15/10 (2006.01)
C12N 9/22 (2006.01)
C07K 19/00 (2006.01)
(56) 对比文件
US 2017268022 A1, 2017.09.21
US 2018127780 A1, 2018.05.10
WO 2016205759 A1, 2016.12.22
审查员 邓钰

(54) 发明名称

用于递送核碱基编辑系统的组合物和方法

(57) 摘要

本发明提供用于递送第一和第二多核苷酸的组合物和方法,所述第一和第二多核苷酸各自编码A-至-G碱基编辑器融合蛋白的片段,其包括一个或多个例如腺苷脱氨酶的脱氨酶和nCas9,其中所述第一多核苷酸编码与断裂内含肽对的内含肽-N融合的nCas9的N末端片段,且所述第二多核苷酸编码与断裂内含肽对的内含肽-C融合的nCas9的C末端片段,以及将这些片段与sgRNA一起递送到细胞的方法(例如,AAV递送),其中在细胞中所述片段通过断裂内含肽系统被剪接在一起,从而在细胞中重构功能性碱基编辑系统。



1. 一种组合物,其包括

(a) 编码融合蛋白的第一多核苷酸,所述融合蛋白包括由N-端至C-端依序是野生型腺苷脱氨酶TadA、进化型版本TadA和Cas9切口酶(nCas9)的N末端片段,其中所述nCas9的N末端片段起始于nCas9的N-端,并且是终止于选自如SEQ ID NO:2中所编号的nCas9的302、309、312、354、455、459、462、465、468、471和473所组成的群组的位置的连续序列,并且其中所述nCas9的N末端片段与断裂内含肽-N融合,以及

(b) 编码所述nCas9的C末端片段的第二多核苷酸,其中所述nCas9的C末端片段起始于如SEQ ID NO:2中所编号的:

- i) 当所述nCas9的N末端片段终止于302的位置时,所述nCas9的S303的位置,
- ii) 当所述nCas9的N末端片段终止于309的位置时,所述nCas9的T310的位置,
- iii) 当所述nCas9的N末端片段终止于312的位置时,所述nCas9的T313的位置,
- iv) 当所述nCas9的N末端片段终止于354的位置时,所述nCas9的S355的位置,
- v) 当所述nCas9的N末端片段终止于455的位置时,所述nCas9的A456的位置,
- vi) 当所述nCas9的N末端片段终止于459的位置时,所述nCas9的S460的位置,
- vii) 当所述nCas9的N末端片段终止于462的位置时,所述nCas9的A463的位置,
- viii) 当所述nCas9的N末端片段终止于465的位置时,所述nCas9的T466的位置,
- ix) 当所述nCas9的N末端片段终止于468的位置时,所述nCas9的S469的位置,
- x) 当所述nCas9的N末端片段终止于471的位置时,所述nCas9的T472的位置,或
- xi) 当所述nCas9的N末端片段终止于473的位置时,所述nCas9的T474的位置,

并且是终止于所述nCas9的C-端的连续序列,其中所述nCas9的C末端片段的N-端残基是取代Ala、Ser或Thr的Cys,且其中所述nCas9的C末端片段与断裂内含肽-C和C末端核定位信号(NLS)融合。

2. 如权利要求1所述的组合物,进一步包括单向导RNA(sgRNA)或编码所述sgRNA的多核苷酸。

3. 如权利要求1所述的组合物,其中所述第一和所述第二多核苷酸被连接起来。

4. 如权利要求1所述的组合物,其中所述N末端和C末端片段是各自表达的。

5. 如权利要求1所述的组合物,其中所述进化型版本TadA是TadA7.10。

6. 如权利要求1所述的组合物,其中所述NLS是二分NLS。

7. 如权利要求1所述的组合物,其中所述nCas9是SpCas9切口酶。

8. 一种组合物,其包括如权利要求1至7中任一项所述的融合蛋白和所述nCas9的C末端片段。

9. 如权利要求1或8所述的组合物,其中所述nCas9的N末端片段与所述脱氨酶通过连接子连接。

10. 如权利要求9所述的组合物,其中所述连接子是肽连接子。

11. 一种载体,其包括如权利要求1至10中任一项所述的第一和第二多核苷酸。

12. 如权利要求11所述的载体,其中所述载体包括启动子。

13. 如权利要求12所述的载体,其中所述启动子是组成型启动子。

14. 如权利要求13所述的载体,其中所述组成型启动子是CMV或CAG启动子。

15. 如权利要求11或12所述的载体,其中所述载体选自逆转录病毒载体、腺病毒载

体、慢病毒载体、疱疹病毒载体和腺相关病毒载体组成的群组。

16. 如权利要求15所述的载体,其中所述载体是腺相关病毒载体。

17. 一种细胞,其包括如权利要求1至5中任一项所述的组合物,或如权利要求11至16中任一项所述的载体。

18. 如权利要求17所述的细胞,其中所述细胞是哺乳动物细胞。

19. 一种组合物于制备用于将碱基编辑器系统递送到细胞的药物的用途,所述组合物包括:

(a) 编码融合蛋白的第一多核苷酸,所述融合蛋白包括由N-端至C-端依序是野生型腺苷脱氨酶TadA、进化型版本TadA和Cas9切口酶(nCas9)的N末端片段,其中所述nCas9的N末端片段起始于nCas9的N-端,并且是终止于选自如SEQ ID NO:2中所编号的nCas9的302、309、312、354、455、459、462、465、468、471和473所组成的群组的位置的连续序列,并且其中所述nCas9的N末端片段与断裂内含肽-N融合,

(b) 编码所述nCas9的C末端片段的第二多核苷酸,其中所述nCas9的C末端片段起始于如SEQ ID NO:2中所编号的:

- i) 当所述nCas9的N末端片段终止于302的位置时,所述nCas9的S303的位置,
- ii) 当所述nCas9的N末端片段终止于309的位置时,所述nCas9的T310的位置,
- iii) 当所述nCas9的N末端片段终止于312的位置时,所述nCas9的T313的位置,
- iv) 当所述nCas9的N末端片段终止于354的位置时,所述nCas9的S355的位置,
- v) 当所述nCas9的N末端片段终止于455的位置时,所述nCas9的A456的位置,
- vi) 当所述nCas9的N末端片段终止于459的位置时,所述nCas9的S460的位置,
- vii) 当所述nCas9的N末端片段终止于462的位置时,所述nCas9的A463的位置,
- viii) 当所述nCas9的N末端片段终止于465的位置时,所述nCas9的T466的位置,
- ix) 当所述nCas9的N末端片段终止于468的位置时,所述nCas9的S469的位置,
- x) 当所述nCas9的N末端片段终止于471的位置时,所述nCas9的T472的位置,或
- xi) 当所述nCas9的N末端片段终止于473的位置时,所述nCas9的T474的位置,

并且是终止于所述nCas9的C-端的连续序列,其中所述nCas9的C末端片段的N-端残基是取代了Ala、Ser或Thr的Cys,且其中所述nCas9的C末端片段与断裂内含肽-C和C末端核定位信号(NLS)融合,以及

(c) 单向导RNA(sgRNA)或编码所述sgRNA的多核苷酸;

其中所述用途包括使细胞与所述组合物接触。

20. 如权利要求19所述的用途,其中所述sgRNA与靶多核苷酸互补。

21. 如权利要求20所述的用途,其中所述靶多核苷酸存在于生物体的基因组中。

22. 如权利要求21所述的用途,其中所述生物体是动物、植物或细菌。

23. 如权利要求19所述的用途,其中所述第一多核苷酸和/或所述第二多核苷酸经由载体与所述细胞接触。

24. 如权利要求23所述的用途,其中所述载体选自逆转录病毒载体、腺病毒载体、慢病毒载体、疱疹病毒载体和腺相关病毒载体组成的群组。

25. 如权利要求24所述的用途,其中所述载体是腺相关病毒载体。

26. 如权利要求19所述的用途,其中所述进化型版本是TadA7.10。

27. 如权利要求19所述的用途,其中所述NLS是二分NLS。

28. 如权利要求19所述的用途,其中所述nCas9的N末端片段和所述nCas9的C末端片段连接,以形成SpCas9切口酶。

29. 一种组合物于制备用于在细胞中编辑靶多核苷酸的药物的用途,所述组合物包括如权利要求1至7中任一项所述的第一和第二多核苷酸接触,其中所述第一或所述第二多核苷酸编码单向导RNA,其中所述用途包括使细胞与所述组合物接触并且在细胞中表达编码的蛋白和单向导RNA。

用于递送核碱基编辑系统的组合物和方法

[0001] 相关申请

[0002] 本申请要求2018年9月7日提交的美国临时专利申请第62/728,703号和2018年12月13日提交的美国临时专利申请第62/779,404号的效益,其全部内容在此通过引用并入本文。

背景技术

[0003] 常间回文重复序列丛集(Clustered Regularly-Interspaced Short Palindromic Repeats,CRISPR)的发现已引领了分子生物学领域的革命。人们的热情大部分都集中在CRISPR/Cas9治疗人类疾病和编辑人类基因组的临床潜力。使用CRISPR或基于CRISPR的系统可能有修复引起疾病的突变的潜力。实现此目标的一个挑战是基因组编辑所需元件的递送。例如,对于CRISPR/Cas9来说,可以在DNA质粒载体中编码SpCas9和sgRNA,并经由腺相关病毒(AAV)递送。然而,由于AAVs的小的包装容量,难以包含其他元件(诸如多肽结构域,启动子,报告基因,荧光标签,多个sgRNAs,或用于HDR的DNA模板)来帮助实现将CRISPR/Cas9组分递送到细胞和/或达到所需的基因编辑目标。

发明内容

[0004] 在一些方面,本文所提供的组合物包括(a)编码融合蛋白的第一多核苷酸,所述融合蛋白包括脱氨酶和Cas9的N末端片段,其中所述Cas9的N末端片段起始于Cas9的N-端,并且是终止于如SEQ ID NO:2中所编号的Cas9的A292-G365之间的位置的连续序列,其中所述Cas9的N末端片段融合到断裂内含肽-N上,以及(b)编码Cas9的C末端片段的第二多核苷酸,其中所述Cas9的C末端片段起始于如SEQ ID NO:2中所编号的Cas9的A292-G365之间的位置,并且是终止于Cas9的C-端的连续序列,且其中所述Cas9的C末端片段是融合到断裂内含肽-C上。

[0005] 在一些方面,本文所提供的组合物包括(a)编码Cas9的N末端片段的第一多核苷酸,其中所述Cas9的N末端片段起始于Cas9的N-端,并且是终止于如SEQ ID NO:2中所编号的Cas9的A292-G365之间的位置的连续序列,其中所述Cas9的N末端片段融合到断裂内含肽-N上,以及(b)编码融合蛋白的第二多核苷酸,所述融合蛋白包括Cas9的C末端片段和脱氨酶,其中所述Cas9的C末端片段起始于如SEQ ID NO:2中所编号的Cas9的A292-G365之间的位置,并且是终止于Cas9的C-端的连续序列,且其中所述Cas9的C末端片段是融合到断裂内含肽-C上。

[0006] 在一些方面,本文所提供的组合物包括(a)编码融合蛋白的第一多核苷酸,所述融合蛋白包括脱氨酶和Cas9的N末端片段,其中所述Cas9的N末端片段起始于Cas9的N-端,并且是终止于如SEQ ID NO:2中所编号的Cas9的A292-G365、F445-K483或E565-T637之间的位置的连续序列,其中所述Cas9的N末端片段融合到断裂内含肽-N上,以及(b)编码Cas9的C末端片段的第二多核苷酸,其中所述Cas9的C末端片段起始于如SEQ ID NO:2中所编号的Cas9的A292-G365、F445-K483或E565-T637之间的位置,并且是终止于Cas9的C-端的连续序列,

其中所述Cas9的C末端片段的N-端残基是取代了Ala、Ser或Thr的Cys,且其中所述Cas9的C末端片段是融合到断裂内含肽-C上。

[0007] 在一些方面,本文所提供的组合物包括(a)编码Cas9的N末端片段的第一多核苷酸,其中所述Cas9的N末端片段起始于Cas9的N-端,并且是终止于如SEQ ID NO:2中所编号的Cas9的A292-G365、F445-K483或E565-T637之间的位置的连续序列,其中所述Cas9的N末端片段融合到断裂内含肽-N上,以及(b)编码融合蛋白的第二多核苷酸,所述融合蛋白包括Cas9的C末端片段和脱氨酶,其中所述Cas9的C末端片段起始于如SEQ ID NO:2中所编号的Cas9的A292-G365、F445-K483或E565-T637之间的位置,并且是终止于Cas9的C-端的连续序列,其中所述Cas9的C末端片段的N-端残基是取代了Ala、Ser或Thr的Cys,且其中所述Cas9的C末端片段是融合到断裂内含肽-C上。

[0008] 在一些实施方式中,Cas9的N末端片段包括最多如SEQ ID NO:2中所编号的氨基酸302、309、312、354、455、459、462、465、471、473、576、588或589。在一些实施方式中,Cas9的C末端片段或Cas9的N末端片段在对应于如SEQ ID NO:2中所编号的氨基酸S303、T310、T313、S355、A456、S460、A463、T466、S469、T472、T474、C574、S577、A589或S590的残基处包括一个Ala/Cys、Ser/Cys或Thr/Cys的突变。在一些实施方式中,组合物进一步包括单向导RNA(sgRNA)或编码所述sgRNA的多核苷酸。在一些实施方式中,第一和第二多核苷酸被连接起来。在一些实施方式中,第一和第二多核苷酸是各自表达的。在一些实施方式中,脱氨酶是腺苷脱氨酶。在一些实施方式中,脱氨酶是野生型TadA或TadA7.10。在一些实施方式中,脱氨酶是TadA二聚体。在一些实施方式中,TadA二聚体包括野生型TadA和TadA 7.10。在一些实施方式中,融合蛋白包括一核定位信号(NLS)。在一些实施方式中,Cas9的N末端片段或Cas9的C末端片段与NLS连接。在一些实施方式中,所述NLS是二分NLS。在一些实施方式中,Cas9的N末端片段和融合蛋白连接起来以形成包括脱氨酶和SpCas9的碱基编辑器蛋白。在一些实施方式中,Cas9的C末端片段和融合蛋白连接起来以形成包括脱氨酶和SpCas9的碱基编辑器蛋白。在一些实施方式中,所述SpCas9具有切口酶活性或者是催化失活的。

[0009] 在一些方面,本文提供了包括本文所揭示的融合蛋白和Cas9的N末端片段的组合物。在一些方面,本文提供了包括本文所揭示的融合蛋白和Cas9的C末端片段的组合物。在一些实施方式中,Cas9的N末端片段或Cas9的C末端片段与脱氨酶通过连接子连接。在一些实施方式中,所述连接子是肽连接子。

[0010] 在一些方面,本文提供了包括本文所揭示的第一和第二多核苷酸的载体。在一些实施方式中,所述载体包括启动子。在一些实施方式中,所述启动子是组成型启动子。在一些实施方式中,所述组成型启动子是CMV或CAG启动子。在一些实施方式中,所述载体选自由逆转录病毒载体、腺病毒载体、慢病毒载体、疱疹病毒载体和腺相关病毒载体组成的群组。在一些实施方式中,所述载体是腺相关病毒载体。

[0011] 在一些方面,本文提供了包括本文所揭示的组合物或本文所揭示的载体的细胞。在一些实施方式中,所述细胞是哺乳动物细胞。

[0012] 在一些方面,本文提供了一种重构的A-至-G碱基编辑器蛋白,其包括Cas9结构域,且所述Cas9结构域包括一Ala/Cys、Ser/Cys或Thr/Cys的突变。在一些实施方式中,所述突变在对应于SpCas9氨基酸S303、T310、T313、S355、A456、S460、A463、T466、S469、T472、T474、C574、S577、A589或S590的残基处。

[0013] 在一些方面,本文提供了包括一个或多个多核苷酸的组合物,所述多核苷酸编码(a) Cas9的N末端片段,其中所述Cas9的N末端片段起始于Cas9的N-端,并且是终止于如SEQ ID NO:2中所编号的Cas9的A292-G365之间的位置的连续序列,且其中Cas9的N末端片段融合到断裂内含肽-N上,以及(b) Cas9的C末端片段,其中所述Cas9的C末端片段起始于如SEQ ID NO:2中所编号的Cas9的A292-G365之间的位置,并且是终止于Cas9的C-端的连续序列,且其中所述Cas9的C末端片段是融合到断裂内含肽-C上。

[0014] 在一些方面,本文提供了包括一个或多个多核苷酸的组合物,所述多核苷酸编码(a) Cas9的N末端片段,其中所述Cas9的N末端片段起始于Cas9的N-端,并且是终止于如SEQ ID NO:2中所编号的Cas9的氨基酸302、309、312、354、455、459、462、465、471、473、576、588或589的连续序列,且其中Cas9的N末端片段融合到断裂内含肽-N上,以及(b) Cas9的C末端片段,其中所述Cas9的C末端片段起始于如SEQ ID NO:2中所编号的Cas9的303、310、313、355、456、460、463、466、472、474、577、589或590,并且是终止于Cas9的C-端的连续序列,且其中所述Cas9的C末端片段是融合到断裂内含肽-C上。

[0015] 在一些实施方式中,Cas9的N末端片段或Cas9的C末端片段与核定位信号(NLS)连接。在一些实施方式中,Cas9的N末端片段和Cas9的C末端片段均与NLS连接。在一些实施方式中,所述NLS是二分NLS。在一些实施方式中,Cas9的N末端片段和Cas9的C末端片段连接起来以形成SpCa9。在一些实施方式中,所述SpCas9具有切口酶活性或者是催化失活的。

[0016] 在一些方面,本文提供了包括本文所揭示的(a)中的Cas9的N末端片段和(b)中的Cas9的C末端片段的组合物。

[0017] 在一些方面,本文提供了包括本文所揭示的一个或多个多核苷酸的载体。在一些实施方式中,所述载体包括启动子。在一些实施方式中,所述启动子是组成型启动子。在一些实施方式中,所述组成型启动子是CMV或CAG启动子。在一些实施方式中,所述载体选自由逆转录病毒载体、腺病毒载体、慢病毒载体、疱疹病毒载体和腺相关病毒载体组成的群组。在一些实施方式中,所述载体是腺相关病毒载体。

[0018] 在一些方面,本文提供了包括本文所揭示的组合物或本文所揭示的载体的细胞。在一些实施方式中,所述细胞是哺乳动物细胞。

[0019] 在一些方面,本文提供了包括一Ala/Cys、Ser/Cys或Thr/Cys的突变的Cas9变体多肽。在一些方面,本文提供了Cas9变体多肽,其在氨基酸303、310、313、355、456、460、463、466、472或474处包括Cys残基。

[0020] 在一些方面,本文提供了一种用于将碱基编辑器系统递送到细胞的方法,所述方法包括使细胞与如下三者接触:(a) 编码融合蛋白的第一多核苷酸,所述融合蛋白包括脱氨酶和Cas9的N末端片段,其中所述Cas9的N末端片段起始于Cas9的N-端,并且是终止于如SEQ ID NO:2中所编号的Cas9的A292-G365之间的位置的连续序列,其中所述Cas9的N末端片段融合到断裂内含肽-N上,(b) 编码Cas9的C末端片段的第二多核苷酸,其中所述Cas9的C末端片段起始于如SEQ ID NO:2中所编号的Cas9的A292-G365之间的位置,并且是终止于Cas9的C-端的连续序列,且其中所述Cas9的C末端片段是融合到断裂内含肽-C上,以及(c) 单向导RNA(sgRNA)或编码所述sgRNA的多核苷酸。

[0021] 在一些方面,本文提供了一种用于将碱基编辑器系统递送到细胞的方法,所述方法包括使细胞与如下三者接触:(a) 编码融合蛋白的第一多核苷酸,所述融合蛋白包括Cas9

的N末端片段,其中所述Cas9的N末端片段起始于Cas9的N-端,并且是终止于如SEQ ID NO:2中所编号的Cas9的A292-G365之间的位置的连续序列,其中所述Cas9的N末端片段融合到断裂内含肽-N上,(b) 编码Cas9的C末端片段和脱氨酶的第二多核苷酸,其中所述Cas9的C末端片段起始于如SEQ ID NO:2中所编号的Cas9的A292-G365之间的位置,并且是终止于Cas9的C-端的连续序列,且其中所述Cas9的C末端片段是融合到断裂内含肽-C上,以及(c) 单向导RNA (sgRNA) 或编码所述sgRNA的多核苷酸。

[0022] 在一些方面,本文提供了一种用于将碱基编辑器系统递送到细胞的方法,所述方法包括使细胞与如下三者接触:(a) 编码融合蛋白的第一多核苷酸,所述融合蛋白包括脱氨酶和Cas9的N末端片段,其中所述Cas9的N末端片段起始于Cas9的N-端,并且是终止于如SEQ ID NO:2中所编号的Cas9的A292-G365、F445-K483或E565-T637之间的位置的连续序列,其中所述Cas9的N末端片段融合到断裂内含肽-N上,(b) 编码Cas9的C末端片段的第二多核苷酸,其中所述Cas9的C末端片段起始于如SEQ ID NO:2中所编号的Cas9的A292-G365、F445-K483或E565-T637之间的位置,并且是终止于Cas9的C-端的连续序列,其中所述Cas9的C末端片段的N-端残基是取代了Ala、Ser或Thr的Cys,且其中所述Cas9的C末端片段是融合到断裂内含肽-C上,以及(c) 单向导RNA (sgRNA) 或编码所述sgRNA的多核苷酸。

[0023] 在一些方面,本文提供了一种用于将碱基编辑器系统递送到细胞的方法,所述方法包括使细胞与如下三者接触:(a) 编码融合蛋白的第一多核苷酸,所述融合蛋白包括Cas9的N末端片段,其中所述Cas9的N末端片段起始于Cas9的N-端,并且是终止于如SEQ ID NO:2中所编号的Cas9的A292-G365、F445-K483或E565-T637之间的位置的连续序列,其中所述Cas9的N末端片段融合到断裂内含肽-N上,(b) 编码Cas9的C末端片段和脱氨酶的第二多核苷酸,其中所述Cas9的C末端片段起始于如SEQ ID NO:2中所编号的Cas9的A292-G365、F445-K483或E565-T637之间的位置,并且是终止于Cas9的C-端的连续序列,其中所述Cas9的C末端片段的N-端残基是取代了Ala、Ser或Thr的Cys,且其中所述Cas9的C末端片段是融合到断裂内含肽-C上,以及(c) 单向导RNA (sgRNA) 或编码所述sgRNA的多核苷酸。

[0024] 在一些实施方式中,sgRNA与靶多核苷酸互补。在一些实施方式中,所述靶多核苷酸存在于生物体的基因组中。在一些实施方式中,所述生物体是动物、植物或细菌。在一些实施方式中,第一多核苷酸,第二多核苷酸,和/或编码sgRNA的多核苷酸经由载体与细胞接触。在一些实施方式中,所述载体选自逆转录病毒载体、腺病毒载体、慢病毒载体、疱疹病毒载体和腺相关病毒载体组成的群组。在一些实施方式中,所述载体是腺相关病毒载体。在一些实施方式中,Cas9的C末端片段或Cas9的N末端片段在对应于如SEQ ID NO:2中所编号的氨基酸S303、T310、T313、S355、A456、S460、A463、T466、S469、T472、T474、C574、S577、A589或S590的残基处包括一个Ala/Cys、Ser/Cys或Thr/Cys的突变。在一些实施方式中,脱氨酶是腺苷脱氨酶。在一些实施方式中,脱氨酶是TadA或其变体。在一些实施方式中,脱氨酶是野生型TadA或Tad7.10。在一些实施方式中,脱氨酶是TadA二聚体。在一些实施方式中,所述TadA二聚体包括野生型TadA和TadA7.10。在一些实施方式中,Cas9的N末端片段或Cas9的C末端片段包括NLS。在一些实施方式中,Cas9的N末端片段和Cas9的C末端片段均包括NLS。在一些实施方式中,所述NLS是二分NLS。在一些实施方式中,Cas9的N末端片段和Cas9的C末端片段连接起来以形成SpCas9。在一些实施方式中,所述SpCas9具有切口酶活性或者是催化失活的。

[0025] 在一些方面,本文提供了编码融合蛋白的多核苷酸,其中所述融合蛋白包括脱氨酶和Cas9的N末端片段,其中所述Cas9的N末端片段起始于Cas9的N-端,并且是终止于如SEQ ID NO:2中所编号的Cas9的A292-G365之间的位置的连续序列,其中Cas9的N末端片段融合到断裂内含肽-N上。在一些方面,本文提供了编码融合蛋白的多核苷酸,其中所述融合蛋白包括脱氨酶和Cas9的N末端片段,其中所述Cas9的N末端片段起始于Cas9的N-端,并且是终止于如SEQ ID NO:2中所编号的Cas9的F445-K483之间的位置的连续序列,且其中Cas9的N末端片段融合到断裂内含肽-N上。在一些方面,本文提供了编码融合蛋白的多核苷酸,其中所述融合蛋白包括脱氨酶和Cas9的N末端片段,其中所述Cas9的N末端片段起始于Cas9的N-端,并且是终止于如SEQ ID NO:2中所编号的Cas9的E565-T637之间的位置的连续序列,且其中Cas9的N末端片段融合到断裂内含肽-N上。在一些方面,本文提供了编码融合蛋白的多核苷酸,其中所述融合蛋白包括脱氨酶和Cas9的C末端片段,其中所述Cas9的C末端片段起始于如SEQ ID NO:2中所编号的Cas9的A292-G365之间的位置,并且是终止于Cas9的C-端的连续序列,且其中Cas9的C末端片段是融合到断裂内含肽-C上。在一些方面,本文提供了编码融合蛋白的多核苷酸,其中所述融合蛋白包括脱氨酶和Cas9的C末端片段,其中所述Cas9的C末端片段起始于如SEQ ID NO:2中所编号的Cas9的F445-K483之间的位置,并且是终止于Cas9的C-端的连续序列,且其中Cas9的C末端片段是融合到断裂内含肽-C上。在一些方面,本文提供了编码融合蛋白的多核苷酸,其中所述融合蛋白包括脱氨酶和Cas9的C末端片段,其中所述Cas9的C末端片段起始于如SEQ ID NO:2中所编号的Cas9的E565-T637之间的位置,并且是终止于Cas9的C-端的连续序列,且其中Cas9的C末端片段是融合到断裂内含肽-C上。在一些方面,本文提供了编码融合蛋白的多核苷酸,其中所述融合蛋白包括脱氨酶和Cas9的C末端片段,其中所述Cas9的C末端片段起始于如SEQ ID NO:2中所编号的Cas9的A292-G365、F445-K483或E565-T637之间的位置,并且是终止于Cas9的C-端的连续序列,其中所述Cas9的C末端片段的N-端残基是取代了Ala、Ser或Thr的Cys,且其中所述Cas9的C末端片段是融合到断裂内含肽-C上。

[0026] 在一些实施方式中,Cas9的C末端片段或Cas9的N末端片段包括一Ala/Cys、Ser/Cys或Thr/Cys的突变。在一些实施方式中,所述突变在对应于如SEQ ID NO:2中所编号的氨基酸S303、T310、T313、S355、A456、S460、A463、T466、S469、T472、T474、C574、S577、A589或S590的残基处。在一些实施方式中,脱氨酶是腺苷脱氨酶。在一些实施方式中,脱氨酶是TadA或其变体。在一些实施方式中,脱氨酶是野生型TadA,或Tad7.10。在一些实施方式中,融合蛋白包括彼此链接的两个脱氨酶。在一些实施方式中,融合蛋白同时包括野生型TadA和TadA7.10。在一些实施方式中,融合蛋白包括NLS。在一些实施方式中,所述NLS是二分NLS。在一些实施方式中,Cas9的N末端片段或Cas9的C末端片段包括SpCas9的氨基酸序列。在一些实施方式中,Cas9的N末端片段或Cas9的C末端片段包括一个或多个与核酸酶活性降低相关联的氨基酸取代。

[0027] 在一些方面,本文提供了与断裂内含肽-N融合的Cas9蛋白的N末端片段,所述N末端片段包括最多至氨基酸302、309、312、354、455、459、462、465、471或473。在一些方面,本文提供了Cas9蛋白的C末端蛋白片段,其中所述C末端片段的N-端氨基酸是在氨基酸303、310、313、355、456、460、463、466、472或474处的Cys取代,并且是融合到断裂内含肽-C上。在一些方面,本文提供了编码A-至-G碱基编辑器融合蛋白的片段的多核苷酸,所述融合蛋白

包括一个或多个脱氨酶和Cas9的N末端片段,其中所述N末端片段融合到断裂内含肽-N上。在一些方面,本文提供了编码A-至-G碱基编辑器融合蛋白的片段的多核苷酸,所述融合蛋白包括一个或多个脱氨酶和Cas9的C末端片段,其中所述C末端片段是融合到断裂内含肽-C上。在一些方面,本文提供了A-至-G碱基编辑器融合蛋白的蛋白片段,所述融合蛋白包括一个或多个脱氨酶和Cas9的N末端片段,其中所述N末端片段融合到断裂内含肽-N上。在一些方面,本文提供了A-至-G碱基编辑器融合蛋白的蛋白片段,所述融合蛋白包括一个或多个脱氨酶和Cas9的C末端片段,其中所述C末端片段是融合到断裂内含肽-C上。

[0028] 在一些方面,本文提供了包括第一和第二多核苷酸的组合物,所述第一和第二多核苷酸各自编码包括一个或多个脱氨酶和Cas9的A-至-G碱基编辑器融合蛋白的片段,其中第一多核苷酸编码融合到断裂内含肽-N的Cas9的N末端片段,且第二多核苷酸编码融合到断裂内含肽-C的Cas9的C末端片段。在一些方面,本文提供了一组合物,其包括A-至-G碱基编辑器融合蛋白的N末端和C末端片段,所述融合蛋白包括一个或多个脱氨酶和SpCas9,其中所述N末端片段包括融合到断裂内含肽-N的SpCas9片段并且所述C末端片段包括融合到断裂内含肽-C的SpCas9剩余部分。

[0029] 本文提供了用于将碱基编辑器系统递送到细胞的方法,所述方法包括使细胞与第一和第二多核苷酸接触,所述第一和第二多核苷酸各自编码包括一个或多个脱氨酶和Cas9的A-至-G碱基编辑器融合蛋白的片段,其中第一多核苷酸编码融合到断裂内含肽-N的Cas9的N末端片段并且第二多核苷酸编码融合到断裂内含肽-C的Cas9的C末端片段,且第一或第二多核苷酸编码单向导RNA。在一些方面,本文提供了用于将碱基编辑器系统递送到细胞的方法,所述方法包括使细胞与包括一个或多个脱氨酶和SpCas9的A-至-G碱基编辑器融合蛋白的N末端和C末端片段接触,其中N末端片段包括融合到断裂内含肽-N的SpCas9片段并且C末端片段包括融合到断裂内含肽-C的SpCas9剩余部分,以及向导RNA。在一些方面,本文提供了用于在细胞中编辑靶多核苷酸的方法,所述方法包括使细胞与第一和第二多核苷酸接触,所述第一和第二多核苷酸各自编码包括一个或多个脱氨酶和Cas9的A-至-G碱基编辑器融合蛋白的片段,其中第一多核苷酸编码融合到断裂内含肽-N的Cas9的N末端片段并且第二多核苷酸编码融合到断裂内含肽-C的Cas9的C末端片段,且第一或第二多核苷酸编码单向导RNA,并且所述方法包括在细胞中表达被编码的蛋白质和单向导RNA。

[0030] 从详细描述和权利要求书看来,本发明的其他特征和优势将显而易见。

[0031] 定义

[0032] 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领域技术人员一般理解的意义。下述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等编撰的《Dictionary of Microbiology and Molecular Biology (1994第二版)》;The Cambridge Dictionary of Science and Technology (Walker编,1988);The Glossary of Genetics (R.Rieger等(编),Springer Verlag (1991));以及Hale和Marham的The Harper Collins Dictionary of Biology (1991)》。本文中,除非另做说明,否则下述术语具有其下方所述的意义。

[0033] “腺苷脱氨酶”是指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,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,腺苷脱氨酶催化在脱氧核糖核酸(DNA)中的腺嘌呤

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)可以来自任何生物体,诸如细菌。在一些实施方式中,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,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,脱氨酶或脱氨酶结构域为至少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,腺苷脱氨酶来自于细菌,诸如大肠杆菌,金黄色葡萄球菌,伤寒沙门菌,腐败希瓦氏菌,流感嗜血杆菌,或新月形杆菌。在一些实施方式中,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。

[0034] 例如,相对于全长ecTadA,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,截短的ecTadA可能缺少1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、6、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,截短的ecTadA可能缺少1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、6、17、18、19或20个C末端氨基酸残基。在一些实施方式中,ecTadA脱氨酶不包括N末端的甲硫氨酸。在一些实施方式中,TadA脱氨酶是N末端截短的TadA。在特定的实施方式中,TadA是在国际第PCT/US2017/045381号专利申请案(其全部内容通过引用并入本文)中所描述的任何一种。

[0035] 在某些实施方式中,腺苷脱氨酶包括氨基酸序列:

[0036] MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRREQEIKAKKAQSSTD,其称为“TadA参考序列”。

[0037] 在一些实施方式中,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,腺苷脱氨酶包括氨基酸序列:

[0038] MRRAFITGVFFLSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRREQEIKAKKAQSSTD

[0039] 然而,应当理解,可用于本发明中的另外的腺苷脱氨酶对本领域技术人员是显而易见的,且所述另外的腺苷脱氨酶在本发明披露的范围内。例如,腺苷脱氨酶可以是作用于tRNA(AD AT)的腺苷脱氨酶的同源物。示例性的AD AT同源物包含但不限于:

[0040] 金黄色葡萄球菌(*Staphylococcus aureus*, S.aureus)TadA:

[0041] MGSMTNDIYFMTLAIIEAKKAAQLGEVPIGAIITKDDEVIAHNLRETLLQPTAH AEHIAIERAAKVLGSRLEGCTLYVTLEPCVMCAGTIVMSRIPRVVYGADDPKGGCSGS LMNLLQQS NFNHRAIVDKG VLKEAC S TLLTTFFKNLRANKKS TN

[0042] 枯草芽孢杆菌(*Bacillus subtilis*, B.subtilis)TadA:

[0043] MTQDELYMKEAIKEAKKAEKGEVPIGAVLVINGEIIARAHNLRETEQRSIAHAEML VIDEACKALGTWRLEGATLYVTLEPCVMCAGAVLSRVEKVVFGAFDPKGGC SGTLMN LLQEERFNHQAEEVSGVLEECCGMLSAFFRELRRKKKAARKNLSE

[0044] 鼠伤寒沙门菌(*Salmonella typhimurium*, S.typhimurium)TadA:

[0045] MPPAFITGVTSLSDELVDHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVLQNYRLDDTTLYVTLEPCVMCAGAMVHSRIGRVVFGARDAKTGAAGSLIDVLHHPGMNHRVEIIIEG

VLRDECATLLSDFFRMRQEIKALKKADRAEGAGPAV

[0046] 腐败希瓦氏菌 (*Shewanella putrefaciens*, *S. putrefaciens*) TadA:

[0047] MDE YWMQVAMQM AEKAEAAAGE VPVGA VLVKDGQQIATGYNLS IS QHDPTAHAEILCLRSAGK
KLENYRLLDATLYITLPCAMCAGAMVHSRIARVVYGARDEKTGAAGTVVNLQHPAFNHQVEVTSGVLAEACSAQ
LSRFFKRRRDEKKALKLAQRAQQGIE流感嗜血杆菌 F3031 (*Haemophilus influenzae* F3031,
H. influenzae) TadA:

[0048] MDAKVRSFDEKMMRYALELADKAEALGEIPVGAVLVDDARNIIGEGWNLSIVQSDPT A H
AEIIALRNG AKNIQN YRLNLS TLY VTLEPCTMC AG AILHS RIKRLVFG ASD YKTGAIGSRFHFFDDY
KMNHTLEITSGVLAECSQKLSTFFQKRREEKKIEKALLKSLSD K

[0049] 新月形杆菌 (*Caulobacter crescentus*, *C. crescentus*) TadA:

[0050] MRTDESEDQDHRMRLALDAARAAAEEAGETPVGAVILDPSTGEVIATAGNGPIAAHDPTAHAEIAAMRA
AAAKLGNRYRLDTLTLVVTLPCAMCAGAI SHARIGRVVFGADDPKGGAVVHGPKFFAQPTCHWRPEVTGGVLADESA
DLLRGFFRARRKAKI

[0051] 硫还原地杆菌 (*Geobacter sulfurreducens*, *G. sulfurreducens*) TadA:

[0052] MSSLKKTPIRDDAYWMGKAIREAAKAAARDEVPIGAVIVRDGAVIGRGNLREGSNDPSAHAEMIAIRQ
AARRSANWRLTGATLYVTLEPCLMCMGAIILARLERVVFGCYDPKGGAAGSLYDLSADPRLNHQVRLSPGVCQEECG
TMSDFFRDLRRRKKAKATPALFIDERKVPPEP

[0053] TadA7.10

[0054] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNLRVIGEGWNRAIGLHDPTAHAEIMALRQGGVL
MQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNRHVEITEGILADECAALLC
YFFRMPRQVFNAQKKAQSSTD

[0055] “剂”是指任何小分子化学化合物,抗体,核酸分子,或多肽,或其片段。

[0056] “改变”是指通过标准的本领域已知的方法,诸如本文所述的方法,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,表达水平的25%变化、40%变化、以及50%或更大变化。

[0057] “类似物”是指不相同但具有类似的功能或结构特征的分子。例如,多肽类似物保留了相应的天然存在的多肽的至少一些生物学活性,同时具有某些序列修饰,所述修饰相对于天然存在的多肽增强了类似物的功能。此种修饰可以增加类似物的蛋白酶抗性、膜渗透性或半衰期,而不会改变例如多核苷酸的结合活性。在另一个例子中,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,所述修饰相对于天然存在的多核苷酸增强了类似物的功能。此种修饰可能增加多核苷酸对DNA的亲合力、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。

[0058] “碱基编辑器(BE)”,或“核碱基编辑器(NBE)”是指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,所述剂是融合蛋白,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,所述具有碱基编辑活性的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,所述碱基编辑器能够使DNA分子内的碱基脱氨基。在一些实施方式中,所述碱基编辑器能够使DNA内的胞嘧啶(C)或腺苷脱氨基。在一些实施方式中,所述碱基编辑器能够使DNA内的胞嘧啶(C)和腺苷(A)脱氨基。在一些实施方式中,所述碱基编辑器是胞苷碱基编辑器

(CBE)。在一些实施方式中,所述碱基编辑器是腺苷碱基编辑器(ABE)。在一些实施方式中,所述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,所述碱基编辑器是融合到腺苷脱氨酶的核酸酶失活Cas9(dCas9)。在一些实施方式中,Cas9是循环排列体(circular permutant)Cas9(例如spCas9或saCas9)。循环排列体Cas9是本领域已知的,且在例如Oakes等,Cell 176,254-267,2019中进行了描述。在一些实施方式中,碱基编辑器融合到碱基切除修复的抑制剂(例如UGI结构域)上。在一些实施方式中,融合蛋白包括融合到脱氨酶上的Cas9切口酶和碱基切除修复的抑制剂(诸如UGI结构域)。在其他实施方式中,碱基编辑器是无碱基的碱基编辑器。

[0059] 碱基编辑器系统的核碱基组分和多核苷酸可编程的核苷酸结合组分可以彼此共价地或非共价地相联。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将脱氨酶结构域靶向到靶核苷酸序列。在一些实施方式中,可以将多核苷酸可编程的核苷酸结合结构域融合到或链接到脱氨酶结构域。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,核碱基编辑组分,例如脱氨酶组分,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,另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,另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,另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,另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,另外的异源部分可能能够与多核苷酸连接子结合。另外的异源部分可以是蛋白结构域。在一些实施方式中,另外的异源部分可以是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。

[0060] 碱基编辑器系统可以进一步包括向导多核苷酸组分。应当理解,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,脱氨酶结构域可以通过向导多核苷酸被靶向到靶核苷酸序列。例如,在一些实施方式中,碱基编辑器系统的核碱基编辑组分,例如脱氨酶组分,可以包括另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结合结构域),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,另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,另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,另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,另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,另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,另外的异源部分可能能够与多核苷酸连接子结合。所述另外的异源部分可能是蛋白结构域。在一些实施方式中,另外的异源部分可以是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。

[0061] 在一些实施方式中,碱基编辑器系统可以包括一个或多个蛋白、融合蛋白、多肽或

其编码多核苷酸。在一些实施方式中,碱基编辑器系统可以包括编码融合蛋白(所述融合蛋白包括脱氨酶和napDNAbp的N末端片段)的第一多核苷酸以及编码napDNAbp的C末端片段的第二多核苷酸。例如,在特定实施方式中,napDNAbp的N末端片段可以融合到内含肽-N上,并且napDNAbp的C末端片段可以融合到内含肽-C上,以使得napDNAbp的N末端片段和C末端片段可以重构以形成碱基编辑器蛋白。

[0062] 在一些实施方式中,碱基编辑器系统可以进一步包括碱基切除修复(BER)组分的抑制剂。在一些实施方式中,碱基编辑器系统可以进一步包括碱基切除修复(BER)组分的抑制剂。应当理解,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。BER组分的抑制剂可以包括碱基切除修复抑制剂。在一些实施方式中,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,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,碱基切除修复的抑制剂可以通过多核苷酸可编程的核苷酸结合结构域被靶向到靶核苷酸序列。在一些实施方式中,可以将多核苷酸可编程的核苷酸结合结构域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,可以将多核苷酸可编程的核苷酸结合结构域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,碱基切除修复的抑制剂组分可以包括另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,可以通过向导多核苷酸将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,碱基切除修复的抑制剂可以包括另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,向导多核苷酸(例如多核苷酸结合结构域,诸如RNA或DNA结合蛋白)的另外的异源部分或结构域可以被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,另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,另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,另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,另外的异源部分可能能够与多核苷酸连接子结合。所述另外的异源部分可能是蛋白结构域。在一些实施方式中,另外的异源部分可以是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。

[0063] “碱基编辑活性”是指达到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。在另一个实施方式中,碱基被从多核苷酸中切除。在一个实施方式中,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,碱基编辑活性是腺苷脱氨酶活性,例如将A·T转换为G·C。

[0064] 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、失活的或部分活性的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(常间回文重复序列丛集)-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素互补的序列,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。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间隔序列互补的线性或环状dsDNA靶。不与crRNA互补的靶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。可以工程化向导RNAs(“sgRNA”或简称“gRNA”)以便将crRNA和tracrRNA的各方面都并入单一RNA种类的中。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本文。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 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.,Nature 471:602-607(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)。已在各种物种中描述了Cas9的异种同源物,包括但不限于化脓链球菌和嗜热链球菌。基于本发明披露,其他的适合的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,Rhodes和Charpentier,“The tracrRNA and Cas9 families of type IICRISPR-Cas immunity systems”(2013)RNA Biology 10:5,726-737中所揭示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。

[0065] 核酸酶-去活化的Cas9蛋白可互换地被称为“dCas9”蛋白(表示核酸酶-“死(dead)”Cas9)或催化失活的Cas9。形成具有失活的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。HNH亚结构域切开与gRNA互补的股,而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完全去活性化化脓链球菌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核酸酶具有失活的(例如去活性化的)DNA切开结构域,就是说,所述Cas9为切口酶,被称为“nCas9”蛋白(表示“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结构域的一:(1)Cas9的gRNA结合结构域;或是(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，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，Cas9变体包括Cas9的片段（例如gRNA结合结构域或DNA-切开结构域），使得所述片段与相应的野生型Cas9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，所述片段是相应的野生型Cas9的氨基酸长度的至少约30%、至少约35%、至少约40%、至少约45%、至少约50%、至少约55%、至少约60%、至少约65%、至少约70%、至少约75%、至少约80%、至少约85%、至少约90%、至少约95%、至少约96%、至少约97%、至少约98%、至少约99%或至少约99.5%。

[0066] 在一些实施方式中，所述片段的长度是至少100个氨基酸。在一些实施方式中，所述片段的长度是至少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。在一些实施方式中，野生型Cas9对应于来自化脓链球菌的Cas9（NCBI参考序列：NC_017053.1，核苷酸和氨基酸序列如下）。

[0067] ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGCGTCGGATGGGCGGTGATCACTGATGATTATAAGGTTCCGTCTAAAAAGTTCAAGGTTCTGGGAAATACAGACCGCCACAGTATCAAAAAAATCTTATAGGGCTCTTTTATTTGGCAGTGGAGAGACAGCGGAAGCGACTCGTCTCAAACGGACAGCTCGTAGAAGGTATACACGTCGGAAGAATCGTATTTGTTATCTACAGGAGATTTTTTCAAATGAGATGGCGAAAGTAGATGATAGTTTCTTTCATCGACTTGAAGAGTCTTTTTTGGTGGAAGAAGACAAGAAGCATGAACGTCATCCTATTTTTTGAAATATAGTAGATGAAGTTGCTTATCATGAGAAATATCCAATCTATCATCTGCGAAAAAATTGGCAGATTCTACTGATAAAGCGGATTTGCGCTTAATCTATTTGGCCTTAGCGCATATGATTAAGTTTCGTGGTCATTTTTTGATTGAGGGAGATTTAAATCCTGATAAGTAGTGATGTGGACAACTATTTATCCAGTTGGTACAAATCTACAATCAATTATTTGAAGAAAACCTATTAACGCAAGTAGAGTAGATGCTAAAGCGATTCTTTCTGCACGATTGAGTAAATCAAGACGATTAGAAAATCTCATTGCTCAGCTCCCCGGTGAGAAGAGAAATGGCTTGTTTGGGAATCTCATTGCTTTGTCATTGGGATTGACCCCTAATTTTAAATCAATTTTGATTTGGCAGAAGATGCTAAATTACAGCTTTCAAAGATACTTACGATGATGATTTAGATAATTTATTGGCGCAAATTGGAGATCAATATGCTGATTTGTTTTTGGCAGCTAAGAATTTATCAGATGCTATTTTACTTTTCAGATATCCTAAGAGTAAATAGTGAAATAACTAAGGCTCCCCTATCAGCTTCAATGATTAAGCGCTACGATGAACATCATCAAGAC TTGACTCTTTTAAAGCTTTAGTTTCGACAACAACCTCCAGAAAAGTATAAAGAAATCTTTT

[0068] TTGATCAATCAAAAAACGGATATGCAGGTTATATTGATGGGGAGCTAGCCAAGAAGAATTTTATAAATTTATCAAACCAATTTTAGAAAAATGGATGGTACTGAGGAATTATTGGTGAACTAAATCGTGAAGATTTGCTGCGCAAGCAACGGACCTTTGACAACGGCTCTATTCCCCATCAATTCACCTGGGTGAGCTGCATGCTATTTTGAGAAGACAAGAAGACTTTTATCCATTTTTTAAAGACAATCGTGAGAAGATTGAAAAATCTTGACTTTTGAATTCCTTATTATGTTGGTCCATTGGCGCGTGGCAATAGTCGTTTTGCATGGATGACTCGGAAGTCTGAAGAAACAATTACCCCATGGAATTTTGAAGAAGTTGTCGATAAAGGTGCTTCAGCTCAATCATTATTGAACGCATGACAACTTTGATAAAAAATCTTCC

AAATGAAAAAGTACTACCAAAACATAGTTTGCTTTATGAGTATTTTACGGTTTATAACGAATTGACAAAGGTCAAAT
ATGTTACTGAGGGAATGCGAAAACAGCATTTCTTTCAGGTGAACAGAAGAAAGCCATTGTTGATTTACTCTTCAAA
ACAAATCGAAAAGTAACCGTTAAGCAATTAAGAAGATTATTTCAAAAAAATAGAATGTTTTGATAGTGTTGAAAT
TTCAGGAGTTGAAGATAGATTTAATGCTTCATTAGGCGCCTACCATGATTTGCTAAAAATTATTAAGATAAAGATT
TTTTGGATAATGAAGAAAATGAAGATATCTTAGAGGATATTGTTTTAACATTGACCTTATTTGAAGATAGGGGGATG
ATTGAGGAAAGACTTAAACATATGCTCACCTCTTTGATGATAAGGTGATGAAACAGCTTAAACGTCGCCGTTATAC
TGGTTGGGGACGTTTGTCTCGAAAATTGATTAATGGTATTAGGGATAAGCAATCTGGCAAAACAATATTAGATTTTT
TGAAATCAGATGGTTTTGCCAATCGCAATTTTATGCAGCTGATCCATGATGATAGTTTGACATTTAAAGAAGATATT
CAAAAAGCACAGGTGTCTGGACAAGGCCATAGTTTACATGAACAGATTGCTAACTTAGCTGGCAGTCTGTCTATTAA
AAAAGGTATTTTACAGACTGTAAAAATTGTTGATGAACTGGTCAAAGTAATGGGGCATAAGCCAGAAAATATCGTTA
TTGAAATGGCACGTGAAAATCAGACAACCTCAAAAGGGCCAGAAAATTCGCGAGAGCGTATGAAACGAATCGAAGAA
GGTATCAAAGAATTAGGAAGTCAGATTCTTAAAGAGCATCCTGTTGAAAATACTCAATTGCAAAATGAAAAGCTCTA
TCTCTATTATCTACAAAATGGAAGAGACATGTATGTGGACCAAGAATTAGATATTAATCGTTTTAAGTGATTATGATG
TCGATCACATTGTTCCACAAAGTTTCATTAAAGACGATTCAATAGACAATAAGGTACTAACGCGTTCTGATAAAAT
CGTGGTAAATCGGATAACGTTCCAAGTGAAGAAGTAGTCAAAAAGATGAAAACTATTGGAGACAACCTTCTAAACGC
CAAGTTAATCACTCAACGTAAGTTTGATAATTTAACGAAAAGCTGAACGTGGAGGTTTGAGTGAACCTGATAAAGCTG
GTTTTATCAAACGCCAATTGGTTGAAACTCGCCAAATCACTAAGCATGTGGCACAAATTTTGATAGTCGCATGAAT
ACTAAATACGATGAAAATGATAAACTTATTCGAGAGGTTAAAGTGATTACCTTAAATCTAAATTAGTTTCTGACTT
CCGAAAAGATTTCCAATTCTATAAAGTACGTGAGATTAACAATTACCATCATGCCATGATGCGTATCTAAATGCCG
TCGTTGGAAGTGCTTTGATTAAGAAATATCCAAAAGTGAATCGGAGTTTGTCTATGGTGATTATAAAGTTTATGAT
GTTTCGTAAAATGATTGCTAAGTCTGAGCAAGAAATAGGCAAAGCAACCGCAAAATATTTCTTTTACTCTAATATCAT
GAACCTCTTCAAAACAGAAATTACACTTGCAAAATGGA

[0069] GAGATTGCGAAACGCCCTCTAATCGAACTAATGGGGAACTGGAGAAATTGTCTGGGATAAAGGGCG
AGATTTTGCCACAGTGCGCAAAGTATTGTCCATGCCCCAAGTCAATATTGTCAAGAAAACAGAAGTACAGACAGGC
GGATTCTCCAAGGAGTCAATTTTACCAAAAAGAAATTCGGACAAGCTTATTGCTCGTAAAAAAGACTGGGATCCAA
AAAAATATGGTGGTTTTGATAGTCCAACGGTAGCTTATTCAGTCCTAGTGGTTGCTAAGGTGGAAAAAGGGAAATC
GAAGAAGTTAAATCCGTTAAAGAGTTACTAGGGATCACAATTATGGAAAGAAGTTCCTTTGAAAAAATCCGATT
GACTTTTTTGAAGCTAAAGGATATAAGGAAGTTAAAAAAGACTTAATCATTAAACTACCTAAATATAGTCTTTTTG
AGTTAGAAAACGGTCGTAAACGGATGCTGGCTAGTGCCGGAAGATTACAAAAAGGAAATGAGCTGGCTCTGCCAAG
CAAATATGTGAATTTTTTATATTTAGCTAGTCATTATGAAAAGTTGAAGGGTAGTCCAGAAGATAACGAACAAAA
CAATTGTTTGTGGAGCAGCATAAGCATTATTTAGATGAGATTATTGAGCAAAATCAGTGAATTTTCTAAGCGTGTTA
TTTTAGCAGATGCCAATTTAGATAAAGTTCTTAGTGATATAACAAACATAGAGACAAACCAATACGTGAACAAGC
AGAAAATATTATTCATTTATTTACGTTGACGAATCTTGAGACTCCCGCTGCTTTTAAATATTTTGATACAACAATT
GATCGTAAACGATATACGTCTACAAAAGAAGTTTTAGATGCCACTCTTATCCATCAATCCATCACTGGTCTTTATG
AAACACGCATTGATTTGAGTCAGCTAGGAGGTGACTGA (SEQ ID NO:1)

MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFGSG
ETAETRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERH
PIFGNIVDEVAYHEKYPTIYHLRKKLADSTDKADRLIYLALAHMIKFRGHFLIEGDLNP
DNSDVDKLFQILVQIYNQLFEENPINASRVDAKAILSARLSKSRLENLIAQLPGEKRNG
LFGNLIASLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNL
SDAILLSDILRVNSEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPKEYKEIFFDQSK
NGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHL
GELHAILRRQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW
NFEEVVDKGASQSFIERMTNFDKNLPNEKVLPHSLLYEYFTVYNELTKVKYVTEGMRK
PAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGAYHD
LLKIIKDKDFLDNEENEDILEDIVLTLTLFEDRGMIEERLKTYAHLFDDKVMKQLKRRRY
TGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFQMLIHDDSLTFKEDIQKAQVSGQ
[0070] GHSLEQIANLAGSPAIKKGILQTVKIVDELVKVMGHKPENIV IEMARENQTTQKGQKNS
RERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYD
VDHIVPQSFIKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF
DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVIT
LKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTA LIKKYPKLESEFVYGDYKVYD
VRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANG
EIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNI VKKTEVQTGGFSKESIL
PKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVAKVEKGKSKKLKSVKELLGITIMER
SSFENPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSK
YVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEII EQISEFSKRVILADANLDKVL
SAYNKHDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATL IHQS
ITGLYETRIDLSQLGGD (SEQ ID NO:2)

[0071] (单底线:HNH结构域;双底线:RuvC结构域)

[0072] 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:

[0073] ATGGATAAAAGTATTCTATTGGTTTAGACATCGGCACTAATCCGTTGGATGGGCTGTCATAACCGAT
GAATACAAAGTACCTTCAAAGAAATTTAAGGTGTTGGGGAACACAGACCGTCATTGATTAATAAGAAATCTTATCGG
TGCCCTCCTATTGATAGTGGCGAAACGGCAGAGGCGACTCGCCTGAAACGAACCGCTCGGAGAAGGTATACACGTC
GCAAGAACCGAATATGTTACTTACAAGAAATTTTAGCAATGAGATGGCCAAAGTTGACGATTCTTTCTTTACCGT
TTGGAAGAGTCCTTCTTGTGCAAGAGGACAAGAAACATGAACGGCACCCCATCTTTGGAACATAGTAGATGAGGT
GGCATATCATGAAAAGTACCAACGATTTATCACCTCAGAAAAAAGCTAGTTGACTCAACTGATAAAGCGGACCTGA
GGTTAATCTACTTGGCTCTTGCCCATATGATAAAGTTCCGTGGGCACTTTCTCATTGAGGGTGATCTAAATCCGGAC

AACTCGGATGTCGACAACTGTTTCATCCAGTTAGTACAAACCTATAATCAGTTGTTTGAAGAGAACCCTATAAATGC
AAGTGGCGTGGATGCGAAGGCTATTCTTAGCGCCCGCCTCTCTAAATCCCGACGGCTAGAAAACCTGATCGCACAAT
TACCCGGAGAGAAGAAAAATGGGTTGTTTCGGTAACCTTATAGCGCTCTCACTAGGCCTGACACCAAATTTTAAGTCG
AACTTCGACTTAGCTGAAGATGCCAAATTGCAGCTTAGTAAGGACACGTACGATGACGATCTCGACAATCTACTGGC
ACAAATTGGAGATCAGTATGCGGACTTATTTTTGGCTGCCAAAAACCTTAGCGATGCAATCCTCCTATCTGACATAC
TGAGAGTTAATACTGAGATTACCAAGGCGCCGTTATCCGCTTCAATGATCAAAAGGTACGATGAACATCACCAAGAC
TTGACACTTCTCAAGGCCCTAGTCCGTCAGCAACTGCCTGAGAAATATAAGGAAATATTCTTTGATCAGTCGAAAA
CGGGTACGCAGGTTATATTGACGGCGGAGCGAGTCAAGAGGAATTCTACAAGTTTATCAAACCCATATTAGAGAAGA
TGGATGGGACGGAAGAGTTGCTTGTA AAACTCAATCGCGAAGATCTACTGCGAAAGCAGCGGACTTTTCGACAACGGT
AGCATTCCACATCAAATCCACTTAGGCGAATTGCATGCTATACTTAGAAGGCAGGAGGATTTTTATCCGTTCTCTAA
AGACAATCGTGAAGAGATTGAGAAAATCCTAACCTTTCGCATACCTTACTATGTGGGACCCCTGGCCCCGAGGGAAC
CTCGGTTTCGCATGGATGACAAGAAAGTCCGAAGAAACGATTACTCCAT

[0074] GGAATTTTGAGGAAGTTGTCGATAAAGGTGCGTCAGTCAATCGTTCATCGAGAGGATGACCAACTTTG
ACAAGAATTTACCGAACGAAAAAGTATTGCCTAAGCACAGTTTACTTTACGAGTATTTACAGTGTACAATGAATC
ACGAAAGTTAAGTATGTCACTGAGGGCATGCGTAAACCCGCCTTTCTAAGCGGAGAACAGAAGAAAGCAATAGTAGA
TCTGTTATTCAAGACCAACCGCAAAGTGACAGTTAAGCAATTGAAAGAGGACTACTTTAAGAAAATTGAATGCTTCG
ATTCTGTCGAGATCTCCGGGTAGAAGATCGATTTAATGCGTCACCTGGTACGTATCATGACCTCCTAAAGATAATT
AAAGATAAGGACTTCCTGGATAACGAAGAGAATGAAGATATCTTAGAAGATATAGTGTGACTCTTACCCTCTTTGA
AGATCGGGAAATGATTGAGGAAAGACTAAAAACATACGCTCACCTGTTGACGATAAGGTTATGAAACAGTTAAAGA
GGCGTCGCTATACGGGCTGGGGACGATTGTGCGGAAACTTATCAACGGGATAAGAGACAAGCAAAGTGGTAAACT
ATTCTCGATTTTCTAAAGAGCGACGGCTTCGCCAATAGGAACTTTATGCAGCTGATCCATGATGACTCTTTAACCTT
CAAAGAGGATATACAAAAGGCACAGGTTTCCGGACAAGGGGACTCATTGCACGAACATATTGCGAATCTTGCTGGTT
CGCCAGCCATCAAAAAGGCATACTCCAGACAGTCAAAGTAGTGGATGAGCTAGTTAAGGTCATGGGACGTCACAAA
CCGGAACATTGTAATCGAGATGGCACGCGAAAAATCAAACGACTCAGAAGGGGCAAAAAACAGTCGAGAGCGGAT
GAAGAGAATAGAAGAGGGTATTAAAGAACTGGGCAGCCAGATCTTAAAGGAGCATCCTGTGGAATAATCCCAATTGC
AGAACGAGAACTTTACCTCTATTACCTACAAAATGGAAGGGACATGTATGTTGATCAGGAACTGGACATAAACCGT
TTATCTGATTACGACGTCGATCACATTGTACCCCAATCCTTTTTGAAGGACGATTCAATCGACAATAAAGTGCTTAC
ACGCTCGGATAAGAACCAGGGGAAAAGTGACAATGTTCCAAGCGAGGAAGTCGTAAAGAAAATGAAGAACTATTGGC
GGCAGCTCCTAAATGCGAACTGATAACGCAAGAAAGTTTCGATAACTTAAAGCTGAGAGGGGTGGCTTGCT
GAACTTGACAAGGCCGATTTATTAAACGTCAGCTCGTGGAACCCGCCAAATCACAAAGCATGTTGCACAGATACT
AGATTCCCGAATGAATACGAAATACGACGAGAACGATAAGCTGATTCCGGGAAGTCAAAGTAATCACTTTAAAGTCAA
AATTGGTGTGCGACTTCAGAAAGGATTTTCAATTCTATAAAGTTAGGGAGATAAATAACTACCACCATGCGCACGAC
GCTTATCTTAATGCCGTCGTAGGGACCGCACTCATTAAAGAAATACCCGAAGCTAGAAAGTGAGTTTGTGTATGGTGA
TTACAAAGTTTATGACGTCCGTAAGATGATCGCGAAAAAGCGAACAGGAGATAGGCAAGGCTACAGCCAAATACTTCT
TTTATTCTAACATTATGAATTTCTTTAAGACGGAATCACTCTGGCAAACGGAGAGATACGCAAACGACCTTTAATT
GAAACCAATGGGGAGACAGGTGAAATCGTATGGGATAAGGGCCGGGACTTCGCGACGGTGAGAAAAGTTTTGTCCAT
GCCCCAAGTCAACATAGTAAAGAAAACCTGAGGTGCAGACCGGAGGGTTTTCAAAGGAATCGATTCTTCAAAAAGGA
ATAGTGATAAGCTCATCGCTCGTAAAAAGGACTGGGACCCGAAAAAGTACGGTGGCTTCGATAGCCCTACAGTTGCC
TATTCTGTCCTAGTAGTGGCAAAAGTTGAGAAGGGAAAAATCCAAGAACTGAAGTCAGTCAAAGAATTATTGGGGAT

AACGATTATGGAGCGCTCGTCTTTTAAAAAGAACCCC

[0075] ATCGACTTCCTTGAGGCGAAAGGTTACAAGGAAGTAAAAAAGGATCTCATAATTAACTACCAAAGTAT
AGTCTGTTTGAGTTAGAAAATGGCCGAAAACGGATGTTGGCTAGCGCCGGAGAGCTTCAAAGGGGAACGAACCTCGC
ACTACCGTCTAAATACGTGAATTTCTGTATTTAGCGTCCCATTACGAGAAGTTGAAAGGTTACCTGAAGATAACG
AACAGAAGCAACTTTTTGTTGAGCAGCACAAACATTATCTCGACGAAATCATAGAGCAAATTTTCGGAATTCAGTAAG
AGAGTCATCCTAGCTGATGCCAATCTGGACAAAAGTATTAAGCGCATACAACAAGCACAGGGATAAACCCATACGTGA
GCAGGCGGAAAATATTATCCATTTGTTTACTCTTACCAACCTCGGCGCTCCAGCCGCATTCAAGTATTTTGACACAA
CGATAGATCGCAAACGATACACTTCTACCAAGGAGGTGCTAGACGCGACACTGATTCACCAATCCATCACGGGATTA
TATGAAACTCGGATAGATTTGTACAGCTTGGGGGTGACGGATCCCCCAAGAAGAAGAGGAAAGTCTCGAGCGACTA
CAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGGCTGCAGGA

MDKKYSIGLAIGTNSVGWAVITDEYKVPSSKKFKVLGNTDRHSIKKNLIGALLFDSG
ETAETRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERH
PIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADRLIYLALAHMIKFRGHFLIEGDLNP
DNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKN
LFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNL
SDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSK
NGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHL
GELHAILRRQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW
NFEEVVDKGASQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRK
PAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHD
[0076] LLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRY
TGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQ
GDSLHEHIANLAGSPAIIKKGILQTVKVDELVKVMGRHKPENIVIAMARENQTTQKGQKN
SRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDY
DVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNPSEEVVKMKNYWRQLLNAKLITQRK
FDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVI
TLKSKLVSDFRKDFQFYKVBREINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVY
DVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDK
GRDFATVRKVLSPQVNIIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDS
PTVAYSVLVVAKEVGKSKKLKSVKELLGITIMERSSEKNPIDFLEAKGYKEVKKDLII
KLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFYLYASHYEKLKGSPEDEQKQ
[0077] LFVEQHKHYLDEIIIEQISEFSKRIVILADANLDKVL SAYNKHDKPI

[0078] REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDTLIIHQSI TGLYETRIDLSQLGGD

[0079] (单底线:HNH结构域;双底线:RuvC结构域)

[0080] 在一些实施方式中,野生型Cas9对应于来自化脓链球菌的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)。

[0081] ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGCGTCGGATGGGCGGTGATCACTGATGAATATAAGGTTCCGTCTAAAAAGTTCAAGGTTCTGGGAAATACAGACCGCCACAGTATCAAAAAAATCTTATAGGGCTCTTTTATTTGACAGTGGAGAGACAGCGGAAGCGACTCGTCTCAAACGGACAGCTCGTAGAAGGTATACACGTCGGAAGAATCGTATTTGTTATCTACAGGAGATTTTTTCAAATGAGATGGCGAAAGTAGATGATAGTTTCTTTCATCGACTTGAAGAGTCTTTTTTGGTGGAAGAAGACAAGAAGCATGAACGTCATCCTATTTTTGGAAATATAGTAGATGAAGTTGCTTATCATGAGAAATATCCAATCTATCATCTGCGAAAAAATTGGTAGATTCTACTGATAAAGCGGATTTGCGCTTAATCTATTTGGCCTTAGCGCATATGATTAAAGTTTCGTGGTCATTTTTTGATTGAGGGAGATTTAAATCCTGATAAGTGATGTGGACAACTATTTATCCAGTTGGTACAAACCTACAATCAATTATTTGAAGAAAACCTATTAACGCAAGTGAGTAGATGCTAAAGCGATTCTTTCTGCACGATTGAGTAAATCAAGACGATTAGAAAATCTCATTGCTCAGCTCCCCGGTGAGAAGAAAAATGGCTTATTTGGGAATCTCATTGCTTTGTCATTGGGTTTGACCCCTAATTTTAAATCAATTTTGATTGTCAGAGATGCTAAATTACAGCTTTCAAAGATACTTACGATGATGATTTAGATAATTTATTGGCGCAAATTGGAGATCAATATGCTGATTTGTTTTTGGCAGCTAAGAATTTATCAGATGCTATTTTACTTTTACAGATATCCTAAGAGTAAATACTGAAATAACTAAGGCTCCCTATCAGCTTCAATGATTAAACGCTACGATGAACATCATCAAGAC TTGACTCTTTTAAAAGCTTTAGTTTCGACAACAACTTCCAGAAAAGTATAAAGAAATCTTTTTTGATCAATCAAAAAACGGATATGCAGGTTATATTGATGGGGGAGCTAGCCAGAAGAATTTTATAAATTTATCAAACCAATTTTAGAAAAATGGATGGTACTGAGGAATTATTGGTGAAACTAAATCGTGAAGATTGCTGCGCAAGCAACGGACCTTTGACAACGGCTCTATTCCCATCAAATTCACCTTGGGTGAGCTGCATGCTATTTTGAGAAGACAAGAAGACTTTTATCCATTTTAAAGACAATCGTGAGAAGATTGAAAAATCTTGACTTTTCGAATTCCTTATTATGTTGGTCCATTGGCGCGTGGCAATAGTCGTTTTGTCATGGATGACTCGGAAGTCTGAAGAAACAATTACCCCATGGAATTTTGAAGAAGTTGTCGATAAAGGTGCTTCAGCTCAATCATTTATTGAACGCATGACAACTTTGATAAAAAATCTTCCAAATGAAAAAGTACTACCAAAACATAGTTTGCTTTATGAGTATTTTACGGTTTATAACGAATTGACAAAGGTCAAATATGTTACTGAAGGAATGCGAAAAC CAG

[0082] CATTTCTTTCAGGTGAACAGAAGAAAGCCATTGTTGATTTACTCTTCAAACAAATCGAAAAGTAACCGTTAAGCAATTAAGAAGATTATTTCAAAAAATAGAATGTTTTGATAGTGTGAAATTTTCAGGAGTTGAAGATAGATTTAATGCTTCATTAGGTACCTACCATGATTTGCTAAAAATTATTAAGATAAAGATTTTTTGGATAATGAAGAAATGAAGATATCTTAGAGGATATTGTTTTAACATTGACCTTATTTGAAGATAGGGAGATGATTGAGGAAAGACTTAAACATATGCTCACCTCTTTGATGATAAGGTGATGAAACAGCTTAAACGTCGCCGTTATACTGGTTGGGGACGTTTGTCTCGAAAATTGATTAATGGTATTAGGGATAAGCAATCTGGCAAAACAATATTAGATTTTTTGAATCAGATGGTTTTGCAATCGCAATTTTATGCAGCTGATCCATGATGATAGTTTGACATTTAAAGAAGACATTCAAAAAGCACAAGTGTCTGGACAAGGCGATAGTTTACATGAACATATTGCAAATTTAGCTGGTAGCCCTGCTATTAAAAAAGGTATTTTACAGACTGTAAAAGTTGTTGATGAATTGGTCAAAGTAATGGGGCGGCATAAGCCAGAAAATATCGTTATTGAAATGGCACGTGAAATCAGACAACCTCAAAGGGCCAGAAAAATTCGCGAGAGCGTATGAAACGAATCGAAGAAGGTATCAAAGAATTAGGAAGTCAGATTCTTAAAGAGCATCCTGTTGAAAAACTCAATTGCAAAATGAAAAGCTCTATCTCTATTATCTCCAAATGGAAGAGACATGTATGTGGACCAAGAATTAGATATTAATCGTTTAAAGTGATTATGATGTCGATCACATTGTTCCACAAAGTTTCCTTAAAGACGATTCAATAGACAATAAGGTCTTAACGCGTTCTGATAAAAAATCGTGGTAAATCGGATAACGTTCCAAGTGAAGAAGTAGTCAAAAAGATGAAAACTATTGGAGACAACCTCTAAACGCCAAGTTAATCACTCAACGTAAGTTTGATAATTTAACGAAAGCTGAACGTGGAGGTTTGAGTGAACCTGATAAAGCTGGTTTTATCAAACGCCA

ATTGGTTGAAACTCGCCAAATCACTAAGCATGTGGCACAAATTTTGGATAGTCGCATGAATACTAAATACGATGAAA
ATGATAAACTTATTCGAGAGGTTAAAGTGATTACCTTAAAAATCTAAATTAGTTTCTGACTTCCGAAAAGATTTCCAA
TTCTATAAAGTACGTGAGATTAACAATTACCATCATGCCCATGATGCGTATCTAAATGCCGTCGTTGGAAGTCTTT
GATTAAGAAATATCCAAAACCTGAATCGGAGTTTGTCTATGGTGATTATAAAGTTTATGATGTTCTGTAAGTATGATTG
CTAAGTCTGAGCAAGAAATAGGCAAAGCAACCGCAAAATATTTCTTTTACTCTAATATCATGAACCTTCTTCAAAACA
GAAATTACACTTGCAAATGGAGAGATTCGCAAACGCCCTCTAATCGAACTAATGGGGAAACTGGAGAAATTGTCTG
GGATAAAGGGCGAGATTTTGGCACAGTGCGCAAAGTATTGTCCATGCCCCAAGTCAATATTGTCAAGAAAACAGAAG
TACAGACAGGCGGATTCTCCAAGGAGTCAATTTTACCAAAAAGAAATTCGGACAAGCTTATTGCTCGTAAAAAGAC
TGGGATCCAAAAAATATGGTGGTTTTGATAGTCCAACGGTAGCTTATTCAGTCCTAGTGTTGCTAAGGTGGAAAA
AGGGAAATCGAAGAAGTTAAATCCGTTAAAGAGTTACTAGGGATCACAATTATGGAAAGAAGTTCCTTTGAAAAAA
ATCCGATTGACTTTTTAGAAGCTAAAGGATATAAGGAAGTTAAAAAAGACTTAATCATTAACCTACCTAAATATAGT
CTTTTTGAGTTAGAAAACGGTCGTAAACGGATGCTGGCTAGTGCCGGAGAATTACAAAAGGAAATGAGCTGGCTCT
GCCAAGCAAATATGTGAATTTTTTATATTTAGCTAGT

[0083] CATTATGAAAAGTTGAAGGGTAGTCCAGAAGATAACGAACAAAAACAATTGTTTGTGGAGCAGCATAAG
CATTATTTAGATGAGATTATTGAGCAAATCAGTGAATTTTCTAAGCGTGTTATTTTAGCAGATGCCAATTTAGATAA
AGTTCTTAGTGATATAACAAACATAGAGACAAACCAATACGTGAACAAGCAGAAAATATTATTCATTTATTTACGT
TGACGAATCTTGGAGCTCCCGCTGCTTTTAAATATTTTGATACAACAATTGATCGTAAACGATATACGTCTACAAAA
GAAGTTTTAGATGCCACTCTTATCCATCAATCCATCACTGGTCTTTATGAAACACGCATTGATTTGAGTCAGCTAGG
AGGTGACTGA

MDKKYSIGLDIGTNSVGWAVITDEYKVPSSKFKVLGNTDRHSIKKNLIGALLFDSG
ETAETRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERH
PIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADRLIYLALAHMIKFRGHFLIEGDLNP
DNSDVDKLFILVQTYNQLFEENPINASGVDAKAILSARLSKSRRENLI AQLPGEKKNG
LFGNLI ALSGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNL
SDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPKEYKEIFFDQSK
NGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHL

[0084] GELHAILRRQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW
NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPHSLLYEYFTVYNELTKVKYVTEGMRK
PAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHD
LLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRY
TGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQ
GDSLHEHIANLAGSPAIKKGILQTVKVVDLVKVMGRHKPENIVIEMARENQTTQKGQKN
SRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDY
DVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKMKKNYWRQLLNAKLITQRK

FDNLTKAERC GLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVI
TLKSKLVSDFRKDFQFYKVVREINNYHHAHDAYLNAVVGTA LKKYPKLESEFVYGDYKVI
DVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDK
GRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDS
PTVAYSVLVVAKEVGKSKKLKSVKELLGITIMERSSEKPNIDFLEAKGYKEVKKDLII
KLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLGSPEDNEQKQ
LFVEQHKHYLDEIIIEQISEFSKRVI LADANLDKVL SAYNKHDKPIREQAENI IHLFTLT
NLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQSI TGLYETRIDLSQLGGD (单底线:

HNH结构域;双底线:RuvC结构域)

[0085] 在一些实施方式中,Cas9是指来自: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*) (NCBI Ref:NC_021846.1);海豚链球菌 (*Streptococcus iniae*) (NCBI Ref:NC_021314.1);含羞草伯克霍尔德氏菌 (*Belliella baltica*) (NCBI Ref:NC_018010.1);扭曲冷弯曲菌I (*Psychroflexus torquis*) (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)的Cas9,或来自任何其他生物体的Cas9。

[0086] 在一些实施方式中,dCas9对应于或包括,具有一个或多个去活性化Cas9核酸酶活性突变的Cas9氨基酸序列的一部分或全部。例如,在一些实施方式中,dCas9结构域包括D10A和H840A的突变,或另一个Cas9中的对应突变。在一些实施方式中,dCas9包括dCas9 (D10A和H840A)的氨基酸序列:

[0087] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSG
ETAETRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERH

[0088]

PIFGNIVDEVAYHEKYPTIYHLRKKLV DSTDKADRLIYLALAHMIKFRGHFLIEGDLNP
DNSDVDKLF IQLVQTYNQ LFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKN
LFGNLIASLGLTPNFKS NFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNL
SDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLEPKYKEIFFDQSK
NGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHL
GELHAILRRQEDFYFPLKDNREKIEKILTFRIPIYYVGPLARGNSRFAWMTRKSEETITPW
NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRK
PAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHD
LLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRY
TGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFQMLIHDDSLTFKEDIQKAQVSGQ
GDSLHEHIANLAGSPAICKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKN
SRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDY
DVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKMKKNYWRQLLNAKLITQRK
FDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVI
TLKSKLVSDFRKDFQFYKVBREINNYHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVY
DVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDK
GRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDS
PTVAYSVLVVAKEVGKSKKLKSVKELLGITIMERSSEKPNIDFLEAKGYKEVKKDLII
KLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQ
LFVEQHKHYLDEIIEQISEFSKRVILADANLDKVL SAYNKHHRDKPI

[0089] REQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQSI TGLYETRIDLSQLGGD

[0090] (单底线: HNH结构域; 双底线: RuvC结构域)。

[0091] 在一些实施方式中, Cas9结构域包括D10A的突变, 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。

[0092] 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, 导致核酸酶去活化的Cas9 (dCas9)。此种突变, 举例来说, 包含在D10和H840处其他的氨基酸取代, 或在Cas9的核酸酶结构域内的其他取代 (例如, 在HNH核酸酶亚结构域和/或RuvC1亚结构域内的取代)。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。

[0093] 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序列的一。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。

[0094] 本文提供了适合的Cas9结构域和Cas9片段的示例性的氨基酸序列,并且另外的适合的Cas9结构域和片段的序列对本领域技术人员而言是显而易见的。

[0095] 在一些实施方式中,Cas9是指来自如下生物体的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);栖蝇蝇螺原体(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);亚马逊江豚链球菌/鱼型链球菌(NCBI Ref:NC_021314.1);波罗的海*Belliella*属菌(*Belliella baltica*)(NCBI Ref:NC_018010.1);扭曲冷弯曲菌I(*Psychroflexus torquis*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)。

[0096] 应当理解,另外的Cas9蛋白(例如核酸酶失效的Cas9(dCas9)、Cas9切口酶(nCas9)或核酸酶有活性的Cas9),包括其变体和同源物,是在本发明披露的范围的內。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,所述Cas9蛋白是核酸酶失效的Cas9(dCas9)。在一些实施方式中,所述Cas9蛋白是Cas9切口酶(nCas9)。在一些实施方式中,所述Cas9蛋白是核酸酶有活性的Cas9。

[0097] 示例性的催化失活的Cas9(dCas9):

[0098] DKKYSIGLAIGTNSVGWAVITDEYKVPSSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR
RYTRRKNRICYLQEIFSNEMAKVDDSFHRLSESLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVLDST
KADRLRIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLEN
LIAQLPGKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDQYDDDLNLLAQIGDQYADLFLAAKNLSAIL
LSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
ILEKMDGTEELLVKNLREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPLKDNREKIEKILTFRIPYYVGPL
ARGNSRFAWMTRKSEETITPWNFEVVDKGASQSFIERMTNFDKNLPNEKVLPHKSLLYEYFTVYNELTKVKYVTE
GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDN
EENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSD
GFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIIKKGILQTVKVVDELVKVMGRHKPENIVIE
ARENQTTQKGQKNSRERMKRIIEGIELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVDA
IVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFI
KRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKIVITLKSCLVSDFRKDFQFYKREINNYHHAHDAYLNAVVG
TALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGE
IVWDKGRDFATVRKVLSPQVNVKKTVEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVAK
VEKGKSKKLKSVKELLGITIMERSSEFKNPIDFLEAKGYKEVKKDLIIKLPKYSLEFENGRKRMLASAGELQKGNE
LALPSKYVNFLYLASHYEKLKGSPEDEQKQLFVEQHKHYLDEIIIEQISEFSKRVLADANLDKVL SAYNKHDKPI
REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDTLIIHQSI TGLYETRIDLSQLGGD

[0099] 示例性的催化的Cas9切口酶(nCas9):

[0100] DKKYSIGLAIGTNSVGWAVITDEYKVPSSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR

RYTRRNKRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLV DSTD
KADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLF IQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLEN
LIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKD TYDDDLNLLAQIGDQYADFLAAKNLSDA IL
LSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLEPKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKP
ILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPYYVGPL
ARGNSRFAWMTRKSEETITPWNFEEVVDKGASASQFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE
GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDN
EENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSD
GFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA IKKGILQTVKVDELVKVMGRHKPENIV IEM
ARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYD VDH
IVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFI
KRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS KLVSDFRKDFQFYK VREINNYHHAHDAYLNAVVG
TALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGE
IVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLV VAK
VEKGKSKKLKSVKELLGITIMERS SFEKNPIDFLEAKGYKEVKKDLIIKL PKYSLFELENGRKRMLASAGELQKGNE
LALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIEQISEFSKRVI LADANLDKVL SAYNKH RDKPI
REQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSI TGLYETRIDLSQLGGD

[0101] 示例性的催化Cas9:

[0102] DKKYSIGLDIGTNSVGWAVITDEYKVP SKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR
RYTRRNKRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLV DSTD
KADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLF IQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLEN
LIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKD TYDDDLNLLAQIGDQYADFLAAKNLSDA IL
LSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLEPKY

[0103] KEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHL
LGELHAILRRQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASASQF
IERMTNFDKNLPNEKVL PKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKED
YFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLF
DDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDS
LHEHIANLAGSPA IKKGILQTVKVDELVKVMGRHKPENIV IEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI
LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYD VDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVP
SEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDEND
KLIREVKVITLKS KLVSDFRKDFQFYK VREINNYHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVYDVRKMI
KSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTE
VQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLV VAKVEKGKSKKLKSVKELLGITIMERS SFE
KNPIDFLEAKGYKEVKKDLIIKL PKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPED
NEQKQLFVEQHKHYLDEIEQISEFSKRVI LADANLDKVL SAYNKH RDKPIREQAENI IHLFTLTNLGAPAAFKYF
DTTIDRKRYTSTKEVLDATLIHQSI TGLYETRIDLSQLGGD.

[0104] 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,“New CRISPR-Cas systems from uncultivated microbes.”Cell Res.2017Feb 21.doi:10.1038/cr.2017.21,其全部内容在此通过引用并入本文。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,发现了两个的前未知的系统,CRISPR-CasX和CRISPR-CasY,其为迄今为止所发现的最致密紧凑的系统的一。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。应所述理解,其他RNA-导向DNA结合蛋白可用作核酸可编程的DNA结合蛋白(napDNAbp),并且在本发明披露的范围内。

[0105] 在一些实施方式中,所述napDNAbp是Cas9结构域,例如核酸酶有活性的Cas9、Cas9切口酶(nCas9)或核酸酶失活的Cas9(dCas9)。核酸可编程的DNA结合蛋白的非限制性的例子包含,Cas9(例如dCas9和nCas9)、II型Cas效应蛋白、V型Cas效应蛋白、VI型Cas效应蛋白、CARF、DinG、其同源物或经修饰的或经工程化的版本。其他核酸可编程的DNA结合蛋白也在此公开披露的范围内,尽管它们可能在本发明披露中未被明确列出。参见例如Makarova等“Classification and Nomenclature of CRISPR-Cas Systems:Where from Here?”CRISPR J.2018Oct;1:325-336.doi:10.1089/crispr.2018.0033;Yan等,“Functionally diverse type V CRISPR-Cas systems”Science.2019Jan 4;363(6422):88-91.doi:10.1126/science.aav7271,其每个的全部内容在此通过引用并入本文。

[0106] 在一些实施方式中,本文所提供的任一融合蛋白的核酸可编程的DNA结合蛋白(napDNAbp)可以是CasX或CasY蛋白。在一些实施方式中,所述napDNAbp是CasX蛋白。在一些实施方式中,所述napDNAbp是CasY蛋白。在一些实施方式中,所述napDNAbp包括的氨基酸序列与天然存在的CasX或CasY蛋白的同一性是至少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,所述napDNAbp是天然存在的CasX或CasY蛋白。在一些实施方式中,所述napDNAbp包括的氨基酸序列与本文所描述的任何CasX或CasY蛋白的同一性是至少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%。应所述理解,来自其他细菌种类的CasX和CasY也可以根据本发明披露而被使用。

[0107] CasX(uniprot.org/uniprot/F0NN87;uniprot.org/uniprot/F0NH53)

[0108] >tr|F0NN87|F0NN87_SULIH CRISPR-关联Casx蛋白OS=冰岛硫化叶菌(Sulfolobus islandicus)(菌株HVE10/4)GN=SiH_0402PE=4SV=1

[0109] MEVPLYNIFGDNYIIQVATEAENSTIYNNKVEIDDEELRNVLNLAYKIAKNEDAAAERRGKAKKKKG
EEGETTTSNIIPLSGNDKNPWTETLKCYNFPTTVALSEVFNFSQVKECEEVSAPSFVKPEFYEFGRSPGMVERT
RRVKLEVEPHYLIIAAGWVLTGLGAKVSEG DYGVNVFTPTRGILYSLIQNVNGIVPGIKPETAFLWIARKVV
SSVTNPVSVVRIYTISDAVGQNPTTINGGFSIDLTKLLEKRYLLSERLEAIARNALSISSNMRRERYIVLANIYE
YLTG SKRLEDLLYFANRDLIMNLNSDDGKVRDLKLISAYVNGELIRGEG

[0110] >tr|F0NH53|F0NH53_SULIR CRISPR关联蛋白,Casx OS=冰岛硫化叶菌(菌株REY15A)GN=SiRe_0771 PE=4 SV=1

[0111] MEVPLYNIFGDNYIIQVATEAENSTIYNNKVEIDDEELRNVLNLAYKIAKNEDAAAERRGKAKKKKG
EEGETTTSNIIPLSGNDKNPWTETLKCYNFPTTVALSEVFNFSQVKECEEVSAPSFVKPEFYKFGRSPGMVERTRR

VKLEVEPHYLIMAAAGWVLTGLGKAKVSEGDYVGVNVFTPTRGILYSLIQNVNGIVPGIKPETAFLWIARKVVSSV
TNPVSVVSIYTTISDAVGQNPTTINGGFSIDLTKLLEKRDLLSERLEAIARNALSISSNMRRERYIVLANIYIYEYLTG
SKRLEDLLYFANRDLIMNLNSDDGKVRDLKLISAYVNGELIRGEG

[0112] CasY(ncbi.nlm.nih.gov/protein/APG80656.1)

[0113] >APG80656.1 CRISPR-关联蛋白CasY[未经培养的俭菌菌群细菌]

[0114] MSKRHPRISGVKGYRLHAQRLEYTGKSGAMRTIKYPLYSSPSGGRTVPREIVSAINDDYVGLYGLSNF
DDLNAEKRNEEKVYSVLDFWYDCVQYGAVFSYAPGLLKNVAEVRGGSYELTKTLKGSHLYDELQIDKVIKFLNK
KEISRANGSLDKLKKDIIDCFKAERYRERHKDQCNKLADDIKNAKKDAGASLGERQKKLFRDFFGISEQSENDKPSF
TNPLNLTCCLLPFDTVNNRNRGEVLFNKLKEYAQKLDKNEGSLEMWEYIGIGNSGTAFSNFLGEGFLGRLRENKI
TELKKAMMDITDAWRGQEQUEEELEKRLRILAALTIKLREPKFDNHWGGYRSDINGKLSSWLQNYINQTVKIKEDLK
GHKKDLKKAKEMINRFGESDTKEEAVVSSLLESIEKIVPDDSDADEKPDIPAIAIYRRFLSDGRLTLNRFVQREDV
QEALIKERLEAEKKKKPKKRKKKSDAEDEKETIDFKELFPHLAKPLKLVNPFYGD SKRELYKKYKNAAIYTDALWK
AVEKIYKSAFSSSLKNSFFD TDFDKDFFIKRLQKIFSVYRRFNTDKWKPIVKNSFAPYCDIVSLAENEVLYKPKQS
RSRKSAIDKNRVRLPSTENIAKAGIALARELSVAGFDWKDLLKKEEHEEYIDLIELHKTALALLAVTETQLDIS
ALDFVENGTVKDFMKTRDGNLVLEGRFLEMFSQSIVFSELRLAGLMSRKEFITRSAIQTMNGKQAELLYIPHEFQ
SAKITTPKEMSRAFLDLAPAEFATSLEPESLSEKSLKQMRYYPHYFGYELTRTGQIDGGVAENALRLEKSPV
KKREIKCKQYKTLGRGQNKIVLYVRSSYYQTQFLEWFLHRPKNVQTDVAVSGSFLIDEKKVKTRWNYDALTVALEP
VSGSERVFVSQPFTIFPEKSAEEEGQRYLGIDIGEYGIAYTALEITGDSAKILDQNFISDPQLKTLREEVKGLKLD
QRRGTFAMPSTKIARIRESLVHSLRNRIHHLALKHKAKIVYELEVS RFEEGKQKIKKVYATLKKADVYSEIDADKN
LQTTVWGKLAVASEISASYTSQFCGACKKLWRAEMQVDETITTQELIGTVRVIKGGTLIDAIKDFMRPPIFDENDT
PFPKYRDFCDKHHISKMRGNSCLFICPFCRANADADIQASQTIALLRVYVKEEKKVEDYFERFRKLKN
IKVLGQMKKI

[0115] 术语“CRISPR-Cas结构域”或“CRISPR-Cas DNA结合结构域”是指RNA-导向的蛋白，其包括CRISPR关联(Cas)蛋白或其片段(例如，一蛋白，其包括活性的、失活的或有部分活性的Cas蛋白的DNA切开结构域，和/或Cas蛋白的gRNA结合结构域)。CRISPR簇被转录并加工成为CRISPR RNA(crRNA)。CRISPR簇被转录并加工成为CRISPR RNA(crRNA)。在一些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Cas蛋白。所述tracrRNA作为核糖核酸酶3-辅助的pre-crRNA的加工的向导。随后，Cas9/crRNA和/或tracrRNA以核酸内裂解的方式切开与间隔序列互补的线性或环状dsDNA靶。实际上，DNA-结合和切开可能需要两个RNA。然而，可以工程化单个的向导RNAs(“sgRNA”或简称“gRNA”)以便将crRNA和tracrRNA的各方面都并入单一RNA种类的中。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本文。Cas蛋白识别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。CRISPR-Cas蛋白包含但不限于Cas9、CasX、CasY、Cpf1、C2c1和C2c3或其活性片段。基于本发明披露，另外的合适CRISPR-Cas蛋白和序列对于本领域技术人员将是显而易见的。

[0116] 核酸酶-去活化的CRISPR-Cas蛋白可互换地被称为“dCas”蛋白(表示核酸酶-“死(dead)”Cas)或催化失活的Cas。形成具有失活的DNA切开结构域的Cas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。HNH亚结构域切开与gRNA互补的股, 而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完全去活性化化脓链球菌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, Cas核酸酶具有失活的(例如去活性化的)DNA切开结构域, 即是说, 所述Cas为一切口酶, 被称为“nCas”蛋白(表示“切口酶(nickase)”Cas)。Cas变体与CRISPR-Cas蛋白或其片段共享同源性。Cas变体与野生型CRISPR-Cas蛋白是至少约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CRISPR-Cas蛋白相比, 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, Cas变体包括CRISPR-Cas的片段(例如gRNA可编程的DNA结合结构域或DNA-切开结构域), 使得所述片段与相应的野生型CRISPR-Cas蛋白是至少约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, 所述片段是相应的野生型CRISPR-Cas蛋白的氨基酸长度的至少约30%、至少约35%、至少约40%、至少约45%、至少约50%、至少约55%、至少约60%、至少约65%、至少约70%、至少约75%、至少约80%、至少约85%、至少约90%、至少约95%、至少约96%、至少约97%、至少约98%、至少约99%或至少约99.5%。在一些实施方式中, 所述片段的长度是至少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] 在此公开披露中, “包括(comprises)”, “包括(comprising)”, “含有(containing)”, “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“基本上由...组成(consists essentially)”同样地具有美国专利法中所赋予的含义, 并且所述术语是开放式的, 即允许存在多于本文所列举的, 只要其并不会改变本文所列举的基本或新颖的特征, 但不包括现有技术的实施方式。

[0118] “胞苷脱氨酶”是指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, 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, 其源自海七鳃鳗(海七鳃鳗胞嘧啶脱氨酶, 即“PmCDA1”), AID(活化-诱导的胞苷脱氨酶; AICDA), 其源自哺乳动物(例如, 人类、猪、牛、马、猴等), 以及APOBEC是示例性的胞苷脱氨酶。

[0119] 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如下所示:

[0120] >tr|A5H718|A5H718_PETMA胞嘧啶脱氨酶OS=海七鳃鳗OX=7757PE=2SV=1

[0121] MTDAEYVRIHEKLDIYTFKKQFFNNKKS VSHRCYVLFELKRRGERRACFWGYAVNKPQSGTERGIHAEI

FSIRKVEEYLRDNPQGFTINWYSSWSPCADCAEKILEWYNQELRNGHTLKIWACKLYYEKNARNQIGLWNLRDNGV
GLNVMVSEHYQCCRKIFIQSSHNQLNENRWLEKTLKRAEKRRSELSIMIQQVILHTTKSPAV

[0122] >EF094822.1海七鳃鳗分离物PmCDA.21胞嘧啶脱氨酶mRNA,完整的cds

[0123] TGACACGACACAGCCGTGTATATGAGGAAGGGTAGCTGGATGGGGGGGGGGGAATACGTTTCAGAGAGG
ACATTAGCGAGCGTCTTGTTGGTGGCCTTGAGTCTAGACACCTGCAGACATGACCGACGCTGAGTACGTGAGAATCC
ATGAGAAGTTGGACATCTACACGTTTAAGAAACAGTTTTTCAACAACAAAAAATCCGTGTCGCATAGATGCTACGTT
CTCTTTGAATTAACGACGGGGTGAACGTAGAGCGTGTGTTTGGGGCTATGCTGTGAATAAACACAGAGCGGGAC
AGAACGTGGAATTCACGCCGAAATCTTTAGCATTAGAAAAGTCGAAGAATACCTGCGCGACAACCCCGGACAATTCA
CGATAAATTGGTACTCATCCTGGAGTCCTTGTCAGATTGCGCTGAAAAGATCTTAGAATGGTATAACCAGGAGCTG
CGGGGGAACGGCCACACTTTGAAAATCTGGGCTTGCAAACTCTATTACGAGAAAAATGCGAGGAATCAAATTGGGCT
GTGGAACCTCAGAGATAACGGGGTTGGGTGAATGTAATGGTAAGTGAACACTACCAATGTTGCAGGAAAAATATTCA
TCCAATCGTCGCACAATCAATTGAATGAGAATAGATGGCTTGAGAAGACTTTGAAGCGAGCTGAAAAACGACGGAGC
GAGTTGTCCATTATGATTACAGGTAAAAATACTCCACACCACTAAGAGTCCTGCTGTTTAAGAGGCTATGCGGATGGT
TTTC

[0124] >tr|Q6QJ80|Q6QJ80_人类活化-诱导的胞苷脱氨酶OS=智人OX=9606GN=AICDA
PE=2SV=1

[0125] MDSLLMNRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNKGHVLELLFLRYISDWD
LDPGRCYRVTFWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLYFCEDRKAEPGLRRLHRAGVQIAIMTFKAPV

[0126] >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

[0127] AGAGAACCATCATTAATTGAAGTGAGATTTTTCTGGCCTGAGACTTGCAGGGAGGCAAGAAGACACTCT
GGACACCACTATGGACAGGTAAAGAGGCAGTCTTCTCGTGGGTGATTGCACTGGCCTTCCTCTCAGAGCAAATCTGA
GTAATGAGACTGGTAGCTATCCCTTTCTCTCATGTAAGTGTCTGACTGATAAGATCAGCTTGATCAATATGCATATA
TATTTTTTGATCTGTCTCCTTTTCTTCTATTGATCTTATACGCTGTCAGCCCAATTCTTTCTGTTTCAGACTTCT
CTTGATTTCCCTCTTTTTCATGTGGCAAAAAGAGTAGTGCGTACAATGTACTGATTCGTCCTGAGATTTGTACCATG
GTTGAAACTAATTTATGGTAATAATATTAACATAGCAAATCTTTAGAGACTCAAATCATGAAAAGGTAATAGCAGTA
CTGTACTAAAAACGGTAGTGCTAATTTTCGTAATAATTTGTAAATATTCAACAGTAAACAACCTGAAGACACACT
TTCCTAGGGAGGCGTTACTGAAATAATTTAGCTATAGTAAGAAAATTTGTAATTTTAGAAATGCCAAGCATTCTAAA
TTAATTGCTTGAAAGTCACTATGATTGTGTCCATTATAAGGAGACAAATTCATTCAAGCAAGTTATTTAATGTTAAA
GGCCCAATTGTTAGGCAGTTAATGGCACTTTTACTATTAATAATCTTTCCATTTGTTTCAGACGTAGCTTAACCTAC
CTCTTAGGTGTGAATTTGGTTAAGGTCCTCATAATGTCTTTATGTGCAGTTTTTGATAGGTTATTGTCATAGAAGTT
ATTCTATTCCCTACATTTATGATTACTATGGATGTATGAGAATAACACCTAATCCTTATACTTTACCTCAATTTAACT
CCTTTATAAAGAACTTACATTACAGAATAAAGATTTTTTAAAAATATATTTTTTTGTAGAGACAGGGTCTTAGCCCA
GCCGAGGCTGGTCTCTAAGTCCTGGCCCAAGCGATCCTCCTGCCTGGGCCTCCTAAAGTGCTGGAATTATAGACATG
AGCCATCACATCCAATATACAGAATAAAGATTTTAAATGGAGGATTTAATGTTCTTCAGAAAATTTTCTTGAGGTCA
GACAATGTCAAATGTCTCCTCAGTTTACACTGAGATTTTGAACAAGTCTGAGCTATAGGTCCTTGTAAGGGTCC
ATTGGAATACTTGTTCAAAGTAAATGGAAAGCAAAGGTAAATCAGCAGTTGAAATTCAGAGAAAGACAGAAAAG
GAGAAAAGATGAAATTCAACAGGACAGAAGGGAAATATATTATCATTAAAGGAGGACAGTATCTGTAGAGCTCATTAG
TGATGGCAAAATGACTTGGTCAGGATTATTTTAACCCGCTGTTTCTGGTTTGACGGCTGGGGATGCAGCTAGGG

TTCTGCCTCAGGGAGCACAGCTGTCCAGAGCAGCTGTCAGCCTGCAAGCCTGAAACACTCCCCTCGGTAAAGTCCTTC
CTACTCAGGACAGAAATGACGAGAACAGGGAGCTGGAACAGGCCCTAACCCAGAGAAGGGAAGTAATGGATCAACA
AAGTTAACTAGCAGGTCAGGATCACGCAATTCATTTCACTCTGACTGGTAACATGTGACAGAAACAGTGTAGGCTTA
TTGTATTTTCATGTAGAGTAGGACCCAAAAATCCACCCAAAGTCCTTTATCTATGCCACATCCTTCTTATCTATACT
TCCAGGACACTTTTTCTTCCTTATGATAAGGCTCTCTCTCTCTCCACACACACACACACACACACACACACACAC
ACACACACACACAAACACACACCCCGCCAACCAAGGTGCATGTAAAAAGATGTAGATTCCCTCTGCCTTTCTCATCTA
CACAGCCCAGGAGGGTAAGTTAATATAAGAGGGATTTATTGGTAAGAGATGATGCTTAATCTGTTTAACTGGGCC
TCAAAGAGAGAATTTCTTTCTTCTGTACTTATTAAGCACCTATTATGTGTTGAGCTTATATATACAAAGGGTTATT
ATATGCTAATATAGTAATAGTAATGGTGGTTGGTACTATGGTAATTACCATAAAAAATTATTATCCTTTTAAAAATAAA
GCTAATTATTATTGGATCTTTTTTAGTATTCATTTTATGTTTTTATGTTTTTGATTTTTTAAAAGACAATCTCACC
CTGTTACCCAGGCTGGAGTGCAGTGGTGCATCATAGCTTTCTGCAGTCTGAACTCCTGGGCTCAAGCAATCCTCC
TGCCTTGGCCTCCCAAAGTGTTGGGATACAGTCATGAGCCACTGCATCTGGCCTAGGATCCATTTAGATTAAATAT
GCATTTTAAATTTTAAAAATAATATGGCTAATTTTACCTTATGTAATGTGTATACTGGCAATAAATCTAGTTTGCTG
CCTAAAGTTTAAAGTGCTTTCCAGTAAGCTTCATGTACGTGAGGGGAGACATTTAAAGTGAAACAGACAGCCAGGTG
TGGTGGCTCACGCCTGTAATCCCAGCACTCTGGGAGGCTGAGGTGGGTGGATCGCTTGAGCCCTGGAGTTCAAGACC
AGCCTGAGCAACATGGCAAAACGCTGTTTCTATAACAAAAAATTAGCCGGGCATGGTGGCATGTGCCTGTGGTCCCAG
CTACTAGGGGGCTGAGGCAGGAGAATCGTTGGAGCCCAGGAGGTCAAGGCTGCACTGAGCAGTGCTTGCGCCACTGC
ACTCCAGCCTGGGTGACAGGACCAGACCTTGCCTCAAAAAAATAAGAAGAAAAATTAAAAATAAATGGAAACAATA
CAAAGAGCTGTTGTCTAGATGAGCTACTTAGTTAGGCTGATATTTGGTATTTAACTTTTAAAGTCAGGGTCTGTC
ACCTGCACTACATTATTAATAATATCAATTCTCAATGTATATCCACACAAAGACTGGTACGTGAATGTTCATAGTACC
TTTATTACAAAAACCCCAAAGTAGAGACTATCCAAATATCCATCAACAAGTGAACAAATAAACAAAATGTGCTATAT
CCATGCAATGGAATACCACCCTGCAGTACAAAGAAGCTACTTGGGGATGAATCCCAAAGTCATGACGCTAAATGAAA
GAGTCAGACATGAAGGAGGAGATAATGTATGCCATACGAAATCTAGAAAATGAAAGTAATTATAGTTACAGAAAG
CAAATCAGGGCAGGCATAGAGGCTCACACCTGTAATCCCAGCACTTTGAGAGGCCACGTGGGAAGATTGCTAGAACT
CAGGAGTTCAAGACCAGCCTGGGCAACACAGTGAACTCCATTCTCCACAAAAATGGGAAAAAAGAAAGCAAATCA
GTGTTGTCTGTGGGGAGGGGAAGGACTGCAAGAGGGGAAGAAGCTCTGGTGGGGTGAGGGTGGTGATTGAGGTTT
TGTATCCTGACTGTGGTAGCAGTTTGGGGTGTTTACATCCAAAAATATTCTGTAAGTAATTATGCATCTTAAATGGGTGG
AGTTTACTGTATGTAAATTATACCTCAATGTAAAGAAAAAATAATGTGTAAGAAAACCTTTCAATTCTCTTGCCAGCAA
ACGTTATTCAAATTCCTGAGCCCTTTACTTCGCAAAATCTCTGCACTTCTGCCCCGTACCATTAGGTGACAGCACTA
GCTCCACAAATTGGATAAATGCATTTCTGGAAAAGACTAGGGACAAAATCCAGGCATCACTTGTGCTTTTCATATCAA
CCATGCTGTACAGCTTGTGTTGCTGTCTGCAGCTGCAATGGGGACTCTTGATTTCTTTAAGGAACTTGGGTTACCA
GAGTATTTCCACAAATGCTATTCAAATTAGTGCTTATGATATGCAAGACACTGTGCTAGGAGCCAGAAAACAAAGAG
GAGGAGAAATCAGTCATTATGTGGGAACAACATAGCAAGATATTTAGATCATTTTGAAGTAAAGGAGCAGCAGA
GTACAAAATCACACATGCAATCAGTATAATCCAAATCATGTAAATATGTGCCTGTAGAAAGACTAGAGGAATAAACA
CAAGAATCTTAACAGTCATTGTGATTAGACACTAAGTCTAATTATTATTATTAGACACTATGATATTTGAGATTTAA
AAAATCTTTAATATTTTAAATTTAGAGCTCTTCTATTTTCCATAGTATTCAAGTTTGACAATGATCAAGTATTAC
TCTTTCTTTTTTTTTTTTTTTTTTTTTTTTTTTTGGAGATGGAGTTTGGTCTTGTGCCCATGCTGGAGTGGAATGGCAT
GACCATAGCTCACTGCAACCTCCACCTCCTGGGTTCAAGCAAAGCTGTCGCCTCAGCCTCCCGGGTAGATGGGATTA
CAGGCGCCACCACCACACTCGGCTAATGTTTGTATTTTAGTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGGT

CTCAAACCTCCTGACCTCAGAGGATCCACCTGCCTCAGCCTCCCAAAGTGCTGGGATTACAGATGTAGGCCACTGCGC
CCGGCCAAGTATTGCTCTTATACATTAAAAAACAGGTGTGAGCCACTGCGCCAGCCAGGTATTGCTCTTATACATT
AAAAAATAGGCCGGTGCAGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAAGCCAAGGCGGGCAGAACACCCGAGG
TCAGGAGTCCAAGGCCAGCCTGGCCAAGATGGTGAACCCCGTCTCTATTAAAAATACAAACATTACCTGGGCATGA
TGGTGGGCGCCTGTAATCCCAGCTACTCAGGAGGCTGAGGCAGGAGGATCCGCGGAGCCTGGCAGATCTGCCTGAGC
CTGGGAGGTTGAGGCTACAGTAAGCCAAGATCATGCCAGTATACTTCAGCCTGGGCGACAAAGTGAGACCGTAACAA
AAAAAAAAAATTTAAAAAAGAAATTTAGATCAAGATCCAAGTGTAAAAAGTGGCCTAAACACCACATTAAAGAGT
TTGGAGTTTATTCTGCAGGCAGAAGAGAACCATCAGGGGGTCTTCAGCATGGGAATGGCATGGTGCACCTGGTTTTT
GTGAGATCATGGTGGTGACAGTGTGGGGAATGTTATTTTGGAGGGACTGGAGGCAGACAGACCGTTAAAAGGCCAG
CACAACAGATAAGGAGGAAGAAGATGAGGGCTTGGACCGAAGCAGAGAAGAGCAAACAGGGAAGGTACAAATTCAAG
AAATATTGGGGGGTTTGAATCAACACATTTAGATGATTAATTAATATGAGGACTGAGGAATAAGAAATGAGTCAAG
GATGGTTCCAGGCTGCTAGGCTGCTTACCTGAGGTGGCAAAGTCGGGAGGAGTGGCAGTTTAGGACAGGGGGCAGTT
GAGGAATATTGTTTTGATCATTTTTGAGTTTGGGTACAAAGTTGGACACTTAGGTAAAGACTGGAGGGGAAATCTGAA
TATACAATTATGGGACTGAGGAACAAGTTTATTTATTTTTGTTTCGTTTTCTTGTGAAGAACAAATTTAATTGT
AATCCCAAGTCATCAGCATCTAGAAGACAGTGGCAGGAGGTGACTGTCTTGTGGGTAAGGGTTTGGGGTCCTTGATG
AGTATCTCTCAATTGGCCTTAAATATAAGCAGGAAAAGGAGTTTATGATGGATTCCAGGCTCAGCAGGGCTCAGGAG
GGCTCAGGCAGCCAGCAGAGGAAGTCAGAGCATCTTCTTGGTTTAGCCCAAGTAATGACTTCCTTAAAAAGCTGAA
GGAAAATCCAGAGTGACCAGATTATAAACTGTACTCTTGCATTTTCTCTCCCTCCTCTACCCACAGCCTCTTGATG
AACCGGAGGAAGTTTCTTTACCAATTCAAAAATGTCCGCTGGGCTAAGGGTCGGCGTGAGACCTACCTGTGCTACGT
AGTGAAGAGGCGTGACAGTGCTACATCCTTTTCACTGGACTTTGGTTATCTTCGCAATAAGGTATCAATTAAAGTCG
GCTTTGCAAGCAGTTTAATGGTCAACTGTGAGTGCTTTTAGAGCCACCTGCTGATGGTATTACTTCCATCCTTTTTT
GGCATTGTGTCTCTATCACATTCTCAAATCCTTTTTTTTTATTTCTTTTTCCATGTCCATGCACCCATATTAGACA
TGGCCCAAAATATGTGATTTAATTCCTCCCCAGTAATGCTGGGCACCCTAATACCACTCCTTCCTTCAGTGCCAAGA
ACAACTGCTCCCAAAGTGTACCAGCTTTCCTCAGCATCTGAATTGCCTTTGAGATTAATTAAGCTAAAAGCATTT
TTATATGGGAGAATATTATCAGCTTGTCCAAGCAAAAATTTAAATGTGAAAAACAAATTGTGTCTTAAGCATTTTT
GAAATTAAGGAAGAAGAATTTGGGAAAAAATTAACGGTGGCTCAATTCGTCTTCCAAATGATTTCTTTTCCCTCC
TACTCACATGGGTCGTAGGCCAGTGAATACATTCAACATGGTGATCCCCAGAAAAGTCAAGAGAAGCCTCGGCTGATG
ATTAATTAATTTGATCTTTTCGGCTACCCGAGAGAATTACATTTCCAAGAGACTTCTTACCAAAAATCCAGATGGGTT
TACATAAACTTCTGCCCACGGGTATCTCCTCTCTCCTAACACGCTGTGACGTCTGGGCTTGGTGGAATCTCAGGGAA
GCATCCGTGGGGTGGAAGGTCATCGTCTGGCTCGTTGTTGATGGTTATATTACCATGCAATTTTCTTTGCCTACAT
TTGTATTGAATACATCCCAATCTCCTTCCTATTCGGTGACATGACACATTCTATTTCAGAAGGCTTTGATTTTATCA
AGCACTTTTCAATTTACTTCTCATGGCAGTGCTATTACTTCTCTTACAATACCCATCTGTCTGCTTTACCAAAATCTA
TTTCCCTTTTTCAGATCCTCCCAAATGGTCCTCATAAACTGTCCTGCCTCCACCTAGTGGTCCAGGTATATTTCCAC
AATGTTACATCAACAGGCACCTTCTAGCCATTTTCTTCTCAAAAGGTGCAAAAAGCAACTTCATAAACACAAATTA
ATCTTCGGTGAGGTAGTGTGATGCTGCTTCTCCTCCCAACTCAGCGCACTTCGTCTTCTCATTCCACAAAAACCCATA
GCCTTCTTCACTCTGCAGGACTAGTGCTGCCAAGGGTTCAGCTCTACCTACTGGTGTGCTCTTTTGAGCAAGTTGC
TTAGCCTCTCTGTAACACAAGGACAATAGCTGCAAGCATCCCCAAAGATCATTGCAGGAGACAATGACTAAGGCTAC
CAGAGCCGCAATAAAAGTCAGTGAATTTTAGCGTGGTCTCTCTGTCTCTCCAGAACGGCTGCCACGTGGAATTGCT
CTTCTCCGCTACATCTCGGACTGGGACCTAGACCCTGGCCGCTGCTACCGCGTCACCTGGTTCACCTCCTGGAGCC

CCTGCTACGACTGTGCCCCGACATGTGGCCGACTTTCTGCGAGGGAACCCCAACCTCAGTCTGAGGATCTTCACCGCG
CGCCTCTACTTCTGTGAGGACCGCAAGGCTGAGCCCGAGGGGCTGCGGCGGCTGCACCGCGCCGGGGTGCAAATAGC
CATCATGACCTTCAAAGGTGCGAAAGGGCCTTCCGCGCAGGCGCAGTGCAGCAGCCCGCATTTCGGGATTGCGATGCG
GAATGAATGAGTTAGTGGGGAAGCTCGAGGGGAAGAAGTGGGCGGGGATTCTGGTTCACCTCTGGAGCCGAAATTAA
AGATTAGAAGCAGAGAAAAAGAGTGAATGGCTCAGAGACAAGGCCCGAGGAAATGAGAAAATGGGGCCAGGGTTGCT
TCTTTCCCTCGATTTGGAACCTGAACTGTCTTCTACCCCCATATCCCCGCCTTTTTTTCCTTTTTTTTTTTTTGAA
GATTATTTTTACTGCTGGAATACTTTTGTAGAAAAACCACGAAAGAACTTTCAAAGCCTGGGAAGGGCTGCATGAAAA
TTCAGTTCGTCTCTCCAGACAGCTTCGGCGCATCCTTTTGGTAAGGGGCTTCCTCGCTTTTTAAATTTTCTTTCTTT
CTCTACAGTCTTTTTTGGAGTTTCGTATATTTCTTATATTTTCTTATTGTTCAATCACTCTCAGTTTTTCATCTGATG
AAAACTTTATTTCTCCTCCACATCAGCTTTTTCTTCTGCTGTTTCACCATTCAAGAGCCCTCTGCTAAGGTTCTTTTT
CCCTCCCTTTTTCTTTCTTTTGTGTTTTCACATCTTTAAATTTCTGTCTCTCCCCAGGGTTGCGTTTCTTCTGCTC
AGAATTCTTTTCTCCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTAAACAAACAAACAAAAACCCAAAAAACTCTTTC
CCAATTTACTTTCTTCCAACATGTTACAAAGCCATCCACTCAGTTTAGAAGACTCTCCGGCCCCACCGACCCCAAC
CTCGTTTTGAAGCCATTCACTCAATTTGCTTCTCTCTTCTCTACAGCCCCTGTATGAGGTTGATGACTTACGAGAC
GCATTTTCGTACTTTGGGACTTTGATAGCAACTTCCAGGAATGTCACACACGATGAAATATCTCTGCTGAAGACAGTG
GATAAAAAACAGTCCTTCAAGTCTTCTCTGTTTTTATTCTTCAACTCTCACTTTCTTAGAGTTTACAGAAAAATAT
TTATATACGACTCTTTAAAAAGATCTATGTCTTGAATAAGAGAAGGAACACAGGTCTGGCCAGGGACGTGCTGCAA
TTGGTGCAGTTTTGAATGCAACATTGTCCCCTACTGGGAATAACAGAACTGCAGGACCTGGGAGCATCCTAAAGTGT
CAACGTTTTTCTATGACTTTTAGGTAGGATGAGAGCAGAAGGTAGATCCTAAAAAGCATGGTGAGAGGATCAAATGT
TTTTATATCAACATCCTTTATTATTTGATTCATTTGAGTTAACAGTGGTGTAGTGATAGATTTTTCTATTCTTTTC
CCTTGACGTTTACTTTCAAGTAACACAAACTCTTCCATCAGGCCATGATCTATAGGACCTCCTAATGAGAGTATCTG
GGTGATTGTGACCCCAAAACCATCTCTCCAAAGCATTAAATATCCAATCATGCGCTGTATGTTTTAATCAGCAGAAGCA
TGTTTTTATGTTTGTACAAAAGAAGATTGTTATGGGTGGGGATGGAGGTATAGACCATGCATGGTCACCTTCAAGCT
ACTTTAATAAAGGATCTTAAAAATGGGCAGGAGGACTGTGAACAAGACACCCTAATAATGGGTTGATGTCTGAAGTAG
CAAATCTTCTGGAAACGCAAACTCTTTAAGGAAGTCCCTAATTTAGAAACACCCACAACTTCACATATCATAATT
AGCAAACAATTGGAAGGAAGTTGCTTGAATGTTGGGGAGAGGAAAATCTATTGGCTCTCGTGGGTCTCTTCATCTCA
GAAATGCCAATCAGGTCAAGGTTTGCTACATTTTGTATGTGTGTGATGCTTCTCCCAAAGGTATATTAATACTATATAA
GAGAGTTGTGACAAAACAGAATGATAAAGCTGCGAACCGTGGCACACGCTCATAGTTCTAGCTGCTTGGGAGGTTGA
GGAGGGAGGATGGCTTGAACACAGGTGTTCAAGGCCAGCCTGGGCAACATAACAAGATCCTGTCTCTCAAAAAAAAA
AAAAAAAAAAAAAGAGAGAGAGAGAGGGCCGGCGTGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCGAGCCGG
GCGGATCACCTGTGGTCAGGAGTTTGAGACCAGCCTGGCCAACATGGCAAAACCCCGTCTGTACTCAAAATGCAAAA
ATTAGCCAGGCGTGGTAGCAGGCACCTGTAATCCCAGCTACTTGGGAGGCTGAGGCAGGAGAATCGCTTGAACCCAG
GAGGTGGAGGTTGCAGTAAGCTGAGATCGTGCCGTTGCACTCCAGCCTGGGCGACAAGAGCAAGACTCTGTCTCAGA
AAAAAAAAAAAAAGAGAGAGAGAGAGAAAAGAGAACAATATTTGGGAGAGAAGGATGGGGAAGCATTGCAAGGAAA
TTGTGCTTTATCCAACAAAATGTAAGGAGCCAATAAGGGATCCCTATTTGTCTCTTTTGGTGTCTATTTGTCCCTAA
CAACTGTCTTTGACAGTGAGAAAAATATTCAGAATAACCATATCCCTGTGCCGTTATTACCTAGCAACCCTTGCAAT
GAAGATGAGCAGATCCACAGGAAAACCTGAATGCACAACTGTCTATTTTAATCTTATTGTACATAAGTTTGTA AAA
GAGTTAAAAATTGTTACTTCATGTATTCAATTATATTTATATTTTTCGCTCTAATGATTTTTTTATTAACATGAT
TTCCTTTTCTGATATATTGAAATGGAGTCTCAAAGCTTCATAAATTTATAACTTTAGAAATGATTCTAATAACAACG

TATGTAATTGTAACATTGCAGTAATGGTGCTACGAAGCCATTTCTCTTGATTTTTAGTAACTTTTATGACAGCAAA
TTTGCTTCTGGCTCACTTTCAATCAGTTAAATAAATGATAAATAATTTTGAAGCTGTGAAGATAAAATACCAAATA
AAATAATATAAAAGTGATTATATGAAGTTAAAAATAAAAAATCAGTATGATGGAATAAACTTG

[0128] 载脂蛋白B mRNA编辑酶催化多肽样(APOBEC)蛋白是进化上保守的胞苷脱氨酶家族。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,所述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。一些修饰过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、以及YEE-BE3,其可从Addgene(质粒85169、85170、85171、85172、85173、85174、85175、85176、85177)获得。

[0129] 下面提供了根据本发明披露的方面可以是融合到Cas9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,没有定位信号的结构域(核定位序列,其无核输出信号,细胞质定位信号)。

[0130] 人类AID:

MDSLLMNRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNKGCH

[0131] VELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGPNLSLRIFTARLYFC
EDRKAEPGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHENSVRLSRQL
RRILLPLYEVDDLRDAFRTLGL (单底线:核定位序列;双底线:核输出信号)

[0132] 小鼠AID:

MDSLLMKQKKFLYHFKNVRWAKGRHETLYCYVVKRRDSATSCSLDFGHLRNKSGCHVELLFLRYIS

[0133] DWDLDPGRCYRVTWFTSWSPCYDCARHVAEFLRWPNLSLRIFTARLYFCEDRKAEPGLRRLHRAGVQI
GIMTFKDYFYCWNTFVENRERTFKAWGLHENSVRLTRQLRRILLPLYEVDDLRDAFRMLGF (单底
线:核定位序列;双底线:核输出信号)

[0134] 犬AID:

MDSLLMKQRKFLYHFKNVRWAKGRHETLYCYVVKRRDSATSFSLDFGHLRNKSGCHVELLFLRYIS

[0135] DWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGYPNLSLRIFAARLYFCEDRKAEPGLRRLHRAGVQI
AIMTFKDYFYCWNTFVENREKTFKAWEGLHENSVRLSRQLRRILLPLYEVDDLRDAFRTLGL (单底
线:核定位序列;双底线:核输出信号)

[0136] 牛AID:

MDSLLKKQRQFLYQFKNVRWAKGRHETLYCYVVKRRDSPTSFSLDFGHLRNKAGCHVELLFLRYIS

[0137] DWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGYPNLSLRIFTARLYFCDKERKAEPGLRRLHRAGVQ
IAIMTFKDYFYCWNTFVENHERTFKAWEGLHENSVRLSRQLRRILLPLYEVDDLRDAFRTLGL (单底
线:核定位序列;双底线:核输出信号)

[0138] 大鼠AID

- [0139] MAVGSKPKAALVGPHWERERIWCFLCSTGLGTQQTGQTSRWLRPAATQDPVSPPRSLLMKQRKFLY
HFKNVRWAKGRHETLYCYVVKRRDSATSFSLDFGYLRNKSCHVELLFLRYISDWDLDPGRCYRVTWFTS
WSPCYDCARHVADFLRGNPNLSLRIFTARLTGWGALPAGLMSPARPSDYFYCWNTFVENHERTFKAWEGL
HENSVRLSRRLRRILLPLYEVDDLRLDAFRTLGL
- [0140] (单底线:核定位序列;双底线:核输出信号)
- [0141] 小鼠APOBEC-3
MGPFCLGCSHRKCYSPIRNLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVSLHHGVFKNKDNIHAE
ICFLYWFHDKVLKVLSPREEFKITWYMSWSPCFECAEQIVRFLATHHNLSLDIFSSRLYNVQDPETQQNLCLVQEG
- [0142] AQVAAMDLYEFKKCKWKFDVNGGRRFRPWKRLTNFRYQDSKLQEILRPCYIPVPSSSSSTLSNICLTGKLPETRFC
VEGRRMDPLSEEFYSQFYNQVRVKHLCYYHRMKPYLCYQLEQFNGQAPLKGCLLSEKGGKQHAETILFDKIRSMELSQ
VTITCYLTWSPCPNCAWQLAAFKRDRPDLILHIYTSRLYFHWKRPQKGLCSLWQSGILVDVMDLPQFTDCWTNFVN
PKRPFWPWKGLEIISRRTQRRLLRIKESWGLQDLVNDFGNLQLGPPMS (斜体:核酸编辑结构域)
- [0143] 大鼠APOBEC-3:
MGPFCLGCSHRKCYSPIRNLISQETFKFHFKNRLRYAIDRKDTFLCYEVTRKDCDSPVSLHHGVFKNKDNIHAE
ETCFLYWFHDKVLKVLSPREEFKITWYMSWSPCFECAEQVLRFLATHHNLSLDIFSSRLYNIRDPENQQNLCLVQEG
- [0144] GAQVAAMDLYEFKKCKWKFDVNGGRRFRPWKLLTNFRYQDSKLQEILRPCYIPVPSSSSSTLSNICLTGKLPETRF
CVERRRVHLLSEEFYSQFYNQVRVKHLCYYHGVKPYLCYQLEQFNGQAPLKGCLLSEKGGKQHAETILFDKIRSMELS
QVIITCYLTWSPCPNCAWQLAAFKRDRPDLILHIYTSRLYFHWKRPQKGLCSLWQSGILVDVMDLPQFTDCWTNFV
NPKRPFWPWKGLEIISRRTQRRLLHRIKESWGLQDLVNDFGNLQLGPPMS (斜体:核酸编辑结构域)
- [0145] 恒河猕猴 (Rhesus macaque) APOBEC-3G:
- [0146] MVEPMDPRTFVSNFNNRPILSGLNTVWLCCEVKT KDPSGPPLDAKIFQGKVYSKAKYHPMRFLRWFH
KWRQLHHDQEYKVTWYVSWSPCTRCANSVATFLAKDPKVTLTIFVARLYYFWKPDYQQALRILCQKRGGPHATMKI
MNYNEFQDCWNKFVDGRGKPFKPRNNLPKHYTLLQATLGELLRHLMDPGTFTSNFNKPWVSGQHETLYCYKVERL
HNDTWVPLNQHRGFLRNQAPNIHGFPKGRHAELCFLDLIPFWKLDGQQYRVTCFTSWSPCFSCAQEMAKFISNNEH
VSLCIFAARIYDDQGRYQEGRLALHRDGAKIAMMNYSEFEYCWDTFVDRQGRPFQPWDGLDEHSQALSGRLRAI (斜
体:核酸编辑结构域;单底线:细胞质定位信号)
- [0147] 黑猩猩APOBEC-3G:
MKPHFRNPVERMYQDTFSDNFYNRPILSHRNTVWLCYEVKTGKPSRPPLDAKIFRGQVYSKLKYHPEMRFFHW
FSKWRKLHRDQEYEVTWYISWSPCTKCTRDVATFLAEDPKVTLTIFVARLYYFWDPDYQEALRSLCQKRDGPATMK
IMNYDEFQHCWSKFVYSQRELFEPPWNNLPKYYILLHIMLGEILRHSMPPPTFTSNFNELWVRGRHETLYCYEVERL
- [0148] HNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDVIPFWKLDLHQDYRVTCFTSWSPCFSCAQEMAKFISNNKH
VSLCIFAARIYDDQGRQCQEGRLTLAKAGAKISIMTYSEFKHCWDTFVDHQGCPFPWDGLEEHSQALSGRLRAILQN
QGN
- [0149] (斜体:核酸编辑结构域;单底线:细胞质定位信号)
- [0150] 绿猴APOBEC-3G:

- [0151] MNPQIRNMVEQMEPDIFVYYFNNRPILSGRNTVWLCYEVKTKDPSGPPLDANIFQGKLYPEAKDHP
FRKWRQLHRDQEYEVTWYVSWSPCTRCANSVATFLAEDPKVTLTIFVARLYYFWKPDYQQALRILCQERGGPHATMK
IMNYNEFQHCWNEFDGQGKPFKPRKNLPKHYTELLHATLGELLRHVMDPGTFTS
NFNNKPWVSGQRETYLCYKVERSHNDTWVLLNQHRGFLRNQAPDRHGFPKGRHAELCFLDLIPFWKLDDQQYR
VTCFTSWSPCFSCAQKMAKFISNNKHVSLCIFAARIYDDQGRCQEGLRTLHRDGAIAVMNYSEFEYCWDTFVDRQG
RPFQPWDGLDEHSQALSGRLRAI
- [0152] (斜体:核酸编辑结构域;单底线:细胞质定位信号)
- [0153] 人类APOBEC-3G:
MKPHFRNTVERMYRDTFSYNFYNNRPILSRRNTVWLCYEVKTKGPSRPPLDAKIFRGQVYSELKYHPEMRFFHW
FSKWRKLHRDQEYEVWYISWSPCTKCTRDMATFLAEDPKVTLTIFVARLYYFWDPDYQEALRSLCQKRDGPRAITMK
IMNYDEFQHCWSKFVYSQRELFEWNNLPKYYILLHIMLGEILRHSMPTFTFNNNEPWVRGRHETYLCEVERM
HNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDVIPFWKLDDQDYRVTCFTSWSPCFSCAQEMAKFISNNKH
VSLCIFTARIYDDQGRCQEGLRTLAEGAKISIMTYSEFKHCWDTFVDHQGCPFPWDGLDEHSQDLSGRLRAILQN
QEN
- [0156] (斜体:核酸编辑结构域;单底线:细胞质定位信号)
- [0157] 人类APOBEC-3F:
MKPHFRNTVERMYRDTFSYNFYNNRPILSRRNTVWLCYEVKTKGPSRPRLDAKIFRGQVYSQPEHHAEMCFLSW
FCGNQLPAYKCFQITWVSWTPCPCVAKLAFLAEHPNVTLTISAARLYYWERDYRRALCRLSQAGARVKIMDE
EFAYCWENFVYSEGQPFMPWYKFDDNYAFLHRTLKEILRNPMEMYPHIFYFHFKNLRKAYGRNESWLCFTMEVVKH
HSPVSWKRGVFRNQVDPETHCHAERCFLSWFCDDILSPNTNYEVTWYTSWSPCPECAGEVAEFLARHSNNLTIFTA
RLYYFWDTDYQGLRSLSQEGASVEIMGYKDFKYCWENFVYNDDEPFKPKWGLKYNFLFLDSKLQEILE
- [0159] (斜体:核酸编辑结构域)
- [0160] 人类APOBEC-3B:
MNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEVKIKRGRSNLLWDTGVFRGQVYFKPKQYHAEMCFLS
WFCGNQLPAYKCFQITWVSWTPCPCVAKLAFLSEHPNVTLTISAARLYYWERDYRRALCRLSQAGARVTIMDY
EEFAYCWENFVYNEGQFMPWYKFDDNYAFLHRTLKEILRYLMDPDTFTFNNNDPLVLRRTYLCYEVERLDNGT
WVLMQHMGLCNEAKNLLCGFYGRHAELRFLDLVPSLQLDPAQIYRVTWFTSWSPCFSWGCAGEVRAFLQENTHVR
LRIFAARTYDYDPLYKEALQMLRDAGAQVSIMTYDEFECWDTFVYRQGCPFPWDGLEEHSQALSGRLRAILQNQG
N
- [0162] (斜体:核酸编辑结构域)
- [0163] 大鼠APOBEC-3B:
- [0164] MQPQGLGPNAGMPVCLGCSHRRPYSPIRNPLKKLYQQTFYFHFKNVRYAWGRKNNFLCYEVNGMDCAL
PVPLRQGCVFRKQGGHIAELCFIYWFHDKVLRVLSPMEEFKVTWYMSWSPCSKCAEQVARFLAAHRNLSLAIFSSRLY
YYLRNPNYQQKLCRLIQEGVHVAAMDLPFCKKCNKFVDNDGQPFPRWMLRINFYDCKLQEIFSRMNLREDVF
YLQFNNSHRVKPVQNRYYRRKSYLCYQLERANGQEPLKGYLLYKKGEQH
- [0165] VEILFLEKMRSMELSQVRITCYLTWSPCPNCARQLAAFKKDHPDLILRIYTSRLYFWRKKFQKGLCTLW
RSGIHVDVMDLPQFADCWTNFVNPQRPFPRPWNELEKNSWRIQRRRLRIKESWGL
- [0166] 牛APOBEC-3B:
- [0167] DGWEVAFRSGTVLKAGVLGVSMTEGWAGSGHPGQGACVWTPGTRNTMNLLEVLFKQQFGNQPRVPAPY

YRRKTYLCYQLKQRNDLTLDRGCFRNKKQRHAERFIDKINSLDLNPSQSYKICITYITWSPCPNCANELVNFITRNNH
LKLEIFASRLYFHWIKSFKMGLQDLQAGISVAVMTHTEFEDCWEQFVDNQSRPFQPWDKLEQYSASIRRLQRILT
API

[0168] 黑猩猩APOBEC-3B:

[0169] MNPQIRNPMEMYQRTFYFNFENEPILYGRSYTWLCYEVKIRRGHSNLLWDTGVFRGQMYSQPEHHAEM
CFLSWFCGNQLSAYKCFQITWVSWTPCPDCVAKLAKFLAEHPNVTLTISAARLYYYWERDYRRALCRLSQAGARVK
IMDDEEFAYCWENFVYNEGQPFMPWYKFDDNYAFLHRTLKEIIRHLMDDPTFTFNFNNDPLVLRHQTYLCYEVERL
DNGTWVLMQDQHMGLCNEAKNLLCGFYGRHAELRFLDLVPSLQLDPAQIYRVTFWISWSPCFSWGCAGQVRAFLQEN
THVRLRIFAARIYDYDPLYKEALQMLRDAGAQVSIMTYDEFYCWDTFVYRQGCQFQPWDGLEEHSQALSGLRLAIL
QVRASSLCMVPHRPPPPPPQSPGPCLPLCSEPPLGSLPTGRPAPSLPFLLTASFSPPPASLPPLPSLSLSPGHLPV
PSFHSLTSCSIQPPCSSRIRETEGWASVSKEGRDLG

[0170] 人类APOBEC-3C:

MNPQIRNPMKAMYPGTFYFQFKNLWEANDRNETWLCFTVEGIRRSVSVSWKTGVFRNQVDSETH*CHAERCFLS*
[0171] *WFCDDILSPNTKYQVTWYTSWSPCPDCAGEVAEFLARHSNVNLTIFTARLYYFQYPCYQEGRLSLSQEGVAVEIMDY*
EDFKYCWENFVYNDNEPFKPKWGLKTNFRLLKRRLRESLQ

[0172] (斜体:核酸编辑结构域)

[0173] 大猩猩APOBEC3C

MNPQIRNPMKAMYPGTFYFQFKNLWEANDRNETWLCFTVEGIRRSVSVSWKTGVFRNQVDSETH*CHAERCFLS*
[0174] *WECDDILSPNTNYQVTWYTSWSPCECAGEVAEFLARHSNVNLTIFTARLYYFQDTDYQEGRLSLSQEGVAVKIMDY*
KDFKYCWENFVYNDDEPFKPKWGLKYNFRFLKRRLQEILE

[0175] (斜体:核酸编辑结构域)

[0176] 人类APOBEC-3A:

MEASPASGRHLMDPHIFTSNFNNGIGRHKTYLCYEVERLDNGTSVKMDQHRGFLHNQAKNLLCGFYGRHAEL
[0177] *RFLDLVPSLQLDPAQIYRVTFWISWSPCFSWGCAGEVRAFLQENTHVRLRIFAARIYDYDPLYKEALQMLRDAGAQV*
SIMTYDEFKHCWDTFVDHQGCQFQPWDGLDEHSQALSGLRLAILQNQGN

[0178] (斜体:核酸编辑结构域)

[0179] 普通猕猴APOBEC-3A:

MDGSPASRPHLMDPNTFTFNFNNDLSVRGRHQTYLCYEVERLDNGTWVPMDERRGFLCNKAKNVPCGDYGC/H
[0180] *VELRFLCEVPSWQLDPAQTYRVTFWISWSPCFRRGCAGQVRVFLQENKHVRLRIFAARIYDYDPLYQEALRTLRLDAG*
AQVSIPTYEEFKHCWDTFVDRQGRPFQPWDGLDEHSQALSGLRLAILQNQGN

[0181] (斜体:核酸编辑结构域)

[0182] 牛APOBEC-3A:

MDEYTFTENFNQGWPSKTYLCYEMERLDGDATIPLDEYKGFVRNKGLDQPEKPCHAELYFLGKIHSWNLDRN
[0183] *QHYRLTCFISWSPCYDCAQKLTTFLKENHHISLHILASRIYTHNRFCHQSGLCELQAAGARITIMTFEDFKHCWET*
FVDHKGKPFQWEGLVKSQALCTELQAILKTQQN

[0184] (斜体:核酸编辑结构域)

[0185] 人类APOBEC-3H:

- MALLTAETFRLQFNNKRRRLRRPYPRKALLCYQLTPQNGSTPTRGYFENKKKCHAEICFINEIKSMGLDETQC
- [0186] *YQVTCYLTWSPCSSCAWELVDFIKAHDHLNLGIFASRLYYHWCKPQQKGLRLLCGSQVPVEVMGFPKFADCWENFVD*
HEKPLSFNPYKMLEELDKNSRAIKRRLERIKIPGVRAQGRYMDILCDAEV
- [0187] (斜体:核酸编辑结构域)
- [0188] 恒河猕猴APOBEC-3H:
- [0189] MALLTAKTFSLQFNNKRRVKNPYPRKALLCYQLTPQNGSTPTRGHLKNKKKDHAEIRFINKIKSMGLD
ETQCYQVTCYLTWSPCSCAGELVDFIKAHRHLNLRIFASRLYYHWRPNYQEGLLLLCGSQVPVEVMGLPEFTDCWE
NFVDHKEPPSFNPSEKLEELDKNSQAIKRRLERIKRSRVDVLENGLRSLQLGPVTPSSSIRNSR
- [0190] 人类APOBEC-3D:
- MNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEVKIKRGRSNLLWDTGVFRGPVLPKRQSNHRQEVYF*
RFENHAEMCFLSWFCGNRLPANRRFQITWVFSWNPCLPCVVKVTKFLAEHPNVTLTISAARLYYRDRDWRWVLLRL
[0191] *HKAGARVKIMDYEDFAYCWENFVCNEGQPFMPWYKFDNYASLHRTLKEILRNPMEAMYPHIFYHFKNLLKACGRN*
ESWLCTMEVTKHSAVFRKRGVFRNQDPETHCHAERCFLSWFCDDILSPNTNYEVTWYTSWSPCECAGEVAEFL
ARHSNVNLTIFTARLCYFWDTDYQEGLCSLSQEGASVKIMGYKDFVSCWKNFVYSDDEPFKPKWGLQTNFRLLKRRLL
REILQ
- [0192] (斜体:核酸编辑结构域)
- [0193] 人类APOBEC-1:
- [0194] MTSEKGPSTGDPTLRRRIEPWEFDVFDYDPRELKREACLLYEIKWGMRSRKIWRSSGKNTTNHVEVNFIEK
FTSERDFHPSMSCSITWFLSWSPCWECSQAIREFLSRHPGVTLVIIYVARLFWHMDQQNRQGLRDLVNSGVTIQIMRA
SEYYHCWRNFVNYPPGDEAHPQYPPLWMMLYALELHCIILSLPPCLKISRRWQNHLTFFRLHLQNCHYQTIIPPHIL
LATGLIHPSVAWR
- [0195] 小鼠APOBEC-1:
- [0196] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELKRETCCLYEINWGGRHSSVRHTSQNTSNHVEVNFLEK
FTTERYFRPNTRCSITWFLSWSPCGECSRAITEFLSRHPVTLFIYIARLYHHTDQRNRQGLRDLISSGVTIQIMTE
QEYCYCWRNFVNYPPSNEAYWPRYPHLWVKLYVLELYCIIILGLPPCLKILRRKQPQLTFFTITLQTCHYQRIPPHLL
WATGLK
- [0197] 大鼠APOBEC-1:
- [0198] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELKRETCCLYEINWGGRHSSVRHTSQNTNKHVEVNFIEK
FTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHTLFIYIARLYHHADPRNRQGLRDLIS
- [0199] SGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIIILGLPPCLNILRRKQPQLTF
FTIALQSCHYQRLPPHILWATGLK
- [0200] 人类APOBEC-2:
- [0201] MAQKEEAAVATEAASQNGEDLENLDDPEKLKELIELPPFEIVTGERLPANFFKFQFRNVEYSSGRNKTF
LCYVVEAQGGQVQASRGYLEDEHAAAHAEEAFFNTILPAFDPALRYNVTWYVSSSPCAACADRIKTLSTKTNLR
LLILVGRLFMWEEPEIQAALKKLKEAGCKLRIMKPQDFEYVWQNFVEQEEGESKAFQPWEDIQENFLYEEKLADIL
K
- [0202] 小鼠APOBEC-2:
- [0203] MAQKEEAAEAAAPASQNGDDLENLEDPEKLKELIDLPPFEIVTGVRLPVNFFKFQFRNVEYSSGRNKTF
LCYVVEVQSKGGQAQATQGYLEDEHAGAHAAEEAFFNTILPAFDPALKYNVTWYVSSSPCAACADRILKTLSTKTNLR

LLILVSRLFMWEEPEVQAALKKLKEAGCKLRIMKPQDFEYIWQNFVEQEEGESKAFEPWEDIQENFLYYEEKLADIL
K

[0204] 大鼠APOBEC-2:

[0205] MAQKEEAAEAAAPASQNGDDLENLEDPEKLKELIDLPPFEIVTGVRLPVNFFKFQFRNVEYSSGRNKTF
LCYVVEAQSKGGQVQATQGYLEDEHAGAHAEAEFFNTILPAFDPALKYNVTWYVSSSPCAACADRILKTLSTKNLR
LLILVSRLFMWEEPEVQAALKKLKEAGCKLRIMKPQDFEYLWQNFVEQEEGESKAFEPWEDIQENFLYYEEKLADIL
K

[0206] 牛APOBEC-2:

[0207] MAQKEEAAAAAEPASQNGEEVENLEDPEKLKELIELPPFEIVTGERLPAHYFKFQFRNVEYSSGRNKTF
LCYVVEAQSKGGQVQASRGYLEDEHATNHAEAEFFNSIMPTFDPALRYMVTWYVSSSPCAACADRIVKTLNKTKNLR
LLILVGRLFMWEEPEIQAALRKLKEAGCRLRIMKPQDFEYIWQNFVEQEEGESKAFEPWEDIQENFLYYEEKLADIL
K

[0208] 海七鳃鳗CDA1 (pmCDA1)

[0209] MTD AEYVRIHEKLDIYTFKKQFFNNKKS VSHRCYVLFELKRRGERRACFWGYAVNKPQSGTERGIHAEI
FSIRKVEEYLRDNPQGFTINWYSSWSPCADCAEKILEWYNQELRGNGHTLKIWACKLYYEKNARNQIGLWNLRDNGV
GLNVMVSEHYQCCRKIFIQSSHNQLNENRWLEKTLKRAEKRRSELSFMIQVKILHTTKSPAV

[0210] 人类APOBEC3G D316R D317R

[0211] MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPPLDAKIFRGQVYSELKYHPEMR
FFHWFSSKWRKLHRDQEYEV TWYISWSPCTKCTRD MATFLAEDPKVTLTIFVARLYYFWDPDYQEALRSLCQKRDGPR
ATMKFNYDEFQHCWSKFVYSQREL FEPWNNLPKYYILLHFMLGEILRHSM DPPTFTFNFNNEPWVRGRHET YLCYEV
ERMHNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDVIPFWKLDLDQDYRVTCFTSWSPCFSCAQEMAKFISK
KHVSLCIFTARIYRRQGRCQEGLRTLAEAGAKISFTYSEFKHCWDTFVDHQGCPFPQPDG LDEHSQDLSGRLRAILQ
NQEN

[0212] 人类APOBEC3G链A

[0213] MDPPTFTFNFNNEPWVRGRHET YLCYEVERMHNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDV
IPFWKLDLDQDYRVTCFTSWSPCFSCAQEMAKFISKNHVSLCIFTARIYDDQGRCQEGLRTLAEAGAKISFTYSE
FKHCWDTFVDHQGCPFPQPDGLD EHSQDLSGRLRAILQ

[0214] 人类APOBEC3G链A D120R D121R

[0215] MDPPTFTFNFNNEPWVRGRHET YLCYEVERMHNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDV
IPFWKLDLDQDYRVTCFTSWSPCFSCAQEMAKFISKNHVSLCIFTARIYRRQGRCQEGLRTLAEAGAKISFMTYSE
FKHCWDTFVDHQGCPFPQPDGLD EHSQDLSGRLRAILQ

[0216] 术语“脱氨酶”或“脱氨酶结构域”是指催化脱氨反应的蛋白或其片段。

[0217] “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。

[0218] “可检测标记”是指当与目标分子链接时,经由分光的、光化学的,生物化学的,免疫化学的,或化学的手段使得目标分子可被检测到的组合物。例如,有用的标记包括放射性同位素、磁珠、金属珠、胶体颗粒、荧光染料、电子致密试剂、酶(例如,像ELISA中常用的)、生物素、地高辛或半抗原。

[0219] “疾病”是指损害或干扰细胞、组织或器官的正常功能的任何病况或病症。在特定

的实施方式中,可接受用本发明的组合物治疗的疾病与点突变,剪接事件,提前的终止密码子,或错误折叠事件相关联。

[0220] “DNA结合蛋白结构域”是指结合DNA的多肽或其片段。在一些实施方式中,DNA结合蛋白结构域是具有序列特异性DNA结合活性的锌指或TALE结构域。在其他实施方式中,DNA结合蛋白结构域是结合DNA的CRISPR-Cas蛋白(例如,Cas9)的结构域,包含例如结合原间隔序列临近基序(PAM)。在一些实施方式中,DNA结合蛋白结构域与多核苷酸(例如,单向导RNA)形成复合体,并且所述复合体结合由gRNA和原间隔序列临近基序所指定的DNA序列。在一些实施方式中,DNA结合蛋白结构域包括切口酶活性(例如,nCas9)或是催化失活的(例如,dCas9,锌指结构域,TALE)。在其他实施方式中,DNA结合蛋白结构域是归巢核酸内切酶I-SceI的催化失活的变体,或是TALE蛋白AvrBs4的DNA-结合结构域。参见例如,Gabsalilow等,Nucleic Acids Research,Volume 41,Issue 7,1April 2013,e83页。在一些实施方式中,DNA结合蛋白结构域融合到具有催化活性的结构域上(例如,FokI,MutH)。在特定的实施方式中,锌指结构域融合到核酸内切酶FokI的催化结构域上。在其他实施方式中,TALE融合到MutH上,所述MutH包括位点-特异性DNA切口活性。

[0221] 如本文所用,术语“有效量”是指足以引起所期望的生物应答的生物活性剂的量。在特定的实施方式中,有效量是两个或更多个质粒的量,所述质粒包括碱基编辑器系统的各部分,而所述量则足以在用所述质粒转染的细胞中表达有活性的碱基编辑系统。如本领域技术人员将理解的,剂(例如融合蛋白)的有效量可能取决于各种因素而变化,例如所期望的生物应答,例如待编辑的特定等位基因、基因组或靶位点,正被靶向的细胞或组织,以及正被使用的剂。

[0222] “片段”是指多肽或核酸分子的一部分。这部分含有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%。片段可以含有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。

[0223] “杂交”是指互补核碱基之间的氢键键结,其可以是沃森-克里克(Watson-Crick)、胡斯坦(Hoogsteen)或反向胡斯坦氢键键结。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。

[0224] 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,所述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、T7Endo1、T4PDG、UDG、hSMUG1和hAAG的抑制剂。在一些实施方式中,所述IBR是Endo V或hAAG的抑制剂。在一些实施方式中,所述IBR是催化失活的EndoV或催化失活的hAAG。

[0225] “内含肽”是蛋白质的片段,其能够切除自身并在名为蛋白质剪接的过程中将剩余的片段(外显肽)用肽键连接在一起。内含肽也被称为“蛋白质内含子”。内含肽切除自身并连接蛋白质剩余的的部分的过程在本文中称为“蛋白质剪接”或“内含肽-介导的蛋白质剪接”。在一些实施方式中,前体蛋白(先前于内含肽-介导的蛋白剪接的含有内含肽的蛋白)的内含肽来自两个基因。此种内含肽在本文中称为断裂内含肽(例如,断裂内含肽-N和断裂内含肽-C)。例如在蓝细菌中,DnaE,即DNA聚合酶III的催化亚基,是由两个分开的基因dnaE-n和dnaE-c编码。由dnaE-n基因编码的内含肽在本文中可称为“内含肽-N”。由dnaE-c基因编码的内含肽在本文中可称为“内含肽-C”。

[0226] 亦可使用其他内含肽系统。例如,基于dnaE内含肽的合成内含肽,即Cfa-N(例如,断裂内含肽-N)和Cfa-C(例如,断裂内含肽-C)内含肽对,已被描述(例如,在Stevens等,J Am Chem Soc.2016Feb.24;138(7):2162-5中,通过引用并入本文)。可以根据本发明披露使用的内含肽对的非限制性实例包含:Cfa DnaE内含肽、Ssp GyrB内含肽、Ssp DnaX内含肽、Ter DnaE3内含肽、Ter ThyX内含肽、Rma DnaB内含肽和Cne Prp8内含肽(例如,如美国专利号8,394,604中所述,通过引用并入本文)。

[0227] 提供了内含肽的示例性的核苷酸和氨基酸序列。

[0228] DnaE内含肽-N DNA:

[0229] TGCCTGTCATACGAAACCGAGATACTGACAGTAGAATATGGCCTTCTGCCAATCGGGAAGATTGTGGA
GAAACGGATAGAATGCACAGTTTACTCTGTCTGATAACAATGGTAACATTTATACTCAGCCAGTTGCCCAGTGGCAC
GACCGGGGAGAGCAGGAAGTATTCGAATACTGTCTGGAGGATGGAAGTCTCATTAGGGCCACTAAGGACCACAAAT
TTATGACAGTCGATGGCCAGATGCTGCCTATAGACGAAATCTTTGAGCGAGAGTTGGACCTCATGCGAGTTGACAA
CCTT CCTAAT

[0230] DnaE内含肽-N蛋白:

[0231] CLSYETEILTVEYGLLPIGKIVEKRIECTVYSVDNNGNIYTPVAQWHDRGEQEVFEYCLEDGSLIRA
TKDHFMTVDGQMLPIDEIFERELDLMRVDNL PN

[0232] DnaE内含肽-C DNA:

[0233] ATGATCAAGATAGCTACAAGGAAGTATCTTGGCAAACAAAACGTTTATGATATTGGAGTCGAAAGAGA
TCACAACCTTGCTCTGAAGAACGGATTCATAG CTCTAAT

[0234] 内含肽-C:MIKIATRKYLKGQNVYDIGVERDHNFKNGFIASN

[0235] Cfa-N DNA:

[0236] TGCCTGTCTTATGATACCGAGATACTTACCGTTGAATATGGCTTCTTGCCTATTGGAAAGATTGTCTGA
AGAGAGAATTGAATGCACAGTATATACTGTAGACAAGAATGGTTTCGTTTACACACAGCCCATTGCTCAATGGCAC
AATCGCGGCGAACAAGAAGTATTTGAGTACTGTCTCGAGGATGGAAGCATCATACGAGCAACTAAAGATCATAAAT
TCATGACCACTGACGGGCAGATGTTGCCAATAGATGAGATATTCGAGCGGGGCTTGGATCTCAAACAAGTGGATGG
ATTG CCA

[0237] Cfa-N蛋白:

[0238] CLSYDTEILTVEYGFLPIGKIVEERIECTVYTVDKNGFVYTQPIAQWHNRGEQEVFEYCLEDGSIIRA
TKDHFMTTDTGQMLPIDEIFERGLDLKQVDGL P

[0239] Cfa-C DNA:

[0240] ATGAAGAGGACTGCCGATGGATCAGAGTTTGAATCTCCCAAGAAGAAGAGGAAAAGTAAAGATAATATC
TCGAAAAAGTCTTGGTACCCAAAATGTCTATGATATTGGAGTGGAGAAAGATCACAACTTCCTTCTCAAGAACGGT
CTCGTA GCCAGCAAC

[0241] Cfa-C蛋白:

[0242] MKRTADGSEFESPKKKRKVKIISRKSLGTQNVYDIGVEKDNHFLKNGLV ASN

[0243] 内含肽-N和内含肽-C可以分别融合到断裂Cas9的N末端部分和断裂Cas9的C末端部分,以连接断裂Cas9N末端部分和断裂Cas9的C末端部分。例如,在一些实施方式中,内含肽-N融合到断裂Cas9的N末端部分的C-端,即,形成N-[断裂Cas9的N末端部分]-[内含肽-N]-C的结构。在一些实施方式中,内含肽-C融合到断裂Cas9的C末端部分的N-端,即,形成

N-[内含肽-C]--[断裂Cas9的C末端部分]-C的结构。用于将内含肽所融合到的蛋白(例如,断裂Cas9)连接起来的内含肽-介导的蛋白质剪接的机制是本领域已知的,例如,如在Shah等,Chem Sci.2014;5(1):446-461中所描述的,在此通过引用并入本文。用于设计和使用内含肽的方法在本领域中是已知的并且被描述,例如国际专利公开案第W02014004336号、国际专利公开案第W02017132580号、美国专利公开案第US20150344549号、和美国专利公开案第US20180127780号,其每个的全部内容通过引用并入本文。

[0244] 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、病毒材料或培养基;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。对于可以被修饰的蛋白质,例如磷酸化或糖基化,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。

[0245] “分离的核酸”是指核酸(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所述核酸两侧翼的基因。因此,所述术语包含例如,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。另外,所述术语包含从DNA分子转录而来的RNA分子,以及编码附加的多肽序列的作为杂合基因一部分的重组DNA。

[0246] “分离的多肽”是指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,所述多肽即视为分离的。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。可以获得分离的本发明的多肽,例如,通过从天然来源提取,通过表达编码这种多肽的重组核酸;或通过蛋白质的化学合成。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。

[0247] 如本文所用,术语“连接子”是指键(例如,共价键),化学基团或链接两个分子或部分体(例如一融合蛋白的两个结构域)的分子。在一些实施方式中,连接子连接了RNA-可编程的核酸酶的gRNA结合结构域(其包含Cas9核酸酶结构域),和核酸编辑蛋白的催化结构域。在一些实施方式中,连接子连接了dCas9和核酸编辑蛋白。通常,连接子是位于两个基团、分子或其他部分体之间或两侧,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,所述连接子是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,所述连接子是有机分子、基团、聚合物或化学部分体。在一些实施方式中,所述连接子为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、60、65、70、70、75、80、85、90、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。更长或更短的连接子也在预期中。在一些实施方式中,连接子包括氨基酸序列SGSETPGTSESATPES,其也可称为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基序或这些中的任一组合,其中n独立地是1至30之间的整数,并且其中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。

[0248] 在一些实施方式中,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融合:SGGSSGSETPGTSESATPESSGGS、SGGSSGGSSGSETPGTSESATPESSGGSSGGS或GGSGGSPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTE PSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGGSSGGS。在一些实施方式中,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融合:SGSETPGTSESATPES,其也可称为XTEN连接子。在一些实施方式中,所述连接子的长度是24个氨基酸。在一些实施方式中,所述连接子包括氨基酸序列SGGSSGGSSGSETPGTSESATPES。在一些实施方式中,所述连接子的长度是40个氨基酸。在一些实施方式中,所述连接子包括氨基酸序列SGGSSGGSSGSETPGTSESATPESSGGSSGGSSGGSSGGS。在一些实施方式中,所述连接子的长度是64个氨基酸。在一些实施方式中,所述连接子包括氨基酸序列SGGSSGGSSGSETPGTSESATPESSGGSSGGSSGGSSGSETPGTSESATPESSGGS SGGS。在一些实施方式中,所述连接子的长度是92个氨基酸。在一些实施方式中,所述连接子包括氨基酸序列PGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGGSSGGS。

[0249] “标记物”是指具有与疾病或病症相关的表达水平或活性的改变的任何蛋白质或多核苷酸。

[0250] 如本文所用,术语“突变”是指序列(例如核酸或氨基酸序列)中的一残基被另一残基取代,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,随后指出序列中所述残基的位置,然后指出新取代的残基的身份。制造出本文所提供的氨基酸取代(突变)的各种方法在本领域中是众所周知的,例如由Green和Sambrook,《分子克隆:实验室手册》(第4版,冷泉港实验室出版社,冷泉港,纽约(2012))提供。

[0251] 如本文所用,术语“核酸”和“核酸分子”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的情况下。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,不具有磷酸二酯主干的类似物。核酸可以从天然来源中纯化,使用重组表达系统生产,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,诸如具有化学修饰过的碱基或糖,以及主干修饰。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、0(6)-甲基鸟嘌呤和2-硫胞苷)；化学修饰过的碱基；生物修饰过的碱基(例如,甲基化的碱基)；嵌和的碱基；修饰过的糖(例如,2'-氟核糖、核糖、2'-脱氧核糖、阿拉伯糖和己糖)；和/或修饰过的磷酸基团(例如,硫代磷酸和5'-N-亚磷酸酐键合)。

[0252] 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等的国际PCT申请,PCT/EP2000/011690,其提交于2000年11月23日,以W0/2001/038547公布于2001年5月31日,针对其示例性核定位序列的公开披露,其内容在此通过引用并入本文。在其他实施方式中,NLS是优化的NLS,其被描述于例如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、KRGINDRNFWRGENGKTR、RKSGKIAAIVVKRPRK、PKKKRKV或MDSLLMNRKFLYQFKNVRWAKGRRETYLC。

[0253] 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)相联的蛋白质,诸如向导核酸,其将napDNAbp导向至特定核酸序列。例如,Cas9蛋白可以与向导RNA相联,所述向导RNA将Cas9蛋白导向至与向导RNA互补的特定DNA序列。在一些实施方式中,所述napDNAbp是Cas9结构域,例如,核酸酶活性的Cas9、Cas9切口酶(nCas9)或核酸酶失活的Cas9(dCas9)。核酸可编程的DNA结合蛋白的例子包括但不限于,Cas9(例如,dCas9和nCas9)、CasX、CasY、Cpf1、C2c1和C2c3。其他核酸可编程的DNA结合蛋白也在本发明披露的范围内,尽管它们可能未在本发明披露中具体列出。

[0254] 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所述剂。

[0255] 如本文所用,“患者”或“受试者”是指一受试者,例如,被诊断或被怀疑具有或正在发展某一疾病或病症的哺乳动物受试者。在一些实施方式中,术语“患者”是指具有高于平均水平的发展出一疾病或病症的可能性的哺乳动物受试者。示例性的患者可以是人类,非人类的灵长类动物,猫,狗,猪,牛,马,山羊,绵羊,啮齿类动物(例如小鼠,兔子,大鼠或豚鼠)和其他可以从本文所揭示的治疗中受益的哺乳动物。示例性的人类患者可以是男性和/或女性。“需要其的患者”或“需要其的受试者”在本文中是指被诊断或被怀疑具有某一疾病或病症的患者。

[0256] 术语“RNA-可编程的核酸酶”和“RNA-导向的核酸酶”与不是切开靶的一种或多种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,可以称其为核酸酶:RNA复合体。通常,结合的RNA被称为向导RNA(gRNA)。gRNA可以以具有两个或多个RNA的复合体形式存在,也可以以单个RNA分子形式存在。以单个RNA分子形式存在的gRNA可以被称为单导向RNA(sgRNA),尽管“gRNA”可互换使用以指代以单个分子形式存在或以具有两个或多个分子的复合体形式存在的向导RNA。通常,以单个RNA种类存在的gRNA包含两个结构域:(1)与靶核酸共享同源性的结构域(例如,其并导引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)对应于称

为tracrRNA的序列,并包含茎-环结构。例如,在一些实施方式中,结构域(2)与Jinek等, Science 337:816-821(2012)中所提供的tracrRNA相同或同源,其全部内容在此通过引用并入本文。gRNA的其他例子(例如,包括结构域2的那些)可以在2013年9月6日提交的名称为“Switchable Cas9Nucleases And Uses Thereof”的美国临时专利申请第U.S.S.N.61/874,682号中,以及在2013年9月6日提交的名称为“Delivery System For Functional Nucleases”美国临时专利申请第U.S.S.N.61/874,746号中找到,其每个的全部内容在此通过引用并入本文。在一些实施方式中,gRNA包括两个或多个结构域域(1)和(2),并可以称为“延伸的gRNA”。例如,如本文所述,延伸的gRNA将,例如,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。所述gRNA包括与靶位点互补的核苷酸序列,其介导核酸酶/RNA复合体结合到所述靶位点,并提供了核酸酶:RNA复合体的序列特异性。在一些实施方式中,所述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.,Nature 471:602-607(2011))。

[0257] 如本文在蛋白质或核酸的上下文中使用的,术语“重组”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,所述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。

[0258] “减少”是指至少10%、25%、50%、75%或100%的负向改变。

[0259] “参考”是指一标准或对照条件。在一个实施方式中,参考是以单一质粒所表达的全长核碱基编辑器的活性,其与以若干片段(所述片段包括用于内含肽依赖重组的内含肽)所表达的核碱基编辑器是在相同细胞中且在相同条件下表达。

[0260] “参考序列”是用作序列比对的基准的界定的序列。参考序列可以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,优选至少约20个氨基酸,更优选至少约25个氨基酸,甚至再更优选约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,以及约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。

[0261] 术语“单核苷酸多态性(SNP)”是在基因组中特定位置发生的单一核苷酸中的变异,其中每个变异在群体中以某种可观的程度存在(例如,>1%)。例如,在人类基因组的特定碱基位置,C核苷酸可能出现在大多数个体中,但在少数个体中,所述位置被A占据。这意味着在所述特定位置存在一个SNP,并且这两个可能的核苷酸变异,C或A,被认为是此位置的等位基因。SNPs构成疾病易感性差异的基础。疾病的严重程度以及人体对治疗的反应方式也是遗传变异的临床表征。SNPs可以落入基因的编码区域、基因的非编码区域或位于基

因间区域(基因之间的区域)。在一些实施方式中,由于遗传密码的简并性,编码序列内的SNPs不一定改变所生产蛋白质的氨基酸序列。编码区内的SNPs有两种类型:同义和非同义SNPs。同义SNPs不会影响所述蛋白质序列,而非同义SNPs改变蛋白质的氨基酸序列。非同义SNPs有两种类型:错义和无义。不在蛋白质-编码区的SNPs仍可能影响基因剪接、转录因子结合、信使RNA降解或非编码RNA的序列。受这种类型SNP影响的基因表达称为eSNP(表达SNP),并可以在基因的上游或下游。单核苷酸变体(SNV)是单一核苷酸中的变异,其不受频率的限制并且可以在体细胞中发生。体细胞单核苷酸变异亦可称为单-核苷酸改变。

[0262] “特异性结合”是指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本(例如生物学样本)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。

[0263] 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所述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所述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”是指在各种严格条件下,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)《酶学方法》152:507)。

[0264] 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,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°C,更优选至少约37°C,最优选至少约42°C的温度。变化其他参数,例如杂交时间、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各种严格程度。在一个实施方式中,杂交将发生在30°C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°C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°C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。

[0265] 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°C的温度,更优选至少约42°C,甚至更优选至少约68°C。在一优选的实施方式中,清洗步骤将发生在25°C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°C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°C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,在这些条件上的其他变化是显而易见的。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)。

[0266] “断裂”是指分为两个或更多个片段。

[0267] “断裂Cas9蛋白”或“断裂Cas9”是指一种Cas9蛋白,其被提供为由两个单独的核苷酸序列编码的N末端片段和C末端片段。对应于Cas9蛋白的N末端部分和C末端部分可以被剪接以形成一“重构的”Cas9蛋白。在特定的实施方式中,Cas9蛋白在所述蛋白的无序区域内被分成两个片段,例如,如Nishimasu等,Cell,Volume 156,Issue 5,pp.935-949,2014中所述,或如Jiang等.(2016) Science351:867-871.PDB file:5F9R中所述,其每个的全部内容通过引用并入本文。可以通过本领域已知的一种或多种蛋白质结构确定技法来确定无序区域,所述技术包含但不限于X-射线结晶学,NMR光谱,电子显微镜(例如cryoEM),和/或计算机模拟蛋白质建模。在一些实施方式中,在大约氨基酸A292-G365,F445-K483或E565-T637之间的SpCas9的一区域内,或在任何其他Cas9,Cas9变体(例如nCas9,dCas9)或其他napDNAbp的对应位置上,所述蛋白在任一C、T、A或S处被分为两个片段。在一些实施方式中,蛋白质在SpCas9的T310、T313、A456、S469或C574处被分为两个片段。在一些实施方式中,将蛋白质分为两个片段的过程称为蛋白质的“断裂过程”。

[0268] “基本上相同”是指与参考氨基酸序列(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或核酸水平上是至少60%、80%或85%、90%、95%或甚至99%相同的。

[0269] 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(Sequence Analysis Software Package of the Genetics Computer Group),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)。此种软件通过将同源性程度指定为各种取代、缺失和/或其他修饰而匹配相同或相似的序列。保守性取代通常包含以下组的组内取代:甘氨酸,丙氨酸;缬氨酸,异亮氨酸,亮氨酸;天冬氨酸,谷氨酸,天冬酰胺,谷氨酰胺;丝氨酸,苏氨酸;赖氨酸,精氨酸;和苯丙氨酸,酪氨酸。在确定同一性程度的示例性方法中,可以使用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。

[0270] “受试者”是指哺乳动物,包含但不限于人类或非人类的哺乳动物,诸如牛、马、犬、绵羊或猫。受试者包含牲畜、饲养以生产劳力和提供货品(诸如食物)的家养动物,包含但不限于牛、山羊、鸡、马、猪、兔和绵羊。

[0271] 术语“靶位点”是指核酸分子内被核碱基编辑器所修饰的序列。在一个实施方式中,所述靶位点被脱氨酶或包括脱氨酶的融合蛋白(例如胞苷或腺嘌呤脱氨酶)脱氨基。

[0272] 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, Multiplex genome engineering using CRISPR/Cas systems, Science 339, 819-823(2013); Mali, P.等, RNA-guided human genome engineering via Cas9, Science 339, 823-826(2013); Hwang, W.Y.等, Efficient genome editing in zebrafish using aCRISPR-Cas system, Nature biotechnology 31, 227-229(2013); Jinek, M.等, RNA-programmed genome editing in human cells, eLife 2, e00471(2013); Dicarlo, J.E.等, Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems, Nucleic acids research(2013); Jiang, W.等, 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)。

[0273] 本文所提供的范围应理解为所述范围内的所有数值的简略表达。例如, 1至50的范围应被理解为包含由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所组成的群组中的任何数字、数字的组合, 或子范围。

[0274] 如本文所用, 术语“治疗(treat)”, “(进行)治疗(treating)”, “治疗(过程)(treatment)”等是指减少或改善一种病症和/或伴随其的症状。应知悉, 尽管不排除, 对病症或病况的治疗不需要所述病症、病况或伴随其的症状被完全消除。

[0275] 如本文所用, 术语“尿嘧啶糖基化酶抑制剂”或“UGI”是指能够抑制尿嘧啶-DNA糖基化酶碱基-切除修复酶的蛋白质。在一些实施方式中, UGI结构域包括野生型UGI或其修饰过的版本。在一些实施方式中, 本文所提供的UGI蛋白包含UGI的片段和与UGI或UGI片段同源的蛋白。例如, 在一些实施方式中, UGI结构域包括本文下面列出的氨基酸序列的片段。在一些实施方式中, UGI片段包括的氨基酸序列包括至少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100%的本文所提供的示例性UGI序列。在一些实施方式中, UGI包括与本文下面列出的氨基酸序列同源的氨基酸序列, 或与本文下面列出的氨基酸序列的片段同源的氨基酸序列。在一些实施方式中, 包括UGI或UGI片段或UGI同源物或UGI片段同源物的蛋白质被称为“UGI变体”。UGI变体与UGI或其片段共享同源性。例如, UGI变体与野生型UGI或本文下面列出的UGI是至少70%相同、至少75%相同、至少80%相同、至少85%相同、至少90%相同、至少95%相同、至少96%相同、至少97%相同、至少98%相同、至少99%相同、至少99.5%相同或至少99.9%相同。在一些实施方式中, UGI变体包括UGI的片段, 使得所述片段野生型UGI的对应的片段或本文下面列出的UGI的对应的片段是至少70%相同、至少80%相同、至少90%相同、至少95%相同、至少96%相同、至少97%相同、至少98%相同、至少99%相同、至少99.5%相同或至少99.9%相同。在一些实施方式中, 所述UGI包括含以下氨基酸序列:

[0276] >sp1P14739IUNGI_BPPB2尿嘧啶-DNA糖基化酶抑制剂

[0277] MTNLSDIIEKETGKQLVIQESILMLPEEVVEEVIGNKPESDILVHTAYDESTDENVM LLT SD APE YKPW ALVIQDS NGENKIKML

[0278] 除非特别说明或从上下文中显而易见, 否则如本文所用, 术语“或(or)”应被理解为包含在内。除非特别说明或从上下文中显而易见, 否则如本文所用, “一(a)”, “一(an)”和“所述(the)”应被理解为单数或复数。

[0279] 除非特别说明或从上下文中显而易见,否则如本文所用,术语“约”应被理解为在本领域的正常容许偏差范围内,例如在平均值的2个标准差内。约可以被理解为在所载明的数值的10%、9%、8%、7%、6%、5%、4%、3%、2%、1%、0.5%、0.1%、0.05%或0.01%的范围内。除非上下文中明确说明,否则本文提供的所有数字数值全部由术语约修饰。

[0280] 本文中,在变量的任何定义中所列举的化学基团列表包含将所述变量定义为所表列出的基团的任何单个基团或组合。本文中,对某变量或某方面的实施方式的列举包含所述实施方式作为任何单个实施方式或与任何其他实施方式或其部分的组合。

[0281] 本文所提供的任何组合物或方法可与本文所提供的任何其他组合物和方法中的一种或多种合并。

附图说明

[0282] 图1是A-至-G碱基编辑器(ABE)融合蛋白核碱基编辑器的示意图,所述融合蛋白核碱基编辑器包括与具有C末端二分核定位信号(NLS)的化脓链球菌(Sp)Cas9切口酶(nCas9)融合的两个腺苷脱氨酶结构域,即野生型(wt) TadA和进化型(evo)版本的TadA。本示意图亦指出了所述Cas9蛋白的三个非结构化区域,其为所述融合蛋白可能断裂的处,而所述融合蛋白断裂得到的N末端片段和C末端片段可以使用断裂内含肽系统(即,分别将内含肽-N与内含肽-C融合到所述N末端片段与C末端片段)使N末端片段和C末端片段重构。

[0283] 图2复制了图1中所描述的融合蛋白并提供了三个图表,其定量了碱基编辑系统的碱基编辑活性,所述碱基编辑系统包含一个核碱基编辑器融合蛋白(即ABE),而所述融合蛋白包括一个剪接过的nCas9。ABE在所示的氨基酸位置(例如,参照SpCas9氨基酸序列的T310、T313、A456、S469和C574)断裂。ABE的N末端片段和C末端片段分别融合到内含肽-N和内含肽-C上。这些片段和所示的向导RNA,各自在单独的质粒上于培养的HEK 293细胞中表达,所述HEK 293细胞表达具有具有5882G>A突变的ABCA4基因的蛋白质。将重构的ABE在ABCA4 5882A>G靶标上的碱基编辑活性与对照ABE的活性进行了比较。碱基编辑活性取决于ABE的N末端片段和C末端片段这二者的存在。当仅表达ABE的N末端片段或C末端片段的一时,未观察到碱基编辑活性。

[0284] 图3是使用图2中所描述的21-nt向导(即向导RNA)而确认了碱基编辑活性的图表。请注意图3中的实验是以异于图2中的型式而执行的。所有重构的ABEs均显示了出良好的碱基编辑活性。所述活性取决于nCas9的N末端片段和C末端片段这二者的存在。当仅表达nCas9的N末端片段或C末端片段时,未观察到活性。将重构的ABE的活性与对照ABE7.09和ABE7.10融合蛋白的活性进行了比较。

[0285] 图4是量化了图2中所描述的碱基编辑活性的图表。在图4中,使用了20个核苷酸(nt)的向导RNA,其包含锤头型核酶(HRz)。将各种重构的ABEs的活性与对照ABE7.09和ABE7.10融合蛋白的活性进行比较。

[0286] 图5A-5D显示了对于ARPE-19细胞的AAV2共-感染的感染复数(MOI)的测定。图5A是显示在各种病毒载量(vg/细胞)下AAV2/CMV-mCherry和AAV2/CMV-EmGFP的1:1共感染的图表。图5B描绘了检测到EmGFP(左)和mCherry(中)的荧光图像;以及显示EmGFP和mCherry共同-定位的合并图像(右)。图5C是描绘了当被AAV2/CMV-mCherry感染时,在各种病毒载量(vg/细胞)下表达mCherry的细胞的百分比的图表。图5D是描绘了当被AAV2/EmGFP感染时,

在各种病毒载量 (vg/细胞) 下表达EmGFP的细胞的百分比的图表。

[0287] 图6是一系列图表,其显示经由双重AAV2感染将断裂编辑器递送至ARPE-19细胞中而在ABCA4 5882A处产生了高的A>G转换。所显示的双重感染的感染复数(MOI)为20,000vg/细胞(上,左);30,000vg/细胞(顶部,右);40,000vg/细胞(底部,左侧);和60,000vg/细胞(底部,右)。

具体实施方式

[0288] 如下所述,本发明提供了用于递送一碱基编辑系统的组合物和方法。本发明至少部分地基于以下发现:可以使A-至-G核碱基编辑器(ABE)“断裂”并使用断裂内含肽将其重构。编码ABE的N末端片段和C末端片段的多核苷酸被分别融合到断裂内含肽对的内含肽-N和内含肽-C上,然后与单向导RNA一起,在分开的载体上被递送到细胞中。编码的ABE片段被剪接在一起以重构功能性核碱基编辑融合蛋白,其对于核酸序列的靶向编辑尤其有用。

[0289] 内含肽

[0290] 内含肽(内含蛋白)是在多种多样的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。

[0291] 在蛋白质剪接中,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接合侧翼的外显肽(外部蛋白)序列。这种重排发生在翻译的后(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。

[0292] 大约5%的内含肽是断裂内含肽,其被转录并翻译为两种单独的多肽,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,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。所述过程的启动是通过活化连接内含肽N-端和N-外显肽的肽键。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,所述中间体通过对第一个C-外显肽残基(+1)(所述残基是半胱氨酸、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所述过程,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内含肽,并留下通过非-肽键合附接的外显肽。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。

[0293] 腺苷脱氨酶

[0294] 在一些实施方式中,本发明的融合蛋白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将腺嘌呤脱氨基。在一些实施方式中,本文所提供的腺苷脱氨酶能够将DNA的脱氧腺苷残基中的腺嘌呤脱氨基。腺苷脱氨酶可以来源于任何适合的生物(例如大肠杆菌)。在一些实施方式中,腺嘌呤脱氨酶是一种天然存在的腺苷脱氨酶,其包含一个或多个与本文所提供的任何突变(例如ecTadA中的突变)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,识别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腺苷脱氨酶(例如具有与ecTadA的同源性)中形成与本文所述(例如在ecTadA中鉴别出的任何突变)的任何突变相对应的突变。在一些实施方式中,腺苷脱氨酶来自原核生物。在一些实施方式中,腺苷脱氨酶来自细菌。在一些实施方式中,腺苷脱氨酶来自大肠杆菌,金黄色葡萄球菌,伤寒沙门菌,腐烂希瓦氏菌,流感嗜血杆菌,新月形杆菌,或枯草芽孢杆菌。在一些实施方式中,腺苷脱氨酶来自大肠杆菌。

[0295] 在一个实施方式中,本发明的融合蛋白包含与TadA7.10链接的野生型TadA,而所述TadA7.10则链接到Cas9切口酶上。在特定实施方式中,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,ABE7.10编辑器包括能够形成异二聚体的TadA7.10和TadA(wt)。相关序列如下:

[0296] TadA(wt):

[0297] SEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVM
QNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLSD
FFRMRRQEIKAKKAQSSTD

[0298] TadA7.10:

[0299] SEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVM
QNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCY
FFRMPRQVFNAQKKAQSSTD

[0300] 在一些实施方式中,以同二聚体或单体,提供了TadA(例如,具有双-股底物活性)或TadA7.10。

[0301] 在一些实施方式中,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以包含一个或多个突变(例如,本文所提供的任何突变)。本发明披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,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,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个相同的连续氨基酸残基。

[0302] 在一些实施方式中,腺苷脱氨酶包括在TadA参考序列中的一D108X突变,或在另一腺苷脱氨酶中的一相应突变,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,腺苷脱氨酶包括在TadA参考序列中的一D108G、D108N、D108V、D108A或D108Y突变,或在另一腺苷脱氨酶中的一相应突变。然而,应当理解,可以类似地对另外的脱氨酶进行比对,以鉴别出同源氨基酸残基,其可如本文所述被突变。

[0303] 在一些实施方式中,腺苷脱氨酶包括在TadA参考序列中的一A106X突变,或在另一

腺苷脱氨酶中的一相应突变,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,腺苷脱氨酶包括在TadA参考序列中的一A106V突变,或在另一腺苷脱氨酶中的一相应突变。

[0304] 在一些实施方式中,腺苷脱氨酶包括在TadA参考序列中的一E155X突变,或在另一腺苷脱氨酶中的一相应突变,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,腺苷脱氨酶包括在TadA参考序列中的一E155D、E155G或E155V突变,或在另一腺苷脱氨酶中的一相应突变。

[0305] 在一些实施方式中,腺苷脱氨酶包括在TadA参考序列中的一D147X突变,或在另一腺苷脱氨酶中的一相应突变,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,腺苷脱氨酶包括在TadA参考序列中的一D147Y突变,或在另一腺苷脱氨酶中的一相应突变。

[0306] 应当理解,可以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,例如金黄色葡萄球菌TadA(saTadA),或其他腺苷脱氨酶(例如细菌的腺苷脱氨酶)。对于本领域技术人员而言显而易见的是如何制备与ecTadA中的突变残基同源的突变。因此,可以在其他具有同源氨基酸残基的腺苷脱氨酶中制备于ecTadA中鉴别出的任何突变。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,腺苷脱氨酶可以含有在TadA参考序列中的D108N、A106V、E155V和/或D147Y的突变,或在另一腺苷脱氨酶中的相应突变。在一些实施方式中,一腺苷脱氨酶包括在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E155V;D108N、A106V,和D147Y;D108N、E155V,和D147Y;A106V、E155V,和D147Y;和D108N、A106V、E155V,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。

[0307] 在一些实施方式中,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、K110X、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,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、I95L、V102A、F104L、A106V、R107C或R107H或R107P、D108G或D108N或D108V或D108A或D108Y、K110L、M118K、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。

[0308] 在一些实施方式中,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,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。

[0309] 在一些实施方式中,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,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。

[0310] 在一些实施方式中,腺苷脱氨酶包括在TadA参考序列中的一、二、三、四、五或六个突变,其选自由以下所组成的群组:H8X、D108X、N127X、D147X、R152X和Q154X,或在另一腺苷脱氨酶中的一个相应突变或多个相应突变,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,腺苷脱氨酶包括在TadA参考序列中的一、二、三、四、五、六、七或八个突变,其选自由以下所组成的群组:H8X、M61X、M70X、D108X、N127X、Q154X、E155X和Q163X,或在另一腺苷脱氨酶中的一个相应突变或多个相应突变,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,腺苷脱氨酶包括在TadA参考序列中的一、二、三、四或五个突变,其选自由以下所组成的群组:H8X、D108X、N127X、E155X和T166X,或在另一腺苷脱氨酶中的一个相应突变或多个相应突变,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,腺苷脱氨酶包括在TadA参考序列中的一、二、三、四、五或六个突变,其选自由以下所组成的群组:H8X、A106X、D108X在另一腺苷脱氨酶中的一个相应突变或多个相应突变,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,腺苷脱氨酶包括在TadA参考序列中的一、二、三、四、五、六、七或八个突变,其选自由以下所组成的群组:H8X、R126X、L68X、D108X、N127X、D147X和E155X,或在另一腺苷脱氨酶中的一个相应突变或多个相应突变,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,腺苷脱氨酶包括在TadA参考序列中的一、二、三、四或五个突变,其选自由以下所组成的群组:H8X、D108X、A109X、N127X和E155X,或在另一腺苷脱氨酶中的一个相应突变或多个相应突变,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。

[0311] 在一些实施方式中,腺苷脱氨酶包括在TadA参考序列中的一、二、三、四、五或六个突变,其选自由以下所组成的群组:H8Y、D108N、N127S、D147Y、R152C和Q154H,或在另一腺苷脱氨酶中的一个相应突变或多个相应突变。在一些实施方式中,腺苷脱氨酶包括在TadA参考序列中的一、二、三、四、五、六、七或八个突变,其选自由以下所组成的群组:H8Y、M61I、M70V、D108N、N127S、Q154R、E155G和Q163H,或在另一腺苷脱氨酶中的一个相应突变或多个相应突变。在一些实施方式中,腺苷脱氨酶包括在TadA参考序列中的一、二、三、四或五个突变,其选自由以下所组成的群组:H8Y、D108N、N127S、E155V和T166P,或在另一腺苷脱氨酶中的一个相应突变或多个相应突变。在一些实施方式中,腺苷脱氨酶包括在TadA参考序列中的一、二、三、四、五或六个突变,其选自由以下所组成的群组:H8Y、A106T、D108N、N127S、E155D和K161Q,或在另一腺苷脱氨酶中的一个相应突变或多个相应突变。在一些实施方式中,腺苷脱氨酶包括在TadA参考序列中的一、二、三、四、五、六、七或八个突变,其选自由以下所组成的群组:H8Y、R126W、L68Q、D108N、N127S、D147Y和E155V,或在另一腺苷脱氨酶中的一个相应突变或多个相应突变。在一些实施方式中,腺苷脱氨酶包括在TadA参考序列中的一、二、三、四或五个突变,其选自由以下所组成的群组:H8Y、D108N、A109T、N127S和E155G,或在另一腺苷脱氨酶中的一个相应突变或多个相应突变。

[0312] 在一些实施方式中,其腺苷脱氨酶包括一个或多个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,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,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。在一些实施方式中,腺苷脱氨酶包括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,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,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,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,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,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。

[0313] 在一些实施方式中,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,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。

[0314] 在一些实施方式中,腺苷脱氨酶包括一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,腺苷脱氨酶包括在TadA参考序列中的一L84F突变,或在另一腺苷脱氨酶中的一个相应突变。

[0315] 在一些实施方式中,腺苷脱氨酶包括在TadA参考序列中的一H123X突变,或在另一腺苷脱氨酶中的一个相应突变,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。在一些实施方式中,腺苷脱氨酶包括在TadA参考序列中的一H123Y突变,或在另一腺苷脱氨酶中的一个相应突变。

[0316] 在一些实施方式中,腺苷脱氨酶包括在TadA参考序列中的一I157X突变,或在另一腺苷脱氨酶中的一个相应突变,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。在一些实施方式中,腺苷脱氨酶包括在TadA参考序列中的一I157F突变,或在另一腺苷脱氨酶中的一个相应突变。

[0317] 在一些实施方式中,腺苷脱氨酶包括在TadA参考序列中的一、二、三、四、五、六或七个突变,其选自由以下所组成的群组:L84X、A106X、D108X、H123X、D147X、E155X和I156X,或在另一腺苷脱氨酶中的一个相应突变或多个相应突变,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,腺苷脱氨酶包括在TadA参考序列中的一、二、三、四、五或六个突变,其选自由以下所组成的群组:S2X、I49X、A106X、D108X、D147X和E155X,或在另一腺苷脱氨酶中的一个相应突变或多个相应突变,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,腺苷脱氨酶包括在TadA参考序列中的一、二、三、四或五个突变,其选自由以下所组成的群组:H8X、A106X、D108X、N127X和K160X,或在另一腺苷脱氨酶中的一个相应突变或多个相应突变,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。

[0318] 在一些实施方式中,腺苷脱氨酶包括在TadA参考序列中的一、二、三、四、五、六或七个突变,其选自由以下所组成的群组:L84F、A106V、D108N、H123Y、D147Y、E155V和I156F,或在另一腺苷脱氨酶中的一个相应突变或多个相应突变。在一些实施方式中,腺苷脱氨酶包括在TadA参考序列中的一、二、三、四、五或六个突变,其选自由以下所组成的群组:S2A、I49F、A106V、D108N、D147Y和E155V。

[0319] 在一些实施方式中,腺苷脱氨酶包括在TadA参考序列中的一、二、三、四或五个突变,其选自由以下所组成的群组:H8Y、A106T、D108N、N127S和K160S,或在另一腺苷脱氨酶中的一个相应突变或多个相应突变。

[0320] 在一些实施方式中,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,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,腺苷脱氨酶包括一个或多个本文所述的相应于TadA参考序列中的突变,或在另一腺苷脱氨酶中的一个或多个相应突变。

[0321] 在一些实施方式中,腺苷脱氨酶包括在TadA参考序列中的一个E25X突变,或在另一腺苷脱氨酶中的一个相应突变,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。在一些实施方式中,腺苷脱氨酶包括在TadA参考序列中的一个E25M、E25D、E25A、E25R、E25V、E25S或E25Y突变,或在另一腺苷脱氨酶中的一个相应突变。

[0322] 在一些实施方式中,腺苷脱氨酶包括在TadA参考序列中的一个R26X突变,或在另一腺苷脱氨酶中的一个相应突变,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。在一些实施方式中,腺苷脱氨酶包括在TadA参考序列中的一个R26G、R26N、R26Q、R26C、R26L或R26K突变,或在另一腺苷脱氨酶中的一个相应突变。

[0323] 在一些实施方式中,腺苷脱氨酶包括在TadA参考序列中的一个R107X突变,或在另一腺苷脱氨酶中的一个相应突变,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。在一些实施方式中,腺苷脱氨酶包括在TadA参考序列中的一个R107P、R07K、R107A、R107N、R107W、R107H或R107S突变,或在另一腺苷脱氨酶中的一个相应突变。

[0324] 在一些实施方式中,腺苷脱氨酶包括在TadA参考序列中的一个A142X突变,或在另一腺苷脱氨酶中的一个相应突变,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。在一些实施方式中,腺苷脱氨酶包括在TadA参考序列中的一个A142N、A142D、A142G突变,或在另一腺苷脱氨酶中的一个相应突变。

[0325] 在一些实施方式中,腺苷脱氨酶包括在TadA参考序列中的一个A143X突变,或在另一腺苷脱氨酶中的一个相应突变,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。在一些实施方式中,腺苷脱氨酶包括在TadA参考序列中的一个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相应突变。

[0326] 在一些实施方式中,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,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。

[0327] 在一些实施方式中,腺苷脱氨酶包括在TadA参考序列中的一个H36X突变,或在另一腺苷脱氨酶中的一个相应突变,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。在一些实施方式中,腺苷脱氨酶包括在TadA参考序列中的一个H36L突变,或在另一腺苷脱氨酶中的一个相应突变。

[0328] 在一些实施方式中,腺苷脱氨酶包括在TadA参考序列中的一个N37X突变,或在另一腺苷脱氨酶中的一个相应突变,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。在一些实施方式中,腺苷脱氨酶包括在TadA参考序列中的一个N37T或N37S突变,或在另一腺苷脱氨酶中的一个相应突变。

[0329] 在一些实施方式中,腺苷脱氨酶包括在TadA参考序列中的一个P48X突变,或在另一腺苷脱氨酶中的一个相应突变,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。在一些实施方式中,腺苷脱氨酶包括在TadA参考序列中的一个P48T或P48L突变,或在另一腺苷脱氨酶中的一个相应突变。

[0330] 在一些实施方式中,腺苷脱氨酶包括在TadA参考序列中的一个R51X突变,或在另一腺苷脱氨酶中的一个相应突变,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。在一些实施方式中,包括在TadA参考序列中的一个R51H或R51L突变,或在另一腺苷脱氨酶中的一个相应突变。

[0331] 在一些实施方式中,腺苷脱氨酶包括在TadA参考序列中的一个S146X突变,或在另一腺苷脱氨酶中的一个相应突变,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。在一些实施方式中,包括在TadA参考序列中的一个S146R或S146C突变,或在另一腺苷脱氨酶中的一个相应突变。

[0332] 在一些实施方式中,腺苷脱氨酶包括在TadA参考序列中的一个K157X突变,或在另一腺苷脱氨酶中的一个相应突变,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。在一些实施方式中,腺苷脱氨酶包括在TadA参考序列中的一个K157N突变,或在另一腺苷脱氨酶中的一个相应突变。

[0333] 在一些实施方式中,腺苷脱氨酶包括在TadA参考序列中的一个P48X突变,或在另一腺苷脱氨酶中的一个相应突变,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。在一些实施方式中,腺苷脱氨酶包括在TadA参考序列中的一个P48S、P48T或P48A突变,或在另一腺苷脱氨酶中的一个相应突变。

[0334] 在一些实施方式中,腺苷脱氨酶包括在TadA参考序列中的一个A142X突变,或在另一腺苷脱氨酶中的一个相应突变,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。在一些实施方式中,腺苷脱氨酶包括在TadA参考序列中的一个A142N突变,或在另一腺苷脱氨酶中的一个相应突变。

[0335] 在一些实施方式中,腺苷脱氨酶包括在TadA参考序列中的一个W23X突变,或在另一腺苷脱氨酶中的一个相应突变,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。在一些实施方式中,腺苷脱氨酶包括在TadA参考序列中的一个W23R或W23L突变,或在另一腺苷脱氨酶中的一个相应突变。

[0336] 在一些实施方式中,腺苷脱氨酶包括在TadA参考序列中的一个R152X突变,或在另一腺苷脱氨酶中的一个相应突变,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。在一些实施方式中,腺苷脱氨酶包括在TadA参考序列中的一个R152P或R52H突变,或在另一腺苷脱氨酶中的一个相应突变。

[0337] 在一个实施方式中,腺苷脱氨酶可能包括突变H36L、R51L、L84F、A106V、D108N、H123Y、S146C、D147Y、E155V、I156F和K157N。在一些实施方式中,腺苷脱氨酶包括以下相对于TadA参考序列的突变组合,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_S127S_D 147Y_Q154H), (H8Y_R24W_D108N_N127S_D147Y_E155V),

[0338] (D108N_D147Y_E155V), (H8Y_D108N_S127S),

[0339] (H8Y_D108N_N127S_D147Y_Q154H), (A106V D108N D147Y E155V) (D108Q D147Y E155V) (D108M_D147Y_E155V),

[0340] (D108L_D147Y_E155V), (D108K_D147Y_E155V),

[0341] (D108I_D147Y_E155V), (D108F_D147Y_E155V),

[0342] (A106V_D108N_D147Y), (A106V_D108M_D147Y_E155V),

[0343] (E59A_A106V_D108N_D147Y_E155V), (E59A cat dead (催化失活)_A106V_D108N_D147Y_E155V),

[0344] (L84F_A106V_D108N_H123Y_D147Y_E155V_I156Y),

[0345] (L84F_A106V_D108N_H123Y_D147Y_E155V_I156F), (D103A_D014N),

[0346] (G22P_D 103 A_D 104N), (G22P_D 103 A_D 104N_S 138 A), (D 103 A_D 104N_S 138A),

[0347] (R26G_L84F_A106V_R107H_D108N_H123Y_A142N_A143D_D147Y_E155V_I156F),

[0348] (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),

[0349] (E25M_R26G_L84F_A106V_R107P_D108N_H123Y_A142N_A143D_D147Y_E155V_I156F),

[0350] (R26C_L84F_A106V_R107H_D108N_H123Y_A142N_D147Y_E155V_I156F), (L84F_A106V_D108N_H123Y_A142N_A143L_D147Y_E155V_I156F),

[0351] (R26G_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F),

[0352] (E25A_R26G_L84F_A106V_R107N_D108N_H123Y_A142N_A143E_D147Y_E155V_I156F),

[0353] (R26G_L84F_A106V_R107H_D108N_H123Y_A142N_A143D_D147Y_E155V_I156F),

[0354] (A106V_D108N_A142N_D147Y_E155V),

[0355] (R26G_A106V_D108N_A142N_D147Y_E155V),

[0356] (E25D_R26G_A106V_R107K_D108N_A142N_A143G_D147Y_E155V),

[0357] (R26G_A106V_D108N_R107H_A142N_A143D_D147Y_E155V),

- [0358] (E25D_R26G_A106V_D108N_A142N_D147Y_E155V) ,
[0359] (A106V_R107K_D108N_A142N_D147Y_E155V) ,
[0360] (A106V_D108N_A142N_A143G_D147Y_E155V) ,
[0361] (A106V_D108N_A142N_A143L_D147Y_E155V) ,
[0362] (H36L_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N) ,
[0363] (N37T_P48T_M70L_L84F_A106V_D108N_H123Y_D147Y_I49V_E155V_I156F) ,
[0364] (N37S_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_K161T) ,
[0365] (H36L_L84F_A106V_D108N_H123Y_D147Y_Q154H_E155V_I156F) ,
[0366] (N72S_L84F_A106V_D108N_H123Y_S 146R_D147Y_E155V_I156F) ,
[0367] (H36L_P48L_L84F_A106V_D108N_H123Y_E134G_D147Y_E155V_I156F) ,
[0368] 57N) ,
[0369] (H36L_L84F_A106V_D108N_H123Y_S 146C_D147Y_E155V_I156F) ,
[0370] (L84F_A106V_D108N_H123Y_S 146R_D147Y_E155V_I156F_K161T) ,
[0371] (N37S_R51H_D77G_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F) ,
[0372] (R51L_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_K157N) ,
[0373] (D24G_Q71R_L84F_H96L_A106V_D108N_H123Y_D147Y_E155V_I156F_K160E) ,
[0374] (H36L_G67V_L84F_A106V_D108N_H123Y_S146T_D147Y_E155V_I156F) ,
[0375] (Q71L_L84F_A106V_D108N_H123Y_L137M_A143E_D147Y_E155V_I156F) ,
[0376] (E25G_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_Q159L) ,
[0377] (L84F_A91T_F104I_A106V_D108N_H123Y_D147Y_E155V_I156F) ,
[0378] (N72D_L84F_A106V_D108N_H123Y_G125A_D147Y_E155V_I156F) ,
[0379] (P48S_L84F_S97C_A106V_D108N_H123Y_D147Y_E155V_I156F) ,
[0380] (W23G_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F) ,
[0381] (D24G_P48L_Q71R_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_Q159L) ,
[0382] (L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F) ,
[0383] (H36L_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_E155V_I156F_K157N) , (N37S_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F_K161T) ,
[0384] (L84F_A106V_D108N_D147Y_E155V_I156F) ,
[0385] (R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N_K161T) ,
[0386] (L84F_A106V_D108N_H123Y_S 146C_D147Y_E155V_I156F_K161T) ,
[0387] (L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N_K160E_K161T) ,
[0388] (L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N_K160E) , (R74Q
[0389] L84F_A106V_D108N_H123Y_D147Y_E155V_I156F) ,
[0390] (R74A_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F) ,
[0391] (L84F_A106V_D108N_H123Y_D147Y_E155V_I156F) ,
[0392] (R74Q_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F) ,
[0393] (L84F_R98Q_A106V_D108N_H123Y_D147Y_E155V_I156F) ,
[0394] (L84F_A106V_D108N_H123Y_R129Q_D147Y_E155V_I156F) ,
[0395] (P48S_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F) , (P48S_A142N) ,

[0396] (P48T_I49V_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F_L157N) ,
[0397] (P48T_I49V_A142N) ,
[0398] (H36L_P48S_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N) ,
[0399] (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) ,
[0400] (H36L_P48T_I49V_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_E155V_I156F_K157N) ,
[0401] (H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N) ,
[0402] (H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_E155V_I156F_K157N) ,
[0403] (H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_A142N_D147Y_E155V_I156F_K157N) ,
[0404] (W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N) ,
[0405] (W23R_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N) ,
[0406] (W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146R_D147Y_E155V_I156F_K161T) ,
[0407] (H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152H_E155V_I156F_K157N) ,
[0408] (H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152P_E155V_I156F_K157N) ,
[0409] (W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152P_E155V_I156F_K157N) ,
[0410] (W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142A_S146C_D147Y_E155 V_I156F_K157N) ,
[0411] (W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142A_S146C_D147Y_R152P_E155V_I156F_K157N) ,
[0412] (W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146R_D147Y_E155V_I156F_K161T) ,
[0413] (W23R_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152P_E155V_I156F_K157N) ,
[0414] (H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_R152P_E155 V_I156F_K157N) .

[0415] 胞苷脱氨酶

[0416] 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,本文所提供的胞嘧啶脱氨酶能够使DNA中的胞嘧啶脱氨基。胞苷脱氨酶可以来源于任何适合的生物体。在一些实施方式中,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,识别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形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,胞苷脱氨酶来自原核生物。在一些实施方式中,胞苷脱氨酶来自细菌。在一些实施方式中,胞苷脱氨酶来自哺乳动物(例如人类)。

[0417] 在一些实施方式中,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以包含一个或多个突变(例如,本文所提供的任何突变)。本发明披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,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,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个相同的连续氨基酸残基。

[0418] 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,所述核酸编辑结构域可以催化C至U的碱基变化。在一些实施方式中,所述核酸编辑结构域是脱氨酶结构域。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,所述脱氨酶是载脂蛋白B mRNA-编辑复合体(APOBEC)家族的脱氨酶。在一些实施方式中,所述脱氨酶是APOBEC1脱氨酶。在一些实施方式中,所述脱氨酶是APOBEC2脱氨酶。在一些实施方式中,所述脱氨酶是APOBEC3脱氨酶。在一些实施方式中,所述脱氨酶是APOBEC3 A脱氨酶。

[0419] 在一些实施方式中,所述脱氨酶是APOBEC3B脱氨酶。在一些实施方式中,所述脱氨酶是APOBEC3C脱氨酶。在一些实施方式中,所述脱氨酶是APOBEC3D脱氨酶。在一些实施方式中,所述脱氨酶是APOBEC3E脱氨酶。在一些实施方式中,所述脱氨酶是APOBEC3F脱氨酶。在一些实施方式中,所述脱氨酶是APOBEC3G脱氨酶。在一些实施方式中,所述脱氨酶是APOBEC3H脱氨酶。在一些实施方式中,所述脱氨酶是APOBEC4脱氨酶。在一些实施方式中,所述脱氨酶是活化-诱导的脱氨酶(AID)。在一些实施方式中,所述脱氨酶是脊椎动物脱氨酶。在一些实施方式中,所述脱氨酶是无脊椎动物脱氨酶。在一些实施方式中,所述脱氨酶是人类、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,所述脱氨酶是人类脱氨酶。在一些实施方式中,所述脱氨酶是大鼠脱氨酶,例如rAPOBEC1。在一些实施方式中,所述脱氨酶是海七鳃鳗胞苷脱氨酶1(pmCDA1)。在一些实施方式中,所述脱氨酶是人类APOBEC3G。在一些实施方式中,所述脱氨酶是人类APOBEC3G的片段。在一些实施方式中,所述脱氨酶是包括D316R D317R突变的人类APOBEC3G变体。在一些实施方式中,所述脱氨酶是人类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,核酸编辑结构域与本文所述任何脱氨酶的脱氨酶结构域相比为至少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。

[0420] 在某些实施方式中,本文所提供的融合蛋白包括一种或多种改善融合蛋白的碱基编辑活性的特征。例如,本文所提供的任何融合蛋白可以包括具有降低的核酸酶活性的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以具有一个没有核酸酶活性的Cas9结构域(dCas9),或切割双股螺旋DNA分子的一股的Cas9结构域,其被称为Cas9切口酶(nCas9)。

[0421] 其他核碱基编辑器

[0422] 本发明提供了核碱基编辑器融合蛋白,其中几乎本领域已知的任何核碱基编辑器都可以取代本发明的融合蛋白中的胞苷脱氨酶或腺苷脱氨酶结构域。

[0423] 核碱基编辑器的Cas9结构域

[0424] 在一些方面,核酸可编程的DNA结合蛋白(napDNAbp)是一个Cas9结构域。本文提供了非限制性的示例性Cas9结构域。所述Cas9结构域可以是核酸酶活性Cas9结构域、核酸酶失活Cas9结构域或Cas9切口酶。在一些实施方式中,所述Cas9结构域是核酸酶活性结构域。例如,所述Cas9结构域可以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,所述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,所述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,所述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,所述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个相同的连续氨基酸残基。

[0425] 在一些实施方式中,Cas9结构域是核酸酶-失活的Cas9结构域(dCas9)。例如,所述dCas9结构域可以结合到双股螺旋核酸分子上(例如,经由gRNA分子),而不切割所述双股螺旋核酸分子的任一股。在一些实施方式中,所述核酸酶-失活的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为任何氨基酸的变化。在一些实施方式中,所述核酸酶-失活的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-失活的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。

[0426] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHE

[0427] RHPIFGNIVDEVAYHEKYPTIYHLRKKLV DSTDKADRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIASLGLTPNFKSNFDLAEDAKLQLSKD TYDDLDNLLAQIGDQYADFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI

PHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEVVDKGA
SAQSFIERMTNFDKNLPNEKVLPKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVK
QLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKT
YAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVS
GQGDSLHEHIANLAGSPAICKGILQTVKVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKE
LGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGK
SDNVPSEEVVKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTK
YDENDKLIREVKVITLKSCLVSDFRKDFQFYKREINNYHHAHDAYLNAVVGITALIKKYPKLESEFVYGDYKVYDV
RKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNI
VKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKEKGKSKKLKSVKELLGITIME
RSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLEFENGRKRLASAGELQKGNELALPSKYVNFYLYASHYEKLK
GSPEDNEQKQLFVEQHKHYLDEIEQISEFSKRVILADANLDKVL SAYNKHDKPIREQAENI IHLFTLTNLGAPA
AFKYFDTTIDRKRYTSTKEVLDTLIIHQSI TGLYETRIDLSQLGGD (参见例如, Qi等, Repurposing
CRISPR as an RNA-guided platform for sequence-specific control of gene
expression, Cell. 2013; 152 (5) : 1173-83, 其全部内容通过引用并入本文)。

[0428] 基于本发明披露和本领域的知识, 其他适合的核酸酶-失活的dCas9结构域对本领域技术人员而言将是显而易见的, 并且在本发明披露的范围内。此类另外的示例性的适合的核酸酶-失活的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, CAS9 transcriptional 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, 所述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, 所述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, 所述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个相同的连续氨基酸残基。

[0429] 在一些实施方式中, 所述Cas9结构域是一个Cas9切口酶。所述Cas9切口酶可以是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, 所述Cas9切口酶切开双股螺旋核酸分子的标靶股, 这意味着所述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位置具有组氨酸。在一些实施方式中, Cas9切口酶切开双股螺旋核酸分子的非-标靶、非-碱基-被编辑的那一股, 这意味着所述Cas9切口酶切开了未与gRNA (例如sgRNA) 碱基配对的那一股, 而所述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10位置或相应的突变处具有一天冬氨酸残基。在一些实施方式中,所述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。基于本发明披露和本领域的知识,其他适合的Cas9切口酶对本领域技术人员而言将是显而易见的,并且在本发明披露的范围内。

[0430] PAM排他性降低的Cas9结构域

[0431] 在一个特定的实施方式中,本发明特写了一种核碱基编辑器,其包括断裂成为两个片段的Cas9结构域,而每个片段具有末端内含肽,即,所述N末端片段在其C末端融合到内含肽系统的一个成员,而所述C末端片段在其N末端具有内含肽系统的一个成员:

[0432] 通常,Cas9蛋白,诸如来自化脓链球菌的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Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage”Nature 533,420-424(2016),其全部内容在此通过引用并入本文。因此,在一些实施方式中,本文所提供的任何融合蛋白都可以含有Cas9结构域,其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”Nature Biotechnology 33,1293-1298(2015);其每个的全部内容在此通过引用并入本文。下表1中描述了几种PAM变体:

[0433] 表1.Cas9蛋白和相应的PAM序列

[0434]

变体	PAM
spCas9	NGG
spCas9-VRQR	NGA
spCas9-VRER	NGCG
SpCas9-MQKFRAER	NGC
xCas9(sp)	NGN
saCas9	NNRRRT
saCas9-KKH	NNNRRT
spCas9-MQKSER	NGCG
spCas9-MQKSER	NGCN
spCas9-LRKIQK	NGTN
spCas9-LRVSQK	NGTN
spCas9-LRVSQL	NGTN

SpyMacCas9	NAA
Cpf1	5' (TTTV)

[0435] 在一些实施方式中,所述PAM是NGC。在一些实施方式中,所述NGC PAM被Cas9变体所识别。在一些实施方式中,所述NGC PAM变体包含一个或多个氨基酸取代,所述取代是选自D1135M、S1136Q、G1218K、E1219F、A1322R、D1332A、R1335E和T1337R (共同地被称为“MQKFRAER”)。

[0436] 在一些实施方式中,所述Cas9结构域是金黄色葡萄球菌的Cas9结构域 (SaCas9)。在一些实施方式中,所述SaCas9是核酸酶活性的SaCas9,核酸酶失活的 (SaCas9d),或SaCas9切口酶 (SaCas9n)。在一些实施方式中,所述SaCas9在本文所提供的任何氨基酸序列中包括N579A突变或相应突变。

[0437] 在一些实施方式中,SaCas9结构域、SaCas9d结构域或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,SaCas9结构域、SaCas9d结构域或SaCas9n结构域可以结合具有NGRRT PAM序列的核酸序列。在一些实施方式中,所述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是任何氨基酸。在一些实施方式中,所述SaCas9结构域包括E781K、N967K和R1014H突变中的一个或多个,或者在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,所述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。

[0438] 示例性的SaCas9序列

[0439] K R N Y I L G L D I G I T S V G Y G I I D Y E T R D V I D A G V R L F K E A N V E N N E G R R S K R G A R R L K R R R R H R I Q R V K K L L F D Y N L L T D H S E L S G I N P Y E A R V K G L S Q K L S E E E F S A A L L H L A K R R G V H N V N E V E E D T G N E L S T K E Q I S R N S K A L E E K Y V A E L Q L E R L K K D G E V R G S I N R F K T S D Y V K E A K Q L L K V Q K A Y H Q L D Q S F I D T Y I D L L E T R R T Y Y E G P G E G S P F G W K D I K E W Y E M L M G H C T Y F P E E L R S V K Y A Y N A D L Y N A L N D L N N L V I T R D E N E K L E Y Y E K F Q I I E N V F K Q K K K P T L K Q I A K E I L V N E E D I K G Y R V T S T G K P E F T N L K V Y H D I K D I T A R K E I I E N A

ELLDQIAKILTIYQSSEDIQEELTNLNSLTQEEIEQISNLKGYTGTHNLSLKAIN
LILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVI
NAIIKKYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEII RTTGKENAKYLI
EKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNKVLVKQEENS
KKGNRTPFQYLSSSDSKISYETFKKHILNLAKGKGRI SKTKKEYLLEERDINRFSVQKDF
INRNLDTRYATRGLMNLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHI
[0440] AEDALIIANADFIFKEWKKLDKAKKVMENQMFEKQAESMPEIETE QEYKEIFITPHQIK
HIKDFKDYKYSHRVDKKPNRELINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLLKLLIN
KSPEKLLMYHHPQTYQKLKLIMEQYGDEKNPLYKYEEETGNYLTKYSKKDNGPVIKKIK
YYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYE
VNSKCYEEAKKLKKISNQAEFIASFYNNDLIKINGELYRVIGVNNDLLNRIEVNMIDITY
REYLENMNDKRPPRIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG

[0441] 上方的残基N579,其带有底线和以粗体显示,可以对其突变(例如,变为A579)以产生SaCas9切口酶。

[0442] 示例性的SaCas9n序列

KRNYILGLDIGITSVGYGIIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLK
RRRRHRIQRVKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRG
VHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTS DYV
KEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHC
TYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYEKFQIIENVFKQKKKPTLK
QIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEI IENAELLDQIAKILTI
YQSSEDIQEELTNLNSLTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAI
FNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAI IKKYGLPNDIIIE
[0443] LAREKNSKDAQKMINEMQKRNRQTNERIEEII RTTGKENAKYLIEKIKLHDMQEGKCLYS
LEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNKVLVKQEEASKKGNRTPFQYLSSSDS
KISYETFKKHILNLA KGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLM
NLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKERNKGYKHAEDALI IANADFI FKE
WKKLDKAKKVMENQMFEKQAESMPEIETE QEYKEIFITPHQIKHIKDFKDYKYSHRVDK
KPNRELINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLLKLINKSPEKLLMYHHPQTY
QKLKLIMEQYGDEKNPLYKYEEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDY
PNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKIS
NQAEFIASFYNNDLIKINGELYR

[0444] VIGVNDLLNRIEVMIDITYREYLENMNDKRPPRIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQI
IKKG

[0445] 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。

[0446] 示例性的SaKKH Cas9

KRNYILGLDIGITSVGYGIIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLK
RRRRHRIQRVKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRG
[0447] VHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTS DYV
KEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHC

TYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYEKFQIIENVFKQKKKPTLK
QIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEI IENAELLDQIAKILTI
YQSSEDIQEELTNLNSLTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAI
FNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAI IKKYGLPNDIIIE
LAREKNSKDAQKMINEMQKRNRQTNERIEE IIRTTGKENAKYLIEKIKLHDMQEGKCLYS
LEAIPLEDLLNNPFNYEVDHII PRSVSFDNSFNKVLVKQEEASKKGNRTPFQYLSSSDS
[0448] KISYETFKKHILNLA KGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLM
NLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHAEDALI IANADFI FKE
WKKLDKAKKVMENQMFEKQAESMPEIETE QEYKEIFITPHQIKHIKDFKDYKYSHRVDK
KPNRLLINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHPQTY
QKLKLIMEQYGDEKNPLYKYEEETGNLYTKYSKKDNGPVIKKIKYYGNKLNHLDI TDDY
PNSRNKVVKLSLKPYPFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKIS
NQAEFIASFYANDLIKINGELYRVIGVNNDLLNRIEVNMIDITYREYLENMNDKRPPHI
KTIASKTQSIKKYSTDILGNLYEVKSKKHPQII KKG.

[0449] 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。

[0450] 在一些实施方式中,所述Cas9结构域是化脓链球菌的Cas9结构域(SpCas9)。在一些实施方式中,所述SpCas9结构域是核酸酶活性的SpCas9,核酸酶失活的SpCas9(SpCas9d),或SpCas9切口酶(SpCas9n)。在一些实施方式中,所述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的外的任何氨基酸。在一些实施方式中,所述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,SpCas9结构域、SpCas9d结构域或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,SpCas9结构域、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。在一些实施方式中,所述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是任何氨基酸。在一些实施方式中,所述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,所述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,所述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是任何氨基酸。在一些实施方式中,所述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,所述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,所述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是任何氨基酸。在一些实施方式中,所述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,所述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。

[0451] 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。

[0452] 示例性的SpCas9

[0453] DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR
RYTRRKNRICYLQEIFSNEMAKVDDSSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTD
KADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFILQVQTYNQLFEEENPINASGVDAKAILSARLSKSRLEN
LIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAIL
LSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPKEYKEIFFDQSKNGYAGYIDGGASQEEFYKFIFK
ILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPYYVGPL
ARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLEYFTVYNELTKVKYVTE
GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKIIKDKDFLDN
EENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKITLDLFLKSD
GFANRNFQMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIIKKGILQTVKVVDELVKVMGRHKPENIVIE
ARENQTTQKGQKNSRERMKRIEEGKELGSQLKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVDH
IVPQSFLKDDSIDNKVLRSDKNRGKSDNVPSEEVVKMKKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFI
KRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVG
TALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEI GKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGE
IVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
VEKGKSKKLKSVKELLGITIMERSSFENPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNE
LALPSKYVNFLYLASHYEKLKGPEDNEQKQLFVEQHKHYLDEII EQISEFSKRVILADANLDKVL SAYNKHDKPI
REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDTLIIHQSI TGLYETRIDLSQLGGD

[0454] 示例性的SpCas9n

[0455] DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR
RYTRRKNRICYLQEIFSNEMAKVDDSSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTD
KADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFILQVQTYNQLFEEENPINASGVDAKAILSARLSKSRLEN
LIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAIL
LSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPKEY

[0456] KEIFFDQSKNGYAGYIDGGASQEEFYKFIFKILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHL
GELHAILRRQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIE
RMTNFDKNLPNEKVLPKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFK

KIECFDSVEISGVEDRFNASLGTYHDLKI IKDKDFLDNEENEDI EDIVLTLTLFEDREMIEERLKYAHLFDDKV
MKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQDLSLHEHI
ANLAGSPAIIKKGILQTVKVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIE EG IKELGSQILKEHPV
ENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVK
MKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKV
ITLKSKLVSDFRKDFQFYKVR EINNYHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVYDVRKMI AKSEQEIGKA
TAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKES
ILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKEKGKSKKLKSVKELLGITIMERS SFEKNPIDFLEAKG
YKEVKKDLIIKLPKYSLFELNGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHK
HYLDEII EQISEFSKRVILADANLDKVL SAYNKH RDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTK
EVL DATLIHQSI TGLYETRIDLSQLGGD

[0457] 示例性的SpEQR Cas9

[0458] DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR
RYTRRNRI CYLQEIFS NEMAKVDD SFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLV DSTD
KADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLF IQLVQTYNQLFEENP INASGVDAKAILSARLSKSRLEN
LIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDA IL
LSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPKEYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKP
ILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPIYVGPL
ARGNSRFAWMTRKSEETITPWNFEVVVDKGASAQSFIERMTNFDKNLPNEKVL PKHSLLEYFTVYNELTKVKYVTE
GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKI IKDKDFLDN
EENEDI EDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSD
GFANRNFMLIHDDSLTFKEDIQKAQVSGQDLSLHEHIANLAGSPAIIKKGILQTVKVDELVKVMGRHKPENIVIEM
ARENQTTQKGQKNSRERMKRIE EG IKELGSQILKEHPVENTQL

QNEKLYLYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDK

[0459] NRGKSDNVPSEEVVKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLV
ETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVR EINNYH
HAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVYDVRKMI AKSEQEIGKATAKYFFYSNI
MNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQT
GGFSKESILPKRNSDKLIARKKDWDPKKYGGFESPTVAYSVLVVAKEKGKSKKLKSVKE

[0460] LLGITIMERS SFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELNGRKRMLASAGELQK
GNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEII EQISEFSKRVIL
ADANLDKVL SAYNKH RDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRK **QYR**STKEV
LDATLIHQSI TGLYETRIDLSQLGGD

[0461] 上方的残基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。

[0462] 示例性的SpVQR Cas9

DKKYSIGLAIGTNSVGWAVITDEYKVPSSKKFKVLGNTDRHSIKKNLIGALLFDSGE
TAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHP
IFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPD
NSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGL
FGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDQYDDDLNLLAQIGDQYADLFLAAKNLS
DAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPKEYKEIFFDQSKN
GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLG
ELHAILRRQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWN
FEEVVDKGASAQSFIERMTNFDKNLPNEKVLPHKSLLEYEFTVYNELTKVKYVTEGMRKP
[0463] AFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL
LKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYT
GWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQG
DSLHEHIANLAGSPAIIKKGILQTVKVVDLVKVMGRHKPENIVIEMARENQTTQKGQKNS
RERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYD
VDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF
DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVIT
LKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVYD
VRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKG
RDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSP
TVAYSVLVV

AKVEKGKSKKLKSVKELLGITIMERSSSFENPIDFLEAKGYKEVKKDLIIKLPKYS
[0464] LFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDEQKQLFVEQH
KHYLDEIIIEQISEFSKRVLADANLDKVL SAYNKHDKPIREQAENI IHLFTLTNLGAPA
AFKYFDTTIDRKQYRSTKEVL DATLIHQSI TGLYETRIDLSQLGGD

[0465] 上方的残基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。

[0466] 示例性的SpVRER Cas9

DKKYSIGLAIGTNSVGWAVITDEYKVPSSKKFKVLGNTDRHSIKKNLIGALLFDSGE
TAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHP
IFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPD
NSDVDKLFILVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNG
FGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDITYDDDLNLLAQIGDQYADLFLAAKNLS
DAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLEPKYKEIFFDQSKN
GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLG
ELHAILRRQEDFYFPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWN
FEEVVDKGASAQSFIERMTNFDKNLPNEKVLPHKSLLEYEFTVYNELTKVKYVTEGMRKP
AFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL
LKI IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYT
[0467] GWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFQMLIHDDSLTFKEDIQKAQVSGQG
DSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEARENQTTQKGQKNS
RERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYD
VDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKMKMKNYWRQLLNAKLITQRKF
DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVIT
LKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVYD
VRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKG
RDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSP
TVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSEKPNIDFLEAKGYKEVKKDLIIK
LPKYSLFELENGRKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIIEQISEFSKRVLADANLDKVL SAYNKH RDKPIREQAENI IHLFTLTN
LGAPAAFKYFDTTIDRKE~~Y~~RSTKEVLDATLIHQSI TGLYETRIDLSQLGGD.

[0468] 上方的残基V1134, R1217, Q1334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。

[0469] 在特定的实施方式中,本发明的融合蛋白包括结合正准PAM序列的dCas9结构域和结合非-正准PAM序列(例如,表1中鉴别出的非-正准PAM)的nCas9结构域。在另一个实施方式中,本发明的融合蛋白包括结合标准PAM序列的nCas9结构域和结合非-标准PAM序列(例如,表1中鉴别出的非-标准PAM)的dCas9结构域。

[0470] 高保真的Cas9结构域

[0471] 本发明披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,所述突变降低了Cas9结构域和DNA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,所述突变降低了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,所述突变使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%。

[0472] 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是任何氨基酸。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,所述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,且对本领域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“无High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects”Nature 529,490-495 (2016);以及Slaymaker,I.M.等“Rationally engineered Cas9 nucleases with improved specificity”Science 351,84-88(2015);其每个的全部内容通过引用并入本文。

[0473] 高保真Cas9结构域的(相对于Cas9的)突变以粗体和底线显示

DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGE
TAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHP
IFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPD
NSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGL
FGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDITYDDDLNLLAQIGDQYADLFLAAKNLS
DAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLEPKYKEIFFDQSKN
GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLG
ELHAILRRQEDFYFPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWN
FEEVVDKGASAQSFIERMTAFDKNLPNEKVLPHKSLLEYEFTVYNELTKVKYVTEGMRKP
AFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL
LKI IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYT
[0474] GWGALSRKLINGIRDKQSGKTILDFLKSDGFANRNFMALIHDDSLTFKEDIQKAQVSGQG
DSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIV IEMARENQTTQKGQKNS
RERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYD
VDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF
DNLTKAERGGLSELDKAGFIKRQLVETRAIITKHVAQILDSRMNTKYDENDKLIREVKVIT
LKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVYD
VRKMI AKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKG
RDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSP
TVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSEKNPIDFLEAKGYKEVKKDLIIK
LPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDEQKQL
FVEQHKHYLDEIIIEQISEFSKRVLADANLDKVL SAYNKH RDKPIREQAENI IHLFTLTN
LGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQSI TGLYETRIDLSQLGGD

[0475] Cas9核酸酶具有两个功能性核酸内切酶结构域：RuvC和HNH。一经与靶结合，Cas9经历构象变化，其使核酸酶结构域定位以切开靶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）。然后通过两种一般的修复途径的一修复所得的DSB：

[0476] (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。

[0477] 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形成切开产物，而产物与底物的比率可用于计算百分比。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)](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)。

[0478] 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测定法可用于产生切开产物,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在原始断裂位点形成小的随机插入或缺失(indels))。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的部分(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. 2013 Sep. 12; 154(6):1380-9;以及Ran等, Nat Protoc. 2013 Nov.; 8(11):2281-2308)。

[0479] 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(indels)。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在靶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,所述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。

[0480] 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)可用于形成特定的核苷酸变化,其范围从可单核苷酸的变化到大的插入,如添加荧光团或标签。

[0481] 为了将HDR用于基因编辑,可以将含有所期望序列的DNA修复模板与gRNA(复数个)和Cas9或Cas9切口酶一起递送入目标细胞类型中。所述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(称为左和右同源臂)。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。所述修复模板可以是单-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。涉及NHEJ的化学性或遗传性抑制基因也可以增加HDR频率。

[0482] 在一些实施方式中,Cas9是修饰过的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其他位点。这些位点称为脱靶位点,并在设计gRNA时需要考虑这些位点。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的合并的活性以形成双股断裂(DSB)。Cas9切口酶,SpCas9的D10A突变体,保留了一个核酸酶结构域并形成DNA切口而非DSB。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。

[0483] 在一些情况下,Cas9是的一种变体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相差一个氨基酸(例如,具有缺失、插入、取代、融合)。在一些实例中,所述变体Cas9多肽具有降低Cas9多肽的核酸酶活性的一氨基酸变化(例如,缺失、插入或取代)。例如,例如,在一些实例中,所述变体Cas9多肽,与相应的野生型Cas9蛋白相比,具有小于50%,小于40%,小于30%,小于20%,小于10%,小于5%或小于1%的核酸酶活性。在一些情况下,所述变体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”。

[0484] 在一些情况下,变体Cas9蛋白具有减少的核酸酶活性。例如,变体Cas9蛋白展现出野生型Cas9蛋白(例如野生型Cas9蛋白)核酸内切酶活性的约小于20%、约小于15%、约小于10%、约小于5%、约小于1%或约小于0.1%。

[0485] 在一些情况下,变体Cas9蛋白可以切开向导标靶序列的互补股,但是其切开双股向导标靶序列的非-互补股的能力是降低的。例如,所述变体Cas9蛋白可能具有降低RuvC结构域功能的突变(氨基酸取代)。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标靶序列的互补股,但是其切开双股向导标靶序列的非-互补股的能力是降低的(因此,当所述变体Cas9蛋白切开双股标靶核酸时,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)。

[0486] 在一些情况下,变体Cas9蛋白可以切开双股向导标靶序列的非-互补股,但是其切开向导标靶序列的互补股的能力是降低的。例如,所述变体Cas9蛋白可能具有降低HNH结构域(RuvC/HNH/RuvC结构域基序)功能的突变(氨基酸取代)。作为非限制性的例子,在一些实施方式中,所述变体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标靶序列的非-互补股,但是其切开向导标靶序列的互补股的能力是降低的(因此,当所述变体Cas9蛋白切开双股标靶核酸时,会得到SSB而不是DSB)。此种Cas9蛋白切开向导标靶序列(例如,单股向导标靶序列)的能力是降低的,但是保留了结合向导标靶序列(例如,单股向导标靶序列)的能力。

[0487] 在一些情况下,变体Cas9蛋白切开双股标靶DNA互补和非-互补股的能力均降低。作为非限制性的例子,在一些情况下,所述变体Cas9蛋白带有D10A和H840A两个突变,使得所述多肽切开双股标靶DNA互补和非-互补股的能力均降低。此种Cas9蛋白切开标靶DNA(例如,单股标靶DNA)的能力是降低的,但是保留了结合标靶DNA(例如,单股标靶DNA)的能力。

[0488] 作为另一个非限制性的例子,在一些情况下,所述变体Cas9蛋白带有W476A和W1126A两个突变,使得所述多肽切开标靶DNA的能力是降低的。此种Cas9蛋白切开标靶DNA(例如,单股标靶DNA)的能力是降低的,但是保留了结合标靶DNA(例如,单股标靶DNA)的能力。

[0489] 作为另一个非限制性的例子,在一些情况下,所述变体Cas9蛋白带有P475A、W476A、N477A、D1125A、W1126A和D1127A突变,使得所述多肽切开标靶DNA的能力是降低的。此种Cas9蛋白切开标靶DNA(例如,单股标靶DNA)的能力是降低的,但是保留了结合标靶DNA(例如,单股标靶DNA)的能力。

[0490] 作为另一个非限制性的例子,在一些情况下,所述变体Cas9蛋白带有H840A、W476A和W1126A突变,使得所述多肽切开标靶DNA的能力是降低的。此种Cas9蛋白切开标靶DNA(例如,单股标靶DNA)的能力是降低的,但是保留了结合标靶DNA(例如,单股标靶DNA)的能力。作为另一个非限制性的例子,在一些情况下,所述变体Cas9蛋白带有H840A、D10A、W476A和W1126A突变,使得所述多肽切开标靶DNA的能力是降低的。此种Cas9蛋白切开标靶DNA(例如,单股标靶DNA)的能力是降低的,但是保留了结合标靶DNA(例如,单股标靶DNA)的能力。在一些实施方式中,所述变体Cas9已在Cas9 HNH结构域的位置840处复原了催化性的His残基(A840H)。

[0491] 作为另一个非限制性的例子,在一些情况下,所述变体Cas9蛋白带有H840A、

P475A、W476A、N477A、D1125A、W1126A和D1127A突变,使得所述多肽切开靶DNA的能力是降低的。此种Cas9蛋白切开靶DNA(例如,单股靶DNA)的能力是降低的,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,所述变体Cas9蛋白带有D10A、H840A、P475A、W476A、N477A、D1125A、W1126A和D1127A突变,使得所述多肽切开靶DNA的能力是降低的。此种Cas9蛋白切开靶DNA(例如,单股靶DNA)的能力是降低的,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,或当所述变体Cas9蛋白带有P475A、W476A、N477A、D1125A、W1126A和D1127A突变时,所述变体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,所述方法不需要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,所述方法可以包含向导RNA,但所述方法可以在不存在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,使一个或另外一个核酸酶部分去活化)。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,被取代)。同样地,除丙氨酸取代以外的突变也是适合的。

[0492] 在一些实施方式中,具有降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,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,所述变体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被靶向至靶DNA序列)。

[0493] 在一些实施方式中,所述变体Cas蛋白可以是spCas9、spCas9-VRQR、spCas9-VRER、xCas9(sp)、saCas9、saCas9-KKH、spCas9-MQKSER、spCas9-LRKIQK或spCas9-LRVSQL。

[0494] 化脓链球菌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*)和弗朗西丝菌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丝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,所述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系统的某些局限性。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的参差的切开模式也许提供了定向基因移转的可能性,类似于传统的限制性内切酶克隆,其可以提高基因编辑的效率。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。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。Cpf1蛋白具有RuvC样核酸内切酶结构域,其类似于Cas9的RuvC结构域。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Cpf1有独特的功能,其被归类为2类V型CRISPR系统。所述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,Cpf1-crRNA复合体通过识别原间隔序列邻近基序5'-YTN-3'而切开靶DNA或RNA。在识别PAM的后,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。

[0495] 原间隔序列邻近基序

[0496] 术语“原间隔序列邻近基序(PAM)”或PAM样基序是指紧随在CRISPR细菌适应性免疫系统中被Cas9核酸酶所靶向的DNA序列后的2-6个碱基对的DNA序列。在一些实施方式中,所述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,所述PAM可以是3' PAM(即,位于原间隔序列5'末端的下游)。

[0497] PAM序列对于结合靶是必需的,但是确切的序列取决于Cas蛋白的类型。

[0498] 本文所提供的碱基编辑器可以包括CRISPR蛋白-衍生的结构域,所述结构域能够结合含有标准的或非-标准的原间隔序列邻近基序(PAM)序列的核苷酸序列。PAM位点是临近靶多核苷酸序列的核苷酸序列。本发明披露的一些方面提供了碱基编辑器,其包括具有不同PAM特异性的全部或部分的CRISPR蛋白质。例如,Cas9蛋白通常,诸如化脓链球菌的Cas9(spCas9),需要标准的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2-6个核苷酸之间。表1中描述了几种PAM变体。

[0499] 在一些实施方式中,所述SpCas9对PAM核酸序列5'-NGC-3'或5'-NGG-3'具有特异性。在以上方面的各种实施方式中,所述SpCas9是Cas9或表1所列出的Cas9变体。在以上方面的各种实施方式中,所述修饰过的SpCas9是spCas9-MQKFRAER。在一些实施方式中,所述变体Cas蛋白可能是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,使用了修饰过的SpCas9,其包含氨基酸取代D1135M、S1136Q、G1218K、E1219F、A1322R、D1332A、R1335E和T1337R(SpCas9-MQKFRAER)并对改变的PAM 5'-NGC-3'具有特异性。

[0500] 在一些实施方式中,所述PAM是NGT。在一些实施方式中,所述NGT PAM是一变体。在一些实施方式中,所述NGT PAM变体是通过在一个或多个1335、1337、1135、1136、1218和/或1219残基处的靶向突变而创建的。在一些实施方式中,所述NGT PAM变体是通过在一个或多个1219、1335、1337、1218残基处的靶向突变而创建的。在一些实施方式中,所述NGT PAM变体是通过在1135、1136、1218、1219和1335残基处的一个或多个的靶向突变而创建的。在一些实施方式中,所述NGT PAM变体选自下表2和3中所提供的靶向突变。

[0501] 表2:在残基1219,1335,1337,1218处的NGT PAM变体突变

[0502]

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	

[0503] 表3:在残基1135,1136,1218,1219和1335处的NGT PAM变体突变

[0504]

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

[0505]

54					Y
55	N1286Q	I1331F			

[0506] 在一些实施方式中,所述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进的NGT PAM识别能力。

[0507] 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,所述NGT PAM变体选自下表4中所提供的变体,以得到具有改进的识别能力的突变。

[0508] 表4:在残基1219、1335、1337和1218处的NGT PAM变体突变

[0509]

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

[0510] 在一些实施方式中,所述NGT PAM选自下表5所提供的变体。

[0511] 表5.NGT PAM变体

[0512]

	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

[0513] 在一些实施方式中,所述Cas9结构域是化脓链球菌的Cas9结构域(SpCas9)。在一些实施方式中,所述SpCas9结构域是核酸酶活性的SpCas9,核酸酶失活的SpCas9(SpCas9d),或SpCas9切口酶(SpCas9n)。在一些实施方式中,所述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,所述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,所述SpCas9结构域,SpCas9d结构域或SpCas9n结构域可以结合具有非-标准PAM的核酸序列。在一些实施方式中,所述SpCas9结构域,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。

[0514] 在一些实施方式中,所述SpCas9结构域包括D1135X、R1335X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是任何氨基酸。在一些实施方式中,所述SpCas9结构域包括D1135E、R1335Q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,所述SpCas9结构域包括D1135E、R1335Q和T1336R的突变,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,所述SpCas9结构域包括D1135X、R1335X和T1336X的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是任何氨基酸。在一些实施方式中,所述SpCas9结构域包括D1135V、R1335Q和T1336R的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,所述SpCas9结构域包括D1135V、R1335Q和T1336R的突变,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,所述SpCas9结构域包括D1135X、G1217X、R1335X和T1336X的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是任何氨基酸。在一些实施方式中,所述SpCas9结构域包括D1135V、G1217R、R1335Q和T1336R的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,所述SpCas9结构域包括D1135V、G1217R、R1335Q和T1336R的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。

[0515] 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结构域是由本文所述的任何Cas9多肽的氨基酸序列所组成。

[0516] 在一些例子中,被本文所揭示的碱基编辑器的CRISPR蛋白-衍生的结构域所识别的PAM可以提供给细胞,且所述PAM是在相对于编码前述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,因为在与靶序列相同的多核苷酸上不存在临近的PAM。

[0517] 在一种实施方式中,化脓链球菌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,可以使用一不同的核酸内切酶以靶向某些基因组靶。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-SpCas9s”可以结合多种亦可用于本发明披露的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)可能导致携带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,所述SaCas9核酸内切酶能够在体外情况下的哺乳动物细胞中和在体内情况下的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'-NNNGATT)的那些PAM,也可以在靶基因的临近处被找到。

[0518] 在一些实施方式中,对于化脓链球菌系统,靶基因序列可以居先于(即,为5'至)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至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PAM上游的紧邻,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:

[0519] 示例性的结合-PAM的SpCas9的氨基酸序列如下:

[0520] MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLV DST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRL E
NLIAQLPGEKKNGLFGNLI ALSGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASQSFIERMTNFDKNLPNEKVLPKHSLLEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFQMLIHDDSLTFKEDIQKAQVSGQDLSLHEHIANLAGSPA IKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYD VD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSEL DKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVR EINNYHHAHDAYLNAV
[0521] VGTALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEI GKATAKYFFYSNIMNFFKTEITLANGEIRKR
PLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDS
PTVAYSVLVVAKEVGKSKKLKSVKELLGITIMERS SFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKR
MLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEII EQISEFSKRVILADANLD
KVL SAYNKHDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQSI TGLYETRIDLSQ
LGGD.

[0522] 示例性的结合-PAM的SpCas9n的氨基酸序列如下:

[0523] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERH
[0524] PIFGNIVDEVAYHEKYPTIYHLRKKLV DSTDKADLRLIYLALAHMIKFRGHFLIEGDLNP
[0525] DNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRL ENL I AQLPGEKKNG
[0526] LFGNLI ALSGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNL
[0527] SDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSK
[0528] NGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHL
[0529] GELHAILRRQEDFYFPLKDNREKIEKILTFRIPYYVG PLARGNSRFAWMTRKSEETITPW
[0530] NFEEVVDKGASQSFIERMTNFDKNLPNEKVLPKHSLLEYFTVYNELTKVKYVTEGMRK
[0531] PAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHD
[0532] LLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRY
[0533] TGWGRLSRKLINGIRDKQSGKTILDFLKS DGFANRNFQMLIHDDSLTFKEDIQKAQVSGQ
[0534] GDLSLHEHIANLAGSPA IKKGILQTVKVVDELVKVMGRHKPENIVIE MARENQTTQKGQKN
[0535] SRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDY

- [0536] DVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRK
[0537] FDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVI
[0538] TLKSKLVSDFRKDFQFYKVBREINNYHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVV
[0539] DVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDK
[0540] GRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDS
[0541] PTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSSFENPIDFLEAKGYKEVKKDLII
[0542] KLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLGSPEDNEQKQ
[0543] LFVEQHKHYLDEIIIEQISEFSKRVILADANLDKVL SAYNKH RDKPIREQAENI IHLFTLT
[0544] NLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQSI TGLYETRIDLSQLGGD.
[0545] 示例性的结合-PAM的SpEQR Cas9的氨基酸序列如下:

MDKKYSIGLAIGTNSVGWAVITDEYKVP SKKFKVLGNTDRHSIKKNLIGALLFDSG
ETAETRLKRTARRRYTRRKNRICYLQE IFSNEMAKVDDSFHRLEESFVEEDKKHERHP
IFGNIVDEVAYHEKYPTIYHLRKKLV DSTDKADLR LIYLALAHMIKFRGHFLIEGDLNPD
NSDVDKLF IQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGL
[0546] FGNLIALSLGLTPNFKSNFDLAEDAKLQLSKD TYDDDLNLLAQIGDQYADLFLAAKNLS
DAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLP EKYKEIFFDQSKN
GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLG
ELHAILRRQEDFYFPLKDNREKIEKILTFRIPIYYVGPLARGNSRFAMTRKSEETITPWN
FEEVVDKGASAQSFIERMTNFDKNLPNEKVL PKHSLLYEYFTVYNELTKVKYVTEGMRKP
AFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL
LKI IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYT
GWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQG
DSLHEHIANLAGSPA IKKGILQTVKVVD ELVKVMGRHKPENIV IEMARENQTTQKGQKNS
RERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYD
VDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF
DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVIT
LKSKLVSDFRKDFQFYKVBREINNYHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVV
VRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKG
RDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFESP
TVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSSFENPIDFLEAKGYKEVKKDLIIK
LPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLGSPEDNEQKQL
FVEQHKHYLDEIIIEQISEFSKRVILADANLDKVL SAYNKH RDKPIREQAENI IHLFTLTN
LGAPAAFKYFDTTIDRKQYRSTKEVL DATLIHQSI TGLYETRIDLSQLGGD。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SpEQRCas9,带有底线和以粗体显示。

[0547] 示例性的结合-PAM的SpVQR Cas9的氨基酸序列如下:

[0548] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSITDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDFLFIQLVQTYNQLFEEINPINASGVDAKAILSARLSKSRRLLENLIAQLPGEKKNGLFGNLIALLSLGLTPNFKSNDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASQSFIERMTNFDKNLPNEKVLPHSLLYEYFTVYNELTKVKYVT

EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKKIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF

[0549] LKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIIKKGIL

QTVKVVDDELVKVMGRHKPENIVIMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLT
RSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK
RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVREI
NNYHHAHDAYLNAVVGITALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFF
YSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIKKTE
EVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKEVGKSKKLLK
SVKELLGITIMERSSSFENPIDFLEAKGYKEVKKDLIKLPKYSLELENGRKRMLASAG
ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDEQKQLFVEQHKHYLDEITIEQISEFSK
RVILADANLDKVL SAYNKHRRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRS

TKEVLDTLIHQSI TGLYETRIDLSQLGGD。在这个序列中,残基V1135,Q1335和R1336,其可以由D1135,R1335, and T1336突变而来以产生SpVQR Cas9,带有底线和以粗体显示。

[0550] 示例性的结合-PAM的SpVRER Cas9的氨基酸序列如下:

[0551] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSITDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDFLFIQLVQTYNQLFEEINPINASGVDAKAILSARLSKSRRLLENLIAQLPGEKKNGLFGNLIALLSLGLTPNFKSNDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASQSFIERMTNFDKNLPNEKVLPHSLLYEYFTVYNELTKVKYVT

EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKKIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS

DGFANRNFQMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYD
HIVPQSFLKDDSIDNKVLRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVIENNYHHAHDAYLNAV
GTALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEI GKATAYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDK

GRDFATVRKVL SMPQVNI VKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPPKKYGGFVS
PTVAYSVLVVAKEVGKSKKLKSVKELLGITIMERSSEKPNIDFLEAKGYKEVKKDLII
KLPKYSLELNGRKRMLASARELQKGNELALPSKYVNFLYLAS

[0552]

HYEKLKGPEDNEQKQLFVEQHKKHYLDEII EQISEFSKRVILADANLDKVL SAYNK
HRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKEEYRSTKEVLDATLIHQSI TGLY
ETRIDLSQLGGD.

[0553] 在一些实施方式中,所述Cas9结构域是重组Cas9结构域。在一些实施方式中,所述重组Cas9结构域是SpyMacCas9结构域。在一些实施方式中,所述SpyMacCas9结构域是核酸酶活性的SpyMacCas9,核酸酶失活的SpyMacCas9 (SpyMacCas9d),或SpyMacCas9切口酶(SpyMacCas9n)。在一些实施方式中,SaCas9结构域,SaCas9d结构域或SaCas9n结构域可以结合具有非-标准PAM的核酸序列。在一些实施方式中, SpyMacCas9结构域, SpCas9d结构域或SpCas9n结构域可以结合具有NAA PAM序列的核酸序列。

[0554] 示例性的SpyMacCas9

[0555] MDKKYSIGLDIGTNSVGWAVITDDYKVP SKKFKVLGNTDRHS IKKNLIGALLFGSGETAETRLKRTA
RRRYTRRKNRICYLQEIFS NEMAKVDDSFHRL EESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLAD
STDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQIYNQLFEENPINASRVDAKAILSARLSKSR
RLENLIAQLPGEKRNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKD TYDDDLDNLLAQIGDQYADFLAANKL
SDAILLS DILRVNSEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKF IKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRI
PYVVGPLARGNSRFAWMTRKSEETITPWNFEVV DKGASAQSFIERMTNFDKNLPNEKVL PKHSLLYEYFTVYNEL
TKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGAYHDLLKI
IKDKDFLDNEENEDILEDIVLTTLTFEDRGMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG
KTILDFLKSDGFANRNFQMLIHDDSLTFKEDIQKAQVSGQGHSLEHIANLAGSPAIIKKGILQTVKIVDELVKVMG
HKPENIVIE MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELD
INRLSDYDVDHIVPQSFIKDDSIDNKVLRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAER
GGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVIENNY
HHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEI GKATAYFFYSNIMNFFKTEITLANGEI
RKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNI VKKTEIQTVGQNGGLFDDNPKSPLEVTPSKLVPLKKELN
PKKYGGYQKPTTAYPVLLITDTKQLIPISVMNKKQFEQNPVKFLRDRGYQQVGKNDFIKLPKYTLVDIGDGIKRLW
ASSKEIHKGNQLVVS KKSQILLYHAHHLSDLSNDYLQNHNNQFDFVLFNEIISFSKKCKLGKEHIQKIENVYSNKK
NSASIEELAESFIKLLGFTQLGATSPFNFLGVKLNQKQYKGGKYILPCTEGTLIRQSITGLYETRVDLSKIGED.

[0556] 在一些情况下,变体Cas9蛋白带有H840A,P475A,W476A,N477A,D1125A,W1126A和

D1218A突变,使得所述多肽切开靶DNA或RNA的能力是降低的。此种Cas9蛋白切开靶DNA(例如,单股靶DNA)的能力是降低的,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,所述变体Cas9蛋白带有D10A、H840A、P475A、W476A、N477A、D1125A、W1126A和D1218A突变,使得所述多肽切开靶DNA的能力是降低的。此种Cas9蛋白切开靶DNA(例如,单股靶DNA)的能力是降低的,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,或者当所述变体Cas9蛋白带有P475A、W476A、N477A、D1125A、W1126A和D1218A突变时,所述变体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,所述方法不需要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,所述方法可以包含向导RNA,但所述方法可以在不存在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,使一个或另外一个核酸酶部分去活性化)。作为非限制性的例子,残基D10、G12、G17、E762、H840、N854、N863、H982、H983、A984、D986和/或A98可以被改变(即,被取代)。同样地,除丙氨酸取代以外的突变也是适合的。

[0557] 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标准PAM序列(NGG)的Cas9蛋白的全部或一部分。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-标准的PAM序列。此种序列已经在本领域中描述,并且对本领域技术人员而言是显而易见的。例如,结合非-标准PAM序列的Cas9结构域已在如下文献中描述: Kleinstiver, B.P.等,“具有改变的PAM特异性的工程化的CRISPR-Cas9核酸酶”*Nature* 523,481-485(2015);以及Kleinstiver, B.P.等,“通过修饰PAM的识别来扩大金黄色葡萄球菌CRISPR-Cas9的靶向范围”*Nature Biotechnology* 33,1293-1298(2015);其每个的全部内容在此通过引用并入本文。

[0558] 包括核定位序列(NLS)的融合蛋白

[0559] 在一些实施方式中,本文所提供的融合蛋白进一步包括一个或多个(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,使用了二分NLS。在一些实施方式中,NLS包括促进包括NLS的蛋白质输入细胞核中(例如,通过核转运)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进一步包括核定位序列(NLS)。在一些实施方式中,NLS融合到融合蛋白的N-端。在一些实施方式中,NLS融合到融合蛋白的C-端。在一些实施方式中,NLS融合到Cas9结构域的N-端。在一些实施方式中,NLS融合到nCas9结构域或dCas9结构域的C-端。在一些实施方式中,NLS融合到脱氨酶的N-端。在一些实施方式中,NLS融合到脱氨酶的C-端。在一些实施方式中,NLS经由一个或多个连接子融合到融合蛋白。在一些实施方式中,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。另外的核定位序列在本领域中是已知的,并且对本领域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披露的引用而并入本文。在一些实施方式中,NLS包括氨基酸序列PKKKRKVEGADKRTADGSEFES PKKKRKV, KRTADGSEFESPKKKRKV, KRPAATKKAGQAKKKK, KKTELQTTNAENKTKKL, KRGINDRNFWRGENGRKTR, RKSGKIAAIVVKRPRKPKKKRKV或MDSLLMNRRKFLYQFKNVRWAKGRRETYLC。

[0560] 在一些实施方式中,NLS存在于连接子中,或连接子位于NLS的两侧翼,例如,本文

所述的连接子。在一些实施方式中,N-端或C-端NLS是二分NLS。二分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,二分-2部分,而单分NLSs则不是)。核质蛋白的NLS,KR[PAATKKAGQA]KKKK,是普遍存在的二分信号的原型: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。示例性的二分NLS的序列如下:

[0561] PKKKRKVEGADKRTADGSEFES PKKKRKV

[0562] 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,在一个或多个结构域或蛋白之间的连接子是存在的。

[0563] 应当理解,本发明披露的融合蛋白可以包括一个或多个附加的特征。例如,在一些实施方式中,所述融合蛋白可以包括抑制剂,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,以及可用于所述融合蛋白的增溶、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。其他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,所述融合蛋白包括一个或多个His标签。

[0564] 连接子

[0565] 在某些实施方式中,连接子可用于链接本发明的任何肽或肽结构域。连接子可以简单如共价键,或者它可以是长度为许多原子的聚合连接子。连接子可以是肽连接子或非-肽连接子。在某些实施方式中,连接子可以是UV-可切开的连接子。在一些实施方式中,连接子可以是多核苷酸连接子,例如,RNA连接子。在某些实施方式中,连接子是多肽或是基于多个氨基酸。在其他实施方式中,连接子不是肽-样的。在某些实施方式中,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,连接子是酰氨键合的碳-氮键。在某些实施方式中,连接子是环状的或无环的,取代的或未取代的,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,连接子是聚合的(例如,聚乙烯,聚乙二醇,聚酰氨,聚酯等)。在某些实施方式中,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,连接子包括一氨基链烷酸(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,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,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,连接子包括聚乙二醇部分体(PEG)。在其他实施方式中,连接子包括氨基酸。在某些实施方式中,连接子包括肽。在某些实施方式中,连接子包括芳基和杂芳基部分体。在某些实施方式中,连接子是基于苯环。连接子可以包含官能化的部分体,以促进亲核基(例如,硫醇基,氨基)从肽到连接子的附接。任何亲电子试剂都可用作连接子的一部分。示例性的亲电子试剂包含但不限于活化的酯,活化的酰氨,麦可(Michael)受体(活化的烯烃),烷基卤化物,芳基卤化物,酰基卤化物,和异硫氰酸酯。

[0566] 在一些实施方式中,连接子是一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,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,连接子是约3至约104个(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、61、62、63、64、65、70、75、80、85、90、95或100个)氨基酸的长度。

[0567] 具有向导RNAs的Cas9复合体

[0568] 本发明披露的一些方面提供了包括本文所提供的任何融合蛋白的复合体,并且可以采用向导RNA(例如,范围从非常柔韧的构成为(GGGG)_n, (GGGG)_n和(G)_n的连接子,到更刚性的构成为(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的连接子)以获得对核碱基编辑器活性最佳的长度。在一些实施方式中,n为1、2、3、4、5、6、7、8、9、10、11、12、13、14或15。在一些实施方式中,连接子包括一(GGS)_n基序,其中n为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的连接子融合。

[0569] 在一些实施方式中,向导核酸(例如,向导RNA)是从15到100个核苷酸的长度,并且包括一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,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,向导RNA包括一与靶序列互补的至少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个接连核苷酸的序列。在一些实施方式中,靶序列是DNA序列。在一些实施方式中,靶序列是细菌,酵母,真菌,昆虫,植物,或动物基因组中的序列。在一些实施方式中,靶序列是人类基因组中的序列。在一些实施方式中,靶序列的3'末端紧邻一标准的PAM序列(NGG)。在一些实施方式中,靶序列的3'末端紧邻一非-标准的PAM序列(例如,表1中所列出的序列)。在一些实施方式中,向导核酸(例如,向导RNA)和与疾病或病症相关联的序列是互补的。

[0570] 本发明披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,本发明披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以及至少一个向导RNA接触,其中所述向导RNA是约15到100个核苷酸的长度,并且包括一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,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,靶序列的3'末端紧邻NGA、NGCG、NGN、NNGRRT、NNNRRT、NGCG、NGCN、NGTN、NGTN、NGTN或5' (TTTV) 序列。

[0571] 在一些实施方式中,本发明的融合蛋白用于诱变处理目标的靶。这些突变可能影响靶的功能。例如,当用核碱基编辑器靶向调节区域时,所述调节区域的功能被改变并且下游蛋白的表达减少。

[0572] 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。

[0573] 对本领域技术人员而言将是显而易见的是,为了将本文所揭示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-表达所述融合蛋白。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。备选地,所述向导RNA和tracrRNA

可以作为两个核酸分子而分开提供。在一些实施方式中,所述向导RNA包括一结构,其中所述向导序列包括与靶序列互补的序列。所述向导序列通常为20核苷酸的长度。基于本发明披露,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。

[0574] 使用包括一胞苷脱氨酶、腺苷脱氨酶和Cas9结构域的融合蛋白的方法

[0575] 本发明披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,本发明披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以及至少一个向导RNA接触,其中所述向导RNA是约15到100个核苷酸的长度,并且包括一与靶序列互补的至少10个核苷酸的序列。在一些实施方式中,靶序列的3'末端紧邻标准的PAM序列(NGG)。在一些实施方式中,靶序列的3'末端不直接与标准的PAM序列(NGG)相邻。在一些实施方式中,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,靶序列的3'末端紧邻NGA、NGCG、NGN、NNGRRT、NNNRRT、NGCG、NGCN、NGTN、NGTN或5' (TTTT)序列。

[0576] 在一些实施方式中,本发明的融合蛋白用于诱变处理目的靶。特别地,本文所述的多-效应器核碱基编辑器能够在靶序列内制造多个突变。这些突变可能影响靶的功能。例如,当用多-效应器核碱基编辑器靶向调节区域时,所述调节区域的功能被改变并且下游蛋白的表达减少。

[0577] 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。

[0578] 对本领域技术人员而言将是显而易见的是,为了将如本文所揭示的包括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-表达所述融合蛋白。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。备选地,所述向导RNA和tracrRNA可以作为两个核酸分子而分开提供。在一些实施方式中,所述向导RNA包括一结构,其中所述向导序列包括与靶序列互补的序列。所述向导序列通常为20核苷酸的长度。基于本发明披露,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。

[0579] 碱基编辑器效率

[0580] 本发明的融合蛋白,通过修饰特定核苷酸碱基而不形成显著比例的插入/缺失,改进了碱基编辑器效率。如本文所用,“插入/缺失”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,期望形成有效修饰(例如,使突变)核酸内的特定核苷酸而不在核酸内形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形成比插入/缺失更大比例的预期修饰(例如,突变)。在一些实施方式中,本文所提供的碱基编辑器能够形成的预期修饰与插入/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形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/缺失的数目可以使用任何适合的方法来确定。

[0581] 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,所述区域是在被碱基编辑器所靶向的核苷酸处,或在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/缺失的数目可以取决于核酸(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)暴露于碱基编辑器的后至少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。

[0582] 本发明披露的一些方面是基于以下认识:本文所提供的任何碱基编辑器能够在核酸(例如,受试者的基因组内的核酸)内有效地形成预期突变而不形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结合的特定的碱基编辑器所形成的突变,而所述gRNA经特别设计以产生所述预期突变。在一些实施方式中,所述预期突变是在基因的编码区域内形成终止密码子(例如提前的终止密码子)的突变。在一些实施方式中,所述预期突变消除一终止密码子的突变。在一些实施方式中,所述预期突变是改变基因的剪接的突变。在一些实施方式中,所述预期突变是改变基因的调节序列(例如,基因启动子或基因阻遏剂)的突变。在一些实施方式中,本文所提供的任何碱基编辑器能够形成大于1:1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)。在一些实施方式中,本文所提供的任何碱基编辑器能够形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或更大。应当理解,在本文的“碱基编辑器效率”节段中所述的碱基编辑器的特征,可以应用于本文所提供的任何融合蛋白或使用所述融合蛋白的方法。

[0583] 用于编辑核酸的方法

[0584] 本发明披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,所述方法是用于编辑核酸(例如,双股DNA序列的碱基对)的核碱基的方法。在一些实施方式中,所述方法包括以下步骤:a)使核酸的标靶区域(例如,双股DNA序列)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)将所述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nCas9切割所述靶区域且所述切割不超过一股,其中与第一核碱基互补的第三核碱基被与第二核碱基互补的第四核碱基替换。在一些实施方式中,所述方法造成核酸中少于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,所述方法造成少于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,所述方法进一步包括用与第四核碱基互补的第五核碱基替换第二核碱基,从而形成预期的编辑过的碱基对(例如,G • C至A • T)。在一些实施方式中,至少5%的预期的碱基对是编辑过的。在一些实施方式中,至少10%、15%、20%、25%、30%、35%、40%、45%或50%的预期的碱基对是编辑过的。

[0585] 在一些实施方式中,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,被切割的单股(带切口的股)杂交到向导核酸上。在一些实施方式中,被切割的单股与包括第一核碱基的股是相对的。在一些实施方式中,所述碱基编辑器包括dCas9结构域。在一些实施方式中,所述碱基编辑器保护或结合未编辑过的股。在一些实施方式中,所述预期的编辑过的碱基对在PAM位点的上游。在一些实施方式中,所述预期的编辑过的碱基对在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,所述预期的编辑过的碱基对在PAM位点的下游。在一些实施方式中,所述预期的编辑过的碱基对在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,所述方法不需要标准的(例如,NGG)PAM位点。在一些实施方式中,所述核碱基编辑器包括连接子。在一些实施方式中,所述连接子是1至25个氨基酸的长度。在一些实施方式中,所述连接子是5至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,所述连接子是32个氨基酸的长度。在另一个实施方式中,一“长连接子”是至少约60个氨基酸的长度。在其他实施方式中,所述连接子是约3至100个氨基酸之间的长度。在一些实施方式中,所述靶区域包括目标窗,其中所述目标窗包括靶核碱基对。在一些实施方式中,所述目标窗包括1至10个核苷酸。在一些实施方式中,所述目标窗是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。在一些实施方式中所述目标窗是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,所述预期的编辑过的碱基对是在目标窗内。在一些实施方式中,所述目标窗包括所述预期的编辑过的碱基对。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所述方法。

[0586] 在一些实施方式中,本发明披露提供了用于编辑核苷酸的方法。在一些实施方式中,本发明披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,所述方法包括a)使双股DNA序列的靶区域(例如,双股DNA序列)与包括碱基编辑器和向导核酸(例如,gRNA)的复合体接触,其中所述靶区域包括靶核碱基对,b)诱导所述靶区域的股分离,c)将所述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)切割所述靶区域且所述切割不超过一股,其中与第一核碱基互补的第三核碱

基被与第二核碱基互补的第四核碱基替换,并且所述第二核碱基被与所述第四核碱基互补的第五核碱基替换,从而形成预期的编辑过的碱基对,其中形成所述预期的编辑过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%的预期碱基对是编辑过的。在一些实施方式中,至少10%、15%、20%、25%、30%、35%、40%、45%或50%预期的碱基对是编辑过的。在一些实施方式中,所述方法造成少于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,被切割的单股(带切口的股)杂交到向导核酸上。在一些实施方式中,被切割的单股与包括第一核碱基的股是相对的。在一些实施方式中,所述预期的编辑过的碱基对在PAM位点的上游。在一些实施方式中,所述预期的编辑过的碱基对在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、个20核苷酸。在一些实施方式中,所述预期的编辑过的碱基对在PAM位点的下游。在一些实施方式中,所述预期的编辑过的碱基对在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,所述方法不需要标准的(例如,NGG) PAM位点。在一些实施方式中,所述连接子是1至25个氨基酸的长度。在一些实施方式中,所述连接子是5至20个氨基酸的长度。在一些实施方式中,所述连接子是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,所述靶区域包括一目标窗,其中所述目标窗包括所述靶核碱基对。在一些实施方式中,所述目标窗包括1至10个核苷酸。在一些实施方式中,所述目标窗是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,所述目标窗是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,所述预期的编辑过的碱基对出现在目标窗内。在一些实施方式中,所述目标窗包括所述预期的编辑过的碱基对。在一些实施方式中,所述核碱基编辑器是本文所提供的任一碱基编辑器。

[0587] 宿主细胞中融合蛋白的表达

[0588] 使用本领域技术人员已知的常规方法,本发明的融合蛋白可以在几乎任何目标宿主细胞中表达,包括但不限于细菌,酵母,真菌,昆虫,植物和动物细胞。通过将编码具有核碱基修饰活性(例如,腺苷脱氨酶或胞苷脱氨酶)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形成融合蛋白。

[0589] 可以通过本领域已知的任何方法来获得编码本文所述的蛋白质结构域的DNA,诸如通过化学合成所述DNA链,通过PCR或通过Gibson组装(Gibson Assembly)方法。通过化学合成或通过化学合成与PCR方法或Gibson组装方法的组合以构建全长DNA的优势在于,可以优化密码子以确保融合蛋白在宿主细胞中是以高水平表达。可以使用遗传密码使用频率数据库(<http://www.kazusa.or.jp/codon/index.html>)来选择优化过的密码子,所述数据库在Kazusa DNA研究院的主页上公开。一经获得,即将编码融合蛋白的多核苷酸并入适合的表达载体中。

[0590] 适合的表达载体包含大肠杆菌-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);适于在昆虫细胞中表达的质粒(例如pFast-Bac);适于在哺乳动物细胞中表达的质粒(例如pA1-11,pXT1,pRc/CMV,pRc/RSV,pcDNA1/Neo);还有噬菌体,诸如 λ 噬菌体等;其他的可用载体,包含昆虫病毒载体,诸如杆状病毒等(例如BmNPV,AcNPV);以及适于在哺乳动物细胞中表达的病毒载体,诸如逆转录病毒,牛痘病毒,腺病毒等。

[0591] 编码融合蛋白的多核苷酸通常在适合的启动子的控制下表达,所述启动子可用于在所需宿主细胞中的表达。例如,当所述宿主是动物细胞时,使用以下任一启动子:SR α 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。在一个实施方式中,所述启动子是CMV启动子或SR α 启动子。当宿主细胞是大肠杆菌时,可以使用以下任何启动子:trp启动子,lac启动子,recA启动子, λ pL启动子,lpp启动子,T7启动子等。当宿主是芽孢杆菌属时,可以使用任何以下启动子:SP01启动子,SP02启动子,penP启动子等。当宿主是酵母时,可以使用任何以下启动子:Gal1/10启动子,PH05启动子,PGK启动子,GAP启动子,ADH启动子等。当宿主是昆虫细胞时,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。

[0592] 如果需要,所述表达载体亦包含增强子,剪接信号,终止子,polyA附加信号,选择标记物(例如抗药性基因,营养缺陷型互补基因等)或复制起点中的任何一个或多个。

[0593] 可以通过,例如通过在体外转录系统中转录mRNA,来制备编码本文所述的蛋白结构域的RNA。

[0594] 可以通过将编码融合蛋白的表达载体引入宿主细胞内并培养所述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。

[0595] 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。

[0596] 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。

[0597] 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。

[0598] 使用例如病毒载体,诸如AcNPV,在昆虫细胞中表达融合蛋白。昆虫宿主细胞包含任何以下细胞是:甘蓝菜行军虫幼虫-衍生的确立细胞是(草地贪夜蛾细胞;Sf细胞),衍生自粉纹夜蛾(*Trichoplusia ni*)中肠的MG1细胞,衍生自粉纹夜蛾的卵(巢)的“High Five(击掌相庆)”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,使用家蚕-衍生的细胞是(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。

[0599] 关于昆虫,使用家蚕、果蝇、蟋蟀等的幼虫来表达融合蛋白[Nature,315,592 (1985)]。

[0600] 哺乳动物细胞是可用于表达融合蛋白。此种细胞是包含猴COS-7细胞,猴Vero细胞,中国仓鼠卵巢(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。使用分化多能的干细胞,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。

[0601] 可以使用本领域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(例如水稻,小麦,玉米,番茄,黄瓜,茄子,康乃馨,洋桔梗,烟草,拟南芥)中制备的。

[0602] 所有上述宿主细胞可以是单倍体(单套体),或多倍体(例如,二倍体,三倍体,四倍体等)。

[0603] 使用任何转染方法(例如,使用溶菌酶,PEG,CaCl₂共沉淀,电穿孔,显微注射,粒子枪,脂转染,农杆菌等)将编码本发明的融合蛋白的表达载体引入宿主细胞。基于待转染的宿主细胞而选择转染方法。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。转导芽孢杆菌属的方法描述于例如Molecular&General Genetics,168,111(1979)。

[0604] 使用例如Methods in Enzymology,194,182-187(1991)和Proc.Natl.Acad.Sci.美国,75,1929(1978)等中所述方法转导酵母细胞。

[0605] 使用例如Bio/Technology,6,47-55(1988)等中所述方法转染昆虫细胞。

[0606] 使用例如,New Cell Engineering Experiment Protocol,263-267(1995)(Shujunsha出版)和Virology,52,456(1973)中所述方法转染哺乳动物细胞。

[0607] 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。

[0608] 例如,当培养大肠杆菌或芽孢杆菌属细胞时,使用液体培养基。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。所述培养基亦可以含有酵母提取物、维生素、生长促进因子等。所述培养基的pH优选地是在约5至约8之间。

[0609] 作为用于培养大肠杆菌的培养基,例如,使用了含有葡萄糖,酪蛋白氨基酸的M9培养基[Journal of Experiments in Molecular Genetics,431-433,Cold Spring Harbor Laboratory,纽约1972]。大肠杆菌通常在约15-约43°C下培养。必要时,可执行曝气和搅拌。

[0610] 芽孢杆菌属通常在约30到约40°C下培养。必要时,可执行曝气和搅拌。

[0611] 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。所述培养基的pH优选地是约5-约8。所述培养通常是在约20°C到约35°C下执行。必要时,可执行曝气和搅拌。

[0612] 作为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'昆虫培养基[Nature,195,788(1962)]。所述培养基的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。

[0613] 哺乳动物细胞的培养是在如下培养基中进行,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。所述培养基的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。

[0614] 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。所述培养基的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。

[0615] 可以使用诱导型启动子(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)来调节融合蛋白的表达,在适当的阶段将诱导剂添加到培养基中(或从培养基中去除)以诱导融合蛋白的表达。

[0616] 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括但不限于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。

[0617] 基于-核酸的核碱基编辑器的递送

[0618] 可以通过本领域已知的方法或如本文所述将编码本发明披露的核碱基编辑器的核酸施用于受试者或递送进入细胞(例如细菌、酵母、真菌、昆虫、植物和动物细胞)。例如,核碱基编辑器可以通过例如载体(例如病毒或非-病毒载体),基于非-载体的方法(例如使用裸DNA或DNA复合体),或其组合来递送。

[0619] 编码核碱基编辑器的核酸能以裸DNA或RNA的形式直接递送进入细胞(例如细菌、酵母、真菌、昆虫、植物和动物细胞),例如通过转染或电穿孔的手段,或偶联到可促进靶细胞摄取的分子(例如,N-乙酰半乳糖氨)上。也可以使用核酸载体,诸如以上所述载体。

[0620] 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,所述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),以及一个或多个脱氨酶。

[0621] 核酸载体亦可包含任何适合数量的调节/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。

[0622] 本发明披露的核酸载体包含重组病毒载体。示例性的病毒载体在上文中列出。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/或肽形式存在的基因组编辑系统组分。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标靶组织特异性。

[0623] 除病毒载体外,非-病毒载体还可用于递送编码本发明的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发明披露的某些实施方式中的递送载体。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。

[0624] 表6

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	
1, 2-二油酰基-sn-甘油-3-磷脂酰乙醇氨	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)酰氨基甘氨酸精氨酸	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(精氨酸酰氨基-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	
[0625] N-(2-羟基乙基)-N, N-二甲基-2, 3-双(十四烷基氧基)-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	
3 β -[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)氨-胆固醇	BGTC	阳离子	
1, 3-二脱氧-2-(6-羧基-精氨酸)-丙基酰氨	DOSPER	阳离子	
二(甲基十八烷基)溴化铵	DDAB	阳离子	
二(十八烷基)酰氨基甘氨酸亚精氨酸	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	
1, 2-二肉豆蔻酰基-三甲基铵 丙烷	DMTAP	阳离子	
0, 0'-二肉豆蔻基-N-赖氨酸天冬氨酸	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-精氨酸	CCS	阳离子	

用于基因移转的脂质			
	脂质	缩写	特性
	N-t-叔丁基-N0-十四烷基-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]咪唑鎓氯盐	DOTIM	阳离子
[0626]	N1 -胆固醇羰氧基-3, 7-二氮杂壬烷-1, 9-二氨	CDAN	阳离子
	2-(3-[双(3-氨基-丙基)-氨基]丙基氨基)-N-二(十四烷基)氨基甲酰基甲基乙基乙酰氨	RPR209120	阳离子
	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-DMA	阳离子
	二亚油基-甲基-4-二甲基氨基丁酸酯	DLin-MC3-DMA	阳离子
[0627]	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。		
[0628]	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
聚乙烯亚氨	PEI
二硫代(琥珀酸亚氨丙酸酯)	DSP
二甲基-3, 3'-二硫代双丙亚氨酸二甲酯	DTBP
聚(乙烯亚氨)双氨基甲酸酯	PEIC
聚(L-赖氨酸)	PLL
组氨酸修饰过的 PLL	
聚(N-乙烯基吡咯烷酮)	PVP
聚(丙烯亚氨)	PPI
聚(酰氨基氨)	PAMAM
聚(酰氨基乙烯亚氨)	SS-PAEI
三乙烯四氨	TETA
聚(β -氨基酯)	
聚(4-羟基-L-脯氨酸酯)	PHP
聚(烯丙基氨)	
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-羟丙基甲基丙烯酰氨)	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-精氨	D-SPM

[0629]

[0630] 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

[0631] 表8

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
[0632]	病毒的	逆转录病毒	否	稳定的	是 RNA
	慢病毒	是	稳定的	是/否(修饰 过的可为非 整合型载体)	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	减毒的细菌	是	瞬时的	否 核酸
	递送载体	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
	生物性脂质体: 红细胞血影与 外泌体	是	瞬时的	否	核酸

[0633] 在一些方面,本发明披露涉及使用例如病毒载体的融合蛋白的病毒递送。示例性的病毒载体包含逆转录病毒载体(例如Moloney小鼠白血病病毒,MML-V),腺病毒载体(例如AD100),慢病毒载体(基于HIV和FIV的载体),疱疹病毒载体(例如HSV-2)和腺相关病毒载体。

[0634] 腺相关病毒载体

[0635] AAV是一种小的单股DNA依赖病毒,其属于细小病毒家族。所述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,并且所述基因组的两侧翼都是145-bp的反向末端重复序列(ITRs)。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。

[0636] 与野生型AAV相似,重组AAV(rAAV)利用顺式-作用的145bp ITRs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。感染的后,rAAV可以表达本发明的融合蛋白,并且不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。

[0637] AAV载体的小的包装能力使得超过所述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衣壳蛋白上的核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)。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。

[0638] 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。

[0639] 在一些实施方式中,核酸酶的一部分或片段(例如脱氨酶的片段,诸如胞苷脱氨酶,腺苷脱氨酶或Cas9的片段)与内含肽融合。核酸酶可以被融合到内含肽的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蛋白并且融合到AAV衣壳蛋白。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,并且将内含肽的C-端融合至AAV衣壳蛋白的N-端。

[0640] 在一个实施方式中,通过将大的转基因表达盒断裂成两个分开的两半(5'和3'末端,或头和尾)来形成双重AAV载体,其中盒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载体共同感染同一细胞,然后进行以下操作即可实现全长转基因表达盒的重新-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。

[0641] 筛选核碱基编辑器

[0642] 可以用本文所述的各种筛选手段评估包含“分裂的”和重新组装的Cas9的核碱基编辑器融合蛋白的适宜性。将待测试的融合蛋白的每个片段与少量编码报告基因(例如,GFP)的载体一起递送到单个目标细胞中(例如,细菌、酵母、真菌、昆虫、植物或动物细胞)。

这些细胞可以在人类细胞是(诸如293T,K562或U20S)中永生化,或者,可以使用原代人细胞。此种细胞可能与最终的细胞标靶有关。

[0643] 可以使用病毒载体执行递送。在一个实施方式中,可以使用脂质转染(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,以确认一致和高水平的转染。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。

[0644] 如本文所述评估核碱基编辑器的活性,即,通过对细胞的基因组测序以检测标靶序列的改变。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(方法或试剂盒)小量制备并使用单个引物测序。测序也可以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的后,可以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)。

[0645] 可以选择在初始测试中诱导了最高水平的标靶特异性改变的融合蛋白,以供进一步评估。

[0646] 在特定的实施方式中,核碱基编辑器用于靶向目的多核苷酸。在一个实施方式中,核碱基编辑器用于靶向调节序列,包括但不限于剪接位点、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所述调节元件控制的基因的表达式的影响。

[0647] 在其他实施方式中,本发明的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。

[0648] 在其他实施方式中,本发明的核碱基编辑器用于靶向生物体(例如细菌,酵母,真菌,昆虫,植物和动物)的基因组内的目的多核苷酸。在一个实施方式中,将本发明的核碱基编辑器与向导RNA文库一起递送到细胞中,所述向导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。

[0649] 核碱基编辑器的应用

[0650] 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表达的改变。在一个实施方式中,核碱基编辑器用于修饰非-编码或调节序列,包括但不限于剪接位点、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所述调节元件控制的基因的表达式的影响。在一个特定的实施方式中,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,这可以在基因组标靶序列中不形成双股断裂的情况下完成。

[0651] 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。就突变诱发的情况而言,例如,与容易出错的PCR和其他基于聚合酶的方法相比,核碱基编辑器具有一些优势。不同于容易出错的PCR,其诱导整个多核苷酸的随机改变,本发明的核碱基编辑器可用于靶向目标蛋白质的限定区域内的特定氨基酸。

[0652] 在其他实施方式中,本发明的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,所述生物体是微生物组中的细菌(例如,拟杆菌门、疣微菌门、厚壁菌门; γ -变形菌纲、 α -变形菌纲、拟杆菌纲、梭菌纲、丹毒丝菌纲、芽孢杆菌纲;肠杆菌目、拟杆菌目、疣微菌目、梭菌目、丹毒丝菌目、乳杆菌目;肠杆菌科、拟杆菌科、丹毒丝菌科、普

雷沃氏菌科、红螯菌科和产碱菌科；埃希氏菌属、拟杆菌属、理研菌科的Alistipes(另枝菌)属、疣微菌科的Akkermansia(阿克曼)属、梭菌属、乳杆菌属)。在另一个实施方式中,所述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、棉花、油菜、水稻、烟草、苹果、葡萄、桃子、李子、樱桃)。在一个实施方式中,将本发明的核碱基编辑器与向导RNA文库一起递送到细胞中,所述向导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。

[0653] 可以在各种蛋白质的任一种中进行突变,以促进结构功能分析或改变蛋白质的内源活性。可以例如在酶(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进行突变。在一个实施方式中,核碱基编辑器靶向编码所述酶的活性位点的核酸分子,编码所述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、降低或废除酶的活性。突变对酶的影响在酶活性测定中被表征,所述测定包含本领域已知和/或对本领域技术人员显而易见的一些测定中的任何一种。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,所述测定法包含本领域已知和/或对本领域技术人员显而易见的一些测定法中的任何一种。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。备选地,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。

[0654] 本发明披露提供了用于治疗被诊断为患有与基因突变相关联的疾病或由基因突变所引起的疾病的受试者的方法,所述基因突变包含归因于点突变的基因转换、影响剪接(例如,改变剪接供体或受体位点)和反常的或错误折叠的蛋白质,所述点突变可以被本文所提供的碱基编辑器系统所更正。例如,在一些实施方式中,提供了一种方法,所述方法包括对具有此种疾病(例如,由基因转换或其他遗传性突变所引起的疾病)的受试者施用有效量的核碱基编辑器(例如,腺苷脱氨酶碱基编辑器或胞苷脱氨酶碱基编辑器,其包含一个或多个DNA结合蛋白结构域),所述核碱基编辑器直接地或间接地编辑与疾病相关联基因内的突变相关联的核苷碱基。在某些方面,提供了用于治疗其他疾病或病症的方法,例如,所述其他疾病或病症是与一种点突变相关联或由其引起的,而所述点突变可以被脱氨酶介导的基因编辑所更正或改善。本文描述了一些此类疾病,并且基于本发明披露,对于本领域技术人员而言将是显而易见的是,可以用本文所提供的策略和融合蛋白来治疗其他适合的疾病。

[0655] 药物组合物

[0656] 本发明披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。如本文所用,术语“药物组合物”是指配制用于药物用途的组合物。在一些实施方式中,所述药物组合物进一步包括药学上可接受的载体。在一些实施方式中,所述药物组合物包括另外的试剂(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)。

[0657] 如本文所用,术语“药学上可接受的载体”是指药学上可接受的材料、组合物或媒介物,诸如液体或固体填充剂,稀释剂,赋形剂,制造助剂(例如润滑剂、镁滑石、硬脂酸钙或硬脂酸锌或硬脂酸),或涉及将所述化合物从身体的一个位置(例如,递送位置)携带或运输

到另一位置(例如,人体的器官、组织或身体部位)的溶剂封装材料。药学上可接受的载体在与制剂的其他成分相容并且对受试者的组织无害(例如,生理相容性,无菌,生理性的pH等)的意义上是“可接受的”。

[0658] 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)林格氏溶液;(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。制剂中也可以存在润湿剂,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。诸如“赋形剂”,“载体”,“药学上可接受的载体”,“媒介物”的类的术语在本文中可互换使用。

[0659] 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,所述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。所述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。

[0660] 药物组合物亦可含有一种或多种渗透调节剂,即,将所述制剂的渗透特性(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。所述渗透调节剂可以是不螯合钙离子的剂。所述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。本领域技术人员可以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;以及这些试剂和/或试剂类型中的一种或多种的混合物。所述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。

[0661] 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。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、透皮、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、肌内、静脉内、血管内、骨内、眼周、肿瘤内、脑内和脑室内给药。

[0662] 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,所述植入物为多孔,无孔或凝胶状的材料,其包含膜,诸如硅橡胶或纤维。

[0663] 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,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Medical Applications of Controlled Release (Langer和Wise编,CRC出版社,佛罗里达州博卡拉顿市,1974);Controlled Drug Bioavailability,Drug Product Design and Performance (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,Science228:190;During等,1989,Ann.Neurol.25:351;Howard等,1989,J.Neurosurg.71:105。)其他的控释系统如前述在Langer的论文中被讨论。

[0664] 在一些实施方式中,根据常规程序将所述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无菌等渗溶液中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)的溶液。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封容器诸如安瓿或小药囊中。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。如果通过注射施用所述药物组合物,可以提供一安瓿瓶的无菌注射用水或生理盐水,以便可以在施用的前将各成分混合。

[0665] 用于全身给药的药物组合物可以是液体,例如无菌生理盐、乳酸林格氏液或Hank's溶液。另外,所述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。亦考虑了冻干形式。所述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,其也适合于肠胃外给药。改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“稳定的质粒-脂质微粒”(SPLP)中,所述微粒含有致融类脂质二油酰基磷脂酰乙醇氨(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(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,美国专利第4,880,635号;美国专利第4,906,477号;美国专利第4,911,928号;美国专利第4,917,951号;美国专利第4,920,016号;和美国专利第4,921,757号;其每个都通过引用并入本文。

[0666] 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。当用于本发明披露的药物组合物时,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。

[0667] 此外,可以将所述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(例如用于重构或稀释冻干的本发明的化合物的无菌的)的第二容器。可选的伴随此种容器(复数个)的可以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,所述行政命令反映了所述机构对此用于人体给药的试剂盒制造、使用或出售的批准。

[0668] 在另一方面,包含一种制品,其含有可用于治疗上述疾病的材料。在一些实施方式中,所述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。所述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,所述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,所述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。所述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所述组合物用于治疗所选的疾病。所述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。

[0669] 在一些实施方式中,本文描述的任何融合蛋白、gRNA和/或复合体作为药物组合物的一部分提供。在一些实施方式中,所述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,所述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,所述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。

[0670] 在一些实施方式中,本文所述的任何融合蛋白,gRNA,系统和/或复合体均作为药物组合物的一部分提供。在一些实施方式中,所述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,所述药物组合物包括本文所提供的任何系统或复合体。在一些实施方式中,所述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)或其片段。在一些实施方式中,所述药物组合物包括核糖核蛋白复合体,其包括多种可编程的DNA结合蛋白(例如Cas9,锌指,TALE,TALE-N蛋白或其片段)。所述可编程的DNA结合蛋白可能包括核酸酶活性、切口酶活性或无核酸酶活性。在一些实施方式中,所述药物组合物包括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。本文所述的药物组合物可任选地包括一种或更多种其他治疗活性物质。

[0671] 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,细胞获自受试者,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,可选地在细胞中已经实现或检测到所需的基因组修饰的后,将从受试者中取出并与药物组合物离体接触的细胞重新引入到受试者体内。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,例如在美国专利第6,453,242号;美国专利第6,503,717号;美国专利第6,534,261号;美国专利第6,599,692号;美国专利第6,607,882;美国专利第6,689,558号;美国专利第6,824,978号;美国专利第6,933,113号;美国专利第6,979,539号;美国专利第7,013,219号;和美国专利第7,163,824号中,其所有的公开披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,但是本领域技术人员应理解,此类组合物通常适合于对各种动物或生物体施用。为了使所述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/或执行这种修饰。预期被施用所述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。

[0672] 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。

[0673] 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的The Science and Practice of Pharmacy,第21版,A.R.Gennaro (Lippincott, Williams & Wilkins, Baltimore, MD, 2006;其全部内容通过引用并入本文) 揭示了用于配制药物流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物流合物的其他适合的方法、试剂、赋形剂和溶剂,亦请参见国际专利申请第PCT/US2010/055131号(国际专利公开第W02011053982 A8号,2010年11月2日提交),其全部内容通过引用并入本文。除非任何常规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所述物流合物的任何其他成分相互作用,否则预期其使用范围是在本发明的范围的內。在一些实施方式中,依据本发明的组合物可用于治疗各种疾病,病症和/或病症中的任何一种。

[0674] 试剂盒,载体,细胞

[0675] 本发明的各个方面提供了包括核酸构建体的试剂盒,所述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,所述核碱基编辑器融合蛋白包括脱氨酶,被“断裂”成为两个片段的SpnCas9结构域(所述N末端片段包括内含肽系统的一个成员,而所述C末端片段包括内含肽系统的另一个成员),和NLS;其中所述核苷酸序列是在驱动所述融合蛋白表达的异源性启动子的控制的下。

[0676] 本发明披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/融合蛋白的细胞(例如细菌、酵母、真菌、昆虫、植物和动物细胞)。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。

[0677] 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,其完全在本领域技术人员知识范围内。这些技法在文献中有充分的解释,诸如,“Molecular Cloning: A Laboratory Manual”,第二版(Sambrook, 1989); “Oligonucleotide Synthesis” (Gait, 1984);

[0678] “Animal Cell Culture” (Freshney, 1987); “Methods in Enzymology” “Handbook of Experimental Immunology” (Weir, 1996); “Gene Transfer Vectors for Mammalian Cells” (Miller和Calos, 1987); Current Protocols in Molecular Biology” (Ausubel, 1987); “PCR: The Polymerase Chain Reaction” (Mullis, 1994); “Current Protocols in Immunology” (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。

[0679] 提出以下实施例以向本领域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披露和描述,并非意图限制发明人所视其发明的范围。

[0680] 实施例

[0681] 实施例1:断裂和重新组装的Cas9在碱基编辑系统中保留了功能和活性

[0682] 病毒载体提供了有吸引力的递送系统,但是目前的技术递送较大的多核苷酸的能力有限。为了探索与碱基编辑系统有关的这些挑战,形成了一种融合蛋白,其从N-端到C-端包括:野生型腺苷脱氨酶TadA,融合到TadA的一进化型版本上,融合到与二分NLS融合的nCas9上。所述融合蛋白在SpCas9的非结构化区域内被断裂成为N末端片段和C末端片段,并与编码向导RNA的第三个质粒一起在两个载体中被递送。nCas9的N末端片段和C末端片段分别融合到内含肽-N and内含肽-C上。编码这些内含肽的氨基酸序列如下:

[0683] 于融合物中的Cfa-N内含肽:

[0684] CLSYDTEILTVEYGFLPIGKIVEERIECTVYTVDKNGFVYTQPIAQWHNRGEQEVF EYCLEDGSIIR
ATKDHKFMTTDGQMLPIDEIFERGLDLKQVDGLP

[0685] 于融合物中的Cfa (GEP) -C内含肽:

[0686] MVKII SRKSLGTQNVYDIGVGEPHNFLKNGLVASNC

[0687] 编码这些内含肽的多核苷酸序列如下:

[0688] 于融合物中的Cfa-N内含肽:

[0689] TGCCTGAGCTACGATACCGAGATCCTGACCGTGGAATACGGCTTCCTGCCTATCGGCAAGATCGTCGAG
GAACGGATCGAGTGCACAGTGTACACCGTGGATAAGAATGGCTTCGTGTACACCCAGCCTATCGCTCAGTGGCACAA
CAGAGGCGAGCAAGAGGTGTTTCGAGTACTGCCTGGAAGATGGCAGCATCATCCGGGCCACCAAGGACCACAAGTTTA
TGACCACCGACGGCCAGATGCTGCCCATCGACGAGATCTTTGAGAGAGGCCTGGACCTGAAACAGGTGGACGGACTG
CCT

[0690] 于融合物中的Cfa (GEP) -C内含肽:

[0691] ATGGTCAAGATCATCAGCAGAAAGAGCCTGGGCACCCAGAACGTGTACGATATCGG AGTGGGCGAGC
CCCACAACCTTCTGCTCAAGAATGGCCTGGTGGCCAGCAACTGC

[0692] 选择了spCas9的三个区域,其中所述ABE融合蛋白在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,且所述融合是在图2-4中所示的氨基酸位置处,即S303、T310、T313、S355、A456、S460、A463、T466、S469、T472、T474、C574、S577、A589和S590,其在以下序列中以粗体大写字母示出。

1 mdkkysigld igtntsvgwav itdeykvpsk kfkvlgntr hsikknliga
llfdsgetae

61 atrlkrtrr rytrrkrric ylqeifsnem akvddsffhr leesflveed
kkherhpifg

121 nivdevayhe kyptiyhlrk klvdstdkad lrliylalah
mikfrghfli egdlndnsd

[0693] 181 vdklfiqlvq tynqlfeenp inasgvdaka ilsarlsksr
rlenliaqlp gekknlgfng

241 lialslgltp nfksnfdlae daklqlskdt yddldnlla
qigdqyadlf laaknlsdai

301 llSdilrvnT eiTkaplsas mikrydehhq dltllkalvr
qqlpekykei ffdqSkngya

361 gyidggasqe efykfikpil ekmdgteell vklnredllr

	kqrtfdngsi phqihlgelh			
	421 ailrrqedfy	pflkdnreki	ekiltfripy	yvgplArgnS
	rfAwmTrkSe eTiTpwnfee			
	481 vvdkgasaqs	fiermtnfdk	nlpnekvlpk	hsllyeyftv
	yneltkvkyv tegmrkpafI			
	541 sgeqkkaivd	llfktnrkvt	vkqlkedyfk	kieCfdSvei
	sgvedrfnAS lgtyhdlIki			
	601 ikdkdfldne	enedilediv	ltltlIfedre	mieerlktya
	hlfdkvmkq lkrrrytgwg			
	661 rlsrklingi	rdkqsgktil	dfIksdgfan	rnfmqlihdd
	sltfkediql aqvsgqgdsI			
	721 hehianlags	paikkgilqt	vkvdelvkv	mgrhkpeniv
	iemarenqtt qkgqknsrer			
	781 mkrieegike	lgsqilkehp	ventqlqnek	lylyylqngr
	dmyvdqeldi nrlsdydvdh			
[0694]	841 ivpqsflkdd	sidnkvltrs	dknrgksdnv	pseevvkkmk
	nywrqlInak litqrkfdnl			
	901 tkaergglse	ldkagfikrq	lvetrqitkh	vaqildsrnm
	tkydendkli revkvitlks			
	961 klvsdfrkdf	qfykvreinn	yhhahdayIn	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 qhkhyldiei	eqisefskrv	iladanldkv	lsaynkhrdk
	pireqaenii hlftltnlga			
[0695]	1321 paafkyfdtt	idrkyrtstk	evldatlihq	sitglyetri dlsqIggd
[0696]	用质粒转染HEK 293细胞。当N末端片段和C末端片段在培养细胞中(所述两片段与			

向导RNA相结合)被表达并被剪接在一起时,使用断裂内含肽系统重构了融合蛋白,从而形成了功能性的核碱基编辑系统。所使用的向导RNA的序列如下:

[0697] 20-nt向导原间隔序列 (PAM:AGG)

[0698] 5'-TGTCGAAGTTCGCCCTGGAG-3'

[0699] 锤头型核酶(下面的序列)融合到上面的原间隔序列的5'-端:

[0700] 5'-GTCGACACTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTC-3'

[0701] 21-nt ABCA4向导原间隔序列 (PAM:AGG)

[0702] 5'-GTGTCGAAGTTCGCCCTGGAG-3'

[0703] HEK2向导原间隔序列 (PAM:GGG)

[0704] 5'-GAACACAAAGCATAGACTGC-3'

[0705] 完全的向导序列是所述原间隔序列直接附加到以下序列的5'-侧:

[0706] 5'-GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAAC TTGAAAAAGTGGCA
CCGAGTCGGTGCTTTTTTT-3'

[0707] 对于所述20-nt向导,在锤头切开的前的完全的向导转录本如下:

[0708] 20-nt向导

[0709] 5'-GTCGACACTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCTGTCGA AGTTCGCCCTGGAG
GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTAT CAACTTGAAAAAGTGGCACCGAGTCGGTG
CTTTTTT-3'

[0710] 21-nt向导

[0711] 5'-GTGTCGAAGTTCGCCCTGGAGGTTTTAGAGCTAGAAATAGCAAGTTAAAA TAAGGCTAGTCCGT
TATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTT-3'

[0712] HEK2向导

[0713] 5'-GAACACAAAGCATAGACTGCGTTTTAGAGCTAGAAATAGCAAGTTAAAAT AAGGCTAGTCCGTT
ATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTT-3'

[0714] 通过检测ABCA4基因的改变来测定碱基-编辑活性。使用高通量测序(HTS)检测到了碱基编辑后的序列变化。将重构的ABE的活性与ABE7.10的活性进行比较,后者在相同细胞类型中表达,但使用单个质粒载体进行递送。这些实验的结果示于图2至图4。

[0715] 图2显示了在Cas9的第一区域(氨基酸292-365)内将ABE断裂成为两个片段对重构的融合蛋白的碱基编辑活性的影响。令人惊讶地,所述重构的碱基编辑器在21-核苷酸、20-核苷酸和HEK2向导RNA的存在下表现出与在单个载体中所表达的对照碱基编辑器相当的活性。

[0716] 图3显示了在SpCas9的以下区域内(F445-K483和E565-T637)的氨基酸处将ABE断裂成为两个片段对重构的融合蛋白的碱基编辑活性的影响。不论ABE断裂的位置如何,当使用内含肽系统重构融合蛋白时,其显示的活性均与ABE7.10相当。当仅表达片段化融合蛋白的N末端部分或C末端部分时,未观察到碱基编辑活性。因此,编辑取决于Cas9的N末端片段和C末端片段二者都存在。如图4所示,当20-核苷酸向导RNA包含锤头型核酶时,使用断裂内含肽系统重构的融合蛋白也显示出活性。以与图3和图4中的型式不同的型式执行了图2中的实验。

[0717] 本文报道的结果是使用以下方法和材料进行的。

[0718] 将碱基编辑系统转染到培养的HEK 293T细胞中,所述细胞含有慢病毒整合的ABCA4多核苷酸,所述多核苷酸含有5882G>A突变(HEK/ABCA4/5882G>A细胞)。每 μg 转染的DNA用1.5 μL Lipofectamine2000进行转染。所述碱基编辑系统包含融合蛋白ABE7.10,其包括野生型腺苷脱氨酶TadA,融合到TadA的一进化型版本上,融合到与C末端二分NLS融合的SpCas9上。

[0719] 将ABE7.10的N末端片段和C末端片段各自克隆到pCAG质粒中,在其中它们的表达由CAG启动子驱动。作为参考,在HEK 293T细胞中表达了在CMV启动子控制下的ABE7.10。每个上述核酸序列都是密码子优化过的。用三个质粒转染细胞,第一个质粒编码融合到内含肽-N的ABE7.10的N末端片段,第二个质粒编码融合到内含肽-C的ABE7.10的C末端片段,以及一个质粒表达向导RNA。以等摩尔比(每一半为 $9.05 \times 10^{-14} \text{mol}$;863ng总的编辑器DNA)转染编码碱基编辑器的两个质粒。关于向导(RNA),在每次转染中使用127ng的向导质粒($9.05 \times 10^{-14} \text{mol}$)。

[0720] 对照:490ng ($9.05 \times 10^{-14} \text{mol}$)的pCMV-ABE7.10,其带有二分NLS(C末端)和GeneArt密码子优化+pNMG-B8(在哺乳动物细胞中不表达的非-相关质粒,其用于归一化转染DNA的数量)。

[0721] 图2至图4中所示的编码全长ABE7.10和ABE7.10“断裂”Cas9的多核苷酸和氨基酸序列如下:

[0722] 完整无损的ABE7.10:

[0723] ATGAGCGAGGTGGAATTCAGCCACGAGTACTGGATGCGGCACGCCCTGACACTGGCCAAAAGAGCTTGG
GACGAGAGGGAAGTGCCTGTGGGAGCTGTGCTGGTGCACAACAACAGAGTGATCGGCGAAGGCTGGAACAGACCCAT
CGGCAGACACGATCCTACAGCTCACGCCGAGATCATGGCCCTGAGACAAGGCGGACTGGTCATGCAGAACTACCGGC
TGATCGACGCCACACTGTACGTGACCCTGGAACCTTGCGTGATGTGTGCCGGCGCTATGATCCACAGCAGAATCGGC
AGAGTGGTGTTCGGCGCCAGAGATGCCAAAACAGGCGCTGCCGGAAGCCTGATGGATGTGCTGCATACCCCGGCAT
GAACCACAGAGTGGAATCACCGAGGGCATCCTGGCCGATGAATGTGCCGCTCTGCTGAGCGACTTCTTCCGGATGC
GGCGGCAAGAGATCAAGGCCCAGAAGAAGGCCAGTCCAGCACAGATAGCGGCGGATCTAGCGGAGGCAGCTCTGGA
TCTGAGACACCTGGCACAAGCGAGAGCGCCACACCTGAAAGTTCTGGCGGTTCTTCTGGCGGCAGCAGCGAGGTGCA
GTTCTCTCACGAATATTGGATGAGACACGCTCTCACCTGGCTAAGAGAGCCAGGGACGAAAGAGAGGTGCCAGTTG
GCGCTGTCTGGTGTGAACAATCGCGTCATCGGAGAAGGATGGAATCGCGCCATTGGCCTGCACGATCCAACCGCA
CATGCCGAAATTATGGCTCTGCGGCAAGGCGGCCTCGTGATGCAAAATTACAGACTGATCGATGCTACCCTCTACGT
CACCTTCGAGCCCTGTGTCATGTGTGCTGGGGCAATGATTCACTCCCGGATTGGCCGCGTGGTGTGTTGGAGTGCGGA
ATGCCAAGACTGGCGCCGCTGGATCTCTGATGGACGTCCTGCACTATCCTGGGATGAACCACCGGGTCGAGATCACA
GAGGGAATTCTGGCTGACGAGTGCGCTGCCCTGCTGTGCTACTTCTTTAGAATGCCAGACAGGTGTTCAACGCCCA
GAAAAAGCTCAGAGCAGCACCGATTCCGGCGGAAGCAGCGGAGGATCTTCTGGAAGCGAAACCCAGGCACCAGCG
AGTCTGCCACACCAGAATCATCTGGCGGTAGCTCCGGCGGCAGCGACAAGAAGTATTCTATCGGA

[0724] CTGGCCATCGGCACCAACTCTGTTGGATGGGCCGTGATCACCGACGAGTACAAGGTGCCAGCAAGAAA
TTCAAGGTGCTGGGCAACACCGACAGGCACAGCATCAAGAAGAACCTGATCGGCGCACTGCTGTTGCACTCTGGCGA
AACAGCCGAGGCCACCAGACTGAAGAGAACAGCCCGCAGACGGTACACCAGAAGAAAGAACCGGATCTGCTACCTCC
AAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTCCACAGACTGGAAGAGTCCTTCTGTTGGAA
GAGGACAAGAAGCACGAGAGACACCCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACCACGAGAAGTACCCAC

CATCTACCACCTGAGAAAGAACTGGTGGACAGCACCGACAAGGCCGACCTGAGACTGATCTATCTGGCCCTGGCTC
ACATGATCAAGTTCGGGGGCACTTCTGATCGAGGGCGACCTGAATCCTGACAACAGCGACGTGGACAAGCTGTTC
ATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCCATCAACGCCAGCGGAGTGGATGCCAAGGCCAT
CCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGCTGCCTGGCGAGAAGAAGAATGGCC
TGTTTCGGCAACCTGATTGCCCTGAGCCTGGGCTGACACCTAACTTCAAGAGCAACTTCGACCTGGCCGAGGACGCC
AAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAATCTGCTGGCCCAGATCGGCGATCAGTACGCCGA
CTTGTTTTCTGGCCGCAAGAATCTGAGCGACGCCATCCTGCTGTCCGACATCCTGAGAGTGAACACCGAGATCACCA
AGGCACCTCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGATCTGACCCTGCTGAAGGCCCTCGTT
AGACAGCAGCTGCCAGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCGGCTACATTGATGG
CGGAGCCAGCCAAGAGGAATTCTACAAGTTCATCAAGCCCATCCTCGAGAAGATGGACGGCACCGAGGAAGTCTGG
TCAAGCTGAACAGAGAGGACCTGCTGAGAAAGCAGAGAACCTTCGACAACGGCAGCATCCCTCACCAGATCCACCTG
GGAGAACTGCACGCCATTCTGCGGAGACAAGAGGACTTTTACCCATTCTGAAGGACAACCGGGAAAAGATCGAGAA
AATCCTGACCTTCAGGATCCCCTACTACGTGGGACCACTGGCCAGAGGCAATAGCAGATTGCGCTGGATGACCAGAA
AGAGCGAGGAAACCATCACTCCCTGGAACCTCGAGGAAGTGGTGGACAAGGGCGCCAGCGCTCAGTCCTTCATCGAG
CGGATGACCAACTTCGATAAGAACCTGCCTAACGAGAAGGTGCTGCCAAGCACAGCCTGCTGTACGAGTACTTCAC
CGTGTACAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTTCTGAGCGGCGAGCAGA
AAAAGGCCATCGTGGATCTGCTGTTCAAGACCAACCGGAAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAG
AAAATCGAGTGCTTCGACAGCGTCGAGATCTCCGGCGTGGAAGATCGGTTCAATGCCAGCCTGGGCACATACCACGA
TCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACA

[0725] ACGAAGAGAACGAGGACATCCTTGAGGACATCGTGCTGACACTGACCCTGTTTGAGGACAGAGAGATGA
TCGAGGAACGGCTGAAAACATACGCCACCTGTTTCGACGACAAAGTGATGAAGCAACTGAAGCGGCGGAGATACACC
GGCTGGGGCAGACTGTCTCGGAAGCTGATCAACGGCATCCGGGATAAGCAGTCCGGCAAGACCATCCTGGACTTTCT
GAAGTCCGACGGCTTCGCCAACAGAACTTCATGCAGCTGATTACGACGACAGCCTCACCTTCAAAGAGGATATCC
AGAAAGCCCAGGTGTCCGGCCAGGGCGATTCTCTGCATGAGCACATTGCCAACCTGGCCGGCTCTCCCGCCATTAAG
AAAGGCATCCTGCAGACAGTGAAGGTGGTGGACGAGCTTGTGAAAGTGATGGGCAGACACAAGCCCCGAGAATCGT
GATCGAAATGGCCAGAGAGAACCAGACCACAGAAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAG
AGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAACACCCAGCTGCAGAACGAGAAGCTG
TACCTGTACTACCTGCAGAATGGACGGGATATGTACGTGGACCAAGAGCTGGACATCAACAGACTGTCCGACTACGA
TGTGGACCATATCGTGCCCCAGTCTTTTCTGAAGGACGACTCCATCGACAACAAGGTCCTGACCAGATCCGACAAGA
ATCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTGGTCAAGAAGATGAAGAATACTGGCGACAGCTGCTGAAC
GCCAAGCTGATTACCCAGCGGAAGTTCGACAATCTGACCAAGGCCGAAAGAGGCGGCCTGAGCGAACTGGATAAGGC
CGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATTCTGGACTCTCGGATGA
ACACTAAGTACGACGAGAACGACAACTGATCCGCGAAGTGAAAGTCATACCCTGAAGTCCAAGCTGGTGTCCGAT
TTCCGGAAGGATTTCCAGTTCTACAAAGTGCGCGAGATCAACAATACTACCATCACGCCCACGACGCTACCTGAATGC
CGTTGTTGGAACAGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACG
ACGTGCGGAAGATGATCGCCAAGAGCGAGCAAGAGATTGGCAAGGCAACCGCCAAGTACTTCTTCTACAGCAACATC
ATGAACTTTTTCAAGACAGAGATCACCTCGCCAACGGCGAGATCAGAAAGCGGCCTCTGATCGAGACAAACGGCGA
AACCGGCGAGATTGTGTGGGATAAGGGCAGAGACTTTGCCACAGTGGGAAAGTGCTGAGCATGCCCCAAGTGAATA
TCGTGAAGAAAACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCTAAGCGGAACTCCGACAAGCTG

ATCGCCAGAAAGAAGGACTGGGACCCCAAGAAGTACGGCGGCTTCGATTCTCCTACCGTGGCCTATAGCGTGCTGGT
GGTGGCCAAAGTGGAAAAGGGCAAGTCCAAGAACTCAAGAGCGTGAAAGAGCTGCTGGGGATCACCATCATGGAAA
GAAGCAGCTTCGAGAAGAATCCGATCGATTTCTCGAGGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATC
AAGCTCCCCAAGTACTCCCTGTTTCGAGCTGGAAAACGG

[0726] CCGGAAGAGAATGCTGGCCTCTGCTGGCGAACTGCAGAAGGGAAACGAACTGGCCCTGCCTAGCAAATA
TGTGAACTTCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCAGCCCCGAGGACAATGAGCAAAAGCAGCTGT
TTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTTAGCAAGAGAGTGATTCTGGCC
GACGCCAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGACAAGCCTATCAGAGAGCAGGCCGAGAATAT
CATCCACCTGTTTACCCTGACCAACCTGGGAGCCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGC
GGTACACCTCCACCAAGAGGTGCTGGACGCCACTCTGATCCACCAGTCTATCACCGGCCTGTACGAGACACGGATC
GACCTGTCTCAACTCGGAGGCGACGAAGGCGCCGATAAGAGAACCGCCGATGGCTCTGAGTTCGAGAGCCCCAAGAA
AAAGCGCAAAGTGTGA

[0727] ABE7.10_Cfa-N_断裂_S303C:

[0728] ATGAGCGAGGTGGAATTCAGCCACGAGTACTGGATGCGGCACGCCCTGACACTGGCCAAAAGAGCTTGG
GACGAGAGGGAAGTGCTGTGGGAGCTGTGCTGGTGACAAACAACAGAGTGATCGGCGAAGGCTGGAACAGACCCAT
CGGCAGACACGATCCTACAGCTCACGCCGAGATCATGGCCCTGAGACAAGGCGGACTGGTCATGCAGAACTACCGGC
TGATCGACGCCACACTGTACGTGACCCTGGAACCTTGCGTGATGTGTGCCGGCGCTATGATCCACAGCAGAATCGGC
AGAGTGGTGTTCGGCGCCAGAGATGCCAAAACAGGCGCTGCCGGAAGCCTGATGGATGTGCTGCATCACCCCGGCAT
GAACCACAGAGTGGAATCACCGAGGGCATCCTGGCCGATGAATGTGCCGCTCTGCTGAGCGACTTCTTCCGGATGC
GGCGGCAAGAGATCAAGGCCCAGAAGAAGGCCCAGTCCAGCACAGATAGCGGCGGATCTAGCGGAGGCAGCTCTGGA
TCTGAGACACCTGGCACAAGCGAGAGCGCCACACCTGAAAGTTCTGGCGGTTCTTCTGGCGGCAGCAGCGAGGTGCA
GTTCTCTCACGAATATTGGATGAGACACGCTCTCACCTGGCTAAGAGAGCCAGGGACGAAAGAGAGGTGCCAGTTG
GCGCTGTCTGGTGTGTAACAATCGCGTCATCGGAGAAGGATGGAATCGCGCCATTGGCCTGCACGATCCAACCGCA
CATGCCGAAATTATGGCTCTGCGGCAAGGCGGCCTCGTGATGCAAAATTACAGACTGATCGATGCTACCCTCTACGT
CACCTTCGAGCCCTGTGTCATGTGTGCTGGGGCAATGATTCACTCCCGGATTGGCCGCGTGGTGTGTTGGAGTGCGGA
ATGCCAAGACTGGCGCCGCTGGATCTCTGATGGACGTCTGCACTATCCTGGGATGAACCACCGGGTCGAGATCACA
GAGGGAATTCTGGCTGACGAGTGCGCTGCCCTGCTGTGCTACTTCTTTAGAATGCCAGACAGGTGTTCAACGCCCA
GAAAAAGCTCAGAGCAGCACCGATTCCGGCGGAAGC

[0729] AGCGGAGGATCTTCTGGAAGCGAAACCCAGGCACCAGCGAGTCTGCCACACCAGAATCATCTGGCGGT
AGCTCCGGCGGCAGCGACAAGAAGTATTCTATCGGACTGGCCATCGGCACCAACTCTGTTGGATGGGCCGTGATCAC
CGACGAGTACAAGGTGCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACAGGCACAGCATCAAGAAGAACCTGA
TCGGCGCACTGCTGTTTCGACTCTGGCGAAACAGCCGAGGCCACCAGACTGAAGAGAACAGCCCGCAGACGGTACACC
AGAAGAAAGAACCGGATCTGCTACCTCCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCA
CAGACTGGAAGAGTCCTTCTGGTGGAAGAGGACAAGAAGCACGAGAGACACCCCATCTTCGGCAACATCGTGGACG
AGGTGGCCTACCACGAGAAGTACCCACCATCTACCACCTGAGAAAGAACTGGTGGACAGCACCGACAAGGCCGAC
CTGAGACTGATCTATCTGGCCCTGGCTCATATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAATCC
TGACAACAGCGACGTGGACAAGCTGTTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTTCGAGGAAAACCCCATCA
ACGCCAGCGGAGTGATGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAATCTGATCGCC
CAGCTGCCTGGCGAGAAGAAGAATGGCCTGTTTCGGCAACCTGATTGCCCTGAGCCTGGGCCTGACACCTAACTTCAA

GAGCAACTTCGACCTGGCCGAGGACGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAATCTGC
TGGCCAGATCGGCGATCAGTACGCCGACTTGTCTTCTGGCCGCCAAGAATCTGAGCGACGCCATCCTGCTGTGCCTG
AGCTACGATACCGAGATCCTGACCGTGGAATACGGCTTCCTGCCTATCGGCAAGATCGTCGAGGAACGGATCGAGTG
CACAGTGACACCGTGGAATAAGAATGGCTTCGTGTACACCCAGCCTATCGCTCAGTGGCACAACAGAGGCGAGCAAG
AGGTGTTTCGAGTACTGCCTGGAAGATGGCAGCATCATCCGGGCCACCAAGGACCACAAGTTTATGACCACCGACGGC
CAGATGCTGCCATCGACGAGATCTTTGAGAGAGGCCTGGACCTGAAACAGGTGGACGGACTGCCTTGA

[0730] ABE7.10_Cfa-N_断裂_T310C:

[0731] ATGAGCGAGGTGGAATTCAGCCACGAGTACTGGATGCGGCACGCCCTGACACTGGCCAAAAGAGCTTGG
GACGAGAGGGAAGTGCTGTGGGAGCTGTGCTGGTGACACAACAGAGTGATCGGCGAAGGCTGGAACAGACCCAT
CGGCAGACACGATCCTACAGCTCACGCCGAGATCATGGCCCTGAGACAAGGCGGACTGGTCATGCAGAACTACCGGC
TGATCGACGCCACACTGTACGTGACCTGGAACCTTGCCTGATGTGTGCCGGCGCTATGATCCACAGCAGAATCGGC
AGAGTGGTGTTCGGCGCCAGAGATGCCAAAACAGGCGCTGCCGGAAGCCTGATGGATGTGCTGCATCA

[0732] CCCCGCATGAACCACAGAGTGGAATCACCGAGGGCATCCTGGCCGATGAATGTGCCGCTCTGCTGAG
CGACTTCTTCCGGATGCGGCGCAAGAGATCAAGGCCCAGAAGAAGGCCAGTCCAGCACAGATAGCGGCGGATCTA
GCGGAGGCAGCTCTGGATCTGAGACACCTGGCACAAGCGAGAGCGCCACACCTGAAAGTTCTGGCGTTCTTCTGGC
GGCAGCAGCGAGGTGAGTTCTCTACGAATATTGGATGAGACACGCTCTCACCTGGCTAAGAGAGCCAGGGACGA
AAGAGAGGTGCCAGTTGGCGCTGTCTGGTGTTGAACAATCGCGTCATCGGAGAAGGATGGAATCGCGCCATTGGCC
TGCACGATCCAACCGCACATGCCGAAATTATGGCTCTGCGGCAAGGCGGCCTCGTGATGCAAAATTACAGACTGATC
GATGCTACCCTCTACGTACCTTCGAGCCCTGTGTATGTGTGCTGGGGCAATGATTCACTCCCGGATTGGCCGCGT
GGTGTTCGGAGTGCGGAATGCCAAGACTGGCGCCGCTGGATCTCTGATGGACGTCCTGCACTATCCTGGGATGAACC
ACCGGGTCGAGATCACAGAGGGAATTCTGGCTGACGAGTGCGCTGCCCTGCTGTGCTACTTCTTTAGAATGCCCAGA
CAGGTGTTCAACGCCCAGAAAAAGCTCAGAGCAGCACCGATTCCGGCGGAAGCAGCGAGGATCTTCTGGAAGCGA
AACCCAGGCACAGCGAGTCTGCCACACCAGAATCATCTGGCGGTAGCTCCGGCGGCAGCGACAAGAAGTATTCTA
TCGGACTGGCCATCGGCACCAACTCTGTTGGATGGGCCGTGATCACCGACGAGTACAAGGTGCCAGCAAGAAATTC
AAGGTGCTGGGCAACACCGACAGGCACAGCATCAAGAAGAACCTGATCGGCGCACTGCTGTTGACTCTGGCGAAAC
AGCCGAGGCCACAGACTGAAGAGAACAGCCCGCAGACGGTACACCAGAAGAAAGAACCGGATCTGCTACCTCCAAG
AGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCTGGTGAAGAG
GACAAGAAGCAGAGAGACCCCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACCACGAGAAGTACCCACCAT
CTACCACCTGAGAAAGAACTGGTGGACAGCACCGACAAGGCCGACCTGAGACTGATCTATCTGGCCCTGGCTCACA
TGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAATCCTGACAACAGCGACGTGGACAAGCTGTTTCATC
CAGCTGGTGCAGACCTACAACCAGCTGTTTCGAGGAAAAACCCATCAACGCCAGCGGAGTGATGCCAAGGCCATCCT
GTCTGCCAGACTGAGCAAGAGCAGACGGCTGAAAAATCTGATCGCCAGCTGCCTGGCGAGAAGAAGAATGGCCTGT
TCGGCAACCTGATTGCCCTGAGCCTGGGCCTGACACCTAACTTCAAGAGCAACTTCGACCTGGCCGAGGACGCCAAA
CTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAATCTGCTGGCCCAGATCGGCGATCAGTACGCCGACTT
GTTTCTGGCCGCCAAGAATCTGAGCGACGCCATCCTGCTGTCCGACATCCTGAGAGTGAAGTGCCTGAGCTACGATA
CCGAGATCCTGACCGTGGAATACGGCTTCCTGC

[0733] CTATCGGCAAGATCGTCGAGGAACGGATCGAGTGACAGTGTACACCGTGGATAAGAATGGCTTCGTGT
ACACCCAGCCTATCGCTCAGTGGCACAACAGAGGCGAGCAAGAGGTGTTTCGAGTACTGCCTGGAAGATGGCAGCATC
ATCCGGGCCACCAAGGACCACAAGTTTATGACCACCGACGGCCAGATGCTGCCATCGACGAGATCTTTGAGAGAGG

CCTGGACCTGAAACAGGTGGACGGACTGCCTTGA

[0734] ABE7.10_Cfa-N_断裂_T313C:

[0735] ATGAGCGAGGTGGAATTCAGCCACGAGTACTGGATGCGGCACGCCCTGACACTGGCCAAAAGAGCTTGG
GACGAGAGGGAAGTGCCTGTGGGAGCTGTGCTGGTGCACAACAACAGAGTGATCGGCGAAGGCTGGAACAGACCCAT
CGGCAGACACGATCCTACAGCTCACGCCGAGATCATGGCCCTGAGACAAGGCGGACTGGTCATGCAGAACTACCGGC
TGATCGACGCCACACTGTACGTGACCCTGGAACCTTGCGTGATGTGTGCCGGCGCTATGATCCACAGCAGAATCGGC
AGAGTGGTGTTCGGCGCCAGAGATGCCAAAACAGGCGCTGCCGGAAGCCTGATGGATGTGCTGCATACCCCCGGCAT
GAACCACAGAGTGGAATCACCAGGGGCATCCTGGCCGATGAATGTGCCGCTCTGCTGAGCGACTTCTTCCGGATGC
GGCGGCAAGAGATCAAGGCCCAGAAGAAGGCCCAGTCCAGCACAGATAGCGGCGGATCTAGCGGAGGCAGCTCTGGA
TCTGAGACACCTGGCACAAGCGAGAGCGCCACACCTGAAAGTTCTGGCGGTTCTTCTGGCGGCAGCAGCGAGGTCTGA
GTTCTCTCACGAATATTGGATGAGACACGCTCTCACCTGGCTAAGAGAGCCAGGGACGAAAGAGAGGTGCCAGTTG
GCGCTGTCTGGTGTGTAACAATCGCGTCATCGGAGAAGGATGGAATCGCGCCATTGGCCTGCACGATCCAACCGCA
CATGCCGAAATTATGGCTCTGCGGCAAGGCGGCCTCGTGATGCAAAATTACAGACTGATCGATGCTACCCTCTACGT
CACCTTCGAGCCCTGTGTCTGTGTGCTGGGGCAATGATTCACTCCCGGATTGGCCGCGTGGTGTGTTGGAGTGCGGA
ATGCCAAGACTGGCGCCGCTGGATCTCTGATGGACGTCCTGCACTATCCTGGGATGAACCACCGGGTCGAGATCACA
GAGGGAATTCTGGCTGACGAGTGCGCTGCCCTGCTGTGCTACTTCTTTAGAATGCCAGACAGGTGTTCAACGCCCA
GAAAAAGCTCAGAGCAGCACCGATTCCGGCGGAAGCAGCGGAGGATCTTCTGGAAGCGAAACCCAGGCACCGCG
AGTCTGCCACACCAGAATCATCTGGCGGTAGCTCCGGCGGCAGCGACAAGAAGTATTCTATCGGACTGGCCATCGGC
ACCAACTCTGTTGGATGGGCCGTGATCACCAGCAGGTACAAGGTGCCAGCAAGAAATTCAAGGTGCTGGGCAACAC
CGACAGGCACAGCATCAAGAAGAACCTGATCGGCGCACTGCTGTTGACTCTGGCGAAACAGCCGAGGCCACCAGAC
TGAAGAGAACAGCCCGCAGACGGTACACCAGAAGAAAGAACCGGAT

[0736] CTGCTACCTCCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGA
AGAGTCCTTCTGGTGGAAAGAGGACAAGAAGCACGAGAGACACCCCATCTTCGGCAACATCGTGGACGAGGTGGCCT
ACCACGAGAAGTACCCACCATCTACCACCTGAGAAAGAACTGGTGGACAGCACCGACAAGGCCGACCTGAGACTG
ATCTATCTGGCCCTGGCTCACATGATCAAGTTCCGGGGGCACTTCTGATCGAGGGCGACCTGAATCCTGACAACAG
CGACGTGGACAAGCTGTTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTTCGAGGAAAACCCCATCAACGCCAGCG
GAGTGGATGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAATCTGATCGCCAGCTGCCT
GGCGAGAAGAAGAATGGCCTGTTTCGGCAACCTGATTGCCCTGAGCCTGGGCCTGACACCTAATTCAAGAGCAACTT
CGACCTGGCCGAGGACGCCAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAATCTGCTGGCCCAGA
TCGGCGATCAGTACGCCGACTTGTTCGCGGCCAAGAATCTGAGCGACGCCATCCTGCTGTCCGACATCCTGAGA
GTGAACACCGAGATCTGCCTGAGCTACGATACCGAGATCCTGACCGTGGAATACGGCTTCCTGCCTATCGGCAAGAT
CGTCGAGGAACGGATCGAGTGACAGTGTACACCGTGGATAAGAATGGCTTCGTGTACACCCAGCCTATCGCTCAGT
GGCACAACAGAGGCGAGCAAGAGGTGTTTCGAGTACTGCCTGGAAGATGGCAGCATCATCCGGGCCACCAAGGACCAC
AAGTTTATGACCACCGACGCCAGATGCTGCCCATCGACGAGATCTTTGAGAGAGGCCTGGACCTGAAACAGGTGGA
CGGACTGCCTTGA

[0737] ABE7.10_Cfa-N_断裂_S355C:

[0738] ATGAGCGAGGTGGAATTCAGCCACGAGTACTGGATGCGGCACGCCCTGACACTGGCCAAAAGAGCTTGG
GACGAGAGGGAAGTGCCTGTGGGAGCTGTGCTGGTGCACAACAACAGAGTGATCGGCGAAGGCTGGAACAGACCCAT
CGGCAGACACGATCCTACAGCTCACGCCGAGATCATGGCCCTGAGACAAGGCGGACTGGTCATGCAGAACTACCGGC

TGATCGACGCCACACTGTACGTGACCCTGGAACCTTGCGTGATGTGTGCCGGCGCTATGATCCACAGCAGAATCGGC
AGAGTGGTGTTCGGCGCCAGAGATGCCAAAACAGGCGCTGCCGGAAGCCTGATGGATGTGCTGCATCACCCCGGCAT
GAACCACAGAGTGGAATCACCAGGGGCATCCTGGCCGATGAATGTGCCGCTCTGCTGAGCGACTTCTTCCGGATGC
GGCGGCAAGAGATCAAGGCCAGAAGAAGGCCAGTCCAGCACAGATAGCGGCGGATCTAGCGGAGGCAGCTCTGGA
TCTGAGACACCTGGCACAAGCGAGAGCGCCACACCTGAAAGTTCTGGCGGTTCTTCTGGCGGCAGCAGCGAGGTCTGA
GTTCTCTCACGAATATTGGATGA

[0739] GACACGCTCTACCCTGGCTAAGAGAGCCAGGGACGAAAGAGAGGTGCCAGTTGGCGCTGTCCTGGTGT
TGAACAATCGCGTCATCGGAGAAGGATGGAATCGCGCCATTGGCCTGCACGATCCAACCGCACATGCCGAAATTATG
GCTCTGCGGCAAGGGCGCTCGTGATGCAAAATTACAGACTGATCGATGTACCTCTACGTCACCTTCGAGCCCTG
TGTCATGTGTGCTGGGGCAATGATTCACTCCCGGATTGGCCGCGTGGTGTGGAGTGCGGAATGCCAAGACTGGCG
CCGCTGGATCTCTGATGGACGTCCTGCACTATCCTGGGATGAACCACCGGGTCGAGATCACAGAGGGAATTCTGGCT
GACGAGTGCCTGCCCTGCTGTGCTACTTCTTTAGAATGCCCAGACAGGTGTTCAACGCCCAGAAAAAGCTCAGAG
CAGCACCGATTCCGGCGGAAGCAGCGGAGGATCTTCTGGAAGCGAAACCCAGGCACCAGCGAGTCTGCCACACCAG
AATCATCTGGCGGTAGCTCCGGCGGCAGCGACAAGAAGTATTCTATCGGACTGGCCATCGGCACCAACTCTGTTGGA
TGGGCCGTGATCACCGACGAGTACAAGGTGCCCAGCAAGAAATCAAGGTGCTGGGCAACACCGACAGGCACAGCAT
CAAGAAGAACCTGATCGGCGCACTGCTGTTGCACTCTGGCGAAACAGCCGAGGCCACCAGACTGAAGAGAACAGCCC
GCAGACGGTACACCAGAAGAAAGAACCGGATCTGCTACCTCCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGAC
GACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGAAAGAGGACAAGAAGCACGAGAGACACCCATCTTCGG
CAACATCGTGGACGAGGTGGCCTACCACGAGAAGTACCCACCATCTACCACCTGAGAAAGAACTGGTGGACAGCA
CCGACAAGGCCGACCTGAGACTGATCTATCTGGCCCTGGCTCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAG
GGCGACCTGAATCCTGACAACAGCGACGTGGACAAGCTGTTATCCAGCTGGTGCAGACCTACAACCAGCTGTTCTGA
GGAAAACCCCATCAACGCCAGCGGAGTGGATGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGG
AAAATCTGATCGCCCAGCTGCCTGGCGAGAAGAAGATGGCCTGTTCCGGCAACCTGATTGCCCTGAGCCTGGGCCTG
ACACCTAACTTCAAGAGCAACTTCGACCTGGCCGAGGACGCCAACTGCAGCTGAGCAAGGACACCTACGACGACGA
CCTGGACAATCTGCTGGCCAGATCGGCGATCAGTACGCCGACTTGTCTTCTGGCCGCAAGAATCTGAGCGACGCCA
TCCTGCTGTCCGACATCCTGAGAGTGAACACCGAGATCACCAAGGCACCTCTGAGCGCCTCTATGATCAAGAGATAC
GACGAGCACCACCAGGATCTGACCCTGCTGAAGGCCCTCGTTAGACAGCAGCTGCCAGAGAAGTACAAAGAGATTTT
CTTCGACCAGTGCCTGAGCTACGATACCGAGATCCTGACCGTGGAATACGGCTTCCTGCCTATCGGCAAGATCGTCG
AGGAACGGATCGAGTGCACAGTGTACACCGTGGATAAGAATGGCTTCGTGTACACCCAGCCTATCGCTCAGTGGCAC
AACAGAGGCGAGCAAGAGGTGTTGAGTACTGCCTG

[0740] GAAGATGGCAGCATCATCCGGGCCACCAAGGACCACAAGTTTATGACCACCGACGGCCAGATGCTGCCC
ATCGACGAGATCTTTGAGAGAGGCCTGGACCTGAAACAGGTGGACGGACTGCCTTGA

[0741] ABE7.10_Cfa-N_断裂_A456C:

[0742] ATGAGCGAGGTGGAATTCAGCCACGAGTACTGGATGCGGCACGCCCTGACACTGGCCAAAAGAGCTTGG
GACGAGAGGGAAGTGCCTGTGGGAGCTGTGCTGGTGCACAACAACAGAGTGATCGGCGAAGGCTGGAACAGACCCAT
CGGCAGACACGATCCTACAGCTCACGCCGAGATCATGGCCCTGAGACAAGGCGGACTGGTCATGCAGAACTACCGGC
TGATCGACGCCACACTGTACGTGACCCTGGAACCTTGCGTGATGTGTGCCGGCGCTATGATCCACAGCAGAATCGGC
AGAGTGGTGTTCGGCGCCAGAGATGCCAAAACAGGCGCTGCCGGAAGCCTGATGGATGTGCTGCATCACCCCGGCAT
GAACCACAGAGTGGAATCACCAGGGGCATCCTGGCCGATGAATGTGCCGCTCTGCTGAGCGACTTCTTCCGGATGC

GGCGGCAAGAGATCAAGGCCAGAAGAAGGCCAGTCCAGCACAGATAGCGGCGGATCTAGCGGAGGCAGCTCTGGA
TCTGAGACACCTGGCACAAGCGAGAGCGCCACACCTGAAAGTTCTGGCGGTTCTTCTGGCGGCAGCAGCGAGGTCTGA
GTTCTCTCACGAATATTGGATGAGACACGCTCTCACCTGGCTAAGAGAGCCAGGGACGAAAGAGAGGTGCCAGTTG
GCGCTGTCTGGTGTGAACAATCGCGTCATCGGAGAAGGATGGAATCGCGCCATTGGCCTGCACGATCCAACCGCA
CATGCCGAAATTATGGCTCTGCGGCAAGGCGGCCTCGTGATGCAAAATTACAGACTGATCGATGCTACCCTCTACGT
CACCTTCGAGCCCTGTGTCATGTGTGCTGGGGCAATGATTCACTCCCGGATTGGCCGCGTGGTGTGTTGGAGTGCGGA
ATGCCAAGACTGGCGCCGCTGGATCTCTGATGGACGTCCTGCACTATCCTGGGATGAACCACCGGGTCGAGATCACA
GAGGGAATTCTGGCTGACGAGTGCGCTGCCCTGCTGTGCTACTTCTTTAGAATGCCAGACAGGTGTTCAACGCCCA
GAAAAAGCTCAGAGCAGCACCGATTCCGGCGGAAGCAGCGGAGGATCTTCTGGAAGCGAAACCCAGGCACCGCG
AGTCTGCCACACCAGAATCATCTGGCGGTAGCTCCGGCGGCAGCGACAAGAAGTATTCTATCGGACTGGCCATCGGC
ACCAACTCTGTTGGATGGGCGGTGATCACCGACGAGTACAAGGTGCCAGCAAGAAATTCAAGGTGCTGGGCAACAC
CGACAGGCACAGCATCAAGAAGAACCTGATCGGCGCACTGCTGTTGACTCTGGCGAAACAGCCGAGGCCACCAGAC
TGAAGAGAACAGCCCGCAGACGGTACACCAGAAGAAAGAACCGGATCTGCTACCTCCAAGAGATCTTCAGCAACGAG
ATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGAAAGAGGACAAGAAGCACGAG

[0743] AGACACCCCATCTTCGGCAACATCGTGACGAGGTGGCCTACCACGAGAAGTACCCACCATCTACCAC
CTGAGAAAGAACTGGTGGACAGCACCGACAAGGCCGACCTGAGACTGATCTATCTGGCCCTGGCTCACATGATCAA
GTTCCGGGGCCACTTCTGATCGAGGGCGACCTGAATCCTGACAACAGCGACGTGGACAAGCTGTTTCATCCAGCTGG
TGCAGACCTACAACCAGCTGTTTCGAGGAAAACCCCATCAACGCCAGCGGAGTGGATGCCAAGGCCATCCTGTCTGCC
AGACTGAGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGCTGCCTGGCGAGAAGAAGAATGGCCTGTTTCGGCAA
CCTGATTGCCCTGAGCCTGGGCTGACACCTAACTTCAAGAGCAACTTCGACCTGGCCGAGGACGCCAACTGCAGC
TGAGCAAGGACACCTACGACGACGACCTGGACAATCTGCTGGCCCAGATCGGCGATCAGTACGCCGACTTGTTTCTG
GCCGCCAAGAATCTGAGCGACGCCATCCTGCTGTCCGACATCCTGAGAGTGAACACCGAGATCACCAAGGCACCTCT
GAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGATCTGACCCTGCTGAAGGCCCTCGTTAGACAGCAGC
TGCCAGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCGGCTACATTGATGGCGGAGCCAGC
CAAGAGGAATTCTACAAGTTCATCAAGCCATCCTCGAGAAGATGGACGGCACCGAGGAAGTCTGGTCAAGCTGAA
CAGAGAGGACCTGCTGAGAAAGCAGAGAACCCTTCGACAACGGCAGCATCCCTCACCAGATCCACCTGGGAGAACTGC
ACGCCATTCTGCGGAGACAAGAGGACTTTTACCCATTCTGAAGGACAACCGGAAAAGATCGAGAAAATCCTGACC
TTCAGGATCCCCTACTACGTGGGACCACTGTGCCTGAGCTACGATACCGAGATCCTGACCGTGAATACGGCTTCT
GCCTATCGGCAAGATCGTCGAGGAACGGATCGAGTGACAGTGTACACCGTGGATAAGAATGGCTTCGTGTACACCC
AGCCTATCGCTCAGTGGCACAACAGAGGCGAGCAAGAGGTGTTTCGAGTACTGCCTGGAAGATGGCAGCATCATCCGG
GCCACCAAGGACCACAAGTTTATGACCACCGACGGCCAGATGCTGCCCATCGACGAGATCTTTGAGAGAGGCCTGGA
CCTGAAACAGGTGGACGGACTGCCTTGA

[0744] ABE7.10_Cfa-N_断裂_S460C:

[0745] ATGAGCGAGGTGGAATTCAGCCACGAGTACTGGATGCGGCACGCCCTGACACTGGCCAAAAGAGCTTGG
GACGAGAGGGAAGTGCCTGTGGGAGCTGTGCTGGTGCACAACAACAGAGTGATCGGCGAAGGCTGGAACAGACCCAT
CGGCAGACACGATCCTACAGCTCACGCCGAGATCATGGCCCTGAGACAAGGCGGACTGGTCATGCAGAACTACCGGC
TGATCGACGCCACACTGTACGTGACCCTGGAACCTTGCGT

[0746] GATGTGTGCCGGCGCTATGATCCACAGCAGAATCGGCAGAGTGGTGTTCGGCGCCAGAGATGCCAAAAC
AGGCGCTGCCGGAAGCCTGATGGATGTGCTGCATACCCCGGCATGAACCACAGAGTGGAATCACCGAGGGCATCC

TGGCCGATGAATGTGCCGCTCTGCTGAGCGACTTCTTCCGGATGCGGCGGCAAGAGATCAAGGCCAGAAGAAGGCC
CAGTCCAGCACAGATAGCGGCGGATCTAGCGGAGGCAGCTCTGGATCTGAGACACCTGGCACAAGCGAGAGCGCCAC
ACCTGAAAGTTCTGGCGGTTCTTCTGGCGGCAGCAGCGAGGTCGAGTTCTCTCACGAATATTGGATGAGACACGCTC
TCACCCTGGCTAAGAGAGCCAGGGACGAAAGAGAGGTGCCAGTTGGCGCTGTCCTGGTGTGAACAATCGCGTCATC
GGAGAAGGATGGAATCGCGCCATTGGCCTGCACGATCCAACCGCACATGCCGAAATTATGGCTCTGCGGCAAGGCGG
CCTCGTGATGCAAAATTACAGACTGATCGATGCTACCCTCTACGTCACCTTCGAGCCCTGTGTATGTGTGCTGGGG
CAATGATTCACTCCCGGATTGGCCGCGTGGTGTGGAGTGCGGAATGCCAAGACTGGCGCCGCTGGATCTCTGATG
GACGTCTGCACTATCCTGGGATGAACCACCGGGTCGAGATCACAGAGGGAATTCTGGCTGACGAGTGCGCTGCCCT
GCTGTGCTACTTCTTTAGAATGCCAGACAGGTGTTCAACGCCAGAAAAAGCTCAGAGCAGCACCATTCCGGCG
GAAGCAGCGGAGGATCTTCTGGAAGCGAAACCCAGGCACCAGCGAGTCTGCCACACCAGAATCATCTGGCGGTAGC
TCCGGCGGCAGCGACAAGAAGTATTCTATCGGACTGGCCATCGGCACCAACTCTGTTGGATGGGCCGTGATACCCGA
CGAGTACAAGGTGCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACAGGCACAGCATCAAGAAGAACCTGATCG
GCGCACTGCTGTTGCACTCTGGCGAAACAGCCGAGGCCACCAGACTGAAGAGAACAGCCCGCAGACGGTACACCAGA
AGAAAGAACCGGATCTGCTACCTCCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAG
ACTGGAAGAGTCCTTCTGCTGGAAGAGGACAAGAAGCACGAGAGACACCCCATCTTCGGCAACATCGTGACGAGG
TGGCCTACCACGAGAAGTACCCACCATCTACCACCTGAGAAAGAACTGGTGGACAGCACCGACAAGGCCGACCTG
AGACTGATCTATCTGGCCCTGGCTCACATGATCAAGTTCGGGGGCCACTTCTGATCGAGGGCGACCTGAATCCTGA
CAACAGCGACGTGGACAAGCTGTTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTTCGAGGAAAACCCCATCAACG
CCAGCGGAGTGATGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAATCTGATCGCCAG
CTGCCTGGCGAGAAGAAGAATGGCCTGTTTCGGCAACCTGATTGCCCTGAGCCTGGGCCTGACACCTAATTCAAGAG
CAACTTCGACCTGGCCGAGGACGCCAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAATCTGCTGG
CCCAGATCGGCGATCAGTACGCCGACTTGTTT

[0747] CTGGCCGCCAAGAATCTGAGCGACGCCATCCTGCTGTCCGACATCCTGAGAGTGAACACCGAGATCACC
AAGGCACCTCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGATCTGACCCTGCTGAAGGCCCTCGT
TAGACAGCAGCTGCCAGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCGGCTACATTGATG
GCGGAGCCAGCCAAGAGGAATTCTACAAGTTCATCAAGCCCATCCTCGAGAAGATGGACGGCACCAGGAAGTCTG
GTCAAGCTGAACAGAGAGGACCTGCTGAGAAAGCAGAGAACCTTCGACAACGGCAGCATCCCTCACCAGATCCACCT
GGGAGAAGTGCACGCCATTCTGCGGAGACAAGAGGACTTTTACCCATTCTGAAGGACAACCGGGAAAAGATCGAGA
AAATCCTGACCTTCAGGATCCCTACTACGTGGGACCACTGGCCAGAGGCAATTGCCTGAGCTACGATACCGAGATC
CTGACCGTGGAATACGGCTTCCTGCCTATCGGCAAGATCGTCGAGGAACGGATCGAGTGCACAGTGTACACCGTGGA
TAAGAATGGCTTCGTGTACACCCAGCCTATCGCTCAGTGGCACAACAGAGGCGAGCAAGAGGTGTTTCGAGTACTGCC
TGGAAGATGGCAGCATCATCCGGGCCACCAAGGACCACAAGTTTATGACCACCGACGGCCAGATGCTGCCCATCGAC
GAGATCTTTGAGAGAGGCCTGGACCTGAAACAGGTGGACGGACTGCCTTGA

[0748] ABE7.10_Cfa-N_断裂_A463C:

[0749] ATGAGCGAGGTGGAATTCAGCCACGAGTACTGGATGCGGCACGCCCTGACACTGGCCAAAAGAGCTTGG
GACGAGAGGGAAGTGCTGTGGGAGCTGTGCTGGTGCACAACAACAGAGTGATCGGCGAAGGCTGGAACAGACCCAT
CGGCAGACACGATCCTACAGCTCACGCCGAGATCATGGCCCTGAGACAAGGCGGACTGGTCATGCAGAACTACCGGC
TGATCGACGCCACACTGTACGTGACCCTGGAACCTTGCGTGATGTGTGCCGGCGCTATGATCCACAGCAGAATCGGC
AGAGTGGTGTTCGGCGCCAGAGATGCCAAAACAGGCGCTGCCGGAAGCCTGATGGATGTGCTGCATACCCCGGCAT

GAACCACAGAGTGGAAATCACCGAGGGCATCCTGGCCGATGAATGTGCCGCTCTGCTGAGCGACTTCTTCCGGATGC
GGCGGCAAGAGATCAAGGCCAGAAGAAGGCCAGTCCAGCACAGATAGCGGCGGATCTAGCGGAGGCAGCTCTGGA
TCTGAGACACCTGGCACAAGCGAGAGCGCCACACCTGAAAGTTCTGGCGGTTCTTCTGGCGGCAGCAGCGAGGTCTGA
GTTCTCTCACGAATATTGGATGAGACACGCTCTCACCTGGCTAAGAGAGCCAGGGACGAAAGAGAGGTGCCAGTTG
GCGCTGTCTGGTGTGAACAATCGCGTCATCGGAGAAGGATGGAATCGCGCCATTGGCCTGCACGATCCAACCGCA
CATGCCGAAATTATGGCTCTGCGGCAAGGCGGCCTCGTGATGCAAAATTACAGACTGATCGATGCTACCCTCTACGT
CAC

[0750] CTTGAGCCCTGTGTCTGTGTGCTGGGGCAATGATTCACTCCCGGATTGGCCGCGTGGTGTGTTGGAGT
GCGGAATGCCAAGACTGGCGCCGCTGGATCTCTGATGGACGTCCTGCACTATCCTGGGATGAACCACCGGGTCGAGA
TCACAGAGGGAATTCTGGCTGACGAGTGCCTGCCCTGCTGTGCTACTTCTTTAGAATGCCAGACAGGTGTTCAAC
GCCCAGAAAAAGCTCAGAGCAGCACCGATTCCGGCGGAAGCAGCGGAGGATCTTCTGGAAGCGAAACCCAGGCAC
CAGCGAGTCTGCCACACCAGAATCATCTGGCGGTAGCTCCGGCGGCAGCGACAAGAAGTATTCTATCGGACTGGCCA
TCGGCACCAACTCTGTTGGATGGGCCGTGATCACCGACGAGTACAAGGTGCCAGCAAGAAATCAAGGTGCTGGGC
AACACCGACAGGCACAGCATCAAGAAGAACCTGATCGGCGCACTGCTGTTGACTCTGGCGAAACAGCCGAGGCCAC
CAGACTGAAGAGAACAGCCCGCAGACGGTACACCAGAAGAAAGAACCGGATCTGCTACCTCCAAGAGATCTTCAGCA
ACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGAAGAGGACAAGAAGCAC
GAGAGACACCCCATCTTCGGCAACATCGTGACGAGGTGGCCTACCACGAGAAGTACCCACCATCTACCACCTGAG
AAAGAACTGGTGGACAGCACCGACAAGGCCGACCTGAGACTGATCTATCTGGCCCTGGCTCACATGATCAAGTTCC
GGGGCCACTTCCTGATCGAGGGCGACCTGAATCCTGACAACAGCGACGTGGACAAGCTGTTTCATCCAGCTGGTGCAG
ACCTACAACCAGCTGTTTCGAGGAAAACCCCATCAACGCCAGCGGAGTGGATGCCAAGGCCATCCTGTCTGCCAGACT
GAGCAAGAGCAGACGGCTGAAAAATCTGATCGCCAGCTGCCTGGCGAGAAGAAGAATGGCCTGTTTCGGCAACCTGA
TTGCCCTGAGCCTGGGCCTGACACCTAACTTCAAGAGCAACTTCGACCTGGCCGAGGACGCCAACTGCAGCTGAGC
AAGGACACCTACGACGACGACCTGGACAATCTGCTGGCCCAGATCGGCGATCAGTACGCCGACTTGTCTTCTGGCCGC
CAAGAATCTGAGCGACGCCATCCTGCTGTCCGACATCCTGAGAGTGAACACCGAGATCACCAAGGCACCTCTGAGCG
CCTCTATGATCAAGAGATACGACGAGCACCACCAGGATCTGACCCTGCTGAAGGCCCTCGTTAGACAGCAGCTGCCA
GAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCGGCTACATTGATGGCGGAGCCAGCCAAGA
GGAATTCTACAAGTTCATCAAGCCCATCCTCGAGAAGATGGACGGCACCGAGGAAGTCTGGTCAAGCTGAACAGAG
AGGACCTGCTGAGAAAGCAGAGAACCTTCGACAACGGCAGCATCCCTCACCAGATCCACCTGGGAGAACTGCACGCC
ATTCTGCGGAGACAAGAGGACTTTTACCCATTCTGAAGGACAACCGGGAAAAGATCGAGAAAATCCTGACCTTCAG
GATCCCTACTACGTGGGACCACTGGCCAGAGGCAATAGCAGATTCTGCCTGAGCTACGATACCGAGATCCTGACCG
TGGAATACGGCTTCCTGCCTATCGGCAAGATCGTCG

[0751] AGGAACGGATCGAGTGCACAGTGACACCGTGGATAAGAATGGCTTCGTGTACACCCAGCCTATCGCTC
AGTGGCACAACAGAGGCGAGCAAGAGGTGTTTCGAGTACTGCCTGGAAGATGGCAGCATCATCCGGGCCACCAAGGAC
CACAAGTTTATGACCACCGACGGCCAGATGCTGCCCATCGACGAGATCTTTGAGAGAGGCCTGGACCTGAAACAGGT
GGACGGACTGCCTTGA

[0752] ABE7.10_Cfa-N_断裂_T466C:

[0753] ATGAGCGAGGTGGAATTCAGCCACGAGTACTGGATGCGGCACGCCCTGACACTGGCCAAAAGAGCTTGG
GACGAGAGGGAAGTGCCTGTGGGAGCTGTGCTGGTGACAAACAACAGAGTGATCGGCGAAGGCTGGAACAGACCCAT
CGGCAGACACGATCCTACAGCTCACGCCGAGATCATGGCCCTGAGACAAGGCGGACTGGTCATGCAGAACTACCGGC

TGATCGACGCCACACTGTACGTGACCCTGGAACCTTGCGTGATGTGTGCCGGCGCTATGATCCACAGCAGAATCGGC
AGAGTGGTGTTCGGCGCCAGAGATGCCAAAACAGGCGCTGCCGGAAGCCTGATGGATGTGCTGCATCACCCCGGCAT
GAACCACAGAGTGGAATCACCGAGGGCATCCTGGCCGATGAATGTGCCGCTCTGCTGAGCGACTTCTTCCGGATGC
GGCGGCAAGAGATCAAGGCCAGAAGAAGGCCAGTCCAGCACAGATAGCGGCGGATCTAGCGGAGGCAGCTCTGGA
TCTGAGACACCTGGCACAAGCGAGAGCGCCACACCTGAAAGTTCTGGCGGTTCTTCTGGCGGCAGCAGCGAGGTGCA
GTTCTCTCACGAATATTGGATGAGACACGCTCTCACCTGGCTAAGAGAGCCAGGGACGAAAGAGAGGTGCCAGTTG
GCGCTGTCTGGTGTGAACAATCGCGTCATCGGAGAAGGATGGAATCGCGCCATTGGCCTGCACGATCCAACCGCA
CATGCCGAAATTATGGCTCTGCGGCAAGGCGGCCTCGTGATGCAAAATTACAGACTGATCGATGCTACCCTCTACGT
CACCTTCGAGCCCTGTGTCATGTGTGCTGGGGCAATGATTCACTCCCGGATTGGCCGCGTGGTGTGGAGTGCGGA
ATGCCAAGACTGGCGCCGCTGGATCTCTGATGGACGTCCTGCACTATCCTGGGATGAACCACCGGGTCGAGATCACA
GAGGGAATTCTGGCTGACGAGTGCGCTGCCCTGCTGTGCTACTTCTTTAGAATGCCAGACAGGTGTTCAACGCCCA
GAAAAAGCTCAGAGCAGCACCGATTCCGGCGGAAGCAGCGGAGGATCTTCTGGAAGCGAAACCCAGGCACCAGCG
AGTCTGCCACACCAGAATCATCTGGCGGTAGCTCCGGCGGCAGCGACAAGAAGTATTCTATCGGACTGGCCATCGGC
ACCAACTCTGTTGGATGGGCCGTGATCACCGACGAGTACAAGGTGCCAGCAAGAAATTCAAGGTGCTGGGCAACAC
CGACAGGCACAGCATCAAGAAGAACCTGATCGGCGCACTGCTGTTGACTCTGGCGAAACAGCCGAGGCCACCAGAC
TGAAGAGAACAGCCCGCAGACGGTACACCAGAAGAAAGAACCGGAT

[0754] CTGCTACCTCCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGA
AGAGTCCTTCTGGTGAAGAGGACAAGAAGCACGAGAGACACCCCATCTTCGGCAACATCGTGGACGAGGTGGCCT
ACCACGAGAAGTACCCACCATCTACCACCTGAGAAAAGAACTGGTGGACAGCACCGACAAGGCCGACCTGAGACTG
ATCTATCTGGCCCTGGCTCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAATCCTGACAACAG
CGACGTGGACAAGCTGTTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTTCGAGGAAAACCCCATCAACGCCAGCG
GAGTGGATGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGAAAATCTGATCGCCAGCTGCCT
GGCGAGAAGAAGTGGCCTGTTTCGGCAACCTGATTGCCCTGAGCCTGGGCCTGACACCTAATTCAAGAGCAACTT
CGACCTGGCCGAGGACGCCAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAATCTGCTGGCCCAGA
TCGGCGATCAGTACGCCGACTTGTCTGCGGCCAAGAATCTGAGCGACGCCATCCTGCTGTCCGACATCCTGAGA
GTGAACACCGAGATCACCAAGGCACCTCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGATCTGAC
CCTGCTGAAGGCCCTCGTTAGACAGCAGCTGCCAGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCT
ACGCCGGCTACATTGATGGCGGAGCCAGCCAAGAGGAATTCTACAAGTTCATCAAGCCCATCCTCGAGAAGATGGAC
GGCACCGAGGAAGTCTGGTCAAGCTGAACAGAGAGGACCTGCTGAGAAAGCAGAGAACCTTCGACAACGGCAGCAT
CCCTCACCAGATCCACCTGGGAGAACTGCACGCCATTCTGCGGAGACAAGAGGACTTTTACCCATTCTGAAGGACA
ACCGGGAAAAGATCGAGAAAATCCTGACCTTCAGGATCCCCTACTACGTGGGACCACTGGCCAGAGGCAATAGCAGA
TTCGCTGGATGTGCTGAGCTACGATACCGAGATCCTGACCGTGAATACGGCTTCCTGCCTATCGGCAAGATCGT
CGAGGAACGGATCGAGTGCACAGTGTACACCGTGGATAAGAATGGCTTCGTGTACACCCAGCCTATCGCTCAGTGGC
ACAACAGAGGCGAGCAAGAGGTGTTTCGAGTACTGCCTGGAAGATGGCAGCATCATCCGGGCCACCAAGGACCACAAG
TTTATGACCACCGACGGCCAGATGCTGCCCATCGACGAGATCTTTGAGAGAGGCCTGGACCTGAAACAGGTGGACGG
ACTGCCTTGA

[0755] ABE7.10_Cfa-N_断裂_S469C:

[0756] ATGAGCGAGGTGGAATTCAGCCACGAGTACTGGATGCGGCACGCCCTGACACTGGCCAAAAGAGCTTGG
GACGAGAGGGAAGTGCTGTGGGAGCTGTGCTGGTGCACAACAACAGAGTGATCGGCGAAGGCTGGAACAGACCCAT

CGGCAGACACGATCCTACAGCTACGCCGAGATCATGGCCCTGAGACAAGGCGGACTGGTCATG

[0757] CAGAACTACCGGCTGATCGACGCCACACTGTACGTGACCCTGGAACCTTGCGTGATGTGTGCCGGCGCT
ATGATCCACAGCAGAATCGGCAGAGTGGTGTTCGGCGCCAGAGATGCCAAAACAGGCGCTGCCGGAAGCCTGATGGA
TGTGCTGCATACCCCGGCATGAACCACAGAGTGGAAATCACCGAGGGCATCCTGGCCGATGAATGTGCCGCTCTGC
TGAGCGACTTCTTCCGGATGCGGCGCAAGAGATCAAGGCCAGAAGAAGGCCAGTCCAGCACAGATAGCGGCGGA
TCTAGCGGAGGCAGCTCTGGATCTGAGACACCTGGCACAAGCGAGAGCGCCACACCTGAAAGTTCTGGCGGTTCTTC
TGGCGGCAGCAGCGAGGTGAGTTCTCTCACGAATATTGGATGAGACACGCTCTCACCTGGCTAAGAGAGCCAGGG
ACGAAAGAGAGGTGCCAGTTGGCGCTGTCCTGGTGTGAACAATCGCGTCATCGGAGAAGGATGGAATCGCGCCATT
GGCCTGCACGATCCAACCGCACATGCCGAAATTATGGCTCTGCGGCAAGGCGGCCTCGTGATGCAAAATTACAGACT
GATCGATGCTACCCTCTACGTCACCTTCGAGCCCTGTGTCATGTGTGCTGGGGCAATGATTCACTCCCGGATTGGCC
GCGTGGTGTGTTGGAGTGCAGGATGCCAAGACTGGCGCCGCTGGATCTCTGATGGACGTCCTGCACTATCTGGGATG
AACCACCGGGTCGAGATCACAGAGGGAATTCTGGCTGACGAGTGCCTGCCCCTGCTGTGCTACTTCTTTAGAATGCC
CAGACAGGTGTTCAACGCCCAGAAAAAGCTCAGAGCAGCACCGATTCCGGCGGAAGCAGCGGAGGATCTTCTGGAA
GCGAAACCCAGGCACCAGCGAGTCTGCCACACCAGAATCATCTGGCGGTAGCTCCGGCGGCAGCGACAAGAAGTAT
TCTATCGGACTGGCCATCGGCACCAACTCTGTTGGATGGGCCGTGATCACCGACGAGTACAAGGTGCCAGCAAGAA
ATTCAAGGTGCTGGGCAACACCGACAGGCACAGCATCAAGAAGAACCTGATCGGCGCACTGCTGTTGACTCTGGCG
AAACAGCCGAGGCCACCAGACTGAAGAGAACAGCCCGCAGACGGTACACCAGAAGAAAGAACCGGATCTGCTACCTC
CAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTCCACAGACTGGAAGAGTCCTTCTGGTGGGA
AGAGGACAAGAAGCACGAGAGACACCCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACCACGAGAAGTACCCCA
CCATCTACCACCTGAGAAAGAACTGGTGGACAGCACCGACAAGGCCGACCTGAGACTGATCTATCTGGCCCTGGCT
CACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAATCCTGACAACAGCGACGTGGACAAGCTGTT
CATCCAGCTGGTGCAGACCTACAACCAGCTGTTGAGGAAAACCCCATCAACGCCAGCGGAGTGGATGCCAAGGCCA
TCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGAAAAATCTGATCGCCAGCTGCCTGGCGAGAAGAAGAATGGC
CTGTTCCGGCAACCTGATTGCCCTGAGCCTGGCCTGACACCTAACTTCAAGAGCAACTTCGACCTGGCCGAGGACGC
CAAACCTGCAGCTGAGCAAGGACACCTACGACG

[0758] ACGACCTGGACAATCTGCTGGCCAGATCGGCGATCAGTACGCCGACTTGTTTCTGGCCGCCAAGAATC
TGAGCGACGCCATCCTGCTGTCCGACATCCTGAGAGTGAACACCGAGATCACCAAGGCACCTCTGAGCGCCTCTATG
ATCAAGAGATACGACGAGCACCACCAGGATCTGACCTGCTGAAGGCCCTCGTTAGACAGCAGCTGCCAGAGAAGTA
CAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCGGCTACATTGATGGCGGAGCCAGCCAAGAGGAATTCT
ACAAGTTCATCAAGCCCATCCTCGAGAAGATGGACGGCACCGAGGAAGTCTGGTCAAGCTGAACAGAGAGGACCTG
CTGAGAAAAGCAGAGAACCTTCGACAACGGCAGCATCCCTCACCAGATCCACCTGGGAGAACTGCACGCCATTCTGCG
GAGACAAGAGGACTTTTACCCATTCTGAAGGACAACCGGGAAAAGATCGAGAAAATCCTGACCTTCAGGATCCCT
ACTACGTGGGACCACTGGCCAGAGGCAATAGCAGATTCGCCTGGATGACCAGAAAGTGCCTGAGCTACGATACCGAG
ATCCTGACCGTGGAATACGGCTTCCTGCCTATCGGCAAGATCGTCGAGGAACGGATCGAGTGCACAGTGTACACCGT
GGATAAGAATGGCTTCGTGTACCCAGCCTATCGCTCAGTGGCACAACAGAGGCGAGCAAGAGGTGTTGAGTACT
GCCTGGAAGATGGCAGCATCATCCGGGCCACCAAGGACCACAAGTTTATGACCACCGACGGCCAGATGCTGCCCATC
GACGAGATCTTTGAGAGAGGCCTGGACCTGAAACAGGTGGACGGACTGCCTTGA

[0759] ABE7.10_Cfa-N_断裂_T472C:

[0760] ATGAGCGAGGTGGAATTCAGCCACGAGTACTGGATGCGGCACGCCCTGACACTGGCCAAAAGAGCTTGG

GACGAGAGGGAAGTGCCTGTGGGAGCTGTGCTGGTGCACAACAACAGAGTGATCGGCGAAGGCTGGAACAGACCCAT
CGGCAGACACGATCCTACAGCTCACGCCGAGATCATGGCCCTGAGACAAGGCGGACTGGTCATGCAGAACTACCGGC
TGATCGACGCCACACTGTACGTGACCCTGGAACCTTGCGTGATGTGTGCCGGCGCTATGATCCACAGCAGAATCGGC
AGAGTGGTGTTCGGCGCCAGAGATGCCAAAACAGGCGCTGCCGGAAGCCTGATGGATGTGCTGCATACCCCGGCAT
GAACCACAGAGTGGAATCACCAGGGGCATCCTGGCCGATGAATGTGCCGCTCTGCTGAGCGACTTCTTCCGGATGC
GGCGGCAAGAGATCAAGGCCCAGAAGAAGGCCAGTCCAGCACAGATAGCGGCGGATCTAGCGGAGGCAGCTCTGGA
TCTGAGACACCTGGCACAAGCGAGAGCGCCACACCTGAAAGTTCTGGCGGTTCTTCTGGCGGCAGCAGCGAGGTGCA
GTTCTCTCACGAATATTGGATGAGACACGCTCTCACCTGGCTAAGAGAGCCAGGGACGAAAGAGAGGTGCCAGTTG
GCGCTGTCTGGTGTGAACAATCGCGTCATCGGAGAAGGATGGAATCGCGCCATTGGCCTGCACGATCCAACCGCA
CATGCCGAAATTATGGCTCTGCGGCAA

[0761] GGCGGCCTCGTGATGCAAAATTACAGACTGATCGATGCTACCCTCTACGTACCTTCGAGCCCTGTGTC
ATGTGTGCTGGGGCAATGATTCACTCCCGGATTGGCCGCGTGGTGTGGAGTGCGGAATGCCAAGACTGGCGCCGC
TGGATCTCTGATGGACGTCTGCACTATCCTGGGATGAACCACCGGGTCGAGATCACAGAGGGAATTCTGGCTGACG
AGTGCGCTGCCCTGCTGTGCTACTTCTTTAGAATGCCAGACAGGTGTTCAACGCCCAGAAAAAGCTCAGAGCAGC
ACCGATTCCGGCGGAAGCAGCGGAGGATCTTCTGGAAGCGAAACCCAGGCACCAGCGAGTCTGCCACACCAGAATC
ATCTGGCGGTAGCTCCGGCGGCAGCGACAAGAAGTATTCTATCGGACTGGCCATCGGCACCAACTCTGTTGGATGGG
CCGTGATCACCGACGAGTACAAGGTGCCAGCAAGAAATCAAGGTGCTGGGCAACACCGACAGGCACAGCATCAAG
AAGAACCTGATCGGCGCACTGCTGTTCTGACTCTGGCGAAACAGCCGAGGCCACCAGACTGAAGAGAACAGCCCGCAG
ACGGTACACCAGAAGAAAGAACCGGATCTGCTACCTCCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACA
GCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGAAAGAGGACAAGAAGCACGAGAGACACCCATCTTCGGCAAC
ATCGTGGACGAGGTGGCCTACCACGAGAAGTACCCACCATCTACCACCTGAGAAAGAACTGGTGGACAGCACCGA
CAAGGCCGACCTGAGACTGATCTATCTGGCCCTGGCTCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCG
ACCTGAATCCTGACAACAGCGACGTGGACAAGCTGTTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTTCGAGGAA
AACCCCATCAACGCCAGCGGAGTGGATGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAA
TCTGATCGCCCAGCTGCCTGGCGAGAAGAAGATGGCCTGTTTCGGCAACCTGATTGCCCTGAGCCTGGGCCTGACAC
CTAACTTCAAGAGCAACTTCGACCTGGCCGAGGACGCCAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTG
GACAATCTGCTGGCCAGATCGGCGATCAGTACGCCGACTTGTTTCTGGCCGCCAAGAATCTGAGCGACGCCATCCT
GCTGTCCGACATCCTGAGAGTGAACACCGAGATCACCAAGGCACCTCTGAGCGCCTCTATGATCAAGAGATACGACG
AGCACCACCAGGATCTGACCCTGCTGAAGGCCCTCGTTAGACAGCAGCTGCCAGAGAAGTACAAAGAGATTTTCTTC
GACCAGAGCAAGAACGGCTACGCCGGCTACATTGATGGCGGAGCCAGCCAAGAGGAATTCTACAAGTTCATCAAGCC
CATCCTCGAGAAGATGGACGGCACCGAGGAACTGCTGGTCAAGCTGAACAGAGAGGACCTGCTGAGAAAGCAGAGAA
CCTTCGACAACGGCAGCATCCCTCACCAGATCCACCTGGGAGAACTGCACGCCATTCTGCGGAGACAAGAGGACTTT
TACCCATTCTGAAGGACAACCGGGAAAAAGATCGAGAAAATCCTGACCTTCAGGATCCCCTACTACGTGGGACCACT
GGCCAGAGGCAATAGCAGATTGCGCTGGATGACCAG

[0762] AAAGAGCGAGGAATGCCTGAGCTACGATACCGAGATCCTGACCGTGAATACGGCTTCCTGCCTATCGG
CAAGATCGTCGAGGAACGGATCGAGTGACAGTGTACACCGTGGATAAGAATGGCTTCGTGTACACCCAGCCTATCG
CTCAGTGGCACAACAGAGGCGAGCAAGAGGTGTTTCGAGTACTGCCTGGAAGATGGCAGCATCATCCGGGCCACCAAG
GACCACAAGTTTATGACCACCGACGGCCAGATGCTGCCCATCGACGAGATCTTTGAGAGAGGCCTGGACCTGAAACA
GGTGGACGGACTGCCTTGA

[0763] ABE7.10_Cfa-N_断裂_T474C:

[0764] ATGAGCGAGGTGGAATTCAGCCACGAGTACTGGATGCGGCACGCCCTGACACTGGCCAAAAGAGCTTGG
GACGAGAGGGAAGTGCCTGTGGGAGCTGTGCTGGTGCACAACAACAGAGTGATCGGCGAAGGCTGGAACAGACCCAT
CGGCAGACACGATCCTACAGCTCACGCCGAGATCATGGCCCTGAGACAAGGCGGACTGGTCATGCAGAACTACCGGC
TGATCGACGCCACACTGTACGTGACCCTGGAACCTTGCGTGATGTGTGCCGGCGCTATGATCCACAGCAGAATCGGC
AGAGTGGTGTTCGGCGCCAGAGATGCCAAAACAGGCGCTGCCGGAAGCCTGATGGATGTGCTGCATCACCCCGGCAT
GAACCACAGAGTGGAATCACCGAGGGCATCCTGGCCGATGAATGTGCCGCTCTGCTGAGCGACTTCTTCCGGATGC
GGCGGCAAGAGATCAAGGCCCAGAAGAAGGCCAGTCCAGCACAGATAGCGGCGGATCTAGCGGAGGCAGCTCTGGA
TCTGAGACACCTGGCACAAGCGAGAGCGCCACACCTGAAAGTTCTGGCGGTTCTTCTGGCGGCAGCAGCGAGGTGCA
GTTCTCTCACGAATATTGGATGAGACACGCTCTCACCTGGCTAAGAGAGCCAGGGACGAAAGAGAGGTGCCAGTTG
GCGCTGTCTGGTGTGAACAATCGCGTCATCGGAGAAGGATGGAATCGCGCCATTGGCCTGCACGATCCAACCGCA
CATGCCGAAATTATGGCTCTGCGGCAAGGCGGCCTCGTGATGCAAAATTACAGACTGATCGATGCTACCCTCTACGT
CACCTTCGAGCCCTGTGTCATGTGTGCTGGGGCAATGATTCACTCCCGGATTGGCCGCGTGGTGTGTTGGAGTGCGGA
ATGCCAAGACTGGCGCCGCTGGATCTCTGATGGACGTCCTGCACTATCCTGGGATGAACCACCGGGTCGAGATCACA
GAGGGAATTCTGGCTGACGAGTGCGCTGCCCTGCTGTGCTACTTCTTTAGAATGCCAGACAGGTGTTCAACGCCCA
GAAAAAGCTCAGAGCAGCACCGATTCCGGCGGAAGCAGCGGAGGATCTTCTGGAAGCGAAACCCAGGCACCAGCG
AGTCTGCCACACCAGAATCATCTGGCGGTAGCTCCGGCGGCAGCGACAAGAAGTATTCTATCGGACTGGCCATCGGC
ACCAACTCTGTTGGATGGGCCGTGATCACCGACGAGTACAAGGTGCCAGCAAGAAATTCAAGGTGCTGGGCAACAC
CGACAGGCACAGCATCA

[0765] AGAAGAACCTGATCGGCGCACTGCTGTTGCACTCTGGCGAAACAGCCGAGGCCACCAGACTGAAGAGAA
CAGCCCGCAGACGGTACACCAGAAGAAAGAACCGGATCTGCTACCTCCAAGAGATCTTCAGCAACGAGATGGCCAAG
GTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGAAGAGGACAAGAAGCACGAGAGACACCCCAT
CTTCGGCAACATCGTGGACGAGGTGGCCTACCACGAGAAGTACCCACCATCTACCACCTGAGAAAGAACTGGTGG
ACAGCACCGACAAGGCCGACCTGAGACTGATCTATCTGGCCCTGGCTCACATGATCAAGTTCCGGGGCCACTTCCTG
ATCGAGGGCGACCTGAATCCTGACAACAGCGACGTGGACAAGCTGTTTCATCCAGCTGGTGCAGACCTACAACCAGCT
GTTTCGAGGAAAACCCCATCAACGCCAGCGGAGTGGATGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGAC
GGCTGGAAAATCTGATCGCCCAGCTGCCTGGCGAGAAGAAGATGGCCTGTTTCGGCAACCTGATTGCCCTGAGCCTG
GGCCTGACACCTAACTTCAAGAGCAACTTCGACCTGGCCGAGGACGCCAACTGCAGCTGAGCAAGGACACCTACGA
CGACGACCTGGACAATCTGCTGGCCCAGATCGGCGATCAGTACGCCGACTTGTTTCTGGCCGCCAAGAATCTGAGCG
ACGCCATCCTGCTGTCCGACATCCTGAGAGTGAACACCGAGATCACCAAGGCACCTCTGAGCGCCTCTATGATCAAG
AGATACGACGAGCACCACCAGGATCTGACCCTGCTGAAGGCCCTCGTTAGACAGCAGCTGCCAGAGAAGTACAAAGA
GATTTTCTTCGACCAGAGCAAGAACGGCTACGCCGGCTACATTGATGGCGGAGCCAGCCAAGAGGAATTCTACAAGT
TCATCAAGCCCATCCTCGAGAAGATGGACGGCACCGAGGAACTGCTGGTCAAGCTGAACAGAGAGGACCTGCTGAGA
AAGCAGAGAACCTTCGACAACGGCAGCATCCCTCACCAGATCCACCTGGGAGAACTGCACGCCATTCTGCGGAGACA
AGAGGACTTTTACCCATTCTGAAGGACAACCGGGAAAAGATCGAGAAAATCCTGACCTTCAGGATCCCCTACTACG
TGGGACCACTGGCCAGAGGCAATAGCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCTGCCTGAGCTAC
GATACCGAGATCCTGACCGTGGAATACGGCTTCCTGCCTATCGGCAAGATCGTCGAGGAACGGATCGAGTGCACAGT
GTACACCGTGGATAAGAATGGCTTCGTGTACCCCAGCCTATCGCTCAGTGGCACAACAGAGGCGAGCAAGAGGTGT
TCGAGTACTGCCTGGAAGATGGCAGCATCATCCGGGCCACCAAGGACCACAAGTTTATGACCACCGACGGCCAGATG

CTGCCCATCGACGAGATCTTTGAGAGAGGCCTGGACCTGAAACAGGTGGACGGACTGCCTTGA

[0766] ABE7.10_Cfa-N_断裂_C574C:

[0767] ATGAGCGAGGTGGAATTCAGCCACGAGTACTGGATGCGGCACGCCCTGACACT

[0768] GGCCAAAAGAGCTTGGGACGAGAGGGAAGTGCTGTGGGAGCTGTGCTGGTGCACAACAACAGAGTGAT
CGGCGAAGGCTGGAACAGACCCATCGGCAGACACGATCCTACAGCTCACGCCGAGATCATGGCCCTGAGACAAGGCG
GACTGGTCATGCAGAACTACCGGCTGATCGACGCCACACTGTACGTGACCCTGGAACCTTGCGTGATGTGTGCCGGC
GCTATGATCCACAGCAGAATCGGCAGAGTGGTGTTCGGCGCCAGAGATGCCAAAACAGGCGCTGCCGGAAGCCTGAT
GGATGTGCTGCATCACCCCGGCATGAACCACAGAGTGGAATCACCGAGGGCATCCTGGCCGATGAATGTGCCGCTC
TGCTGAGCGACTTCTTCCGGATGCGGCGCAAGAGATCAAGGCCCAGAAGAAGGCCCAGTCCAGCACAGATAGCGGC
GGATCTAGCGGAGGCAGCTCTGGATCTGAGACACCTGGCACAAGCGAGAGCGCCACACCTGAAAGTTCTGGCGGTTC
TTCTGGCGGCAGCAGCGAGGTCGAGTTCTCTCACGAATATTGGATGAGACACGCTCTCACCTGGCTAAGAGAGCCA
GGGACGAAAGAGAGGTGCCAGTTGGCGCTGTCTGGTGTGAACAATCGCGTCATCGGAGAAGGATGGAATCGCGCC
ATTGGCCTGCACGATCCAACCGCACATGCCGAAATTATGGCTCTGCGGCAAGGCGGCCTCGTGATGCAAAATTACAG
ACTGATCGATGCTACCCTCTACGTACCTTCGAGCCCTGTGTCATGTGTGCTGGGGCAATGATTCACTCCCGGATTG
GCCGCGTGGTGTGTTGGAGTGCGGAATGCCAAGACTGGCGCCGCTGGATCTCTGATGGACGTCCTGCACTATCCTGGG
ATGAACCACCGGGTCGAGATCACAGAGGGAATTCTGGCTGACGAGTGCCTGCCCTGCTGTGCTACTTCTTTAGAAT
GCCCAGACAGGTGTTCAACGCCCAGAAAAAGCTCAGAGCAGCACCGATTCCGGCGGAAGCAGCGGAGGATCTTCTG
GAAGCGAAACCCAGGCACCGAGTCTGCCACACCAGAATCATCTGGCGGTAGCTCCGGCGGCAGCGACAAGAAG
TATTCTATCGGACTGGCCATCGGCACCAACTCTGTTGGATGGGCCGTGATCACCGACGAGTACAAGGTGCCAGCAA
GAAATTCAAGGTGCTGGGCAACACCGACAGGCACAGCATCAAGAAGAACCTGATCGGCGCACTGCTGTTGACTCTG
GCGAAACAGCCGAGGCCACCAGACTGAAGAGAACAGCCCGCAGACGGTACACCAGAAGAAAGAACCGGATCTGCTAC
CTCCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGT
GGAAGAGGACAAGAAGCACGAGAGACACCCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACCACGAGAAGTACC
CCACCATCTACCACCTGAGAAAGAACTGGTGGACAGCACCGACAAGGCCGACCTGAGACTGATCTATCTGGCCCTG
GCTCACATGATCAAGTTCCGGGGCCACTTCTGATCGAGGGCGACCTGAATCCTGACAACAGCGACGTGGACAAGCT
GTTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTTCGAGGAAAACCCCATCAACGCCAGCGAGTGGATGCCAAGG
CCATCCTGTCTGCCAGACTGAGCAAGAGCAG

[0769] ACGGCTGGAAAATCTGATCGCCCAGCTGCCTGGCGAGAAGAAGATGGCCTGTTCCGGCAACCTGATTGC
CCTGAGCCTGGGCCTGACACCTAACTTCAAGAGCAACTTCGACCTGGCCGAGGACGCCAACTGCAGCTGAGCAAGG
ACACCTACGACGACGACCTGGACAATCTGCTGGCCCAGATCGGCGATCAGTACGCCGACTTGTCTTGGCCGCCAAG
AATCTGAGCGACGCCATCCTGCTGTCCGACATCCTGAGAGTGAACACCGAGATCACCAGGCACCTCTGAGCGCCTC
TATGATCAAGAGATACGACGAGCACCACAGGATCTGACCCTGCTGAAGGCCCTCGTTAGACAGCAGCTGCCAGAGA
AGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCGGCTACATTGATGGCGGAGCCAGCCAAGAGGAA
TTCTACAAGTTCATCAAGCCCATCCTCGAGAAGATGGACGGCACCGAGGAAGTCTGGTCAAGCTGAACAGAGAGGA
CCTGCTGAGAAAGCAGAGAACCTTCGACAACGGCAGCATCCCTCACCAGATCCACCTGGGAGAACTGCACGCCATT
TGCGGAGACAAGAGGACTTTTACCCATTCTGAAGGACAACGGGAAAAGATCGAGAAAATCCTGACCTTCAGGATC
CCCTACTACGTGGGACCACTGGCCAGAGGCAATAGCAGATTGCCTGGATGACCAGAAAGAGCGAGGAAACCATCAC
TCCCTGGAACCTTCGAGGAAGTGGTGGACAAGGGCGCCAGCGCTCAGTCCTTCATCGAGCGGATGACCAACTTCGATA
AGAACCTGCCTAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTACAACGAGCTGACC

AAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTTCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGATCT
GCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCCTGAGCT
ACGATACCGAGATCCTGACCGTGAATACGGCTTCCTGCCTATCGGCAAGATCGTCGAGGAACGGATCGAGTGCACA
GTGTACACCGTGGATAAGAATGGCTTCGTGTACACCCAGCCTATCGCTCAGTGGCACAACAGAGGCGAGCAAGAGGT
GTTTCGAGTACTGCCTGGAAGATGGCAGCATCATCCGGGCCACCAAGGACCACAAGTTTATGACCACCGACGGCCAGA
TGCTGCCCATCGACGAGATCTTTGAGAGAGGCCTGGACCTGAAACAGGTGGACGGACTGCCTTGA

[0770] ABE7.10_Cfa-N_断裂_S577C:

[0771] ATGAGCGAGGTGGAATTCAGCCACGAGTACTGGATGCGGCACGCCCTGACACTGGCCAAAAGAGCTTGG
GACGAGAGGGAAGTGCCTGTGGGAGCTGTGCTGGTGCACAACAACAGAGTGATCGGCGAAGGCTGGAACAGACCCAT
CGGCAGACACGATCCTACAGCTCACGCCGAGATCATGGCCCTGAGACAAGGCGGACTGGTCATGCAGAACTACCGGC
TGATCGACGCCACACTGTACGTGACCCTGGAACCTTGCCTGATGTGTGCCGGCGCTATGATCCACAGCAGAATCGGC
AGAGTGGTGTTCGGC

[0772] GCCAGAGATGCCAAAACAGGCGCTGCCGGAAGCCTGATGGATGTGCTGCATACCCCGGCATGAACCAC
AGAGTGGAATCACCAGGGGCATCCTGGCCGATGAATGTGCCGCTCTGCTGAGCGACTTCTTCCGGATGCGGCGGCA
AGAGATCAAGGCCCAGAAGAAGGCCAGTCCAGCACAGATAGCGGCGGATCTAGCGGAGGCAGCTCTGGATCTGAGA
CACCTGGCACAAGCGAGAGCGCCACACCTGAAAGTTCTGGCGGTTCTTCTGGCGGCAGCAGCGAGGTCGAGTTCTCT
CACGAATATTGGATGAGACACGCTCTCACCTGGCTAAGAGAGCCAGGGACGAAAGAGAGGTGCCAGTTGGCGCTGT
CCTGGTGTGTAACAATCGCGTCATCGGAGAAGGATGGAATCGCGCCATTGGCCTGCACGATCCAACCGCACATGCCG
AAATTATGGCTCTGCGGCAAGGCGGCCTCGTGATGCAAAATTACAGACTGATCGATGCTACCCTCTACGTCACCTTC
GAGCCCTGTGTATGTGTGCTGGGGCAATGATTCACTCCCGGATTGGCCGCGTGGTGTGAGTGCAGGATGCCAA
GACTGGCGCCGCTGGATCTCTGATGGACGTCCTGCACTATCCTGGGATGAACCACCGGGTCGAGATCACAGAGGGAA
TTCTGGCTGACGAGTGCCTGCCCTGCTGTGCTACTTCTTTAGAATGCCAGACAGGTGTTCAACGCCCAGAAAAAA
GCTCAGAGCAGCACCGATTCCGGCGGAAGCAGCGGAGGATCTTCTGGAAGCGAAACCCAGGCACCAGCGAGTCTGC
CACACCAGAATCATCTGGCGGTAGCTCCGGCGGCAGCGACAAGAAGTATTCTATCGGACTGGCCATCGGCACCAACT
CTGTTGGATGGGCCGTGATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAGGTGCTGGGCAACACCGACAGG
CACAGCATCAAGAAGAACCTGATCGGCGCACTGCTGTTGACTCTGGCGAAACAGCCGAGGCCACCAGACTGAAGAG
AACAGCCCGCAGACGGTACACCAGAAGAAAGAACCGGATCTGCTACCTCCAAGAGATCTTCAGCAACGAGATGGCCA
AGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCTGGTGGAGAGGACAAGAAGCACGAGAGACACCCC
ATCTTCGGCAACATCGTGGACGAGGTGGCCTACCACGAGAAGTACCCACCATCTACCACCTGAGAAAGAACTGGT
GGACAGCACCGACAAGGCCGACCTGAGACTGATCTATCTGGCCCTGGCTCACATGATCAAGTTCCGGGGCCACTTCC
TGATCGAGGGCGACCTGAATCCTGACAACAGCGACGTGGACAAGCTGTTTCATCCAGCTGGTGCAGACCTACAACCAG
CTGTTTCGAGGAAAACCCCATCAACGCCAGCGGAGTGATGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAG
ACGGCTGGAAAATCTGATCGCCCAGCTGCCTGGCGAGAAGAAGATGGCCTGTTTCGGCAACCTGATTGCCCTGAGCC
TGGGCCTGACACCTAACTTCAAGAGCAACTTCGACCTGGCCGAGGACGCCAACTGCAGCTGAGCAAGGACACCTAC
GACGACGACCTGGACAATCTGCTGGCCCAGATCGGCGATCAGTACGCCGACTTGTTTCTGGCCGCCAAGAATCTGAG
CGACGCCATCCTGCTGTCCGACATCCTGAGAGT

[0773] GAACACCGAGATACCAAGGCACCTCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGA
TCTGACCCTGCTGAAGGCCCTCGTTAGACAGCAGCTGCCAGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGA
ACGGCTACGCCGGCTACATTGATGGCGGAGCCAGCCAAGAGGAATTCTACAAGTTCATCAAGCCCATCCTCGAGAAG

ATGGACGGCACCGAGGAAGTCTGGTCAAGCTGAACAGAGAGGACCTGCTGAGAAAGCAGAGAACCTTCGACAACGG
CAGCATCCCTCACCAGATCCACCTGGGAGAACTGCACGCCATTCTGCGGAGACAAGAGGACTTTTACCCATTCTGA
AGGACAACCGGAAAAAGATCGAGAAAAATCCTGACCTTCAGGATCCCCTACTACGTGGGACCACTGGCCAGAGGCAAT
AGCAGATTTCGCTGGATGACCAGAAAGAGCGAGGAAACCATCACTCCCTGGAAGTTCGAGGAAGTGGTGGACAAGGG
CGCCAGCGCTCAGTCCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCTAACGAGAAGGTGCTGCCCAAGC
ACAGCCTGCTGTACGAGTACTTCACCGTGTACAACGAGCTGACCAAAGTGAATACGTGACCGAGGGAATGAGAAAG
CCCGCCTTTCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGATCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAA
GCAGCTGAAAGAGGACTACTTCAAGAAAAATCGAGTGCTTCGACTGCCTGAGCTACGATACCGAGATCCTGACCGTGG
AATACGGCTTCTGCCTATCGGCAAGATCGTCGAGGAACGGATCGAGTGCACAGTGTACACCGTGGATAAGAATGGC
TTCGTGTACACCCAGCCTATCGCTCAGTGGCACAACAGAGGCGAGCAAGAGGTGTTTCGAGTACTGCCTGGAAGATGG
CAGCATCATCCGGGCCACCAAGGACCACAAGTTTATGACCACCGACGGCCAGATGCTGCCCATCGACGAGATCTTTG
AGAGAGGCCTGGACCTGAAACAGGTGGACGGACTGCCTTGA

[0774] ABE7.10_Cfa-N 断裂 A589C:

[0775] ATGAGCGAGGTGGAATTCAGCCACGAGTACTGGATGCGGCACGCCCTGACACTGGCCAAAAGAGCTTGG
GACGAGAGGGAAGTGCCTGTGGGAGCTGTGCTGGTGACAAACAACAGAGTGATCGGCGAAGGCTGGAACAGACCCAT
CGGCAGACACGATCCTACAGCTCACGCCGAGATCATGGCCCTGAGACAAGGCGGACTGGTCATGCAGAACTACCGGC
TGATCGACGCCACACTGTACGTGACCCTGGAACCTTGCGTGATGTGTGCCGGCGCTATGATCCACAGCAGAATCGGC
AGAGTGGTGTTCGGCGCCAGAGATGCCAAAACAGGCGCTGCCGGAAGCCTGATGGATGTGCTGCATCACCCCGGCAT
GAACCACAGAGTGGAATCACCGAGGGCATCCTGGCCGATGAATGTGCCGCTCTGCTGAGCGACTTCTTCCGGATGC
GGCGGCAAGAGATCAAGGCCCAGAAGAAGGCCCAGTCCAGCACAGATAGCGGCGGATCTAGCGGAGGCAGCTCTGGA
TCTGAGACACCTGGCACAAGCGAGAGCGCCACACCTGAAAGTTCTGG

[0776] CGGTTCTTCTGGCGGCAGCAGCGAGGTCGAGTTCTCTCACGAATATTGGATGAGACACGCTCTCACCT
GGCTAAGAGAGCCAGGGACGAAAGAGAGGTGCCAGTTGGCGCTGTCTGGTGTGAACAATCGCGTCATCGGAGAAG
GATGGAATCGCGCCATTGGCCTGCACGATCCAACCGCACATGCCGAAATTATGGCTCTGCGGCAAGGCGGCCTCGTG
ATGCAAAATTACAGACTGATCGATGCTACCTCTACGTACCTTCGAGCCCTGTGTCATGTGTGCTGGGGCAATGAT
TCACTCCCGGATTGGCCGCGTGGTGTGTTGGAGTGCGGAATGCCAAGACTGGCGCCGCTGGATCTCTGATGGACGTCC
TGCACTATCCTGGGATGAACCACCGGGTCGAGATCACAGAGGGAATTCTGGCTGACGAGTGCCTGCCCTGCTGTGC
TACTTCTTTAGAATGCCAGACAGGTGTTCAACGCCCAGAAAAAGCTCAGAGCAGCACCGATTCCGGCGGAAGCAG
CGGAGGATCTTCTGGAAGCGAAACCCAGGCACCAGCGAGTCTGCCACACCAGAATCATCTGGCGGTAGCTCCGGCG
GCAGCGACAAGAAGTATTCTATCGGACTGGCCATCGGCACCAACTCTGTTGGATGGGCCGTGATCACCGACGAGTAC
AAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACAGGCACAGCATCAAGAAGAACCTGATCGGCGCACT
GCTGTTTCGACTCTGGCGAAACAGCCGAGGCCACCAGACTGAAGAGAACAGCCCGCAGACGGTACACCAGAAGAAAGA
ACCGGATCTGCTACCTCCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAA
GAGTCCTTCTGGTGAAGAGGACAAGAAGCACGAGAGACACCCATCTTCGGCAACATCGTGGACGAGGTGGCCTA
CCACGAGAAGTACCCACCATCTACCACCTGAGAAAGAACTGGTGGACAGCACCGACAAGGCCGACCTGAGACTGA
TCTATCTGGCCCTGGCTCATATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAATCCTGACAACAGC
GACGTGGACAAGCTGTTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTTCGAGGAAAACCCCATCAACGCCAGCGG
AGTGGATGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAATCTGATCGCCAGCTGCCTG
GCGAGAAGAAGAATGGCCTGTTTCGGCAACCTGATTGCCCTGAGCCTGGGCCTGACACCTAACTTCAAGAGCAACTTC

GACCTGGCCGAGGACGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAATCTGCTGGCCAGAT
CGGCGATCAGTACGCCGACTTGTTTCTGGCCGCCAAGAATCTGAGCGACGCCATCCTGCTGTCCGACATCCTGAGAG
TGAACACCGAGATCACCAAGGCACCTCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGATCTGACC
CTGCTGAAGGCCCTCGTTAGACAGCAGCTGCCAGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTA
CGCCGGCTACATTGATGGCGGAGCCAGCCAAGAGGAATTCTACAAGTTCATCAAGCCCATCCTCGAGAAGATGGACG
GCACCGAGGAAGTCTGGTCAAGCTGAACAGAGAG

[0777] GACCTGCTGAGAAAGCAGAGAACCTTCGACAACGGCAGCATCCCTCACCAGATCCACCTGGGAGAACTG
CACGCCATTCTGCGGAGACAAGAGGACTTTTACCCATTCCTGAAGGACAACCGGGAAAAGATCGAGAAAATCCTGAC
CTTCAGGATCCCCTACTACGTGGGACCACTGGCCAGAGGCAATAGCAGATTTCGCCTGGATGACCAGAAAGAGCGAGG
AAACCATCACTCCCTGGAAGTTCGAGGAAGTGGTGGACAAGGGCGCCAGCGCTCAGTCCTTCATCGAGCGGATGACC
AACTTCGATAAGAACCTGCCTAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTACAA
CGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTTCTGAGCGGCGAGCAGAAAAAGGCCA
TCGTGGATCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAG
TGCTTCGACAGCGTCGAGATCTCCGGCGTGGAAGATCGGTTCAATTGCCTGAGCTACGATACCGAGATCCTGACCGT
GGAATACGGCTTCCTGCCTATCGGCAAGATCGTCGAGGAACGGATCGAGTGCACAGTGTACACCGTGGATAAGAATG
GCTTCGTGTACACCCAGCCTATCGCTCAGTGGCACAACAGAGGCGAGCAAGAGGTGTTGAGTACTGCCTGGAAGAT
GGCAGCATCATCCGGGCCACCAAGGACCACAAGTTTATGACCACCGACGGCCAGATGCTGCCCATCGACGAGATCTT
TGAGAGAGGCCTGGACCTGAAACAGGTGGACGGACTGCCTTGA

[0778] ABE7.10_Cfa-N_断裂_S590C:

[0779] ATGAGCGAGGTGGAATTCAGCCACGAGTACTGGATGCGGCACGCCCTGACACTGGCCAAAAGAGCTTGG
GACGAGAGGGAAAGTGCTGTGGGAGCTGTGCTGGTGCACAACAACAGAGTGATCGGCGAAGGCTGGAACAGACCCAT
CGGCAGACACGATCCTACAGCTCACGCCGAGATCATGGCCCTGAGACAAGGCGGACTGGTCATGCAGAACTACCGGC
TGATCGACGCCACACTGTACGTGACCCTGGAACCTTGCGTGATGTGTGCCGGCGCTATGATCCACAGCAGAATCGGC
AGAGTGGTGTTCGGCGCCAGAGATGCCAAAACAGGCGCTGCCGGAAGCCTGATGGATGTGCTGCATACCCCGGCAT
GAACCACAGAGTGGAATCACCGAGGGCATCCTGGCCGATGAATGTGCCGCTCTGCTGAGCGACTTCTTCCGGATGC
GGCGGCAAGAGATCAAGGCCCAGAAGAAGGCCCAGTCCAGCACAGATAGCGGCGGATCTAGCGGAGGCAGCTCTGGA
TCTGAGACACCTGGCACAAGCGAGAGCGCCACACCTGAAAGTTCTGGCGGTTCTTCTGGCGGCAGCAGCGAGGTGCA
GTTCTCTCACGAATATTGGATGAGACACGCTCTCACCTGGCTAAGAGAGCCAGGGACGAAAGAGAGGTGCCAGTTG
GCGCTGTCTGGTGTGAACAATCGCGTCATCGGAGAAGGATGGAATCGCGCCATTGGCCTGCACGATCCAACCGCA
CATGCCGAAATTATGGCTCTGCGGCAA

[0780] GGCGGCCTCGTGATGCAAAATTACAGACTGATCGATGCTACCCTCTACGTCACCTTCGAGCCCTGTGTC
ATGTGTGCTGGGGCAATGATTCACCTCCCGGATTGGCCGCGTGGTGTGGAGTGCGGAATGCCAAGACTGGCGCCGC
TGGATCTCTGATGGACGTCCTGCACTATCCTGGGATGAACCACCGGGTCGAGATCACAGAGGGAATTCTGGCTGACG
AGTGCGCTGCCCTGCTGTGCTACTTCTTTAGAATGCCCAGACAGGTGTTCAACGCCCAGAAAAAGCTCAGAGCAGC
ACCGATTCCGGCGGAAGCAGCGGAGGATCTTCTGGAAGCGAAACCCAGGCACCAGCGAGTCTGCCACACCAGAATC
ATCTGGCGGTAGCTCCGGCGGCAGCGACAAGAAGTATTCTATCGGACTGGCCATCGGCACCAACTCTGTTGGATGGG
CCGTGATCACCGACGAGTACAAGGTGCCCAGCAAGAAATCAAGGTGCTGGGCAACACCGACAGGCACAGCATCAAG
AAGAACCTGATCGGCGCACTGCTGTTGACTCTGGCGAAACAGCCGAGGCCACCAGACTGAAGAGAACAGCCCGCAG
ACGGTACACCAGAAGAAAGAACCGGATCTGCTACCTCCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACA

GCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGAAGAGGACAAGAAGCACGAGAGACACCCCATCTTCGGCAAC
ATCGTGGACGAGGTGGCCTACCACGAGAAGTACCCACCATCTACCACCTGAGAAAGAACTGGTGGACAGCACC
CAAGGCCGACCTGAGACTGATCTATCTGGCCCTGGCTCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCG
ACCTGAATCCTGACAACAGCGACGTGGACAAGCTGTTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTTCGAGGAA
AACCCCATCAACGCCAGCGGAGTGGATGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAA
TCTGATCGCCCAGCTGCCTGGCGAGAAGAAGATGGCCTGTTTCGGCAACCTGATTGCCCTGAGCCTGGGCCTGACAC
CTAACTTCAAGAGCAACTTCGACCTGGCCGAGGACGCCAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTG
GACAATCTGCTGGCCCAGATCGGCGATCAGTACGCCGACTTGTTTCTGGCCGCCAAGAATCTGAGCGACGCCATCCT
GCTGTCCGACATCCTGAGAGTGAACACCGAGATCACCAAGGCACCTCTGAGCGCCTCTATGATCAAGAGATACGACG
AGCACCACCAGGATCTGACCCTGCTGAAGGCCCTCGTTAGACAGCAGCTGCCAGAGAAGTACAAAGAGATTTTCTTC
GACCAGAGCAAGAACGGCTACGCCGGCTACATTGATGGCGGAGCCAGCCAAGAGGAATTCTACAAGTTCATCAAGCC
CATCCTCGAGAAGATGGACGGCACCGAGGAACTGCTGGTCAAGCTGAACAGAGAGGACCTGCTGAGAAAGCAGAGAA
CCTTCGACAACGGCAGCATCCCTACCAGATCCACCTGGGAGAACTGCACGCCATTCTGCGGAGACAAGAGGACTTT
TACCCATTCTGAAGGACAACCGGGAAAAAGATCGAGAAAAATCCTGACCTTCAGGATCCCCTACTACGTGGGACCACT
GGCCAGAGGCAATAGCAGATTTCGCTGGATGACCAG

[0781] AAAGAGCGAGGAAACCATCACTCCCTGGAACCTCGAGGAAGTGGTGGACAAGGGCGCCAGCGCTCAGTC
CTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCTAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACG
AGTACTTCACCGTGTACAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTTCTGAGC
GGCGAGCAGAAAAAGGCCATCGTGGATCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAGAGGA
CTACTTCAAGAAAATCGAGTGCTTCGACAGCGTCGAGATCTCCGGCGTGGAAGATCGGTTCAATGCCTGCCTGAGCT
ACGATACCGAGATCCTGACCGTGAATACGGCTTCCTGCCTATCGGCAAGATCGTCGAGGAACGGATCGAGTGCACA
GTGTACACCGTGGATAAGAATGGCTTCGTGTACACCCAGCCTATCGCTCAGTGGCACAACAGAGGCGAGCAAGAGGT
GTTTCGAGTACTGCCTGGAAGATGGCAGCATCATCCGGGCCACCAAGGACCACAAGTTTATGACCACCGACGGCCAGA
TGCTGCCCATCGACGAGATCTTTGAGAGAGGCCTGGACCTGAAACAGGTGGACGGACTGCCTTGA

[0782] ABE7.10_Cfa-N_断裂_S-1C:

[0783] ATGAGCGAGGTGGAATTCAGCCACGAGTACTGGATGCGGCACGCCCTGACACTGGCCAAAAGAGCTTGG
GACGAGAGGGAAGTGCCTGTGGGAGCTGTGCTGGTGCACAACAACAGAGTGATCGGCGAAGGCTGGAACAGACCCAT
CGGCAGACACGATCCTACAGCTCACGCCGAGATCATGGCCCTGAGACAAGGCGGACTGGTTCATGCAGAACTACCGGC
TGATCGACGCCACACTGTACGTGACCCTGGAACCTTGCCTGATGTGTGCCGGCGCTATGATCCACAGCAGAATCGGC
AGAGTGGTGTTCGGCGCCAGAGATGCCAAAACAGGCGCTGCCGGAAGCCTGATGGATGTGCTGCATACCCCGGCAT
GAACCACAGAGTGGAATCACCGAGGGCATCCTGGCCGATGAATGTGCCGCTCTGCTGAGCGACTTCTTCCGGATGC
GGCGGCAAGAGATCAAGGCCCAGAAGAAGGCCAGTCCAGCACAGATAGCGGCGGATCTAGCGGAGGCAGCTCTGGA
TCTGAGACACCTGGCACAAGCGAGAGCGCCACACCTGAAAGTTCTGGCGGTTCTTCTGGCGGCAGCAGCGAGGTGGA
GTTCTCTCACGAATATTGGATGAGACACGCTCTCACCTGGCTAAGAGAGCCAGGGACGAAAGAGAGGTGCCAGTTG
GCGCTGTCTGGTGTGTAACAATCGCGTCATCGGAGAAGGATGGAATCGCGCCATTGGCCTGCACGATCCAACCGCA
CATGCCGAAATTATGGCTCTGCGGCAAGGCGGCCTCGTGATGCAAAATTACAGACTGATCGATGCTACCCTCTACGT
CACCTTCGAGCCCTGTGTATGTGTGCTGGGGCAATGATTCACTCCCGGATTGGCCGCGTGGTGTGTTGGAGTGCGGA
ATGCCAAGACTGGCGCCGCTGGATCTCTGATG

[0784] GACGTCCTGCACTATCCTGGGATGAACCACCGGGTCGAGATCACAGAGGGAATTCTGGCTGACGAGTGC

GCTGCCCTGCTGTGCTACTTCTTTAGAAATGCCAGACAGGTGTTCAACGCCAGAAAAAGCTCAGAGCAGCACCGA
TTCCGGCGGAAGCAGCGGAGGATCTTCTGGAAGCGAAACCCAGGCACCAGCGAGTCTGCCACACCAGAATCATCTG
GCGGTAGCTCCGGCGGCTGCCTGAGCTACGATACCGAGATCCTGACCGTGAATACGGCTTCCTGCCTATCGGCAAG
ATCGTCGAGGAACGGATCGAGTGACAGTGTACACCGTGGATAAGAATGGCTTCGTGTACACCCAGCCTATCGCTCA
GTGGCACAACAGAGGCGAGCAAGAGGTGTTTCGAGTACTGCCTGGAAGATGGCAGCATCATCCGGGCCACCAAGGACC
ACAAGTTTATGACCACCGACGGCCAGATGCTGCCCATCGACGAGATCTTTGAGAGAGGCCTGGACCTGAAACAGGTG
GACGGACTGCCTTGA

[0785] ABE7.10_Cfa-N_断裂_S-32C:

[0786] ATGAGCGAGGTGGAATTCAGCCACGAGTACTGGATGCGGCACGCCCTGACACTGGCCAAAAGAGCTTGG
GACGAGAGGGAAGTGCCTGTGGGAGCTGTGCTGGTGCACAACAACAGAGTGATCGGCGAAGGCTGGAACAGACCCAT
CGGCAGACACGATCCTACAGCTCACGCCGAGATCATGGCCCTGAGACAAGGCGGACTGGTCATGCAGAACTACCGGC
TGATCGACGCCACACTGTACGTGACCCTGGAACCTTGCCTGATGTGTGCCGGCGCTATGATCCACAGCAGAATCGGC
AGAGTGGTGTTCGGCGCCAGAGATGCCAAAACAGGCGCTGCCGGAAGCCTGATGGATGTGCTGCATACCCCGGCAT
GAACCACAGAGTGGAATCACCGAGGGCATCCTGGCCGATGAATGTGCCGCTCTGCTGAGCGACTTCTTCCGGATGC
GGCGGCAAGAGATCAAGGCCCAGAAGAAGGCCAGTCCAGCACAGATAGCGGCGGATCTAGCGGAGGCAGCTCTGGA
TCTGAGACACCTGGCACAAGCGAGAGCGCCACACCTGAAAGTTCTGGCGGTTCTTCTGGCGGCAGCAGCGAGGTGCA
GTTCTCTCACGAATATTGGATGAGACACGCTCTCACCTGGCTAAGAGAGCCAGGGACGAAAGAGAGGTGCCAGTTG
GCGCTGTCTGGTGTGTAACAATCGCGTCATCGGAGAAGGATGGAATCGCGCCATTGGCCTGCACGATCCAACCGCA
CATGCCGAAATTATGGCTCTGCGGCAAGGCGGCCTCGTGATGCAAAATTACAGACTGATCGATGCTACCCTCTACGT
CACCTTCGAGCCCTGTGTCATGTGTGCTGGGGCAATGATTCACTCCCGGATTGGCCGCGTGGTGTGTTGGAGTGCGGA
ATGCCAAGACTGGCGCCGCTGGATCTCTGATGGACGTCCTGCACTATCCTGGGATGAACCACCGGGTCGAGATCACA
GAGGGAATTCTGGCTGACGAGTGCGCTGCCCTGCTGTGCTACTTCTTTAGAAATGCCAGACAGGTGTTCAACGCCCA
GAAAAAGCTCAGAGCAGCACCGATTGCCTGAGCTAC

[0787] GATACCGAGATCCTGACCGTGAATACGGCTTCCTGCCTATCGGCAAGATCGTCGAGGAACGGATCGAG
TGCACAGTGACACCGTGGATAAGAATGGCTTCGTGTACACCCAGCCTATCGCTCAGTGGCACAACAGAGGCGAGCA
AGAGGTGTTTCGAGTACTGCCTGGAAGATGGCAGCATCATCCGGGCCACCAAGGACCACAAGTTTATGACCACCGACG
GCCAGATGCTGCCCATCGACGAGATCTTTGAGAGAGGCCTGGACCTGAAACAGGTGGACGGACTGCCTTGA

[0788] ABE7.10_Cfa (GEP) -C_断裂_S303C:

[0789] ATGGTCAAGATCATCAGCAGAAAGAGCCTGGGCACCCAGAACGTGTACGATATCGGAGTGGGCGAGCCC
CACAACCTTTCTGCTCAAGAATGGCCTGGTGGCCAGCAACTGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGC
ACCTCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGATCTGACCCTGCTGAAGGCCCTCGTTAGAC
AGCAGCTGCCAGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCGGCTACATTGATGGCGGA
GCCAGCCAAGAGGAATTCTACAAGTTCATCAAGCCCATCCTCGAGAAGATGGACGGCACCGAGGAACCTGCTGGTCAA
GCTGAACAGAGAGGACCTGCTGAGAAAGCAGAGAACCTTCGACAACGGCAGCATCCCTCACCAGATCCACCTGGGAG
AACTGCACGCCATTCTGCGGAGACAAGAGGACTTTTACCCATTCTGAAGGACAACCGGGAAAAGATCGAGAAAATC
CTGACCTTCAGGATCCCCTACTACGTGGGACCACTGGCCAGAGGCAATAGCAGATTCGCTGGATGACCAGAAAGAG
CGAGGAAACCATCACTCCCTGGAACCTTCGAGGAAGTGGTGGACAAGGGCGCCAGCGCTCAGTCCTTCATCGAGCGGA
TGACCAACTTCGATAAGAACCTGCCTAACGAGAAGGTGCTGCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTG
TACAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTTCTGAGCGGCGAGCAGAAAAA

GGCCATCGTGGATCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAA
TCGAGTGCTTCGACAGCGTCGAGATCTCCGGCGTGGAAGATCGGTTCAATGCCAGCCTGGGCACATACCACGATCTG
CTGAAAATTATCAAGGACAAGGACTTCCTGGACAACGAAGAGAACGAGGACATCCTTGAGGACATCGTGCTGACACT
GACCCTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACATACGCCCACCTGTTGACGACAAAAGTGATGA
AGCAACTGAAGCGGCGGAGATACACCGGCTGGGGCAGACTGTCTCGGAAGCTGATCAACGGCATCCGGGATAAGCAG
TCCGGCAAGACCATCCTGGACTTTCTGAAGTCCGACGGCTTCGCCAACAGAACTTCATGCAGCTGATTACGACGA
CAGCCTCACCTTCAAAGAGGATATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATTCTCTGCATGAGCACATTGCCA
A

[0790] CCTGGCCGGCTCTCCCGCCATTAAGAAAGGCATCCTGCAGACAGTGAAGGTGGTGGACGAGCTTGTGAA
AGTGATGGGCAGACACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACACAGAAGGGACAGA
AGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCC
GTGGAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGGACGGGATATGTACGTGGACCA
AGAGCTGGACATCAACAGACTGTCCGACTACGATGTGGACCATATCGTGCCCCAGTCTTTTCTGAAGGACGACTCCA
TCGACAACAAGGTCTTGACCAGATCCGACAAGAATCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTGGTCAAG
AAGATGAAGAACTACTGGCGACAGCTGCTGAACGCCAAGCTGATTACCCAGCGGAAGTTCGACAATCTGACCAAGGC
CGAAAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGCAGATCACAA
AGCACGTGGCACAGATTCTGGACTCTCGGATGAACACTAAGTACGACGAGAACGACAACTGATCCGCGAAGTGAAA
GTCATCACCTGAAGTCCAAGCTGGTGTCCGATTTCGGGAAGGATTTCAGTTCTACAAAGTGCGCGAGATCAACAA
CTACCATCACGCCACGACGCCTACCTGAATGCCGTTGTTGGAACAGCCCTGATCAAAAAGTACCCTAAGCTGGAAA
GCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAGAGCGAGCAAGAGATTGGCAAG
GCAACCGCCAAGTACTTCTTCTACAGCAACATCATGAACTTTTCAAGACAGAGATCACCTCGCCAACGGCGAGAT
CAGAAAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGCGAGATTGTGTGGGATAAGGGCAGAGACTTTGCCACAG
TGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAGAAAACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAG
TCTATCCTGCCTAAGCGGAACTCCGACAAGCTGATCGCCAGAAAGAAGGACTGGGACCCCAAGAAGTACGGCGGCTT
CGATTCTCCTACCGTGGCCTATAGCGTGCTGGTGGTGGCCAAAGTGGAAGGGCAAGTCCAAGAACTCAAGAGCG
TGAAAGAGCTGCTGGGGATCACCATCATGGAAGAAGCAGCTTCGAGAAGAATCCGATCGATTTCCTCGAGGCCAAG
GGCTACAAAGAAGTAAAAAGGACCTGATCATCAAGCTCCCCAAGTACTCCCTGTTGAGCTGGAACCGGCCGGAA
GAGAATGCTGGCCTCTGCTGGCGAACTGCAGAAGGGAAACGAACTGGCCCTGCCTAGCAAATATGTGAACTTCTGT
ACCTGGCCAGCCACTATGAGAAGCTGAAGGGCAGCCCCGAGGACAATGAGCAAAAGCAGCTGTTTGTGGAACAGCAC
AAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTTAGCAAGAGAGTGATTCTGGCCGACGCCAATCTGGA
CAAAGTGCTGTCCGCCTACAACAAGCACCGGGACAAGCCTATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTA
CCCTGACCAACCTGGGAGCCCCTGCCGCCTTCAAGTA

[0791] CTTTGACACCACCATCGACCGGAAGCGGTACACCTCCACCAAAGAGGTGCTGGACGCCACTCTGATCCA
CCAGTCTATCACCGGCCTGTACGAGACACGGATCGACCTGTCTCAACTCGGAGGCGACGAAGGCGCCGATAAGAGAA
CCGCCGATGGCTCTGAGTTCGAGAGCCCCAAGAAAAAGCGCAAAGTGTA

[0792] ABE7.10_Cfa (GEP) -C_断裂_T310C:

[0793] ATGGTCAAGATCATCAGCAGAAAGAGCCTGGGCACCCAGAACGTGTACGATATCGGAGTGGGCGAGCCC
CACAACCTTTCTGCTCAAGAATGGCCTGGTGGCCAGCAACTGCGAGATACCAAGGCACCTCTGAGCGCCTCTATGAT
CAAGAGATACGACGAGCACCACCAGGATCTGACCCTGCTGAAGGCCCTCGTTAGACAGCAGCTGCCAGAGAAGTACA

AAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCGGCTACATTGATGGCGGAGCCAGCCAAGAGGAATTCTAC
AAGTTCATCAAGCCCATCCTCGAGAAGATGGACGGCACCAGGAACTGCTGGTCAAGCTGAACAGAGAGGACCTGCT
GAGAAAGCAGAGAACCTTCGACAACGGCAGCATCCCTCACCAGATCCACCTGGGAGAACTGCACGCCATTCTGCGGA
GACAAGAGGACTTTTACCCATTCTTGAAGGACAACCGGGAAAAGATCGAGAAAATCCTGACCTTCAGGATCCCCTAC
TACGTGGGACCACTGGCCAGAGGCAATAGCAGATTTCGCTGGATGACCAGAAAGAGCGAGGAAACCATCACTCCCTG
GAACTTCGAGGAAGTGGTGGACAAGGGCGCCAGCGCTCAGTCCCTTCATCGAGCGGATGACCAACTTCGATAAGAACC
TGCCTAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTACAACGAGCTGACCAAAGTG
AAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTTCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGATCTGCTGTT
CAAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACAGCGTCG
AGATCTCCGGCGTGGAAGATCGGTTCAATGCCAGCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAG
GACTTCCTGGACAACGAAGAGAACGAGGACATCCTTGAGGACATCGTGCTGACACTGACCCTGTTTGAGGACAGAGA
GATGATCGAGGAACGGCTGAAAACATACGCCACCTGTTTCGACGACAAAGTGATGAAGCAACTGAAGCGGCGGAGAT
ACACCGGCTGGGGCAGACTGTCTCGGAAGCTGATCAACGGCATCCGGGATAAGCAGTCCGGCAAGACCATCCTGGAC
TTTCTGAAGTCCGACGGCTTCGCCAACAGAACTTCATGCAGCTGATTCACGACGACAGCCTCACCTTCAAAGAGGA
TATCCAGAAAGCCAGGTGTCCGGCCAGGGCGATTCTCTGCATGAGCACATTGCCAACCTGGCCGGCTCTCCCGCCA
TTAAGAAAGGCATCCTGCAGACAGTGAAGGTGGTGGACGAGCTTGTGAAAGTGATGGGCAGACACAAGCCCGAGAAC
ATCGTGATCGAAATGGCCAGAGAGAACCA

[0794] GACCACACAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCT
GGGCAGCCAGATCCTGAAAGAACACCCCGTGAAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGC
AGAATGGACGGGATATGTACGTGGACCAAGAGCTGGACATCAACAGACTGTCCGACTACGATGTGGACCATATCGTG
CCCCAGTCTTTTCTGAAGGACGACTCCATCGACAACAAGGTCCTGACCAGATCCGACAAGAATCGGGGCAAGAGCGA
CAACGTGCCCTCCGAAGAGGTGGTCAAGAAGATGAAGAACTACTGGCGACAGCTGCTGAACGCCAAGCTGATTACCC
AGCGGAAGTTGACAATCTGACCAAGGCCGAAAGAGGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGA
CAGCTGGTGGAAACCCGGCAGATCACAAGCACGTGGCACAGATTCTGGACTCTCGGATGAACACTAAGTACGACGA
GAACGACAAACTGATCCGCAAGTGAAAGTCATCACCTGAAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCC
AGTTCTACAAAGTGCGCGAGATCAACAACCTACCATCACGCCACGACGCCTACCTGAATGCCGTTGTTGGAACAGCC
CTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGAT
CGCCAAGAGCGAGCAAGAGATTGGCAAGGCAACCGCCAAGTACTTCTTCTACAGCAACATCATGAACTTTTTCAAGA
CAGAGATCACCTCGCCAACGGCGAGATCAGAAAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGCGAGATTGTG
TGGGATAAGGGCAGAGACTTTGCCACAGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAGAAAACCGA
GGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCTAAGCGGAACTCCGACAAGCTGATCGCCAGAAAGAAGG
ACTGGGACCCCCAAGAAGTACGGCGGCTTCGATTCTCCTACCGTGGCCTATAGCGTGCTGGTGGTGGCCAAAGTGGA
AAGGGCAAGTCCAAGAACTCAAGAGCGTGAAAGAGCTGCTGGGGATCACCATCATGGAAGAAGCAGCTTCGAGAA
GAATCCGATCGATTTCTCGAGGCCAAGGGCTACAAAGAAGTAAAAAGGACCTGATCATCAAGCTCCCCAAGTACT
CCCTGTTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCTGGCGAACTGCAGAAGGGAAACGAACTGGCC
CTGCCTAGCAAATATGTGAACTTCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCAGCCCCGAGGACAATGA
GCAAAAGCAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTTAGCAAGA
GAGTGATTCTGGCCGACGCCAATCTGGACAAAGTGCTGTCCGCTACAACAAGCACCGGACAAGCCTATCAGAGAG
CAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAACCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCAC

CATCGACCGGAAGCGGTACACCTCCACCAAAGAGGTGCTGGACGCCACTCTGATCCACCAGTCTATCACCGGCCTGT
ACGAGACACGGATCGACCTGTCTCAACTCGGAGGCGACG

[0795] AAGGCGCCGATAAGAGAACCGCCGATGGCTCTGAGTTCGAGAGCCCCAAGAAAAAGCGCAAAGTGTGA

[0796] ABE7.10_Cfa (GEP) -C_断裂_T313C:

[0797] ATGGTCAAGATCATCAGCAGAAAGAGCCTGGGCACCCAGAACGTGTACGATATCGGAGTGGGCGAGCCC
CACAACCTTCTGCTCAAGAATGGCCTGGTGGCCAGCAACTGCAAGGCACCTCTGAGCGCCTCTATGATCAAGAGATA
CGACGAGCACCACCAGGATCTGACCCTGTGAAGGCCCTCGTTAGACAGCAGCTGCCAGAGAAGTACAAAGAGATTT
TCTTCGACCAGAGCAAGAACGGCTACGCCGGCTACATTGATGGCGGAGCCAGCCAAGAGGAATTCTACAAGTTCATC
AAGCCCATCCTCGAGAAGATGGACGGCACCAGGAACTGCTGGTCAAGCTGAACAGAGAGGACCTGCTGAGAAAGCA
GAGAACCTTCGACAACGGCAGCATCCCTCACCAGATCCACCTGGGAGAACTGCACGCCATTCTGCGGAGACAAGAGG
ACTTTTACCCATTCTGAAGGACAACCGGAAAAAGATCGAGAAAATCCTGACCTTCAGGATCCCCTACTACGTGGGA
CCACTGGCCAGAGGCAATAGCAGATTGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACTCCCTGGAACCTTCGA
GGAAGTGGTGGACAAGGGCGCCAGCGCTCAGTCCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCTAACG
AGAAGGTGCTGCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTACAACGAGCTGACCAAAGTGAAATACGTG
ACCGAGGGAATGAGAAAGCCCGCCTTTCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGATCTGCTGTTCAAGACCAA
CCGGAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACAGCGTCGAGATCTCCG
GCGTGAAGATCGGTTCAATGCCAGCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTG
GACAACGAAGAGAACGAGGACATCCTTGAGGACATCGTGCTGACACTGACCCTGTTTGAGGACAGAGAGATGATCGA
GGAACGGCTGAAAACATACGCCACCTGTTTCGACGACAAAAGTGATGAAGCAACTGAAGCGGCGGAGATACACCGGT
GGGGCAGACTGTCTCGGAAGCTGATCAACGGCATCCGGGATAAGCAGTCCGGCAAGACCATCCTGGACTTTCTGAAG
TCCGACGGCTTCGCCAACAGAACTTCATGCAGCTGATTCACGACGACAGCCTCACCTTCAAAGAGGATATCCAGAA
AGCCCAGGTGTCCGGCCAGGGCGATTCTCTGCATGAGCACATTGCCAACCTGGCCGGCTCTCCCGCCATTAAGAAAG
GCATCCTGCAGACAGTGAAGGTGGTGGACGAGCTTGTGAAAGTGATGGGCAGACACAAGCCCCGAGAACATCGTGATC
GAAATGGCCAGAGAGAACCAGACCACACAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGG
CATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGAAAAACACCCAG

[0798] CTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGGACGGGATATGTACGTGGACCAAGAGCTG
GACATCAACAGACTGTCCGACTACGATGTGGACCATATCGTGCCCCAGTCTTTTCTGAAGGACGACTCCATCGACAA
CAAGGTCCTGACCAGATCCGACAAGAATCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTGGTCAAGAAGATGA
AGAACTACTGGCGACAGCTGCTGAACGCCAAGCTGATTACCCAGCGGAAGTTCGACAATCTGACCAAGGCCGAAAGA
GGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGCAGATCACAAAGCACGT
GGCACAGATTCTGGACTCTCGGATGAACACTAAGTACGACGAGAACGACAACTGATCCGCGAAGTGAAAGTCATCA
CCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTCCAGTTCTACAAAGTGC GCGAGATCAACAACTACCAT
CACGCCCACGACGCCTACCTGAATGCCGTTGTTGGAACAGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTT
CGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAGAGCGAGCAAGAGATTGGCAAGGCAACCG
CCAAGTACTTCTTCTACAGCAACATCATGAACTTTTCAAGACAGAGATCACCTCGCCAACGGCGAGATCAGAAAG
CGGCCTCTGATCGAGACAAACGGCGAAACCGGCGAGATTGTGTGGGATAAGGGCAGAGACTTTGCCACAGTGC GGAA
AGTGCTGAGCATGCCCCAAGTGAATATCGTGAAGAAAACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCC
TGCCTAAGCGGAACTCCGACAAGCTGATCGCCAGAAAGAAGGACTGGGACCCCAAGAAGTACGGCGGCTTCGATTCT
CCTACCGTGGCCTATAGCGTGCTGGTGGTGGCCAAAGTGAAAAAGGGCAAGTCCAAGAACTCAAGAGCGTGAAAGA

GCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCGATCGATTTCCTCGAGGCCAAGGGCTACA
AAGAAGTGAAAAAGGACCTGATCATCAAGCTCCCCAAGTACTCCCTGTTTCGAGCTGGAAAACGGCCGGAAGAGAATG
CTGGCCTCTGCTGGCGAACTGCAGAAGGGAAACGAACTGGCCCTGCCTAGCAAATATGTGAACTTCCTGTACCTGGC
CAGCCACTATGAGAAGCTGAAGGGCAGCCCCGAGGACAATGAGCAAAAGCAGCTGTTTGTGGAACAGCACAAGCACT
ACCTGGACGAGATCATCGAGCAGATCAGCGAGTTTAGCAAGAGAGTGATTCTGGCCGACGCCAATCTGGACAAAGTG
CTGTCCGCCTACAACAAGCACCGGGACAAGCCTATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGAC
CAACCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGCGGTACACCTCCACCAAAGAGG
TGCTGGACGCCACTCTGATCCACCAGTCTATCACCGGCCTGTACGAGACACGGATCGACCTGTCTCAACTCGGAGGC
GACGAAGGCGCCGATAAGAGAACCGCCGATGGCTCTGAGTTCGAGAGCCCCAAGAAAAAGCGCAAAGTGTGA

[0799] ABE7.10_Cfa (GEP) -C_断裂_S355C:

[0800] ATGGTCAAGATCATCAGCAGAAAGAGCCTGGGCACCCAGAACGTGTACGATATCGGAGTGGGCGAGCCC
CACAACCTTTCTGCTCAAGAATGGCCTGGTGGCCAGCAACTGCAAGAACGGCTACGCCGGCTACATTGATGGCGGAGC
CAGCCAAGAGGAATTCTACAAGTTCATCAAGCCATCCTCGAGAAGATGGACGGCACCGAGGAAGTCTGGTCAAGC
TGAACAGAGAGGACCTGCTGAGAAAGCAGAGAACCTTCGACAACGGCAGCATCCCTCACCAGATCCACCTGGGAGAA
CTGCACGCCATTCTGCGGAGACAAGAGGACTTTTACCCATTCTGAAGGACAACCGGGAAAAGATCGAGAAAATCCT
GACCTTCAGGATCCCCTACTACGTGGGACCACTGGCCAGAGGCAATAGCAGATTCGCCTGGATGACCAGAAAGAGCG
AGGAAACCATCACTCCCTGGAACCTTCGAGGAAGTGGTGGACAAGGGCGCCAGCGCTCAGTCCTTCATCGAGCGGATG
ACCAACTTCGATAAGAACCTGCCTAACGAGAAGGTGCTGCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTA
CAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCTGAGCGGCGAGCAGAAAAAGG
CCATCGTGGATCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATC
GAGTGCTTCGACAGCGTCGAGATCTCCGGCGTGGAAGATCGGTTCAATGCCAGCCTGGGCACATACCACGATCTGCT
GAAAATTATCAAGGACAAGGACTTCCTGGACAACGAAGAGAACGAGGACATCCTTGAGGACATCGTGCTGACACTGA
CCCTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACATACGCCCACCTGTTTCGACGACAAAGTGATGAAG
CAACTGAAGCGGCGGAGATACACCGGCTGGGGCAGACTGTCTCGGAAGCTGATCAACGGCATCCGGGATAAGCAGTC
CGGCAAGACCATCCTGGACTTTCTGAAGTCCGACGGCTTCGCCAACAGAACTTCATGCAGCTGATTACGACGACA
GCCTCACCTTCAAAGAGGATATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATTCTCTGCATGAGCACATTGCCAAC
CTGGCCGGCTCTCCCGCCATTAAGAAAGGCATCCTGCAGACAGTGAAGGTGGTGGACGAGCTTGTGAAAGTGATGGG
CAGACACAAGCCCAGAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACACAGAAGGGACAGAAGAACAGCC
GCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGAAAAAC
ACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGGACGGGATATGTACGTGGACCAAGAGCTGGA
CATCAACAGACTGTCCGACTACGATGTGGACCATATCGTGCCCCAGTCTTTTCTGAAGGACGACTCCATCGACAACA
AGGTCCTGACCAGATCCGACAAGAATCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTGGTCAAGAAGATGAAG
AACTA

[0801] CTGGCGACAGCTGCTGAACGCCAAGCTGATTACCCAGCGGAAGTTCGACAATCTGACCAAGGCCGAAAG
AGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGCAGATCACAAAGCACG
TGGCACAGATTCTGGACTCTCGGATGAACACTAAGTACGACGAGAACGACAAACTGATCCGCGAAGTGAAAGTCATC
ACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTCTACAAAGTGCGCGAGATCAACAACTACCA
TCACGCCCACGACGCCTACCTGAATGCCGTTGTTGGAACAGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGT
TCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAGAGCGAGCAAGAGATTGGCAAGGCAACC

GCCAAGTACTTCTTCTACAGCAACATCATGAACTTTTTCAAGACAGAGATCACCTCGCCAACGGCGAGATCAGAAA
GCGGCCTCTGATCGAGACAAACGGCGAAACCGGCGAGATTGTGTGGGATAAGGGCAGAGACTTTGCCACAGTGCGGA
AAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAGAAAACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATC
CTGCCTAAGCGGAACTCCGACAAGCTGATCGCCAGAAAGAAGGACTGGGACCCCAAGAAGTACGGCGGCTTCGATTC
TCCTACCGTGGCCTATAGCGTGCTGGTGGTGCCAAAGTGGAAAAGGGCAAGTCCAAGAACTCAAGAGCGTGAAAG
AGCTGCTGGGGATCACCATCATGGAAAAGCAGCTTCGAGAGAATCCGATCGATTCCTCGAGGCCAAGGGCTAC
AAAGAAGTGAAAAGGACCTGATCATCAAGCTCCCCAAGTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAAT
GCTGGCCTCTGCTGGCGAACTGCAGAAGGGAAACGAACTGGCCCTGCCTAGCAAATATGTGAACTTCCTGTACCTGG
CCAGCCACTATGAGAAGCTGAAGGGCAGCCCCGAGGACAATGAGCAAAAGCAGCTGTTTGTGGAACAGCACAAGCAC
TACCTGGACGAGATCATCGAGCAGATCAGCGAGTTTAGCAAGAGAGTGATTCTGGCCGACGCCAATCTGGACAAAGT
GCTGTCCGCCTACAACAAGCACCCGGGACAAGCCTATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGA
CCAACCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGCGGTACACCTCCACCAAAGAG
GTGCTGGACGCCACTCTGATCCACCAGTCTATCACCGCCTGTACGAGACACGGATCGACCTGTCTCAACTCGGAGG
CGACGAAGGCGCCGATAAGAGAACCGCCGATGGCTCTGAGTTCGAGAGCCCCAAGAAAAAGCGCAAAGTGTGA

[0802] ABE7.10_Cfa (GEP) -C_断裂_A456C:

[0803] ATGGTCAAGATCATCAGCAGAAAGAGCCTGGGCACCCAGAACGTGTACGATATCGGAGTGGGCGAGCCC
CACAACCTTTCTGCTCAAGAATGGCCTGGTGGCCAGC

[0804] AACTGCAGAGGCAATAGCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACTCCCTGGAAC
TTCGAGGAAGTGGTGGACAAGGGCGCCAGCGCTCAGTCCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCC
TAACGAGAAGGTGCTGCCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTACAACGAGCTGACCAAAGTGAAT
ACGTGACCGAGGGAATGAGAAAAGCCCGCCTTTCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGATCTGCTGTTCAAG
ACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACAGCGTCGAGAT
CTCCGGCGTGGAAGATCGGTTCAATGCCAGCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACT
TCCTGGACAACGAAGAGAACGAGGACATCCTTGAGGACATCGTGCTGACACTGACCCTGTTTGAGGACAGAGAGATG
ATCGAGGAACGGCTGAAAACATACGCCCACCTGTTTCGACGACAAAGTGATGAAGCAACTGAAGCGGCGGAGATACAC
CGGCTGGGGCAGACTGTCTCGGAAGCTGATCAACGGCATCCGGGATAAGCAGTCCGGCAAGACCATCCTGGACTTTC
TGAAGTCCGACGGCTTCGCCAACAGAACTTCATGCAGCTGATTACGACGACAGCCTCACCTTCAAAGAGGATATC
CAGAAAGCCCAGGTGTCCGGCCAGGGCGATTCTCTGCATGAGCACATTGCCAACCTGGCCGGCTCTCCCGCCATTAA
GAAAGGCATCCTGCAGACAGTGAAGGTGGTGGACGAGCTTGTGAAAGTGATGGGCAGACACAAGCCCGAGAACATCG
TGATCGAAATGGCCAGAGAGAACCAGACCACACAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAA
GAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAACACCCAGCTGCAGAACGAGAAGCT
GTACCTGTACTACCTGCAGAATGGACGGGATATGTACGTGGACCAAGAGCTGGACATCAACAGACTGTCCGACTACG
ATGTGGACCATATCGTGCCCCAGTCTTTTCTGAAGGACGACTCCATCGACAACAAGTCTTGACCAGATCCGACAAG
AATCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTGGTCAAGAAGATGAAGAACTACTGGCGACAGCTGCTGAA
CGCCAAGCTGATTACCCAGCGGAAGTTCGACAATCTGACCAAGGCCGAAAGAGGCGGCCTGAGCGAACTGGATAAGG
CCGGCTTCATCAAGAGACAGCTGGTGGAAAACCCGGCAGATCAGAAAGCACGTGGCACAGATTCTGGACTCTCGGATG
AACACTAAGTACGACGAGAACGACAACTGATCCGCGAAGTGAAAGTCATCACCTGAAGTCCAAGCTGGTGTCCGA
TTTCCGGAAGGATTTCCAGTTCTACAAAGTGCGCGAGATCAACAACTACCATCACGCCACGACGCCTACCTGAATG
CCGTTGTTGGAACAGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTAC

GACGTGCGGAAGATGATCGCCAAGAGCGAGCAAGAGATTGGCAAGGCAACCGCCAAGTACTTCTTCTACAGCAACAT
CATGAACCTTTTCAAGACAGAGATCACCTCGCCAACGG

[0805] CGAGATCAGAAAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGCGAGATTGTGTGGGATAAGGGCAG
AGACTTTGCCACAGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAGAAAACCGAGGTGCAGACAGGCG
GCTTCAGCAAAGAGTCTATCCTGCCTAAGCGGAACCTCGACAAGCTGATCGCCAGAAAGAAGGACTGGGACCCCAAG
AAGTACGGCGGCTTCGATTCTCTACCGTGGCCTATAGCGTGCTGGTGGTGGCCAAAGTGGAAGGGCAAGTCCAA
GAAACTCAAGAGCGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCGATCGATT
TCCTCGAGGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTCCCCAAGTACTCCCTGTTTCGAGCTG
GAAACCGGCCGAAGAGAATGCTGGCCTCTGCTGGCGAACTGCAGAAGGGAAACGAACTGGCCCTGCCTAGCAAATA
TGTGAACTTCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCAGCCCCGAGGACAATGAGCAAAAGCAGCTGT
TTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTTAGCAAGAGAGTGATTCTGGCC
GACGCCAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGACAAGCCTATCAGAGAGCAGGCCGAGAATAT
CATCCACCTGTTTACCCTGACCAACCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGC
GGTACACCTCCACCAAGAGGTGCTGGACGCCACTCTGATCCACCAGTCTATCACCGGCCTGTACGAGACACGGATC
GACCTGTCTCAACTCGGAGGCGACGAAGGCGCCGATAAGAGAACCGCCGATGGCTCTGAGTTCGAGAGCCCCAAGAA
AAAGCGCAAAGTGTGA

[0806] ABE7.10_Cfa (GEP) -C_断裂_S460C:

[0807] ATGGTCAAGATCATCAGCAGAAAGAGCCTGGGCACCCAGAACGTGTACGATATCGGAGTGGGCGAGCCC
CACAACCTTCTGCTCAAGAATGGCCTGGTGGCCAGCAACTGCAGATTGCCTGGATGACCAGAAAGAGCGAGGAAAC
CATCACTCCCTGGAACCTTCGAGGAAGTGGTGGACAAGGGCGCCAGCGCTCAGTCCTTCATCGAGCGGATGACCAACT
TCGATAAGAACCTGCCTAACGAGAAGGTGCTGCCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTACAACGAG
CTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCTGAGCGGCGAGCAGAAAAAGGCCATCGT
GGATCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCT
TCGACAGCGTCGAGATCTCCGGCGTGGAAGATCGGTTCAATGCCAGCCTGGGCACATACCACGATCTGCTGAAAATT
ATCAAGGACAAGGACTTCTTGACAACGAAGAGAACGAGGACATCCTTGAGGACATCGTGCTGACACTGACCCTGTT
TGAGGACAGAGAGATGATCGAGGAAC

[0808] GGCTGAAAACATACGCCCACCTGTTTCGACGACAAAGTGATGAAGCAACTGAAGCGGCGGAGATACACCG
GCTGGGGCAGACTGTCTCGGAAGCTGATCAACGGCATCCGGGATAAGCAGTCCGGCAAGACCATCTGGACTTTCTG
AAGTCCGACGGCTTCGCCAACAGAACTTCATGCAGCTGATTCACGACGACAGCCTCACCTTCAAAGAGGATATCCA
GAAAGCCCAGGTGTCCGGCCAGGGCGATTCTCTGCATGAGCACATTGCCAACCTGGCCGGCTCTCCCGCCATTAAGA
AAGGCATCCTGCAGACAGTGAAGGTGGTGGACGAGCTTGTGAAAGTGATGGGCAGACACAAGCCCGAGAACATCGTG
ATCGAAATGGCCAGAGAGAACCAGACCACACAGAAGGGACAGAAGAAGCCGCGAGAGAATGAAGCGGATCGAAGA
GGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACCCCCGTGGAAGAACCCAGCTGCAGAACGAGAAGCTGT
ACCTGTACTACCTGCAGAATGGACGGGATATGTACGTGGACCAAGAGCTGGACATCAACAGACTGTCCGACTACGAT
GTGGACCATATCGTGCCCCAGTCTTTTCTGAAGGACGACTCCATCGACAACAAGTCTTGACCAGATCCGACAAGAA
TCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTGGTCAAGAAGATGAAGAACTACTGGCGACAGCTGCTGAACG
CCAAGCTGATTACCCAGCGGAAGTTCGACAATCTGACCAAGGCCGAAAGAGGCGGCCTGAGCGAACTGGATAAGGCC
GGCTTCATCAAGAGACAGCTGGTGGAAACCCGGCAGATCACAAGCACGTGGCACAGATTCTGGACTCTCGGATGAA
CACTAAGTACGACGAGAACGACAAACTGATCCGCGAAGTGAAAGTCATCACCTGAAGTCCAAGCTGGTGTCCGATT

TCCGGAAGGATTTCCAGTTCTACAAAGTGCGGAGATCAACAACCTACCATCACGCCCACGACGCCTACCTGAATGCC
GTTGTTGGAACAGCCCTGATCAAAAAGTACCCTAAGCTGGAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGA
CGTGCGGAAGATGATCGCCAAGAGCGAGCAAGAGATTGGCAAGGCAACCGCCAAGTACTTCTTCTACAGCAACATCA
TGAACCTTTTTCAAGACAGAGATCACCTCGCCAACGGCGAGATCAGAAAGCGGCCTCTGATCGAGACAAACGGCGAA
ACCGGCGAGATTGTGTGGGATAAGGGCAGAGACTTTGCCACAGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATAT
CGTGAAGAAAACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCTAAGCGGAACCTCCGACAAGCTGA
TCGCCAGAAAGAAGGACTGGGACCCCAAGAAGTACGGCGGCTTCGATTCTCCTACCGTGGCCTATAGCGTGCTGGTG
GTGGCCAAAGTGGAAGGGCAAGTCCAAGAAACTCAAGAGCGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAG
AAGCAGCTTCGAGAAGAATCCGATCGATTTCTCGAGGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCA
AGCTCCCCAAGTACTCCCTGTTTCGAGCTGGAACCGCCGGAAGAGAATGCTGGCCTCTGCTGGCGAACTGCAGAAG
GGAAACGAACTGGCCCTGCCTAGCAAATATGTGAACTTC

[0809] CTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCAGCCCCGAGGACAATGAGCAAAAGCAGCTGTTT
GTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTTAGCAAGAGAGTGATTCTGGCCGA
CGCCAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGACAAGCCTATCAGAGAGCAGGCCGAGAATATCA
TCCACCTGTTTACCCTGACCAACCTGGGAGCCCCTGCCGCCTCAAGTACTTTGACACCACCATCGACCGGAAGCGG
TACACCTCCACCAAGAGGTGCTGGACGCCACTCTGATCCACCAGTCTATCACCGGCCTGTACGAGACACGGATCGA
CCTGTCTCAACTCGGAGGCGACGAAGGCGCCGATAAGAGAACCGCCGATGGCTCTGAGTTCGAGAGCCCCAAGAAAA
AGCGCAAAGTGTGA

[0810] ABE7.10_Cfa (GEP) -C_断裂_A463C:

[0811] ATGGTCAAGATCATCAGCAGAAAGAGCCTGGGCACCCAGAACGTGTACGATATCGGAGTGGGCGAGCCC
CACAACCTTTCTGCTCAAGAATGGCCTGGTGGCCAGCAACTGCTGGATGACCAGAAAGAGCGAGGAAACCATCACTCC
CTGGAACCTTCGAGGAAGTGGTGGACAAGGGCGCCAGCGCTCAGTCCTTCATCGAGCGGATGACCAACTTCGATAAGA
ACCTGCCTAACGAGAAGGTGCTGCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTACAACGAGCTGACCAAA
GTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTTCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGATCTGCT
GTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACAGCG
TCGAGATCTCCGGCGTGGAAGATCGGTTCAATGCCAGCCTGGGCACATACCAGATCTGCTGAAAATTATCAAGGAC
AAGGACTTCCTGGACAACGAAGAGAACGAGGACATCCTTGAGGACATCGTGCTGACACTGACCCTGTTTGAGGACAG
AGAGATGATCGAGGAACGGCTGAAAACATACGCCACCTGTTCGACGACAAAGTGATGAAGCAACTGAAGCGGCGGA
GATACACCGGCTGGGGCAGACTGTCTCGGAAGCTGATCAACGGCATCCGGGATAAGCAGTCCGGCAAGACCATCCTG
GACTTTCTGAAGTCCGACGGCTTCGCCAACAGAACTTCATGCAGCTGATTACGACGACAGCCTCACCTTCAAAGA
GGATATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATTCTCTGCATGAGCACATTGCCAACCTGGCCGGCTCTCCCG
CCATTAAGAAAGGCATCCTGCAGACAGTGAAGGTGGTGGACGAGCTTGTGAAAGTGATGGGCAGACACAAGCCCGAG
AACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACACAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCG
GATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAAGAACCCAGCTGCAGAACG
AGAAGCTGTACCTGTACTACCTGCAGAATGGACGGGATATGTACGTGGACCAAGAGCTGGACATCAACA

[0812] GACTGTCCGACTACGATGTGGACCATATCGTGCCCCAGTCTTTTCTGAAGGACGACTCCATCGACAACA
AGGTCTTGACCAGATCCGACAAGAATCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTGGTCAAGAAGATGAAG
AACTACTGGCGACAGCTGCTGAACGCCAAGCTGATTACCCAGCGGAAGTTCGACAATCTGACCAAGGCCGAAAGAGG
CGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTGAAACCCGGCAGATCACAAAGCACGTGG

CACAGATTCTGGACTCTCGGATGAACACTAAGTACGACGAGAACGACAACTGATCCGCGAAGTGAAAGTCATCACC
CTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTCTACAAAGTGC GCGAGATCAACA ACTACCATCA
CGCCCACGACGCCTACCTGAATGCCGTTGTTGGAACAGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCG
TGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAGAGCGAGCAAGAGATTGGCAAGGCAACCGCC
AAGTACTTCTTCTACAGCAACATCATGAACTTTTTCAAGACAGAGATCACCTCGCCAACGGCGAGATCAGAAAGCG
GCCTCTGATCGAGACAAACGGCGAAACCGGCGAGATTGTGTGGGATAAGGGCAGAGACTTTGCCACAGTGC GGAAG
TGCTGAGCATGCCCCAAGTGAATATCGTGAAGAAAACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTG
CCTAAGCGGAACTCCGACAAGCTGATCGCCAGAAAAGGACTGGGACCCCAAGAAGTACGGCGGCTTCGATTCTCC
TACCGTGGCCTATAGCGTGCTGGTGGTGGCCAAAGTGAAAAAGGGCAAGTCCAAGAACTCAAGAGCGTGAAAGAGC
TGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCGATCGATTTCTCGAGGCCAAGGGCTACAAA
GAAGTGAAAAAGGACCTGATCATCAAGCTCCCCAAGTACTCCCTGTTTCGAGCTGGAAAACGGCCGGAAGAGAATGCT
GGCCTCTGCTGGCGAACTGCAGAAGGGAAACGAACTGGCCCTGCCTAGCAAATATGTGAACTTCCTGTACCTGGCCA
GCCACTATGAGAAGCTGAAGGGCAGCCCCGAGGACAATGAGCAAAAGCAGCTGTTTGTGGAACAGCACAAGCACTAC
CTGGACGAGATCATCGAGCAGATCAGCGAGTTTAGCAAGAGAGTGATTCTGGCCGACGCCAATCTGGACAAAGTGCT
GTCCGCCTACAACAAGCACCGGGACAAGCCTATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCA
ACCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGCGGTACACCTCCACCAAAGAGGTG
CTGGACGCCACTCTGATCCACCAGTCTATCACCGGCTGTACGAGACACGGATCGACCTGTCTCAACTCGGAGGCGA
CGAAGGCGCCGATAAGAGAACCGCCGATGGCTCTGAGTTCGAGAGCCCCAAGAAAAAGCGCAAAGTGTGA

[0813] ABE7.10_Cfa (GEP) -C_断裂_T446C:

[0814] ATGGTCAAGATCATCAGCAGAAAGAGCCTGGGCACCCAGAACGTGTACGATATCGGAGTGGGCGAGCCC
CACA ACTTTCTGCTCAAGAATGGCCTGGTGGCCAGCAACTGCAGAAAGAGCGAGGAAACCATCACTCCCTGGA ACTT
CGAGGAAGTGGTGGACAAGGGCGCCAGCGCTCAGTCCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCTA
ACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTACAACGAGCTGACCAAAGTGAAATAC
GTGACCGAGGGAATGAGAAAGCCCGCCTTTCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGATCTGCTGTTCAAGAC
CAACCGGAAAGTGACCGTGAAGCAGCTGAAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACAGCGTCGAGATCT
CCGGCGTGGAAGATCGGTTCAATGCCAGCCTGGGCACATACCAGATCTGCTGAAAATTATCAAGGACAAGGACTTC
CTGGACAACGAAGAGAACGAGGACATCCTTGAGGACATCGTGCTGACACTGACCCTGTTTGAGGACAGAGAGATGAT
CGAGGAACGGCTGAAAACATACGCCACCTGTTTCGACGACAAAAGTGATGAAGCAACTGAAGCGGCGGAGATACACCG
GCTGGGGCAGACTGTCTCGGAAGCTGATCAACGGCATCCGGGATAAGCAGTCCGGCAAGACCATCCTGGACTTTCTG
AAGTCCGACGGCTTCGCCAACAGAACTTCATGCAGCTGATTCACGACGACAGCCTCACCTTCAAAGAGGATATCCA
GAAAGCCCAGGTGTCCGGCCAGGGCGATTCTCTGCATGAGCACATTGCCAACCTGGCCGGCTCTCCCGCCATTAAGA
AAGGCATCCTGCAGACAGTGAAGGTGGTGGACGAGCTTGTGAAAGTGATGGGCAGACACAAGCCCGAGAACATCGTG
ATCGAAATGGCCAGAGAGAACCAGACCACACAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGA
GGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACCCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGT
ACCTGTACTACCTGCAGAATGGACGGGATATGTACGTGGACCAAGAGCTGGACATCAACAGACTGTCCGACTACGAT
GTGGACCATATCGTGCCCCAGTCTTTTCTGAAGGACGACTCCATCGACAACAAGTCTTGACCAGATCCGACAAGAA
TCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTGGTCAAGAAGATGAAGAACTACTGGCGACAGCTGCTGAACG
CCAAGCTGATTACCCAGCGGAAGTTCGACAATCTGACCAAGGCCGAAAGAGGCGGCCTGAGCGAACTGGATAAGGCC
GGCTTCATCAAGAGACAGCTGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATTCTGGACTCTCGGATGAA

CACTAAGTACGACGAGAACGACAAACTGATCCGCGAAGTGAAAGTCATCACCCCTGAAGTCCAAGCTGGTGTCCGATT
TCCGGAAGGATTTCCAGTTCTACAAAGTGC GCGAGATCAACAACTACCATCACGCCCACGACGCCTACCTGAATGCC
GTTGTTGGAACAGCCCTGATCAAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGA
CGTGCGGAAGATGATCGCCAAGAGCGAGCAAGAGATTGG

[0815] CAAGGCAACCGCCAAGTACTTCTTCTACAGCAACATCATGAACTTTTTCAAGACAGAGATCACCCCTCGC
CAACGGCGAGATCAGAAAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGCGAGATTGTGTGGGATAAGGGCAGAG
ACTTTGCCACAGTGC GGAAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAGAAAACCGAGGTGCAGACAGGCGGC
TTCAGCAAAGAGTCTATCCTGCCTAAGCGGAACTCCGACAAGCTGATCGCCAGAAAGAAGGACTGGGACCCCAAGAA
GTACGGCGGCTTCGATTCTCTACCGTGGCCTATAGCGTGCTGGTGGTGGCCAAAGTGGAAGGGCAAGTCCAAGA
AACTCAAGAGCGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCGATCGATTTC
CTCGAGGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTCCCCAAGTACTCCCTGTTTCGAGCTGGA
AAACGGCCGGAAGAGAATGCTGGCCTCTGCTGGCGAACTGCAGAAGGGAAACGAAGTGGCCCTGCCTAGCAAATATG
TGAAGTTCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCAGCCCCGAGGACAATGAGCAAAAGCAGCTGTTT
GTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTTAGCAAGAGAGTGATTCTGGCCGA
CGCCAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGACAAGCCTATCAGAGAGCAGGCCGAGAATATCA
TCCACCTGTTTACCCTGACCAACCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGCGG
TACACCTCCACCAAGAGGTGCTGGACGCCACTCTGATCCACCAGTCTATCACCGGCCTGTACGAGACACGGATCGA
CCTGTCTCAACTCGGAGGCGACGAAGGCGCCGATAAGAGAACCGCCGATGGCTCTGAGTTCGAGAGCCCCAAGAAAA
AGCGCAAAGTGTGA

[0816] ABE7.10_Cfa (GEP) -C_断裂_T469C:

[0817] ATGGTCAAGATCATCAGCAGAAAGAGCCTGGGCACCCAGAACGTGTACGATATCGGAGTGGGCGAGCCC
CACAAGTTTCTGCTCAAGAATGGCCTGGTGGCCAGCAACTGCGAGGAAACCATCACTCCCTGGAAGTTCGAGGAAGT
GGTGGACAAGGGCGCCAGCGCTCAGTCCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCTAACGAGAAGG
TGCTGCCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTACAACGAGCTGACCAAAGTGAAATACGTGACCGAG
GGAATGAGAAAGCCCGCCTTTCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGATCTGCTGTTCAAGACCAACCGGAA
AGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACAGCGTCGAGATCTCCGGCGTGG
AAGATCGGTTCAATGCCAGCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAAC
GAAGAGAACGAGGACATCCTTGAGGACATCGTGCTGACACTGACCCTGTTTG

[0818] AGGACAGAGAGATGATCGAGGAACGGCTGAAAACATACGCCCACCTGTTTCGACGACAAAGTGATGAAGC
AACTGAAGCGGCGGAGATACACCGGCTGGGGCAGACTGTCTCGGAAGCTGATCAACGGCATCCGGGATAAGCAGTCC
GGCAAGACCATCCTGGACTTTCTGAAGTCCGACGGCTTCGCCAACAGAACTTCATGCAGCTGATTCACGACGACAG
CCTCACCTTCAAAGAGGATATCCAGAAAGCCAGGTGTCGGGCCAGGGCGATTCTCTGCATGAGCACATTGCCAACC
TGGCCGGCTCTCCCGCCATTAAGAAAGGCATCCTGCAGACAGTGAAGGTGGTGGACGAGCTTGTGAAAGTGATGGGC
AGACACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACAGAAAGGGACAGAAGAACAGCCG
CGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGAAAAACA
CCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAAATGGACGGGATATGTACGTGGACCAAGAGCTGGAC
ATCAACAGACTGTCCGACTACGATGTGGACCATATCGTGCCCCAGTCTTTTCTGAAGGACGACTCCATCGACAACAA
GGTCCTGACCAGATCCGACAAGAATCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTGGTCAAGAAGATGAAGA
ACTACTGGCGACAGCTGCTGAACGCCAAGCTGATTACCCAGCGGAAGTTCGACAATCTGACCAAGGCCGAAAGAGGC

GGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGCAGATCACAAAGCACGTGGC
ACAGATTCTGGACTCTCGGATGAACACTAAGTACGACGAGAACGACAACTGATCCGCGAAGTGAAAGTCATCACCC
TGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTCTACAAAGTGCGCGAGATCAACAACTACCATCAC
GCCCACGACGCCTACCTGAATGCCGTTGTTGGAACAGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCTG
GTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAGAGCGAGCAAGAGATTGGCAAGGCAACCGCCA
AGTACTTCTTCTACAGCAACATCATGAACTTTTTCAAGACAGAGATCACCCCTCGCCAACGGCGAGATCAGAAAGCGG
CCTCTGATCGAGACAAACGGCGAAACCGGCGAGATTGTGTGGGATAAGGGCAGAGACTTTGCCACAGTGCGGAAAGT
GCTGAGCATGCCCCAAGTGAATATCGTGAAGAAAACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGC
CTAAGCGGAACTCCGACAAGCTGATCGCCAGAAAAGAGGACTGGGACCCCAAGAAGTACGGCGGCTTCGATTCTCTCT
ACCGTGGCCTATAGCGTGCTGGTGGTGGCCAAAGTGGAAGGGCAAGTCCAAGAACTCAAGAGCGTGAAAGAGCT
GCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCGATCGATTTCTCGAGGCCAAGGGCTACAAAG
AAGTGAAAAAGGACCTGATCATCAAGCTCCCCAAGTACTCCCTGTTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTG
GCCTCTGCTGGCGAACTGCAGAAGGGAAACGAACTGGC

[0819] CCTGCCTAGCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCAGCCCCGA
GGACAATGAGCAAAAGCAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGT
TTAGCAAGAGAGTGATTCTGGCCGACGCCAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGACAAGCCT
ATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAACCTGGGAGCCCCTGCCGCCTTCAAGTACTT
TGACACCACCATCGACCGGAAGCGGTACACCTCCACCAAAGAGGTGCTGGACGCCACTCTGATCCACCAGTCTATCA
CCGGCCTGTACGAGACACGGATCGACCTGTCTCAACTCGGAGGCGACGAAGGCGCCGATAAGAGAACCGCCGATGGC
TCTGAGTTCGAGAGCCCCAAGAAAAAGCGCAAAGTGTTGA

[0820] ABE7.10_Cfa (GEP) -C_断裂_T472C:

[0821] ATGGTCAAGATCATCAGCAGAAAGAGCCTGGGCACCCAGAACGTGTACGATATCGGAGTGGGCGAGCCC
CACAACTTTCTGCTCAAGAATGGCCTGGTGGCCAGCAACTGCATCACTCCCTGGAACCTCGAGGAAGTGGTGACAA
GGGCGCCAGCGCTCAGTCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCTAACGAGAAGGTGCTGCCCA
AGCACAGCCTGCTGTACGAGTACTTCACCGTGTACAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGA
AAGCCCGCCTTTCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGATCTGCTGTTCAAGACCAACCGGAAAGTGACCGT
GAAGCAGCTGAAAGAGGACTACTTCAAGAAAAATCGAGTGCTTCGACAGCGTCGAGATCTCCGGCGTGGAAGATCGGT
TCAATGCCAGCCTGGGCACATACCAGATCTGCTGAAAAATTATCAAGGACAAGGACTTCCTGGACAACGAAGAGAAC
GAGGACATCCTTGAGGACATCGTGCTGACACTGACCCTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAAC
ATACGCCACCTGTTTCGACGACAAAGTGATGAAGCAACTGAAGCGGCGGAGATACACCGGCTGGGGCAGACTGTCTC
GGAAGCTGATCAACGGCATCCGGGATAAGCAGTCCGGCAAGACCATCCTGGACTTTCTGAAGTCCGACGGCTTCGCC
AACAGAACTTCATGCAGCTGATTCACGACGACAGCCTCACCTTCAAAGAGGATATCCAGAAAGCCCAGGTGTCCGG
CCAGGGCGATTCTCTGCATGAGCACATTGCCAACCTGGCCGGCTCTCCCGCCATTAAGAAAGGCATCCTGCAGACAG
TGAAGGTGGTGGACGAGCTTGTGAAAGTGATGGGCAGACACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAG
AACCAGACCACACAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGG
CAGCCAGATCCTGAAAGAACACCCCGTGGAACACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGA
ATGGACGGGATATGTACG

[0822] TGGACCAAGAGCTGGACATCAACAGACTGTCCGACTACGATGTGGACCATATCGTGCCCCAGTCTTTTC
TGAAGGACGACTCCATCGACAACAAGGTCCTGACCAGATCCGACAAGAATCGGGGCAAGAGCGACAACGTGCCCTCC

GAAGAGGTGGTCAAGAAGATGAAGAACTACTGGCGACAGCTGCTGAACGCCAAGCTGATTACCCAGCGGAAGTTCGA
CAATCTGACCAAGGCCGAAAGAGGCGGCCTGAGCGAACTGGATAAGGCCGCTTCATCAAGAGACAGCTGGTGGAAA
CCCGGCAGATCACAAAGCACGTGGCACAGATTCTGGACTCTCGGATGAACACTAAGTACGACGAGAACGACAAACTG
ATCCGCGAAGTGAAAGTCATCACCCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTCTACAAAGT
GCGCGAGATCAACAACCTACCATCACGCCACGACGCCTACCTGAATGCCGTTGTTGGAACAGCCCTGATCAAAAAGT
ACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAGAGCGAG
CAAGAGATTGGCAAGGCAACCGCCAAGTACTTCTTCTACAGCAACATCATGAACTTTTTCAAGACAGAGATCACCCCT
CGCCAACGGCGAGATCAGAAAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGCGAGATTGTGTGGGATAAGGGCA
GAGACTTTGCCACAGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAGAAAACCGAGGTGCAGACAGGC
GGCTTCAGCAAAGAGTCTATCCTGCCTAAGCGGAACTCCGACAAGCTGATCGCCAGAAAGAAGGACTGGGACCCCAA
GAAGTACGGCGGCTTCGATTCTCTACCGTGGCCTATAGCGTGCTGGTGGTGGCCAAAGTGGAAGGGCAAGTCCA
AGAACTCAAGAGCGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCGATCGAT
TTCTCGAGGCCAAGGGCTACAAAGAAGTAAAAAGGACCTGATCATCAAGCTCCCCAAGTACTCCCTGTTTCGAGCT
GGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCTGGCGAACTGCAGAAGGGAAACGAACTGGCCCTGCCTAGCAAAT
ATGTGAACTTCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCAGCCCCGAGGACAATGAGCAAAAGCAGCTG
TTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTTAGCAAGAGAGTGATTCTGGC
CGACGCCAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGACAAGCCTATCAGAGAGCAGGCCGAGAATA
TCATCCACCTGTTTACCCTGACCAACCTGGGAGCCCCCTGCCGCTTCAAGTACTTTGACACCACCATCGACCGGAAG
CGGTACACCTCCACCAAAGAGGTGCTGGACGCCACTCTGATCCACCAGTCTATCACCAGCCTGTACGAGACACGGAT
CGACCTGTCTCAACTCGGAGGCGACGAAGGCGCCGATAAGAGAACCGCCGATGGCTCTGAGTTCGAGAGCCCCAAGA
AAAAGCGCAAAGTGTGA

[0823] ABE7.10_Cfa (GEP) -C_断裂_T474C:

[0824] ATGGTCAAGATCATCAGCAGAAAGAGCCTGGGCACCCAGAACGTGTACGATATCGGAGTGGGCGAGCCC
CACAACCTTTCTGCTCAAGAATGGCCTGGTGGCCAGCAACTGCCCCCTGGAACCTCGAGGAAGTGGTGGACAAGGGCGC
CAGCGCTCAGTCCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCTAACGAGAAGGTGCTGCCCAAGCACA
GCCTGCTGTACGAGTACTTCACCGTGTACAACGAGCTGACCAAAGTGAATACGTGACCGAGGGAATGAGAAAGCCC
GCCTTTCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGATCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCA
GCTGAAAGAGGACTACTTCAAGAAAAATCGAGTGCTTCGACAGCGTCGAGATCTCCGGCGTGGAAGATCGGTTCAATG
CCAGCCTGGGCACATACCACGATCTGCTGAAAAATTATCAAGGACAAGGACTTCCTGGACAACGAAGAGAACGAGGAC
ATCCTTGAGGACATCGTGCTGACACTGACCTGTTTGAAGACAGAGAGATGATCGAGGAACGGCTGAAAACATACGC
CCACCTGTTTCGACGACAAAGTGATGAAGCAACTGAAGCGGCGGAGATACACCGGCTGGGGCAGACTGTCTCGGAAGC
TGATCAACGGCATCCGGGATAAGCAGTCCGGCAAGACCATCCTGGACTTTCTGAAGTCCGACGGCTTCGCCAACAGA
AACTTCATGCAGCTGATTCACGACGACAGCCTCACCTTCAAAGAGGATATCCAGAAAGCCCAGGTGTCCGGCCAGGG
CGATTCTCTGCATGAGCACATTGCCAACCTGGCCGGCTCTCCCGCCATTAAGAAAGGCATCCTGCAGACAGTGAAGG
TGGTGGACGAGCTTGTGAAAGTGATGGGCAGACACAAGCCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACCAG
ACCACACAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCA
GATCCTGAAAGAACACCCCGTGAAAAACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGGAC
GGGATATGTACGTGGACCAAGAGCTGGACATCAACAGACTGTCCGACTACGATGTGGACCATATCGTGCCCCAGTCT
TTTCTGAAGGACGACTCCATCGACAACAAGGTCCTGACCAGATCCGACAAGAATCGGGGCAAGAGCGACAACGTGCC

CTCCGAAGAGGTGGTCAAGAAGATGAAGAACTACTGGCGACAGCTGCTGAACGCCAAGCTGATTACCCAGCGGAAGT
TCGACAATCTGACCAAGGCCGAAAGAGGCGGCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTG
GAAACCCGGCAGATCACAAAGCACGTGGCACAGATTCTGGACTCTCGGATGAACACTAAGTACGACGAGAACGACAA
ACTGATCCGCGAAGTGAAAATCATCACCCCTGAAGTCCAAGCTGGTGTCCGATTTCGGAAGGATTTCAGTTCTACA
AAGTGC GCGAGATCAACAACCTACCATCACGCCACGACGCCTACCTGAATGCCGTTGTTGGAACAGCCCTGATCAAA
AAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGG

[0825] AAGATGATCGCCAAGAGCGAGCAAGAGATTGGCAAGGCAACCGCCAAGTACTTCTTCTACAGCAACATC
ATGAACTTTTTCAAGACAGAGATCACCCCTCGCCAACGGCGAGATCAGAAAGCGGCCTCTGATCGAGACAAACGGCGA
AACC GGCGAGATTGTGTGGGATAAGGGCAGAGACTTTGCCACAGTGGCGAAAGTGCTGAGCATGCCCCAAGTGAATA
TCGTGAAGAAAACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCTAAGCGGAACTCCGACAAGCTG
ATCGCCAGAAAGAAGGACTGGGACCCCCAAGAAGTACGGCGGCTTCGATTCTCCTACCGTGGCCTATAGCGTGTGGT
GGTGGCCAAAGTGGAAGGGCAAGTCCAAGAACTCAAGAGCGTGAAAGAGCTGCTGGGGATCACCATCATGGAAA
GAAGCAGCTTCGAGAAGAATCCGATCGATTTCCTCGAGGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATC
AAGCTCCCCAAGTACTCCCTGTTTCGAGCTGGAAAAACGGCCGGAAGAGAATGCTGGCCTCTGCTGGCGAACTGCAGAA
GGGAAACGAACTGGCCCTGCCTAGCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCA
GCCCCGAGGACAATGAGCAAAAGCAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATC
AGCGAGTTTAGCAAGAGAGTGATTCTGGCCGACGCCAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACC GGGA
CAAGCCTATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAACCTGGGAGCCCCTGCCGCCTTCA
AGTACTTTGACACCACCATCGACCGGAAGCGGTACACCTCCACCAAAGAGGTGCTGGACGCCACTCTGATCCACCAG
TCTATCACCGGCCTGTACGAGACACGGATCGACCTGTCTCAACTCGGAGGCGACGAAGGCGCCGATAAGAGAACCGC
CGATGGCTCTGAGTTCGAGAGCCCCAAGAAAAAGCGCAAAAGTGTA

[0826] ABE7.10_Cfa (GEP) -C_断裂_C574C:

[0827] ATGGTCAAGATCATCAGCAGAAAGAGCCTGGGCACCCAGAACGTGTACGATATCGGAGTGGGCGAGCCC
CACAAC TTTCTGCTCAAGAATGGCCTGGTGGCCAGCAACTGCTTCGACAGCGTCGAGATCTCCGGCGTGGAAGATCG
GTTCAATGCCAGCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAACGAAGAGA
ACGAGGACATCCTTGAGGACATCGTGCTGACACTGACCCTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAA
ACATACGCCCACCTGTTTCGACGACAAAGTGATGAAGCAACTGAAGCGGCGGAGATACACCGGCTGGGGCAGACTGTC
TCGGAAGCTGATCAACGGCATCCGGGATAAGCAGTCCGGCAAGACCATCCTGGACTTTCTGAAGTCCGACGGCTTCG
CCAACAGAACTTCATGCAGCTGATTCACGACGACAGCCTCACCTTCAAAGAGGATATCCAGAAAGCCCAGGTGTC

[0828] CGGCCAGGGCGATTCTCTGCATGAGCACATTGCCAACCTGGCCGGCTCTCCCGCCATTAAGAAAGGCAT
CCTGCAGACAGTGAAGGTGGTGGACGAGCTTGTGAAAGTGATGGGCAGACACAAGCCCCGAGAACATCGTGATCGAAA
TGGCCAGAGAGAACCAGACCACACAGAAGGGACAGAAGAAGAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATC
AAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAAGAACCCAGCTGCAGAACGAGAAGCTGTACCTGTA
CTACCTGCAGAATGGACGGGATATGTACGTGGACCAAGAGCTGGACATCAACAGACTGTCCGACTACGATGTGGACC
ATATCGTGCCCCAGTCTTTTCTGAAGGACGACTCCATCGACAACAAGGTCCTGACCAGATCCGACAAGAATCGGGGC
AAGAGCGACAACGTGCCCTCCGAAGAGGTGGTCAAGAAGATGAAGAACTACTGGCGACAGCTGCTGAACGCCAAGCT
GATTACCCAGCGGAAGTTCGACAATCTGACCAAGGCCGAAAGAGGCGGCTGAGCGAACTGGATAAGGCCGGCTTCA
TCAAGAGACAGCTGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATTCTGGACTCTCGGATGAACACTAAG
TACGACGAGAACGACAACTGATCCGCGAAGTGAAAGTCATCACCCCTGAAGTCCAAGCTGGTGTCCGATTTCGGAA

GGATTTCAGTTCTACAAAGTGC GCGAGATCAACA ACTACCATCACGCCCACGACGCCTACCTGAATGCCGTTGTTG
GAACAGCCCTGATCAAAAAGTACCCTAAGCTGGAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGG
AAGATGATCGCCAAGAGCGAGCAAGAGATTGGCAAGGCAACCGCCAAGTACTTCTTCTACAGCAACATCATGAACTT
TTTCAAGACAGAGATCACCTCGCCAACGGCGAGATCAGAAAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGCG
AGATTGTGTGGGATAAGGGCAGAGACTTTGCCACAGTGC GGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAG
AAAACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCTAAGCGGAACTCCGACAAGCTGATCGCCAG
AAAGAAGGACTGGGACCCCCAAGAAGTACGGCGGCTTCGATTCTCCTACCGTGGCCTATAGCGTGCTGGTGGTGGCCA
AAGTGAAAAGGGCAAGTCCAAGAACTCAAGAGCGTGAAAGAGCTGCTGGGGATCACCATCATGGAAGAAGCAGC
TTCGAGAAGAATCCGATCGATTTCTCGAGGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTCCC
CAAGTACTCCCTGTTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCTGGCGAACTGCAGAAGGGAAACG
AACTGGCCCTGCCTAGCAAATATGTGAACTTCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCAGCCCCGAG
GACAATGAGCAAAAGCAGCTGTTTGTGGAACAGCACAAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTT
TAGCAAGAGAGTGATTCTGGCCGACGCCAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGACAAGCCTA
TCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACC

[0829] CTGACCAACCTGGGAGCCCCCTGCCGCTTCAAGTACTTTGACACCACCATCGACCGGAAGCGGTACACC
TCCACCAAAGAGGTGCTGGACGCCACTCTGATCCACCAGTCTATCACCGGCCTGTACGAGACACGGATCGACCTGTC
TCAACTCGGAGGCGACGAAGGCGCCGATAAGAGAACC GCCGATGGCTCTGAGTTCGAGAGCCCCAAGAAAAAGCGCA
AAGTGTGA

[0830] ABE7.10_Cfa (GEP) -C_断裂_S577C:

[0831] ATGGTCAAGATCATCAGCAGAAAGAGCCTGGGCACCCAGAACGTGTACGATATCGGAGTGGGCGAGCCC
CACAAC TTTCTGCTCAAGAATGGCCTGGTGGCCAGCAACTGCGTCGAGATCTCCGGCGTGGAAGATCGGTTCAATGC
CAGCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAACGAAGAGAACGAGGACA
TCCTTGAGGACATCGTGCTGACACTGACCCTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACATACGCC
CACCTGTTTCGACGACAAAGTGATGAAGCAACTGAAGCGGCGGAGATACACCGCTGGGGCAGACTGTCTCGGAAGCT
GATCAACGGCATCCGGGATAAGCAGTCCGGCAAGACCATCCTGGACTTTCTGAAGTCCGACGGCTTCGCCAACAGAA
ACTTCATGCAGCTGATTCACGACGACAGCCTCACCTTCAAAGAGGATATCCAGAAAGCCCAGGTGTCCGGCCAGGGC
GATTCTCTGCATGAGCACATTGCCAACCTGGCCGGCTCTCCCGCCATTAAGAAAGGCATCCTGCAGACAGTGAAGGT
GGTGGACGAGCTTGTGAAAGTGATGGGCAGACACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACCAGA
CCACACAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAG
ATCCTGAAAGAACACCCCGTGAAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGGACG
GGATATGTACGTGGACCAAGAGCTGGACATCAACAGACTGTCCGACTACGATGTGGACCATATCGTGCCCCAGTCTT
TTCTGAAGGACGACTCCATCGACAACAAGGTCTGACCAGATCCGACAAGAATCGGGGCAAGAGCGACAACGTGCCC
TCCGAAGAGGTGGTCAAGAAGATGAAGAACTACTGGCGACAGCTGCTGAACGCCAAGCTGATTACCCAGCGGAAGTT
CGACAATCTGACCAAGGCCGAAAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTGG
AAACCCGGCAGATCACAAGCACGTGGCACAGATTCTGGACTCTCGGATGAACACTAAGTACGACGAGAACGACAAA
CTGATCCGCGAAGTGAAAGTCATCACCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTCTACAA
AGTGCGCGAGATCAACA ACTACCATCACGCCCACGACGCCTACCTGAATGCCGTTGTTGGAACAGCCCTGATCAAAA
AGTACCCTAAGCTGGAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGA

[0832] CGTGCGGAAGATGATCGCCAAGAGCGAGCAAGAGATTGGCAAGGCAACCGCCAAGTACTTCTTCTACAG

CAACATCATGAACTTTTTCAAGACAGAGATCACCCCTCGCCAACGGCGAGATCAGAAAGCGGCCTCTGATCGAGACAA
ACGGCGAAACCGGCGAGATTGTGTGGGATAAGGGCAGAGACTTTGCCACAGTGCGGAAAGTGCTGAGCATGCCCCAA
GTGAATATCGTGAAGAAAACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCTAAGCGGAACTCCGA
CAAGCTGATCGCCAGAAAGAAGGACTGGGACCCCAAGAAGTACGGCGGCTTCGATTCTCCTACCGTGGCCTATAGCG
TGCTGGTGGTGGCCAAAGTGGAAGAGGGCAAGTCCAAGAACTCAAGAGCGTGAAAGAGCTGCTGGGGATCACCATC
ATGGAAAGAAGCAGCTTCGAGAAGAATCCGATCGATTTCCCTCGAGGCCAAGGGCTACAAAGAAGTGAAAAAGGACCT
GATCATCAAGCTCCCCAAGTACTCCCTGTTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCTGGCGAAC
TGCAGAAGGGAAACGAACTGGCCCTGCCTAGCAAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTG
AAGGGCAGCCCCGAGGACAATGAGCAAAAGCAGCTGTTTGTGGAACAGCACAAAGCACTACCTGGACGAGATCATCGA
GCAGATCAGCGAGTTTAGCAAGAGAGTGATTCTGGCCGACGCCAATCTGGACAAAGTGCTGTCCGCCTACAACAAGC
ACCGGGACAAGCCTATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAACCTGGGAGCCCCCTGCC
GCCTTCAAGTACTTTGACACCACCATCGACCGGAAGCGGTACACCTCCACCAAAGAGGTGCTGGACGCCACTCTGAT
CCACAGTCTATACCGGCCTGTACGAGACACGGATCGACCTGTCTCAACTCGGAGGCGACGAAGGCGCCGATAAGA
GAACCGCCGATGGCTCTGAGTTCGAGAGCCCCAAGAAAAAGCGCAAAGTGTGA

[0833] ABE7.10_Cfa (GEP) -C_断裂_A589C:

[0834] ATGGTCAAGATCATCAGCAGAAAGAGCCTGGGCACCCAGAACGTGTACGATATCGGAGTGGGCGAGCCC
CACAACCTTTCTGCTCAAGAATGGCCTGGTGGCCAGCAACTGCAGCCTGGGCACATACCACGATCTGCTGAAAATTAT
CAAGGACAAGGACTTCCTGGACAACGAAGAGAACGAGGACATCCTTGAGGACATCGTGCTGACACTGACCCTGTTTG
AGGACAGAGAGATGATCGAGGAACGGCTGAAAACATACGCCACCTGTTTCGACGACAAAGTGATGAAGCAACTGAAG
CGGCGGAGATACACCGGCTGGGGCAGACTGTCTCGGAAGCTGATCAACGGCATCCGGGATAAGCAGTCCGGCAAGAC
CATCCTGGACTTTCTGAAGTCCGACGGCTTCGCCAACAGAACTTCATGCAGCTGATTCACGACGACAGCCTCACCT
TCAAAGAGGATATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATTCTCTGCATGAGCACATTGCCAACCTGGCC

[0835] GGCTCTCCCGCCATTAAGAAAGGCATCCTGCAGACAGTGAAGGTGGTGGACGAGCTTGTGAAAGTGATG
GGCAGACACAAGCCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACACAGAAGGGACAGAAGAACAG
CCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACCCCCGTGGAAA
ACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGGACGGGATATGTACGTGGACCAAGAGCTG
GACATCAACAGACTGTCCGACTACGATGTGGACCATATCGTGCCCCAGTCTTTTCTGAAGGACGACTCCATCGACAA
CAAGGTCCTGACCAGATCCGACAAGAATCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTGGTCAAGAAGATGA
AGAACTACTGGCGACAGCTGCTGAACGCCAAGCTGATTACCCAGCGGAAGTTCGACAATCTGACCAAGGCCGAAAGA
GGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGCAGATCACAAAGCACGT
GGCACAGATTCTGGACTCTCGGATGAACACTAAGTACGACGAGAACGACAACTGATCCGCGAAGTGAAAGTCATCA
CCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTCTACAAAGTGC GCGAGATCAACAACTACCAT
CACGCCCACGACGCCTACCTGAATGCCGTTGTTGGAACAGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTT
CGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAGAGCGAGCAAGAGATTGGCAAGGCAACCG
CCAAGTACTTCTTCTACAGCAACATCATGAACTTTTTCAAGACAGAGATCACCCCTCGCCAACGGCGAGATCAGAAAG
CGGCCTCTGATCGAGACAAACGGCGAAACCGGCGAGATTGTGTGGGATAAGGGCAGAGACTTTGCCACAGTGCGGAA
AGTGCTGAGCATGCCCCAAGTGAATATCGTGAAGAAAACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCC
TGCCTAAGCGGAACTCCGACAAGCTGATCGCCAGAAAGAAGGACTGGGACCCCAAGAAGTACGGCGGCTTCGATTCT
CCTACCGTGGCCTATAGCGTGCTGGTGGTGGCCAAAGTGGAAGAGGGCAAGTCCAAGAACTCAAGAGCGTGAAAGA

GCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCGATCGATTTCCTCGAGGCCAAGGGCTACA
AAGAAGTGAAAAAGGACCTGATCATCAAGCTCCCCAAGTACTCCCTGTTTCGAGCTGGAAAACGGCCGAAGAGAATG
CTGGCCTCTGCTGGCGAACTGCAGAAGGGAAACGAACTGGCCCTGCCTAGCAAATATGTGAACTTCCTGTACCTGGC
CAGCCACTATGAGAAGCTGAAGGGCAGCCCCGAGGACAATGAGCAAAAGCAGCTGTTTGTGGAACAGCACAAGCACT
ACCTGGACGAGATCATCGAGCAGATCAGCGAGTTTAGCAAGAGAGTGATTCTGGCCGACGCCAATCTGGACAAAGTG
CTGTCCGCCTACAACAAGCACCGGGACAAGCCTATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGAC
CAACCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACAC

[0836] CACCATCGACCGGAAGCGGTACACCTCCACCAAAGAGGTGCTGGACGCCACTCTGATCCACCAGTCTAT
CACCGGCCTGTACGAGACACGGATCGACCTGTCTCAACTCGGAGGCGACGAAGGCGCCGATAAGAGAACCGCCGATG
GCTCTGAGTTCGAGAGCCCCAAGAAAAAGCGCAAAGTGTGA

[0837] ABE7.10_Cfa (GEP) -C_断裂_S590C:

[0838] ATGGTCAAGATCATCAGCAGAAAGAGCCTGGGCACCCAGAACGTGTACGATATCGGAGTGGGCGAGCCC
CACAACCTTTCTGCTCAAGAATGGCCTGGTGGCCAGCAACTGCCTGGGCACATACCACGATCTGCTGAAAATTATCAA
GGACAAGGACTTCCTGGACAACGAAGAGAACGAGGACATCCTTGAGGACATCGTGCTGACACTGACCCTGTTTGAGG
ACAGAGAGATGATCGAGGAACGGCTGAAAACATACGCCACCTGTTTCGACGACAAAGTGATGAAGCAACTGAAGCGG
CGGAGATACACCGGTGGGGCAGACTGTCTCGGAAGCTGATCAACGGCATCCGGGATAAGCAGTCCGGCAAGACCAT
CCTGGACTTTCTGAAGTCCGACGGCTTCGCCAACAGAACTTCATGCAGCTGATTACGACGACAGCCTCACCTTCA
AAGAGGATATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATTCTCTGCATGAGCACATTGCCAACCTGGCCGGCTCT
CCCGCCATTAAGAAAGGCATCCTGCAGACAGTGAAGGTGGTGGACGAGCTTGTGAAAGTGATGGGCAGACACAAGCC
CGAGAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACACAGAAGGGACAGAAGAACAGCCGCGAGAGAATGA
AGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGAAAAACACCCAGCTGCAG
AACGAGAAGCTGTACCTGTACTACCTGCAGAATGGACGGGATATGTACGTGGACCAAGAGCTGGACATCAACAGACT
GTCCGACTACGATGTGGACCATATCGTGCCCCAGTCTTTTCTGAAGGACGACTCCATCGACAACAAGGTCCTGACCA
GATCCGACAAGAATCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTGGTCAAGAAGATGAAGAACTACTGGCGA
CAGCTGCTGAACGCCAAGCTGATTACCCAGCGGAAGTTCGACAATCTGACCAAGGCCGAAAGAGGCGGCCTGAGCGA
ACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATTCTGG
ACTCTCGGATGAACACTAAGTACGACGAGAACGACAACTGATCCGCGAAGTGAAAGTCATCACCTGAAGTCCAAG
CTGGTGTCCGATTTCCGGAAGGATTTCCAGTTCTACAAAGTGCGCGAGATCAACAACCTACCATCACGCCACGACGC
CTACCTGAATGCCGTTGTTGGAACAGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACT
ACAAGGTGTACGACGTGCGGAAGATGATCGCCAAGAGCGAGCAAGAGATTGGCAAGGCAACCGCCAAGTACTTCTTC
TACAGCAACATCATGAACTTTTTCAAGAC

[0839] AGAGATCACCTCGCCAACGGCGAGATCAGAAAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGCGA
GATTGTGTGGGATAAGGGCAGAGACTTTGCCACAGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAGA
AAACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCTAAGCGGAACCTCCGACAAGCTGATCGCCAGA
AAGAAGGACTGGGACCCCAAGAAGTACGGCGGCTTCGATTCTCCTACCGTGGCCTATAGCGTGCTGGTGGTGCCAA
AGTGGAAGGGCAAGTCCAAGAACTCAAGAGCGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCT
TCGAGAAGAATCCGATCGATTTCTCGAGGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTCCCC
AAGTACTCCCTGTTTCGAGCTGGAAAACGGCCGAAGAGAATGCTGGCCTCTGCTGGCGAACTGCAGAAGGGAAACGA
ACTGGCCCTGCCTAGCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCAGCCCCGAGG

ACAATGAGCAAAAGCAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTT
AGCAAGAGAGTGATTCTGGCCGACGCCAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGACAAGCCTAT
CAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAACCTGGGAGCCCCTGCCGCCTTCAAGTACTTTG
ACACCACCATCGACCGGAAGCGGTACACCTCCACCAAAGAGGTGCTGGACGCCACTCTGATCCACCACTTATCACC
GGCCTGTACGAGACACGGATCGACCTGTCTCAACTCGGAGGCGACGAAGGCGCCGATAAGAGAACCGCCGATGGCTC
TGAGTTCGAGAGCCCCAAGAAAAAGCGCAAAGTGTGA

[0840] 完整无损的ABE7.10:

[0841] MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLS
DFFRMRRQEIKAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGSSSEVEFSHEYWMRHALTLAKRARDE
REVPVGAVLVHNNRVIGEGWNRPIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRV
VFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDSGGSSGGSSGSE
TPGTSESATPESSGGSSGSDKKYSIGLAIGTNSVGWAVITDEYKVPSSKKFKVLGNTDRHSIKKNLIGALLFDSGET
AEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTI
YHLRKKLVSDTKADRLIYLALAHMIKFRGHFLIEGDLNPDNSVDKLFQVLVQTYNQLFEENPINASGVDAKAIL
SARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDITYDDDLNLLA

[0842] QIGDQYADLFLAAKNLSAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKE
IFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQ
EDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRFAMTRKSEETITPWNFEVVDKGASAQSFIERMTNFDKNLP
NEKVLPHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEI
SGVEDRFNASLGTYHDLKIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYT
GWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA
IKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEELIKELGSQILKEHPVENTQLQNEKL
YLYYLQNGRDMYVDQELDINRLSDYDVHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLN
AKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSD
FRKDFQFYKVINNYHHAHDAYLNAVVGTA
LIKYPKLESEFVYGDYKVYDVRKMIKSEQEI
GKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL
SMPQVNI
VKKTEVQTGGFSKESILPKRNSDKLIARKKDWD
PKKYGGFDSPTVAYSVLVVAKEKGSKKLKSVKELLGITIMERS
SFEKNPIDFLEAKGYKEVKDLIIKLPKYS
LFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGS
PEDNEQKQLFVEQHKHYLDEII
EQISEFSKRVILADANLDKVL
SAYNKH
RDKPIREQAENI
IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL
DATLIHQ
SITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[0843] ABE7.10_Cfa-N_断裂_S303C:

[0844] MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLS
DFFRMRRQEIKAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGSSSEVEFSHEYWMRHALTLAKRARDE
REVPVGAVLVHNNRVIGEGWNRPIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRV
VFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDSGGSSGGSSGSE
TPGTSESATPESSGGSSGSDKKYSIGLAIGTNSVGWAVITDEYKVPSSKKFKVLGNTDRHSIKKNLIGALLFDSGET
AEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTI

YHLRKKLV DSTDKADLR LIYLALAHMIKFRGHFLIEGDLNPDNSDVKLF IQLVQTYNQLFEENPINASGVDAKAIL
SARLSKSRLENL I AQLPGEKKNGLFGNLI ALSLGLTPNFKSNFDLAEDAKLQLSKD TYDDDLNLLA

[0845] QIGDQYADLFLAAKNLSDA ILLCLS YDTEILTVEYGF LPIGKIVEERIECTVYTVDKNGFVYTQPIAQW
HNRGEQEVFEYCLEDGSIIRATKDHKFMTTDGQMLPIDEIFERGLDLKQVDGLP

[0846] ABE7.10_Cfa-N_断裂_T310C:

[0847] MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLS
DFFRMRRQEIKAKKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGSSSEVEFSHEYWMRHALTLAKRARDE
REVPVGAVLVHNNRVIGEGWNRALGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRV
VFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDSGGSSGGSSGSE
TPGTSESATPESSGGSSGSDKKYSIGLAIGTNSVGWAVITDEYKVP SKKFKVLGNTDRHSIKKNLIGALLFDSGET
AEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTI
YHLRKKLV DSTDKADLR LIYLALAHMIKFRGHFLIEGDLNPDNSDVKLF IQLVQTYNQLFEENPINASGVDAKAIL
SARLSKSRLENL I AQLPGEKKNGLFGNLI ALSLGLTPNFKSNFDLAEDAKLQLSKD TYDDDLNLLA QIGDQYADL
FLAAKNLSDA ILLSDILRVNCLS YDTEILTVEYGF LPIGKIVEERIECTVYTVDKNGFVYTQPIAQWHNRGEQEVFE
YCLEDGSIIRATKDHKFMTTDGQMLPIDEIFERGLDLKQVDGLP

[0848] ABE7.10_Cfa-N_断裂_T313C:

[0849] MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLS
DFFRMRRQEIKAKKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGSSSEVEFSHEYWMRHALTLAKRARDE
REVPVGAVLVHNNRVIGEGWNRALGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRV
VFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDSGGSSGGSSGSE
TPGTSESATPESSGGSSGSDKKYSIGLAIGTNSVGWAVITDEYKVP SKKFKVLGNTDRHSIKKNLIGALLFDSGET
AEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTI
YHLRKKLV DSTDKADLR LIYLALAHMIKFRGHFLIEGDLNPDNSDVKLF IQLVQTYNQLFEENPINASGVDAKAIL
SARLSKSRLENL I AQLPGEKKNGLFGNLI ALSLGLTPNFKSNFDLAEDAKLQLSKD TYDDDLNLLA

[0850] QIGDQYADLFLAAKNLSDA ILLSDILRVNTEICLS YDTEILTVEYGF LPIGKIVEERIECTVYTVDKNG
FVYTQPIAQWHNRGEQEVFEYCLEDGSIIRATKDHKFMTTDGQMLPIDEIFERGLDLKQVDGLP

[0851] ABE7.10_Cfa-N_断裂_S355C:

[0852] MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLS
DFFRMRRQEIKAKKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGSSSEVEFSHEYWMRHALTLAKRARDE
REVPVGAVLVHNNRVIGEGWNRALGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRV
VFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDSGGSSGGSSGSE
TPGTSESATPESSGGSSGSDKKYSIGLAIGTNSVGWAVITDEYKVP SKKFKVLGNTDRHSIKKNLIGALLFDSGET
AEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTI
YHLRKKLV DSTDKADLR LIYLALAHMIKFRGHFLIEGDLNPDNSDVKLF IQLVQTYNQLFEENPINASGVDAKAIL
SARLSKSRLENL I AQLPGEKKNGLFGNLI ALSLGLTPNFKSNFDLAEDAKLQLSKD TYDDDLNLLA QIGDQYADL
FLAAKNLSDA ILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQCLS YDTEILTVE

YGFLPIGKIVEERIECTVYTVDKNGFVYTQPIAQWHNRGEQEVFEYCLEDGSIIRATKDHKFMTTDQGMLPIDEIFE
RGLDLKQVDGLP

[0853] ABE7.10_Cfa-N_断裂_A456C:

[0854] MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLS
DFFRMRRQEIKAKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGSSSEVEFSHEYWMRHALTLAKRARDE
REVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRV
VFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDSGGSSGGSSGSE
TPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGET
AEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTI
YHLRKKLV DSTDKADRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAIL
SARLSKSRRLL

[0855] ENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFL
AAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGAS
QEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILT
FRIPYYVGPLCLSYDTEILTVEYGFLPIGKIVEERIECTVYTVDKNGFVYTQPIAQWHNRGEQEVFEYCLEDGSIIR
ATKDHKFMTTDQGMLPIDEIFERGLDLKQVDGLP

[0856] ABE7.10_Cfa-N_断裂_S460C:

[0857] MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLS
DFFRMRRQEIKAKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGSSSEVEFSHEYWMRHALTLAKRARDE
REVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRV
VFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDSGGSSGGSSGSE
TPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGET
AEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTI
YHLRKKLV DSTDKADRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAIL
SARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADL
FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGG
ASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKI
LTFRIPYYVGPLARGNCLSYDTEILTVEYGFLPIGKIVEERIECTVYTVDKNGFVYTQPIAQWHNRGEQEVFEYCLE
DGSIIIRATKDHKFMTTDQGMLPIDEIFERGLDLKQVDGLP

[0858] ABE7.10_Cfa-N_断裂_A463C:

[0859] MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLS
DFFRMRRQEIKAKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGSSSEVEFSHEYWMRHALTLAKRARDE
REVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRV
VFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDSGGSSGGSSGSE
TPGTSESATP

[0860] ESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAE

ATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYH
LRKKLV DSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLF IQLVQTYNQLFEENPINASGVDAKAILSA
RLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKD TYDDDLNLLAQIGDQYADLFL
AAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGAS
QEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILT
FRIPYYVGPLARGNSRFCLSYDTEILTVEYGFLPIGKIVEERIECTVYTVDKNGFVYTQPIAQWHNRGEQEVFEYCL
EDGSIIRATKDHKFM TTDGQMLPIDEIFERGLDLKQVDGLP

[0861] ABE7.10_Cfa-N_断裂_T466C:

[0862] MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLS
DFFRMRRQEIKAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGSSSEVEFSHEYWMRHALTLAKRARDE
REVPVGAVLVHNNRVIGEGWNRPIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRV
VFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDSGGSSGGSSGSE
TPGTSESATPESSGGSSGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGET
AEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTI
YHLRKKLV DSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLF IQLVQTYNQLFEENPINASGVDAKAIL
SARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKD TYDDDLNLLAQIGDQYADL
FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGG
ASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKI
LTFRIPYYVGPLARGNSRFAWMCLSYDTEILTVEYGFLPIGKIVEERIECTVYTVDKNGFVYTQPIAQWHNRGEQEV
FEYCLEGDSIIRATKDHKFM TTDGQMLPIDEIFERGLDLKQVDGLP

[0863] ABE7.10_Cfa-N_断裂_S469C:

[0864] MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAK

[0865] TGAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTDSGGSSGGSSGSE
TPGTSESATPESSGGSSGSSSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVHNNRVIGEGWNRPIGLHDPTAHAE
IMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEG
ILADECAALLCYFFRMPRQVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGSDKKYSIGLAIGTN
SVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMA
KVDDSFHRLSEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLV DSTDKADLRLIYLALAHMIKFRGHF
LIEGDLNPDNSDVKLF IQLVQTYNQLFEENPINASGVDAKAIL SARLSKSRLENLIAQLPGEKKNGLFGNLIALS
LGLTPNFKSNFDLAEDAKLQLSKD TYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMI
KRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLL
RKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKCLSYDTEI
LTVEYGFLPIGKIVEERIECTVYTVDKNGFVYTQPIAQWHNRGEQEVFEYCLEGDSIIRATKDHKFM TTDGQMLPID
EIFERGLDLKQVDGLP

[0866] ABE7.10_Cfa-N_断裂_T472C:

[0867] MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLS

DFFRMRRQEIKAQKKAQSSTDGSGSSGSSGSETPGTSESATPESSGGSSGSSSEVEFSHEYWMRHALTLAKRARDE
REVPVGAVLVNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRV
VFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGSGSSGSSGSE
TPGTSESATPESSGGSSGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGET
AEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTI
YHLRKKLV DSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAIL
SARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKS NFDAEDAKLQLSKD TYDDDLNLLAQIGDQYADL
FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGG
ASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKI
LTFRIPIYYVGPLARGNSRFAWMTRKSEEC LS YDTEILTVEYGFLPIGKIVEERIECTVYTVDKNGFVYTQPIAQWHN
RGEQEVFEYCLEGDSIIIRATKDHKFM TTDGQMLPIDEIFERGLDLKQV

[0868] DGLP

[0869] ABE7.10_Cfa-N_断裂_T474C:

[0870] MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLS
DFFRMRRQEIKAQKKAQSSTDGSGSSGSSGSETPGTSESATPESSGGSSGSSSEVEFSHEYWMRHALTLAKRARDE
REVPVGAVLVNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRV
VFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGSGSSGSSGSE
TPGTSESATPESSGGSSGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGET
AEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTI
YHLRKKLV DSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAIL
SARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKS NFDAEDAKLQLSKD TYDDDLNLLAQIGDQYADL
FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGG
ASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKI
LTFRIPIYYVGPLARGNSRFAWMTRKSEETICLSYDTEILTVEYGFLPIGKIVEERIECTVYTVDKNGFVYTQPIAQW
HNRGEQEVFEYCLEGDSIIIRATKDHKFM TTDGQMLPIDEIFERGLDLKQVDGLP

[0871] ABE7.10_Cfa-N_断裂_C574C:

[0872] MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLS
DFFRMRRQEIKAQKKAQSSTDGSGSSGSSGSETPGTSESATPESSGGSSGSSSEVEFSHEYWMRHALTLAKRARDE
REVPVGAVLVNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRV
VFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGSGSSGSSGSE
TPGTSESATPESSGGSSGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGET
AEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTI
YHLRKKLV DSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAIL
SARLSKSRR

[0873] ENLIAQLPGEKKNGLFGNLIALSLGLTPNFKS NFDAEDAKLQLSKD TYDDDLNLLAQIGDQYADLFL
AAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGAS
QEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILT

FRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYN
ELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECLSYDTEILTVEYGFLPIGKIVEERI
ECTVYTVDKNGFVYTQPIAQWHNRGEQEVFEYCLEDGSIIRATKDHKFM TTDGQMLPIDEIFERGLDLKQVDGLP

[0874] ABE7.10_Cfa-N_断裂_S577C:

[0875] MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLS
DFFRMRRQEIKAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGSSSEVEFSHEYWMRHALTLAKRARDE
REVPVGAVLVLNRRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRV
VFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDSGGSSGGSSGSE
TPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSSKKFKVLGNTDRHSIKKNLIGALLFDSGET
AEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTI
YHLRKKLV DSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAIL
SARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKD TYDDDLNLLAQIGDQYADL
FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGG
ASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHA ILRRQEDFYFPLKDNREKIEKI
LTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTV
YNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDCLSYDTEILTVEYGFLPIGKI
VEERIECTVYTVDKNGFVYTQPIAQWHNRGEQEVFEYCLEDGSIIRATKDHKFM TTDGQMLPIDEIFERGLDLKQVD
GLP

[0876] ABE7.10_Cfa-N_断裂_A589C:

[0877] MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLS
DFFRMRRQEIKAQKKAQSST

[0878] DSGGSSGGSSGSETPGTSESATPESSGGSSGSSSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLN
NRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGVRNAKTGA
GSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPES
SGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR
RYTRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLV DSTD
KADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLEN
LIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKD TYDDDLNLLAQIGDQYADLFLAAKNLSDAIL
LSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHA ILRRQEDFYFPLKDNREKIEKILTFRIPYYVG
PLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE
GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNCLSYDTEILTVEYGFLPIGKI
VEERIECTVYTVDKNGFVYTQPIAQWHNRGEQEVFEYCLEDGSIIRATKDHKFM TTDGQMLPIDEIFERGLDLKQVD
GLP

[0879] ABE7.10_Cfa-N_断裂_S590C:

[0880] MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLS

DDFRMRRQEIKAQKKAQSSTDSSGSSGGSSGSETPGTSESATPESSGGSSGSSSEVEFSHEYWMRHALTLAKRARDE
REVPVGAVLVNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRV
VFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDSSGSSGGSSGSE
TPGTSESATPESSGGSSGSDKKYSIGLAIGTNSVGWAVITDEYKVPSSKFKVLGNTDRHSIKKNLIGALLFDSGET
AEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTI
YHLRKKLV DSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAIL
SARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNDLAEDAKLQLSKD TYDDDLNLLAQIGDQYADL
FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGG
ASQEEFYKF IKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKI
LTFRIPYYVGPL

[0881] ARGNSRFAWMTRKSEETITPWNFEEVVDKGASASFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNEL
TKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNACLSYDTEILTVE
YGFLPIGKIVEERIECTVYTVDKNGFVYTQPIAQWHNRGEQEVFEYCLEDSIIRATKDHKFMTTDGQMLPIDEIFE
RGLDLKQVDGLP

[0882] ABE7.10_Cfa (GEP) -C_断裂_S303C:

[0883] MVKII SRKSLGTQNVYDIGVGEPHNFLKNGLVASNC DILRVNTEITKAPLSASMIKRYDEHHQDLTLL
KALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKF IKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPH
QIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKE
DYFKKIECFDSVEISGVEDRFNASLGT YHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLF
DDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDFANRNFMLIHDDSLTFKEDIQKAQVSGQDLSL
HEHIANLAGSPA IKKGILQTVKVVDLVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIE EG IKELGSQILK
EHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDV DHI V PQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE
VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIR
EVKVITLKSCLVSDFRKDFQFYKVR EINNYHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVYDVRKMIKSEQE
IGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNI VKKTEVQTGGF
SKESILPKRNSDKLIARKKDWDPKKG GFDSP TVAYSVLVAKVEKGSKKLKSVKELLGITIMERS SFEKNPIDFL
EAKGYKEVKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLGKSPEDNEQKQLFV
EQHKHYLDEIIEQISEFSKRVI LADANLDKVL SAYNKH RDKPIREQAENI IHLFTLTNLGAPAAF KYFDTTIDRKRY
TSTKEVL DATLIHQSI TGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[0884] ABE7.10_Cfa (GEP) -C_断裂_T310C:

[0885] MVKII SRKSLGTQNVYDIGVGEPHNFLKNGLVASNC EITKAPLSASMIKRYDEHHQDLTLLKALVRQQ
LPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKF IKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGEL
HAILRRQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASASFIERMT
NFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKV

[0886] TVKQLKEDYFKKIECFDSVEISGVEDRFNASLGT YHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFED
REMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDFANRNFMLIHDDSLTFK
EDIQKAQVSGQDLSLHEHIANLAGSPA IKKGILQTVKVVDLVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMK
RIE EG IKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDV DHI V PQSFLKDDSIDNKVLTR

SDKNRGKSDNPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILD
SRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKREINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDY
KVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL
SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKEKGKSKKLKSVKELLGIT
IMERSSEKPNIDFLEAKGYKEVKKDLIIKLPKYSLEFLENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEK
LKGSPEDNEQKQLFVEQHKHYLDEIEQISEFSKRVILADANLDKVL SAYNKH RDKPIREQAENIIHLFTLTNLGAP
AAFKYFDTTIDRKRYTSTKEVL DATLIHQ SITGLYETRIDLSQLGGDEGADKRTADGSEFESP KKKRKV

[0887] ABE7.10_Cfa (GEP) -C_断裂_T313C:

[0888] MVKII SRKSLGTQNVYDIGVGEPHFLKNGLVASNCKAPLSASMIKRYDEHHQDLTLLKALVRQQLPE
KYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAI
LRRQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEVVVDKGASQSFIERMTNFD
KNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLS GEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFD
SVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKR
RRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGS
PAIKKGILQTVKVVDELVKVMGRHKPENIVIAMARENQTTQKGQKNSRERMKRIE EG I KELGSQILKEHPVENTQLQ
NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNPSEEVVKKMKNYWR
QLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS
CLVSDFRKDFQFYKREINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFF
YSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRN
SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKEKGKSKKLKSVKELLGITIMERSSEKPNIDFLEAKGYKEVKK
DLIIKLPKYSLEFLENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKG

[0889] SPEDNEQKQLFVEQHKHYLDEIEQISEFSKRVILADANLDKVL SAYNKH RDKPIREQAENIIHLFTLT
NLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQ SITGLYETRIDLSQLGGDEGADKRTADGSEFESP KKKRKV

[0890] ABE7.10_Cfa (GEP) -C_断裂_S355C:

[0891] MVKII SRKSLGTQNVYDIGVGEPHFLKNGLVASNCKNGYAGYIDGGASQEEFYKFIKPILEKMDGTE
ELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRFAW
MTRKSEETITPWNFEVVVDKGASQSFIERMTNFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLS
GEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILED
IVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFML
LIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIIKKGILQTVKVVDELVKVMGRHKPENIVIAMARENQTTQK
GQKNSRERMKRIE EG I KELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLK
DSDIDNKVLTRSDKNRGKSDNPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQ
ITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKREINNYHHAHDAYLNAVVGTAIIKKYPK
LESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDF
ATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKEKGKSKKL
KSVKELLGITIMERSSEKPNIDFLEAKGYKEVKKDLIIKLPKYSLEFLENGRKRMLASAGELQKGNELALPSKYVN
FLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIEQISEFSKRVILADANLDKVL SAYNKH RDKPIREQAENIIH
LFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQ SITGLYETRIDLSQLGGDEGADKRTADGSEFESP KKKR
KV

[0892] ABE7.10_Cfa (GEP) -C_断裂_A456C:

[0893] MVKIIISRKSLGTQNVYDIGVGEPHNFLKNGLVASNCRGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSE

[0894] EVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLTKSLVSDFRKDFQFYKVVREINNYHHAHDAYLNAVVGITALIKKYPKLESEFVYGDYKVDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVAKVEKGSKKLKSVKELLGITIMERSSEFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLEFENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIIEQISEFSKRVIILADANLDKVL SAYNKHDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQSI TGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[0895] ABE7.10_Cfa (GEP) -C_断裂_S460C:

[0896] MVKIIISRKSLGTQNVYDIGVGEPHNFLKNGLVASNCRFAMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLTKSLVSDFRKDFQFYKVVREINNYHHAHDAYLNAVVGITALIKKYPKLESEFVYGDYKVDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVAKVEKGSKKLKSVKELLGITIMERSSEFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLEFENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPE DNEQKQLFVEQH KHYLDEIIIEQISEFSKRVIILADANLDKVL SAYNKHDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQSI TGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[0897] ABE7.10_Cfa (GEP) -C_断裂_A463C:

[0898] MVKIIISRKSLGTQNVYDIGVGEPHNFLKNGLVASNCWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRR

[0899] RYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLTKSLVSDFRKDFQFYKVVREINNYHHAHDAYLNAVVGITALIKKYPKLESEFVYGDYKVDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIKKTEVQTGGFSKE

SILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVAKVEKGKSKKLKSVKELLGITIMERSSSFENPIDFLEAK
GYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQH
KHYLDEIIIEQISEFSKRVIADANLDKVL SAYNKHDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTST
KEVLDA TLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPCKKKRKV

[0900] ABE7.10_Cfa (GEP) -C_断裂_T446C:

[0901] MVKII SRKSLGTQNVYDIGVGEPHNFLKNGLVASNCRKSEETITPWNFEVVVDKGASQSFIERMTNF
DKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECF
DSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLK
RRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFQMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAG
SPAIIKKGILQTVKVVDELVKVMGRHKPENIVIAMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYW
RQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS
KLVSDFRKDFQFYKVRINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQEI GKATAKYF
FYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNI VKKTEVQTGGFSKESILPKR
NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVAKVEKGKSKKLKSVKELLGITIMERSSSFENPIDFLEAKGYKEVK
KDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDE
IIIEQISEFSKRVIADANLDKVL SAYNKHDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDA
TLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPCKKKRKV

[0902] ABE7.10_Cfa (GEP) -C_断裂_T469C:

[0903] MVKII SRKSLGTQNVYDIGVGEPHNFLKNGLVASNCEETITPWNFEVVVDKGASQSFIERMTNFDKN
LPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSV
EISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRR
YTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFQMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA
IKKGILQTVKVVDELVKVMGRHKPENIVIAMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNE
KLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQL
LNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS
SDFRKDFQFYKVRINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQEI GKATAKYFFYS
NIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNI VKKTEVQTGGFSKESILPKRNSD
KLIARKKDWDPKKYGGFDSPTVAYSVLVAKVEKGKSKKLKSVKELLGITIMERSSSFENPIDFLEAKGYKEVKKDL
IIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEII
EQISEFSKRVIADANLDKVL SAYNKHDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDA
TLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPCKKKRKV

[0904] ABE7.10_Cfa (GEP) -C_断裂_T472C:

[0905] MVKII SRKSLGTQNVYDIGVGEPHNFLKNGLVASNCITPWNFEVVVDKGASQSFIERMTNFDKNLPN
EKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEIS
GVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTG
WGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFQMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIIK
GILQTVKVVDELVKVMGRHKPENIVIAMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLY
LYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNA

KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVR
EINNYHHAHDAYLNAVVGITALIKKYPKLESEFVYGDYKVDVRKMIKSEQEI GKATAKYFFYSNIMNFFKTEITLANGEIRKRPL
IETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLI ARKKDWDPKKYGGFDSPTVAYS
VLVAKVEKGKSKKLKSVKELLGITIMERSSEFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELNGRKRMLASAGELQKGNELAL
PSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIIEQISEFSKRVLADANLDKVL

[0906] AYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQSI TGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[0907] ABE7.10_Cfa (GEP) -C_断裂_T474C:

[0908] MVKII SRKSLGTQNVYDIGVGEPHNFLKNGLVASNCWFEEVVDKGASQSFIERMTNFDKNLPNEKVLPHKSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDR
FNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWRLSRKLINGIRDKQSGKT
ILDFLKSDGFANRNFQMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIIKGGILQTVKVVDELVKVMGRHKPENIV
IEMARENQTTQKGQKNSRERMKRIEEG IKELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDV
DHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQL
VETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVR EINNYHHAHDAYLNAVVGITALIKKYP
KLESEFVYGDYKVDVRKMIKSEQEI GKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL
SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVAKVEKGKSKKLKSVKELLGITIMERS
SEFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELNGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIIEQISEFSKRVLADANLDKVL SAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEV
LDATLIHQSI TGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[0909] ABE7.10_Cfa (GEP) -C_断裂_C574C:

[0910] MVKII SRKSLGTQNVYDIGVGEPHNFLKNGLVASNCVDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDI
LEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFQMLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIIKGGILQTVKVVDELVKVMGRHKPENIV IEMARENQTTQKGQKNSRERM
KRIEEG IKELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDV DHIVPQSFLKDDSIDNKVLTRSDKNR
GKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDEND
KLIREVKVITLKSCLVSDFRKDFQFYKVR EINNYHHAHDAYLNAVVGITALIKKYPKLESEFVYGDYKVDVRKMIKSEQEI GK
ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNS
DKLIARKKDWDPKKYGGFDSPTVAYSVLVAKVEKGKSKKLKSVKELLGITIMERSSEFEKNPIDFLEAKGYKEVKKDLIIKLPK
YSLFELNGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIIEQISEFSKRVLADAN
LDKVL SAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQSI TGLYETRIDLSQLGGDEGAD
KRTADGSEFESPKKKRKV

[0911] ITIMERSSEFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELNGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKL
KGSPEDNEQKQLFVEQHKHYLDEIIIEQISEFSKRVLADANLDKVL SAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTT
IDRKRYTSTKEVL DATLIHQSI TGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[0912] ABE7.10_Cfa (GEP) -C_断裂_S577C:

[0913] MVKII SRKSLGTQNVYDIGVGEPHNFLKNGLVASNCVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDI
LEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWRLSRKLINGIRDKQSGKTILDFLKSD

GFANRNFQMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIIKKGILQTVKVVDDELVKVMGRHKPENIVIE
ARENQTTQKGQKNSRERMKRIIEGKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYD
IVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFI
KRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVG
TALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGE
IVWDKGRDFATVRKVLSPQVNIIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
VEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLEFENGRKRMLASAGELQKGN
LALPSKYVNFLYLASHYEKLKGSPEDEQKQLFVEQHKHYLDEIIIEQISEFSKRVILADANLDKVL SAYNKH
RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDTLIIHQSIITGLYETRIDLSQLGGDEGADKRTADGS
EFESPKKKRKV

[0914] ABE7.10_Cfa (GEP) -C_断裂_A589C:

[0915] MVKIIISRKSLGTQNVYDIGVGEPHNFLKNGLVASNCSLGTYHDLLKIIKDKDFLDNEENEDI
LEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFQMLIH
DDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIIKKGILQTVKVVDDELVKVMGRHKPENIVIEARENQTTQKGQK
NSRERMKRIIEGKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYD
IVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITK
HVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVG
TALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATV
RKVLSPQVNIIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDF

[0916] LEAKGYKEVKKDLIIKLPKYSLEFENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPE
DNEQKQLFVEQHKHYLDEIIIEQISEFSKRVILADANLDKVL SAYNKH
RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDTLIIHQSIITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[0917] ABE7.10_Cfa (GEP) -C_断裂_S590C:

[0918] MVKIIISRKSLGTQNVYDIGVGEPHNFLKNGLVASNCCLGTYHDLLKIIKDKDFLDNEENEDI
LEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFQMLIH
DSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIIKKGILQTVKVVDDELVKVMGRHKPENIVIEARENQTTQKGQK
NSRERMKRIIEGKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYD
IVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
VAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVG
TALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATV
RKVLSPQVNIIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLEFENGRKRMLASAGELQKGNELALPSKYVNFLYL
ASHYEKLKGSPEDEQKQLFVEQHKHYLDEIIIEQISEFSKRVILADANLDKVL SAYNKH
RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDTLIIHQSIITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[0919] 实施例2:经由双重AAV感染的断裂A-至-G碱基编辑器 (ABE) 的递送产出了高A>G转换率。

[0920] 评估了融合到内含肽-N和内含肽-C的断裂A-至-G碱基编辑器的AAV递送,包含断裂ABE在共-感染细胞中重新组装和重构碱基编辑活性的能力。使用荧光报告基因构建物

执行初步研究以确定视网膜细胞中AAV2共-感染的最佳感染复数(MOI)。含有CMV-mCherry和CMV-EmGFP的AAV2表达载体用于监测共-感染。流式细胞仪显示ARPE-19细胞的AAV2共感染的最佳MOI为60,000-100,000vg/细胞(图5A)。在共-感染后第3天,在60,000vg/细胞的条件下,荧光显微镜显示EmGFP和mCherry的共-定位(图5B)。在50,000vg/细胞的条件下,CMV-mCherry(图5C)和CMV-EmGFP(图5D)的单一感染亦都显示大量(>95%)具有荧光报告基因的细胞表达。荧光细胞的百分比取决于总的病毒载量。

[0921] 经由双重AAV2感染将断裂编辑器递送到ARPE-19视网膜细胞在ABCA4 5882A处产出了高的A>G转换(图6)。在所述实验中,评估了在Cas9 T310处(N-端在Cas9氨基酸位置N309处融合至内含肽-N,而C-端在Cas9氨基酸位置T310处融合到内含肽-C)断裂的ABE。将以下构建体包装到AAV2载体中:

[0922] AAV2-N: AAV2/P_{CMV} 断裂_ABE7.10^{N(T310)} - Int^N-rGpA | P_{U6}sgRNA

[0923] AAV2-C: AAV2/P_{CMV} Int^C-断裂_ABE7.10^{C(T310)} -rGpA

[0924] 是根据AAV包装的常规方法(病毒载体核心设施,马萨诸塞大学医学院)。视网膜ARPE-19细胞是:用两种AAV载体共-感染;用每个AAV载体单个感染;或未经处理。如在从20,000-60,000vg/细胞的范围内的各种MOI(双重感染)所示,在ABCA4 5882A处观察到A>G转换。A>G转换随双重感染MOI的增加而增加(在20,000vg/细胞时为约13% -在60,000vg/细胞时为约30%)。A>G转换伴随着低水平(<.1%)的插入/缺失形成。在单独用N末端片段或C末端片段感染的细胞中或在未经处理的细胞中,在靶位点处均未观察到A>G转换。

[0925] 其他实施方式

[0926] 从前面的描述中将显而易见的是,可以对本文所描述的本发明进行变化和修改以将其适用于各种用途和条件。此种实施方式也在以下权利要求的范围内。

[0927] 本文对变量的任何定义中的元件列表的列举,包含了将所述变量定义为所列元件的任何单个元件或其组合(或其子组合)。本文对实施方式的列举,包含了所述实施方式作为任何单个实施方式或与任何其他实施方式或其部分的组合。

[0928] 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被通过引用并入本文。

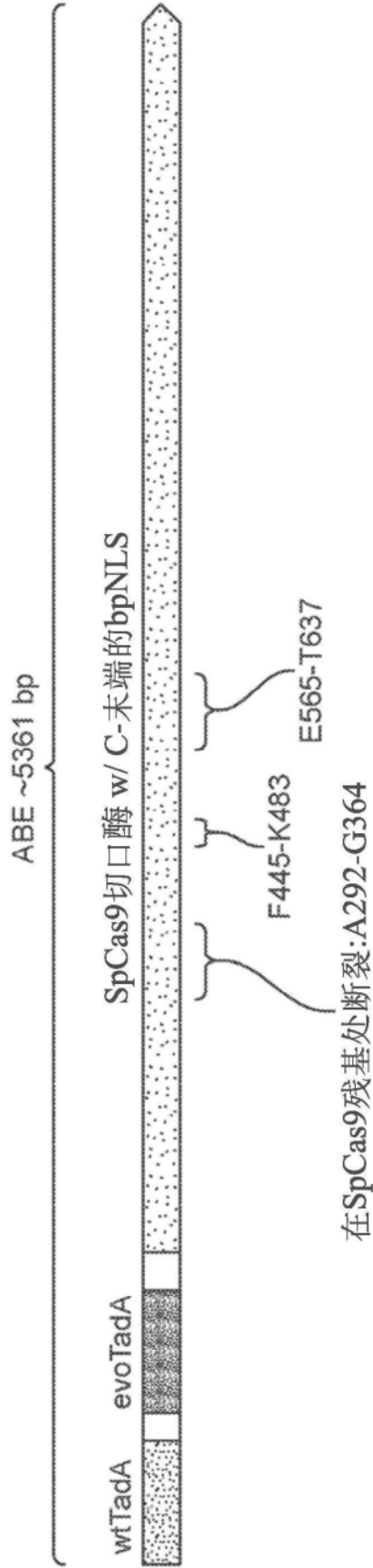


图1

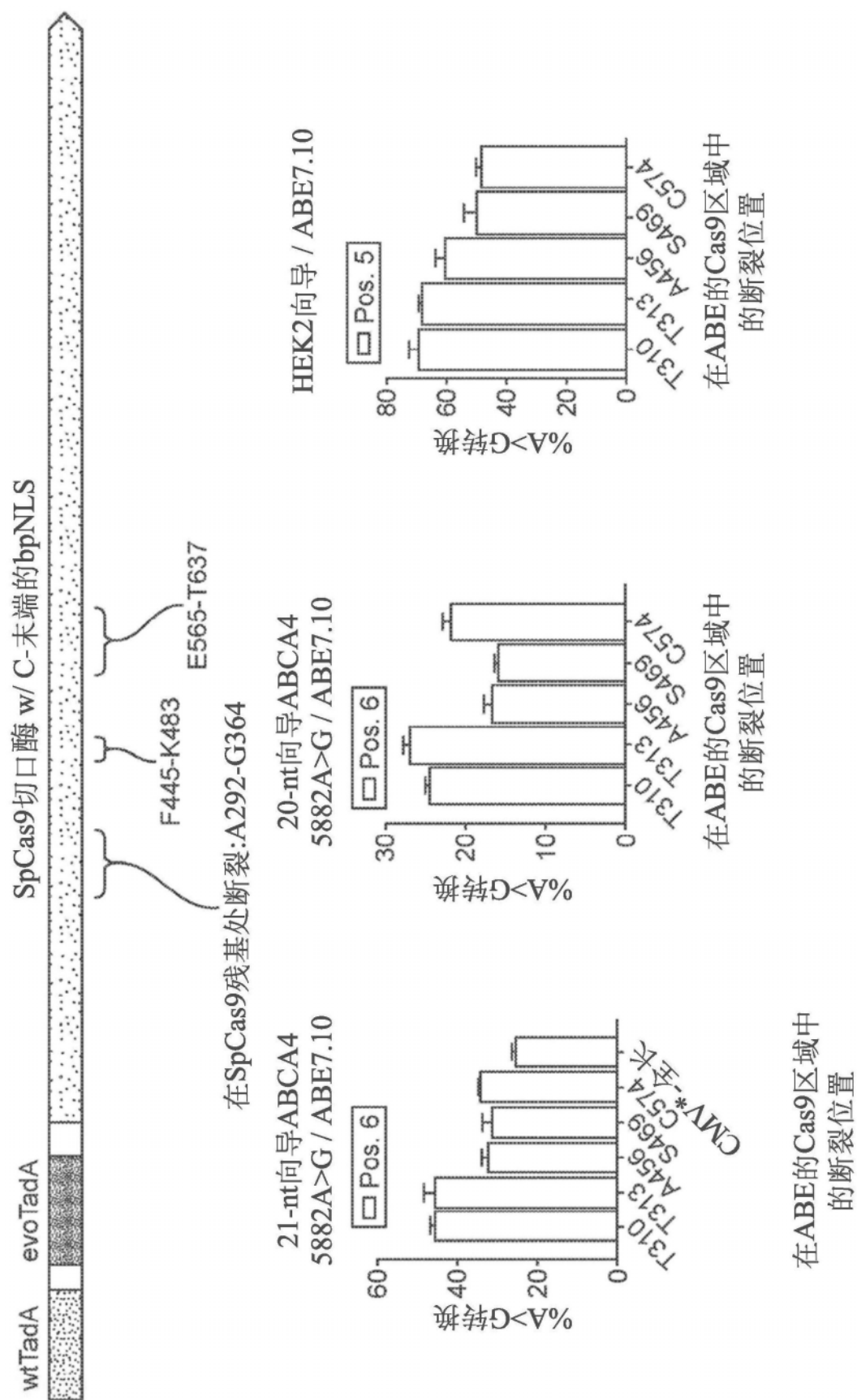


图2

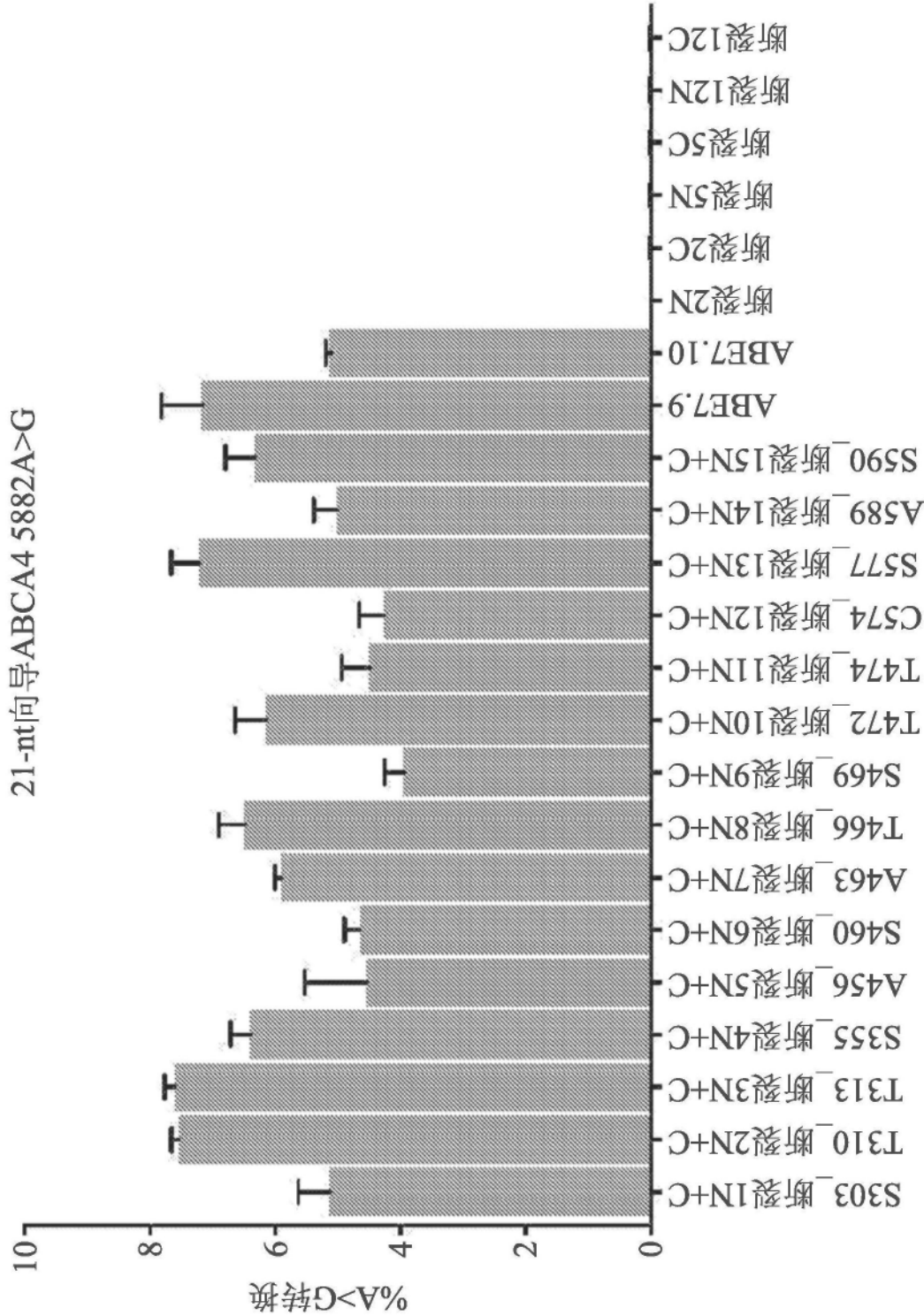


图3

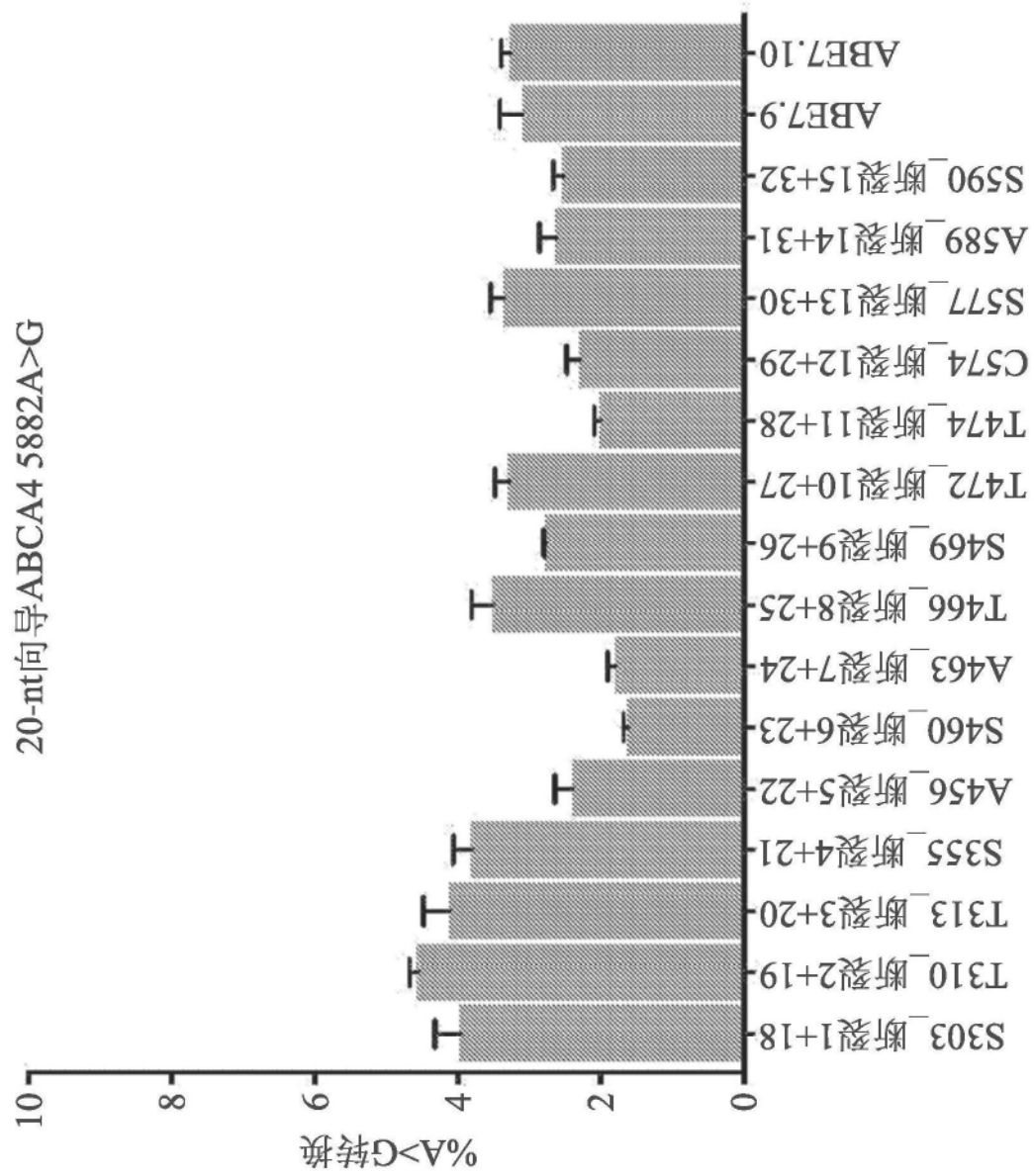


图4

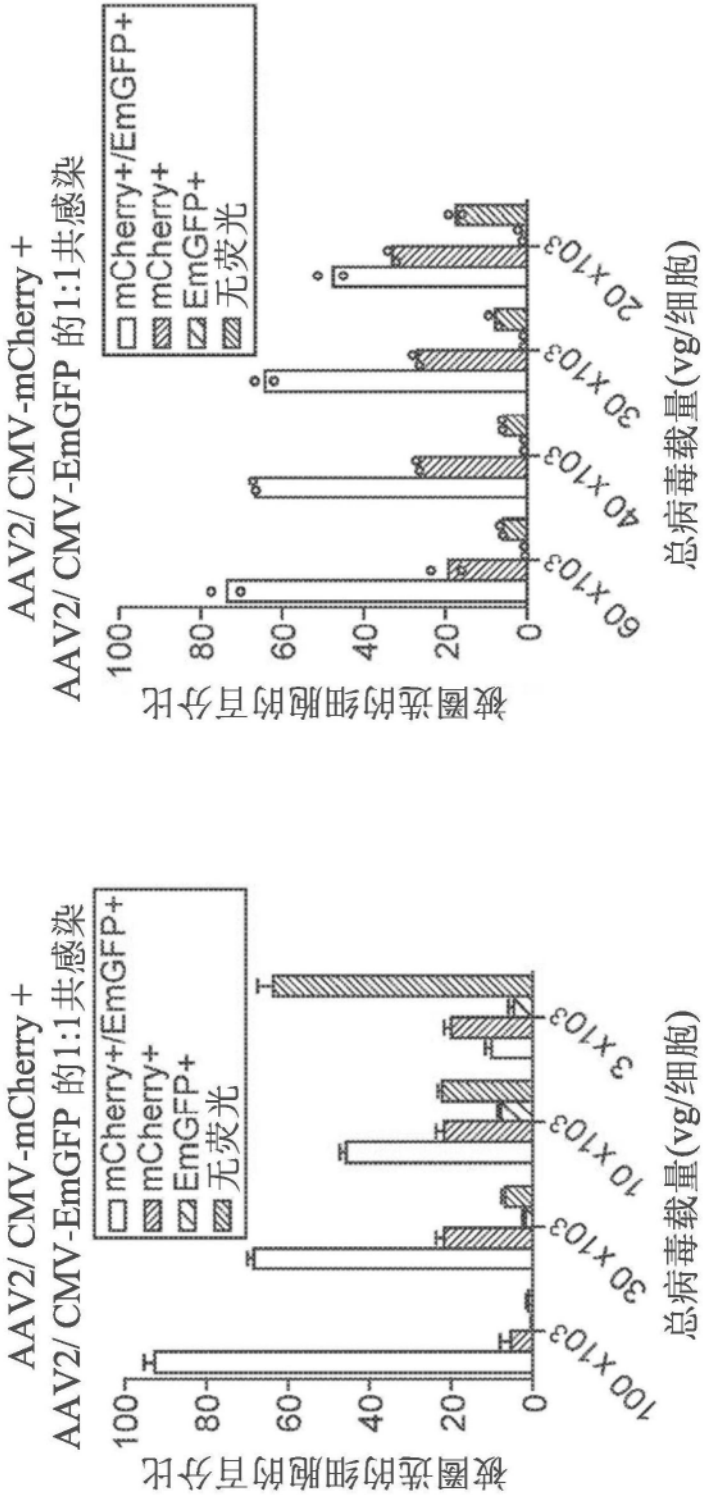


图5A

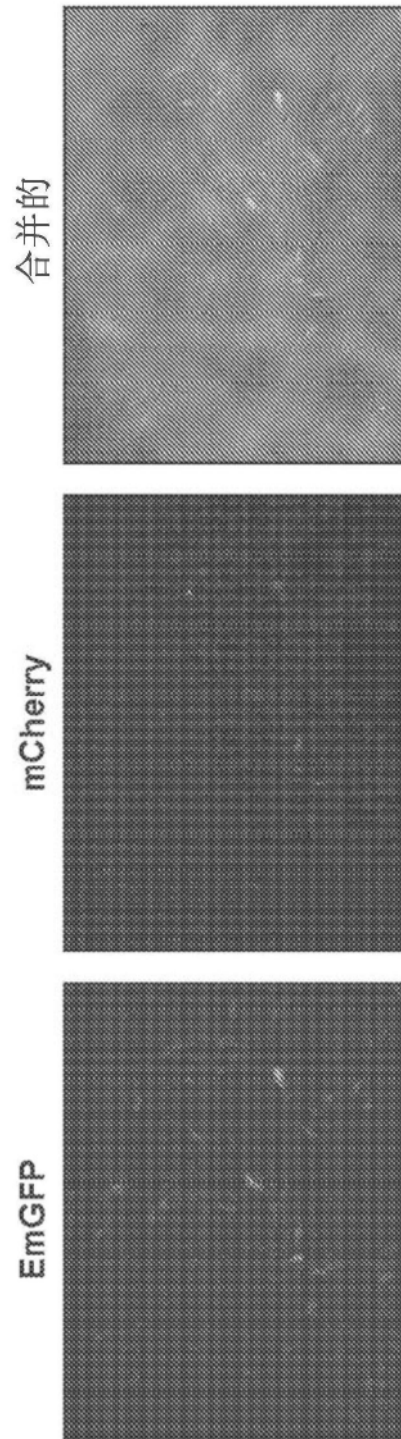


图5B

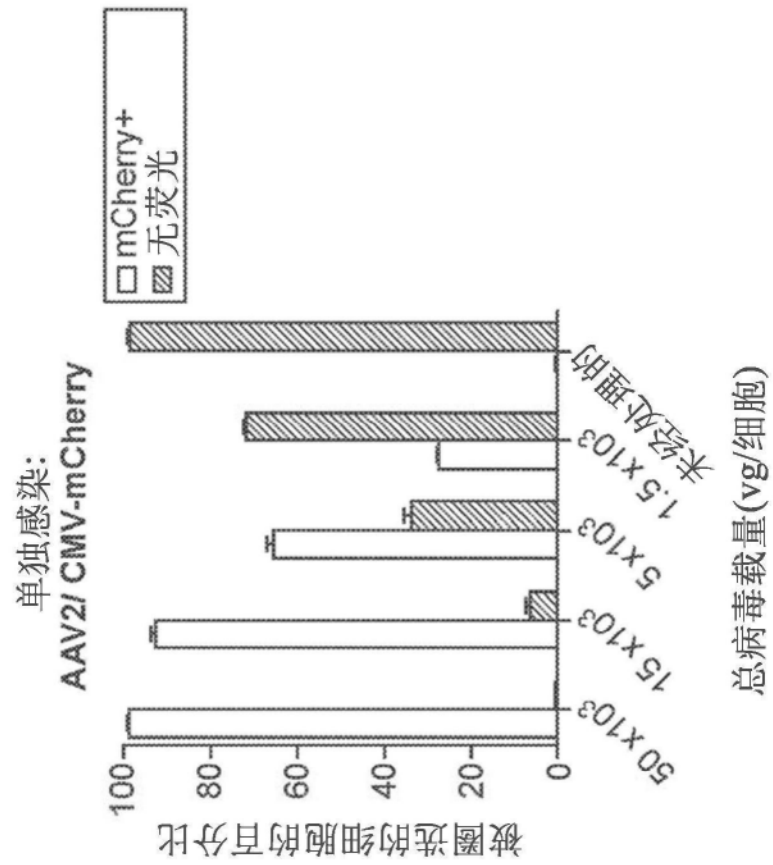


图5C

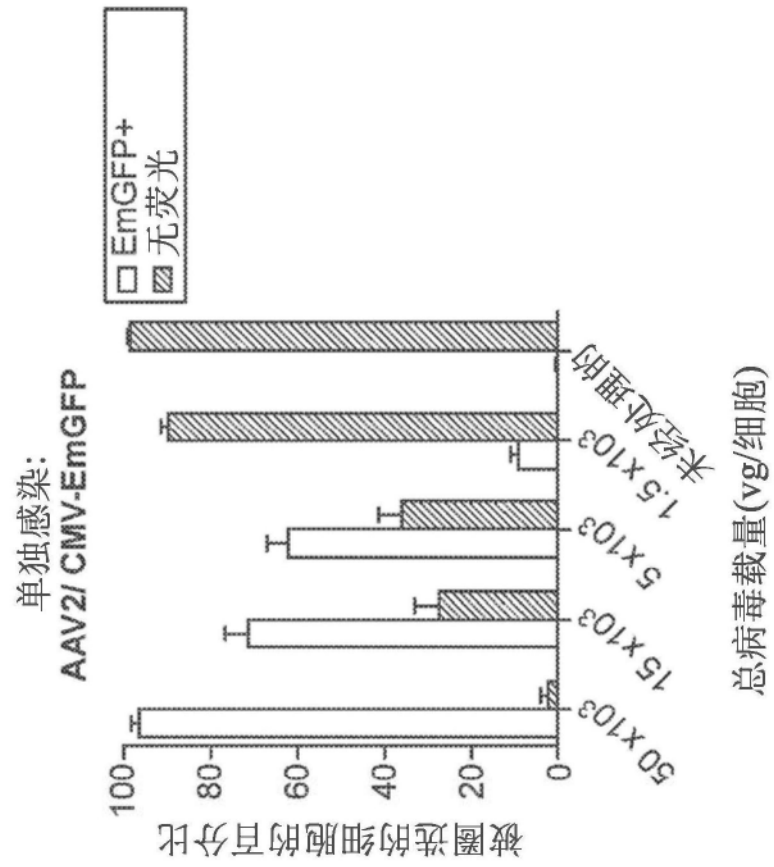


图5D

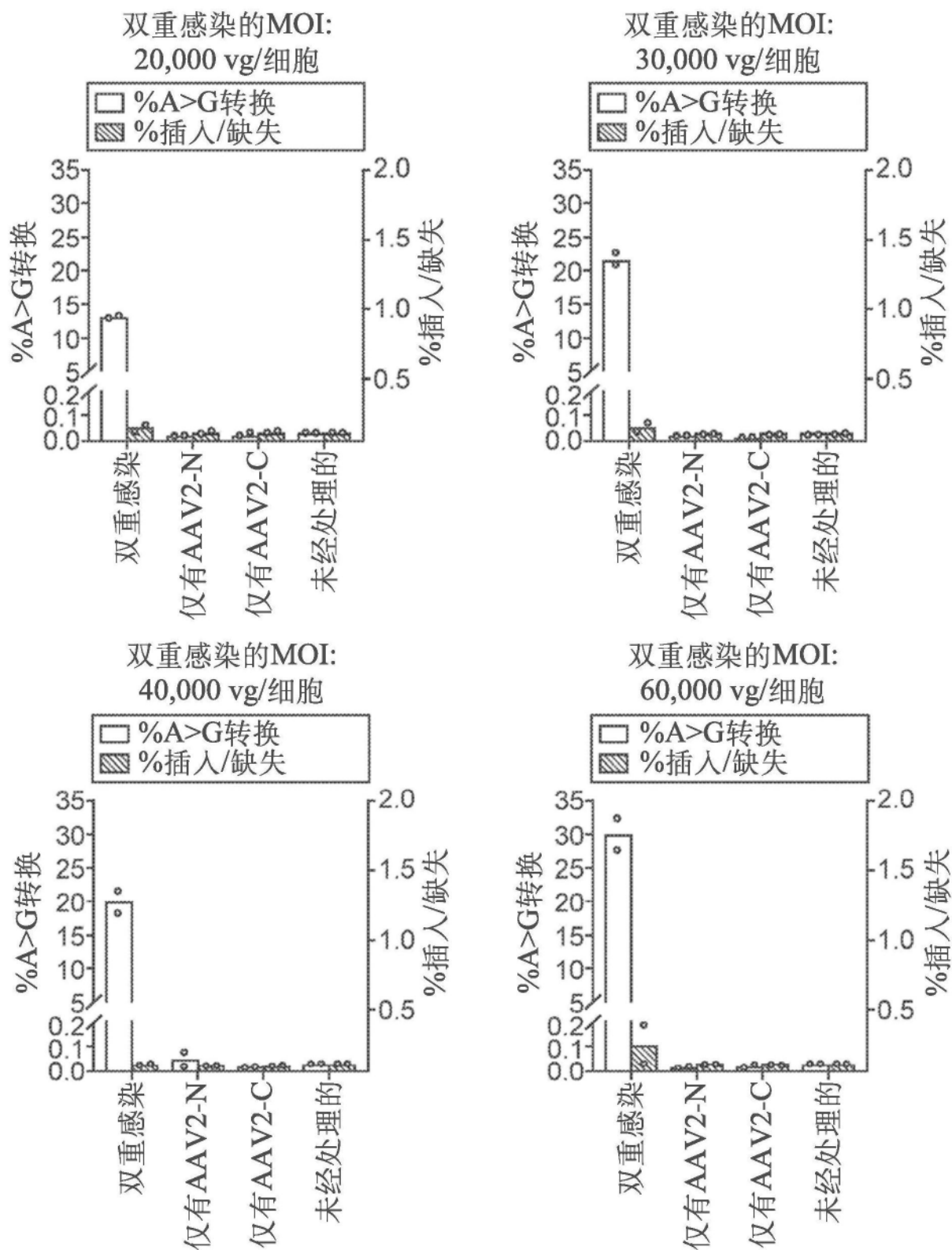


图6