(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 118318037 A
(43)申请公布日 2024.07.09

(21)申请号 202280073061.3

(22)申请日 2022.09.07

(30)优先权数据
63/241,931 2021.09.08 US

(85)PCT国际申请进入国家阶段日
2024.04.29

(86)PCT国际申请的申请数据
PCT/US2022/076024 2022.09.07

(87)PCT国际申请的公布数据
WO2023/039407 EN 2023.03.16

(71)申请人 旗舰先锋创新VI有限责任公司
地址 美国马萨诸塞州

(72)发明人 R·C·阿尔特舒勒 A·H·博思默 R·J·奇特里克 C·G·S·科特拉-拉穆西诺 K·金 R·M·科特拉 S·库马尔 A·雷 N·罗奎特 J·R·鲁本斯 W·E·萨洛蒙 B·E·斯坦伯格 Z·J·王 E·阿金奇

(74)专利代理机构 北京市柳沈律师事务所 11105
专利代理师 秦剑

(51)Int.Cl.
C12N 9/22(2006.01)
C07K 14/195(2006.01)

权利要求书3页 说明书292页
序列表（电子公布） 附图3页

(54)发明名称

调节基因组的方法和组合物

(57)摘要

披露了调节靶基因组的方法和组合物。

CN 118318037 A

1.一种基因修饰多肽,其包含:

Cas结构域(例如,Cas切口酶结构域,例如,Cas9切口酶结构域);

表1或表23的聚合酶(Pol)结构域,或与其具有至少70%、75%、80%、85%、90%、95%、97%、98%或99%同一性的序列,其中该Pol结构域位于该Cas结构域的C末端;以及

布置在该Pol结构域和该Cas结构域之间的接头。

2.如权利要求1所述的基因修饰多肽,其中该接头具有来自表6的序列或与其具有至少70%、75%、80%、85%、90%、95%、97%、98%或99%同一性的序列。

3.如权利要求1或2所述的基因修饰多肽,其中该Pol结构域具有与表1或23的Pol结构域具有至少90%同一性的序列。

4.如前述权利要求中任一项所述的基因修饰多肽,其中该Pol结构域具有与表1或23的Pol结构域具有至少95%同一性的序列。

5.如前述权利要求中任一项所述的基因修饰多肽,其中该Pol结构域具有与表1或23的Pol结构域具有至少98%同一性的序列。

6.如前述权利要求中任一项所述的基因修饰多肽,其中该Pol结构域具有与表1或23的Pol结构域具有至少99%同一性的序列。

7.如前述权利要求中任一项所述的基因修饰多肽,其中该Pol结构域具有与表1或23的Pol结构域具有100%同一性的序列。

8.如前述权利要求中任一项所述的基因修饰多肽,其中该接头具有与来自表6的接头序列具有至少90%同一性的序列。

9.如前述权利要求中任一项所述的基因修饰多肽,其中该接头具有与来自表6的接头序列具有至少95%同一性的序列。

10.如前述权利要求中任一项所述的基因修饰多肽,其中该接头具有与来自表6的接头序列具有至少97%同一性的序列。

11.如前述权利要求中任一项所述的基因修饰多肽,其中该接头具有与来自表6的接头序列具有100%同一性的序列。

12.如前述权利要求中任一项所述的基因修饰多肽,其中该Cas结构域包含表4的序列或与其具有至少70%、75%、80%、85%、90%、95%、98%或99%同一性的序列。

13.如前述权利要求中任一项所述的基因修饰多肽,其中该Cas结构域是Cas切口酶结构域。

14.如前述权利要求中任一项所述的基因修饰多肽,其中该Cas结构域是Cas9切口酶结构域。

15.如前述权利要求中任一项所述的基因修饰多肽,其中该Cas结构域包含N863A突变。

16.如前述权利要求中任一项所述的基因修饰多肽,其包含NLS,例如,其中该基因修饰多肽包含两个NLS。

17.如前述权利要求中任一项所述的基因修饰多肽,其包含该Cas9结构域的N末端的NLS。

18.如前述权利要求中任一项所述的基因修饰多肽,其包含该Pol结构域的C末端的NLS。

19.如前述权利要求中任一项所述的基因修饰多肽,其包含位于该Cas9结构域的N末端

的第一NLS和位于该Pol结构域的C末端的第二NLS。

20.一种核酸(例如,DNA或RNA,例如,mRNA),其编码如前述权利要求中任一项所述的基因修饰多肽。

21.一种细胞,其包含如权利要求1-19中任一项所述的基因修饰多肽或如权利要求20所述的核酸。

22.一种系统,其包含:

i)如权利要求1-19中任一项所述的基因修饰多肽,和

ii)模板核酸(例如,模板RNA),其包含:

a)与部分靶核酸序列互补的gRNA间隔子;

b)与该基因修饰多肽的Cas结构域结合的gRNA支架;

c)异源对象序列;以及

d)引物结合位点序列(PBS序列)。

23.如权利要求22所述的系统,其中该模板核酸包括RNA。

24.如权利要求22或23所述的系统,其中该模板核酸包括DNA。

25.如权利要求22所述的系统,其中该模板核酸包括DNA和RNA。

26.如权利要求22-25中任一项所述的系统,其中该模板核酸包含如表24中所列的核酸序列,或与其具有至少75%、80%、85%、90%、95%、96%、97%、98%或99%序列同一性的核酸序列。

27.如权利要求26所述的系统,其中该模板核酸的gRNA间隔子包含如表24中所列的间隔子序列,或与其具有至少75%、80%、85%、90%、95%、96%、97%、98%或99%序列同一性的核酸序列。

28.如权利要求26或27所述的系统,其中该模板核酸的gRNA支架包含如表24中所列的支架序列,或与其具有至少75%、80%、85%、90%、95%、96%、97%、98%或99%序列同一性的核酸序列。

29.如权利要求26-28中任一项所述的系统,其中该模板核酸的PBS序列包含如表24中所列的PBS序列,或与其具有至少75%、80%、85%、90%、95%、96%、97%、98%或99%序列同一性的核酸序列。

30.如权利要求26所述的系统,其中该模板核酸序列包含以下中的一种或多种(例如,1、2或全部3种),例如按照5'至3'的顺序:

(i)如表24中所列的间隔子序列,或与其具有至少75%、80%、85%、90%、95%、96%、97%、98%或99%序列同一性的核酸序列;

(ii)如表24中所列的支架序列,或与其具有至少75%、80%、85%、90%、95%、96%、97%、98%或99%序列同一性的核酸序列;和/或

(iii)如表24中所列的PBS序列,或与其具有至少75%、80%、85%、90%、95%、96%、97%、98%或99%序列同一性的核酸序列。

31.如权利要求26-30中任一项所述的系统,其中该模板核酸包含如表24中所列的全模板分子序列,或与其具有至少75%、80%、85%、90%、95%、96%、97%、98%或99%序列同一性的核酸序列。

32.如权利要求22-31中任一项所述的系统,其中该gRNA间隔子和该gRNA支架包含RNA。

33.如权利要求22-32中任一项所述的系统,其中该异源对象序列包含DNA并且PBS序列包含RNA。

34.如权利要求22-32中任一项所述的系统,其中该异源对象序列和PBS序列包含DNA。

35.如权利要求22-34中任一项所述的系统,其中该基因修饰多肽包含如表23中所列的氨基酸序列,或与其具有至少75%、80%、85%、90%、95%、96%、97%、98%或99%序列同一性的氨基酸序列。

36.如权利要求22-35中任一项所述的系统,其中该基因修饰多肽包含nCas9-UL-Polθ_L、nCas9-UL-Polθ_M、nCas9-UL-Polθ_4x0q、nCas9-FL-Polθ_M或nCas9-FL-Polθ_4x0q中任一个的氨基酸序列,或与其具有至少75%、80%、85%、90%、95%、96%、97%、98%或99%序列同一性的氨基酸序列。

37.如权利要求22-36中任一项所述的系统,其中该基因修饰多肽包含如表23中所列的基因修饰多肽的Cas结构域氨基酸序列,或与其具有至少75%、80%、85%、90%、95%、96%、97%、98%或99%序列同一性的氨基酸序列。

38.如权利要求22-37中任一项所述的系统,其中该基因修饰多肽包含nCas9-UL-Polθ_L、nCas9-UL-Polθ_M、nCas9-UL-Polθ_4x0q、nCas9-FL-Polθ_M或nCas9-FL-Polθ_4x0q中任一个的Cas结构域氨基酸序列,或与其具有至少75%、80%、85%、90%、95%、96%、97%、98%或99%序列同一性的氨基酸序列。

39.如权利要求22-38中任一项所述的系统,其中该基因修饰多肽包含如表23中所列的基因修饰多肽的Pol结构域氨基酸序列,或与其具有至少75%、80%、85%、90%、95%、96%、97%、98%或99%序列同一性的氨基酸序列。

40.如权利要求22-39中任一项所述的系统,其中该基因修饰多肽包含nCas9-UL-Polθ_L、nCas9-UL-Polθ_M、nCas9-UL-Polθ_4x0q、nCas9-FL-Polθ_M或nCas9-FL-Polθ_4x0q中任一个的Pol结构域氨基酸序列,或与其具有至少75%、80%、85%、90%、95%、96%、97%、98%或99%序列同一性的氨基酸序列。

41.如权利要求22-40中任一项所述的系统,其中该基因修饰多肽包含如表23中所列的基因修饰多肽的Cas结构域氨基酸序列和Pol结构域氨基酸序列,或与其具有至少75%、80%、85%、90%、95%、96%、97%、98%或99%序列同一性的氨基酸序列。

42.如权利要求22-41中任一项所述的系统,其中该基因修饰多肽包含nCas9-UL-Polθ_L、nCas9-UL-Polθ_M、nCas9-UL-Polθ_4x0q、nCas9-FL-Polθ_M或nCas9-FL-Polθ_4x0q中任一个的Cas结构域氨基酸序列和Pol结构域氨基酸序列,或与其具有至少75%、80%、85%、90%、95%、96%、97%、98%或99%序列同一性的氨基酸序列。

43.一种用于修饰细胞(例如,人细胞)中的靶核酸的方法,该方法包括使该细胞与如权利要求22-42中任一项所述的系统或编码其的核酸接触,从而修饰该靶核酸。

44.一种用于治疗患有与遗传缺陷相关的疾病或病况的受试者的方法,该方法包括:

向该受试者施用如前述权利要求中任一项所述的系统、多肽、模板RNA或编码其的DNA,从而治疗该患有与遗传缺陷相关的疾病或病况的受试者。

45.如权利要求44所述的方法,其中该与遗传缺陷相关的疾病或病况是表12-15中任一个中所列的适应症,和/或其中该遗传缺陷是表12-15中任一个中所列的基因的缺陷。

46.如权利要求44或45所述的方法,其中该受试者是人患者。

调节基因组的方法和组合物

[0001] 相关申请

[0002] 本申请要求2021年9月8日提交的美国序列号63/241,931的优先权,将其中每个的全部内容通过援引并入本文。

[0003] 序列表

[0004] 本申请含有序列表,该序列表已以XML格式电子提交,并通过援引以其全文特此并入。所述XML副本创建于2021年9月8日,名为V2065-7020WO_SL.xml,大小为11,288,576字节。

背景技术

[0005] 在没有专门的蛋白质来促进插入事件的情况下,目的核酸整合到基因组中的频率较低且位点特异性极低。一些现有的方法,如CRISPR/Cas9,更适合依赖宿主修复途径的小型编辑,并且在整合较长序列时效率较低。其他现有的方法如Cre/loxP需要第一步先将loxP位点插入基因组中,然后第二步将目的序列插入loxP位点中。本领域需要改善的组合物(例如,蛋白质和核酸)和方法,以在基因组中插入、改变、或缺失目的序列。

发明内容

[0006] 本披露涉及用于体内或体外改变宿主细胞、组织或受试者中一个或多个位置处的基因组的新型组合物、系统和方法。特别地,本发明的特征在于用于在宿主基因组中插入、改变或缺失目的序列的组合物、系统和方法。例如,本披露提供了能够调节(例如,插入、改变或缺失目的序列)基因活性的系统和治疗疾病的方法,这些方法通过以下进行:施用一种或多种这样的系统来改变核苷酸处的基因组序列,以纠正导致疾病的致病突变。

[0007] 这些组合物或方法的特征可包括以下列举的实施例中的一个或多个。

[0008] 1.一种基因修饰多肽,其包含:

[0009] 与靶核酸序列结合的DNA结合结构域(DBD),和

[0010] 表1或23的聚合酶(Pol)结构域,或与其具有至少70%、75%、80%、85%、90%、95%、97%、98%或99%同一性的序列;

[0011] 其中该DBD与该Pol结构域异源;以及

[0012] 布置在该Pol结构域和该DBD之间的接头。

[0013] 2.一种基因修饰多肽,其包含:

[0014] Cas结构域(例如,Cas切口酶结构域,例如,Cas9切口酶结构域);

[0015] 表1或23的聚合酶(Pol)结构域,或与其具有至少70%、75%、80%、85%、90%、95%、97%、98%或99%同一性的序列,其中该Pol结构域位于该Cas结构域的C末端;以及

[0016] 布置在该Pol结构域和该Cas结构域之间的接头。

[0017] 3.如实施例1或2所述的基因修饰多肽,其中该接头具有来自表6的序列,或与其具有至少70%、75%、80%、85%、90%、95%、97%、98%或99%同一性的序列。

[0018] 4.如前述实施例所述的基因修饰多肽,其中该Pol结构域具有与表1或23的Pol结

构域具有至少90％同一性的序列。

[0019] 5.如前述实施例中任一项所述的基因修饰多肽,其中该Pol结构域具有与表1或23的Pol结构域具有至少95％同一性的序列。

[0020] 6.如前述实施例中任一项所述的基因修饰多肽,其中该Pol结构域具有与表1或23的Pol结构域具有至少98％同一性的序列。

[0021] 7.如前述实施例中任一项所述的基因修饰多肽,其中该Pol结构域具有与表1或23的Pol结构域具有至少99％同一性的序列。

[0022] 8.如前述实施例中任一项所述的基因修饰多肽,其中该Pol结构域具有与表1或23的Pol结构域具有100％同一性的序列。

[0023] 9.如前述实施例中任一项所述的基因修饰多肽,其中该接头具有与来自表6的接头序列具有至少90％同一性的序列。

[0024] 10.如前述实施例中任一项所述的基因修饰多肽,其中该接头具有与来自表6的接头序列具有至少95％同一性的序列。

[0025] 11.如前述实施例中任一项所述的基因修饰多肽,其中该接头具有与来自表6的接头序列具有至少97％同一性的序列。

[0026] 12.如前述实施例中任一项所述的基因修饰多肽,其中该接头具有与来自表6的接头序列具有100％同一性的序列。

[0027] 13.如前述实施例中任一项所述的基因修饰多肽,其中该Cas结构域包含表4的序列,或与其具有至少70％、75％、80％、85％、90％、95％、98％或99％同一性的序列。

[0028] 14.如前述实施例中任一项所述的基因修饰多肽,其中该Cas结构域是Cas切口酶结构域。

[0029] 15.如前述实施例中任一项所述的基因修饰多肽,其中该Cas结构域是Cas9切口酶结构域。

[0030] 16.如前述实施例中任一项所述的基因修饰多肽,其中该Cas结构域包含N863A突变。

[0031] 17.如前述实施例中任一项所述的基因修饰多肽,其包含NLS,例如其中该基因修饰多肽包含两个NLS。

[0032] 18.如前述实施例中任一项所述的基因修饰多肽,其包含该Cas9结构域的N末端的NLS。

[0033] 19.如前述实施例中任一项所述的基因修饰多肽,其包含该Pol结构域的C末端的NLS。

[0034] 20.如前述实施例中任一项所述的基因修饰多肽,其包含位于该Cas9结构域的N末端的第一NLS和位于该Pol结构域的C末端的第二NLS。

[0035] 21.一种核酸(例如,DNA或RNA,例如,mRNA),其编码如前述实施例中任一项所述的基因修饰多肽。

[0036] 22.一种细胞,其包含如实施例1-20中任一项所述的基因修饰多肽或如实施例21所述的核酸。

[0037] 23.一种系统,其包含:

[0038] i)如实施例1-20中任一项所述的基因修饰多肽,和

[0039] ii)模板核酸(例如,模板RNA),其包含:

[0040] a)与部分靶核酸序列互补的gRNA间隔子;

[0041] b)与该基因修饰多肽的Cas结构域结合的gRNA支架;

[0042] c)异源对象序列;以及

[0043] d)引物结合位点序列(PBS序列)。

[0044] 24.如实施例23所述的系统,其中该模板核酸包括RNA。

[0045] 25.如实施例23或24所述的系统,其中该模板核酸包括DNA。

[0046] 26.如实施例23-25中任一项所述的系统,其中该gRNA间隔子和该gRNA支架包含RNA。

[0047] 27.如实施例23-26中任一项所述的系统,其中该异源对象序列包含DNA并且PBS序列包含RNA。

[0048] 28.如实施例23-26中任一项所述的系统,其中该异源对象序列和PBS序列包含DNA。

[0049] 29.一种用于修饰细胞(例如,人细胞)中的靶核酸的方法,该方法包括使该细胞与如实施例23-28中任一项所述的系统或编码其的核酸接触,从而修饰该靶核酸。

[0050] 30.一种用于治疗患有与遗传缺陷相关的疾病或病况的受试者的方法,该方法包括:

[0051] 向该受试者施用如前述实施例中任一项所述的系统、多肽、模板RNA或编码其的DNA,从而治疗该患有与遗传缺陷相关的疾病或病况的受试者。

[0052] 31.如实施例30所述的方法,其中该与遗传缺陷相关的疾病或病况是表12-15中任一个中所列的适应症,和/或其中该遗传缺陷是表12-15中任一个中所列的基因的缺陷。

[0053] 32.如实施例30或31所述的方法,其中该受试者是人患者。

[0054] 一方面,本披露涉及用于修饰基因的系统,该系统包含(a)编码能够靶向引发的逆转录的基因修饰多肽的核酸,该多肽包含(i)聚合酶(Pol)结构域和(ii)结合DNA并具有核酸内切酶活性的Cas9切口酶;和(b)模板RNA、DNA或在同一链中同时具有核糖核苷酸和脱氧核糖核苷酸残基的杂交体,该模板RNA、DNA或杂交体包含(i)与靶基因的第一部分互补的gRNA间隔子,(ii)结合该多肽的gRNA支架,(iii)包含突变区以修饰基因的异源对象序列,以及(iv)引物结合位点(PBS)序列,其包含与该模板RNA的3′端处的靶DNA链具有100%同源性的至少3、4、5、6、7或8个碱基。

[0055] gRNA间隔子可以包含与模板RNA的5’端处的靶DNA具有100%同一性的至少15个碱基。模板RNA可以进一步包含PBS序列,该PBS序列包含与靶DNA链具有至少80%同源性的至少5个碱基。模板RNA可以包含一个或多个化学修饰。

[0056] 基因修饰多肽的结构域可以通过肽接头连接。该多肽可包含一个或多个肽接头。基因修饰多肽还可包含核定位信号。该多肽可包含多于一个核定位信号,例如,多个相邻的核定位信号或在该多肽的不同区域中的一个或多个核定位信号,例如,在该多肽的N末端中的一个或多个核定位信号以及在该多肽的C末端中的一个或多个核定位信号。编码基因修饰多肽的核酸可以编码一个或多个内含肽结构域。

[0057] 将该系统引入靶细胞可以导致插入至少1、2、3、4、5、10、15、20、25、30、35、40、45、50、60、70、80、90、100、150、200、250、300、350、400、500或1000个碱基对的外源DNA。将该系

统引入靶细胞可导致缺失,其中该缺失是该插入上游或下游的基因组DNA的少于2、3、4、5、10、50或100个碱基对。将该系统引入靶细胞可导致取代,例如1、2或3个核苷酸(例如连续核苷酸)的取代。

[0058] 异源对象序列可以是至少5、10、25、50、100、150、200、250、300、400、500、600或700个碱基对。

[0059] 一方面,本披露涉及一种药物组合物,其包含上述系统和药学上可接受的赋形剂或载剂,其中该药学上可接受的赋形剂或载剂选自由以下组成的组:质粒载体、病毒载体、囊泡和脂质纳米颗粒。一方面,本披露涉及一种药物组合物,其包含上述系统和多种药学上可接受的赋形剂或载剂,其中该药学上可接受的赋形剂或载剂选自由以下组成的组:质粒载体、病毒载体、囊泡和脂质纳米颗粒,例如,其中上述系统由两种不同的赋形剂或载剂递送,例如两种脂质纳米颗粒、两种病毒载体、或一种脂质纳米颗粒和一种病毒载体。该病毒载体可以是腺相关病毒(AAV)。

[0060] 一方面,本披露涉及一种宿主细胞(例如哺乳动物细胞,例如人细胞),其包含上述系统。

[0061] 一方面,本披露涉及纠正细胞、组织或受试者中的人类基因中的突变的方法,该方法包括向该细胞、组织或受试者施用上述系统。该系统可以体内、体外、离体或原位引入。(a)的核酸可以整合到宿主细胞的基因组中。在一些实施例中,(a)的核酸未整合到宿主细胞的基因组中。在一些实施例中,异源对象序列仅插入宿主细胞基因组中的一个靶位点。异源对象序列可以插入宿主细胞基因组中的两个或更多个靶位点,例如,插入两个同源染色体中的相同对应位点,或相同或不同染色体上的两个不同位点。异源对象序列可以编码哺乳动物多肽或其片段或变体。该系统的组分可以在1、2、3、4或更多个不同的核酸分子上递送。可以通过电穿孔或通过使用选自质粒载体、病毒载体、囊泡和脂质纳米颗粒的至少一种媒介物将该系统引入宿主细胞中。

附图说明

[0062] 图1描绘了如本文所述的基因修饰系统。左手图显示基因修饰多肽,其包含通过接头连接的Cas切口酶结构域(例如,spCas9 N863A)和Pol1结构域。右手图显示模板核酸,其从5′至3′包含gRNA间隔子、gRNA支架、异源对象序列和引物结合位点序列(PBS序列)。异源对象序列可以包含突变区,该突变区包含相对于靶位点的一个或多个序列差异。异源对象序列还可以包含位于该突变区侧翼的编辑前同源区和编辑后同源区。不希望受理论束缚,认为模板核酸的gRNA间隔子与基因组中的靶位点的第二链结合,而模板核酸的gRNA支架与基因修饰多肽结合,例如将基因修饰多肽定位到基因组中的靶位点。认为基因修饰多肽的Cas结构域在靶位点(例如,靶位点的第一链)产生切口,例如,允许PBS序列与跟靶位点的第一链上待改变位点相邻的序列结合。认为基因修饰多肽的Pol1结构域使用与互补序列(包含模板核酸的PBS序列)结合的靶位点的第一链作为引物,并以模板核酸的异源对象序列作为模板,例如聚合与异源对象序列互补的序列。不希望受理论束缚,认为DNA聚合然后可以通过编辑前同源区,然后通过突变区,然后通过编辑后同源区进行,从而产生包含异源对象序列指定的突变的DNA链。

[0063] 图2是显示人DNA聚合酶θ的示例性截短的图。

[0064] 图3A-3B的一系列图显示HEK293细胞(图3A)和U2OS细胞(图3B)中Cas-Pol基因修饰多肽对所指示的模板核酸分子的编辑活性。

具体实施方式

[0065] 定义

[0066] 如本文所用,术语“表达盒”是指包含足以表达本发明的核酸分子的核酸元件的核酸构建体。

[0067] 如本文所用,“gRNA间隔子”是指与靶核酸具有互补性并且可以与gRNA支架一起将Cas蛋白靶向靶核酸的核酸部分。

[0068] 如本文所用,“gRNA支架”是指可以结合Cas蛋白并且可以与gRNA间隔子一起将Cas蛋白靶向靶核酸的核酸部分。在一些实施例中,gRNA支架包含crRNA序列、四环和tracrRNA序列。

[0069] 如本文所用,“基因修饰多肽”是指包含聚合酶或逆转录病毒逆转录酶的多肽,或包含与聚合酶或逆转录病毒逆转录酶具有至少75%、80%、85%、90%、95%、96%、97%、98%或99%氨基酸序列同一性的氨基酸序列的多肽,其能够将核酸序列(例如,模板核酸上提供的序列)整合到靶DNA分子中(例如,在哺乳动物宿主细胞中,如宿主细胞中的基因组DNA分子)。在一些实施例中,基因修饰多肽能够在基本上不依赖宿主机器的情况下整合序列。在一些实施例中,基因修饰多肽将序列整合到基因组中的随机位置,并且在一些实施例中,基因修饰多肽将序列整合到特定靶位点。在一些实施例中,基因修饰多肽包含一个或多个结构域,它们共同促进1)结合模板核酸,2)结合靶DNA分子,和3)促进模板核酸的至少一部分的整合进入靶DNA。基因修饰多肽包括天然存在的多肽以及前述多肽的工程改造变体,例如,这些变体相对于天然存在的序列具有一个或多个氨基酸取代。基因修饰多肽还包括异源构建体,例如,其中一个或多个上述结构域彼此异源,无论是通过在其他方面是野生型的结构域的异源融合(或其他缀合物),以及经修饰的结构域的融合,例如,通过异源子结构域或其他经取代的结构域的替代或融合。可用于本文提供的方法的示例性基因修饰多肽、包含它们的系统以及使用它们的方法例如描述于PCT/US 2021/020948,其关于包含逆转录病毒逆转录酶结构域的基因修饰多肽通过援引并入本文。在一些实施例中,基因修饰多肽将序列整合到基因中。在一些实施例中,基因修饰多肽将序列整合到基因外的序列中。如本文所用,“基因修饰系统”是指包含基因修饰多肽和模板核酸的系统。

[0070] 如本文所用,术语“结构域”是指有助于生物分子的特定功能的生物分子的结构。结构域可以包含生物分子的连续区域(例如,连续序列)或不同的非连续区域(例如,非连续序列)。蛋白质结构域的实例包括但不限于核酸内切酶结构域、DNA结合结构域、聚合酶(Pol)结构域、募集结构域、逆转录结构域;核酸的结构域的实例是调节结构域,例如转录因子结合结构域。在一些实施例中,结构域(例如,Cas结构域)可以包含两个或更多个较小的结构域(例如,DNA结合结构域和核酸内切酶结构域)。

[0071] 如本文所用,术语“外源的”,当相对于生物分子(例如核酸序列或多肽)使用时,意指通过人工将生物分子引入宿主基因组、细胞或生物中。例如,使用重组DNA技术或其他方法添加到现有基因组、细胞、组织或受试者中的核酸对于现有核酸序列、细胞、组织或受试者而言是外源的。

[0072] 如本文所用,用于描述靶DNA的单个DNA链的“第一链”和“第二链”基于Pol结构域启动聚合的链来区分两条DNA链,例如,基于靶引发的合成启动的地方。第一链是指靶DNA的链,Pol结构域在该链上启动聚合,例如,在靶引发的合成启动的地方。第二链是指靶DNA的另一条链。第一和第二链名称在其他方面没有描述靶位点DNA链;例如,在一些实施例中,第一链和第二链被本文所述的多肽切口,但“第一”和“第二”链的名称与此类切口发生的顺序无关。

[0073] “基因组安全港位点”(GSH位点)是宿主基因组中的位点,该位点能够容纳新遗传材料的整合,例如,使得插入的遗传元件不会引起宿主基因组的显著改变对宿主细胞或生物构成风险。GSH位点通常满足以下标准中的1、2、3、4、5、6、7、8或9项:(i)距癌症相关基因>300kb;(ii)距miRNA/其他功能性小RNA>300kb;(iii)距5′基因末端>50kb;(iv)距复制起点>50kb;(v)距任何极保守元件>50kb;(vi)转录活性低(即无mRNA +/-25kb);(vii)不在拷贝数可变区中;(viii)在开放染色质中;和/或(ix)是唯一的,在人基因组中有1个拷贝。满足一些或所有这些标准的人基因组中GSH位点的实例包括:(i)腺相关病毒位点1(AAVS1),它是AAV病毒在19号染色体上整合的天然存在的位点;(ii)趋化因子(C-C基序)受体5(CCR5)基因,一种被称为HIV-1共同受体的趋化因子受体基因;(iii)小鼠Rosa26基因座的人直系同源物;(iv)核糖体DNA(“rDNA”)基因座。另外的GSH位点是已知的,并且描述于例如Pellenz等人,2018年8月20日电子公开(https://doi.org/10.1101/396390)中。

[0074] 当本文用于参考第二元件来描述第一元件时,术语“异源”意指第一元件和第二元件在自然界中不以如所描述的布置存在。例如,异源多肽、核酸分子、构建体或序列是指(a)对于表达其的细胞而言不是天然的多肽、核酸分子或多肽或核酸分子序列的一部分,(b)相对于其天然状态已发生改变或突变的多肽或核酸分子或多肽或核酸分子的一部分,或(c)具有与在类似条件下的天然表达水平相比改变的表达的多肽或核酸分子。例如,异源调节序列(例如启动子、增强子)可以用于调节基因或核酸分子的表达,其方式不同于基因或核酸分子通常在自然界中表达的方式。在另一个实例中,多肽或核酸序列的异源结构域(例如,多肽的DNA结合结构域或编码多肽的DNA结合结构域的核酸)可以相对于其他结构域布置,或者可以是不同的序列或相对于多肽的其他结构域或部分或其编码核酸来自不同来源。在某些实施例中,异源核酸分子可以存在于天然宿主细胞基因组中,但是可以具有改变的表达水平或具有不同的序列或两者。在其他实施例中,异源核酸分子对于宿主细胞或宿主基因组可能不是内源的,而是通过转化(例如,转染、电穿孔)引入宿主细胞的,其中所添加的分子可以整合到宿主基因组中,或可以作为染色体外遗传材料短暂存在(例如,mRNA)或半稳定存在超过一代(例如,游离病毒载体、质粒或其他自我复制载体)。

[0075] 如本文所用,将序列“插入”靶位点是指在靶位点处DNA序列的净添加,例如,在未编辑的靶位点中没有同源位置的异源对象序列中存在新的核苷酸的情况。在一些实施例中,PBS序列和异源对象序列与靶核酸序列的核苷酸比对将导致靶核酸序列中的比对空位。

[0076] 如本文所用,异源对象序列在靶位点产生的“缺失”是指靶位点处DNA序列的净缺失,例如,在异源对象序列中没有同源位置的未编辑的靶位点中存在核苷酸的情况。在一些实施例中,PBS序列和异源对象序列与靶核酸序列的核苷酸比对将导致包含PBS序列和异源对象序列的分子中出现比对空位。

[0077] 如本文所用,术语“反向末端重复序列”或“ITR”是指AAV病毒顺式元件,因其对称

性而如此命名。这些元件促进AAV基因组的有效倍增。假设ITR功能的最小元件是Rep结合位点(RBS;5′-GCGCGCTCGCTCGCTC-3′,对于AAV2)和末端解离位点(TRS;5′-AGTTGG-3′,对于AAV2)加上允许发夹形成的可变回文序列。根据本发明,ITR至少包含这三个元件(RBS、TRS和允许形成发夹的序列)。此外,在本发明中,术语"ITR"是指已知天然AAV血清型的ITR(例如血清型1、2、3、4、5、6、7、8、9、10或11AAV的ITR)、由源自不同血清型的ITR元件融合形成的嵌合ITR,及其功能变体。"功能变体"是指与已知ITR具有至少80%、85%、90%、优选至少95%序列同一性的序列,允许包含所述ITR的序列在Rep蛋白存在下倍增。

[0078] 如本文所用,术语"突变区"是指模板核酸中相对于靶核酸中的相应序列具有一个或多个序列差异的区域。序列差异可以包括例如取代、插入、移码或缺失。

[0079] 当应用于核酸序列时,术语"突变的"意指与参考(例如天然)核酸序列相比,核酸序列中的核苷酸被插入、缺失或改变。可以在基因座处进行单个改变(点突变),或者可以在单个基因座处插入、缺失或改变多个核苷酸。另外,可以在核酸序列内的任何数目的基因座处进行一个或多个改变。核酸序列可以通过本领域已知的任何方法进行突变。

[0080] 核酸分子是指RNA和DNA分子两者,包括但不限于互补DNA("cDNA")、基因组DNA("gDNA")和信使RNA("mRNA"),并且还包括合成的核酸分子,例如化学合成或重组产生的核酸分子,例如如本文所述的RNA模板。核酸分子可以是双链或单链、环状或线性的。如果是单链,则核酸分子可以是有义链或反义链。除非另有说明,并且作为本文以通用格式"SEQ ID NO:"描述的所有序列的实例,"包含SEQ ID NO:1的核酸"是指其至少一部分具有(i)SEQ ID NO:1的序列,或(ii)与SEQ ID NO:1互补的序列的核酸。两者之间的选择取决于使用SEQ ID NO:1的上下文。例如,如果将核酸用作探针,则两者之间的选择取决于探针与期望的靶互补的要求。如本领域技术人员将容易理解的,本披露的核酸序列可以经化学或生物化学修饰或可以含有非天然或衍生的核苷酸碱基。此类修饰包括例如标签,甲基化,用类似物取代一个或多个天然存在的核苷酸,核苷酸间修饰,例如不带电荷的连接(例如,甲基膦酸酯、磷酸三酯、氨基磷酸酯、氨基甲酸酯等)、带电荷的连接(例如,硫代磷酸酯、二硫代磷酸酯等),侧链部分(例如,多肽),嵌入剂(例如,吖啶、补骨脂素等),螯合剂,烷基化剂和经修饰的连接(例如,α异头核酸等等)。还包括化学修饰的碱基(参见例如表9,下文)、主链(参见例如表10,下文)和经修饰的帽(参见例如表11,下文)。还包括合成的分子,它们模拟多核苷酸经由氢键和其他化学相互作用与指定序列结合的能力。此类分子是本领域已知的,并且包括例如其中肽连接替代分子主链中的磷酸连接的那些,例如肽核酸(PNA)。其他修饰可以包括,例如,其中核糖环含有桥接部分或其他结构(例如在"锁"核酸(LNA)中发现的修饰)的类似物。在各个实施例中,核酸与另外的遗传元件(例如一个或多个组织特异性表达控制序列(例如,组织特异性启动子和组织特异性微小RNA识别序列))以及另外的元件(例如反向重复序列(例如,反向末端重复序列,例如来自或源自病毒的元件,例如,AAV ITR)和串联重复序列、反向重复序列/直接重复序列、同源区(与靶DNA具有不同同源程度的区段)、非翻译区(UTR)(5'、3'或5'和3'UTR))以及前述的各种组合可操作地关联。本发明提供的系统的核酸元件能以多种拓扑结构提供,包括单链、双链、环状、线性、具有开放末端的线性、具有封闭末端的线性,以及这些的特定版本,例如狗骨DNA(doggybone DNA,dbDNA)、封闭末端DNA(ceDNA)。

[0081] 如本文所用,"基因表达单元"是核酸序列,其包含与至少一个效应子序列可操作

地连接的至少一个调节核酸序列。当第一核酸序列被放置成与第二核酸序列有功能关系时,该第一核酸序列与该第二核酸序列可操作地连接。例如,如果启动子或增强子影响编码序列的转录或表达,则该启动子或增强子与该编码序列可操作地连接。可操作地连接的DNA序列可以是连续的或非连续的。在需要连接两个蛋白质编码区的情况下,可操作地连接的序列可以在同一阅读框中。

[0082] 如本文所用,术语“宿主基因组”或“宿主细胞”是指已将蛋白质和/或遗传材料引入其中的细胞和/或其基因组。应当理解,这样的术语不仅旨在指特定的受试者细胞和/或基因组,而且还指这样的细胞的子代和/或这样的细胞的子代的基因组。因为由于突变或环境影响,某些修饰可能在后代中发生,所以这样的子代实际上可能与亲本细胞不同,但仍包括在如本文所用的术语“宿主细胞”的范围内。宿主基因组或宿主细胞可以是在培养物中生长的分离的细胞或细胞系,或者是从这种细胞或细胞系分离的基因组材料,或者可以是构成活组织或生物的宿主细胞或宿主基因组。在一些情况下,宿主细胞可以是动物细胞或植物细胞,例如,如本文所述。在某些情况下,宿主细胞可以是哺乳动物细胞、人细胞、禽类细胞、爬行动物细胞、牛细胞、马细胞、猪细胞、山羊细胞、绵羊细胞、鸡细胞或火鸡细胞。在某些情况下,宿主细胞可以是玉米细胞、大豆细胞、小麦细胞或稻细胞。

[0083] 如本文所用,“可操作的关联”描述了两个核酸序列(例如1)启动子和2)异源对象序列)之间的功能关系,并且在这样的实例中意味着启动子和异源对象序列(例如目的基因)的取向使得在合适的条件下,启动子驱动异源对象序列的表达。例如,携带启动子和异源对象序列的模板核酸可以是单链的,例如(+)或(-)取向。该模板中启动子与异源对象序列之间的“可操作的关联”意指,无论模板核酸是否以特定状态转录,当其处于合适的状态时(例如,处于(+)取向,在需要的催化因子和NTP等存在的情况下),就被准确地转录。可操作的关联类似地适用于其他核酸对,包括其他组织特异性表达控制序列(例如增强子、阻遏物和微小RNA识别序列)、IR/DR、ITR、UTR或同源区和异源对象序列或编码逆转录病毒RT结构域的序列。

[0084] 如本文所用,术语“引物结合位点序列”或“PBS序列”是指能够与靶核酸序列中包含的区域结合的模板核酸的一部分。在一些情况下,PBS序列是包含与靶核酸序列中包含的区域具有100%同一性的至少3、4、5、6、7或8个碱基的核酸序列。在一些实施例中,引物区包含与靶核酸序列中包含的区域具有100%同一性的至少5、6、7、8个碱基。不希望受理论束缚,在一些实施例中,当模板核酸包含PBS序列和异源对象序列时,PBS序列与靶核酸序列中包含的区域结合,从而允许Pol结构域使用该区域作为DNA聚合的引物,并使用异源对象序列作为模板。

[0085] 如本文所用,“茎环序列”是指具有足够的自互补性以形成茎-环的核酸序列(例如,RNA序列),例如,具有的茎包含至少两个(例如,3、4、5、6、7、8、9或10个)碱基对,以及具有的环具有至少三个(例如,四个)碱基对。茎可能包含错配或凸起。

[0086] 如本文所用,“组织特异性表达控制序列”意指在靶组织中以组织特异性方式例如相对于一个或多个脱靶组织优先在一个或多个中靶组织中增加或降低包含异源对象序列的转录本水平的核酸元件。在一些实施例中,组织特异性表达控制序列优先在靶组织中以组织特异性方式例如相对于一个或多个脱靶组织优先在一个或多个中靶组织中驱动或抑制包含异源对象序列的转录本的转录、活性或半衰期。示例性组织特异性表达控制序列包

括组织特异性启动子、阻遏物、增强子或其组合,以及组织特异性微小RNA识别序列。组织特异性是指中靶(期望或耐受模板核酸的表达或活性的一个或多个组织)和脱靶(不期望或不耐受模板核酸的表达或活性的一个或多个组织)。例如,相对于脱靶组织,组织特异性启动子优先驱动中靶组织中的表达。相反,相对于中靶组织,结合组织特异性微小RNA识别序列的微小RNA优先在脱靶组织中表达,从而减少脱靶组织中模板核酸的表达。因此,关于组织中相关联序列的转录、活性或半衰期,对同一组织(例如靶组织)特异的启动子和微小RNA识别序列具有不同的功能(分别促进和抑制,具有一致的表达水平,即脱靶组织中的高水平微小RNA和中靶组织中的低水平,而启动子驱动中靶组织中的高表达和脱靶组织中的低表达)。

[0087] 目录

[0088] 1)引言

[0089] 2)基因修饰系统

[0090] a)基因修饰系统的多肽组分

[0091] i)书写结构域

[0092] ii)核酸内切酶结构域和DNA结合结构域

[0093] (1)包含Cas结构域的基因修饰多肽

[0094] (2)TAL效应子和锌指核酸酶

[0095] iii)接头

[0096] iv)基因修饰系统的定位序列

[0097] v)基因修饰多肽和系统的进化变体

[0098] vi)内含肽

[0099] vii)另外的结构域

[0100] b)模板核酸

[0101] i)gRNA间隔子和gRNA支架

[0102] ii)异源对象序列

[0103] iii)PBS序列

[0104] iv)示例性模板序列

[0105] c)具有诱导活性的gRNA

[0106] d)基因修饰系统中的环状RNA和核酶

[0107] e)靶核酸位点

[0108] f)第二链切口

[0109] 3)组合物和系统的产生

[0110] 4)应用

[0111] a)治疗性应用

[0112] b)应用于植物

[0113] 5)施用和递送

[0114] a)组织特异性活性/施用

[0115] i)启动子

[0116] ii)微小RNA

[0117] b)病毒载体及其组分

[0118] c)AAV施用

[0119] d)脂质纳米颗粒

[0120] 6)试剂盒、制品和药物组合物

[0121] 7)化学、制造和控制(CMC)

[0122] 引言

[0123] 本披露涉及用于例如体内或体外靶向、编辑、修饰或操纵细胞、组织或受试者中DNA序列中的一个或多个位置处的DNA序列(例如,将异源对象序列插入哺乳动物基因组的靶位点)的组合物、系统和方法。异源对象DNA序列可以包括例如取代、缺失、插入,例如编码序列、调节序列或基因表达单元。

[0124] 更特别地,本披露提供了用于改变目的基因组DNA序列的基于DNA聚合酶(Pol)的系统,例如,通过向目的序列中插入一个或多个核苷酸、使目的序列缺失一个或多个核苷酸或取代目的序列中的一个或多个核苷酸。

[0125] Cas9相关功能与聚合酶功能的融合可用于驱动基因组DNA的修饰。聚合酶功能可以是例如从核酸模板合成DNA的DNA聚合酶。核酸模板可以是例如DNA或RNA。在可以使用RNA模板的DNA聚合酶(例如,RNA依赖性DNA聚合酶,例如逆转录酶)的情况下,可以向基因修饰多肽组分提供例如上述的模板RNA。一个这样的实例是DNA聚合酶θ(由POLQ编码,其多肽产物在本文中可以称为"POLQ"或"Polθ"),其是真核DNA聚合酶,已被证明使用DNA或RNA作为模板。Chandramouly等人2021(DOI:10.1126/sciadv.abf1771)。因此,当将与POLQ(或POLQ的组分)融合(任选地通过接头)的Cas9功能施用到具有模板RNA的生物或细胞时,该Cas9功能可以用作基因组修饰的驱动子,该模板RNA靶向期望修饰(经由gRNA间隔子)的基因组位点、募集Cas9功能(经由gRNA支架)、并对DNA合成进行引发和模板化(经由模板RNA)。

[0126] 本披露部分地提供了包含基因修饰多肽组分和模板核酸(例如,模板RNA)组分的基因修饰系统。在一些实施例中,基因修饰系统可用于将改变引入基因组中的靶位点。在一些实施例中,基因修饰多肽组分包含书写结构域(例如,逆转录酶结构域)、DNA结合结构域和核酸内切酶结构域(例如,切口酶结构域)。在一些实施例中,模板核酸(例如,模板RNA)包含结合基因组中的靶位点(例如,结合靶位点的第二链)的序列(例如,gRNA间隔子)、结合基因修饰多肽组分的序列(例如,gRNA支架)、异源对象序列和PBS序列。不希望受理论束缚,认为模板核酸(例如模板RNA)结合基因组中靶位点的第二链,并结合基因修饰多肽组分(例如,将多肽组分定位于基因组中的靶位点)。认为基因修饰多肽组分的核酸内切酶(例如,切口酶)切割靶位点(例如,靶位点的第一链),例如,允许PBS序列与跟靶位点的第一链上待改变位点相邻的序列结合。认为多肽组分的书写结构域(例如,逆转录酶结构域)使用与包含模板核酸的PBS序列作为引物和模板核酸的异源对象序列作为模板以例如聚合与异源对象序列互补的序列的互补序列结合的靶位点的第一链。不希望受理论束缚,认为选择合适的异源对象序列可导致在靶位点处取代、缺失和/或插入一个或多个核苷酸。

[0127] 基因修饰系统

[0128] 在一些实施例中,本文所述的基因修饰系统包含:(A)基因修饰多肽或编码该基因修饰多肽的核酸,其中该基因修饰多肽包含(i)逆转录酶结构域和(x)含有DNA结合功能的核酸内切酶结构域或(y)核酸内切酶结构域和单独的DNA结合结构域;和(B)模板RNA。在一

些实施例中,基因修饰多肽作为基本上自主的蛋白质机器,能够将模板核酸序列整合到靶DNA分子中(例如,在哺乳动物宿主细胞中,例如宿主细胞中的基因组DNA分子),基本上不依赖于宿主机器。例如,基因修饰蛋白可包含DNA结合结构域、逆转录酶结构域和核酸内切酶结构域。在一些实施例中,DNA结合功能可涉及将蛋白质引导至DNA序列(例如gRNA间隔子)的RNA组分。在其他实施例中,基因修饰多肽可包含逆转录酶结构域和核酸内切酶结构域。基因修饰系统的RNA模板元件通常与基因修饰多肽元件异源,并提供要插入(逆转录)到宿主基因组中的对象序列。在一些实施例中,基因修饰多肽能够靶向引发的逆转录。在一些实施例中,基因修饰多肽能够进行第二链合成。

[0129] 在一些实施例中,基因修饰系统与第二多肽组合。在一些实施例中,第二多肽可包含核酸内切酶结构域。在一些实施例中,第二多肽可包含聚合酶结构域,例如逆转录酶结构域。在一些实施例中,第二多肽可包含DNA依赖性DNA聚合酶结构域。在一些实施例中,第二多肽有助于完成基因组编辑,例如,通过有助于第二链合成或DNA修复解离。

[0130] 功能性基因修饰多肽可以由不相关的DNA结合结构域、逆转录结构域和核酸内切酶结构域构成。这种模块化结构允许组合功能性结构域,例如dCas9(DNA结合)、MMLV逆转录酶(逆转录)、FokI(核酸内切酶)。在一些实施例中,多个功能性结构域可以来自单一蛋白质,例如,Cas9或Cas9切口酶(DNA结合、核酸内切酶)。

[0131] 在一些实施例中,基因修饰多肽包含一个或多个结构域,它们共同促进1)结合模板核酸,2)结合靶DNA分子,和3)促进模板核酸的至少一部分的整合进入靶DNA。在一些实施例中,基因修饰多肽是工程改造的多肽,例如,相对于天然存在的序列具有一个或多个氨基酸取代。在一些实施例中,基因修饰多肽包含两个或更多个相对于彼此异源的结构域,例如,通过在其他方面是野生型的结构域的异源融合(或其他缀合物),或经修饰的结构域的融合,例如,通过异源子结构域或其他经取代的结构域的替代或融合。例如,在一些实施例中,以下中的一项或多项:该RT结构域与该DBD异源;该DBD与该核酸内切酶结构域异源;或该RT结构域与该核酸内切酶结构域异源。

[0132] 在一些实施例中,用于该系统中的模板RNA分子从5′至3′包含(1)gRNA间隔子;(2)gRNA支架;(3)异源对象序列;(4)引物结合位点(PBS)序列。在一些实施例中:

[0133] (1)是约18-22nt(例如,20nt)的gRNA间隔子

[0134] (2)是包含一个或多个发夹环(例如,1、2、或3个环)的gRNA支架,用于使模板与Cas结构域例如切口酶Cas9结构域相关联。在一些实施例中,gRNA支架从5′到3′包含序列GTTTTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCC(SEQ ID NO:4008)。

[0135] (3)在一些实施例中,异源对象序列长度是例如7-74,例如10-20、20-30、30-40、40-50、50-60、60-70、或70-80nt或80-90nt。在一些实施例中,序列的第一个(最5')碱基不是C。

[0136] (4)在一些实施例中,在切口发生后结合靶引发序列的PBS序列是例如3-20nt,例如7-15nt,例如12-14nt。在一些实施例中,PBS序列具有40%-60%的GC含量。

[0137] 在一些实施例中,与系统相关联的第二gRNA可能有助于驱动完全整合。在一些实施例中,第二gRNA可以靶向距第一链切口0-200nt,例如距第一链切口0-50、50-100、100-200nt的位置。在一些实施例中,第二gRNA只能在进行编辑后结合其靶序列,例如,gRNA结合

存在于异源对象序列中但不存在于初始靶序列中的序列。

[0138] 在一些实施例中,本文所述的基因修饰系统用于在HEK293、K562、U2OS、或HeLa细胞中进行编辑。在一些实施例中,基因修饰系统用于在原代细胞(例如,来自E18.5小鼠的原代皮层神经元)中进行编辑。

[0139] 在一些实施例中,如本文所述的基因修饰多肽包含含有MoMLV RT序列或其变体的逆转录酶或RT结构域(例如,如本文所述)。在实施例中,MoMLV RT序列包含一种或多种选自以下的突变:D200N、L603W、T330P、T306K、W313F、D524G、E562Q、D583N、P51L、S67R、E67K、T197A、H204R、E302K、F309N、L435G、N454K、H594Q、D653N、R110S、和K103L。在实施例中,MoMLV RT序列包含突变(例如D200N、L603W和T330P)的组合,任选地还包括T306K和/或W313F。

[0140] 在一些实施例中,核酸内切酶结构域(例如,如本文所述)包含nCAS9,例如,包含H840A突变。

[0141] 在一些实施例中,异源对象序列(例如,如本文所述的系统的)长度是约1-50、50-100、100-200、200-300、300-400、400-500、500-600、600-700、700-800、800-900、900-1000或更多个核苷酸。

[0142] 在一些实施例中,RT和核酸内切酶结构域通过柔性接头连接,例如,包含氨基酸序列SGGSSGGSSGSETPGTSESATPESSGGSSGGSS(SEQ ID NO:4006)。

[0143] 在一些实施例中,核酸内切酶结构域相对于RT结构域在N末端。在一些实施例中,核酸内切酶结构域相对于RT结构域在C末端。

[0144] 在一些实施例中,该系统通过TPRT将异源对象序列掺入靶位点,例如,如本文所述。

[0145] 在一些实施例中,基因修饰多肽包含DNA结合结构域。在一些实施例中,基因修饰多肽包含RNA结合结构域。在一些实施例中,RNA结合结构域包含B-盒蛋白、MS2外壳蛋白、dCas、或本文表中序列的元件的RNA结合结构域。在一些实施例中,RNA结合结构域能够以比参考RNA结合结构域更大的亲和力结合模板RNA。

[0146] 在一些实施例中,基因修饰系统能够在靶位点中产生至少45、50、55、60、65、70、75、80、85、90、95、或100个核苷酸(并且任选地没有超过500、400、300、200或100个核苷酸)插入。在一些实施例中,基因修饰系统能够在靶位点中产生至少1、2、3、4、5、6、7、8、9、10、15、20、25、30、35、40、45、50、55、60、65、70、75、80、85、90、95、或100个核苷酸(并且任选地不超过500、400、300、200或100个核苷酸)插入。在一些实施例中,基因修饰系统能够在靶位点中产生至少0.2、0.3、0.4、0.5、0.6、0.7、0.8、0.9、1、1.5、2、2.5、3、3.5、4、4.5、5、5.5、6、6.5、7、7.5、8、8.5、9、9.5或10千碱基(并且任选地不超过1、5、10或20千碱基)插入。在一些实施例中,基因修饰系统能够产生至少81、85、90、95、100、110、120、130、140、150、160、170、180、190、或200个核苷酸(并且任选地不超过500、400、300或200个核苷酸)缺失。在一些实施例中,基因修饰系统能够产生至少81、85、90、95、100、110、120、130、140、150、160、170、180、190、或200个核苷酸(并且任选地不超过500、400、300或200个核苷酸)缺失。在一些实施例中,基因修饰系统能够产生至少1、2、3、4、5、6、7、8、9、10、15、20、25、30、35、40、45、50、55、60、65、70、75、80、85、90、95、100、110、120、130、140、150、160、170、180、190、或200个核苷酸(并且任选不超过500、400、300或200个核苷酸)缺失。在一些实施例中,基因修饰系统

能够产生至少0.2、0.3、0.4、0.5、0.6、0.7、0.8、0.9、1、1.5、2、2.5、3、3.5、4、4.5、5、5.5、6、6.5、7、7.5、8、8.5、9、9.5或10千碱基(并且任选不超过1、5、10、或20千碱基)缺失。在一些实施例中,基因修饰系统能够在靶位点中产生至少1、2、3、4、5、6、7、8、9、10、15、20、25、30、35、40、45、50、60、70、80、90、或100个或更多个核苷酸取代。在一些实施例中,基因修饰系统能够在靶位点中产生1-2、2-3、3-4、4-5、5-10、10-15、15-20、20-30、30-40、40-50、50-60、60-70、70-80、80-90或90-100个核苷酸取代。

[0147] 在一些实施例中,取代是转位突变。在一些实施例中,取代是颠换突变。在一些实施例中,取代将腺嘌呤转化为胸腺嘧啶,腺嘌呤转化为鸟嘌呤,腺嘌呤转化为胞嘧啶,鸟嘌呤转化为胸腺嘧啶,鸟嘌呤转化为胞嘧啶,鸟嘌呤转化为腺嘌呤,胸腺嘧啶转化为胞嘧啶,胸腺嘧啶转化为腺嘌呤,胸腺嘧啶转化为鸟嘌呤,胞嘧啶转化为腺嘌呤,胞嘧啶转化为鸟嘌呤,或胞嘧啶转化为胸腺嘧啶。

[0148] 在一些实施例中,插入、缺失、取代或其组合增加或减少基因的表达(例如转录或翻译)。在一些实施例中,插入、缺失、取代或其组合通过改变、添加或缺失启动子或增强子中的序列(例如结合转录因子的序列)来增加或减少基因的表达(例如转录或翻译)。在一些实施例中,插入、缺失、取代或其组合改变基因的翻译(例如改变氨基酸序列),插入或缺失起始或终止密码子,改变或固定基因的翻译框架。在一些实施例中,插入、缺失、取代或其组合改变基因的剪接,例如通过插入、缺失或改变剪接受体或供体位点。在一些实施例中,插入、缺失、取代或其组合改变转录本或蛋白质半衰期。在一些实施例中,插入、缺失、取代或其组合改变细胞中的蛋白质定位(例如从细胞质到线粒体,从细胞质到细胞外空间(例如添加分泌标签))。在一些实施例中,插入、缺失、取代或其组合改变(例如改善)蛋白质折叠(例如以防止错误折叠蛋白质的积累)。在一些实施例中,插入、缺失、取代或其组合改变、增加、降低基因的活性,例如由基因编码的蛋白质的活性。

[0149] 示例性的基因修饰多肽、包含它们的系统以及使用它们的方法例如描述于PCT/US2021/020948,其关于逆转录病毒RT结构域(包括其中的氨基酸和核酸序列)通过援引并入本文。

[0150] 示例性的基因修饰多肽和逆转录病毒RT结构域序列也描述于例如2021年3月4日提交的国际申请号PCT/US 21/20948,例如其中的表30、表31和表44;整个申请关于逆转录病毒RT通过援引并入本文,例如在所述序列和表中。因此,本文所述的基因修饰多肽可包含根据本段提及的任何表的氨基酸序列或其结构域(例如,逆转录病毒RT结构域),或前述中任一个的功能片段或变体,或与其具有至少70%、80%、85%、90%、95%或99%同一性的氨基酸序列。

[0151] 在一些实施例中,用于本文所述的任何系统中的多肽可以是基于多个同源蛋白的对齐的多肽序列的分子重建或遗传重建。在一些实施例中,用于本文所述任何系统的逆转录酶结构域可以是分子重建或遗传重建,或者可以基于来自相同或不同来源的逆转录酶结构域的比对在特定残基处进行修饰。基于本文提供的登录号,技术人员可以例如通过使用常规序列分析工具(如基本局部比对搜索工具(BLAST)或CD-搜索(用于保守结构域分析))来比对多肽或核酸序列。可以基于共有序列创建分子重建,例如使用在Ivics等人,Cell[细胞]1997,501-510;Wagstaff等人,Molecular Biology and Evolution[分子生物学与进化]2013,88-99中描述的方法。

[0152] 基因修饰系统的多肽组分

[0153] 在一些实施例中,基因修饰多肽具有DNA靶位点结合、模板核酸(例如RNA)结合、DNA靶位点切割和模板核酸(例如RNA)书写(例如逆转录)的功能。在一些实施例中,每个功能都包含在不同的结构域内。在一些实施例中,功能可以归属于两个或更多个结构域(例如,两个或更多个结构域一起展示该功能)。在一些实施例中,两个或更多个结构域可以具有相同或相似的功能(例如,两个或更多个结构域各自独立地具有DNA结合功能,例如对于两个不同的DNA序列)。在其他实施例中,一个或多个结构域可能能够实现一种或多种功能,例如,Cas9结构域能够同时实现DNA结合和靶位点切割。在一些实施例中,这些结构域都位于单个多肽内。在一些实施例中,第一结构域在一个多肽中并且第二结构域在第二多肽中。例如,在一些实施例中,序列可以在第一多肽和第二多肽之间断裂,例如,其中第一多肽包含逆转录酶(RT)结构域并且其中第二多肽包含DNA结合结构域和核酸内切酶结构域,例如切口酶结构域。作为另一个实例,在一些实施例中,第一多肽和第二多肽各自包含DNA结合结构域(例如,第一DNA结合结构域和第二DNA结合结构域)。在一些实施例中,第一和第二多肽可以通过断裂内含肽在翻译后结合在一起以形成单个基因修饰多肽。

[0154] 在一些方面,本文所述的基因修饰多肽包含(例如,本文所述的系统包含基因修饰多肽,其包含):1)Cas结构域(例如,Cas切口酶结构域,例如,Cas9切口酶结构域);2)表D的逆转录酶(RT)结构域,或与其具有至少70%、75%、80%、85%、90%、95%、97%、98%或99%同一性的序列,其中该RT结构域位于该Cas结构域的C末端;和位于该RT结构域和该Cas结构域之间的接头,其中该接头具有来自表D中与该RT结构域同一行的序列,或者与其具有至少70%、75%、80%、85%、90%、95%、97%、98%或99%同一性的序列。

[0155] 在一些实施例中,RT结构域具有与表D的RT结构域具有100%同一性的序列,并且接头具有来自表D中与该RT结构域同一行的接头序列具有100%同一性的序列。在一些实施例中,Cas结构域包含表4的序列或与其具有至少70%、75%、80%、85%、90%、95%、98%或99%同一性的序列。在一些实施例中,基因修饰多肽包含根据序列表中SEQ ID NO:1-3332中任一个的氨基酸序列,或与其具有至少70%、75%、80%、85%、90%、95%、97%、98%或99%同一性的序列。

[0156] 在一些实施例中,基因修饰多肽包含Cas结构域和接头之间的GG氨基酸序列、RT结构域和第二NLS之间的AG氨基酸序列、和/或接头和RT结构域之间的GG氨基酸序列。在一些实施例中,基因修饰多肽包含含有第一NLS和Cas结构域的SEQ ID NO:4000的序列,或与其具有至少70%、75%、80%、85%、90%、95%、98%或99%同一性的序列。在一些实施例中,基因修饰多肽包含含有第二NLS的SEQ ID NO:4001的序列,或与其具有至少70%、75%、80%、85%、90%、95%、98%或99%同一性的序列。

[0157] 示例性N末端NLS-Cas9结构域

[0158] MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNR

EKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL
LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRK
LINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTV
KVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQ
NGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLIT
QRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKD
FQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNF
FKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIA
RKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIK
LPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQI
SEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH
QSITGLYETRIDLSQLGGDGG(SEQ ID NO:4000)

[0159] 示例性C末端序列

[0160] AGKRTADGSEFEKRTADGSEFESPKKKAKVE(SEQ ID NO:4001)

[0161] 书写结构域

[0162] 在本发明的某些方面,基因修饰系统的书写结构域利用聚合酶功能来驱动对基因组DNA的修饰。聚合酶功能可以是例如从核酸模板合成DNA的DNA聚合酶。核酸模板可以是例如DNA或RNA。在可以使用RNA模板的DNA聚合酶(例如,RNA依赖性DNA聚合酶,例如逆转录酶)的情况下,可以向基因修饰多肽组分提供例如上述的模板RNA。一个这样的实例是DNA聚合酶θ(由POLQ编码,其多肽产物在本文中可以称为"POLQ"或"Polθ"),其是真核DNA聚合酶,已被证明使用DNA或RNA作为模板。Chandramouly等人2021(DOI:10.1126/sciadv.abf1771)。因此,当将与POLQ(或POLQ的组分)融合(任选地通过接头)的Cas9功能施用到具有模板RNA的生物或细胞时,该Cas9功能可以用作基因组修饰的驱动子,该模板RNA靶向期望修饰(经由gRNA间隔子)的基因组位点、募集Cas9功能(经由gRNA支架)、并对DNA合成进行引发和模板化(经由模板RNA)。

[0163] 使用DNA为模板的DNA聚合酶也可以掺入具有Cas9功能的融合体中,从而实现基因组修饰。在这种情况下,模板核酸是sgRNA和DNA模板的融合体,通过共价键或接头末端到末端彼此连接。Polθ(或其组分)也可以以这种方式发挥作用,因为Polθ可以从DNA模板合成DNA。应当理解,如本文所述,涉及模板RNA的实施例可以包括包含核糖核苷酸的模板核酸,或包含核糖核苷酸和脱氧核糖核苷酸的模板核酸(例如,包含与一个或多个DNA区域偶联的一个或多个RNA区域的模板RNA)。在一些实施例中,本文所述的基因修饰多肽包含具有根据表1的氨基酸序列,或与其具有至少70%、80%、85%、90%、95%、97%、98%或99%同一性的序列的聚合酶结构域。在一些实施例中,本文所述的核酸编码具有根据表1的氨基酸序列,或与其具有至少70%、80%、85%、90%、95%、97%、98%或99%同一性的序列的聚合酶结构域。

[0164] 应当理解,如本文所述,涉及逆转录酶或逆转录酶结构域的实施例可以包括如表1中所列的聚合酶。

[0165] 表1:用于基因组工程多肽的示例性聚合酶。

[0166]

行	Pol	序列	长度
1	Eta（η）	MATGQDRVVALVDMDCFFVQVEQRQNPHLRNKPCAVVQYKSWKGGGIIAVSYEARAFGVTRSMWADDAKKLCPDLLLAQVRESR GKANLTKYREASVEVMEIMSRFAVIERASIDEAYVDLTSAVQERLQKLQGQPISADLLPSTYIEGLPQGPTTAEETVQKEGMRKQGLF QWLDSLQIDNLTSPDLQLTVGAVIVEEMRAAIERETGFQCSAGISHNKVLAKLACGLNKPNRQTLVSHGSVPQLFSQMPIRKIRSLGG KLGASVIEILGIEYMGELTQFTESQLQSHFGEKNGSWLYAMCRGIEHDPVKPRQLPKTIGCSKNFPGKTALATREQVQWWLLQLAQ ELEERLTKDRNDNDRVATQLVVSIRVQGDKRLSSLRRCCALTRYDAHKMSHDAFTVIKNCNTSGIQTEWSPPLTMLFLCATKFSASA PSSSTDITSFLSSDPSSLPKVPVTSSEAKTQGSGPAVTATKKATTSLESFFQKAAERQKVKEASLSSLTAPTQAPMSNSPSKPSLPF QTSQSTGTEPFFKQKSLLLKQKQLNNSSVSSPQQNPWSNCKALPNSLPTEYPGCVPVCEGVSKLEESSKATPAEMDLAHNSQSM HASSASKSVLEVTQKATPNPSLLAAEDQVPCEKCGSLVPVWDMPEHMDYHFALELQKSFLQPHSSNPQVVSAVSHQGKRNPKSPL ACTNKRPRPEGMQTLESFFKPLTH	713
2	Iota（ι）	MEKLGVEPEEEGGGDDDEEDAEAWAMELADVGAAASSQGVHDQVLPTPNASSRVIVHVDLDCFYAQVEMISNPELKDKPLGVQQ KYLVVTCNYEARKLGVKKLMNVRDAKEKCPQLVLVNGEDLTRYREMSYKVTELLEEFSPVVERLGFDENFVDLTEMVEKRLQQLQS DELSAVTVSGHVYNNQSINLLDVLHIRLLVGSQIAAEMREAMYNQLGLTGCAGVASNKLLAKLVSGVFKPNQQTVLLPESCQHLIHSL NHIKEIPGIGYKTAKCLEALGINSVRDLQTFSPKILEKELGISVAQRIQKLSFGEDNSPVILSGPPQSFSEEDSFKKCSSEVEAKNKIEEL LASLLNRVCQDGRKPHTVRLIIRRYSSEKHYGRESRQCPIPSHVIQKLGTGNYDVMTPMVDILMKLFRNMVNVKMPFHLTLLSVCFC NLKALNTAKKGLIDYYLMPSLSTTSRSGKHSFKMKDTHMEDFPKDKETNRDFLPSGRIESTRTRESPLDTTNFSKEKDINEFPLCSLP EGVDQEVFKQLPVDIQEEILSGKSREKFQGKGSVSCPLHASRGVLSFFSKKQMQDIPINPRDHLSSSKQVSSVSPCEPGTSGFNSS SSSYMSSQKDYSYYLDNRLKDERISQGPKEPQGFHFTNSNPAVSAFHSFPNLQSEQLFSRNHTTDSHKQTVATDSHEGLTENREP DSVDEKITFPSDIDPQVFYELPEAVQKELLAEWKRAGSDFHIGHK	740
3	Kappa（κ）	MDSTKEKCDSYKDDLLLRMGLNDNKAGMEGLDKEKINKIIMEATKGSRFYGNELKKEKQVNQRIENMMQQKAQITSQQLRKAQLQV DRFAMELEQSRNLSNTIVHIDMDAFYAAVEMRDNPELKDKPIAVGSMSMLSTSNYHARRFGVRAAMPGFIAKRLCPQLIIVPPNFDK YRAVSKEVKEILADYDPNFMAMSLDEAYLNITKHLEERQNWPEDKRRYFIKMGSSVENDNPGKEVNKLSEHERSISPLLFEESPSDV QPPGDPFQVNFEEQNNPQILQNSVVFGTSAQEVVKEIRFRIEQKTTLTASAGIAPNTMLAKVCSDKNKPNGQYQILPNRQAVMDFIK DLPIRKVSGIGKVTEKMLKALGIITCTELYQQRALLSLLFSETSWHYFLHISLGLGSTHLTRDGERKSMSVERTFSEINKAEEQYSLCQ ELCSELAQDLQKERLKGRTVTIKLKNVNFEVKTRASTVSSVVSTAEEIFAIAKELLKTEIDADFPHPLRLRLMGVRISSFPNEEDRKHQ QRSIIGFLQAGNQALSATECTLEKTDKDKFVKPLEMSHKKSFFDKKRSERKWSHQDTFKCEAVNKQSFQTSQPFQVLKKKMNENLE ISENSDDCQILTCPVCFRAQGCISLEALNKHVDECLDGPSISENFKMFSCSHVSATKVNKKENVPASSLCEKQDYEAHPKIKEISSVD CIALVDTIDNSSKAESIDALSNKHSKEECSSLPSKSFNIEHCHQNSSSTVSLENEDVGSFRQEYRQPYLCEVKTGQALVCPVCNVEQ KTSDLTLFNVHVDVCLNKSFIQELRKDKFNPVNQPKESSRSTGSSSGVQKAVTRTKRPGLMTKYSTSKKIKPNNPKHTLDIFFK	870
4	REV1	MRRGGWRKRAENDGWETWGGYMAAKVQKLEEQFRSDAAMQKDGTSSTIFSGVAIYVNGYTDPSAEELRKLMMLHGGQYHVYYS RSKTTHIIATNLPNAKIKELKGEKVIRPEWIVESIKAGRLLSYIPYQLYTKQSSVQKGLSFNPVCRPEDPLPGPSNIAKQLNNRVNHIVK KIETENEVKVNGMNSWNEEDENNDFSFVDLEQTSPGRKQNGIPHPRGSTAIFNGHTPSSNGALKTQDCLVPMVNSVASRLSPAFS QEEDKAEKSSTDFRDCTLQQLQQSTRNTDALRNPHRTNSFSLSPLHSNTKINGAHHSTVQGPSSTKSTSSVSTFSKAAPSVPSKPS DCNFISNFYSHSRLHHISMWKCELTEFVNTLQRQSNGIFPGREKLKKMKTGRSALVVTDTGDMSVLNSPRHQSCIMHVDMDCFFVS VGIRNRPDLKGKPVAVTSNRGTGRAPLRPGANPQLEWQYYQNKILKGKAADIPDSSLWENPDSAQANGIDSVLSRAEIASCSYEAR QLGIKNGMFFGHAKQLCPNLQAVPYDFHAYKEVAQTLYETLASYTHNIEAVSCDEALVDITEILAETKLTPDEFANAVRMEIKDQTKC AASVGIGSNILLARMATRKAKPDGQYHLKPEEVDDFIRGQLVTNLPGVGHSMESKLASLGIKTCGDLQYMTMAKLQKEFGPKTGQM LYRFCRGLDDRPVRTEKERKSVSAEINYGIRFTQPKEAEAFLLSLSEEIQRRLEATGMKGKRLTLKIMVRKPGAPVETAKFGGHGICD NIARTVTLDQATDNAKIIGKAMLNMFHTMKLNISDMRGVGIHVNQLVPTNLNPSTCPSRPSVQSSHFPSGSYSVRDVFQVQKAKKST EEEHKEVFRAAVDLEISSASRTCTFLPPFPAHLPTSPDTNKAESSGKWNGLHTPVSVQSRLNLSIEVPSPSQLDQSVLEALPPDLRE QVEQVCAVQQAESHGDKKKEPVNGCNTGILPQPVGTVLLQIPEPQESNSDAGINLIALPAFSQVDPEVFAALPAELQRELKAAYDQR QRQGENSTHQQSASASVPKNPLLHLKAAVKEKKRNKKKKTIGSPKRIQSPLNNKLLNSPAKTLPGACGSPQKLIDGFLKHEGPPAEK PLEELSASTSGVPGLSSLQSDPAGCVRPPAPNLAGAVEFNDVKTLLREWITTISDPMEEDILQVVKYCTDLIEEKDLEKLDLVIKYMKR LMQQSVESVWNMAFDFILDNVQVVLQQTYGSTLKVT	1251
5	Beta（β）	MSKRKAPQETLNGGITDMLTELANFEKNVSQAIHKYNAYRKAASVIAKYPHKIKSGAEAKKLPGVGTKIAEKIDEFLATGKLRKLEKIR QDDTSSSINFLTRVSGIGPSAARKFVDEGIKTLEDLRKNEDKLNHHQRIGLKYFGDFEKRIPREEMLQMQDIVLNEVKKVDSEYIATV CGSFRRGAESSGDMDVLLTHPSFTSESTKQPKLLHQVVEQLQKVHFITDTLSKGETKFMGVCQLPSKNDEKEYPHRRIDIRLIPKDQ YYCGVLYFTGSDIFNKNMRAHALEKGFTINEYTIRPLGVTGVAGEPLPVDSEKDIFDYIQWKYREPKDRSE	335
6	Lambda(λ)	MDPRGILKAFPKRQKIHADASSKVLAKIPRREEGEEAEEWLSSLRAHVVRTGIGRARAELFEKQIVQHGGQLCPAQGPGVTHIVVDE GMDYERALRLLRLPQLPPGAQLVKSAWLSLCLQERRLVDVAGFSIFIPSRYLDHPQPSKAEQDASIPPGTHEALLQTALSPPPPPTR PVSPPQKAKEAPNTQAQPISDDEASDGEETQVSAADLEALISGHYPTSLEGDCEPSPAPAVLDKWVCAQPSSQKATNHNLHITEKL EVLAKAYSVQGDKWRALGYAKAINALKSFHKPVTSYQEACSIPGIGKRMAEKIIEILESGHLRKLDHISESVPVLELFSNIWGAGTKTA QMWYQQGFRSLEDIRSQASLTTQQAIGLKHYSDFLERMPREEATEIEQTVQKAAQAFNSGLLCVACGSYRRGKATCGDVDVLITHP DGRSHRGIFSRLLDSLRQEGFLTDDLVSQEENGQQQKYLGVCRLPGPGRRHRRLDIIVVPYSEFACALLYFTGSAHFNRSMRALAK TKGMSLSEHALSTAVVRNTHGCKVGPGRVLPTPTEKDVFRLLGLPYREPAERDW	575
7	Mu（μ）	MLPKRRRARVGSPSGDAASSTPPSTRFPGVAIYLVEPRMGRSRRAFLTGLARSKGFRVLDACSSEATHVVMEETSAEEAVSWQER RMAAAPPGCTPPALLDISWLTESLGAGQPVPVECRHRLEVAGPRKGPLSPAWMPAYACQRPTPLTHHNTGLSEALEILAEAAGFEG SEGRLLTFCRAASVLKALPSPVTTLSQLQGLPHFGEHSSRVVQELLEHGVCEEVERVRRSERYQTMKLFTQIFGVGVKTADRWYRE GLRTLDDLREQPQKLTQQQKAGLQHHQDLSTPVLRSDVDALQQVVEEAVGQALPGATVTLTGGFRRGKLQGHDVDFLITHPKEGQ EAGLLPRVMCRLQDQGLILYHQHQHSCCESPTRLAQQSHMDAFERSFCIFRLPQPPGAAVGGSTRPCPSWKAVRVDLVVAPVSQF PFALLGWTGSKLFQRELRRFSRKEKGLWLNSHGLFDPEQKTFFQAASEEDIFRHLGLEYLPPEQRNA	494
8	TdT	MDPPRASHLSPRKKRPRQTGALMASSPQDIKFQDLVVFILEKKMGTTRRAFLMELARRKGFRVENELSDSVTHIVAENNSGSDVLE WLQAQKVQVSSQPELLDVSWLIECIRAGKPVEMTGKHQLVVRRDYSDSTNPGPPKTPPIAVQKISQYACQRRTTLNNCNQIFTDAF DILAENCEFRENEDSCVTFMRAASVLKSLPFTIISMKDTEGIPCLGSKVKGIIEEIIEDGESSEVKAVLNDERYQSFKLFTSVFGVGLKT SEKWFRMGFRTLSKVRSDKSLKFTRMQKAGFLYYEDLVSCVTRAEAEAVSVLVKEAVWAFLPDAFVTMTGGFRRGKKKMGHDVDF LITSPGSTEDEEQLLQKVMNLWEKKGLLLYYDLVESTFEKLRLPSRKVDALDHFQKCFLIFKLPRQRVDSDQSSWQEGKTWKAIRVD LVLCPYERRAFALLGWTGSRQFERDLRRYATHERKMILDNHALYDKTKRIFLKAESEEEIFAHLGLDYIEPWERNA	509
9	Theta（θ）	MNLLRRSGKRRRSESGSDSFSGSGGDSSASPQFLSGSVLSPPPGLGRCLKAAAAGECKPTVPDYERDKLLLANWGLPKAVLEKYH SFGVKKMFEWQAECLLLGQVLEGKNLVYSAPTSAGKTLVAELLILKRVLEMRKKALFILPFVSVAKEKKYYLQSLFQEVGIKVDGYMG STSPSRHFSSLDIAVCTIERANGLINRLIEENKMDLLGMVVVDELHMLGDSHRGYLLELLLTKICYITRKSASCQADLASSLSNAVQIVG MSATLPNLELVASWLNAELYHTDFRPVPLLESVKVGNSIYDSSMKLVREFEPMLQVKGDEDHVVSLCYETICDNHSVLLFCPSKKW CEKLADIIAREFYNLHHQAEGLVKPSECPPVILEQKELLEVMDQLRRLPSGLDSVLQKTVPWGVAFHHAGLTFEERDIIEGAFRQGLI RVLAATSTLSSGVNLPARRVIIRTPIFGGRPLDILTYKQMVGRAGRKGVDTVGESILICKNSEKSKGIALLQGSLKPVRSCLQRREGEE VTGSMIRAILEIIVGGVASTSQDMHTYAACTFLAASMKEGKQGIQRNQESVQLGAIEACVMWLLENEFIQSTEASDGTEGKVYHPTH LGSATLSSSLSPADTLDIFADLQRAMKGFVLENDLHILYLVTPMFEDWTTIDWYRFFCLWEKLPTSMKRVAELVGVEEGFLARCVKG KVVARTERQHRQMAIHKRFFTSLVLLDLISEVPLREINQKYGCNRGQIQSLQQSAAVYAGMITVFSNRLGWHNMELLLSQFQKRLTF	2590

[0167]

行	Pol	序列	长度
		GIQRELCDLVRVSLLNAQRARVLYASGFHTVADLARANIVEVEVILKNAVPFKSARKAVDEEEEAVEERRNMRTIWVTGRKGLTERE AAALIVEEARMILQQDLVEMGVQWNPCALLHSSTCSLTHSESEVKEHTFISQTKSSYKKLTSKNKSNTIFSDSYIKHSPNIVQDLNKSR EHTSSFNCNFQNGNQEHQTCSIFRARKRASLDINKEKPGASQNEGKTSDKKVVQTFSQKTKKAPLNFNSEKMSRSFRSWKRRKHL KRSRDSSPLKDSGACRIHLQGQTLSNPSLCEDPFTLDEKKTEFRNSGPFAKNVSLSGKEKDNKTSFPLQIKQNCSWNITLTNDNFVE HIVTGSQSKNVTCQATSVVSEKGRGVAVEAEKINEVLIQNGSKNQNVYMKHHDIHPINQYLRKQSHEQTSTITKQKNIIERQMPCEAV SSYINRDSNVTINCERIKLNTEENKPSHFQALGDDISRTVIPSEVLPSAGAFSKSEGQHENFLNISRLQEKTGTYTTNKTKNNHVSDLG LVLCDFEDSFYLDTQSEKIIQQMATENAKLGAKDTNLAAGIMQKSLVQQNSMNSFQKECHIPFPAEQHPLGATKIDHLDLKTVGTMK QSSDSHGVDILTPESPIFHSPILLEENGLFLKKNEVSVTDSQLNSFLQGYQTQETVKPVILLIPQKRTPTGVEGECLPVPETSLNMSDS LLFDSFSDDYLVKEQLPDMQMKEPLPSEVTSNHFSDSLCLQEDLIKKSNVNENQDTHQQLTCSNDESIIFSEMDSVQMVEALDNVDI FPVQEKNHTVVSPRALELSDPVLDEHHQGDQDGGDQDERAEKSKLTGTRQNHSFIWSGASFDLSPGLQRILDKVSSPLENEKLKS MTINFSSLNRKNTELNEEQEVISNLETKQVQGISFSSNNEVKSKIEMLENNANHDETSSLLPRKESNIVDDNGLIPPTPIPTSASKLTFP GILETPVNPWKTNNVLQPGESYLFGSPSDIKNHDLSPGSRNGFKDNSPISDTSFSLQLSQDGLQLTPASSSSESLSIIDVASDQNLFQ TFIKEWRCKKRFSISLACEKIRSLTSSKTATIGSRFKQASSPQEIPIRDDGFPIKGCDDTLVVGLAVCWGGRDAYYFSLQKEQKHSEIS ASLVPPSLDPSLTLKDRMWYLQSCLRKESDKECSVVIYDFIQSYKILLLSCGISLEQSYEDPKVACWLLDPDSQEPTLHSIVTSFLPHE LPLLEGMETSQGIQSLGLNAGSEHSGRYRASVESILIFNSMNQLNSLLQKENLQDVFRKVEMPSQYCLALLELNGIGFSTAECESQK HIMQAKLDAIETQAYQLAGHSFSFTSSDDIAEVLFLELKLPPNREMKNQGSKKTLGSTRRGIDNGRKLRLGRQFSTSKDVLNKLKALH PLPGLILEWRRITNAITKVVFPLQREKCLNPFLGMERIYPVSQSHTATGRITFTEPNIQNVPRDFEIKMPTLVGESPPSQAVGKGLLPM GRGKYKKGFSVNPRCQAQMEERAADRGMPFSISMRHAFVPFPGGSILAADYSQLELRILAHLSHDRRLIQVLNTGADVFRSIAAEW KMIEPESVGDDLRQQAKQICYGIIYGMGAKSLGEQMGIKENDAACYIDSFKSRYTGINQFMTETVKNCKRDGFVQTILGRRRYLPGIK DNNPYRKAHAERQAINTIVQGSAADIVKIATVNIQKQLETFHSTFKSHGHREGMLQSDQTGLSRKRKLQGMFCPIRGGFFILQLHDEL LYEVAEEDVVQVAQIVKNEMESAVKLSVKLKVKVKIGASWGELKDFDV	
10	Nu（ν）	MENYEALVGFDLCNTPLSSVAQKIMSAMHSGDLVDSKTWGKSTETMEVINKSSVKYSVQLEDRKTQSPEKKDLKSLRSQTSRGSA KLSPQSFSVRLTDQLSADQKQKSISSLTLSSCLIPQYNQEASVLQKKGHKRKHFLMENINNENKGSINLKRKHITYNNLSEKTSKQMA LEEDTDDAEGYLNSGNSGALKKHFCDIRHLDDWAKSQLIEMLKQAAALVITVMYTDGSTQLGADQTPVSSVRGIVVLVKRQAEGGH GCPDAPACGPVLEGFVSDDPCIYIQIEHSAIWDQEQEAHQQFARNVLFQTMKCKCPVICFNAKDFVRIVLQFFGNDGSWKHVADFIG LDPRIAAWLIDPSDATPSFEDLVEKYCEKSITVKVNSTYGNSSRNIVNQNVRENLKTLYRLTMDLCSKLKDYGLWQLFRTLELPLIPIL AVMESHAIQVNKEEMEKTSALLGARLKELEQEAHFVAGERFLITSNNQLREILFGKLKLHLLSQRNSLPRTGLQKYPSTSEAVLNALR DLHPLPKIILEYRQVHKIKSTFVDGLLACMKKGSISSTWNQTGTVTGRLSAKHPNIQGISKHPIQITTPKNFKGKEDKILTISPRAMFVS SKGHTFLAADFSQIELRILTHLSGDPELLKLFQESERDDVFSTLTSQWKDVPVEQVTHADREQTKKVVYAVVYGAGKERLAACLGVP IQEAAQFLESFLQKYKKIKDFARAAIAQCHQTGCVVSIMGRRRPLPRIHAHDQQLRAQAERQAVNFVVQGSAADLCKLAMIHVFTAV AASHTLTARLVAQIHDELLFEVEDPQIPECAALVRRTMESLEQVQALELQLQVPLKVSLSAGRSWGHLVPLQEAWGPPPGPCRTES PSNSLAAPGSPASTQPPPLHFSPSFCL	900

[0168] 表1a:表1中列出的聚合酶的特性

[0169]

行	家族	Pol	宿主	基因符号	UniProt登录号	UniProt名称	催化亚基质量（kDa）	另外的亚基	生物途径	模板核酸
1	Y	Eta（η）	智人	POLH	Q9Y253	POLH_人类	78	单体	TLS	
2	Y	Iota（ι）	智人	POLI	Q9UNA4	POLI_人类	80	单体	TLS	
3	Y	Kappa（κ）	智人	POLK	Q9UBT6	POLK_人类	76	单体	TLS	
4	Y	REV1	智人	REV1	Q9UBZ9	REV1_人类	138	单体	TLS	
5	X	Beta（β）	智人	POLB	P06746	DPOLB_人类	39	单体	BER	
6	X	Lambda（λ）	智人	POLL	Q9UGP5	DPOLL_人类	66	单体	BER、NHEJ、TLS	
7	X	Mu（μ）	智人	POLM	Q9NP87	DPOLM_人类	55	单体	NHEJ、V(D)J	
8	X	TdT	智人	DNTT	P04053	TDT_人类	56	单体	V(D)J	不依赖模板
9	A	Theta（θ）	智人	POLQ	O75417	DPOLQ_人类	290	单体	TLS、MMEJ、RTDR	DNA和/或RNA
10	A	Nu（ν）	智人	POLN	Q7Z5Q5	DPOLN_人类	100	单体	ICL修复	

[0170] 对于生物途径,TLS＝跨损伤合成;BER＝碱基切除修复;NHEJ＝非同源末端连接;V(D)J＝V(D)J重组过程;MMEJ＝微同源介导的末端连接;RTDR＝RNA模板DNA修复。

[0171] 在本发明的某些方面,基因修饰系统的书写结构域具有逆转录酶活性,也称为逆转录酶结构域(RT结构域)。在一些实施例中,RT结构域包含RT催化部分和RNA结合区(例如,结合模板RNA的区域)。

[0172] 在一些实施例中,编码逆转录酶的核酸从其天然序列改变为具有改变的密码子使用,例如,针对人细胞进行改善。在一些实施例中,逆转录酶结构域是来自逆转录病毒的异源逆转录酶。在一些实施例中,包含基因修饰多肽的RT结构域已从其原始氨基酸序列突变,例如,具有至少1、2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90或100个取代。在一些实施例中,RT结构域源自逆转录病毒的RT,例如HIV-1RT、莫洛尼鼠白血病病毒(MMLV)RT、禽成髓细胞瘤病毒(AMV)RT、或劳斯肉瘤病毒(RSV)RT。

[0173] 在一些实施例中,逆转录病毒逆转录酶(RT)结构域表现出靶引发的逆转录(TPRT)起始的增强的严格性,例如,相对于内源RT结构域。在一些实施例中,当靶位点中紧邻第一链切口上游的3nt,例如引发RNA模板的基因组DNA,与RNA模板中的同源3nt具有至少66%或100%的互补性时,RT结构域启动TPRT。在一些实施例中,当模板RNA同源性和靶DNA引发逆转录之间存在少于5nt错配(例如少于1、2、3、4或5nt错配)时,RT结构域启动TPRT。在一些实施例中,修饰RT结构域使得TPRT反应引发中的错配的严格性增加,例如,其中相对于野生型(例如,未修饰的)RT结构域,RT结构域不容许任何错配或容许引发区域中更少的错配。在一些实施例中,RT结构域包含HIV-1RT结构域。在实施例中,HIV-1RT结构域启动较低水平的合成,即使相对于替代RT结构域具有三个核苷酸错配(例如,如Jamburuthugoda和Eickbush J Mol Biol[分子生物学杂志]407(5):661-672(2011)所述;其通过援引以其全文并入本文)。在一些实施例中,RT结构域形成二聚体(例如,异二聚体或同二聚体)。在一些实施例中,RT结构域是单体的。在一些实施例中,RT结构域天然地作为单体或二聚体(例如,异二聚体或同二聚体)起作用。在一些实施例中,RT结构域天然地作为单体起作用,例如,源自病毒,其中它作为单体起作用。在实施例中,RT结构域选自来自以下的RT结构域,或其功能片段或变体(例如,与其具有至少70%、80%、90%、95%或99%同一性的氨基酸序列):鼠白血病病毒(MLV;有时称为MoMLV)(例如,P03355)、猪内源逆转录病毒(PERV)(例如,UniProt Q4VFZ2)、小鼠乳腺肿瘤病毒(MMTV)(例如,UniProt P03365)、梅森-菲舍(Mason-Pfizer)猴病毒(MPMV)(例如,UniProt P07572)、牛白血病病毒(BLV)(例如,UniProt P03361)、人T细胞白血病病毒-1(HTLV-1)(例如UniProt P03362)、人泡沫病毒(HFV)(例如UniProt P14350)、猿泡沫病毒(SFV)(例如UniProt P23074)或牛泡沫/合胞病毒(BFV/BSV)(例如UniProt O41894)。在一些实施例中,RT结构域在其天然功能上是二聚体。在一些实施例中,RT结构域源自病毒,其中它作为二聚体起作用。在实施例中,RT结构域选自来自以下的RT结构域:禽肉瘤/白血病病毒(ASLV)(例如,UniProt A0A142BKH1)、劳斯肉瘤病毒(RSV)(例如,UniProt P03354)、禽成髓细胞瘤病毒(AMV)(例如,UniProt Q83133)、人免疫缺陷病毒I型(HIV-1)(例如,UniProt P03369)、人免疫缺陷病毒II型(HIV-2)(例如,UniProt P15833)、猿免疫缺陷病毒(SIV)(例如,UniProt P05896)、牛免疫缺陷病毒(BIV)(例如,UniProt P19560)、马传染性贫血病毒(EIAV)(例如,UniProt P03371)或猫免疫缺陷病毒(FIV)(例如,UniProt P16088)(Herschhorn和Hizi Cell Mol Life Sci[细胞和分子生命科学]67(16):2717-2747(2010)),或其功能片段或变体(例如,与其具有至少70%、80%、90%、95%或99%同一性的氨基酸序列)。在一些实施例中,天然异二聚体RT结构域也可以作为同二聚体起作用。在一些实施例中,二聚体RT结构域被表达为融合蛋白,例如,同二聚体融合蛋白或异二聚体融合蛋白。在一些实施例中,系统的RT功能由多个RT结构域实现(例如,如本文所述)。在另外的实施例中,多个RT结构域是融合的或分开的,例如,可以在相同的多肽上或在不同的多肽上。

[0174] 在一些实施例中,本文所述的基因修饰系统包含整合酶结构域,例如,其中整合酶结构域可以是RT结构域的一部分。在一些实施例中,RT结构域(例如,如本文所述)包含整合酶结构域。在一些实施例中,RT结构域(例如,如本文所述)缺少整合酶结构域,或包含已通过突变或缺失失活的整合酶结构域。在一些实施例中,本文所述的基因修饰系统包含RNA酶H结构域,例如,其中RNA酶H结构域可以是RT结构域的一部分。在一些实施例中,RNA酶H结构

域不是RT结构域的一部分并且通过柔性接头共价连接。在一些实施例中,RT结构域(例如,如本文所述)包含RNA酶H结构域,例如,内源RNA酶H结构域或异源RNA酶H结构域。在一些实施例中,RT结构域(例如,如本文所述)缺少RNA酶H结构域。在一些实施例中,RT结构域(例如,如本文所述)包含异源RNA酶H结构域的添加、缺失、突变或交换的RNA酶H结构域。在一些实施例中,多肽包含灭活的内源RNA酶H结构域。在一些实施例中,从多肽的其他结构域之一中遗传去除内源RNA酶H结构域,使得它不包含在多肽中,例如,内源RNA酶H结构域从包含结构域中部分或完全截短。在一些实施例中,RNA酶H结构域的突变产生表现出更低RNA酶活性的多肽,例如,如通过Kotewicz等人Nucleic Acids Res[核酸研究]16(1):265-277(1988)(其通过援引以其全文并入本文)描述的方法所确定的,例如与没有该突变的在其他方面类似的结构域相比降低至少10%、20%、30%、40%、50%、60%、70%、80%或90%。在一些实施例中,RNA酶H活性被消除。

[0175] 在一些实施例中,与没有突变的其他方面类似结构域相比,RT结构域被突变以增加保真度。例如,在一些实施例中,RT结构域中(例如,逆转录酶中)的YADD或YMDD基序被YVDD替换。在实施例中,替换YADD或YMDD或YVDD导致逆转录病毒逆转录酶活性的保真度更高(例如,如Jamburuthugoda和Eickbush J Mol Biol[分子生物学杂志]2011中所述;其通过援引以其全文并入本文)。

[0176] 在一些实施例中,本文所述的基因修饰多肽包含具有根据表2的氨基酸序列,或与其具有至少70%、80%、85%、90%、95%、97%、98%或99%同一性的序列的RT结构域。在一些实施例中,本文所述的核酸编码具有根据表2的氨基酸序列,或与其具有至少70%、80%、85%、90%、95%、97%、98%或99%同一性的序列的RT结构域。

[0177] 表2:来自逆转录病毒的示例性逆转录酶结构域

[0178]

RT 名称	RT 氨基酸序列
AVIRE_P03360	TAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPTLFDEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVREFLGTIGYCRLWIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAGWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGLLTAGGKAIKNAPEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATIS
AVIRE_P03360_3mut	TAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPTLFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVREFLGTIGYCRLWIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAGWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKNAPEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATIS
AVIRE_P03360_3mutA	TAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPTLFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVREFLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAGWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALELWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKNAPEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATIS
BAEVM_P10272	TVSLQDEHRLFDIPVTTSLPDVWLQDFPQAWAETGGLGRAKCQAPIIIDLKPTAVPVSIKQYPMSLEAHMGIRQHIIKFLELGVLRPCRSPWNTPLLPVKKPGTQDYRPVQDLREINKRTVDIHPTVPNPYNLLSTLKPDYSWYTVLDLKDAFFCLPLAPQSQELFAFEWKDPERGISGQLTWTRLPQGFKNSPTLFDEALHRDLTDFRTQHPEVTLLQYVDDLLLAAPTKKACTQGTRHLLQELGEKGYRASAKKAQICQTKVTYLGYILSEGKRWLTPGRIETVARIPPPRNPREVREFLGTAGFCRLWIPGFAELAAPLYALTKESTPFTWQTEHQLAFEALKKALLSAPALGLPDTSKPFTLFLDERQGIAKGVLTQKLGPWKRPVAYLSKKLDPVAAGWPPCLRIMAATAMLVKDSAKLTLGQPLTVITPHTLEAIVRQPPDRWITNARLTHYQALLLDTDRVQFGPPVTLNPATLLPVPENQPSPHDCRQVLAETHGTREDLKDQELPDADHTWYTDGSSYLDSGTRRAGAAVVDGHNTIWAQSLPPGTSAQKAELIALTKALELSKGKKANIYTDSRYAFATAHTHGSIYERRGLLTSEGKEIKNKAEIIALLKALFLPQEVAIIHCPGHQKGQDPVAVGNRQADRVARQAAMAEVLTLATEPDNTSHIT
BAEVM_P10272_3mut	TVSLQDEHRLFDIPVTTSLPDVWLQDFPQAWAETGGLGRAKCQAPIIIDLKPTAVPVSIKQYPMSLEAHMGIRQHIIKFLELGVLRPCRSPWNTPLLPVKKPGTQDYRPVQDLREINKRTVDIHPTVPNPYNLLSTLKPDYSWYTVLDLKDAFFCLPLAPQSQELFAFEWKDPERGISGQLTWTRLPQGFKNSPT

[0179]

RT 名称	RT 氨基酸序列
	LFNEALHRDLTDFRTQHPEVTLLQYVDDLLLAAPTKKACTQGTRHLLQELGEKGYRASAKKAQICQTKVTYLGYILSEGKRWLTPGRIETVARIPPPR NPREVREFLGTAGFCRLWIPGFAELAAPLYALTKPSTPFTWQTEHQLAFEALKKALLSAPALGLPDTSKPFTLFLDERQGIAKGVLTQKLGPWKRPV AYLSKKLDPVAAGWPPCLRIMAATAMLVKDSAKLTLGQPLTVITPHTLEAIVRQPPDRWITNARLTHYQALLLDTDRVQFGPPVTLNPATLLPVPENQ PSPHDCRQVLAETHGTREDLKDQELPDADHTWYTDGSSYLDSGTRRAGAAVVDGHNTIWAQSLPPGTSAQKAELIALTKALELSKGKKANIYTDSR YAFATAHTHGSIYERRGWLTSEGKEIKNKAEIIALLKALFLPQEVAIIHCPGHQKGQDPVAVGNRQADRVARQAAMAEVLTLATEPDNTSHIT
BAEVM_P10272_3mutA	TVSLQDEHRLFDIPVTTSLPDVWLQDFPQAWAETGGLGRAKCQAPIIIDLKPTAVPVSIKQYPMSLEAHMGIRQHIIKFLELGVLRPCRSPWNTPLLPV KKPGTQDYRPVQDLREINKRTVDIHPTVPNPYNLLSTLKPDYSWYTVLDLKDAFFCLPLAPQSQELFAFEWKDPERGISGQLTWTRLPQGFKNSPT LFNEALHRDLTDFRTQHPEVTLLQYVDDLLLAAPTKKACTQGTRHLLQELGEKGYRASAKKAQICQTKVTYLGYILSEGKRWLTPGRIETVARIPPPR NPREVREFLGKAGFCRLFIPGFAELAAPLYALTKPSTPFTWQTEHQLAFEALKKALLSAPALGLPDTSKPFTLFLDERQGIAKGVLTQKLGPWKRPV AYLSKKLDPVAAGWPPCLRIMAATAMLVKDSAKLTLGQPLTVITPHTLEAIVRQPPDRWITNARLTHYQALLLDTDRVQFGPPVTLNPATLLPVPENQ PSPHDCRQVLAETHGTREDLKDQELPDADHTWYTDGSSYLDSGTRRAGAAVVDGHNTIWAQSLPPGTSAQKAELIALTKALELSKGKKANIYTDSR YAFATAHTHGSIYERRGWLTSEGKEIKNKAEIIALLKALFLPQEVAIIHCPGHQKGQDPVAVGNRQADRVARQAAMAEVLTLATEPDNTSHIT
BLVAU_P25059	GVLDAPPSHIGLEHLPPPPEVPQFPLNLERLQALQDLVHRSLEAGYISPWDGPGNNPVFPVRKPNGAWRFVHDLRVTNALTKPIPALSPGPPDLTAI PTHLPHIICLDLKDAFFQIPVEDRFRSYFAFTLPTPGGLQPHRRFAWRVLPQGFINSPALFERALQEPLRQVSAAFSQSLLVSYMDDILYVSPTEEQR LQCYQTMAAHLRDLGFQVASEKTRQTPSPVPFLGQMVHERMVTYQSLPTLQISSPISLHQLQTVLGDLQWVSRGTPTTRRPLQLLYSSLKGIDDPR AIIHLSPEQQQGIAELRQALSHNARSRYNEQEPLLAYVHLTRAGSTLVLFQKGAQFPLAYFQTPLTDNQASPWGLLLLLGCQYLQAQALSSYAKTILK YYHNLPKTSLDNWIQSSEDPRVQELLQLWPQISSQGIQPPGPWKTLVTRAEVFLTPQFSPEPIPAALCLFSDGAARRGAYCLWKDHLLDFQAVPAP ESAQKGELAGLLAGLAAAPPEPLNIWVDSKYLYSLLRTLVLGAWLQPDPVPSYALLYKSLLRHPAIFVGHVRSHSSASHPIASLNNYVDQL
BLVAU_P25059_2mut	GVLDAPPSHIGLEHLPPPPEVPQFPLNLERLQALQDLVHRSLEAGYISPWDGPGNNPVFPVRKPNGAWRFVHDLRVTNALTKPIPALSPGPPDLTAI PTHLPHIICLDLKDAFFQIPVEDRFRSYFAFTLPTPGGLQPHRRFAWRVLPQGFINSPALFQRALQEPLRQVSAAFSQSLLVSYMDDILYVSPTEEQR LQCYQTMAAHLRDLGFQVASEKTRQTPSPVPFLGQMVHERMVTYQSLPTLQISSPISLHQLQTVLGDLQWVSRGTPTTRRPLQLLYSSLKPIDDPR AIIHLSPEQQQGIAELRQALSHNARSRYNEQEPLLAYVHLTRAGSTLVLFQKGAQFPLAYFQTPLTDNQASPWGLLLLLGCQYLQAQALSSYAKTILK YYHNLPKTSLDNWIQSSEDPRVQELLQLWPQISSQGIQPPGPWKTLVTRAEVFLTPQFSPEPIPAALCLFSDGAARRGAYCLWKDHLLDFQAVPAP ESAQKGELAGLLAGLAAAPPEPLNIWVDSKYLYSLLRTLVLGAWLQPDPVPSYALLYKSLLRHPAIFVGHVRSHSSASHPIASLNNYVDQL
BLVJ_P03361	GVLDTPPSHIGLEHLPPPPEVPQFPLNLERLQALQDLVHRSLEAGYISPWDGPGNNPVFPVRKPNGAWRFVHDLRATNALTKPIPALSPGPPDLTAI PTHPPHIICLDLKDAFFQIPVEDRFRFYLSFTLPSPGGLQPHRRFAWRVLPQGFINSPALFQRALQEPLRQVSAAFSQSLLVSYMDDILYASPTEEQR SQCYQALAARLRDLGFQVASEKTSQTPSPVPFLGQMVHEQIVTYQSLPTLQISSPISLHQLQAVLGDLQWVSRGTPTTRRPLQLLYSSLKRHHDPR AIIQLSPEQLQGIAELRQALSHNARSRYNEQEPLLAYVHLTRAGSTLVLFQKGAQFPLAYFQTPLTDNQASPWGLLLLLGCQYLQTQALSSYAKPILK YYHNLPKTSLDNWIQSSEDPRVQELLQLWPQISSQGIQPPGPWKTLITRAEVFLTPQFSPDPIPAALCLFSDGATGRGAYCLWKDHLLDFQAVPAPE SAQKGELAGLLAGLAAAPPEPVNIWVDSKYLYSLLRTLVLGAWLQPDPVPSYALLYKSLLRHPAIVVGHVRSHSSASHPIASLNNYVDQL
BLVJ_P03361_2mut	GVLDTPPSHIGLEHLPPPPEVPQFPLNLERLQALQDLVHRSLEAGYISPWDGPGNNPVFPVRKPNGAWRFVHDLRATNALTKPIPALSPGPPDLTAI PTHPPHIICLDLKDAFFQIPVEDRFRFYLSFTLPSPGGLQPHRRFAWRVLPQGFINSPALFNRALQEPLRQVSAAFSQSLLVSYMDDILYASPTEEQR SQCYQALAARLRDLGFQVASEKTSQTPSPVPFLGQMVHEQIVTYQSLPTLQISSPISLHQLQAVLGDLQWVSRGTPTTRRPLQLLYSSLKRHHDPR AIIQLSPEQLQGIAELRQALSHNARSRYNEQEPLLAYVHLTRAGSTLVLFQKGAQFPLAYFQTPLTDNQASPWGLLLLLGCQYLQTQALSSYAKPILK YYHNLPKTSLDNWIQSSEDPRVQELLQLWPQISSQGIQPPGPWKTLITRAEVFLTPQFSPDPIPAALCLFSDGATGRGAYCLWKDHLLDFQAVPAPE SAQKGELAGLLAGLAAAPPEPVNIWVDSKYLYSLLRTWVLGAWLQPDPVPSYALLYKSLLRHPAIVVGHVRSHSSASHPIASLNNYVDQL
BLVJ_P03361_2mutB	GVLDTPPSHIGLEHLPPPPEVPQFPLNLERLQALQDLVHRSLEAGYISPWDGPGNNPVFPVRKPNGAWRFVHDLRATNALTKPIPALSPGPPDLTA PPTHPPHIICLDLKDAFFQIPVEDRFRFYLSFTLPSPGGLQPHRRFAWRVLPQGFINSPALFQRALQEPLRQVSAAFSQSLLVSYMDDILYASPTEEQ RSQCYQALAARLRDLGFQVASEKTSQTPSPVPFLGQMVHEQIVTYQSLPTLQISSPISLHQLQAVLGDLQWVSRGTPTTRRPLQLLYSSLKRHHDP RAIIQLSPEQLQGIAELRQALSHNARSRYNEQEPLLAYVHLTRAGSTLVLFQKGAQFPLAYFQTPLTDNQASPWGLLLLLGCQYLQTQALSSYAKPIL KYYHNLPKTSLDNWIQSSEDPRVQELLQLWPQISSQGIQPPGPWKTLITRAEVFLTPQFSPDPIPAALCLFSDGATGRGAYCLWKDHLLDFQAVPAP ESAQKGELAGLLAGLAAAPPEPVNIWVDSKYLYSLLRTWVLGAWLQPDPVPSYALLYKSLLRHPAIVVGHVRSHSSASHPIASLNNYVDQL
FFV_O93209	MDLLKPLTVERKGVKIKGYWNSQADITCVPKDLLQGEEPVRQQNVTTIHGTQEGDVYYVNLKIDGRRINTEVIGTTLDYAIITPGDVPWILKKPLELTIK LDLEEQQGTLLNNSILSKKGKEELKQLFEKYSALWQSWENQVGHRRIRPHKIATGTVKPTPQKQYHINPKAKPDIQIVINDLLKQGVLIQKESTMNTP VYPVPKPNGRWRMVLDYRAVNKVTPLIAVQNQHSYGILGSLFKGRYKTTIDLSNGFWAHPIVPEDYWITAFTWQGKQYCWTVLPQGFLNSPGLFT GDVVDLLQGIPNVEVYVDDVYISHDSEKEHLEYLDILFNRLKEAGYIISLKKSNIANSIVDFLGFQITNEGRGLTDTFKEKLENITAPTTLKQLQSILGLLN FARNFIPDFTELIAPLYALIPKSTKNYVPWQIEHSTTLETLITKLNGAEYLQGRKGDKTLIMKVNASYTTGYIRYYNEGEKKPISYVSIVFSKTELKFTELE KLLTTVHKGLLKALDLSMGQNIHVYSPIVSMQNIQKTPQTAKKALASRWLSWLSYLEDPRIRFFYDPQMPALKDLPAVDTGKDNKKHPSNFQHIFYT DGSAITSPTKEGHLNAGMGIVYFINKDGNLQKQQEWSISLGNHTAQFAEIAAFEFALKKCLPLGGNILVVTDSNYVAKAYNEELDVWASNGFVNNRK KPLKHISKWKSVADLKRLRPDVVVTHEPGHQKLDSSPHAYGNNLADQLATQASFKVH
FFV_O93209_2mut	MDLLKPLTVERKGVKIKGYWNSQADITCVPKDLLQGEEPVRQQNVTTIHGTQEGDVYYVNLKIDGRRINTEVIGTTLDYAIITPGDVPWILKKPLELTIK LDLEEQQGTLLNNSILSKKGKEELKQLFEKYSALWQSWENQVGHRRIRPHKIATGTVKPTPQKQYHINPKAKPDIQIVINDLLKQGVLIQKESTMNTP VYPVPKPNGRWRMVLDYRAVNKVTPLIAVQNQHSYGILGSLFKGRYKTTIDLSNGFWAHPIVPEDYWITAFTWQGKQYCWTVLPQGFLNSPGLFN GDVVDLLQGIPNVEVYVDDVYISHDSEKEHLEYLDILFNRLKEAGYIISLKKSNIANSIVDFLGFQITNEGRGLTDTFKEKLENITAPTTLKQLQSILGLLN FARNFIPDFTELIAPLYALIPKSPKNYVPWQIEHSTTLETLITKLNGAEYLQGRKGDKTLIMKVNASYTTGYIRYYNEGEKKPISYVSIVFSKTELKFTEL EKLLTTVHKGLLKALDLSMGQNIHVYSPIVSMQNIQKTPQTAKKALASRWLSWLSYLEDPRIRFFYDPQMPALKDLPAVDTGKDNKKHPSNFQHIFY TDGSAITSPTKEGHLNAGMGIVYFINKDGNLQKQQEWSISLGNHTAQFAEIAAFEFALKKCLPLGGNILVVTDSNYVAKAYNEELDVWASNGFVNNR KKPLKHISKWKSVADLKRLRPDVVVTHEPGHQKLDSSPHAYGNNLADQLATQASFKVH
FFV_O93209_2mutA	MDLLKPLTVERKGVKIKGYWNSQADITCVPKDLLQGEEPVRQQNVTTIHGTQEGDVYYVNLKIDGRRINTEVIGTTLDYAIITPGDVPWILKKPLELTIK LDLEEQQGTLLNNSILSKKGKEELKQLFEKYSALWQSWENQVGHRRIRPHKIATGTVKPTPQKQYHINPKAKPDIQIVINDLLKQGVLIQKESTMNTP VYPVPKPNGRWRMVLDYRAVNKVTPLIAVQNQHSYGILGSLFKGRYKTTIDLSNGFWAHPIVPEDYWITAFTWQGKQYCWTVLPQGFLNSPGLFN GDVVDLLQGIPNVEVYVDDVYISHDSEKEHLEYLDILFNRLKEAGYIISLKKSNIANSIVDFLGFQITNEGRGLTDTFKEKLENITAPTTLKQLQSILGKLN FARNFIPDFTELIAPLYALIPKSPKNYVPWQIEHSTTLETLITKLNGAEYLQGRKGDKTLIMKVNASYTTGYIRYYNEGEKKPISYVSIVFSKTELKFTEL EKLLTTVHKGLLKALDLSMGQNIHVYSPIVSMQNIQKTPQTAKKALASRWLSWLSYLEDPRIRFFYDPQMPALKDLPAVDTGKDNKKHPSNFQHIFY TDGSAITSPTKEGHLNAGMGIVYFINKDGNLQKQQEWSISLGNHTAQFAEIAAFEFALKKCLPLGGNILVVTDSNYVAKAYNEELDVWASNGFVNNR KKPLKHISKWKSVADLKRLRPDVVVTHEPGHQKLDSSPHAYGNNLADQLATQASFKVH
FFV_O93209-Pro	VPWILKKPLELTIKLDLEEQQGTLLNNSILSKKGKEELKQLFEKYSALWQSWENQVGHRRIRPHKIATGTVKPTPQKQYHINPKAKPDIQIVINDLLKQ GVLIQKESTMNTPVYPVPKPNGRWRMVLDYRAVNKVTPLIAVQNQHSYGILGSLFKGRYKTTIDLSNGFWAHPIVPEDYWITAFTWQGKQYCWTVL PQGFLNSPGLFTGDVVDLLQGIPNVEVYVDDVYISHDSEKEHLEYLDILFNRLKEAGYIISLKKSNIANSIVDFLGFQITNEGRGLTDTFKEKLENITAPT TLKQLQSILGLLNFARNFIPDFTELIAPLYALIPKSTKNYVPWQIEHSTTLETLITKLNGAEYLQGRKGDKTLIMKVNASYTTGYIRYYNEGEKKPISYVSI VFSKTELKFTELEKLLTTVHKGLLKALDLSMGQNIHVYSPIVSMQNIQKTPQTAKKALASRWLSWLSYLEDPRIRFFYDPQMPALKDLPAVDTGKDNK KHPSNFQHIFYTDGSAITSPTKEGHLNAGMGIVYFINKDGNLQKQQEWSISLGNHTAQFAEIAAFEFALKKCLPLGGNILVVTDSNYVAKAYNEELDV WASNGFVNNRKKPLKHISKWKSVADLKRLRPDVVVTHEPGHQKLDSSPHAYGNNLADQLATQASFKVH
FFV_O93209-Pro_2mut	VPWILKKPLELTIKLDLEEQQGTLLNNSILSKKGKEELKQLFEKYSALWQSWENQVGHRRIRPHKIATGTVKPTPQKQYHINPKAKPDIQIVINDLLKQ GVLIQKESTMNTPVYPVPKPNGRWRMVLDYRAVNKVTPLIAVQNQHSYGILGSLFKGRYKTTIDLSNGFWAHPIVPEDYWITAFTWQGKQYCWTVL PQGFLNSPGLFNGDVVDLLQGIPNVEVYVDDVYISHDSEKEHLEYLDILFNRLKEAGYIISLKKSNIANSIVDFLGFQITNEGRGLTDTFKEKLENITAPT TLKQLQSILGLLNFARNFIPDFTELIAPLYALIPKSPKNYVPWQIEHSTTLETLITKLNGAEYLQGRKGDKTLIMKVNASYTTGYIRYYNEGEKKPISYVSI VFSKTELKFTELEKLLTTVHKGLLKALDLSMGQNIHVYSPIVSMQNIQKTPQTAKKALASRWLSWLSYLEDPRIRFFYDPQMPALKDLPAVDTGKDNK KHPSNFQHIFYTDGSAITSPTKEGHLNAGMGIVYFINKDGNLQKQQEWSISLGNHTAQFAEIAAFEFALKKCLPLGGNILVVTDSNYVAKAYNEELDV WASNGFVNNRKKPLKHISKWKSVADLKRLRPDVVVTHEPGHQKLDSSPHAYGNNLADQLATQASFKVH

[0180]

RT 名称	RT 氨基酸序列
FFV_O93209-Pro_2mutA	VPWILKKPLELTIKLDLEEQQGTLLNNSILSKKGKEELKQLFEKYSALWQSWENQVGHRRIRPHKIATGTVKPTPQKQYHINPKAKPDIQIVINDLLKQGVLIQKESTMNTPVYPVPKPNGRWRMVLDYRAVNKVTPLIAVQNQHSYGILGSLFKGRYKTTIDLSNGFWAHPIVPEDYWITAFTWQGKQYCWTVLPQGFLNSPGLFNGDVVDLLQGIPNVEVYVDDVYISHDSEKEHLEYLDILFNRLKEAGYIISLKKSNIANSIVDFLGFQITNEGRGLTDTFKEKLENITAPTTLKQLQSILGKLNFARNFIPDFTELIAPLYALIPKSPKNYVPWQIEHSTTLETLITKLNGAEYLQGRKGDKTLIMKVNASYTTGYIRYYNEGEKKPISYVSIVFSKTELKFTELEKLLTTVHKGLLKALDLSMGQNIHVYSPIVSMQNIQKTPQTAKKALASRWLSWLSYLEDPRIRFFYDPQMPALKDLPAVDTGKDNKKHPSNFQHIFYTDGSAITSPTKEGHLNAGMGIVYFINKDGNLQKQQEWSISLGNHTAQFAEIAAFEFALKKCLPLGGNILVVTDSNYVAKAYNEELDVWASNGFVNNRKKPLKHISKWKSVADLKRLRPDVVVTHEPGHQKLDSSPHAYGNNLADQLATQASFKVH
FLV_P10273	TLQLEEEYRLFEPESTQKQEMDIWLKNFPQAWAETGGMGTAHCQAPVLIQLKATATPISIRQYPMPHEAYQGIKPHIRRMLDQGILKPCQSPWNTPLLPVKKPGTEDYRPVQDLREVNKRVEDIHPTVPNPYNLLSTLPPSHPWYTVLDLKDAFFCLRLHSESQLLFAFEWRDPEIGLSGQLTWTRLPQGFKNSPTLFDEALHSDLADFRVRYPALVLLQYVDDLLLAAATRTECLEGTKALLETLGNKGYRASAKKAQICLQEVTYLGYSLKDGQRWLTKARKEAILSIPVPKNSRQVREFLGTAGYCRLWIPGFAELAAPLYPLTRPGTLFQWGTEQQLAFEDIKKALLSSPALGLPDITKPFELFIDENSGFAKGVLVQKLGPWKRPVAYLSKKLDTVASGWPPCLRMVAAIAILVKDAGKLTLGQPLTILTSHPVEALVRQPPNKWLSNARMTHYQAMLLDAERVHFGPTVSLNPATLLPLPSGGNHHDCLQILAETHGTRPDLTDQPLPDADLTWYTDGSSFIRNGEREAGAAVTTESEVIWAAPLPPGTSAQRAELIALTQALKMAEGKKLTVYTDSRYAFATTHVHGEIYRRRGLLTSEGKEIKNKNEILALLEALFLPKRLSIIHCPGHQKGDSPQAKGNRLADDTAKKAATETHSSLTVLP
FLV_P10273_3mut	TLQLEEEYRLFEPESTQKQEMDIWLKNFPQAWAETGGMGTAHCQAPVLIQLKATATPISIRQYPMPHEAYQGIKPHIRRMLDQGILKPCQSPWNTPLLPVKKPGTEDYRPVQDLREVNKRVEDIHPTVPNPYNLLSTLPPSHPWYTVLDLKDAFFCLRLHSESQLLFAFEWRDPEIGLSGQLTWTRLPQGFKNSPTLFNEALHSDLADFRVRYPALVLLQYVDDLLLAAATRTECLEGTKALLETLGNKGYRASAKKAQICLQEVTYLGYSLKDGQRWLTKARKEAILSIPVPKNSRQVREFLGTAGYCRLWIPGFAELAAPLYPLTRPGTLFQWGTEQQLAFEDIKKALLSSPALGLPDITKPFELFIDENSGFAKGVLVQKLGPWKRPVAYLSKKLDTVASGWPPCLRMVAAIAILVKDAGKLTLGQPLTILTSHPVEALVRQPPNKWLSNARMTHYQAMLLDAERVHFGPTVSLNPATLLPLPSGGNHHDCLQILAETHGTRPDLTDQPLPDADLTWYTDGSSFIRNGEREAGAAVTTESEVIWAAPLPPGTSAQRAELIALTQALKMAEGKKLTVYTDSRYAFATTHVHGEIYKNKNEILALLEALFLPKRLSIIHCPGHQKGDSPQAKGNRLADDTAKKAATETHSSLTVLP
FLV_P10273_3mutA	TLQLEEEYRLFEPESTQKQEMDIWLKNFPQAWAETGGMGTAHCQAPVLIQLKATATPISIRQYPMPHEAYQGIKPHIRRMLDQGILKPCQSPWNTPLLPVKKPGTEDYRPVQDLREVNKRVEDIHPTVPNPYNLLSTLPPSHPWYTVLDLKDAFFCLRLHSESQLLFAFEWRDPEIGLSGQLTWTRLPQGFKNSPTLFNEALHSDLADFRVRYPALVLLQYVDDLLLAAATRTECLEGTKALLETLGNKGYRASAKKAQICLQEVTYLGYSLKDGQRWLTKARKEAILSIPVPKNSRQVREFLGKAGYCRLFIPGFAELAAPLYPLTRPGTLFQWGTEQQLAFEDIKKALLSSPALGLPDITKPFELFIDENSGFAKGVLVQKLGPWKRPVAYLSKKLDTVASGWPPCLRMVAAIAILVKDAGKLTLGQPLTILTSHPVEALVRQPPNKWLSNARMTHYQAMLLDAERVHFGPTVSLNPATLLPLPSGGNHHDCLQILAETHGTRPDLTDQPLPDADLTWYTDGSSFIRNGEREAGAAVTTESEVIWAAPLPPGTSAQRAELIALTQALKMAEGKKLTVYTDSRYAFATTHVHGEIYRRRGWLTSEGKEIKNKNEILALLEALFLPKRLSIIHCPGHQKGDSPQAKGNRLADDTAKKAATETHSSLTVLP
FOAMV_P14350	MNPLQLLQPLPAEIKGTKLLAHWNSGATITCIPESFLEDEQPIKKTLIKTIHGEKQQNVYYVTFKVKGRKVEAEVIASPYEYILLSPTDVPWLTQQPLQLTILVPLQEYQEKILSKTALPEDQKQQLKTLFVKYDNLWQHWENQVGHRKIRPHNIATGDYPPRPQKQYPINPKAKPSIQIVIDDLLKQGVLTPQNSTMNTPVYPVPKPDGRWRMVLDYREVNKTIPLTAAQNQHSAGILATIVRQKYKTTLDLANGFWAHPITPESYWLTAFTWQGKQYCWTRLPQGFLNSPALFTADVVDLLKEIPNVQVYVDDIYLSHDDPKEHVQQLEKVFQILLQAGYVVSLKKSEIGQKTVEFLGFNITKEGRGLTDTFKTKLLNITPPKDLKQLQSILGLLNFARNFIPNFAELVQPLYNLIAPAKGKYIEWSEENTKQLNMVIEALNTASNLEERLPEQRLVIKVNTSPSAGYVRYYNETGKKPIMYLNYVFSKAELKFSMLEKLLTTMHKALIKAMDLAMGQEILVYSPIVSMTKIQKTPLPERKALPIRWITWMTYLEDPRIQFHYDKTLPELKHIPDVYTSSQSPVKHPSQYEGVFYTDGSAIKSPDPTKSNNAGMGIVHATYKPEYQVLNQWSIPLGNHTAQMAEIAAVEFACKKALKIPGPVLVITDSFYVAESANKELPYWKSNGFVNNKKKPLKHISKWKSIAECLSMKPDITIQHEKGISLQIPVFILKGNALADKLATQGSYVVN	
FOAMV_P14350_2mut	MNPLQLLQPLPAEIKGTKLLAHWNSGATITCIPESFLEDEQPIKKTLIKTIHGEKQQNVYYVTFKVKGRKVEAEVIASPYEYILLSPTDVPWLTQQPLQLTILVPLQEYQEKILSKTALPEDQKQQLKTLFVKYDNLWQHWENQVGHRKIRPHNIATGDYPPRPQKQYPINPKAKPSIQIVIDDLLKQGVLTPQNSTMNTPVYPVPKPDGRWRMVLDYREVNKTIPLTAAQNQHSAGILATIVRQKYKTTLDLANGFWAHPITPESYWLTAFTWQGKQYCWTRLPQGFLNSPAGFLNADVVDLLKEIPNVQVYVDDIYLSHDDPKEHVQQLEKVFQILLQAGYVVSLKKSEIGQKTVEFLGFNITKEGRGLTDTFKTKLLNITPPKDLKQLQSILGLLNFARNFIPNFAELVQPLYNLIAPAKGKYIEWSEENTKQLNMVIEALNTASNLEERLPEQRLVIKVNTSPSAGYVRYYNETGKKPIMYLNYVFSKAELKFSMLEKLLTTMHKALIKAMDLAMGQEILVYSPIVSMTKIQKTPLPERKALPIRWITWMTYLEDPRIQFHYDKTLPELKHIPDVYTSSQSPVKHPSQYEGVFYTDGSAIKSPDPTKSNNAGMGIVHATYKPEYQVLNQWSIPLGNHTAQMAEIAAVEFACKKALKIPGPVLVITDSFYVAESANKELPYWKSNGFVNNKKKPLKHISKWKSIAECLSMKPDITIQHEKGISLQIPVFILKGNALADKLATQGSYVVN	
FOAMV_P14350_2mutA	MNPLQLLQPLPAEIKGTKLLAHWNSGATITCIPESFLEDEQPIKKTLIKTIHGEKQQNVYYVTFKVKGRKVEAEVIASPYEYILLSPTDVPWLTQQPLQLTILVPLQEYQEKILSKTALPEDQKQQLKTLFVKYDNLWQHWENQVGHRKIRPHNIATGDYPPRPQKQYPINPKAKPSIQIVIDDLLKQGVLTPQNSTMNTPVYPVPKPDGRWRMVLDYREVNKTIPLTAAQNQHSAGILATIVRQKYKTTLDLANGFWAHPITPESYWLTAFTWQGKQYCWTRLPQGFLNSPAGFLNADVVDLLKEIPNVQVYVDDIYLSHDDPKEHVQQLEKVFQILLQAGYVVSLKKSEIGQKTVEFLGFNITKEGRGLTDTFKTKLLNITPPKDLKQLQSILGKLNFARNFIPNFAELVQPLYNLIAPAKGKYIEWSEENTKQLNMVIEALNTASNLEERLPEQRLVIKVNTSPSAGYVRYYNETGKKPIMYLNYVFSKAELKFSMLEKLLTTMHKALIKAMDLAMGQEILVYSPIVSMTKIQKTPLPERKALPIRWITWMTYLEDPRIQFHYDKTLPELKHIPDVYTSSQSPVKHPSQYEGVFYTDGSAIKSPDPTKSNNAGMGIVHATYKPEYQVLNQWSIPLGNHTAQMAEIAAVEFACKKALKIPGPVLVITDSFYVAESANKELPYWKSNGFVNNKKKPLKHISKWKSIAECLSMKPDITIQHEKGISLQIPVFILKGNALADKLATQGSYVVN	
FOAMV_P14350-Pro	VPWLTQQPLQLTILVPLQEYQEKILSKTALPEDQKQQLKTLFVKYDNLWQHWENQVGHRKIRPHNIATGDYPPRPQKQYPINPKAKPSIQIVIDDLLKQGVLTPQNSTMNTPVYPVPKPDGRWRMVLDYREVNKTIPLTAAQNQHSAGILATIVRQKYKTTLDLANGFWAHPITPESYWLTAFTWQGKQYCWTRLPQGFLNSPALFTADVVDLLKEIPNVQVYVDDIYLSHDDPKEHVQQLEKVFQILLQAGYVVSLKKSEIGQKTVEFLGFNITKEGRGLTDTFKTKLLNITPPKDLKQLQSILGLLNFARNFIPNFAELVQPLYNLIAPAKGKYIEWSEENTKQLNMVIEALNTASNLEERLPEQRLVIKVNTSPSAGYVRYYNETGKKPIMYLNYVFSKAELKFSMLEKLLTTMHKALIKAMDLAMGQEILVYSPIVSMTKIQKTPLPERKALPIRWITWMTYLEDPRIQFHYDKTLPELKHIPDVYTSSQSPVKHPSQYEGVFYTDGSAIKSPDPTKSNNAGMGIVHATYKPEYQVLNQWSIPLGNHTAQMAEIAAVEFACKKALKIPGPVLVITDSFYVAESANKELPYWKSNGFVNNKKKPLKHISKWKSIAECLSMKPDITIQHEKGISLQIPVFILKGNALADKLATQGSYVVN	
FOAMV_P14350-Pro_2mut	VPWLTQQPLQLTILVPLQEYQEKILSKTALPEDQKQQLKTLFVKYDNLWQHWENQVGHRKIRPHNIATGDYPPRPQKQYPINPKAKPSIQIVIDDLLKQGVLTPQNSTMNTPVYPVPKPDGRWRMVLDYREVNKTIPLTAAQNQHSAGILATIVRQKYKTTLDLANGFWAHPITPESYWLTAFTWQGKQYCWTRLPQGFLNSPALFNADVVDLLKEIPNVQVYVDDIYLSHDDPKEHVQQLEKVFQILLQAGYVVSLKKSEIGQKTVEFLGFNITKEGRGLTDTFKTKLLNITPPKDLKQLQSILGLLNFARNFIPNFAELVQPLYNLIAPAKGKYIEWSEENTKQLNMVIEALNTASNLEERLPEQRLVIKVNTSPSAGYVRYYNETGKKPIMYLNYVFSKAELKFSMLEKLLTTMHKALIKAMDLAMGQEILVYSPIVSMTKIQKTPLPERKALPIRWITWMTYLEDPRIQFHYDKTLPELKHIPDVYTSSQSPVKHPSQYEGVFYTDGSAIKSPDPTKSNNAGMGIVHATYKPEYQVLNQWSIPLGNHTAQMAEIAAVEFACKKALKIPGPVLVITDSFYVAESANKELPYWKSNGFVNNKKKPLKHISKWKSIAECLSMKPDITIQHEKGISLQIPVFILKGNALADKLATQGSYVVN	
FOAMV_P14350-Pro_2mutA	VPWLTQQPLQLTILVPLQEYQEKILSKTALPEDQKQQLKTLFVKYDNLWQHWENQVGHRKIRPHNIATGDYPPRPQKQYPINPKAKPSIQIVIDDLLKQGVLTPQNSTMNTPVYPVPKPDGRWRMVLDYREVNKTIPLTAAQNQHSAGILATIVRQKYKTTLDLANGFWAHPITPESYWLTAFTWQGKQYCWTRLPQGFLNSPALFNADVVDLLKEIPNVQVYVDDIYLSHDDPKEHVQQLEKVFQILLQAGYVVSLKKSEIGQKTVEFLGFNITKEGRGLTDTFKTKLLNITPPKDLKQLQSILGKLNFARNFIPNFAELVQPLYNLIAPAKGKYIEWSEENTKQLNMVIEALNTASNLEERLPEQRLVIKVNTSPSAGYVRYYNETGKKPIMYLNYVFSKAELKFSMLEKLLTTMHKALIKAMDLAMGQEILVYSPIVSMTKIQKTPLPERKALPIRWITWMTYLEDPRIQFHYDKTLPELKHIPDVYTSSQSPVKHPSQYEGVFYTDGSAIKSPDPTKSNNAGMGIVHATYKPEYQVLNQWSIPLGNHTAQMAEIAAVEFACKKALKIPGPVLVITDSFYVAESANKELPYWKSNGFVNNKKKPLKHISKWKSIAECLSMKPDITIQHEKGISLQIPVFILKGNALADKLATQGSYVVN	
GALV_P21414	VLNLEEEYRLHEKPVPSSIDPSWLQLFPTVWAERAGMGLANQVPPVVVELRSGASPVAVRQYPMSKEAREGIRPHIQKFLDLGVLVPCRSPWNTPLLPVKKPGTNDYRPVQDLREINKRVQDIHPTVPNPYNLLSSLPPSYTWYSVLDLKDAFFCLRLHPNSQPLFAFEWKDPEKGNTGQLTWTRLPQGFKNSPTLFDEALHRDLAPFRALNPQVVLLQYVDDLLVAAPTYEDCKKGTQKLLQELSKLGYRVSAKKAQLCQREVTYLGYLLKEGKRWLTPARKATVMKIPVPTTPRQVREFLGTAGFCRLWIPGFASLAAPLYPLTKESIPFIWTEEHQQAFDHIKKALLSAPALALPDLTKPFTLYIDERAGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPTCLKAVAAVALLLKDADKLTLGQNVTVIASHSLESIVRQPPDRWMTNARMTHYQSLLLNERVSFAPPAVLNPATLLPVESEATPVHRCSEILAEETGTRRDLEDQPLPGVPTWYTDGSSFITEGKRRAGAPIVDGKRTVWASSLPEGTSAQKAELVALTQALRLAEGKNINIYTDSRYAFATAHIHGAIYKQRGLLTSAGKDIKNKEEILALLEAIHLPRRVAIIHCPGHQRGSNPVATGNRRADEAAKQAALSTRVLAGTTKP

[0181]

RT 名称	RT 氨基酸序列
GALV_P21414_3mut	VLNLEEEYRLHEKPVPSSIDPSWLQLFPTVWAERAGMGLANQVPPVVVELRSGASPVAVRQYPMSKEAREGIRPHIQKFLDLGVLVPCRSPWNTP LLPVKKPGTNDYRPVQDLREINKRVQDIHPTVPNPYNLLSSLPPSYTWYSVLDLKDAFFCLRLHPNSQPLFAFEWKDPEKGNTGQLTWTRLPQGFK NSPTLFNEALHRDLAPFRALNPQVVLLQYVDDLLVAAPTYEDCKKGTQKLLQELSKLGYRVSAKKAQLCQREVTYLGYLLKEGKRWLTPARKATVM KIPVPTTPRQVREFLGTAGFCRLWIPGFASLAAPLYPLTKPSIPFIWTEEHQQAFDHIKKALLSAPALALPDLTKPFTLYIDERAGVARGVLTQTLGPW RRPVAYLSKKLDPVASGWPTCLKAVAAVALLLKDADKLTLGQNVTVIASHSLESIVRQPPDRWMTNARMTHYQSLLLNERVSFAPPAVLNPATLLPV ESEATPVHRCSEILAEETGTRRDLEDQPLPGVPTWYTDGSSFITEGKRRAGAPIVDGKRTVWASSLPEGTSAQKAELVALTQALRLAEGKNINIYTD SRYAFATAHIHGAIYKQRGWLTSAGKDIKNKEEILALLEAIHLPRRVAIIHCPGHQRGSNPVATGNRRADEAAKQAALSTRVLAGTTKP
GALV_P21414_3mutA	VLNLEEEYRLHEKPVPSSIDPSWLQLFPTVWAERAGMGLANQVPPVVVELRSGASPVAVRQYPMSKEAREGIRPHIQKFLDLGVLVPCRSPWNTP LLPVKKPGTNDYRPVQDLREINKRVQDIHPTVPNPYNLLSSLPPSYTWYSVLDLKDAFFCLRLHPNSQPLFAFEWKDPEKGNTGQLTWTRLPQGFK NSPTLFNEALHRDLAPFRALNPQVVLLQYVDDLLVAAPTYEDCKKGTQKLLQELSKLGYRVSAKKAQLCQREVTYLGYLLKEGKRWLTPARKATVM KIPVPTTPRQVREFLGKAGFCRLFIPGFASLAAPLYPLTKPSIPFIWTEEHQQAFDHIKKALLSAPALALPDLTKPFTLYIDERAGVARGVLTQTLGPW RRPVAYLSKKLDPVASGWPTCLKAVAAVALLLKDADKLTLGQNVTVIASHSLESIVRQPPDRWMTNARMTHYQSLLLNERVSFAPPAVLNPATLLPV ESEATPVHRCSEILAEETGTRRDLEDQPLPGVPTWYTDGSSFITEGKRRAGAPIVDGKRTVWASSLPEGTSAQKAELVALTQALRLAEGKNINIYTD SRYAFATAHIHGAIYKQRGWLTSAGKDIKNKEEILALLEAIHLPRRVAIIHCPGHQRGSNPVATGNRRADEAAKQAALSTRVLAGTTKP
HTL1A_P03362	AVLGLEHLPRPPQISQFPLNPERLQALQHLVRKALEAGHIEPYTGPGNNPVFPVKKANGTWRFIHDLRATNSLTIDLSSSSPGPPDLSSLPTTLAHLQ TIDLRDAFFQIPLPKQFQPYFAFTVPQQCNYGPGTRYAWKVLPQGFKNSPTLFEMQLAHILQPIRQAFPQCTILQYMDDILLASPSHEDLLLLSEATM ASLISHGLPVSENKTQQTPGTIKFLGQIISPNHLTYDAVPTVPIRSRWALPELQALLGEIQWVSKGTPTLRQPLHSLYCALQRHTDPRDQIYLNPSQV QSLVQLRQALSQNCRSRLVQTLPLLGAIMLTLTGTTTVVFQSKEQWPLVWLHAPLPHTSQCPWGQLLASAVLLLDKYTLQSYGLLCQTIHHNISTQT FNQFIQTSDHPSVPILLHHSHRFKNLGAQTGELWNTFLKTAAPLAPVKALMPVFTLSPVIINTAPCLFSDGSTSRAAYILWDKQILSQRSFPLPPPHKS AQRAELLGLLHGLSSARSWRCLNIFLDSKYLYHYLRTLALGTFQGRSSQAPFQALLPRLLSRKVVYLHHVRSHTNLPDPISRLNALTDALLITPVLQL
HTL1A_P03362_2mut	AVLGLEHLPRPPQISQFPLNPERLQALQHLVRKALEAGHIEPYTGPGNNPVFPVKKANGTWRFIHDLRATNSLTIDLSSSSPGPPDLSSLPTTLAHLQ TIDLRDAFFQIPLPKQFQPYFAFTVPQQCNYGPGTRYAWKVLPQGFKNSPTLFQMQLAHILQPIRQAFPQCTILQYMDDILLASPSHEDLLLLSEATM ASLISHGLPVSENKTQQTPGTIKFLGQIISPNHLTYDAVPTVPIRSRWALPELQALLGEIQWVSKGTPTLRQPLHSLYCALQPHTDPRDQIYLNPSQV QSLVQLRQALSQNCRSRLVQTLPLLGAIMLTLTGTTTVVFQSKEQWPLVWLHAPLPHTSQCPWGQLLASAVLLLDKYTLQSYGLLCQTIHHNISTQT FNQFIQTSDHPSVPILLHHSHRFKNLGAQTGELWNTFLKTAAPLAPVKALMPVFTLSPVIINTAPCLFSDGSTSRAAYILWDKQILSQRSFPLPPPHKS AQRAELLGLLHGLSSARSWRCLNIFLDSKYLYHYLRTLALGTFQGRSSQAPFQALLPRLLSRKVVYLHHVRSHTNLPDPISRLNALTDALLITPVLQL
HTL1A_P03362_2mutB	AVLGLEHLPRPPQISQFPLNPERLQALQHLVRKALEAGHIEPYTGPGNNPVFPVKKANGTWRFIHDLRATNSLTIDLSSSSPGPPDLSSPPTTLAHLQ TIDLRDAFFQIPLPKQFQPYFAFTVPQQCNYGPGTRYAWKVLPQGFKNSPTLFQMQLAHILQPIRQAFPQCTILQYMDDILLASPSHEDLLLLSEATM ASLISHGLPVSENKTQQTPGTIKFLGQIISPNHLTYDAVPTVPIRSRWALPELQALLGEIQWVSKGTPTLRQPLHSLYCALQPHTDPRDQIYLNPSQV QSLVQLRQALSQNCRSRLVQTLPLLGAIMLTLTGTTTVVFQSKEQWPLVWLHAPLPHTSQCPWGQLLASAVLLLDKYTLQSYGLLCQTIHHNISTQT FNQFIQTSDHPSVPILLHHSHRFKNLGAQTGELWNTFLKTAAPLAPVKALMPVFTLSPVIINTAPCLFSDGSTSRAAYILWDKQILSQRSFPLPPPHKS AQRAELLGLLHGLSSARSWRCLNIFLDSKYLYHYLRTLALGTFQGRSSQAPFQALLPRLLSRKVVYLHHVRSHTNLPDPISRLNALTDALLITPVLQL
HTL1C_P14078	AVLGLEHLPRPPEISQFPLNPERLQALQHLVRKALEAGHIEPYTGPGNNPVFPVKKANGTWRFIHDLRATNSLTIDLSSSSPGPPDLSSLPTTLAHLQ TIDLKDAFFQIPLPKQFQPYFAFTVPQQCNYGPGTRYAWRVLPQGFKNSPTLFQMQLAHILQPIRQAFPQCTILQYMDDILLASPSHADLQLLSEATM ASLISHGLPVSENKTQQTPGTIKFLGQIISPNHLTYDAVPKVPIRSRWALPELQALLGEIQWVSKGTPTLRQPLHSLYCALQRHTDPRDQIYLNPSQV QSLVQLRQALSQNCRSRLVQTLPLLGAIMLTLTGTTTVVFQSKQQWPLVWLHAPLPHTSQCPWGQLLASAVLLLDKYTLQSYGLLCQTIHHNISTQT FNQFIQTSDHPSVPILLHHSHRFKNLGAQTGELWNTFLKTTAPLAPVKALMPVFTLSPVIINTAPCLFSDGSTSQAAYILWDKHILSQRSFPLPPPHKS AQRAELLGLLHGLSSARSWRCLNIFLDSKYLYHYLRTLALGTFQGRSSQAPFQALLPRLLSRKVVYLHHVRSHTNLPDPISRLNALTDALLITPVLQL
HTL1C_P14078_2mut	AVLGLEHLPRPPEISQFPLNPERLQALQHLVRKALEAGHIEPYTGPGNNPVFPVKKANGTWRFIHDLRATNSLTIDLSSSSPGPPDLSSLPTTLAHLQ TIDLKDAFFQIPLPKQFQPYFAFTVPQQCNYGPGTRYAWRVLPQGFKNSPTLFQMQLAHILQPIRQAFPQCTILQYMDDILLASPSHADLQLLSEATM ASLISHGLPVSENKTQQTPGTIKFLGQIISPNHLTYDAVPKVPIRSRWALPELQALLGEIQWVSKGTPTLRQPLHSLYCALQPHTDPRDQIYLNPSQV QSLVQLRQALSQNCRSRLVQTLPLLGAIMLTLTGTTTVVFQSKQQWPLVWLHAPLPHTSQCPWGQLLASAVLLLDKYTLQSYGLLCQTIHHNISTQT FNQFIQTSDHPSVPILLHHSHRFKNLGAQTGELWNTFLKTTAPLAPVKALMPVFTLSPVIINTAPCLFSDGSTSQAAYILWDKHILSQRSFPLPPPHKS AQRAELLGLLHGLSSARSWRCLNIFLDSKYLYHYLRTLALGTFQGRSSQAPFQALLPRLLSRKVVYLHHVRSHTNLPDPISRLNALTDALLITPVLQL
HTL1L_P0C211	GLEHLPRPPEISQFPLNPERLQALQHLVRKALEAGHIEPYTGPGNNPVFPVKKANGTWRFIHDLRATNSLTVDLSSSSPGPPDLSSLPTTLAHLQTID LKDAFFQIPLPKQFQPYFAFTVPQQCNYGPGTRYAWKVLPQGFKNSPTLFEMQLASILQPIRQAFPQCVILQYMDDILLASPSPEDLQQLSEATMAS LISHGLPVSQDKTQQTPGTIKFLGQIISPNHITYDAVPTVPIRSRWALPELQALLGEIQWVSKGTPTLRQPLHSLYCALQGHTDPRDQIYLNPSQVQSL MQLQQALSQNCRSRLAQTLPLLGAIMLTLTGTTTVVFQSKQQWPLVWLHAPLPHTSQCPWGQLLASAVLLLDKYTLQSYGLLCQTIHHNISIQTFNQ FIQTSDHPSVPILLHHSHRFKNLGAQTGELWNTFLKTAAPLAPVKALTPVFTLSPIIINTAPCLFSDGSTSQAAYILWDKHILSQRSFPLPPPHKSAQQA ELLGLLHGLSSARSWHCLNIFLDSKYLYHYLRTLALGTFQGKSSQAPFQALLPRLLAHKVIYLHHVRSHTNLPDPISKLNALTDALLITPIL
HTL1L_P0C211_2mut	GLEHLPRPPEISQFPLNPERLQALQHLVRKALEAGHIEPYTGPGNNPVFPVKKANGTWRFIHDLRATNSLTVDLSSSSPGPPDLSSLPTTLAHLQTID LKDAFFQIPLPKQFQPYFAFTVPQQCNYGPGTRYAWKVLPQGFKNSPTLFQMQLASILQPIRQAFPQCVILQYMDDILLASPSPEDLQQLSEATMAS LISHGLPVSQDKTQQTPGTIKFLGQIISPNHITYDAVPTVPIRSRWALPELQALLGEIQWVSKGTPTLRQPLHSLYCALQGHTDPRDQIYLNPSQVQSL MQLQQALSQNCRSRLAQTLPLLGAIMLTLTGTTTVVFQSKQQWPLVWLHAPLPHTSQCPWGQLLASAVLLLDKYTLQSYGLLCQTIHHNISIQTFNQ FIQTSDHPSVPILLHHSHRFKNLGAQTGELWNTFLKTAAPLAPVKALTPVFTLSPIIINTAPCLFSDGSTSQAAYILWDKHILSQRSFPLPPPHKSAQQA ELLGLLHGLSSARSWHCLNIFLDSKYLYHYLRTLAWGTFQGKSSQAPFQALLPRLLAHKVIYLHHVRSHTNLPDPISKLNALTDALLITPIL
HTL1L_P0C211_2mutB	GLEHLPRPPEISQFPLNPERLQALQHLVRKALEAGHIEPYTGPGNNPVFPVKKANGTWRFIHDLRATNSLTVDLSSSSPGPPDLSSPPTTLAHLQTID LKDAFFQIPLPKQFQPYFAFTVPQQCNYGPGTRYAWKVLPQGFKNSPTLFQMQLASILQPIRQAFPQCVILQYMDDILLASPSPEDLQQLSEATMAS LISHGLPVSQDKTQQTPGTIKFLGQIISPNHITYDAVPTVPIRSRWALPELQALLGEIQWVSKGTPTLRQPLHSLYCALQGHTDPRDQIYLNPSQVQSL MQLQQALSQNCRSRLAQTLPLLGAIMLTLTGTTTVVFQSKQQWPLVWLHAPLPHTSQCPWGQLLASAVLLLDKYTLQSYGLLCQTIHHNISIQTFNQ FIQTSDHPSVPILLHHSHRFKNLGAQTGELWNTFLKTAAPLAPVKALTPVFTLSPIIINTAPCLFSDGSTSQAAYILWDKHILSQRSFPLPPPHKSAQQA ELLGLLHGLSSARSWHCLNIFLDSKYLYHYLRTLAWGTFQGKSSQAPFQALLPRLLAHKVIYLHHVRSHTNLPDPISKLNALTDALLITPIL
HTL32_Q0R5R2	GLEHLPPPPEVSQFPLNPERLQALTDLVSRALEAKHIEPYQGPGNNPIFPVKKPNGKWRFIHDLRATNSVTRDLASPSPGPPDLTSLPQGLPHLRTI DLTDAFFQIPLPTIFQPYFAFTLPQPNNYGPGTRYSWRVLPQGFKNSPTLFEQQLSHILTPVRKTFPNSLIIQYMDDILLASPAPGELAALTDKVTNALT KEGLPLSPEKTQATPGPIHFLGQVISQDCITYETLPSINVKSTWSLAELQSMLGELQWVSKGTPVLRSSLHQLYLALRGHRDPRDTIKLTSIQVQALR TIQKALTLNCRSRLVNQLPILALIMLRPTGTTAVLFQTKQKWPLVWLHTPHPATSLRPWGQLLANAVIILDKYSLQHYGQVCKSFHHNISNQALTYYL HTSDQSSVAILLQHSHRFHNLGAQPSGPWRSLLQMPQIFQNIDVLRPPFTISPVVINHAPCLFSDGSASKAAFIIWDRQVIHQQVLSLPSTCSAQAGE LFGLLAGLQKSQPWVALNIFLDSKFLIGHLRRMALGAFPGPSTQCELHTQLLPLLQGKTVYVHHVRSHTLLQDPISRLNEATDALMLAPLLPL
HTL32_Q0R5R2_2mut	GLEHLPPPPEVSQFPLNPERLQALTDLVSRALEAKHIEPYQGPGNNPIFPVKKPNGKWRFIHDLRATNSVTRDLASPSPGPPDLTSLPQGLPHLRTI DLTDAFFQIPLPTIFQPYFAFTLPQPNNYGPGTRYSWRVLPQGFKNSPTLFQQQLSHILTPVRKTFPNSLIIQYMDDILLASPAPGELAALTDKVTNAL TKEGLPLSPEKTQATPGPIHFLGQVISQDCITYETLPSINVKSTWSLAELQSMLGELQWVSKGTPVLRSSLHQLYLALRGHRDPRDTIKLTSIQVQAL RTIQKALTLNCRSRLVNQLPILALIMLRPTGTTAVLFQTKQKWPLVWLHTPHPATSLRPWGQLLANAVIILDKYSLQHYGQVCKSFHHNISNQALTYY LHTSDQSSVAILLQHSHRFHNLGAQPSGPWRSLLQMPQIFQNIDVLRPPFTISPVVINHAPCLFSDGSASKAAFIIWDRQVIHQQVLSLPSTCSAQAG ELFGLLAGLQKSQPWVALNIFLDSKFLIGHLRRMAWGAFPGPSTQCELHTQLLPLLQGKTVYVHHVRSHTLLQDPISRLNEATDALMLAPLLPL
HTL32_Q0R5R2_2mutB	GLEHLPPPPEVSQFPLNPERLQALTDLVSRALEAKHIEPYQGPGNNPIFPVKKPNGKWRFIHDLRATNSVTRDLASPSPGPPDLTSPPQGLPHLRTI DLTDAFFQIPLPTIFQPYFAFTLPQPNNYGPGTRYSWRVLPQGFKNSPTLFQQQLSHILTPVRKTFPNSLIIQYMDDILLASPAPGELAALTDKVTNAL TKEGLPLSPEKTQATPGPIHFLGQVISQDCITYETLPSINVKSTWSLAELQSMLGELQWVSKGTPVLRSSLHQLYLALRGHRDPRDTIKLTSIQVQAL RTIQKALTLNCRSRLVNQLPILALIMLRPTGTTAVLFQTKQKWPLVWLHTPHPATSLRPWGQLLANAVIILDKYSLQHYGQVCKSFHHNISNQALTYY LHTSDQSSVAILLQHSHRFHNLGAQPSGPWRSLLQMPQIFQNIDVLRPPFTISPVVINHAPCLFSDGSASKAAFIIWDRQVIHQQVLSLPSTCSAQAG ELFGLLAGLQKSQPWVALNIFLDSKFLIGHLRRMAWGAFPGPSTQCELHTQLLPLLQGKTVYVHHVRSHTLLQDPISRLNEATDALMLAPLLPL

[0182]

RT 名称	RT 氨基酸序列
HTL3P_Q4U0X6	GLEHLPPPPEVSQFPLNPERLQALTDLVSRALEAKHIEPYQGPGNNPIFPVKKPNGKWRFIHDLRATNSLTRDLASPSPGPPDLTSLPQDLPHLRTID LTDAFFQIPLPAVFQPYFAFTLPQPNNHGPGTRYSWRVLPQGFKNSPTLFEQQLSHILAPVRKAFPNSLIIQYMDDILLASPALRELTALTDKVTNALT KEGLPMSLEKTQATPGSIHFLGQVISPDCITYETLPSIHVKSIWSLAELQSMLGELQWVSKGTPVLRSSLHQLYLALRGHRDPRDTIELTSTQVQALK TIQKALALNCRSRLVSQLPILALIILRPTGTTAVLFQTKQKWPLVWLHTPHPATSLRPWGQLLANAIITLDKYSLQHYGQICKSFHHNISNQALTYYLHT SDQSSVAILLQHSHRFHNLGAQPSGPWRSLLQVPQIFQNIDVLRPPFIISPVVIDHAPCLFSDGATSKAAFILWDKQVIHQQVLPLPSTCSAQAGELF GLLAGLQKSKPWPALNIFLDSKFLIGHLRRMALGAFLGPSTQCDLHARLFPLLQGKTVYVHHVRSHTLLQDPISRLNEATDALMLAPLLPL
HTL3P_Q4U0X6_2mut	GLEHLPPPPEVSQFPLNPERLQALTDLVSRALEAKHIEPYQGPGNNPIFPVKKPNGKWRFIHDLRATNSLTRDLASPSPGPPDLTSLPQDLPHLRTID LTDAFFQIPLPAVFQPYFAFTLPQPNNHGPGTRYSWRVLPQGFKNSPTLFQQQLSHILAPVRKAFPNSLIIQYMDDILLASPALRELTALTDKVTNALT KEGLPMSLEKTQATPGSIHFLGQVISPDCITYETLPSIHVKSIWSLAELQSMLGELQWVSKGTPVLRSSLHQLYLALRGHRDPRDTIELTSTQVQALK TIQKALALNCRSRLVSQLPILALIILRPTGTTAVLFQTKQKWPLVWLHTPHPATSLRPWGQLLANAIITLDKYSLQHYGQICKSFHHNISNQALTYYLHT SDQSSVAILLQHSHRFHNLGAQPSGPWRSLLQVPQIFQNIDVLRPPFIISPVVIDHAPCLFSDGATSKAAFILWDKQVIHQQVLPLPSTCSAQAGELF GLLAGLQKSKPWPALNIFLDSKFLIGHLRRMAWGAFLGPSTQCDLHARLFPLLQGKTVYVHHVRSHTLLQDPISRLNEATDALMLAPLLPL
HTL3P_Q4U0X6_2mutB	GLEHLPPPPEVSQFPLNPERLQALTDLVSRALEAKHIEPYQGPGNNPIFPVKKPNGKWRFIHDLRATNSLTRDLASPSPGPPDLTSPPQDLPHLRTID LTDAFFQIPLPAVFQPYFAFTLPQPNNHGPGTRYSWRVLPQGFKNSPTLFQQQLSHILAPVRKAFPNSLIIQYMDDILLASPALRELTALTDKVTNALT KEGLPMSLEKTQATPGSIHFLGQVISPDCITYETLPSIHVKSIWSLAELQSMLGELQWVSKGTPVLRSSLHQLYLALRGHRDPRDTIELTSTQVQALK TIQKALALNCRSRLVSQLPILALIILRPTGTTAVLFQTKQKWPLVWLHTPHPATSLRPWGQLLANAIITLDKYSLQHYGQICKSFHHNISNQALTYYLHT SDQSSVAILLQHSHRFHNLGAQPSGPWRSLLQVPQIFQNIDVLRPPFIISPVVIDHAPCLFSDGATSKAAFILWDKQVIHQQVLPLPSTCSAQAGELF GLLAGLQKSKPWPALNIFLDSKFLIGHLRRMAWGAFLGPSTQCDLHARLFPLLQGKTVYVHHVRSHTLLQDPISRLNEATDALMLAPLLPL
HTLV2_P03363_2mut	HLPPPPQVDQFPLNLPERLQALNDLVSKALEAGHIEPYSGPGNNPVFPVKKPNGKWRFIHDLRATNAITTTLTSPSPGPPDLTSLPTALPHLQTIDLT DAFFQIPLPKQYQPYFAFTIPQPCNYGPGTRYAWTVLPQGFKNSPTLFQQQLAAVLNPMRKMFPTSTIVQYMDDILLASPTNEELQQLSQLTLQALT THGLPISQEKTQQTPGQIRFLGQVISPNHITYESTPTIPIKSQWTLTELQVILGEIQWVSKGTPILRKHLQSLYSALHPYRDPRACITLTPQQLHALHAIQ QALQHNCRGRLNPALPLLGLISLSTSGTTSVIFQPKQNWPLAWLHTPHPPTSLCPWGHLLACTILTLDKYTLQHYGQLCQSFHHNMSKQALCDFLR NSPHPSVGILIHHMGRFHNLGSQPSGPWKTLLHLPTLLQEPRLLRPIFTLSPVVLDTAPCLFSDGSPQKAAYVLWDQTILQQDITPLPSHETHSAQKG ELLALICGLRAAKPWPSLNIFLDSKYLIKYLHSLAIGAFLGTSAHQTLQAALPPLLQGKTIYLHHVRSHTNLPDPISTFNEYTDSLILAPLVPL
JSRV_P31623	PLGTSDSPVTHADPIDWKSEEPVWVDQWPLTQEKLSAAQQLVQEQLRLGHIEPSTSAWNSPIFVIKKKSGKWRLLQDLRKVNETMMHMGALQPGL PTPSAIPDKSYIIIVIDLKDCFYTIPLAPQDCKRFAFSLPSVNFKEPMQRYQWRVLPQGMTNSPTLCQKFVATAIAPVRQRFPQLYLVHYMDDILLAHTD EHLLYQAFSILKQHLSLNGLVIADEKIQTHFPYNYLGFSLYPRVYNTQLVKLQTDHLKTLNDFQKLLGDINWIRPYLKLPTYTLQPLFDILKGDSDPASP RTLSLEGRTALQSIEEAIRQQQITYCDYQRSWGLYILPTPRAPTGVLYQDKPLRWIYLSATPTKHLLPYYELVAKIIAKGRHEAIQYFGMEPPFICVPYA LEQQDWLFQFSDNWSIAFANYPGQITHHYPSDKLLQFASSHAFIFPKIVRRQPIPEATLIFTDGSSNGTAALIINHQTYYAQTSFSSAQVVELFAVHQA LLTVPTSFNLFTDSSYVVGALQMIETVPIIGTTSPEVLNLFTLIQQVLHCRQHPCFFGHIRAHSTLPGALVQGNHTADVLTKQVFFQS
JSRV_P31623_2mutB	PLGTSDSPVTHADPIDWKSEEPVWVDQWPLTQEKLSAAQQLVQEQLRLGHIEPSTSAWNSPIFVIKKKSGKWRLLQDLRKVNETMMHMGALQPGL PTPSPIPDKSYIIIVIDLKDCFYTIPLAPQDCKRFAFSLPSVNFKEPMQRYQWRVLPQGMTNSPTLCQKFVATAIAPVRQRFPQLYLVHYMDDILLAHTD EHLLYQAFSILKQHLSLNGLVIADEKIQTHFPYNYLGFSLYPRVYNTQLVKLQTDHLKTLNDFQKLLGDINWIRPYLKLPTYTLQPLFDILKGDSDPASP RTLSLEGRTALQSIEEAIRQQQITYCDYQRSWGLYILPTPRAPTGVLYQDKPLRWIYLSATPTKHLLPYYELVAKIIAKGRHEAIQYFGMEPPFICVPYA LEQQDWLFQFSDNWSIAFANYPGQITHHYPSDKLLQFASSHAFIFPKIVRRQPIPEATLIFTDGSSNGTAALIINHQTYYAQTSFSSAQVVELFAVHQA LLTVPTSFNLFTDSSYVVGALQMIETVPIIGTTSPEVLNLFTLIQQVLHCRQHPCFFGHIRAHSTLPGALVQGNHTADVLTKQVFFQS
KORV_Q9TTC1	TLGDQGSRGSDPLPEPRVTLTVEGIPTEFLVNTGAEHSVLTKPMGKMGSKRTVVAGATGSKVYPWTTKRLLKIGQKQVTHSFLVIPECPAPLLGRD LLTKLKAQIQFSTEGPQVTWEDRPAMCLVLNLEEEYRLHEKPVPPSIDPSWLQLFPMVWAEKAGMGLANQVPPVVVELKSDASPVAVRQYPMSKE AREGIRPHIQRFLDLGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLSSLPPSHTWYSVLDLKDAFFCLKLHPNSQP LFAFEWRDPEKGNTGQLTWTRLPQGFKNSPTLFDEALHRDLASFRALNPQVVMLQYVDDLLVAAPTYRDCKEGTRRLLQELSKLGYRVSAKKAQL CREEVTYLGYLLKGGKRWLTPARKATVMKIPTPTTPRQVREFLGTAGFCRLWIPGFASLAAPLYPLTRKVPFTWTEAHQEAFGRIKEALLSAPALA LPDLTKPFALYVDEKEGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPTCLKAIAAVALLLKDADKLTLGQNVLVIAPHNLESIVRQPPDRWMTNA RMTHYQSLLLNERVSFAPPAILNPATLLPVESDDTPIHICSEILAEETGTRPDLRDQPLPGVPAWYTDGSSFIMDGRRQAGAAIVDNKRTVWASNLPE GTSAQKAELIALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGLLTSAGKDIKNKEEILALLEAIHLPKRVAIIHCPGHQRGTDPVATGNRKADE AAKQAAQSTRILTETTKN
KORV_Q9TTC1_3mut	TLGDQGSRGSDPLPEPRVTLTVEGIPTEFLVNTGAEHSVLTKPMGKMGSKRTVVAGATGSKVYPWTTKRLLKIGQKQVTHSFLVIPECPAPLLGRD LLTKLKAQIQFSTEGPQVTWEDRPAMCLVLNLEEEYRLHEKPVPPSIDPSWLQLFPMVWAEKAGMGLANQVPPVVVELKSDASPVAVRQYPMSKE AREGIRPHIQRFLDLGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLSSLPPSHTWYSVLDLKDAFFCLKLHPNSQP LFAFEWRDPEKGNTGQLTWTRLPQGFKNSPTLFNEALHRDLASFRALNPQVVMLQYVDDLLVAAPTYRDCKEGTRRLLQELSKLGYRVSAKKAQL CREEVTYLGYLLKGGKRWLTPARKATVMKIPTPTTPRQVREFLGTAGFCRLWIPGFASLAAPLYPLTRPKVPFTWTEAHQEAFGRIKEALLSAPALA LPDLTKPFALYVDEKEGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPTCLKAIAAVALLLKDADKLTLGQNVLVIAPHNLESIVRQPPDRWMTNA RMTHYQSLLLNERVSFAPPAILNPATLLPVESDDTPIHICSEILAEETGTRPDLRDQPLPGVPAWYTDGSSFIMDGRRQAGAAIVDNKRTVWASNLPE GTSAQKAELIALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGKDIKNKEEILALLEAIHLPKRVAIIHCPGHQRGTDPVATGNRKAD EAAKQAAQSTRILTETTKN
KORV_Q9TTC1_3mutA	TLGDQGSRGSDPLPEPRVTLTVEGIPTEFLVNTGAEHSVLTKPMGKMGSKRTVVAGATGSKVYPWTTKRLLKIGQKQVTHSFLVIPECPAPLLGRD LLTKLKAQIQFSTEGPQVTWEDRPAMCLVLNLEEEYRLHEKPVPPSIDPSWLQLFPMVWAEKAGMGLANQVPPVVVELKSDASPVAVRQYPMSKE AREGIRPHIQRFLDLGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLSSLPPSHTWYSVLDLKDAFFCLKLHPNSQP LFAFEWRDPEKGNTGQLTWTRLPQGFKNSPTLFNEALHRDLASFRALNPQVVMLQYVDDLLVAAPTYRDCKEGTRRLLQELSKLGYRVSAKKAQL CREEVTYLGYLLKGGKRWLTPARKATVMKIPTPTTPRQVREFLGKAGFCRLFIPGFASLAAPLYPLTRPKVPFTWTEAHQEAFGRIKEALLSAPALAL PDLTKPFALYVDEKEGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPTCLKAIAAVALLLKDADKLTLGQNVLVIAPHNLESIVRQPPDRWMTNA RMTHYQSLLLNERVSFAPPAILNPATLLPVESDDTPIHICSEILAEETGTRPDLRDQPLPGVPAWYTDGSSFIMDGRRQAGAAIVDNKRTVWASNLPE GTSAQKAELIALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGKDIKNKEEILALLEAIHLPKRVAIIHCPGHQRGTDPVATGNRKAD EAAKQAAQSTRILTETTKN
KORV_Q9TTC1-Pro	LLGRDLLTKLKAQIQFSTEGPQVTWEDRPAMCLVLNLEEEYRLHEKPVPPSIDPSWLQLFPMVWAEKAGMGLANQVPPVVVELKSDASPVAVRQY PMSKEAREGIRPHIQRFLDLGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLSSLPPSHTWYSVLDLKDAFFCLKLH PNSQPLFAFEWRDPEKGNTGQLTWTRLPQGFKNSPTLFDEALHRDLASFRALNPQVVMLQYVDDLLVAAPTYRDCKEGTRRLLQELSKLGYRVSA KKAQLCREEVTYLGYLLKGGKRWLTPARKATVMKIPTPTTPRQVREFLGTAGFCRLWIPGFASLAAPLYPLTREKVPFTWTEAHQEAFGRIKEALLS APALALPDLTKPFALYVDEKEGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPTCLKAIAAVALLLKDADKLTLGQNVLVIAPHNLESIVRQPPDR WMTNARMTHYQSLLLNERVSFAPPAILNPATLLPVESDDTPIHICSEILAEETGTRPDLRDQPLPGVPAWYTDGSSFIMDGRRQAGAAIVDNKRTVW ASNLPEGTSAQKAELIALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGLLTSAGKDIKNKEEILALLEAIHLPKRVAIIHCPGHQRGTDPVATG NRKADEAAKQAAQSTRILTETTKN
KORV_Q9TTC1-Pro_3mut	LLGRDLLTKLKAQIQFSTEGPQVTWEDRPAMCLVLNLEEEYRLHEKPVPPSIDPSWLQLFPMVWAEKAGMGLANQVPPVVVELKSDASPVAVRQY PMSKEAREGIRPHIQRFLDLGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLSSLPPSHTWYSVLDLKDAFFCLKLH PNSQPLFAFEWRDPEKGNTGQLTWTRLPQGFKNSPTLFNEALHRDLASFRALNPQVVMLQYVDDLLVAAPTYRDCKEGTRRLLQELSKLGYRVSA KKAQLCREEVTYLGYLLKGGKRWLTPARKATVMKIPTPTTPRQVREFLGTAGFCRLWIPGFASLAAPLYPLTRPKVPFTWTEAHQEAFGRIKEALLS APALALPDLTKPFALYVDEKEGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPTCLKAIAAVALLLKDADKLTLGQNVLVIAPHNLESIVRQPPDR WMTNARMTHYQSLLLNERVSFAPPAILNPATLLPVESDDTPIHICSEILAEETGTRPDLRDQPLPGVPAWYTDGSSFIMDGRRQAGAAIVDNKRTVW ASNLPEGTSAQKAELIALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGKDIKNKEEILALLEAIHLPKRVAIIHCPGHQRGTDPVATG NRKADEAAKQAAQSTRILTETTKN
KORV_Q9TT	LLGRDLLTKLKAQIQFSTEGPQVTWEDRPAMCLVLNLEEEYRLHEKPVPPSIDPSWLQLFPMVWAEKAGMGLANQVPPVVVELKSDASPVAVRQY

[0183]

RT 名称	RT 氨基酸序列
C1-Pro_3mutA	PMSKEAREGIRPHIQRFLDLGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLSSLPPSHTWYSVLDLKDAFFCLKLH PNSQPLFAFEWRDPEKGNTGQLTWTRLPQGFKNSPTLFNEALHRDLASFRALNPQVVMLQYVDDLLVAAPTYRDCKEGTRRLLQELSKLGYRVSA KKAQLCREEVTYLGYLLKGGKRWLTPARKATVMKIPTPTTPRQVREFLGKAGFCRLFIPGFASLAAPLYPLTRPKVPFTWTEAHQEAFGRIKEALLS APALALPDLTKPFALYVDEKEGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPTCLKAIAAVALLLKDADKLTLGQNVLVIAPHNLESIVRQPPDR WMTNARMTHYQSLLLNERVSFAPPAILNPATLLPVESDDTPIHICSEILAEETGTRPDLRDQPLPGVPAWYTDGSSFIMDGRRQAGAAIVDNKRTVW ASNLPEGTSAQKAELIALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGKDIKNKEEILALLEAIHLPKRVAIIHCPGHQRGTDPVATG NRKADEAAKQAAQSTRILTETTKN
MLVAV_P03356	TLNLEDEYRLYETSAEPEVSPGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEAKLGIKPHIQRLLDQGILVPCQSPWNTP LLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHRWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGMGISGQLTWTRLPQGF KNSPTLFDEALHRDLADFRIQHPDLILLQYVDDILLAATSELDCQQGTRALLLTLGNLGYRASAKKAQLCQKQVKYLGYLLKEGQRWLTEARKETVM GQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKL GPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLRKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQAMLLDTDRVQFGPVVALN PATLLPLPEEGAPHDCLEILAETHGTRPDLTDQPIPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVIWARALPAGTSAQRAELIALTQALKMAEGK RLNVYTDSRYAFATAHIHGEIYRRRGLLTSEGREIKNKSEILALLKALFLPKRLSIIHCLGHQKGDSAEARGNRLADQAAREAAIKTPPDTSTLL
MLVAV_P03356_3mut	TLNLEDEYRLYETSAEPEVSPGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEAKLGIKPHIQRLLDQGILVPCQSPWNTP LLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHRWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGMGISGQLTWTRLPQGF KNSPTLFNEALHRDLADFRIQHPDLILLQYVDDILLAATSELDCQQGTRALLLTLGNLGYRASAKKAQLCQKQVKYLGYLLKEGQRWLTEARKETVM GQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKL GPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLRKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQAMLLDTDRVQFGPVVALN PATLLPLPEEGAPHDCLEILAETHGTRPDLTDQPIPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVIWARALPAGTSAQRAELIALTQALKMAEGK RLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGREIKNKSEILALLKALFLPKRLSIIHCLGHQKGDSAEARGNRLADQAAREAAIKTPPDTSTLL
MLVAV_P03356_3mutA	TLNLEDEYRLYETSAEPEVSPGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEAKLGIKPHIQRLLDQGILVPCQSPWNTP LLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHRWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGMGISGQLTWTRLPQGF KNSPTLFNEALHRDLADFRIQHPDLILLQYVDDILLAATSELDCQQGTRALLLTLGNLGYRASAKKAQLCQKQVKYLGYLLKEGQRWLTEARKETVM GQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKL GPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLRKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQAMLLDTDRVQFGPVVALN PATLLPLPEEGAPHDCLEILAETHGTRPDLTDQPIPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVIWARALPAGTSAQRAELIALTQALKMAEGK RLNVYTDSRYAFATAHIHGEIYRRRGLLTSEGREIKNKSEILALLKALFLPKRLSIIHCLGHQKGDSAEARGNRLADQAAREAAIKTPPDTSTLL
MLVBM_Q7SVK7	TLGIEDEYRLHETSTEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIQQYPMSHEARLGIKPHIQRLLDQGILVPCQSPWNTPL LPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGMGISGQLTWTRLPQGFK NSPTLFDEALHRDLADFRIQHPDLILLQYVDDILLAATSELDCQQGTRALLQTLGDLGYRASAKKAQICQKQVKYLGYLLREGQRWLTEARKETVMG QPVPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFSWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLG PWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQAMLLDTDRVQFGPVVALNPA TLLPLPEEGAPHDCLEILAETHGTRPDLTDQPIPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVIWAGALPAGTSAQRAELIALTQALKMAEGKRL NVYTDSRYAFATAHIHGEIYRRRGLLTSEGREIKNKSEILALLKALFLPKRLSIIHCLGHQKGDSAEARGNRLADQAAREAAIKTPPDTSTLL
MLVBM_Q7SVK7_3mut	TLGIEDEYRLHETSTEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIQQYPMSHEARLGIKPHIQRLLDQGILVPCQSPWNTPL LPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGMGISGQLTWTRLPQGFK NSPTLFNEALHRDLADFRIQHPDLILLQYVDDILLAATSELDCQQGTRALLQTLGDLGYRASAKKAQICQKQVKYLGYLLREGQRWLTEARKETVMG QPVPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKPGTLFSWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLG PWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQAMLLDTDRVQFGPVVALNPA TLLPLPEEGAPHDCLEILAETHGTRPDLTDQPIPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVIWAGALPAGTSAQRAELIALTQALKMAEGKRL NVYTDSRYAFATAHIHGEIYRRRGWLTSEGREIKNKSEILALLKALFLPKRLSIIHCLGHQKGDSAEARGNRLADQAAREAAIKTPPDTSTLL
MLVBM_Q7SVK7_3mut	TLGIEDEYRLHETSTEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIQQYPMSHEARLGIKPHIQRLLDQGILVPCQSPWNTPL LPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGMGISGQLTWTRLPQGFK NSPTLFNEALHRDLADFRIQHPDLILLQYVDDILLAATSELDCQQGTRALLQTLGDLGYRASAKKAQICQKQVKYLGYLLREGQRWLTEARKETVMG QPVPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKPGTLFSWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLG PWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQAMLLDTDRVQFGPVVALNPA TLLPLPEEGAPHDCLEILAETHGTRPDLTDQPIPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVIWAGALPAGTSAQRAELIALTQALKMAEGKRL NVYTDSRYAFATAHIHGEIYRRRGWLTSEGREIKNKSEILALLKALFLPKRLSIIHCLGHQKGDSAEARGNRLADQAAREAAIKTPPDTSTLL
MLVBM_Q7SVK7_3mutA_WS	LGIEDEYRLHETSTEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIQQYPMSHEARLGIKPHIQRLLDQGILVPCQSPWNTPLL PVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGMGISGQLTWTRLPQGFKN SPTLFNEALHRDLADFRIQHPDLILLQYVDDILLAATSELDCQQGTRALLQTLGDLGYRASAKKAQICQKQVKYLGYLLREGQRWLTEARKETVMGQ PVPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFSWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGP WRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQAMLLDTDRVQFGPVVALNPAT LLPLPEEGAPHDCLEILAETHGTRPDLTDQPIPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVIWAGALPAGTSAQRAELIALTQALKMAEGKRLN VYTDSRYAFATAHIHGEIYRRRGWLTSEGREIKNKSEILALLKALFLPKRLSIIHCLGHQKGDSAEARGNRLADQAAREAAIKTPPDTSTLLI
MLVBM_Q7SVK7_3mutA_WS	LGIEDEYRLHETSTEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIQQYPMSHEARLGIKPHIQRLLDQGILVPCQSPWNTPLL PVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGMGISGQLTWTRLPQGFKN SPTLFNEALHRDLADFRIQHPDLILLQYVDDILLAATSELDCQQGTRALLQTLGDLGYRASAKKAQICQKQVKYLGYLLREGQRWLTEARKETVMGQ PVPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFSWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGP WRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQAMLLDTDRVQFGPVVALNPAT LLPLPEEGAPHDCLEILAETHGTRPDLTDQPIPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVIWAGALPAGTSAQRAELIALTQALKMAEGKRLN VYTDSRYAFATAHIHGEIYRRRGWLTSEGREIKNKSEILALLKALFLPKRLSIIHCLGHQKGDSAEARGNRLADQAAREAAIKTPPDTSTLLI
MLVCB_P08361	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPL LPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFK NSPTLFDEALHRDLAGFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGDLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMG QPIPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPDQQKAFQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGP WRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQAMLLDTDRVQFGPVVALNPAT LLPLPEEGLQHDCLDILAEAHGTRSDLMDQPLPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVIWARALPAGTSAQRAELIALTQALKMAEGKKL NVYTDSRYAFATAHIHGEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGNSAEARGNRMADQAAREVATRETPETSTLL
MLVCB_P08361_3mut	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPL LPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFK NSPTLFNEALHRDLAGFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGDLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMG

[0184]

RT 名称	RT 氨基酸序列
	QPIPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAFQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGP WRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPAT LLPLPEEGLQHDCLDILAEAHGTRSDLMDQPLPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVIWARALPAGTSAQRAELIALTQALKMAEGKKL NVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGNSAEARGNRMADQAAREVATRETPETSTLL
MLVCB_P08361_3mutA	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPL LPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFK NSPTLFNEALHRDLAGFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGDLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMG QPIPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAFQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGP WRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPAT LLPLPEEGLQHDCLDILAEAHGTRSDLMDQPLPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVIWARALPAGTSAQRAELIALTQALKMAEGKKL NVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGNSAEARGNRMADQAAREVATRETPETSTLL
MLVF5_P26810	TLNIEDEYRLHETSKGPDVPLGSTWLSDFPQAWAETGGMGLAFRQAPLIISLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPL LPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQSLFAFEWKDPEMGISGQLTWTRLPQGFK NSPTLFDEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGDLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMG QPTPKTPRQLREFLGTAGLCRLWIPGFAEMAAPLYPLTKTGTLFKWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLG PWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDVGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPIVALNPAT LLPLPEEGLQHDCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSFLQEGQRRAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAAGKKL NVYTDSRYAFATAHIHGEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGNHAEARGNRMADQAAREVATRETPETSTLL
MLVF5_P26810_3mut	TLNIEDEYRLHETSKGPDVPLGSTWLSDFPQAWAETGGMGLAFRQAPLIISLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPL LPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQSLFAFEWKDPEMGISGQLTWTRLPQGFK NSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGDLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMG QPTPKTPRQLREFLGTAGLCRLWIPGFAEMAAPLYPLTKPGTLFKWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLG PWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDVGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPIVALNPAT LLPLPEEGLQHDCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSFLQEGQRRAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAAGKKL NVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGNHAEARGNRMADQAAREVATRETPETSTLL
MLVF5_P26810_3mutA	TLNIEDEYRLHETSKGPDVPLGSTWLSDFPQAWAETGGMGLAFRQAPLIISLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPL LPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQSLFAFEWKDPEMGISGQLTWTRLPQGFK NSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGDLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMG QPTPKTPRQLREFLGKAGLCRLFIPGFAEMAAPLYPLTKPGTLFKWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGP WRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDVGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPIVALNPATL LPLPEEGLQHDCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSFLQEGQRRAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAAGKKLN VYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGNHAEARGNRMADQAAREVATRETPETSTLL
MLVFF_P26809_3mut	TLNIEDEYRLHETSKGPDVPLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPL LPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQSLFAFEWRDPEMGISGQLTWTRLPQGFK NSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGDLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMG QPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKPGTLFEWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLG PWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPIVALNPAT LLPLPEEGLQHDCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVVWAKALPAGTSAQRAELIALTQALKMAEGKKL NVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGNRAEARGNRMADQAAREVATRETPETSTLL
MLVFF_P26809_3mutA	TLNIEDEYRLHETSKGPDVPLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPL LPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQSLFAFEWRDPEMGISGQLTWTRLPQGFK NSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGDLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMG QPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFEWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGP WRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPIVALNPATL LPLPEEGLQHDCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVVWAKALPAGTSAQRAELIALTQALKMAEGKKLN VYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGNRAEARGNRMADQAAREVATRETPETSTLL
MLVMS_P03355	TLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPL LPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFK NSPTLFDEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMG QPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLG PWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPA TLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKL NVYTDSRYAFATAHIHGEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLL
MLVMS_P03355_3mut	TLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPL LPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFK NSPTLFDEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMG QPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLG PWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPA TLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKL NVYTDSRYAFATAHIHGEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLL
MLVMS_P03355_3mutA_WS	TLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPL LPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFK NSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMG QPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGP WRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPAT LLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLN

[0185]

RT 名称	RT 氨基酸序列
	VYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLL
MLVMS_P03355_3mutA_WS	TLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPL LPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFK NSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMG QPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGP WRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPAT LLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLN VYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLL	
MLVMS_P03355_PLV919	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPL LPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFK NSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMG QPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGP WRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPAT LLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLN VYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSPS GGSKRTADGSEFE
MLVMS_P03355_PLV919	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPL LPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFK NSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMG QPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGP WRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPAT LLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLN VYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSPS GGSKRTADGSEFE
MLVRD_P11227	TLNIEDEYRLHEISTEPDVSPGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEAKLGIKPHIQRLLDQGILVPCQSPWNTPL LPVKKPGTNDYRPVQGLREVNKRVEDIHPTVPNPYNLLSGLPTSHRWYTVLDLKDAFFCLRLHPTSQPLFASEWRDPGMGISGQLTWTRLPQGFK NSPTLFDEALHRGLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLKTLGNLGYRASAKKAQICQKQVKYLGYLLREGQRWLTEARKETVMG QPTPKTPRQLREFLGTAGFCRLWIPRFAEMAAPLYPLTKTGTLFNWGPDQQKAYHEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLG PWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQAMLLDTDRVQFGPVVALNPA TLLPLPEEGAPHDCLEILAETHGTEPDLTDQPIPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVIWARALPAGTSAQRAELIALTQALKMAEGKRL NVYTDSRYAFATAHIHGEIYKRRGLLTSEGREIKNKSEILALLKALFLPKRLSIIHCLGHQKGDSAEARGNRLADQAAREAAIKTPPDTSTLL
MLVRD_P11227_3mut	TLNIEDEYRLHEISTEPDVSPGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEAKLGIKPHIQRLLDQGILVPCQSPWNTPL LPVKKPGTNDYRPVQGLREVNKRVEDIHPTVPNPYNLLSGLPTSHRWYTVLDLKDAFFCLRLHPTSQPLFASEWRDPGMGISGQLTWTRLPQGFK NSPTLFNEALHRGLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLKTLGNLGYRASAKKAQICQKQVKYLGYLLREGQRWLTEARKETVMG QPTPKTPRQLREFLGTAGFCRLWIPRFAEMAAPLYPLTKPGTLFNWGPDQQKAYHEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLG PWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQAMLLDTDRVQFGPVVALNPA TLLPLPEEGAPHDCLEILAETHGTEPDLTDQPIPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVIWARALPAGTSAQRAELIALTQALKMAEGKRL NVYTDSRYAFATAHIHGEIYKRRGWLTSEGREIKNKSEILALLKALFLPKRLSIIHCLGHQKGDSAEARGNRLADQAAREAAIKTPPDTSTLL
MMTVB_P03365	WVQEISDSRPMLHIYLNGRRFLGLLNTGADKTCIAGRDWPANWPIHQTESSLQGLGMACGVARSSQPLRWQHEDKSGIIHPFVIPTLPFTLWGRDI MKDIKVRLMTDSPDDSQDLMIGAIESNLFADQISWKSDQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWRLLQD LRAVNATMHDMGALQPGLPSPVAVPKGWEIIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFKRPYQRFQWKVLPQGMKNSPTLCQKFVDKAILTVRD KYQDSYIVHYMDDILLAHPSRSIVDEILTSMIQALNKHGLVVSTEKIQKYDNLKYLGTHIQGDSVSYQKLQIRTDKLRTLNDFQKLLGNINWIRPFLKLT TGELKPLFEILNGDSNPISTRKLTPEACKALQLMNERLSTARVKRLDLSQPWSLCILKTEYTPTACLWQDGVVEWIHLPHISPKVITPYDIFCTQLIIKG RHRSKELFSKDPDYIVVPYTKVQFDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTFTLQTAIIFPHMTSTTPLEKGIVIFTDGSANGRSVTYIQGREPII KENTQNTAQQAEIVAVITAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPRTKIYTELKHLQRLIHKRQEKFYIGHIRGHTGLPGPLAQGNAYADSLTRI LT
MMTVB_P03365	WVQEISDSRPMLHIYLNGRRFLGLLNTGADKTCIAGRDWPANWPIHQTESSLQGLGMACGVARSSQPLRWQHEDKSGIIHPFVIPTLPFTLWGRDI MKDIKVRLMTDSPDDSQDLMIGAIESNLFADQISWKSDQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWRLLQD LRAVNATMHDMGALQPGLPSPVAVPKGWEIIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFKRPYQRFQWKVLPQGMKNSPTLCQKFVDKAILTVRD KYQDSYIVHYMDDILLAHPSRSIVDEILTSMIQALNKHGLVVSTEKIQKYDNLKYLGTHIQGDSVSYQKLQIRTDKLRTLNDFQKLLGNINWIRPFLKLT TGELKPLFEILNGDSNPISTRKLTPEACKALQLMNERLSTARVKRLDLSQPWSLCILKTEYTPTACLWQDGVVEWIHLPHISPKVITPYDIFCTQLIIKG RHRSKELFSKDPDYIVVPYTKVQFDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTFTLQTAIIFPHMTSTTPLEKGIVIFTDGSANGRSVTYIQGREPII KENTQNTAQQAEIVAVITAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPRTKIYTELKHLQRLIHKRQEKFYIGHIRGHTGLPGPLAQGNAYADSLTRI LT
MMTVB_P03365_2mut	WVQEISDSRPMLHIYLNGRRFLGLLNTGADKTCIAGRDWPANWPIHQTESSLQGLGMACGVARSSQPLRWQHEDKSGIIHPFVIPTLPFTLWGRDI MKDIKVRLMTDSPDDSQDLMIGAIESNLFADQISWKSDQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWRLLQD LRAVNATMHDMGALQPGLPSPVAVPKGWEIIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFKRPYQRFQWKVLPQGMKNSPTLCQKFVDKAILTVRD KYQDSYIVHYMDDILLAHPSRSIVDEILTSMIQALNKHGLVVSTEKIQKYDNLKYLGTHIQGDSVSYQKLQIRTDKLRTLNDFQKLLGNINWIRPFLKLT TGELKPLFEILNPDSNPISTRKLTPEACKALQLMNERLSTARVKRLDLSQPWSLCILKTEYTPTACLWQDGVVEWIHLPHISPKVITPYDIFCTQLIIKG RHRSKELFSKDPDYIVVPYTKVQFDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTFTLQTAIIFPHMTSTTPLEKGIVIFTDGSANGRSVTYIQGREPII KENTQNTAQQAEIVAVITAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPRTKIYTELKHLQRLIHKRQEKFYIGHIRGHTGLPGPLAQGNAYADSLTRI LT
MMTVB_P03365_2mut_WS	VQEISDSRPMLHIYLNGRRFLGLLDTGADKTCIAGRDWPANWPIHQTESSLQGLGMACGVARSSQPLRWQHEDKSGIIHPFVIPTLPFTLWGRDIMK DIKVRLMTDSPDDSQDLMIGAIESNLFADQISWKSDQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWRLLQDLR AVNATMHDMGALQPGLPSPVAVPKGWEIIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFKRPYQRFQWKVLPQGMKNSPTLCQKFVDKAILTVRDKY QDSYIVHYMDDILLAHPSRSIVDEILTSMIQALNKHGLVVSTEKIQKYDNLKYLGTHIQGDSVSYQKLQIRTDKLRTLNDFQKLLGNINWIRPFLKLTTG ELKPLFEILNPDSNPISTRKLTPEACKALQLMNERLSTARVKRLDLSQPWSLCILKTEYTPTACLWQDGVVEWIHLPHISPKVITPYDIFCTQLIIKGRH RSKELFSKDPDYIVVPYTKVQFDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTFTLQTAIIFPHMTSTTPLEKGIVIFTDGSANGRSVTYIQGREPIIKE NTQNTAQQAEIVAVITAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPRTKIYTELKHLQRLIHKRQEKFYIGHIRGHTGLPGPLAQGNAYADSLTRILT A
MMTVB_P03365_2mut_WS	VQEISDSRPMLHIYLNGRRFLGLLDTGADKTCIAGRDWPANWPIHQTESSLQGLGMACGVARSSQPLRWQHEDKSGIIHPFVIPTLPFTLWGRDIMK DIKVRLMTDSPDDSQDLMIGAIESNLFADQISWKSDQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWRLLQDLR AVNATMHDMGALQPGLPSPVAVPKGWEIIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFKRPYQRFQWKVLPQGMKNSPTLCQKFVDKAILTVRDKY QDSYIVHYMDDILLAHPSRSIVDEILTSMIQALNKHGLVVSTEKIQKYDNLKYLGTHIQGDSVSYQKLQIRTDKLRTLNDFQKLLGNINWIRPFLKLTTG ELKPLFEILNPDSNPISTRKLTPEACKALQLMNERLSTARVKRLDLSQPWSLCILKTEYTPTACLWQDGVVEWIHLPHISPKVITPYDIFCTQLIIKGRH RSKELFSKDPDYIVVPYTKVQFDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTFTLQTAIIFPHMTSTTPLEKGIVIFTDGSANGRSVTYIQGREPIIKE NTQNTAQQAEIVAVITAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPRTKIYTELKHLQRLIHKRQEKFYIGHIRGHTGLPGPLAQGNAYADSLTRILT A
MMTVB_P03365_2mutB	WVQEISDSRPMLHIYLNGRRFLGLLNTGADKTCIAGRDWPANWPIHQTESSLQGLGMACGVARSSQPLRWQHEDKSGIIHPFVIPTLPFTLWGRDI MKDIKVRLMTDSPDDSQDLMIGAIESNLFADQISWKSDQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWRLLQD

[0186]

RT 名称	RT 氨基酸序列
	LRAVNATMHDMGALQPGLPSPVAPPKGWEIIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFKRPYQRFQWKVLPQGMKNSPTLCQKFVDKAILTVRD KYQDSYIVHYMDDILLAHPSRSIVDEILTSMIQALNKHGLVVSTEKIQKYDNLKYLGTHIQGDSVSYQKLQIRTDKLRTLNDFQKLLGNINWIRPFLKLT TGELKPLFEILNPDSNPISTRKLTPEACKALQLMNERLSTARVKRLDLSQPWSLCILKTEYTPTACLWQDGVVEWIHLPHISPKVITPYDIFCTQLIIKG RHRSKELFSKDPDYIVVPYTKVQFDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTFTLQTAIIFPHMTSTTPLEKGIVIFTDGSANGRSVTYIQGREPII KENTQNTAQQAEIVAVITAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPRTKIYTELKHLQRLIHKRQEKFYIGHIRGHTGLPGPLAQGNAYADSLTRI LT
MMTVB_P03365_2mutB	WVQEISDSRPMLHIYLNGRRFLGLLNTGADKTCIAGRDWPANWPIHQTESSLQGLGMACGVARSSQPLRWQHEDKSGIIHPFVIPTLPFTLWGRDI MKDIKVRLMTDSPDDSQDLMIGAIESNLFADQISWKSDQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWRLLQD LRAVNATMHDMGALQPGLPSPVAPPKGWEIIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFKRPYQRFQWKVLPQGMKNSPTLCQKFVDKAILTVRD KYQDSYIVHYMDDILLAHPSRSIVDEILTSMIQALNKHGLVVSTEKIQKYDNLKYLGTHIQGDSVSYQKLQIRTDKLRTLNDFQKLLGNINWIRPFLKLT TGELKPLFEILNPDSNPISTRKLTPEACKALQLMNERLSTARVKRLDLSQPWSLCILKTEYTPTACLWQDGVVEWIHLPHISPKVITPYDIFCTQLIIKG RHRSKELFSKDPDYIVVPYTKVQFDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTFTLQTAIIFPHMTSTTPLEKGIVIFTDGSANGRSVTYIQGREPII KENTQNTAQQAEIVAVITAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPRTKIYTELKHLQRLIHKRQEKFYIGHIRGHTGLPGPLAQGNAYADSLTRI LT
MMTVB_P03365_2mutB_WS	VQEISDSRPMLHIYLNGRRFLGLLDTGADKTCIAGRDWPANWPIHQTESSLQGLGMACGVARSSQPLRWQHEDKSGIIHPFVIPTLPFTLWGRDIMK DIKVRLMTDSPDDSQDLMIGAIESNLFADQISWKSDQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWRLLQDLR AVNATMHDMGALQPGLPSPPAVPKGWEIIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFKRPYQRFQWKVLPQGMKNSPTLCQKFVDKAILTVRDKY QDSYIVHYMDDILLAHPSRSIVDEILTSMIQALNKHGLVVSTEKIQKYDNLKYLGTHIQGDSVSYQKLQIRTDKLRTLNDFQKLLGNINWIRPFLKLTTG ELKPLFEILNPDSNPISTRKLTPEACKALQLMNERLSTARVKRLDLSQPWSLCILKTEYTPTACLWQDGVVEWIHLPHISPKVITPYDIFCTQLIIKGRH RSKELFSKDPDYIVVPYTKVQFDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTFTLQTAIIFPHMTSTTPLEKGIVIFTDGSANGRSVTYIQGREPIIKE NTQNTAQQAEIVAVITAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPRTKIYTELKHLQRLIHKRQEKFYIGHIRGHTGLPGPLAQGNAYADSLTRILT A
MMTVB_P03365_2mutB_WS	VQEISDSRPMLHIYLNGRRFLGLLDTGADKTCIAGRDWPANWPIHQTESSLQGLGMACGVARSSQPLRWQHEDKSGIIHPFVIPTLPFTLWGRDIMK DIKVRLMTDSPDDSQDLMIGAIESNLFADQISWKSDQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWRLLQDLR AVNATMHDMGALQPGLPSPPAVPKGWEIIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFKRPYQRFQWKVLPQGMKNSPTLCQKFVDKAILTVRDKY QDSYIVHYMDDILLAHPSRSIVDEILTSMIQALNKHGLVVSTEKIQKYDNLKYLGTHIQGDSVSYQKLQIRTDKLRTLNDFQKLLGNINWIRPFLKLTTG ELKPLFEILNPDSNPISTRKLTPEACKALQLMNERLSTARVKRLDLSQPWSLCILKTEYTPTACLWQDGVVEWIHLPHISPKVITPYDIFCTQLIIKGRH RSKELFSKDPDYIVVPYTKVQFDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTFTLQTAIIFPHMTSTTPLEKGIVIFTDGSANGRSVTYIQGREPIIKE NTQNTAQQAEIVAVITAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPRTKIYTELKHLQRLIHKRQEKFYIGHIRGHTGLPGPLAQGNAYADSLTRILT A
MMTVB_P03365_WS	VQEISDSRPMLHIYLNGRRFLGLLDTGADKTCIAGRDWPANWPIHQTESSLQGLGMACGVARSSQPLRWQHEDKSGIIHPFVIPTLPFTLWGRDIMK DIKVRLMTDSPDDSQDLMIGAIESNLFADQISWKSDQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWRLLQDLR AVNATMHDMGALQPGLPSPVAVPKGWEIIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFKRPYQRFQWKVLPQGMKNSPTLCQKFVDKAILTVRDKY QDSYIVHYMDDILLAHPSRSIVDEILTSMIQALNKHGLVVSTEKIQKYDNLKYLGTHIQGDSVSYQKLQIRTDKLRTLNDFQKLLGNINWIRPFLKLTTG ELKPLFEILNGDSNPISTRKLTPEACKALQLMNERLSTARVKRLDLSQPWSLCILKTEYTPTACLWQDGVVEWIHLPHISPKVITPYDIFCTQLIIKGRH RSKELFSKDPDYIVVPYTKVQFDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTFTLQTAIIFPHMTSTTPLEKGIVIFTDGSANGRSVTYIQGREPIIKE NTQNTAQQAEIVAVITAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPRTKIYTELKHLQRLIHKRQEKFYIGHIRGHTGLPGPLAQGNAYADSLTRILT A
MMTVB_P03365_WS	VQEISDSRPMLHIYLNGRRFLGLLDTGADKTCIAGRDWPANWPIHQTESSLQGLGMACGVARSSQPLRWQHEDKSGIIHPFVIPTLPFTLWGRDIMK DIKVRLMTDSPDDSQDLMIGAIESNLFADQISWKSDQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWRLLQDLR AVNATMHDMGALQPGLPSPVAVPKGWEIIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFKRPYQRFQWKVLPQGMKNSPTLCQKFVDKAILTVRDKY QDSYIVHYMDDILLAHPSRSIVDEILTSMIQALNKHGLVVSTEKIQKYDNLKYLGTHIQGDSVSYQKLQIRTDKLRTLNDFQKLLGNINWIRPFLKLTTG ELKPLFEILNGDSNPISTRKLTPEACKALQLMNERLSTARVKRLDLSQPWSLCILKTEYTPTACLWQDGVVEWIHLPHISPKVITPYDIFCTQLIIKGRH RSKELFSKDPDYIVVPYTKVQFDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTFTLQTAIIFPHMTSTTPLEKGIVIFTDGSANGRSVTYIQGREPIIKE NTQNTAQQAEIVAVITAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPRTKIYTELKHLQRLIHKRQEKFYIGHIRGHTGLPGPLAQGNAYADSLTRILT A
MMTVB_P03365-Pro	GRDIMKDIKVRLMTDSPDDSQDLMIGAIESNLFADQISWKSDQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWR LLQDLRAVNATMHDMGALQPGLPSPVAVPKGWEIIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFKRPYQRFQWKVLPQGMKNSPTLCQKFVDKAIL TVRDKYQDSYIVHYMDDILLAHPSRSIVDEILTSMIQALNKHGLVVSTEKIQKYDNLKYLGTHIQGDSVSYQKLQIRTDKLRTLNDFQKLLGNINWIRPF LKLTTGELKPLFEILNGDSNPISTRKLTPEACKALQLMNERLSTARVKRLDLSQPWSLCILKTEYTPTACLWQDGVVEWIHLPHISPKVITPYDIFCTQL IIKGRHRSKELFSKDPDYIVVPYTKVQFDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTFTLQTAIIFPHMTSTTPLEKGIVIFTDGSANGRSVTYIQGR EPIIKENTQNTAQQAEIVAVITAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPRTKIYTELKHLQRLIHKRQEKFYIGHIRGHTGLPGPLAQGNAYADSL TRILT
MMTVB_P03365-Pro	GRDIMKDIKVRLMTDSPDDSQDLMIGAIESNLFADQISWKSDQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWR LLQDLRAVNATMHDMGALQPGLPSPVAVPKGWEIIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFKRPYQRFQWKVLPQGMKNSPTLCQKFVDKAIL TVRDKYQDSYIVHYMDDILLAHPSRSIVDEILTSMIQALNKHGLVVSTEKIQKYDNLKYLGTHIQGDSVSYQKLQIRTDKLRTLNDFQKLLGNINWIRPF LKLTTGELKPLFEILNGDSNPISTRKLTPEACKALQLMNERLSTARVKRLDLSQPWSLCILKTEYTPTACLWQDGVVEWIHLPHISPKVITPYDIFCTQL IIKGRHRSKELFSKDPDYIVVPYTKVQFDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTFTLQTAIIFPHMTSTTPLEKGIVIFTDGSANGRSVTYIQGR EPIIKENTQNTAQQAEIVAVITAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPRTKIYTELKHLQRLIHKRQEKFYIGHIRGHTGLPGPLAQGNAYADSL TRILT
MMTVB_P03365-Pro_2mut	GRDIMKDIKVRLMTDSPDDSQDLMIGAIESNLFADQISWKSDQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWR LLQDLRAVNATMHDMGALQPGLPSPVAVPKGWEIIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFKRPYQRFQWKVLPQGMKNSPTLCQKFVDKAIL TVRDKYQDSYIVHYMDDILLAHPSRSIVDEILTSMIQALNKHGLVVSTEKIQKYDNLKYLGTHIQGDSVSYQKLQIRTDKLRTLNDFQKLLGNINWIRPF LKLTTGELKPLFEILNPDSNPISTRKLTPEACKALQLMNERLSTARVKRLDLSQPWSLCILKTEYTPTACLWQDGVVEWIHLPHISPKVITPYDIFCTQL IIKGRHRSKELFSKDPDYIVVPYTKVQFDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTFTLQTAIIFPHMTSTTPLEKGIVIFTDGSANGRSVTYIQGR EPIIKENTQNTAQQAEIVAVITAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPRTKIYTELKHLQRLIHKRQEKFYIGHIRGHTGLPGPLAQGNAYADSL TRILT
MMTVB_P03365-Pro_2mut	GRDIMKDIKVRLMTDSPDDSQDLMIGAIESNLFADQISWKSDQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWR LLQDLRAVNATMHDMGALQPGLPSPVAVPKGWEIIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFKRPYQRFQWKVLPQGMKNSPTLCQKFVDKAIL TVRDKYQDSYIVHYMDDILLAHPSRSIVDEILTSMIQALNKHGLVVSTEKIQKYDNLKYLGTHIQGDSVSYQKLQIRTDKLRTLNDFQKLLGNINWIRPF LKLTTGELKPLFEILNPDSNPISTRKLTPEACKALQLMNERLSTARVKRLDLSQPWSLCILKTEYTPTACLWQDGVVEWIHLPHISPKVITPYDIFCTQL IIKGRHRSKELFSKDPDYIVVPYTKVQFDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTFTLQTAIIFPHMTSTTPLEKGIVIFTDGSANGRSVTYIQGR EPIIKENTQNTAQQAEIVAVITAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPRTKIYTELKHLQRLIHKRQEKFYIGHIRGHTGLPGPLAQGNAYADSL TRILT
MMTVB_P03365-Pro_2mutB	GRDIMKDIKVRLMTDSPDDSQDLMIGAIESNLFADQISWKSDQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWR LLQDLRAVNATMHDMGALQPGLPSPVAPPKGWEIIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFKRPYQRFQWKVLPQGMKNSPTLCQKFVDKAIL TVRDKYQDSYIVHYMDDILLAHPSRSIVDEILTSMIQALNKHGLVVSTEKIQKYDNLKYLGTHIQGDSVSYQKLQIRTDKLRTLNDFQKLLGNINWIRPF LKLTTGELKPLFEILNPDSNPISTRKLTPEACKALQLMNERLSTARVKRLDLSQPWSLCILKTEYTPTACLWQDGVVEWIHLPHISPKVITPYDIFCTQL IIKGRHRSKELFSKDPDYIVVPYTKVQFDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTFTLQTAIIFPHMTSTTPLEKGIVIFTDGSANGRSVTYIQGR EPIIKENTQNTAQQAEIVAVITAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPRTKIYTELKHLQRLIHKRQEKFYIGHIRGHTGLPGPLAQGNAYADSL

[0187]

RT 名称	RT 氨基酸序列
	TRILT
MMTVB_P03365-Pro_2mutB	GRDIMKDIKVRLMTDSPDDSQDLMIGAIESNLFADQISWKSDQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWRLLQDLRAVNATMHDMGALQPGLPSPVAPPKGWEIIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFKRPYQRFQWKVLPQGMKNSPTLCQKFVDKAILTVRDKYQDSYIVHYMDDILLAHPSRSIVDEILTSMIQALNKHGLVVSTEKIQKYDNLKYLGTHIQGDSVSYQKLQIRTDKLRTLNDFQKLLGNINWIRPFLKLTTGELKPLFEILNPDSNPISTRKLTPEACKALQLMNERLSTARVKRLDLSQPWSLCILKTEYTPTACLWQDGVVEWIHLPHISPKVITPYDIFCTQLIIKGRHRSKELFSKDPDYIVVPYTKVQFDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTFTLQTAIIFPHMTSTTPLEKGIVIFTDGSANGRSVTYIQGREPIIKENTQNTAQQAEIVAVITAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPRTKIYTELKHLQRLIHKRQEKFYIGHIRGHTGLPGPLAQGNAYADSLTRILT	
MPMV_P07572	LTAAIDILAPQQCAEPITWKSDEPVWVDQWPLTNDKLAAAQQLVQEQLEAGHITESSSPWNTPIFVIKKKSGKWRLLQDLRAVNATMVLMGALQPGLPSPVAIPQGYLKIIIDLKDCFFSIPLHPSDQKRFAFSLPSTNFKEPMQRFQWKVLPQGMANSPTLCQKYVATAIHKVRHAWKQMYIIHYMDDILIAGKDGQQVLQCFDQLKQELTAAGLHIAPEKVQLQDPYTYLGFELNGPKITNQKAVIRKDKLQTLNDFQKLLGDINWLRPYLKLTTGDLKPLFDTLKGDSDPNSHRSLSKEALASLEKVETAIAEQFVTHINYSLPLIFLIFNTALTPTGLFWQDNPIMWIHLPASPKKVLLPYYDAIADLIILGRDHSKKYFGIEPSTIIQPYSKSQIDWLMQNTEMWPIACASFVGILDNHYPPNKLIQFCKLHTFVFPQIISKTPLNNALLVFTDGSSTGMAAYTLTDTTIKFQTNLNSAQLVELQALIAVLSAFPNQPLNIYTDSAYLAHSIPLLETVAQIKHISETAKLFLQCQQLIYNRSIPFYIGHVRAHSGLPGPIAQGNQRADLATKIVASNINT	
MPMV_P07572_2mutB	LTAAIDILAPQQCAEPITWKSDEPVWVDQWPLTNDKLAAAQQLVQEQLEAGHITESSSPWNTPIFVIKKKSGKWRLLQDLRAVNATMVLMGALQPGLPSPVAPPQGYLKIIIDLKDCFFSIPLHPSDQKRFAFSLPSTNFKEPMQRFQWKVLPQGMANSPTLCQKYVATAIHKVRHAWKQMYIIHYMDDILIAGKDGQQVLQCFDQLKQELTAAGLHIAPEKVQLQDPYTYLGFELNGPKITNQKAVIRKDKLQTLNDFQKLLGDINWLRPYLKLTTGDLKPLFDTLKPDSDPNSHRSLSKEALASLEKVETAIAEQFVTHINYSLPLIFLIFNTALTPTGLFWQDNPIMWIHLPASPKKVLLPYYDAIADLIILGRDHSKKYFGIEPSTIIQPYSKSQIDWLMQNTEMWPIACASFVGILDNHYPPNKLIQFCKLHTFVFPQIISKTPLNNALLVFTDGSSTGMAAYTLTDTTIKFQTNLNSAQLVELQALIAVLSAFPNQPLNIYTDSAYLAHSIPLLETVAQIKHISETAKLFLQCQQLIYNRSIPFYIGHVRAHSGLPGPIAQGNQRADLATKIVASNINT	
PERV_Q4VFZ2	TLQLDDEYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFDEALHRDLANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGTAGFCRLWIPGFATLAAPLYPLTKEKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGREIKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLL	
PERV_Q4VFZ2	TLQLDDEYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFDEALHRDLANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGTAGFCRLWIPGFATLAAPLYPLTKEKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGREIKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLL	
PERV_Q4VFZ2_3mut	TLQLDDEYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRDLANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGTAGFCRLWIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGREIKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLL	
PERV_Q4VFZ2_3mut	TLQLDDEYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRDLANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGTAGFCRLWIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGREIKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLL	
PERV_Q4VFZ2_3mutA_WS	LDDEYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRDLANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGKAGFCRLFIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGREIKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLP	
PERV_Q4VFZ2_3mutA_WS	LDDEYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRDLANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGKAGFCRLFIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGREIKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLP	
SFV1_P23074	MDPLQLLQPLEAEIKGTKLKAHWNSGATITCVPEAFLEDERPIQTMLIKTIHGEKQQDVYYLTFKVQGRKVEAEVLASPYDYILLNPSDVPWLMKKPLQLTVLVPLHEYQERLLQQTALPKEQKELLQKLFLKYDALWQHWENQVGHRRIKPHNIATGTLAPRPQKQYPINPKAKPSIQIVIDDLLKQGVLIQQNSTMNTPVYPVPKPDGKWRMVLDYREVNKTIPLIAAQNQHSAGILSSIYRGKYKTTLDLTNGFWAHPITPESYWLTAFTWQGKQYCWTRLPQGFLNSPALFTADVVDLLKEIPNVQAYVDDIYISHDDPQEHLEQLEKIFSILLNAGYVVSLKKSEIAQREVEFLGFNITKEGRGLTDTFKQKLLNITPPKDLKQLQSILGLLNFARNFIPNYSELVKPLYTIVANANGKFISWTEDNSNQLQHIISVLNQADNLEERNPETRLIIKVNSSPSAGYIRYYNEGSKRPIMYVNYIFSKAEAKFTQTEKLLTTMHKGLIKAMDLAMGQEILVYSPIVSMTKIQRTPLPERKALPVRWITWMTYLEDPRIQFHYDKSLPELQQIPNVTEDVIAKTKHPSEFAMVFYTDGSAIKHPDVNKSHSAGMGIAQVQFIPEYKIVHQWSIPLGDHTAQLAEIAAVEFACKKALKISGPVLIVTDSFYVAESANKELPYWKSNGFLNNKKKPLRHVSKWKSIAECLQLKPDIIIMHEKGHQQPMTTLHTEGNNLADKLATQGSYVVH	
SFV1_P23074_2mut	MDPLQLLQPLEAEIKGTKLKAHWNSGATITCVPEAFLEDERPIQTMLIKTIHGEKQQDVYYLTFKVQGRKVEAEVLASPYDYILLNPSDVPWLMKKPLQLTVLVPLHEYQERLLQQTALPKEQKELLQKLFLKYDALWQHWENQVGHRRIKPHNIATGTLAPRPQKQYPINPKAKPSIQIVIDDLLKQGVLIQQNSTMNTPVYPVPKPDGKWRMVLDYREVNKTIPLIAAQNQHSAGILSSIYRGKYKTTLDLTNGFWAHPITPESYWLTAFTWQGKQYCWTRLPQGFLNSPALFTADVVDLLKEIPNVQAYVDDIYISHDDPQEHLEQLEKIFSILLNAGYVVSLKKSEIAQREVEFLGFNITKEGRGLTDTFKQKLLNITPPKDLKQLQSILGLLNFARNFIPNYSELVKPLYTIVAPANGKFISWTEDNSNQLQHIISVLNQADNLEERNPETRLIIKVNSSPSAGYIRYYNEGSKRPIMYVNYIFSKAEAKFTQTEKLLTTMHKGLIKAMDLAMGQEILVYSPIVSMTKIQRTPLPERKALPVRWITWMTYLEDPRIQFHYDKSLPELQQIPNVTEDVIAKTKHPSEFAMVFYTDGSAIKHPDVNKSHSAGMGIAQVQFIPEYKIVHQWSIPLGDHTAQLAEIAAVEFACKKALKISGPVLIVTDSFYVAESANKELPYWKSNGFLNNKKKPLRHVSKWKSIAECLQLKPDIIIMHEKGHQQPMTTLHTEGNNLADKLATQGSYVVH	
SFV1_P23074_2mutA	MDPLQLLQPLEAEIKGTKLKAHWNSGATITCVPEAFLEDERPIQTMLIKTIHGEKQQDVYYLTFKVQGRKVEAEVLASPYDYILLNPSDVPWLMKKPLQLTVLVPLHEYQERLLQQTALPKEQKELLQKLFLKYDALWQHWENQVGHRRIKPHNIATGTLAPRPQKQYPINPKAKPSIQIVIDDLLKQGVLIQQNS	

[0188]

RT 名称	RT 氨基酸序列
	TMNTPVYPVPKPDGKWRMVLDYREVNKTIPLIAAQNQHSAGILSSIYRGKYKTTLDLTNGFWAHPITPESYWLTAFTWQGKQYCWTRLPQGFLNSP ALFNADVVDLLKEIPNVQAYVDDIYISHDDPQEHLEQLEKIFSILLNAGYVVSLKKSEIAQREVEFLGFNITKEGRGLTDTFKQKLLNITPPKDLKQLQSI LGKLNFARNFIPNYSELVKPLYTIVAPANGKFISWTEDNSNQLQHIISVLNQADNLEERNPETRLIIKVNSSPSAGYIRYYNEGSKRPIMYVNYIFSKAE AKFTQTEKLLTTMHKGLIKAMDLAMGQEILVYSPIVSMTKIQRTPLPERKALPVRWITWMTYLEDPRIQFHYDKSLPELQQIPNVTEDVIAKTKHPSEF AMVFYTDGSAIKHPDVNKSHSAGMGIAQVQFIPEYKIVHQWSIPLGDHTAQLAEIAAVEFACKKALKISGPVLIVTDSFYVAESANKELPYWKSNGFL NNKKKPLRHVSKWKSIAECLQLKPDIIIMHEKGHQQPMTTLHTEGNNLADKLATQGSYVVH
SFV1_P2307 4-Pro	VPWLMKKPLQLTVLVPLHEYQERLLQQTALPKEQKELLQKLFLKYDALWQHWENQVGHRRIKPHNIATGTLAPRPQKQYPINPKAKPSIQIVIDDLLK QGVLIQQNSTMNTPVYPVPKPDGKWRMVLDYREVNKTIPLIAAQNQHSAGILSSIYRGKYKTTLDLTNGFWAHPITPESYWLTAFTWQGKQYCWTR LPQGFLNSPALFTADVVDLLKEIPNVQAYVDDIYISHDDPQEHLEQLEKIFSILLNAGYVVSLKKSEIAQREVEFLGFNITKEGRGLTDTFKQKLLNITPP KDLKQLQSILGLLNFARNFIPNYSELVKPLYTIVANANGKFISWTEDNSNQLQHIISVLNQADNLEERNPETRLIIKVNSSPSAGYIRYYNEGSKRPIMY VNYIFSKAEAKFTQTEKLLTTMHKGLIKAMDLAMGQEILVYSPIVSMTKIQRTPLPERKALPVRWITWMTYLEDPRIQFHYDKSLPELQQIPNVTEDVIA KTKHPSEFAMVFYTDGSAIKHPDVNKSHSAGMGIAQVQFIPEYKIVHQWSIPLGDHTAQLAEIAAVEFACKKALKISGPVLIVTDSFYVAESANKELPY WKSNGFLNNKKKPLRHVSKWKSIAECLQLKPDIIIMHEKGHQQPMTTLHTEGNNLADKLATQGSYVVH
SFV1_P2307 4-Pro_2mut	VPWLMKKPLQLTVLVPLHEYQERLLQQTALPKEQKELLQKLFLKYDALWQHWENQVGHRRIKPHNIATGTLAPRPQKQYPINPKAKPSIQIVIDDLLK QGVLIQQNSTMNTPVYPVPKPDGKWRMVLDYREVNKTIPLIAAQNQHSAGILSSIYRGKYKTTLDLTNGFWAHPITPESYWLTAFTWQGKQYCWTR LPQGFLNSPALFNADVVDLLKEIPNVQAYVDDIYISHDDPQEHLEQLEKIFSILLNAGYVVSLKKSEIAQREVEFLGFNITKEGRGLTDTFKQKLLNITPP KDLKQLQSILGLLNFARNFIPNYSELVKPLYTIVAPANGKFISWTEDNSNQLQHIISVLNQADNLEERNPETRLIIKVNSSPSAGYIRYYNEGSKRPIMY VNYIFSKAEAKFTQTEKLLTTMHKGLIKAMDLAMGQEILVYSPIVSMTKIQRTPLPERKALPVRWITWMTYLEDPRIQFHYDKSLPELQQIPNVTEDVIA KTKHPSEFAMVFYTDGSAIKHPDVNKSHSAGMGIAQVQFIPEYKIVHQWSIPLGDHTAQLAEIAAVEFACKKALKISGPVLIVTDSFYVAESANKELPY WKSNGFLNNKKKPLRHVSKWKSIAECLQLKPDIIIMHEKGHQQPMTTLHTEGNNLADKLATQGSYVVH
SFV1_P2307 4-Pro_2mutA	VPWLMKKPLQLTVLVPLHEYQERLLQQTALPKEQKELLQKLFLKYDALWQHWENQVGHRRIKPHNIATGTLAPRPQKQYPINPKAKPSIQIVIDDLLK QGVLIQQNSTMNTPVYPVPKPDGKWRMVLDYREVNKTIPLIAAQNQHSAGILSSIYRGKYKTTLDLTNGFWAHPITPESYWLTAFTWQGKQYCWTR LPQGFLNSPALFNADVVDLLKEIPNVQAYVDDIYISHDDPQEHLEQLEKIFSILLNAGYVVSLKKSEIAQREVEFLGFNITKEGRGLTDTFKQKLLNITPP KDLKQLQSILGKLNFARNFIPNYSELVKPLYTIVAPANGKFISWTEDNSNQLQHIISVLNQADNLEERNPETRLIIKVNSSPSAGYIRYYNEGSKRPIMY VNYIFSKAEAKFTQTEKLLTTMHKGLIKAMDLAMGQEILVYSPIVSMTKIQRTPLPERKALPVRWITWMTYLEDPRIQFHYDKSLPELQQIPNVTEDVIA KTKHPSEFAMVFYTDGSAIKHPDVNKSHSAGMGIAQVQFIPEYKIVHQWSIPLGDHTAQLAEIAAVEFACKKALKISGPVLIVTDSFYVAESANKELPY WKSNGFLNNKKKPLRHVSKWKSIAECLQLKPDIIIMHEKGHQQPMTTLHTEGNNLADKLATQGSYVVH
SFV3L_P274 01	MDPLQLLQPLEAEIKGTKLKAHWNSGATITCVPQAFLEEEVPIKNIWIKTIHGEKEQPVYYLTFKIQGRKVEAEVISSPYDYILVSPSDIPWLMKKPLQL TTLVPLQEYEERLLKQTMLTGSYKEKLQSLFLKYDALWQHWENQVGHRRIKPHHIATGTVNPRPQKQYPINPKAKASIQTVINDLLKQGVLIQQNSIM NTPVYPVPKPDGKWRMVLDYREVNKTIPLIAAQNQHSAGILSSIFRGKYKTTLDLSNGFWAHSITPESYWLTAFTWLGQQYCWTRLPQGFLNSPAL FTADVVDLLKEVPNVQVYVDDIYISHDDPREHLEQLEKVFSLLLNAGYVVSLKKSEIAQHEVEFLGFNITKEGRGLTETFKQKLLNITPPRDLKQLQSIL GLLNFARNFIPNFSELVKPLYNIIATANGKYITWTTDNSQQLQNIISMLNSAENLEERNPEVRLIMKVNTSPSAGYIRFYNEFAKRPIMYLNYVYTKAEV KFTNTEKLLTTIHKGLIKALDLGMGQEILVYSPIVSMTKIQKTPLPERKALPIRWITWMSYLEDPRIQFHYDKTLPELQQVPTVTDDIIAKIKHPSEFSMV FYTDGSAIKHPNVNKSHNAGMGIAQVQFKPEFTVINTWSIPLGDHTAQLAEVAAVEFACKKALKIDGPVLIVTDSFYVAESVNKELPYWQSNGFFNN KKKPLKHVSKWKSIADCIQLKPDIIIIHEKGHQPTASTFHTEGNNLADKLATQGSYVVN
SFV3L_P274 01_2mut	MDPLQLLQPLEAEIKGTKLKAHWNSGATITCVPQAFLEEEVPIKNIWIKTIHGEKEQPVYYLTFKIQGRKVEAEVISSPYDYILVSPSDIPWLMKKPLQL TTLVPLQEYEERLLKQTMLTGSYKEKLQSLFLKYDALWQHWENQVGHRRIKPHHIATGTVNPRPQKQYPINPKAKASIQTVINDLLKQGVLIQQNSIM NTPVYPVPKPDGKWRMVLDYREVNKTIPLIAAQNQHSAGILSSIFRGKYKTTLDLSNGFWAHSITPESYWLTAFTWLGQQYCWTRLPQGFLNSPAL FNADVVDLLKEVPNVQVYVDDIYISHDDPREHLEQLEKVFSLLLNAGYVVSLKKSEIAQHEVEFLGFNITKEGRGLTETFKQKLLNITPPRDLKQLQSIL GLLNFARNFIPNFSELVKPLYNIIATAPGKYITWTTDNSQQLQNIISMLNSAENLEERNPEVRLIMKVNTSPSAGYIRFYNEFAKRPIMYLNYVYTKAEV KFTNTEKLLTTIHKGLIKALDLGMGQEILVYSPIVSMTKIQKTPLPERKALPIRWITWMSYLEDPRIQFHYDKTLPELQQVPTVTDDIIAKIKHPSEFSMV FYTDGSAIKHPNVNKSHNAGMGIAQVQFKPEFTVINTWSIPLGDHTAQLAEVAAVEFACKKALKIDGPVLIVTDSFYVAESVNKELPYWQSNGFFNN KKKPLKHVSKWKSIADCIQLKPDIIIIHEKGHQPTASTFHTEGNNLADKLATQGSYVVN
SFV3L_P274 01_2mutA	MDPLQLLQPLEAEIKGTKLKAHWNSGATITCVPQAFLEEEVPIKNIWIKTIHGEKEQPVYYLTFKIQGRKVEAEVISSPYDYILVSPSDIPWLMKKPLQL TTLVPLQEYEERLLKQTMLTGSYKEKLQSLFLKYDALWQHWENQVGHRRIKPHHIATGTVNPRPQKQYPINPKAKASIQTVINDLLKQGVLIQQNSIM NTPVYPVPKPDGKWRMVLDYREVNKTIPLIAAQNQHSAGILSSIFRGKYKTTLDLSNGFWAHSITPESYWLTAFTWLGQQYCWTRLPQGFLNSPAL FNADVVDLLKEVPNVQVYVDDIYISHDDPREHLEQLEKVFSLLLNAGYVVSLKKSEIAQHEVEFLGFNITKEGRGLTETFKQKLLNITPPRDLKQLQSIL GKLNFARNFIPNFSELVKPLYNIIATAPGKYITWTTDNSQQLQNIISMLNSAENLEERNPEVRLIMKVNTSPSAGYIRFYNEFAKRPIMYLNYVYTKAEV KFTNTEKLLTTIHKGLIKALDLGMGQEILVYSPIVSMTKIQKTPLPERKALPIRWITWMSYLEDPRIQFHYDKTLPELQQVPTVTDDIIAKIKHPSEFSMV FYTDGSAIKHPNVNKSHNAGMGIAQVQFKPEFTVINTWSIPLGDHTAQLAEVAAVEFACKKALKIDGPVLIVTDSFYVAESVNKELPYWQSNGFFNN KKKPLKHVSKWKSIADCIQLKPDIIIIHEKGHQPTASTFHTEGNNLADKLATQGSYVVN
SFV3L_P274 01-Pro	IPWLMKKPLQLTTLVPLQEYEERLLKQTMLTGSYKEKLQSLFLKYDALWQHWENQVGHRRIKPHHIATGTVNPRPQKQYPINPKAKASIQTVINDLLK QGVLIQQNSIMNTPVYPVPKPDGKWRMVLDYREVNKTIPLIAAQNQHSAGILSSIFRGKYKTTLDLSNGFWAHSITPESYWLTAFTWLGQQYCWTRL PQGFLNSPALFNADVVDLLKEVPNVQVYVDDIYISHDDPREHLEQLEKVFSLLLNAGYVVSLKKSEIAQHEVEFLGFNITKEGRGLTETFKQKLLNITP PRDLKQLQSILGLLNFARNFIPNFSELVKPLYNIIATANGKYITWTTDNSQQLQNIISMLNSAENLEERNPEVRLIMKVNTSPSAGYIRFYNEFAKRPIM YLNYVYTKAEVKFTNTEKLLTTIHKGLIKALDLGMGQEILVYSPIVSMTKIQKTPLPERKALPIRWITWMSYLEDPRIQFHYDKTLPELQQVPTVTDDIIA KIKHPSEFSMVFYTDGSAIKHPNVNKSHNAGMGIAQVQFKPEFTVINTWSIPLGDHTAQLAEVAAVEFACKKALKIDGPVLIVTDSFYVAESVNKELP YWQSNGFFNNKKKPLKHVSKWKSIADCIQLKPDIIIIHEKGHQPTASTFHTEGNNLADKLATQGSYVVN
SFV3L_P274 01-Pro_2mut	IPWLMKKPLQLTTLVPLQEYEERLLKQTMLTGSYKEKLQSLFLKYDALWQHWENQVGHRRIKPHHIATGTVNPRPQKQYPINPKAKASIQTVINDLLK QGVLIQQNSIMNTPVYPVPKPDGKWRMVLDYREVNKTIPLIAAQNQHSAGILSSIFRGKYKTTLDLSNGFWAHSITPESYWLTAFTWLGQQYCWTRL PQGFLNSPALFNADVVDLLKEVPNVQVYVDDIYISHDDPREHLEQLEKVFSLLLNAGYVVSLKKSEIAQHEVEFLGFNITKEGRGLTETFKQKLLNITP PRDLKQLQSILGLLNFARNFIPNFSELVKPLYNIIATAPGKYITWTTDNSQQLQNIISMLNSAENLEERNPEVRLIMKVNTSPSAGYIRFYNEFAKRPIM YLNYVYTKAEVKFTNTEKLLTTIHKGLIKALDLGMGQEILVYSPIVSMTKIQKTPLPERKALPIRWITWMSYLEDPRIQFHYDKTLPELQQVPTVTDDIIA KIKHPSEFSMVFYTDGSAIKHPNVNKSHNAGMGIAQVQFKPEFTVINTWSIPLGDHTAQLAEVAAVEFACKKALKIDGPVLIVTDSFYVAESVNKELP YWQSNGFFNNKKKPLKHVSKWKSIADCIQLKPDIIIIHEKGHQPTASTFHTEGNNLADKLATQGSYVVN
SFV3L_P274 01-Pro_2mut A	IPWLMKKPLQLTTLVPLQEYEERLLKQTMLTGSYKEKLQSLFLKYDALWQHWENQVGHRRIKPHHIATGTVNPRPQKQYPINPKAKASIQTVINDLLK QGVLIQQNSIMNTPVYPVPKPDGKWRMVLDYREVNKTIPLIAAQNQHSAGILSSIFRGKYKTTLDLSNGFWAHSITPESYWLTAFTWLGQQYCWTRL PQGFLNSPALFNADVVDLLKEVPNVQVYVDDIYISHDDPREHLEQLEKVFSLLLNAGYVVSLKKSEIAQHEVEFLGFNITKEGRGLTETFKQKLLNITP PRDLKQLQSILGKLNFARNFIPNFSELVKPLYNIIATAPGKYITWTTDNSQQLQNIISMLNSAENLEERNPEVRLIMKVNTSPSAGYIRFYNEFAKRPIM YLNYVYTKAEVKFTNTEKLLTTIHKGLIKALDLGMGQEILVYSPIVSMTKIQKTPLPERKALPIRWITWMSYLEDPRIQFHYDKTLPELQQVPTVTDDIIA KIKHPSEFSMVFYTDGSAIKHPNVNKSHNAGMGIAQVQFKPEFTVINTWSIPLGDHTAQLAEVAAVEFACKKALKIDGPVLIVTDSFYVAESVNKELP YWQSNGFFNNKKKPLKHVSKWKSIADCIQLKPDIIIIHEKGHQPTASTFHTEGNNLADKLATQGSYVVN
SFVCP_Q87 040	MNPLQLLQPLPAEVKGTKLLAHWNSGATITCIPESFLEDEQPIKQTLIKTIHGEKQQNVYYLTFKVKGRKVEAEVIASPYEYILLSPTDVPWLTQQPLQ LTILVPLQEYQDRILNKTALPEEQKQQLKALFTKYDNLWQHWENQVGHRKIRPHNIATGDYPPRPQKQYPINPKAKPSIQIVIDDLLKQGVLTPQNST MNTPVYPVPKPDGRWRMVLDYREVNKTIPLTAAQNQHSAGILATIVRQKYKTTLDLANGFWAHPITPDSYWLTAFTWQGKQYCWTRLPQGFLNSP ALFTADAVDLLKEVPNVQVYVDDIYLSHDNPHEHIQQLEKVFQILLQAGYVVSLKKSEIGQRTVEFLGFNITKEGRGLTDTFKTKLLNVTPPKDLKQLQ SILGLLNFARNFIPNFAELVQTLYNLIASSKGKYIEWTEDNTKQLNKVIEALNTASNLEERLPDQRLVIKVNTSPSAGYVRYYNESGKKPIMYLNYVFSK AELKFSMLEKLLTTMHKALIKAMDLAMGQEILVYSPIVSMTKIQKTPLPERKALPIRWITWMTYLEDPRIQFHYDKTLPELKHIPDVYTSSIPPLKHPSQ YEGVFCTDGSAIKSPDPTKSNNAGMGIVHAIYNPEYKILNQWSIPLGHHTAQMAEIAAVEFACKKALKVPGPVLVITDSFYVAESANKELPYWKSNGF VNNKKEPLKHISKWKSIAECLSIKPDITIQHEKGHQPINTSIHTEGNALADKLATQGSYVVN

[0189]

RT 名称	RT 氨基酸序列
SFVCP_Q87040_2mut	MNPLQLLQPLPAEVKGTKLLAHWNSGATITCIPESFLEDEQPIKQTLIKTIHGEKQQNVYYLTFKVKGRKVEAEVIASPYEYILLSPTDVPWLTQQPLQ LTILVPLQEYQDRILNKTALPEEQKQQLKALFTKYDNLWQHWENQVGHRKIRPHNIATGDYPPRPQKQYPINPKAKPSIQIVIDDLLKQGVLTPQNST MNTPVYPVPKPDGRWRMVLDYREVNKTIPLTAAQNQHSAGILATIVRQKYKTTLDLANGFWAHPITPDSYWLTAFTWQGKQYCWTRLPQGFLNSP ALFNADAVDLLKEVPNVQVYVDDIYLSHDNPHEHIQQLEKVFQILLQAGYVVSLKKSEIGQRTVEFLGFNITKEGRGLTDTFKTKLLNVTPPKDLKQLQ SILGLLNFARNFIPNFAELVQTLYNLIASSPGKYIEWTEDNTKQLNKVIEALNTASNLEERLPDQRLVIKVNTSPSAGYVRYYNESGKKPIMYLNYVFSK AELKFSMLEKLLTTMHKALIKAMDLAMGQEILVYSPIVSMTKIQKTPLPERKALPIRWITWMTYLEDPRIQFHYDKTLPELKHIPDVYTSSIPPLKHPSQ YEGVFCTDGSAIKSPDPTKSNNAGMGIVHAIYNPEYKILNQWSIPLGHHTAQMAEIAAVEFACKKALKVPGPVLVITDSFYVAESANKELPYWKSNGF VNNKKEPLKHISKWKSIAECLSIKPDITIQHEKGHQPINTSIHTEGNALADKLATQGSYVVN
SFVCP_Q87040_2mutA	MNPLQLLQPLPAEVKGTKLLAHWNSGATITCIPESFLEDEQPIKQTLIKTIHGEKQQNVYYLTFKVKGRKVEAEVIASPYEYILLSPTDVPWLTQQPLQ LTILVPLQEYQDRILNKTALPEEQKQQLKALFTKYDNLWQHWENQVGHRKIRPHNIATGDYPPRPQKQYPINPKAKPSIQIVIDDLLKQGVLTPQNST MNTPVYPVPKPDGRWRMVLDYREVNKTIPLTAAQNQHSAGILATIVRQKYKTTLDLANGFWAHPITPDSYWLTAFTWQGKQYCWTRLPQGFLNSP ALFNADAVDLLKEVPNVQVYVDDIYLSHDNPHEHIQQLEKVFQILLQAGYVVSLKKSEIGQRTVEFLGFNITKEGRGLTDTFKTKLLNVTPPKDLKQLQ SILGKLNFARNFIPNFAELVQTLYNLIASSPGKYIEWTEDNTKQLNKVIEALNTASNLEERLPDQRLVIKVNTSPSAGYVRYYNESGKKPIMYLNYVFS KAELKFSMLEKLLTTMHKALIKAMDLAMGQEILVYSPIVSMTKIQKTPLPERKALPIRWITWMTYLEDPRIQFHYDKTLPELKHIPDVYTSSIPPLKHPS QYEGVFCTDGSAIKSPDPTKSNNAGMGIVHAIYNPEYKILNQWSIPLGHHTAQMAEIAAVEFACKKALKVPGPVLVITDSFYVAESANKELPYWKSN GFVNNKKEPLKHISKWKSIAECLSIKPDITIQHEKGHQPINTSIHTEGNALADKLATQGSYVVN
SFVCP_Q87040-Pro	VPWLTQQPLQLTILVPLQEYQDRILNKTALPEEQKQQLKALFTKYDNLWQHWENQVGHRKIRPHNIATGDYPPRPQKQYPINPKAKPSIQIVIDDLLK QGVLTPQNSTMNTPVYPVPKPDGRWRMVLDYREVNKTIPLTAAQNQHSAGILATIVRQKYKTTLDLANGFWAHPITPDSYWLTAFTWQGKQYCWT RLPQGFLNSPALFTADAVDLLKEVPNVQVYVDDIYLSHDNPHEHIQQLEKVFQILLQAGYVVSLKKSEIGQRTVEFLGFNITKEGRGLTDTFKTKLLNV TPPKDLKQLQSILGLLNFARNFIPNFAELVQTLYNLIASSKGKYIEWTEDNTKQLNKVIEALNTASNLEERLPDQRLVIKVNTSPSAGYVRYYNESGKK PIMYLNYVFSKAELKFSMLEKLLTTMHKALIKAMDLAMGQEILVYSPIVSMTKIQKTPLPERKALPIRWITWMTYLEDPRIQFHYDKTLPELKHIPDVYT SSIPPLKHPSQYEGVFCTDGSAIKSPDPTKSNNAGMGIVHAIYNPEYKILNQWSIPLGHHTAQMAEIAAVEFACKKALKVPGPVLVITDSFYVAESANK ELPYWKSNGFVNNKKEPLKHISKWKSIAECLSIKPDITIQHEKGHQPINTSIHTEGNALADKLATQGSYVVN
SFVCP_Q87040-Pro_2mut	VPWLTQQPLQLTILVPLQEYQDRILNKTALPEEQKQQLKALFTKYDNLWQHWENQVGHRKIRPHNIATGDYPPRPQKQYPINPKAKPSIQIVIDDLLK QGVLTPQNSTMNTPVYPVPKPDGRWRMVLDYREVNKTIPLTAAQNQHSAGILATIVRQKYKTTLDLANGFWAHPITPDSYWLTAFTWQGKQYCWT RLPQGFLNSPALFNADAVDLLKEVPNVQVYVDDIYLSHDNPHEHIQQLEKVFQILLQAGYVVSLKKSEIGQRTVEFLGFNITKEGRGLTDTFKTKLLNV TPPKDLKQLQSILGLLNFARNFIPNFAELVQTLYNLIASSPGKYIEWTEDNTKQLNKVIEALNTASNLEERLPDQRLVIKVNTSPSAGYVRYYNESGKK PIMYLNYVFSKAELKFSMLEKLLTTMHKALIKAMDLAMGQEILVYSPIVSMTKIQKTPLPERKALPIRWITWMTYLEDPRIQFHYDKTLPELKHIPDVYT SSIPPLKHPSQYEGVFCTDGSAIKSPDPTKSNNAGMGIVHAIYNPEYKILNQWSIPLGHHTAQMAEIAAVEFACKKALKVPGPVLVITDSFYVAESANK ELPYWKSNGFVNNKKEPLKHISKWKSIAECLSIKPDITIQHEKGHQPINTSIHTEGNALADKLATQGSYVVN
SFVCP_Q87040-Pro_2mutA	VPWLTQQPLQLTILVPLQEYQDRILNKTALPEEQKQQLKALFTKYDNLWQHWENQVGHRKIRPHNIATGDYPPRPQKQYPINPKAKPSIQIVIDDLLK QGVLTPQNSTMNTPVYPVPKPDGRWRMVLDYREVNKTIPLTAAQNQHSAGILATIVRQKYKTTLDLANGFWAHPITPDSYWLTAFTWQGKQYCWT RLPQGFLNSPALFNADAVDLLKEVPNVQVYVDDIYLSHDNPHEHIQQLEKVFQILLQAGYVVSLKKSEIGQRTVEFLGFNITKEGRGLTDTFKTKLLNV TPPKDLKQLQSILGKLNFARNFIPNFAELVQTLYNLIASSPGKYIEWTEDNTKQLNKVIEALNTASNLEERLPDQRLVIKVNTSPSAGYVRYYNESGKK PIMYLNYVFSKAELKFSMLEKLLTTMHKALIKAMDLAMGQEILVYSPIVSMTKIQKTPLPERKALPIRWITWMTYLEDPRIQFHYDKTLPELKHIPDVYT SSIPPLKHPSQYEGVFCTDGSAIKSPDPTKSNNAGMGIVHAIYNPEYKILNQWSIPLGHHTAQMAEIAAVEFACKKALKVPGPVLVITDSFYVAESANK ELPYWKSNGFVNNKKEPLKHISKWKSIAECLSIKPDITIQHEKGHQPINTSIHTEGNALADKLATQGSYVVN
SMRVH_P03364	PRSRAIDIPVPHADKISWKITDPVWVDQWPLTYEKTLAAIALVQEQLAAGHIEPTNSPWNTPIFIIKKKSGSWRLLQDLRAVNKVMVPMGALQPGLPS PVAIPLNYHKIVIDLKDCFFTIPLHPEDRPYFAFSVPQINFQSPMPRYQWKVLPQGMANSPTLCQKFVAAAIAPVRSQWPEAYILHYMDDILLACDSA EAAKACYAHIISCLTSYGLKIAPDKVQVSEPFSYLGFELHHQQVFTPRVCLKTDHLKTLNDFQKLLGDIQWLRPYLKLPTSALVPLNNILKGDPNPLSV RALTPEAKQSLALINKAIQNQSVQQISYNLPLVLLLLPTPHTPTAVFWQPNGTDPTKNGSPLLWLHLPASPSKVLLTYPSLLAMLIIKGRYTGRQLFGR DPHSIIIPYTQDQLTWLLQTSDEWAIALSSFTGDIDNHYPSDPVIQFAKLHQFIFPKITKCAPIPQATLVFTDGSSNGIAAYVIDNQPISIKSPYLSAQLVE LYAILQVFTVLAHQPFNLYTDSAYIAQSVPLLETVPFIKSSTNATPLFSKLQQLILNRQHPFFIGHLRAHLNLPGPLAEGNALADAATQIFPIISD
SMRVH_P03364_2mut	PRSRAIDIPVPHADKISWKITDPVWVDQWPLTYEKTLAAIALVQEQLAAGHIEPTNSPWNTPIFIIKKKSGSWRLLQDLRAVNKVMVPMGALQPGLPS PVAIPLNYHKIVIDLKDCFFTIPLHPEDRPYFAFSVPQINFQSPMPRYQWKVLPQGMANSPTLCQKFVAAAIAPVRSQWPEAYILHYMDDILLACDSA EAAKACYAHIISCLTSYGLKIAPDKVQVSEPFSYLGFELHHQQVFTPRVCLKTDHLKTLNDFQKLLGDIQWLRPYLKLPTSALVPLNNILKPDPNPLSV RALTPEAKQSLALINKAIQNQSVQQISYNLPLVLLLLPTPHTPTAVFWQPNGTDPTKNGSPLLWLHLPASPSKVLLTYPSLLAMLIIKGRYTGRQLFGR DPHSIIIPYTQDQLTWLLQTSDEWAIALSSFTGDIDNHYPSDPVIQFAKLHQFIFPKITKCAPIPQATLVFTDGSSNGIAAYVIDNQPISIKSPYLSAQLVE LYAILQVFTVLAHQPFNLYTDSAYIAQSVPLLETVPFIKSSTNATPLFSKLQQLILNRQHPFFIGHLRAHLNLPGPLAEGNALADAATQIFPIISD
SMRVH_P03364_2mutB	PRSRAIDIPVPHADKISWKITDPVWVDQWPLTYEKTLAAIALVQEQLAAGHIEPTNSPWNTPIFIIKKKSGSWRLLQDLRAVNKVMVPMGALQPGLPS PVAPPLNYHKIVIDLKDCFFTIPLHPEDRPYFAFSVPQINFQSPMPRYQWKVLPQGMANSPTLCQKFVAAAIAPVRSQWPEAYILHYMDDILLACDSA EAAKACYAHIISCLTSYGLKIAPDKVQVSEPFSYLGFELHHQQVFTPRVCLKTDHLKTLNDFQKLLGDIQWLRPYLKLPTSALVPLNNILKPDPNPLSV RALTPEAKQSLALINKAIQNQSVQQISYNLPLVLLLLPTPHTPTAVFWQPNGTDPTKNGSPLLWLHLPASPSKVLLTYPSLLAMLIIKGRYTGRQLFGR DPHSIIIPYTQDQLTWLLQTSDEWAIALSSFTGDIDNHYPSDPVIQFAKLHQFIFPKITKCAPIPQATLVFTDGSSNGIAAYVIDNQPISIKSPYLSAQLVE LYAILQVFTVLAHQPFNLYTDSAYIAQSVPLLETVPFIKSSTNATPLFSKLQQLILNRQHPFFIGHLRAHLNLPGPLAEGNALADAATQIFPIISD
SRV2_P51517	LATAVDILAPQRYADPITWKSDEPVWVDQWPLTQEKLAAAQQLVQEQLQAGHIIESNSPWNTPIFVIKKKSGKWRLLQDLRAVNATMVLMGALQPG LPSPVAIPQGYFKIVIDLKDCFFTIPLQPVDQKRFAFSLPSTNFKQPMKRYQWKVLPQGMANSPTLCQKYVAAAIEPVRKSWAQMYIIHYMDDILIAG KLGEQVLQCFAQLKQALTTTGLQIAPEKVQLQDPYTYLGFQINGPKITNQKAVIRRDKLQTLNDFQKLLGDINWLRPYLHLTTGDLKPLFDILKGDSNP NSPRSLSEAALASLQKVETAIAEQFVTQIDYTQPLTFLIFNTTLTPTGLFWQNNPVMWVHLPASPKKVLLPYYDAIADLIILGRDNSKKYFGLEPSTIIQ PYSKSQIHWLMQNTETWPIACASYAGNIDNHYPPNKLIQFCKLHAVVFPRIISKTPLDNALLVFTDGSSTGIAAYTFEKTTVRFKTSHTSAQLVELQALI AVLSAFPHRALNVYTDSAYLAHSIPLLETVSHIKHISDTAKFFLQCQQLIYNRSIPFYLGHIRAHSGLPGPLSQGNHITDLATKVVATTLTT
SRV2_P51517_2mutB	LATAVDILAPQRYADPITWKSDEPVWVDQWPLTQEKLAAAQQLVQEQLQAGHIIESNSPWNTPIFVIKKKSGKWRLLQDLRAVNATMVLMGALQPG LPSPVAPPQGYFKIVIDLKDCFFTIPLQPVDQKRFAFSLPSTNFKQPMKRYQWKVLPQGMANSPTLCQKYVAAAIEPVRKSWAQMYIIHYMDDILIAG KLGEQVLQCFAQLKQALTTTGLQIAPEKVQLQDPYTYLGFQINGPKITNQKAVIRRDKLQTLNDFQKLLGDINWLRPYLHLTTGDLKPLFDILKGDSNP NSPRSLSEAALASLQKVETAIAEQFVTQIDYTQPLTFLIFNTTLTPTGLFWQNNPVMWVHLPASPKKVLLPYYDAIADLIILGRDNSKKYFGLEPSTIIQ PYSKSQIHWLMQNTETWPIACASYAGNIDNHYPPNKLIQFCKLHAVVFPRIISKTPLDNALLVFTDGSSTGIAAYTFEKTTVRFKTSHTSAQLVELQALI AVLSAFPHRALNVYTDSAYLAHSIPLLETVSHIKHISDTAKFFLQCQQLIYNRSIPFYLGHIRAHSGLPGPLSQGNHITDLATKVVATTLTT
WDSV_O92815	SCQTKNTLNIDEYLLQFPDQLWASLPTDIGRMLVPPITIKIKDNASLPSIRQYPLPKDKTEGLRPLISSLENQGILIKCHSPCNTPIFPIKKAGRDEYRMIH DLRAINNIVAPLTAVVASPTTVLSNLAPSLHWFTVIDLSNAFFSVPIHKDSQYLFAFTFEGHQYTWTVLPQGFIHSPTLFSQALYQSLHKIKFKISSEICI YMDDVLIASKDRDTNLKDTAVMLQHLASEGHKVSKKKLQLCQQEVVYLGQLLTPEGRKILPDRKVTVSQFQQPTTIRQIRAFLGLVGYCRHWIPEFSI HSKFLEKQLKKDTAEPFQLDDQQVEAFNKLKHAITTAPVLVVPDPAKPFQLYTSHSEHASIAVLTQKHAGRTRPIAFLSSKFDAIESGLPPCLKACASI HRSLTQADSFILGAPLIIYTTHAICTLLQRDRSQLVTASRFSKWEADLLRPELTFVACSAVSPAHLYMQSCENNIPPHDCVLLTHTISRPRPDLSDLPIP DPDMTLFSDGSYTTGRGGAAVVMHRPVTDDFIIIHQQPGGASAQTAELLALAAACHLATDKTVNIYTDSRYAYGVVHDFGHLWMHRGFVTSAGTPI KNHKEIEYLLKQIMKPKQVSVIKIEAHTKGVSMEVRGNAAADEAAKNAVFLVQR
WDSV_O92815_2mut	SCQTKNTLNIDEYLLQFPDQLWASLPTDIGRMLVPPITIKIKDNASLPSIRQYPLPKDKTEGLRPLISSLENQGILIKCHSPCNTPIFPIKKAGRDEYRMIH DLRAINNIVAPLTAVVASPTTVLSNLAPSLHWFTVIDLSNAFFSVPIHKDSQYLFAFTFEGHQYTWTVLPQGFIHSPTLFNQALYQSLHKIKFKISSEICI YMDDVLIASKDRDTNLKDTAVMLQHLASEGHKVSKKKLQLCQQEVVYLGQLLTPEGRKILPDRKVTVSQFQQPTTIRQIRAFLGLVGYCRHWIPEFSI HSKFLEKQLKPDTAEPFQLDDQQVEAFNKLKHAITTAPVLVVPDPAKPFQLYTSHSEHASIAVLTQKHAGRTRPIAFLSSKFDAIESGLPPCLKACASI HRSLTQADSFILGAPLIIYTTHAICTLLQRDRSQLVTASRFSKWEADLLRPELTFVACSAVSPAHLYMQSCENNIPPHDCVLLTHTISRPRPDLSDLPIP DPDMTLFSDGSYTTGRGGAAVVMHRPVTDDFIIIHQQPGGASAQTAELLALAAACHLATDKTVNIYTDSRYAYGVVHDFGHLWMHRGFVTSAGTPI

[0190]

RT 名称	RT 氨基酸序列
	KNHKEIEYLLKQIMKPKQVSVIKIEAHTKGVSMEVRGNAAADEAAKNAVFLVQR
WDSV_O92815_2mutA	SCQTKNTLNIDEYLLQFPDQLWASLPTDIGRMLVPPITIKIKDNASLPSIRQYPLPKDKTEGLRPLISSLENQGILIKCHSPCNTPIFPIKKAGRDEYRMIH DLRAINNIVAPLTAVVASPTTVLSNLAPSLHWFTVIDLSNAFFSVPIHKDSQYLFAFTFEGHQYTWTVLPQGFIHSPTLFNQALYQSLHKIKFKISSEICI YMDDVLIASKDRDTNLKDTAVMLQHLASEGHKVSKKKLQLCQQEVVYLGQLLTPEGRKILPDRKVTVSQFQQPTTIRQIRAFLGKVGYCRHFIPEFSI HSKFLEKQLKPDTAEPFQLDDQQVEAFNKLKHAITTAPVLVVPDPAKPFQLYTSHSEHASIAVLTQKHAGRTRPIAFLSSKFDAIESGLPPCLKACASI HRSLTQADSFILGAPLIIYTTHAICTLLQRDRSQLVTASRFSKWEADLLRPELTFVACSAVSPAHLYMQSCENNIPPHDCVLLTHTISRPRPDLSDLPIP DPDMTLFSDGSYTTGRGGAAVVMHRPVTDDFIIIHQQPGGASAQTAELLALAAACHLATDKTVNIYTDSRYAYGVVHDFGHLWMHRGFVTSAGTPI KNHKEIEYLLKQIMKPKQVSVIKIEAHTKGVSMEVRGNAAADEAAKNAVFLVQR
WMSV_P03359	VLNLEEEYRLHEKPVPSSIDPSWLQLFPTVWAERAGMGLANQVPPVVVELRSGASPVAVRQYPMSKEAREGIRPHIQRFLDLGVLVPCQSPWNTP LLPVKKPGTNDYRPVQDLREINKRVQDIHPTVPNPYNLLSSLPPSHTWYSVLDLKDAFFCLKLHPNSQPLFAFEWRDPEKGNTGQLTWTRLPQGFK NSPTLFDEALHRDLAPFRALNPQVVLLQYVDDLLVAAPTYRDCKEGTQKLLQELSKLGYRVSAKKAQLCQKEVTYLGYLLKEGKRWLTPARKATVM KIPPPTTPRQVREFLGTAGFCRLWIPGFASLAAPLYPLTKESIPFIWTEEHQKAFDRIKEALLSAPALALPDLTKPFTLYVDERAGVARGVLTQTLGPW RRPVAYLSKKLDPVASGWPTCLKAVAAVALLLKDADKLTLGQNVTVIASHSLESIVRQPPDRWMTNARMTHYQSLLLNERVSFAPPAVLNPATLLPV ESEATPVHRCSEILAEETGTRRDLKDQPLPGVPAWYTDGSSFIAEGKRRAGAAIVDGKRTVWASSLPEGTSAQKAELVALTQALRLAEGKDINIYTD SRYAFATAHIHGAIYKQRGLLTSAGKDIKNKEEILALLEAIHLPKRVAIIHCPGHQKGNDPVATGNRRADEAAKQAALSTRVLAETTKP
WMSV_P03359_3mut	VLNLEEEYRLHEKPVPSSIDPSWLQLFPTVWAERAGMGLANQVPPVVVELRSGASPVAVRQYPMSKEAREGIRPHIQRFLDLGVLVPCQSPWNTP LLPVKKPGTNDYRPVQDLREINKRVQDIHPTVPNPYNLLSSLPPSHTWYSVLDLKDAFFCLKLHPNSQPLFAFEWRDPEKGNTGQLTWTRLPQGFK NSPTLFNEALHRDLAPFRALNPQVVLLQYVDDLLVAAPTYRDCKEGTQKLLQELSKLGYRVSAKKAQLCQKEVTYLGYLLKEGKRWLTPARKATVM KIPPPTTPRQVREFLGTAGFCRLWIPGFASLAAPLYPLTKPSIPFIWTEEHQKAFDRIKEALLSAPALALPDLTKPFTLYVDERAGVARGVLTQTLGPW RRPVAYLSKKLDPVASGWPTCLKAVAAVALLLKDADKLTLGQNVTVIASHSLESIVRQPPDRWMTNARMTHYQSLLLNERVSFAPPAVLNPATLLPV ESEATPVHRCSEILAEETGTRRDLKDQPLPGVPAWYTDGSSFIAEGKRRAGAAIVDGKRTVWASSLPEGTSAQKAELVALTQALRLAEGKDINIYTD SRYAFATAHIHGAIYKQRGWLTSAGKDIKNKEEILALLEAIHLPKRVAIIHCPGHQKGNDPVATGNRRADEAAKQAALSTRVLAETTKP
WMSV_P03359_3mutA	VLNLEEEYRLHEKPVPSSIDPSWLQLFPTVWAERAGMGLANQVPPVVVELRSGASPVAVRQYPMSKEAREGIRPHIQRFLDLGVLVPCQSPWNTP LLPVKKPGTNDYRPVQDLREINKRVQDIHPTVPNPYNLLSSLPPSHTWYSVLDLKDAFFCLKLHPNSQPLFAFEWRDPEKGNTGQLTWTRLPQGFK NSPTLFNEALHRDLAPFRALNPQVVLLQYVDDLLVAAPTYRDCKEGTQKLLQELSKLGYRVSAKKAQLCQKEVTYLGYLLKEGKRWLTPARKATVM KIPPPTTPRQVREFLGKAGFCRLFIPGFASLAAPLYPLTKPSIPFIWTEEHQKAFDRIKEALLSAPALALPDLTKPFELFVDEKQGYAKGVLTQKL GPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQAMLLDTDRVQFGPVVALN PATLLPLPEKEAPHDCLEILAETHGTRPDLTDQPIPDADYTWYTDGSSFLQEGQRRAGAAVTTETEVIWARALPAGTSAQRAELIALTQALKMAEGK KLNVYTDSRYAFATAHVHGEIYRRRGLLTSEGREIKNKNEILALLKALFLPKRLSIIHCPGHQKGNSAEARGNRMADQAARAAMKAVLETSTLL
XMRV6_A1Z651	TLNIEDEYRLHETSKEPDVPLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPL LPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFK NSPTLFDEALHRDLADFRIQHPDLILLQYVDDLLLAATSEQDCQRGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVM GQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKL GPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQAMLLDTDRVQFGPVVALN PATLLPLPEKEAPHDCLEILAETHGTRPDLTDQPIPDADYTWYTDGSSFLQEGQRRAGAAVTTETEVIWARALPAGTSAQRAELIALTQALKMAEGK KLNVYTDSRYAFATAHVHGEIYRRRGLLTSEGREIKNKNEILALLKALFLPKRLSIIHCPGHQKGNSAEARGNRMADQAARAAMKAVLETSTLL
XMRV6_A1Z651_3mut	TLNIEDEYRLHETSKEPDVPLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPL LPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFK NSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSEQDCQRGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVM GQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKL GPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQAMLLDTDRVQFGPVVALN PATLLPLPEKEAPHDCLEILAETHGTRPDLTDQPIPDADYTWYTDGSSFLQEGQRRAGAAVTTETEVIWARALPAGTSAQRAELIALTQALKMAEGK KLNVYTDSRYAFATAHVHGEIYRRRGWLTSEGREIKNKNEILALLKALFLPKRLSIIHCPGHQKGNSAEARGNRMADQAARAAMKAVLETSTLL
XMRV6_A1Z651_3mutA	TLNIEDEYRLHETSKEPDVPLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPL LPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFK NSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSEQDCQRGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVM GQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKL GPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQAMLLDTDRVQFGPVVALN PATLLPLPEKEAPHDCLEILAETHGTRPDLTDQPIPDADYTWYTDGSSFLQEGQRRAGAAVTTETEVIWARALPAGTSAQRAELIALTQALKMAEGK KLNVYTDSRYAFATAHVHGEIYRRRGWLTSEGREIKNKNEILALLKALFLPKRLSIIHCPGHQKGNSAEARGNRMADQAARAAMKAVLETSTLL

[0191] 在一些实施例中,逆转录酶结构域被修饰,例如通过位点特异性突变。在一些实施例中,将逆转录酶结构域工程改造以具有改善的特性,例如源自MMLV RT的SuperScript IV(SSIV)逆转录酶。在一些实施例中,可以将逆转录酶结构域工程改造以具有较低的错误率,例如,如WO 2001068895(通过援引并入本文)中所述。在一些实施例中,可以将逆转录酶结构域工程改造以更耐热。在一些实施例中,可以将逆转录酶结构域工程改造以更具持续合成能力。在一些实施例中,可以将逆转录酶结构域工程改造以对抑制剂具有耐受性。在一些实施例中,可以将逆转录酶结构域工程改造为更快。在一些实施例中,可以将逆转录酶结构域工程改造以更好地耐受RNA模板中的经修饰的核苷酸。在一些实施例中,可以将逆转录酶结构域工程改造以插入经修饰的DNA核苷酸。在一些实施例中,将逆转录酶结构域工程改造以结合模板RNA。在一些实施例中,一种或多种突变选自鼠白血病病毒逆转录酶RT结构域中的D200N、L603W、T330P、D524G、E562Q、D583N、P51L、S67R、E67K、T197A、H204R、E302K、F309N、W313F、L435G、N454K、H594Q、L671P、E69K或D653N或另一个RT结构域的对应位置的相应突变。

[0192] 在一些实施例中,基因修饰多肽包含来自逆转录病毒逆转录酶的RT结构域,例如野生型M-MLV RT,例如,包含以下序列:

[0193] M-MLV(WT)：

[0194] TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFDEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLI(SEQ ID NO:4012)

[0195] 在一些实施例中,基因修饰多肽包含来自逆转录病毒逆转录酶的RT结构域,例如M-MLV RT,例如,包含以下序列：

[0196] TLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFDEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLL(SEQ ID NO:4013)

[0197] 在一些实施例中,基因修饰多肽包含来自逆转录病毒逆转录酶的RT结构域,该逆转录酶包含NP_057933的氨基酸659-1329的序列。在实施例中,基因修饰多肽在NP_057933的氨基酸659-1329的序列的N末端进一步包含一个另外的氨基酸,例如,如下所示：

[0198] TLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIP LKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLP**VK KPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVL DLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSP TLFDEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRAL LQTLGNLGYRASAKKAQICQKQVKYLGY**LLKEGQRWLTEARKETVM GQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPD QQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWR RPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHA VEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEE GLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGA AVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAF ATAHIHGEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQK GHSAEARGNRMADQAARKAA（SEQ ID NO: 4014）

[0199] **核心 RT**（粗体），按上述注释

[0200] RNA酶H（下划线），按上述注释

[0201] 在实施例中，基因修饰多肽在NP_057933的氨基酸659-1329序列的C末端进一步包含一个另外的氨基酸。在实施例中，基因修饰多肽包含RNA酶H1结构域（例如，NP_057933的氨基酸1178-1318）。

[0202] 在一些实施例中，逆转录病毒逆转录酶结构域，例如M-MLV RT，可以包含野生型序列的一个或多个突变，其可以改善RT的特征，例如热稳定性、持续合成能力和/或模板结合。在一些实施例中，M-MLV RT结构域相对于上述M-MLV(WT)序列包含一个或多个突变，例如选自D200N、L603W、T330P、T306K、W313F、D524G、E562Q、D583N、P51L、S67R、E67K、T197A、H204R、E302K、F309N、L435G、N454K、H594Q、D653N、R110S、K103L，例如突变的组合，例如D200N、L603W，和T330P，任选地进一步包括T306K和W313F。在一些实施例中，本文使用的M-MLV RT包含突变D200N、L603W、T330P、T306K和W313F。在实施例中，突变M-MLV RT包含以下氨基酸序列：

[0203] M-MLV(PE2)：

[0204] TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQ EARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQ WYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILL QYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKT PRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDE KQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPP DRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTD GSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRR

RGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLI(SEQ ID NO:4015)

[0205] 在一些实施例中,书写结构域(例如,RT结构域)包含RNA结合结构域,例如,其特异性结合RNA序列。在一些实施例中,模板RNA包含由书写结构域的RNA结合结构域特异性结合的RNA序列。

[0206] 在一些实施例中,逆转录结构域仅识别和逆转录特定模板,例如系统的模板RNA。在一些实施例中,模板包含能够被逆转录结构域识别和逆转录的序列或结构。在一些实施例中,模板包含能够与本文所述基因组工程系统的多肽组分的RNA结合结构域关联的序列或结构。在一些实施例中,基因组工程系统优选地逆转录包含关联序列的模板,而非缺少关联序列的模板。

[0207] 书写结构域还可包含DNA依赖性DNA聚合酶活性,例如,包含能够将DNA从模板DNA序列书写入基因组的酶活性。在一些实施例中,采用DNA依赖性DNA聚合来完成靶位点编辑的第二链合成。在一些实施例中,DNA依赖性DNA聚合酶活性由多肽中的DNA聚合酶结构域提供。在一些实施例中,DNA依赖性DNA聚合酶活性由逆转录酶结构域提供,该逆转录酶结构域也能够进行DNA依赖性DNA聚合,例如第二链合成。在一些实施例中,DNA依赖性DNA聚合酶活性由系统中的第二多肽提供。在一些实施例中,DNA依赖性DNA聚合酶活性由内源宿主细胞聚合酶提供,该聚合酶任选地由基因组工程系统的组分募集到靶位点。

[0208] 在一些实施例中,相对于参考逆转录酶结构域,逆转录酶结构域在体外具有较低的过早终止率概率(P_{off})。在一些实施例中,参考逆转录酶结构域是病毒逆转录酶结构域,例如来自M-MLV的RT结构域。

[0209] 在一些实施例中,逆转录酶结构域具有低于约5x 10^{-3}/nt、5x 10^{-4}/nt或5x 10^{-6}/nt的体外过早终止率(P_{off})的较低概率,例如如在1094nt RNA上测量。在实施例中,体外过早终止率如Bibillo和Eickbush(2002)J Biol Chem[生物化学杂志]277(38):34836-34845(其通过援引以其全文并入本文)中所述确定。

[0210] 在一些实施例中,逆转录酶结构域能够在细胞中完成至少约30%或50%的整合。完全整合的百分比可以通过将基本上全长整合事件(例如,包含至少98%的预期整合序列的基因组位点)的数量除以细胞群体中总(包括基本上全长和部分)整合事件的数量来测量。在实施例中,使用长读段扩增子测序确定细胞中的整合(例如,跨整合位点),例如,如Karst等人(2020)bioRxiv doi.org/10.1101/645903(其通过援引以其全文并入本文)中所述。

[0211] 在实施例中,定量细胞中的整合包括计数包含至少约75%、80%、85%、90%、95%、96%、97%、98%、99%或100%的对应于模板RNA(例如长度至少为0.05、0.1、0.5、0.6、0.7、0.8、0.9、1、1.5、2、3、4或5kb的模板RNA,例如长度在0.5-0.6、0.6-0.7、0.7-0.8、0.8-0.9、1.0-1.2、1.2-1.4、1.4-1.6、1.6-1.8、1.8-2.0、2-3、3-4或4-5kb)的DNA序列的整合部分。

[0212] 在一些实施例中,逆转录酶结构域能够在体外聚合dNTP。在实施例中,逆转录酶结构域能够以0.1-50nt/sec(例如0.1-1、1-10或10-50nt/sec)的速率在体外聚合dNTP。在实施例中,通过单分子测定法测量逆转录酶结构域对dNTP的聚合,例如,如Schwartz和Quake(2009)PNAS[美国国家科学院院刊]106(48):20294-20299(其通过援引以其全文并入)中所

述。

[0213] 在一些实施例中,逆转录酶结构域的体外错误率(例如,核苷酸的错误掺入)是1x 10^{-3}-1x 10^{-4}或1x 10^{-4}-1x 10^{-5}个取代/nt,例如,如Yasukawa等人(2017)Biochem Biophys Res Commun[生物化学与生物物理研究通讯]492(2):147-153(其通过援引以其全文并入本文)中所述。在一些实施例中,逆转录酶结构域在细胞(例如,HEK293T细胞)中具有的错误率(例如,核苷酸的错误掺入)是1x 10^{-3}-1x 10^{-4}或1x 10^{-4}-1x 10^{-5}个取代/nt,例如,通过长读段扩增子测序,例如,如Karst等人(2020)bioRxiv doi.org/10.1101/645903(其通过援引以其全文并入本文)中所述。

[0214] 在一些实施例中,逆转录酶结构域能够在体外进行靶RNA的逆转录。在一些实施例中,逆转录酶需要至少3个核苷酸的引物来启动模板的逆转录。在一些实施例中,通过检测来自靶RNA的cDNA来确定靶RNA的逆转录(例如,当提供有ssDNA引物时,例如,其与靶在3'端退火至少3、4、5、6、7、8、9或10nt),例如,如Bibillo和Eickbush(2002)J Biol Chem[生物化学杂志]277(38):34836-34845(其通过援引以其全文并入本文)中所述。

[0215] 在一些实施例中,与缺乏蛋白质结合基序(例如,3'UTR)的RNA模板相比,例如当将其RNA模板转化为cDNA时,逆转录酶结构域执行逆转录的效率至少高5或10倍(例如,通过cDNA产生)。在实施例中,逆转录效率如Yasukawa等人(2017)Biochem Biophys Res Commun[生物化学与生物物理研究通讯]492(2):147-153(其通过援引以其全文并入本文)中所述测量。

[0216] 在一些实施例中,逆转录酶结构域以比任何内源细胞RNA(例如,当在细胞(例如,HEK293T细胞)中表达时)更高的频率(例如,高约5或10倍的频率)特异性结合特定的RNA模板。在实施例中,通过CLIP-seq测量逆转录酶结构域和模板RNA之间的特异性结合频率,例如,如Lin和Miles(2019)Nucleic Acids Res[核酸研究]47(11):5490-5501(其通过援引以其全文并入本文)中所述。

[0217] 模板核酸结合结构域

[0218] 基因修饰多肽通常包含能够与模板核酸(例如,模板RNA)相关联的区域。在一些实施例中,模板核酸结合结构域是RNA结合结构域。在一些实施例中,RNA结合结构域是可与含有特定特征(例如结构基序)的RNA分子相关联的模块化结构域。在其他实施例中,模板核酸结合结构域(例如,RNA结合结构域)包含在逆转录结构域内,例如,逆转录酶来源的组分具有已知的RNA偏好特征。

[0219] 在其他实施例中,模板核酸结合结构域(例如,RNA结合结构域)包含在靶DNA结合结构域内。例如,在一些实施例中,DNA结合结构域是识别包含gRNA的模板核酸(例如,模板RNA)的结构的CRISPR相关蛋白。在一些实施例中,基因修饰多肽包含DNA结合结构域,其包含与gRNA支架相关联的CRISPR相关蛋白,该支架允许DNA结合结构域结合靶基因组DNA序列。在一些实施例中,gRNA支架和gRNA间隔子包含在模板核酸(例如,模板RNA)内,因此DNA结合结构域也是模板核酸结合结构域。在一些实施例中,多肽在多个结构域中具有RNA结合功能,例如,可以结合CRISPR相关DNA结合结构域中的gRNA结构和逆转录酶结构域中另外的序列或结构。

[0220] 在一些实施例中,RNA结合结构域能够以比参考RNA结合结构域更大的亲和力结合模板RNA。在一些实施例中,参考RNA结合结构域是来自化脓性链球菌(S.pyogenes)的Cas9

的RNA结合结构域。在一些实施例中,RNA结合结构域能够以100pM-10nM(例如,100pM-1nM或1nM-10nM)的亲和力结合模板RNA。在一些实施例中,RNA结合结构域对其模板RNA的亲和力在体外测量,例如通过热泳,例如,如Asmari等人Methods[方法]146:107-119(2018)(将其通过援引以其全文并入本文)中所述。在一些实施例中,RNA结合结构域对其模板RNA的亲和力在细胞中测量(例如,通过FRET或CLIP-Seq)。

[0221] 在一些实施例中,RNA结合结构域与模板RNA在体外以比乱序RNA高至少约5倍或10倍的频率相关联。在一些实施例中,RNA结合结构域与模板RNA或乱序RNA之间的结合频率通过CLIP-seq测量,例如,如Lin和Miles(2019)Nucleic Acids Res[核酸研究]47(11):5490-5501(其通过援引以其全文并入本文)中所述。在一些实施例中,RNA结合结构域与模板RNA在细胞(例如,HEK293T细胞)中以比乱序RNA高至少约5倍或10倍的频率相关联。在一些实施例中,RNA结合结构域与模板RNA或乱序RNA之间的关联频率通过CLIP-seq测量,例如,如Lin和Miles(2019)同上中所述。

[0222] 核酸内切酶结构域和DNA结合结构域

[0223] 在一些实施例中,基因修饰多肽具有通过核酸内切酶结构域切割DNA靶位点的功能。在一些实施例中,基因修饰多肽包含DNA结合结构域,例如用于结合靶核酸。在一些实施例中,基因修饰多肽的结构域(例如,Cas结构域)包含两个或更多个较小的结构域(例如,DNA结合结构域和核酸内切酶结构域)。应当理解,当DNA结合结构域(例如,Cas结构域)被描述为与靶核酸序列结合时,在一些实施例中,该结合是由gRNA介导的。

[0224] 在一些实施例中,结构域具有两种功能。例如,在一些实施例中,核酸内切酶结构域也是DNA结合结构域。在一些实施例中,核酸内切酶结构域也是模板核酸(例如,模板RNA)结合结构域。例如,在一些实施例中,多肽包含CRISPR相关的核酸内切酶结构域,其结合包含gRNA的模板RNA,结合靶DNA序列(例如,与gRNA的一部分互补),并切割靶DNA序列。在一些实施例中,来自异源的核酸内切酶结构域或核酸内切酶/DNA结合结构域可被用于或可在本文所述的基因修饰系统中被修饰(例如,通过插入、缺失或取代一个或多个残基)。

[0225] 在一些实施例中,编码核酸内切酶结构域或核酸内切酶/DNA结合结构域的核酸被从其天然序列改变为具有改变的密码子使用,例如,针对人细胞进行改善。在一些实施例中,核酸内切酶元件是异源核酸内切酶元件,例如Cas核酸内切酶(例如,Cas9)、II型限制性核酸内切酶(例如,Fok1)、大范围核酸酶(例如,I-SceI)或其他核酸内切酶结构域。

[0226] 在某些方面,选择、设计或构建本文所述的基因修饰多肽的DNA结合结构域以结合所期望的宿主DNA靶序列。在某些实施例中,多肽的DNA结合结构域是异源DNA结合元件。在一些实施例中,异源DNA结合元件是锌指元件或TAL效应子元件,例如锌指或TAL多肽或其功能片段。在一些实施例中,异源DNA结合元件是序列指导的DNA结合元件,例如Cas9、Cpf1或其他已被改变为不具有核酸内切酶活性的CRISPR相关蛋白。在一些实施例中,异源DNA结合元件保留核酸内切酶活性。在一些实施例中,异源DNA结合元件保留部分核酸内切酶活性以切割ssDNA,例如,具有切口酶活性。在特定实施例中,异源DNA结合结构域可以是Cas9、TAL结构域、ZF结构域、Myb结构域、其组合或其倍数中的任何一个或多个。

[0227] 在一些实施例中,例如通过位点特异性突变、增加或减少DNA结合元件(例如锌指的数量和/或特异性)等来修饰DNA结合结构域,以改变DNA结合特异性和亲和力。在一些实施例中,编码DNA结合结构域的核酸序列从其天然序列改变为具有改变的密码子使用,例

如,针对人细胞进行改善。在实施例中,该DNA结合结构域相对于野生型DNA结合结构域包含一个或多个修饰、例如经由定向进化(例如,噬菌体辅助的连续进化(PACE))的修饰。

[0228] 在一些实施例中,DNA结合结构域包含大范围核酸酶结构域(例如,如本文所述,例如,在核酸内切酶结构域部分中),或其功能片段。在一些实施例中,大范围核酸酶结构域具有核酸内切酶活性、例如双链切割和/或切口酶活性。在其他实施例中,大范围核酸酶结构域具有降低的活性,例如,缺乏核酸内切酶活性,例如,该大范围核酸酶无催化活性。在一些实施例中,无催化活性的大范围核酸酶用作DNA结合结构域,例如,如Fonfara等人Nucleic Acids Res[核酸研究]40(2):847-860(2012)中所述,该文献通过援引以其全文并入本文。

[0229] 在一些实施例中,基因修饰多肽包含对DNA结合结构域的修饰,例如,相对于野生型多肽。在一些实施例中,DNA结合结构域包含对原始DNA结合结构域的氨基酸序列的添加、缺失、替换或修饰。在一些实施例中,DNA结合结构域被修饰以包括特异性结合目的靶核酸(例如DNA)序列的异源功能结构域。在一些实施例中,功能结构域替换多肽的先前DNA结合结构域的至少一部分(例如,全部)。在一些实施例中,功能结构域包含锌指(例如,特异性结合目的靶核酸(例如,DNA)序列的锌指)。在一些实施例中,功能结构域包含Cas结构域(例如,特异性结合目的靶核酸(例如,DNA)序列的Cas结构域。在一些实施例中,Cas结构域包含Cas9或其突变体或变体(例如,如本文所述)。在实施例中,Cas结构域与指导RNA(gRNA)相关联,例如,如本文所述。在实施例中,Cas结构域被gRNA导向目的靶核酸(例如,DNA)序列。在实施例中,Cas结构域与gRNA在相同的核酸(例如,RNA)分子中编码。在实施例中,Cas结构域与gRNA在不同的核酸(例如,RNA)分子中编码。

[0230] 在一些实施例中,DNA结合结构域能够以比参考DNA结合结构域更大的亲和力结合靶序列(例如,dsDNA靶序列)。在一些实施例中,参考DNA结合结构域是来自化脓性链球菌的Cas9的DNA结合结构域。在一些实施例中,DNA结合结构域能够以100pM-10nM(例如,100pM-1nM或1nM-10nM之间)之间的亲和力结合靶序列(例如,dsDNA靶序列)。

[0231] 在一些实施例中,DNA结合结构域对其靶序列(例如,dsDNA靶序列)的亲和力在体外测量,例如,通过热泳,例如,如Asmari等人Methods[方法]146:107-119(2018)(通过援引以其全文并入本文)中所述。

[0232] 在实施例中,在例如约100倍摩尔过量的摩尔过量的乱序序列竞争者dsDNA存在的情况下,DNA结合结构域能够例如以100pM-10nM(例如,100pM-1nM或1nM-10nM之间)之间的亲和力结合其靶序列(例如,dsDNA靶序列)。

[0233] 在一些实施例中,发现DNA结合结构域与其靶序列(例如,dsDNA靶序列)相关联的频率高于靶细胞(例如,人靶细胞)基因组中的任何其他序列,例如,如通过ChIP-seq测量的(例如,在HEK293T细胞中),例如,如He和Pu(2010)Curr.Protoc Mol Biol[分子生物学最新方案]第21章(其通过援引以其全文并入本文)中所述。在一些实施例中,发现DNA结合结构域与其靶序列(例如,dsDNA靶序列)以比靶细胞的基因组中任何其他序列更频繁至少约5倍或10倍的频率相关联,例如,如通过ChIP-seq(例如,在HEK293T细胞中)测量的,例如,如He和Pu(2010),同上中所述。

[0234] 在一些实施例中,核酸内切酶结构域具有切口酶活性并且切割靶DNA的一条链。在一些实施例中,切口酶活性减少了靶位点处双链断裂的形成。在一些实施例中,核酸内切酶结构域在靶DNA的第一链和第二链中产生交错的切口结构。在一些实施例中,交错的切口结

构在靶位点产生游离3′突出端。在一些实施例中,靶位点处的游离3′突出端提高编辑效率,例如,通过增强模板核酸的3′同源区的访问和退火。在一些实施例中,交错的切口结构减少了靶位点处双链断裂的形成。

[0235] 在一些实施例中,核酸内切酶结构域切割靶DNA的两条链,例如导致靶的平端切割,并且在切割位点的两侧没有ssDNA突出端。本文所述的基因修饰系统的核酸内切酶结构域的氨基酸序列可以与本文所述的核酸内切酶结构域的氨基酸序列至少约50%、至少约60%、至少约70%、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%相同。

[0236] 在某些实施例中,异源核酸内切酶是Fok1或其功能片段。在某些实施例中,异源核酸内切酶是霍利迪(Holliday)连接解离酶或其同源物,例如来自硫磺矿硫化叶菌(Sulfolobus solfataricus)-Ssol Hje的霍利迪连接解离酶(Govindaraju等人,Nucleic Acids Research[核酸研究]44:7,2016)。在某些实施例中,异源核酸内切酶是剪接体蛋白诸如Prp8的大片段的核酸内切酶(Mahbub等人,Mobile DNA[移动DNA]8:16,2017)。在某些实施例中,异源核酸内切酶源自CRISPR相关蛋白,例如Cas9。在某些实施例中,异源核酸内切酶被工程改造为仅具有ssDNA切割活性,例如仅具有切口酶活性,例如是Cas9切口酶,例如具有D10A、H840A或N863A突变的SpCas9。表3提供了与切口酶活性相关的示例性Cas蛋白和突变。在又其他实施例中,同源核酸内切酶结构域被修饰,例如通过位点特异性突变,以改变DNA核酸内切酶活性。在又其他实施例中,对核酸内切酶结构域进行修饰以降低DNA序列特异性,例如通过截短以去除赋予DNA序列特异性的结构域或通过突变以灭活赋予DNA序列特异性的区域。

[0237] 在一些实施例中,核酸内切酶结构域具有切口酶活性并且不形成双链断裂。在一些实施例中,核酸内切酶结构域以比双链断裂更高的频率形成单链断裂,例如,至少90%、95%、96%、97%、98%、或99%的断裂是单链断裂,或少于10%、5%、4%、3%、2%、或1%的断裂是双链断裂。在一些实施例中,核酸内切酶基本上不形成双链断裂。在一些实施例中,核酸内切酶不形成可检测水平的双链断裂。

[0238] 在一些实施例中,核酸内切酶结构域具有对第一链的靶位点DNA进行切口的切口酶活性;例如,在一些实施例中,核酸内切酶切割基因组DNA的靶位点,该靶位点在将被书写结构域延伸的链上的改变位点附近。在一些实施例中,核酸内切酶结构域具有对第一链的靶位点DNA进行切口并且不对第二链的靶位点DNA进行切口的切口酶活性。例如,当多肽包含具有切口酶活性的CRISPR相关核酸内切酶结构域时,在一些实施例中,所述CRISPR相关核酸内切酶结构域对含有PAM位点的靶位点DNA链进行切口(例如,并且不对不含有PAM位点的靶位点DNA链进行切口)。作为另一个实例,当多肽包含具有切口酶活性的CRISPR相关核酸内切酶结构域时,在一些实施例中,所述CRISPR相关核酸内切酶结构域对不含有PAM位点的靶位点DNA链进行切口(例如,并且不对包含PAM位点的靶位点DNA链进行切口)。

[0239] 在一些其他实施例中,核酸内切酶结构域具有切口酶活性,其对第一链和第二链的靶位点DNA进行切口。不希望受理论束缚,在本文所述的多肽的书写结构域(例如,RT结构域)从模板核酸(例如,模板RNA)的异源对象序列聚合(例如,逆转录)之后,细胞DNA修复机制必须修复第一DNA链上的切口。靶位点DNA现在包含两个不同的第一DNA链序列:一个对应于原始基因组DNA(例如,具有游离5′端),并且第二个对应于从异源对象序列聚合而来的那

个(例如,具有游离3′端)。认为这两个不同的序列相互平衡,第一个与第二链杂交,然后另一个,并且细胞DNA修复装置掺入其修复的靶位点中的序列可以是随机过程。不希望受理论束缚,认为向第二链引入另外的切口可能使细胞DNA修复机制偏向于比原始基因组序列更频繁地采用基于异源对象序列的序列(Anzalone等人Nature[自然]576:149-157(2019))。在一些实施例中,另外的切口位于靶位点修饰(例如,插入、缺失或取代)或第一链上的切口的5′或3′的至少10、15、20、25、30、35、40、45、50、55、60、65、70、75、80、85、90、95、100、105、110、115、120、125、130、135、140、145、或150个核苷酸处。

[0240] 可替代地或另外地,不希望受理论束缚,认为第二链的另外切口可促进第二链合成。在一些实施例中,当基因修饰系统已插入或取代了第一链的一部分时,需要合成对应于第二链中的插入/取代的新序列。

[0241] 在一些实施例中,多肽包含具有核酸内切酶活性的单个结构域(例如,单个核酸内切酶结构域)并且所述结构域对第一链和第二链进行切口。例如,在这样的实施例中,核酸内切酶结构域可以是CRISPR相关核酸内切酶结构域,并且模板核酸(例如,模板RNA)包含指导对第一链进行切口的gRNA间隔子和指导对第二链进行切口的另外gRNA间隔子。在一些实施例中,多肽包含多个具有核酸内切酶活性的结构域,并且第一核酸内切酶结构域对第一链进行切口并且第二核酸内切酶结构域对第二链进行切口(任选地,第一核酸内切酶结构域不(例如,不能)对第二链进行切口,并且第二核酸内切酶结构域不(例如,不能)对第一链进行切口)。

[0242] 在一些实施例中,核酸内切酶结构域能够对第一链和第二链进行切口。在一些实施例中,第一和第二链切口出现在靶位点中的相同位置但在相对的链上。在一些实施例中,第二链切口出现在第一切口的交错位置,例如上游或下游。在一些实施例中,如果第二链切口在第一链切口的上游,则核酸内切酶结构域产生靶位点缺失。在一些实施例中,如果第二链切口在第一链切口的下游,则核酸内切酶结构域产生靶位点重复。在一些实施例中,如果第一和第二链的切口出现在靶位点的相同位置,则核酸内切酶结构域不会产生重复和/或缺失。在一些实施例中,核酸内切酶结构域具有改变的活性,这取决于蛋白质构象或RNA结合状态,例如,这促进第一或第二链的切口(例如,如Christensen等人PNAS[美国国家科学院院刊]2006中所述;其通过援引以其全文并入本文)。

[0243] 在一些实施例中,核酸内切酶结构域包含大范围核酸酶或其功能片段。在一些实施例中,核酸内切酶结构域包含归巢核酸内切酶或其功能片段。在一些实施例中,核酸内切酶结构域包含来自LAGLIDADG、GIY-YIG、HNH、His-Cys盒或PD-(D/E)XK家族的大范围核酸酶,或其功能片段或变体,例如,这些功能片段或变体具有例如如家族名称所示的保守氨基酸基序。在一些实施例中,核酸内切酶结构域包含大范围核酸酶或其片段,其选自例如I-SmaMI(Uniprot F7WD42)、I-SceI(Uniprot P03882)、I-AniI(Uniprot P03880)、I-DmoI(Uniprot P21505)、I-CreI(Uniprot P05725)、I-TevI(Uniprot P13299)、I-OnuI(Uniprot Q4VWW5)、或I-BmoI(Uniprot Q9ANR6)。在一些实施例中,大范围核酸酶呈其功能形式时是天然单体,例如I-SceI、I-TevI,或二聚体,例如I-CreI。例如,具有单个LAGLIDADG基序拷贝的LAGLIDADG大范围核酸酶通常形成同二聚体,而具有两个LAGLIDADG基序拷贝的成员通常作为单体被发现。在一些实施例中,通常以二聚体形式形成的大范围核酸酶被表达为融合体,例如,两个亚基作为单个ORF表达并且任选地通过接头连接,例如I-CreI二聚体融合体

(Rodriguez-Fornes等人Gene Therapy[基因疗法]2020;通过援引以其全文并入本文)。在一些实施例中,改变大范围核酸酶或其功能片段以有利于双链DNA分子的一条链的切口酶活性,例如I-SceI(K122I和/或K223I)(Niu等人J Mol Biol[分子生物学杂志]2008)、I-AniI(K227M)(McConnell Smith等人PNAS[美国国家科学院院刊]2009)、I-DmoI(Q42A和/或K120M)(Molina等人J Biol Chem[生物化学杂志]2015)。在一些实施例中,具有这种对单链切割的偏好的大范围核酸酶或其功能片段被用作核酸内切酶结构域,例如,具有切口酶活性。在一些实施例中,核酸内切酶结构域包含大范围核酸酶或其功能片段,其天然靶向或工程改造以靶向安全港位点,例如靶向SH6位点的I-CreI(Rodriguez-Fornes等人,同上)。在一些实施例中,核酸内切酶结构域包含大范围核酸酶或其功能片段,其具有序列耐受催化结构域,例如,识别最小基序CNNNG的I-TevI(Kleinstiver等人PNAS[美国国家科学院院刊]2012)。在一些实施例中,将靶序列耐受性催化结构域融合至DNA结合结构域,例如以指导活性,例如通过将I-TevI融合至:(i)锌指以产生Tev-ZFE(Kleinstiver等人PNAS[美国国家科学院院刊]2012),(ii)其他大范围核酸酶以产生MegaTevs(Wolfs等人Nucleic Acids Res[核酸研究]2014),和/或(iii)Cas9以产生TevCas9(Wolfs等人PNAS[美国国家科学院院刊]2016)。

[0244] 在一些实施例中,核酸内切酶结构域包含限制酶,例如,IIS型或IIP型限制酶。在一些实施例中,核酸内切酶结构域包含IIS型限制酶,例如FokI,或其片段或变体。在一些实施例中,核酸内切酶结构域包含IIP型限制酶,例如PvuII,或其片段或变体。在一些实施例中,二聚体限制酶表达为融合体,从而其作为单链发挥作用,例如,FokI二聚融合体(Minczuk等人Nucleic Acids Res[核酸研究]36(12):3926-3938(2008))。

[0245] 例如,在Guha和Edgell Int J Mol Sci[国际分子科学杂志]18(22):2565(2017)中描述了另外的核酸内切酶结构域的使用,该文献通过援引以其全文并入本文。

[0246] 在一些实施例中,基因修饰多肽包含对核酸内切酶结构域的修饰,例如,相对于野生型Cas蛋白。在一些实施例中,核酸内切酶结构域包含对野生型Cas蛋白的氨基酸序列的添加、缺失、替换或修饰。在一些实施例中,核酸内切酶结构域被修饰以包括异源功能结构域,其特异性结合和/或诱导目的靶核酸(例如,DNA)序列的核酸内切酶切割。在一些实施例中,核酸内切酶结构域包含锌指。在实施例中,包含Cas结构域的核酸内切酶结构域与例如如本文所述的指导RNA(gRNA)相关联。在一些实施例中,核酸内切酶结构域被修饰以包括不靶向特定靶核酸(例如,DNA)序列的功能结构域。在实施例中,核酸内切酶结构域包含Fok1结构域。

[0247] 在一些实施例中,核酸内切酶结构域与靶dsDNA在体外以比乱序dsDNA高至少约5倍或10倍的频率相关联。在一些实施例中,核酸内切酶结构域与靶dsDNA在体外以比乱序dsDNA高至少约5倍或10倍的频率相关联,例如在细胞(例如,HEK293T细胞)中。在一些实施例中,核酸内切酶结构域与靶DNA或乱序DNA之间的关联频率通过ChIP-seq测量,例如,如He和Pu(2010)Curr.Protoc Mol Biol[分子生物学最新方案]第21章(其通过援引以其全文并入本文)中所述。

[0248] 在一些实施例中,核酸内切酶结构域可以催化在靶序列处形成切口,例如相对于非靶序列(例如,相对于靶细胞基因组中的任何其他基因组序列)增加至少约5倍或10倍。在一些实施例中,使用NickSeq确定切口形成的水平,例如,如Elacqua等人(2019)bioRxiv

doi.org/10.1101/867937(其通过援引以其全文并入本文)中所述。

[0249] 在一些实施例中,核酸内切酶结构域能够在体外对DNA进行切口。在实施例中,切口导致暴露的碱基。在实施例中,暴露的碱基可以使用核酸酶敏感性测定来检测,例如,如Chaudhry和Weinfeld(1995)Nucleic Acids Res[核酸研究]23(19):3805-3809(其通过援引以其全文并入本文)中所述。在实施例中,暴露的碱基的水平(例如,通过核酸酶敏感性测定检测)相对于参考核酸内切酶结构域增加至少10%、50%或更多。在一些实施例中,参考核酸内切酶结构域是来自化脓性链球菌的Cas9的核酸内切酶结构域。

[0250] 在一些实施例中,核酸内切酶结构域能够在细胞中对DNA进行切口。在实施例中,核酸内切酶结构域能够在HEK293T细胞中对DNA进行切口。在实施例中,在没有Rad51的情况下经历复制的未修复的切口导致切口部位的NHEJ率增加,这可以例如通过使用Rad51抑制测定来检测,例如,如Bothmer等人(2017)Nat Commun[自然通讯]8:13905(将其通过援引以其全文并入本文)中所述。在实施例中,NHEJ率增加至0-5%以上。在实施例中,例如在Rad51抑制后,NHEJ率增加至20%-70%(例如,在30%-60%或40%-50%)。

[0251] 在一些实施例中,核酸内切酶结构域在切割后释放靶标。在一些实施例中,通过评估酶的多次周转间接指示靶标的释放,例如,如Yourik等人RNA 25(1):35-44(2019)(其通过援引以其全文并入本文)中所述并如图2所示。在一些实施例中,如通过这样的方法测量的,核酸内切酶结构域的k_{exp}为1x10^{-3}-1x 10^{-}5min-1。

[0252] 在一些实施例中,核酸内切酶结构域在体外具有大于约1x 10^{8}s^{-1}M^{-1}的催化效率(k_{cat}/K_m)。在实施例中,核酸内切酶结构域在体外具有大于约1x 10^{5}、1x 10^{6}、1x 10^{7}或1x 10^{8}s^{-1}M^{-1}的催化效率。在实施例中,催化效率如Chen等人(2018)Science[科学]360(6387):436-439(其通过援引以其全文并入本文)所述确定。在一些实施例中,核酸内切酶结构域在细胞中具有大于约1x 10^{8}s^{-1}M^{-1}的催化效率(k_{cat}/K_m)。在实施例中,核酸内切酶结构域的催化效率在细胞中大于约1x 10^{5}、1x 10^{6}、1x 10^{7}或1x 10^{8},s^{-1}M^{-1}。

[0253] 包含Cas结构域的基因修饰多肽

[0254] 在一些实施例中,本文所述的基因修饰多肽包含Cas结构域。在一些实施例中,Cas结构域可以将基因修饰多肽引导至gRNA间隔子指定的靶位点,从而“顺式”修饰靶核酸序列。在一些实施例中,基因修饰多肽与Cas结构域融合。在一些实施例中,基因修饰多肽包含CRISPR/Cas结构域(在本文中也称为CRISPR相关蛋白)。在一些实施例中,CRISPR/Cas结构域包含参与成簇的调节间隔短回文重复序列(CRISPR)系统的蛋白质(例如Cas蛋白),并且任选地结合指导RNA,例如单指导RNA(sgRNA)。

[0255] CRISPR系统是最初在细菌和古细菌中发现的自适应防御系统。CRISPR系统使用称为CRISPR相关或“Cas”核酸内切酶的RNA引导性核酸酶(例如,Cas9或Cpf1)来切割外来DNA。例如,在典型的CRISPR-Cas系统中,核酸内切酶通过靶向单链或双链DNA序列的序列特异性的非编码“指导RNA”定向到靶核苷酸序列(例如,基因组中待序列编辑的位点)。已经鉴定了三类(I-III)CRISPR系统。II类CRISPR系统使用单个Cas核酸内切酶(而不是多个Cas蛋白)。一种II类CRISPR系统包括II型Cas核酸内切酶,例如Cas9、CRISPR RNA(“crRNA”)和反式激活crRNA(“tracrRNA”)。crRNA含有“间隔子”序列,即通常对应于靶DNA序列的约20个核苷酸的RNA序列(“原间隔子”)。在野生型系统和一些工程改造的系统中,crRNA还包含与tracrRNA结合的区域,以形成被RNA酶III切割的部分双链结构,产生crRNA/tracrRNA杂交

分子。然后,crRNA/tracrRNA杂交体指导Cas核酸内切酶识别并切割靶DNA序列。靶DNA序列通常与“原间隔子相邻基序”(“PAM”)相邻,该基序对于给定的Cas核酸内切酶具有特异性,并且对于与crRNA间隔子匹配的靶位点处的切割活性是需要的。从不同原核物种鉴定的CRISPR核酸内切酶具有独特的PAM序列要求,例如,如表3中针对示例性Cas酶所列出的;PAM序列的实例包括5′-NGG(化脓性链球菌(Streptococcus pyogenes))、5′-NNAGAA(嗜热链球菌(Streptococcus thermophilus)CRISPR1)、5′-NGGNG(嗜热链球菌CRISPR3)、和5′-NNNGATT(奈瑟氏脑膜炎双球菌(Neisseria meningiditis))。一些核酸内切酶(例如Cas9核酸内切酶)与富含G的PAM位点(例如5′-NGG)相关联,并在距PAM位点上游(5′)3个核苷酸处对靶DNA进行钝端切割。另一个II类CRISPR系统包括V型核酸内切酶Cpf1,它比Cas9小;实例包括AsCpf1(来自氨基酸球菌属物种(Acidaminococcus sp.))和LbCpf1(来自毛螺旋菌属物种(Lachnospiraceae sp.))。Cpf1相关CRISPR阵列被处理成成熟crRNA,而不需要tracrRNA;换言之,在一些实施例中,Cpf1系统仅包含Cpf1核酸酶和crRNA以切割靶DNA序列。Cpf1核酸内切酶通常与富含T的PAM位点例如5′-TTN相关联。Cpf1也可以识别5′-CTA PAM基序。Cpf1通常通过引入具有4或5个核苷酸的5′突出端的错位或交错的双链断裂来切割靶DNA,例如切割如下靶DNA,该靶DNA中的5个核苷酸的错位或交错的切割位于距离编码链上的PAM位点下游(3′)18个核苷酸的位置处和距离互补链上的PAM位点下游23个核苷酸的位置处;由这样的错位切割产生的5个核苷酸的突出端使得通过同源重组的DNA插入比在平端切割的DNA的插入更精确地进行基因组编辑。参见例如,Zetsche等人(2015)Cell[细胞],163:759-771。

[0256] 多种CRISPR相关(Cas)基因或蛋白可以用于本披露提供的技术中,并且Cas蛋白的选择将取决于该方法的具体条件。Cas蛋白的具体实例包括II类系统,包括Cas1、Cas2、Cas3、Cas4、Cas5、Cas6、Cas7、Cas8、Cas9、Cas10、Cpf1、C2C1或C2C3。在一些实施例中,Cas蛋白(例如,Cas9蛋白)可以来自多种原核物种中的任一种。在一些实施例中,特定Cas蛋白(例如,特定Cas9蛋白)被选择以识别特定的原间隔子邻近基序(PAM)序列。在一些实施例中,DNA结合结构域或核酸内切酶结构域包括靶向多肽(例如Cas蛋白,例如Cas9)的序列。在某些实施例中,Cas蛋白(例如,Cas9蛋白)可以从细菌或古细菌中获得或使用已知方法合成。在某些实施例中,Cas蛋白可以来自革兰氏阳性细菌或革兰氏阴性细菌。在某些实施例中,Cas蛋白可以来自链球菌属(例如,化脓性链球菌或嗜热链球菌)、弗朗西斯菌属(例如,新凶手弗朗西斯菌)、葡萄球菌属(例如,金黄色葡萄球菌)、氨基酸球菌属(例如,氨基酸球菌属物种BV3L6)、奈瑟氏球菌(例如,脑膜炎奈瑟球菌)、隐球菌属、棒状杆菌属、嗜血杆菌属、真细菌属、巴斯德氏菌属、普雷沃氏菌属、韦荣球菌属或海洋杆菌属。

[0257] 在一些实施例中,基因修饰多肽可包含如表3或4中所列的Cas结构域或其功能片段,或与其具有至少70%、75%、80%、85%、90%、95%、96%、97%、98%、99%同一性的序列。

[0258] 表3.CRISPR/Cas蛋白、物种和突变

[0259]

名称	酶	物种	AA 的数量	PAM	改变 PAM 识别的突变	导致催化失活的突变
FnCas9	Cas9	新凶手弗朗西斯菌	1629	5'-NGG-3'	Wt	D11A/H969A/N995A
FnCas9 RHA	Cas9	新凶手弗朗西斯菌	1629	5'-YG-3'	E1369R/E1449H/R1556A	D11A/H969A/N995A
SaCas9	Cas9	金黄色葡萄球菌（Staphylococcus aureus）	1053	5'-NNGRRT-3'	Wt	D10A/H557A
SaCas9 KKH	Cas9	金黄色葡萄球菌	1053	5'-NNNRRT-3'	E782K/N968K/R1015H	D10A/H557A
SpCas9	Cas9	化脓性链球菌	1368	5'-NGG-3'	Wt	D10A/D839A/H840A/N863A
SpCas9 VQR	Cas9	化脓性链球菌	1368	5'-NGA-3'	D1135V/R1335Q/T1337R	D10A/D839A/H840A/N863A
AsCpf1 RR	Cpf1	氨基酸球菌属物种 BV3L6	1307	5'-TYCV-3'	S542R/K607R	E993A
AsCpf1 RVR	Cpf1	氨基酸球菌属物种 BV3L6	1307	5'-TATV-3'	S542R/K548V/N552R	E993A
FnCpf1	Cpf1	新凶手弗朗西斯菌	1300	5'-NTTN-3'	Wt	D917A/E1006A/D1255A
NmCas9	Cas9	脑膜炎奈瑟球菌（Neisseria meningitidis）	1082	5'-NNNGATT-3'	Wt	D16A/D587A/H588A/N611A

[0260] 表4CRISPR/Cas蛋白的氨基酸序列、物种和突变

[0261]

变体	亲本宿主	蛋白质序列	切口酶（HNH）	切口酶（HNH）	切口酶（RuvC）
Nme2Cas9	脑膜炎奈瑟球菌	MAAFKPNPINYILGLDIGIASVGWAMVEIDEEENPIRLIDLGVRVFERAEVPKTGDSLAMARRLARSVRRLTRRRAHRLLRARRLLKREGVLQAADFDENGLIKSLPNTPWQLRAAALDRKLTPLEWSAVLLHLIKHRGYLSQRKNEGETADKELGALLKGVANNAHALQTGDFRTPAELALNKFEKESGHIRNQRGDYSHTFSRKDLQAELILLFEKQKEFGNPHVSGGLKEGIETLLMTQRPALSGDAVQKMLGHCTFEPAEPKAAKNTYTAERFIWLTKLNNLRILEQGSERPLTDTERATLMDEPYRKSKLTYAQARKLLGLEDTAFFKGLRYGKDNAEASTLMEMKAYHAISRALEKEGLKDKKSPLNLSSELQDEIGTAFSLFKTDEDITGRLKDRVQPEILEALLKHISFDKFVQISLKALRRIVPLMEQGKRYDEACAEIYGDHYGKKNTEEKIYLPPIPADEIRNPVVLRALSQARKVINGVVRRYGSPARIHIETAREVGKSFKDRKEIEKRQEENRKDREKAAAKFREYFPNFVGEPKSKDILKLRLYEQQHGKCLYSGKEINLVRLNEKGYVEIDHALPFSRTWDDSFNNKVLVLGSENQNKGNQTPYEYFNGKDNSREWQEFKARVETSRFPRSKKQRILLQKFDEDGFKECNLNDTRYVNRFLCQFVADHILLTGKGKRRVFASNGQITNLLRGFWGLRKVRAENDRHHALDAVVVACSTVAMQQKITRFVRYKEMNAFDGKTIDKETGKVLHQKTHFPQPWEFFAQEVMIRVFGKPDGKPEFEEADTPEKLRTLLAEKLSSRPEAVHEYVTPLFVSRAPNRKMSGAHKDTLRSAKRFVKHNEKISVKRVWLTEIKLADLENMVNYKNGREIELYEALKARLEAYGGNAKQAFDPKDNPFYKKGGQLVKAVRVEKTQESGVLLNKKNAYTIADNGDMVRVDVFCKVDKKGKNQYFIVPIYAWQVAENILPDIDCKGYRIDDSYTFCFSLHKYDLIAFQKDEKSKVEFAYYINCDSSNGRFYLAWHDKGSKEQQFRISTQNLVLIQKYQVNELGKEIRPCRLKKRPPVR	N611A	H588A	D16A
PpnCas9	侵肺巴斯德氏菌（Pasteurella pneumotropica）	MQNNPLNYILGLDLGIASIGWAVVEIDEESSPIRLIDVGVRTFERAEVAKTGESLALSRRLARSSRRLIKRRAERLKKAKRLLKAEKILHSIDEKLPINVWQLRVKGLKEKLERQEWAAVLLHLSKHRGYLSQRKNEGKSDNKELGALLSGIASNHQMLQSSEYRTPAEIAVKKFQVEEGHIRNQRGSYTHTFSRLDLLAEMELLFQRQAELGNSYTSTTLLENLTALLMWQKPALAGDAILKMLGKCTFEPSEYKAAKNSYSAERFVWLTKLNNLRILENGTERALNDNERFALLEQPYEKSKLTYAQVRAMLALSDNAIFKGVRYLGEDKKTVESKTTLIEMKFYHQIRKTLGSAELKKEWNELKGNSDLLDEIGTAFSLYKTDDDICRYLEGKLPERVLNALLENLNFDKFIQLSLKALHQILPLMLQGGQRYDEAVSAIYGDHYGKKSTETTRLLPTIPADEIRNPVVLRTLTQA	N605A	H582A	D13A

[0262]

		RKVINAVVRLYGSPARIHIETAREVGKSYQDRKKLEKQQEDNRKQRESAVKK FKEMFPHFVGEPKGKDILKMRLYELQQAKCLYSGKSLELHRLLEKGYVEVDH ALPFSRTWDDSFNNKVLVLANENQNKGNLTPYEWLDGKNNSERWQHFVV RVQTSGFSYAKKQRILNHKLDEKGFIERNLNDTRYVARFLCNFIADNMLLVG KGKRNVFASNGQITALLRHRWGLQKVREQNDRHHALDAVVVACSTVAMQ QKITRFVRYNEGNVFSGERIDRETGEIIPLHFPSPWAFFKENVEIRIFSENPKLE LENRLPDYPQYNHEWVQPLFVSRMPTRKMTGQGHMETVKSAKRLNEGLS VLKVPLTQLKLSDLERMVNRDREIALYESLKARLEQFGNDPAKAFAEPFYKKG GALVKAVRLEQTQKSGVLVRDGNGVADNASMVRVDVFTKGGKYFLVPIYT WQVAKGILPNRAATQGKDENDWDIMDEMATFQFSLCQNDLIKLVTKKKTI FGYFNGLNRATSNINIKEHDLDKSKGKLGIYLEVGVKLAISLEKYQVDELGKNI RPCRPTKRQHVR			
SauCas9	金黄色葡萄球菌	MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGA RRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSA ALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKK DGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYE GPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLN NLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTST GKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSEL TQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKVD LSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSKD AQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLE AIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQYL SSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLV DTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGY KHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQE YKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRELINDTLYSTRKDDKGNTLIVN NLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPL YKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKL SLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQA EFIASFYNNDLIKINGELYRVIGVNNDLLNRIEVNMIDITYREYLENMNDKRPP RIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG	N580A	H557A	D10A
SauCas9-KKH	金黄色葡萄球菌	MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGA RRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSA ALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKK DGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYE GPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLN NLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTST GKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSEL TQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKVD LSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSKD AQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLE AIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQYL SSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLV DTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGY KHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQE YKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRKLINDTLYSTRKDDKGNTLIVN NLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPL YKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKL SLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQA EFIASFYKNDLIKINGELYRVIGVNNDLLNRIEVNMIDITYREYLENMNDKRPP HIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG	N580A	H557A	D10A
SauriCas9	耳葡萄球菌	MQENQQKQNYILGLDIGITSVGYGLIDSKTREVIDAGVRLFPEADSENNSNR RSKRGARRLKRRRIHRLNRVKDLLADYQMIDLNNVPKSTDPYTIRVKGLREPL TKEEFAIALLHIAKRRGLHNISVSMGDEEQDNELSTKQQLQKNAQQLQDKY VCELQLERLTNINKVRGEKNRFKTEDFVKEVKQLCETQRQYHNIDDQFIQQY IDLVSTRREYFEGPGNGSPYGWDGDLLKWYEKLMGRCTYFPEELRSVKYAYS ADLFNALNDLNNLVVTRDDNPKLEYYEKYHIIENVFKQKKNPTLKQIAKEIGV QDYDIRGYRITKSGKPQFTSFKLYHDLKNIFEQAKYLEDVEMLDEIAKILTIYQ DEISIKKALDQLPELLTESEKSQIAQLTGYTGTHRLSLKCIHIVIDELWESPENQ MEIFTRLNLKPKKVEMSEIDSIPTTLVDEFILSPVVKRAFIQSIKVINAVINRFGL PEDIIIELAREKNSKDRRKFINKLQKQNEATRKKIEQLLAKYGNTNAKYMIEKI KLHDMQEGKCLYSLEAIPLEDLLSNPTHYEVDHIIPRSVSFDNSLNNKVLVKQ SENSKKGNRTPYQYLSSNESKISYNQFKQHILNLSKAKDRISKKKRDMLLEER DINKFEVQKEFINRNLVDTRYATRELSNLLKTYFSTHDYAVKVKTINGGFTNH LRKVWDFKKHRNHGYKHHAEDALVIANADFLFKTHKALRRTDKILEQPGLE VNDTTVKVDTEEKYQELFETPKQVKNIKQFRDFKYSHRVDKKPNRQLINDTL YSTREIDGETYVVQTLKDLYAKDNEKVKKLFTERPQKILMYQHDPKTFEKLM TILNQYAEAKNPLAAYYEDKGEYVTKYAKKGNGPAIHKIKYIDKKLGSYLDVS NKYPETQNKLVKLSLKSFRFDIYKCEQGYKMVSIGYLDVLKKDNYYYIPKDKYE AEKQKKIKESDLFVGSFYYNDLIMYEDELFRVIGVNSDINNLVELNMVDITY KDFCEVNNVTGEKRIKKTIGKRVVLIEKYTTDILGNLYKTPLPKKPQLIFKRGEL			
SauriCas9-	耳葡萄球菌	MQENQQKQNYILGLDIGITSVGYGLIDSKTREVIDAGVRLFPEADSENNSNR	N588A	H565A	D15A

[0263]

KKH		RSKRGARRLKRRRIHRLNRVKDLLADYQMIDLNNVPKSTDPYTIRVKGLREPL TKEEFAIALLHIAKRRGLHNISVSMGDEEQDNELSTKQQLQKNAQQLQDKY VCELQLERLTNINKVRGEKNRFKTEDFVKEVKQLCETQRQYHNIDDQFIQQY IDLVSTRREYFEGPGNGSPYGWDGDLLKWYEKLMGRCTYFPEELRSVKYAYS ADLFNALNDLNNLVVTRDDNPKLEYYEKYHIIENVFKQKKNPTLKQIAKEIGV QDYDIRGYRITKSGKPQFTSFKLYHDLKNIFEQAKYLEDVEMLDEIAKILTIYQ DEISIKKALDQLPELLTESEKSQIAQLTGYTGTHRLSLKCIHIVIDELWESPENQ MEIFTRLNLKPKKVEMSEIDSIPTTLVDEFILSPVVKRAFIQSIKVINAVINRFGL PEDIIIELAREKNSKDRRKFINKLQKQNEATRKKIEQLLAKYGNTNAKYMIEKI KLHDMQEGKCLYSLEAIPLEDLLSNPTHYEVDHIIPRSVSFDNSLNNKVLVKQ SENSKKGNRTPYQYLSSNESKISYNQFKQHILNLSKAKDRISKKKRDMLLEER DINKFEVQKEFINRNLVDTRYATRELSNLLKTYFSTHDYAVKVKTINGGFTNH LRKVWDFKKHRNHGYKHHAEDALVIANADFLFKTHKALRRTDKILEQPGLE VNDTTVKVDTEEKYQELFETPKQVKNIKQFRDFKYSHRVDKKPNRKLINDTL YSTREIDGETYVVQTLKDLYAKDNEKVKKLFTERPQKILMYQHDPKTFEKLM TILNQYAEAKNPLAAYYEDKGEYVTKYAKKGNGPAIHKIKYIDKKLGSYLDVS NKYPETQNKLVKLSLKSFRFDIYKCEQGYKMVSIGYLDVLKKDNYYYIPKDKYE AEKQKKKIKESDLFVGSFYKNDLIMYEDELFRVIGVNSDINNLVELNMVDITY KDFCEVNNVTGEKHIKKTIGKRVVLIEKYTTDILGNLYKTPLPKKPQLIFKRGEL			
ScaCas9-Sc++	犬链球菌	MEKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTNRKSIKKNLMGALL FDSGETAEATRLKRTARRRYTRRKNRIRYLQEIFANEMAKLDDSFFQRLEESF LVEEDKKNERHPIFGNLADEVAYHRNYPTIYHLRKKLADSPEKADLRLIYLALA HIIKFRGHFLIEGKLNAENSDVAKLFYQLIQTYNQLFEESPLDEIEVDAKGILSA RLSKSKRLEKLIAVFPNEKKNGLFGNIIALALGLTPNFKSNFDLTEDAKLQLSKD TYDDDLDELLGQIGDQYADLFSAAKNLSDAILLSDILRSNSEVTKAPLSASMV KRYDEHHQDLALLKTLVRQQFPEKYAEIFKDDTKNGYAGYVGADKKLRKRS GKLATEEEFYKFIKPILEKMDGAEELLAKLNRDDLLRKQRTFDNGSIPHQIHLK ELHAILRRQEEFYPFLKENREKIEKILTFRIPYYVGPLARGNSRFAWLTRKSEEA ITPWNFEEVVDKGASAQSFIERMTNFDEQLPNKKVLPKHSLLYEYFTVYNEL TKVKYVTERMRKPEFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDS VEIIGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIE ERLKTYAHLFDDKVMKQLKRRHYTGWGRLSRKMINGIRDKQSGKTILDFLKS DGFSNRNFMQLIHDDSLTFKEEIEKAQVSGQGDSLHEQIADLAGSPAIKKGIL QTVKIVDELVKVMGHKPENIVIEMARENQTTTKGLQQSRERKKRIEEGIKELE SQILKENPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVP QSFIKDDSIDNKVLTRSVENRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQ RKFDNLTKAERGGLSEADKAGFIKRQLVETRQITKHVARILDSRMNTKRDKN DKPIREVKVITLKSKLVSDFRKDFQLYKVRDINNYHHAHDAYLNAVVGTALIK KYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKRFFYSNIMNFFKTEVKL ANGEIRKRPLIETNGETGEVVWNKEKDFATVRKVLAMPQVNIVKKTEVQTG GFSKESILSKRESAKLIPRKKGWDTRKYGGFGSPTVAYSILVVAKVEKGKAKKL KSVKVLVGITIMEKGSYEKDPIGFLEAKGYKDIKKELIFKLPKYSLFELENGRRR MLASAKELQKANELVLPQHLVRLLYYTQNISATTGSNNLGYIEQHREEFKEIF EKIIDFSEKYILKNKVNSNLKSSFDEQFAVSDSILLSNSFVSLLKYTSFGASGGFT FLDLDVKQGRLRYQTVTEVLDATLIYQSITGLYETRTDLSQLGGD	N872A	H849A	D10A
SpyCas9	化脓性链球菌	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFL VEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAH MIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILS ARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLS KDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQ EDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANR NFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQ ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSF LKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPK LESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEI RKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKES ILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKE LLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASA GELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEII EQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAF KYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	N863A	H840A	D10A
SpyCas9-NG	化脓性链球菌	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFL	N863A	H840A	D10A

[0264]

		VEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAH MIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILS ARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLS KDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQ EDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANR NFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQ ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSF LKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPK LESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEI RKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKES IRPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKE LLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASA RFLQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEII EQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPRAF KYFDTTIDRKVYRSTKEVLDATLIHQSITGLYETRIDLSQLGGD			
SpyCas9-SpRY	化脓性链球菌	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF DSGETAERTRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFL VEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAH MIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILS ARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLS KDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQ EDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANR NFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQ ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSF LKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPK LESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEI RKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKES IRPKRNSDKLIARKKDWDPKKYGGFLWPTVAYSVLVVAKVEKGKSKKLKSVK ELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS AKQLQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDE IIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTRLGAPRAF KYFDTTIDPKQYRSTKEVLDATLIHQSITGLYETRIDLSQLGGD	N863A	H840A	D10A
St1Cas9	嗜热链球菌	MSDLVLGLDIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAENNLVRRTNRQG RRLARRKKHRRVRLNRLFEESGLITDFTKISINLNPYQLRVKGLTDELSNEELFI ALKNMVKHRGISYLDDASDDGNSSVGDYAQIVKENSKQLETKTPGQIQLER YQTYGQLRGDFTVEKDGKKHRLINVFPTSAYRSEALRILQTQQEFNPQITDEF INRYLEILTGKRKYYHGPGNEKSRTDYGRYRTSGETLDNIFGILIGKCTFYPDEF RAAKASYTAQEFNLLNDLNNLTVPTETKKLSKEQKNQIINYVKNEKAMGPAK LFKYIAKLLSCDVADIKGYRIDKSGKAEIHTFEAYRKMKTLETLDIEQMDRETL DKLAYVLTLNTEREGIQEALEHEFADGSFSQKQVDELVQFRKANSSIFGKGW HNFSVKLMMELIPELYETSEEQMTILTRLGKQKTTSSSNKTKYIDEKLLTEEIY NPVVAKSVRQAIKIVNAAIKEYGDFDNIVIEMARETNEDDEKKAIQKIQKAN KDEKDAAMLKAANQYNGKAELPHSVFHGHKQLATKIRLWHQQGERCLYT GKTISIHDLINNSNQFEVDHILPLSITFDDSLANKVLVYATANQEKGQRTPYQ ALDSMDDAWSFRELKAFVRESKTLSNKKKEYLLTEEDISKFDVRKKFIERNLV DTRYASRVVLNALQEHFRAHKIDTKVSVVRGQFTSQLRRHWGIEKTRDTYH HHAVDALIIAASSQLNLWKKQKNTLVSYSEDQLLDIETGELISDDEYKESVFKA PYQHFVDTLKSKEFEDSILFSYQVDSKFNRKISDATIYATRQAKVGKDKADET YVLGKIKDIYTQDGYDAFMKIYKKDKSKFLMYRHDPQTFEKVIEPILENYPNK QINEKGKEVPCNPFLKYKEEHGYIRKYSKKGNGPEIKSLKYYDSKLGNHIDITP KDSNNKVVLQSVSPWRADVYFNKTTGKYEILGLKYADLQFEKGTGTYKISQE KYNDIKKKEGVDSDSEFKFTLYKNDLLLVKDTETKEQQLFRFLSRTMPKQKHY VELKPYDKQKFEGGEALIKVLGNVANSGQCKKGLGKSNISIYKVRTDVLGNQ HIIKNEGDKPKLDF	N622A	H599A	D9A
BlatCas9	侧孢短芽孢	MAYTMGIDVGIASCGWAIVDLERQRIIDIGVRTFEKAENPKNGEALAVPRRE	N607A	H584A	D8A

[0265]

	杆菌	ARSSRRRLRRKKHRIERLKHMFVRNGLAVDIQHLEQTLRSQNEIDVWQLRV DGLDRMLTQKEWLRVLIHLAQRRGFQSNRKTDGSSEDGQVLVNVTENDRL MEEKDYRTVAEMMVKDEKFSDHKRNKNGNYHGVVSRSSLLVEIHTLFETQ RQHHNSLASKDFELEYVNIWSAQRPVATKDQIEKMIGTCTFLPKEKRAPKAS WHFQYFMLLQTINHIRITNVQGTRSLNKEEIEQVVNMALTKSKVSYHDTRKI LDLSEEYQFVGLDYGKEDEKKKVESKETIIKLDDYHKLNKIFNEVELAKGETWE ADDYDTVAYALTFFKDDEDIRDYLQNKYKDSKNRLVKNLANKEYTNELIGKV STLSFRKVGHLSLKALRKIIPFLEQGMTYDKACQAAGFDFQGISKKKRSVVLP VIDQISNPVVNRALTQTRKVINALIKKYGSPETIHIETARELSKTFDERKNITKD YKENRDKNEHAKKHLSELGIINPTGLDIVKYKLWCEQQGRCMYSNQPISFER LKESGYTEVDHIIPYSRSMNDSYNNRVLVMTRENREKGNQTPFEYMGNDT QRWYEFEQRVTTNPQIKKEKRQNLLLKGFTNRRELEMLERNLNDTRYITKYL SHFISTNLEFSPSDKKKKVVNTSGRITSHLRSRWGLEKNRGQNDLHHAMDAI VIAVTSDSFIQQVTNYYKRKERRELNGDDKFPLPWKFFREEVIARLSPNPKEQ IEALPNHFYSEDELADLQPIFVSRMPKRSITGEAHQAQFRRVVGKTKEGKNIT AKKTALVDISYDKNGDFNMYGRETDPATYEAIKERYLEFGGNVKKAFSTDLH KPKKDGTKGPLIKSVRIMENKTLVHPVNKGKGVVYNSSIVRTDVFQRKEKYY LLPVYVTDVTKGKLPNKVIVAKKGYHDWIEVDDSFTFLFSLYPNDLIFIRQNPK KKISLKKRIESHSISDSKEVQEIHAYYKGVDSSTAAIEFIIHDGSYYAKGVGVQN LDCFEKYQVDILGNYFKVKGEKRLELETSDSNHKGKDVNSIKSTSR			
cCas9-v16	金黄色葡萄球菌	MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGA RRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSA ALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKK DGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYE GPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLN NLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTST GKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSEL TQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKVD LSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSKD AQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLE AIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQYL SSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLV DTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGY KHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQE YKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRKLINDTLYSTRKDDKGNTLIVN NLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPL YKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKL SLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQA EFIASFYKNDLIKINGELYRVIGVNSDKNNLIEVNMIDITYREYLENMNDKRPP HIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG	N580A	H557A	D10A
cCas9-v17	金黄色葡萄球菌	MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGA RRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSA ALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKK DGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYE GPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLN NLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTST GKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSEL TQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKVD LSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSKD AQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLE AIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQYL SSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLV DTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGY KHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQE YKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRKLINDTLYSTRKDDKGNTLIVN NLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPL YKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKL SLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQA EFIASFYKNDLIKINGELYRVIGVNNSTRNIVELNMIDITYREYLENMNDKRPP HIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG	N580A	H557A	D10A
cCas9-v21	金黄色葡萄球菌	MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGA RRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSA ALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKK DGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYE GPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLN NLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTST GKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSEL TQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKVD LSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSKD AQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLE AIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQYL SSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLV DTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGY	N580A	H557A	D10A

[0266]

		KHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQE YKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRKLINDTLYSTRKDDKGNTLIVN NLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPL YKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKL SLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQA EFIASFYKNDLIKINGELYRVIGVNSDDRNIIELNMIDITYREYLENMNDKRPP HIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG			
cCas9-v42	金黄色葡萄球菌	MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGA RRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSA ALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKK DGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYE GPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLN NLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTST GKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSEL TQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKVD LSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSKD AQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLE AIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQYL SSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLV DTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGY KHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQE YKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRKLINDTLYSTRKDDKGNTLIVN NLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPL YKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKL SLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQA EFIASFYKNDLIKINGELYRVIGVNNNRLNKIELNMIDITYREYLENMNDKRPP HIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG	N580A	H557A	D10A
CdiCas9	白喉棒状杆菌（Corynebacterium diphtheriae）	MKYHVGIDVGTFSVGLAAIEVDDAGMPIKTLSLVSHIHDSGLDPDEIKSAVTR LASSGIARRTRRLYRRKRRRLQQLDKFIQRQGWPVIELEDYSDPLYPWKVRA ELAASYIADEKERGEKLSVALRHIARHRGWRNPYAKVSSLYLPDGPSDAFKAI REEIKRASGQPVPETATVGQMVTLCELGTLKLRGEGGVLSARLQQSDYAREI QEICRMQEIGQELYRKIIDVVFAAESPKGSASSRVGKDPLQPGKNRALKASD AFQRYRIAALIGNLRVRVDGEKRILSVEEKNLVFDHLVNLTPKKEPEWVTIAEI LGIDRGQLIGTATMTDDGERAGARPPTHDTNRSIVNSRIAPLVDWWKTAS ALEQHAMVKALSNAEVDDFDSPEGAKVQAFFADLDDDVHAKLDSLHLPVG RAAYSEDTLVRLTRRMLSDGVDLYTARLQEFGIEPSWTPPTPRIGEPVGNPA VDRVLKTVSRWLESATKTWGAPERVIIEHVREGFVTEKRAREMDGDMRRR AARNAKLFQEMQEKLNVQGKPSRADLWRYQSVQRQNCQCAYCGSPITFS NSEMDHIVPRAGQGSTNTRENLVAVCHRCNQSKGNTPFAIWAKNTSIEGV SVKEAVERTRHWVTDTGMRSTDFKKFTKAVVERFQRATMDEEIDARSMES VAWMANELRSRVAQHFASHGTTVRVYRGSLTAEARRASGISGKLKFFDGV GKSRLDRRHHAIDAAVIAFTSDYVAETLAVRSNLKQSQAHRQEAPQWREFT GKDAEHRAAWRVWCQKMEKLSALLTEDLRDDRVVVMSNVRLRLGNGSA HKETIGKLSKVKLSSQLSVSDIDKASSEALWCALTREPGFDPKEGLPANPERHI RVNGTHVYAGDNIGLFPVSAGSIALRGGYAELGSSFHHARVYKITSGKKPAF AMLRVYTIDLLPYRNQDLFSVELKPQTMSMRQAEKKLRDALATGNAEYLG WLVVDDELVVDTSKIATDQVKAVEAELGTIRRWRVDGFFSPSKLRLRPLQM SKEGIKKESAPELSKIIDRPGWLPAVNKLFSDGNVTVVRRDSLGRVRLESTAH LPVTWKVQ	N597A	H573A	D8A
CjeCas9	空肠弯曲杆菌（Campylobacter jejuni）	MARILAFDIGISSIGWAFSENDELKDCGVRIFTKVENPKTGESLALPRRLARSA RKRLARRKARLNHLKHLIANEFKLNYEDYQSFDESLAKAYKGSLISPYELRFRA LNELLSKQDFARVILHIAKRRGYDDIKNSDDKEKGAILKAIKQNEEKLANYQS VGEYLYKEYFQKFKENSKEFTNVRNKKESYERCIAQSFLKDELKLIFKKQREFG FSFSKKFEEEVLSVAFYKRALKDFSHLVGNCSFFTDEKRAPKNSPLAFMFVAL TRIINLLNNLKNTEGILYTKDDLNALLNEVLKNGTLTYKQTKKLLGLSDDYEFK GEKGTYFIEFKKYKEFIKALGEHNLSQDDLNEIAKDITLIKDEIKLKKALAKYDLN QNQIDSLSKLEFKDHLNISFKALKLVTPLMLEGKKYDEACNELNLKVAINEDK KDFLPAFNETYYKDEVTNPVVLRAIKEYRKVLNALLKKYGKVHKINIELAREVG KNHSQRAKIEKEQNENYKAKKDAELECEKLGLKINSKNILKLRLFKEQKEFCAY SGEKIKISDLQDEKMLEIDHIYPYSRSFDDSYMNKVLVFTKQNQEKLNQTPFE AFGNDSAKWQKIEVLAKNLPTKKQKRILDKNYKDKEQKNFKDRNLNDTRYI ARLVLNYTKDYLDFLPLSDDENTKLNDTQKGSKVHVEAKSGMLTSALRHTW GFSAKDRNNHLHHAIDAVIIAYANNSIVKAFSDFKKEQESNSAELYAKKISELD YKNKRKFFEPFSGFRQKVLDKIDEIFVSKPERKKPSGALHEETFRKEEEFYQSY GGKEGVLKALELGKIRKVNGKIVKNGDMFRVDIFKHKKTNKFYAVPIYTMDF ALKVLPNKAVARSKKGEIKDWILMDENYEFCFSLYKDSLILIQTKDMQEPEFV YYNAFTSSTVSLIVSKHDNKFETLSKNQKILFKNANEKEVIAKSIGIQNLKVFEK YIVSALGEVTKAEFRQREDFKK	N582A	H559A	D8A
GeoCas9	嗜热脂肪地芽孢杆菌	MRYKIGLDIGITSVGWAVMNLDIPRIEDLGVRIFDRAENPQTGESLALPRRLA RSARRRLRRRKHRLERIRRLVIREGILTKEELDKLFEEKHEIDVWQLRVEALDR KLNNDELARVLLHLAKRRGFKSNRKSERSNKENSTMLKHIEENRAILSSYRTV GEMIVKDPKFALHKRNKGENYTNTIARDDLEREIRLIFSKQREFGNMSCTEEF ENEYITIWASQRPVASKDDIEKKVGFCTFEPKEKRAPKATYTFQSFIAWEHIN	N605A	H582A	D8A

[0267]

		KLRLISPSGARGLTDEERRLLYEQAFQKNKITYHDIRTLLHLPDDTYFKGIVYDR GESRKQNENIRFLELDAYHQIRKAVDKVYGKGKSSSFLPIDFDTFGYALTLFKD DADIHSYLRNEYEQNGKRMPNLANKVYDNELIEELLNLSFTKFGHLSLKALRS ILPYMEQGEVYSSACERAGYTFTGPKKKQKTMLLPNIPPIANPVVMRALTQA RKVVNAIIKKYGSPVSIHIELARDLSQTFDERRKTKKEQDENRKKNETAIRQL MEYGLTLNPTGHDIVKFKLWSEQNGRCAYSLQPIEIERLLEPGYVEVDHVIPY SRSLDDSYTNKVLVLTRENREKGNRIPAEYLGVGTERWQQFETFVLTNKQFS KKKRDRLLRLHYDENEETEFKNRNLNDTRYISRFFANFIREHLKFAESDDKQK VYTVNGRVTAHLRSRWEFNKNREESDLHHAVDAVIVACTTPSDIAKVTAFY QRREQNKELAKKTEPHFPQPWPHFADELRARLSKHPKESIKALNLGNYDDQ KLESLQPVFVSRMPKRSVTGAAHQETLRRYVGIDERSGKIQTVVKTKLSEIKL DASGHFPMYGKESDPRTYEAIRQRLLEHNNDPKKAFQEPLYKPKKNGEPGP VIRTVKIIDTKNQVIPLNDGKTVAYNSNIVRVDVFEKDGKYYCVPVYTMDIM KGILPNKAIEPNKPYSEWKEMTEDYTFRFSLYPNDLIRIELPREKTVKTAAGEE INVKDVFVYYKTIDSANGGLELISHDHRFSLRGVGSRTLKRFEKYQVDVLGNI YKVRGEKRVGLASSAHSKPGKTIRPLQSTRD			
iSpyMacCas9	链球菌属物种	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFL VEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAH MIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILS ARLSKSRKLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLS KDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF YKFIKPILEKMDGTEELLVKLKREDLLRKQRTFDNGSIPHQIHLGELHAILRRQ EDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANR NFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQ ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSF LKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPK LESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEI RKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEIQTVGQNGG LFDDNPKSPLEVTPSKLVPLKKELNPKKYGGYQKPTTAYPVLLITDTKQLIPISV MNKKQFEQNPVKFLRDRGYQQVGKNDFIKLPKYTLVDIGDGIKRLWASSKEI HKGNQLVVSKKSQILLYHAHHLDSDLSNDYLQNHNQQFDVLFNEIISFSKKC KLGKEHIQKIENVYSNKKNSASIEELAESFIKLLGFTQLGATSPFNFLGVKLNQ KQYKGKKDYILPCTEGTLIRQSITGLYETRVDLSKIGEDSGGSGGSKRTADGSE FES	N863A	H840A	D10A
NmeCas9	脑膜炎奈瑟球菌	MAAFKPNSINYILGLDIGIASVGWAMVEIDEEENPIRLIDLGVRVFERAEVPK TGDSLAMARRLARSVRRLTRRRAHRLLRTRRLLKREGVLQAANFDENGLIKS LPNTPWQLRAAALDRKLTPLEWSAVLLHLIKHRGYLSQRKNEGETADKELG ALLKGVAGNAHALQTGDFRTPAELALNKFEKESGHIRNQRSDYSHTFSRKDL QAELILLFEKQKEFGNPHVSGGLKEGIETLLMTQRPALSGDAVQKMLGHCTF EPAEPKAAKNTYTAERFIWLTKLNNLRILEQGSERPLTDTERATLMDEPYRKS KLTYAQARKLLGLEDTAFFKGLRYGKDNAEASTLMEMKAYHAISRALEKEGL KDKKSPLNLSPELQDEIGTAFSLFKTDEDITGRLKDRIQPEILEALLKHISFDKFV QISLKALRRIVPLMEQGKRYDEACAEIYGDHYGKKNTEEKIYLPPIPADEIRNP VVLRALSQARKVINGVVRRYGSPARIHIETAREVGKSFKDRKEIEKRQEENRK DREKAAAKFREYFPNFVGEPKSKDILKLRLYEQQHGKCLYSGKEINLGRLNEK GYVEIDHALPFSRTWDDSFNNKVLVLGSENQNKGNQTPYEYFNGKDNSRE WQEFKARVETSRFPRSKKQRILLQKFDEDGFKERNLNDTRYVNRFLCQFVA DRMRLTGKGKKRVFASNGQITNLLRGFWGLRKVRAENDRHHALDAVVVA CSTVAMQQKITRFVRYKEMNAFDGKTIDKETGEVLHQKTHFPQPWEFFAQ EVMIRVFGKPDGKPEFEEADTLEKLRTLLAEKLSSRPEAVHEYVTPLFVSRAP NRKMSGQGHMETVKSAKRLDEGVSVLRVPLTQLKLKDLEKMVNREREPKL YEALKARLEAHKDDPAKAFAEPFYKYDKAGNRTQQVKAVRVEQVQKTGVW VRNHNGIADNATMVRVDVFEKGDKYYLVPIYSWQVAKGILPDRAVVQGKD EEDWQLIDDSFNFKFSLHPNDLVEVITKKARMFGYFASCHRGTGNINIRIHD LDHKIGKNGILEGIGVKTALSFQKYQIDELGKEIRPCRLKKRPPVR	N611A	H588A	D16A
ScaCas9	犬链球菌	MEKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTNRKSIKKNLMGALL FDSGETAEATRLKRTARRRYTRRKNRIRYLQEIFANEMAKLDDSFFQRLEESF LVEEDKKNERHPIFGNLADEVAYHRNYPTIYHLRKKLADSPEKADLRLIYLALA HIIKFRGHFLIEGKLNAENSDVAKLFYQLIQTYNQLFEESPLDEIEVDAKGILSA RLSKSKRLEKLIAVFPNEKKNGLFGNIIALALGLTPNFKSNFDLTEDAKLQLSKD TYDDDLDELLGQIGDQYADLFSAAKNLSDAILLSDILRSNSEVTKAPLSASMV KRYDEHHQDLALLKTLVRQQFPEKYAEIFKDDTKNGYAGYVDGIGIKHRKRTT KLATQEEFYKFIKPILEKMDGAEELLAKLNRDDLLRKQRTFDNGSIPHQIHLKE LHAILRRQEEFYPFLKENREKIEKILTFRIPYYVGPLARGNSRFAWLTRKSEEAI	N872A	H849A	D10A

[0268]

		TPWNFEEVVDKGASAQSFIERMTNFDEQLPNKKVLPKHSLLYEYFTVYNELT KVKYVTERMRKPEFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSV EIIGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE RLKTYAHLFDDKVMKQLKRRHYTGWGRLSRKMINGIRDKQSGKTILDFLKS DGFSNRNFMQLIHDDSLTFKEEIEKAQVSGQGDSLHEQIADLAGSPAIKKGIL QTVKIVDELVKVMGHKPENIVIEMARENQTTTKGLQQSRERKKRIEEGIKELE SQILKENPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVP QSFIKDDSIDNKVLTRSVENRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQ RKFDNLTKAERGGLSEADKAGFIKRQLVETRQITKHVARILDSRMNTKRDKN DKPIREVKVITLKSKLVSDFRKDFQLYKVRDINNYHHAHDAYLNAVVGTALIK KYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKRFFYSNIMNFFKTEVKL ANGEIRKRPLIETNGETGEVVWNKEKDFATVRKVLAMPQVNIVKKTEVQTG GFSKESILSKRESAKLIPRKKGWDTRKYGGFGSPTVAYSILVVAKVEKGKAKKL KSVKVLVGITIMEKGSYEKDPIGFLEAKGYKDIKKELIFKLPKYSLFELENGRRR MLASATELQKANELVLPQHLVRLLYYTQNISATTGSNNLGYIEQHREEFKEIF EKIIDFSEKYILKNKVNSNLKSSFDEQFAVSDSILLSNSFVSLLKYTSFGASGGFT FLDLDVKQGRLRYQTVTEVLDATLIYQSITGLYETRTDLSQLGGD			
ScaCas9-HiFi-Sc++	犬链球菌	MEKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTNRKSIKKNLMGALL FDSGETAEATRLKRTARRRYTRRKNRIRYLQEIFANEMAKLDDSFFQRLEESF LVEEDKKNERHPIFGNLADEVAYHRNYPTIYHLRKKLADSPEKADLRLIYLALA HIIKFRGHFLIEGKLNAENSDVAKLFYQLIQTYNQLFEESPLDEIEVDAKGILSA RLSKSKRLEKLIAVFPNEKKNGLFGNIIALALGLTPNFKSNFDLTEDAKLQLSKD TYDDDLDELLGQIGDQYADLFSAAKNLSDAILLSDILRSNSEVTKAPLSASMV KRYDEHHQDLALLKTLVRQQFPEKYAEIFKDDTKNGYAGYVGADKKLRKRS GKLATEEEFYKFIKPILEKMDGAEELLAKLNRDDLLRKQRTFDNGSIPHQIHLK ELHAILRRQEEFYPFLKENREKIEKILTFRIPYYVGPLARGNSRFAWLTRKSEEA ITPWNFEEVVDKGASAQSFIERMTNFDEQLPNKKVLPKHSLLYEYFTVYNEL TKVKYVTERMRKPEFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDS VEIIGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIE ERLKTYAHLFDDKVMKQLKRRHYTGWGRLSRKMINGIRDKQSGKTILDFLKS DGFSNANFMQLIHDDSLTFKEEIEKAQVSGQGDSLHEQIADLAGSPAIKKGIL QTVKIVDELVKVMGHKPENIVIEMARENQTTTKGLQQSRERKKRIEEGIKELE SQILKENPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVP QSFIKDDSIDNKVLTRSVENRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQ RKFDNLTKAERGGLSEADKAGFIKRQLVETRQITKHVARILDSRMNTKRDKN DKPIREVKVITLKSKLVSDFRKDFQLYKVRDINNYHHAHDAYLNAVVGTALIK KYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKRFFYSNIMNFFKTEVKL ANGEIRKRPLIETNGETGEVVWNKEKDFATVRKVLAMPQVNIVKKTEVQTG GFSKESILSKRESAKLIPRKKGWDTRKYGGFGSPTVAYSILVVAKVEKGKAKKL KSVKVLVGITIMEKGSYEKDPIGFLEAKGYKDIKKELIFKLPKYSLFELENGRRR MLASAKELQKANELVLPQHLVRLLYYTQNISATTGSNNLGYIEQHREEFKEIF EKIIDFSEKYILKNKVNSNLKSSFDEQFAVSDSILLSNSFVSLLKYTSFGASGGFT FLDLDVKQGRLRYQTVTEVLDATLIYQSITGLYETRTDLSQLGGD	N872A	H849A	D10A
SpyCas9-3var-NRRH	化脓性链球菌	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFL VEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAH MIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILS ARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLS KDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS MVKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEE FYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGIIPHQIHLGELHAILRRQ GDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA HLFDDKVMKQLKRLRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRN FMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV DELVKVMGGHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSF LKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPK LESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEI RKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKES ILPKGNSDKLIARKKDWDPKKYGGFNSPTAAYSVLVVAKVEKGKSKKLKSVK ELLGITIMERSSFEKNPIGFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS AGVLHKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDE IIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGVPAA FKYFDTTIDKKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	N863A	H840A	D10A
SpyCas9-3var-NRTH	化脓性链球菌	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFL VEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAH MIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILS	N863A	H840A	D10A

[0269]

		ARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLS KDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS MVKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEE FYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGIIPHQIHLGELHAILRRQ GDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA HLFDDKVMKQLKRLRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRN FMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV DELVKVMGGHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSF LKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPK LESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEI RKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKES ILPKGNSDKLIARKKDWDPKKYGGFNSPTVAYSVLVVAKVEKGKSKKLKSVK ELLGITIMERSSFEKNPIGFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS ASVLHKGNELALPSKYVNFLYLASHYEKLKGSSEDNKQKQLFVEQHKHYLDEI IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGASAAF KYFDTTIGRKLYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD			
SpyCas9-3var-NRCH	化脓性链球菌	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFL VEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAH MIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILS ARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLS KDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS MVKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEE FYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGIIPHQIHLGELHAILRRQ GDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA HLFDDKVMKQLKRLRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRN FMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV DELVKVMGGHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSF LKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPK LESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEI RKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKES ILPKGNSDKLIARKKDWDPKKYGGFNSPTVAYSVLVVAKVEKGKSKKLKSVK ELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS AGVLQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDE IIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAA FKYFDTTINRKQYNTTKEVLDATLIRQSITGLYETRIDLSQLGGD	N863A	H840A	D10A
SpyCas9-HF1	化脓性链球菌	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFL VEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAH MIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILS ARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLS KDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQ EDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANR NFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQ ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSF LKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPK LESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEI RKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKES ILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKE LLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASA GELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEII EQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAF	N863A	H840A	D10A

[0270]

		KYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD			
SpyCas9-QQR1	化脓性链球菌	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVLDATLIHQSITGLYETRIDLSQLGGD	N863A	H840A	D10A
SpyCas9-SpG	化脓性链球菌	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFLWPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAKQLQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVLDATLIHQSITGLYETRIDLSQLGGD	N863A	H840A	D10A
SpyCas9-VQR	化脓性链球菌	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVLDATLIHQSITGLYETRIDLSQLGGD	N863A	H840A	D10A

[0271]

		RKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKES ILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKE LLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASA GELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEII EQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAF KYFDTTIDRKQYRSTKEVLDATLIHQSITGLYETRIDLSQLGGD			
SpyCas9-VRER	化脓性链球菌	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFL VEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAH MIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILS ARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLS KDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQ EDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANR NFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQ ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSF LKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPK LESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEI RKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKES ILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKE LLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASA RELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEII EQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAF KYFDTTIDRKEYRSTKEVLDATLIHQSITGLYETRIDLSQLGGD	N863A	H840A	D10A
SpyCas9-xCas	化脓性链球菌	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFL VEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAH MIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILS ARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDTKLQLS KDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS MIKLYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGIIPHQIHLGELHAILRRQE DFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEK VVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE GMRKPAFLSGDQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANR NFIQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSF LKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPK LESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEI RKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKES ILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKE LLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASA GVLQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEII EQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAF KYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	N863A	H840A	D10A
SpyCas9-xCas-NG	化脓性链球菌	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFL VEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAH MIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILS ARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDTKLQLS KDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS MIKLYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGIIPHQIHLGELHAILRRQE DFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEK VVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE GMRKPAFLSGDQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANR NFIQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI	N863A	H840A	D10A

[0272]

		LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSF LKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPK LESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEI RKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKES IRPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKE LLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASA RFLQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEII EQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPRAF KYFDTTIDRKVYRSTKEVLDATLIHQSITGLYETRIDLSQLGGD			
St1Cas9-CNRZ1066	嗜热链球菌	MSDLVLGLDIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAENNLVRRTNRQG RRLARRKKHRRVRLNRLFEESGLITDFTKISINLNPYQLRVKGLTDELSNEELFI ALKNMVKHRGISYLDDASDDGNSSVGDYAQIVKENSKQLETKTPGQIQLER YQTYGQLRGDFTVEKDGKKHRLINVFPTSAYRSEALRILQTQQEFNPQITDEF INRYLEILTGKRKYYHGPGNEKSRTDYGRYRTSGETLDNIFGILIGKCTFYPDEF RAAKASYTAQEFNLLNDLNNLTVPTETKKLSKEQKNQIINYVKNEKAMGPAK LFKYIAKLLSCDVADIKGYRIDKSGKAEIHTFEAYRKMKTLETLDIEQMDRETL DKLAYVLTLNTEREGIQEALEHEFADGSFSQKQVDELVQFRKANSSIFGKGW HNFSVKLMMELIPELYETSEEQMTILTRLGKQKTTSSSNKTKYIDEKLLTEEIY NPVVAKSVRQAIKIVNAAIKEYGDFDNIVIEMARETNEDDEKKAIQKIQKAN KDEKDAAMLKAANQYNGKAELPHSVFHGHKQLATKIRLWHQQGERCLYT GKTISIHDLINNSNQFEVDHILPLSITFDDSLANKVLVYATANQEKGQRTPYQ ALDSMDDAWSFRELKAFVRESKTLSNKKKEYLLTEEDISKFDVRKKFIERNLV DTRYASRVVLNALQEHFRAHKIDTKVSVVRGQFTSQLRRHWGIEKTRDTYH HHAVDALIIAASSQLNLWKKQKNTLVSYSEEQLLDIETGELISDDEYKESVFKA PYQHFVDTLKSKEFEDSILFSYQVDSKFNRKISDATIYATRQAKVGKDKKDETY VLGKIKDIYTQDGYDAFMKIYKKDKSKFLMYRHDPQTFEKVIEPILENYPNKQ MNEKGKEVPCNPFLKYKEEHGYIRKYSKKGNGPEIKSLKYYDSKLLGNPIDITP ENSKNKVVLQSLKPWRTDVYFNKATGKYEILGLKYADLQFEKGTGTYKISQE KYNDIKKKEGVDSDSEFKFTLYKNDLLLVKDTETKEQQLFRFLSRTLPKQKHY VELKPYDKQKFEGGEALIKVLGNVANGGQCIKGLAKSNISIYKVRTDVLGNQ HIIKNEGDKPKLDF	N622A	H599A	D9A
St1Cas9-LMG1831	嗜热链球菌	MSDLVLGLDIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAENNLVRRTNRQG RRLARRKKHRRVRLNRLFEESGLITDFTKISINLNPYQLRVKGLTDELSNEELFI ALKNMVKHRGISYLDDASDDGNSSVGDYAQIVKENSKQLETKTPGQIQLER YQTYGQLRGDFTVEKDGKKHRLINVFPTSAYRSEALRILQTQQEFNPQITDEF INRYLEILTGKRKYYHGPGNEKSRTDYGRYRTSGETLDNIFGILIGKCTFYPDEF RAAKASYTAQEFNLLNDLNNLTVPTETKKLSKEQKNQIINYVKNEKAMGPAK LFKYIAKLLSCDVADIKGYRIDKSGKAEIHTFEAYRKMKTLETLDIEQMDRETL DKLAYVLTLNTEREGIQEALEHEFADGSFSQKQVDELVQFRKANSSIFGKGW HNFSVKLMMELIPELYETSEEQMTILTRLGKQKTTSSSNKTKYIDEKLLTEEIY NPVVAKSVRQAIKIVNAAIKEYGDFDNIVIEMARETNEDDEKKAIQKIQKAN KDEKDAAMLKAANQYNGKAELPHSVFHGHKQLATKIRLWHQQGERCLYT GKTISIHDLINNSNQFEVDHILPLSITFDDSLANKVLVYATANQEKGQRTPYQ ALDSMDDAWSFRELKAFVRESKTLSNKKKEYLLTEEDISKFDVRKKFIERNLV DTRYASRVVLNALQEHFRAHKIDTKVSVVRGQFTSQLRRHWGIEKTRDTYH HHAVDALIIAASSQLNLWKKQKNTLVSYSEEQLLDIETGELISDDEYKESVFKA PYQHFVDTLKSKEFEDSILFSYQVDSKFNRKISDATIYATRQAKVGKDKKDETY VLGKIKDIYTQDGYDAFMKIYKKDKSKFLMYRHDPQTFEKVIEPILENYPNKQ MNEKGKEVPCNPFLKYKEEHGYIRKYSKKGNGPEIKSLKYYDSKLLGNPIDITP ENSKNKVVLQSLKPWRTDVYFNKNTGKYEILGLKYADLQFEKKTGTYKISQE KYNGIMKEEGVDSDSEFKFTLYKNDLLLVKDTETKEQQLFRFLSRTMPNVKY YVELKPYSKDKFEKNESLIEILGSADKSGRCIKGLGKSNISIYKVRTDVLGNQHII KNEGDKPKLDF	N622A	H599A	D9A
St1Cas9-MTH17CL396	嗜热链球菌	MSDLVLGLDIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAENNLVRRTNRQG RRLARRKKHRRVRLNRLFEESGLITDFTKISINLNPYQLRVKGLTDELSNEELFI ALKNMVKHRGISYLDDASDDGNSSVGDYAQIVKENSKQLETKTPGQIQLER YQTYGQLRGDFTVEKDGKKHRLINVFPTSAYRSEALRILQTQQEFNPQITDEF INRYLEILTGKRKYYHGPGNEKSRTDYGRYRTSGETLDNIFGILIGKCTFYPDEF RAAKASYTAQEFNLLNDLNNLTVPTETKKLSKEQKNQIINYVKNEKAMGPAK LFKYIAKLLSCDVADIKGYRIDKSGKAEIHTFEAYRKMKTLETLDIEQMDRETL DKLAYVLTLNTEREGIQEALEHEFADGSFSQKQVDELVQFRKANSSIFGKGW HNFSVKLMMELIPELYETSEEQMTILTRLGKQKTTSSSNKTKYIDEKLLTEEIY NPVVAKSVRQAIKIVNAAIKEYGDFDNIVIEMARETNEDDEKKAIQKIQKAN KDEKDAAMLKAANQYNGKAELPHSVFHGHKQLATKIRLWHQQGERCLYT GKTISIHDLINNSNQFEVDHILPLSITFDDSLANKVLVYATANQEKGQRTPYQ ALDSMDDAWSFRELKAFVRESKTLSNKKKEYLLTEEDISKFDVRKKFIERNLV DTRYASRVVLNALQEHFRAHKIDTKVSVVRGQFTSQLRRHWGIEKTRDTYH HHAVDALIIAASSQLNLWKKQKNTLVSYSEDQLLDIETGELISDDEYKESVFKA PYQHFVDTLKSKEFEDSILFSYQVDSKFNRKISDATIYATRQAKVGKDKADET YVLGKIKDIYTQDGYDAFMKIYKKDKSKFLMYRHDPQTFEKVIEPILENYPNK QINEKGKEVPCNPFLKYKEEHGYIRKYSKKGNGPEIKSLKYYDSKLGNHIDITP	N622A	H599A	D9A

[0273]

		KDSNNKVVLQSLKPWRTDVYFNKNTGKYEILGLKYSDMQFEKGTGKYSISKE QYENIKVREGVDENSEFKFTLYKNDLLLLKDSENGEQILLRFTSRNDTSKHYVE LKPYNRQKFEGSEYLIKSLGTVAKGGQCIKGLGKSNISIYKVRTDVLGNQHIIK NEGDKPKLDF			
St1Cas9-TH 1477	嗜热链球菌	MSDLVLGLDIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAENNLVRRTNRQG RRLARRKKHRRVRLNRLFEESGLITDFTKISINLNPYQLRVKGLTDELSNEELFI ALKNMVKHRGISYLDDASDDGNSSVGDYAQIVKENSKQLETKTPGQIQLER YQTYGQLRGDFTVEKDGKKHRLINVFPTSAYRSEALRILQTQQEFNPQITDEF INRYLEILTGKRKYYHGPGNEKSRTDYGRYRTSGETLDNIFGILIGKCTFYPDEF RAAKASYTAQEFNLLNDLNNLTVPTETKKLSKEQKNQIINYVKNEKAMGPAK LFKYIAKLLSCDVADIKGYRIDKSGKAEIHTFEAYRKMKTLETLDIEQMDRETL DKLAYVLTLNTEREGIQEALEHEFADGSFSQKQVDELVQFRKANSSIFGKGW HNFSVKLMMELIPELYETSEEQMTILTRLGKQKTTSSSNKTKYIDEKLLTEEIY NPVVAKSVRQAIKIVNAAIKEYGDFDNIVIEMARETNEDDEKKAIQKIQKAN KDEKDAAMLKAANQYNGKAELPHSVFHGHKQLATKIRLWHQQGERCLYT GKTISIHDLINNSNQFEVDHILPLSITFDDSLANKVLVYATANQEKGQRTPYQ ALDSMDDAWSFRELKAFVRESKTLSNKKKEYLLTEEDISKFDVRKKFIERNLV DTRYASRVVLNALQEHFRAHKIDTKVSVVRGQFTSQLRRHWGIEKTRDTYH HHAVDALIIAASSQLNLWKKQKNTLVSYSEDQLLDIETGELISDDEYKESVFKA PYQHFVDTLKSKEFEDSILFSYQVDSKFNRKISDATIYATRQAKVGKDKADET YVLGKIKDIYTQDGYDAFMKIYKKDKSKFLMYRHDPQTFEKVIEPILENYPNK QINEKGKEVPCNPFLKYKEEHGYIRKYSKKGNGPEIKSLKYYDSKLGNHIDITP KDSNNKVVLQSLKPWRTDVYFNKNTGKYEILGLKYSDMQFEKGTGKYSISKE QYENIKVREGVDENSEFKFTLYKNDLLLLKDSENGEQILLRFTSRNDTSKHYVE LKPYNRQKFEGSEYLIKSLGTVVKGGRCIKGLGKSNISIYKVRTDVLGNQHIIK NEGDKPKLDF	N622A	H599A	D9A
sRGN3.1	葡萄球菌属物种	MNQKFILGLDIGITSVGYGLIDYETKNIIDAGVRLFPEANVENNEGRRSKRGS RRLKRRRIHRLERVKLLLTEYDLINKEQIPTSNNPYQIRVKGLSEILSKDELAIAL LHLAKRRGIHNVDVAADKEETASDSLSTKDQINKNAKFLESRYVCELQKERLE NEGHVRGVENRFLTKDIVREAKKIIDTQMQYYPEIDETFKEKYISLVETRREYF EGPGQGSPFGWNGDLKKWYEMLMGHCTYFPQELRSVKYAYSADLFNALN DLNNLIIQRDNSEKLEYHEKYHIIENVFKQKKKPTLKQIAKEIGVNPEDIKGYRI TKSGTPEFTSFKLFHDLKKVVKDHAILDDIDLLNQIAEILTIYQDKDSIVAELGQ LEYLMSEADKQSISELTGYTGTHSLSLKCMNMIIDELWHSSMNQMEVFTYL NMRPKKYELKGYQRIPTDMIDDAILSPVVKRTFIQSINVINKVIEKYGIPEDIIIE LARENNSDDRKKFINNLQKKNEATRKRINEIIGQTGNQNAKRIVEKIRLHDQ QEGKCLYSLESIPLEDLLNNPNHYEVDHIIPRSVSFDNSYHNKVLVKQSENSK KSNLTPYQYFNSGKSKLSYNQFKQHILNLSKSQDRISKKKKEYLLEERDINKFE VQKEFINRNLVDTRYATRELTNYLKAYFSANNMNVKVKTINGSFTDYLRKV WKFKKERNHGYKHHAEDALIIANADFLFKENKKLKAVNSVLEKPEIETKQLDI QVDSEDNYSEMFIIPKQVQDIKDFRNFKYSHRVDKKPNRQLINDTLYSTRKK DNSTYIVQTIKDIYAKDNTTLKKQFDKSPEKFLMYQHDPRTFEKLEVIMKQYA NEKNPLAKYHEETGEYLTKYSKKNNGPIVKSLKYIGNKLGSHLDVTHQFKSST KKLVKLSIKNYRFDVYLTEKGYKFVTIAYLNVFKKDNYYYIPKDKYQELKEKKKI KDTDQFIASFYKNDLIKLNGDLYKIIGVNSDDRNIIELDYYDIKYKDYCEINNIK GEPRIKKTIGKKTESIEKFTTDVLGNLYLHSTEKAPQLIFKRGL	N585A	H562A	D10A
sRGN3.3	葡萄球菌属物种	MNQKFILGLDIGITSVGYGLIDYETKNIIDAGVRLFPEANVENNEGRRSKRGS RRLKRRRIHRLERVKLLLTEYDLINKEQIPTSNNPYQIRVKGLSEILSKDELAIAL LHLAKRRGIHNVDVAADKEETASDSLSTKDQINKNAKFLESRYVCELQKERLE NEGHVRGVENRFLTKDIVREAKKIIDTQMQYYPEIDETFKEKYISLVETRREYF EGPGQGSPFGWNGDLKKWYEMLMGHCTYFPQELRSVKYAYSADLFNALN DLNNLIIQRDNSEKLEYHEKYHIIENVFKQKKKPTLKQIAKEIGVNPEDIKGYRI TKSGTPEFTSFKLFHDLKKVVKDHAILDDIDLLNQIAEILTIYQDKDSIVAELGQ LEYLMSEADKQSISELTGYTGTHSLSLKCMNMIIDELWHSSMNQMEVFTYL NMRPKKYELKGYQRIPTDMIDDAILSPVVKRTFIQSINVINKVIEKYGIPEDIIIE LARENNSDDRKKFINNLQKKNEATRKRINEIIGQTGNQNAKRIVEKIRLHDQ QEGKCLYSLESIPLEDLLNNPNHYEVDHIIPRSVSFDNSYHNKVLVKQSENSK KSNLTPYQYFNSGKSKLSYNQFKQHILNLSKSQDRISKKKKEYLLEERDINKFE VQKEFINRNLVDTRYATRELTSYLKAYFSANNMDVKVKTINGSFTNHLRKV WRFDKYRNHGYKHHAEDALIIANADFLFKENKKLQNTNKILEKPTIENNTKK VTVEKEEDYNNVFETPKLVEDIKQYRDYKFSHRVDKKPNRQLINDTLYSTRM KDEHDYIVQTITDIYGKDNTNLKKQFNKNPEKFLMYQNDPKTFEKLSIIMKQ YSDEKNPLAKYYEETGEYLTKYSKKNNGPIVKKIKLLGNKVGNHLDVTNKYEN STKKLVKLSIKNYRFDVYLTEKGYKFVTIAYLNVFKKDNYYYIPKDKYQELKEKK KIKDTDQFIASFYKNDLIKLNGDLYKIIGVNSDDRNIIELDYYDIKYKDYCEINNI KGEPRIKKTIGKKTESIEKFTTDVLGNLYLHSTEKAPQLIFKRGL	N585A	H562A	D10A

[0274] 在一些实施例中,Cas蛋白需要原间隔子邻近基序(PAM)存在于靶DNA序列中或邻近靶DNA序列,以便Cas蛋白结合和/或发挥功能。在一些实施例中,PAM是或包含从5′至3′的NGG、YG、NNGRRT、NNNRRT、NGA、TYCV、TATV、NTTN或NNNGATT,其中N代表任何核苷酸,Y代表C或T,R代表A或G,并且V代表A或C或G。在一些实施例中,Cas蛋白是表3或4中列出的蛋白。在一些实施例中,Cas蛋白包含一个或多个改变其PAM的突变。在一些实施例中,Cas蛋白包含E1369R、E1449H和R1556A突变或对应于所述位置的氨基酸的类似取代。在一些实施例中,

Cas蛋白包含E782K、N968K和R1015H突变或对应于所述位置的氨基酸的类似取代。在一些实施例中,Cas蛋白包含D1135V、R1335Q和T1337R突变或对应于所述位置的氨基酸的类似取代。在一些实施例中,Cas蛋白包含S542R和K607R突变或对应于所述位置的氨基酸的类似取代。在一些实施例中,Cas蛋白包含S542R、K548V和N552R突变或对应于所述位置的氨基酸的类似取代。工程改造Cas酶以识别改变的PAM序列的示例性进展综述于Collias等人Nature Communications[自然通讯]12:555(2021),其通过援引以其全文并入本文。

[0275] 在一些实施例中,Cas蛋白具有催化活性并切割靶DNA位点的一条或两条链。在一些实施例中,切割靶DNA位点之后形成改变,例如插入或缺失,例如通过细胞修复机制。

[0276] 在一些实施例中,Cas蛋白被修饰以失活或部分失活核酸酶,例如,核酸酶缺陷型Cas9。而在由gRNA靶向的特异性DNA序列上,野生型Cas9产生双链断裂(DSB),具有修饰的功能性的许多CRISPR核酸内切酶是可得的,例如:部分失活的Cas9“切口酶”版本仅产生单链断裂;无催化活性的Cas9(“dCas9”)不会切割靶DNA。在一些实施例中,dCas9与DNA序列的结合可以通过空间位阻干扰该位点处的转录。在一些实施例中,dCas9与锚定序列的结合可以干扰(例如,减少或阻止)基因组复合物(例如,ASMC)的形成和/或维持。在一些实施例中,DNA结合结构域包含无催化活性的Cas9,例如dCas9。许多无催化活性的Cas9蛋白是本领域已知的。在一些实施例中,dCas9包含Cas蛋白的每个核酸内切酶结构域中的突变,例如D10A和H840A或N863A突变。在一些实施例中,无催化活性或部分无催化活性的CRISPR/Cas结构域包含Cas蛋白,该Cas蛋白包含一个或多个突变,例如表3中列出的一个或多个突变。在一些实施例中,在表3的给定行中描述的Cas蛋白包含在表3的同一行中列出的突变中的一个、两个、三个或所有。在一些实施例中,例如未在表3中描述的Cas蛋白包含在表3的行中列出的突变中的一个、两个、三个或所有或在该Cas蛋白中相应位点处的相应突变。

[0277] 在一些实施例中,无催化活性的例如dCas9或部分失活的Cas9蛋白包含D11突变(例如D11A突变)或对应于所述位置的氨基酸的类似取代。在一些实施例中,无催化活性的Cas9蛋白例如dCas9或部分失活的Cas9蛋白包含H969突变(例如H969A突变)或对应于所述位置的氨基酸的类似取代。在一些实施例中,无催化活性的Cas9蛋白例如dCas9或部分失活的Cas9蛋白包含N995突变(例如N995A突变)或对应于所述位置的氨基酸的类似取代。在一些实施例中,无催化活性的Cas9蛋白例如dCas9包含在位置D11、H969和N995中的一个、两个或三个处的突变(例如,D11A、H969A和N995A突变)或对应于所述位置的氨基酸的类似取代。

[0278] 在一些实施例中,无催化活性的Cas9蛋白例如dCas9或部分失活的Cas9蛋白包含D10突变(例如D10A突变)或对应于所述位置的氨基酸的类似取代。在一些实施例中,无催化活性的Cas9蛋白例如dCas9或部分失活的Cas9蛋白包含H557突变(例如H557A突变)或对应于所述位置的氨基酸的类似取代。在一些实施例中,无催化活性的Cas9蛋白例如dCas9包含D10突变(例如,D10A突变)和H557突变(例如,H557A突变)或对应于所述位置的氨基酸的类似取代。

[0279] 在一些实施例中,无催化活性的Cas9蛋白例如dCas9或部分失活的Cas9蛋白包含D839突变(例如D839A突变)或对应于所述位置的氨基酸的类似取代。在一些实施例中,无催化活性的Cas9蛋白例如dCas9或部分失活的Cas9蛋白包含H840突变(例如H840A突变)或对应于所述位置的氨基酸的类似取代。在一些实施例中,无催化活性的Cas9蛋白例如dCas9或部分失活的Cas9蛋白包含N863突变(例如N863A突变)或对应于所述位置的氨基酸的类似取

代。在一些实施例中,无催化活性的Cas9蛋白例如dCas9包含D10突变(例如D10A)、D839突变(例如D839A)、H840突变(例如H840A)和N863突变(例如N863A)或对应于所述位置的氨基酸的类似取代。

[0280] 在一些实施例中,无催化活性的Cas9蛋白例如dCas9或部分失活的Cas9蛋白包含E993突变(例如E993A突变)或对应于所述位置的氨基酸的类似取代。

[0281] 在一些实施例中,无催化活性的Cas9蛋白例如dCas9或部分失活的Cas9蛋白包含D917突变(例如D917A突变)或对应于所述位置的氨基酸的类似取代。在一些实施例中,无催化活性的Cas9蛋白例如dCas9或部分失活的Cas9蛋白包含E1006突变(例如E1006A突变)或对应于所述位置的氨基酸的类似取代。在一些实施例中,无催化活性的Cas9蛋白例如dCas9或部分失活的Cas9蛋白包含D1255突变(例如D1255A突变)或对应于所述位置的氨基酸的类似取代。在一些实施例中,无催化活性的Cas9蛋白例如dCas9包含D917突变(例如D917A)、E1006突变(例如E1006A)和D1255突变(例如D1255A)或对应于所述位置的氨基酸的类似取代。

[0282] 在一些实施例中,无催化活性的Cas9蛋白例如dCas9或部分失活的Cas9蛋白包含D16突变(例如D16A突变)或对应于所述位置的氨基酸的类似取代。在一些实施例中,无催化活性的Cas9蛋白例如dCas9或部分失活的Cas9蛋白包含D587突变(例如D587A突变)或对应于所述位置的氨基酸的类似取代。在一些实施例中,部分失活的Cas结构域具有切口酶活性。在一些实施例中,部分失活的Cas9结构域是Cas9切口酶结构域。在一些实施例中,无催化活性的Cas结构域或失活Cas结构域不产生可检测的双链断裂形成。在一些实施例中,无催化活性的Cas9蛋白例如dCas9或部分失活的Cas9蛋白包含H588突变(例如H588A突变)或对应于所述位置的氨基酸的类似取代。在一些实施例中,无催化活性的Cas9蛋白例如dCas9或部分失活的Cas9蛋白包含N611突变(例如N611A突变)或对应于所述位置的氨基酸的类似取代。在一些实施例中,无催化活性的Cas9蛋白例如dCas9包含D16突变(例如D16A)、D587突变(例如D587A)、H588突变(例如H588A)和N611突变(例如N611A)或对应于所述位置的氨基酸的类似取代。

[0283] 在一些实施例中,DNA结合结构域或核酸内切酶结构域可以包含Cas分子,该Cas分子包含或连接(例如,共价地)gRNA(例如,模板核酸,例如,包含gRNA的模板RNA)。

[0284] 在一些实施例中,核酸内切酶结构域或DNA结合结构域包含化脓性链球菌Cas9(SpCas9)或其功能片段或变体。在一些实施例中,核酸内切酶结构域或DNA结合结构域包含经修饰的SpCas9。在实施例中,经修饰的SpCas9包含改变了原间隔子邻近基序(PAM)特异性的修饰。在实施例中,PAM对核酸序列5′-NGT-3′具有特异性。在实施例中,经修饰的SpCas9包含例如在位置L1111、D1135、G1218、E1219、A1322、或R1335中的一个或多个处的一个或多个氨基酸取代,例如,该一个或多个氨基酸取代选自L1111R、D1135V、G1218R、E1219F、A1322R、R1335V。在实施例中,经修饰的SpCas9包含氨基酸取代T1337R和一个或多个另外的氨基酸取代,例如,该一个或多个另外的氨基酸取代选自L1111、D1135L、S1136R、G1218S、E1219V、D1332A、D1332S、D1332T、D1332V、D1332L、D1332K、D1332R、R1335Q、T1337、T1337L、T1337Q、T1337I、T1337V、T1337F、T1337S、T1337N、T1337K、T1337H、T1337Q、和T1337M,或其对应的氨基酸取代。在实施例中,经修饰的SpCas9包含:(i)一个或多个氨基酸取代,其选自D1135L、S1136R、G1218S、E1219V、A1322R、R1335Q、和T1337;以及(ii)一个或多个氨基酸取

代,其选自L1111R、G1218R、E1219F、D1332A、D1332S、D1332T、D1332V、D1332L、D1332K、D1332R、T1337L、T1337I、T1337V、T1337F、T1337S、T1337N、T1337K、T1337R、T1337H、T1337Q、和T1337M,或其对应的氨基酸取代。

[0285] 在一些实施例中,核酸内切酶结构域或DNA结合结构域包含Cas结构域,例如Cas9结构域。在实施例中,核酸内切酶结构域或DNA结合结构域包含核酸酶活性Cas结构域、Cas切口酶(nCas)结构域或无核酸酶活性Cas(dCas)结构域。在实施例中,核酸内切酶结构域或DNA结合结构域包含核酸酶活性Cas9结构域、Cas9切口酶(nCas9)结构域或无核酸酶活性Cas9(dCas9)结构域。在一些实施例中,核酸内切酶结构域或DNA结合结构域包含Cas9的结构域Cas9(例如,dCas9和nCas9)、Cas12a/Cpf1、Cas12b/C2c1、Cas12c/C2c3、Cas12d/CasY、Cas12e/CasX、Cas12g、Cas12h、或Cas12i。在一些实施例中,核酸内切酶结构域或DNA结合结构域包含Cas9(例如,dCas9和nCas9)、Cas12a/Cpf1、Cas12b/C2c1、Cas12c/C2c3、Cas12d/CasY、Cas12e/CasX、Cas12g、Cas12h、或Cas12i。在一些实施例中,核酸内切酶结构域或DNA结合结构域包含化脓性链球菌或嗜热链球菌Cas9,或其功能片段。在一些实施例中,核酸内切酶结构域或DNA结合结构域包含Cas9序列,例如,如Chylinski、Rhun,和Charpentier(2013)RNA Biology[RNA生物学]10:5、726-737中所述;该文献通过援引并入本文。在一些实施例中,核酸内切酶结构域或DNA结合结构域包含Cas的HNH核酸酶亚结构域和/或RuvC1亚结构域,例如,如本文所述的Cas9,或其变体。在一些实施例中,核酸内切酶结构域或DNA结合结构域包含Cas12a/Cpf1、Cas12b/C2c1、Cas12c/C2c3、Cas12d/CasY、Cas12e/CasX、Cas12g、Cas12h、或Cas12i。在一些实施例中,核酸内切酶结构域或DNA结合结构域包含Cas多肽(例如酶)或其功能片段。在实施例中,Cas多肽(例如,酶)选自Cas1、Cas1B、Cas2、Cas3、Cas4、Cas5、Cas5d、Cas5t、Cas5h、Cas5a、Cas6、Cas7、Cas8、Cas8a、Cas8b、Cas8c、Cas9(例如,Csn1或Csx12)、Cas10、Cas10d、Cas12a/Cpf1、Cas12b/C2c1、Cas12c/C2c3、Cas12d/CasY、Cas12e/CasX、Cas12g、Cas12h、Cas12i、Csy1、Csy2、Csy3、Csy4、Cse1、Cse2、Cse3、Cse4、Cse5e、Csc1、Csc2、Csa5、Csn1、Csn2、Csm1、Csm2、Csm3、Csm4、Csm5、Csm6、Cmr1、Cmr3、Cmr4、Cmr5、Cmr6、Csb1、Csb2、Csb3、Csx17、Csx14、Csx10、Csx16、CsaX、Csx3、Csx1、Csx1S、Csx11、Csf1、Csf2、Cs0、Csf4、Csd1、Csd2、Cst1、Cst2、Csh1、Csh2、Csa1、Csa2、Csa3、Csa4、Csa5、II型Cas效应子蛋白、V型Cas效应子蛋白、VI型Cas效应子蛋白、CARF、DinG、Cpf1、Cas12b/C2c1、Cas12c/C2c3、Cas12b/C2c1、Cas12c/C2c3、SpCas9(K855A)、eSpCas9(1.1)、SpCas9-HF1、超精确的Cas9变体(HypaCas9)、其同源物、其经修饰的或经工程改造的版本、和/或其功能片段。在实施例中,Cas9包含例如选自H840A、D10A、P475A、W476A、N477A、D1125A、W1126A、和D1127A的一个或多个取代。在实施例中,Cas9包含在选自以下的位置处的一个或多个突变:D10、G12、G17、E762、H840、N854、N863、H982、H983、A984、D986、和/或A987,例如,选自D10A、G12A、G17A、E762A、H840A、N854A、N863A、H982A、H983A、A984A、和/或D986A的一个或多个取代。在一些实施例中,核酸内切酶结构域或DNA结合结构域包含来自以下的Cas(例如,Cas9)序列或其片段或变体:溃疡棒状杆菌(Corynebacterium ulcerans)、白喉棒状杆菌、梅毒螺原体(Spiroplasma syrphidicola)、中间普雷沃氏菌(Prevotella intermedia)、台湾螺原体(Spiroplasma taiwanense)、海豚链球菌(Streptococcus iniae)、波罗的海贝尔氏菌(Belliella baltica)、扭曲冷弯曲菌(Psychroflexus torquis)、嗜热链球菌、无害李斯特菌(Listeria innocua)、空肠弯曲杆菌、脑膜炎奈瑟球

菌、化脓性链球菌或金黄色葡萄球菌。

[0286] 在一些实施例中,核酸内切酶结构域或DNA结合结构域包含例如包含一个或多个取代(例如,在位置D917、E1006A、D1255处)或其任何组合的Cpf1结构域,该一个或多个取代例如选自D917A、E1006A、D1255A、D917A/E1006A、D917A/D1255A、E1006A/D1255A、和D917A/E1006A/D1255A。

[0287] 在一些实施例中,核酸内切酶结构域或DNA结合结构域包含spCas9、spCas9-VRQR(SEQ ID NO:4019)、spCas9-VRER(SEQ ID NO:4020)、xCas9(sp)、saCas9、saCas9-KKH、spCas9-MQKSER(SEQ ID NO:4021)、spCas9-LRKIQK(SEQ ID NO:4022)或spCas9-LRVSQL(SEQ ID NO:4023)。

[0288] 在一些实施例中,基因修饰多肽具有包含Cas9切口酶例如Cas9 H840A的核酸内切酶结构域。在实施例中,Cas9 H840A具有以下氨基酸序列:Cas9切口酶(H840A):

[0289] DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR
RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTD
KADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLEN
LIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAIL
LSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKP
ILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPL
ARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE
GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDN
EENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSD
GFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEM
ARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDA
IVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFI
KRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVG
TALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGE
IVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAK
VEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNE
LALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPI
REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[0290] 在一些实施例中,基因修饰多肽包含含有D10A和/或H840A突变的dCas9序列,例如,以下序列:

[0291] SMDKKYSIGLAIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRT
ARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLV
DSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKS
RRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKN
LSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEE
FYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFR
IPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNE
LTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLK

IIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS GKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVM GRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQE LDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKA ERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREIN NYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANG EIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKY GGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELE NGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILA DANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETR IDLSQLGGD(SEQ ID NO:4007)

[0292] TAL效应子和锌指核酸酶

[0293] 在一些实施例中,核酸内切酶结构域或DNA结合结构域包含TAL效应子分子。TAL效应子分子,例如特异性结合DNA序列的TAL效应子分子,通常包含多个TAL效应子结构域或其片段,以及任选地天然存在的TAL效应子的一个或多个附加部分(例如,多个TAL效应子结构域的N和/或C末端)。许多TAL效应子是本领域技术人员已知的并且是可商购的,例如从赛默飞世尔科技公司(Thermo Fisher Scientific)商购。

[0294] 天然存在的TALE是由多种细菌病原体(包括植物病原体黄单胞菌属(Xanthomonas))分泌的天然效应子蛋白,其调节宿主植物中的基因表达并促进细菌定植和存活。TAL效应子的特异性结合基于串联排列的几乎相同的典型33或34个氨基酸重复序列的中心重复结构域(重复可变二残基,RVD结构域)。

[0295] TAL效应子家族的成员主要在其重复序列的数量和顺序上不同。重复序列的数量范围通常为1.5至33.5个重复,并且C末端重复通常长度较短(例如,约20个氨基酸),并且通常被称为“半重复”。TAL效应子的每个重复序列通常具有一个重复序列对一个碱基对的相关性,其中不同的重复序列类型表现出不同的碱基对特异性(一个重复序列识别靶基因序列上的一个碱基对)。通常,重复序列数量越少,蛋白质-DNA相互作用越弱。已证明6.5个重复序列的数量足以激活报告基因的转录(Scholze等人,2010)。

[0296] 重复序列至重复序列的变异主要发生在氨基酸位置12和13处,因此它们被称为“高变的”,并负责与靶DNA启动子序列相互作用的特异性,如表5所示,其列出了示例性重复序列可变双残基(RVD)及其与核酸碱基靶标的对应关系。

[0297] 表5-RVD和核酸碱基特异性

[0298]

靶标	可能的RVD氨基酸组合												
A	NI	NN	CI	HI	KI								
G	NN	GN	SN	VN	LN	DN	QN	EN	HN	RH	NK	AN	FN
C	HD	RD	KD	ND	AD								
T	NG	HG	VG	IG	EG	MG	YG	AA	EP	VA	QG	KG	RG

[0299] 因此,有可能修饰TAL效应子的重复序列以靶向特定的DNA序列。进一步的研究表

明,RVD NK可以靶向G。TAL效应子的靶位点也倾向于包括在被第一重复序列靶向的5'碱基侧翼的T,但这种识别的确切机制尚不清楚。迄今已知超过113种TAL效应子序列。来自黄单胞菌属的TAL效应子的非限制性实例包括Hax2、Hax3、Hax4、AvrXa7、AvrXa10和AvrBs3。

[0300] 相应地,本文所述的TAL效应子分子的TAL效应子结构域可以源自来自任何细菌物种(例如黄单胞菌属物种,例如米糠黄单胞菌(Xanthomonas oryzae pv.Oryzae)的非洲菌株(Yu等人2011)、野油菜黄单胞菌萝卜致病变种(Xanthomonas campestris pv.raphani)菌株756C和水稻细菌性条斑病菌(Xanthomonas oryzae pv.oryzicola)菌株BLS256(Bogdanove等人2011))的TAL效应子。在一些实施例中,TAL效应子结构域包含RVD结构域以及也来自天然存在的TAL效应子的一个或多个侧翼序列(RVD结构域的N末端和/或C末端侧上的序列)。它可以包含比天然存在的TAL效应子的RVD更多或更少的重复序列。TAL效应子分子可以被设计成基于上述编码和本领域已知的其他编码来靶向给定的DNA序列。TAL效应子结构域(例如,重复序列(单体或模块))的数量及其特定序列可基于所期望的DNA靶序列来选择。例如,为了适应特定的靶序列,可以去除或添加TAL效应子结构域,例如重复序列。在一个实施例中,本发明的TAL效应子分子包含6.5至33.5个TAL效应子结构域,例如重复序列。在一个实施例中,本发明的TAL效应子分子包含8至33.5个TAL效应子结构域,例如重复序列,例如10至25个TAL效应子结构域,例如重复序列,例如10至14个TAL效应子结构域,例如重复序列。

[0301] 在一些实施例中,TAL效应子分子包含对应于与DNA靶序列完全匹配的TAL效应子结构域。在一些实施例中,允许DNA靶序列上的重复序列和靶碱基对之间的错配,只要它允许包含TAL效应子分子的多肽的功能。通常,TALE结合与错配数量呈负相关。在一些实施例中,本发明的多肽的TAL效应子分子与靶DNA序列包含不超过7个错配、6个错配、5个错配、4个错配、3个错配、2个错配或1个错配,并且任选地没有错配。不希望受理论束缚,一般来说,TAL效应子分子中TAL效应子结构域的数量越少,将被容许的错配数量就越少,并且仍然允许包含TAL效应子分子的多肽的功能。结合亲和力被认为取决于匹配的重复-DNA组合的总和。例如,具有25个或更多个TAL效应子结构域的TAL效应子分子可能能够耐受多达7个错配。

[0302] 除了TAL效应子结构域之外,本发明的TAL效应子分子还可以包含源自天然存在的TAL效应子的另外序列。包含在TAL效应子分子的TAL效应子结构域部分每一侧上的一个或多个C末端和/或N末端序列的长度可以变化,并且由本领域技术人员选择,例如基于Zhang等人(2011)的研究。Zhang等人已经表征了Hax3来源的基于TAL效应子的蛋白质中的许多C末端和N末端截短突变体,并且已经鉴定了有助于与靶序列最佳结合并因此激活转录的关键元件。通常,发现转录活性与N末端的长度呈负相关。关于C末端,鉴定了Hax 3序列前68个氨基酸内DNA结合残基的重要元件。因此,在一些实施例中,天然存在的TAL效应子的TAL效应子结构域的C末端侧上的前68个氨基酸包括在TAL效应子分子中。因此,在实施例中,TAL效应子分子包含1)一个或多个源自天然存在的TAL效应子的TAL效应子结构域;2)至少70、80、90、100、110、120、130、140、150、170、180、190、200、220、230、240、250、260、270、280个或更多个来自TAL效应子结构域N末端侧上的天然存在的TAL效应子的氨基酸;和/或3)至少68、80、90、100、110、120、130、140、150、170、180、190、200、220、230、240、250、260个或更多个来自TAL效应子结构域C末端侧上的天然存在的TAL效应子的氨基酸。

[0303] 在一些实施例中,核酸内切酶结构域或DNA结合结构域是或包含Zn指分子。Zn指分子包含Zn指蛋白,例如天然存在的Zn指蛋白或工程化的Zn指蛋白、或其片段。许多Zn指蛋白是本领域技术人员已知的并且是可商购的,例如从西格玛奥德里奇公司(Sigma-Aldrich)商购。

[0304] 在一些实施例中,Zn指分子包含非天然存在的Zn指蛋白,其被工程改造以与选择的靶DNA序列结合。例如,参见Beerli等人(2002)Nature Biotechnol.[自然生物技术]20:135-141;Pabo等人(2001)Ann.Rev.Biochem.[生物化学年度综述]70:313-340;Isalan等人(2001)Nature Biotechnol.[自然生物技术]19:656-660;Segal等人(2001)Curr.Opin.Biotechnol.[生物技术当前观点]12:632-637;Choo等人(2000)Curr.Opin.Struct.Biol.[当代结构生物学观点]10:411-416;美国专利号6,453,242、6,534,261、6,599,692、6,503,717、6,689,558、7,030,215、6,794,136、7,067,317、7,262,054、7,070,934、7,361,635、7,253,273;以及美国专利公开号2005/0064474、2007/0218528、2005/0267061,均通过援引以其全文并入本文。

[0305] 与天然存在的Zn指蛋白相比,工程改造的Zn指蛋白可能具有新型结合特异性。工程改造方法包括但不限于合理设计和各种类型的选择。合理设计包括,例如,使用包含三联体(或四联体)核苷酸序列和单个Zn指氨基酸序列的数据库,其中每个三联体或四联体核苷酸序列与结合特定三联体或四联体序列的锌指的一个或多个氨基酸序列相关联。参见例如,美国专利号6,453,242和6,534,261,其通过援引以其全文并入本文。

[0306] 示例性选择方法(包括噬菌体展示和双杂交系统)披露于以下中:美国专利号5,789,538、5,925,523、6,007,988、6,013,453、6,410,248、6,140,466、6,200,759、和6,242,568;以及国际专利公开号WO 98/37186、WO 98/53057、WO 00/27878、以及WO 01/88197和GB 2,338,237。另外,增强锌指蛋白的结合特异性已经例如,在国际专利公开号WO 02/077227中描述。

[0307] 另外,如这些和其他参考文献中所披露的,锌指结构域和/或多指锌指蛋白可以使用任何合适的接头序列(包括例如,长度为5个或更多个氨基酸的接头)连接在一起。另参见美国专利号6,479,626、6,903,185、和7,153,949的示例性接头序列长度为6个或更多个氨基酸。本文所述的蛋白质可以包括蛋白质的单个锌指之间的合适接头的任何组合。另外,增强锌指结合结构域的结合特异性已经例如,在共同拥有的国际专利公开号WO 02/077227中描述。

[0308] Zn指蛋白和用于设计和构建融合蛋白(和编码其的多核苷酸)的方法是本领域技术人员已知的,并在以下中详细描述:美国专利号6,140,0815、789,538、6,453,242、6,534,261、5,925,523、6,007,988、6,013,453、和6,200,759;国际专利公开号WO 95/19431、WO 96/06166、WO 98/53057、WO 98/54311、WO 00/27878、WO 01/60970、WO 01/88197、WO 02/099084、WO 98/53058、WO 98/53059、WO 98/53060、WO 02/016536、和WO 03/016496。

[0309] 另外,如这些和其他参考文献中所披露的,Zn指蛋白和/或多指Zn指蛋白可以使用任何合适的接头序列(包括例如,长度为5个或更多个氨基酸的接头)连接在一起,例如作为融合蛋白。另参见美国专利号6,479,626;6,903,185、和7,153,949的示例性接头序列长度为6个或更多个氨基酸。本文所述的Zn指分子可以包括Zn指分子的单个锌指蛋白和/或多指Zn指蛋白之间的合适接头的任何组合。

[0310] 在某些实施例中,DNA结合结构域或核酸内切酶结构域包含Zn指分子,该Zn指分子包含与靶DNA序列结合(以序列特异性方式)的工程化锌指蛋白。在一些实施例中,Zn指分子包含一种Zn指蛋白或其片段。在其他实施例中,Zn指分子包含多种Zn指蛋白(或其片段),例如2、3、4、5、6或更多种Zn指蛋白(并且任选地,不超过12、11、10、9、8、7、6、5、4、3或2种Zn指蛋白)。在一些实施例中,Zn指分子包含至少三种Zn指蛋白。在一些实施例中,Zn指分子包含四个、五个或六个指。在一些实施例中,Zn指分子包含8、9、10、11或12个指。在一些实施例中,包含三种Zn指蛋白的Zn指分子识别包含9或10个核苷酸的靶DNA序列。在一些实施例中,包含四种Zn指蛋白的Zn指分子识别包含12至14个核苷酸的靶DNA序列。在一些实施例中,包含六种Zn指蛋白的Zn指分子识别包含18至21个核苷酸的靶DNA序列。

[0311] 在一些实施例中,Zn指分子包含双手Zn指蛋白。双手锌指蛋白是这样的蛋白质,其中两簇锌指蛋白被间插氨基酸分开,使得两个锌指结构域与两个不连续的靶DNA序列结合。双手型锌指结合蛋白的实例是SIP1,其中四种锌指蛋白的簇位于蛋白质的氨基末端处,并且三种Zn指蛋白的簇位于羧基末端处(参见Remade等人(1999)EMBO Journal[欧洲分子生物学杂志]18(18):5073-5084)。这些蛋白质中的每一簇锌指均能够与独特的靶序列结合,并且这两个靶序列之间的间隔可以包含许多核苷酸。

[0312] 接头

[0313] 在一些实施例中,基因修饰多肽可以包含接头,例如肽接头,例如表6中描述的接头。在一些实施例中,基因修饰多肽在N末端至C末端方向上包含Cas结构域(例如,表3的Cas结构域)、表6的接头(或与其具有至少70%、80%、85%、90%、95%或99%同一性的序列)、以及RT结构域(例如,表2的RT结构域)。在一些实施例中,基因修饰多肽包含在核酸内切酶和RT结构域之间的柔性接头,例如,包含氨基酸序列SGGSSGGSSGSETPGTSESATPESSGGSSGGSS的接头。在一些实施例中,基因修饰多肽的RT结构域可以位于核酸内切酶结构域的C末端。在一些实施例中,基因修饰多肽的RT结构域可以位于核酸内切酶结构域的N末端。

[0314] 表6示例性接头序列

[0315]

氨基酸序列	SEQ ID NO
GGS	4101
GGSGGS	4102
GGSGGSGGS	4103
GGSGGSGGSGGS	4104
GGSGGSGGSGGSGGS	4105
GGSGGSGGSGGSGGSGGS	4106
GGGGS	4107
GGGGSGGGGS	4108
GGGGSGGGGSGGGGS	4109
GGGGSGGGGSGGGGSGGGGS	4110
GGGGSGGGGSGGGGSGGGGSGGGGS	4111
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	4112
GGG	4113
GGGG	4114
GGGGG	4115
GGGGGG	4116
GGGGGGG	4117
GGGGGGGG	4118
GSS	4119
GSSGSS	4120
GSSGSSGSS	4121
GSSGSSGSSGSS	4122
GSSGSSGSSGSSGSS	4123

[0316]

氨基酸序列	SEQ ID NO
GSSGSSGSSGSSGSSGSS	4124
EAAAK	4125
EAAAKEAAAK	4126
EAAAKEAAAKEAAAK	4127
EAAAKEAAAKEAAAKEAAAK	4128
EAAAKEAAAKEAAAKEAAAKEAAAK	4129
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	4130
PAP	4131
PAPAP	4132
PAPAPAP	4133
PAPAPAPAP	4134
PAPAPAPAPAP	4135
PAPAPAPAPAPAP	4136
GGSGGG	4137
GGGGGS	4138
GGSGSS	4139
GSSGGS	4140
GGSEAAAK	4141
EAAAKGGS	4142
GGSPAP	4143
PAPGGS	4144
GGGGSS	4145
GSSGGG	4146
GGGEAAAK	4147
EAAAKGGG	4148
GGGPAP	4149
PAPGGG	4150
GSSEAAAK	4151
EAAAKGSS	4152

[0317]

氨基酸序列	SEQ ID NO
GSSPAP	4153
PAPGSS	4154
EAAAKPAP	4155
PAPEAAAK	4156
GGSGGGGSS	4157
GGSGSSGGG	4158
GGGGGSGSS	4159
GGGGSSGGS	4160
GSSGGSGGG	4161
GSSGGGGGS	4162
GGSGGGEAAAK	4163
GGSEAAAKGGG	4164
GGGGGSEAAAK	4165
GGGEAAAKGGS	4166
EAAAKGGSGGG	4167
EAAAKGGGGGS	4168
GGSGGGPAP	4169
GGSPAPGGG	4170
GGGGGSPAP	4171
GGGPAPGGS	4172
PAPGGSGGG	4173
PAPGGGGGS	4174
GGSGSSEAAAK	4175
GGSEAAAKGSS	4176
GSSGGSEAAAK	4177
GSSEAAAKGGS	4178
EAAAKGGSGSS	4179
EAAAKGSSGGS	4180
GGSGSSPAP	4181

[0318]

氨基酸序列	SEQ ID NO
GGSPAPGSS	4182
GSSGGSPAP	4183
GSSPAPGGS	4184
PAPGGSGSS	4185
PAPGSSGGS	4186
GGSEAAAKPAP	4187
GGSPAPEAAAK	4188
EAAAKGGSPAP	4189
EAAAKPAPGGS	4190
PAPGGSEAAAK	4191
PAPEAAAKGGS	4192
GGGGSSEAAAK	4193
GGGEAAAKGSS	4194
GSSGGGEAAAK	4195
GSSEAAAKGGG	4196
EAAAKGGGGSS	4197
EAAAKGSSGGG	4198
GGGGSSPAP	4199
GGGPAPGSS	4200
GSSGGGPAP	4201
GSSPAPGGG	4202
PAPGGGGSS	4203
PAPGSSGGG	4204
GGGEAAAKPAP	4205
GGGPAPEAAAK	4206
EAAAKGGGPAP	4207
EAAAKPAPGGG	4208
PAPGGGEAAAK	4209
PAPEAAAKGGG	4210

[0319]

氨基酸序列	SEQ ID NO
GSSEAAAKPAP	4211
GSSPAPEAAAK	4212
EAAAKGSSPAP	4213
EAAAKPAPGSS	4214
PAPGSSEAAAK	4215
PAPEAAAKGSS	4216
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	4217
GGGGSEAAAKGGGGS	4218
EAAAKGGGGSEAAAK	4219
SGSETPGTSESATPES	4220
GSAGSAAGSGEF	4221
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	4222

[0320] 在一些实施例中,基因修饰多肽的接头包含选自以下的基序:$(SGGS)_n$(SEQ ID NO:4025)、$(GGGS)_n$(SEQ ID NO:4026)、$(GGGGS)_n$(SEQ ID NO:4027)、$(G)_n$、$(EAAAK)_n$(SEQ ID NO:4028)、$(GGS)_n$或$(XP)_n$。

[0321] 通过池化筛选选择基因修饰多肽

[0322] 可以筛选候选基因修饰多肽来评价候选物的基因编辑能力。例如,可以使用为靶向编辑人基因组中的编码序列而设计的RNA基因修饰系统。在某些实施例中,这样的基因修饰系统可以与池化筛选方法结合使用。

[0323] 例如,可以将基因修饰多肽候选物文库和模板指导RNA(tgRNA)引入哺乳动物细胞,以通过池化筛选方法测试候选物的基因编辑能力。在特别的实施例中,将基因修饰多肽候选物文库引入哺乳动物细胞,然后将tgRNA引入细胞。

[0324] 可用于筛选的哺乳动物细胞的代表性非限制性实例包括HEK293T细胞、U2OS细胞、HeLa细胞、HepG2细胞、Huh7细胞、K562细胞、或iPS细胞。

[0325] 基因修饰多肽候选物可以包含1)Cas核酸酶,例如野生型Cas核酸酶(例如野生型Cas9核酸酶)、突变Cas核酸酶(例如Cas切口酶,例如Cas9切口酶,例如Cas9 N863A切口酶)、或选自表3或表4的Cas核酸酶,2)肽接头,例如来自表D或表6的序列,其可能表现出不同程度的长度、柔性、疏水性和/或二级结构;和3)逆转录酶(RT),例如来自表D或表2的RT结构域。基因修饰多肽候选文库包含:多个不同的基因修饰多肽候选物,其在Cas核酸酶、肽接头或RT结构域组分中的一个、两个或全部三个方面彼此不同;或多个编码此类基因修饰多肽候选物的核酸表达载体。

[0326] 为了筛选基因修饰多肽候选物,可以使用包含基因修饰多肽组分和tgRNA组分的双组分系统。基因修饰组分可以包括例如表达载体,例如表达质粒或慢病毒载体,其编码基因修饰多肽候选物,例如包含人密码子优化的核酸,其编码基因修饰多肽候选物,例如如上所述的Cas-接头-RT融合体。在特定实施例中,利用慢病毒盒,其包含:(i)用于在哺乳动物

细胞中表达的启动子,例如CMV启动子;(ii)基因修饰文库候选物,例如包含表3或表4的Cas核酸酶、表6的肽接头和表2的RT的Cas-接头-RT融合体,例如如表D的Cas-接头-RT融合体;(iii)自切割多肽,例如T2A肽;(iv)能够在哺乳动物细胞中选择的标志物,例如嘌呤霉素抗性基因;和(v)终止信号,例如聚A尾。

[0327] tgRNA组分可以包含tgRNA或表达载体,例如表达质粒,其产生tgRNA,例如利用U6启动子来驱动tgRNA的表达,其中tgRNA是被Cas识别并将其定位至目的基因组基因座的非编码RNA序列,并且还通过RT结构域作为将所期望编辑逆转录到基因组中的模板。

[0328] 为了制备表达基因修饰多肽文库候选物的细胞池,可以用基因修饰候选多肽文库的池化基因修饰多肽候选表达载体制剂(例如慢病毒制剂)转导哺乳动物细胞(例如HEK293T或U2OS细胞)。在特定实施例中,利用慢病毒质粒,并将HEK293 Lenti-X细胞接种于15cm平板中(约12x 10^6个细胞)后再进行慢病毒质粒转染。在这样的实施例中,可以使用慢病毒包装混合物(博塞塔公司(Biosettia))进行慢病毒质粒转染,并根据制造商的方案在第二天使用Lipofectamine 2000和Opti-MEM培养基对基因修饰候选文库的质粒DNA进行转染。在这样的实施例中,可以通过第二天的完全培养基更换去除细胞外DNA,并且可以在48小时后收获含有病毒的培养基。慢病毒培养基可使用Lenti-X浓缩液(宝生物科学公司(TaKaRa Biosciences))浓缩,然后可制备5mL慢病毒等分试样并储存于-80℃。慢病毒滴度测定是通过在选择后(例如,在嘌呤霉素选择后)计数菌落形成单位来进行的。

[0329] 为了监测靶DNA的基因编辑,可以利用携带靶DNA的哺乳动物细胞,例如HEK293T或U2OS细胞。在监测靶DNA基因编辑的其他实施例中,可以利用携带靶DNA基因组着陆垫的哺乳动物细胞,例如HEK293T或U2OS细胞。在特定实施例中,靶DNA基因组着陆垫可以包含要编辑以治疗目的疾病或病症的基因。在其他特定实施例中,靶DNA是表达表现出可检测特征的蛋白质的基因序列,可以监测这些特征以确定是否发生了基因编辑。例如,在某些实施例中,利用表达蓝色荧光蛋白(BFP)或绿色荧光蛋白(GFP)的基因组着陆垫。在某些实施例中,将包含靶DNA(例如靶DNA基因组着陆垫)的哺乳动物细胞(例如HEK293T或U2OS细胞)以每基因修饰文库候选物500x-3000x个细胞接种在培养板中,并以0.2-0.3的感染复数(MOI)进行转导,以最大程度地减少每个细胞的多重感染。感染后48小时可加入嘌呤霉素(2.5ug/mL)以选择受感染的细胞。在这样的实施例中,细胞可在嘌呤霉素选择下保持至少7天,然后扩大规模以引入tgRNA,例如,tgRNA电穿孔。

[0330] 为了确定是否发生基因编辑,可以用基因修饰多肽文库候选物感染含有要编辑的靶DNA的哺乳动物细胞,然后用设计用于编辑靶DNA的tgRNA转染。随后,可以分析细胞以确定靶基因座的编辑是否根据设计的结果发生,或者是否无编辑发生或发生不完美编辑,例如通过使用细胞分选和序列分析。

[0331] 在特定实施例中,为了确定是否发生基因组编辑,可以用基因修饰文库候选物感染表达BFP或GFP的哺乳动物细胞(例如HEK293T或U2OS细胞),然后用tgRNA质粒或RNA转染或电穿孔,例如通过使用200ng tgRNA质粒对250,000个细胞/孔进行电穿孔,该质粒设计用于将BFP转化为GFP或GFP转化为BFP,其中细胞计数确保每个文库候选物的覆盖率>250x-1000x。在这样的实施例中,可以通过在电穿孔后4-10天针对颜色转化的荧光蛋白(FP)的表达用荧光激活细胞分选(FACS)对细胞进行分选来评估该测定中各种构建体的基因组编辑能力。对细胞进行分选和收获,分为未编辑细胞(显示原始荧光蛋白信号)、已编辑细胞(显

示转化的荧光蛋白信号)和不完美编辑细胞(不显示荧光蛋白信号)的不同群体。还可以收获未分选的细胞样品作为输入群体,以确定分析过程中的候选富集。

[0332] 为了确定哪些基因修饰文库候选物在测定中表现出基因组编辑能力,从分选的细胞群体中收获基因组DNA(gDNA),并通过对每个群体中的基因修饰文库候选物进行测序进行分析。简而言之,可以使用针对基因修饰多肽表达载体(例如慢病毒盒)具有特异性的引物从基因组中扩增基因修饰候选物,在第二轮PCR中扩增以稀释基因组DNA,然后进行测序,例如通过下一代测序平台进行测序。在对测序读段进行质量控制之后,将至少约1500个核苷酸并且通常不超过约3200个核苷酸的读段映射到基因修饰多肽文库序列,并且那些与文库序列至少约80%匹配的读段被认为已成功与给定候选物比对,以用于此池化筛选。为了识别能够在测定中进行基因编辑(例如,BFP到GFP、或GFP到BFP编辑)的候选物,将编辑群体中每个文库候选物的读取计数与初始未分选群体中的读取计数进行比较。

[0333] 为了进行池化筛选,具有基因组编辑能力的基因修饰候选物是根据编辑的(转化的FP)群体相对于未分选(输入)细胞的富集来识别的。在一些实施例中,相对于输入至少1.0、1.5、2.0、2.5、3.0、4.0、5.0、6.0、7.0、8.0、9.0、10、15、20、25、30、40、50、60、70、80、90或至少100倍的富集表明潜在有用的基因编辑活性,例如至少2倍的富集。在一些实施例中,通过对富集率取以2为底的对数,将富集转化为对数值。在一些实施例中,log2富集分数为至少0、1、2、3、4、5、5.5、6.0、6.2、6.3、6.4、6.5、或至少6.6表明潜在有用的基因编辑活性,例如log2富集分数为至少1.0。在特定实施例中,可以利用参考(例如元件ID号:17380)将观察到的基因修饰候选物的富集值与在类似条件下观察到的富集值进行比较。

[0334] 在一些实施例中,可以使用多个tgRNA来筛选基因修饰候选文库。在特定实施例中,可以利用多个tgRNA来优化模板/Cas-接头-RT融合对,例如用于特定靶基因的基因编辑,例如用于治疗疾病的基因靶标。在特别的实施例中,可以使用阵列形式的多种不同的tgRNA来进行筛选基因修饰候选物的池化方法。

[0335] 在一些实施例中,可以使用多种类型的编辑,例如不同长度的插入、取代和/或缺失,来筛选基因修饰候选文库。

[0336] 在一些实施例中,可以使用多个靶序列(例如不同的荧光蛋白)来筛选基因修饰候选文库。在一些实施例中,可以使用多个靶序列(例如不同的荧光蛋白)来筛选基因修饰候选文库。在一些实施例中,可以使用多种细胞类型,例如HEK293T或U2OS,来筛选基因修饰候选文库。本领域的普通技术人员将理解,给定的候选物可表现出改变的编辑能力,甚至在不同条件下获得或丧失任何可观察或有用的活性,包括tgRNA序列(例如,核苷酸修饰、PBS长度、RT模板长度)、靶序列、靶位置、编辑类型、相对于基因修饰多肽的第一链切口的突变位置或细胞类型。因此,在一些实施例中,跨多个参数筛选基因修饰文库候选物,例如,使用至少两种细胞类型中的至少两种不同的tgRNA,并且通过在任何单一条件下的富集来识别基因编辑活性。在其他实施例中,通过在至少两种条件下(例如,在所有筛选条件下)的富集来识别在不同tgRNA和细胞类型中具有更强活性的候选物。为清楚起见,在任何给定条件下表现出很少或没有富集的候选物不被认为是在所有条件下都无活性的,并且可以用不同的参数进行筛选或在多肽水平上重新配置,例如通过交换、改组或进化结构域(例如,RT结构域)、接头或其他信号(例如,NLS)。

[0337] 示例性Cas9-接头-RT融合体序列

[0338] 在一些实施例中,基因修饰多肽包含接头序列和RT序列。在一些实施例中,基因修饰多肽包含如表D中所列的接头序列,或与其具有至少75%、80%、85%、90%、95%、96%、97%、98%或99%同一性的氨基酸序列。在一些实施例中,基因修饰多肽包含如表D中所列的RT结构域的氨基酸序列,或与其具有至少75%、80%、85%、90%、95%、96%、97%、98%或99%同一性的氨基酸序列。在一些实施例中,基因修饰多肽包含如表D中所列的接头序列,或与其具有至少75%、80%、85%、90%、95%、96%、97%、98%或99%同一性的氨基酸序列;以及如表D中所列的RT结构域的氨基酸序列,或与其具有至少75%、80%、85%、90%、95%、96%、97%、98%或99%同一性的氨基酸序列。在一些实施例中,基因修饰多肽包含:(i)如表D的行中所列的接头序列,或与其具有至少75%、80%、85%、90%、95%、96%、97%、98%或99%同一性的氨基酸序列;以及(ii)如表D的同一行中所列的RT结构域的氨基酸序列,或与其具有至少75%、80%、85%、90%、95%、96%、97%、98%或99%同一性的氨基酸序列。

[0339] 基因修饰系统的定位序列

[0340] 在某些实施例中,基因编辑器系统RNA进一步包含细胞内定位序列,例如,核定位序列(NLS)。在一些实施例中,基因修饰多肽包含如SEQ ID NO:4000和/或SEQ ID NO:4001中包含的NLS,或具有与其具有至少75%、80%、85%、90%、95%、96%、97%、98%或99%同一性的氨基酸序列的NLS。

[0341] 核定位序列可以是促进RNA输入细胞核中的RNA序列。在某些实施例中,核定位信号位于模板RNA上。在某些实施例中,基因修饰多肽在第一RNA上编码,并且模板RNA是第二单独RNA,并且核定位信号位于模板RNA上而不是在编码基因修饰多肽的RNA上。尽管不希望受理论的束缚,但是在一些实施例中,编码基因修饰多肽的RNA主要靶向细胞质以促进其翻译,而模板RNA主要靶向细胞核以促进其插入基因组。在一些实施例中,核定位信号在模板RNA的3'端、5'端或内部区域。在一些实施例中,核定位信号在异源序列的3'(例如,直接在异源序列的3')或在异源序列的5'(例如,直接在异源序列的5')。在一些实施例中,核定位信号被置于模板RNA的5′UTR之外或3′UTR之外。在一些实施例中,核定位信号放置在5'UTR和3'UTR之间,其中任选地,核定位信号不随转基因转录(例如,核定位信号是反义取向或在转录终止信号或聚腺苷酸化信号的下游)。在一些实施例中,核定位序列位于内含子内部。在一些实施例中,多个相同或不同的核定位信号在RNA中,例如在模板RNA中。在一些实施例中,核定位信号的长度小于5、10、25、50、75、100、150、200、250、300、350、400、450、500、600、700、800、900或1000bp。可以使用各种RNA核定位序列。例如,Lubelsky和Ulitsky,Nature[自然]555(107-111),2018描述了RNA序列,其驱动RNA定位进入细胞核。在一些实施例中,核定位信号是SINE来源的核RNA定位(SIRLOIN)信号。在一些实施例中,核定位信号结合核富集蛋白。在一些实施例中,核定位信号结合HNRNPK蛋白。在一些实施例中,核定位信号富含嘧啶,例如是富含C/T、富含C/U、富含C、富含T或富含U的区域。在一些实施例中,核定位信号源自长非编码RNA。在一些实施例中,核定位信号源自MALAT1长非编码RNA或是MALAT1的600个核苷酸的M区(在Miyagawa等人,RNA 18,(738-751),2012中描述)。在一些实施例中,核定位信号源自BORG长非编码RNA或为AGCCC基序(在Zhang等人,Molecular and Cellular Biology[分子和细胞生物学]34,2318-2329(2014))中描述。在一些实施例中,核定位序列在Shukla等人,The EMBO Journal[EMBO杂志]e98452(2018)中描述。在一些实施例中,核定

位信号源自逆转录病毒。

[0342] 在一些实施例中,本文所述的多肽包含一个或多个(例如,2、3、4、5个)核靶向序列,例如核定位序列(NLS)。在一些实施例中,NLS是两组分NLS。在一些实施例中,NLS促进了包含NLS的蛋白质导入到细胞核中。在一些实施例中,将NLS与本文所述的基因修饰多肽的N末端融合。在一些实施例中,将NLS与基因修饰多肽的C末端融合。在一些实施例中,将NLS与Cas结构域的N末端或C末端融合。在一些实施例中,在NLS与基因修饰多肽的邻近结构域之间布置接头序列。

[0343] 在一些实施例中,NLS包含氨基酸序列MDSLLMNRRKFLYQFKNVRWAKGRRETYLC(SEQ ID NO:4009)、PKKRKVEGADKRTADGSEFESPKKKRKV(SEQ ID NO:4010)、RKSGKIAAIWKRPRKPKKKRKV(SEQ ID NO:4011)、KRTADGSEFESPKKKRKV(SEQ ID NO:4012)、KKTELQTTNAENKTKKL(SEQ ID NO:4013)、或KRGINDRNFWRGENGRKTR(SEQ ID NO:4014)、KRPAATKKAGQAKKKK(SEQ ID NO:4015),或其功能片段或变体。示例性NLS序列还描述于PCT/EP 2000/011690中,该专利的内容针对其对示例性核定位序列的披露通过援引并入本文。在一些实施例中,NLS包含如表7中披露的氨基酸序列。该表的NLS可以与多肽的一个或多个拷贝在多肽中一个或多个位置使用,例如N末端结构域中、肽结构域之间、C末端结构域中或多个位置的组合中的1、2、3个或多个NLS拷贝,以改善细胞核的亚细胞定位。可以在单个多肽中使用多个独特的序列。序列可以是天然的单组分或两组分的,例如,具有一段或两段碱性氨基酸,或者可以用作嵌合两组分序列。序列参考对应于UniProt登录号,除非针对使用亚细胞定位预测算法挖掘的序列指示为SeqNLS(Lin等人BMC Bioinformat[BMC生物信息学]13:157(2012),通过援引以其全文并入本文)。

[0344] 表7用于基因修饰系统的示例性核定位信号

[0345]

序列	序列参考	SEQ ID No.
AHFKISGEKRPSTDPGKKAKNPKKKKKKDP	Q76IQ7	4223
AHRAKKMSKTHA	P21827	4224
ASPEYVNLPINGNG	SeqNLS	4225
CTKRPRW	O88622、Q86W56、Q9QYM2、O02776	4226
DKAKRVSRNKSEKKRR	O15516、Q5RAK8、Q91YB2、Q91YB0、Q8QGQ6、O08785、Q9WVS9、Q6YGZ4	4227
EELRLKEELLKGIYA	Q9QY16、Q9UHL0、Q2TBP1、Q9QY15	4228
EEQLRRRKNSRLNNTG	G5EFF5	4229
EVLKVIRTGKRKKKAWKRMVTKVC	SeqNLS	4230
HHHHHHHHHHHHQPH	Q63934、G3V7L5、Q12837	4231
HKKKHPDASVNFSEFSK	P10103、Q4R844、P12682、B0CM99、A9RA84、Q6YKA4、	4232

[0346]

序列	序列参考	SEQ ID No.
	P09429、P63159、Q08IE6、P63158、Q9YH06、B1MTB0	
HKRTKK	Q2R2D5	4233
IINGRKLKLKKSRRRSSQTSNNSFTSRRS	SeqNLS	4234
KAEQERRK	Q8LH59	4235
KEKRKRREELFIEQKKRK	SeqNLS	4236
KKGKDEWFSRGKKP	P30999	4237
KKGPSVQKRKKT	Q6ZN17	4238
KKKTVINDLLHYKKEK	SeqNLS、P32354	4239
KKNGGKGKNKPSAKIKK	SeqNLS	4240
KKPKWDDFKKKKK	Q15397、Q8BKS9、Q562C7	4241
KKRKKD	SeqNLS、Q91Z62、Q1A730、Q969P5、Q2KHT6、Q9CPU7	4242
KKRRKRRRK	SeqNLS	4243
KKRRRRARK	Q9UMS6、D4A702、Q91YE8	4244
KKSKRGR	Q9UBS0	4245
KKSRKRGS	B4FG96	4246
KKSTALSRELGKIMRRR	SeqNLS、P32354	4247
KKSYQDPEIIAHSRPRK	Q9U7C9	4248
KKTGKNRKLKSKRVKTR	Q9Z301、O54943、Q8K3T2	4249
KKVSIAGQSGKLWRWKR	Q6YUL8	4250
KKYENVVIKRSPRKRGRPRK	SeqNLS	4251
KNKKRK	SeqNLS	4252
KPKKKR	SeqNLS	4253
KRAMKDDSHGNSTSPKRRK	Q0E671	4254
KRANSNLVAAYEKAKKK	P23508	4255
KRASEDTTSGSPPKKSSAGPKR	Q9BZZ5、Q5R644	4256
KRFKRRWMVRKMKTKK	SeqNLS	4257
KRGLNSSFETSPKKVK	Q8IV63	4258
KRGNSSIGPNDLSKRKQRKK	SeqNLS	4259
KRIHSVSLSQSQIDPSKKVKRAK	SeqNLS	4260
KRKGKLKNKGSKRKK	O15381	4261
KRRRRRREKRKR	Q96GM8	4262
KRSNDRTYSPEEEKQRRA	Q91ZF2	4263
KRTVATNGDASGAHRAKKMSK	SeqNLS	4264
KRVYNKGEDEQEHLPKGKKR	SeqNLS	4265
KSGKAPRRRAVSMDNSNK	Q9WVH4、O43524	4266
KVNFLDMSLDDIIIYKELE	Q9P127	4267
KVQHRIAKKTTRRRR	Q9DXE6	4268
LSPSLSPL	Q9Y261、P32182、P35583	4269
MDSLLMNRRKFLYQFKNVRWAKGRRETYLC	Q9GZX7	4270
MPQNEYIELHRKRYGYRLDYHEKKRKKESRE AHERSKKAKKMIGLKAKLYHK	SeqNLS	4271
MVQLRPRASR	SeqNLS	4272
NNKLLAKRRKGGASPKDDPMDDIK	Q965G5	4273
NYKRPMDGTYGPPAKRHEGE	O14497、A2BH40	4274
PDTKRAKLDSSETTMVKKK	SeqNLS	4275
PEKRTKI	SeqNLS	4276
PGGRGKKK	Q719N1、Q9UBP0、A2VDN5	4277
PGKMDKGEHRQERRDRPY	Q01844、Q61545	4278
PKKGDKYDKTD	Q45FA5	4279
PKKKSRK	O35914、Q01954	4280
PKKNKPE	Q22663	4281
PKKRAKV	P04295、P89438	4282
PKPKKLKVE	P55263、P55262、P55264、Q64640	4283
PKRGRGR	Q9FYS5、Q43386	4284
PKRRLVDDA	P0C797	4285
PKRRRTY	SeqNLS	4286
PLFKRR	A8X6H4、Q9TXJ0	4287
PLRKAKR	Q86WB0、Q5R8V9	4288
PPAKRKCIF	Q6AZ28、O75928、Q8C5D8	4289
PPARRRRL	Q8NAG6	4290
PPKKKRKV	Q3L6L5、P03070、P14999、P03071	4291
PPNKRMKVKH	Q8BN78	4292
PPRIYPQLPSAPT	P0C799	4293
PQRSPFPKSSVKR	SeqNLS	4294
PRPRKVPR	P0C799	4295
PRRRVQRKR	SeqNLS、Q5R448、Q5TAQ9	4296
PRRVRLK	Q58DJ0、P56477、Q13568	4297
PSRKRPR	Q62315、Q5F363、Q92833	4298
PSSKKRKV	SeqNLS	4299

[0347]

序列	序列参考	SEQ ID No.
PTKKRVK	P07664	4300
QRPGPYDRP	SeqNLS	4301
RGKGGKGLGKGGAKRHRK	SeqNLS	4302
RKAGKGGGGHKTTKKRSAKDEKVP	B4FG96	4303
RKIKLKRAK	A1L3G9	4304
RKIKRKRAK	B9X187	4305
RKKEAPGPREELRSRGR	O35126、P54258、Q5IS70、P54259	4306
RKKRKGK	SeqNLS、Q29243、Q62165、Q28685、O18738、Q9TSZ6、Q14118	4307
RKKRRQRRR	P04326、P69697、P69698、P05907、P20879、P04613、P19553、P0C1J9、P20893、P12506、P04612、Q73370、P0C1K0、P05906、P35965、P04609、P04610、P04614、P04608、P05905	4308
RKKSIPLSIKNLKRKHKRKKNKITR	Q9C0C9	4309
RKLVKPKNTKMKTKLRTNPY	Q14190	4310
RKRLILSDKGQLDWKK	SeqNLS、Q91Z62、Q1A730、Q2KHT6、Q9CPU7	4311
RKRLKSK	Q13309	4312
RKRRVRDNM	Q8QPH4、Q809M7、A8C8X1、Q2VNC5、Q38SQ0、O89749、Q6DNQ9、Q809L9、Q0A429、Q20NV3、P16509、P16505、Q6DNQ5、P16506、Q6XT06、P26118、Q2ICQ2、Q2RCG8、Q0A2D0、Q0A2H9、Q9IQ46、Q809M3、Q6J847、Q6J856、B4URE4、A4GCM7、Q0A440、P26120、P16511,	4313
RKRSPKDKKEKDLDGAGKRRKT	Q7RTP6	4314
RKRTPRVDGQTGENDMNKRRRK	O94851	4315
RLPVRRRRRR	P04499、P12541、P03269、P48313、P03270	4316
RLRFRKPKSK	P69469	4317
RQQRKR	Q14980	4318
RRDLNSSFETSPKKVK	Q8K3G5	4319
RRDRAKLR	Q9SLB8	4320
RRGDGRRR	Q80WE1、Q5R9B4、Q06787、P35922	4321
RRGRKRKAEKQ	Q812D1、Q5XXA9、Q99JF8、Q8MJG1、Q66T72、O75475	4322
RRKKRR	Q0VD86、Q58DS6、Q5R6G2、Q9ERI5、Q6AYK2、Q6NYC1	4323
RRKRSKSEDMDSVESKRRR	Q7TT18	4324
RRKRSR	Q99PU7、D3ZHS6、Q92560、A2VDM8	4325
RRPKGKTLQKRKPK	Q6ZN17	4326
RRRGFERFGPDNMGRKRK	Q63014、Q9DBR0	4327
RRRGKNKVAAQNCRK	SeqNLS	4328
RRRKRR	Q5FVH8、Q6MZT1、Q08DH5、Q8BQP9	4329
RRRQKQKGGASRRR	SeqNLS	4330
RRRREGPRARRRR	P08313、P10231	4331
RRTIRLKLVYDKCDRSCKIQKKNRNKCQYCR FHKCLSVGMSHNAIRFGRMPRSEKAKLKAE	SeqNLS	4332
RRVPQRKEVSRCRKCRK	Q5RJN4、Q32L09、Q8CAK3、Q9NUL5	4333
RVGGRRQAVECIEDLLNEPGQPLDLSCKRPR P	P03255	4334
RVVKLRIAP	P52639、Q8JMN0	4335
RVVRRR	P70278	4336
SKRKTKISRKTR	Q5RAY1、O00443	4337
SYVKTVPNRTRTYIKL	P21935	4338
TGKNEAKKRKIA	P52739、Q8K3J5、Q5RAU9	4339
TLSPASSPSSVSCPVIPASTDESPGSALNI	SeqNLS	4340
VSKKQRTGKKIH	P52739、Q8K3J5、Q5RAU9	4341
SPKKKRKVE		4342
KRTAD GSEFE SPKKKRKVE		4343
PAAKRVKLD		4344
PKKKRKV		4345
MDSLLMNRRKFLYQFKNVRWAKGRRETYLC		4346
SPKKKRKVEAS		4347
MAPKKKRKVGIHRGVP		4348

[0348] 在一些实施例中,NLS是两组分NLS。两组分NLS典型地包含由间隔子序列(其长度可以是例如约10个氨基酸)间隔开的两个碱性氨基酸簇。单组分NLS典型地缺乏间隔子。两组分NLS的实例是核浆素NLS,具有序列KR[PAATKKAGQA]KKKK(SEQ ID NO:4015),其中间隔子置于括号内。另一个示例性两组分NLS具有序列PKKKRKVEGADKRTADGSEFESP KKKRKV(SEQ ID NO:4016)。示例性NLS描述于国际申请WO 2020051561中,该申请通过援引以其全文并入本文,包括其关于核定位序列的披露。

[0349] 在某些实施例中,基因编辑器系统多肽(例如,如本文所述的基因修饰多肽)进一

步包含细胞内定位序列,例如,核定位序列和/或核仁定位序列。核定位序列和/或核仁定位序列可以是促进蛋白质输入到核和/或核仁中的氨基酸序列,其中它可以促进异源序列整合到基因组中。在某些实施例中,基因编辑器系统多肽(例如,如本文所述的基因修饰多肽)进一步包含核仁定位序列。在某些实施例中,基因修饰多肽在第一RNA上编码,模板RNA是第二单独RNA,并且核仁定位信号在编码基因修饰多肽的RNA上编码,而不在模板RNA上。在一些实施例中,核仁定位信号位于多肽的N末端、C末端或内部区域。在一些实施例中,使用多个相同或不同的核仁定位信号。在一些实施例中,核定位信号的长度小于5、10、25、50、75或100个氨基酸。可以使用各种多肽核仁定位信号。例如,Yang等人,Journal of Biomedical Science[生物化学科学杂志]22,33(2015)描述了一种核定位信号,其也起着核仁定位信号的作用。在一些实施例中,核仁定位信号也可以是核定位信号。在一些实施例中,核仁定位信号可以与核定位信号重叠。在一些实施例中,核仁定位信号可包含碱性残基区段。在一些实施例中,核仁定位信号可以富含精氨酸和赖氨酸残基。在一些实施例中,核仁定位信号可以源自在核仁中富集的蛋白质。在一些实施例中,核仁定位信号可以源自在核糖体RNA基因座处富集的蛋白质。在一些实施例中,核仁定位信号可以源自结合rRNA的蛋白质。在一些实施例中,核仁定位信号可以源自MSP58。在一些实施例中,核仁定位信号可以是单组分基序。在一些实施例中,核仁定位信号可以是两组分基序。在一些实施例中,核仁定位信号可以由多个单组分或两组分基序组成。在一些实施例中,核仁定位信号可以由单组分和两组分基序的混合物组成。在一些实施例中,核仁定位信号可以是双重两组分基序。在一些实施例中,核仁定位基序可以是KRASSQALGTIPKRRSSSRFIKRKK(SEQ ID NO:4017)。在一些实施例中,核仁定位信号可以源自核因子-κB诱导激酶。在一些实施例中,核仁定位信号可以是RKKRKKK基序(SEQ ID NO:4018)(在Birbach等人,Journal of Cell Science[细胞科学杂志],117(3615-3624),2004中描述)。

[0350] 基因修饰多肽和系统的进化变体

[0351] 在一些实施例中,本发明提供了如本文所述的基因修饰多肽的进化变体。在一些实施例中,进化变体可以通过对参考基因修饰多肽或其中包含的片段或结构域之一进行诱变处理而产生。在一些实施例中,一个或多个结构域(例如,逆转录酶结构域)进化。在一些实施例中,可以使一个或多个此类进化变体结构域单独进化或与其他结构域一起进化。在一些实施例中,可以将一个或多个进化变体结构域与一个或多个未进化的同源组分或一个或多个同源组分的进化的变体组合,例如,该一个或多个同源组分的进化的变体能以并行或连续方式进化。

[0352] 在一些实施例中,对参考基因修饰多肽或其片段或结构域进行诱变处理的过程包括对参考基因修饰多肽或其片段或结构域进行诱变处理。在实施例中,诱变包括连续进化方法(例如,PACE)或非连续进化方法(例如,PANCE),例如,如本文所述。在一些实施例中,进化的基因修饰多肽或其片段或结构域包含相对于参考基因修饰多肽或其片段或结构域的氨基酸序列引入其氨基酸序列中的一个或多个氨基酸变异。在实施例中,氨基酸序列变异可以包括参考基因修饰多肽的氨基酸序列内的一个或多个突变的残基(例如,保守取代、非保守取代、或其组合),例如,该一个或多个突变的残基是由于编码基因修饰多肽的核苷酸序列的变化(例如,该编码序列中任何特定位置处密码子的变化),该变化引起一个或多个氨基酸(例如,截短的蛋白质)的缺失、一个或多个氨基酸的插入或前述内容的任何组合。进

化变体基因修饰多肽可以包括基因修饰多肽的一个或多个组分或结构域中的变体(例如,引入逆转录酶结构域的变体)。

[0353] 在一些方面,本披露提供了使用或包含基因修饰多肽的进化变体的基因修饰多肽、系统、试剂盒和方法,例如,采用了基因修饰多肽的进化变体或由PACE或PANCE生产或可由其生产的基因修饰多肽。在实施例中,未进化的参考基因修饰多肽是如本文披露的基因修饰多肽。

[0354] 如本文所用,术语"噬菌体辅助的连续进化(PACE)"通常是指采用噬菌体作为病毒载体的连续进化。PACE技术的实例已描述于例如以下中:2009年9月8日提交的国际PCT申请号PCT/US 2009/056194,其于2010年3月11日公开为WO 2010/028347;2011年12月22日提交的国际PCT申请PCT/US 2011/066747,其于2012年6月28日公开为WO 2012/088381;2015年5月5日发布的美国专利号9,023,594;2017年9月26日发布的美国专利号9,771,574;2016年7月19日发布的美国专利号9,394,537;2015年1月20日提交的国际PCT申请PCT/US 2015/012022,其于2015年9月11日公开为WO 2015/134121;2019年1月15日发布的美国专利号10,179,911;以及2016年4月15日提交的国际PCT申请PCT/US 2016/027795,其于2016年10月20日公开为WO 2016/168631,其中每个的全部内容通过援引并入本文。

[0355] 如本文所用,术语"噬菌体辅助的非连续进化(PANCE)"通常是指采用噬菌体作为病毒载体的非连续进化。PANCE技术的实例已描述于例如Suzuki T.等人,Crystal structures reveal an elusive functional domain of pyrrolysyl-tRNA synthetase[晶体结构揭示了吡咯赖氨酰tRNA合成酶的难以捉摸的功能结构域],Nat Chem Biol.[自然化学生物学]13(12):1261-1266(2017)中,该文献通过援引以其全文并入本文。简言之,PANCE是一种使用进化中的选择噬菌体(SP)的连续烧瓶转移进行快速体内定向进化的技术,其中含有要在新鲜宿主细胞(例如,大肠杆菌细胞)中进化的目的基因。宿主细胞内的基因可能保持不变,而SP中含有的基因则连续进化。在噬菌体生长之后,可以使用等分的受感染细胞转染随后的含有宿主大肠杆菌的烧瓶。这一过程可以重复和/或继续,直到期望的表型实现进化,例如,持续所需的转移次数。

[0356] 技术人员通过参考(尤其是)前述文献可以容易地理解将PACE和PANCE应用于基因修饰多肽的方法。用于例如使用噬菌体颗粒例如在宿主细胞群体中引导基因组修饰蛋白或系统的连续进化的另外示例性方法可用于产生基因修饰多肽或其片段或亚结构域的进化变体。此类方法的非限制性实例描述于以下中:2009年9月8日提交的国际PCT申请PCT/US 2009/056194,其于2010年3月11日公开为WO 2010/028347;2011年12月22日提交的国际PCT申请PCT/US 2011/066747,其于2012年6月28日公开为WO 2012/088381;2015年5月5日发布的美国专利号9,023,594;2017年9月26日发布的美国专利号9,771,574;2016年7月19日发布的美国专利号9,394,537;2015年1月20日提交的国际PCT申请PCT/US 2015/012022,其于2015年9月11日公开为WO 2015/134121;2019年1月15日发布的美国专利号10,179,911;2019年6月14日提交的国际申请号PCT/US 2019/37216;2019年1月31日公开的国际专利公开WO 2019/023680;2016年4月15日提交的国际PCT申请PCT/US 2016/027795,其于2016年10月20日公开为WO 2016/168631;以及2019年8月23日提交的国际专利公开号PCT/US 2019/47996;其中每个通过援引以其全文并入本文。

[0357] 在一些非限制性说明性实施例中,进化变体基因修饰多肽、或其片段或结构域的

进化的方法包括:(a)使宿主细胞群体与包含目的基因(起始基因修饰多肽或其片段或结构域)的病毒载体群体接触,其中:(1)宿主细胞易于被病毒载体感染;(2)宿主细胞对产生病毒颗粒所需的病毒基因进行表达;(3)产生感染性病毒颗粒所需的至少一种病毒基因的表达取决于目的基因的功能;和/或(4)病毒载体允许蛋白质在宿主细胞中表达,并且可以被宿主细胞复制和包装成病毒颗粒。在一些实施例中,该方法包括(b)使宿主细胞与诱变剂接触,其使用具有提高突变率的突变的宿主细胞(例如,通过携带突变质粒或一些基因组修饰—例如,校对受损的DNA聚合酶、SOS基因,例如UmuC、UmuD'、和/或RecA,如果与质粒结合,这些突变可能在诱导型启动子的控制下)或其组合。在一些实施例中,该方法包括(c)在允许病毒复制和产生病毒颗粒的条件下孵育宿主细胞群体,其中从宿主细胞群体中去除宿主细胞,并将新鲜的、未感染的宿主细胞引入到宿主细胞群体中,从而补充宿主细胞群体并产生宿主细胞流。在一些实施例中,将细胞在允许目的基因获得突变的条件下孵育。在一些实施例中,该方法进一步包括(d)从宿主细胞群体中分离病毒载体的突变版本,该突变版本编码进化的基因产物(例如,进化变体基因修饰多肽、或其片段或结构域)。

[0358] 技术人员将理解在上述框架内可采用的各种特征。例如,在一些实施例中,病毒载体或噬菌体是丝状噬菌体,例如M13噬菌体,例如M13选择噬菌体。在某些实施例中,产生感染性病毒颗粒所需的基因是M13基因III(gIII)。在实施例中,噬菌体可能缺乏功能性gIII,但不同的是包含gI、gII、gIV、gV、gVI、gVII、gVIII、gIX、和gX。在一些实施例中,感染性VSV颗粒的产生涉及包膜蛋白VSV-G。各种实施例可以使用不同的逆转录病毒载体,例如鼠白血病病毒载体或慢病毒载体。在实施例中,利用VSV-G包膜蛋白(例如,作为病毒的天然包膜蛋白的替代物)可以有效包装逆转录病毒载体。

[0359] 在一些实施例中,根据合适数量的病毒生命周期孵育宿主细胞,例如至少10、至少20、至少30、至少40、至少50、至少100、至少200、至少300、至少400、至少500、至少600、至少700、至少800、至少900、至少1000、至少1250、至少1500、至少1750、至少2000、至少2500、至少3000、至少4000、至少5000、至少7500、至少10000或更多个连续的病毒生命周期,在M13噬菌体的说明性和非限制性实例中,每个病毒生命周期为10-20分钟。类似地,可以调节条件以调整宿主细胞在宿主细胞群体中保留的时间,例如约10、约11、约12、约13、约14、约15、约16、约17、约18、约19、约20、约21、约22、约23、约24、约25、约30、约35、约40、约45、约50、约55、约60、约70、约80、约90、约100、约120、约150、或约180分钟。可以部分地通过宿主细胞的密度来控制宿主细胞群体,或者在一些实施例中,流入物中的宿主细胞密度为例如10^3个细胞/ml、约10^4个细胞/ml、约10^5个细胞/ml、约5-10^5个细胞/ml、约10^6个细胞/ml、约5-10^6个细胞/ml、约10^7个细胞/ml、约5-10^7个细胞/ml、约10^8个细胞/ml、约5-10^8个细胞/ml、约10^9个细胞/ml、约5•10^9个细胞/ml、约10^{10}个细胞/ml、或约5•10^{10}个细胞/ml。

[0360] 内含肽

[0361] 在一些实施例中,如下文更详细描述的,内含肽-N(intN)结构域可融合至本文所述的基因修饰多肽的第一结构域的N末端部分,并且内含肽-C(intC)结构域可融合至本文所述的基因修饰多肽的第二结构域的C末端部分用于将N末端部分连接到C末端部分,从而连接第一和第二结构域。在一些实施例中,第一和第二结构域各自独立地选自DNA结合结构域、RNA结合结构域、RT结构域和核酸内切酶结构域。

[0362] 内含肽可作为自剪接蛋白内含子(例如肽)发生,例如,其连接侧翼N末端和C末端

外显肽(例如,待连接的片段)。在一些情况下,内含肽可以包含蛋白质的片段,该片段能够在称为蛋白质剪接的过程中自我切除并将剩余的片段(外显肽)与肽键连接。内含肽也称为“蛋白内含子”。本文将内含肽自我切除并将蛋白质的剩余部分连接的过程称为“蛋白质剪接”或“内含肽介导的蛋白质剪接”。

[0363] 在一些实施例中,前体蛋白(在内含肽介导的蛋白质剪接之前的含内含肽的蛋白质)的内含肽来自两个基因。这种内含肽在本文称为断裂内含肽(例如,断裂内含肽-N和断裂内含肽-C)。因此,可以使用基于内含肽的方法将第一多肽序列和第二多肽序列连接在一起。例如,在蓝细菌中,DNA聚合酶III的催化亚基a(即DnaE)由两个分开的基因dnaE-n和dnaE-c编码。当被定位为第一多肽序列的一部分时,内含肽-N结构域(例如,由dnaE-n基因编码)可以将第一多肽序列与第二多肽序列连接,其中该第二多肽序列包括内含肽-C结构域(例如由dnaE-c基因编码)。因此,在一些实施例中,可以通过提供编码第一和第二多肽序列的核酸(例如,其中第一核酸分子编码第一多肽序列并且第二核酸分子编码第二多肽序列)来制备蛋白质,并且在允许产生第一和第二多肽序列并且允许通过基于内含肽的机制将第一多肽序列连接至第二多肽序列的条件下将核酸引入细胞。

[0364] 以下中描述了使用内含肽连接异源蛋白质片段:例如,Wood等人,J.Biol.Chem.[生物化学杂志]289(21);14512-9(2014)(其通过援引以其全文并入本文)。例如,当与分开的蛋白质片段融合时,内含肽IntN和IntC可以彼此识别,自我剪除,和/或同时连接它们所融合的蛋白质片段的侧翼N末端和C末端外显肽,从而从两个蛋白质片段重构全长蛋白质。

[0365] 在一些实施例中,使用基于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。此类内含肽的实例已在以下中进行了描述:例如Stevens等人,J Am Chem Soc.[美国化学学会杂志]2016年2月24日;138(7):2162-5(其通过援引以其全文并入本文)。根据本披露可以使用的内含肽对的非限制性实例包括: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,该专利通过援引并入本文)。

[0366] 在一些涉及断裂Cas9的实施例中,可以将内含肽-N结构域和内含肽-C结构域分别与断裂Cas9的N末端部分和断裂Cas9的C末端部分融合,以便将断裂Cas9的N末端部分和断裂Cas9的C末端部分连接。例如,在一些实施例中,内含肽-N融合至分裂型Cas9的N末端部分的C末端,即形成N—[分裂型Cas9的N末端部分]-[内含肽-N]~C的结构。在一些实施例中,内含肽-C融合到分裂型Cas9的C末端部分的N末端,即,形成N-[内含肽-C]~[分裂型Cas9的C末端部分]-C的结构。用于连接与内含肽融合的蛋白质(例如,分裂型Cas9)的内含肽介导的蛋白质剪接机制在以下中进行描述:Shah等人,Chem Sci.[化学科学]2014;5(1):446-461,其通过援引并入本文。用于设计和使用内含肽的方法在本领域已知,并且例如由WO 2020051561、WO2014004336、WO 2017132580、US 20150344549、和US 20180127780进行了描述,其中每个通过援引以其全文并入本文。

[0367] 在一些实施例中,断裂是指分成两个或更多个片段。在一些实施例中,断裂Cas9蛋白或断裂Cas9包含Cas9蛋白,该蛋白作为由两个分开的核苷酸序列编码的N末端片段和C末端片段来提供。可以对与Cas9蛋白的N末端部分和C末端部分对应的多肽进行剪接以形成重构的Cas9蛋白。在实施例中,Cas9蛋白质在蛋白质的无序区内被分成两个片段,例如,如以

下中描述:Nishimasu等人,Cell[细胞],第156卷,第5期,第935-949页,2014,或Jiang等人(2016)Science[科学]351:867-871和PDB文件:5F9R(其中每个通过援引以其全文并入本文)。无序区可通过本领域已知的一种或多种蛋白质结构确定技术确定,包括但不限于X射线晶体学、NMR光谱学、电子显微术(例如,cryoEM)和/或计算机蛋白质建模。在一些实施例中,将蛋白质在例如氨基酸A292-G364、F445-K483、或E565-T637之间的SpCas9的区域内的任何C、T、A、或S处,或在任何其他Cas9、Cas9变体(例如,nCas9、dCas9)或其他napDNAbp中的对应位置处分成两个片段。在一些实施例中,将蛋白质在SpCas9 T310、T313、A456、S469、或C574处分成两个片段。在一些实施例中,将蛋白质分成两个片段的过程称为对蛋白质的断裂。

[0368] 在一些实施例中,蛋白质片段的长度范围为约2-1000个氨基酸(例如,2-10、10-50、50-100、100-200、200-300、300-400、400-500、500-600、600-700、700-800、800-900、或900-1000个之间的氨基酸)。在一些实施例中,蛋白质片段的长度范围为约5-500个氨基酸(例如,5-10、10-50、50-100、100-200、200-300、300-400、或400-500个之间的氨基酸)。在一些实施例中,蛋白质片段的长度范围为约20-200个氨基酸(例如,20-30、30-40、40-50、50-100、或100-200个之间的氨基酸)。

[0369] 在一些实施例中,将基因修饰多肽的部分或片段与内含肽融合。可以将核酸酶与内含肽的N末端或C末端融合。在一些实施例中,将融合蛋白的部分或片段与内含肽融合并与AAV衣壳蛋白融合。可以将内含肽、核酸酶和衣壳蛋白以任何排列方式(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)融合在一起。在一些实施例中,将内含肽的N末端与融合蛋白的C末端融合,并将内含肽的C末端与AAV衣壳蛋白的N末端融合。

[0370] 在一些实施例中,核酸内切酶结构域(例如,切口酶Cas9结构域)与内含肽-N融合,并且包含RT结构域的多肽与内含肽-C融合。

[0371] 下面提供了内含肽-N结构域和相容的内含肽-C结构域的示例性核苷酸和氨基酸序列:

[0372] DnaE内含肽-N DNA:

[0373] TGCCTGTCATACGAAACCGAGATACTGACAGTAGAATATGGCCTTCTGCCAATCGGGAAGATTGTGGAGAAACGGATAGAATGCACAGTTTACTCTGTCGATAACAATGGTAACATTTATACTCAGCCAGTTGCCCAGTGGCACGACCGGGGAGAGCAGGAAGTATTCGAATACTGTCTGGAGGATGGAAGTCTCATTAGGGCCACTAAGGACCACAAATTTATGACAGTCGATGGCCAGATGCTGCCTATAGACGAAATCTTTGAGCGAGAGTTGGACCTCATGCGAGTTGACAACCTTCCTAAT(SEQ ID NO:4029)

[0374] DnaE内含肽-N蛋白质:

[0375] CLSYETEILTVEYGLLPIGKIVEKRIECTVYSVDNNGNIYTQPVAQWHDRGEQEVFEYCLEDGSLIRATKDHKFMTVDGQMLPIDEIFERELDLMRVDNLPN(SEQ ID NO:4030)

[0376] DnaE内含肽-C DNA:

[0377] ATGATCAAGATAGCTACAAGGAAGTATCTTGGCAAACAAAACGTTTATGATATTGGAGTCGAAAGAGATCACAACTTTGCTCTGAAGAACGGATTCATAGCTTCTAAT(SEQ ID NO:4031)

[0378] DnaE内含肽-C蛋白质:

[0379] MIKIATRKYLGKQNVYDIGVERDHNFALKNGFIASN(SEQ ID NO:4032)Cfa-N DNA:

[0380] TGCCTGTCTTATGATACCGAGATACTTACCGTTGAATATGGCTTCTTGCCTATTGGAAAGATTGTCGAAGAGAGAATTGAATGCACAGTATATACTGTAGACAAGAATGGTTTCGTTTACACACAGCCCATTGCTCAATGGCACAATCGCGGCGAACAAGAAGTATTTGAGTACTGTCTCGAGGATGGAAGCATCATACGAGCAACTAAAGATCATAAATTCATGACCACTGACGGGCAGATGTTGCCAATAGATGAGATATTCGAGCGGGGCTTGGATCTCAAACAAGTGGATGGATTG CCA(SEQ ID NO:4033)

[0381] Cfa-N蛋白质:

[0382] CLSYDTEILTVEYGFLPIGKIVEERIECTVYTVDKNGFVYTQPIAQWHNRGEQEVFEYCLEDGSIIRATKDHKFMTTDGQMLPIDEIFERGLDLKQVDGLP(SEQ ID NO:4034)

[0383] Cfa-C DNA:

[0384] ATGAAGAGGACTGCCGATGGATCAGAGTTTGAATCTCCCAAGAAGAAGAGGAAAGTAAAGATAATATCTCGAAAAAGTCTTGGTACCCAAAATGTCTATGATATTGGAGTGGAGAAAGATCACAACTTCCTTCTCAAGAACGGTCTCGTAGCCAGCAAC(SEQ ID NO:4035)

[0385] Cfa-C蛋白质:

[0386] MKRTADGSEFESPKKKRKVKIISRKSLGTQNVYDIGVEKDHNFLLKNGLVASN(SEQ ID NO:4036)

[0387] 另外的结构域

[0388] 基因修饰多肽可以结合靶DNA序列和模板核酸(例如模板RNA)、对靶位点进行切口并将模板书写(例如逆转录)入DNA中,从而产生靶位点的修饰。在一些实施例中,可以向多肽添加另外的结构域以提高过程的效率。在一些实施例中,基因修饰多肽可包含另外的DNA连接结构域以将逆转录的DNA连接至靶位点的DNA。在一些实施例中,多肽可以包含异源RNA结合结构域。在一些实施例中,多肽可包含具有5′至3′外切核酸酶活性的结构域(例如,其中5′至3′外切核酸酶活性增加靶位点的改变的修复,例如有利于改变原始基因组序列)。在一些实施例中,多肽可包含具有3′至5′外切核酸酶活性,例如校对活性的结构域。在一些实施例中,书写结构域,例如RT结构域,具有3′至5′外切核酸酶活性,例如校对活性。

[0389] 模板核酸

[0390] 本文所述的基因修饰系统可以使用模板核酸序列修饰宿主靶DNA位点。在一些实施例中,本文所述的基因修饰系统通过靶引发的逆转录(TPRT)将RNA序列模板转录到宿主靶DNA位点中。通过将RNA序列模板直接逆转录到宿主基因组中来修饰一个或多个DNA序列,基因修饰系统可以将对象序列插入靶基因组中,而不需要将外源DNA序列引入宿主细胞中(不同于例如CRISPR系统)以及消除外源DNA插入步骤。基因修饰系统还可以从靶基因组中缺失序列或使用对象序列引入取代。因此,基因修饰系统提供了使用定制的RNA序列模板的平台,该模板包含对象序列,例如,包含异源基因编码和/或功能信息的序列。

[0391] 在一些实施例中,模板核酸包含一个或多个与基因修饰多肽结合的序列(例如,2个序列)。

[0392] 在一些实施例中,模板核酸包含在同一链中同时具有核糖核苷酸和脱氧核糖核苷酸残基的杂交体。

[0393] 在一些实施例中,本文所述的系统或方法包含单个模板核酸(例如,模板RNA)。在一些实施例中,本文所述的系统或方法包含多个模板核酸(例如,模板RNA)。例如,本文所述的系统包含第一RNA和第二RNA(例如,模板RNA),该第一RNA包含(例如,从5′至3′)结合基因修饰多肽的序列(例如,DNA结合结构域和/或核酸内切酶结构域,例如,gRNA)和结合靶位点

(例如,靶基因组中位点的第二链)的序列,该第二RNA包含(例如,从5′至3′)任选地结合基因修饰多肽(例如,特异性结合RT结构域)的序列、异源对象序列和PBS序列。在一些实施例中,当系统包含多个核酸时,每个核酸包含缀合结构域。在一些实施例中,缀合结构域使得核酸分子能够相关联,例如,通过互补序列的杂交。例如,在一些实施例中,第一RNA包含第一缀合结构域并且第二RNA包含第二缀合结构域,并且第一和第二缀合结构域能够例如在严格条件下彼此杂交。在一些实施例中,杂交的严格条件包括在约65C在4x氯化钠/柠檬酸钠(SSC)中杂交,然后在约65C在1xSSC中洗涤。

[0394] 在一些实施例中,模板核酸包含RNA。在一些实施例中,模板核酸包含DNA(例如,单链或双链DNA)。在一些实施例中,模板核酸包含在同一链中同时具有核糖核苷酸和脱氧核糖核苷酸残基的杂交体。

[0395] 在一些实施例中,模板核酸包含一个或多个(例如2个)与靶序列具有同源性的同源结构域。在一些实施例中,同源结构域的长度约为10-20、20-50或50-100个核苷酸。

[0396] 在一些实施例中,模板RNA可以包含gRNA序列,例如,以将基因修饰多肽引导至目的靶位点。在一些实施例中,模板RNA包含(例如,从5′至3′):(i)任选地结合靶位点(例如,靶基因组中位点的第二链)的gRNA间隔子,(ii)任选地结合本文所述的多肽(例如,基因修饰多肽或Cas多肽)的gRNA支架,(iii)包含突变区的异源对象序列(任选地,异源对象序列从5’至3’包含第一同源区、突变区和第二同源区),以及(iv)包含3′靶同源结构域的引物结合位点(PBS)序列。

[0397] 本文所述的基因组编辑系统的模板核酸(例如,模板RNA)组分通常能够结合系统的基因修饰多肽。在一些实施例中,模板核酸(例如,模板RNA)具有3′区,其能够结合基因修饰多肽。结合区(例如3′区)可以是结构化的RNA区,例如具有至少1、2或3个发夹环,其能够结合系统的基因修饰多肽。结合区可以将模板核酸(例如,模板RNA)与任何多肽模块相关联。在一些实施例中,模板核酸(例如,模板RNA)的结合区可以与多肽中的RNA结合结构域相关联。在一些实施例中,模板核酸(例如,模板RNA)的结合区可以与基因修饰多肽的逆转录结构域相关联(例如,特异性结合RT结构域)。在一些实施例中,模板核酸(例如,模板RNA)可以与多肽的DNA结合结构域相关联,例如,gRNA与Cas9来源的DNA结合结构域相关联。在一些实施例中,结合区还可以提供DNA靶识别,例如gRNA与靶DNA序列杂交并结合多肽,例如Cas9结构域。在一些实施例中,模板核酸(例如,模板RNA)可以与多肽的多个组分(例如,DNA结合结构域和逆转录结构域)相关联。

[0398] 在一些实施例中,模板RNA在3′端具有聚A尾。在一些实施例中,模板RNA在3′端不具有聚A尾。

[0399] 在一些实施例中,可以定制模板RNA以纠正靶细胞基因组DNA中的给定突变(例如,离体或体内,例如在靶组织或器官中,例如在受试者中)。例如,突变可以是相对于野生型序列的疾病相关突变。不希望受理论束缚,任何给定的靶位点和编辑都将具有用于在基因修饰系统中使用的大量可能的模板RNA分子,这些模板RNA分子将产生一系列编辑效率和保真度。为了部分减轻这种筛选负担,经验参数组有助于确保模板RNA或其部分的最佳初始计算机模拟设计。作为非限制性说明性实例,对于选择的突变,可以采用以下设计参数。在一些实施例中,通过获取突变任一侧上大约500bp(例如,高达50、100、150、200、250、300、350、400、450、500、550、600、650、或700bp并且任选地至少20、30、40、50、100、150、200、250、300、

350、400、450、500、550、600、或650bp)侧翼序列启动设计以作为靶区域。在一些实施例中,模板核酸包含gRNA。在一些实施例中,gRNA包含结合靶位点的序列(例如,CRISPR间隔子)。在一些实施例中,通过以下来选择结合靶位点的序列(例如CRISPR间隔子)用于将模板核酸靶向靶区域:考虑所使用的特定基因修饰多肽(例如,核酸内切酶结构域或书写结构域,例如,包含CRISPR/Cas结构域)(例如,对于Cas9,紧邻20个核苷酸的gRNA结合区3′的NGG的原间隔子邻近基序(PAM))。在一些实施例中,通过首先根据PAM是否将被基因修饰系统诱导的编辑破坏来选择CRISPR间隔子。在一些实施例中,PAM的中断可以提高编辑效率。在一些实施例中,可以通过以下来破坏PAM:还在基因修饰期间在靶位点中(例如,作为基因组DNA中靶位点的另一种修饰的一部分或除此之外再)引入沉默突变(例如,不改变由靶核酸序列编码的氨基酸残基的突变,如果有的话)。在一些实施例中,通过以下来选择CRISPR间隔子:根据其相应基因组位点与所期望编辑位置的接近程度对序列进行排序。在一些实施例中,gRNA包含gRNA支架。在一些实施例中,使用的gRNA支架可以是标准支架(例如,对于Cas9,5′-GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC-3′),或者可以包含一个或多个核苷酸取代。在一些实施例中,异源对象序列与第一链切口的3′靶位点(例如,紧邻第一链切口的3′或第一链切口的3′高达1、2、3、4或5个核苷酸)具有至少90%同一性,例如至少90%、95%、98%、99%或100%同一性,或包含不超过1、2、3、4或5个非同一性的位置,可以通过基因修饰写入靶位点的任何插入、取代或缺失除外。在一些实施例中,3′靶同源结构域与第一链切口的5′靶位点(例如,紧邻第一链切口的5′或第一链切口的3′高达1、2、3、4或5个核苷酸)具有至少90%同一性,例如至少90%、95%、98%、99%或100%同一性,或包含不超过1、2、3、4或5个非同一性的位置。

[0400] 在一些实施例中,模板核酸是模板RNA。在一些实施例中,模板RNA包含一个或多个经修饰的核苷酸。例如,在一些实施例中,模板RNA包含一个或多个脱氧核糖核苷酸。在一些实施例中,模板RNA的区域被DNA核苷酸替代,例如,以增强分子的稳定性。例如,模板的3′端可包含DNA核苷酸,而模板的其余部分包含可以逆转录的RNA核苷酸。例如,在一些实施例中,异源对象序列主要或完全由RNA核苷酸(例如,至少90%、95%、98%或99%RNA核苷酸)构成。在一些实施例中,PBS序列主要或完全由DNA核苷酸(例如,至少90%、95%、98%或99%DNA核苷酸)构成。在其他实施例中,用于书写进入基因组的异源对象序列可以包含DNA核苷酸。在一些实施例中,模板中的DNA核苷酸通过能够具有DNA依赖性DNA聚合酶活性的结构域复制到基因组中。在一些实施例中,DNA依赖性DNA聚合酶活性由多肽中的DNA聚合酶结构域提供。在一些实施例中,DNA依赖性DNA聚合酶活性由逆转录酶结构域提供,该逆转录酶结构域也能够进行DNA依赖性DNA聚合,例如第二链合成。在一些实施例中,模板分子仅由DNA核苷酸构成。在一些实施例中,模板核酸包含在同一链中同时具有核糖核苷酸和脱氧核糖核苷酸残基的杂交体。

[0401] 在一些实施例中,本文所述系统包含两种核酸,其共同构成本文所述模板RNA的序列。在一些实施例中,两种核酸以非共价方式彼此关联,例如,直接彼此关联(例如,通过碱基配对),或作为包含一个或多个另外分子的复合物的一部分间接关联。

[0402] 本文所述的模板RNA从5′至3′可包含:(1)gRNA间隔子;(2)gRNA支架;(3)异源对象序列;(4)引物结合位点(PBS)序列。现在将更详细地描述这些组分中的每一种。

[0403] gRNA间隔子和gRNA支架

[0404] 本文所述的模板RNA可以包含将基因修饰系统引导至靶核酸的gRNA间隔子,和促进模板RNA与基因修饰多肽的Cas结构域相关联的gRNA支架。本文所述的系统还可以包含不属于模板核酸一部分的gRNA。例如,包含gRNA间隔子和gRNA支架但不包含异源对象序列或PBS序列的gRNA可用于例如诱导第二链切口,例如,如本文中标题为“第二链切口”的部分中所描述的。

[0405] 在一些实施例中,gRNA是由参与CRISPR相关蛋白结合的支架序列和针对基因组靶标的用户定义的约20个核苷酸的靶向序列构成的短合成RNA。Nishimasu等人Cell[细胞]156、第935-949页(2014)描述了完整gRNA的结构。gRNA(也称为单指导RNA的sgRNA)由crRNA和tracrRNA来源的序列组成,这些序列通过人工四环连接。crRNA序列可分为指导区(20nt)和重复序列区(12nt),而tracrRNA序列可分为抗重复序列区(14nt)和三个tracrRNA茎环(Nishimasu等人Cell[细胞]156,第935-949页(2014))。在实践中,指导RNA序列通常被设计为具有17-24个核苷酸(例如19、20或21个核苷酸)的长度,并且与靶核酸序列互补。定制gRNA生成器和算法可通过商业途径获得,用于设计有效的指导RNA。在一些实施例中,gRNA包含来自天然CRISPR系统的两种RNA组分,例如crRNA和tracrRNA。如本领域公知的,gRNA还可以包含嵌合的单指导RNA(sgRNA),其含有来自tracrRNA(以结合核酸酶)和至少一个crRNA(以将核酸酶引导至被靶向进行编辑/结合的序列)的序列。化学修饰的sgRNA也已被证明可有效地与CRISPR相关蛋白一起使用;参见,例如,Hendel等人(2015)Nature Biotechnol.[自然生物技术],985-991。在一些实施例中,gRNA间隔子包含与靶基因相关联的DNA序列互补的核酸序列。

[0406] 在一些实施例中,包含gRNA的模板核酸(例如模板RNA)的区域采用结合到靶DNA的gRNA的下绕带状结构(例如,如以下所述:Mulepati等人Science[科学]2014年9月19日:第345卷,第6203期,第1479-1484页)。不希望受理论束缚,这种非典型结构被认为是通过每六个核苷酸轮换出RNA-DNA杂交体来促进的。因此,在一些实施例中,包含gRNA的模板核酸(例如模板RNA)的区域可以耐受以某个间隔(例如每六个碱基)与靶位点增加的错配。在一些实施例中,包含与靶位点同源的gRNA的模板核酸(例如模板RNA)区域可以具有以规则间隔(例如每六个碱基)的摆动位置,其不需要与靶位点进行碱基配对。

[0407] 在一些实施例中,模板核酸(例如,模板RNA)具有与靶位点具有至少80%、85%、90%、95%、99%或100%同源性的至少15、16、17、18、19、20、21、22、23或24个碱基,例如在5’末端,例如包含长度适合于基因修饰多肽的Cas9结构域(表3)的gRNA间隔子序列。

[0408] 在一些实施例中,可以在基因修饰多肽中使用具有增强活性的Cas9衍生物。在一些实施例中,Cas9衍生物可以包含改善HNH核酸内切酶结构域活性的突变,例如SpyCas9 R221K、N394K,或改善R环形成的突变,例如SpyCas9 L1245V,或包含此类突变的组合,例如SpyCas9 R221K/N394K、SpyCas9 N394K/L1245V、SpyCas9 R221K/L1245V或SpyCas9R221K/N394K/L1245V(参见例如,Spencer和Zhang Sci Rep[科学报告]7:16836(2017),Cas9衍生物及其包含突变的内容通过援引并入本文)。在一些实施例中,Cas9衍生物可以包括本文所述的一种或多种类型的突变,例如,PAM修饰突变、蛋白质稳定突变、活性增强突变和/或相对于亲本酶部分或完全灭活一个或两个核酸内切酶结构域的突变(例如,一个或多个突变以消除针对靶DNA的一条或两条链的核酸内切酶活性,例如,切口酶或催化失活酶)。在一些实施例中,本文所述系统中使用的Cas9酶除了包含提高催化效率的突变(例如,SpyCas9

R221K、N394K和/或L1245V)之外,还可以包含赋予酶切口酶活性的突变(例如,SpyCas9 N863A或H840A)。在一些实施例中,本文所述系统中使用的Cas9酶是SpyCas9酶或衍生物,其除了包含提高催化效率的R221K和N394K突变之外还进一步包含赋予切口酶活性的N863A突变。

[0409] 表8提供了定义用于设计gRNA和/或模板RNA的组分的参数,以将表3中列出的Cas变体应用于基因修饰。切割位点指示经验证或预测的原间隔子邻近基序(PAM)要求、经验证或预测的切割位点位置(相对于PAM位点的最上游碱基)。给定酶的gRNA可以通过连接crRNA、四环和tracrRNA序列,并进一步在间隔子(min)和间隔子(max)内添加长度与靶位点的原间隔子匹配的5′间隔子来组装。此外,ssDNA切口在靶标上的预测位置对于设计模板RNA的PBS序列(其可立即与切口5′的序列退火,以启动靶引发的逆转录)很重要。在一些实施例中,本文所述的gRNA支架包含在5′至3′方向上包含以下的核酸序列或与其具有至少70%、80%、85%、90%、95%或99%同一性的序列:表8的crRNA、来自表8的同一行的四环、以及来自表8的同一行的tracrRNA。在一些实施例中,包含支架的gRNA或模板RNA进一步包含gRNA间隔子,其长度在表8的同一行中指示的间隔子(min)和间隔子(max)内。在一些实施例中,进一步包含基因修饰多肽的系统包含具有根据表8的序列的gRNA或模板RNA,其中该基因修饰多肽包含表8的同一行中描述的Cas结构域。

[0410] 表8定义用于设计gRNA和/或模板RNA的组分的参数以将表3中列出的Cas变体应用于基因修饰系统

[0411]

变体	PAM	切割	等级	间隔子(min)	间隔子(max)	crRNA	四环	tracrRNA
Nme2Cas9	NNNNCC	-3	1	22	24	GTTGTAGCTCCCTTTCTCATTTCG	GAAA	CGAAATGAGAACCGTTGCTACAATAAGGCCGTCTGAAAAGATGTGCCGCAACGCTCTGCCCCTTAAAGCTTCTGCTTTAAGGGGC

[0412]

								ATCGTTTA
PpnCas9	NNNNRTT		1	21	24	GTTGTAGCTCCCTTTTTCATTTCGC	GAAA	GCGAAATGAAAAACGTTGTTACAATAAGAGATGAATTTCTCGCAAAGCTCTGCCTCTTGAAATTTCGGTTTCAAGAGGCATC
SauCas9	NNGRR;NNGRRT	-3	1	21	23	GTTTTAGTACTCTG	GAAA	CAGAATCTACTAAAACAAGGCAAAATGCCGTGTTTATCTCGTCAACTTGTTGGCGAGA
SauCas9-KKH	NNNRR;NNNRRT	-3	1	21	21	GTTTTAGTACTCTG	GAAA	CAGAATCTACTAAAACAAGGCAAAATGCCGTGTTTATCTCGTCAACTTGTTGGCGAGA
SauriCas9	NNGG	-3	1	21	21	GTTTTAGTACTCTG	GAAA	CAGAATCTACTAAAACAAGGCAAAATGCCGTGTTTATCTCGTCAACTTGTTGGCGAGA
SauriCas9-KKH	NNRG	-3	1	21	21	GTTTTAGTACTCTG	GAAA	CAGAATCTACTAAAACAAGGCAAAATGCCGTGTTTATCTCGTCAACTTGTTGGCGAGA
ScaCas9-Sc++	NNG	-3	1	20	20	GTTTTAGAGCTA	GAAA	TAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
SpyCas9	NGG	-3	1	20	20	GTTTTAGAGCTA	GAAA	TAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
SpyCas9-NG	NG (NGG=NGA=NGT>NGC)	-3	1	20	20	GTTTTAGAGCTA	GAAA	TAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
SpyCas9-SpRY	NRN>NYN	-3	1	20	20	GTTTTAGAGCTA	GAAA	TAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
St1Cas9	NNAGAAW>NNAGGAW=NNGGAAW	-3	1	20	20	GTCTTTGTACTCTG	GTAC	CAGAAGCTACAAAGATAAGGCTTCATGCCGAAATCAACACCCTGTCATTTTATGGCAGGGTGTTTT
BlatCas9	NNNNCNAA>NNNNCNDD>NNNNC	-3	1	19	23	GCTATAGTTCCTTACT	GAAA	GGTAAGTTGCTATAGTAAGGGCAACAGACCCGAGGCGTTGGGGATCGCCTAGCCCGTGTTTACGGGCTCTCCCCATATTCAAAATAATGACAGACGAGCACCTTGGAGCATTTATCTCCGAGGTGCT
cCas9-v16	NNVACT;NNVATGM;NNVATT;NNVGCT;NNVGTG;NN	-3	2	21	21	GTCTTAGTACTCTG	GAAA	CAGAATCTACTAAGACAAGGCAAAATGCCGTGTTTATCTCGTCAACTTGTTGGCGAGA

[0413]

	VGTT							
cCas9-v17	NNVRRN	-3	2	21	21	GTCTTAGTACTCTG	GAAA	CAGAATCTACTAAGACAAGGCAAAATGCCGTGTTTATCTCGTCAACTTGTTGGCGAGA
cCas9-v21	NNVACT;NNVATGM;NNVATT;NNVGCT;NNVGTG;NNVGTT	-3	2	21	21	GTCTTAGTACTCTG	GAAA	CAGAATCTACTAAGACAAGGCAAAATGCCGTGTTTATCTCGTCAACTTGTTGGCGAGA
cCas9-v42	NNVRRN	-3	2	21	21	GTCTTAGTACTCTG	GAAA	CAGAATCTACTAAGACAAGGCAAAATGCCGTGTTTATCTCGTCAACTTGTTGGCGAGA
CdiCas9	NNRHHHY;NNRAAAY		2	22	22	ACTGGGGTTCAG	GAAA	CTGAACCTCAGTAAGCATTGGCTCGTTTCCAATGTTGATTGCTCCGCCGGTGCTCCTTATTTTTAAGGGCGCCGGC
CjeCas9	NNNNRYAC	-3	2	21	23	GTTTTAGTCCCT	GAAA	AGGGACTAAAATAAAGAGTTTGCGGGACTCTGCGGGGTTACAATCCCCTAAAACCGC
GeoCas9	NNNNCRAA		2	21	23	GTCATAGTTCCCCTGA	GAAA	TCAGGGTTACTATGATAAGGGCTTTCTGCCTAAGGCAGACTGACCCGCGGCGTTGGGGATCGCCTGTCGCCCGCTTTTGGCGGGCATTCCCCATCCTT
iSpyMacCas9	NAAN	-3	2	19	21	GTTTTAGAGCTA	GAAA	TAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
NmeCas9	NNNNGAYT;NNNNGYTT;NNNNGAYA;NNNNGTCT	-3	2	20	24	GTTGTAGCTCCCTTTCTCATTTCG	GAAA	CGAAATGAGAACCGTTGCTACAATAAGGCCGTCTGAAAAGATGTGCCGCAACGCTCTGCCCCTTAAAGCTTCTGCTTTAAGGGGCATCGTTTA
ScaCas9	NNG	-3	2	20	20	GTTTTAGAGCTA	GAAA	TAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
ScaCas9-HiFi-Sc++	NNG	-3	2	20	20	GTTTTAGAGCTA	GAAA	TAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
SpyCas9-3var-NRRH	NRRH	-3	2	20	20	GTTTAAGAGCTATGCTG	GAAA	CAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
SpyCas9-3var-NRTH	NRTH	-3	2	20	20	GTTTAAGAGCTATGCTG	GAAA	CAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGT

[0414]

								GC
SpyCas9-3var-NRCH	NRCH	-3	2	20	20	GTTTAAGAGCTATGCTG	GAAA	CAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
SpyCas9-HF1	NGG	-3	2	20	20	GTTTTAGAGCTA	GAAA	TAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
SpyCas9-QQR1	NAAG	-3	2	20	20	GTTTTAGAGCTA	GAAA	TAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
SpyCas9-SpG	NGN	-3	2	20	20	GTTTTAGAGCTA	GAAA	TAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
SpyCas9-VQR	NGAN	-3	2	20	20	GTTTTAGAGCTA	GAAA	TAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
SpyCas9-VRER	NGCG	-3	2	20	20	GTTTTAGAGCTA	GAAA	TAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
SpyCas9-xCas	NG; GAA; GAT	-3	2	20	20	GTTTAAGAGCTATGCTG	GAAA	CAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
SpyCas9-xCas-NG	NG	-3	2	20	20	GTTTAAGAGCTATGCTG	GAAA	CAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
St1Cas9-CNRZ1066	NNACAA	-3	2	20	20	GTCTTTGTACTCTG	GTAC	CAGAAGCTACAAAGATAAGGCTTCATGCCGAAATCAACACCCTGTCATTTTATGGCAGGGTGTTTT
St1Cas9-LMG1831	NNGCAA	-3	2	20	20	GTCTTTGTACTCTG	GTAC	CAGAAGCTACAAAGATAAGGCTTCATGCCGAAATCAACACCCTGTCATTTTATGGCAGGGTGTTTT
St1Cas9-MTH17CL396	NNAAAA	-3	2	20	20	GTCTTTGTACTCTG	GTAC	CAGAAGCTACAAAGATAAGGCTTCATGCCGAAATCAACACCCTGTCATTTTATGGCAGGGTGTTTT
St1Cas9-TH1477	NNGAAA	-3	2	20	20	GTCTTTGTACTCTG	GTAC	CAGAAGCTACAAAGATAAGGCTTCATGCCGAAATCAACACCCTGTCATTTTATGGCAGGGTGTTTT
sRGN3.1	NNGG		1	21	23	GTTTTAGTACTCTG	GAAA	CAGAATCTACTGAAACAAGACAATATGTCGTGTTTATCCCATCAATTTATTGGTGGGATTTT
sRGN3.3	NNGG		1	21	23	GTTTTAGTACTCTG	GAAA	CAGAATCTACTGAAACAAGACAATATGTCGTGTTTATCCCATCAATTTATTGGTGGGATTTT

[0415] 此外,应理解,可以任选地从tracrRNA序列中添加或去除末端U和T,并且当以RNA提供时可以进行修饰或不进行修饰。在不希望受实例约束的情况下,表8中所例示的gRNA支架序列的替代版本也可以与表4中所例示的不同Cas9酶或其衍生物一起发挥作用,例如,具有核苷酸添加、取代或缺失的替代gRNA支架序列,例如,添加或去除茎环结构的序列。本文

预期,gRNA支架序列代表基因修饰系统的组分,可以针对给定系统、Cas-RT融合多肽、适应症、靶突变、模板RNA或递送媒介物进行类似优化。

[0416] 本文中,当认为RNA序列(例如,模板RNA序列)包含含有胸腺嘧啶(T)的特定序列(例如,表8的序列或其一部分)时,当然应理解,RNA序列可以(并且确实经常)包含尿嘧啶(U)来代替T。例如,RNA序列在表8中的序列中显示为T的每个位置处都可以包含U。更特别地,本披露提供了根据表8的每个gRNA支架序列的RNA序列,其中该RNA序列具有U代替表8中的序列中的每个T。

[0417] 异源对象序列

[0418] 本文所述的模板RNA可以包含异源对象序列,基因修饰多肽可将其用作逆转录的模板,以将所期望的序列写入靶核酸。在一些实施例中,异源对象序列从5′到3′包含编辑后同源区、突变区和编辑前同源区。不希望受理论束缚,对模板RNA进行逆转录的RT首先逆转录编辑前同源区,然后是突变区,然后是编辑后同源区,从而产生包含所需突变的DNA链,其两侧均具有同源区。

[0419] 在一些实施例中,异源对象序列的长度是至少32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、120、140、160、180、200、500或1,000个核苷酸(nt),或长度是至少1、1.5、2、2.5、3、3.5、4、4.5、5、5.5、6、6.5、7、7.5、8、8.5、9、9.5或10千碱基。在一些实施例中,异源对象序列的长度是不超过33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、120、140、160、180、200、500、1,000或2000个核苷酸(nt),或长度是不超过20、15、10、9、8、7、6、5、4或3千碱基。在一些实施例中,异源对象序列的长度是30-1000、40-1000、50-1000、60-1000、70-1000、74-1000、75-1000、76-1000、77-1000、78-1000、79-1000、80-1000、85-1000、90-1000、100-1000、120-1000、140-1000、160-1000、180-1000、200-1000、500-1000、30-500、40-500、50-500、60-500、70-500、74-500、75-500、76-500、77-500、78-500、79-500、80-500、85-500、90-500、100-500、120-500、140-500、160-500、180-500、200-500、30-200、40-200、50-200、60-200、70-200、74-200、75-200、76-200、77-200、78-200、79-200、80-200、85-200、90-200、100-200、120-200、140-200、160-200、180-200、30-100、40-100、50-100、60-100、70-100、74-100、75-100、76-100、77-100、78-100、79-100、80-100、85-100或90-100个核苷酸(nt),或长度是1-20、1-15、1-10、1-9、1-8、1-7、1-6、1-5、1-4、1-3、1-2、2-20、2-15、2-10、2-9、2-8、2-7、2-6、2-5、2-4、2-3、3-20、3-15、3-10、3-9、3-8、3-7、3-6、3-5、3-4、4-20、4-15、4-10、4-9、4-8、4-7、4-6、4-5、5-20、5-15、5-10、5-9、5-8、5-7、5-6、6-20、6-15、6-10、6-9、6-8、6-7、7-20、7-15、7-10、7-9、7-8、8-20、8-15、8-10、8-9、9-20、9-15、9-10、10-15、10-20或15-20千碱基。在一些实施例中,异源对象序列的长度为10-100、10-90、10-80、10-70、10-60、10-50、10-40、10-30或10-20nt,例如长度为10-80、10-50或10-20nt,例如,长度为约10-20nt。在一些实施例中,异源对象序列的长度为8-30、9-25、10-20、11-16或12-15个核苷酸,例如长度为11-16nt。不希望受理论束缚,在一些实施例中,更大的插入大小、更大的编辑区域(例如,靶区域中第一编辑/取代和第二编辑/取代之间的

距离)和/或更多数量的所期望编辑(例如,异源对象序列与靶基因组的错配)可以产生更长的最佳异源对象序列。

[0420] 在某些实施例中,模板核酸包含定制的RNA序列模板,其可以鉴定、设计、工程改造和构建定制的RNA序列模板,以包含改变或指定宿主基因组功能的序列,例如通过将异源编码区引入基因组;影响或引起外显子结构/可变剪接,例如导致一个或多个外显子的外显子跳过;引起内源基因破坏,例如造成基因敲除;引起内源基因的转录激活;引起内源DNA的表观遗传调节;引起一个或多个可操作地连接的基因的上调,例如导致基因激活或过表达;引起一个或多个可操作地连接的基因的上调,例如造成基因敲除;等。在某些实施例中,可以将定制的RNA序列模板工程改造以包含编码外显子和/或转基因的序列,提供与转录因子激活剂、阻遏物、增强子等及其组合的结合位点。在一些实施例中,可以工程改造定制模板来编码核酸或肽标签,以在可操作地连接到靶位点的内源RNA转录本或内源蛋白质中表达。在其他实施例中,编码序列可以进一步用剪接供体位点、剪接受体位点、或聚A尾定制。

[0421] 系统的模板核酸(例如模板RNA)通常包含用于将所期望的序列写入靶DNA的对象序列(例如,异源对象序列)。对象序列可以是编码的或非编码的。模板核酸(例如,模板RNA)可以设计成在靶DNA基因座处产生插入、突变或缺失。在一些实施例中,模板核酸(例如,模板RNA)可被设计成导致插入靶DNA。例如,模板核酸(例如,模板RNA)可以含有异源序列,其中逆转录将导致异源序列插入靶DNA中。在其他实施例中,RNA模板可以设计为将缺失引入靶DNA。例如,模板核酸(例如,模板RNA)可以在所期望缺失的上游和下游匹配靶DNA,其中逆转录将导致从模板核酸(例如模板RNA)上游和下游序列的复制,而没有间插序列,例如导致间插序列的缺失。在其他实施例中,模板核酸(例如,模板RNA)可被设计为将编辑引入靶DNA。例如,模板RNA可以在一个或多个核苷酸除外的情况下匹配靶DNA序列,其中逆转录将导致这些编辑复制到靶DNA中,例如导致突变,例如转位或颠换突变。

[0422] 在一些实施例中,将对象序列写入靶位点会导致核苷酸的取代,例如,其中对象序列的全长对应于具有一个或多个错配碱基的靶位点的匹配长度。在一些实施例中,异源对象序列可以设计成使得可以发生序列改变的组合,例如同时添加和缺失、添加和取代、或缺失和取代。

[0423] 在一些实施例中,异源对象序列可包含开放阅读框或开放阅读框的片段。在一些实施例中,异源对象序列具有Kozak序列。在一些实施例中,异源对象序列具有内部核糖体进入位点。在一些实施例中,异源对象序列具有自切割肽,例如T2A或P2A位点。在一些实施例中,异源对象序列具有起始密码子。在一些实施例中,模板RNA具有剪接受体位点。在一些实施例中,模板RNA具有剪接供体位点。示例性剪接受体和剪接供体位点在WO 2016044416中进行了描述,其通过援引以其全文并入本文。示例性剪接受体位点序列是本领域技术人员已知的。在一些实施例中,模板RNA在终止密码子的下游具有微小RNA结合位点。在一些实施例中,模板RNA在开放阅读框的终止密码子下游具有聚A尾。在一些实施例中,模板RNA包含一个或多个外显子。在一些实施例中,模板RNA包含一个或多个内含子。在一些实施例中,模板RNA包含真核转录终止子。在一些实施例中,模板RNA包含增强的翻译元件或翻译增强元件。在一些实施例中,RNA包含人T细胞白血病病毒(HTLV-1)R区。在一些实施例中,RNA包含增强核输出的转录后调节元件,例如乙型肝炎病毒(HPRE)或土拨鼠肝炎病毒(WPRE)的转录后调节元件。

[0424] 在一些实施例中,异源对象序列可以含有非编码序列。例如,模板核酸(例如,模板RNA)可以包含调节元件,例如,启动子或增强子序列或miRNA结合位点。在一些实施例中,对象序列在靶位点的整合将导致内源基因的上调。在一些实施例中,对象序列在靶位点的整合将导致内源基因的下调。在一些实施例中,模板核酸(例如,模板RNA)包含组织特异性启动子或增强子,其中的每个可以是单向的或双向的。在一些实施例中,启动子是RNA聚合酶I启动子、RNA聚合酶II启动子或RNA聚合酶III启动子。在一些实施例中,启动子包含TATA元件。在一些实施例中,启动子包含B识别元件。在一些实施例中,启动子具有针对转录因子的一个或多个结合位点。

[0425] 在一些实施例中,模板核酸(例如,模板RNA)包含协调表观遗传修饰的位点。在一些实施例中,模板核酸(例如,模板RNA)包含染色质绝缘子。例如,模板核酸(例如模板RNA)包含CTCF位点或靶向用于DNA甲基化的位点。

[0426] 在一些实施例中,模板核酸(例如,模板RNA)包含由至少一个可操作地连接至效应子序列的调节区构成的基因表达单元。效应子序列可以是转录成RNA的序列(例如,编码序列或非编码序列,例如编码微小RNA的序列)。

[0427] 在一些实施例中,将模板核酸(例如,模板RNA)的异源对象序列插入靶基因组的内源内含子中。在一些实施例中,将模板核酸(例如,模板RNA)的异源对象序列插入靶基因组中,从而充当新的外显子。在一些实施例中,将异源对象序列插入靶基因组导致天然外显子的替换或天然外显子的跳过。

[0428] 在一些实施例中,将模板核酸(例如,模板RNA)的异源对象序列插入靶基因组的基因组安全港位点中,例如AAVS1、CCR5、ROSA26或白蛋白基因座中。在一些实施例中,使用基因修饰将CAR整合到T细胞受体α恒定(TRAC)基因座中(Eyquem等人Nature[自然]543,113-117(2017))。在一些实施例中,使用基因修饰系统将CAR整合到T细胞受体β恒定(TRBC)基因座中。许多其他安全港已通过计算方法鉴定(Pellenz等人Hum Gen Ther[人类基因疗法]30,814-828(2019))并且可用于基因修饰系统介导的整合。在一些实施例中,将模板核酸(例如,模板RNA)的异源对象序列添加到基因组的基因间或基因内区域中。在一些实施例中,将模板核酸(例如,模板RNA)的异源对象序列添加到基因组的内源活性基因的5′或3′的0.1kb、0.25kb、0.5kb、0.75,kb,1kb、2kb、3kb、4kb,5kb、7.5kb、10kb、15kb、20kb、25kb、50、75kb或100kb之内。在一些实施例中,将模板核酸(例如,模板RNA)的异源对象序列添加到基因组的内源启动子或增强子的5′或3′的0.1kb、0.25kb、0.5kb、0.75,kb,1kb、2kb、3kb、4kb,5kb、7.5kb、10kb、15kb、20kb、25kb、50、75kb或100kb之内。在一些实施例中,模板核酸(例如模板RNA)的异源对象序列可以是例如在50-50,000个碱基对之间(例如,在50-40,000bp之间,在500-30,000bp之间,在500-20,000bp之间,在100-15,000bp之间,在500-10,000bp之间,在50-10,000bp之间,在50-5,000bp之间。

[0429] 模板核酸(例如,模板RNA)可以设计成在靶DNA基因座处产生插入、突变或缺失。在一些实施例中,模板核酸(例如,模板RNA)可被设计成导致插入靶DNA。例如,模板核酸(例如,模板RNA)可以含有异源对象序列,其中逆转录将导致异源对象序列插入靶DNA中。在其他实施例中,RNA模板可以设计为将缺失写入靶DNA。例如,模板核酸(例如,模板RNA)可以在所期望缺失的上游和下游匹配靶DNA,其中逆转录将导致从模板核酸(例如模板RNA)上游和下游序列的复制,而没有间插序列,例如导致间插序列的缺失。在其他实施例中,模板核酸

(例如,模板RNA)可被设计为将编辑写入靶DNA。例如,模板RNA可以在一个或多个核苷酸除外的情况下匹配靶DNA序列,其中逆转录将导致这些编辑复制到靶DNA中,例如导致突变,例如转位或颠换突变。

[0430] 在一些实施例中,编辑前同源结构域包含与包含在靶核酸分子中的核酸序列具有至少100%序列同一性的核酸序列。

[0431] 在一些实施例中,编辑后同源结构域包含与包含在靶核酸分子中的核酸序列具有至少100%序列同一性的核酸序列。

[0432] PBS序列

[0433] 在一些实施例中,模板核酸(例如,模板RNA)包含PBS序列。在一些实施例中,PBS序列位于异源对象序列的3′并且与跟待由本文所述系统修饰的位点相邻的序列互补,或与跟待由系统/基因修饰多肽修饰的位点相邻的序列互补的序列包含不超过1、2、3、4或5个错配。在一些实施例中,PBS序列在靶核酸分子中切口位点的1、2、3、4、5、6、7、8、9或10个核苷酸内结合。在一些实施例中,PBS序列与靶核酸分子的结合允许启动靶引发的逆转录(TPRT),例如,3′同源结构域充当TPRT的引物。在一些实施例中,PBS序列的长度为3-5、5-10、10-30、10-25、10-20、10-19、10-18、10-17、10-16、10-15、10-14、10-13、10-12、10-11、11-30、11-25、11-20、11-19、11-18、11-17、11-16、11-15、11-14、11-13、11-12、12-30、12-25、12-20、12-19、12-18、12-17、12-16、12-15、12-14、12-13、13-30、13-25、13-20、13-19、13-18、13-17、13-16、13-15、13-14、14-30、14-25、14-20、14-19、14-18、14-17、14-16、14-15、15-30、15-25、15-20、15-19、15-18、15-17、15-16、16-30、16-25、16-20、16-19、16-18、16-17、17-30、17-25、17-20、17-19、17-18、18-30、18-25、18-20、18-19、19-30、19-25、19-20、20-30、20-25或25-30个核苷酸,例如长度为10-17、12-16或12-14个核苷酸。在一些实施例中,PBS序列的长度为5-20、8-16、8-14、8-13、9-13、9-12或10-12个核苷酸,例如长度为9-12个核苷酸。

[0434] 模板核酸(例如,模板RNA)可以与靶DNA有一些同源性。在一些实施例中,模板核酸(例如,模板RNA)的PBS序列结构域可以用作靶DNA的退火区,使得靶DNA被定位以引发模板核酸(例如,模板RNA)的逆转录。在一些实施例中,模板核酸(例如,模板RNA)具有在RNA的3′端的与靶DNA完全同源的至少2、3、4、5、6、7、8、9、10、11、12、13、14、15、20、25、30、35、40、45、50、60、70、80、90、100、110、120、130、140、150、175、200或更多个碱基。在一些实施例中,模板核酸(例如,模板RNA)具有例如在模板核酸(例如,模板RNA)的5′端的与靶DNA至少50%、60%、70%、80%、85%、90%、95%、97%、98%、99%或100%同源的至少2、3、4、5、6、7、8、9、10、11、12、13、14、15、20、25、30、35、40、45、50、60、70、80、90、100、110、120、130、140、150、175、200或更多个碱基。

[0435] 具有诱导活性的gRNA

[0436] 在一些实施例中,本文所述的gRNA(例如,作为模板RNA的一部分的gRNA或用于第二链切口的gRNA)具有诱导活性。可通过模板核酸例如模板RNA实现诱导活性,该模板核酸(除gRNA之外)还包含阻断结构域,其中部分或全部阻断结构域的序列至少部分互补于gRNA的一部分或全部。因此,阻断结构域能够与gRNA的一部分或全部杂交或基本上杂交。在一些实施例中,阻断结构域和诱导活性gRNA布置在模板核酸例如模板RNA上,使得gRNA可以采用第一构象(其中阻断结构域与gRNA杂交或基本上杂交)和第二构象(其中阻断结构域不与

gRNA杂交或基本上不杂交)。在一些实施例中,在第一构象中,gRNA不能结合基因修饰多肽(例如,模板核酸结合结构域、DNA结合结构域或核酸内切酶结构域(例如,CRISPR/Cas蛋白))或以与缺乏阻断结构域的其他方面相似的模板RNA相比亲和力显著降低的方式结合。在一些实施例中,在第二构象中,gRNA能结合基因修饰多肽(例如,模板核酸结合结构域、DNA结合结构域或核酸内切酶结构域(例如,CRISPR/Cas蛋白))。在一些实施例中,gRNA是处于第一构象还是第二构象可以影响基因修饰多肽(例如,基因修饰多肽包含的CRISPR/Cas蛋白)的DNA结合或核酸内切酶活性是否活跃。

[0437] 在一些实施例中,协调第二切口的gRNA具有诱导活性。在一些实施例中,协调第二切口的gRNA在模板被逆转录后诱导。在一些实施例中,gRNA与阻断结构域的杂交可以使用开放分子破坏。在一些实施例中,开放分子包含与gRNA的部分或全部或阻断结构域结合并抑制gRNA与阻断结构域杂交的药剂。在一些实施例中,开放分子包含核酸,例如,包含与gRNA、阻断结构域或两者部分地或完全地互补的序列。通过选择或设计合适的开放分子,提供的开放分子可以促进gRNA构象的变化,使其可以与CRISPR/Cas蛋白相关联并提供CRISPR/Cas蛋白的相关功能(例如,DNA结合和/或核酸内切酶活性)。不希望受理论束缚,在选定的时间和/或位置提供开放分子可以允许对gRNA、CRISPR/Cas蛋白或包含它们的基因修饰系统的活性进行空间和时间控制。在一些实施例中,开放分子包含对包含基因修饰多肽和/或模板核酸的细胞而言是外源的。在一些实施例中,开放分子包含内源药剂(例如,对于包含基因修饰多肽和/或模板核酸(其包含gRNA和阻断结构域)的细胞而言是内源的)。例如,可以选择诱导型gRNA、阻断结构域和开放分子,使得开放分子是在靶细胞或组织中表达的内源药剂,例如,从而确保基因修饰系统在靶细胞或组织中的活性。作为另一个实例,可以选择诱导型gRNA、阻断结构域和开放分子,使得开放分子在一个或多个非靶细胞或组织中不存在或基本上不表达,例如,从而确保基因修饰系统的活性在一种或多种非靶细胞或组织中不发生或基本上不发生,或与靶细胞或组织相比以降低的水平发生。示例性的阻断结构域、开放分子及其用途描述于PCT申请公开WO 2020044039 A1,其通过援引以其全文并入本文。在一些实施例中,模板核酸(例如模板RNA)可以包含一个或多个序列或结构,用于由基因修饰多肽的一个或多个组分(例如逆转录酶或RNA结合结构域和gRNA)结合。在一些实施例中,gRNA促进与基因修饰多肽的模板核酸结合结构域(例如,RNA结合结构域)的相互作用。在一些实施例中,gRNA将基因修饰多肽引导至匹配的靶序列,例如在靶细胞基因组中。

[0438] 基因修饰系统中的环状RNA和核酶

[0439] 预期在靶细胞内的配制、递送或基因修饰反应期间采用环状和/或线性RNA状态可能是有用的。因此,在本文所述的任何方面的一些实施例中,基因修饰系统包含一个或多个环状RNA(circRNA)。在本文所述任何方面的一些实施例中,基因修饰系统包含一种或多种线性RNA。在一些实施例中,本文所述的核酸(例如,模板核酸、编码基因修饰多肽的核酸分子或两者)是circRNA。在一些实施例中,环状RNA分子编码基因修饰多肽。在一些实施例中,将编码基因修饰多肽的circRNA分子递送至宿主细胞。在一些实施例中,环状RNA分子编码重组酶,例如,如本文所述。在一些实施例中,将编码重组酶的circRNA分子递送至宿主细胞。在一些实施例中,编码基因修饰多肽的circRNA分子在翻译之前被线性化(例如,在宿主细胞中,例如,在宿主细胞的细胞核中)。

[0440] 已发现环状RNA(circRNA)天然存在于细胞中,并且已发现其具有不同的功能,包括在人细胞中的非编码和蛋白编码作用。已显示,可以通过将自剪接内含子掺入RNA分子(或编码RNA分子的DNA),导致RNA环化来工程改造circRNA,并且工程改造的circRNA可以具有增强的蛋白质产生和稳定性(Wesselhoeft等人Nature Communications[自然通讯]2018)。在一些实施例中,基因修饰多肽由circRNA编码。在某些实施例中,模板核酸是DNA,例如dsDNA或ssDNA。在某些实施例中,circDNA包含模板RNA。

[0441] 在一些实施例中,circRNA包含一个或多个核酶序列。在一些实施例中,核酶序列被激活用于例如在宿主细胞中自切割,例如,从而导致circRNA的线性化。在一些实施例中,当镁浓度达到足以切割的水平时,核酶被激活,例如在宿主细胞中。在一些实施例中,在递送至宿主细胞之前,circRNA保持在低镁环境中。在一些实施例中,核酶是蛋白质反应性核酶。在一些实施例中,核酶是核酸反应性核酶。在一些实施例中,circRNA包含切割位点。在一些实施例中,circRNA包含第二切割位点。

[0442] 在一些实施例中,circRNA在靶细胞的细胞核中被线性化。在一些实施例中,细胞的细胞核中circRNA的线性化涉及细胞的细胞核中存在的组分,例如以激活切割事件。在一些实施例中,对核元件(例如核蛋白,例如基因组相互作用蛋白,例如表观遗传修饰因子,例如EZH2)有反应的核酶(例如来自B2或ALU元件的核酶)掺入例如基因修饰系统的circRNA中。在一些实施例中,circRNA的核定位导致核酶的自催化活性增加和circRNA的线性化。

[0443] 在一些实施例中,核酶与基因修饰系统的一个或多个其他组分是异源的。在一些实施例中,可诱导的核酶(例如,在本文所述的circRNA中)是合成产生的,例如,通过利用蛋白质配体反应性适体设计。已描述了利用烟草环斑病毒锤头状核酶的卫星RNA与MS2外壳蛋白适体的系统(Kennedy等人Nucleic Acids Res[核酸研究]42(19):12306-12321(2014),其通过援引以其全文并入本文),其在MS2外壳蛋白的存在下导致核酶活性的激活。在实施例中,这样的系统对定位于细胞质或细胞核的蛋白质配体产生反应。在一些实施例中,蛋白质配体不是MS2。已经描述了用于产生靶标配体的RNA适体的方法,例如,基于通过指数富集的配体系统进化(SELEX)(Tuerk和Gold,Science[科学]249(4968):505-510(1990);Ellington和Szostak,Nature[自然]346(6287):818-822(1990);其中每个的方法通过援引并入本文)并且在一些情况下得到计算机设计的帮助(Bell等人PNAS[美国国家科学院院刊]117(15):8486-8493,其方法通过援引并入本文)。因此,在一些实施例中,产生用于靶配体的适体并将其掺入合成核酶系统中,例如引发核酶介导的切割和circRNA线性化,例如在蛋白质配体存在下。在一些实施例中,circRNA线性化在细胞质中被引发,例如,使用与细胞质中的配体相关联的适体。在一些实施例中,circRNA线性化在细胞核中被引发,例如,使用与细胞核中的配体相关联的适体。在实施例中,细胞核中的配体包含表观遗传修饰因子或转录因子。在一些实施例中,引发线性化的配体以高于脱靶细胞的水平存在于中靶细胞中。

[0444] 还预期核酸反应性核酶系统可用于circRNA线性化。例如,在例如Penchovsky(Biotechnology Advances[生物技术进展]32(5):1015-1027(2014),通过援引并入本文)中描述了感测确定的靶核酸分子以引发核酶激活的生物传感器。通过这些方法,核酶自然折叠成非活性状态,并且仅在存在确定的靶核酸分子(例如,RNA分子)的情况下才被激活。在一些实施例中,基因修饰系统的circRNA包含在确定的靶核酸(例如RNA,例如mRNA、miRNA、指导RNA、gRNA、sgRNA、ncRNA、lncRNA、tRNA、snRNA或mtRNA)存在下被激活的核酸反

应性核酶。在一些实施例中,引发线性化的核酸以高于脱靶细胞的水平存在于中靶细胞中。

[0445] 在本文任一方面的一些实施例中,基因修饰系统掺入了一种或多种对目的靶组织或靶细胞具有可诱导特异性的核酶,例如,由目的靶组织或靶细胞中以较高水平存在的配体或核酸激活的核酶。在一些实施例中,基因修饰系统掺入对亚细胞区室(例如细胞核、核仁、细胞质或线粒体)具有可诱导特异性的核酶。在一些实施例中,核酶被以较高水平存在于靶亚细胞区室中的配体或核酸激活。在一些实施例中,基因修饰系统的RNA组分以circRNA的形式提供,例如其通过线性化激活。在一些实施例中,编码基因修饰多肽的circRNA的线性化激活了该分子进行翻译。在一些实施例中,激活基因修饰系统的circRNA组分的信号在中靶细胞或组织中以更高水平存在,例如,使得该系统在这些细胞中被特异性地激活。

[0446] 在一些实施例中,基因修饰系统的RNA组分以circRNA的形式提供,例如其通过线性化灭活。在一些实施例中,编码基因修饰多肽的circRNA通过切割和降解而灭活。在一些实施例中,编码基因修饰多肽的circRNA通过将翻译信号与多肽的编码序列分离的切割而灭活。在一些实施例中,灭活基因修饰系统的circRNA组分的信号在脱靶细胞或组织中以更高水平存在,使得该系统在这些细胞中被特异性地灭活。

[0447] 靶核酸位点

[0448] 在一些实施例中,在基因修饰之后,编辑序列周围的靶位点例如在少于约50%或10%的编辑事件中包含有限数量的插入或缺失,例如,如通过对靶位点的长读段扩增子测序所确定的,例如,如Karst等人(2020)bioRxiv doi.org/10.1101/645903(通过援引以其全文并入本文)中所述。在一些实施例中,靶位点不显示多个连续编辑事件,例如头对尾或头对头重复,例如,如通过靶位点的长读段扩增子测序确定的,例如,如Karst等人bioRxiv doi.org/10.1101/645903(2020)(其通过援引以其全文并入本文)中所述。在一些实施例中,靶位点包含对应于模板RNA的整合序列。在一些实施例中,靶位点在超过约1%或10%的事件中不包含由内源RNA产生的插入,例如,如通过靶位点的长读段扩增子测序所确定的,例如,如Karst等人bioRxiv doi.org/10.1101/645903(2020)(其通过援引以其全文并入本文)中所述。在一些实施例中,靶位点包含对应于模板RNA的整合序列。

[0449] 在本发明的某些方面,由基因修饰系统整合的宿主DNA结合位点可以在基因中、内含子中、外显子中、ORF中、在任何基因的编码区之外、在基因的调节区内、或在基因的调节区外。在其他方面,多肽可以结合一个或多于一个宿主DNA序列。

[0450] 在一些实施例中,基因修饰系统用于编辑多个等位基因中的靶基因座。在一些实施例中,基因修饰系统被设计用于编辑特定等位基因。例如,基因修饰多肽可以针对仅存在于一个等位基因上的特定序列,例如包含与靶等位基因(例如,gRNA或退火结构域)具有同源性的模板RNA,但不针对第二同源等位基因。在一些实施例中,基因修饰系统可以改变单倍型特异性等位基因。在一些实施例中,靶向特定等位基因的基因修饰系统优先靶向该等位基因,例如,对靶等位基因具有至少2、4、6、8或10倍的偏好。

[0451] 第二链切口

[0452] 在一些实施例中,本文所述的基因修饰系统包括对第一链进行切口的切口酶活性(例如,在基因修饰多肽中),和对靶DNA的第二链进行切口的切口酶活性(例如,在与基因修饰多肽分开的多肽中)。如本文所讨论的,不希望受理论束缚,对靶位点DNA的第一链进行切

口被认为提供了可被RT结构域用于逆转录模板RNA的序列(例如,异源对象序列)的3′OH。不希望受理论束缚,认为向第二链引入另外的切口可能使细胞DNA修复机制偏向于比原始基因组序列更频繁地采用基于异源对象序列的序列。在一些实施例中,第二链的另外切口由与第一链的切口相同的核酸内切酶结构域(例如,切口酶结构域)产生。在一些实施例中,相同的基因修饰多肽既对第一链进行切口,又对第二链进行切口。在一些实施例中,基因修饰多肽包含CRISPR/Cas结构域并且第二链的另外切口由另外的核酸(例如包含引导CRISPR/Cas结构域对第二链进行切口的第二gRNA)引导。在其他实施例中,另外的第二链切口由与第一链的切口不同的核酸内切酶结构域(例如,切口酶结构域)产生。在一些实施例中,该不同的核酸内切酶结构域位于另外的多肽中(例如,本发明的系统进一步包含另外的多肽),与基因修饰多肽分开。在一些实施例中,另外的多肽包含本文所述的核酸内切酶结构域(例如,切口酶结构域)。在一些实施例中,另外的多肽包含例如本文所述的DNA结合结构域。

[0453] 在本文中预期第二链切口相对于第一链切口出现的位置可影响以下一项或多项的程度:获得所期望的基因修饰DNA修饰,出现不期望的双链断裂(DSB),出现不期望的插入,或出现不期望的缺失。不希望受理论束缚,第二链切口可能以两个总体取向发生:向内切口和向外切口。

[0454] 在一些实施例中,在向内切口取向,RT结构域进行聚合(例如,使用模板RNA(例如,异源对象序列))远离第二链切口。在一些实施例中,在向内切口取向,第一链的切口的位置和第二链的切口的位置位于第一PAM位点和第二PAM位点之间(例如,在其中两个切口都由包含CRISPR/Cas结构域的多肽(例如,基因修饰多肽)产生的情况下)。在一些实施例中,在向内切口取向,第一链的该切口的位置和第二链的切口的位置在多肽和另外的多肽与靶DNA结合的位点之间。在一些实施例中,在向内切口取向,第二链的切口的位置相对于第一链的切口的位置位于多肽和另外的多肽的结合位点的同一侧。在一些实施例中,在向内切口取向,第一链的切口的位置和第二链的切口的位置位于PAM位点和距靶位点一定距离的位点之间。

[0455] 提供向内切口取向的基因修饰系统的实例包括:包含CRISPR/Cas结构域的基因修饰多肽、包含引导对第一链上的靶位点DNA进行切口的gRNA的模板RNA和另外的核酸(其包含在距离第一切口位置一定距离的位点处引导切口的另外的gRNA),其中第一切口的位置和第二切口的位置在两个gRNA引导基因修饰多肽所至的位点的PAM位点之间。作为另一个实例,提供向内切口取向的另一种基因修饰系统包含含有锌指分子和第一切口酶结构域的基因修饰多肽,其中锌指分子以引导第一切口酶结构域对靶位点的第一链进行切口的方式结合靶DNA;包含CRISPR/Cas结构域的另外的多肽,和包含gRNA的另外的核酸,该gRNA引导另外的多肽在第二链上与靶位点DNA相距一段距离的位点进行切口,其中第一切口的位置和第二切口的位置位于PAM位点和锌指分子结合的位点之间。作为另一个实例,提供向内切口取向的另一种基因修饰系统包含含有锌指分子和第一切口酶结构域的基因修饰多肽,其中锌指分子以引导第一切口酶结构域对靶位点的第一链进行切口的方式结合靶DNA;包含TAL效应子分子和第二切口酶结构域的另外的多肽,其中TAL效应子分子以引导另外的多肽对第二链进行切口的方式结合到距靶位点一定距离的位点,其中第一切口的位置和第二个切口的位置在TAL效应子分子结合的位点和锌指分子结合的位点之间。

[0456] 在一些实施例中,在向外切口取向,RT结构域进行聚合(例如,使用模板RNA(例如,

异源对象序列))朝向第二链切口。在一些实施例中,在向内切口取向,当第一和第二切口均由包含CRISPR/Cas结构域的多肽(例如,基因修饰多肽)产生时,第一PAM位点和第二PAM位点位于第一链的切口的位置和第二链的切口的位置之间。在一些实施例中,在向内切口取向,多肽(例如,基因修饰多肽)和另外的多肽结合至靶DNA上位于第一链的切口的位置和第二链的切口的位置之间的位点。在一些实施例中,在向内切口取向,第二链的切口的位置相对于第一链的切口的位置位于多肽和另外的多肽的结合位点的相对侧。在一些实施例中,在向内取向,PAM位点和距靶位点一定距离的位点位于第一链的切口的位置和第二链的切口的位置之间。

[0457] 提供向外切口取向的基因修饰系统的实例包括:包含CRISPR/Cas结构域的基因修饰多肽、包含引导对第一链上的靶位点DNA进行切口的gRNA的模板RNA和另外的核酸(其包含在距离第一切口位置一定距离的位点处引导切口的另外的gRNA),其中第一切口的位置和第二切口的位置在两个gRNA引导基因修饰多肽所至的位点的PAM位点之外(即PAM位点位于第一切口的位置和第二切口的位置之间)。作为另一个实例,提供向外切口取向的另一种基因修饰系统包含含有锌指分子和第一切口酶结构域的基因修饰多肽,其中锌指分子以引导第一切口酶结构域对靶位点的第一链进行切口的方式结合靶DNA;包含CRISPR/Cas结构域的另外的多肽,和包含gRNA的另外的核酸,该gRNA引导另外的多肽在第二链上与靶位点DNA相距一段距离的位点进行切口,其中第一切口的位置和第二个切口在PAM位点和锌指分子结合的位点之外(即PAM位点和锌指分子结合的位点在第一切口的位置和第二切口的位置之间)。作为另一个实例,提供向外切口取向的另一种基因修饰系统包含含有锌指分子和第一切口酶结构域的基因修饰多肽,其中锌指分子以引导第一切口酶结构域对靶位点的第一链进行切口的方式结合靶DNA;包含TAL效应子分子和第二切口酶结构域的另外的多肽,其中TAL效应子分子以引导另外的多肽对第二链进行切口的方式结合到距靶位点一定距离的位点,其中第一切口的位置和第二切口的位置在TAL效应子分子结合的位点和锌指分子结合的位点之外(即TAL效应子分子结合的位点和锌指分子结合的位点在第一切口的位置和第二切口的位置之间)。

[0458] 不希望受理论束缚,认为对于提供第二链切口的基因修饰系统,在一些实施例中优选向外切口取向。如本文所述,与向外切口取向相比,向内切口可产生更多数量的双链断裂(DSB)。DSB可以被细胞核中的DSB修复途径识别,这可能导致不期望的插入和缺失。向外切口取向可提供降低的DSB形成风险,和相应更少量的不期望的插入和缺失。在一些实施例中,不期望的插入和缺失是不由异源对象序列编码的插入和缺失,例如由与异源对象序列编码的修饰无关的双链断裂修复途径产生的插入或缺失。在一些实施例中,所期望的基因修饰包含对由异源对象序列编码(例如,以及通过基因修饰将异源对象序列写入靶位点来实现)的靶DNA的改变(例如,取代、插入或缺失)。在一些实施例中,第一链切口和第二链切口处于向外取向。

[0459] 此外,第一链切口和第二链切口之间的距离可能影响以下一项或多项的程度:获得所期望的基因修饰系统DNA修饰,出现不期望的双链断裂(DSB),出现不期望的插入,或出现不期望的缺失。不希望受理论束缚,认为第二链切口的益处,即DNA修复偏向于将异源对象序列掺入靶DNA中,随着第一链切口和第二链切口之间的距离减小而增加。然而,认为DSB形成风险也随着第一链切口和第二链切口之间的距离减小而增加。相应地,认为不期望的

插入和/或缺失的数量可能随着第一链切口和第二链切口之间的距离减小而增加。在一些实施例中,选择第一链切口和第二链切口之间的距离以平衡偏向将异源对象序列掺入靶DNA中的DNA修复的益处和DSB形成和不希望的缺失和/或插入的风险。在一些实施例中,相对于第一切口和第二切口相隔小于阈值距离的在其他方面类似的向内切口取向系统,第一链切口和第二链切口相隔至少阈值距离的系统具有增加水平的所期望基因修饰系统修饰结果、降低水平的不期望的缺失和/或降低水平的不期望的插入。在一些实施例中,一个或多个阈值距离在下面给出。

[0460] 在一些实施例中,第一切口和第二切口相隔至少20、25、30、35、40、45、50、55、60、65、70、75、80、85、90、95、100、110、120、130、140、150、160、170、180、190或200个核苷酸。在一些实施例中,第一切口和第二切口相隔不超过25、30、35、40、45、50、55、60、65、70,75、80、85、90、95、100、110、120、130、140、150、160、170、180、190、200或250个核苷酸。在一些实施例中,第一切口和第二切口相隔20-200、30-200、40-200、50-200、60-200、70-200、80-200、90-200、100-200、110-200、120-200、130-200、140-200、150-200、160-200、170-200、180-200、190-200、20-190、30-190、40-190、50-190、60-190、70-190、80-190、90-190、100-190、110-190、120-190、130-190、140-190、150-190、160-190、170-190、180-190、20-180、30-180、40-180、50-180、60-180、70-180、80-180、90-180、100-180、110-180、120-180、130-180、140-180、150-180、160-180、170-180、20-170、30-170、40-170、50-170、60-170、70-170、80-170、90-170、100-170、110-170、120-170、130-170、140-170、150-170、160-170、20-160、30-160、40-160、50-160、60-160、70-160、80-160、90-160、100-160、110-160、120-160、130-160、140-160、150-160、20-150、30-150、40-150、50-150、60-150、70-150、80-150、90-150、100-150、110-150、120-150、130-150、140-150、20-140、30-140、40-140、50-140、60-140、70-140、80-140、90-140、100-140、110-140、120-140、130-140、20-130、30-130、40-130、50-130、60-130、70-130、80-130、90-130、100-130、110-130、120-130、20-120、30-120、40-120、50-120、60-120、70-120、80-120、90-120、100-120、110-120、20-110、30-110、40-110、50-110、60-110、70-110、80-110、90-110、100-110、20-100、30-100、40-100、50-100、60-100、70-100、80-100、90-100、20-90、30-90、40-90、50-90、60-90、70-90、80-90、20-80、30-80、40-80、50-80、60-80、70-80、20-70、30-70、40-70、50-70、60-70、20-60、30-60、40-60、50-60、20-50、30-50、40-50、20-40、30-40或20-30个核苷酸。在一些实施例中,第一切口和第二切口相隔40-100个核苷酸。

[0461] 不希望受理论束缚,认为对于提供第二链切口并选择向内切口取向的基因修饰系统,增加第一链切口和第二链切口之间的距离可以是优选的。如本文所述,向内切口取向可以比向外切口取向产生更多数量的DSB,并且可以导致比向外切口取向更多量的不期望的插入和缺失,但是增加切口之间的距离可以减轻DSB、不期望的缺失和/或不期望的插入的这种增加。在一些实施例中,相对于第一切口和第二切口相隔小于阈值距离的在其他方面类似的向内切口取向系统,其中第一切口和第二切口相隔至少阈值距离的向内切口取向具有增加水平的所期望基因修饰系统修饰结果、减少水平的不期望的缺失和/或减少水平的不期望的插入。在一些实施例中,阈值距离在下面给出。

[0462] 在一些实施例中,第一链切口和第二链切口处于向内取向。在一些实施例中,第一链切口和第二链切口处于向内取向,并且第一链切口和第二链切口相隔至少100、110、120、

130、140、150、160、170、180、190、200、220、240、260、280、300、350、400、450或500个核苷酸，例如至少100个核苷酸，(并且任选地相隔不超过500、400、300、200、190、180、170，160、150、140、130或120个核苷酸)。在一些实施例中，第一链切口和第二链切口处于向内取向，并且第一链切口和第二链切口相隔100-200、110-200、120-200、130-200、140-200、150-200、160-200、170-200、180-200、190-200、100-190、110-190、120-190、130-190、140-190、150-190、160-190、170-190、180-190、100-180、110-180、120-180、130-180、140-180、150-180、160-180、170-180、100-170、110-170、120-170、130-170、140-170、150-170、160-170、100-160、110-160、120-160、130-160、140-160、150-160、100-150、110-150、120-150、130-150、140-150、100-140、110-140、120-140、130-140、100-130、110-130、120-130、100-120、110-120或100-110个核苷酸。

[0463] 化学修饰的核酸和核酸末端特征

[0464] 本文所述的核酸(例如模板核酸，例如模板RNA；或编码基因修饰多肽的核酸(例如，mRNA)；或gRNA)可以包含未修饰或经修饰的核碱基。天然存在的RNA从四种基本核糖核苷酸合成：ATP、CTP、UTP和GTP，但可以含有转录后修饰的核苷酸。此外，已经在RNA中鉴定了大约一百种不同的核苷修饰(Rozenski，J，Crain，P，和McCloskey，J.(1999).The RNA Modification Database:1999update.[RNA修饰数据库：1999年更新]Nucl Acids Res[核酸研究]27:196-197)。RNA还可包含自然界中不存在的完全合成核苷酸。

[0465] 在一些实施例中，化学修饰是在以下中提供的化学修饰：WO/2016/183482、美国专利公开号20090286852、国际申请号WO/2012/019168、WO/2012/045075、WO/2012/135805、WO/2012/158736、WO/2013/039857、WO/2013/039861、WO/2013/052523、WO/2013/090648、WO/2013/096709、WO/2013/101690、WO/2013/106496、WO/2013/130161、WO/2013/151669、WO/2013/151736、WO/2013/151672、WO/2013/151664、WO/2013/151665、WO/2013/151668、WO/2013/151671、WO/2013/151667、WO/2013/151670、WO/2013/151666、WO/2013/151663、WO/2014/028429、WO/2014/081507、WO/2014/093924、WO/2014/093574、WO/2014/113089、WO/2014/144711、WO/2014/144767、WO/2014/144039、WO/2014/152540、WO/2014/152030、WO/2014/152031、WO/2014/152027、WO/2014/152211、WO/2014/158795、WO/2014/159813、WO/2014/164253、WO/2015/006747、WO/2015/034928、WO/2015/034925、WO/2015/038892、WO/2015/048744、WO/2015/051214、WO/2015/051173、WO/2015/051169、WO/2015/058069、WO/2015/085318、WO/2015/089511、WO/2015/105926、WO/2015/164674、WO/2015/196130、WO/2015/196128、WO/2015/196118、WO/2016/011226、WO/2016/011222、WO/2016/011306、WO/2016/014846、WO/2016/022914、WO/2016/036902、WO/2016/077125、或WO/2016/077123，其中每个通过援引以其全文并入本文。应当理解，将化学修饰的核苷酸掺入多核苷酸可以导致将修饰掺入核碱基、主链或二者，这取决于该修饰在核苷酸中的位置。在一些实施例中，主链修饰是EP 2813570中提供的修饰，将其通过援引以其全文并入本文。在一些实施例中，经修饰的帽是美国专利公开号20050287539(其通过援引以其全文并入本文)中提供的帽。

[0466] 在一些实施例中，化学修饰的核酸(例如，RNA，例如，mRNA)包含一种或多种ARCA：抗反向帽类似物(m27.3′-OGP3G)、GP3G(未甲基化帽类似物)、m7GP3G(单甲基化帽类似物)、m32.2.7GP3G(三甲基化帽类似物)、m5CTP(5′-甲基-胞苷三磷酸)、m6ATP(N6-甲基-腺苷-

5′-三磷酸)、s2UTP(2-硫代-尿苷三磷酸和Ψ(假尿苷三磷酸)。

[0467] 在一些实施例中,化学修饰的核酸包含5′帽,例如:7-甲基鸟苷帽(例如,O-Me-m7G帽);超甲基化帽类似物;NAD+来源的帽类似物(例如,如Kiledjian,Trends in Cell Biology[细胞生物学趋势]28,454-464(2018)中所述);或经修饰的,例如生物素化的帽类似物(例如,Bednarek等人,Phil Trans R Soc B[伦敦皇家学会哲学汇刊b辑-生物科学]373,20180167(2018))中所述)。

[0468] 在一些实施例中,化学修饰的核酸包含选自以下中的一种或多种的3′特征:聚A尾;16个核苷酸长的茎环结构,其两侧为未配对的5个核苷酸(例如,Mannironi等人,Nucleic Acid Research[核酸研究]17,9113-9126(1989)中所述);三螺旋结构(例如,Brown等人,PNAS[美国国家科学院院刊]109,19202-19207(2012)所述);tRNA、Y RNA或穹窿RNA结构(例如,如Labno等人,Biochemica et Biophysica Acta[生物化学和生物物理学报]1863,3125-3147(2016)所述);掺入一个或多个脱氧核糖核苷酸三磷酸(dNTP)、2'O-甲基化NTP或硫代磷酸酯-NTP;单核苷酸化学修饰(例如,将3′末端核糖氧化为反应性醛,然后缀合醛反应性修饰的核苷酸);或化学连接到另一个核酸分子。

[0469] 在一些实施例中,核酸(例如,模板核酸)包含一个或多个经修饰的核苷酸,例如选自二氢尿苷、肌苷、7-甲基鸟苷、5-甲基胞苷(5mC)、5'磷酸核糖胸核苷、2'-O-甲基核糖胸核苷、2'-O-乙基核糖胸核苷、2'-氟核糖胸核苷、C-5丙炔基-脱氧胞苷(pdC)、C-5丙炔基-脱氧尿苷(pdU)、C-5丙炔基胞苷(pC)、C-5丙炔基尿苷(pU)、5-甲基胞苷、5-甲基尿苷、5-甲基脱氧胞苷、5-甲基脱氧尿苷甲氧基、2,6-二氨基嘌呤、5'-二甲氧基三苯甲基-N4-乙基-2'-脱氧胞苷、C-5丙炔基-f-胞苷(pfC)、C-5丙炔基-f-尿苷(pfU)、5-甲基f-胞苷、5-甲基f-尿苷、C-5丙炔基-m-胞苷(pmC)、C-5丙炔基-f-尿苷(pmU)、5-甲基m-胞苷、5-甲基m-尿苷、LNA(锁核酸)、MGB(小沟结合剂)假尿苷(Ψ)、1-N-甲基假尿苷(1-Me-Ψ)、或5-甲氧基尿苷(5-MO-U)。

[0470] 在一些实施例中,核酸包含主链修饰,例如对主链中的糖或磷酸基团的修饰。在一些实施例中,核酸包含核碱基修饰。

[0471] 在一些实施例中,核酸包含表9的一个或多个化学修饰的核苷酸、表10的一个或多个化学主链修饰、表11的一个或多个化学修饰的帽。例如,在一些实施例中,核酸包含两个或更多个(例如,3、4、5、6、7、8、9或10或更多个)不同类型的化学修饰。例如,核酸可以包含例如如本文所述的例如在表9中的两个或更多个(例如,3、4、5、6、7、8、9或10或更多个)不同类型的修饰的核碱基。可替代地或组合地,核酸可包含两个或更多个(例如,3、4、5、6、7、8、9或10或更多个)不同类型的主链修饰,例如,如本文所述,例如表10中所述。可替代地或组合地,核酸可包含一个或多个经修饰的帽,例如,如本文所述,例如表11中所述。例如,在一些实施例中,核酸包含一种或多种类型的经修饰的核碱基和一种或多种类型的主链修饰;一种或多种类型的经修饰的核碱基和一个或多个经修饰的帽;一种或多种类型的经修饰的帽和一种或多种类型的主链修饰;或一种或多种类型的经修饰的核碱基、一种或多种类型的主链修饰和一种或多种类型的经修饰的帽。

[0472] 在一些实施例中,核酸包含一个或多个(例如,2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90、100、150、200、250、300、350、400、450、500、600、700、800、900、1000、或更多个)经修饰的核碱基。在一些实施例中,核酸的所有核碱基都被修饰。在一些实施例中,在

主链中的一个或多个(例如,2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90、100、150、200、250、300、350、400、450、500、600、700、800、900、1000或更多个)位置修饰核酸。在一些实施例中,核酸的所有主链位置都被修饰。

[0473] 表9.修饰的核苷酸

[0474]

5-氮杂-尿苷	N2-甲基-6-硫代-鸟苷
2-硫代-5-氮杂-尿苷	N2,N2-二甲基-6-硫代-鸟苷
2-硫代尿苷	吡啶-4-酮核糖核苷
4-硫代-假尿苷	2-硫代-5-氮杂-尿苷
2-硫代-假尿苷	2-硫代尿苷
5-羟基尿苷	4-硫代-假尿苷
3-甲基尿苷	2-硫代-假尿苷
5-羧甲基-尿苷	3-甲基尿苷
1-羧甲基-假尿苷	1-丙炔基-假尿苷
5-丙炔基-尿苷	1-甲基-1-去氮杂-假尿苷
1-丙炔基-假尿苷	2-硫代-1-甲基-1-去氮杂-假尿苷

[0475]

5-牛磺基甲基尿苷	4-甲氧基-假尿苷
1-牛磺基甲基-假尿苷	5'-O-(1-硫代磷酸)-腺苷
5-牛磺基甲基-2-硫代-尿苷	5'-O-(1-硫代磷酸)-胞苷
1-牛磺基甲基-4-硫代-尿苷	5'-O-(1-硫代磷酸)-鸟苷
5-甲基-尿苷	5'-O-(1-硫代磷酸)-尿苷
1-甲基-假尿苷	5'-O-(1-硫代磷酸)-假尿苷
4-硫代-1-甲基-假尿苷	2'-O-甲基-腺苷
2-硫代-1-甲基-假尿苷	2'-O-甲基-胞苷
1-甲基-1-去氮杂-假尿苷	2'-O-甲基-鸟苷
2-硫代-1-甲基-1-去氮杂-假尿苷	2'-O-甲基-尿苷
二氢尿苷	2'-O-甲基-假尿苷
二氢假尿苷	2'-O-甲基-肌苷
2-硫代-二氢尿苷	2-甲基腺苷
2-硫代-二氢假尿苷	2-甲基硫代-N6-甲基腺苷
2-甲氧基尿苷	2-甲基硫代-N6 异戊烯基腺苷
2-甲氧基-4-硫代-尿苷	2-甲基硫代-N6-(顺式-羟基异戊烯基)腺苷
4-甲氧基-假尿苷	N6-甲基-N6-苏氨酰基氨甲酰基腺苷
4-甲氧基-2-硫代-假尿苷	N6-羟基正缬氨酰基氨甲酰基腺苷
5-氮杂-胞苷	2-甲基硫代-N6-羟基正缬氨酰基氨甲酰基腺苷
假异胞苷	
3-甲基-胞苷	2'-O-核糖基腺苷（磷酸）
N4-乙酰基胞苷	1,2'-O-二甲基肌苷
5-甲酰基胞苷	5,2'-O-二甲基胞苷
N4-甲基胞苷	N4-乙酰基-2'-O-甲基胞苷
5-羟基甲基胞苷	赖胞苷
1-甲基-假异胞苷	7-甲基鸟苷
吡咯并胞苷	N2,2'-O-二甲基鸟苷
吡咯并假异胞苷	N2,N2,2'-O-三甲基鸟苷
2-硫代-胞苷	2'-O-核糖基鸟苷（磷酸）
2-硫代-5-甲基-胞苷	怀丁苷
4-硫代-假异胞苷	过氧怀丁苷

[0476]

4-硫代-1-甲基-假异胞苷	羟基怀丁苷
4-硫代-1-甲基-1-去氮杂-假异胞苷	改性不足的（undermodified）羟基怀丁苷
1-甲基-1-去氮杂-假异胞苷	甲基怀俄苷
泽布拉林（zebularine）	辫苷
5-氮杂-泽布拉林	环氧辫苷
5-甲基-泽布拉林	半乳糖基-辫苷
5-氮杂-2-硫代-泽布拉林	甘露糖基-辫苷
2-硫代-泽布拉林	7-氰基-7-去氮杂鸟苷
2-甲氧基-胞苷	7-氨甲基-7-去氮杂鸟苷
2-甲氧基-5-甲基-胞苷	古嘌苷
4-甲氧基-假异胞苷	5,2'-O-二甲基尿苷
4-甲氧基-1-甲基-假异胞苷	4-硫代尿苷
2-氨基嘌呤	5-甲基-2-硫代尿苷
2,6-二氨基嘌呤	2-硫代-2'-O-甲基尿苷
7-去氮杂-腺嘌呤	3-(3-氨基-3-羧丙基)尿苷
7-去氮杂-8-氮杂-腺嘌呤	5-甲氧基尿苷
7-去氮杂-2-氨基嘌呤	尿苷 5-氧乙酸
7-去氮杂-8-氮杂-2-氨基嘌呤	尿苷 5-氧乙酸甲酯
7-去氮杂-2,6-二氨基嘌呤	5-(羧基羟甲基)尿苷）
7-去氮杂-8-氮杂-2,6-二氨基嘌呤	5-(羧基羟甲基)尿苷甲酯
1-甲基腺苷	5-甲氧基羰基甲基尿苷
N6-异戊烯基腺苷	5-甲氧基羰基甲基-2'-O-甲基尿苷
N6-(顺式-羟基异戊烯基)腺苷	5-甲氧基羰基甲基-2-硫代尿苷
2-甲基硫代-N6-(顺式-羟基异戊烯基)腺苷	5-氨甲基-2-硫代尿苷
	5-甲基氨甲基尿苷
N6-甘氨酰基氨甲酰基腺苷	5-甲基氨甲基-2-硫代尿苷
N6-苏氨酰基氨甲酰基腺苷	5-甲基氨甲基-2-硒尿苷
2-甲基硫代-N6-苏氨酰基氨甲酰基腺苷	5-氨甲酰基甲基尿苷
	5-氨甲酰基甲基-2'-O-甲基尿苷
N6,N6-二甲基腺苷	5-羧甲基氨甲基尿苷
7-甲基腺嘌呤	5-羧甲基氨甲基-2'-O-甲基尿苷

[0477]

2-甲基硫代-腺嘌呤	5-羧甲基氨甲基-2-硫代尿苷
2-甲氧基-腺嘌呤	N4,2'-O-二甲基胞苷
肌苷	5-羧甲基尿苷
1-甲基-肌苷	N6,2'-O-二甲基腺苷
怀俄苷	N,N6,O-2'-三甲基腺苷
怀丁苷	N2,7-二甲基鸟苷
7-去氮杂-鸟苷	N2,N2,7-三甲基鸟苷
7-去氮杂-8-氮杂-鸟苷	3,2'-O-二甲基尿苷
6-硫代-鸟苷	5-甲基二氢尿苷
6-硫代-7-去氮杂-鸟苷	5-甲酰基-2'-O-甲基胞苷
6-硫代-7-去氮杂-8-氮杂-鸟苷	1,2'-O-二甲基鸟苷
7-甲基-鸟苷	4-去甲基怀俄苷
6-硫代-7-甲基-鸟苷	异怀俄苷
7-甲基肌苷	N6-乙酰基腺苷
6-甲氧基-鸟苷	
1-甲基鸟苷	
N2-甲基鸟苷	
N2,N2-二甲基鸟苷	
8-氧代-鸟苷	
7-甲基-8-氧代-鸟苷	
1-甲基-6-硫代-鸟苷	

[0478] 表10.主链修饰

[0479]

2'-O-甲基主链
肽核酸（PNA）主链
硫代磷酸酯主链
吗啉代主链
氨基甲酸酯主链
硅氧烷主链
硫化物主链
亚砜主链

[0480]

砜主链
甲酰乙酰基主链
硫代甲酰乙酰基主链
亚甲基甲酰乙酰基主链
核糖乙酰基主链
含烯烃的主链
氨基磺酸酯主链
磺酸酯主链
磺酰胺主链
亚甲基亚氨基主链
亚甲基肼基主链
酰胺主链

[0481] 表11.经修饰的帽

[0482]

m7GpppA
m7GpppC
m2,7GpppG
m2,2,7GpppG
m7Gpppm7G
m7,2'OmeGpppG
m72'dGpppG
m7,3'OmeGpppG
m7,3'dGpppG
GppppG
m7GppppG
m7GppppA
m7GppppC
m2,7GppppG
m2,2,7GppppG
m7Gppppm7G
m7,2'OmeGppppG
m72'dGppppG

[0483]

m7,3'OmeGppppG
m7,3'dGppppG

[0484] 构成基因修饰系统的模板的核苷酸可以是天然碱基或经修饰的碱基,或其组合。例如,模板可以包含假尿苷、二氢尿苷、肌苷、7-甲基鸟苷或其他经修饰的碱基。在一些实施例中,模板可以包含锁核酸核苷酸。在一些实施例中,模板中使用的经修饰的碱基不抑制模板的逆转录。在一些实施例中,模板中使用的经修饰的碱基可以提高逆转录,例如特异性或保真度。

[0485] 在一些实施例中,系统的RNA组分(例如,模板RNA或gRNA)包含一个或多个核苷酸修饰。在一些实施例中,与未修饰或末端修饰的指导物相比,gRNA的修饰模式可显著影响体内活性(例如,如以下所示:来自Finn等人Cell Rep[细胞报道]22(9):2227-2235(2018)的图1D;其通过援引以其全文并入本文)。不希望受理论束缚,该过程可能至少部分归因于修饰赋予的RNA稳定性。这种修饰的非限制性实例可以包括2'-O-甲基(2'-O-Me)、2'-O-(2-甲氧基乙基)(2'-O-MOE)、2'-氟(2'-F)、核苷酸之间的硫代磷酸酯(PS)键、G-C取代以及核苷酸及其等价物之间的反向无碱基连接。

[0486] 在一些实施例中,模板RNA(例如,在其结合靶位点的部分)或指导RNA包含5′末端区域。在一些实施例中,模板RNA或指导RNA不包含5′末端区域。在一些实施例中,5′末端区域包含gRNA间隔子区,例如,如Briner AE等人,Molecular Cell[分子细胞]56:333-339(2014)中关于sgRNA所描述的(通过援引以其全文并入本文;适用于本文,例如,对于所有指导RNA)。在一些实施例中,5′末端区域包含5′端修饰。在一些实施例中,具有或不具有间隔子区的5′末端区域可以与crRNA、trRNA、sgRNA和/或dgRNA相关联。在一些情况下,gRNA间隔子区可以包含指导区、指导结构域或靶向结构域。

[0487] 在一些实施例中,本文所述的模板RNA(例如,在其结合靶位点的部分处)或指导RNA包含WO 2018107028 A1(通过援引以其全文并入本文)的表4中所示的任何序列。在一些实施例中,当序列显示指导区和/或间隔子区时,组合物可以包含或不包含该区域。在一些实施例中,指导RNA包含WO 2018107028 A1的表4中所示的任何序列(例如,如其中由SEQ ID NO表示)的一个或多个修饰。在实施例中,核苷酸可以相同或不同,和/或所示的修饰模式可以与WO 2018107028 A1的表4中所示的指导序列的修饰模式相同或相似。在一些实施例中,修饰模式包括gRNA或gRNA区域(例如5′末端区域、下部茎区、凸起区、上部茎区、连结区、发夹1区、发夹2区,3′末端区域)的修饰的相对位置和同一性。在一些实施例中,修饰模式包含WO 2018107028 A1的表4的序列栏中所示的任一序列的修饰和/或在该序列的一个或多个区域上的修饰的至少50%、55%、60%、70%、75%、80%、85%、90%、95%、96%、97%、98%、99%、或100%。在一些实施例中,修饰模式与WO 2018107028 A1的表4的序列栏中所示的任一序列的修饰模式至少50%、55%、60%、70%、75%、80%、85%、90%、95%、96%、97%、98%、99%或100%相同。在一些实施例中,修饰模式在WO 2018107028 A1的表4中所示序列的一个或多个区域(例如,在5'末端区域、下部茎区、凸起区、上部茎区、连结区、发夹1区、发夹2区,和/或3'末端区域)上至少50%、55%、60%、70%、75%、80%、85%、90%、95%、96%、97%、98%、99%、或100%相同。在一些实施例中,修饰模式与5'末端区域上序列的修饰模式至少50%、55%、60%、70%、75%、80%、85%、90%、95%、96%、97%、98%、99%、或100%相同。在一些实施例中,修饰模式在下部茎上至少50%、55%、60%、70%、75%、80%、85%、90%、95%、96%、97%、98%、99%、或100%相同。在一些实施例中,修饰模式在凸起上至少50%、55%、60%、70%、75%、80%、85%、90%、95%、96%、97%、98%、

99％或100％相同。在一些实施例中,修饰模式在上部茎上至少50％、55％、60％、70％、75％、80％、85％、90％、95％、96％、97％、98％、99％、或100％相同。在一些实施例中,修饰模式在连结上至少50％、55％、60％、70％、75％、80％、85％、90％、95％、96％、97％、98％、99％、或100％相同。在一些实施例中,修饰模式在发夹1上至少50％、55％、60％、70％、75％、80％、85％、90％、95％、96％、97％、98％、99％或100％相同。在一些实施例中,修饰模式在发夹2上至少50％、55％、60％、70％、75％、80％、85％、90％、95％、96％、97％、98％、99％或100％相同。在一些实施例中,修饰模式在3'末端上至少50％、55％、60％、70％、75％、80％、85％、90％、95％、96％、97％、98％、99％、或100％相同。在一些实施例中,修饰模式与WO 2018107028A1的表4的序列或这样的序列的区域(例如5′末端、下部茎、凸起、上部茎、连结、发夹1、发夹2、3′末端)的修饰模式例如在0、1、2、3、4、5、6或更多个核苷酸处不同。在一些实施例中,gRNA包含修饰,这些修饰与WO 2018107028A1的表4的序列的修饰例如在0、1、2、3、4、5、6或更多个核苷酸处不同。在一些实施例中,gRNA包含修饰,这些修饰与WO 2018107028 A1的表4的序列的区域(例如,5'末端、下部茎、凸起、上部茎、连结、发夹1、发夹2、3'末端)的修饰例如在0、1、2、3、4、5、6或更多个核苷酸处不同。

[0488] 在一些实施例中,模板RNA(例如,在其结合靶位点的部分)或gRNA包含2'-O-甲基(2'-O-Me)修饰的核苷酸。在一些实施例中,gRNA包含2'-O-(2-甲氧基乙基)(2'-O-moe)修饰的核苷酸。在一些实施例中,gRNA包含2'-氟(2'-F)修饰的核苷酸。在一些实施例中,gRNA包含核苷酸之间的硫代磷酸酯(PS)键。在一些实施例中,gRNA包含5'端修饰、3'端修饰或5'和3'端修饰。在一些实施例中,5'端修饰包含核苷酸之间的硫代磷酸酯(PS)键。在一些实施例中,5'端修饰包含2'-O-甲基(2'-O-Me)、2'-O-(2-甲氧基乙基)(2'-O-MOE)和/或2'-氟(2'-F)修饰的核苷酸。在一些实施例中,5'端修饰包含至少一个硫代磷酸酯(PS)键和2'-O-甲基(2'-O-Me)、2'-O-(2-甲氧基乙基)(2'-O-MOE)和/或2'-氟(2'-F)修饰的核苷酸中的一个或多个。端修饰可以包含硫代磷酸酯(PS)、2'-O-甲基(2'-O-Me)、2'-O-(2-甲氧基乙基)(2'-O-MOE)和/或2'-氟(2'-F)修饰。等效的端修饰也包含在本文所述的实施例中。在一些实施例中,模板RNA或gRNA包含端修饰与模板RNA或gRNA的一个或多个区域的修饰组合。用于保护RNA(例如gRNA)的其他示例性修饰和方法及其式在WO 2018126176 A1(其通过援引以其全文并入本文)中进行了描述。

[0489] 在一些实施例中,结构指导的且系统的方法用于将修饰(例如,2′-OMe-RNA、2′-F-RNA、和PS修饰)引入模板RNA或指导RNA,例如,如在Mir等人Nat Commun[自然通讯]9:2641(2018)(通过援引以其全文并入本文)中描述。在一些实施例中,2'-F-RNA的掺入增加了RNA:RNA或RNA:DNA双链体的热稳定性和核酸酶稳定性,例如,同时对C3'-内糖褶皱的干扰最小。在一些实施例中,在2'-OH对RNA:DNA双链体稳定性很重要的位置,2'-F可能比2′-OMe具有更好的耐受性。在一些实施例中,crRNA包含一个或多个不降低Cas9活性的修饰,例如C10、C20或C21(完全修饰的),例如,如Mir等人Nat Commun[自然通讯]9:2641(2018)(通过援引以其全文并入本文)的补充表1中所述。在一些实施例中,tracrRNA包含一个或多个不降低Cas9活性的修饰,例如Mir等人Nat Commun[自然通讯]9:2641(2018)的补充表1中所述的T2、T6、T7或T8(完全修饰的)。在一些实施例中,包含一个或多个修饰(例如,如本文所述)的crRNA可以与包含一个或多个修饰(例如C20和T2)的tracrRNA配对。在一些实施例中,gRNA包含例如crRNA和tracrRNA的嵌合体(例如,Jinek等人Science[科学]337(6096):816-

821(2012))。在实施例中,来自crRNA和tracrRNA的修饰被映射到单指导嵌合体上,例如,以产生具有增强稳定性的经修饰的gRNA。

[0490] 在一些实施例中,gRNA分子可以通过添加或减少天然存在的结构组分例如发夹来修饰。在一些实施例中,gRNA可包含缺失了一个或多个3'发夹元件的gRNA,例如,如WO 2018106727(通过援引以其全文并入本文)中所述。在一些实施例中,gRNA可以包含添加的发夹结构,例如,在间隔子区中添加的发夹结构,其在Kocak等人Nat Biotechnol[自然生物技术]37(6):657-666(2019)的教导中显示增加CRISPR-Cas系统的特异性。另外的修饰,包括缩短的gRNA和提高体内活性的特定修饰的实例可以在US 20190316121(通过援引以其全文并入本文)中找到。

[0491] 在一些实施例中,结构指导的且系统的方法(例如,如以下中所述:Mir等人Nat Commun[自然通讯]9:2641(2018);通过援引以其全文并入本文)用于寻找模板RNA的修饰。在实施例中,通过包含或排除模板RNA的指导区来鉴定修饰。在一些实施例中,与模板RNA结合的多肽结构用于确定RNA的非蛋白质接触核苷酸,然后可以选择这些核苷酸进行修饰,例如,其中破坏RNA与多肽结合的风险较低。模板RNA中的二级结构也可以通过软件工具在计算机上预测,例如RNAstructure工具可在以下获得:rna.urmc.rochester.edu/RNAstructureWeb(Bellaousov等人Nucleic Acids Res[核酸研究]41:W471-W474(2013);通过援引以其全文并入本文),例如,以确定用于选择修饰的二级结构,例如发夹、茎和/或凸起。

[0492] 组合物和系统的产生

[0493] 如本领域技术人员将理解的那样,设计和构建核酸构建体和蛋白质或多肽(例如本文所述的系统、构建体和多肽)的方法在本领域中是常规的。通常,可以使用重组方法。通常,参见Smales和James(编辑),Therapeutic Proteins:Methods and Protocols[治疗性蛋白:方法和方案](Methods in Molecular Biology[分子生物学方法]),Humana Press[胡玛纳出版社](2005);以及Crommelin,Sindelar和Meibohm(编辑),Pharmaceutical Biotechnology:Fundamentals and Applications[药物生物技术:基础与应用],Springer[斯普林格出版社](2013)。设计、制备、评价、纯化和操纵核酸组合物的方法描述于Green和Sambrook(编辑),Molecular Cloning:A Laboratory Manual[分子克隆:实验室手册](第四版),Cold Spring Harbor Laboratory Press[冷泉港实验室出版社](2012)。

[0494] 本披露部分地提供了编码本文所述的基因修饰多肽、本文所述的模板核酸、或两者的核酸(例如,载体)。在一些实施例中,载体包含选择性标志物,例如,抗生素抗性标志物。在一些实施例中,抗生素抗性标志物是卡那霉素抗性标志物。在一些实施例中,抗生素抗性标志物不赋予对β-内酰胺抗生素的抗性。在一些实施例中,载体不包含氨苄西林抗性标志物。在一些实施例中,载体包含卡那霉素抗性标志物而不包含氨苄西林抗性标志物。在一些实施例中,将编码基因修饰多肽的载体整合到靶细胞基因组中(例如,在施用于靶细胞、组织、器官或受试者后)。在一些实施例中,不将编码基因修饰多肽的载体整合到靶细胞基因组中(例如,在施用于靶细胞、组织、器官或受试者后)。在一些实施例中,编码模板核酸(例如,模板RNA)的载体没有整合到靶细胞基因组中(例如,在施用于靶细胞、组织、器官或受试者后)。在一些实施例中,如果将载体整合到靶细胞基因组中的靶位点中,则不将选择性标志物整合到基因组中。在一些实施例中,如果将载体整合到靶细胞基因组中的靶位点

中,则不将参与载体维持的基因或序列(例如,质粒维持基因)整合到基因组中。在一些实施例中,如果将载体整合到靶细胞基因组中的靶位点中,则不将转移调节序列(例如,反向末端重复序列,例如,来自AAV)整合到基因组中。在一些实施例中,向靶细胞、组织、器官或受试者施用载体(例如,编码本文所述的基因修饰多肽、本文所述的模板核酸、或两者的载体)可使载体的部分整合到所述靶细胞、组织、器官或受试者的一个或多个基因组中的一个或多个靶位点中。在一些实施例中,包含整合材料的少于99%、95%、90%、80%、70%、60%、50%、40%、30%、20%、10%、5%、4%、3%、2%、或1%的靶位点(例如,没有靶位点)包含来自载体的选择性标志物(例如,抗生素抗性基因)、转移调节序列(例如,反向末端重复序列,例如,来自AAV)、或两者。

[0495] 用于产生本文所述的治疗性药物蛋白质或多肽的示例性方法涉及在哺乳动物细胞中表达,尽管也可以使用昆虫细胞、酵母、细菌、或其他细胞,在适当的启动子控制下,产生重组蛋白。哺乳动物表达载体可以包含非转录元件,如复制起点、合适的启动子、以及其他5'或3'侧翼非转录序列;以及5'或3'非翻译序列,如必要的核糖体结合位点、聚腺苷酸化位点、剪接供体和受体位点、以及终止序列。源自SV40病毒基因组的DNA序列,例如SV40起点、早期启动子、剪接和聚腺苷酸化位点可以用于提供异源DNA序列表达所需的其他遗传元件。在以下文献中描述了用于与细菌、真菌、酵母、和哺乳动物细胞宿主一起使用的适当的克隆和表达载体:Green和Sambrook,Molecular Cloning:A Laboratory Manual[分子克隆:实验室手册](第四版),Cold Spring Harbor Laboratory Press[冷泉港实验室出版社](2012)。

[0496] 各种哺乳动物细胞培养系统可以用于表达和制造重组蛋白。哺乳动物表达系统的实例包括CHO、COS、HEK293、HeLA和BHK细胞系。在以下文献中描述了用于生产蛋白治疗剂的宿主细胞培养的过程:Zhou和Kantardjieff(编辑),Mammalian Cell Cultures for Biologics Manufacturing[用于生物制品制造的哺乳动物细胞培养](Advances in Biochemical Engineering/Biotechnology[生物化学工程/生物科技的进展]),Springer[斯普林格出版社](2014)。本文所述的组合物可包括载体,例如编码重组蛋白的病毒载体,例如慢病毒载体。在一些实施例中,载体,例如病毒载体,可以包含编码重组蛋白的核酸。

[0497] 在以下文献中描述了蛋白治疗剂的纯化:Franks,Protein Biotechnology:Isolation,Characterization,and Stabilization[蛋白生物技术:分离、表征、和稳定化],Humana Press[胡玛纳出版社](2013);以及Cutler,Protein Purification Protocols[蛋白纯化方案](Methods in Molecular Biology[分子生物学方法]),Humana Press[胡玛纳出版社](2010)。

[0498] 本披露还提供了用于产生对基因修饰多肽和/或基因组靶位点具有特异性的模板核酸分子(例如,模板RNA)的组合物和方法。在一方面,该方法包括产生RNA区段,包括上游同源区段、异源对象序列区段、基因修饰多肽结合基序和gRNA区段。

[0499] 应用

[0500] 治疗性应用

[0501] 在一些实施例中,如本文所述的基因修饰系统可用于修饰细胞(例如,动物细胞、植物细胞或真菌细胞)。在一些实施例中,如本文所述的基因修饰系统可用于修饰哺乳动物细胞(例如,人细胞)。在一些实施例中,如本文所述的基因修饰系统可用于修饰来自家畜动

物(例如,牛、马、绵羊、山羊、猪、美洲驼、羊驼、骆驼、牦牛、鸡、鸭、鹅或鸵鸟)的细胞。在一些实施例中,如本文所述的基因修饰系统可用作实验室工具或研究工具,或用于实验室方法或研究方法中,例如以修饰动物细胞例如哺乳动物细胞(例如,人细胞)、植物细胞或真菌细胞。

[0502] 通过将编码基因整合到RNA序列模板中,基因修饰系统可以满足治疗需求,例如,通过在具有功能丧失性突变的个体中提供治疗性转基因的表达,通过以正常转基因代替功能获得性突变,通过提供调节序列以消除功能获得性突变表达,和/或通过控制可操作地连接的基因、转基因及其系统的表达。在某些实施例中,RNA序列模板编码对宿主细胞的治疗需要具有特异性的启动子区,例如组织特异性启动子或增强子。在又其他实施例中,启动子可以可操作地连接至编码序列。

[0503] 在一些实施例中,插入、缺失、取代或其组合增加或减少靶基因的表达(例如转录或翻译)。在一些实施例中,插入、缺失、取代或其组合通过改变、添加或缺失启动子或增强子中的序列(例如结合转录因子的序列)来增加或减少靶基因的表达(例如转录或翻译)。在一些实施例中,插入、缺失、取代或其组合改变靶基因的翻译(例如改变氨基酸序列),插入或缺失起始或终止密码子,改变或固定基因的翻译框架。在一些实施例中,插入、缺失、取代或其组合改变靶基因的剪接,例如通过插入、缺失或改变剪接受体或供体位点。在一些实施例中,插入、缺失、取代或其组合改变转录本或蛋白质半衰期。在一些实施例中,插入、缺失、取代或其组合改变、增加、降低靶基因的活性,例如由靶基因编码的蛋白质的活性。

[0504] 补偿性编辑

[0505] 在一些实施例中,本文提供的系统或方法可用于引入补偿性编辑。在一些实施例中,补偿性编辑位于与疾病或障碍相关的基因的位置,该位置不同于引起疾病的突变的位置。在一些实施例中,补偿性突变不在包含致病突变的基因中。在一些实施例中,补偿性编辑可以取消或补偿引起疾病的突变。在一些实施例中,可以通过本文提供的系统或方法引入补偿性编辑以抑制或逆转引起疾病的突变的突变效应。

[0506] 调节性编辑

[0507] 在一些实施例中,本文提供的系统或方法可用于引入调节性编辑。在一些实施例中,将调节性编辑引入基因的调节序列,例如基因启动子、基因增强子、基因阻遏物或调节基因剪接的序列。在一些实施例中,调节性编辑增加或降低靶基因的表达水平。在一些实施例中,靶基因与含有引起疾病的突变的基因相同。在一些实施例中,靶基因不同于含有引起疾病的突变的基因。

[0508] 重复序列扩增疾病

[0509] 在一些实施例中,本文提供的系统或方法可用于治疗重复序列扩增疾病。在一些实施例中,本文提供的系统或方法,例如包含基因修饰多肽的系统或方法,可用于通过根据定制的RNA模板重置基因座处的重复序列数量来治疗重复序列扩增疾病。

[0510] 治疗适应症

[0511] 在一些实施例中,本文提供的系统或方法可用于治疗下表12-15中任一个的适应症。例如,在一些实施例中,基因修饰系统修饰细胞中基因组DNA中的靶位点,其中靶位点在表12-15中任一个的基因中,例如在具有表12-15中任一个中列出的相应适应症的受试者中的基因。在一些实施例中,细胞是肝细胞并且靶位点在表12的基因中,例如在具有表12中列

出的相应适应症的受试者中的基因。在一些实施例中,细胞是造血干细胞(HSC)并且靶位点在表13的基因中,例如在具有表13中列出的相应适应症的受试者中的基因。在一些实施例中,细胞是中枢神经系统(CNS)细胞并且靶位点在表14的基因中,例如在具有表14中列出的相应适应症的受试者中的基因。在一些实施例中,细胞是眼睛的细胞并且靶位点在表15的基因中,例如在具有表15中列出的相应适应症的受试者中的基因。在一些实施例中,靶位点在基因的编码区中。在一些实施例中,靶位点在启动子中。在一些实施例中,靶位点在表12-15中任一个的基因的5'UTR或3'UTR中。在一些实施例中,靶位点在基因的内含子或外显子中。在一些实施例中,基因修饰系统纠正基因中的突变。在一些实施例中,基因修饰多肽插入已经从基因中缺失(例如,通过引起疾病的突变)的序列。在一些实施例中,基因修饰系统缺失已在基因中复制(例如,通过引起疾病的突变)的序列。在一些实施例中,基因修饰系统用相应的野生型序列替换突变(例如,引起疾病的突变)。在一些实施例中,突变是取代、插入、缺失或倒位。

[0512] 表12:适应症和遗传靶标,例如在肝脏中

[0513]

疾病	受影响的基因
急性间歇性卟啉症	HMBS
α-1-抗胰蛋白酶缺乏症（AAT）	SERPINA1
精氨酸酶缺乏症	ARG1
精氨琥珀酸裂解酶缺乏症	ASL
氨甲酰磷酸合成酶 I 缺乏症	CPS1
希特林缺乏症（Citrin deficiency）	SLC25A13
瓜氨酸血症 I 型	ASS1
克里格勒-纳贾尔综合征（高胆红素血症）	UGT1A1
法布里病	GLA
家族性高胆固醇血症 4（纯合子家族性胆固醇血症）	LDLRAP1
戊二酸血症 I	GCDH
戊二酸血症 II（多酰基-CoA 脱氢酶缺乏症）	GA IIA：ETFA GA IIB：ETFB GA IIC：ETFDH
IV 型糖原贮积病	GBE1
A 型血友病	F8
B 型血友病	F9
遗传性血色素沉着症	HFE
同型胱氨酸尿症	CBS
枫糖尿病（MSUD）	Ia 型：BCKDHA Ib 型：BCKDHB II 型：DBT
甲基丙二酸血症（甲基丙二酰-CoA 变位酶缺乏症）	MMUT
MPS 1S（施艾综合征（Scheie syndrome））	IDUA
MPS2	IDS
MPS3（圣菲利波综合征（San Filippo Syndrome））	IIIa 型：SGSH

[0514]

疾病	受影响的基因
	IIIb 型：NAGLU IIIc 型：HGSNAT IIId 型：GNS
MPS4	IVA 型：GALNS IVB 型：GLB1
MPS6	ARSB
MPS7	GUSB
鸟氨酸转氨甲酰酶缺乏症	OTC
苯丙酮尿症（苯丙氨酸羟化酶缺乏症）	PAH
多囊肝病	PCLD1：PRKCSH PCLD2：SEC63 PLCD3：ALG8 PCLD4：LRP5
庞贝病	GAA
原发性高草酸尿症 1（草酸中毒）	AGXT
进行性家族性肝内胆汁淤积 1 型	ATP8B1
进行性家族性肝内胆汁淤积 2 型	ABCB11
进行性家族性肝内胆汁淤积 3 型	ABCB4
丙酸血症	PCCB；PCCA
丙酮酸羧化酶缺乏症	PC
酪氨酸血症 I 型	FAH
威尔森氏病（Wilson's disease）	ATP7B

[0515] 表13:HSC的适应症和遗传靶标

[0516]

疾病	受影响的基因
肾上腺脑白质营养不良（CALD）	ABCD1
α-甘露糖苷贮积症	MAN2B1
布-戴二氏（Blackfan-Diamond）贫血	
先天性无巨核细胞的血小板过低症	MPL

[0517]

疾病	受影响的基因
先天性角化不良	TERC
范科尼贫血	FANC
戈谢病	GBA
球形细胞脑白质营养不良（克拉伯病）	GALC
噬血细胞性淋巴组织细胞增生症	PRF1；STX11；STXBP2；UNC13D
恶性婴儿骨硬化症-常染色体隐性骨硬化症	涉及许多基因
异染性脑白质营养不良	PSAP
MPS 1S（施艾综合征）	IDUA
MPS2	IDS
MPS7	GUSB
黏脂贮积症 II	GNPTAB
尼曼-皮克病 A 和 B	SMPD1
尼曼-皮克病 C	NPC1
庞贝病	GAA
丙酮酸激酶缺陷（PKD）	PKLR
镰状细胞病（SCD）	HBB
泰-萨二氏	HEXA
地中海贫血	HBB

[0518] 表14:CNS的适应症和遗传靶标

[0519]

疾病	受影响的基因
α-甘露糖苷贮积症	MAN2B1
共济失调毛细血管扩张症	ATM
CADASIL	NOTCH3
卡纳万（Canavan）病	ASPA
氨基甲酰磷酸合成酶 1 缺乏症	CPS1
CLN1 病	PPT1
CLN2 疾病	TPP1

[0520]

疾病	受影响的基因
CLN3 病（幼年神经元蜡样脂褐质储积症，贝敦氏（Batten）症）	CLN3
科-勒综合征	RPS6KA3
先天性肌无力综合征 5	COLQ
德朗热综合征（NIPBL）	NIPBL
德朗热综合征（SMC1A）	SMC1A
Dravet 综合征（SCN1A）	SCN1A
甘氨酸脑病（GLDC）	GLDC
GM1 神经节苷脂沉积症	GLB1
亨廷顿病	HTT
脑积水伴中脑导水管狭窄	L1CAM
亚急性坏死性脑脊髓病	SURF1
异染性脑白质营养不良（ARSA）	ARSA
MPS 2 型	IDS
MPS 3 型	类型 3a：SGSH 类型 3b：NAGLU
黏脂贮积症 IV 型	MCOLN1
1 型神经纤维瘤病	NF1
2 型神经纤维瘤病	NF2
泛酸激酶神经退行性疾病	PANK2
吡哆醇依赖性癫痫	ALDH7A1
雷特（Rett）综合征（MECP2）	MECP2
桑德霍夫（Sandhoff）病	HEXB
语义痴呆（额颞叶痴呆）	MAPT
脊髓小脑性共济失调伴轴突神经病变（动眼不能共济失调综合征）	SETX
泰-萨氏病（Tay-Sachs disease）	HEXA
X 连锁肾上腺脑白质营养不良	ABCD1

[0521] 表15:眼睛的适应症和遗传靶标

[0522]

疾病	受影响的基因
色盲	CNGB3
先天性黑蒙（LCA1）	GUCY2D
先天性黑蒙（LCA10）	CEP290
先天性黑蒙（LCA2）	RPE65
先天性黑蒙（LCA8）	CRB1
无脉络膜血症	CHM
锥体杆体营养不良（ABCA4）	ABCA4
锥体杆体营养不良（GUCY2D）	GUCY2D
非肾病性胱氨酸眼病	CTNS
多英蜂窝状视网膜营养不良（DHRD）	EFEMP1
家族性肾淀粉样变性	TTR
角膜炎-鱼鳞病-耳聋（KID）	GJB2
晶格状角膜营养不良 I 型	TGFBI
斑点状角膜营养不良（MCD）	CHST6
Meesmann 角膜营养不良	KRT12；KRT3
视神经萎缩	OPA1
色素性视网膜炎（AR）	USH2A
色素性视网膜炎（AD）	RHO
索斯比眼底营养不良（Sorsby Fundus Dystrophy）	TIMP3
斯特格（Stargardt）病	ABCA4

[0523] 应用于植物

[0524] 在一些实施例中,本文提供的系统或方法可用于修饰植物或植物部分(例如,叶、根、花、果实或种子)例如以增加植物的适应度。

[0525] 向植物的递送

[0526] 本文提供了将本文所述的基因修饰系统递送至植物的方法。包括用于通过使植物或其一部分与基因修饰系统接触而将基因修饰系统递送至植物的方法。这些方法可用于修饰植物以例如增加植物的适应度。

[0527] 更特别地,在一些实施例中,本文所述的核酸(例如,编码基因修饰系统的核酸)可以在载体中编码,例如邻近植物启动子(例如,植物载体(例如,pHUC411)中的玉蜀黍泛素启

动子(ZmUBI))而插入。在一些实施例中,本文所述的核酸通过农杆菌被引入植物(例如,粳稻)或植物的一部分(例如,植物的愈伤组织)中。在一些实施例中,本文所述的系统和方法可以通过用无效等位基因(例如,在起始密码子处含有碱基取代)替换植物基因(例如,潮霉素磷酸转移酶(HPT))而用于植物。以下中描述了用于修饰植物基因组的系统和方法:Xu等人Development of plant prime-editing systems for precise genome editing,[开发用于精确基因组编辑的植物先导编辑系统],2020,Plant Communications[植物通讯]。

[0528] 一方面,本文提供了一种增加植物的适应度的方法,该方法包括向该植物递送本文所述的基因修饰系统(例如,以有效的量和持续时间)以相对于未经处理的植物(例如,未递送该基因修饰系统的植物)增加该植物的适应度。

[0529] 由于递送基因修饰系统而产生的植物适应度的增加能以多种方式表现出来,例如,从而导致植物的更好的生产,例如改善的产率,改善的植物活力或从植物中收获的产物的质量,农业或园艺业所希望的收获前或收获后的性状(例如,味道、外观、货架期)的改善,或在其他方面使人类受益的性状(例如,减少变应原产生)的改善。改善的植物产率涉及相对于在相同条件下但不应用本发明组合物而生产的植物的相同产品的产率或与应用常规植物修饰剂相比,按可测量量计植物的产品的产率的增加(例如,如通过植物生物质、谷物、种子或果实产率、蛋白质含量、碳水化合物或油含量或叶面积测量的)。例如,产率可以增加至少约0.5%、约1%、约2%、约3%、约4%、约5%、约10%、约20%、约30%、约40%、约50%、约60%、约70%、约80%、约90%、约100%、或大于100%。在一些情况下,相对于未处理的植物,该方法有效地将产率增加约2x倍、5x倍、10x倍、25x倍、50x倍、75x倍、100x倍、或大于100x倍。产率可以以在某种基础上植物或植物的产品的重量或体积计的量来表示。基础可以以时间、生长面积、生产的植物的重量、或所用原材料的量来表示。例如,这样的方法可以增加植物组织的产率,这些植物组织包括但不限于:种子、果实、仁、圆荚、块茎、根和叶。

[0530] 由于递送基因修饰系统而产生的植物适应度的增加也可以通过其他手段来测量,如相对于在相同条件下但不施用本发明组合物或应用常规植物修饰剂而生产的植物的相同因素,以活力等级的增加或改善、植株密度(stand)(每单位面积的植物数量)、植物高度、秆围、秆长、叶数量、叶尺寸、植物冠层、视觉外观(诸如更绿的叶颜色)、根等级、出苗、蛋白质含量、分蘖的增加、更大的叶、更多的叶、更少的死的基生叶、分蘖更强、所需肥料更少、所需种子更少、分蘖更多产、开花更早、提早的谷物或种子成熟度、更少的植物节(verse)(倒伏)、芽生长的增加、更早萌发、或这些因素的任何组合,按可测量或可察觉的量来测量。

[0531] 因此,本文提供了一种修饰植物的方法,该方法包括向植物递送有效量的本文提供的基因修饰系统中的任一种,其中该方法修饰该植物并由此相对于未经处理的植物引入或增加该植物中的有益性状(例如,增加约1%、2%、5%、10%、20%、30%、40%、50%、60%,70%、80%、90%、100%或大于100%)。特别地,该方法相对于未经处理的植物可以增加植物的适应度(例如,增加约1%、2%、5%、10%、20%、30%、40%、50%、60%、70%、80%、90%、100%或大于100%)。

[0532] 在一些情况下,植物适应度的增加是以下方面的增加(例如,增加约1%、2%、5%、10%、20%、30%、40%、50%、60%、70%、80%、90%、100%或大于100%):抗病性、耐旱性、耐热性、耐寒性、耐盐性、金属耐受性、除草剂耐受性、化学耐受性、水分利用效率、氮利用、对氮胁迫的抗性、固氮、有害生物抗性、草食动物抗性、病原体抗性、产率、限水条件下的产

率、活力、生长、光合能力、营养、蛋白质含量、碳水化合物含量、油含量、生物质、芽长、根长、根结构、种子重量、或可收获产物的量。

[0533] 在一些情况下,适应度的增加是发育、生长、产率、对非生物胁迫源的抗性或对生物胁迫源的抗性增加(例如,增加约1%、2%、5%、10%、20%、30%、40%、50%、60%、70%、80%、90%、100%或大于100%)。非生物胁迫是指植物或植物部分所经受的环境胁迫条件,包括例如干旱胁迫、盐胁迫、热胁迫、冷胁迫和低营养胁迫。生物胁迫是指植物或植物部分所经受的环境胁迫条件,包括例如线虫胁迫、食草昆虫胁迫、真菌病原体胁迫、细菌病原体胁迫或病毒病原体胁迫。胁迫可以是暂时的,例如几个小时,几天,几个月或永久的,例如持续植物的一生。

[0534] 在一些情况下,植物适应度的增加是从植物收获的产物质量增加(例如,增加约1%、2%、5%、10%、20%、30%、40%、50%、60%、70%、80%、90%、100%或大于100%)。例如,植物适应度的增加可以是从植物收获的产物的商业上有利的特征(例如,味道或外观)的改善。在其他情况下,植物适应度的增加是从植物收获的产物的货架期的增加(例如,增加约1%、2%、5%、10%、20%、30%、40%、50%、60%、70%、80%、90%、100%或大于100%)。

[0535] 可替代地,适应度的增加可以是对人类或动物健康有益的性状的改变,例如变应原产生的减少。例如,适应度的增加可以是刺激动物(例如人)中免疫应答的变应原(例如花粉)的产生减少(例如,减少约1%、2%、5%、10%、20%、30%、40%、50%、60%、70%、80%、90%、100%或大于100%)。

[0536] 植物的修饰(例如,适应度的增加)可能来自一个或多个植物部分的修饰。例如,可以通过接触植物的叶、种子、花粉、根、果实、芽、花、细胞,原生质体或组织(例如分生组织)来修饰植物。因此,在另一方面,本文提供了一种增加植物的适应度的方法,该方法包括使植物的花粉与有效量的本文中植物修饰组合物中的任一种接触,其中相对于未经处理的植物,该方法使植物的适应度增加(例如,增加约1%、2%、5%、10%、20%、30%、40%、50%、60%、70%、80%、90%、100%或大于100%)。

[0537] 在又另一方面,本文提供了一种增加植物的适应度的方法,该方法包括使植物的种子与有效量的本文披露的基因修饰系统中的任一种接触,其中相对于未经处理的植物,该方法使植物的适应度增加(例如,增加约1%、2%、5%、10%、20%、30%、40%、50%、60%、70%、80%、90%、100%或大于100%)。

[0538] 在另一方面,本文提供了一种包括使植物的原生质体与有效量的本文所述的基因修饰系统中的任一种接触的方法,其中相对于未经处理的植物,该方法使植物的适应度增加(例如,增加约1%、2%、5%、10%、20%、30%、40%、50%、60%、70%、80%、90%、100%,或大于100%)。

[0539] 在另外的方面,本文提供了一种增加植物的适应度的方法,该方法包括使植物的植物细胞与有效量的本文所述的基因修饰系统中的任一种接触,其中相对于未经处理的植物,该方法使植物的适应度增加(例如,增加约1%、2%、5%、10%、20%、30%、40%、50%、60%、70%、80%、90%、100%,或大于100%)。

[0540] 在另一方面,本文提供了一种增加植物的适应度的方法,该方法包括使植物的分生组织与有效量的本文的植物修饰组合物中的任一种接触,其中相对于未经处理的植物,

该方法使植物的适应度增加(例如,增加约1%、2%、5%、10%、20%、30%、40%、50%、60%、70%、80%、90%、100%,或大于100%)。

[0541] 在另一方面,本文提供了一种增加植物的适应度的方法,该方法包括使植物的胚与有效量的本文的植物修饰组合物中的任一种接触,其中相对于未经处理的植物,该方法使植物的适应度增加(例如,增加约1%、2%、5%、10%、20%、30%、40%、50%、60%、70%、80%、90%、100%,或大于100%)。

[0542] 植物施用方法

[0543] 本文所述的植物可以以允许将组合物递送或施用于植物的任何合适方式暴露于本文所述的任何基因修饰系统组合物。基因修饰系统可以单独递送或与其他活性(例如,肥料剂)或非活性物质组合递送,并且可以通过例如喷雾、注射(例如显微注射)、通过植物、倾倒、浸渍,以浓缩液体、凝胶、溶液、悬浮液、喷雾、粉剂、丸剂、块剂、砖剂等(配制成递送有效浓度的植物修饰组合物)的形式来应用。应用本文所述的组合物的量和位置通常取决于植物的习性、植物可被植物修饰组合物靶向的生命周期阶段、将施用的位置、以及植物修饰组合物的物理和功能特征。

[0544] 在一些情况下,通过例如背包喷雾、空中喷雾、作物喷雾/尘剂等将组合物直接喷雾到植物(例如作物)上。在将基因修饰系统递送至植物的情况下,接受基因修饰系统的植物可以处于植物生长的任何阶段。例如,配制的植物修饰组合物可以在植物生长的早期阶段以种子包衣或根处理剂的形式或在作物周期的后期阶段以总植物处理剂的形式来应用。在一些情况下,植物修饰组合物可以作为局部剂应用于植物。

[0545] 此外,可以将基因修饰系统(例如,在植物生长的土壤中,或在用于浇灌植物的水中)作为通过植物的组织而吸收和分布的内吸剂(systemic agent)应用。在一些情况下,植物或食物生物可以经遗传转化以表达基因修饰系统。

[0546] 延迟释放或持续释放也可以通过以下方式完成:向基因修饰系统或具有一种或多种植物修饰组合物的组合物包覆可溶解或生物可侵蚀的包衣层(诸如明胶),该包衣层在使用环境中溶解或侵蚀,从而然后使植物修饰系统位置可用,或者通过将药剂分散在可溶解或可侵蚀的基质中。这样的持续释放和/或分配方式装置可有利地用于始终维持本文所述的一种或多种植物修饰组合物的有效浓度。

[0547] 在一些情况下,将基因修饰系统递送至植物的一部分,例如叶、种子、花粉、根、果实、芽、或花,或其组织、细胞或原生质体。在一些情况下,将基因修饰系统递送至植物的细胞。在一些情况下,将基因修饰系统递送至植物的原生质体。在一些情况下,将基因修饰系统递送至植物的组织。例如,可以将组合物递送至植物的分生组织(例如,顶端分生组织、侧生分生组织或间生分生组织)。在一些情况下,将组合物递送至植物的永久组织(例如,简单组织(例如,薄壁组织、厚角组织或厚壁组织)或复杂的永久组织(例如,木质部或韧皮部))。在一些情况下,将基因修饰系统递送至植物胚。

[0548] 植物

[0549] 可以将多种植物递送至本文所述的基因修饰系统或用其处理。可以根据本发明方法递送基因修饰系统(即,“处理的”)的植物包括整株植物及其部分,包括但不限于芽营养器官/结构(例如,叶、茎和块茎)、根、花和花器官/结构(例如,苞片、萼片、花瓣、雄蕊、心皮、花药和胚珠)、种子(包括胚、胚乳、子叶、和种皮)和果实(成熟的子房)、植物组织(例如,维

管组织、基本组织等)和细胞(例如,保卫细胞、卵细胞等)及其子代。植物部分可以进一步指如以下的植物部分:芽、根、茎、种子、托叶、叶、花瓣、花、胚珠、苞片、枝、叶柄、节间、树皮、短柔毛、分蘖、根茎、叶状体(frond)、叶片、花粉、雄蕊等。

[0550] 可以在本文披露的方法中处理的植物的类别包括高等植物和低等植物类别,包括被子植物(单子叶植物和双子叶植物)、裸子植物、蕨类、木贼类植物、裸蕨植物、石松类植物、苔藓植物、和藻类(例如,多细胞藻类或单细胞藻类)。可以根据本发明方法处理的植物进一步包括任何维管植物,例如单子叶植物或双子叶植物或裸子植物,包括但不限于苜蓿、苹果、拟南芥属、香蕉、大麦、卡诺拉油菜、蓖麻籽、菊花、三叶草、可可、咖啡、棉花、棉籽、玉米、海甘蓝、蔓越莓、黄瓜、石斛兰、薯蓣、桉树、羊茅草、亚麻、唐菖蒲、百合科、亚麻籽、粟、甜瓜、芥菜、燕麦、油棕、油菜、番木瓜、花生、菠萝、观赏植物、菜豆、马铃薯、油菜籽、水稻、黑麦、黑麦草、红花、芝麻、高粱、大豆、甜菜、甘蔗、向日葵、草莓、烟草、番茄、草皮草、小麦和蔬菜作物(如莴苣、芹菜、西兰花、花椰菜、葫芦);水果树和坚果树,如苹果、梨、桃、橙子、葡萄柚、柠檬、酸橙、扁桃、山核桃、胡桃、榛子;藤本植物,如葡萄(例如,葡萄园)、猕猴桃、蛇麻子(hop);水果灌木和悬钩子,如覆盆子、黑莓、醋栗;林木,如水曲柳、松树、冷杉、枫树、橡树、栗树、杨树(popular);与苜蓿、卡诺拉油菜、蓖麻籽、玉米、棉花、海甘蓝、亚麻、亚麻籽、芥菜、油棕、油菜、花生、马铃薯、水稻、红花、芝麻、大豆、甜菜、向日葵、烟草、番茄、和小麦。可以根据本发明的方法处理的植物包括任何作物植物,例如,草料作物、油籽作物、谷物作物、水果作物、蔬菜作物、纤维作物、香料作物、坚果作物、草皮作物、糖作物、饮料作物、和森林作物。在某些情况下,在该方法中处理的作物植物是大豆植物。在其他某些情况下,作物植物是小麦。在某些情况下,作物植物是玉米。在某些情况下,作物植物是棉花。在某些情况下,作物植物是苜蓿。在某些情况下,作物植物是甜菜。在某些情况下,作物植物是水稻。在某些情况下,作物植物是马铃薯。在某些情况下,作物植物是番茄。

[0551] 在某些情况下,植物是作物。这样的作物植物的实例包括但不限于单子叶植物和双子叶植物,包括但不限于饲养料或草料豆类、观赏植物、食物作物、树木、或灌木,选自枫属物种(Acer spp.)、葱属物种(Allium spp.)、苋属物种(Amaranthus spp.)、凤梨(Ananas comosus)、旱芹(Apium graveolens)、花生属物种(Arachis spp)、石刁柏(Asparagus officinalis)、甜菜(Beta vulgaris)、芸苔属物种(Brassica spp.)(例如,欧洲油菜(Brassica napus)、芜菁(Brassica rapa ssp.)(卡诺拉油菜、油菜、白菜型油菜(turnip rape))、野茶树(Camellia sinensis)、美人蕉(Canna indica)、大麻(Cannabis saliva)、辣椒属物种(Capsicum spp.)、栗属物种(Castanea spp.)、栽培菊苣(Cichorium endivia)、西瓜(Citrullus lanatus)、柑橘属物种(Citrus spp.)、椰子属物种(Cocos spp.)、咖啡属物种(Coffea spp.)、芫荽(Coriandrum sativum)、榛属物种(Corylus spp.)、山楂属物种(Crataegus spp.)、南瓜属物种(Cucurbita spp.)、黄瓜属物种(Cucumis spp.)、胡萝卜(Daucus carota)、水青冈属物种(Fagus spp.)、无花果(Ficus carica)、草莓属物种(Fragaria spp.)、银杏(Ginkgo biloba)、大豆属物种(Glycine spp.)(例如,大豆(Glycine max)、黄豆(Soja hispida)或大豆(Soja max))、陆地棉(Gossypium hirsutum)、向日葵属物种(Helianthus spp.)(例如,向日葵)、木槿属物种(Hibiscus spp.)、大麦属物种(Hordeum spp.)(例如,大麦(Hordeum vulgare))、甘薯(Ipomoea batatas)、胡桃属物种(Juglans spp.)、莴苣(Lactuca sativa)、亚麻(Linum

usitatissimum)、荔枝(Litchi chinensis)、莲属物种(Lotus spp.)、棱角丝瓜(Luffa acutangula)、羽扇豆属物种(Lupinus spp.)、蕃茄属物种(Lycopersicon spp.)(例如,番茄(Lycopersicon esculenturn))、圣女果(Lycopersicon lycopersicum)、梨形番茄(Lycopersicon pyriforme)、苹果属物种(Malus spp.)、紫花苜蓿(Medicago sativa)、薄荷属物种(Mentha spp.)、芒(Miscanthus sinensis)、黑桑(Morus nigra)、芭蕉属物种(Musa spp.)、烟草属物种(Nicotiana spp.)、木犀榄属物种(Olea spp.)、稻属物种(Oryza spp.)(例如,稻(Oryza sativa))、宽叶野生稻(Oryza latifolia)、黍稷(Panicum miliaceum)、柳枝稷(Panicum virgatum)、西番莲(Passiflora edulis)、欧芹(Petroselinumcrispum)、菜豆属物种(Phaseolus spp.)、松属物种(Pinus spp.)、阿月浑子(Pistacia vera)、豌豆属物种(Pisum spp.)、早熟禾属物种(Poa spp.)、杨属物种(Populus spp.)、李属物种(Prunus spp.)、西洋梨(Pyrus communis)、栎属物种(Quercus spp.)、萝卜(Raphanus sativus)、波叶大黄(Rheum rhabarbarum)、茶藨子属物种(Ribes spp.)、蓖麻(Ricinus communis)、悬钩子属物种(Rubus spp.)、甘蔗属物种(Saccharum spp.)、柳属物种(Salix sp.)、接骨木属物种(Sambucus spp.)、黑麦(Secale cereale)、胡麻属物种(Sesamum spp.)、白芥属物种(Sinapis spp.)、茄属物种(Solanum spp.)(例如,马铃薯(Solanum tuberosum)、红茄(Solanumintegrifolium)或番茄(Solanum lycopersicum))、双色高粱(Sorghum bicolor)、石茅(Sorghum halepense)、菠菜属物种(Spinacia spp.)、罗晃子(Tamarindus indica)、可可树(Theobroma cacao)、三叶草属物种(Trifolium spp.)、小黑麦(Triticosecale rimpaui)、小麦属物种(Triticum spp.)(例如,普通小麦(Triticum aestivum)、硬粒小麦(Triticum durum)、圆锥小麦(Triticum turgidum)、Triticum hybernum、马卡小麦(Triticum macha)、Triticum sativum或Triticum vulgare)、越橘属物种(Vaccinium spp.)、蚕豆属物种(Vicia spp.)、豇豆属物种(Vigna spp.)、香堇菜(Viola odorata)、葡萄属物种(Vitis spp.)、和玉米(Zea mays)。在某些实施例中,作物植物是水稻、油菜、卡诺拉油菜、大豆、玉米(玉蜀黍)、棉花、甘蔗、苜蓿、高粱、或小麦。

[0552] 用于本发明的植物或植物部分包括任何植物发育阶段的植物。在某些情况下,可以在萌发、幼苗生长、营养生长、和繁殖生长的阶段进行递送。在某些情况下,向植物的递送在营养生长和繁殖生长阶段期间进行。在一些情况下,将组合物递送至植物的花粉。在一些情况下,将组合物递送至植物的种子。在一些情况下,将组合物递送至植物的原生质体。在一些情况下,将组合物递送至植物的组织。例如,可以将组合物递送至植物的分生组织(例如,顶端分生组织、侧生分生组织或间生分生组织)。在一些情况下,将组合物递送至植物的永久组织(例如,简单组织(例如,薄壁组织、厚角组织或厚壁组织)或复杂的永久组织(例如,木质部或韧皮部))。在一些情况下,将组合物递送至植物胚。在一些情况下,将组合物递送至植物细胞。营养生长和繁殖生长阶段在本文中也称为“成株”或“成熟”植物。

[0553] 在将基因修饰系统递送至植物部分的情况下,可以通过植物修饰剂对植物部分进行修饰。可替代地,基因修饰系统可以被分布到植物的其他部分(例如,通过植物的循环系统),其他部分随后被植物修饰剂修饰。

[0554] 施用和递送

[0555] 本文所述的组合物和系统可以在体外或体内使用。在一些实施例中,例如在体外

或体内将系统或系统的组分递送至细胞(例如,哺乳动物细胞,例如人细胞)。在一些实施例中,细胞是真核细胞,例如多细胞生物的细胞,例如植物或动物,例如哺乳动物(例如人、猪、牛)、鸟(例如家禽,例如鸡、火鸡、或鸭)或鱼。在一些实施例中,细胞是非人动物细胞(例如,实验动物、家畜动物或伴侣动物)。在一些实施例中,细胞是干细胞(例如,造血干细胞)、成纤维细胞或T细胞。在一些实施例中,细胞是免疫细胞,例如,T细胞(例如,Treg、CD4、CD8、γδ或记忆T细胞)、B细胞(例如,记忆B细胞或浆细胞)或NK细胞。在一些实施例中,细胞是非分裂细胞,例如非分裂成纤维细胞或非分裂T细胞。在一些实施例中,细胞是HSC,并且p53没有被上调或被上调少于10%、5%、2%或1%、例如,如根据PCT/US2019/048607的实例30中所述的方法测定。技术人员将理解,能以多肽、核酸(例如,DNA、RNA)及其组合的形式递送基因修饰系统的组分。

[0556] 在一个实施例中,系统和/或系统的组分以核酸的形式递送。例如,基因修饰多肽能以编码该多肽的DNA或RNA的形式递送,并且模板RNA能以RNA或其有待转录成RNA的互补DNA的形式递送。在一些实施例中,系统或系统的组分在1、2、3、4或更多个不同的核酸分子上递送。在一些实施例中,系统或系统的组分作为DNA和RNA的组合递送。在一些实施例中,系统或系统的组分作为DNA和蛋白质的组合递送。在一些实施例中,系统或系统的组分作为RNA和蛋白质的组合递送。在一些实施例中,基因修饰多肽作为蛋白质递送。

[0557] 在一些实施例中,使用载体将系统或系统的组分递送至细胞,例如哺乳动物细胞或人细胞。载体可以是例如质粒或病毒。在一些实施例中,递送是体内、体外、离体或原位的。在一些实施例中,病毒是腺相关病毒(AAV)、慢病毒或腺病毒。在一些实施例中,系统或系统的组分与病毒样颗粒或病毒体一起被递送至细胞。在一些实施例中,递送使用一种以上的病毒、病毒样颗粒或病毒体。

[0558] 在一个实施例中,本文所述的组合物和系统可以配制在脂质体或其他类似的囊泡中。脂质体是球形囊泡结构,这些球形囊泡结构由围绕内部水性隔室的单层或多层的脂质双层和相对不可渗透的外部亲脂性磷脂双层构成。脂质体可以是阴离子的、中性的或阳离子的。脂质体具有生物相容性,无毒,可以递送亲水性和亲脂性药物分子,保护其货物免受血浆酶的降解,并将其负载运输穿过生物膜和血脑屏障(BBB)(有关综述,参见,例如,Spuch和Navarro,Journal of Drug Delivery[药物递送杂志],第2011卷,文章ID 469679,第12页,2011.doi:10.1155/2011/469679)。

[0559] 囊泡可以由若干种不同类型的脂质制成;然而,磷脂最常用于生成脂质体作为药物载剂。用于制备多层囊泡脂质的方法是本领域已知的(参见例如美国专利号6,693,086,其关于多层囊泡脂质制备的传授内容通过援引并入本文)。尽管当脂质膜与水溶液混合时,囊泡形成可以是自发的,但也可以通过经由使用均质器、超声波仪或挤压设备以振荡的形式施加力来加快囊泡形成(关于综述,参见例如,Spuch和Navarro,Journal of Drug Delivery[药物递送杂志],第2011卷,文章ID 469679,第12页,2011.doi:10.1155/2011/469679)。可以通过挤出通过具有减小尺寸的过滤器来制备挤出的脂质,如Templeton等人,Nature Biotech[自然生物技术],15:647-652,1997中所述,该文献关于挤出脂质制备的传授内容通过援引并入本文。

[0560] 多种纳米颗粒可用于递送,例如脂质体、脂质纳米颗粒、阳离子脂质纳米颗粒、可电离脂质纳米颗粒、聚合物纳米颗粒、金纳米颗粒、树枝状大分子、环糊精纳米颗粒、胶束或

上述的组合。

[0561] 脂质纳米颗粒是为本文所述的药物组合物提供生物相容性且可生物降解的递送系统的载剂的实例。纳米结构化的脂质载剂(NLC)是经修饰的固体脂质纳米颗粒(SLN),这些经修饰的固体脂质纳米颗粒保留了SLN的特征、改善了药物稳定性和负载能力、并且防止了药物泄漏。聚合物纳米颗粒(PNP)是药物递送的重要组成部分。这些纳米颗粒可以有效地将药物递送引导至特定靶标并且改善药物稳定性和受控的药物释放。也可以使用脂质聚合物纳米颗粒(PLN),即一种组合了脂质体和聚合物的载剂。这些纳米颗粒具有PNP和脂质体的互补优势。PLN由核-壳结构构成;聚合物核提供了稳定的结构,并且磷脂壳提供了良好的生物相容性。这样,这两种组分增加了药物包封效率、促进了表面修饰、并且防止了水溶性药物的泄漏。对于综述,参见例如,Li等人2017,Nanomaterials[纳米材料]7,122;doi:10.3390/nano7060122。

[0562] 外泌体也可用作本文所述的组合物和系统的药物递送媒介物。对于综述,参见Ha等人2016年7月.Acta Pharmaceutica Sinica B[药学学报]第6卷第4期,第287-296页;https://doi.org/10.1016/j.apsb.2016.02.001。

[0563] 融合体与靶细胞相互作用并融合,并因此可用作多种分子的递送媒介物。它们通常由封闭管腔或腔的两亲性脂质双层和与两亲性脂质双层相互作用的融合剂组成。融合剂组分已被证明是可工程化的,以便为融合和载荷递送赋予靶细胞特异性,从而允许创建具有可编程细胞特异性的递送媒介物(参见例如专利申请WO 2020014209,其涉及融合体设计、制备和使用的教导通过援引并入本文)。

[0564] 在一些实施例中,基因修饰系统的一种或多种蛋白质组分可以与模板核酸(例如,模板RNA)预先关联。例如,在一些实施例中,基因修饰多肽可以首先与模板核酸(例如,模板RNA)组合以形成核糖核蛋白(RNP)复合物。在一些实施例中,可通过例如转染、核转染、病毒、囊泡、LNP、外泌体、融合体将RNP递送至细胞。

[0565] 可以将基因修饰系统引入细胞、组织和多细胞生物中。在一些实施例中,系统或系统的组分经由机械手段或物理手段递送至细胞。

[0566] 以下文献中描述了蛋白治疗剂的配制品:Meyer(编辑),Therapeutic Protein Drug Products:Practical Approaches to formulation in the Laboratory,Manufacturing,and the Clinic[治疗性蛋白药物产品:实验室、制造和临床中配制品的实践方法],Woodhead Publishing Series[伍德海德出版系列](2012)。

[0567] 组织特异性活性/施用

[0568] 在一些实施例中,本文所述的系统可以利用一个或多个特征(例如,启动子或微小RNA结合位点)来限制脱靶细胞或组织中的活性。

[0569] 在一些实施例中,本文所述的核酸(例如,模板RNA或编码模板RNA的DNA)包含启动子序列,例如组织特异性启动子序列。在一些实施例中,组织特异性启动子用于增加基因修饰系统的靶细胞特异性。例如,可以基于启动子在靶细胞类型中有活性但在非靶细胞类型中无活性(或在较低水平上有活性)来选择启动子。因此,即使启动子整合到非靶细胞的基因组中,它也不会驱动整合基因的表达(或仅驱动低水平表达)。如本文所述,在模板RNA中具有组织特异性启动子序列的系统也可与微小RNA结合位点(例如在模板RNA或编码基因修饰蛋白的核酸中,例如,如本文所述)组合使用。在模板RNA中具有组织特异性启动子序列的

系统也可与由组织特异性启动子驱动的编码基因修饰多肽的DNA组合使用,例如,以在靶细胞中获得比非靶细胞中更高水平的基因修饰蛋白。在一些实施例中,例如,对于肝脏适应症,组织特异性启动子选自WO 2020014209(通过援引并入本文)的表3。

[0570] 在一些实施例中,本文所述的核酸(例如,模板RNA或编码模板RNA的DNA)包含微小RNA结合位点。在一些实施例中,微小RNA结合位点用于增加基因修饰系统的靶细胞特异性。例如,可以基于在非靶细胞类型中存在但在靶细胞类型中不存在(或相对于非靶细胞而言以降低的水平存在)的miRNA的识别来选择微小RNA结合位点。因此,当模板RNA存在于非靶细胞中时,它将与miRNA结合,而当模板RNA存在于靶细胞中时,它将不会与miRNA结合(或结合,但相对于非靶细胞而言以降低的水平结合)。尽管不希望受到理论的束缚,但miRNA与模板RNA的结合可以干扰其活性,例如,可以干扰异源对象序列插入基因组。因此,该系统将比其编辑非靶细胞的基因组更有效地编辑靶细胞的基因组,例如,异源对象序列将比插入非靶细胞的基因组更有效地插入靶细胞的基因组,或者插入或缺失在靶细胞中比在非靶细胞中更有效地产生。在模板RNA(或编码它的DNA)中具有微小RNA结合位点的系统也可以与编码基因修饰多肽的核酸组合使用,其中基因修饰多肽的表达受第二微小RNA结合位点的调节,例如如本文所述。在一些实施例中,例如,对于肝适应症,miRNA选自WO 2020014209(通过援引并入本文)的表4。

[0571] 在一些实施例中,模板RNA包含微小RNA序列、siRNA序列、指导RNA序列、或piwiRNA序列。

[0572] 启动子

[0573] 在一些实施例中,一种或多种启动子或增强子元件可操作地连接至编码基因修饰蛋白的核酸或模板核酸,例如,其控制异源对象序列的表达。在某些实施例中,该一个或多个启动子或增强子元件包含细胞类型或组织特异性元件。在一些实施例中,启动子或增强子是相同的或源自天然地控制异源对象序列表达的启动子或增强子。例如,鸟氨酸转氨甲酰酶启动子和增强子可用于在本发明提供的系统或方法中控制鸟氨酸转氨甲酰酶基因的表达以便纠正鸟氨酸转氨甲酰酶缺陷。在一些实施例中,启动子是表16或17中的启动子或其功能片段或变体。

[0574] 例如,可在统一资源定位器(例如,www.invivogen.com/tissue-specific-promoters)中找到可商购的示例性组织特异性启动子。在一些实施例中,启动子是天然启动子或最小启动子,例如由来自给定基因的5'区域的单个片段组成。在一些实施例中,天然启动子包括核心启动子及其天然5'UTR。在一些实施例中,5'UTR包含内含子。在其他实施例中,这些包括复合型启动子,这些复合型启动子组合了起点不同的启动子元件,或由远端增强子与起点相同的最小启动子组装而产生。

[0575] 示例性细胞或组织特异性启动子在下表中提供,并且编码它们的示例性核酸序列是本领域已知的并且可以使用多种资源容易地获得,例如NCBI数据库,包括RefSeq,以及真核启动子数据库(//epd.epfl.ch//index.php)。

[0576] 表16.示例性细胞或组织特异性启动子

[0577]

启动子	靶细胞
B29 启动子	B 细胞
CD14 启动子	单核细胞
CD43 启动子	白细胞和血小板
CD45 启动子	造血细胞
CD68 启动子	巨噬细胞
结蛋白启动子	肌肉细胞

[0578]

启动子	靶细胞
弹性蛋白酶-1 启动子	胰腺腺泡细胞
内皮糖蛋白启动子	内皮细胞
纤连蛋白启动子	分化细胞、愈合组织
Flt-1 启动子	内皮细胞
GFAP 启动子	星形胶质细胞
GPIIB 启动子	巨核细胞
ICAM-2 启动子	内皮细胞
INF-β 启动子	造血细胞
Mb 启动子	肌肉细胞
Nphs1 启动子	足细胞
OG-2 启动子	成骨细胞、成牙本质细胞
SP-B 启动子	肺
Syn1 启动子	神经元
WASP 启动子	造血细胞
SV40/bAlb 启动子	肝
SV40/bAlb 启动子	肝
SV40/Cd3 启动子	白细胞和血小板
SV40/CD45 启动子	造血细胞
NSE/RU5'启动子	成熟神经元

[0579] 表17.另外的示例性细胞或组织特异性启动子

[0580]

启动子	基因描述	基因特异性
APOA2	载脂蛋白 A-II	肝细胞（来自肝细胞祖细胞）
SERPINA1（hAAT）	丝氨酸蛋白酶肽酶抑制剂，进化分枝 A（α-1 抗蛋白酶、抗胰蛋白酶），成员 1（又称为 α1 抗胰蛋白酶）	肝细胞（来自定形内胚层阶段）
CYP3A	细胞色素 P450，家族 3，亚家族 A，多肽	成熟肝细胞
MIR122	微小 RNA 122	肝细胞（来自早期阶段胚胎肝脏细胞）和内胚层

[0581]

启动子	基因描述	基因特异性
胰腺特异性启动子		
启动子	**基因描述**	**基因特异性**
INS	胰岛素	胰腺β细胞（来自定形内胚层阶段）
IRS2	胰岛素受体底物2	胰腺β细胞
Pdx1	胰腺和十二指肠同源异型盒1	胰腺（来自定形内胚层阶段）
Alx3	无芒样同源异型盒3	胰腺β细胞（来自定形内胚层阶段）
Ppy	胰多肽	PP胰腺细胞（γ细胞）
心脏特异性启动子		
启动子	**基因描述**	**基因特异性**
Myh6（aMHC）	肌球蛋白，重链6，心肌，α	心肌细胞的晚期分化标志物（心房特异性）
MYL2（MLC-2v）	肌球蛋白，轻链2，调节性，心脏，慢	心肌细胞的晚期分化标志物（心室特异性）
ITNNl3（cTnl）	肌钙蛋白I型3（心脏）	心肌细胞（来自不成熟状态）
ITNNl3（cTnl）	肌钙蛋白I型3（心脏）	心肌细胞（来自不成熟状态）
NPPA（ANF）	利钠肽前体A（也称为心房利钠因子）	成人细胞的心房特异性
Slc8a1（Ncx1）	溶质载剂家族8（钠/钙交换剂），成员1	来自早期发育阶段的心肌细胞
CNS特异性启动子		
SYN1（hSyn）	突触蛋白I	神经元
GFAP	胶质原纤维酸性蛋白	星形胶质细胞
INA	内联蛋白神经元中间丝蛋白，α（a-内联蛋白）	神经祖细胞
NES	巢蛋白	神经祖细胞和外胚层
MOBP	髓鞘相关少突胶质细胞碱性蛋白	少突胶质细胞
MBP	髓鞘碱性蛋白	少突胶质细胞
TH	酪氨酸羟化酶	多巴胺能神经元
FOXA2（HNF3β）	叉头盒A2	多巴胺能神经元（还用作内胚

启动子	基因描述	基因特异性
		层的标志物）
皮肤特异性启动子		
启动子	**基因描述**	**基因特异性**
FLG	丝聚合蛋白	来自颗粒层的角质形成细胞
K14	角蛋白 14	来自颗粒层和基底层的角质形成细胞
TGM3	转谷氨酰胺酶 3	来自颗粒层的角质形成细胞
免疫细胞特异性启动子		
启动子	**基因描述**	**基因特异性**
ITGAM（CD11B）	整合素，αM（补体组分 3 受体 3 亚基）	单核细胞、巨噬细胞、粒细胞、自然杀伤细胞
泌尿生殖细胞特异性启动子		
启动子	**基因描述**	**基因特异性**
Pbsn	Probasin	前列腺上皮细胞
Upk2	尿溶蛋白 2	膀胱
Sbp	精胺结合蛋白	前列腺癌
Fer1l4	Fer-1-样 4	膀胱
内皮细胞特异性启动子		
启动子	**基因描述**	**基因特异性**
ENG	内皮糖蛋白	内皮细胞
多能和胚胎细胞特异性启动子		
启动子	**基因描述**	**基因特异性**
Oct4（POU5F1）	POU 5 类同源异型盒 1	多能细胞（生殖细胞、ES 细胞、iPS 细胞）
NANOG	Nanog 同源异型盒	多能细胞（ES 细胞、iPS 细胞）
合成 Oct4	基于 Oct-4 核心增强子元件的合成启动子	多能细胞（ES 细胞、iPS 细胞）
T brachyury	Brachyury	中胚层
NES	巢蛋白	神经祖细胞和外胚层
SOX17	SRY（性别决定区 Y）-盒 17	内胚层
FOXA2（HNFJ β）	叉头盒 A2	内胚层（还用作多巴胺能神经元的标志物）
MIR122	微小 RNA 122	内胚层和肝细胞（来自早期阶段胚胎肝细胞）

[0582]

[0583] 取决于所利用的宿主/载体系统，可以在表达载体中使用许多合适的转录和翻译控制元件中的任一种，包括组成型和诱导型启动子、转录增强子元件、转录终止子等（参见

例如,Bitter等人(1987)Methods in Enzymology[酶学方法],153:516-544;其通过援引以其全文并入本文)。

[0584] 在一些实施例中,编码基因修饰蛋白或模板核酸的核酸与控制元件(例如,转录控制元件,例如启动子)可操作地连接。在一些实施例中,转录控制元件可以在以下中起作用:真核细胞例如哺乳动物细胞;或原核细胞(例如细菌或古细菌细胞)。在一些实施例中,编码多肽的核苷酸序列与例如允许该编码多肽的核苷酸序列在原核和真核细胞中表达的多个控制元件可操作地连接。

[0585] 出于说明目的,空间上受限的启动子的实例包括但不限于神经元特异性启动子、脂肪细胞特异性启动子、心肌细胞特异性启动子、平滑肌特异性启动子、光感受器特异性启动子等。神经元特异性空间上受限的启动子包括但不限于神经元特异性烯醇化酶(NSE)启动子(参见例如EMBL HSENO2、X51956);芳香族氨基酸脱羧酶(AADC)启动子、神经丝启动子(参见例如,GenBank HUMNFL,L04147);突触蛋白启动子(参见例如,GenBank HUMSYNIB,M55301);thy-1启动子(参见例如,Chen等人(1987)Cell[细胞]51:7-19;以及Llewellyn,等人(2010)Nat.Med.[自然·医学]16(10):1161-1166);5-羟色胺受体启动子(参见例如,GenBank S62283);酪氨酸羟化酶启动子(TH)(参见例如,Oh等人(2009)Gene Ther[基因疗法]16:437;Sasaoka等人(1992)Mol.Brain Res.[分子脑研究]16:274;Boundy等人(1998)J.Neurosci.[神经科学杂志]18:9989;以及Kaneda等人(1991)Neuron[神经元]6:583-594);GnRH启动子(参见例如,Radovick等人(1991)Proc.Natl.Acad.Sci.USA[美国国家科学院院刊]88:3402-3406);L7启动子(参见例如,Oberdick等人(1990)Science[科学]248:223-226);DNMT启动子(参见例如,Bartge等人(1988)Proc.Natl.Acad.Sci.USA[美国国家科学院院刊]85:3648-3652);脑啡肽启动子(参见例如,Comb等人(1988)EMBO J.[欧洲分子生物学学会杂志]17:3793-3805);髓鞘碱性蛋白(MBP)启动子;Ca2+-钙调蛋白依赖性蛋白激酶II-α(CamKIIα)启动子(参见例如,Mayford等人(1996)Proc.Natl.Acad.Sci.USA[美国国家科学院院刊]93:13250;以及Casanova等人(2001)Genesis[遗传]31:37);CMV增强子/血小板源性生长因子-β启动子(参见例如,Liu等人(2004)Gene Therapy[基因疗法]11:52-60);等。

[0586] 脂肪细胞特异性的空间上受限的启动子包括但不限于:aP2基因启动子/增强子,例如,人aP2基因的-5.4kb至+21bp区域(参见例如,Tozzo等人(1997)Endocrinol[内分泌学].138:1604;Ross等人(1990)Proc.Natl.Acad.Sci.USA[美国国家科学院院刊]87:9590;以及Pavjani等人(2005)Nat.Med.[自然·医学]11:797);葡萄糖转运蛋白-4(GLUT4)启动子(参见例如,Knight等人(2003)Proc.Natl.Acad.Sci.USA[美国国家科学院院刊]100:14725);脂肪酸转位酶(FAT/CD36)启动子(参见例如Kuriki等人(2002)Biol.Pharm.Bull.[生物和医药学报]25:1476;以及Sato等人(2002)J.Biol.Chem.[生物化学杂志]277:15703);硬脂酰基-辅酶A去饱和酶-1(SCD1)启动子(Tabor等人(1999)J.Biol.Chem.[生物化学杂志]274:20603);瘦素启动子(参见例如,Mason等人(1998)Endocrinol.[内分泌学]139:1013;以及Chen等人(1999)Biochem.Biophys.Res.Comm.[生物化学与生物物理研究通讯]262:187);脂连蛋白启动子(参见例如,Kita等人(2005)Biochem.Biophys.Res.Comm.[生物化学与生物物理研究通讯]331:484;以及Chakrabarti(2010)Endocrinol.[内分泌学]151:2408);降脂蛋白启动子(参见例如,Platt等人(1989)Proc.Natl.Acad.Sci.USA[美

国国家科学院院刊]86:7490);抗胰岛素蛋白启动子(参见例如,Seo等人(2003)Molec.Endocrinol.[分子内分泌学]17:1522);等。

[0587] 心肌细胞特异性的空间上受限的启动子包括但不限于源自以下基因的控制序列:肌球蛋白轻链-2、α-肌球蛋白重链、AE3、心肌肌钙蛋白C、心肌肌动蛋白等。Franz等人(1997)Cardiovasc.Res.[心血管研究]35:560-566;Robbins等人(1995)Ann.N.Y.Acad.Sci.[纽约科学院年鉴]752:492-505;Linn等人(1995)Circ.Res.[循环研究]76:584-591;Parmacek等人(1994)Mol.Cell.Biol.[分子细胞生物学]14:1870-1885;Hunter等人(1993)Hypertension[高血压]22:608-617;以及Sartorelli等人(1992)Proc.Natl.Acad.Sci.USA[美国国家科学院院刊]89:4047-4051。

[0588] 平滑肌特异性空间受限启动子包括但不限于SM22α启动子(参见,例如,Akyürek等人(2000)Mol.Med.[分子医学]6:983;和美国专利号7,169,874);平滑肌细胞分化特异性抗原(smoothelin)启动子(参见例如,WO 2001/018048);α-平滑肌肌动蛋白启动子;等。例如,已显示SM22α启动子的0.4kb区域(其中包含两个CArG元件)介导血管平滑肌细胞特异性表达(参见,例如,Kim,等人(1997)Mol.Cell.Biol.[分子细胞生物学]17,2266-2278;Li,等人,(1996)J.Cell Biol.[细胞生物学杂志]132,849-859;和Moessler,等人(1996)Development[发育]122,2415-2425)。

[0589] 光感受器特异性的空间上受限的启动子包括但不限于视紫红质启动子;视紫红质激酶启动子(Young等人(2003)Ophthalmol.Vis.Sci.[眼科和视觉科学]44:4076);β磷酸二酯酶基因启动子(Nicoud等人(2007)J.Gene Med.[基因医学杂志]9:1015);视网膜色素变性基因启动子(Nicoud等人(2007)同上);光感受器间类视黄醇结合蛋白(IRBP)基因增强子(Nicoud等人(2007)同上);IRBP基因启动子(Yokoyama等人(1992)Exp Eye Res.[实验眼科研究杂志]55:225);等。

[0590] 在一些实施例中,基因修饰系统,例如编码基因修饰多肽的DNA、编码模板RNA的DNA或编码异源对象序列的DNA或RNA,被设计成使得一个或多个元件可操作地连接到组织特异性启动子,例如在T细胞中有活性的启动子。在另外的实施例中,T细胞活性启动子在其他细胞类型例如B细胞、NK细胞中是无活性的。在一些实施例中,T细胞活性启动子源自编码T细胞受体组分(例如TRAC、TRBC、TRGC、TRDC)的基因的启动子。在一些实施例中,T细胞活性启动子源自编码T细胞特异性分化蛋白簇(例如CD3,例如CD3D、CD3E、CD3G、CD3Z)的组分的基因的启动子。在一些实施例中,通过比较跨细胞类型的公开可用的基因表达数据并从在T细胞中具有增强的表达的基因中选择启动子来发现基因修饰系统中的T细胞特异性启动子。在一些实施例中,可以根据所期望的表达宽度选择启动子,例如仅在T细胞中具有活性的启动子、仅在NK细胞中具有活性的启动子、在T细胞和NK细胞中都具有活性的启动子。

[0591] 本领域已知的细胞特异性启动子可用于引导基因修饰蛋白的表达,例如,如本文所述。非限制性示例性哺乳动物细胞特异性启动子已被表征并用于以细胞特异性方式表达Cre重组酶的小鼠。某些非限制性示例性哺乳动物细胞特异性启动子列于US 9845481的表1中,该文献通过援引并入本文。

[0592] 在一些实施例中,如本文所述的载体包含表达盒。典型地,表达盒包含本发明的与启动子序列可操作地连接的核酸分子。例如,当启动子能够影响编码序列的表达时,则该启动子与该编码序列可操作地连接(例如,该编码序列在该启动子的转录控制之下)。编码序

列能以有义或反义取向与调节序列可操作地连接。在某些实施例中,启动子是异源启动子。在某些实施例中,表达盒可以包含另外的元件,例如,内含子、增强子、聚腺苷酸化位点、土拨鼠反应元件(WRE)和/或已知影响编码序列表达水平的其他元件。启动子典型地控制编码序列或功能性RNA的表达。在某些实施例中,启动子序列包含近端和更远端上游元件,并可以进一步包含增强子元件。增强子典型地可以刺激启动子的活性,且可以是启动子的固有元件或插入以增强启动子的水平或组织特异性的异源元件。在某些实施例中,启动子整体源自天然基因。在某些实施例中,启动子由源自不同天然存在的启动子的不同元件构成。在某些实施例中,启动子包含合成的核苷酸序列。本领域技术人员将理解,不同的启动子将引导基因在不同的组织或细胞类型中、或在不同的发育阶段、或应答于不同的环境条件或应答于药物或转录辅助因子的存在或不存在的表达。无处不在的、细胞类型特异性的、组织特异性的、发育阶段特异性的和条件性的启动子,例如,药物反应性启动子(例如,四环素反应性启动子)为本领域技术人员所熟知。示例性启动子包括但不限于:磷酸甘油酸激酶(PKG)启动子、CAG(CMV增强子、鸡β肌动蛋白启动子(CBA)和兔β珠蛋白内含子的复合物)、NSE(神经元特异性烯醇化酶)、突触蛋白或NeuN启动子、SV40早期启动子、小鼠乳腺肿瘤病毒LTR启动子;腺病毒主要晚期启动子(Ad MLP)、单纯疱疹病毒(HSV)启动子、巨细胞病毒(CMV)启动子例如CMV立即早期启动子区(CMVIE)、SFFV启动子、劳斯肉瘤病毒(RSV)启动子、合成启动子、杂合启动子等。其他启动子可以是人类来源的或来自其他物种(包括来自小鼠)。常见的启动子包括例如:人巨细胞病毒(CMV)立即早期基因启动子、SV40早期启动子、劳斯肉瘤病毒长末端重复序列、[β]-肌动蛋白、大鼠胰岛素启动子、磷酸甘油酸激酶启动子、人α-1抗胰蛋白酶(hAAT)启动子、甲状腺素转运蛋白启动子、TBG启动子和其他肝特异性启动子、结蛋白启动子和类似的肌肉特异性启动子、EF1-α启动子、具有多组织特异性的杂合启动子、对神经元特异的启动子(如突触蛋白)和甘油醛-3-磷酸脱氢酶启动子,所有这些都是本领域技术人员熟知且容易获得的启动子,可用于获得目的编码序列的高水平表达。另外,源自非病毒基因(如鼠金属硫蛋白基因)的序列也将在本文找到用途。此类启动子序列可商购自例如Stratagene公司(Stratagene)(加利福尼亚州圣地亚哥(San Diego,CA))。另外的示例性启动子序列描述于例如WO 2018213786 A1(其通过援引以其全文并入本文)中。

[0593] 在一些实施例中,载脂蛋白E增强子(ApoE)或其功能片段用于例如促使在肝中的表达。在一些实施例中,使用两个拷贝的ApoE增强子或其功能片段。在一些实施例中,ApoE增强子或其功能片段与启动子(例如,人α-1抗胰蛋白酶(hAAT)启动子)组合使用。

[0594] 在一些实施例中,调节序列赋予组织特异性基因表达能力。在一些情况下,组织特异性调节序列结合以组织特异性方式诱导转录的组织特异性转录因子。各种组织特异性调节序列(例如,启动子、增强子等)是本领域已知的。示例性组织特异性调节序列包括但不限于以下组织特异性启动子:肝特异性甲状腺素结合球蛋白(TBG)启动子、胰岛素启动子、胰高血糖素启动子、生长抑素启动子、胰多肽(PPY)启动子、突触蛋白-1(Syn)启动子、肌酸激酶(MCK)启动子、哺乳动物结蛋白(DES)启动子、α-肌球蛋白重链(a-MHC)启动子或心肌肌钙蛋白T(cTnT)启动子。其他示例性启动子包括:β-肌动蛋白启动子、乙型肝炎病毒核心启动子,Sandig等人,Gene Ther.[基因疗法],3:1002-9(1996);甲胎蛋白(AFP)启动子,Arbuthnot等人,Hum.Gene Ther.[人类基因疗法],7:1503-14(1996)),骨钙素启动子(Stein等人,Mol.Biol.Rep.[分子生物学报告],24:185-96(1997));骨唾液蛋白启动子

(Chen等人,J.Bone Miner.Res.[骨与矿物质研究杂志]11:654-64(1996)),CD2启动子(Hansal等人,J.Immunol.[免疫学杂志],161:1063-8(1998);免疫球蛋白重链启动子;T细胞受体α链启动子,神经元例如神经元特异性烯醇化酶(NSE)启动子(Andersen等人Cell.Mol.Neurobiol.[细胞和分子神经生物学],13:503-15(1993)),神经丝轻链基因启动子(Piccioli等人,Proc.Natl.Acad.Sci.USA[美国国家科学院院刊],88:5611-5(1991)),和神经元特异性vgf基因启动子(Piccioli等人,Neuron[神经元],15:373-84(1995)),以及其他。另外的示例性启动子序列描述于例如美国专利号10300146(其通过援引以其全文并入本文)中。在一些实施例中,组织特异性调节元件(例如,组织特异性启动子)选自已知与在给定组织中高度表达的基因可操作地连接的一种,例如,如通过RNA-seq或蛋白质表达数据、或其组合所测量的。用于通过表达分析组织特异性的方法教授于Fagerberg等人Mol Cell Proteomics[分子与细胞蛋白质组学]13(2):397-406(2014)中,该文献通过援引以其全文并入本文。

[0595] 在一些实施例中,本文所述的载体是多顺反子表达构建体。多顺反子表达构建体包括例如携带第一表达盒和第二表达盒的构建体,该第一表达盒例如包含第一启动子和第一编码核酸序列,该第二表达盒例如包含第二启动子和第二编码核酸序列。在一些情况下,此类多顺反子表达构建体可特别用于递送非翻译基因产物(例如发夹RNA)以及多肽(例如,基因修饰多肽和基因修饰模板)。在一些实施例中,多顺反子表达构建体可以表现出一种或多种所包括的转基因的降低的表达水平,例如,这是因为启动子干扰或存在非常接近的不相容的核酸元件。如果多顺反子表达构建体是病毒载体的一部分,则在一些情况下,自互补核酸序列的存在可能会干扰病毒繁殖或包装所需结构的形成。

[0596] 在一些实施例中,序列编码含发夹的RNA。在一些实施例中,发夹RNA是指导RNA、模板RNA、shRNA、或微小RNA。在一些实施例中,第一启动子是RNA聚合酶I启动子。在一些实施例中,第一启动子是RNA聚合酶II启动子。在一些实施例中,第二启动子是RNA聚合酶III启动子。在一些实施例中,第二启动子是U6或H1启动子。

[0597] 不希望受理论束缚,与含有仅一个顺反子的表达系统相比,多顺反子表达构建体可能无法实现最佳的表达水平。利用包含两个或更多个启动子元件的多顺反子表达构建体实现的表达水平降低的所认为的原因之一是启动子干扰现象(参见例如,Curtin J A,Dane A P,Swanson A,Alexander I E,Ginn S L.Bidirectional promoter interference between two widely used internal heterologous promoters in a late-generation lentiviral construct[晚期慢病毒构建体中两个广泛使用的内部异源启动子之间的双向启动子干扰].Gene Ther.[基因疗法]2008年3月;15(5):384-90;和Martin-Duque P,Jezzard S,Kaftansis L,Vassaux G.Direct comparison of the insulating properties of two genetic elements in an adenoviral vector containing two different expression cassettes[在含有两个不同表达盒的腺病毒载体中对两个遗传元件的绝缘特性的直接比较].Hum Gene Ther.[人类基因疗法]2004年10月;15(10):995-1002;两个参考文献均通过援引并入本文以披露启动子干扰现象)。在一些实施例中,可以通过以下克服启动子干扰的问题,例如通过产生包含仅一个启动子的多顺反子表达构建体,该启动子促进由内部核糖体进入位点分开的多个编码核酸序列的转录;或通过将包含具有转录绝缘子元件的自身启动子的顺反子分开。在一些实施例中,多个顺反子的单启动

子驱动的表达可能导致顺反子的不均匀表达水平。在一些实施例中,不能有效地分离启动子并且分离元件可能与一些基因转移载体(例如,一些逆转录病毒载体)不相容。

[0598] 微小RNA

[0599] 微小RNA(miRNA)和其他小干扰核酸通常经由靶RNA转录物切割/降解或靶信使RNA(mRNA)的翻译抑制来调节基因表达。在一些情况下,miRNA可以天然表达,典型地作为最终的19-25个非翻译RNA产物。miRNA通常通过与靶mRNA的3′非翻译区(UTR)的序列特异性相互作用来表现出它们的活性。这些内源表达的miRNA可形成发夹前体,随后被加工成miRNA双链体,并进一步加工成成熟的单链miRNA分子。这种成熟的miRNA通常会指导多蛋白复合物miRISC,它根据靶mRNA与成熟miRNA的互补性来识别靶mRNA的3′UTR区。有用的转基因产物可以包括例如调节连接的多肽表达的miRNA或miRNA结合位点。miRNA基因的非限制性列表;例如,在如US 10300146,22:25-25:48(其通过援引并入本文)中所列的那些方法的方法中,这些基因及其同源物的产物可用作转基因或用作小干扰核酸(例如,miRNA海绵、反义寡核苷酸)的靶标。在一些实施例中,将一个或多个前述miRNA的一个或多个结合位点掺入转基因(例如,由rAAV载体递送的转基因)中,例如以抑制转基因在携带该转基因的动物的一种或多种组织中的表达。在一些实施例中,可以选择结合位点以便以组织特异性方式控制转基因的表达。例如,可以将肝脏特异性miR-122的结合位点掺入转基因中以抑制该转基因在肝脏中的表达。另外的示例性miRNA序列描述于例如美国专利号10,300,146(其通过援引以其全文并入本文)中。

[0600] miR抑制剂或miRNA抑制剂通常是阻断miRNA表达和/或加工的药剂。此类药剂的实例包括但不限于:抑制miRNA与Drosha复合物相互作用的微小RNA拮抗剂、微小RNA特异性反义、微小RNA海绵和微小RNA寡核苷酸(双链、发夹、短寡核苷酸)。微小RNA抑制剂,例如miRNA海绵,可以在细胞中从转基因表达(例如,如Ebert,M.S.Nature Methods[自然方法],2007年8月12日电子出版中所述;其通过援引以其全文并入本文)。在一些实施例中,微小RNA海绵或其他miR抑制剂与AAV一起使用。微小RNA海绵通常通过互补七聚体种子序列特异性抑制miRNA。在一些实施例中,可以使用单个海绵序列沉默整个miRNA家族。其他用于在细胞中沉默miRNA功能(miRNA靶标的去阻抑)的方法对于本领域普通技术人员来说将是显而易见的。

[0601] 在一些实施例中,本文所述的基因修饰系统、模板RNA或多肽被施用至靶组织(例如第一组织)或在靶组织(例如第一组织)中具有活性(例如,在其中更具活性)。在一些实施例中,基因修饰系统、模板RNA或多肽未施用于非靶组织或在非靶组织中活性较低(例如,在其中不具有活性)。在一些实施例中,本文所述的基因修饰系统、模板RNA或多肽可用于修饰靶组织(例如第一组织)中的DNA(例如,并且不修饰非靶组织中的DNA)。

[0602] 在一些实施例中,基因修饰系统包含(a)本文所述的多肽或编码其的核酸,(b)本文所述的模板核酸(例如,模板RNA),和(c)对靶组织特异性的一个或多个第一组织特异性表达控制序列,其中对靶组织特异性的一个或多个第一组织特异性表达控制序列与(a)、(b)、或(a)和(b)可操作地关联,其中,当与(a)关联时,(a)包含编码多肽的核酸。

[0603] 在一些实施例中,(b)中的核酸包含RNA。

[0604] 在一些实施例中,(b)中的核酸包含DNA。

[0605] 在一些实施例中,(b)中的核酸:(i)是单链区段或包含单链区段,例如,是单链DNA

或包含单链区段和一个或多个双链区段;(ii)具有反向末端重复序列;或(iii)(i)和(ii)两者。

[0606] 在一些实施例中,(b)中的核酸是双链区段或包含双链区段。

[0607] 在一些实施例中,(a)包含编码多肽的核酸。

[0608] 在一些实施例中,(a)中的核酸包含RNA。

[0609] 在一些实施例中,(a)中的核酸包含DNA。

[0610] 在一些实施例中,(a)中的核酸:(i)是单链区段或包含单链区段,例如,是单链DNA或包含单链区段和一个或多个双链区段;(ii)具有反向末端重复序列;或(iii)(i)和(ii)两者。

[0611] 在一些实施例中,(a)中的核酸是双链区段或包含双链区段。

[0612] 在一些实施例中,(a)、(b)、或(a)和(b)中的核酸是线性的。

[0613] 在一些实施例中,(a)、(b)、或(a)和(b)中的核酸是环状的,例如质粒或小环。

[0614] 在一些实施例中,异源对象序列与第一启动子可操作地关联。

[0615] 在一些实施例中,一个或多个第一组织特异性表达控制序列包含组织特异性启动子。

[0616] 在一些实施例中,组织特异性启动子包含与以下可操作地关联的第一启动子:(i)异源对象序列,(ii)编码逆转录病毒RT的核酸,或(iii)(i)和(ii)。

[0617] 在一些实施例中,一个或多个第一组织特异性表达控制序列包含与以下可操作地关联的组织特异性微小RNA识别序列:(i)异源对象序列,(ii)编码逆转录病毒RT结构域的核酸,或(iii)(i)和(ii)。

[0618] 在一些实施例中,系统包含组织特异性启动子,并且该系统进一步包含一种或多种组织特异性微小RNA识别序列,其中:(i)组织特异性启动子与以下可操作地关联:(I)异源对象序列,(II)编码逆转录病毒RT结构域的核酸,或(III)(I)和(II);和/或(ii)一种或多种组织特异性微小RNA识别序列与以下可操作地关联:(I)异源对象序列,(II)编码逆转录病毒RT的核酸,或(III)(I)和(II)。

[0619] 在一些实施例中,其中(a)包含编码多肽的核酸,该核酸包含与编码多肽的核酸可操作地关联的启动子。

[0620] 在一些实施例中,编码多肽的核酸包含对靶组织具有特异性的、与多肽编码序列可操作地关联的一个或多个第二组织特异性表达控制序列。

[0621] 在一些实施例中,一个或多个第二组织特异性表达控制序列包含组织特异性启动子。

[0622] 在一些实施例中,组织特异性启动子是与编码多肽的核酸可操作地关联的启动子。

[0623] 在一些实施例中,一个或多个第二组织特异性表达控制序列包含组织特异性微小RNA识别序列。

[0624] 在一些实施例中,与编码多肽的核酸可操作地关联的启动子是组织特异性启动子,该系统进一步包含一个或多个组织特异性微小RNA识别序列。

[0625] 在一些实施例中,本发明提供的系统的核酸组分序列(例如,编码多肽或包含异源对象序列)的侧翼是修饰蛋白质表达水平的非翻译区(UTR)。各种5'和3' UTR会影响蛋白质

表达。例如,在一些实施例中,编码序列之前可以是修饰RNA稳定性或蛋白质翻译的5′UTR。在一些实施例中,序列之后可以是修饰RNA稳定性或翻译的3′UTR。在一些实施例中,序列之前可以是5′UTR,然后是修饰RNA稳定性或翻译的3′UTR。在一些实施例中,5’和/或3’UTR可选自补体因子3(C3)(CACTCCTCCCCATCCTCTCCCTCTGTCCCTCTGTCCCTCTGACCCTGCACTGTCCCAGCACC)或血清类黏蛋白1(ORM1)(CAGGACACAGCCTTGGATCAGGACAGAGACTTGGGGGCCATCCTGC CCCTCCAACCCGACATGTGTACCTCAGCTTTTTCCCTCACTTGCATCAATAAAGCTTCTGTGTTTGGAACAGCTAA)的5’和3’UTR(Asrani等人RNA Biology[核糖核酸学]2018)。在某些实施例中,5′UTR是来自C3的5′UTR并且3′UTR是来自ORM1的3′UTR。在某些实施例中,用于蛋白质表达(例如基因修饰多肽或异源对象序列的mRNA(或编码RNA的DNA))的5′UTR和3′UTR包含优化的表达序列。在一些实施例中,5’UTR包含GGGAAAUAAGAGAGAAAAGAAGAGUA AGAAGAAAUAUAAGAGCCACC和/或3’UTR包含UGAUAAUAGGCUGGAGCCUCGGUGGCCAUGCUUCUUGCCCCUUGGG CCUCCCCCCCAGCCCCUCCUCCCCUUCCUGCACCCGUACCCCCCGUGGU CUUUGAAUAAAGUCUGA,例如,如Richner等人Cell[细胞]168(6):第1114-1125页(2017)中所述,其中的序列通过援引并入本文。

[0626] 在一些实施例中,可以选择5’和/或3’UTR以增强蛋白质表达。在一些实施例中,可以选择5′和/或3′UTR来修饰蛋白质表达,从而最大程度地减少过度生产抑制。在一些实施例中,UTR在编码序列周围,例如在编码序列之外以及在其他实施例中靠近编码序列。在一些实施例中,UTR中包含另外的调节元件(例如,miRNA结合位点、顺式调节位点)。

[0627] 在一些实施例中,基因修饰系统的开放阅读框,例如,编码基因修饰多肽的mRNA(或编码mRNA的DNA)的ORF或异源对象序列的mRNA(或编码mRNA的DNA)的一个或多个ORF,侧翼有增强其表达的5′和/或3′非翻译区(UTR)。在一些实施例中,系统的mRNA组分(或从DNA组分产生的转录本)的5’UTR包含序列5’-GGGAAAUAAGAGAGAAAAGAAGAGUAAGAAGAAAUAUAAGAGCCACC-3’。在一些实施例中,系统的mRNA组分(或从DNA组分产生的转录本)的3’UTR包含序列5’-UGAUAAUAGGCUGGAGCCUCGGUGGCCAUGCUUCUUGCCCCUUG GGCCUCCCCCCAGCCCCUCCUCCCCUUCCUGCACCCGUACCCCCGUG GUCUUUGAAUAAAGUCUGA-3’。已经由以下证明这种5′UTR和3′UTR的组合可导致可操作地连接的ORF的期望表达:Richner等人Cell[细胞]168(6):第1114-1125页(2017),其教导和序列通过援引并入本文。在一些实施例中,本文所述的系统包含编码转录本的DNA,其中该DNA包含对应的5′UTR和3′UTR序列,其中T取代以上所列的序列中的U。在一些实施例中,用于产生系统的RNA组分的DNA载体进一步包含用于启动体外转录的5′UTR上游的启动子,例如T7、T3或SP6启动子。以上5′UTR以GGG开头,这对于使用T7 RNA聚合酶优化转录是合适的开始。对于调整转录水平和改变转录起始位点核苷酸以适应替代性的5′UTR,Davidson等人.Pac Symp Biocomput[Pac Symp生物计算]433-443(2010)的传授内容描述了满足这两个特征的T7启动子变体及其发现方法。

[0628] 病毒载体及其组分

[0629] 除了本文所述的相关酶或结构域的来源,例如作为本文使用的聚合酶和聚合酶功能(例如DNA-依赖性DNA聚合酶、RNA-依赖性RNA聚合酶、RNA-依赖性DNA聚合酶、DNA-依赖性RNA聚合酶、逆转录酶)的来源,病毒还是本文所述系统的有用的递送媒剂来源。一些酶,例如逆转录酶,可能具有多种活性,例如能够进行RNA-依赖性DNA聚合和DNA-依赖性DNA聚合,例如第一和第二链合成。在一些实施例中,用作基因修饰递送系统或其组分来源的病毒可

选自如Baltimore Bacteriol Rev[细菌综述]35(3):235-241(1971)所述的组。

[0630] 在一些实施例中,病毒选自组I病毒,例如,该病毒是DNA病毒并将dsDNA包装成病毒体。在一些实施例中,组I病毒选自例如腺病毒、疱疹病毒、痘病毒。

[0631] 在一些实施例中,病毒选自组II病毒,例如,该病毒是DNA病毒并将ssDNA包装成病毒体。在一些实施例中,组II病毒选自例如细小病毒。在一些实施例中,细小病毒是依赖性细小病毒,例如腺相关病毒(AAV)。

[0632] 在一些实施例中,病毒选自组III病毒,例如,该病毒是RNA病毒并将dsRNA包装成病毒体。在一些实施例中,组III病毒选自例如呼肠孤病毒。在一些实施例中,包含在此类病毒体中的dsRNA的一条或两条链是编码分子,在转导至宿主细胞后能够直接用作mRNA,例如,在转导至宿主细胞后可以直接翻译成蛋白质而不需要任何干预性核酸复制或聚合步骤。

[0633] 在一些实施例中,病毒选自组IV病毒,例如,该病毒是RNA病毒并将ssRNA(+)包装成病毒体。在一些实施例中,组IV病毒选自例如冠状病毒、小RNA病毒、披膜病毒。在一些实施例中,包含在此类病毒体中的ssRNA(+)是编码分子,在转导至宿主细胞后能够直接用作mRNA,例如,在转导至宿主细胞后可以直接翻译成蛋白质而不需要任何干预性核酸复制或聚合步骤。

[0634] 在一些实施例中,病毒选自组V病毒,例如,该病毒是RNA病毒并将ssRNA(-)包装成病毒体。在一些实施例中,组V病毒选自例如正黏病毒、弹状病毒。在一些实施例中,具有ssRNA(-)基因组的RNA病毒还在病毒体内携带酶,该酶被转导至具有病毒基因组的宿主细胞,例如RNA依赖性RNA聚合酶,能够将ssRNA(-)拷贝到可以由宿主直接翻译的ssRNA(+)。

[0635] 在一些实施例中,病毒选自组VI病毒,例如,该病毒是逆转录病毒并将ssRNA(+)包装成病毒体。在一些实施例中,组VI病毒选自例如逆转录病毒。在一些实施例中,逆转录病毒是慢病毒,例如,HIV-1、HIV-2、SIV、BIV。在一些实施例中,逆转录病毒是泡沫病毒属(spumavirus),例如泡沫病毒(foamy virus),例如HFV、SFV、BFV。在一些实施例中,包含在此类病毒体中的ssRNA(+)是编码分子,在转导至宿主细胞后能够直接用作mRNA,例如,在转导至宿主细胞后可以直接翻译成蛋白质而不需要任何干预性核酸复制或聚合步骤。在一些实施例中,ssRNA(+)首先被逆转录并拷贝以产生dsDNA基因组中间体,mRNA可以由该基因组中间体在宿主细胞中得以转录。在一些实施例中,具有ssRNA(+)基因组的RNA病毒还在病毒体内携带酶,该酶被转导至具有病毒基因组的宿主细胞,例如RNA依赖性DNA聚合酶,能够将ssRNA(+)拷贝到可以转录为mRNA并由宿主翻译的dsDNA。在一些实施例中,来自组VI逆转录病毒的逆转录酶作为基因修饰多肽的逆转录酶结构域掺入。

[0636] 在一些实施例中,病毒选自组VII病毒,例如,该病毒是逆转录病毒并将dsRNA包装成病毒体。在一些实施例中,组VII病毒选自例如嗜肝DNA病毒。在一些实施例中,包含在此类病毒体中的dsRNA的一条或两条链是编码分子,在转导至宿主细胞后能够直接用作mRNA,例如,在转导至宿主细胞后可以直接翻译成蛋白质而不需要任何干预性核酸复制或聚合步骤。在一些实施例中,这样的病毒体中包含的dsRNA的一条或两条链首先被逆转录并拷贝以产生dsDNA基因组中间体,mRNA可以由该基因组中间体在宿主细胞中得以转录。在一些实施例中,具有dsRNA基因组的RNA病毒还在病毒体内携带酶,该酶被转导至具有病毒基因组的宿主细胞,例如RNA依赖性DNA聚合酶,能够将dsRNA拷贝到可以转录为mRNA并由宿主翻译的

dsDNA。在一些实施例中,来自VII组逆转录病毒的逆转录酶作为基因修饰多肽的逆转录酶结构域掺入。

[0637] 在一些实施例中,本发明中用于递送核酸的病毒体还可以携带参与基因修饰过程的酶。例如,逆转录病毒病毒体可以包含与核酸一起被递送到宿主细胞中的逆转录酶结构域。在一些实施例中,RNA模板可以与病毒体内的基因修饰多肽相关联,从而在从病毒颗粒转导核酸后两者共同递送至靶细胞。在一些实施例中,病毒体中的核酸可以包含DNA,例如线性ssDNA、线性dsDNA、环状ssDNA、环状dsDNA、小环DNA、dbDNA、ceDNA。在一些实施例中,病毒体中的核酸可以包含RNA,例如线性ssRNA、线性dsRNA、环状ssRNA、环状dsRNA。在一些实施例中,病毒基因组可以在转导至宿主细胞后环化,例如,线性ssRNA分子可以经历共价连接以形成环状ssRNA,线性dsRNA分子可以经历共价连接以形成环状dsRNA或一个或多个环状ssRNA。在一些实施例中,病毒基因组可以通过在宿主细胞中的滚环复制来复制。在一些实施例中,病毒基因组可以包含单个核酸分子,例如,包含非分段基因组。在一些实施例中,病毒基因组可以包含两个或更多个核酸分子,例如,包含分段基因组。在一些实施例中,病毒体中的核酸可以与一种或蛋白质相关联。在一些实施例中,病毒体中的一种或多种蛋白质可在转导后被递送至宿主细胞。在一些实施例中,可通过向靶核酸添加病毒体包装信号而使天然病毒适于核酸递送,其中宿主细胞用于包装含有包装信号的靶核酸。

[0638] 在一些实施例中,用作递送媒介物的病毒体可以包含共生人类病毒。在一些实施例中,用作递送媒介物的病毒体可以包含指环病毒,其用途描述于WO 2018232017 A1中,该文献通过援引以其全文并入本文。

[0639] AAV施用

[0640] 在一些实施例中,腺相关病毒(AAV)与本文所述的系统、模板核酸和/或多肽联合使用。在一些实施例中,AAV用于递送、施用或包装本文所述的系统、模板核酸和/或多肽。在一些实施例中,AAV是重组AAV(rAAV)。

[0641] 在一些实施例中,系统包含(a)本文所述的多肽或编码其的核酸,(b)本文所述的模板核酸(例如,模板RNA),和(c)对靶组织特异性的一个或多个第一组织特异性表达控制序列,其中对靶组织特异性的一个或多个第一组织特异性表达控制序列与(a)、(b)、或(a)和(b)可操作地关联,其中,当与(a)关联时,(a)包含编码多肽的核酸。

[0642] 在一些实施例中,本文所述的系统还包含第一重组腺相关病毒(rAAV)衣壳蛋白;其中(a)或(b)中的至少一个与第一rAAV衣壳蛋白相关联,其中(a)或(b)中的至少一个侧翼为AAV反向末端重复序列(ITR)。

[0643] 在一些实施例中,(a)和(b)与第一rAAV衣壳蛋白相关联。

[0644] 在一些实施例中,(a)和(b)在单个核酸上。

[0645] 在一些实施例中,该系统进一步包含第二rAAV衣壳蛋白,其中(a)或(b)中的至少一个与第二rAAV衣壳蛋白相关联,并且其中与第二rAAV衣壳蛋白相关联的(a)或(b)中的至少一个和与第一rAAV衣壳蛋白相关联的(a)或(b)中的至少一个不同。

[0646] 在一些实施例中,(a)或(b)中的至少一个与第一或第二rAAV衣壳蛋白相关联分散在第一或第二rAAV衣壳蛋白的内部,该第一或第二rAAV衣壳蛋白是以AAV衣壳颗粒的形式。

[0647] 在一些实施例中,该系统还包含纳米颗粒,其中该纳米颗粒与(a)或(b)中的至少一个相关联。

[0648] 在一些实施例中,(a)和(b)分别与以下相关联:a)第一rAAV衣壳蛋白和第二rAAV衣壳蛋白;b)纳米颗粒和第一rAAV衣壳蛋白;c)第一rAAV衣壳蛋白;d)第一腺病毒衣壳蛋白;e)第一纳米颗粒和第二纳米颗粒;或f)第一纳米颗粒。

[0649] 病毒载体可用于递送本发明提供的系统的全部或部分,例如用于本发明提供的方法中。源自不同病毒的系统已被用于递送多肽或核酸;例如:整合酶缺陷型慢病毒、腺病毒、腺相关病毒(AAV)、单纯疱疹病毒和杆状病毒(在Hodge等人Hum Gene Ther[人类基因疗法]2017;Narayanavari等人Crit Rev Biochem Mol Biol[生物化学和分子生物学评论]2017;Boehme等人Curr Gene Ther[当今基因疗法]2015)中进行了综述。

[0650] 腺病毒是常见的病毒,由于具有明确的生物学特性、遗传稳定性、高转导效率并易于大规模生产,其已被用作基因递送媒剂(例如,参见Lee等人Genes&Diseases[基因与疾病]2017中的综述)。它们具有线性dsDNA基因组,并有多种血清型,在组织和细胞嗜性方面有所不同。为了防止感染性病毒在受体细胞中复制,用于包装的腺病毒基因组被缺失了一些或全部内源病毒蛋白,这些内源病毒蛋白在病毒生产细胞中以反式形式提供。这使得基因组依赖于辅助功能,这意味着它们只能在所谓的辅助功能提供的缺失组分存在的情况下被复制并包装成病毒颗粒。去除所有病毒ORF的辅助依赖性腺病毒系统可与包装高达约37kb的外源DNA兼容(Parks等人J Virol[病毒学杂志]1997)。在一些实施例中,腺病毒载体用于递送对应于基因修饰系统的多肽或模板组分的DNA,或两者都包含在单独或相同的腺病毒载体上。在一些实施例中,腺病毒是不能自包装的辅助依赖性腺病毒(HD-AdV)。在一些实施例中,腺病毒是高容量腺病毒(HC-AdV),其已缺失了全部或大部分内源病毒ORF,同时保留了包装成腺病毒颗粒所需的序列组分。对于这种类型的载体,基因组包装所需的唯一腺病毒序列是非编码序列:两末端的反向末端重复序列(ITR)和5′末端的包装信号(Jager等人Nat Protoc[自然实验手册]2009)。在一些实施例中,腺病毒基因组还包含填充DNA以满足用于最佳生产和稳定性的最小基因组大小(参见,例如,Hausl等人Mol Ther[分子疗法]2010)。在一些实施例中,腺病毒用于将基因修饰系统递送至肝。

[0651] 在一些实施例中,腺病毒用于将基因修饰系统递送至HSC,例如HDAd5/35++。HDAd5/35++是具有经修饰的血清型35纤维(其将载体从肝去靶向)的腺病毒(Wang等人Blood Adv[血液研究进展]2019)。在一些实施例中,将基因修饰系统递送至HSC的腺病毒利用在原始HSC上特异性表达的受体,例如,CD46。

[0652] 腺相关病毒(AAV)属于细小病毒科,更特别地构成依赖性细小病毒属。AAV基因组由线性单链DNA分子构成,该分子包含约4.7千碱基(kb)并且由两个主要的开放阅读框(ORF)(编码非结构Rep(复制)和结构Cap(衣壳)蛋白)组成。cap基因内的第二ORF被鉴定为编码组装激活蛋白(AAP)。AAV编码区两侧的DNA是两个顺式作用反向末端重复(ITR)序列,长度约为145个核苷酸,具有间断的回文序列,这些回文序列可折叠成能量稳定的发夹结构,这些发夹结构用作DNA复制的引物。除了它们在DNA复制中的作用外,ITR序列已被证明与病毒DNA整合到细胞基因组中、从宿主基因组或质粒中拯救以及将病毒核酸包裹到成熟病毒体中有关(Muzyczka,(1992)Curr.Top.Micro.Immunol.[微生物学和免疫学的当前主题]158:97-129).在一些实施例中,一种或多种基因修饰核酸组分的侧翼是源自AAV的ITR,用于病毒包装。参见,例如,WO 2019113310。

[0653] 在一些实施例中,基因修饰系统的一种或多种组分通过至少一种AAV载体携带。在

一些实施例中,针对特定细胞、组织、生物的嗜性选择至少一种AAV载体。在一些实施例中,AAV载体是假型的,例如AAV2/8,其中AAV2描述了构建体的设计,但衣壳蛋白被来自AAV8的蛋白替换。应当理解,任何所述载体可以是假型衍生物,其中用于包装AAV基因组的衣壳蛋白源自不同AAV血清型的衣壳蛋白。不希望受限于载体选择,示例性AAV血清型的列表可在表18中找到。在一些实施例中,用于基因修饰的AAV可针对新型细胞或组织嗜性进行进化,如文献中已证明的(例如,Davidsson等人Proc Natl Acad Sci U S A[美国国家科学院院刊]2019)。

[0654] 在一些实施例中,AAV递送载体是具有两个AAV反向末端重复序列(ITR)和目的核苷酸序列(例如,编码基因修饰多肽或DNA模板,或两者的序列)的载体,所述ITR中的每个具有中断(或非连续)回文序列,即由三个片段构成的序列:第一个区段和最后一个片段在5'→3'读取时是相同的,但在彼此相对放置时会杂交,以及一个不同的区段将相同的区段分开。参见,例如,WO 2012123430。

[0655] 通常,通过引入一个或多个编码rAAV或scAAV基因组、Rep蛋白和Cap蛋白的质粒来产生带有衣壳的AAV病毒体(Grimm等人,1998)。在反式引入这些辅助质粒后,AAV基因组从宿主基因组中被"拯救"(即释放并随后回收),并进一步包裹以产生感染性AAV。在一些实施例中,通过将侧翼为ITR的核酸与辅助功能一起引入包装细胞中,将一种或多种基因修饰核酸包装到AAV颗粒中。

[0656] 在一些实施例中,AAV基因组是所谓的自互补基因组(称为scAAV),使得位于ITR之间的序列包含所期望的核酸序列(例如,编码基因修饰多肽或模板的DNA,或两者)以及所期望的核酸序列的反向互补序列,使得这两种组分可以折叠和自杂交。在一些实施例中,自互补模块由允许DNA自身折叠的间插序列分开,例如,形成茎环。scAAV的优势在于在进入细胞核后准备好进行转录,而不是首先依赖ITR引发和第二链合成来形成dsDNA。在一些实施例中,一种或多种基因修饰组分被设计为scAAV,其中AAV ITR之间的序列包含两个反向互补模块,它们可以自杂交以产生dsDNA。

[0657] 在一些实施例中,递送至细胞的核酸(例如,编码多肽或模板,或两者)是封闭末端的线性双链体DNA(CELiD DNA或ceDNA)。在一些实施例中,ceDNA源自AAV基因组的复制形式(Li等人PLoS One[公共科学图书馆·综合]2013)。在一些实施例中,核酸(例如,编码多肽或模板DNA,或两者)的侧翼为ITR,例如AAV ITR,其中至少一个ITR包含末端解离位点和复制蛋白结合位点(有时称为复制型蛋白结合位点)。在一些实施例中,ITR源自腺相关病毒,例如AAV1、AAV2、AAV3、AAV4、AAV5、AAV6、AAV7、AAV8、AAV9、AAV10、AAV11、AAV12或其组合。在一些实施例中,ITR是对称的。在一些实施例中,ITR是不对称的。在一些实施例中,提供至少一种Rep蛋白以使构建体能够复制。在一些实施例中,至少一种Rep蛋白源自腺相关病毒,例如AAV1、AAV2、AAV3、AAV4、AAV5、AAV6、AAV7、AAV8、AAV9、AAV10、AAV11、AAV12或其组合。在一些实施例中,通过向生产细胞提供(i)侧翼为ITR(例如AAV ITR)的DNA,和(ii)ITR依赖性复制所需的组分,例如AAV蛋白Rep78和Rep52(或编码蛋白的核酸)来产生ceDNA。在一些实施例中,ceDNA不含任何衣壳蛋白,例如,未包装到感染性AAV颗粒中。在一些实施例中,ceDNA被配制成LNP(参见例如WO 2019051289 A1)。

[0658] 在一些实施例中,ceDNA载体由两个自互补序列组成,例如本文定义的不对称或对称或基本对称的ITR,位于所述表达盒的侧翼,其中ceDNA载体不与衣壳蛋白相关联。在一些

实施例中,ceDNA载体包含在AAV基因组中发现的两个自互补序列,其中至少一个ITR包含AAV的可操作的Rep结合元件(RBE)(在本文中有时也称为“RBS”)和末端解离位点(trs)或RBE的功能变体。参见,例如,WO 2019113310。

[0659] 在一些实施例中,AAV基因组包含分别编码四种复制蛋白和三种衣壳蛋白的两个基因。在一些实施例中,基因的侧翼中任何一侧有145-bp的反向末端重复序列(ITR)。在一些实施例中,病毒体包含例如以1:1:10比率产生的多达三种衣壳蛋白(Vp1、Vp2、和/或Vp3)。在一些实施例中,衣壳蛋白产生自相同的开放阅读框和/或差异剪接(Vp1)和替代性的翻译起始位点(分别为Vp2和Vp3)。通常,Vp3是病毒体中最丰富的亚基,并参与细胞表面的受体识别,从而定义了病毒的嗜性。在一些实施例中,Vp1在Vp1的N末端包含例如在病毒感染性方面起作用的磷脂酶结构域。

[0660] 在一些实施例中,病毒载体的包装能力限制了可以包装到载体中的基因修饰系统的大小。例如,AAV的包装能力可以是约4.5kb(例如,约3.0、3.5、4.0、4.5、5.0、5.5、或6.0kb),例如,包括一个或两个反向末端重复序列(ITR),例如,145个碱基ITR。

[0661] 在一些实施例中,重组AAV(rAAV)包含在载体转基因盒侧翼的顺式作用145-bp ITR,例如,提供高达4.5kb用于外源DNA的包装。感染后,在一些情况下,rAAV可以表达本发明的融合蛋白,并且通过以环状头对尾多联体的附加体形式持续存在而不整合到宿主基因组中。rAAV可例如在体外和体内使用。在一些实施例中,AAV介导的基因递送要求基因的编码序列的长度在大小上等于或大于野生型AAV基因组。

[0662] 超过该大小的基因的AAV递送和/或大的生理调节元件的使用可以例如通过将要递送的一种或多种蛋白质分成两个或更多个片段来完成。在一些实施例中,N末端片段与内含肽-N序列融合。在一些实施例中,C末端片段与内含肽-C序列融合。在实施例中,将片段包装到两个或更多个AAV载体中。

[0663] 在一些实施例中,通过将大的转基因表达盒分成两个单独的半部分(5′和3′端,或头和尾)来产生双重AAV载体,例如,其中盒的每一半被包装在单个AAV载体中(其<5kb)。在一些实施例中,然后可以在通过两个双重AAV载体对同一细胞进行的共感染后实现全长转基因表达盒的重新组装。在一些实施例中,共感染之后是以下中的一项或多项:(1)5′和3′基因组之间的同源重组(HR)(双重AAV重叠载体);(2)5′和3′基因组的ITR介导的尾对头连环化(双重AAV反式剪接载体);和/或(3)这两种机制的组合(双重AAV杂合载体)。在一些实施例中,体内使用双重AAV载体导致全长蛋白质的表达。在一些实施例中,双重AAV载体平台的使用代表了用于大小大于约4.0、4.1、4.2、4.3、4.4、4.5、4.6、4.7、4.8、4.9、或5.0kb的转基因的有效且可行的基因转移策略。在一些实施例中,AAV载体还可用于例如在核酸和肽的体外生产中用靶核酸转导细胞。在一些实施例中,AAV载体可用于体内和离体基因疗法程序(参见,例如,West等人,Virology[病毒学]160:38-47(1987);美国专利号4,797,368;WO 93/24641;Kotin,Human Gene Therapy[人类基因疗法]5:793-801(1994);Muzyczka,J.Clin.Invest.[临床研究期刊]94:1351(1994);其各自通过援引以其全文并入本文)。重组AAV载体的构建描述于许多公开物中,包括美国专利号5,173,414;Tratschin等人,Mol.Cell.Biol.[分子细胞生物学]5:3251-3260(1985);Tratschin,等人,Mol.Cell.Biol.[分子细胞生物学]4:2072-2081(1984);Hermonat和Muzyczka,PNAS[美国国家科学院院刊]81:6466-6470(1984);以及Samulski等人,J.Virol.[病毒学杂志]63:03822-3828(1989)

(其通过援引以其全文并入本文)。

[0664] 在一些实施例中,本文所述的基因修饰多肽(例如,具有或不具有一种或多种指导核酸)可以使用AAV、慢病毒、腺病毒或其他质粒或病毒载体类型进行递送,特别是使用来自以下文献的配制品和剂量:例如,美国专利号8,454,972(针对腺病毒的配制品、剂量)、美国专利号8,404,658(针对AAV的配制品、剂量)和美国专利号5,846,946(针对DNA质粒的配制品、剂量)以及来自临床试验和关于涉及慢病毒、AAV和腺病毒的临床试验的公开物。例如,对于AAV,施用途径、配制品和剂量可如美国专利号8,454,972和涉及AAV的临床试验中所述。对于腺病毒,施用途径、配制品和剂量可如美国专利号8,404,658和涉及腺病毒的临床试验中所述。对于质粒递送,施用途径、配制品和剂量可如美国专利号5,846,946和涉及质粒的临床研究中所述。剂量可以基于或外推为平均70kg的个体(例如男性成人),并且可以针对患者、受试者、不同重量和物种的哺乳动物进行调整。施用频率在医学或兽医学从业者(例如医师、兽医师)的范围之内,其取决于常规因素,包括患者或受试者的年龄、性别、一般健康状况、其他状况以及着手解决的特定病症或症状。在一些实施例中,可以将病毒载体注射到目的组织中。对于细胞类型的特异性基因修饰,在一些实施例中,基因修饰多肽和任选的指导核酸的表达可以由细胞类型的特异性启动子驱动。

[0665] 在一些实施例中,例如,由于纯化方法不需要可以激活免疫反应的细胞颗粒的超速离心,AAV允许低毒性。在一些实施例中,AAV允许引起插入诱变的可能性低,原因是例如它基本上不整合到宿主基因组中。

[0666] 在一些实施例中,AAV具有约4.4、4.5、4.6、4.7、或4.75kb的包装限制。在一些实施例中,基因修饰多肽编码序列、启动子和转录终止子可以装配到单个病毒载体中。在一些情况下,SpCas9(4.1kb)可能难以包装成AAV。因此,在一些实施例中,使用长度比其他基因修饰多肽编码序列或碱基编辑器短的基因修饰多肽编码序列。在一些实施例中,基因修饰多肽编码序列小于约4.5kb、4.4kb、4.3kb、4.2kb、4.1kb、4kb、3.9kb、3.8kb、3.7kb、3.6kb、3.5kb、3.4kb、3.3kb、3.2kb、3.1kb、3kb、2.9kb、2.8kb、2.7kb、2.6kb、2.5kb、2kb、或1.5kb。

[0667] AAV可以是AAV1、AAV2、AAV5或其任何组合。在一些实施例中,AAV的类型是根据要靶向的细胞来选择的;例如,可选择AAV血清型1、2、5或杂合衣壳AAV1、AAV2、AAV5或其任何组合用于靶向脑或神经元细胞;或者可以选择AAV4用于靶向心脏组织。在一些实施例中,选择AAV8用于递送至肝脏。关于这些细胞的示例性AAV血清型描述于例如Grimm,D.等人,J.Virol.[病毒学杂志]82:5887-5911(2008)(其通过援引以其全文并入本文)中。在一些实施例中,AAV是指所有血清型、亚型和天然存在的AAV以及重组AAV。AAV可用于指代病毒本身或其衍生物。在一些实施例中,AAV包括AAV1、AAV2、AAV3、AAV3B、AAV4、AAV5、AAV6、AAV6.2、AAV7、AAVrh.64R1、AAVhu.37、AAVrh.8、AAVrh.32.33、AAV8、AAV9、AAV-DJ、AAV2/8、AAVrh10、AAVLK03、AV10、AAV11、AAV 12、rh10、和其杂合体,禽AAV、牛AAV、犬AAV、马AAV、灵长类动物AAV、非灵长类动物AAV、和羊AAV。各种AAV血清型的基因组序列,以及天然末端重复序列(TR)、Rep蛋白和衣壳亚基的序列是本领域已知的。此类序列可在文献或公共数据库如GenBank中找到。表18中列出了另外的示例性AAV血清型。

[0668] 表18.示例性AAV血清型。

[0669]

靶组织	媒介物	参考文献
肝	AAV（AAV8[1]、AAVrh.8[1]、AAVhu.37[1]、AAV2/8、AAV2/rh10[2]、AAV9、AAV2、NP40[3]、NP59[2,3]、AAV3B[5]、AAV-DJ[4]、AAV-LK01[4]、AAV-LK02[4]、AAV-LK03[4]、AAV-LK19[4]、AAV5[7] 腺病毒（Ad5，HC-AdV[6]）	1. Wang 等人, *Mol. Ther.*[分子疗法] **18**, 118-25 (2010) 2. Ginn 等人, *JHEP Reports* [JHEP 报告], 100065 (2019) 3. Paulk 等人, *Mol. Ther.*[分子疗法] **26**, 289-303 (2018). 4. L. Lisowski 等人, *Nature.*[自然] **506**, 382-6 (2014). 5. L. Wang 等人, *Mol. Ther.*[分子疗法] **23**, 1877-87

[0670]

		(2015). 6. Hausl *Mol Ther* [分子疗法] (2010) 7.Davidoff 等人, *Mol. Ther.*[分子疗法]11, 875-88 (2005)
肺	AAV（AAV4、AAV5、AAV6[1]、AAV9、H22[2]）腺病毒（Ad5、Ad3、Ad21、Ad14）[3]	1. Duncan 等人, *Mol Ther Methods Clin Dev* [分子疗法方法与临床发展] (2018) 2. Cooney 等人, Am J Respir Cell Mol Biol [美国呼吸道细胞与分子生物学杂志] (2019) 3. Li 等人, *Mol Ther Methods Clin Dev* [分子疗法方法与临床发展] (2019)
皮肤	AAV (AAV6[1], AAV-LK19[2])	1. Petek 等人, *Mol. Ther.*[分子疗法] (2010) 2. L. Lisowski 等人, *Nature.*[自然] **506**, 382-6 (2014).
HSC	腺病毒（HDAd5/35^{++}）	Wang 等人 *Blood Adv* [血液研究进展] (2019)

[0671] 在一些实施例中,药物组合物(例如,包含如本文所述的AAV的药物组合物)具有少于10%的空衣壳、少于8%的空衣壳、少于7%的空衣壳、少于5%的空衣壳、少于3%的空衣壳、或少于1%的空衣壳。在一些实施例中,药物组合物具有少于约5%的空衣壳。在一些实施例中,空衣壳的数量低于检测限。在一些实施例中,药物组合物具有少量空衣壳是有利的,原因是例如空衣壳可能产生例如很少或没有实质性的治疗益处的不良反应(例如,免疫反应、炎性反应、肝脏反应和/或心脏反应)。

[0672] 在一些实施例中,药物组合物中的残余宿主细胞蛋白(rHCP)少于或等于100ng/ml rHCP/1x 10^{13}vg/ml,例如,少于或等于40ng/ml rHCP/1x 10^{13}vg/ml或1-50ng/ml rHCP/1x 10^{13}vg/ml。在一些实施例中,药物组合物包含少于10ng rHCP/1.0x 10^{13}vg、或少于5ng rHCP/1.0x 10^{13}vg、少于4ng rHCP/1.0x 10^{13}vg、或少于3ng rHCP/1.0x 10^{13}vg,或介于之间的任何浓度。在一些实施例中,药物组合物中的残余宿主细胞DNA(hcDNA)少于或等于5x 10^{6}pg/ml hcDNA/1x 10^{13}vg/ml、少于或等于1.2x 10^{6}pg/ml hcDNA/1x10^{13}vg/ml、或1x

10^5pg/ml hcDNA/1x 10^{13}vg/ml。在一些实施例中,所述药物组合物中的残余宿主细胞DNA少于5.0x 10^5pg/1x 10^{13}vg、少于2.0x 10^5pg/1.0x 10^{13}vg、少于1.1x 10^5pg/1.0x 10^{13}vg、少于1.0x 10^5pg hcDNA/1.0x10^{13}vg、少于0.9x 10^5pg hcDNA/1.0x 10^{13}vg、少于0.8x 10^5pg hcDNA/1.0x10^{13}vg,或介于之间的任何浓度。

[0673] 在一些实施例中,药物组合物中的残余质粒DNA少于或等于1.7x 10^5pg/ml/1.0x 10^{13}vg/ml、或1x 10^5pg/ml/1x 1.0x 10^{13}vg/ml、或1.7x 10^6pg/ml/1.0x 10^{13}vg/ml。在一些实施例中,药物组合物中的残余DNA质粒少于10.0x 10^5pg/1.0x 10^{13}vg、少于8.0x 10^5pg/1.0x 10^{13}vg或少于6.8x 10^5pg/1.0x 10^{13}vg。在实施例中,药物组合物包含少于0.5ng/1.0x 10^{13}vg、少于0.3ng/1.0x 10^{13}vg、少于0.22ng/1.0x 10^{13}vg或少于0.2ng/1.0x 10^{13}vg或任何中间浓度的牛血清白蛋白(BSA)。在实施例中,药物组合物中的全能核酸酶(benzonase)为少于0.2ng/1.0x 10^{13}vg、少于0.1ng/1.0x 10^{13}vg、少于0.09ng/1.0x 10^{13}vg、少于0.08ng/1.0x 10^{13}vg或任何中间浓度。在实施例中,药物组合物中的泊洛沙姆188(Poloxamer 188)为约10至150ppm、约15至100ppm或约20至80ppm。在实施例中,药物组合物中的铯为少于50pg/g(ppm)、少于30pg/g(ppm)或少于20pg/g(ppm)或任何中间浓度。

[0674] 在实施例中,药物组合物包含少于10%、少于8%、少于7%、少于6%、少于5%、少于4%、少于3%、少于2%或介于之间的任何百分比的总杂质,例如,如通过SDS-PAGE测定。在实施例中,例如,如通过SDS-PAGE测定的总纯度为大于90%、大于92%、大于93%、大于94%、大于95%、大于96%、大于97%、大于98%、或介于之间的任何百分比。在实施例中,例如,如通过SDS-PAGE测量的,没有单一的未命名相关杂质多于5%、多于4%、多于3%或多于2%、或介于之间的任何百分比。在实施例中,药物组合物包含的填充的衣壳相对于总衣壳(例如,如通过分析型超速离心测量的峰1+峰2)的百分比为大于85%、大于86%、大于87%、大于88%、大于89%、大于90%、大于91%、大于91.9%、大于92%、大于93%,或介于之间的任何百分比。在药物组合物的实施例中,通过分析型超速离心在峰1中测量的填充的衣壳的百分比为20-80%、25-75%、30-75%、35-75%或37.4-70.3%。在药物组合物的实施例中,通过分析型超速离心在峰2中测量的填充的衣壳的百分比为20%-80%、20%-70%、22%-65%、24%-62%、或24.9%-60.1%。

[0675] 在一个实施例中,药物组合物包含1.0至5.0x 10^{13}vg/mL、1.2至3.0x10^{13}vg/mL或1.7至2.3x 10^{13}vg/ml的基因组效价。在一个实施例中,药物组合物显示出小于5CFU/mL、小于4CFU/mL、小于3CFU/mL、小于2CFU/mL或小于1CFU/mL或任何中间浓度的生物负载。在实施例中,根据USP,例如USP<85>(通过援引以其全文并入)的内毒素的量少于1.0EU/mL、少于0.8EU/mL或少于0.75EU/mL。在实施例中,根据USP,例如USP<785>(通过援引以其全文并入)的药物组合物的渗透压摩尔浓度为350至450mOsm/kg、370至440mOsm/kg或390至430mOsm/kg。在实施例中,药物组合物含有少于1200个大于25μm的颗粒/容器、少于1000个大于25μm的颗粒/容器、少于500个大于25μm的颗粒/容器或任何中间值。在实施例中,药物组合物含有少于10,000个大于10μm的颗粒/容器、少于8000个大于10μm的颗粒/容器、或少于600个大于10pm的颗粒/容器。

[0676] 在一个实施例中,药物组合物具有0.5至5.0x 10^{13}vg/mL、1.0至4.0x10^{13}vg/mL、1.5至3.0x 10^{13}vg/ml或1.7至2.3x 10^{13}vg/ml的基因组效价。在一个实施例中,本文所述的药物组合物包含以下中的一项或多项:小于约0.09ng全能核酸酶/1.0x 10^{13}vg,小于约

30pg/g(ppm)的铯,约20至80ppm泊洛沙姆188,小于约0.22ng BSA/1.0x 10^{13}vg,小于约6.8x 10^{5}pg的残余DNA质粒/1.0x 10^{13}vg,小于约1.1x 10^{5}pg的残余hcDNA/1.0x10^{13}vg,小于约4ng的rHCP/1.0x 10^{13}vg,pH 7.7至8.3,约390至430mOsm/kg,小于约600个大小>25μm的颗粒/容器,小于约6000个大小>10μm的颗粒/容器,约1.7x 10^{13}-2.3x 10^{13}vg/mL基因组效价,约3.9x 10^{8}至8.4x10^{10}IU/1.0x 10^{13}vg的感染效价,约100-300pg/1.0x 10^{13}vg的总蛋白,在约7.5x 10^{13}vg/kg剂量的病毒载体情况下A7SMA小鼠>24天的平均存活,根据基于体外细胞的测定的约70%至130%相对效力和/或小于约5%空衣壳。在各种实施例中,本文所述的药物组合物包含本文讨论的任何病毒颗粒,该药物组合物保留了参考标准品的±20%之间、±15%之间、±10%之间、或±5%内的效力。在一些实施例中,使用合适的体外细胞测定或体内动物模型来测量效力。

[0677] WO 2019094253中传授了制备、表征和给予AAV颗粒的另外的方法,该文献通过援引以其全文并入本文。

[0678] 可与本发明一致使用的其他rAAV构建体包括Wang等人2019中描述的那些,可在以下网址获得://doi.org/10.1038/s41573-019-0012-9,包括其表1,将该文献通过援引以其全文并入。

[0679] 脂质纳米颗粒

[0680] 本文提供的方法和系统可以采用任何合适的载剂或递送形式,在某些实施例中包括脂质纳米颗粒(LNP)。在一些实施例中,脂质纳米颗粒包含一种或多种离子脂质,诸如非阳离子脂质(例如,中性或阴离子或两性离子脂质);一种或多种缀合脂质(如WO 2019217941的表5中描述的PEG缀合脂质或缀合至聚合物的脂质;其通过援引以其全文并入本文);一种或多种固醇(例如,胆固醇);以及,任选地,一种或多种靶向分子(例如,缀合的受体、受体配体、抗体);或前述内容的组合。

[0681] 可用于形成纳米颗粒(例如,脂质纳米颗粒)的脂质包括例如WO 2019217941(通过援引并入)的表4中描述的那些—例如,含脂质的纳米颗粒可包含WO 2019217941的表4中的一种或多种脂质。脂质纳米颗粒可以包括另外的要素,如聚合物,如通过援引并入的WO 2019217941的表5中描述的聚合物。

[0682] 在一些实施例中,缀合脂质,当存在时,可以包括以下的一种或多种:PEG-二酰基甘油(DAG)(如1-(单甲氧基-聚乙二醇)-2,3-二肉豆蔻酰甘油(PEG-DMG))、PEG-二烷氧基丙基(DAA)、PEG-磷脂、PEG-神经酰胺(Cer)、聚乙二醇化磷脂酰乙醇胺(PEG-PE)、PEG琥珀酸二酰基甘油(PEGS-DAG)(如4-O-(2',3'-二(十四烷酰氧基)丙基-1-O-(w-甲氧基(聚乙氧基)乙基)丁二酸酯(PEG-S-DMG))、PEG二烷氧基丙基氨基甲酸酯、N-(羰基-甲氧基聚乙二醇2000)-1,2-二硬脂酰-sn-甘油-3-磷酸乙醇胺钠盐,以及在WO 2019051289的表2中描述的那些(通过援引并入)和前述的组合。

[0683] 在一些实施例中,可掺入脂质纳米颗粒中的固醇包括胆固醇或胆固醇衍生物中的一种或多种,如通过援引并入的WO2009/127060或US 2010/0130588中的那些。另外的示例性固醇包括植物固醇,包括通过援引并入本文的Eygeris等人(2020),dx.doi.org/10.1021/acs.nanolett.0c01386中描述的那些。

[0684] 在一些实施例中,脂质颗粒包含可电离脂质、非阳离子脂质、抑制颗粒聚集的缀合脂质和固醇。这些组分的量可以独立地变化,以获得所需特性。例如,在一些实施例中,脂质

纳米颗粒包含:可电离脂质,其量是总脂质的约20mol％至约90mol％(在其他实施例中,它可以是存在于脂质纳米颗粒中的总脂质的20％-70％(mol)、30％-60％(mol)或40％-50％(mol);约50mol％至约90mol％);非阳离子脂质,其量是总脂质的约5mol％至约30mol％;缀合脂质,其量是总脂质的约0.5mol％至约20mol％,以及固醇,其量是总脂质的约20mol％至约50mol％。总脂质与核酸(例如,编码基因修饰多肽或模板核酸)的比率可以根据需要而变化。例如,总脂质与核酸(质量或重量)的比率可为约10:1至约30:1。

[0685] 在一些实施例中,可电离脂质可以是阳离子脂质、可电离阳离子脂质,例如可以根据pH以带正电荷的形式或中性形式存在的阳离子脂质,或可以容易地质子化的含胺脂质。在一些实施例中,阳离子脂质是例如在生理条件下能够带正电的脂质。示例性的阳离子脂质包括一个或多个带有正电荷的胺基。在一些实施例中,脂质颗粒包含阳离子脂质与中性脂质、可电离含胺脂质、生物可降解炔烃脂质、类固醇、包括多不饱和脂质的磷脂、结构脂质(例如固醇)、PEG、胆固醇和聚合物缀合脂质一起配制。在一些实施例中,阳离子脂质可以是可电离的阳离子脂质。如本文所披露的示例性阳离子脂质可具有超过6.0的有效pKa。在实施例中,脂质纳米颗粒可包含具有与第一阳离子脂质不同的有效pKa(例如,大于第一有效pKa)的第二阳离子脂质。脂质纳米颗粒可包含40mol％至60mol％的阳离子脂质、中性脂质、类固醇、聚合物缀合脂质和治疗剂,例如本文所述的核酸(例如RNA)(例如模板核酸或编码基因修饰多肽的核酸),包封在脂质纳米颗粒内或与脂质纳米颗粒相关联。在一些实施例中,核酸与阳离子脂质共同配制。核酸可以吸附到LNP(例如包含阳离子脂质的LNP)的表面。在一些实施例中,核酸可以包封在LNP(例如包含阳离子脂质的LNP)中。在一些实施例中,脂质纳米颗粒可包含靶向部分,例如用靶向剂包被的靶向部分。在实施例中,LNP配制品是生物可降解的。在一些实施例中,包含一种或多种本文所述的脂质(例如式(i)、(ii)、(ii)、(vii)和/或(ix))的脂质纳米颗粒包封至少1％、至少5％、至少10％、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92％、至少95％、至少97％、至少98％或100％的RNA分子,例如,模板RNA和/或编码基因修饰多肽的mRNA。

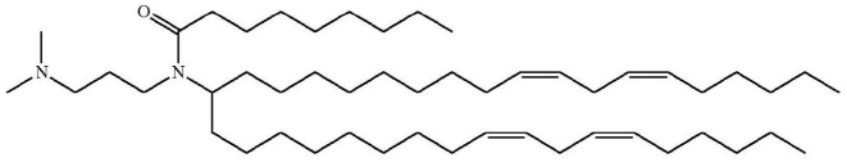
[0686] 在一些实施例中,脂质与核酸的比率(质量/质量比率;w/w比率)可以在以下范围中:约1:1至约25:1、约10:1至约14:1、约3:1至约15:1、约4:1至约10:1、约5:1至约9:1、或约6:1至约9:1。可以调节脂质和核酸的量以提供所需的N/P比,例如3、4、5、6、7、8、9、10或更高的N/P比。通常,脂质纳米颗粒配制品的总脂质含量可在约5mg/ml至约30mg/mL的范围内。

[0687] 可用于脂质纳米颗粒配制品中的示例性可电离脂质包括但不限于通过援引并入本文的WO 2019051289的表1中所列的那些。另外的示例性脂质包括但不限于下式中的一种或多种:US 2016/0311759的X;US 20150376115或US 2016/0376224中的I;US 20160151284的I、II或III;US 20170210967的I、IA、II或IIA;US 20150140070的I-c;US 2013/0178541的A;US 2013/0303587或US 2013/0123338的I;US 2015/0141678的I;US 2015/0239926的II、III、IV或V;US 2017/0119904的I;WO 2017/117528的I或II;US 2012/0149894的A;US 2015/0057373的A;WO 2013/116126的A;US 2013/0090372的A;US 2013/0274523的A;US 2013/0274504的A;US 2013/0053572的A;WO2013/016058的A;WO2012/162210的A;US 2008/042973的I;US 2012/01287670的I、II、III或IV;US 2014/0200257的I或II;US 2015/0203446的I、II或III;US 2015/0005363的I或III;US 2014/0308304的I、IA、IB、IC、ID、II、IIA、IIB、IIC、IID或III-XXIV;US 2013/0338210;WO2009/132131的I、II、III或IV;US

2012/01011478的A;US 2012/0027796的I或XXXV;US 2012/0058144的XIV或XVII;US 2013/0323269的;US 2011/0117125的I;US 2011/0256175的I、II或III;US 2012/0202871的I、II、III、IV、V、VI、VII、VIII、IX、X、XI、XII;US 2011/0076335的I、II、III、IV、V、VI、VII、VIII、X、XII、XIII、XIV、XV或XVI;US 2006/008378的I或II;US 2013/0123338的I;US 2015/0064242的I或X-A-Y-Z;US 2013/0022649的XVI、XVII或XVIII;US 2013/0116307的I、II或III;US 2013/0116307的I、II或III;US 2010/0062967的I或II;US 2013/0189351的I-X;US 2014/0039032的I;US 2018/0028664的V;US2016/0317458的I;US 2013/0195920的I;US 10,221,127的5、6或10;WO 2018/081480的III-3;WO 2020/081938的I-5或I-8;US 9,867,888的18或25;US 2019/0136231的A;WO 2020/219876的II;US 2012/0027803的1;US 2019/0240349的OF-02;US 10,086,013的23;Miao等人(2020)的cKK-E12/A6;WO 2010/053572的C12-200;Dahlman等人(2017)的7C1;Whitehead等人的304-013或503-013;US 9,708,628的TS-P4C2;WO 2020/106946的I;WO 2020/106946的I。

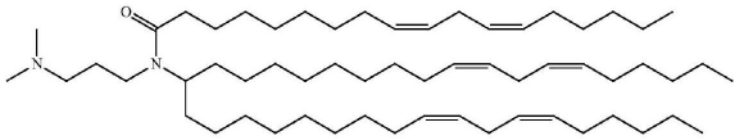
[0688] 在一些实施例中,可电离脂质是MC3(6Z,9Z,28Z,31Z)-三十七烷-6,9,28,31-四烯-19-基-4-(二甲基氨基)丁酸酯(DLin-MC3-DMA或MC3),例如,如WO 2019051289A9(通过援引以其全文并入本文)的实例9中所述。在一些实施例中,可电离脂质是脂质ATX-002,例如,如WO 2019051289A9(通过援引以其全文并入本文)的实例10中所述。在一些实施例中,可电离脂质是(13Z,16Z)-A,A-二甲基-3-壬基二十二-13,16-二烯-1-胺(化合物32),例如,如WO 2019051289 A9(通过援引以其全文并入本文)的实例11中所述。在一些实施例中,可电离脂质是化合物6或化合物22,例如,如WO 2019051289A9(通过援引以其全文并入本文)的实例12中所述。在一些实施例中,可电离脂质是十七烷-9-基8-((2-羟乙基)(6-氧代-6-(十一烷氧基)己基)氨基)辛酸酯(SM-102);例如,如US9,867,888(其通过援引以其全文并入本文)的实例1中所述。在一些实施例中,可电离脂质是9Z,12Z)-3-((4,4-双(辛基氧基)丁酰基)氧基)-2-((((3-(二乙基氨基)丙氧基)羰基)氧基)甲基)丙基十八碳-9,12-二烯酸酯(LP01),例如,如WO 2015/095340(其通过援引以其全文并入本文)的实例13中合成的。在一些实施例中,可电离脂质是9-((4-二甲基氨基)丁酰基)氧基)十七烷二酸二((Z)-壬-2-烯-1-基)酯(L319),例如如US 2012/0027803(通过援引以其全文并入本文)的实例7、8或9中合成的。在一些实施例中,可电离脂质是1,1'-((2-(4-(2-((2-(双(2-羟基十二烷基)氨基)乙基)(2-羟基十二烷基)氨基)乙基)哌嗪-1-基)乙基)氮烷二基)双(十二烷-2-醇)(C12-200),例如,如WO 2010/053572(其通过援引以其全文并入本文)的实例14和16中合成的。在一些实施例中,可电离脂质是;咪唑胆固醇酯(ICE)脂质(3S,10R,13R,17R)-10,13-二甲基-17-((R)-6-甲基庚-2-基)-2,3,4,7,8,9,10,11,12,13,14,15,16,17-十四氢-1H-环戊[a]菲-3-基3-(1H-咪唑-4-基)丙酸酯,例如来自WO 2020/106946(其通过援引以其全文并入本文)的结构(I)。

[0689] 可用于(例如,与其他脂质组分组合)形成用于递送本文所述的组合物,例如本文所述的核酸(例如,RNA)(例如,模板核酸或编码基因修饰多肽的核酸)的脂质纳米颗粒的脂质化合物的一些非限制性实例包括:

[0690] 

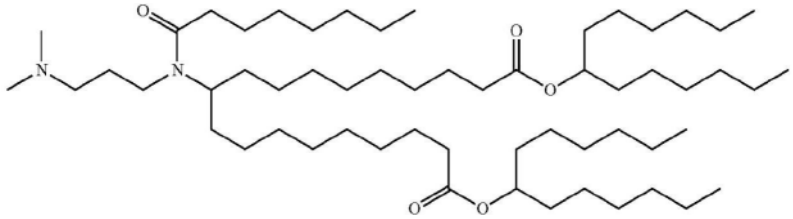
(i)

[0691] 在一些实施例中,包含式(i)的LNP用于将本文所述的基因修饰组合物递送至肝和/或肝细胞。

[0692] 

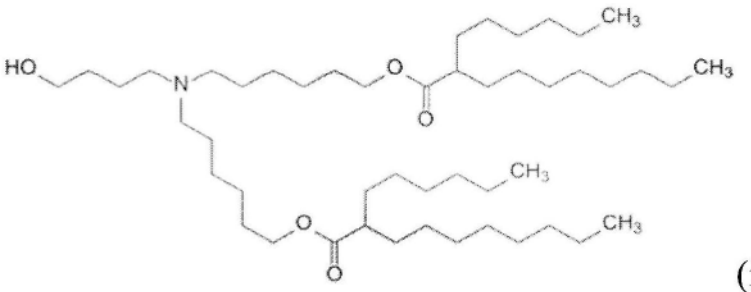
(ii)

[0693] 在一些实施例中,包含式(ii)的LNP用于将本文所述的基因修饰组合物递送至肝和/或肝细胞。

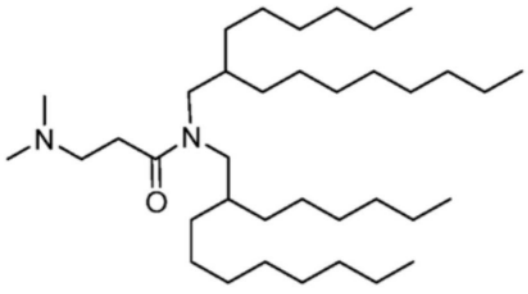
[0694] 

(iii)

[0695] 在一些实施例中,包含式(iii)的LNP用于将本文所述的基因修饰组合物递送至肝和/或肝细胞。

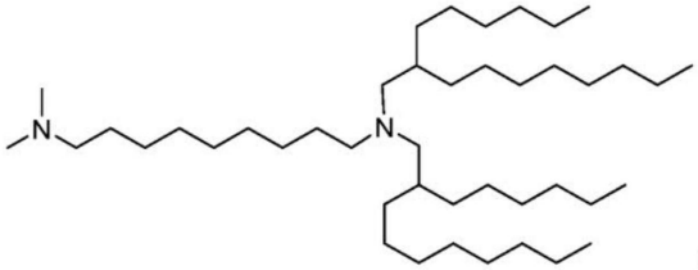
[0696] 

(iv)

[0697] 

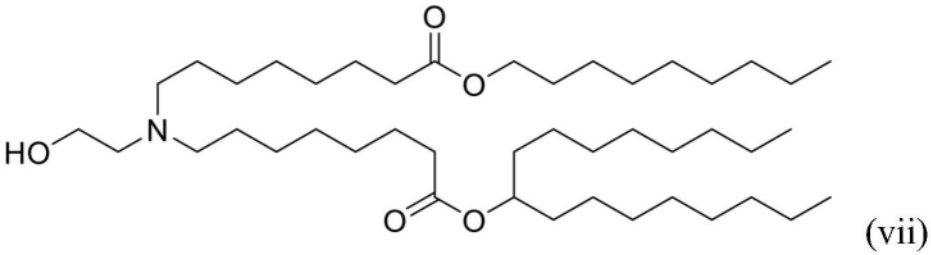
(v)

[0698] 在一些实施例中,包含式(v)的LNP用于将本文所述的基因修饰组合物递送至肝和/或肝细胞。

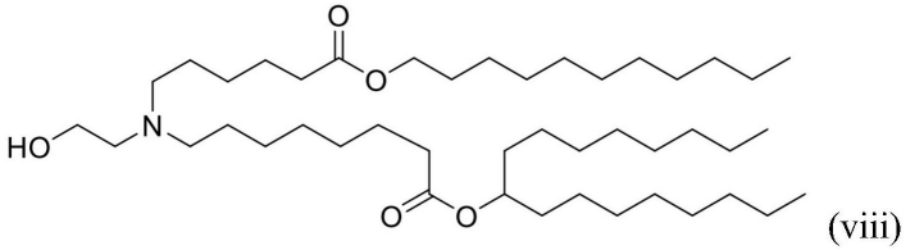
[0699] 

(vi)

[0700] 在一些实施例中,包含式(vi)的LNP用于将本文所述的基因修饰组合物递送至肝和/或肝细胞。

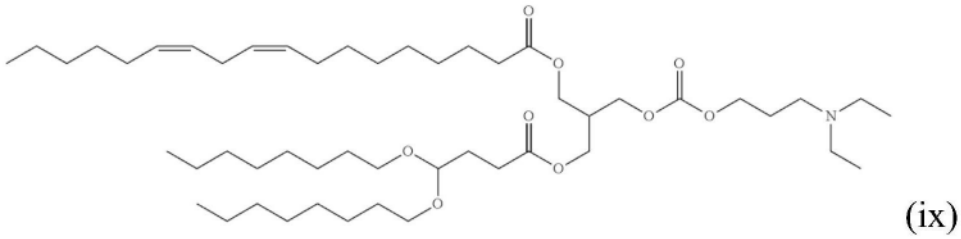
[0701] 

(vii)



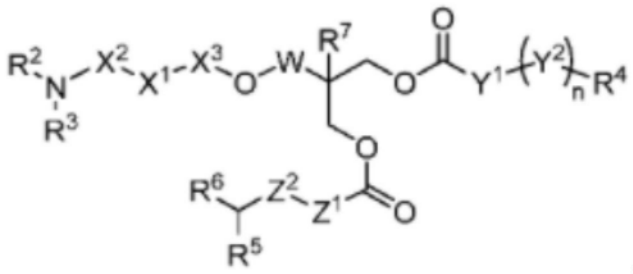
(viii)

[0702] 在一些实施例中,包含式(viii)的LNP用于将本文所述的基因修饰组合物递送至肝和/或肝细胞。

[0703] 

(ix)

[0704] 在一些实施例中,包含式(ix)的LNP用于将本文所述的基因修饰组合物递送至肝和/或肝细胞。

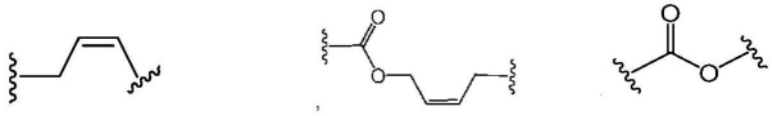
[0705] 

(x)

[0706] 其中

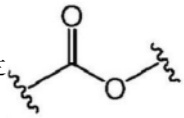
[0707] X^1是O、NR^1或直接键,X^2是C2-5亚烷基,X^3是C(=O)或直接键,R^1是H或Me,R^3是Ci-3烷基,R^2是Ci-3烷基,或R^2与它所附接的氮原子和X^2的1-3个碳原子一起形成4元、5元或6元环,或X^1是NR^1,R^1和R^2与它们所附接的氮原子一起形成5元或6元环,或R^2与R^3和它们所附接

的氮原子一起形成5元、6元或7元环，Y^1是C2-12亚烷基，Y^2选自

[0708] 

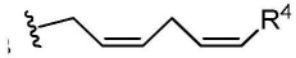
[0709] (在任一取向上)，(在任一取向上)，(在任一取向上)，

[0710] n是0至3，R^4是Ci-15烷基，Z^1是Ci-6亚烷基或直接键，

[0711] Z^2是 

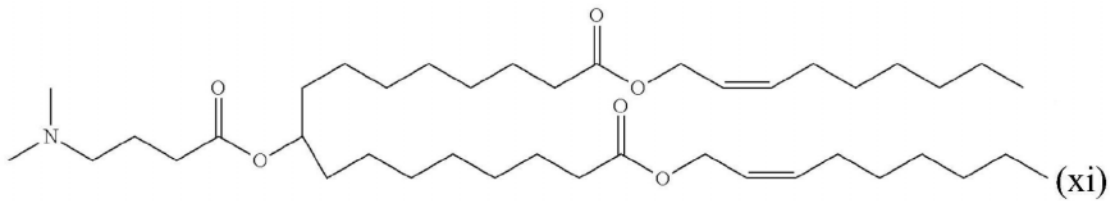
[0712] (在任一取向上)或不存在，条件是如果Z^1是直接键，则Z^2不存在；

[0713] R^5是C5-9烷基或C6-10烷氧基，R^6是C5-9烷基或C6-10烷氧基，W是亚甲基或直接键，并且R^7是H或Me，或其盐，条件是如果R^3和R^2是C2烷基，X^1是O，X^2是直链C3亚烷基，X^3是C(=O)，Y^1是直链Ce亚烷基，$(Y^2)n-R^4$是

[0714] 

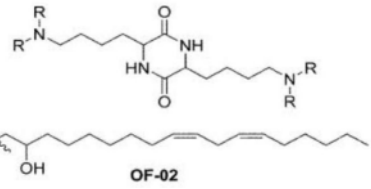
[0715] ，R^4是直链C5烷基，Z^1是C2亚烷基，Z^2不存在，W是亚甲基，并且R^7是H，则R^5和R^6不是Cx烷氧基。

[0716] 在一些实施例中，包含式(xii)的LNP用于将本文所述的基因修饰组合物递送至肝和/或肝细胞。

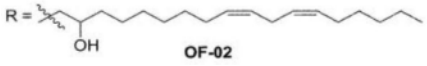
[0717] 

(xi)

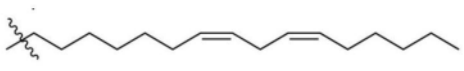
[0718] 在一些实施例中，包含式(xi)的LNP用于将本文所述的基因修饰组合物递送至肝和/或肝细胞。

[0719] 

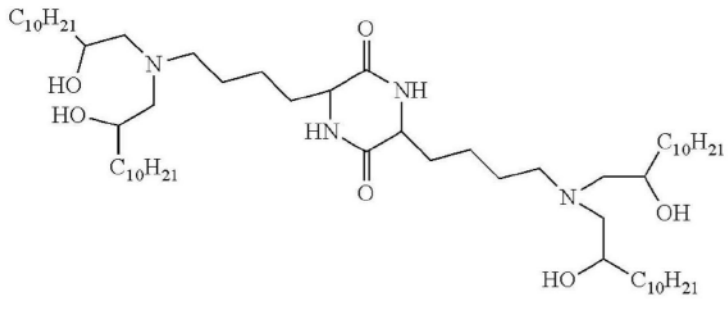
其中



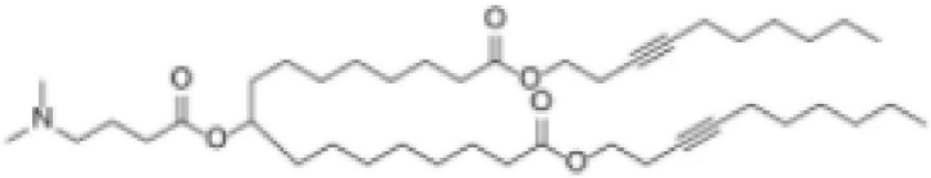
OF-02

R= 

(xii)



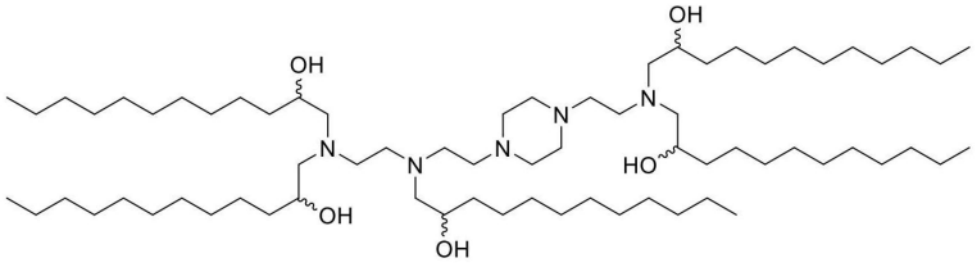
(xiii)



(xiv)

[0720] 在一些实施例中,LNP包含式(xiii)的化合物和式(xiv)的化合物。

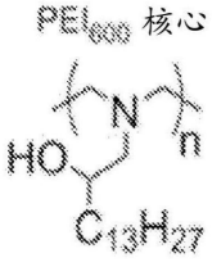
[0721]



(xv)

[0722] 在一些实施例中,包含式(xv)的LNP用于将本文所述的基因修饰组合物递送至肝和/或肝细胞。

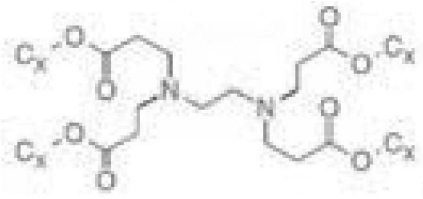
[0723]



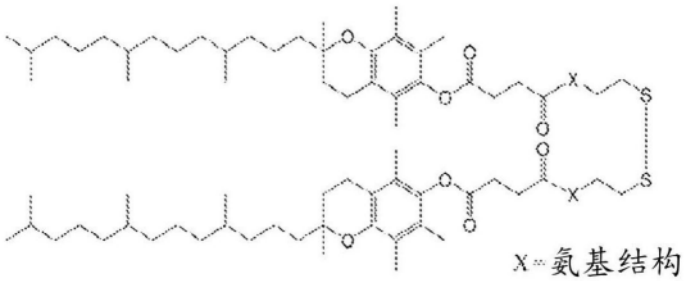
(xvi)

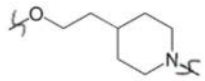
[0724] 在一些实施例中,包含式(xvi)的配制品的LNP用于将本文所述的基因修饰组合物递送至肺内皮细胞。

[0725]

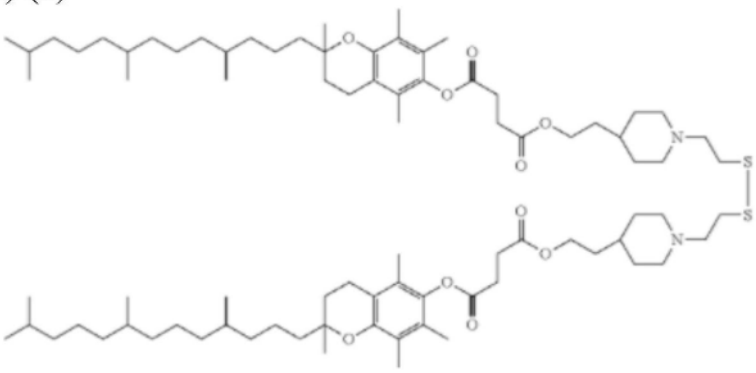


(xvii)

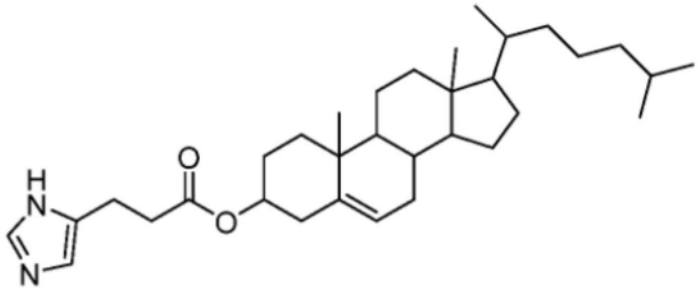
[0726] 

其 中 X= 

(xviii) (a)

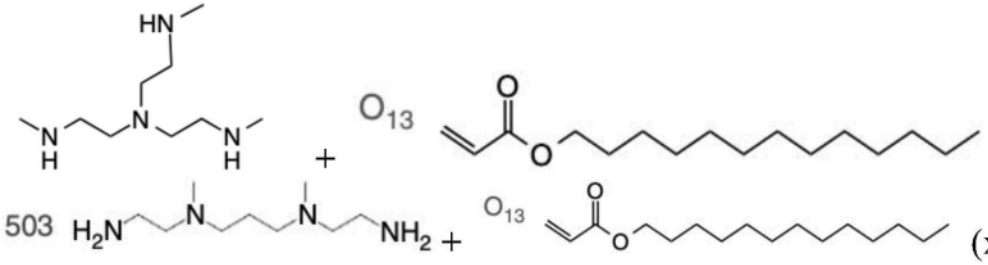


(xviii)(b)



(xix)

[0727] 在一些实施例中,用于形成用于递送本文所述组合物(例如本文所述的核酸(例如,RNA)(例如模板核酸或编码基因修饰多肽的核酸))的脂质纳米颗粒的脂质化合物通过以下反应之一制备:

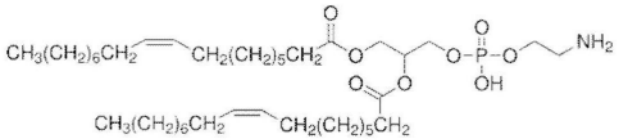
[0728] 

(xx) (a)

(xx)(b)

[0729] 示例性的非阳离子脂质包括但不限于二硬脂酰-sn-甘油-磷酸乙醇胺、二硬脂酰磷脂酰胆碱(DSPC)、二油酰磷脂酰胆碱(DOPC)、二棕榈酰磷脂酰胆碱(DPPC)、二油酰磷脂酰甘油(DOPG)、二棕榈酰磷脂酰甘油(DPPG)、二油酰-磷脂酰乙醇胺(DOPE)、1,2-二油酰-sn-甘油-3-磷酸乙醇胺(DOPE)、棕榈酰油酰磷脂酰胆碱(POPC)、棕榈酰油酰磷脂酰乙醇胺(POPE)、二油酰-磷脂酰乙醇胺4-(N-马来酰亚胺甲基)-环已烷-1-甲酸盐(DOPE-mal)、二棕榈酰磷脂酰乙醇胺(DPPE)、二肉豆蔻酰磷酸乙醇胺(DMPE)、二硬脂酰-磷脂酰-乙醇胺(DSPE)、单甲基-磷脂酰乙醇胺(例如16-O-单甲基PE)、二甲基-磷脂酰乙醇胺(例如16-O-二甲基PE)、18-1-反式PE,1-硬脂酰-2-油酰-磷脂酰乙醇胺(SOPE)、氢化大豆磷脂酰胆碱

(HSPC)、蛋磷脂酰胆碱(EPC)、二油酰磷脂酰丝氨酸(DOPS)、鞘磷脂(SM)、二肉豆蔻酰磷脂酰胆碱(DMPC)、二肉豆蔻酰磷脂酰甘油(DMPG)、二硬脂酰磷脂酰甘油(DSPG)、二芥子酰磷脂酰胆碱(DEPC)、棕榈酰油酰磷脂酰甘油(POPG)、二反油酰-磷脂酰乙醇胺(DEPE)、卵磷脂、磷脂酰乙醇胺、溶血卵磷脂、溶血磷脂酰乙醇胺、磷脂酰丝氨酸、磷脂酰肌醇、鞘磷脂、卵鞘磷脂(ESM)、脑磷脂、心磷脂、磷脂酸、脑苷脂、双十六烷基磷酸、溶血磷脂酰胆碱、二亚油酰磷脂酰胆碱、或其混合物。应当理解,也可以使用其他二酰基磷脂酰胆碱和二酰基磷脂酰乙醇胺磷脂。这些脂质中的酰基基团优选为源自具有C10-C24碳链的脂肪酸的酰基基团,例如月桂酰基、肉豆蔻酰基、棕榈酰基、硬脂酰基或油酰基。在某些实施例中,另外的示例性脂质包括但不限于通过援引并入本文的Kim等人(2020)dx.doi.org/10.1021/acs.nanolett.0c01386中描述的那些。在一些实施例中,这样的脂质包括发现会改善用mRNA进行肝脏转染的植物脂质(例如DGTS)。在一些实施例中,非阳离子脂质可以具有以下结构,

[0730] 

(xxi)

[0731] 适合用于脂质纳米颗粒中的非阳离子脂质的其他实例包括但不限于非磷脂质,例如硬脂胺、十二烷基胺、十六烷基胺、乙酰基棕榈酸酯、蓖麻酸甘油酯、硬脂酸十六烷基酯、肉豆蔻酸异丙酯、两性丙烯酸聚合物、三乙醇胺-月桂基硫酸酯、烷基-芳基硫酸酯、聚乙氧基化脂肪酸酰胺、双十八烷基二甲基溴化铵、神经酰胺、鞘磷脂等。其他非阳离子脂质在WO 2017/099823或美国专利公开US 2018/0028664中描述,其内容通过援引以其全文并入本文。

[0732] 在一些实施例中,非阳离子脂质是油酸或通过援引以其全文并入本文的US 2018/0028664的式I、II或IV的化合物。非阳离子脂质可以占脂质纳米颗粒中存在的总脂质的例如0-30%(mol)。在一些实施例中,非阳离子脂质含量是脂质纳米颗粒中存在的总脂质的5%-20%(mol)或10%-15%(mol)。在实施例中,可电离脂质与中性脂质的摩尔比为约2:1至约8:1(例如,约2:1、3:1、4:1、5:1、6:1、7:1或8:1)。

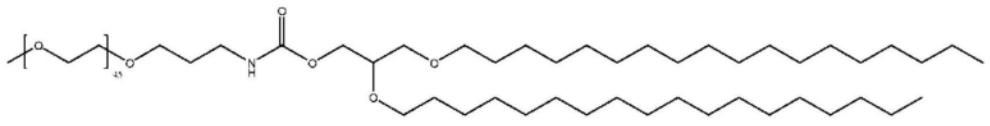
[0733] 在一些实施例中,脂质纳米颗粒不包含任何磷脂。

[0734] 在一些实施例中,脂质纳米颗粒可进一步包含诸如固醇的组分以提供膜完整性。可用于脂质纳米颗粒中的一种示例性固醇是胆固醇及其衍生物。胆固醇衍生物的非限制性实例包括极性类似物,诸如5a-胆甾烷醇、53-粪甾烷醇、胆甾醇基-(2,-羟基)-乙基醚、胆甾醇基-(4'-羟基)-丁基醚和6-酮胆甾烷醇;非极性类似物,诸如5a-胆甾烷、胆甾烯酮、5a-胆甾烷酮、5p-胆甾烷酮和胆甾醇癸酸酯;及其混合物。在一些实施例中,胆固醇衍生物是极性类似物,例如,胆甾醇基-(4'-羟基)-丁基醚。示例性的胆固醇衍生物在PCT公开WO 2009/127060和美国专利公开US 2010/0130588中描述,其中每个通过援引以其全文并入本文。

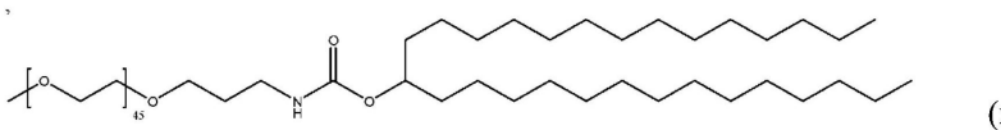
[0735] 在一些实施例中,提供膜完整性的组分,诸如固醇,可占脂质纳米颗粒中存在的总脂质的0-50%(mol)(例如,0-10%、10%-20%、20%-30%、30%-40%或40%-50%)。在一些实施例中,这样的组分是脂质纳米颗粒的总脂质含量的20%-50%(mol)、30%-40%(mol)。

[0736] 在一些实施例中,脂质纳米颗粒可包含聚乙二醇(PEG)或缀合的脂质分子。通常,这些用于抑制脂质纳米颗粒的聚集和/或提供空间稳定。示例性的缀合脂质包括但不限于PEG-脂质缀合物、聚噁唑啉(POZ)-脂质缀合物、聚酰胺-脂质缀合物(如ATTA-脂质缀合物)、阳离子聚合物脂质(CPL)缀合物及其混合物。在一些实施例中,缀合脂质分子是PEG-脂质缀合物,例如(甲氧基聚乙二醇)缀合脂质。

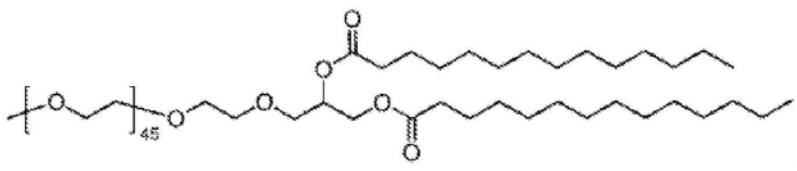
[0737] 示例性的PEG-脂质缀合物包括但不限于PEG-二酰基甘油(DAG)(诸如1-(单甲氧基-聚乙二醇)-2,3-二肉豆蔻酰甘油(PEG-DMG))、PEG-二烷氧基丙基(DAA)、PEG-磷脂、PEG-神经酰胺(Cer)、聚乙二醇化磷脂酰乙醇胺(PEG-PE)、1,2-二肉豆蔻酰基-sn-甘油,甲氧基聚乙二醇(DMG-PEG-2K)、PEG琥珀酸二酰基甘油(PEGS-DAG)(诸如4-O-(2',3'-二(十四烷酰基氧基)丙基-1-O-(w-甲氧基(聚乙氧基)乙基)丁二酸酯(PEG-S-DMG))、PEG二烷氧基丙基氨基甲酸酯、N-(羰基-甲氧基聚乙二醇2000)-1,2-二硬脂酰基-sn-甘油-3-磷酸乙醇胺钠盐或其混合物。另外的示例性PEG-脂质缀合物例如在US 5,885,613、US6,287,591、US 2003/0077829、US 2003/0077829、US 2005/0175682、US 2008/0020058、US 2011/0117125、US 2010/0130588、US 2016/0376224、US 2017/0119904和US/099823中描述,所有这些的内容通过援引以其全文并入本文。在一些实施例中,PEG-脂质是US 2018/0028664的式III、III-a-I、III-a-2、III-b-1、III-b-2或V的化合物,其内容通过援引以其全文并入本文。在一些实施例中,PEG-脂质具有US 20150376115或US2016/0376224的式II,两者的内容通过援引以其全文并入本文。在一些实施例中,PEG-DAA缀合物可以是例如PEG-二月桂基氧基丙基、PEG-二肉豆蔻基氧基丙基、PEG-二棕榈基氧基丙基或PEG-二硬脂基氧基丙基。PEG-脂质可以是以下的一种或多种:PEG-DMG、PEG-二月桂基甘油、PEG-二棕榈酰甘油、PEG-二硬脂基甘油、PEG-二月桂基甘油脂酰胺、PEG-二肉豆蔻基甘油脂酰胺、PEG-二棕榈酰甘油脂酰胺、PEG-二硬脂基甘油脂酰胺、PEG-胆固醇(1-[8'-(胆甾-5-烯-3[β]-氧基)甲酰胺基-3',6'-二氧杂辛基]氨基甲酰基-[ω]-甲基-聚(乙二醇))、PEG-DMB(3,4-双十四烷氧基苄基-[ω]-甲基-聚(乙二醇)醚)和1,2-二肉豆蔻酰基-sn-甘油-3-磷酸乙醇胺-N-[甲氧基(聚乙二醇)-2000]。在一些实施例中,PEG-脂质包含PEG-DMG、1,2-二肉豆蔻酰基-sn-甘油-3-磷酸乙醇胺-N-[甲氧基(聚乙二醇)-2000]。在一些实施例中,PEG-脂质包含选自以下的结构:

[0738] 

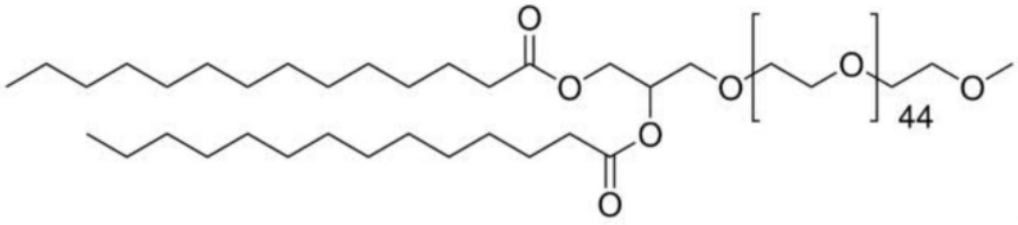
(xxii),



(xxiii),



(xxiv),和

[0739] 

(xxv)。

[0740] 在一些实施例中,与PEG以外的分子缀合的脂质也可用于代替PEG-脂质。例如,聚噁唑啉(POZ)-脂质缀合物、聚酰胺-脂质缀合物(如ATTA-脂质缀合物)和阳离子聚合物脂质(GPL)缀合物可用于代替PEG-脂质或与PEG-脂质一起使用。

[0741] 示例性缀合脂质,即PEG-脂质、(POZ)-脂质缀合物、ATTA-脂质缀合物和阳离子聚合物-脂质在WO 2019051289 A9和WO 2020106946 A1的表2中列出的PCT和LIS专利申请中描述,所有这些的内容通过援引以其全文并入本文。

[0742] 在一些实施例中,LNP包含式(xix)化合物、式(xxi)化合物和式(xxv)化合物。在一些实施例中,包含式(xix)、式(xxi)和式(xxv)的配制品的LNP用于将本文所述的基因修饰组合物递送至肺或肺细胞。

[0743] 在一些实施例中,脂质纳米颗粒可包含一种或多种选自式(i)、式(ii)、式(iii)、式(vii)和式(ix)的阳离子脂质。在一些实施例中,LNP可以进一步包含一种或多种中性脂质,例如DSPC、DPPC、DMPC、DOPC、POPC、DOPE、SM,类固醇,例如胆固醇,和/或一种或多种聚合物缀合的脂质,例如聚乙二醇化脂质,例如PEG-DAG、PEG-PE、PEG-S-DAG、PEG-cer或PEG二烷氧基丙基氨基甲酸酯。

[0744] 在一些实施例中,PEG或缀合脂质可以占脂质纳米颗粒中存在的总脂质的0-20%(mol)。在一些实施例中,PEG或缀合脂质的含量为脂质纳米颗粒中存在的总脂质的0.5%-10%或2%-5%(mol)。可电离脂质、非阳离子脂质、固醇和PEG/缀合脂质的摩尔比可以根据需要变化。例如,脂质颗粒可包含按组合物的摩尔或总重量计30%-70%的可电离脂质,按组合物的摩尔或总重量计0-60%的胆固醇,按组合物的摩尔或总重量计0-30%的非阳离子脂质和按组合物的摩尔或总重量计1%-10%的缀合脂质。优选地,组合物包含按组合物的摩尔或总重量计30%-40%的可电离脂质,按组合物的摩尔或总重量计40%-50%的胆固醇,和按组合物的摩尔或总重量计10%-20%的非阳离子脂质。在一些其他实施例中,该组合物是按组合物的摩尔或总重量计50%-75%的可电离脂质,按组合物的摩尔或总重量计20%-40%的胆固醇和按组合物的摩尔或总重量计5%至10%的非阳离子脂质以及按组合物的摩尔或总重量计1%-10%的缀合脂质。该组合物可以含有按组合物的摩尔或总重量计60%-70%的可电离脂质,按组合物的摩尔或总重量计25%-35%的胆固醇,以及按组合物的摩尔或总重量计5%-10%的非阳离子脂质。该组合物还可含有按组合物的摩尔或总重量计高达90%的可电离脂质和按组合物的摩尔或总重量计2%至15%的非阳离子脂质。配制品也可以是脂质纳米颗粒配制品,例如包含按组合物的摩尔或总重量计8%-30%的可电离脂质,按组合物的摩尔或总重量计5%-30%的非阳离子脂质,以及按组合物的摩尔或总重量计0-20%的胆固醇;按组合物的摩尔或总重量计4%-25%的可电离脂质,按组合物的摩尔或总重量计4%-25%的非阳离子脂质,按组合物的摩尔或总重量计2%至25%的胆固醇,按组合物的摩尔或总重量计10%至35%的缀合脂质,以及按组合物的摩尔或总重量计5%的胆固醇;或按组合物的摩尔或总重量计2%-30%的可电离脂质,按组合物的摩尔或总重

量计2%-30%的非阳离子脂质,按组合物的摩尔或总重量计1%至15%的胆固醇,按组合物的摩尔或总重量计2%至35%的缀合脂质,以及按组合物的摩尔或总重量计1%-20%的胆固醇;或按组合物的摩尔或总重量计甚至高达90%的可电离脂质和按组合物的摩尔或总重量计2%-10%的非阳离子脂质,或按组合物的摩尔或总重量计甚至100%的阳离子脂质。在一些实施例中,脂质颗粒配制品包含摩尔比为50:10:38.5:1.5的可电离脂质、磷脂、胆固醇和聚乙二醇化脂质。在一些其他实施例中,脂质颗粒配制品包含摩尔比为60:38.5:1.5的可电离脂质、胆固醇和聚乙二醇化脂质。

[0745] 在一些实施例中,脂质颗粒包含可电离脂质、非阳离子脂质(例如磷脂)、固醇(例如胆固醇)和聚乙二醇化脂质,其中可电离脂质的脂质摩尔比在20至70摩尔%的范围内,目标为40-60摩尔%,非阳离子脂质的摩尔百分比在0至30摩尔%的范围内,目标为0至15摩尔%,固醇的摩尔百分比在20至70摩尔%的范围内,目标为30至50摩尔%,并且聚乙二醇化脂质的摩尔百分比在1至6摩尔%的范围内,目标为2至5摩尔%。

[0746] 在一些实施例中,脂质颗粒包含摩尔比为50:10:38.5:1.5的可电离脂质/非阳离子脂质/固醇/缀合脂质。

[0747] 在一方面,本披露提供了包含磷脂、卵磷脂、磷脂酰胆碱和磷脂酰乙醇胺的脂质纳米颗粒配制品。

[0748] 在一些实施例中,还可以包括一种或多种另外的化合物。那些化合物可以单独施用,或者另外的化合物可以包括在本发明的脂质纳米颗粒中。换言之,除核酸或至少第二核酸之外,脂质纳米颗粒可含有不同于第一核酸的其他化合物。非限制性地,其他另外的化合物可以选自由以下组成的组:小的或大的有机分子或无机分子、单糖、二糖、三糖、寡糖、多糖、肽、蛋白质、其肽类似物和衍生物、肽模拟物、核酸、核酸类似物和衍生物、由生物材料制成的提取物,或其任何组合。

[0749] 在一些实施例中,脂质纳米颗粒(或包含脂质纳米颗粒的配制品)缺乏反应性杂质(例如,醛或酮),或包含低于预选水平的反应性杂质(例如,醛或酮)。虽然不希望受理论约束,但在一些实施例中,脂质试剂用于制备脂质纳米颗粒配制品,并且脂质试剂可包含污染性反应性杂质(例如,醛或酮)。可以基于具有低于预选水平的反应性杂质(例如,醛或酮)来选择用于制造的脂质试剂。不希望受理论束缚,在一些实施例中,醛可引起RNA的修饰和损伤,例如,碱基之间的交联和/或脂质与RNA的共价缀合(例如,形成脂质-RNA加合物)。在一些情况下,这可能导致逆转录酶反应失败和/或例如在一个或多个病变的一个或多个位点掺入不适当的碱基,例如新合成的靶DNA中的突变。

[0750] 在一些实施例中,脂质纳米颗粒配制品使用包含小于5%、4%、3%、2%、1%、0.9%、0.8%、0.7%、0.6%、0.5%、0.4%、0.3%、0.2%或0.1%的总反应性杂质(例如醛)含量的脂质试剂产生。在一些实施例中,脂质纳米颗粒配制品使用包含小于5%、4%、3%、2%、1%、0.9%、0.8%、0.7%、0.6%、0.5%、0.4%、0.3%、0.2%或0.1%的任何单一反应性杂质(例如醛)物质的脂质试剂产生。在一些实施例中,脂质纳米颗粒配制品使用脂质试剂产生,该脂质试剂包含:(i)小于5%、4%、3%、2%、1%、0.9%、0.8%、0.7%、0.6%、0.5%、0.4%、0.3%、0.2%或0.1%的总反应性杂质(例如醛)含量;和(ii)小于5%、4%、3%、2%、1%、0.9%、0.8%、0.7%、0.6%、0.5%、0.4%、0.3%、0.2%或0.1%的任何单一反应性杂质(例如醛)物质。在一些实施例中,脂质纳米颗粒配制品使用多种脂质试剂产生,

并且多种脂质试剂中的每一种独立地满足本段落中所述的一个或多个标准。在一些实施例中,多种脂质试剂中的每一种满足相同的标准,例如本段落的标准。

[0751] 在一些实施例中,脂质纳米颗粒配制品包含小于5%、4%、3%、2%、1%、0.9%、0.8%、0.7%、0.6%、0.5%、0.4%、0.3%、0.2%或0.1%的总反应性杂质(例如醛)含量。在一些实施例中,脂质纳米颗粒配制品包含小于5%、4%、3%、2%、1%、0.9%、0.8%、0.7%、0.6%、0.5%、0.4%、0.3%、0.2%或0.1%的任何单一反应性杂质(例如醛)物质。在一些实施例中,脂质纳米颗粒配制品包含:(i)小于5%、4%、3%、2%、1%、0.9%、0.8%、0.7%、0.6%、0.5%、0.4%、0.3%、0.2%或0.1%的总反应性杂质(例如醛)含量;和(ii)小于5%、4%、3%、2%、1%、0.9%、0.8%、0.7%、0.6%、0.5%、0.4%、0.3%、0.2%或0.1%的任何单一反应性杂质(例如醛)物质。

[0752] 在一些实施例中,用于如本文所述的脂质纳米颗粒或其配制品的一种或多种或任选地所有脂质试剂包含小于5%、4%、3%、2%、1%、0.9%、0.8%、0.7%、0.6%、0.5%、0.4%、0.3%、0.2%或0.1%的总反应性杂质(例如醛)含量。在一些实施例中,用于如本文所述的脂质纳米颗粒或其配制品的一种或多种或任选地所有脂质试剂包含小于5%、4%、3%、2%、1%、0.9%、0.8%、0.7%、0.6%、0.5%、0.4%、0.3%、0.2%或0.1%的任何单一反应性杂质(例如醛)物质。在一些实施例中,用于本文所述的脂质纳米颗粒或其配制品的一种或多种或任选地所有脂质试剂包含:(i)小于5%、4%、3%、2%、1%、0.9%、0.8%、0.7%、0.6%、0.5%、0.4%、0.3%、0.2%或0.1%的总反应性杂质(例如醛)含量;和(ii)小于5%、4%、3%、2%、1%、0.9%、0.8%、0.7%、0.6%、0.5%、0.4%、0.3%、0.2%或0.1%的任何单一反应性杂质(例如醛)物质。

[0753] 在一些实施例中,总醛含量和/或任何单一反应性杂质(例如醛)种类的量通过液相色谱法(LC),例如与串联质谱法(MS/MS)联用,例如根据PCT/US21/20948的实例40中所述的方法来确定。在一些实施例中,反应性杂质(例如醛)含量和/或反应性杂质(例如醛)物质的量通过检测与例如脂质试剂中反应性杂质(例如醛)的存在相关的核酸分子(例如RNA分子,例如如本文所述)的一个或多个化学修饰来确定。在一些实施例中,反应性杂质(例如醛)含量和/或反应性杂质(例如醛)种类的量通过检测例如脂质试剂中与反应性杂质(例如醛)的存在相关联的核苷酸或核苷(例如核糖核苷酸或核糖核苷,例如包含在如本文所述的模板核酸中或从其分离)的一个或多个化学修饰来确定,例如,根据PCT/US21/20948的实例41中所述的方法来确定。在实施例中,核酸分子、核苷酸或核苷的化学修饰通过测定一个或多个修饰的核苷酸或核苷的存在来检测,例如使用LC-MS/MS分析,例如,根据PCT/US21/20948的实例41中所述的方法来检测。

[0754] 在一些实施例中,本文所述的核酸(例如,RNA)(例如,模板核酸或编码基因修饰多肽的核酸)不包含醛修饰,或包含少于预选量的醛修饰。在一些实施例中,平均每1000个核苷酸,核酸具有少于50、20、10、5、2或1个醛修饰,例如,其中两个核苷酸的单个交联是单个醛修饰。在一些实施例中,醛修饰是RNA加合物(例如脂质-RNA加合物)。在一些实施例中,醛修饰的核苷酸是碱基之间的交联。在一些实施例中,本文所述的核酸(例如RNA)在核苷酸之间包含少于50、20、10、5、2或1个交联。

[0755] 在一些实施例中,通过添加靶向结构域将LNP定向至特定组织。例如,可以将生物配体展示在LNP的表面,以增强与展示同源受体的细胞的相互作用,从而推动与细胞表达受

体的组织的相关联和向其中的载物递送。在一些实施例中,生物配体可以是驱动递送至肝的配体,例如展示GalNAc的LNP促使核酸载物递送至展示无唾液酸糖蛋白受体(ASGPR)的肝细胞。Akinc等人Mol Ther[分子疗法]18(7):1357-1364(2010)的工作传授了将三价GalNAc配体与PEG-脂质缀合(GalNAc-PEG-DSG)以产生依赖于ASGPR的LNP以获得可观察的LNP载物效应(参见例如其中的图6)。其他展示配体的LNP配制品,例如掺入叶酸、转铁蛋白或抗体的配制品,在WO 2017223135中进行了讨论,其通过援引以其全文并入本文,此外还有在其中使用的参考文献也并入本文:即,Kolhatkar等人,Curr Drug Discov Technol[当代药物发现技术].2011 8:197-206;Musacchio和Torchilin,Front Biosci.[生物科学前沿]2011 16:1388-1412;Yu等人,Mol Membr Biol.[分子膜生物学]2010 27:286-298;Patil等人,Crit Rev Ther Drug Carrier Syst[治疗性药物载剂系统的重要评论].2008 25:1-61;Benoit等人,Biomacromolecules[生物大分子].2011 12:2708-2714;Zhao等人,Expert Opin Drug Deliv[药物递送专家观点].2008 5:309-319;Akinc等人,Mol Ther[分子疗法].2010 18:1357-1364;Srinivasan等人,Methods Mol Biol[分子生物学方法].2012 820:105-116;Ben-Arie等人,Methods Mol Biol[分子生物学方法].2012 757:497-507;Peer2010J Control Release[控释杂志].20:63-68;Peer等人,Proc Natl Acad Sci U SA.[美国国家科学院院刊]2007 104:4095-4100;Kim等人,Methods Mol Biol.[分子生物学方法]2011 721:339-353;Subramanya等人,Mol Ther[分子疗法].2010 18:2028-2037;Song等人,Nat Biotechnol.[自然生物技术]200523:709-717;Peer等人,Science[科学].2008 319:627-630;以及Peer和Lieberman,Gene Ther[基因疗法].2011 18:1127-1133。

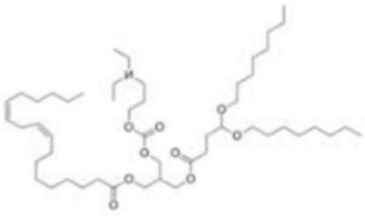
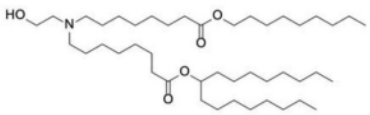
[0756] 在一些实施例中,通过将选择性器官靶向(Selective ORgan Targeting,SORT)分子添加至包含传统组分(例如可电离的阳离子脂质、两亲性磷脂、胆固醇和聚(乙二醇)(PEG))的配制品中来针对组织特异性活性对LNP进行选择。Cheng等人Nat Nanotechnol[自然纳米技术]15(4):313-320(2020)的传授内容证明,添加补充的“SORT”组分可根据SORT分子的百分比和生物物理特性精确地改变体内RNA递送谱并介导组织特异性(例如,肺、肝、脾脏)基因递送和编辑。

[0757] 在一些实施例中,LNP包含生物可降解的可电离脂质。在一些实施例中,LNP包含(9Z,12Z)-3-((4,4-双(辛氧基)丁酰基)氧基)-2-((((3-(二乙基氨基)丙氧基)羰基)氧基)甲基)丙基十八碳-9,12-二烯酸酯,也称为3-((4,4-双(辛氧基)丁酰基)氧基)-2-((((3-(二乙基氨基)丙氧基)羰基)氧基)甲基)丙基(9Z,12Z)-十八碳-9,12-二烯酸酯)或另一种可电离脂质。参见,例如WO 2019/067992、WO/2017/173054、WO 2015/095340和WO 2014/136086,以及其中提供的参考文献的脂质。在一些实施例中,在LNP脂质的上下文中术语阳离子和可电离是可互换的,例如,其中可电离脂质根据pH是阳离子的。

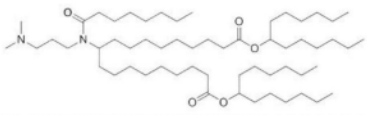
[0758] 在一些实施例中,本文所述的LNP包含表19中所述的脂质。

[0759] 表19示例性脂质

[0760]

脂质 ID	化学名称	分子量	结构
LIPIDV 003	(9Z,12Z)-3-((4,4-双(辛基氧基)丁酰基)氧基)-2-((((3-(二乙基氨基)丙氧基)羰基)氧基)甲基)丙基十八-9,12-二烯酸酯	852.29	
LIPIDV 004	十七烷-9-基 8-((2-羟基乙基)(8-(壬基氧基)-8-氧代辛基)氨基)辛酸酯	710.18	

[0761]

脂质 ID	化学名称	分子量	结构
LIPIDV 005		919.56	

[0762] 在一些实施例中,可以将基因修饰系统的多个组分制备为单一LNP配制品,例如,LNP配制品包含编码基因修饰多肽的mRNA和RNA模板。可以改变核酸组分的比率以便最大化治疗剂的特性。在一些实施例中,RNA模板与编码基因修饰多肽的mRNA的比率为按摩尔比计约1:1至100:1,例如约1:1至20:1、约20:1至40:1、约40:1至60:1、约60:1至80:1、或约80:1至100:1。在其他实施例中,可以由单独的配制品制备多种核酸的系统,例如,包含模板RNA的一种LNP配制品和包含编码基因修饰多肽的mRNA的第二LNP配制品。在一些实施例中,该系统可以包含配制到LNP中的多于两种核酸组分。在一些实施例中,该系统可以包含蛋白质(例如,基因修饰多肽)以及配制到至少一种LNP配制品中的模板RNA。

[0763] 在一些实施例中,LNP配制品的平均LNP直径可以在数十nm和数百nm之间,例如通过动态光散射(DLS)测量的。在一些实施例中,LNP配制品的平均LNP直径可以为约40nm至约150nm,如约40nm、45nm、50nm、55nm、60nm、65nm、70nm、75nm、80nm、85nm、90nm、95nm、100nm、105nm、110nm、115nm、120nm、125nm、130nm、135nm、140nm、145nm或150nm。在一些实施例中,LNP配制品的平均LNP直径可为约50nm至约100nm、约50nm至约90nm、约50nm至约80nm、约50nm至约70nm、约50nm至约60nm、约60nm至约100nm、约60nm至约90nm、约60nm至约80nm、约60nm至约70nm、约70nm至约100nm、约70nm至约90nm、约70nm至约80nm、约80nm至约100nm、约80nm至约90nm或约90nm至约100nm。在一些实施例中,LNP配制品的平均LNP直径可为约70nm至约100nm。在特定实施例中,LNP配制品的平均LNP直径可为约80nm。在一些实施例中,LNP配制品的平均LNP直径可为约100nm。在一些实施例中,LNP配制品的平均LNP直径范围为约1mm至约500mm、约5mm至约200mm、约10mm至约100mm、约20mm至约80mm、约25mm至约60mm、约30mm至约55mm、约35mm至约50mm,或约38mm至约42mm。

[0764] 在一些情况下,LNP可以是相对均质的。多分散性指数可用于指示LNP的均质性,例如脂质纳米颗粒的粒度分布。小的(例如,小于0.3)多分散性指数通常指示窄的粒度分布。LNP的多分散性指数可为约0至约0.25,如0.01、0.02、0.03、0.04、0.05、0.06、0.07、0.08、0.09、0.10、0.11、0.12、0.13、0.14、0.15、0.16、0.17、0.18、0.19、0.20、0.21、0.22、0.23、

0.24或0.25。在一些实施例中,LNP的多分散性指数可为约0.10至约0.20。

[0765] LNP的ζ电位可用于指示组合物的电动电位。在一些实施例中,ζ电位可以描述LNP的表面电荷。具有相对低电荷(正电荷或负电荷)的脂质纳米颗粒通常是期望的,因为更高电荷的物质可能不理想地与体内的细胞、组织和其他元素相互作用。在一些实施例中,LNP的ζ电位可为约-10mV至约+20mV、约-10mV至约+15mV、约-10mV至约+10mV、约-10mV至约+5mV、约-10mV至约0mV、约-10mV至约-5mV、约-5mV至约+20mV、约-5mV至约+15mV、约-5mV至约+10mV、约-5mV至约+5mV、约-5mV至约0mV、约0mV至约+20mV、约0mV至约+15mV、约0mV至约+10mV、约0mV至约+5mV、约+5mV至约+20mV、约+5mV至约+15mV或约+5mV至约+10mV。

[0766] 蛋白质和/或核酸(例如,基因修饰多肽或编码该多肽的mRNA)的包封效率描述了相对于所提供的初始量,在制备后被包封或以其他方式与LNP相关联的蛋白质和/或核酸的量。包封效率理想的是较高(例如,接近100%)。包封效率可以例如通过比较在用一种或多种有机溶剂或去垢剂破碎脂质纳米颗粒之前和之后含有脂质纳米颗粒的溶液中蛋白质或核酸的量来测量。阴离子交换树脂可用于测量溶液中游离蛋白质或核酸(例如RNA)的量。荧光可用于测量溶液中游离蛋白质和/或核酸(例如RNA)的量。对于本文所述的脂质纳米颗粒,蛋白质和/或核酸的包封效率可以是至少50%,例如50%、55%、60%、65%、70%、75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%。在一些实施例中,包封效率可以是至少80%。在一些实施例中,包封效率可以是至少90%。在一些实施例中,包封效率可以是至少95%。

[0767] LNP可以任选地包含一层或多层包衣。在一些实施例中,LNP可以配制在具有包衣的胶囊、膜或片剂中。包含本文所述的组合物的胶囊、膜或片剂可具有任何可用的尺寸、拉伸强度、硬度或密度。

[0768] 另外的示例性脂质、配制品、方法和LNP表征由WO 2020061457传授,其通过援引以其全文并入本文。

[0769] 在一些实施例中,使用Lipofectamine MessengerMax(赛默飞世尔公司(Thermo Fisher))或TransIT-mRNA转染试剂(米卢斯生物公司(Mirus Bio))进行体外或离体细胞脂质转染。在某些实施例中,使用GenVoy_ILM可电离脂质混合物(精密纳米系统(Precision NanoSystems))配制LNP。在某些实施例中,使用2,2-二亚油烯基-4-二甲基氨基乙基-[1,3]-二氧戊环(DLin-KC2-DMA)或二亚油烯基甲基-4-二甲基氨基丁酸酯(DLin-MC3-DMA或MC3)配制LNP,其配制和体内用途在Jayaraman等人Angew Chem Int Ed Engl[德国应用化学]51(34):8529-8533(2012)中传授,其通过援引以其全文并入本文。

[0770] 优化用于递送CRISPR-Cas系统(例如Cas9-gRNA RNP、gRNA、Cas9mRNA)的LNP配制品在两者均通过援引并入的WO 2019067992和WO 2019067910中描述。

[0771] 可用于递送核酸的另外的特定LNP配制品在两者均通过援引并入的US 8158601和US 8168775中描述,其包括帕替西兰(patisiran)中使用的以名称ONPATTRO销售的配制品。

[0772] 基因修饰LNP的示例性给药可包括约0.1、0.25、0.3、0.5、1、2、3、4、5、6、8、10或100mg/kg(RNA)。包含编码系统的一种或多种组分的核酸的AAV的示例性给药可包括约10^{11}、10^{12}、10^{13}和10^{14}vg/kg的MOI。

[0773] 试剂盒、制品和药物组合物

[0774] 在一方面,本披露提供了一种试剂盒,其包含基因修饰多肽或基因修饰系统,例

如,如本文所述。在一些实施例中,试剂盒包含基因修饰多肽(或编码多肽的核酸)和模板RNA(或编码模板RNA的DNA)。在一些实施例中,该试剂盒进一步包含用于将系统引入细胞的试剂,例如转染试剂、LNP等。在一些实施例中,该试剂盒适用于本文所述的任何方法。在一些实施例中,该试剂盒包含一种或多种元件、组合物(例如,药物组合物)、基因修饰多肽和/或基因修饰系统,或其功能片段或组分,它们例如布置在制品中。在一些实施例中,该试剂盒包含其使用说明书。

[0775] 在一方面,本披露提供了一种制品,例如,其中布置有本文所述的试剂盒或其组分。

[0776] 在一方面,本披露提供了一种药物组合物,其包含基因修饰多肽或基因修饰系统,例如,如本文所述。在一些实施例中,药物组合物进一步包含药学上可接受的载剂或赋形剂。在一些实施例中,药物组合物包含模板RNA和/或编码多肽的RNA。在实施例中,药物组合物具有以下特征中的一个或多个(例如,1、2、3或4个):

[0777] (a)相对于模板RNA和/或编码多肽的RNA少于1%(例如少于0.5%、0.4%、0.3%、0.2%或0.1%)的DNA模板,例如,以摩尔计;

[0778] (b)相对于模板RNA和/或编码多肽的RNA少于1%(例如少于0.5%、0.4%、0.3%、0.2%或0.1%)的未加帽RNA,例如,以摩尔计;

[0779] (c)相对于模板RNA和/或编码多肽的RNA少于1%(例如少于0.5%、0.4%、0.3%、0.2%或0.1%)的部分长度RNA,例如,以摩尔计;

[0780] (d)基本上缺乏未反应的帽二核苷酸。

[0781] 化学、制造和控制(CMC)

[0782] 在例如以下文献中描述了蛋白治疗剂的纯化:Franks,Protein Biotechnology:Isolation,Characterization,and Stabilization[蛋白生物技术:分离、表征、和稳定化],Humana Press[胡玛纳出版社](2013);以及Cutler,Protein Purification Protocols[蛋白纯化方案](Methods in Molecular Biology[分子生物学方法]),Humana Press[胡玛纳出版社](2010)。

[0783] 在一些实施例中,基因修饰系统、多肽和/或模板核酸(例如,模板RNA)符合某些质量标准。在一些实施例中,通过本文所述的方法产生的基因修饰系统、多肽和/或模板核酸(例如,模板RNA)符合某些质量标准。因此,在一些方面,本披露涉及制造符合某些质量标准的基因修饰系统、多肽和/或模板核酸(例如,模板RNA)的方法,例如,其中所述质量标准已测定。在一些方面,本披露还涉及在基因修饰系统、多肽和/或模板核酸(例如,模板RNA)中测定所述质量标准的方法。在一些实施例中,质量标准包括但不限于以下中的一项或多项(例如,1、2、3、4、5、6、7、8、9、10、11、或12项):

[0784] (i)模板RNA的长度,例如,模板RNA的长度是否大于参考长度或在参考长度范围内,例如是否存在的模板RNA中的至少80%、85%、90%、95%、96%、97%、98%或99%的长度大于100、125、150、175或200个核苷酸;

[0785] (ii)模板RNA上聚A尾的存在、不存在和/或长度,例如,是否存在的模板RNA中的至少80%、85%、90%、95%、96%、97%、98%或99%含有聚A尾(例如,长度为至少5、10、20、30、50、70、100个核苷酸的聚A尾);

[0786] (iii)模板RNA上5'帽的存在、不存在和/或类型,例如,是否存在的模板RNA中的至

少80％、85％、90％、95％、96％、97％、98％或99％含有5’帽,例如,该帽是否是7-甲基鸟苷帽,例如O-Me-m7G帽;

[0787] (iv)该模板RNA中一个或多个经修饰核苷酸(例如,选自假尿苷、二氢尿苷、肌苷、7-甲基鸟苷、1-N-甲基假尿苷(1-Me-Ψ)、5-甲氧基尿苷(5-MO-U)、5-甲基胞苷(5mC)或锁核苷酸)的存在、不存在和/或类型,例如,是否存在的模板RNA中的至少80％、85％、90％、95％、96％、97％、98％或99％含有一个或多个经修饰核苷酸;

[0788] (v)模板RNA的稳定性(例如,随着时间的推移和/或在预先选择的条件下),例如是否至少80％、85％、90％、95％、96％、97％、98％或99％的模板RNA在稳定性测试后保持完整(例如,长度大于100、125、150、175或200个核苷酸);

[0789] (vi)模板RNA在用于修饰DNA的系统中的效力,例如,在测定包含该模板RNA的系统的效力之后,是否至少1％的靶位点被修饰;

[0790] (vii)多肽、第一多肽或第二多肽的长度,例如,该多肽、第一多肽或第二多肽的长度是否超出参考长度或在参考长度范围内,例如是否存在的至少80％、85％、90％、95％、96％、97％、98％、或99％的多肽、第一多肽或第二多肽的长度大于600、650、700、750、800、850、900、950、1000、1050、1100、1150、1200、1250、1300、1350、1400、1450、1500、1600、1700、1800、1900、或2000个氨基酸(并且任选地,长度不超过2500、2000、1500、1400、1300、1200、1100、1000、900、800、700、或600个氨基酸);

[0791] (viii)多肽、第一多肽或第二多肽上翻译后修饰的存在、不存在和/或类型,例如是否至少80％、85％、90％、95％、96％、97％、98％、或99％的多肽、第一多肽或第二多肽含有磷酸化、甲基化、乙酰化、肉豆蔻酰化、棕榈酰化、异戊二烯化、glipyatyon或脂酰化,或其任何组合;

[0792] (ix)多肽、第一多肽或第二多肽中一种或多种人工、合成或非典型氨基酸(例如,选自鸟氨酸、β-丙氨酸、GABA、δ-氨基乙酰丙酸、PABA、D-氨基酸(例如,D-丙氨酸或D-谷氨酸)、氨基异丁酸、脱氢丙氨酸、胱硫醚、羊毛硫氨酸、甲烯胱氨酸、二氨基庚二酸、高丙氨酸、正缬氨酸、正亮氨酸、高正亮氨酸(Homonorleucine)、高丝氨酸、O-甲基-高丝氨酸和O-乙基-高丝氨酸、乙硫氨酸、硒代半胱氨酸、硒代高半胱氨酸、硒代甲硫氨酸、硒代乙硫氨酸、碲代半胱氨酸或碲代甲硫氨酸)的存在、不存在和/或类型,例如是否存在的至少80％、85％、90％、95％、96％、97％、98％或99％的多肽、第一多肽或第二多肽含有一个或多个人工、合成或非典型氨基酸;

[0793] (x)多肽、第一多肽或第二多肽的稳定性(例如,随着时间的推移和/或在预选条件下),例如是否至少80％、85％、90％、95％、96％、97％、98％或99％的多肽、第一多肽或第二多肽在稳定性测试后保持完整(例如,长度大于600、650、700、750、800、850、900、950、1000、1050、1100、1150、1200、1250、1300、1350、1400、1450、1500、1600、1700、1800、1900、或2000个氨基酸(并且任选地,长度不超过2500、2000、1500、1400、1300、1200、1100、1000、900、800、700、或600个氨基酸));

[0794] (xi)该多肽、第一多肽或第二多肽在用于修饰DNA的系统中的效力,例如在测定了包含该多肽、第一多肽或第二多肽的系统的效力之后是否至少1％的靶位点被修饰;或者

[0795] (xii)热原、病毒、真菌、细菌病原体或宿主细胞蛋白中的一种或多种的存在、不存在、和/或水平,例如,系统是否不含或基本上不含热原、病毒、真菌、细菌病原体或宿主细胞

蛋白污染。

[0796] 在一些实施例中,本文所述的系统或药物组合物不含内毒素。

[0797] 在一些实施例中,对热原、病毒、真菌、细菌病原体和/或宿主细胞蛋白中的一种或多种的存在、不存在、和/或水平进行确定。在实施例中,对系统是否不含或基本上不含热原、病毒、真菌、细菌病原体和/或宿主细胞蛋白污染进行确定。

[0798] 在一些实施例中,如本文所述的药物组合物或系统具有以下特征中的一项或多项(例如,1、2、3或4项):

[0799] (a)相对于模板RNA和/或编码多肽的RNA少于1%(例如少于0.5%、0.4%、0.3%、0.2%或0.1%)的DNA模板,例如,以摩尔计;

[0800] (b)相对于模板RNA和/或编码多肽的RNA少于1%(例如少于0.5%、0.4%、0.3%、0.2%或0.1%)的未加帽RNA,例如,以摩尔计;

[0801] (c)相对于模板RNA和/或编码多肽的RNA少于1%(例如少于0.5%、0.4%、0.3%、0.2%或0.1%)的部分长度RNA,例如,以摩尔计;

[0802] (d)基本上缺乏未反应的帽二核苷酸。

[0803] 实例

[0804] 本发明进一步通过以下实例说明。提供这些实例仅出于说明目的,而不应以任何方式解释为限制本发明的范围或内容。

[0805] 实例1:包含具有各种Pol结构域的Cas-Pol融合体的基因修饰系统实现人细胞中的精确编辑

[0806] 该实例描述了Cas-Pol融合体在人细胞基因组DNA中以可编程方式安装突变的能力。更特别地,改变Cas-Pol融合体的聚合酶结构域,例如本申请中描述的聚合酶结构域,例如,人DNA Polθ,以确定采用新颖聚合酶组合的Cas-Pol融合体的基因组编辑能力。模板核酸与人细胞中的Cas-Pol表达质粒共同递送,以确定Cas-Pol融合体的靶编辑活性。

[0807] 为了生成基因组工程多肽的结构域文库,选择了Cas效应子蛋白;参见表3和表4。进一步选择另外的Cas9结构域用于本文所述的基因组工程多肽中,特征(包括靶序列的PAM要求、赋予切口酶活性的预测突变(例如,SpCas9的D10A、H840A或N863A)和gRNA特征(包括单指导物组成,例如,特定的间隔子参数)和gRNA支架序列(以赋予对于同源Cas酶而言的多肽结合))都能够被确定(表8)。基于对长度、柔性和组成的多样性的搜索,收集连接Cas和Pol结构域的接头序列,以优化融合蛋白(表6)。如本申请中所述,使用文献和聚合酶蛋白结构域特征从各种来源挖掘聚合酶结构域,包括能够具有RNA和/或DNA依赖性DNA聚合活性的野生型聚合酶、具有改善的特性(例如,热稳定性、持续合成能力、保真度)的衍生聚合酶、具有失活或重新激活的功能结构域的衍生聚合酶(例如,赋予5′-3′或3′-5′外切核酸酶活性、校对活性、解旋酶活性或RNA酶活性的失活或重新激活的结构域)以及具有合成地进化的RNA依赖性和/或DNA依赖性DNA聚合酶活性的聚合酶(例如,衍生自Thermococcus kodakarensis(KOD)的PolB的RTX聚合酶,如Ellefson等人Science[科学]352(6293):1590-1593(2016)中所述,其通过援引以其全文并入本文)。

[0808] 特别地,为了评估在基因修饰多肽情况下使用新颖Pol结构域以成功编辑基因组,选择示例性Pol结构域的子集与Cas9(N863A)切口酶融合。简而言之,使用UniProt(The UniProt Consortium Nucleic Acids Res[UniProt联合核酸研究]49(D1):D480-D489

(2021))确定单体人聚合酶的蛋白质序列,并使用InterPro(Blum等人Nucleic Acids Res [核酸研究]49(D1):D344-D354(2021))和InterProScan(Jones等人Bioinformatics[生物信息学]30(9):1236-1240(2014))进一步预测和注释功能结构域(其中的表X)。尽管不希望受这样的实例的限制,但来自聚合酶家族Y、X和A的蛋白质已被描述为包含单个亚基(参见例如Hoitsma等人Cell Mol Life Sci[细胞与分子生命科学]77(1):35-59(2020),其通过援引以其全文并入本文)。

[0809] 为了使用基因组工程系统Cas-Pol融合体产生精确的编辑,模板核酸被构建为通过Pol结构域将编辑进行模板聚合到基因组靶位点。模板核酸被设计为包含(i)用于将Cas-Pol引导至靶区域的gRNA间隔子序列,例如,与HEK3基因座中的20个核苷酸序列互补的序列;(ii)引物结合序列(PBS),其能够与带切口的DNA的单链互补碱基配对以进行靶引发的聚合;(iii)提供聚合模板的异源对象序列,其进一步包含预期的最终靶序列;和(iv)与Cas9-Pol多肽融合体的Cas9结构域相关联的gRNA支架序列。这里使用的构建体特别遵循5'到3'取向(i)、(iv)、(iii)、(ii)。在一些实施例中,(iii)可以包含RNA和/或DNA核苷酸。在一些实施例中,(ii)可以包含RNA和/或DNA核苷酸。不希望受实例限制,(i)和(iv)在这些实验中包含RNA核苷酸。模板组成描述于表21(模板P1、P2、P3)中,其中(ii)和(iii)可各自作为RNA或DNA核苷酸包括在内。模板分子任选地进一步包含5′帽和3′聚A尾。

[0810] 表20:实例1中使用的模板核酸和第二切口gRNA

[0811]

名称	描述	间隔子 (i)	支架 (iv)	Pol模板 (iii)	PBS (ii)	全模板分子
模板P1	HEK3_8PBS_10RT(CTTat1)	GGCCCAGACTGAGCACGTGA	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC	TCTGCCATCAAAG	CGTGCTCA	GGCCCAGACTGAGCACGTGAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTCTGCCATCAAAGCGTGCTCA
模板P2	HEK3_13PBS_10RT(CTTat1)	GGCCCAGACTGAGCACGTGA	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTC	TCTGCCATCAAAG	CGTGCTCAGTCTG	GGCCCAGACTGAGCACGTGAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTCTGCCATCAAAGCGTGCTCAGTCTG

[0812]

名称	描述	间隔子 (i)	支架 (iv)	Pol 模板 (iii)	PBS (ii)	全模板分子
			GGTGC			
模板P3	HEK3_17PBS_10RT（CTTat1）	GGCCCAGACTGAGCACGTGA	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC	TCTGCCATCAAAG	CGTGCTCAGTCTGGGCC	GGCCCAGACTGAGCACGTGAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTCTGCCATCAAAGCGTGCTCAGTCTGGGCC
模板P4	HBB_13PBS_10RT（TtoAat4）	GCATGGTGCACCTGACTCCTG	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC	AGACTTCTCCACAG	GAGTCAGGTGCAC	GCATGGTGCACCTGACTCCTGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCAGACTTCTCCACAGGAGTCAGGTGCAC
2gRNA P5	HEK3_+90	GTCAACCAGTATCCCGGTGC	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC	NA	NA	NA
2gRNA P6	HBB_+72	GCCTTGATACCAACCTGCCCA	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC	NA	NA	NA

[0813] 通过250,000个细胞/孔的电穿孔用约800ng的Cas9-Pol融合(例如,Polθ融合)表达质粒、200ng的化学合成的模板核酸分子和任选地83ng的另外的第二切口gRNA(2gRNA P5(对于模板P1、P2、P3)或2gRNA P6(对于模板P4))转染U2OS或HEK293T细胞(表21)。为了评估Cas-Pol融合体的基因组编辑能力,在转染后第3天收集基因组DNA(gDNA)。通过扩增子测序分析靶基因座(HEK3(对于模板P1、P2、P3)或HBB(对于模板P4))处的预期(如设计的精确和无痕编辑)和非预期(对靶序列的任何非预期改变)编辑的频率。如本文所用,靶位点的扩增子测序包含在靶位点的PCR扩增中使用位点特异性引物,在Illumina MiSeq上对扩增子进行测序,以及使用CRISPResso2管道检测和表征编辑事件(Clement等人Nat Biotechnol[自然生物技术]37(3):224-226(2019))。在一些实施例中,活性Cas-Pol融合体导致可检测水平的编辑,例如,至少0.1%的测序读段表现出靶位点编辑。在一些实施例中,期望的Cas-Pol融合体表现出预期编辑比非预期编辑的频率更高,例如,预期编辑比非预期编辑的频率高至少2倍。

[0814] 实例2:通过接头选择改善Cas-Pol融合体的表达

[0815] 该实例展示了Cas-Pol融合体的优化,以改善哺乳动物细胞中的蛋白质表达。通过取代如上实例1中所述的新功能结构域来构建新颖Cas-Pol融合体可导致基因组工程多肽的低或中度表达。因此,这里预期融合体的经修饰的构型在不同结构域的上下文中可能是有利的。不希望受限于该实例,一种用于改善新融合体的表达和稳定性的方法是通过使用接头文库。在这里,使用接头序列文库,Cas-Pol融合体的Cas和Pol结构域之间的肽接头序列是可变的。更特别地,下表21中的接头用于生成Cas9融合构建体的新变体,并递送至人细胞以筛选改善的Cas-Pol蛋白质表达。

[0816] 一组22个具有不同程度的长度、柔性、疏水性和二级结构的肽接头(表21)首先用于通过取代原始接头来产生Cas-Pol融合蛋白的变体(参见上面提到的实例30)。将HEK293T细胞通过250,000个细胞/孔的电穿孔使用约800ng的每种Cas9-Pol融合体质粒以及200ng

单指导RNA质粒进行转染。为了评估Cas9-Pol融合体的表达水平,在转染后第2天收集细胞裂解物,并使用针对Cas9的一抗通过蛋白质印迹进行分析。

[0817] 表21.用作包含Cas-Pol融合体的基因组工程多肽中Cas和Pol结构域之间的接头的肽序列

[0818]

#	接头序列	注意:
1	GGS	短
2	GGGGS	柔性，短
3	(GGGGS)2	柔性
4	(GGGGS)3	柔性，长
5	(GGGGS)4	柔性，很长
6	(G)6	柔性
7	(G)8	柔性
8	GSAGSAAGSGEF	柔性
9	(GSSGSS)	中等
10	(GSSGSS)2	中等，柔性
11	(GSSGSS)3	中等
12	SGSETPGTSESATPES	XTEN
13	(EAAAK)	刚性螺旋，短
14	(EAAAK)2	刚性螺旋，中等

[0819]

15	(EAAAK)3	刚性螺旋，长
16	PAP	刚性，短
17	PAPAP	刚性，短
18	PAPAPAPAP	刚性，中等
19	A(EAAAK)4ALEA(EAAAK)4A	刚性的，很长，具有螺旋
20	GGGGS(EAAAK)GGGGS	柔性-螺旋-柔性
21	(EAAAK)GGGGS(EAAAK)	螺旋-柔性-螺旋
22	SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	柔性-XTEN-柔性

[0820] 实例3:通过HEK293T和U2OS细胞中的池化筛选进行基因修饰多肽选择

[0821] 该实例描述了使用RNA基因修饰系统来对人基因组中的编码序列进行靶向编辑。更特别地,本实例描述了用基因修饰候选物文库感染HEK293T和U2OS细胞,随后转染模板指

导RNA(tgRNA)以在细胞中进行体外基因修饰,例如,作为通过池化筛选方法评价新的基因修饰多肽在人细胞中的编辑活性的手段。

[0822] 本文测定的基因修饰多肽文库候选物各自包含:1)含有N863A突变的化脓性链球菌(Spy)Cas9切口酶,其灭活一个核酸内切酶活性位点;2)表6所描绘的122个肽接头之一;和3)来自表2中逆转录病毒来源的逆转录酶(RT)结构域。如果预期所利用的特定的逆转录病毒RT结构域作为单体发挥作用,则选择这些结构域。对于每个选定的RT结构域,测试了野生型序列,以及在主要野生型序列中安装了点突变的版本。特别地,测试了143个RT结构域,为野生型或包含各种突变。总共测试了17,446个Cas-接头-RT基因修饰多肽。

[0823] 此处描述的系统是双组分系统,其包含:1)在慢病毒盒内编码人密码子优化的基因修饰多肽文库候选物的表达质粒,2)表达非编码tgRNA序列的tgRNA表达质粒,该序列被Cas识别并将其定位于目的基因组基因座,并且还通过由U6启动子驱动的RT结构域作为将所期望编辑逆转录到基因组中的模板。慢病毒盒包含:(i)用于在哺乳动物细胞中表达的CMV启动子;(ii)如所示的基因修饰多肽文库候选物;(iii)自切割T2A多肽;(iv)能够在哺乳动物细胞中进行选择的嘌呤霉素抗性基因;以及(v)聚A尾终止信号。

[0824] 为了制备表达基因修饰多肽文库候选物的细胞池,用基因修饰候选质粒文库的池化慢病毒制剂转导HEK293T或U2OS细胞。将HEK293 Lenti-X细胞接种于15cm平板中($12x10^6$个细胞)后再进行慢病毒质粒转染。使用慢病毒包装混合物(博塞塔公司,27ug)进行慢病毒质粒转染,并根据制造商的方案在第二天使用Lipofectamine 2000和Opti-MEM培养基对基因修饰候选文库的质粒DNA(27ug)进行转染。通过第二天的完全培养基更换去除细胞外DNA,并且在48小时后收获含有病毒的培养基。慢病毒培养基使用Lenti-X浓缩液(宝生物科学公司)浓缩,制备5mL慢病毒等分试样并储存于-80℃。慢病毒滴度测定是通过在嘌呤霉素选择后计数菌落形成单位来进行的。将携带表达BFP的基因组着陆垫的HEK293T或U2OS细胞以$6x\ 10^7$个细胞接种在培养板中,并以0.3的感染复数(MOI)进行转导,以最大程度地减少每个细胞的多重感染。感染后48小时加入嘌呤霉素(2.5ug/mL)以选择受感染的细胞。将细胞在嘌呤霉素选择下保持至少7天,然后扩大以进行tgRNA电穿孔。

[0825] 为了确定测定中基因修饰文库候选物的基因组编辑能力,然后将感染的表达BFP的HEK293T或U2OS细胞通过电穿孔以250,000个细胞/孔用200ng tgRNA(g4或g10)质粒进行转染,该质粒设计用于将BFP转化为GFP,其中细胞计数足够实现每个文库候选物>1000x的覆盖率。

[0826] g4 tgRNA(5′至3′)如下:20个核苷酸间隔子区(GCCGAAGCACTGCACGCCGT)、支架区(GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC)、模板区(其编码将BFP变为GFP的单碱基对取代(粗体)和在SpyCas9 PAM中引入同义点突变(NGG到NCG)以防止功能性基因修饰反应完成后基因修饰多肽的重新接合的PAM失活(下划线))(ACCCTGACGTACG)和13个核苷酸PBS(GCGTGCAGTGCTT)。

[0827] 类似地,g10 tgRNA(5′至3′)如下:20个核苷酸间隔子区(AGAAGTCGTGCTGCTTCATG)、支架区(GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC)、模板区(其编码将BFP变为GFP的单碱基对取代(粗体)和在SpyCas9 PAM中引入同义点突变(NGG到NGA)以防止功能性基因修饰反应完成后基因修

饰多肽的重新接合的PAM失活(下划线))

(ACCCTGACCTACGGCGTGCAGTGCTTCGGCCGCTACCCCGATCACAT)和13个核苷酸PBS(GAAGCAGCACGAC)。

[0828] 为了评估测定中各种构建体的基因组编辑能力,在电穿孔后6-7天通过荧光激活细胞分选(FACS)对细胞进行分选,以检测GFP表达。对细胞进行分选和收获,分为未编辑(BFP+)细胞、已编辑(GFP+)细胞和不完美编辑(BFP-、GFP-)细胞的不同群体。还收获未分选的细胞样品作为输入群体,以确定分析过程中的富集。

[0829] 为了确定哪些基因修饰文库候选物在测定中具有基因组编辑能力,从分选的和未分选的细胞群体中收获基因组DNA(gDNA),并通过对每个群体中的基因修饰文库候选物进行测序进行分析。简言之,使用针对慢病毒盒的特异性引物从基因组中扩增基因修饰序列,在第二轮PCR中进行扩增以稀释基因组DNA,然后根据制造商的方案使用牛津纳米孔测序技术(Oxford Nanopore Sequencing Technology)进行测序。

[0830] 在对测序读段进行质量控制之后,将至少1500个并且不超过3200个核苷酸的读段映射到基因修饰多肽文库序列,并且那些与文库序列至少80%匹配的读段被认为已成功与给定候选物比对。为了识别能够在测定中进行基因编辑的基因修饰候选物,将编辑群体中每个文库候选物的读段计数与初始未分选群体中的读段计数进行比较。为了该池化筛选的目的,具有基因组编辑能力的基因修饰候选物被选择为相对于未分选的(输入)细胞在经转化的(GFP+)群体中富集的那些候选物,并且其中该富集被确定为等于或高于参考(元件ID号:17380)的富集水平。

[0831] 大量基因修饰多肽候选物被确定在GFP+细胞群体中富集。例如,在测试的17,446个候选物中,超过3,300个在GFP+分选群体中表现出富集(相对于未分选的群体),该富集至少与类似实验条件下的参考值相当(使用g4tgRNA的HEK293T;使用g10 tgRNA的HEK293T细胞;或使用g4 tgRNA的U2OS细胞),在表D中示出。尽管也使用g10 tgRNA在U2OS细胞中测试了17,446个候选物,但在该实验条件下,池化筛选并未产生相对于未分选(输入)细胞在经转化的(GFP+)群体中富集的候选物;需要进一步研究来解释这些结果。

[0832] 表D.筛选的接头和RT序列的组合。该表中每个RT的氨基酸序列提供于表6。

[0833]

接头氨基酸序列	RT 结构域名称
EAAAKGSS	PERV_Q4VFZ2_3mutA_WS
EAAAKEAAAKEAAAKEAAAK	MLVMS_P03355_PLV919
PAPEAAAK	MLVFF_P26809_3mutA
EAAAKPAPGGG	MLVFF_P26809_3mutA
GSSGSSGSSGSSGSSGSS	PERV_Q4VFZ2_3mut
PAPGGGEAAAK	MLVAV_P03356_3mutA
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	MLVMS_P03355_PLV919
GSSEAAAK	MLVFF_P26809_3mutA
EAAAKPAPGGS	MLVFF_P26809_3mutA
GGSGGSGGSGGSGGSGGS	MLVFF_P26809_3mutA
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	XMRV6_A1Z651_3mutA
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	PERV_Q4VFZ2_3mutA_WS
EAAAKEAAAKEAAAK	MLVFF_P26809_3mutA
PAPEAAAKGSS	MLVFF_P26809_3mutA
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	PERV_Q4VFZ2_3mutA_WS
EAAAKEAAAKEAAAK	PERV_Q4VFZ2_3mutA_WS
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	AVIRE_P03360_3mutA
PAPAPAPAPAP	MLVCB_P08361_3mutA
PAPAPAPAPAP	MLVFF_P26809_3mutA
EAAAKGGSPAP	PERV_Q4VFZ2_3mutA_WS

[0834]

接头氨基酸序列	RT 结构域名称
PAP	MLVMS_P03355_PLV919
PAPGGGGSS	WMSV_P03359_3mutA
SGSETPGTSESATPES	MLVFF_P26809_3mutA
PAPEAAAKGSS	XMRV6_A1Z651_3mutA
EAAAKGGSGGG	MLVMS_P03355_PLV919
GGGGSGGGGS	MLVFF_P26809_3mutA
GGGPAPGSS	MLVAV_P03356_3mutA
GGSGGSGGSGGSGGSGGS	XMRV6_A1Z651_3mut
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	MLVCB_P08361_3mutA
GSSPAP	AVIRE_P03360_3mutA
EAAAKGSSPAP	MLVFF_P26809_3mutA
GSSGGGEAAAK	MLVFF_P26809_3mutA
GGSGGSGGSGGSGGSGGS	MLVMS_P03355_3mutA_WS
PAPAPAPAP	MLVFF_P26809_3mutA
EAAAKEAAAKEAAAKEAAAK	XMRV6_A1Z651_3mutA
EAAAKGGSPAP	MLVMS_P03355_3mutA_WS
PAPGGSEAAAK	AVIRE_P03360_3mutA
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	AVIRE_P03360_3mutA
EAAAKGGGGSEAAAK	MLVCB_P08361_3mutA
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	WMSV_P03359_3mutA
GSS	MLVMS_P03355_PLV919
GSSGSSGSSGSS	MLVMS_P03355_PLV919
GSSPAPEAAAK	XMRV6_A1Z651_3mutA
GGSPAPEAAAK	MLVFF_P26809_3mutA
GGGEAAAKGGS	MLVFF_P26809_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAK	PERV_Q4VFZ2_3mutA_WS
GGGGGGGG	PERV_Q4VFZ2_3mut
GGGPAP	MLVCB_P08361_3mutA

[0835]

接头氨基酸序列	RT 结构域名称
PAPAPAPAPAPAP	MLVCB_P08361_3mutA
GGSGGSGGSGGSGGSGGS	MLVCB_P08361_3mutA
PAP	MLVMS_P03355_3mutA_WS
GGSGGSGGSGGSGGSGGS	PERV_Q4VFZ2_3mutA_WS
PAPAPAPAPAPAP	MLVMS_P03355_PLV919
EAAAKPAPGSS	MLVMS_P03355_3mutA_WS
EAAAKEAAAKEAAAKEAAAK	MLVMS_P03355_3mutA_WS
EAAAKGGS	MLVMS_P03355_3mutA_WS
GGGGSEAAAKGGGGS	MLVFF_P26809_3mutA
EAAAKPAPGSS	MLVFF_P26809_3mutA
GGGGSGGGGSGGGGSGGGGS	MLVMS_P03355_PLV919
EAAAKGGGGGS	MLVMS_P03355_PLV919
GGSPAP	XMRV6_A1Z651_3mutA
EAAAKGGGPAP	MLVMS_P03355_PLV919
EAAAKEAAAKEAAAKEAAAKEAAAK	MLVFF_P26809_3mutA
PAP	MLVCB_P08361_3mutA
EAAAK	XMRV6_A1Z651_3mutA
GGSGSSPAP	PERV_Q4VFZ2_3mutA_WS
GSSGSSGSSGSSGSSGSS	MLVMS_P03355_PLV919
GSSEAAAKGGG	MLVAV_P03356_3mutA
GGGEAAAKGGS	XMRV6_A1Z651_3mutA
EAAAKGGGGSEAAAK	MLVAV_P03356_3mutA
GGGGSGGGGSGGGGS	MLVFF_P26809_3mutA
GGGGSGGGGSGGGGSGGGGS	AVIRE_P03360_3mutA
SGSETPGTSESATPES	AVIRE_P03360_3mutA
GGGEAAAKPAP	MLVFF_P26809_3mutA
EAAAKGSSGGG	MLVMS_P03355_3mutA_WS
EAAAKEAAAKEAAAKEAAAKEAAAK	WMSV_P03359_3mut
GGSGGSGGSGGS	XMRV6_A1Z651_3mutA

[0836]

接头氨基酸序列	RT 结构域名称
GGSEAAAKPAP	MLVFF_P26809_3mutA
EAAAKGSSGGG	XMRV6_A1Z651_3mutA
GGGGS	MLVFF_P26809_3mutA
GGGEAAAKGSS	MLVMS_P03355_PLV919
PAPAPAPAPAPAP	MLVAV_P03356_3mutA
GGGGSGGGGSGGGGSGGGGS	MLVCB_P08361_3mutA
GGGEAAAKGSS	MLVCB_P08361_3mutA
PAPGGSGSS	MLVFF_P26809_3mutA
GSAGSAAGSGEF	MLVCB_P08361_3mutA
PAPGGSEAAAK	MLVMS_P03355_3mutA_WS
GGSGSS	XMRV6_A1Z651_3mutA
PAPGGGGSS	MLVMS_P03355_PLV919
GSSGSSGSS	XMRV6_A1Z651_3mut
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	MLVMS_P03355_3mutA_WS
EAAAK	MLVMS_P03355_PLV919
GSSGSSGSSGSS	MLVFF_P26809_3mutA
PAPGGGGSS	MLVCB_P08361_3mutA
GGGEAAAKGGS	MLVCB_P08361_3mutA
PAPGGGEAAAK	MLVMS_P03355_PLV919
GGGGGSPAP	XMRV6_A1Z651_3mutA
EAAAKGGS	XMRV6_A1Z651_3mutA
EAAAKGSSPAP	XMRV6_A1Z651_3mut
PAPEAAAK	MLVAV_P03356_3mutA
GGSGGSGGSGGS	MLVMS_P03355_3mutA_WS
GGGPAPGGS	MLVMS_P03355_PLV919
GSSGSSGSSGSS	PERV_Q4VFZ2_3mutA_WS
EAAAKPAPGGS	MLVCB_P08361_3mutA
GSSGSS	MLVFF_P26809_3mutA

[0837]

接头氨基酸序列	RT 结构域名称
EAAAKEAAAKEAAAKEAAAK	MLVCB_P08361_3mutA
EAAAKEAAAKEAAAKEAAAK	FLV_P10273_3mutA
GSS	MLVFF_P26809_3mutA
EAAAKEAAAK	MLVMS_P03355_3mutA_WS
PAPEAAAKGGG	MLVAV_P03356_3mutA
GGSGSSEAAAK	MLVFF_P26809_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAK	PERV_Q4VFZ2
GSSEAAAKPAP	AVIRE_P03360_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAK	MLVCB_P08361_3mutA
EAAAKGGG	MLVFF_P26809_3mutA
GSSPAPGGG	MLVCB_P08361_3mutA
GGGPAPGSS	MLVMS_P03355_PLV919
GGGGGS	MLVMS_P03355_3mutA_WS
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	PERV_Q4VFZ2_3mut
GGGGSGGGGSGGGGSGGGGSGGGGS	WMSV_P03359_3mutA
EAAAKEAAAKEAAAK	PERV_Q4VFZ2_3mut
PAPAPAPAP	MLVCB_P08361_3mutA
GSSGSSGSSGSSGSS	PERV_Q4VFZ2_3mut
GGGGSSEAAAK	MLVMS_P03355_3mutA_WS
GGSGGSGGSGGS	MLVCB_P08361_3mutA
PAPEAAAKGGS	MLVCB_P08361_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	MLVCB_P08361_3mutA
EAAAKGGGGSEAAAK	MLVMS_P03355_PLV919
EAAAKGGGGSEAAAK	MLVMS_P03355_3mutA_WS
EAAAKGGGPAP	XMRV6_A1Z651_3mut
EAAAKEAAAKEAAAKEAAAKEAAAK	MLVMS_P03355_3mutA_WS
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	FLV_P10273_3mutA
GGSEAAAKGGG	MLVMS_P03355_3mutA_WS

[0838]

接头氨基酸序列	RT结构域名称
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	KORV_Q9TTC1-Pro_3mutA
GGGPAPGGS	MLVCB_P08361_3mutA
PAPAPAPAPAPAP	XMRV6_A1Z651_3mutA
GGSGSSGGG	XMRV6_A1Z651_3mutA
GGSGSSGGG	MLVCB_P08361_3mutA
GGGEAAAKGGS	MLVMS_P03355_3mutA_WS
EAAAK	MLVCB_P08361_3mutA
GGSPAPGSS	MLVMS_P03355_3mutA_WS
GGGGSSEAAAK	PERV_Q4VFZ2_3mut
PAPAPAPAPAP	MLVBM_Q7SVK7_3mut
EAAAKEAAAKEAAAKEAAAK	MLVAV_P03356_3mutA
GGGGGSGSS	MLVCB_P08361_3mutA
EAAAKGSSPAP	MLVMS_P03355_3mutA_WS
PAPAPAPAPAPAP	MLVMS_P03355_3mutA_WS
GSSGGGGGS	MLVMS_P03355_3mutA_WS
PAPGSSGGG	MLVMS_P03355_PLV919
GGSGGGPAP	MLVCB_P08361_3mutA
GGGGGGG	MLVCB_P08361_3mutA
GSSGSSGSSGSSGSSGSS	MLVCB_P08361_3mutA
GGGPAPGGS	MLVFF_P26809_3mutA
EAAAKGGSGGG	PERV_Q4VFZ2_3mut
EAAAKGGGGSS	MLVMS_P03355_3mutA_WS
GSSGSSGSSGSSGSSGSS	MLVMS_P03355_3mut
GGGGSGGGGSGGGGSGGGGS	MLVBM_Q7SVK7_3mutA_WS
PAPAPAPAPAP	MLVMS_P03355_PLV919
GGGEAAAKGGS	MLVMS_P03355_PLV919
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	MLVMS_P03355_3mut

[0839]

接头氨基酸序列	RT结构域名称
GSAGSAAGSGEF	MLVMS_P03355_3mutA_WS
GSSGSSGSSGSSGSS	MLVFF_P26809_3mutA
EAAAKGGSGSS	MLVFF_P26809_3mutA
PAPGGG	MLVFF_P26809_3mutA
GGGPAPGSS	XMRV6_A1Z651_3mutA
PAPEAAAKGGS	AVIRE_P03360_3mutA
PAPGGGEAAAK	MLVFF_P26809_3mut
GGGGSSEAAAK	MLVCB_P08361_3mutA
EAAAK	MLVMS_P03355_PLV919
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	BAEVM_P10272_3mutA
GGSGGGEAAAK	MLVMS_P03355_PLV919
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	MLVFF_P26809_3mutA
GSSPAPGGS	XMRV6_A1Z651_3mutA
GGSGGGPAP	MLVMS_P03355_PLV919
EAAAK	AVIRE_P03360_3mutA
GSS	XMRV6_A1Z651_3mutA
GGSGGSGGS	MLVFF_P26809_3mutA
EAAAKEAAAKEAAAKEAAAK	AVIRE_P03360_3mut
PAPEAAAKGGG	PERV_Q4VFZ2_3mutA_WS
GGGGGSEAAAK	BAEVM_P10272_3mutA
GGSGSSGGG	MLVMS_P03355_3mutA_WS
GGGGGGG	MLVMS_P03355_3mutA_WS
GSSEAAAKPAP	PERV_Q4VFZ2_3mut
GGGGGSEAAAK	WMSV_P03359_3mut
GGGGSGGGGSGGGGSGGGGSGGGGS	MLVFF_P26809_3mut
GGGEAAAKGGS	AVIRE_P03360_3mutA
GGSPAPGGG	AVIRE_P03360_3mutA
GSAGSAAGSGEF	MLVAV_P03356_3mutA

[0840]

接头氨基酸序列	RT结构域名称
EAAAK	MLVAV_P03356_3mutA
EAAAKPAPGSS	WMSV_P03359_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	PERV_Q4VFZ2_3mutA_WS
GGSEAAAKPAP	MLVCB_P08361_3mutA
PAPAPAPAPAPAP	MLVBM_Q7SVK7_3mutA_WS
GGSPAPGGG	MLVMS_P03355_3mutA_WS
GGSEAAAKGGG	MLVMS_P03355_3mut
GGSGGSGGSGGS	MLVFF_P26809_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	MLVFF_P26809_3mutA
GGG	AVIRE_P03360_3mutA
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	PERV_Q4VFZ2_3mut
GGSGGSGGSGGS	MLVMS_P03355_3mutA_WS
GGGEAAAK	MLVCB_P08361_3mutA
GSSGSSGSSGSSGSSGSS	MLVMS_P03355_3mutA_WS
GSSGGGPAP	MLVMS_P03355_3mutA_WS
GSSEAAAKPAP	MLVFF_P26809_3mutA
EAAAKEAAAK	MLVMS_P03355_PLV919
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	MLVCB_P08361_3mut
GGGGGG	MLVMS_P03355_3mutA_WS
GGSGSSGGG	MLVFF_P26809_3mutA
GSSGGGEAAAK	PERV_Q4VFZ2_3mutA_WS
PAPAPAPAPAP	PERV_Q4VFZ2_3mut
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	SFV3L_P27401_2mut
EAAAKGGSGGG	BAEVM_P10272_3mutA
GGGGSSPAP	PERV_Q4VFZ2_3mutA_WS
GGGEAAAKPAP	MLVMS_P03355_PLV919
GGSGGGPAP	BAEVM_P10272_3mutA

[0841]

接头氨基酸序列	RT结构域名称
PAPGSSGGS	MLVMS_P03355_PLV919
GGSGGGPAP	MLVMS_P03355_3mutA_WS
EAAAKGGSPAP	PERV_Q4VFZ2_3mutA_WS
EAAAKGGSGGG	MLVMS_P03355_3mutA_WS
PAPGSSGGG	MLVFF_P26809_3mutA
GSSEAAAKGGS	MLVFF_P26809_3mutA
PAPGSSEAAAK	MLVFF_P26809_3mutA
EAAAKGSSPAP	KORV_Q9TTC1-Pro_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAK	MLVBM_Q7SVK7_3mutA_WS
PAPGSSEAAAK	MLVMS_P03355_PLV919
EAAAKGSSGGG	MLVMS_P03355_3mutA_WS
EAAAKGGGGGS	AVIRE_P03360_3mutA
EAAAKEAAAKEAAAK	MLVMS_P03355_PLV919
PAPAPAPAPAPAP	MLVFF_P26809_3mutA
GGGGSGGGGSGGGGS	MLVCB_P08361_3mutA
PAPGGSEAAAK	MLVCB_P08361_3mutA
PAPGSSEAAAK	MLVBM_Q7SVK7_3mutA_WS
PAPEAAAKGSS	AVIRE_P03360_3mutA
GGSPAPGSS	WMSV_P03359_3mutA
PAPGGSGGG	MLVMS_P03355_PLV919
EAAAKGGSGSS	MLVMS_P03355_3mutA_WS
GGSGGG	MLVFF_P26809_3mutA
GGSEAAAKGSS	KORV_Q9TTC1_3mutA
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	MLVCB_P08361_3mutA
PAPAPAPAPAPAP	PERV_Q4VFZ2_3mutA_WS
PAPEAAAK	MLVMS_P03355_3mutA_WS

[0842]

接头氨基酸序列	RT结构域名称
GGSEAAAKGGG	MLVMS_P03355_PLV919
GSSPAP	MLVMS_P03355_3mutA_WS
GGGGSS	MLVMS_P03355_PLV919
GGGEAAAKPAP	AVIRE_P03360_3mutA
EAAAKPAPGGS	MLVAV_P03356_3mutA
EAAAKGGGPAP	MLVAV_P03356_3mutA
PAPGGSEAAAK	BAEVM_P10272_3mutA
PAPGGSGSS	MLVMS_P03355_3mutA_WS
PAPGGSGSS	AVIRE_P03360_3mutA
GGSGGGPAP	MLVMS_P03355_3mutA_WS
EAAAKEAAAKEAAAKEAAAK	BAEVM_P10272_3mutA
GGGGSGGGGSGGGGSGGGGSGGGGS	MLVMS_P03355_PLV919
GGGGSSPAP	MLVCB_P08361_3mutA
GSSGGGPAP	MLVFF_P26809_3mutA
GGGGSSGGS	MLVMS_P03355_PLV919
GGSGGG	MLVCB_P08361_3mutA
GSSGGGGGS	MLVMS_P03355_PLV919
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	XMRV6_A1Z651_3mutA
GGGGGSGSS	KORV_Q9TTC1_3mut
GGGEAAAKGGS	BAEVM_P10272_3mutA
GGSGGG	BAEVM_P10272_3mutA
PAPAPAP	KORV_Q9TTC1-Pro_3mut
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	SFV3L_P27401_2mutA
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	MLVBM_Q7SVK7_3mutA_WS
GSSGSSGSSGSSGSS	MLVMS_P03355_3mutA_WS
GSSGGGEAAAK	MLVMS_P03355_3mutA_WS
GSSGGSEAAAK	MLVFF_P26809_3mutA

[0843]

接头氨基酸序列	RT 结构域名称
PAP	MLVMS_P03355_PLV919
EAAAKGGGGSEAAAK	MLVBM_Q7SVK7_3mutA_WS
PAPAP	AVIRE_P03360_3mutA
PAP	MLVFF_P26809_3mutA
GSSGGG	MLVMS_P03355_3mut
GSSPAPGGS	MLVFF_P26809_3mutA
PAPAPAPAP	XMRV6_A1Z651_3mutA
EAAAKGSSGGS	PERV_Q4VFZ2_3mut
PAPEAAAKGGG	KORV_Q9TTC1-Pro_3mutA
PAPGGS	MLVCB_P08361_3mutA
EAAAKGGG	MLVCB_P08361_3mutA
GSSEAAAKPAP	MLVMS_P03355_PLV919
PAPGGS	MLVFF_P26809_3mutA
EAAAKGGS	MLVCB_P08361_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	FLV_P10273_3mutA
PAPGGSEAAAK	MLVAV_P03356_3mutA
GSS	MLVCB_P08361_3mutA
GSSGSSGSSGSS	AVIRE_P03360_3mutA
GSSGSSGSS	MLVFF_P26809_3mutA
GSSGGG	MLVMS_P03355_PLV919
EAAAK	MLVFF_P26809_3mutA
GGSPAPEAAAK	MLVCB_P08361_3mutA
GGSGSS	MLVCB_P08361_3mutA
GSSPAPGGG	MLVMS_P03355_PLV919
EAAAKEAAAKEAAAKEAAAKEAAAK	MLVAV_P03356_3mutA
EAAAKGSSPAP	FLV_P10273_3mutA
GGGGSS	XMRV6_A1Z651_3mutA
GGSPAPGSS	MLVMS_P03355_PLV919

[0844]

接头氨基酸序列	RT 结构域名称
EAAAKEAAAKEAAAKEAAAKEAAAK	MLVMS_P03355_3mutA_WS
PAPEAAAKGGG	FLV_P10273_3mutA
EAAAKPAPGGS	XMRV6_A1Z651_3mut
PAPAP	BAEVM_P10272_3mutA
EAAAKEAAAKEAAAKEAAAK	MLVMS_P03355_PLV919
GSSPAPGGG	MLVMS_P03355_PLV919
EAAAKGGGPAP	KORV_Q9TTC1_3mutA
PAPEAAAK	MLVMS_P03355_PLV919
PAPGGGEAAAK	PERV_Q4VFZ2_3mutA_WS
EAAAKGSSGGS	MLVMS_P03355_3mutA_WS
EAAAKEAAAKEAAAK	MLVMS_P03355_PLV919
GSSEAAAK	MLVMS_P03355_3mutA_WS
GSSGSSGSSGSS	MLVMS_P03355_3mutA_WS
GGGGSGGGGSGGGGSGGGGS	MLVMS_P03355_3mutA_WS
EAAAKGGGGSEAAAK	MLVMS_P03355_3mut
GGS	MLVCB_P08361_3mutA
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	XMRV6_A1Z651_3mutA
GGSGSSPAP	MLVCB_P08361_3mutA
GGGGSGGGGSGGGGS	XMRV6_A1Z651_3mutA
PAPAPAPAPAP	BAEVM_P10272_3mutA
PAPAPAPAPAP	MLVMS_P03355_3mutA_WS
EAAAKEAAAKEAAAKEAAAK	MLVBM_Q7SVK7_3mut
GGGGSGGGGSGGGGSGGGGSGGGGS	BAEVM_P10272_3mutA
GGSGGSGGS	MLVMS_P03355_3mutA_WS
EAAAKPAPGSS	MLVMS_P03355_PLV919
GSS	MLVMS_P03355_3mutA_WS
PAPEAAAKGGS	MLVMS_P03355_3mutA_WS
GGGPAPGGS	MLVMS_P03355_3mutA_WS
EAAAKGGGGSS	MLVAV_P03356_3mutA

[0845]

接头氨基酸序列	RT结构域名称
GSSGSSGSSGSSGSS	MLVFF_P26809_3mut
SGSETPGTSESATPES	PERV_Q4VFZ2_3mut
GGSEAAAKGGG	MLVMS_P03355_3mut
GSSGSSGSSGSSGSSGSS	AVIRE_P03360_3mutA
PAPAPAPAPAPAP	AVIRE_P03360_3mut
GGSGGS	XMRV6_A1Z651_3mutA
PAPGSSEAAAK	MLVCB_P08361_3mut
GGSPAPEAAAK	PERV_Q4VFZ2_3mut
EAAAKGGGGGS	MLVCB_P08361_3mutA
GGSGGSGGSGGS	MLVMS_P03355_PLV919
GGGGSSEAAAK	MLVMS_P03355_PLV919
GSSEAAAKGGG	MLVFF_P26809_3mutA
PAPGGS	MLVMS_P03355_3mutA_WS
EAAAKGGSGGG	MLVCB_P08361_3mutA
EAAAKGGG	PERV_Q4VFZ2_3mut
PAPGGS	XMRV6_A1Z651_3mutA
GSSPAPGGG	XMRV6_A1Z651_3mutA
PAPEAAAKGGG	MLVMS_P03355_3mutA_WS
GSSEAAAKGGG	PERV_Q4VFZ2_3mutA_WS
PAPGGSEAAAK	XMRV6_A1Z651_3mutA
GGGGGS	MLVMS_P03355_3mutA_WS
GGSPAPEAAAK	MLVMS_P03355_3mutA_WS
GGGPAP	MLVFF_P26809_3mutA
PAPGSSGGG	XMRV6_A1Z651_3mutA
PAPGSSGGG	MLVBM_Q7SVK7_3mutA_WS
GGGEAAAKGSS	MLVMS_P03355_3mutA_WS
GSSEAAAKGGS	MLVCB_P08361_3mutA
PAPGGSGSS	MLVCB_P08361_3mutA

[0846]

接头氨基酸序列	RT结构域名称
EAAAKGGGGSEAAAK	BAEVM_P10272_3mutA
PAPAPAP	PERV_Q4VFZ2_3mutA_WS
GGGGGG	MLVAV_P03356_3mutA
GSSPAPEAAAK	MLVCB_P08361_3mutA
GGSGGSGGS	MLVMS_P03355_3mutA_WS
GSSGSSGSSGSSGSS	XMRV6_A1Z651_3mut
GGGPAPGGS	XMRV6_A1Z651_3mutA
GGGPAPEAAAK	BAEVM_P10272_3mutA
GGSGGG	AVIRE_P03360_3mutA
SGSETPGTSESATPES	PERV_Q4VFZ2_3mutA_WS
EAAAKGSSPAP	MLVMS_P03355_PLV919
GSSEAAAK	XMRV6_A1Z651_3mut
GSSGGSGGG	MLVFF_P26809_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAK	WMSV_P03359_3mutA
GGGGSEAAAKGGGGS	MLVMS_P03355_PLV919
PAPGGGGSS	MLVMS_P03355_3mutA_WS
SGSETPGTSESATPES	MLVMS_P03355_3mutA_WS
GGSPAPEAAAK	KORV_Q9TTC1-Pro_3mutA
GSSEAAAKGGG	MLVMS_P03355_3mutA_WS
GSSEAAAK	WMSV_P03359_3mutA
GGGGSEAAAKGGGGS	AVIRE_P03360_3mutA
GSS	WMSV_P03359_3mutA
PAPGGSEAAAK	MLVFF_P26809_3mutA
GGGGS	MLVMS_P03355_3mutA_WS
GGGPAP	MLVMS_P03355_3mutA_WS
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	MLVMS_P03355_3mutA_WS
EAAAKPAPGSS	PERV_Q4VFZ2_3mut
EAAAKPAPGSS	MLVCB_P08361_3mutA
GGGGGG	WMSV_P03359_3mutA

[0847]

接头氨基酸序列	RT结构域名称
EAAAKPAPGGS	MLVMS_P03355_PLV919
PAPGGGEAAAK	PERV_Q4VFZ2_3mut
EAAAKEAAAKEAAAKEAAAKEAAAK	AVIRE_P03360_3mutA
GSSEAAAKPAP	XMRV6_A1Z651_3mutA
PAPGGSEAAAK	MLVBM_Q7SVK7_3mutA_WS
PAPGSS	MLVCB_P08361_3mutA
EAAAKGGG	MLVMS_P03355_3mutA_WS
EAAAKPAP	MLVCB_P08361_3mutA
PAPEAAAKGGS	MLVBM_Q7SVK7_3mutA_WS
GGSPAPGGG	MLVCB_P08361_3mutA
PAPGGSGSS	WMSV_P03359_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	MLVMS_P03355_PLV919
GGSGGGPAP	MLVMS_P03355_PLV919
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	MLVMS_P03355
PAPEAAAKGSS	MLVCB_P08361_3mutA
EAAAKGSS	MLVMS_P03355_3mutA_WS
GGSGGS	MLVMS_P03355_3mutA_WS
EAAAKEAAAKEAAAKEAAAKEAAAK	BAEVM_P10272_3mutA
GGGGSEAAAKGGGGS	FLV_P10273_3mutA
GGSEAAAKGGG	MLVCB_P08361_3mutA
GSSGSSGSSGSSGSS	BAEVM_P10272_3mutA
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	MLVFF_P26809_3mutA
EAAAKGGG	PERV_Q4VFZ2_3mut
GGGGGSEAAAK	MLVCB_P08361_3mutA
EAAAKPAPGGS	MLVMS_P03355_3mutA_WS
GGGGGSGSS	XMRV6_A1Z651_3mutA

[0848]

接头氨基酸序列	RT结构域名称
PAPGSSEAAAK	MLVMS_P03355_3mutA_WS
GSSEAAAKPAP	MLVCB_P08361_3mutA
EAAAKGSSPAP	MLVAV_P03356_3mutA
GGGPAPGGS	WMSV_P03359_3mutA
GGSPAP	MLVMS_P03355_3mutA_WS
GGSEAAAKGGG	MLVMS_P03355_3mutA_WS
GGGGGGGG	MLVFF_P26809_3mutA
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	MLVMS_P03355_3mutA_WS
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	MLVBM_Q7SVK7_3mutA_WS
GSSPAPGGG	MLVAV_P03356_3mutA
GGGGGG	AVIRE_P03360_3mutA
GSSGGS	MLVMS_P03355_3mutA_WS
GGSPAPGSS	MLVFF_P26809_3mutA
PAPEAAAKGGG	PERV_Q4VFZ2_3mut
EAAAKGGGPAP	MLVFF_P26809_3mutA
GGGEAAAKGGS	MLVMS_P03355_PLV919
GGSGSSPAP	MLVFF_P26809_3mutA
SGSETPGTSESATPES	WMSV_P03359_3mutA
PAPGGSEAAAK	MLVBM_Q7SVK7_3mutA_WS
GGSGGG	MLVMS_P03355_PLV919
GGGGSSPAP	PERV_Q4VFZ2_3mut
GGGEAAAKGSS	MLVAV_P03356_3mutA
PAPAPAPAPAPAP	MLVMS_P03355_3mutA_WS
EAAAKGGGGSEAAAK	PERV_Q4VFZ2
EAAAKEAAAKEAAAKEAAAKEAAAK	MLVMS_P03355_PLV919
GGGGGSEAAAK	PERV_Q4VFZ2_3mut
PAPGSSEAAAK	MLVCB_P08361_3mutA

[0849]

接头氨基酸序列	RT结构域名称
GSAGSAAGSGEF	PERV_Q4VFZ2_3mutA_WS
EAAAKGGGGSEAAAK	MLVFF_P26809_3mutA
GGSPAPGGG	PERV_Q4VFZ2_3mutA_WS
GSSEAAAKGGG	AVIRE_P03360_3mutA
GGGEAAAKPAP	MLVMS_P03355_3mutA_WS
GGGPAP	AVIRE_P03360_3mutA
GGSEAAAK	MLVCB_P08361_3mutA
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	PERV_Q4VFZ2_3mut
EAAAKPAPGGS	MLVBM_Q7SVK7_3mutA_WS
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	XMRV6_A1Z651_3mut
GGGGGGGG	MLVCB_P08361_3mutA
PAPGSS	PERV_Q4VFZ2_3mut
EAAAK	PERV_Q4VFZ2_3mut
GSAGSAAGSGEF	MLVMS_P03355_3mutA_WS
PAPGGGEAAAK	PERV_Q4VFZ2_3mut
EAAAKGSSGGS	MLVFF_P26809_3mut
GGGGSEAAAKGGGGS	BAEVM_P10272_3mutA
GGGGSGGGGSGGGGS	MLVMS_P03355_PLV919
EAAAKGGGGSEAAAK	BAEVM_P10272_3mut
PAPGGGEAAAK	MLVMS_P03355_3mutA_WS
GGSEAAAKPAP	MLVMS_P03355_3mutA_WS
PAPAP	MLVCB_P08361_3mutA
PAPAP	MLVFF_P26809_3mutA
GGSPAP	AVIRE_P03360_3mutA
EAAAKGSSGGS	MLVCB_P08361_3mutA
PAPGSSGGS	AVIRE_P03360_3mutA
EAAAKGGGGSEAAAK	XMRV6_A1Z651_3mutA

[0850]

接头氨基酸序列	RT 结构域名称
PAPAPAP	BAEVM_P10272_3mutA
GGSGGSGGSGGSGGSGGS	MLVMS_P03355_PLV919
GGGGGSGSS	MLVMS_P03355_PLV919
PAPGSSEAAAK	XMRV6_A1Z651_3mut
GGSEAAAKPAP	XMRV6_A1Z651_3mutA
EAAAKEAAAKEAAAKEAAAK	XMRV6_A1Z651_3mut
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	WMSV_P03359_3mut
GGSGGGEAAAK	XMRV6_A1Z651_3mutA
GGGEAAAK	XMRV6_A1Z651_3mutA
GGGGSGGGGSGGGGS	MLVMS_P03355_3mutA_WS
GGSGGSGGSGGSGGS	MLVFF_P26809_3mutA
GSSGGGGGS	MLVMS_P03355_3mut
PAPGGSEAAAK	MLVMS_P03355_3mutA_WS
GSSGGSPAP	MLVMS_P03355_3mutA_WS
SGSETPGTSESATPES	XMRV6_A1Z651_3mutA
GGGGSGGGGS	MLVMS_P03355_PLV919
PAPAPAPAPAP	MLVMS_P03355_3mut
GSSGSS	XMRV6_A1Z651_3mutA
GSSEAAAKPAP	PERV_Q4VFZ2_3mut
GGSGSSGGG	MLVMS_P03355_3mutA_WS
EAAAKEAAAK	MLVCB_P08361_3mutA
GSSGSSGSSGSS	MLVMS_P03355_3mutA_WS
GSSPAPGGG	PERV_Q4VFZ2_3mutA_WS
EAAAKEAAAKEAAAK	MLVMS_P03355_3mutA_WS
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	SFV1_P23074_2mutA
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	MLVMS_P03355_PLV919
GSAGSAAGSGEF	MLVMS_P03355_PLV919

[0851]

接头氨基酸序列	RT结构域名称
PAPGSSEAAAK	MLVMS_P03355_3mutA_WS
GGSEAAAK	MLVMS_P03355_3mutA_WS
GSSGSSGSSGSSGSS	PERV_Q4VFZ2_3mutA_WS
GGSEAAAKPAP	PERV_Q4VFZ2_3mutA_WS
GGSGGSGGS	MLVCB_P08361_3mutA
EAAAKGGSGSS	MLVCB_P08361_3mutA
GGGGSGGGGSGGGGSGGGGSGGGGS	FLV_P10273_3mutA
EAAAKEAAAKEAAAKEAAAK	MLVBM_Q7SVK7_3mutA_WS
GGSGSSPAP	BAEVM_P10272_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAK	XMRV6_A1Z651_3mutA
GGGGSGGGGSGGGGSGGGGSGGGGS	MLVBM_Q7SVK7_3mutA_WS
GGSGSS	WMSV_P03359_3mutA
PAPEAAAK	MLVCB_P08361_3mutA
EAAAKPAP	BAEVM_P10272_3mutA
GSSPAP	PERV_Q4VFZ2_3mutA_WS
GGGPAP	PERV_Q4VFZ2_3mutA_WS
EAAAKGGSGSS	MLVMS_P03355_3mutA_WS
EAAAKGGGGSEAAAK	AVIRE_P03360_3mutA
GGSGGG	KORV_Q9TTC1-Pro_3mutA
GSSPAP	MLVFF_P26809_3mutA
GGSGSSEAAAK	BAEVM_P10272_3mutA
PAPGSSGGS	BAEVM_P10272_3mutA
GGGGGG	MLVFF_P26809_3mutA
PAPGGSEAAAK	MLVMS_P03355_PLV919
PAPGGS	MLVMS_P03355_PLV919
GGSGGSGGSGGS	BAEVM_P10272_3mutA
GSSPAP	MLVCB_P08361_3mutA

[0852]

接头氨基酸序列	RT结构域名称
PAPAPAPAP	MLVMS_P03355_3mutA_WS
GGGGGG	MLVCB_P08361_3mutA
GSSGSSGSSGSSGSSGSS	KORV_Q9TTC1-Pro_3mutA
GSSEAAAKGGS	BAEVM_P10272_3mutA
GGSEAAAK	FLV_P10273_3mutA
GGSGGSGGSGGSGGS	KORV_Q9TTC1-Pro_3mutA
GSSPAPEAAAK	PERV_Q4VFZ2_3mut
GSSGSSGSSGSSGSS	XMRV6_A1Z651_3mutA
EAAAKPAPGGS	MLVMS_P03355_3mut
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	FLV_P10273_3mut
GGSPAPEAAAK	XMRV6_A1Z651_3mut
EAAAKGGSGGG	MLVFF_P26809_3mutA
EAAAKEAAAKEAAAKEAAAK	MLVFF_P26809_3mutA
GSSPAP	WMSV_P03359_3mutA
PAPAPAPAP	MLVAV_P03356_3mutA
PAPGGSEAAAK	KORV_Q9TTC1_3mut
GGSGSSEAAAK	MLVBM_Q7SVK7_3mutA_WS
GSSGGG	MLVCB_P08361_3mutA
GGGEAAAKGSS	PERV_Q4VFZ2_3mut
PAPGGSGGG	MLVFF_P26809_3mutA
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	FFV_O93209
PAPGGGGSS	MLVMS_P03355_3mutA_WS
EAAAKGGS	MLVAV_P03356_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	MLVBM_Q7SVK7_3mutA_WS
GGSGGSGGS	WMSV_P03359_3mutA
PAPAP	MLVMS_P03355_3mutA_WS

[0853]

接头氨基酸序列	RT 结构域名称
GSSGGGEAAAK	MLVAV_P03356_3mutA
GGGGSSEAAAK	MLVFF_P26809_3mutA
EAAAKGSSGGS	MLVMS_P03355_PLV919
EAAAKGGGGSEAAAK	MLVMS_P03355_3mutA_WS
GGGGGGGG	MLVMS_P03355_PLV919
GSSGSSGSS	MLVMS_P03355_PLV919
GGGEAAAKPAP	PERV_Q4VFZ2_3mutA_WS
GGGGGSGSS	MLVMS_P03355_3mutA_WS
GGGGGGG	MLVMS_P03355_PLV919
GGS	MLVMS_P03355_PLV919
GSSGGG	MLVMS_P03355_3mutA_WS
EAAAKGGSGSS	PERV_Q4VFZ2_3mutA_WS
PAPGSSEAAAK	MLVMS_P03355_PLV919
GSSEAAAKPAP	MLVMS_P03355_PLV919
GGSPAPGSS	BAEVM_P10272_3mutA
GSAGSAAGSGEF	MLVCB_P08361_3mut
GGSPAPGGG	PERV_Q4VFZ2_3mut
GGGGSGGGGSGGGGSGGGGS	MLVMS_P03355_3mut
GSSGSSGSS	PERV_Q4VFZ2_3mutA_WS
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	PERV_Q4VFZ2_3mut
GGGGSEAAAKGGGGS	MLVCB_P08361_3mutA
GGSEAAAKGSS	MLVAV_P03356_3mutA
EAAAKGGGGSEAAAK	MLVCB_P08361_3mut
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	XMRV6_A1Z651_3mutA
PAPGGGEAAAK	MLVMS_P03355_3mutA_WS
GSSGGGEAAAK	PERV_Q4VFZ2_3mutA_WS
GSSGSS	MLVCB_P08361_3mut
PAPAPAPAPAPAP	PERV_Q4VFZ2_3mut
GGSPAPGGG	MLVFF_P26809_3mutA

[0854]

接头氨基酸序列	RT结构域名称
GGSGGSGGSGGSGGS	MLVCB_P08361_3mutA
EAAAKEAAAK	MLVFF_P26809_3mutA
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	GALV_P21414_3mut
PAPAPAPAPAPAP	WMSV_P03359_3mutA
GGGEAAAKGGS	KORV_Q9TTC1_3mutA
EAAAKGGGPAP	KORV_Q9TTC1_3mut
PAPEAAAKGSS	MLVBM_Q7SVK7_3mutA_WS
PAPEAAAKGSS	FLV_P10273_3mutA
PAPGGSEAAAK	MLVMS_P03355_3mut
GSSPAPGGG	BAEVM_P10272_3mutA
GGGEAAAKPAP	KORV_Q9TTC1-Pro_3mutA
GGGGSGGGGS	MLVMS_P03355_PLV919
GGGEAAAKGSS	MLVFF_P26809_3mutA
PAPGGGGSS	MLVBM_Q7SVK7_3mutA_WS
GSSEAAAK	BAEVM_P10272_3mutA
GGGGGGGG	MLVMS_P03355_PLV919
PAPGSSGGS	MLVAV_P03356_3mutA
GGGGSGGGGSGGGGSGGGGS	BAEVM_P10272_3mutA
PAP	MLVMS_P03355_3mut
EAAAKGSSPAP	XMRV6_A1Z651_3mutA
PAPEAAAKGGS	MLVFF_P26809_3mutA
GSSGGGEAAAK	BAEVM_P10272_3mutA
PAPAPAP	MLVMS_P03355_3mutA_WS
GGSEAAAKGGG	MLVMS_P03355_PLV919
GSSEAAAK	PERV_Q4VFZ2_3mut
GGGG	MLVMS_P03355_3mutA_WS

[0855]

接头氨基酸序列	RT结构域名称
GGGGGS	MLVMS_P03355_3mut
GGGGSSEAAAK	PERV_Q4VFZ2_3mut
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	SFV3L_P27401-Pro_2mutA
GGSEAAAKGSS	MLVMS_P03355_3mutA_WS
PAPGSSGGS	XMRV6_A1Z651_3mutA
GGSPAP	MLVMS_P03355_3mutA_WS
GGGGSSEAAAK	BAEVM_P10272_3mut
GGSGGSGGSGGS	AVIRE_P03360_3mutA
PAPGSSGGS	MLVFF_P26809_3mutA
GSSPAPGGG	MLVMS_P03355_3mutA_WS
GGGGGGG	MLVMS_P03355_3mutA_WS
EAAAKGGGGGS	MLVMS_P03355_3mutA_WS
EAAAKGGSGGG	MLVMS_P03355_PLV919
GGGGSSEAAAK	XMRV6_A1Z651_3mutA
GGGGSEAAAKGGGGS	MLVBM_Q7SVK7_3mutA_WS
GSSGSS	MLVMS_P03355_PLV919
GGSGGG	MLVMS_P03355_PLV919
PAPEAAAKGGG	AVIRE_P03360_3mutA
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	FOAMV_P14350-Pro_2mutA
GGGGGSGSS	PERV_Q4VFZ2_3mut
GSSGSSGSSGSSGSS	KORV_Q9TTC1-Pro_3mut
GGGGSEAAAKGGGGS	MLVMS_P03355_3mutA_WS
GGGGGSPAP	FLV_P10273_3mut
GGGEAAAK	MLVMS_P03355_3mutA_WS
GGSGGSGGSGGS	FLV_P10273_3mutA
GGG	MLVMS_P03355_PLV919
GGSPAPEAAAK	BAEVM_P10272_3mutA

[0856]

接头氨基酸序列	RT结构域名称
EAAAKEAAAK	FLV_P10273_3mutA
GGGEAAAKPAP	BAEVM_P10272_3mutA
GGGEAAAKGGS	PERV_Q4VFZ2_3mut
GGSGGSGGS	PERV_Q4VFZ2_3mut
EAAAKGGGPAP	XMRV6_A1Z651_3mutA
EAAAK	MLVBM_Q7SVK7_3mutA_WS
PAPEAAAKGGG	PERV_Q4VFZ2_3mut
EAAAKGSS	MLVCB_P08361_3mutA
GGSEAAAKGGG	MLVBM_Q7SVK7_3mutA_WS
GGGGSGGGGSGGGGSGGGGS	XMRV6_A1Z651_3mutA
GGGGSGGGGSGGGGSGGGGSGGGGS	BAEVM_P10272_3mut
GGGGSSPAP	PERV_Q4VFZ2_3mutA_WS
GGSGGSGGSGGSGGSGGS	PERV_Q4VFZ2_3mut
GGGEAAAKPAP	PERV_Q4VFZ2_3mut
EAAAKEAAAK	BAEVM_P10272_3mutA
GGSGSSEAAAK	XMRV6_A1Z651_3mutA
PAPEAAAKGSS	WMSV_P03359_3mutA
PAPAPAPAPAP	XMRV6_A1Z651_3mutA
GSSGGGEAAAK	MLVMS_P03355_PLV919
GSSPAPGGG	MLVFF_P26809_3mutA
GGSPAPEAAAK	MLVFF_P26809_3mut
PAPGGSEAAAK	PERV_Q4VFZ2_3mut
GGGGSS	MLVFF_P26809_3mutA
GGSGSSGGG	BAEVM_P10272_3mutA
GSSGGGEAAAK	MLVMS_P03355_3mutA_WS
EAAAKGGS	MLVBM_Q7SVK7_3mutA_WS

[0857]

接头氨基酸序列	RT结构域名称
GGGPAPGGS	MLVMS_P03355_PLV919
EAAAKEAAAK	MLVMS_P03355_PLV919
GSSGSSGSS	MLVMS_P03355_PLV919
GGGEAAAKPAP	MLVAV_P03356_3mutA
SGSETPGTSESATPES	FLV_P10273_3mutA
PAPAPAPAPAP	KORV_Q9TTC1-Pro_3mut
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	BAEVM_P10272_3mutA
PAPGSSGGG	MLVMS_P03355_3mutA_WS
GSSGGGEAAAK	XMRV6_A1Z651_3mutA
GGGGSGGGGSGGGGSGGGGSGGGGS	XMRV6_A1Z651_3mutA
GGGGSSPAP	MLVFF_P26809_3mutA
GGSGGGPAP	PERV_Q4VFZ2_3mutA_WS
GSS	PERV_Q4VFZ2_3mut
EAAAKGSSPAP	MLVMS_P03355_3mut
EAAAKGGG	XMRV6_A1Z651_3mutA
GSSGSSGSSGSS	WMSV_P03359_3mutA
PAPEAAAKGSS	MLVMS_P03355_PLV919
GSSEAAAK	AVIRE_P03360_3mutA
EAAAKGGSGSS	AVIRE_P03360_3mutA
GSSEAAAK	MLVMS_P03355_3mut
GGSGSSEAAAK	MLVMS_P03355_PLV919
GGSEAAAKGGG	MLVFF_P26809_3mutA
GGGGSGGGGSGGGGSGGGGS	MLVAV_P03356_3mutA
PAPAPAPAPAPAP	MLVFF_P26809_3mut
EAAAKPAPGSS	KORV_Q9TTC1-Pro_3mut
PAPGSSEAAAK	MLVAV_P03356_3mutA
GGGGSSPAP	WMSV_P03359_3mutA
EAAAKGGGGGS	MLVMS_P03355_3mutA_WS

[0858]

接头氨基酸序列	RT结构域名称
GGGEAAAKGGS	MLVMS_P03355_3mut
GGSGSSGGG	MLVMS_P03355_3mut
GGGPAPGGS	MLVAV_P03356_3mutA
PAPGGGGGS	MLVMS_P03355_PLV919
GGGPAPGSS	PERV_Q4VFZ2_3mut
GGGGGGG	MLVFF_P26809_3mutA
GGSGGGGSS	MLVCB_P08361_3mutA
GGGGGG	FLV_P10273_3mutA
GGSEAAAKGSS	PERV_Q4VFZ2_3mut
GGSPAPGGG	BAEVM_P10272_3mutA
GGSPAPGSS	AVIRE_P03360_3mutA
GGSGGSGGSGGS	KORV_Q9TTC1_3mut
EAAAKEAAAKEAAAKEAAAKEAAAK	MLVBM_Q7SVK7_3mut
PAPGSSGGS	XMRV6_A1Z651_3mut
EAAAKGGGGSS	PERV_Q4VFZ2_3mutA_WS
GGSGGSGGSGGSGGS	PERV_Q4VFZ2_3mutA_WS
PAPGGSGGG	MLVMS_P03355_PLV919
PAPGSSGGG	PERV_Q4VFZ2_3mutA_WS
GSSGSS	BAEVM_P10272_3mutA
EAAAKGSS	MLVFF_P26809_3mutA
GGGPAP	MLVMS_P03355_PLV919
EAAAKGGGGGS	MLVFF_P26809_3mutA
EAAAKGGSPAP	MLVBM_Q7SVK7_3mutA_WS
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	WMSV_P03359_3mutA
GSSPAPGGG	MLVBM_Q7SVK7_3mutA_WS
GGGEAAAKGSS	AVIRE_P03360_3mutA
GGGGSSEAAAK	AVIRE_P03360_3mutA

[0859]

接头氨基酸序列	RT 结构域名称
GGGGGGGG	PERV_Q4VFZ2_3mutA_WS
PAPGSSEAAAK	BAEVM_P10272_3mutA
EAAAKGSS	MLVFF_P26809_3mut
GSSEAAAKGGG	MLVCB_P08361_3mutA
GGSEAAAK	MLVBM_Q7SVK7_3mutA_WS
GSSEAAAKGGG	PERV_Q4VFZ2_3mutA_WS
PAPGGSGGG	WMSV_P03359_3mutA
GSSGGSGGG	MLVCB_P08361_3mutA
EAAAKGSSGGG	FLV_P10273_3mutA
GSSEAAAK	MLVCB_P08361_3mutA
GSSGGGEAAAK	MLVMS_P03355_3mut
GGGGSGGGGS	MLVCB_P08361_3mutA
EAAAKGGGGSEAAAK	MLVBM_Q7SVK7_3mutA_WS
EAAAKGGG	PERV_Q4VFZ2_3mutA_WS
EAAAKGGSPAP	MLVMS_P03355_PLV919
GGGPAPGGS	AVIRE_P03360_3mutA
GSSEAAAK	MLVBM_Q7SVK7_3mutA_WS
GSSGGGEAAAK	PERV_Q4VFZ2_3mut
SGSETPGTSESATPES	MLVMS_P03355_PLV919
GGSGSSPAP	MLVMS_P03355_3mut
GGGGGG	MLVBM_Q7SVK7_3mutA_WS
GGSPAPGGG	XMRV6_A1Z651_3mutA
GGSGSS	PERV_Q4VFZ2_3mutA_WS
PAP	MLVBM_Q7SVK7_3mutA_WS

[0860]

接头氨基酸序列	RT结构域名称
EAAAKPAPGSS	MLVMS_P03355_PLV919
EAAAKGGG	MLVMS_P03355_3mut
GSSEAAAKPAP	PERV_Q4VFZ2_3mutA_WS
GGGGSS	MLVMS_P03355_3mutA_WS
GGSGSSEAAAK	PERV_Q4VFZ2_3mut
GGGGSS	BAEVM_P10272_3mutA
PAPAP	MLVFF_P26809_3mut
PAPEAAAKGGG	BAEVM_P10272_3mutA
EAAAKGGS	MLVMS_P03355_PLV919
PAPAPAPAPAPAP	PERV_Q4VFZ2_3mutA_WS
GGGGGSEAAAK	MLVMS_P03355_3mut
PAPGGS	PERV_Q4VFZ2_3mut
GGGGSS	MLVCB_P08361_3mutA
GGGGS	MLVAV_P03356_3mutA
GSSPAPEAAAK	MLVMS_P03355_PLV919
GGGGSSGGS	MLVFF_P26809_3mutA
PAPEAAAKGSS	MLVMS_P03355_PLV919
GGSGSSEAAAK	MLVMS_P03355_3mutA_WS
EAAAKGGG	MLVAV_P03356_3mutA
PAPGSSEAAAK	FLV_P10273_3mutA
EAAAKGSSGGG	MLVCB_P08361_3mutA
PAPEAAAK	KORV_Q9TTC1-Pro_3mutA
GGSPAPEAAAK	KORV_Q9TTC1-Pro_3mut
GGSGGSGGSGGSGGSGGS	MLVAV_P03356_3mutA
GSSEAAAKPAP	MLVBM_Q7SVK7_3mutA_WS
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	KORV_Q9TTC1-Pro_3mutA
GSSGGGEAAAK	XMRV6_A1Z651_3mut

[0861]

接头氨基酸序列	RT结构域名称
PAPGGSGGG	AVIRE_P03360_3mutA
PAPGGSEAAAK	PERV_Q4VFZ2_3mutA_WS
GGGGS	MLVMS_P03355_3mutA_WS
GGGGSGGGGSGGGGS	MLVBM_Q7SVK7_3mutA_WS
PAPAPAPAPAP	PERV_Q4VFZ2_3mutA_WS
EAAAKEAAAKEAAAKEAAAKEAAAK	MLVMS_P03355_3mut
GSSGGSEAAAK	MLVMS_P03355_3mutA_WS
GGSGGSGGSGGS	WMSV_P03359_3mutA
EAAAKGSSGGG	WMSV_P03359_3mutA
EAAAKGGG	PERV_Q4VFZ2_3mutA_WS
SGSETPGTSESATPES	PERV_Q4VFZ2_3mut
PAPGSSGGS	MLVMS_P03355_3mutA_WS
PAPEAAAKGSS	PERV_Q4VFZ2_3mut
PAPEAAAK	AVIRE_P03360_3mutA
GSSEAAAKGGG	BAEVM_P10272_3mutA
GSSPAP	MLVAV_P03356_3mutA
EAAAKEAAAKEAAAKEAAAK	MLVFF_P26809_3mut
PAPGGSGSS	MLVAV_P03356_3mutA
GGGGSGGGGSGGGGS	PERV_Q4VFZ2_3mutA_WS
GSSGGSEAAAK	MLVCB_P08361_3mutA
EAAAKGGS	KORV_Q9TTC1-Pro_3mutA
EAAAKGGS	MLVFF_P26809_3mutA
GGSPAP	MLVMS_P03355_PLV919
GGSGSS	MLVMS_P03355_PLV919
SGSETPGTSESATPES	WMSV_P03359_3mut
GGGGGGG	WMSV_P03359_3mut
GGSPAPGSS	MLVCB_P08361_3mutA
GGGGSSGGS	WMSV_P03359_3mut

[0862]

接头氨基酸序列	RT结构域名称
PAPGGS	MLVMS_P03355_PLV919
PAPGSSGGS	MLVCB_P08361_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAK	MLVFF_P26809_3mut
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	PERV_Q4VFZ2_3mut
GGSGGSGGSGGSGGS	BAEVM_P10272_3mutA
GSSEAAAK	PERV_Q4VFZ2_3mut
EAAAKEAAAKEAAAKEAAAK	KORV_Q9TTC1-Pro_3mutA
GGSGGSGGSGGSGGS	MLVMS_P03355_3mut
PAPAPAPAPAPAP	MLVMS_P03355_3mut
GGSPAPEAAAK	MLVMS_P03355_PLV919
EAAAK	WMSV_P03359_3mutA
EAAAKGSSGGS	MLVBM_Q7SVK7_3mutA_WS
GGSGGGGSS	MLVMS_P03355_3mutA_WS
GGGEAAAKPAP	MLVMS_P03355_3mut
EAAAKGGSGGG	XMRV6_A1Z651_3mutA
GGGGGSEAAAK	KORV_Q9TTC1-Pro_3mutA
GGGGGG	BAEVM_P10272_3mutA
GGGGGG	MLVMS_P03355_3mut
GGGGGGG	MLVBM_Q7SVK7_3mutA_WS
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	AVIRE_P03360
PAPGSSGGS	PERV_Q4VFZ2_3mut
GGGGGS	XMRV6_A1Z651_3mut
EAAAKPAP	XMRV6_A1Z651_3mutA
GGG	MLVMS_P03355_3mutA_WS
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	FLV_P10273_3mut

[0863]

接头氨基酸序列	RT结构域名称
EAAAKGSSPAP	MLVMS_P03355_3mut
SGSETPGTSESATPES	BAEVM_P10272_3mutA
GGSPAPEAAAK	MLVMS_P03355_3mut
GSSGSSGSSGSS	MLVAV_P03356_3mutA
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	MLVMS_P03355_3mut
GGSPAP	MLVCB_P08361_3mutA
GGGGGSEAAAK	MLVMS_P03355_3mutA_WS
GGGGG	MLVFF_P26809_3mutA
GSSEAAAK	MLVAV_P03356_3mutA
GGS	BAEVM_P10272_3mut
EAAAKGGSPAP	MLVCB_P08361_3mutA
PAPAPAPAP	FLV_P10273_3mutA
PAPGGGEAAAK	MLVCB_P08361_3mutA
GGGGSSEAAAK	MLVMS_P03355_3mutA_WS
GGGGG	PERV_Q4VFZ2_3mutA_WS
GGSGGSGGSGGSGGSGGS	PERV_Q4VFZ2_3mut
GGGGG	MLVMS_P03355_3mut
PAPEAAAKGGG	MLVBM_Q7SVK7_3mutA_WS
GSSGGGPAP	XMRV6_A1Z651_3mutA
GSSGSSGSSGSSGSSGSS	PERV_Q4VFZ2_3mutA_WS
EAAAKGGSPAP	PERV_Q4VFZ2_3mut
GSSGGSEAAAK	MLVMS_P03355_PLV919
GSS	PERV_Q4VFZ2_3mut
EAAAKGGS	WMSV_P03359_3mutA
GGGGGSPAP	PERV_Q4VFZ2_3mutA_WS
EAAAKGSS	MLVMS_P03355_PLV919
EAAAKGGGGSS	KORV_Q9TTC1-Pro_3mutA

[0864]

接头氨基酸序列	RT 结构域名称
PAPGSSGGG	PERV_Q4VFZ2_3mut
GGGGSSEAAAK	MLVFF_P26809_3mut
PAPAPAP	MLVMS_P03355_3mut
GSSGGSEAAAK	XMRV6_A1Z651_3mut
PAPEAAAKGSS	MLVMS_P03355_3mutA_WS
GGSGGSGGSGGSGGS	MLVMS_P03355_3mutA_WS
GGSGSSPAP	XMRV6_A1Z651_3mutA
GGGGSSPAP	MLVMS_P03355_PLV919
GGGGS	MLVCB_P08361_3mutA
EAAAKEAAAKEAAAKEAAAK	PERV_Q4VFZ2_3mutA_WS
EAAAKEAAAK	KORV_Q9TTC1_3mutA
PAPGGGEAAAK	BAEVM_P10272_3mutA
GSSGGSEAAAK	XMRV6_A1Z651_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	FLV_P10273_3mut
GSSEAAAKPAP	MLVMS_P03355_3mutA_WS
EAAAKPAPGSS	PERV_Q4VFZ2_3mutA_WS
GSSGGSPAP	XMRV6_A1Z651_3mutA
GSSEAAAKGGG	PERV_Q4VFZ2_3mut
GGGEAAAKGGS	WMSV_P03359_3mutA
GSSEAAAKGGG	MLVFF_P26809_3mut
PAPAPAP	KORV_Q9TTC1-Pro_3mutA
EAAAKGGSPAP	MLVMS_P03355_3mutA_WS
PAPGGSEAAAK	PERV_Q4VFZ2_3mut
GGGGS	MLVBM_Q7SVK7_3mutA_WS
EAAAKGSSGGG	KORV_Q9TTC1_3mut
EAAAKGGGPAP	MLVCB_P08361_3mutA
EAAAKGSS	BAEVM_P10272_3mutA
GGSPAPGGG	MLVBM_Q7SVK7_3mutA_W

[0865]

接头氨基酸序列	RT 结构域名称
	S
GGGGSEAAAKGGGGS	MLVMS_P03355_3mutA_WS
GGGEAAAKGGS	PERV_Q4VFZ2_3mutA_WS
EAAAKGGGGSS	MLVMS_P03355_3mutA_WS
EAAAKGGGPAP	MLVFF_P26809_3mut
GSSPAP	PERV_Q4VFZ2_3mutA_WS
EAAAKGGS	MLVMS_P03355_3mut
GGGGSS	KORV_Q9TTC1-Pro_3mutA
EAAAKGSSPAP	MLVMS_P03355_3mutA_WS
GGGPAP	PERV_Q4VFZ2_3mut
EAAAKGSSGGS	XMRV6_A1Z651_3mutA
PAPGGG	MLVAV_P03356_3mutA
GSSPAPEAAAK	BAEVM_P10272_3mutA
GGGPAP	MLVBM_Q7SVK7_3mutA_WS
GSSGGGGGS	AVIRE_P03360_3mutA
SGSETPGTSESATPES	MLVMS_P03355_PLV919
GGGPAP	MLVFF_P26809_3mut
EAAAKGGGGSS	XMRV6_A1Z651_3mutA
GGGGSSPAP	XMRV6_A1Z651_3mut
GGGGSEAAAKGGGGS	MLVMS_P03355_3mut
GSSPAP	MLVBM_Q7SVK7_3mutA_WS
GGSGSSEAAAK	FLV_P10273_3mutA
SGSETPGTSESATPES	MLVBM_Q7SVK7_3mutA_WS
PAPGGG	AVIRE_P03360_3mutA
GGGEAAAKPAP	MLVMS_P03355_3mutA_WS
EAAAKGGSGSS	PERV_Q4VFZ2_3mut

[0866]

接头氨基酸序列	RT结构域名称
GGSPAPGGG	MLVAV_P03356_3mutA
PAPGGSGSS	BAEVM_P10272_3mutA
GSSGGSPAP	MLVFF_P26809_3mutA
EAAAKGSSGGG	PERV_Q4VFZ2_3mut
GGGGSGGGGS	PERV_Q4VFZ2_3mutA_WS
GSSGGGGGS	BAEVM_P10272_3mutA
GGGGSSGGS	MLVBM_Q7SVK7_3mutA_WS
EAAAKGGS	PERV_Q4VFZ2_3mutA_WS
GSSGSSGSSGSS	MLVMS_P03355_3mut
GGS	MLVMS_P03355_3mutA_WS
GSSGGSEAAAK	MLVBM_Q7SVK7_3mutA_WS
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	XMRV6_A1Z651
GGGGG	FLV_P10273_3mutA
PAPEAAAKGSS	PERV_Q4VFZ2_3mut
GGGGGG	WMSV_P03359_3mut
EAAAKGGG	BAEVM_P10272_3mutA
GGGGSS	MLVMS_P03355_3mutA_WS
GSSGGGEAAAK	KORV_Q9TTC1_3mut
GGSGSS	AVIRE_P03360_3mutA
EAAAKPAP	MLVMS_P03355_3mut
EAAAKEAAAKEAAAK	FLV_P10273_3mutA
GGGG	XMRV6_A1Z651_3mutA
GSSPAPGGS	BAEVM_P10272_3mutA
GSSGGGGGS	MLVFF_P26809_3mutA
GGGGSSGGS	MLVAV_P03356_3mutA
GGS	PERV_Q4VFZ2_3mut
GGGGG	WMSV_P03359_3mutA

[0867]

接头氨基酸序列	RT 结构域名称
GSSGSSGSSGSSGSSGSS	FLV_P10273_3mutA
PAPGGGGSS	MLVAV_P03356_3mutA
GGGGGGGG	BAEVM_P10272_3mutA
SGSETPGTSESATPES	MLVCB_P08361_3mutA
PAPGGG	BAEVM_P10272_3mutA
GSSGSSGSS	MLVCB_P08361_3mutA
GGSGSS	MLVMS_P03355_3mutA_WS
EAAAKGGGGSEAAAK	WMSV_P03359_3mutA
GGGGGGGG	FLV_P10273_3mutA
GSSGSS	MLVMS_P03355_3mutA_WS
PAPEAAAKGGS	XMRV6_A1Z651_3mutA
EAAAKEAAAK	MLVMS_P03355_3mut
GGGGSGGGGSGGGGS	BAEVM_P10272_3mutA
EAAAKGSSPAP	MLVMS_P03355_PLV919
GGGGSSEAAAK	MLVMS_P03355_3mut
GGGGSSEAAAK	BAEVM_P10272_3mutA
PAPGGSGSS	PERV_Q4VFZ2_3mut
GGSGGGEAAAK	MLVFF_P26809_3mut
PAPEAAAKGGS	PERV_Q4VFZ2_3mut
GGGPAPGSS	AVIRE_P03360_3mut
PAPGGSGGG	PERV_Q4VFZ2_3mutA_WS
GGGGGGGG	PERV_Q4VFZ2_3mutA_WS
GSSEAAAK	MLVMS_P03355_3mutA_WS
GGGGSGGGGSGGGGS	PERV_Q4VFZ2_3mutA_WS
EAAAKGGS	MLVMS_P03355_3mut
GGGGGSGSS	MLVCB_P08361_3mut
GGGPAP	KORV_Q9TTC1-Pro_3mutA
EAAAKPAPGGG	MLVCB_P08361_3mut
GSSGGSPAP	MLVCB_P08361_3mutA

[0868]

接头氨基酸序列	RT 结构域名称
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	MLVMS_P03355_3mut
PAPAPAPAP	MLVMS_P03355_3mut
GSSGGS	XMRV6_A1Z651_3mutA
GSSEAAAKGGG	MLVMS_P03355_3mut
GGSGSSPAP	MLVMS_P03355_3mutA_WS
GSSEAAAKGGS	MLVMS_P03355_PLV919
EAAAKEAAAKEAAAKEAAAKEAAAK	BAEVM_P10272_3mut
PAPGGGGSS	KORV_Q9TTC1_3mutA
EAAAKGSS	MLVMS_P03355_3mutA_WS
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	FFV_O93209_2mut
GGSGGSGGSGGSGGSGGS	BAEVM_P10272_3mutA
GGGGGG	MLVMS_P03355_PLV919
PAPEAAAK	BAEVM_P10272_3mutA
GGSGSSEAAAK	MLVAV_P03356_3mutA
GGG	MLVCB_P08361_3mutA
GGGGG	MLVCB_P08361_3mutA
GGSGGSGGSGGS	KORV_Q9TTC1-Pro_3mutA
GSSGSSGSSGSSGSSGSS	XMRV6_A1Z651_3mutA
GSSEAAAKPAP	FLV_P10273_3mutA
GGGEAAAKPAP	MLVCB_P08361_3mutA
GSSGSSGSS	MLVMS_P03355_3mutA_WS
PAPAPAPAP	MLVMS_P03355_PLV919
EAAAKGGG	MLVMS_P03355_PLV919
PAPAPAPAPAPAP	FLV_P10273_3mutA
EAAAKGGSGSS	MLVMS_P03355_3mut
GGGGGG	PERV_Q4VFZ2_3mutA_WS
PAPGGG	MLVCB_P08361_3mutA
GGGGGSGSS	KORV_Q9TTC1_3mutA

[0869]

接头氨基酸序列	RT结构域名称
GGGGSGGGGSGGGGSGGGGS	XMRV6_A1Z651_3mut
GGSGGSGGS	KORV_Q9TTC1-Pro_3mutA
EAAAKPAPGGG	MLVMS_P03355_3mutA_WS
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	XMRV6_A1Z651
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	FLV_P10273_3mutA
EAAAKGGGGSEAAAK	PERV_Q4VFZ2_3mutA_WS
GGGPAPGSS	AVIRE_P03360_3mutA
GGGGG	MLVMS_P03355_3mutA_WS
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	MLVMS_P03355_3mut
GGGGSGGGGS	MLVMS_P03355_3mutA_WS
EAAAKGGSPAP	XMRV6_A1Z651_3mutA
EAAAKGSSPAP	AVIRE_P03360_3mutA
PAPGGSGSS	KORV_Q9TTC1-Pro_3mutA
GSS	MLVBM_Q7SVK7_3mutA_WS
GSS	WMSV_P03359_3mut
GGGPAPGSS	MLVFF_P26809_3mutA
EAAAKPAP	MLVMS_P03355_3mut
GSSPAPEAAAK	FLV_P10273_3mutA
GGSPAPGSS	MLVBM_Q7SVK7_3mutA_WS
GGGGGSEAAAK	XMRV6_A1Z651_3mut
PAPEAAAKGGG	WMSV_P03359_3mutA
PAPGGG	PERV_Q4VFZ2_3mut
GGSPAPEAAAK	WMSV_P03359_3mutA
GGSGGGGSS	PERV_Q4VFZ2_3mut
EAAAKGGGGSS	PERV_Q4VFZ2_3mut
EAAAKGGSPAP	AVIRE_P03360_3mut

[0870]

接头氨基酸序列	RT结构域名称
GGSGGGGSS	WMSV_P03359_3mutA
PAPGSSEAAAK	MLVFF_P26809_3mut
GSSEAAAK	MLVMS_P03355_PLV919
GSAGSAAGSGEF	AVIRE_P03360_3mutA
EAAAKGGSGSS	MLVMS_P03355_3mut
GGSEAAAKPAP	MLVMS_P03355_PLV919
GGGGSGGGGSGGGGSGGGGSGGGGS	MLVFF_P26809_3mutA
PAPGSSEAAAK	PERV_Q4VFZ2_3mutA_WS
GGGGSSPAP	MLVMS_P03355_3mutA_WS
PAPAPAP	MLVCB_P08361_3mutA
EAAAKPAPGGG	MLVBM_Q7SVK7_3mutA_WS
GGGPAPGSS	BAEVM_P10272_3mutA
PAP	MLVMS_P03355_3mutA_WS
PAPGGSGGG	MLVMS_P03355_3mutA_WS
GGSGGSGGSGGSGGS	MLVBM_Q7SVK7_3mutA_WS
PAPAPAPAP	XMRV6_A1Z651_3mut
GSSPAPGGG	MLVMS_P03355_3mutA_WS
GSSPAPGGG	MLVMS_P03355_3mut
PAPGGG	MLVMS_P03355_PLV919
GGGEAAAKGSS	WMSV_P03359_3mut
EAAAKGSS	KORV_Q9TTC1-Pro_3mutA
EAAAKGGS	PERV_Q4VFZ2_3mut
EAAAKEAAAKEAAAKEAAAKEAAAK	PERV_Q4VFZ2_3mut
PAPEAAAKGGG	MLVMS_P03355_PLV919
EAAAKGSSGGG	MLVFF_P26809_3mut
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	PERV_Q4VFZ2

[0871]

接头氨基酸序列	RT 结构域名称
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	MLVAV_P03356_3mutA
GSSGGSGGG	MLVFF_P26809_3mut
GSSGSSGSSGSS	PERV_Q4VFZ2_3mutA_WS
GGSPAPGGG	MLVMS_P03355_PLV919
GSS	BAEVM_P10272_3mut
GGGPAPGSS	MLVMS_P03355_3mutA_WS
GGGGSS	KORV_Q9TTC1_3mutA
GSSGGSGGG	BAEVM_P10272_3mutA
EAAAKEAAAKEAAAK	MLVCB_P08361_3mutA
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	FLV_P10273_3mutA
PAPGGGGGS	PERV_Q4VFZ2_3mut
PAPAPAPAPAP	KORV_Q9TTC1-Pro_3mutA
EAAAK	MLVMS_P03355_3mutA_WS
GGG	MLVCB_P08361_3mut
GGSEAAAKGGG	BAEVM_P10272_3mutA
GGGGGSGSS	MLVAV_P03356_3mutA
EAAAKGSSPAP	MLVBM_Q7SVK7_3mutA_WS
GGSGGSGGS	XMRV6_A1Z651_3mut
EAAAKPAPGGG	KORV_Q9TTC1-Pro_3mutA
GGGPAPEAAAK	FLV_P10273_3mutA
GGSPAPEAAAK	MLVMS_P03355_3mutA_WS
GGSGGSGGSGGSGGS	MLVFF_P26809_3mut
EAAAKGGSGSS	MLVMS_P03355_PLV919
GGGEAAAKGGS	MLVBM_Q7SVK7_3mutA_WS
PAPAPAPAP	BAEVM_P10272_3mutA
EAAAKEAAAKEAAAKEAAAK	MLVMS_P03355_3mut
EAAAKPAP	XMRV6_A1Z651_3mut

[0872]

接头氨基酸序列	RT 结构域名称
EAAAKEAAAK	MLVBM_Q7SVK7_3mutA_WS
EAAAKGGG	BAEVM_P10272_3mut
EAAAKGSS	MLVAV_P03356_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	MLVFF_P26809_3mut
GGGPAPGSS	PERV_Q4VFZ2_3mutA_WS
GGGG	PERV_Q4VFZ2_3mut
EAAAKGGSGSS	MLVMS_P03355_PLV919
GGGGSGGGGSGGGGS	MLVMS_P03355_3mutA_WS
EAAAK	MLVMS_P03355_3mutA_WS
GGGGSS	PERV_Q4VFZ2
PAPEAAAKGGS	MLVCB_P08361_3mut
GSS	MLVMS_P03355_3mut
GSAGSAAGSGEF	MLVFF_P26809_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	KORV_Q9TTC1-Pro_3mut
GGGGSGGGGS	AVIRE_P03360_3mutA
EAAAK	MLVMS_P03355_3mut
GGGPAPGGS	PERV_Q4VFZ2_3mut
GGGGSGGGGSGGGGS	MLVMS_P03355_PLV919
PAPGGG	MLVMS_P03355_3mutA_WS
GGGEAAAKPAP	PERV_Q4VFZ2_3mutA_WS
EAAAKPAPGSS	KORV_Q9TTC1-Pro_3mutA
PAPGSS	KORV_Q9TTC1_3mutA
GSAGSAAGSGEF	PERV_Q4VFZ2_3mut
PAPGGGGSS	KORV_Q9TTC1-Pro_3mutA
GSSGGGEAAAK	MLVCB_P08361_3mutA
GSS	AVIRE_P03360_3mutA
GSSGSSGSSGSS	XMRV6_A1Z651_3mutA
PAPEAAAKGGG	MLVMS_P03355_PLV919

[0873]

接头氨基酸序列	RT 结构域名称
GGGPAPEAAAK	MLVCB_P08361_3mutA
PAPGGGGGS	MLVCB_P08361_3mutA
EAAAKEAAAKEAAAKEAAAK	PERV_Q4VFZ2_3mutA_WS
GGGGGSPAP	MLVFF_P26809_3mutA
GSSGSSGSSGSSGSS	PERV_Q4VFZ2
GSSPAPEAAAK	MLVMS_P03355_PLV919
GSSGSSGSSGSSGSSGSS	MLVBM_Q7SVK7_3mutA_WS
GSSGSSGSSGSSGSSGSS	MLVMS_P03355_3mutA_WS
GGSPAPEAAAK	MLVAV_P03356_3mutA
GSSGGG	BAEVM_P10272_3mut
EAAAKGSSGGS	KORV_Q9TTC1-Pro_3mutA
GGSGSSEAAAK	MLVMS_P03355_3mutA_WS
GGGPAPEAAAK	MLVFF_P26809_3mutA
GGGPAPGGS	MLVMS_P03355_3mutA_WS
GGGGG	MLVMS_P03355_PLV919
GGGEAAAKPAP	MLVBM_Q7SVK7_3mutA_WS
GGGGSGGGGS	WMSV_P03359_3mut
GGGPAPEAAAK	PERV_Q4VFZ2_3mut
GGSGSSEAAAK	MLVMS_P03355_PLV919
EAAAKGGGPAP	MLVMS_P03355_3mutA_WS
GSSGSSGSSGSSGSS	KORV_Q9TTC1-Pro_3mutA
PAPAP	WMSV_P03359_3mutA
GGSPAPGSS	MLVAV_P03356_3mutA
GGSGGGPAP	MLVMS_P03355_3mut
GGSPAP	MLVMS_P03355_PLV919
EAAAKGGSPAP	PERV_Q4VFZ2_3mut
GSSPAPGGG	KORV_Q9TTC1-Pro_3mutA

[0874]

接头氨基酸序列	RT结构域名称
GSAGSAAGSGEF	MLVMS_P03355_3mut
GGSPAP	PERV_Q4VFZ2_3mut
GSSGSS	KORV_Q9TTC1-Pro_3mut
GGGPAPGSS	MLVMS_P03355_3mutA_WS
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	FOAMV_P14350
PAPGSSGGG	MLVMS_P03355_PLV919
GGSEAAAKPAP	BAEVM_P10272_3mutA
GGGGGS	MLVCB_P08361_3mutA
PAPEAAAKGGS	MLVMS_P03355_3mutA_WS
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	BAEVM_P10272_3mutA
GGSEAAAK	BAEVM_P10272_3mutA
GSSPAPEAAAK	MLVMS_P03355_3mutA_WS
PAPGGG	WMSV_P03359_3mut
EAAAKPAP	PERV_Q4VFZ2_3mut
GSSGSSGSSGSSGSS	WMSV_P03359_3mut
PAPGGG	MLVBM_Q7SVK7_3mutA_WS
GGSGGGEAAAK	BAEVM_P10272_3mutA
PAPGGS	MLVMS_P03355_3mut
GGSGGSGGSGGS	MLVBM_Q7SVK7_3mutA_WS
EAAAKEAAAKEAAAKEAAAK	PERV_Q4VFZ2_3mut
GGSEAAAKGGG	WMSV_P03359_3mutA
GGGPAP	BAEVM_P10272_3mutA
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	XMRV6_A1Z651_3mut
GGSPAPGSS	KORV_Q9TTC1_3mut
GGGPAPGSS	MLVMS_P03355_3mut
GGGGSSGGS	BAEVM_P10272_3mutA

[0875]

接头氨基酸序列	RT结构域名称
GGGEAAAKGSS	KORV_Q9TTC1-Pro_3mutA
PAPAP	MLVBM_Q7SVK7_3mutA_WS
GGSPAPGGG	PERV_Q4VFZ2_3mut
PAPGSS	PERV_Q4VFZ2_3mutA_WS
GSSGGSPAP	MLVBM_Q7SVK7_3mutA_WS
EAAAKGGGGSEAAAK	PERV_Q4VFZ2_3mut
GSSEAAAKGGS	KORV_Q9TTC1-Pro_3mut
PAPAPAPAP	KORV_Q9TTC1-Pro_3mutA
GGSEAAAKPAP	WMSV_P03359_3mutA
PAPGGS	FLV_P10273_3mutA
EAAAKGGGPAP	PERV_Q4VFZ2_3mut
GGSGSSGGG	AVIRE_P03360_3mutA
EAAAKGGSGSS	BAEVM_P10272_3mutA
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	MLVCB_P08361_3mutA
GSSEAAAKGGS	XMRV6_A1Z651_3mutA
GGGGG	BAEVM_P10272_3mutA
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	SFV3L_P27401_2mutA
GGGEAAAKGSS	MLVMS_P03355_PLV919
EAAAKGGGGSEAAAK	KORV_Q9TTC1_3mutA
EAAAKGGG	AVIRE_P03360_3mut
GGSGGG	MLVMS_P03355_3mutA_WS
GGSGSSGGG	MLVMS_P03355_PLV919
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	KORV_Q9TTC1_3mut
GGGGSEAAAKGGGGS	KORV_Q9TTC1_3mutA
PAPAPAPAPAP	FLV_P10273_3mutA
GGS	MLVBM_Q7SVK7_3mutA_WS

[0876]

接头氨基酸序列	RT结构域名称
GGGGGSEAAAK	MLVBM_Q7SVK7_3mutA_WS
GSSGSSGSSGSSGSS	MLVMS_P03355_3mutA_WS
EAAAKEAAAKEAAAKEAAAKEAAAK	MLVMS_P03355_3mut
GGSGSSGGG	PERV_Q4VFZ2_3mut
PAP	MLVFF_P26809_3mut
GSSPAPEAAAK	MLVAV_P03356_3mutA
EAAAKGGGGSS	MLVMS_P03355_3mut
GGGEAAAKGGS	XMRV6_A1Z651_3mut
GGSGGGPAP	MLVBM_Q7SVK7_3mutA_WS
GSAGSAAGSGEF	BAEVM_P10272_3mutA
GSSEAAAK	MLVCB_P08361_3mut
PAPGSS	MLVMS_P03355_3mut
EAAAKEAAAKEAAAK	MLVAV_P03356_3mutA
GSAGSAAGSGEF	XMRV6_A1Z651_3mutA
GSSGSSGSSGSS	BAEVM_P10272_3mutA
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	KORV_Q9TTC1-Pro_3mut
GGGGSSEAAAK	WMSV_P03359_3mut
GSSGGGEAAAK	MLVBM_Q7SVK7_3mutA_WS
EAAAKPAP	MLVFF_P26809_3mutA
GGSPAPGGG	KORV_Q9TTC1_3mutA
PAPEAAAK	FLV_P10273_3mutA
GSSGSSGSS	MLVBM_Q7SVK7_3mutA_WS
GSSGGGEAAAK	FLV_P10273_3mutA
GGSPAP	MLVBM_Q7SVK7_3mutA_W

[0877]

接头氨基酸序列	RT 结构域名称
	S
GSAGSAAGSGEF	KORV_Q9TTC1-Pro_3mutA
PAPGGSEAAAK	MLVMS_P03355_PLV919
GGSPAPEAAAK	MLVBM_Q7SVK7_3mutA_WS
GGGGGSPAP	MLVBM_Q7SVK7_3mutA_WS
EAAAKGSSPAP	WMSV_P03359_3mut
EAAAKGGGPAP	MLVBM_Q7SVK7_3mutA_WS
PAPGSS	KORV_Q9TTC1-Pro_3mutA
GGSGSSGGG	BAEVM_P10272_3mut
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	FFV_O93209-Pro_2mut
GGSGGSGGSGGSGGSGGS	WMSV_P03359_3mutA
GGSGGSGGS	PERV_Q4VFZ2_3mutA_WS
GGGGG	PERV_Q4VFZ2_3mutA_WS
GGGPAP	FLV_P10273_3mutA
PAPGGSGGG	XMRV6_A1Z651_3mutA
GGGGSEAAAKGGGGS	XMRV6_A1Z651_3mut
EAAAKGSSGGG	KORV_Q9TTC1-Pro_3mutA
GSSGGSEAAAK	WMSV_P03359_3mut
EAAAKGGSGSS	PERV_Q4VFZ2_3mut
PAPAPAPAPAP	PERV_Q4VFZ2_3mut
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	MLVMS_P03355_3mutA_WS
GGGGGGG	KORV_Q9TTC1_3mutA
EAAAK	KORV_Q9TTC1-Pro_3mutA
GGGEAAAKGGS	KORV_Q9TTC1-Pro_3mutA
GGGEAAAKGGS	PERV_Q4VFZ2_3mutA_WS
GGGGGSPAP	XMRV6_A1Z651_3mut

[0878]

接头氨基酸序列	RT结构域名称
GGGGSGGGGSGGGGSGGGGS	MLVFF_P26809_3mut
GGGGGGG	MLVFF_P26809_3mut
PAPAPAPAPAPAP	AVIRE_P03360_3mutA
GSSPAPGGG	FLV_P10273_3mutA
GGGGGSPAP	MLVMS_P03355_3mutA_WS
GGGGSGGGGSGGGGS	MLVMS_P03355_3mut
GGGGSGGGGSGGGGS	KORV_Q9TTC1_3mut
GSSEAAAKGGS	MLVAV_P03356_3mutA
GSSGSSGSSGSSGSS	MLVMS_P03355_3mut
EAAAKGGGGGS	PERV_Q4VFZ2_3mutA_WS
GSSGGGGGS	PERV_Q4VFZ2_3mut
GGGEAAAKPAP	MLVMS_P03355_3mut
GSSGGSPAP	PERV_Q4VFZ2_3mutA_WS
GSSGGGPAP	BAEVM_P10272_3mutA
GGGGGSGSS	MLVMS_P03355_PLV919
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	BAEVM_P10272_3mut
PAPEAAAK	MLVMS_P03355_3mut
GGGGSGGGGSGGGGS	FLV_P10273_3mutA
GGSGSSGGG	WMSV_P03359_3mutA
EAAAKGGS	PERV_Q4VFZ2_3mut
EAAAKGSSPAP	MLVCB_P08361_3mut
EAAAKGGSGSS	WMSV_P03359_3mutA
GSSGSS	PERV_Q4VFZ2_3mutA_WS
PAPAPAPAP	MLVMS_P03355_PLV919
GGSGGG	PERV_Q4VFZ2_3mutA_WS
GSS	MLVBM_Q7SVK7_3mutA_WS
PAP	KORV_Q9TTC1-Pro_3mutA

[0879]

接头氨基酸序列	RT 结构域名称
GGSGSSEAAAK	MLVFF_P26809_3mut
PAPEAAAKGSS	KORV_Q9TTC1-Pro_3mutA
GGSGGS	MLVCB_P08361_3mutA
GGGGGGG	PERV_Q4VFZ2_3mutA_WS
GGSPAPEAAAK	MLVBM_Q7SVK7_3mut
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	KORV_Q9TTC1_3mutA
GGSPAP	MLVMS_P03355_3mut
GGSEAAAKGGG	PERV_Q4VFZ2_3mut
GGGGSGGGGS	FLV_P10273_3mutA
GGGEAAAK	BAEVM_P10272_3mutA
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	SFV3L_P27401_2mut
GGSEAAAKPAP	KORV_Q9TTC1-Pro_3mutA
GSSGGGEAAAK	MLVMS_P03355_PLV919
GGGGGSEAAAK	MLVMS_P03355_PLV919
EAAAKGGSGGG	MLVMS_P03355_3mutA_WS
GGGGSSPAP	MLVAV_P03356_3mutA
EAAAKEAAAK	MLVMS_P03355_3mutA_WS
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	SFV3L_P27401_2mut
GSSGSSGSSGSSGSS	MLVMS_P03355_PLV919
GSSGGG	KORV_Q9TTC1-Pro_3mutA
GSSGGS	MLVFF_P26809_3mutA
GGGGSGGGGS	XMRV6_A1Z651_3mutA
PAPGSS	MLVBM_Q7SVK7_3mutA_WS
GGGPAPEAAAK	XMRV6_A1Z651_3mutA
EAAAKGGS	MLVFF_P26809_3mut
GSS	KORV_Q9TTC1_3mutA
GGGG	PERV_Q4VFZ2_3mut

[0880]

接头氨基酸序列	RT 结构域名称
GGGGGSEAAAK	AVIRE_P03360_3mutA
GSSGSSGSSGSSGSS	MLVMS_P03355_PLV919
PAPGGSGGG	PERV_Q4VFZ2_3mut
GGGPAP	PERV_Q4VFZ2_3mut
GGGPAPEAAAK	AVIRE_P03360_3mutA
GGGEAAAK	MLVCB_P08361_3mut
GGG	MLVFF_P26809_3mutA
EAAAKPAPGSS	XMRV6_A1Z651_3mutA
GGSGSSEAAAK	PERV_Q4VFZ2_3mutA_WS
EAAAKGSS	MLVMS_P03355_3mut
GGSGSSEAAAK	BAEVM_P10272_3mut
GGSGGG	MLVBM_Q7SVK7_3mutA_WS
GGGPAP	MLVMS_P03355_PLV919
GGSPAPGGG	PERV_Q4VFZ2_3mutA_WS
GGGGGSEAAAK	MLVFF_P26809_3mutA
EAAAKGSSGGS	MLVBM_Q7SVK7_3mut
PAPAP	XMRV6_A1Z651_3mut
GSSPAPGGS	MLVBM_Q7SVK7_3mutA_WS
GSSEAAAKGGG	WMSV_P03359_3mutA
EAAAKGGGGGS	PERV_Q4VFZ2_3mut
GSSGSSGSSGSSGSS	MLVCB_P08361_3mutA
EAAAKGGGGSS	PERV_Q4VFZ2_3mut
EAAAKGSS	PERV_Q4VFZ2_3mut
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	AVIRE_P03360_3mutA
EAAAKGGS	MLVCB_P08361_3mut
GSSGGSEAAAK	MLVAV_P03356_3mutA
EAAAKPAPGGS	PERV_Q4VFZ2_3mut

[0881]

接头氨基酸序列	RT结构域名称
GGSGGS	MLVAV_P03356_3mutA
EAAAKGSSGGG	AVIRE_P03360_3mutA
GGSGGSGGSGGS	PERV_Q4VFZ2_3mut
GGGGGGGG	KORV_Q9TTC1_3mutA
GGSGSSEAAAK	MLVCB_P08361_3mutA
EAAAKGGG	MLVBM_Q7SVK7_3mutA_WS
GGGGSGGGGSGGGGS	MLVCB_P08361_3mut
GGSGGSGGSGGS	PERV_Q4VFZ2_3mutA_WS
PAPAPAPAPAP	WMSV_P03359_3mut
EAAAKEAAAKEAAAKEAAAK	PERV_Q4VFZ2_3mut
GGSGGSGGS	XMRV6_A1Z651_3mutA
PAPGGGGSS	BAEVM_P10272_3mutA
GSSEAAAKGGS	MLVCB_P08361_3mut
GSSGGGPAP	MLVCB_P08361_3mutA
GGSGSS	MLVBM_Q7SVK7_3mutA_WS
GGGGGSEAAAK	MLVAV_P03356_3mutA
GSSEAAAK	PERV_Q4VFZ2_3mutA_WS
GGGGGSGSS	MLVBM_Q7SVK7_3mutA_WS
EAAAKGGSGSS	MLVFF_P26809_3mut
PAP	FLV_P10273_3mutA
GGGGG	MLVMS_P03355_3mutA_WS
EAAAK	PERV_Q4VFZ2_3mut
GSS	FLV_P10273_3mutA
PAPAPAPAPAPAP	KORV_Q9TTC1-Pro_3mutA
EAAAKEAAAKEAAAKEAAAK	MLVCB_P08361_3mut
EAAAKGGGGSEAAAK	XMRV6_A1Z651_3mut

[0882]

接头氨基酸序列	RT结构域名称
PAPGGSGGG	MLVBM_Q7SVK7_3mutA_WS
GGSGGGPAP	WMSV_P03359_3mutA
GGGGSSEAAAK	MLVBM_Q7SVK7_3mutA_WS
PAPGGGGSS	MLVCB_P08361_3mut
GGSGGSGGSGGS	PERV_Q4VFZ2_3mutA_WS
PAPGGSGGG	MLVMS_P03355_3mutA_WS
GSSPAPGGS	MLVCB_P08361_3mutA
GSSGSSGSS	MLVFF_P26809_3mut
PAPGGGGGS	MLVBM_Q7SVK7_3mutA_WS
GSSPAP	PERV_Q4VFZ2_3mut
GGSGGG	KORV_Q9TTC1-Pro_3mut
EAAAKGGGGSEAAAK	PERV_Q4VFZ2_3mutA_WS
GGSPAPEAAAK	PERV_Q4VFZ2_3mutA_WS
EAAAKPAP	BAEVM_P10272_3mut
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	MLVMS_P03355_3mut
EAAAKGGGGSS	MLVFF_P26809_3mut
EAAAKEAAAK	MLVCB_P08361_3mut
GSSEAAAKGGS	PERV_Q4VFZ2_3mut
GGSPAP	KORV_Q9TTC1-Pro_3mutA
EAAAKEAAAKEAAAKEAAAK	MLVMS_P03355_3mutA_WS
GSSGSSGSSGSSGSS	BAEVM_P10272_3mut
PAPEAAAK	MLVMS_P03355_3mut
GSSGGSPAP	PERV_Q4VFZ2
GGGPAPGGS	BAEVM_P10272_3mutA
EAAAKPAPGGS	MLVMS_P03355_PLV919
GGGGSGGGGS	PERV_Q4VFZ2

[0883]

接头氨基酸序列	RT结构域名称
GGGEAAAK	KORV_Q9TTC1-Pro_3mut
EAAAKGGGGGS	FLV_P10273_3mutA
GGSPAPGSS	MLVMS_P03355_3mut
GSSPAPEAAAK	MLVMS_P03355_3mutA_WS
GSAGSAAGSGEF	MLVBM_Q7SVK7_3mutA_WS
EAAAK	BAEVM_P10272_3mutA
EAAAKGGGGSS	BAEVM_P10272_3mutA
GGG	WMSV_P03359_3mut
GGSGSSPAP	BAEVM_P10272_3mut
GGSEAAAKPAP	MLVBM_Q7SVK7_3mutA_WS
EAAAKGGSGSS	MLVCB_P08361_3mut
PAPGSS	MLVAV_P03356_3mutA
PAPEAAAKGGG	MLVCB_P08361_3mutA
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	FOAMV_P14350-Pro_2mut
GSSGSSGSS	PERV_Q4VFZ2_3mut
PAPGGG	MLVMS_P03355_3mut
PAPGGS	PERV_Q4VFZ2_3mut
GSSGGG	MLVMS_P03355_PLV919
GSSGSSGSSGSSGSSGSS	WMSV_P03359_3mut
PAP	AVIRE_P03360_3mutA
EAAAKGSSPAP	MLVBM_Q7SVK7_3mutA_WS
GSSGSSGSSGSS	MLVMS_P03355_PLV919
GGGGSGGGGSGGGGSGGGGSGGGGS	AVIRE_P03360
GGGGS	PERV_Q4VFZ2_3mut
EAAAKGSSGGG	MLVBM_Q7SVK7_3mutA_W

[0884]

接头氨基酸序列	RT 结构域名称
	S
GGGGGG	KORV_Q9TTC1-Pro_3mut
GGSGSSEAAAK	PERV_Q4VFZ2_3mut
GSSPAPEAAAK	MLVBM_Q7SVK7_3mutA_WS
GGGGSGGGGS	MLVBM_Q7SVK7_3mutA_WS
GSSGGGGGS	MLVAV_P03356_3mutA
GSAGSAAGSGEF	WMSV_P03359_3mutA
GGGEAAAKGSS	BAEVM_P10272_3mutA
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	FFV_O93209-Pro_2mut
PAPGGSGGG	MLVCB_P08361_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAK	SFV3L_P27401_2mut
GGSGSSPAP	MLVMS_P03355_PLV919
GGGGGG	PERV_Q4VFZ2_3mut
EAAAKEAAAKEAAAKEAAAKEAAAK	PERV_Q4VFZ2_3mut
EAAAKGSSPAP	MLVFF_P26809_3mut
GGGPAPGGS	MLVBM_Q7SVK7_3mutA_WS
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	SFV3L_P27401
PAP	PERV_Q4VFZ2_3mut
EAAAKGGS	MLVMS_P03355_PLV919
GSSGGSEAAAK	WMSV_P03359_3mutA
GGSGSSEAAAK	KORV_Q9TTC1-Pro_3mutA
EAAAKEAAAKEAAAK	PERV_Q4VFZ2
GGSGGGEAAAK	MLVMS_P03355_3mutA_WS
GGGGSGGGGSGGGGSGGGGS	BAEVM_P10272_3mut

[0885]

接头氨基酸序列	RT结构域名称
EAAAKGSS	XMRV6_A1Z651_3mutA
GSSGGGGGS	WMSV_P03359_3mutA
GSSGSSGSSGSSGSSGSS	MLVFF_P26809_3mutA
GGSGSS	MLVAV_P03356_3mutA
EAAAKGGGGSEAAAK	MLVMS_P03355_PLV919
EAAAKGGGPAP	PERV_Q4VFZ2
GGSEAAAKGGG	MLVAV_P03356_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	MLVBM_Q7SVK7_3mut
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	KORV_Q9TTC1-Pro_3mutA
GSSPAPEAAAK	MLVFF_P26809_3mutA
GGGGSEAAAKGGGGS	PERV_Q4VFZ2_3mut
GSSGSSGSSGSS	PERV_Q4VFZ2_3mut
GGSEAAAK	MLVFF_P26809_3mutA
GGGGGGGG	MLVMS_P03355_3mut
GSSGGG	XMRV6_A1Z651_3mutA
EAAAKGGS	BAEVM_P10272_3mutA
GGGGS	BAEVM_P10272_3mutA
GGSEAAAKGGG	KORV_Q9TTC1-Pro_3mutA
GGSGSSGGG	KORV_Q9TTC1_3mutA
GGSGSSEAAAK	WMSV_P03359_3mut
EAAAKGGSGSS	MLVBM_Q7SVK7_3mutA_WS
GGS	BAEVM_P10272_3mutA
GGGPAPGSS	WMSV_P03359_3mutA
GSSGSSGSSGSSGSS	AVIRE_P03360_3mut
GGGEAAAKPAP	XMRV6_A1Z651_3mut
GSSGGG	MLVFF_P26809_3mutA
GGSPAPGSS	PERV_Q4VFZ2_3mut
PAPGGS	MLVCB_P08361_3mut

[0886]

接头氨基酸序列	RT结构域名称
PAPAPAPAPAP	KORV_Q9TTC1_3mutA
GSSGGS	MLVCB_P08361_3mutA
GSSGGSEAAAK	PERV_Q4VFZ2_3mut
EAAAKGSSGGS	MLVMS_P03355_PLV919
EAAAKGGG	WMSV_P03359_3mut
PAPGGGGGS	BAEVM_P10272_3mutA
GGGGSEAAAKGGGGS	WMSV_P03359_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	MLVMS_P03355_3mutA_WS
GGS	KORV_Q9TTC1-Pro_3mutA
GSSGGSPAP	BAEVM_P10272_3mutA
GGG	MLVMS_P03355_PLV919
PAPGSS	KORV_Q9TTC1-Pro_3mut
GGSEAAAKGGG	FLV_P10273_3mutA
GGSEAAAKPAP	PERV_Q4VFZ2_3mutA_WS
GGGGSSPAP	XMRV6_A1Z651_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAK	PERV_Q4VFZ2_3mutA_WS
GGGG	PERV_Q4VFZ2_3mutA_WS
GGSEAAAKPAP	MLVMS_P03355_3mut
PAPGSSGGG	MLVMS_P03355_3mutA_WS
PAPEAAAKGGS	AVIRE_P03360_3mut
GGGGSSPAP	MLVMS_P03355_3mutA_WS
GGGGSGGGGSGGGGSGGGGS	PERV_Q4VFZ2_3mut
GGGEAAAK	MLVMS_P03355_3mut
GGGGSS	MLVFF_P26809_3mut
GGSPAPGSS	XMRV6_A1Z651_3mut
GGGGS	KORV_Q9TTC1-Pro_3mutA
EAAAKGSSGGS	FLV_P10273_3mutA
GSS	MLVMS_P03355_PLV919
GGGG	MLVMS_P03355_PLV919

[0887]

接头氨基酸序列	RT结构域名称
GSSGGS	MLVMS_P03355_PLV919
GGSGGSGGSGGS	MLVMS_P03355_3mut
PAPEAAAKGGS	MLVMS_P03355_3mut
EAAAKGSSGGG	BAEVM_P10272_3mutA
GSSEAAAK	KORV_Q9TTC1-Pro_3mutA
GSAGSAAGSGEF	KORV_Q9TTC1_3mutA
GGGGGSEAAAK	MLVCB_P08361_3mut
GGGG	WMSV_P03359_3mut
GGGGSSEAAAK	MLVMS_P03355_PLV919
PAPGGG	WMSV_P03359_3mutA
EAAAKGGSGGG	MLVAV_P03356_3mutA
GGGPAPGGS	MLVMS_P03355_3mut
EAAAKPAP	PERV_Q4VFZ2_3mutA_WS
GSSGSSGSS	KORV_Q9TTC1-Pro_3mutA
GSSPAPGGS	XMRV6_A1Z651_3mut
GGGGGSPAP	BAEVM_P10272_3mutA
GGSGSSGGG	PERV_Q4VFZ2_3mutA_WS
GGGEAAAKGSS	AVIRE_P03360_3mut
GSSEAAAK	FLV_P10273_3mutA
EAAAK	MLVMS_P03355_3mut
EAAAKGGSGSS	WMSV_P03359_3mut
GSSEAAAKGGG	PERV_Q4VFZ2_3mut
PAPGSSGGG	BAEVM_P10272_3mutA
EAAAKGGGGGS	MLVMS_P03355_3mut
GGSEAAAKPAP	AVIRE_P03360_3mut
GGGPAPGGS	XMRV6_A1Z651_3mut
GGGGS	KORV_Q9TTC1_3mutA
GGSGGSGGSGGSGGS	XMRV6_A1Z651_3mut
GGGPAP	KORV_Q9TTC1-Pro_3mut

[0888]

接头氨基酸序列	RT结构域名称
EAAAKPAP	MLVBM_Q7SVK7_3mutA_WS
GGSEAAAK	MLVMS_P03355_PLV919
GSSEAAAKPAP	KORV_Q9TTC1-Pro_3mutA
GGSGSS	MLVMS_P03355_3mut
EAAAKPAPGGG	PERV_Q4VFZ2_3mut
GGSPAPEAAAK	KORV_Q9TTC1_3mutA
GGSEAAAKGGG	AVIRE_P03360_3mutA
GGGGSEAAAKGGGGS	MLVMS_P03355_PLV919
GSSGGGEAAAK	KORV_Q9TTC1-Pro_3mutA
EAAAKGGGPAP	WMSV_P03359_3mut
GSSPAP	XMRV6_A1Z651_3mutA
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	SFV3L_P27401-Pro
GGSEAAAKGSS	MLVMS_P03355_PLV919
GSSGGSEAAAK	KORV_Q9TTC1-Pro_3mutA
GGSEAAAKGSS	KORV_Q9TTC1-Pro_3mutA
EAAAKGGG	AVIRE_P03360_3mutA
GSSGGSEAAAK	BAEVM_P10272_3mutA
GGGGSEAAAKGGGGS	KORV_Q9TTC1-Pro_3mut
PAPGSSEAAAK	MLVMS_P03355_3mut
PAPEAAAK	WMSV_P03359_3mut
PAPGGSGSS	PERV_Q4VFZ2_3mutA_WS
PAPGSS	BAEVM_P10272_3mut
PAPGGGGGS	MLVMS_P03355_3mut
EAAAKPAPGSS	MLVBM_Q7SVK7_3mutA_WS
GSSPAPGGS	MLVMS_P03355_PLV919
GGSGSSEAAAK	MLVMS_P03355_3mut

[0889]

接头氨基酸序列	RT 结构域名称
GGGGGG	KORV_Q9TTC1-Pro_3mutA
EAAAKEAAAKEAAAKEAAAK	MLVBM_Q7SVK7_3mut
GGSPAPGSS	MLVMS_P03355_PLV919
PAPAPAPAPAP	MLVCB_P08361_3mut
GGSGSSPAP	WMSV_P03359_3mutA
EAAAKGGSGGG	PERV_Q4VFZ2_3mutA_WS
GSSGSSGSSGSSGSS	PERV_Q4VFZ2_3mut
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	KORV_Q9TTC1_3mutA
GSSGGGEAAAK	WMSV_P03359_3mutA
GSSGGSEAAAK	FLV_P10273_3mutA
GGGGGGGG	PERV_Q4VFZ2_3mut
PAPGGSEAAAK	FLV_P10273_3mutA
GGGGSSPAP	BAEVM_P10272_3mutA
PAPAPAPAP	WMSV_P03359_3mut
GGSEAAAKPAP	PERV_Q4VFZ2_3mut
PAPGGSGGG	BAEVM_P10272_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	MLVMS_P03355_3mut
GGGGSGGGGSGGGGS	PERV_Q4VFZ2_3mut
GGSGGGPAP	PERV_Q4VFZ2_3mut
GGGPAPEAAAK	MLVFF_P26809_3mut
GGGGGSGSS	MLVMS_P03355_3mutA_WS
GSS	MLVCB_P08361_3mut
GGGGGSPAP	MLVMS_P03355_PLV919
GGSPAP	MLVAV_P03356_3mutA
GGGPAPGGS	KORV_Q9TTC1-Pro_3mutA
PAPGSSGGG	FLV_P10273_3mutA
PAPGSSGGG	WMSV_P03359_3mutA
PAPGGS	MLVBM_Q7SVK7_3mutA_W

[0890]

接头氨基酸序列	RT 结构域名称
	S
GGGEAAAKGSS	PERV_Q4VFZ2_3mutA_WS
GGSEAAAKGSS	MLVBM_Q7SVK7_3mutA_WS
PAPGGSEAAAK	MLVCB_P08361_3mut
GGSEAAAKGGG	XMRV6_A1Z651_3mutA
GGSGGGGSS	WMSV_P03359_3mut
GGGEAAAKPAP	KORV_Q9TTC1_3mutA
EAAAKGSS	KORV_Q9TTC1-Pro_3mut
PAPEAAAKGSS	MLVFF_P26809_3mut
GSAGSAAGSGEF	PERV_Q4VFZ2_3mut
EAAAKGGGGGS	WMSV_P03359_3mut
EAAAKGSSPAP	WMSV_P03359_3mutA
GGGGSEAAAKGGGGS	XMRV6_A1Z651_3mutA
GSSEAAAKPAP	SFV3L_P27401-Pro_2mutA
GGGGGG	PERV_Q4VFZ2_3mutA_WS
PAPGGS	BAEVM_P10272_3mut
PAP	AVIRE_P03360_3mut
PAPAPAP	MLVBM_Q7SVK7_3mutA_WS
GGGG	PERV_Q4VFZ2_3mutA_WS
GSSGGSEAAAK	MLVBM_Q7SVK7_3mut
GGSGGGGSS	MLVFF_P26809_3mut
GGGGSSGGS	AVIRE_P03360_3mutA
GSSPAPGGG	PERV_Q4VFZ2_3mutA_WS
GGSEAAAKPAP	MLVMS_P03355_PLV919
PAP	KORV_Q9TTC1-Pro_3mut
GSSGGS	PERV_Q4VFZ2_3mut
GGGGG	PERV_Q4VFZ2_3mut

[0891]

接头氨基酸序列	RT结构域名称
GSSGGGPAP	FLV_P10273_3mutA
GSSEAAAKGGG	KORV_Q9TTC1-Pro_3mut
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	MLVCB_P08361_3mut
GGSEAAAKPAP	MLVCB_P08361_3mut
PAPAPAPAPAPAP	BAEVM_P10272_3mutA
GGGGSEAAAKGGGGS	MLVMS_P03355_3mut
EAAAKPAPGSS	MLVMS_P03355_3mut
GSSGSSGSSGSSGSS	MLVBM_Q7SVK7_3mutA_WS
PAPEAAAKGSS	MLVAV_P03356_3mut
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	AVIRE_P03360_3mut
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	PERV_Q4VFZ2_3mut
GGSEAAAKGGG	PERV_Q4VFZ2_3mutA_WS
GGSGGGGSS	MLVFF_P26809_3mutA
PAPEAAAKGSS	MLVCB_P08361_3mut
GGG	PERV_Q4VFZ2_3mutA_WS
GGSGGGEAAAK	MLVMS_P03355_3mut
EAAAKGGGGSS	WMSV_P03359_3mut
GSSPAPGGG	WMSV_P03359_3mutA
EAAAKGSSGGG	PERV_Q4VFZ2_3mut
GGSGGGEAAAK	PERV_Q4VFZ2_3mutA_WS
GGSGGSGGSGGSGGS	PERV_Q4VFZ2_3mutA_WS
EAAAKPAPGGS	PERV_Q4VFZ2_3mutA_WS
GGGGGSEAAAK	PERV_Q4VFZ2_3mutA_WS
GSSPAP	MLVFF_P26809_3mut
GGGEAAAKPAP	AVIRE_P03360_3mut
GSSGGSEAAAK	MLVMS_P03355_PLV919

[0892]

接头氨基酸序列	RT结构域名称
EAAAKPAPGGS	WMSV_P03359_3mutA
PAPGGG	KORV_Q9TTC1_3mutA
EAAAKGSSPAP	KORV_Q9TTC1-Pro_3mut
GSSPAPEAAAK	MLVFF_P26809_3mut
GGSGGGEAAAK	MLVFF_P26809_3mutA
GSSGSSGSS	WMSV_P03359_3mutA
EAAAKGGS	BAEVM_P10272_3mut
EAAAKPAPGGS	KORV_Q9TTC1_3mutA
EAAAKPAPGGS	BAEVM_P10272_3mutA
GSSGGGGGS	PERV_Q4VFZ2_3mut
PAPGGGGSS	PERV_Q4VFZ2_3mut
GSSGSSGSS	WMSV_P03359_3mut
EAAAKEAAAKEAAAKEAAAK	WMSV_P03359_3mut
GGS	AVIRE_P03360_3mut
EAAAKPAPGSS	MLVFF_P26809_3mut
EAAAKGGG	KORV_Q9TTC1_3mut
PAPGSSEAAAK	MLVMS_P03355_3mut
PAPGSSGGS	MLVMS_P03355_PLV919
GSSPAPEAAAK	MLVMS_P03355_3mut
GSSGSSGSSGSSGSSGSS	WMSV_P03359_3mutA
GGGGS	BAEVM_P10272_3mut
GSSPAP	MLVMS_P03355_3mut
EAAAKGGGGSEAAAK	KORV_Q9TTC1-Pro_3mutA
EAAAKEAAAK	WMSV_P03359_3mutA
GGGGSSGGS	MLVCB_P08361_3mutA
PAPGGSEAAAK	BAEVM_P10272_3mut
EAAAKGGSPAP	MLVFF_P26809_3mut
GSSGGSGGG	MLVBM_Q7SVK7_3mutA_WS

[0893]

接头氨基酸序列	RT 结构域名称
GSSGGS	PERV_Q4VFZ2_3mut
PAPGGSGSS	PERV_Q4VFZ2_3mutA_WS
EAAAKGGSGSS	KORV_Q9TTC1-Pro_3mutA
PAPAP	MLVCB_P08361_3mut
EAAAKGSSPAP	PERV_Q4VFZ2_3mutA_WS
EAAAKPAPGGG	MLVMS_P03355_PLV919
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	MLVBM_Q7SVK7_3mut
EAAAKGGGGSS	MLVMS_P03355_PLV919
PAPEAAAK	PERV_Q4VFZ2_3mut
EAAAKPAPGSS	BAEVM_P10272_3mutA
GGSPAP	PERV_Q4VFZ2_3mutA_WS
GGSGGS	BAEVM_P10272_3mutA
PAPEAAAKGSS	KORV_Q9TTC1_3mut
PAPGSS	MLVMS_P03355_PLV919
PAPAPAPAPAP	MLVAV_P03356_3mutA
GGG	XMRV6_A1Z651_3mutA
GGGPAP	PERV_Q4VFZ2_3mutA_WS
GSSPAPEAAAK	KORV_Q9TTC1_3mutA
PAP	BAEVM_P10272_3mutA
GGSPAP	BAEVM_P10272_3mutA
PAPEAAAKGGS	MLVMS_P03355_PLV919
PAPGSSGGS	PERV_Q4VFZ2_3mutA_WS
PAPAPAPAPAPAP	PERV_Q4VFZ2_3mut
EAAAKEAAAKEAAAK	MLVCB_P08361_3mut
GGSGGSGGSGGSGGS	MLVMS_P03355_PLV919
EAAAKPAPGGS	MLVMS_P03355_3mut
GGSGGS	MLVMS_P03355_PLV919
EAAAKPAP	MLVMS_P03355_3mutA_WS
GGSEAAAK	XMRV6_A1Z651_3mutA

[0894]

接头氨基酸序列	RT结构域名称
GGSGGG	KORV_Q9TTC1_3mut
GGSGGGEAAAK	PERV_Q4VFZ2_3mut
PAPEAAAKGGG	AVIRE_P03360
PAPAP	PERV_Q4VFZ2_3mut
GSS	KORV_Q9TTC1-Pro_3mutA
EAAAKGSSGGG	MLVAV_P03356_3mutA
GGSPAPGSS	MLVBM_Q7SVK7_3mutA_WS
PAPEAAAK	MLVAV_P03356_3mut
EAAAKGGSPAP	BAEVM_P10272_3mutA
PAPAPAPAP	WMSV_P03359_3mutA
PAPGGSEAAAK	MLVMS_P03355_3mut
GGSGGSGGSGGS	WMSV_P03359_3mut
GGGGGSGSS	XMRV6_A1Z651_3mut
PAPGGSGGG	KORV_Q9TTC1_3mutA
GGS	MLVMS_P03355_3mut
EAAAK	WMSV_P03359_3mut
GGGEAAAKGSS	MLVBM_Q7SVK7_3mutA_WS
GGSPAPGSS	MLVCB_P08361_3mut
GGSEAAAKPAP	PERV_Q4VFZ2_3mut
GGGGSGGGGSGGGGSGGGGSGGGGS	MLVCB_P08361_3mutA
GGSGSS	BAEVM_P10272_3mutA
GGGEAAAKGSS	WMSV_P03359_3mutA
EAAAKGGSPAP	WMSV_P03359_3mut
GSSPAPEAAAK	MLVMS_P03355_3mut
GGSGGSGGSGGS	MLVMS_P03355_PLV919
GSSPAPEAAAK	WMSV_P03359_3mut
GSSGSSGSSGSS	PERV_Q4VFZ2

[0895]

接头氨基酸序列	RT结构域名称
GGSGSSEAAAK	WMSV_P03359_3mutA
GGSGGG	MLVFF_P26809_3mut
GGSPAPGGG	MLVFF_P26809_3mut
GGSGGSGGS	BAEVM_P10272_3mutA
GGGGSSEAAAK	MLVBM_Q7SVK7_3mut
GGSPAPGSS	MLVMS_P03355_3mut
EAAAKPAPGSS	AVIRE_P03360_3mut
GGGGSSGGS	FLV_P10273_3mutA
GGSPAPEAAAK	PERV_Q4VFZ2_3mut
GGSEAAAK	MLVMS_P03355_3mutA_WS
GSSGSSGSSGSS	MLVCB_P08361_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAK	MLVMS_P03355_PLV919
GGGGG	PERV_Q4VFZ2_3mut
GGSEAAAKGSS	MLVCB_P08361_3mutA
GSSGGG	MLVBM_Q7SVK7_3mutA_WS
PAPGSSGGG	KORV_Q9TTC1-Pro_3mutA
GGSGGS	BAEVM_P10272_3mut
EAAAKGGGGGS	MLVBM_Q7SVK7_3mutA_WS
GGSGSSPAP	MLVCB_P08361_3mut
PAPGSSGGG	KORV_Q9TTC1
PAPGGSGGG	MLVMS_P03355_3mut
GGGG	WMSV_P03359_3mutA
EAAAKGGSPAP	MLVCB_P08361_3mut
GSSGSS	FLV_P10273_3mutA
GGSEAAAKPAP	SFV3L_P27401_2mut
EAAAKGSSGGS	MLVAV_P03356_3mutA
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAA	MLVAV_P03356_3mutA

[0896]

接头氨基酸序列	RT 结构域名称
AKA	
EAAAKGGSGSS	PERV_Q4VFZ2_3mutA_WS
GGGGG	MLVCB_P08361_3mut
GGGEAAAK	BAEVM_P10272_3mut
GGSGGSGGSGGS	MLVCB_P08361_3mut
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	PERV_Q4VFZ2
PAPAPAPAPAP	MLVMS_P03355_3mutA_WS
EAAAKEAAAK	XMRV6_A1Z651_3mut
GSSGGSEAAAK	PERV_Q4VFZ2_3mutA_WS
PAPGGSEAAAK	KORV_Q9TTC1-Pro_3mutA
EAAAKGGGPAP	MLVBM_Q7SVK7_3mutA_WS
PAPGGSGSS	PERV_Q4VFZ2
SGSETPGTSESATPES	MLVMS_P03355_3mut
GGSGGS	MLVMS_P03355_PLV919
EAAAKGGS	FLV_P10273_3mut
GGSPAPGSS	MLVMS_P03355_3mutA_WS
EAAAKEAAAKEAAAKEAAAK	FFV_O93209_2mut
GSSGGSGGG	MLVMS_P03355_3mutA_WS
PAPGSSEAAAK	WMSV_P03359_3mut
PAPAPAPAPAPAP	KORV_Q9TTC1_3mutA
GGGGSS	BAEVM_P10272_3mut
GGGGSEAAAKGGGGS	AVIRE_P03360_3mut
GSSPAPEAAAK	KORV_Q9TTC1-Pro_3mutA
PAPEAAAKGGG	MLVBM_Q7SVK7_3mut
EAAAKEAAAK	WMSV_P03359_3mut
EAAAK	SFV3L_P27401-Pro_2mutA
GSSGGSGGG	XMRV6_A1Z651_3mutA
GGGEAAAKPAP	WMSV_P03359_3mutA

[0897]

接头氨基酸序列	RT结构域名称
GGSGGS	MLVFF_P26809_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	FOAMV_P14350_2mutA
GGGGG	MLVAV_P03356_3mutA
GSSGGSEAAAK	BAEVM_P10272_3mut
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	SFV1_P23074
GGSGGGPAP	MLVCB_P08361_3mut
GGSGSS	PERV_Q4VFZ2_3mut
SGSETPGTSESATPES	MLVFF_P26809_3mut
EAAAKGGSPAP	MLVMS_P03355_3mut
PAPAP	PERV_Q4VFZ2_3mut
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	MLVBM_Q7SVK7_3mut
GGGGGS	BAEVM_P10272_3mutA
EAAAKEAAAK	AVIRE_P03360_3mut
GSSGGSEAAAK	PERV_Q4VFZ2_3mut
GGGEAAAK	WMSV_P03359_3mut
GSSGGGEAAAK	AVIRE_P03360_3mutA
GGG	XMRV6_A1Z651_3mut
GGGGSEAAAKGGGGS	BAEVM_P10272_3mut
GGGG	MLVMS_P03355_3mut
GGSGGS	MLVMS_P03355_3mutA_WS
GGSGGGGSS	MLVBM_Q7SVK7_3mutA_WS
GSSPAPGGS	PERV_Q4VFZ2_3mut
GSSPAPEAAAK	PERV_Q4VFZ2_3mutA_WS
EAAAKGGS	WMSV_P03359_3mut
GGSGGSGGSGGS	PERV_Q4VFZ2_3mut
GGGGSSEAAAK	KORV_Q9TTC1-Pro_3mut
PAPAPAPAPAPAP	MLVAV_P03356_3mut

[0898]

接头氨基酸序列	RT 结构域名称
EAAAKGSSGGG	MLVMS_P03355_PLV919
GGGGG	MLVBM_Q7SVK7_3mutA_WS
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	FFV_O93209_2mutA
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	KORV_Q9TTC1-Pro_3mut
GGSPAPGGG	MLVMS_P03355_3mutA_WS
GGGEAAAKGGS	MLVMS_P03355_3mut
GGGEAAAK	PERV_Q4VFZ2_3mut
PAPEAAAKGGG	MLVMS_P03355_3mut
GSSGSSGSSGSSGSSGSS	BAEVM_P10272_3mutA
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	GALV_P21414_3mutA
EAAAKGGSPAP	FFV_O93209-Pro
EAAAKEAAAK	MLVFF_P26809_3mut
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	PERV_Q4VFZ2_3mutA_WS
GGSGGSGGSGGS	MLVAV_P03356_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAK	SFV3L_P27401_2mutA
GSSGSSGSSGSSGSSGSS	BAEVM_P10272_3mut
GGGGS	MLVMS_P03355_PLV919
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	SFV1_P23074
GGGGSGGGGS	KORV_Q9TTC1-Pro_3mutA
GGGGSGGGGS	MLVMS_P03355_3mut
GGSGSS	KORV_Q9TTC1_3mutA
GSSPAPGGG	PERV_Q4VFZ2_3mut
GSSGGSPAP	PERV_Q4VFZ2_3mutA_WS
PAPGGS	PERV_Q4VFZ2_3mutA_WS
GGSPAPEAAAK	FOAMV_P14350_2mutA

[0899]

接头氨基酸序列	RT 结构域名称
GGGPAPGGS	SFV3L_P27401_2mut
PAPGSSGGG	MLVCB_P08361_3mut
GSSGGGEAAAK	AVIRE_P03360_3mut
GSSGGG	XMRV6_A1Z651_3mut
GSSGSS	PERV_Q4VFZ2_3mut
GSSGGG	MLVAV_P03356_3mutA
PAPGGGGGS	PERV_Q4VFZ2_3mut
GSSEAAAK	MLVMS_P03355_3mut
PAPGGG	FLV_P10273_3mutA
GGGGSGGGGS	PERV_Q4VFZ2_3mut
GSSGGS	MLVMS_P03355_PLV919
GGGGSGGGGS	SFV3L_P27401_2mut
EAAAKGGSGSS	FLV_P10273_3mutA
GSSEAAAKGGS	MLVMS_P03355_3mutA_WS
PAPGSSEAAAK	SFV3L_P27401_2mutA
GGGGSGGGGS	SFV3L_P27401-Pro_2mutA
PAPGSSEAAAK	PERV_Q4VFZ2_3mut
PAPGSSEAAAK	PERV_Q4VFZ2
GGSPAPGGG	AVIRE_P03360_3mut
GGGGGS	PERV_Q4VFZ2_3mutA_WS
GGGGSSGGS	PERV_Q4VFZ2_3mut
PAPAPAPAP	AVIRE_P03360_3mutA
GGSGGS	WMSV_P03359_3mutA
GGGPAPGGS	PERV_Q4VFZ2_3mut
GGSGGSGGSGGSGGS	MLVMS_P03355_PLV919
GGSGGG	PERV_Q4VFZ2_3mut
EAAAKEAAAK	SFV3L_P27401_2mut
PAPGSS	XMRV6_A1Z651_3mut
GSSEAAAK	MLVFF_P26809_3mut

[0900]

接头氨基酸序列	RT结构域名称
GGSPAPGGG	MLVMS_P03355_3mut
EAAAKGGG	WMSV_P03359_3mutA
GSSEAAAKGGS	PERV_Q4VFZ2_3mutA_WS
GSSGGSPAP	FFV_O93209
GGGGGS	KORV_Q9TTC1-Pro_3mut
GSSGGG	MLVCB_P08361_3mut
GSSGSS	MLVCB_P08361_3mutA
GGSEAAAKPAP	BAEVM_P10272_3mut
EAAAKGGGGSS	MLVCB_P08361_3mut
EAAAKPAPGGS	KORV_Q9TTC1-Pro_3mutA
GSSGSSGSSGSSGSS	MLVAV_P03356_3mutA
GGGGSEAAAKGGGGS	PERV_Q4VFZ2_3mutA_WS
GGSGSS	KORV_Q9TTC1-Pro_3mut
GSS	SFV3L_P27401-Pro_2mutA
PAPAP	BAEVM_P10272_3mut
EAAAKPAP	BAEVM_P10272
EAAAKEAAAKEAAAKEAAAKEAAAK	KORV_Q9TTC1-Pro_3mut
GGGGGGG	PERV_Q4VFZ2_3mutA_WS
GGGGS	MLVMS_P03355_3mut
GSSGGG	FLV_P10273_3mutA
PAPAPAPAPAP	FLV_P10273_3mut
EAAAKEAAAKEAAAK	WMSV_P03359_3mutA
GSSGGS	MLVBM_Q7SVK7_3mutA_WS
EAAAKPAPGGG	MLVMS_P03355_3mut
GSSPAPGGS	WMSV_P03359_3mut
PAPGSSGGG	PERV_Q4VFZ2_3mutA_WS
GSSGGG	AVIRE_P03360_3mutA
PAPGGSGSS	MLVFF_P26809_3mut

[0901]

接头氨基酸序列	RT结构域名称
PAPGSS	PERV_Q4VFZ2_3mut
GGGGGSGSS	WMSV_P03359_3mutA
EAAAKGGGGSS	MLVBM_Q7SVK7_3mutA_WS
GGGGGGG	BAEVM_P10272_3mut
PAPEAAAKGSS	MLVMS_P03355_3mut
GGSGGGEAAAK	MLVMS_P03355_PLV919
EAAAKGGGGGS	MLVCB_P08361_3mut
PAPGGS	KORV_Q9TTC1-Pro_3mut
GGGG	FLV_P10273_3mutA
EAAAKGGSGSS	MLVBM_Q7SVK7_3mutA_WS
GGGGSSGGS	MLVMS_P03355_3mutA_WS
GGGGGGGG	WMSV_P03359_3mut
GGSGSSGGG	MLVMS_P03355_PLV919
GSSEAAAKGGS	KORV_Q9TTC1-Pro_3mutA
EAAAKPAPGSS	MLVCB_P08361_3mut
GGSPAPGSS	KORV_Q9TTC1_3mutA
PAPGSSGGG	BAEVM_P10272_3mut
EAAAKPAPGSS	WMSV_P03359_3mut
GGSPAPEAAAK	XMRV6_A1Z651_3mutA
GSSPAP	FLV_P10273_3mutA
GSS	BAEVM_P10272_3mutA
EAAAKPAPGGS	FLV_P10273_3mutA
GGSGSSPAP	FLV_P10273_3mutA
PAPGSSGGS	MLVMS_P03355_3mut
GSAGSAAGSGEF	PERV_Q4VFZ2_3mutA_WS
GSSGGSEAAAK	KORV_Q9TTC1_3mutA
GSSGGS	MLVMS_P03355_3mutA_WS

[0902]

接头氨基酸序列	RT结构域名称
EAAAKGGGGSEAAAK	SFV3L_P27401_2mut
GSSGGS	PERV_Q4VFZ2_3mutA_WS
GGSPAPEAAAK	FLV_P10273_3mut
GGSEAAAKGSS	PERV_Q4VFZ2_3mutA_WS
GSSPAPEAAAK	PERV_Q4VFZ2_3mutA_WS
GGSGSSGGG	PERV_Q4VFZ2_3mut
GGGG	AVIRE_P03360_3mutA
GGSEAAAKPAP	WMSV_P03359_3mut
GSSGGSPAP	MLVAV_P03356_3mutA
GSSGGSEAAAK	MLVMS_P03355_3mut
PAPEAAAKGGS	KORV_Q9TTC1-Pro_3mut
GGSPAP	PERV_Q4VFZ2_3mutA_WS
GGSEAAAK	MLVAV_P03356_3mutA
EAAAKGGGGSEAAAK	KORV_Q9TTC1-Pro_3mut
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	MLVMS_P03355_PLV919
GSSEAAAK	KORV_Q9TTC1_3mutA
GGG	AVIRE_P03360
GGSEAAAKGSS	MLVBM_Q7SVK7_3mut
GGSEAAAKGSS	MLVMS_P03355_3mut
GGSPAPEAAAK	MLVCB_P08361_3mut
GGSGGGEAAAK	MLVCB_P08361_3mut
GGSEAAAKPAP	MLVMS_P03355_3mutA_WS
EAAAKGGSGSS	KORV_Q9TTC1-Pro_3mut
GGGEAAAKGGS	MLVCB_P08361_3mut
EAAAKGGGGSEAAAK	FLV_P10273_3mutA
GGSPAP	MLVFF_P26809_3mut
GGGGSSGGS	XMRV6_A1Z651_3mutA
PAP	MLVCB_P08361_3mut
GGS	SFV3L_P27401-Pro_2mutA

[0903]

接头氨基酸序列	RT结构域名称
GGGGSGGGGS	MLVMS_P03355_3mut
GGGEAAAKGGS	MLVAV_P03356_3mutA
GSSGSSGSSGSSGSSGSS	MLVMS_P03355_PLV919
PAPGSS	MLVCB_P08361_3mut
GGSGGSGGS	MLVMS_P03355_PLV919
PAPGGSGGG	FLV_P10273_3mutA
GGGGSGGGGSGGGGS	FLV_P10273_3mut
GGSGSSGGG	KORV_Q9TTC1-Pro_3mutA
GGSGGSGGS	GALV_P21414_3mutA
GGGEAAAKGGS	WMSV_P03359_3mut
SGSETPGTSESATPES	KORV_Q9TTC1_3mutA
EAAAKGGGGGS	KORV_Q9TTC1-Pro_3mut
EAAAKGSSPAP	BAEVM_P10272_3mut
GGGG	MLVCB_P08361_3mut
GGGGSGGGGSGGGGSGGGGSGGGGS	MLVBM_Q7SVK7_3mut
GSSGGSGGG	MLVMS_P03355_PLV919
GGSGSS	MLVFF_P26809_3mut
EAAAKGGS	AVIRE_P03360_3mutA
GSSEAAAKGGS	MLVBM_Q7SVK7_3mutA_WS
EAAAKPAPGGG	WMSV_P03359_3mut
PAPGSSGGG	MLVCB_P08361_3mutA
GGGGSSEAAAK	KORV_Q9TTC1-Pro_3mutA
GSSEAAAKPAP	BAEVM_P10272_3mutA
PAPGGGEAAAK	MLVBM_Q7SVK7_3mutA_WS
GGSGGGEAAAK	MLVCB_P08361_3mutA
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	FFV_O93209
EAAAKGGGGGS	GALV_P21414_3mutA

[0904]

接头氨基酸序列	RT结构域名称
GGSPAPGGG	MLVMS_P03355_3mut
GSSGSSGSS	FLV_P10273_3mutA
EAAAK	MLVBM_Q7SVK7_3mut
GGGGSSGGS	MLVMS_P03355_3mut
GGSGSSPAP	PERV_Q4VFZ2_3mut
EAAAKEAAAKEAAAKEAAAK	BAEVM_P10272_3mut
GGGPAPGSS	MLVMS_P03355_3mut
GSSPAPGGS	PERV_Q4VFZ2_3mutA_WS
PAPAP	FLV_P10273_3mutA
PAPAPAPAP	PERV_Q4VFZ2_3mut
GGGGGSEAAAK	GALV_P21414_3mutA
GGGGGSGSS	BAEVM_P10272_3mutA
GGGEAAAKGSS	KORV_Q9TTC1_3mutA
GGGGGSPAP	AVIRE_P03360_3mut
GGGGGSEAAAK	SFV3L_P27401_2mutA
GGS	KORV_Q9TTC1_3mutA
GGGGGGG	PERV_Q4VFZ2_3mut
SGSETPGTSESATPES	SFV3L_P27401_2mutA
EAAAKGGSGGG	MLVMS_P03355_3mut
GGGGS	MLVFF_P26809_3mut
EAAAKGSSGGG	BAEVM_P10272_3mut
EAAAKPAPGGS	MLVF5_P26810_3mutA
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	SFV3L_P27401_2mutA
GGSPAPGGG	WMSV_P03359_3mutA
GSAGSAAGSGEF	MLVFF_P26809_3mut
GGGGSSGGS	MLVMS_P03355_3mutA_WS
GGGGGGG	MLVCB_P08361_3mut
GSSEAAAK	WMSV_P03359_3mut
PAPGSS	FLV_P10273_3mutA

[0905]

接头氨基酸序列	RT结构域名称
GSSGGG	PERV_Q4VFZ2_3mutA_WS
PAPGGG	MLVFF_P26809_3mut
GGGGGSPAP	MLVMS_P03355_3mut
GGSEAAAK	XMRV6_A1Z651_3mut
GSSGGG	PERV_Q4VFZ2_3mut
GGSGGSGGSGGS	MLVMS_P03355_3mut
PAPAP	AVIRE_P03360_3mut
GGSEAAAK	PERV_Q4VFZ2_3mut
GGGGS	MLVMS_P03355_PLV919
GGGG	BAEVM_P10272_3mutA
EAAAKGGGGSS	MLVCB_P08361_3mutA
EAAAKEAAAKEAAAK	GALV_P21414_3mutA
PAPGGGEAAAK	KORV_Q9TTC1
EAAAKGGSPAP	MLVMS_P03355_3mut
GGSGSSEAAAK	MLVMS_P03355_3mut
GGSPAPEAAAK	FLV_P10273_3mutA
GGGGGGG	PERV_Q4VFZ2_3mut
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	SFV1_P23074_2mutA
EAAAKGSSGGS	MLVMS_P03355_3mut
GSSEAAAKPAP	MLVFF_P26809_3mut
GGGGSS	FLV_P10273_3mutA
EAAAKGGSGGG	AVIRE_P03360_3mutA
GGSGGS	PERV_Q4VFZ2_3mutA_WS
GGGGGSPAP	AVIRE_P03360_3mutA
EAAAKEAAAKEAAAK	XMRV6_A1Z651_3mut
PAPEAAAKGGS	FLV_P10273_3mutA
GSSGGSEAAAK	MLVCB_P08361_3mut
EAAAKGGSGGG	MLVMS_P03355
GGSGGGPAP	MLVMS_P03355_3mut

[0906]

接头氨基酸序列	RT 结构域名称
GGS	XMRV6_A1Z651_3mut
GGSEAAAKPAP	MLVFF_P26809_3mut
EAAAKGGG	MLVMS_P03355_PLV919
GSSGSSGSSGSS	WMSV_P03359_3mut
GGSGSSPAP	PERV_Q4VFZ2_3mut
GGGEAAAK	MLVMS_P03355_3mutA_WS
GSSPAPGGS	KORV_Q9TTC1-Pro_3mutA
GSSEAAAKGGG	SFV3L_P27401_2mut
EAAAKPAPGGS	MLVCB_P08361_3mut
GGSGGGEAAAK	PERV_Q4VFZ2
GGSGSS	MLVCB_P08361_3mut
GGSGGGEAAAK	MLVBM_Q7SVK7_3mutA_WS
GGSGGSGGSGGSGGSGGS	FLV_P10273_3mut
PAPEAAAKGSS	MLVMS_P03355_3mut
EAAAKGSSGGS	WMSV_P03359_3mutA
GGSGSSEAAAK	MLVCB_P08361_3mut
GGSGSSEAAAK	KORV_Q9TTC1_3mutA
GSSGGSGGG	MLVMS_P03355_PLV919
EAAAKGGSGGG	SFV3L_P27401-Pro_2mutA
GGSGGS	AVIRE_P03360_3mutA
GSAGSAAGSGEF	MLVMS_P03355_PLV919
GGSGSS	GALV_P21414_3mutA
GGGG	MLVFF_P26809_3mutA
GGGGSGGGGSGGGGSGGGGS	WMSV_P03359_3mut
SGSETPGTSESATPES	BAEVM_P10272_3mut
EAAAKEAAAKEAAAKEAAAK	FOAMV_P14350_2mutA
GGGEAAAKGGS	FLV_P10273_3mutA
GSSGGSEAAAK	MLVFF_P26809_3mut

[0907]

接头氨基酸序列	RT 结构域名称
EAAAKGGGGSS	MLVAV_P03356_3mut
PAPGGSEAAAK	KORV_Q9TTC1-Pro_3mut
EAAAK	XMRV6_A1Z651_3mut
GSSGSSGSSGSSGSSGSS	PERV_Q4VFZ2_3mut
GGGG	MLVCB_P08361_3mutA
GSSGSS	WMSV_P03359_3mutA
GSSGGSPAP	AVIRE_P03360_3mut
GGSGGSGGS	MLVCB_P08361_3mut
EAAAKGGGPAP	FLV_P10273_3mutA
GGGGSGGGGS	MLVCB_P08361_3mut
GGSEAAAKGSS	PERV_Q4VFZ2_3mutA_WS
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	SFV3L_P27401_2mutA
GGSGSSEAAAK	PERV_Q4VFZ2_3mutA_WS
EAAAKEAAAKEAAAKEAAAK	SFV3L_P27401-Pro_2mutA
GSSEAAAKGGS	FLV_P10273_3mutA
GGSGSS	PERV_Q4VFZ2
GGSGSSEAAAK	SFV3L_P27401-Pro_2mutA
GSSGSSGSS	XMRV6_A1Z651_3mutA
EAAAKGSSPAP	KORV_Q9TTC1_3mutA
EAAAKPAP	FLV_P10273_3mutA
GGSGSSEAAAK	KORV_Q9TTC1-Pro_3mut
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	KORV_Q9TTC1_3mutA
GGGGSGGGGSGGGGS	KORV_Q9TTC1-Pro_3mutA
GGGGGGG	FLV_P10273_3mut
EAAAKGSS	WMSV_P03359_3mut
EAAAKGGGPAP	MLVCB_P08361_3mut
GSSGSS	MLVBM_Q7SVK7_3mutA_WS
EAAAKGGGGGS	MLVFF_P26809_3mut

[0908]

接头氨基酸序列	RT结构域名称
GGSGGGEAAAK	FLV_P10273_3mutA
PAPGSS	MLVFF_P26809_3mutA
PAPGSS	BAEVM_P10272_3mutA
GGSPAPGSS	AVIRE_P03360_3mut
GGGGSSEAAAK	MLVMS_P03355_3mut
GSSGGGGGS	FFV_O93209-Pro
EAAAKGSSPAP	PERV_Q4VFZ2_3mut
GSSPAPGGS	PERV_Q4VFZ2_3mut
GGGGGG	BAEVM_P10272_3mut
EAAAKGGGGSS	PERV_Q4VFZ2_3mutA_WS
PAPGGSEAAAK	KORV_Q9TTC1_3mutA
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	MLVMS_P03355_3mutA_WS
GSSGSSGSSGSS	MLVMS_P03355_3mut
EAAAKGSSGGG	MLVMS_P03355_PLV919
GGSEAAAKPAP	AVIRE_P03360_3mutA
GSSGSSGSSGSSGSS	WMSV_P03359_3mutA
GGGEAAAKPAP	FLV_P10273_3mutA
PAPGSSGGG	KORV_Q9TTC1_3mutA
GSSGSS	MLVMS_P03355_3mutA_WS
PAPEAAAK	BAEVM_P10272_3mut
GGGPAPGSS	PERV_Q4VFZ2
GSSGGSPAP	MLVFF_P26809_3mut
GGGGSS	SFV3L_P27401_2mut
PAPEAAAKGSS	SFV3L_P27401_2mut
GGSGGGPAP	XMRV6_A1Z651_3mutA
PAPGGS	BAEVM_P10272_3mutA
EAAAKGGGGGS	AVIRE_P03360_3mut
GSSGGSPAP	KORV_Q9TTC1-Pro_3mutA
GSSGGGGGS	WMSV_P03359_3mut

[0909]

接头氨基酸序列	RT结构域名称
GGGEAAAKGGS	AVIRE_P03360_3mut
GGGEAAAKGSS	BAEVM_P10272_3mut
PAPEAAAKGSS	MLVAV_P03356_3mutA
GSSGSSGSSGSSGSS	MLVCB_P08361_3mut
GGSPAPGSS	FLV_P10273_3mutA
EAAAKGSSPAP	BAEVM_P10272_3mutA
GGSGGSGGSGGSGGSGGS	PERV_Q4VFZ2
GGGGSSEAAAK	FLV_P10273_3mutA
GGGGSSPAP	FFV_O93209
GSSGGSPAP	MLVMS_P03355_3mut
GGGPAPGSS	MLVMS_P03355_PLV919
PAPGSSGGS	PERV_Q4VFZ2_3mut
GGGGGSPAP	MLVFF_P26809_3mut
SGSETPGTSESATPES	MLVMS_P03355_3mutA_WS
GSSGSSGSSGSSGSS	KORV_Q9TTC1_3mutA
GSSPAPGGG	WMSV_P03359_3mut
PAPAPAPAPAPAP	SFV3L_P27401_2mutA
GGGPAPGGS	MLVMS_P03355_3mut
PAPGGSEAAAK	WMSV_P03359_3mut
GGGGSSEAAAK	FFV_O93209-Pro
GGSPAPGGG	FLV_P10273_3mutA
GSSPAPEAAAK	AVIRE_P03360_3mut
GGGEAAAK	FLV_P10273_3mutA
PAPEAAAKGGG	MLVCB_P08361_3mut
GGSPAPGGG	MLVCB_P08361_3mut
GGSGGGGSS	BAEVM_P10272_3mutA
GSSPAPEAAAK	MLVCB_P08361_3mut
GGSPAPGGG	KORV_Q9TTC1-Pro_3mutA
PAPGGSGSS	KORV_Q9TTC1_3mutA

[0910]

接头氨基酸序列	RT 结构域名称
GSSPAP	KORV_Q9TTC1-Pro_3mutA
SGSETPGTSESATPES	MLVMS_P03355
GSSGSSGSS	MLVAV_P03356_3mutA
PAPGSSGGS	PERV_Q4VFZ2_3mutA_WS
PAPGGS	KORV_Q9TTC1-Pro_3mutA
PAPEAAAKGGG	SFV3L_P27401-Pro_2mutA
GGSGGSGGS	BAEVM_P10272_3mut
PAPGGS	MLVFF_P26809_3mut
GSSGGSPAP	MLVMS_P03355_PLV919
GSSGGGGGS	FLV_P10273_3mutA
GGGGGSPAP	KORV_Q9TTC1-Pro_3mut
EAAAKPAPGSS	SFV3L_P27401-Pro_2mutA
EAAAKGGSPAP	KORV_Q9TTC1-Pro
GGGPAPEAAAK	MLVMS_P03355_PLV919
GGSEAAAKGSS	MLVMS_P03355
PAPEAAAKGSS	KORV_Q9TTC1_3mutA
PAPEAAAKGGS	WMSV_P03359_3mutA
GSSGGG	PERV_Q4VFZ2_3mutA_WS
EAAAKGGGGSS	MLVMS_P03355_PLV919
EAAAKGGSPAP	AVIRE_P03360_3mutA
GGGGSSGGS	MLVMS_P03355_PLV919
PAPEAAAKGSS	PERV_Q4VFZ2_3mutA_WS
EAAAKGGGGGS	BAEVM_P10272_3mut
GSSGGGGGS	MLVMS_P03355_3mut
PAPAPAPAP	KORV_Q9TTC1_3mutA
GGSGGSGGSGGS	MLVAV_P03356_3mut
PAPAPAPAP	SFV3L_P27401_2mut
GSSEAAAKPAP	MLVMS_P03355_3mut
GGSGGGEAAAK	SFV3L_P27401_2mutA

[0911]

接头氨基酸序列	RT结构域名称
GSSGGSGGG	MLVMS_P03355_3mutA_WS
GGGGGSPAP	MLVCB_P08361_3mutA
GGGEAAAKGSS	XMRV6_A1Z651_3mutA
GGGGSSPAP	BAEVM_P10272_3mut
GGSGGG	PERV_Q4VFZ2_3mut
GGGGSS	MLVBM_Q7SVK7_3mutA_WS
EAAAKGSSGGS	PERV_Q4VFZ2_3mutA_WS
GSSGGGGGS	PERV_Q4VFZ2
EAAAKGSSGGS	PERV_Q4VFZ2_3mut
EAAAKEAAAK	MLVAV_P03356_3mut
GSSGGGEAAAK	MLVAV_P03356_3mut
GSSPAPGGG	XMRV6_A1Z651_3mut
GGGGSGGGGSGGGGS	PERV_Q4VFZ2_3mut
EAAAKEAAAKEAAAKEAAAK	KORV_Q9TTC1_3mutA
EAAAKGGSGSS	MLVBM_Q7SVK7_3mut
PAPEAAAK	BLVJ_P03361
GSSGGG	FFV_O93209-Pro
GGSGGGEAAAK	KORV_Q9TTC1-Pro_3mutA
EAAAK	FLV_P10273_3mutA
GGGGSSPAP	MLVMS_P03355_3mut
GSS	SFV3L_P27401-Pro_2mut
PAPEAAAKGSS	BAEVM_P10272_3mut
GGGGGSPAP	PERV_Q4VFZ2_3mut
GSSGSSGSS	BAEVM_P10272_3mutA
GGGGSGGGGSGGGGSGGGGS	SFV1_P23074_2mut
GGGGSSEAAAK	SFV3L_P27401_2mutA
GGGGSGGGGSGGGGSGGGGS	FOAMV_P14350-Pro_2mut
PAPGSSEAAAK	MLVBM_Q7SVK7_3mutA_W

[0912]

接头氨基酸序列	RT 结构域名称
	S
GGGGGSGSS	MLVFF_P26809_3mutA
GGSEAAAKGGG	MLVBM_Q7SVK7_3mut
PAPGSSGGG	PERV_Q4VFZ2
GGS	PERV_Q4VFZ2_3mutA_WS
EAAAKGGSGSS	FLV_P10273_3mut
GGGEAAAK	WMSV_P03359_3mutA
GGSEAAAKPAP	MLVBM_Q7SVK7_3mut
SGSETPGTSESATPES	FOAMV_P14350-Pro_2mutA
EAAAKPAPGGS	AVIRE_P03360_3mut
EAAAKGGGGGS	KORV_Q9TTC1-Pro_3mutA
GGGGS	PERV_Q4VFZ2_3mut
GGSEAAAKGSS	MLVFF_P26809_3mutA
GGSEAAAKGGG	AVIRE_P03360
GGSGGSGGSGGSGGSGGS	SFV3L_P27401_2mut
GGSEAAAKGSS	SFV3L_P27401-Pro_2mutA
GGGEAAAKPAP	MLVCB_P08361_3mut
GGSEAAAK	MLVMS_P03355_PLV919
GGSPAPGSS	KORV_Q9TTC1-Pro_3mutA
GSSPAPEAAAK	WMSV_P03359_3mutA
GGSGSS	KORV_Q9TTC1-Pro_3mutA
PAPGGGGGS	AVIRE_P03360_3mut
PAPEAAAKGSS	FFV_O93209-Pro
GGSGGGEAAAK	WMSV_P03359_3mut
PAPGGG	MLVMS_P03355_3mut
EAAAKGGG	FLV_P10273_3mutA
GSSGSSGSSGSS	MLVCB_P08361_3mut
EAAAKGGSGGG	FFV_O93209
GSSPAPGGS	PERV_Q4VFZ2_3mutA_WS

[0913]

接头氨基酸序列	RT结构域名称
GSSPAPGGS	MLVCB_P08361_3mut
GGGPAP	WMSV_P03359_3mutA
GGGPAP	KORV_Q9TTC1_3mutA
GGSPAPGSS	KORV_Q9TTC1-Pro_3mut
PAPAP	MLVMS_P03355_3mut
GGGGGGG	MLVMS_P03355_3mut
GGGGG	KORV_Q9TTC1-Pro_3mut
GSAGSAAGSGEF	FOAMV_P14350_2mutA
PAPAP	KORV_Q9TTC1-Pro_3mutA
GGSEAAAKGGG	SFV3L_P27401-Pro_2mutA
PAPAP	WMSV_P03359_3mut
GGGGSGGGGSGGGGS	SFV3L_P27401_2mut
PAPGGS	KORV_Q9TTC1_3mutA
GGGEAAAKPAP	FLV_P10273_3mut
GGGGGS	MLVAV_P03356_3mutA
GSSEAAAKGGG	WMSV_P03359_3mut
EAAAKGGGGSS	GALV_P21414_3mutA
GSSGGS	MLVAV_P03356_3mutA
GSSGGG	MLVBM_Q7SVK7_3mut
PAPAPAP	SFV3L_P27401-Pro_2mutA
GGGG	KORV_Q9TTC1_3mutA
EAAAKPAPGGS	MLVFF_P26809_3mut
GGGGSGGGGS	XMRV6_A1Z651_3mut
EAAAKGGG	MLVCB_P08361_3mut
GGGGSSPAP	KORV_Q9TTC1_3mutA
GSSEAAAKGGG	KORV_Q9TTC1-Pro_3mutA
GGGGG	BLVJ_P03361_2mutB
GGGEAAAKGSS	FFV_O93209-Pro
GSSGSSGSS	BAEVM_P10272_3mut

[0914]

接头氨基酸序列	RT 结构域名称
GSSGGSPAP	PERV_Q4VFZ2_3mut
EAAAKGGS	KORV_Q9TTC1_3mut
GGSPAPEAAAK	AVIRE_P03360_3mut
GGSEAAAK	WMSV_P03359_3mut
GSSGGS	KORV_Q9TTC1-Pro_3mutA
GGGPAPEAAAK	KORV_Q9TTC1_3mutA
PAPGSS	WMSV_P03359_3mutA
GGSEAAAKGSS	FLV_P10273_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAK	SFV3L_P27401
GSSEAAAKGGG	SFV3L_P27401-Pro_2mutA
GGGGSEAAAKGGGGS	KORV_Q9TTC1-Pro_3mutA
GGSGGSGGS	WMSV_P03359_3mut
GGGGGSGSS	KORV_Q9TTC1-Pro
GGGGSGGGGSGGGGSGGGGS	MLVMS_P03355_3mut
EAAAKGGG	PERV_Q4VFZ2
GGSEAAAKGGG	KORV_Q9TTC1-Pro_3mut
GSSGGSGGG	PERV_Q4VFZ2_3mutA_WS
GGGGGS	PERV_Q4VFZ2_3mut
GSAGSAAGSGEF	PERV_Q4VFZ2
PAPEAAAKGSS	BAEVM_P10272_3mutA
GSSPAPGGG	MLVCB_P08361_3mut
GGGGSSPAP	KORV_Q9TTC1-Pro_3mutA
PAPGGSGGG	MLVFF_P26809_3mut
GSSPAP	KORV_Q9TTC1_3mutA
PAPGSS	SFV3L_P27401-Pro_2mut
GGSGGGGSS	MLVMS_P03355_PLV919
GSSGGS	WMSV_P03359_3mutA
EAAAKGGGGGS	PERV_Q4VFZ2
GGGGG	KORV_Q9TTC1_3mutA

[0915]

接头氨基酸序列	RT结构域名称
EAAAKGSS	MLVMS_P03355_PLV919
EAAAKEAAAKEAAAKEAAAKEAAAK	FLV_P10273_3mut
EAAAKEAAAKEAAAKEAAAK	SFV3L_P27401-Pro_2mut
GSAGSAAGSGEF	SFV3L_P27401_2mutA
GGGPAPGGS	FLV_P10273_3mutA
GGSEAAAKGGG	MLVCB_P08361_3mut
PAPGGGEAAAK	BAEVM_P10272_3mut
EAAAKPAPGSS	FOAMV_P14350_2mut
GGSEAAAK	KORV_Q9TTC1_3mutA
GGSGSS	AVIRE_P03360
GGSPAPEAAAK	MLVMS_P03355_PLV919
GGGGS	XMRV6_A1Z651_3mut
GGSPAPGGG	XMRV6_A1Z651_3mut
EAAAKPAPGGS	PERV_Q4VFZ2
GSSPAP	BAEVM_P10272_3mut
GGSGSSGGG	FLV_P10273_3mutA
PAPGGG	PERV_Q4VFZ2_3mutA_WS
GSSGGSEAAAK	MLVBM_Q7SVK7_3mut
GGSEAAAK	MLVMS_P03355_3mut
GGGPAPGGS	MLVFF_P26809_3mut
GSAGSAAGSGEF	MLVBM_Q7SVK7_3mutA_WS
EAAAKPAPGGS	SFVCP_Q87040
PAPGGG	PERV_Q4VFZ2_3mutA_WS
GSSPAPEAAAK	MLVBM_Q7SVK7
PAPEAAAK	MLVBM_Q7SVK7_3mut
PAPGGGGGS	AVIRE_P03360_3mutA
GGSEAAAKPAP	MLVBM_Q7SVK7_3mut
EAAAKGSS	WMSV_P03359_3mutA

[0916]

接头氨基酸序列	RT 结构域名称
GGGEAAAK	MLVFF_P26809_3mutA
EAAAKEAAAKEAAAK	MLVMS_P03355_3mut
PAPEAAAKGGG	BAEVM_P10272_3mut
PAPAPAP	MLVCB_P08361_3mut
EAAAKPAPGGS	BAEVM_P10272_3mut
GGGGSGGGGS	FLV_P10273_3mut
GGGGSEAAAKGGGGS	KORV_Q9TTC1_3mut
EAAAK	FLV_P10273_3mut
PAPAPAP	WMSV_P03359_3mut
GGGGSEAAAKGGGGS	FFV_O93209-Pro
GGSPAPEAAAK	MLVMS_P03355_3mut
GGSGSSGGG	XMRV6_A1Z651_3mut
GGSPAPGSS	PERV_Q4VFZ2_3mut
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	SFV3L_P27401-Pro_2mutA
EAAAKGGGPAP	BAEVM_P10272_3mutA
GSSGGSEAAAK	MLVMS_P03355_3mutA_WS
SGSETPGTSESATPES	PERV_Q4VFZ2_3mutA_WS
EAAAKEAAAKEAAAKEAAAKEAAAK	KORV_Q9TTC1-Pro_3mutA
GSSGSSGSS	KORV_Q9TTC1_3mutA
GSSPAPGGG	SFV3L_P27401-Pro_2mutA
GSSGGGEAAAK	KORV_Q9TTC1_3mutA
GGSGGGGSS	PERV_Q4VFZ2_3mutA_WS
GSSGGGEAAAK	MLVCB_P08361_3mut
GSSEAAAKGGG	MLVCB_P08361_3mut
GGSGGGGSS	KORV_Q9TTC1_3mutA
GGSGSSPAP	PERV_Q4VFZ2_3mutA_WS
GSSPAP	MLVMS_P03355_3mut
GGGGSSEAAAK	AVIRE_P03360
GGS	WMSV_P03359_3mut

[0917]

接头氨基酸序列	RT结构域名称
EAAAKEAAAK	PERV_Q4VFZ2_3mut
PAPAPAPAP	MLVAV_P03356_3mut
GGSEAAAKGGG	KORV_Q9TTC1_3mutA
PAPGGG	MLVAV_P03356_3mut
EAAAKGSS	BAEVM_P10272_3mut
GGGGSGGGGS	WMSV_P03359_3mutA
GGSGGSGGS	SFV3L_P27401_2mut
EAAAK	MLVCB_P08361_3mut
GGGGSSGGS	WMSV_P03359_3mutA
GGGPAPEAAAK	MLVAV_P03356_3mutA
EAAAKEAAAKEAAAK	FFV_O93209
GSSEAAAKGGG	MLVBM_Q7SVK7_3mut
GGGPAPGGS	FLV_P10273_3mut
GGSEAAAKGGG	WMSV_P03359_3mut
EAAAKGGGGGS	XMRV6_A1Z651_3mutA
EAAAKGGSGGG	FLV_P10273_3mutA
GGSEAAAKGGG	SFV3L_P27401_2mutA
GGGGS	PERV_Q4VFZ2_3mutA_WS
GSSGGS	MLVMS_P03355_3mut
GSSGSS	MLVAV_P03356_3mutA
GGSPAPGGG	MLVBM_Q7SVK7_3mutA_WS
GSSGGGGGS	MLVF5_P26810_3mut
PAPAPAPAP	MLVCB_P08361_3mut
PAPAP	PERV_Q4VFZ2_3mutA_WS
PAPGSSGGS	KORV_Q9TTC1_3mut
PAPGSSGGG	PERV_Q4VFZ2_3mut
GGGEAAAK	MLVMS_P03355_PLV919
GGSGGSGGSGGSGGS	SFV3L_P27401-Pro_2mutA

[0918]

接头氨基酸序列	RT 结构域名称
GGSGGG	FLV_P10273_3mut
PAPEAAAKGGG	MLVFF_P26809_3mut
PAP	PERV_Q4VFZ2_3mutA_WS
PAPGGSGSS	FFV_O93209_2mut
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	FFV_O93209-Pro_2mut
GSSGSSGSSGSS	FFV_O93209-Pro
GSSGSSGSSGSSGSS	FLV_P10273_3mutA
GGGEAAAKPAP	PERV_Q4VFZ2
PAPGSSGGG	SFV3L_P27401_2mut
PAPGGSGSS	KORV_Q9TTC1-Pro_3mut
PAPAPAPAPAP	GALV_P21414_3mutA
GGSGGGEAAAK	PERV_Q4VFZ2_3mut
GSSPAP	MLVCB_P08361_3mut
EAAAKPAP	MLVF5_P26810_3mut
GGGGSGGGGSGGGGSGGGGS	MLVBM_Q7SVK7_3mut
GGSGGG	WMSV_P03359_3mut
GGSGGSGGS	KORV_Q9TTC1_3mut
GGGGGGGG	MLVFF_P26809_3mut
GGGGSS	MLVAV_P03356_3mut
GSSGGGGGS	SFV3L_P27401_2mut
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	GALV_P21414_3mutA
GSSGSSGSS	PERV_Q4VFZ2_3mut
GSSPAPGGS	MLVFF_P26809_3mut
PAPAPAP	AVIRE_P03360_3mutA
EAAAKEAAAKEAAAKEAAAK	WMSV_P03359_3mutA
PAPAPAPAP	SFV3L_P27401_2mutA
GGGGSS	MLVAV_P03356_3mutA
GSSGSSGSSGSSGSS	SFV3L_P27401_2mutA
PAPGGS	WMSV_P03359_3mutA

[0919]

接头氨基酸序列	RT 结构域名称
GSSEAAAKGGG	PERV_Q4VFZ2
GSSGGSPAP	MLVMS_P03355_PLV919
GSSGSSGSSGSSGSSGSS	SFV3L_P27401_2mutA
GGSGSSGGG	MLVCB_P08361_3mut
GGGPAPGSS	SFV3L_P27401-Pro_2mutA
GSSEAAAKGGS	WMSV_P03359_3mut
GSSEAAAKGGG	MLVAV_P03356_3mut
GGSGGGPAP	FFV_O93209-Pro
GSSGSS	PERV_Q4VFZ2_3mut
PAPGGGGGS	GALV_P21414_3mutA
EAAAKPAPGGS	MLVAV_P03356_3mut
GSSGSS	MLVMS_P03355_3mut
EAAAKPAPGGS	FFV_O93209-Pro
GGGPAPEAAAK	MLVMS_P03355_3mutA_WS
GSSEAAAKGGG	MLVBM_Q7SVK7_3mut
GGGEAAAKGGS	BAEVM_P10272_3mut
GSSGSS	KORV_Q9TTC1-Pro_3mutA
EAAAKEAAAKEAAAK	SFV1_P23074
PAPGSSGGS	KORV_Q9TTC1-Pro_3mut
PAPAPAPAPAP	MLVMS_P03355
GSSEAAAK	SFV3L_P27401_2mut
PAP	PERV_Q4VFZ2_3mut
GGSEAAAKGGG	MLVBM_Q7SVK7_3mut
GGSGGGPAP	MLVBM_Q7SVK7_3mutA_WS
GSSGSS	MLVMS_P03355_3mut
GGSEAAAK	MLVMS_P03355
GSSEAAAKGGS	MLVMS_P03355_PLV919
PAPGGGGGS	MLVFF_P26809_3mut

[0920]

接头氨基酸序列	RT结构域名称
GSSGGG	PERV_Q4VFZ2_3mut
GSSGGS	PERV_Q4VFZ2_3mutA_WS
PAPGGG	BAEVM_P10272_3mut
PAPGSSGGG	MLVBM_Q7SVK7_3mut
GGSEAAAK	SFV3L_P27401_2mut
GSSPAPEAAAK	SFV3L_P27401-Pro_2mut
GSSGGSPAP	BAEVM_P10272_3mut
GGSPAPGSS	PERV_Q4VFZ2_3mutA_WS
GGSGGSGGS	PERV_Q4VFZ2
GGSGGGPAP	FLV_P10273_3mut
GGGPAPEAAAK	SFV3L_P27401_2mutA
GGGGS	FLV_P10273_3mutA
GSSGGSGGG	XMRV6_A1Z651_3mut
EAAAKGGGGSS	PERV_Q4VFZ2
GGSGSSGGG	SFV3L_P27401-Pro_2mutA
GGSGGSGGS	MLVFF_P26809_3mut
GGGPAPEAAAK	FLV_P10273_3mut
GSSGGGEAAAK	MLVMS_P03355_3mut
GGG	SFV3L_P27401_2mut
GSAGSAAGSGEF	WMSV_P03359_3mut
GSSGGGPAP	MLVMS_P03355_PLV919
GGGGSS	KORV_Q9TTC1-Pro_3mut
GGGGSSEAAAK	KORV_Q9TTC1
PAPGGSGGG	SFV3L_P27401_2mut
GSSGSSGSSGSSGSS	FFV_O93209
GSSGGSPAP	MLVMS_P03355_3mut
GGSEAAAK	KORV_Q9TTC1-Pro_3mutA
GGGGSGGGGS	BAEVM_P10272_3mut
GSSEAAAKGGG	AVIRE_P03360_3mut

[0921]

接头氨基酸序列	RT结构域名称
EAAAKPAPGGG	FLV_P10273_3mut
EAAAKGGSPAP	SFV3L_P27401-Pro_2mutA
GSSEAAAKPAP	MLVBM_Q7SVK7_3mut
GGGPAPGGS	MLVCB_P08361_3mut
GGG	SFV3L_P27401_2mutA
EAAAKGGGGSEAAAK	SFV3L_P27401_2mutA
GGSGSSGGG	MLVBM_Q7SVK7_3mut
GSAGSAAGSGEF	BAEVM_P10272_3mut
GGGEAAAK	FOAMV_P14350_2mutA
PAPEAAAKGGS	WMSV_P03359_3mut
PAPAPAPAPAPAP	MLVF5_P26810_3mutA
GGSGGGGSS	FLV_P10273_3mutA
PAPGSSGGS	BAEVM_P10272_3mut
PAPEAAAK	WMSV_P03359_3mutA
GSSGSSGSSGSSGSSGSS	FFV_O93209-Pro_2mut
GGGGGSGSS	FFV_O93209-Pro
GGGGGGGG	SFV3L_P27401-Pro_2mutA
GGGGGG	FLV_P10273_3mut
GSSGGSGGG	MLVAV_P03356_3mutA
GGGGSS	SFV3L_P27401-Pro_2mutA
GGSGGGPAP	FOAMV_P14350_2mut
GSSGSS	AVIRE_P03360_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAK	SFV3L_P27401-Pro_2mutA
EAAAKEAAAK	BAEVM_P10272_3mut
GSSPAPEAAAK	GALV_P21414_3mutA
GGSEAAAKPAP	SFV3L_P27401_2mutA
GGSGGGEAAAK	SFV3L_P27401-Pro_2mutA
EAAAKGSSPAP	FOAMV_P14350_2mut
GGSGSSEAAAK	SFV3L_P27401_2mut

[0922]

接头氨基酸序列	RT结构域名称
GGG	PERV_Q4VFZ2
GGGGGSGSS	FOAMV_P14350_2mut
GGSGGGEAAAK	KORV_Q9TTC1-Pro_3mut
GSSGGSGGG	AVIRE_P03360_3mutA
EAAAKPAPGGG	SFV3L_P27401_2mutA
PAPGGSGGG	KORV_Q9TTC1-Pro_3mut
PAPAPAP	WMSV_P03359_3mutA
GSSEAAAKPAP	SFV1_P23074
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	SRV2_P51517
GSSGGSGGG	PERV_Q4VFZ2_3mutA_WS
GSSGSSGSSGSSGSSGSS	FFV_O93209
GSSGGGPAP	WMSV_P03359_3mut
PAPAPAPAPAPAP	MLVBM_Q7SVK7_3mut
GGGGGSPAP	KORV_Q9TTC1-Pro_3mutA
PAPGSS	MLVBM_Q7SVK7_3mutA_WS
PAPEAAAKGGS	SFV3L_P27401-Pro_2mut
GGGGSSPAP	MLVMS_P03355_3mut
GGSEAAAK	FFV_O93209-Pro
EAAAKPAPGGS	AVIRE_P03360_3mutA
PAPGSS	WMSV_P03359_3mut
PAPGSSGGG	SFV3L_P27401-Pro_2mutA
EAAAKEAAAKEAAAK	SFV3L_P27401_2mut
GGS	MLVRD_P11227_3mut
GGGGS	KORV_Q9TTC1-Pro_3mut
GGSGGGGSS	KORV_Q9TTC1
GGSGGG	MLVMS_P03355_3mutA_WS
GGGEAAAKPAP	BAEVM_P10272_3mut
EAAAKEAAAKEAAAKEAAAKEAAAK	FLV_P10273

[0923]

接头氨基酸序列	RT结构域名称
PAPGGSGGG	KORV_Q9TTC1-Pro_3mutA
GSSGSSGSSGSSGSSGSS	HTL1L_P0C211
GGGEAAAKPAP	WMSV_P03359
GSSGGSPAP	FFV_O93209-Pro
PAPAPAPAPAP	SFV3L_P27401-Pro_2mutA
GSSGGSEAAAK	SFV3L_P27401_2mutA
GGSPAPGSS	SFV3L_P27401_2mut
GGSGGSGGS	KORV_Q9TTC1-Pro_3mut
PAPEAAAKGSS	KORV_Q9TTC1-Pro_3mut
EAAAKGGS	KORV_Q9TTC1_3mutA
EAAAKGGGGSEAAAK	SFV3L_P27401-Pro_2mut
GGGGSSPAP	FFV_O93209-Pro
EAAAK	SFV3L_P27401_2mut
EAAAKGGGGSS	BAEVM_P10272_3mut
GGGGGSEAAAK	MLVBM_Q7SVK7_3mut
GGGG	PERV_Q4VFZ2
GGGGGSEAAAK	FLV_P10273_3mut
EAAAKGGGPAP	KORV_Q9TTC1-Pro
GGGGSGGGGSGGGGSGGGGS	FFV_O93209_2mutA
GSSGGSGGG	PERV_Q4VFZ2_3mut
GGGGSGGGGSGGGGS	GALV_P21414_3mutA
GGSGGGEAAAK	AVIRE_P03360_3mutA
PAPEAAAKGGG	SFV3L_P27401_2mut
GGGGSGGGGS	AVIRE_P03360
GSSGGGEAAAK	SFV3L_P27401_2mutA
GGGGG	AVIRE_P03360_3mutA
GGSGSS	KORV_Q9TTC1_3mut
PAPAPAPAPAPAP	FOAMV_P14350_2mut
GGSEAAAKPAP	KORV_Q9TTC1-Pro_3mut

[0924]

接头氨基酸序列	RT结构域名称
GGGGGG	PERV_Q4VFZ2_3mut
GSSGGGEAAAK	MLVBM_Q7SVK7
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	MLVAV_P03356
GGSPAPGSS	BAEVM_P10272_3mut
GGGGSSPAP	BAEVM_P10272
GGGGSEAAAKGGGGS	SFV3L_P27401_2mut
GGGGGGGG	GALV_P21414_3mutA
PAPAP	MLVAV_P03356_3mut
GGGEAAAK	PERV_Q4VFZ2_3mutA_WS
GSSPAPGGG	FFV_O93209_2mut
GGSGGSGGSGGSGGS	BAEVM_P10272
GGGGGS	MLVF5_P26810_3mutA
PAPGGGGSS	FLV_P10273_3mutA
GGGEAAAK	MLVBM_Q7SVK7_3mut
PAPEAAAKGGG	WMSV_P03359_3mut
GSSEAAAK	MLVBM_Q7SVK7_3mut
EAAAKEAAAK	AVIRE_P03360
EAAAKGGGGGS	MLVBM_Q7SVK7_3mut
GGGEAAAKGGS	SFV3L_P27401-Pro_2mutA
PAPAPAPAPAP	MLVF5_P26810_3mut
PAPGSSEAAAK	SFV3L_P27401-Pro_2mutA
EAAAKEAAAKEAAAK	BAEVM_P10272_3mutA
GGSPAPGSS	MLVMS_P03355
PAPGSSGGS	FLV_P10273_3mutA
EAAAKEAAAKEAAAKEAAAK	FOAMV_P14350-Pro_2mut
EAAAKGGG	KORV_Q9TTC1_3mutA
EAAAKGGSGGG	MLVBM_Q7SVK7_3mut
GGGGGS	KORV_Q9TTC1-Pro_3mutA
PAPGGSGGG	WMSV_P03359_3mut

[0925]

接头氨基酸序列	RT结构域名称
GGGPAPGGS	KORV_Q9TTC1_3mutA
GSS	FFV_O93209
GGSGGSGGS	PERV_Q4VFZ2_3mut
GGGGS	GALV_P21414_3mutA
GGGG	MLVF5_P26810_3mut
GGSEAAAKPAP	FFV_O93209-Pro_2mut
PAPAPAPAP	FFV_O93209-Pro
PAP	MLVF5_P26810_3mut
EAAAKEAAAKEAAAK	FFV_O93209_2mut
EAAAKGSS	MLVCB_P08361_3mut
EAAAKGGG	MLVBM_Q7SVK7_3mut
PAPEAAAKGGG	FFV_O93209_2mut
GSSGGGEAAAK	SFV1_P23074-Pro_2mut
PAPGGGEAAAK	GALV_P21414_3mutA
GGGGSGGGGSGGGGSGGGGS	FOAMV_P14350-Pro_2mutA
GSSGGG	FOAMV_P14350_2mut
GGGGSGGGGSGGGGSGGGGS	SFV3L_P27401_2mutA
GGSGSS	AVIRE_P03360_3mut
GGSGSSEAAAK	MMTVB_P03365_WS
PAPAPAP	MLVAV_P03356_3mutA
GSSGGSPAP	SFV3L_P27401-Pro_2mut
GGSPAP	AVIRE_P03360
GGSGGGPAP	FFV_O93209
GSSEAAAK	PERV_Q4VFZ2
GSSGGGPAP	PERV_Q4VFZ2_3mutA_WS
GGGGSSEAAAK	KORV_Q9TTC1_3mutA
GGSEAAAKPAP	SFVCP_Q87040
GGSGGGPAP	FOAMV_P14350_2mutA
GGGGSGGGGSGGGGSGGGGS	BLVJ_P03361_2mutB

[0926]

接头氨基酸序列	RT 结构域名称
GGGGSSPAP	SFV3L_P27401_2mutA
EAAAKGGS	MLVF5_P26810_3mut
GGSEAAAKGSS	MLVCB_P08361_3mut
GGGGSSEAAAK	SFV3L_P27401_2mut
EAAAKGGSGGG	FOAMV_P14350_2mut
GGSGGS	FLV_P10273_3mut
EAAAKGGG	FFV_O93209-Pro
GSSGSSGSSGSSGSS	SFV3L_P27401
GSSGGGPAP	PERV_Q4VFZ2_3mutA_WS
PAPGGSEAAAK	SFV3L_P27401-Pro_2mutA
GGSPAP	KORV_Q9TTC1
EAAAKPAPGSS	KORV_Q9TTC1_3mutA
SGSETPGTSESATPES	SFV1_P23074
GSSPAP	SFV3L_P27401-Pro_2mutA
GSSPAPGGG	SFV3L_P27401_2mut
GGGEAAAKGSS	SFV1_P23074_2mut
GGGPAPGGS	BAEVM_P10272_3mut
EAAAKGGG	KORV_Q9TTC1-Pro_3mutA
GSSGGG	SFV3L_P27401-Pro_2mut
GGSPAPEAAAK	BAEVM_P10272_3mut
EAAAKGSSPAP	FFV_O93209
EAAAKGGGGSEAAAK	SFV3L_P27401-Pro_2mutA
GSSGSSGSSGSSGSS	SFV1_P23074_2mut
EAAAKGGSPAP	FOAMV_P14350_2mut
GGSGGS	KORV_Q9TTC1-Pro_3mutA
EAAAKGSSGGS	GALV_P21414
GSSGGGPAP	MLVAV_P03356
PAPEAAAKGGS	FOAMV_P14350_2mut
EAAAKPAPGGG	AVIRE_P03360_3mut

[0927]

接头氨基酸序列	RT 结构域名称
GGSPAP	SFV3L_P27401_2mutA
GGGGSGGGGS	SFV3L_P27401_2mutA
GGGGSS	AVIRE_P03360_3mutA
GGSPAPGGG	SFV3L_P27401-Pro_2mutA
EAAAKPAPGSS	SFV3L_P27401
EAAAKPAP	FOAMV_P14350-Pro_2mut
PAPEAAAKGSS	PERV_Q4VFZ2_3mutA_WS
EAAAKGGSGSS	SFV3L_P27401_2mutA
GGGEAAAKGSS	GALV_P21414_3mutA
GGGGSEAAAKGGGGS	PERV_Q4VFZ2_3mut
PAPGGSGSS	FFV_O93209-Pro_2mutA
GGSEAAAKPAP	GALV_P21414_3mutA
GGSGGSGGSGGSGGS	FFV_O93209-Pro
GSSGGSEAAAK	SFV3L_P27401-Pro_2mut
GGS	GALV_P21414_3mutA
PAPGGSEAAAK	MLVMS_P03355
PAPEAAAKGGS	BAEVM_P10272_3mutA
GGSGSSPAP	SFV3L_P27401-Pro_2mutA
GSSPAP	WMSV_P03359_3mut
GGGEAAAK	MMTVB_P03365
GGGGSS	PERV_Q4VFZ2_3mut
GGSPAPGSS	SFV3L_P27401-Pro_2mut
PAPGGS	MLVBM_Q7SVK7_3mut
EAAAKGSSPAP	MLVBM_Q7SVK7_3mut
GGGGSSGGS	PERV_Q4VFZ2_3mut
PAPAPAPAPAPAP	SFV1_P23074
GGSEAAAKGGG	SFV3L_P27401-Pro_2mut
GGSGGS	SFV1_P23074_2mut
GSSGGGGGS	MLVF5_P26810_3mutA

[0928]

接头氨基酸序列	RT结构域名称
EAAAKGGGPAP	SFV3L_P27401
EAAAKEAAAKEAAAKEAAAK	FOAMV_P14350-Pro_2mutA
GGGPAPGSS	SFV3L_P27401_2mutA
GGGGSGGGGSGGGGSGGGGS	SFV3L_P27401_2mut
EAAAKEAAAKEAAAKEAAAK	MMTVB_P03365_WS
PAPGSSGGS	KORV_Q9TTC1-Pro_3mutA
PAPGSSEAAAK	FOAMV_P14350-Pro_2mut
GSSPAPEAAAK	BAEVM_P10272_3mut
EAAAKGGGGSEAAAK	FFV_O93209-Pro
GGSPAP	PERV_Q4VFZ2
GGSGSSEAAAK	XMRV6_A1Z651_3mut
GGSEAAAKGGG	GALV_P21414_3mutA
PAPGGGGSS	AVIRE_P03360_3mutA
GGSGGSGGSGGS	PERV_Q4VFZ2
GGGGSSGGS	PERV_Q4VFZ2_3mutA_WS
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	BAEVM_P10272_3mutA
GGGPAP	MLVAV_P03356_3mut
GGGGSGGGGSGGGGSGGGGS	FFV_O93209_2mut
GSSEAAAK	FFV_O93209
GGSPAPEAAAK	FOAMV_P14350_2mut
GGGGGSEAAAK	FOAMV_P14350_2mut
GSSPAPGGS	MLVBM_Q7SVK7_3mut
GSS	SFVCP_Q87040_2mut
EAAAKPAP	FOAMV_P14350-Pro
EAAAKGGG	SFV3L_P27401_2mut
GGGEAAAK	AVIRE_P03360_3mutA
PAPGSSGGG	WMSV_P03359_3mut
EAAAKGGSPAP	SFV3L_P27401
GSSGGSGGG	SFV3L_P27401-Pro_2mutA

[0929]

接头氨基酸序列	RT结构域名称
GSSGGGEAAAK	GALV_P21414_3mutA
GGGPAPGSS	MLVBM_Q7SVK7_3mutA_WS
PAPGGGEAAAK	FFV_O93209-Pro_2mut
GSSGSSGSSGSS	SFV1_P23074_2mut
GGSEAAAK	PERV_Q4VFZ2_3mutA_WS
GGGEAAAKPAP	SFV3L_P27401_2mut
EAAAKGGGPAP	SFV3L_P27401_2mut
GGGGSSPAP	FLV_P10273_3mut
EAAAKPAPGSS	FFV_O93209_2mut
GGGGSSPAP	SFV3L_P27401_2mut
GSSGSS	KORV_Q9TTC1_3mutA
GGGGSGGGGSGGGGSGGGGSGGGGS	BLVJ_P03361_2mut
GGGGSSGGS	GALV_P21414_3mutA
EAAAKGGSGSS	FFV_O93209-Pro
EAAAKPAP	PERV_Q4VFZ2
GSSGGGEAAAK	MLVBM_Q7SVK7_3mut
PAPGGSGGG	BAEVM_P10272
EAAAKGGGPAP	MLVF5_P26810
GSSGSSGSS	MLVBM_Q7SVK7_3mut
GSSGGS	AVIRE_P03360_3mutA
GGSEAAAKGGG	FOAMV_P14350_2mut
EAAAKGGS	MLVF5_P26810_3mutA
GGSGSSGGG	WMSV_P03359_3mut
EAAAK	SFV1_P23074_2mut
GSSGGSPAP	SFV3L_P27401-Pro_2mutA
GGGGSSGGS	KORV_Q9TTC1_3mut
PAPGGSGGG	FFV_O93209-Pro_2mut
GGGPAPGGS	SFV3L_P27401_2mutA

[0930]

接头氨基酸序列	RT结构域名称
GSSPAPEAAAK	FLV_P10273_3mut
GGSGSSPAP	SFV3L_P27401_2mut
GSSEAAAKGGS	SFV3L_P27401_2mut
PAPGGG	SFV3L_P27401_2mutA
SGSETPGTSESATPES	KORV_Q9TTC1-Pro_3mut
GGGGS	SFV1_P23074-Pro_2mutA
GSSGGGEAAAK	WMSV_P03359
EAAAKGGGGSEAAAK	MLVF5_P26810_3mutA
GSSEAAAKPAP	FFV_O93209
GGGGGG	SFV1_P23074_2mutA
EAAAKEAAAKEAAAK	MMTVB_P03365-Pro
EAAAKPAPGSS	MLVBM_Q7SVK7_3mut
GGSGSSEAAAK	SFV3L_P27401_2mutA
GGSEAAAK	MLVMS_P03355_3mut
GGSPAPEAAAK	SFV3L_P27401_2mut
GGGPAPGSS	SFV1_P23074
GGGGGSEAAAK	MLVBM_Q7SVK7_3mutA_WS
EAAAKPAPGSS	KORV_Q9TTC1-Pro
GSSGSSGSSGSS	SFV3L_P27401_2mut
EAAAKPAP	SFV3L_P27401_2mut
GGGEAAAK	PERV_Q4VFZ2_3mut
GGSGGS	SFV3L_P27401_2mutA
EAAAKGSSGGS	MMTVB_P03365
SGSETPGTSESATPES	SFV3L_P27401
EAAAKGSSGGG	PERV_Q4VFZ2
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	MMTVB_P03365
GGSGGGPAP	KORV_Q9TTC1_3mutA
PAPAPAPAP	SFV3L_P27401

[0931]

接头氨基酸序列	RT结构域名称
GGGEAAAKGGS	SFV1_P23074_2mut
GSSGGSGGG	PERV_Q4VFZ2_3mut
PAPEAAAKGGS	FOAMV_P14350_2mutA
GGGEAAAKGSS	SFV3L_P27401_2mut
GGGGSGGGGSGGGGSGGGGS	MLVBM_Q7SVK7
PAPGSSGGG	FLV_P10273
GGSGSSGGG	FFV_O93209
EAAAKPAPGSS	MLVBM_Q7SVK7
GSSEAAAKGGG	SFV3L_P27401_2mutA
GGSGGSGGSGGSGGS	MLVF5_P26810
GGSEAAAKPAP	SFV3L_P27401-Pro_2mutA
EAAAKGGSPAP	SFV3L_P27401_2mutA
EAAAKGGGGGS	SFV3L_P27401_2mut
GSSPAPEAAAK	SFV3L_P27401_2mutA
PAPAP	MLVBM_Q7SVK7_3mut
PAPGGSEAAAK	KORV_Q9TTC1-Pro
GGSGSS	MLVF5_P26810_3mutA
GGSEAAAKPAP	FFV_O93209_2mut
GSS	MLVMS_P03355
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	SFV3L_P27401-Pro
PAPGGGEAAAK	SFV3L_P27401_2mut
PAPGGGGGS	SFV3L_P27401-Pro_2mut
PAPGGSGSS	BAEVM_P10272_3mut
GSSGGGEAAAK	FFV_O93209
GGSEAAAKPAP	SFV1_P23074_2mut
GGGGG	FLV_P10273_3mut
GGGEAAAKGSS	SFV3L_P27401
GSSGSSGSSGSSGSS	SFV1_P23074-Pro
SGSETPGTSESATPES	AVIRE_P03360

[0932]

接头氨基酸序列	RT结构域名称
PAPGSSGGG	MLVBM_Q7SVK7_3mut
GGGGSSPAP	HTL3P_Q4U0X6_2mut
GGGEAAAK	SFV1_P23074
GGSGGG	AVIRE_P03360
EAAAKGSSGGG	SFV3L_P27401_2mutA
GSSPAPEAAAK	FOAMV_P14350-Pro_2mutA
GGGPAPGSS	WMSV_P03359
EAAAKGSSGGG	MLVMS_P03355
GGGGGSEAAAK	MLVMS_P03355
EAAAKPAPGGS	SFV3L_P27401
EAAAKGSSPAP	SFV3L_P27401
GGGGGGG	FOAMV_P14350_2mutA
EAAAKEAAAKEAAAK	SFV3L_P27401
GSSPAPGGS	FFV_O93209_2mutA
GGGGSSEAAAK	SFV3L_P27401-Pro_2mutA
GGSEAAAKGSS	GALV_P21414_3mutA
GGSEAAAKGSS	BAEVM_P10272_3mutA
EAAAKPAPGGG	MLVCB_P08361
GSSGSSGSSGSSGSSGSS	SFV1_P23074-Pro
GGGGSEAAAKGGGGS	FOAMV_P14350_2mut
GSSPAPGGS	MLVMS_P03355_PLV919
GGGGSGGGGS	FFV_O93209-Pro
GSSGGSPAP	KORV_Q9TTC1_3mutA
GGSGGS	GALV_P21414_3mutA
PAPGSSEAAAK	WMSV_P03359
PAPGGGGSS	MMTVB_P03365-Pro
GGGGSSGGS	PERV_Q4VFZ2_3mutA_WS
GGGGSGGGGS	FFV_O93209_2mut
GGGGSGGGGSGGGGSGGGGS	XMRV6_A1Z651

[0933]

接头氨基酸序列	RT 结构域名称
GGSGSSEAAAK	SFV1_P23074_2mut
GGSGGGGSS	GALV_P21414_3mutA
GGSEAAAKPAP	MLVBM_Q7SVK7
EAAAKGGSPAP	SFV1_P23074_2mutA
PAPAPAPAP	FFV_O93209
GSSGGSPAP	MMTVB_P03365-Pro
GGGGGSPAP	KORV_Q9TTC1_3mutA
EAAAKGGGPAP	PERV_Q4VFZ2
GSSGGSPAP	BAEVM_P10272
GGGGG	FFV_O93209
GGGGGS	FLV_P10273_3mutA
EAAAKEAAAKEAAAK	FOAMV_P14350
PAPGGG	MLVCB_P08361_3mut
GSSGGSEAAAK	FOAMV_P14350_2mutA
GGSPAPGGG	FLV_P10273_3mut
GSSGSSGSSGSSGSSGSS	SFV1_P23074-Pro_2mutA
GGSPAPEAAAK	SFV3L_P27401
PAPGGGGSS	HTL3P_Q4U0X6_2mutB
GGGGSSEAAAK	MMTVB_P03365_2mut_WS
PAPGGS	MLVRD_P11227_3mut
GGSGGSGGSGGSGGS	MMTVB_P03365
GSAGSAAGSGEF	AVIRE_P03360
GSSGGS	BAEVM_P10272_3mutA
GGSGGGGSS	MMTVB_P03365
GGSGGGGSS	WMSV_P03359
PAPEAAAKGSS	SFV1_P23074
GSSGSSGSSGSS	SFV1_P23074-Pro_2mutA
PAPAPAPAPAPAP	SFV3L_P27401
PAPGSSGGG	FLV_P10273_3mut

[0934]

接头氨基酸序列	RT结构域名称
GGSGSSPAP	MLVMS_P03355
GGSGGGPAP	FOAMV_P14350
PAPGGGGGS	KORV_Q9TTC1_3mutA
EAAAKGSSPAP	GALV_P21414_3mutA
GGSGSSPAP	MLVBM_Q7SVK7_3mut
EAAAKGSS	SFV3L_P27401_2mut
GGGGGSEAAAK	WMSV_P03359
GGGGGGGG	SFV1_P23074-Pro
EAAAKEAAAK	MLVBM_Q7SVK7
GGGEAAAKGGS	MLVBM_Q7SVK7
EAAAKGGSPAP	SFV3L_P27401_2mut
GSSEAAAK	XMRV6_A1Z651
PAPGGGEAAAK	MMTVB_P03365_WS
GGSPAP	GALV_P21414_3mutA
GSSPAPGGG	MLVBM_Q7SVK7_3mutA_WS
GGSGSSPAP	SFV1_P23074_2mutA
GGS	HTL32_Q0R5R2_2mut
GGSGGGGSS	MMTVB_P03365-Pro
GGGGSGGGGSGGGGSGGGGS	SFVCP_Q87040_2mutA
EAAAKGGGPAP	FOAMV_P14350_2mut
GSSGGGEAAAK	MMTVB_P03365
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	MLVBM_Q7SVK7_3mutA_WS
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	MMTVB_P03365_WS
EAAAKEAAAK	FOAMV_P14350-Pro_2mut
GSSPAPEAAAK	FOAMV_P14350_2mutA
EAAAKPAPGGS	GALV_P21414_3mutA

[0935]

接头氨基酸序列	RT结构域名称
GSSGGSPAP	KORV_Q9TTC1-Pro_3mut
GGGPAPEAAAK	MLVAV_P03356
GGGEAAAKPAP	SFV1_P23074-Pro_2mut
GGGGGSEAAAK	SFV3L_P27401_2mut
GGGPAPGSS	SFV3L_P27401_2mut
GGSEAAAKPAP	AVIRE_P03360
GSSGSSGSSGSSGSSGSS	SFV1_P23074-Pro_2mut
EAAAKGSSGGS	FOAMV_P14350_2mutA
GGGGGG	MLVBM_Q7SVK7_3mut
GSSPAPGGS	PERV_Q4VFZ2
GGSGSSPAP	GALV_P21414_3mutA
GGGPAPEAAAK	SFV3L_P27401
GGSGGGEAAAK	WMSV_P03359
GSAGSAAGSGEF	SFV1_P23074_2mut
GSSGGGEAAAK	MLVMS_P03355
GGG	MMTVB_P03365-Pro
PAPGSSGGS	FOAMV_P14350_2mut
GGGGSSPAP	FFV_O93209_2mut
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	MMTVB_P03365_WS
GGGGGGG	XMRV6_A1Z651
PAPAPAPAPAP	FOAMV_P14350
GGGGSGGGGSGGGGSGGGGS	MMTVB_P03365_2mut_WS
GGSGGGPAP	SFV3L_P27401_2mut
GGGGGG	SFV1_P23074-Pro
EAAAKPAPGSS	SFV3L_P27401_2mut
GGGGSSGGS	HTL3P_Q4U0X6_2mut
PAPGSSEAAAK	MMTVB_P03365-Pro
GGGGSSPAP	FOAMV_P14350-Pro_2mut
PAPGSSGGS	MMTVB_P03365

[0936]

接头氨基酸序列	RT 结构域名称
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	SRV2_P51517
PAPAPAP	MMTVB_P03365_2mut_WS
PAPGGGGGS	MMTVB_P03365_2mutB
GGGGSS	SFV1_P23074-Pro_2mutA
EAAAKEAAAKEAAAKEAAAK	SFV3L_P27401-Pro
GGSGGSGGSGGSGGS	MMTVB_P03365-Pro
GGGGGGG	SFV3L_P27401_2mut
PAPGGGEAAAK	SFV3L_P27401
PAPGSS	FOAMV_P14350_2mutA
GGGGSGGGGS	SFVCP_Q87040_2mutA
GSSGGSGGG	XMRV6_A1Z651
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	MLVBM_Q7SVK7
GSSEAAAKGGG	FFV_O93209-Pro_2mut
GGSEAAAKPAP	SFV3L_P27401-Pro
GSSGGSGGG	SFV1_P23074_2mut
EAAAKGGGGSS	FOAMV_P14350_2mutA
GGGGGG	SFV3L_P27401_2mut
GGGGG	MLVBM_Q7SVK7_3mut
PAPEAAAKGGG	SFV3L_P27401
EAAAKGGSPAP	KORV_Q9TTC1_3mutA
GGGEAAAKPAP	SFV1_P23074_2mut
GSSGSSGSSGSSGSSGSS	KORV_Q9TTC1-Pro
EAAAKEAAAKEAAAKEAAAK	SFVCP_Q87040
PAPGSSEAAAK	MLVBM_Q7SVK7
GSSGSSGSS	FFV_O93209-Pro_2mut
GSSGGGPAP	SFV3L_P27401-Pro_2mut
GGGPAPEAAAK	WMSV_P03359_3mut
GGGEAAAK	MMTVB_P03365-Pro

[0937]

接头氨基酸序列	RT 结构域名称
GSSGSSGSSGSS	SFV3L_P27401-Pro_2mutA
PAPAPAPAPAP	FFV_O93209-Pro
GGSPAPEAAAK	FFV_O93209-Pro_2mut
GSSGSSGSSGSSGSSGSS	GALV_P21414
EAAAKEAAAKEAAAKEAAAKEAAAK	FOAMV_P14350
GGGPAPEAAAK	MMTVB_P03365-Pro
PAPGGSGGG	MLVF5_P26810_3mutA
PAPGGSGGG	FLV_P10273_3mut
GGGEAAAKGGS	SFV3L_P27401
GSAGSAAGSGEF	MLVBM_Q7SVK7_3mut
GSSPAPGGG	MPMV_P07572_2mutB
GSSGSSGSSGSSGSSGSS	FOAMV_P14350
GGSGGGGSS	BLVJ_P03361_2mut
PAPEAAAKGSS	SFV1_P23074-Pro
GGG	FFV_O93209
EAAAKGGGGSS	SFV1_P23074_2mut
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	SRV2_P51517
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	MMTVB_P03365
GGGEAAAKGGS	MMTVB_P03365_WS
GSSGSS	SFV1_P23074
GSSGGGGGS	SFV3L_P27401
GGGGSSEAAAK	SFV1_P23074
EAAAKGSSGGS	HTL1A_P03362_2mutB
GSSEAAAKGGS	GALV_P21414_3mutA
EAAAKGSSPAP	SFV1_P23074
EAAAKPAPGSS	SFV3L_P27401_2mutA
PAPGSSGGG	SFV3L_P27401-Pro_2mut
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	SFV3L_P27401-Pro
EAAAKEAAAKEAAAKEAAAKEAAAK	MMTVB_P03365_WS

[0938]

接头氨基酸序列	RT 结构域名称
GGGGSSEAAAK	MLVF5_P26810_3mutA
EAAAKGGSPAP	GALV_P21414
PAPEAAAKGSS	MMTVB_P03365_WS
GSSGGGGGS	SFVCP_Q87040_2mut
GGGGSSPAP	SFV1_P23074
EAAAKGGGGSS	XMRV6_A1Z651
PAPAPAPAP	MMTVB_P03365
GGSEAAAKGSS	SFV3L_P27401_2mutA
GSSPAPGGG	MMTVB_P03365_WS
GGGGGG	SFV3L_P27401-Pro
GGSGGSGGS	FOAMV_P14350-Pro_2mut
PAPAPAPAPAPAP	WMSV_P03359
GSSPAP	MLVBM_Q7SVK7
GGGGGSGSS	MMTVB_P03365_2mut_WS
EAAAKGSSGGS	MMTVB_P03365_2mutB_WS
EAAAK	FFV_O93209_2mutA
PAPEAAAK	SFV1_P23074-Pro
EAAAKGGSGSS	SFV3L_P27401
GGSGGSGGS	FFV_O93209-Pro
GSSGGGEAAAK	MMTVB_P03365
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	MLVFF_P26809_3mutA
GGSGGSGGSGGSGGSGGS	HTL1L_P0C211_2mutB
GGGEAAAK	SFV3L_P27401-Pro_2mutA
GGGGGSGSS	MMTVB_P03365
GSSPAPGGS	FOAMV_P14350_2mutA
EAAAKGSS	MLVMS_P03355
GSSGGSGGG	FFV_O93209-Pro
GGSGGGGSS	MMTVB_P03365-Pro_2mut
GGSPAPGSS	FOAMV_P14350_2mut

[0939]

接头氨基酸序列	RT结构域名称
GGSGGSGGSGGSGGSGGS	SFVCP_Q87040-Pro_2mut
GSSEAAAKGGG	FOAMV_P14350_2mutA
GGSGGSGGS	MMTVB_P03365-Pro
GSSGSSGSSGSSGSSGSS	MMTVB_P03365_2mut_WS
GSSGSSGSSGSSGSS	MMTVB_P03365-Pro
PAPEAAAK	WDSV_O92815
GSSGSSGSSGSSGSS	FFV_O93209-Pro_2mut
EAAAKGGGGSEAAAK	MMTVB_P03365-Pro
GGSPAPEAAAK	FOAMV_P14350
GSSGSS	PERV_Q4VFZ2
GGG	MMTVB_P03365-Pro
GGGGSGGGGSGGGGS	FFV_O93209_2mut
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	MMTVB_P03365-Pro
GGSGSSPAP	WMSV_P03359
GGGGGGGG	SFV3L_P27401_2mut
PAPGSSEAAAK	FOAMV_P14350-Pro_2mutA
GGGGSSPAP	FOAMV_P14350_2mut
GSSGGSPAP	MLVBM_Q7SVK7_3mut
GSSGGGGGS	GALV_P21414_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAK	MMTVB_P03365
GSSGGGGGS	SFV1_P23074_2mut
GGGGSEAAAKGGGGS	SFV1_P23074
GGGEAAAKPAP	FFV_O93209
PAPGGGEAAAK	SFV1_P23074
GGSGGGEAAAK	PERV_Q4VFZ2_3mutA_WS
GSSGGG	MMTVB_P03365-Pro
EAAAKGSSGGS	FFV_O93209_2mut
GGGGG	SFV1_P23074_2mut
GGGPAP	SFV3L_P27401

[0940]

接头氨基酸序列	RT结构域名称
GSSGGSEAAAK	FFV_O93209
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	MMTVB_P03365-Pro
GSSGGGEAAAK	SFV1_P23074_2mutA
GSSGSSGSSGSSGSS	SFV3L_P27401_2mut
GGSEAAAKPAP	FLV_P10273
GGGGSGGGGS	FOAMV_P14350-Pro_2mutA
GSSEAAAKPAP	SFV3L_P27401
GGGGSEAAAKGGGGS	MMTVB_P03365-Pro
PAPGSSEAAAK	MLVF5_P26810_3mut
EAAAKGGSGGG	SFV3L_P27401
GGGPAPGGS	SFV3L_P27401
GSSEAAAKGGS	FOAMV_P14350_2mutA
EAAAKGGSGGG	HTL1L_P0C211
GSSGGSPAP	SFV3L_P27401_2mutA
PAPAP	FFV_O93209
PAPGGSGSS	MMTVB_P03365_WS
EAAAKGGGGGS	FOAMV_P14350_2mut
PAPEAAAKGGS	SFV3L_P27401_2mut
GSSEAAAKPAP	MMTVB_P03365-Pro
GGSGGS	PERV_Q4VFZ2_3mut
GSSEAAAKGGG	FFV_O93209-Pro_2mutA
EAAAK	HTL1L_P0C211
GSSPAP	MLVMS_P03355
EAAAKPAPGGG	FFV_O93209-Pro_2mut
GGGGSEAAAKGGGGS	SFV1_P23074-Pro_2mut
EAAAKGSSGGS	SFV3L_P27401
GSAGSAAGSGEF	FFV_O93209_2mutA
PAPEAAAKGGS	MMTVB_P03365_2mutB_WS
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	MMTVB_P03365

[0941]

接头氨基酸序列	RT结构域名称
GGS	MMTVB_P03365
GGSEAAAKPAP	SFV1_P23074
EAAAKGSSGGG	HTLV2_P03363_2mut
GGSEAAAKGGG	MMTVB_P03365_WS
GGSGGS	FFV_O93209-Pro
GSSEAAAKGGS	MMTVB_P03365-Pro
PAPAPAPAPAP	SFV1_P23074_2mutA
GGSEAAAKGGG	MMTVB_P03365_2mutB_WS
PAPAPAPAP	MMTVB_P03365_WS
GGGGSGGGGSGGGGSGGGGSGGGGS	HTL3P_Q4U0X6_2mut
PAPGGSEAAAK	SFV1_P23074-Pro_2mut
GGSGGGPAP	MMTVB_P03365
GSSGSSGSSGSSGSSGSS	MMTVB_P03365-Pro
GGSEAAAKPAP	SFV1_P23074-Pro
GGGEAAAKGSS	SFV3L_P27401_2mutA
GGGPAPGGS	AVIRE_P03360
PAPGGG	MLVRD_P11227
GGSEAAAKGSS	SFV3L_P27401_2mut
GGGEAAAKGSS	FOAMV_P14350_2mut
GGGEAAAKGSS	SFV1_P23074-Pro
EAAAKEAAAKEAAAKEAAAK	MLVAV_P03356
EAAAKGGGPAP	JSRV_P31623_2mutB
EAAAKGGGGSS	FOAMV_P14350_2mut
EAAAKEAAAKEAAAKEAAAKEAAAK	SRV2_P51517
GSSGGGGGS	FFV_O93209
PAPAPAP	FOAMV_P14350_2mutA
GGSGGSGGSGGS	FOAMV_P14350
GGGEAAAK	MMTVB_P03365_WS
GGGGGS	SFV1_P23074_2mutA

[0942]

接头氨基酸序列	RT结构域名称
GGSGGS	WMSV_P03359_3mut
EAAAKGGS	MMTVB_P03365-Pro
GGGGSS	BLVJ_P03361_2mut
PAPAP	MMTVB_P03365-Pro_2mut
PAPGGG	SMRVH_P03364
EAAAKGGGGSS	SFV3L_P27401
PAPAPAPAPAP	MMTVB_P03365
GGGPAP	MMTVB_P03365-Pro
GSSGGSGGG	MMTVB_P03365
EAAAKGGGPAP	FOAMV_P14350_2mutA
GSSGSSGSSGSS	SFV1_P23074
GGGGSGGGGS	SFV3L_P27401
GSSGGSGGG	MLVF5_P26810
GGGEAAAKPAP	MMTVB_P03365-Pro
PAPEAAAK	HTLV2_P03363_2mut
GSSGSSGSSGSS	FOAMV_P14350_2mut
GSSEAAAKPAP	MMTVB_P03365-Pro
PAPEAAAKGGG	HTL3P_Q4U0X6_2mut
GGSEAAAKGSS	MMTVB_P03365-Pro
EAAAKPAPGGS	MMTVB_P03365_2mut_WS
GSSGGSEAAAK	MLVF5_P26810_3mutA
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	MLVF5_P26810_3mut
EAAAKGGGGSS	MMTVB_P03365-Pro
GGGGGSGSS	HTL1A_P03362_2mutB
PAPAP	FFV_O93209-Pro_2mut
GGGGGSPAP	HTL1C_P14078_2mut
GGGPAP	HTLV2_P03363_2mut
EAAAKGGGGSEAAAK	SFVCP_Q87040
GGSEAAAKGGG	FFV_O93209-Pro_2mutA

[0943]

接头氨基酸序列	RT结构域名称
GSSPAPGGS	FOAMV_P14350-Pro_2mut
GGGGGGG	MMTVB_P03365-Pro
EAAAKGSS	SFV3L_P27401_2mutA
EAAAKGGGGSEAAAK	MMTVB_P03365-Pro
GGGGSEAAAKGGGGS	SFV1_P23074-Pro_2mutA
EAAAKGGGGSS	MMTVB_P03365
GGGEAAAKGGS	SFV1_P23074
PAPEAAAKGGG	MLVF5_P26810
GGGGSSGGS	MMTVB_P03365
GGSGSS	MMTVB_P03365
PAPAPAPAPAPAP	KORV_Q9TTC1
EAAAKGGG	SFV1_P23074-Pro_2mut
PAPAPAPAPAPAP	SRV2_P51517
GSSGSSGSSGSSGSS	FFV_O93209-Pro_2mutA
GGGGSS	FOAMV_P14350_2mut
PAPGGGEAAAK	MMTVB_P03365_WS
GGSGGGEAAAK	FFV_O93209-Pro_2mut
PAPAPAPAPAP	MMTVB_P03365_WS
GGGEAAAKGGS	MMTVB_P03365-Pro
GGGEAAAKGSS	MMTVB_P03365_2mutB
GSSPAPEAAAK	MMTVB_P03365_WS
EAAAKEAAAKEAAAKEAAAKEAAAK	SFV1_P23074-Pro_2mutA
PAPGGG	SFV3L_P27401
GSSEAAAKGGG	MMTVB_P03365_WS
GGGGSSEAAAK	FOAMV_P14350_2mut
PAPGSSGGS	SFV1_P23074-Pro_2mut
GSSGSSGSSGSSGSSGSS	SFV3L_P27401
EAAAKGSSGGG	MMTVB_P03365
PAPGGGGSS	WDSV_O92815_2mutA

[0944]

接头氨基酸序列	RT 结构域名称
GGSPAP	MMTVB_P03365-Pro
GGSGGSGGSGGSGGS	SFVCP_Q87040-Pro_2mut
PAPAPAPAP	MMTVB_P03365-Pro
GGGGG	HTL1A_P03362
GGSGGSGGSGGS	SFV1_P23074_2mutA
GSSGSSGSSGSSGSS	FOAMV_P14350-Pro_2mut
PAPGGSEAAAK	MMTVB_P03365_2mutB_WS
PAPAPAPAP	SFV1_P23074_2mut
PAPGGGGSS	MMTVB_P03365
GGSGSS	SFV3L_P27401_2mut
EAAAKEAAAKEAAAKEAAAK	MMTVB_P03365_2mut
EAAAKGGSGGG	HTL3P_Q4U0X6_2mut
PAPGGGGSS	SFVCP_Q87040-Pro_2mutA
EAAAKGGGGGS	MLVAV_P03356
GGGGGS	FOAMV_P14350_2mut
GGGEAAAKGGS	FFV_O93209-Pro_2mutA
EAAAKPAPGGG	MMTVB_P03365_2mutB
GGSGGGPAP	FFV_O93209_2mut
GSSEAAAKPAP	MMTVB_P03365
PAPAPAPAPAPAP	SFV1_P23074_2mut
GGSPAPGGG	MMTVB_P03365-Pro
GGSGGGEAAAK	MMTVB_P03365
PAPAP	SFVCP_Q87040
GSSEAAAK	SFVCP_Q87040
GGGGSGGGGSGGGGS	MMTVB_P03365-Pro
GSSGSSGSS	SFV3L_P27401
EAAAKGGSGGG	MMTVB_P03365-Pro
GSSPAP	SFV1_P23074_2mut
GGGEAAAK	SFV1_P23074-Pro

[0945]

接头氨基酸序列	RT结构域名称
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	MMTVB_P03365-Pro
PAPGGS	HTL1C_P14078_2mut
PAPGSSGGS	SFV1_P23074_2mut
PAPEAAAK	MMTVB_P03365_WS
PAPAP	MMTVB_P03365-Pro
EAAAKGGS	HTL1A_P03362_2mut
GGGGSEAAAKGGGGS	HTL1C_P14078
EAAAKGSSGGS	FOAMV_P14350-Pro
PAPGGSGSS	MMTVB_P03365-Pro
PAPGGSEAAAK	SFV1_P23074_2mut
PAPGSSEAAAK	FFV_O93209-Pro_2mut
PAPGSSGGG	FOAMV_P14350-Pro_2mutA
GSSGGGEAAAK	AVIRE_P03360
GGGGGG	SMRVH_P03364_2mut
PAPEAAAKGGG	MMTVB_P03365-Pro
GGGEAAAKGGS	SFVCP_Q87040_2mutA
PAPAPAPAPAP	SRV2_P51517
GSSGSSGSSGSSGSSGSS	MMTVB_P03365
EAAAKGGGPAP	MLVAV_P03356
PAPAPAPAPAP	FOAMV_P14350-Pro_2mutA
PAPGGSEAAAK	FOAMV_P14350
GSSGGGPAP	HTL32_Q0R5R2_2mutB
GGGGGSPAP	HTL3P_Q4U0X6_2mutB
GSSGGSGGG	MMTVB_P03365-Pro
PAPAP	SFVCP_Q87040-Pro
GSSGGGPAP	MMTVB_P03365-Pro
GGSGSS	MMTVB_P03365-Pro_2mut
GGSPAPEAAAK	SFV1_P23074-Pro_2mut

[0946]

接头氨基酸序列	RT结构域名称
EAAAKGGSGGG	SFV3L_P27401_2mut
GGGGSSEAAAK	MMTVB_P03365_WS
GGGGGSGSS	MMTVB_P03365_2mut
GGGGSSGGS	SFV1_P23074-Pro_2mutA
EAAAKGGGGSEAAAK	MMTVB_P03365_WS
PAPGGGEAAAK	SFV1_P23074-Pro
PAPEAAAKGGG	MMTVB_P03365
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	MMTVB_P03365
GSSGGSEAAAK	FOAMV_P14350-Pro_2mut
GGSPAP	MLVBM_Q7SVK7_3mut
GSSEAAAK	FOAMV_P14350
GSSEAAAK	MMTVB_P03365-Pro
EAAAKGSSGGS	HTL1A_P03362_2mut
GGGEAAAKPAP	FOAMV_P14350-Pro_2mut
EAAAKGGSPAP	FOAMV_P14350
GSSEAAAKPAP	MMTVB_P03365_WS
GSSGSSGSS	FOAMV_P14350_2mut
EAAAKEAAAKEAAAKEAAAK	MMTVB_P03365_WS
EAAAK	MMTVB_P03365
PAPGSS	BAEVM_P10272
PAPGGS	FFV_O93209-Pro_2mut
GGSGGS	SFV1_P23074-Pro_2mutA
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	HTLV2_P03363_2mut
GGSGGGEAAAK	MMTVB_P03365_WS
PAPGSSGGG	HTL1A_P03362
GGSGGS	SFV3L_P27401-Pro
GSSGSS	SFV1_P23074-Pro
PAPGGSEAAAK	MMTVB_P03365

[0947]

接头氨基酸序列	RT 结构域名称
GSAGSAAGSGEF	MMTVB_P03365-Pro
PAPGGG	FOAMV_P14350_2mut
EAAAKGGSGSS	MMTVB_P03365_WS
GSSGGGEAAAK	SFV3L_P27401-Pro
GGSGGGPAP	FOAMV_P14350-Pro_2mut
PAPAPAPAPAPAP	WDSV_O92815
SGSETPGTSESATPES	SFVCP_Q87040-Pro_2mutA
GGSGGSGGS	SFV1_P23074
GGGGSS	SFVCP_Q87040_2mut
GGGGGSEAAAK	MMTVB_P03365
SGSETPGTSESATPES	MMTVB_P03365_WS
PAPAPAP	SFV3L_P27401
PAPEAAAKGSS	MMTVB_P03365_2mutB_WS
GSSGSSGSSGSSGSS	SRV2_P51517
GGGPAPGSS	HTL32_Q0R5R2_2mutB
GGSGGGGSS	MMTVB_P03365-Pro
SGSETPGTSESATPES	SRV2_P51517
EAAAKGSSGGS	MMTVB_P03365-Pro
GSSPAPEAAAK	MMTVB_P03365-Pro
GSSPAPEAAAK	SRV2_P51517
GGGGSSPAP	MMTVB_P03365-Pro
PAPGGGEAAAK	SFV1_P23074-Pro_2mutA
PAPEAAAKGGS	MMTVB_P03365
GSSGSSGSSGSSGSSGSS	FOAMV_P14350-Pro
GGSPAPGSS	SFV3L_P27401
GGGPAPGGS	SFV1_P23074-Pro_2mutA
GGGPAPGSS	MMTVB_P03365-Pro
EAAAKPAP	MLVBM_Q7SVK7
EAAAKEAAAKEAAAK	HTL1C_P14078

[0948]

接头氨基酸序列	RT 结构域名称
GSSGGSEAAAK	SRV2_P51517
PAPGGGGGS	SRV2_P51517
GGGEAAAK	FFV_O93209-Pro_2mut
EAAAKGGGPAP	HTL32_Q0R5R2
GGSGSSGGG	MMTVB_P03365
PAPEAAAKGSS	MMTVB_P03365-Pro
PAPGGGGGS	MMTVB_P03365-Pro
EAAAKGGGGGS	MMTVB_P03365_WS
GGGGGS	MMTVB_P03365-Pro
GGGGSGGGGSGGGGSGGGGSGGGGS	HTL1C_P14078
EAAAKGGSPAP	MMTVB_P03365
GGGGSSPAP	FFV_O93209-Pro_2mut
GGGGSSGGS	MMTVB_P03365-Pro
PAPGSSGGS	MMTVB_P03365-Pro
GGGGGS	SRV2_P51517
GGSGSSGGG	MMTVB_P03365
GSSGGSEAAAK	MMTVB_P03365-Pro
EAAAKEAAAKEAAAKEAAAK	GALV_P21414
GGSEAAAKGGG	MMTVB_P03365-Pro
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	MMTVB_P03365-Pro
GSSEAAAKGGS	MMTVB_P03365
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	HTL3P_Q4U0X6_2mutB
GGGEAAAK	MMTVB_P03365-Pro
PAPAPAPAP	MMTVB_P03365-Pro
PAPGSSGGG	MMTVB_P03365
GSSGSSGSSGSSGSS	GALV_P21414
GGSPAP	MMTVB_P03365_WS
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	MMTVB_P03365-Pro
PAPEAAAK	MMTVB_P03365-Pro

[0949]

接头氨基酸序列	RT 结构域名称
PAPGSSGGG	SFV1_P23074-Pro_2mutA
GGGGGSEAAAK	MMTVB_P03365_2mutB_WS
PAPAPAPAPAP	MMTVB_P03365-Pro
EAAAKGGSGSS	MMTVB_P03365-Pro
EAAAKEAAAKEAAAKEAAAK	MLVRD_P11227_3mut
PAPAPAPAP	FOAMV_P14350_2mutA
GGGPAPGSS	SFVCP_Q87040_2mut
PAPEAAAKGSS	SFVCP_Q87040_2mut
GGSPAPGGG	MMTVB_P03365-Pro
GGGGSGGGGSGGGGSGGGGS	MMTVB_P03365
EAAAKGGS	HTL3P_Q4U0X6_2mut
PAPGSSGGS	MMTVB_P03365_WS
GGGGSGGGGS	MMTVB_P03365
GGSGGS	FOAMV_P14350
EAAAKGGGGSEAAAK	SFVCP_Q87040-Pro_2mut
EAAAKEAAAKEAAAKEAAAK	MMTVB_P03365-Pro_2mutB
PAPGGGEAAAK	SFVCP_Q87040-Pro
GSSGSS	JSRV_P31623_2mutB
EAAAKGGGGGS	MMTVB_P03365_2mut_WS
GSSPAPEAAAK	MMTVB_P03365-Pro
GGGEAAAK	HTL1C_P14078
PAPEAAAKGSS	HTL32_Q0R5R2_2mutB
GGGGSSEAAAK	MMTVB_P03365-Pro
PAPGSSGGS	MMTVB_P03365-Pro
EAAAKGGGGGS	MMTVB_P03365
GGGGSGGGGSGGGGSGGGGS	MMTVB_P03365
EAAAKGGGGSS	HTL3P_Q4U0X6_2mut
GGGEAAAKGGS	SFVCP_Q87040-Pro
GGGGGSPAP	MMTVB_P03365-Pro_2mutB

[0950]

接头氨基酸序列	RT 结构域名称
GGSGGGEAAAK	SFV3L_P27401-Pro
PAPGGGGGS	SFV3L_P27401-Pro
EAAAKGGGGSEAAAK	MMTVB_P03365
PAPEAAAKGSS	MMTVB_P03365-Pro
GGSEAAAKGGG	MMTVB_P03365-Pro
GGSGGSGGSGGSGGS	SMRVH_P03364_2mutB
GGSGGSGGSGGSGGS	HTL1L_P0C211_2mut
GGGGGG	WDSV_O92815
GGGGGSGSS	MMTVB_P03365-Pro
GGSEAAAKPAP	SFV3L_P27401-Pro_2mut
GGGPAPGSS	MMTVB_P03365_2mut_WS
GGGGGS	MMTVB_P03365_WS
GGSPAPEAAAK	MMTVB_P03365
PAPEAAAKGGS	HTL1A_P03362
EAAAKGGSGSS	MMTVB_P03365_2mut_WS
GGGPAPEAAAK	SFV3L_P27401-Pro_2mut
PAPGGGGSS	HTL32_Q0R5R2_2mut
GSSPAPGGG	HTL3P_Q4U0X6_2mut
GGGGSSGGS	BLVAU_P25059_2mut
EAAAKGGGGGS	HTL1L_P0C211
GGSEAAAKGSS	JSRV_P31623_2mutB
GSSGGG	JSRV_P31623
GGSGGSGGSGGS	MMTVB_P03365-Pro
EAAAKPAP	SFV1_P23074-Pro_2mutA
GGGGSSGGS	MMTVB_P03365_WS
GGSGGS	MMTVB_P03365_WS
EAAAKGGGGGS	MMTVB_P03365-Pro
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	MMTVB_P03365
GGSGGSGGS	MMTVB_P03365

[0951]

接头氨基酸序列	RT结构域名称
GGGGGSEAAAK	MLVBM_Q7SVK7
GGSGSSPAP	MMTVB_P03365_WS
EAAAKEAAAKEAAAK	JSRV_P31623
PAPEAAAKGGS	MMTVB_P03365-Pro
GGSGSSEAAAK	FOAMV_P14350
GGGGGSGSS	MMTVB_P03365-Pro_2mut
GGGPAPGGS	MMTVB_P03365
SGSETPGTSESATPES	SFVCP_Q87040_2mut
GSSPAPGGS	SFV1_P23074-Pro_2mutA
GSSGSSGSSGSSGSS	MMTVB_P03365
EAAAKGGGPAP	MMTVB_P03365
GSSGGG	MMTVB_P03365_2mut_WS
GGGEAAAKPAP	MMTVB_P03365
PAPGGSGGG	MMTVB_P03365-Pro
GSSGGSGGG	WDSV_O92815_2mut
GGSGGG	HTL32_Q0R5R2_2mut
EAAAKGGSPAP	HTLV2_P03363_2mut
GGSPAPEAAAK	MMTVB_P03365-Pro
GSSGGSEAAAK	MMTVB_P03365_2mut
GSAGSAAGSGEF	MMTVB_P03365_WS
PAPGGSGSS	FFV_O93209
GGSEAAAKGGG	MMTVB_P03365
GGSPAPGSS	MMTVB_P03365-Pro
GSSGGSGGG	SFV3L_P27401
PAPEAAAKGGG	HTL1A_P03362_2mutB
GGGEAAAKPAP	MMTVB_P03365-Pro
GGSEAAAK	HTL32_Q0R5R2_2mutB
GGGEAAAKGSS	MPMV_P07572
GGGGGSEAAAK	MMTVB_P03365-Pro

[0952]

接头氨基酸序列	RT 结构域名称
PAPAPAPAPAP	SFVCP_Q87040-Pro_2mutA
PAPAPAPAPAP	HTL1L_P0C211_2mut
GGGGSSGGS	HTL3P_Q4U0X6
PAPGGSEAAAK	MMTVB_P03365_2mut_WS
PAPAPAPAPAP	HTL1A_P03362
EAAAKPAPGGG	MMTVB_P03365_2mut_WS
GGSEAAAK	MMTVB_P03365_2mut_WS
GGGEAAAKGSS	SFV1_P23074-Pro_2mutA
GGSPAPGSS	MMTVB_P03365-Pro
GGSEAAAKPAP	MLVBM_Q7SVK7
PAPEAAAKGGG	MMTVB_P03365_2mut_WS
GSSEAAAKPAP	MMTVB_P03365-Pro_2mutB
GGGGSEAAAKGGGGS	MMTVB_P03365-Pro_2mut
GSSEAAAKGGS	MMTVB_P03365-Pro_2mutB
GSSGSSGSSGSSGSS	SRV2_P51517_2mutB
GGGGGSPAP	HTL1L_P0C211_2mut
GGSEAAAK	MMTVB_P03365
GSSPAPEAAAK	SMRVH_P03364_2mutB
GGGPAPGGS	HTL1C_P14078_2mut
GGSPAPEAAAK	MMTVB_P03365_WS
GGSEAAAKPAP	HTL1A_P03362_2mut
PAPAPAPAP	HTLV2_P03363_2mut
GSSPAPGGG	MMTVB_P03365
GSSGSSGSSGSS	MMTVB_P03365-Pro
GGSEAAAKGSS	MMTVB_P03365_WS
GGSGSSGGG	MMTVB_P03365_2mutB
GSSGSSGSSGSSGSSGSS	JSRV_P31623_2mutB
GGSEAAAKPAP	MMTVB_P03365-Pro
GSSGGSGGG	HTLV2_P03363_2mut

[0953]

接头氨基酸序列	RT结构域名称
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	WDSV_O92815_2mut
GGSPAPEAAAK	MMTVB_P03365
GGGGSSEAAAK	MMTVB_P03365
GGSGGGEAAAK	SFV1_P23074-Pro_2mutA
GGGGSEAAAKGGGGS	WDSV_O92815_2mut
GGSGSSEAAAK	MMTVB_P03365_2mutB_WS
GGSEAAAKPAP	MMTVB_P03365_WS
GSSGGGEAAAK	SFVCP_Q87040-Pro
GSSGGS	SFVCP_Q87040-Pro_2mut
GGSEAAAKPAP	SFVCP_Q87040_2mut
GSSGGSEAAAK	SFVCP_Q87040_2mut
GSSPAPEAAAK	SRV2_P51517_2mutB
GGSGGSGGSGGSGGSGGS	BLVAU_P25059
GSSGSSGSSGSSGSS	HTL1C_P14078_2mut
EAAAKGGGGSS	MMTVB_P03365_2mutB
GGGEAAAKGSS	SFVCP_Q87040-Pro

[0954] 实例4:包含具有各种人DNA聚合酶θ结构域的Cas-Pol融合体的基因修饰系统实现人细胞中的精确编辑

[0955] 在该实例中,对包含Cas结构域和聚合酶结构域(Cas-Pol融合体)的示例性基因修饰多肽进行了表征,并确定了它们在人细胞基因组DNA中以可编程方式安装突变的能力。基因修饰多肽包含示例性人聚合酶(人DNA聚合酶θ(Polθ))的不同截短(图2,表23):基因修饰多肽包括聚合酶结构域、外切核酸酶结构域和两个截短的中心结构域(Polθ_M或Polθ_L)中的任一个或缺少中心结构域(Polθ_4x0q)。基因修饰多肽包含示例性Cas结构域SpCas9,其包含N863A突变(nCas9)。使用两种不同的接头序列来连接Cas结构域和Pol结构域:

[0956] 示例性接头FL:

[0957] DNA序列:

[0958] TCTGGAGGATCTAGCGGAGGATCCTCTGGCAGCGAGACACCAGGAACAAGCGAGTCAGCAACACCAGAGAGCAGTGGCGGCAGCAGCGGCGGCAGCAGC(SEQ ID NO:4400)

[0959] 氨基酸序列:

[0960] SGGSSGGSSGSETPGTSESATPESSGGSSGGSS(SEQ ID NO:4401)

[0961] 示例性接头UL:

[0962] DNA序列:

[0963] GCCGAAGCCGCCGCCAAGGAGGCCGCCGCTAAGGAGGCTGCTGCCAAGGAAGCTGCTGCTAAGGCTTT

AGAAGCTGAAGCTGCTGCCAAAGAAGCTGCCGCCAAAGAGGCTGCCGCAAAGGAGGCCGCTGCCAAGGCT(SEQ ID NO:4402)

[0964] 氨基酸序列:

[0965] AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA(SEQ ID NO:4403)

[0966] 测试了六种基因修饰多肽(Cas-Pol融合体):(1)nCas9-FL-Polθ_M、(2)nCas9-FL-Polθ_4x0q、(3)nCas9-UL-Polθ_L、(4)nCas9-UL-Polθ_M、(5)nCas9-UL-Polθ_4x0q和(6)包含nCas9和逆转录酶结构域而不是聚合酶结构域的基准基因修饰多肽。

[0967] 表23:示例性Cas-Pol融合体基因修饰多肽

[0968]

名称	氨基酸序列	SEQ ID NO
Polθ_L	GASFDLSPGLQRILDKVSSPLENEKLKSMTINFSSLNRKNTELNEEQEVISNLETKQVQGISFSSNNEVKS KIEMLENNANHDETSSLLPRKESNIVDDNGLIPPTPIPTSASKLTFPGILETPVNPWKTNNVLQPGESYLFG SPSDIKNHDLSPGSRNGFKDNSPISDTSFSLQLSQDGLQLTPASSSSESLSIIDVASDQNLFQTFIKEWRC KKRFSISLACEKIRSLTSSKTATIGSRFKQASSPQEIPIRDDGFPIKGCDDTLVVGLAVCWGGRDAYYFSL QKEQKHSEISASLVPPSLDPSLTLKDRMWYLQSCLRKESDKECSVVIYDFIQSYKILLLSCGISLEQSYEDP KVACWLLDPDSQEPTLHSIVTSFLPHELPLLEGMETSQGIQSLGLNAGSEHSGRYRASVESILIFNSMNQL NSLLQKENLQDVFRKVEMPSQYCLALLELNGIGFSTAECESQKHIMQAKLDAIETQAYQLAGHSFSFTSS DDIAEVLFLELKLPPNREMKNQGSKKTLGSTRRGIDNGRKLRLGRQFSTSKDVLNKLKALHPLPGLILEW RRITNAITKVVFPLQREKCLNPFLGMERIYPVSQSHTATGRITFTEPNIQNVPRDFEIKMPTLVGESPPSQA VGKGLLPMGRGKYKKGFSVNPRCQAQMEERAADRGMPFSISMRHAFVPFPGGSILAADYSQLELRILAH LSHDRRLIQVLNTGADVFRSIAAEWKMIEPESVGDDLRQQAKQICYGIIYGMGAKSLGEQMGIKENDAAC YIDSFKSRYTGINQFMTETVKNCKRDGFVQTILGRRRYLPGIKDNNPYRKAHAERQAINTIVQGSAADIVKI ATVNIQKQLETFHSTFKSHGHREGMLQSDQTGLSRKRKLQGMFCPIRGGFFILQLHDELLYEVAEEDVVQ VAQIVKNEMESAVKLSVKLKVKVKIGASWGELKDFDV	4404
Polθ_M	GFKDNSPISDTSFSLQLSQDGLQLTPASSSSESLSIIDVASDQNLFQTFIKEWRCKKRFSISLACEKIRSLT SSKTATIGSRFKQASSPQEIPIRDDGFPIKGCDDTLVVGLAVCWGGRDAYYFSLQKEQKHSEISASLVPPS LDPSLTLKDRMWYLQSCLRKESDKECSVVIYDFIQSYKILLLSCGISLEQSYEDPKVACWLLDPDSQEPTL HSIVTSFLPHELPLLEGMETSQGIQSLGLNAGSEHSGRYRASVESILIFNSMNQLNSLLQKENLQDVFRKV EMPSQYCLALLELNGIGFSTAECESQKHIMQAKLDAIETQAYQLAGHSFSFTSSDDIAEVLFLELKLPPNR EMKNQGSKKTLGSTRRGIDNGRKLRLGRQFSTSKDVLNKLKALHPLPGLILEWRRITNAITKVVFPLQREK CLNPFLGMERIYPVSQSHTATGRITFTEPNIQNVPRDFEIKMPTLVGESPPSQAVGKGLLPMGRGKYKKG FSVNPRCQAQMEERAADRGMPFSISMRHAFVPFPGGSILAADYSQLELRILAHLSHDRRLIQVLNTGADV FRSIAAEWKMIEPESVGDDLRQQAKQICYGIIYGMGAKSLGEQMGIKENDAACYIDSFKSRYTGINQFMTE TVKNCKRDGFVQTILGRRRYLPGIKDNNPYRKAHAERQAINTIVQGSAADIVKIATVNIQKQLETFHSTFKS HGHREGMLQSDQTGLSRKRKLQGMFCPIRGGFFILQLHDELLYEVAEEDVVQVAQIVKNEMESAVKLSV KLKVKVKIGASWGELKDFDV	4405
Polθ_4x0q	SSSSESLSIIDVASDQNLFQTFIKEWRCKKRFSISLACEKIRSLTSSKTATIGSRFKQASSPQEIPIRDDGFPI KGCDDTLVVGLAVCWGGRDAYYFSLQKEQKHSEISASLVPPSLDPSLTLKDRMWYLQSCLRKESDKEC	4406

[0969]

	SVVIYDFIQSYKILLLSCGISLEQSYEDPKVACWLLDPDSQEPTLHSIVTSFLPHELPLLEGMETSQGIQSLG LNAGSEHSGRYRASVESILIFNSMNQLNSLLQKENLQDVFRKVEMPSQYCLALLELNGIGFSTAECESQK HIMQAKLDAIETQAYQLAGHSFSFTSSDDIAEVLFLELKLPPNREMKNQGSKKTLGSTRRGIDNGRKLRLG RQFSTSKDVLNKLKALHPLPGLILEWRRITNAITKVVFPLQREKCLNPFLGMERIYPVSQSHTATGRITFTE PNIQNVPRDFEIKMPTLVGESPPSQAVGKGLLPMGRGKYKKGFSVNPRCQAQMEERAADRGMPFSISM RHAFVPFPGGSILAADYSQLELRILAHLSHDRRLIQVLNTGADVFRSIAAEWKMIEPESVGDDLRQQAKQI CYGIIYGMGAKSLGEQMGIKENDAACYIDSFKSRYTGINQFMTETVKNCKRDGFVQTILGRRRYLPGIKDN NPYRKAHAERQAINTIVQGSAADIVKIATVNIQKQLETFHSTFKSHGHREGMLQSDQTGLSRKRKLQGMF CPIRGGFFILQLHDELLYEVAEEDVVQVAQIVKNEMESAVKLSVKLKVKVKIGASWGELKDFDV	
Polθ_4x0q_GS	SSSSESLSIIDVASDQNLFQTFIKEWRCKKRFSISLACEKIRGSGDDTLVVGLAVCWGGRDAYYFSLGGS GGLDPSLTLKDRMWYLQSCLRKESDKECSVVIYDFIQSYKILLLSCGISLEQSYEDPKVACWLLDPDSQE PTLHSIVTSFLPHELPLLEGMETSQGIQSLGLNAGSEHSGRYRASVESILIFNSMNQLNSLLQKENLQDVF RKVEMPSQYCLALLELNGIGFSTAECESQKHIMQAKLDAIETQAYQLAGHSFSFTSSDDIAEVLFLELKLP PGGSGGQFSTSKDVLNKLKALHPLPGLILEWRRITNAITKVVFPLQREKCLNPFLGMERIYPVSQSHTATG RITFTEPNIQNVPRDFEIKMGGSGGMPFSISMRHAFVPFPGGSILAADYSQLELRILAHLSHDRRLIQVLNT GADVFRSIAAEWKMIEPESVGDDLRQQAKQICYGIIYGMGAKSLGEQMGIKENDAACYIDSFKSRYTGIN QFMTETVKNCKRDGFVQTILGRRRYLPGIKDNNPYRKAHAERQAINTIVQGSAADIVKIATVNIQKQLETFH STFKSHGHREGMLQSDGGSGGCPIRGGFFILQLHDELLYEVAEEDVVQVAQIVKNEMESAVKLSVKLKV KVKIGASWGELKDFDV	4407
nCas9	DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRY TRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVD STDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLS KSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYAD LFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAG YIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFL KDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPN EKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECF DSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVM KQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGD SLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRS DKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYP KLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVW DKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVL VVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS AGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANL DKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRID LSQLGGD	4408

[0970] 构建编码直接单碱基对取代的测试模板核酸,当通过示例性基因修饰多肽作为编辑引入时,其起到将外源BFP标志物转化为GFP的作用(如本文所述)。模板核酸包含(i)用于将基因修饰多肽引导至靶区域的gRNA间隔子序列,例如,与BFP基因座中的20个核苷酸序列互补的序列;(ii)引物结合序列(PBS),其能够与带切口的DNA的单链互补碱基配对以进行靶引发的聚合;(iii)提供聚合模板的异源对象序列,其进一步包含预期的最终靶序列;和(iv)与基因修饰多肽的Cas9结构域相关联的gRNA支架序列。示例性模板核酸从5′到3′按照以下顺序排列:(i)-(iv)-(iii)-(ii)。模板组成描述于表24(RNACS-4239-DD和RNACS-4241-DR)中,其中(ii)和(iii)可各自作为RNA或DNA核苷酸包括在内。

[0971] 表24:实例1中使用的模板核酸

[0972]

名称	描述	间隔子 (i)	支架 (iv)	Pol模板 (iii)	PBS (ii)	全模板分子
模板DD	BFP_g4_11RT_13PBS_DD	mG*mC*mC*rGrArArGrCrArCrUrGrCrArCrGrCrCrGrU (SEQ ID NO: 4409)	rGrUrUrUrUrArGrArGrCrUrArCrArGrArArArUrArArCrArArGrUrUrArArArArArUrArArArA	CCTGACGTACG (SEQ ID NO: 4413)	GCGTGCAGTG*mC*mU*m	mG*mC*mC*rGrArArGrCrArCrUrGrCrArCrGrCrCrGrUrGrUrUrUrUrArGrArGrCrUrArArArArArUrArGrCrArA

[0973]

名称	描述	间隔子 (i)	支架 (iv)	Pol模板 (iii)	PBS (ii)	全模板分子
			GrGrCrUrArGrUrCrCrGrUrUrArUrCrArArCrUrUrGrArArArArArGrUrGrGrCrArCrCrGrArGrUrCrGrGrUrGrC（SEQ ID NO: 4411）		U（SEQ ID NO: 4415）	rGrUrUrArArArArUrArArGrGrCrUrArGrUrCrCrGrUrUrArUrCrArArCrUrUrGrArArArArArArArGrUrGrGrCrArCrCrGrArGrUrCrGrGrUrGrCCCTGACGTACGGCGTGCAGTG*mC*mU*mU（SEQ ID NO: 4417）
模板DR	BFP_g4_11RT_13PBS_DR	mG*mC*mC*rGrArArGrCrArCrUrGrCrArCrGrCrCrGrU（SEQ ID NO: 4410）	rGrUrUrUrUrArGrArGrCrArArArArUrArArArArUrArArGrGrCrCrArGrUrCrCrGrUrUrArUrCrArArCrUrUrGrArArArArArArGrUrGrGrCrArCrCrGrArGrUrCrGrGrUrGrC（SEQ ID NO: 4412）	CCTGACGTACG（SEQ ID NO: 4414）	rGrCrGrUrGrCrArGrUrG*mC*mU*mU（SEQ ID NO: 4416）	mG*mC*mC*rGrArArGrCrArCrUrGrCrArCrGrCrCrGrUrGrUrUrUrUrArGrArGrCrUrArGrArArArUrArArGrGrCrCrArGrUrCrCrGrUrUrArUrCrArArCrUrUrGrArArArArArArGrUrGrGrCrArCrCrGrArGrUrCrGrGrUrGrCCCTGACGTACGrGrCrGrUrGrCrArGrUrG*mC*mU*mU（SEQ ID NO: 4418）
模板RR	BFP_g4_11RT_13PBS_RR	mG*mC*mC*rGrArArGrCrArCrUrGrCrArCrGrCrCrGrU（SEQ ID NO: 4419）	rGrUrUrUrUrArGrArGrCrArArArArUrArArArArUrArArGrGrCrCrArGrUrCrCrGrUrUrArUrCrArArCrUrUrGrArArArArArArGrUrGrGrCrArCrCrGrArGrUrCrGrGrUrGrC（SEQ ID NO: 4420）	rCrCrUrGrArCrGrUrArCrG（SEQ ID NO: 4421）	rGrCrGrUrGrCrArGrUrG*mC*mU*mU（SEQ ID NO: 4422）	mG*mC*mC*rGrArArGrCrArCrUrGrCrArCrGrCrCrGrUrGrUrUrUrUrArGrArGrCrUrArGrArArArUrArArGrGrCrCrArGrUrCrCrGrUrUrArUrCrArArCrUrUrGrArArArArArArGrUrGrGrCrArCrCrGrArGrUrCrGrGrUrGrCrCrUrGrArCrGrUrArCrGrGrCrGrUrGrCrArGrUrG*mC*mU*mU（SEQ ID NO: 4423）

[0974] RNA,表示为“r_”,例如rA、rU;DNA,表示为“_”,例如A、T;2'O-甲基RNA碱基表示为“m_”;硫代磷酸酯化2'-O-甲基RNA碱基输入为“m_*”。

[0975] 进行了实验以确定Cas-Pol融合体基因修饰多肽的基因组编辑能力。

[0976] 将U2OS或HEK293T细胞用编码示例性基因修饰多肽并含有抗生素(嘌呤霉素)抗性盒的慢病毒载体感染。通过在补充嘌呤霉素的培养基中培养细胞来选择稳定表达基因修饰多肽的细胞。用2.5uM和5uM的化学合成的示例性模板核酸分子对250,000个细胞/孔(96孔板)进行核转染。核转染后第3天对细胞进行荧光激活细胞分选(FACS,流式细胞术(flow))。预期编辑(BFP阳性细胞转化为GFP阳性细胞)的频率被确定为GFP阳性细胞的百分比相对于BFP阳性细胞百分比(图3A和3B)。

[0977] 结果表明,基因修饰多肽nCas9-UL-Polθ_L、nCas9-UL-Polθ_M、nCas9-UL-Polθ_4x0q、nCas9-FL-Polθ_M和nCas9-FL-Polθ_4x0q表现出编辑活性,例如,在HEK293细胞中编辑效率高达约75%。使用包含DNA编码的异源对象序列和RNA引物结合序列的模板核酸(表示为“DR”),基因修饰多肽显示出高编辑活性(与含有对照逆转录酶的基准基因修饰多肽相当或更高)(图3A-B)。使用包含DNA异源对象序列和DNA引物结合序列的模板核酸(表示为“DD”),基因修饰多肽也表现出的编辑活性,尽管比使用RNA引物结合序列时低(图3A-B)。使用包含RNA异源对象序列和RNA引物结合序列的模板核酸(表示为“RR”),基因修饰多肽没有显示出编辑活性(图3A-B)。编辑水平与基准基因修饰多肽相当或更高,表明使用聚合酶结构域的基因修饰多肽是使用逆转录酶结构域的基因修饰多肽的有用替代物。结果进一步证明,Polθ的各种截短均可用作聚合酶结构域,因为数据显示,包含任何测试截短(包括中心结构域的完全截短)的基因修饰多肽都具有编辑活性。

[0978] 应该理解,对于本申请中描述某个参数的所有数值界限,诸如“约”、“至少”、“小

于”和“大于”而言,该描述还必然涵盖以列举的值为界限的任何范围。因此,例如,描述“至少1、2、3、4或5”还特别地描述了范围1-2、1-3、1-4、1-5、2-3、2-4、2-5、3-4、3-5和4-5等。

[0979] 对于本文引用的所有专利、申请或其他参考文献,例如非专利文献和参考序列信息,应当理解,出于所有目的以及对于所陈述的主张,将它们通过援引以其全文并入本文。如果通过援引并入的文件与本申请之间存在任何冲突,则以本申请为准。与本申请中披露的参考基因序列相关的所有信息,例如GeneID或登录号(通常参考NCBI登录号),包括例如基因组基因座、基因组序列、功能注释、等位基因变体和参考mRNA(包括,例如,外显子边界或响应元件)和蛋白质序列(例如保守结构域结构),以及化学参考(例如PubChem化合物、PubChem物质或PubChem生物测定条目,包括其中的注释,例如结构和测定等),通过援引以其全文并入本文。

[0980] 本申请中使用的标题仅为方便起见并且不影响对本申请的解释。

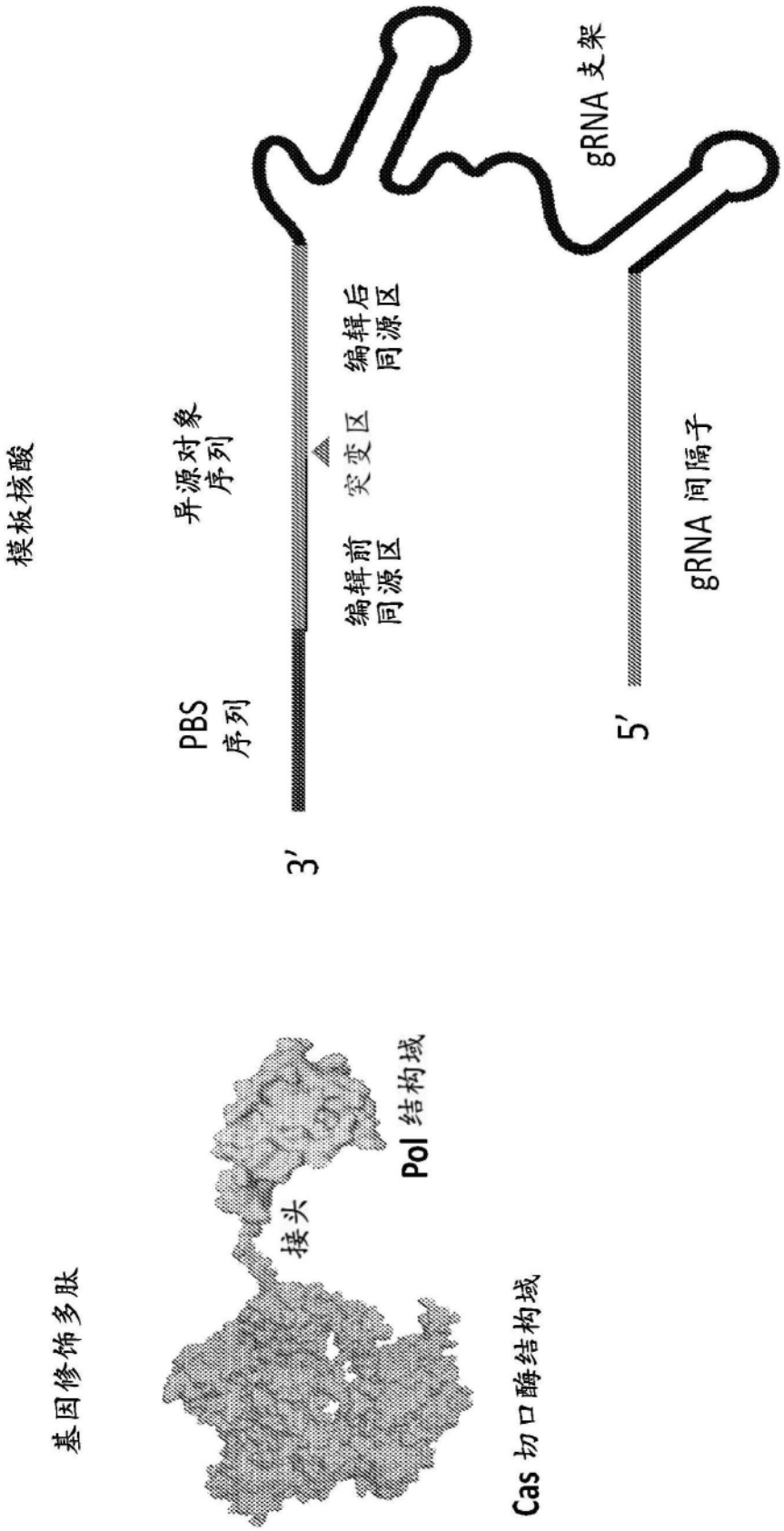

模板核酸
异源对象序列
PBS序列
编辑前同源区
突变区
编辑后同源区
gRNA 支架
gRNA 间隔子
3′
5′
基因修饰多肽
接头
Pol 结构域
Cas 切口酶结构域

图1

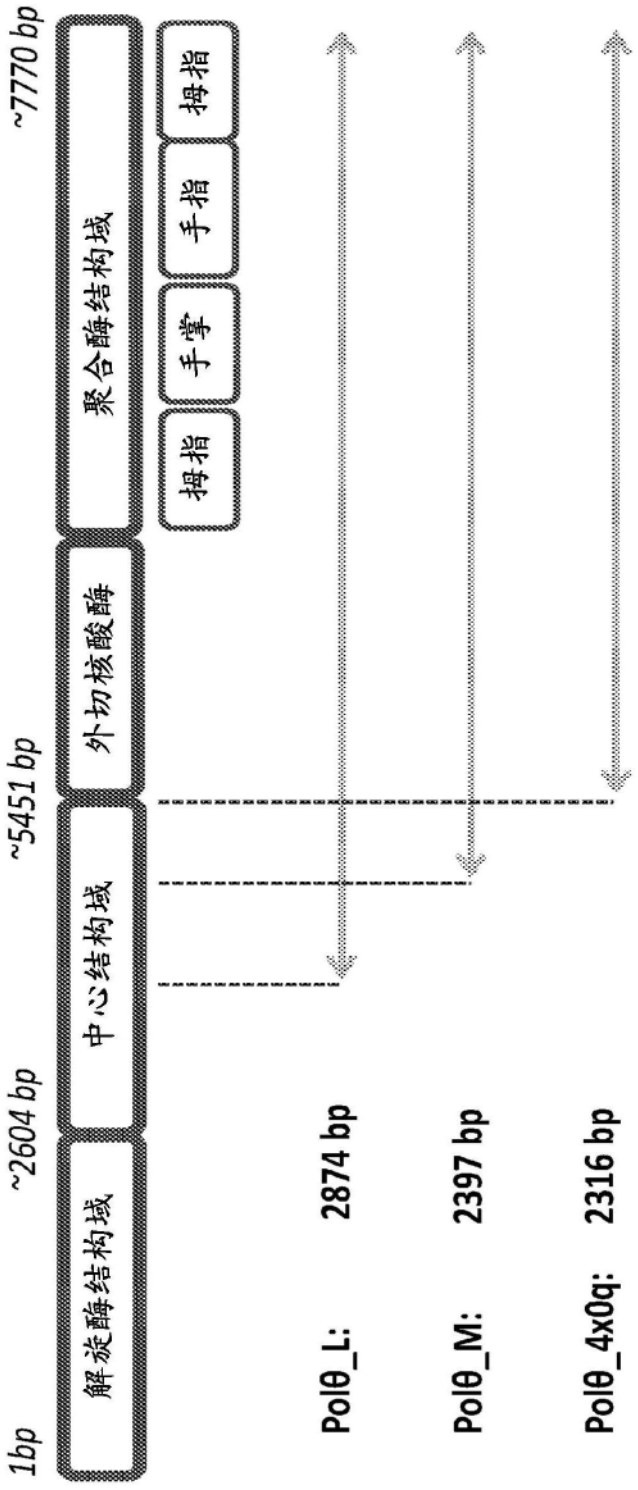

1bp
~2604 bp
~5451 bp
~7770 bp
解旋酶结构域
中心结构域
外切核酸酶
聚合酶结构域
拇指
手掌
手指
拇指
Polθ_L: 2874 bp
Polθ_M: 2397 bp
Polθ_4x0q: 2316 bp

图2

图 3A

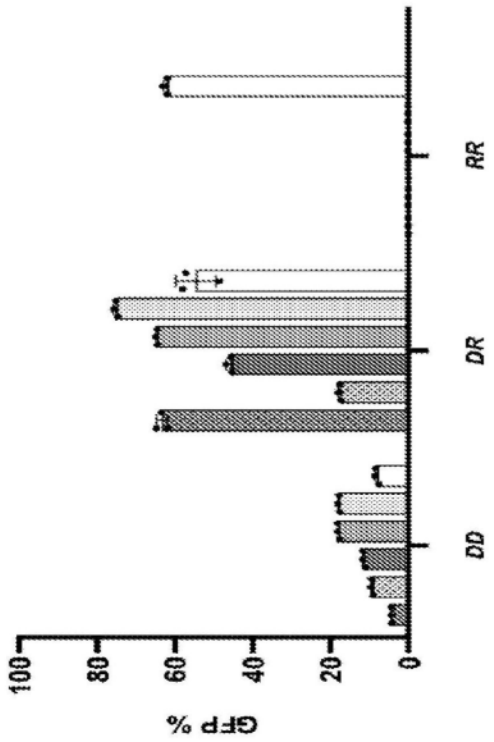
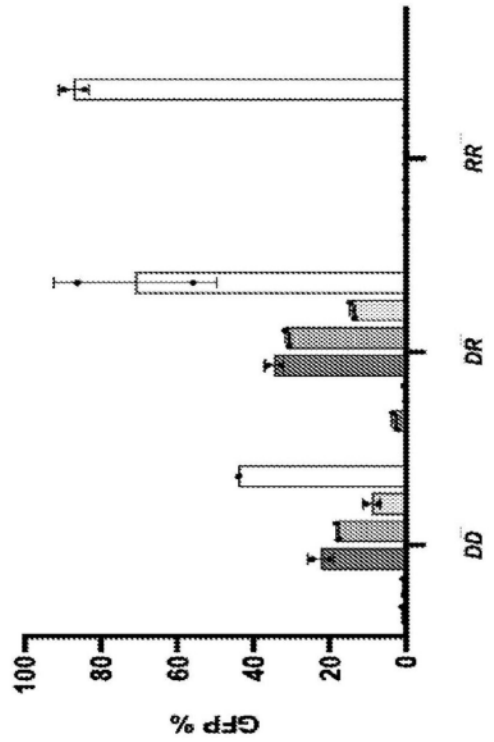
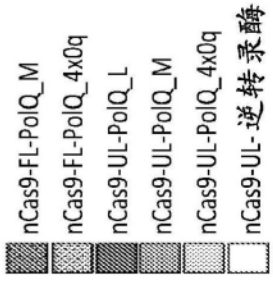

HEK293
GFP %
100
80
60
40
20
0
DD
DR
RR

图 3B


U2OS
GFP %
100
80
60
40
20
0
DD
DR
RR


nCas9-FL-PolQ_M
nCas9-FL-PolQ_4x0q
nCas9-UL-PolQ_L
nCas9-UL-PolQ_M
nCas9-UL-PolQ_4x0q
nCas9-UL-逆转录酶