(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 115698041 A
(43)申请公布日 2023.02.03

(21)申请号 202080096862.2
(22)申请日 2020.12.22
(30)优先权数据
62/952,909 2019.12.23 US
(85)PCT国际申请进入国家阶段日
2022.08.17
(86)PCT国际申请的申请数据
PCT/US2020/066672 2020.12.22
(87)PCT国际申请的公布数据
WO2021/133829 EN 2021.07.01
(71)申请人 加利福尼亚大学董事会
地址 美国加利福尼亚
(72)发明人 J·F·班菲尔德 J·A·多德纳 B·阿尔-沙耶布
(74)专利代理机构 中国贸促会专利商标事务所有限公司 11038
专利代理师 陈晓娜
(51)Int.Cl.
C07K 14/195(2006.01)
C12Q 1/34(2006.01)
C12N 9/22(2006.01)
A61K 38/46(2006.01)

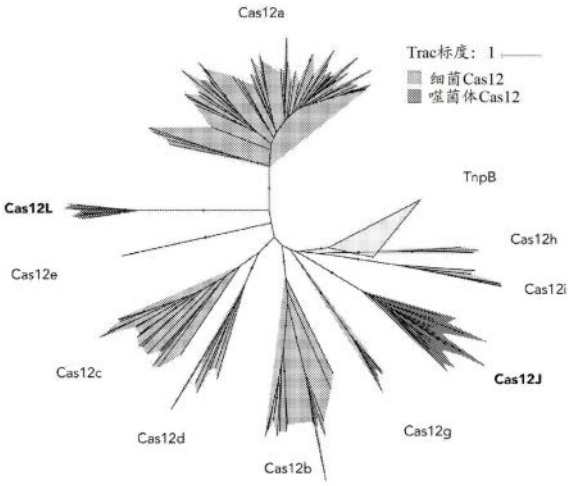
权利要求书9页 说明书134页
序列表165页 附图60页

(54)发明名称

CRISPR-CAS效应多肽及其使用方法

(57)摘要

本公开提供了RNA引导的CRISPR-Cas效应蛋白、编码所述蛋白的核酸以及包含所述蛋白的组合物。本公开提供了核糖核蛋白复合物,其包含:本公开的RNA引导的CRISPR-Cas效应蛋白;和引导RNA。本公开提供了使用本公开的RNA引导的CRISPR-Cas效应蛋白和引导RNA修饰靶核酸的方法。



CN 115698041 A

1.一种组合物,其包含:

a)Cas12L多肽或编码所述Cas12L多肽的核酸分子;和

b)Cas12L引导RNA或编码所述Cas12L引导RNA的一个或多个DNA分子。

2.如权利要求1所述的组合物,其中所述Cas12L多肽包含与图2A-2Z、图2AA-2ZZ和图2AAA-2FFF中任一者中描绘的氨基酸序列具有50%或更多氨基酸序列同一性的氨基酸序列。

3.如权利要求1或权利要求2所述的组合物,其中所述Cas12L引导RNA包含与图7、图2A-2Z、图2AA-2ZZ和图2AAA-2FFF中描绘的crRNA序列中的任一者具有80%、90%、95%、98%、99%或100%核苷酸序列同一性的核苷酸序列。

4.如权利要求1或权利要求2所述的组合物,其中所述Cas12L多肽融合至核定位信号(NLS)。

5.如权利要求1-4中任一项所述的组合物,其中所述组合物包含脂质。

6.如权利要求1-4中任一项所述的组合物,其中a)和b)在脂质体内。

7.如权利要求1-4中任一项所述的组合物,其中a)和b)在颗粒内。

8.如权利要求1-7中任一项所述的组合物,其包含缓冲剂、核酸酶抑制剂和蛋白酶抑制剂中的一种或多种。

9.如权利要求1-8中任一项所述的组合物,其中所述Cas12L多肽包含与图2A-2Z、图2AA-2ZZ和图2AAA-2FFF中任一者中描绘的氨基酸序列具有85%或更多同一性的氨基酸序列。

10.如权利要求1-9中任一项所述的组合物,其中所述Cas12L多肽是仅切割双链靶核酸分子的一条链的切口酶。

11.如权利要求1-9中任一项所述的组合物,其中所述Cas12L多肽是无催化活性的Cas12L多肽(dCas12L)。

12.如权利要求1-11中任一项所述的组合物,其中所述Cas12L多肽具有700个氨基酸至800个氨基酸的长度。

13.如权利要求1-12中任一项所述的组合物,其还包含DNA供体模板。

14.一种Cas12L融合多肽,其包含:与异源多肽融合的Cas12L多肽。

15.如权利要求14所述的Cas12L融合多肽,其中所述Cas12L多肽包含与图2A-2Z、图2AA-2ZZ和图2AAA-2FFF中任一者中描绘的氨基酸序列具有50%或更多同一性的氨基酸序列。

16.如权利要求14所述的Cas12L融合多肽,其中所述Cas12L多肽包含与图2A-2Z、图2AA-2ZZ和图2AAA-2FFF中任一者中描绘的氨基酸序列具有85%或更多同一性的氨基酸序列。

17.如权利要求14-16中任一项所述的Cas12L融合多肽,其中所述Cas12L多肽是仅切割双链靶核酸分子的一条链的切口酶。

18.如权利要求14-17中任一项所述的Cas12L融合多肽,其中所述Cas12L多肽是无催化活性的Cas12L多肽(dCas12L)。

19.如权利要求14-18中任一项所述的Cas12L融合多肽,其中所述Cas12L多肽具有700个氨基酸至800个氨基酸的长度。

20.如权利要求14-19中任一项所述的Cas12L融合多肽,其中所述异源多肽融合到所述Cas12L多肽的N-末端和/或C-末端。

21.如权利要求14-20中任一项所述的Cas12L融合多肽,其包含核定位信号(NLS)。

22.如权利要求14-21中任一项所述的Cas12L融合多肽,其中所述异源多肽是提供与靶细胞或靶细胞类型上的细胞表面部分结合的靶向多肽。

23.如权利要求14-21中任一项所述的Cas12L融合多肽,其中所述异源多肽表现出酶促活性。

24.如权利要求23所述的Cas12L融合多肽,其中所述异源多肽表现出选自以下的一种或多种酶促活性:核酸酶活性、甲基转移酶活性、脱甲基化酶活性、DNA修复活性、DNA损伤活性、脱氨基活性、歧化酶活性、烷基化活性、脱嘌呤活性、氧化活性、嘧啶二聚体形成活性、整合酶活性、转座酶活性、重组酶活性、聚合酶活性、连接酶活性、解旋酶活性、光裂合酶活性和糖基化酶活性。

25.如权利要求23所述的Cas12L融合多肽,其中所述异源多肽表现出选自以下的一种或多种酶促活性:逆转录酶活性、核酸酶活性、甲基转移酶活性、去甲基化酶活性、脱氨基活性、脱嘌呤活性、整合酶活性、转座酶活性和重组酶活性。

26.如权利要求14-21中任一项所述的Cas12L融合多肽,其中所述异源多肽表现出修饰与靶核酸相关的靶多肽的酶促活性。

27.如权利要求26所述的Cas12L融合多肽,其中所述异源多肽表现出组蛋白修饰活性。

28.如权利要求26或权利要求27所述的Cas12L融合多肽,其中所述异源多肽表现出选自以下的一种或多种酶促活性:甲基转移酶活性、脱甲基酶活性、乙酰转移酶活性、脱乙酰酶活性、激酶活性、磷酸酶活性、泛素连接酶活性、脱泛素化活性、腺苷酸化活性、脱腺苷酸化活性、SUMO化活性、脱SUMO化活性、核糖基化活性、脱核糖基化活性、豆蔻酰化活性、脱豆蔻酰化活性、糖基化活性(例如来自O-GlcNAc转移酶)和脱糖基化活性。

29.如权利要求28所述的Cas12L融合多肽,其中所述异源多肽表现出选自以下的一种或多种酶促活性:甲基转移酶活性、脱甲基酶活性、乙酰转移酶活性和脱乙酰酶活性。

30.如权利要求14-21中任一项所述的Cas12L融合多肽,其中所述异源多肽是核内体逃逸多肽。

31.如权利要求30所述的Cas12L融合多肽,其中所述核内体逃逸多肽包含选自GLFXALLXLLXSLWXLLLXA(SEQ ID NO:72)和GLFHALLHLLHSLWHLLLHA(SEQ ID NO:73)的氨基酸序列,其中每个X独立地选自赖氨酸、组氨酸和精氨酸。

32.如权利要求14-21中任一项所述的Cas12L融合多肽,其中所述异源多肽是叶绿体转运肽。

33.如权利要求14-21中任一项所述的Cas12L融合多肽,其中所述异源多肽包含蛋白质转导结构域。

34.如权利要求14-21中任一项所述的Cas12L融合多肽,其中所述异源多肽是增加或减少转录的蛋白质。

35.如权利要求34所述的Cas12L融合多肽,其中所述异源多肽是转录阻遏子结构域。

36.如权利要求34所述的Cas12L融合多肽,其中所述异源多肽是转录激活结构域。

37.如权利要求14-21中任一项所述的Cas12L融合多肽,其中所述异源多肽是蛋白质结

合结构域。

38.一种核酸,其包含编码如权利要求14-37中任一项所述的Cas12L融合多肽的核苷酸序列。

39.如权利要求38所述的核酸,其中编码所述Cas12L融合多肽的所述核苷酸序列可操作地连接至启动子。

40.如权利要求39所述的核酸,其中所述启动子在真核细胞中有功能。

41.如权利要求40所述的核酸,其中所述启动子在以下中的一种或多种中有功能:植物细胞、真菌细胞、动物细胞、无脊椎动物细胞、蝇细胞、脊椎动物细胞、哺乳动物细胞、灵长类动物细胞、非人灵长类动物细胞和人细胞。

42.如权利要求39-41中任一项所述的核酸,其中所述启动子是以下中的一种或多种:组成型启动子、诱导型启动子、细胞类型特异性启动子和组织特异性启动子。

43.如权利要求38-42中任一项所述的核酸,其中所述核酸是重组表达载体。

44.如权利要求43所述的核酸,其中所述重组表达载体是重组腺相关病毒载体、重组逆转录病毒载体或重组慢病毒载体。

45.如权利要求39所述的核酸,其中所述启动子在原核细胞中有功能。

46.如权利要求38所述的核酸,其中所述核酸分子是mRNA。

47.一种或多种核酸,其包含:

(a)编码Cas12L引导RNA的核苷酸序列;和

(b)编码Cas12L多肽的核苷酸序列。

48.如权利要求47所述的一种或多种核酸,其中所述Cas12L多肽包含与图2A-2Z、图2AA-2ZZ和图2AAA-2FFF中任一者中描绘的氨基酸序列具有50%或更多同一性的氨基酸序列。

49.如权利要求47所述的一种或多种核酸,其中所述Cas12L多肽包含与图2A-2Z、图2AA-2ZZ和图2AAA-2FFF中任一者中描绘的氨基酸具有85%或更多同一性的氨基酸序列。

50.如权利要求47-49中任一项所述的一种或多种核酸,其中所述Cas12L引导RNA包含与图7、图2A-2Z、图2AA-2ZZ和图2AAA-2FFF中列出的crRNA序列中的任一者具有80%或更多核苷酸序列同一性的核苷酸序列。

51.如权利要求47-50中任一项所述的一种或多种核酸,其中所述Cas12L多肽融合至一种或多种核定位信号(NLS)。

52.如权利要求47-51中任一项所述的一种或多种核酸,其中编码所述Cas12L引导RNA的所述核苷酸序列可操作地连接至启动子。

53.如权利要求47-52中任一项所述的一种或多种核酸,其中编码所述Cas12L多肽的所述核苷酸序列可操作地连接至启动子。

54.如权利要求52或权利要求53所述的一种或多种核酸,其中可操作地连接至编码所述Cas12L引导RNA的所述核苷酸序列的所述启动子和/或可操作地连接至编码所述Cas12L多肽的所述核苷酸序列的所述启动子,在真核细胞中有功能。

55.如权利要求54所述的一种或多种核酸,其中所述启动子在以下中的一种或多种细胞中有功能:植物细胞、真菌细胞、动物细胞、无脊椎动物细胞、蝇细胞、脊椎动物细胞、哺乳动物细胞、灵长类动物细胞、非人灵长类动物细胞和人细胞。

56.如权利要求53-55中任一项所述的一种或多种核酸,其中所述启动子是以下中的一种或多种:组成型启动子、诱导型启动子、细胞类型特异性启动子和组织特异性启动子。

57.如权利要求47-56中任一项所述的一种或多种核酸,其中所述一种或多种核酸是一种或多种重组表达载体。

58.如权利要求57所述的一种或多种核酸,其中所述一种或多种重组表达载体选自:一种或多种腺相关病毒载体、一种或多种重组逆转录病毒载体或一种或多种重组慢病毒载体。

59.如权利要求53所述的一种或多种核酸,其中所述启动子在原核细胞中有功能。

60.一种真核细胞,其包含以下中的一种或多种:

a)Cas12L多肽或包含编码所述Cas12L多肽的核苷酸序列的核酸,

b)Cas12L融合多肽或包含编码所述Cas12L融合多肽的核苷酸序列的核酸,和

c)Cas12L引导RNA或包含编码所述Cas12L引导RNA的核苷酸序列的核酸。

61.如权利要求60所述的真核细胞,其包含编码所述Cas12L多肽的所述核酸,其中所述核酸整合到所述细胞的基因组DNA中。

62.如权利要求60或权利要求61所述的真核细胞,其中所述真核细胞是植物细胞、哺乳动物细胞、昆虫细胞、蜘蛛细胞、真菌细胞、鸟类细胞、爬行动物细胞、两栖动物细胞、无脊椎动物细胞、小鼠细胞、大鼠细胞、灵长类动物细胞、非人灵长类动物细胞或人细胞。

63.一种细胞,其包含包含Cas12L融合多肽或包含编码所述Cas12L融合多肽的核苷酸序列的核酸。

64.如权利要求63所述的细胞,其中所述细胞是原核细胞。

65.如权利要求63或权利要求64所述的细胞,其包含含编码所述Cas12L融合多肽的核苷酸序列的核酸,其中所述核酸分子整合到所述细胞的基因组DNA中。

66.一种修饰靶核酸的方法,所述方法包括使靶核酸与以下项接触:

a)Cas12L多肽;和

b)包含与所述靶核酸的靶序列杂交的引导序列的Cas12L引导RNA,

其中所述接触导致所述Cas12L多肽对所述靶核酸的修饰。

67.如权利要求66所述的方法,其中所述修饰是所述靶核酸的切割。

68.如权利要求66或权利要求67所述的方法,其中所述靶核酸选自:双链DNA、单链DNA、RNA、基因组DNA和染色体外DNA。

69.如权利要求66-68中任一项所述的方法,其中所述接触在体外在细胞外部发生。

70.如权利要求66-68中任一项所述的方法,其中所述接触在培养的细胞内部发生。

71.如权利要求66-68中任一项所述的方法,其中所述接触在体内细胞内部发生。

72.如权利要求70或权利要求71所述的方法,其中所述细胞是真核细胞。

73.如权利要求72所述的方法,其中所述细胞选自:植物细胞、真菌细胞、哺乳动物细胞、爬行动物细胞、昆虫细胞、禽细胞、鱼细胞、寄生虫细胞、节肢动物细胞、蛛形纲动物细胞、无脊椎动物细胞、脊椎动物细胞、啮齿动物细胞、小鼠细胞、大鼠细胞、灵长类动物细胞、非人灵长类动物细胞和人细胞。

74.如权利要求70或权利要求71所述的方法,其中所述细胞是原核细胞。

75.如权利要求66-74中任一项所述的方法,其中所述接触导致基因组编辑。

76.如权利要求66-75中任一项所述的方法,其中所述接触包括:向细胞中引入:(a)所述Cas12L多肽或包含编码所述Cas12L多肽的核苷酸序列的核酸,和(b)所述Cas12L引导RNA或包含编码所述Cas12L引导RNA的核苷酸序列的核酸。

77.如权利要求76所述的方法,其中所述接触还包括:将DNA供体模板引入到所述细胞中。

78.如权利要求66-77中任一项所述的方法,其中所述Cas12L引导RNA包含与图7、图2A-2Z、图2AA-2ZZ和图2AAA-2FFF中列出的crRNA序列中的任一者具有80%或更多核苷酸序列同一性的核苷酸序列。

79.如权利要求66-78中任一项所述的方法,其中所述Cas12L多肽融合至核定位信号。

80.一种调节靶DNA的转录、修饰靶核酸或修饰与靶核酸相关的蛋白质的方法,所述方法包括使所述靶核酸与以下项接触:

a)包含与异源多肽融合的Cas12L多肽的Cas12L融合多肽;和

b)包含与所述靶核酸的靶序列杂交的引导序列的Cas12L引导RNA。

81.如权利要求80所述的方法,其中所述Cas12L引导RNA包含与图7、图2A-2Z、图2AA-2ZZ和图2AAA-2FFF中列出的crRNA序列中的任一者具有80%或更多核苷酸序列同一性的核苷酸序列。

82.如权利要求80或权利要求81所述的方法,其中所述Cas12L融合多肽包含核定位信号。

83.如权利要求80-82中任一项所述的方法,其中所述修饰不是所述靶核酸的切割。

84.如权利要求80-83中任一项所述的方法,其中所述靶核酸选自:双链DNA、单链DNA、RNA、基因组DNA和染色体外DNA。

85.如权利要求80-84中任一项所述的方法,其中所述接触在体外在细胞外部发生。

86.如权利要求80-84中任一项所述的方法,其中所述接触在培养的细胞内部发生。

87.如权利要求80-84中任一项所述的方法,其中所述接触在体内细胞内部发生。

88.如权利要求86或权利要求87所述的方法,其中所述细胞是真核细胞。

89.如权利要求88所述的方法,其中所述细胞选自:植物细胞、真菌细胞、哺乳动物细胞、爬行动物细胞、昆虫细胞、禽细胞、鱼细胞、寄生虫细胞、节肢动物细胞、蛛形纲动物细胞、无脊椎动物细胞、脊椎动物细胞、啮齿动物细胞、小鼠细胞、大鼠细胞、灵长类动物细胞、非人灵长类动物细胞和人细胞。

90.如权利要求86或权利要求87所述的方法,其中所述细胞是原核细胞。

91.如权利要求80-90中任一项所述的方法,其中所述接触包括:向细胞中引入:(a)所述Cas12L融合多肽或包含编码所述Cas12L融合多肽的核苷酸序列的核酸,和(b)所述Cas12L引导RNA或包含编码所述Cas12L引导RNA的核苷酸序列的核酸。

92.如权利要求80-91中任一项所述的方法,其中所述Cas12L多肽是无催化活性的Cas12L多肽(dCas12L)。

93.如权利要求80-92中任一项所述的方法,其中所述Cas12L多肽具有700个氨基酸至800个氨基酸的长度。

94.如权利要求80-93中任一项所述的方法,其中所述异源多肽表现出酶促活性。

95.如权利要求94所述的方法,其中所述异源多肽表现出选自以下的一种或多种酶促

活性:核酸酶活性、甲基转移酶活性、脱甲基化酶活性、DNA修复活性、DNA损伤活性、脱氨基活性、歧化酶活性、烷基化活性、脱嘌呤活性、氧化活性、嘧啶二聚体形成活性、整合酶活性、转座酶活性、重组酶活性、聚合酶活性、连接酶活性、解旋酶活性、光裂合酶活性和糖基化酶活性。

96.如权利要求94所述的方法,其中所述异源多肽表现出选自以下的一种或多种酶促活性:逆转录酶活性、核酸酶活性、甲基转移酶活性、去甲基化酶活性、脱氨基活性、脱嘌呤活性、整合酶活性、转座酶活性和重组酶活性。

97.如权利要求80-93中任一项所述的方法,其中所述异源多肽表现出修饰与靶核酸相关的靶多肽的酶促活性。

98.如权利要求97所述的方法,其中所述异源多肽表现出组蛋白修饰活性。

99.如权利要求97或权利要求98所述的方法,其中所述异源多肽表现出选自以下的一种或多种酶促活性:甲基转移酶活性、脱甲基酶活性、乙酰转移酶活性、脱乙酰酶活性、激酶活性、磷酸酶活性、泛素连接酶活性、脱泛素化活性、腺苷酸化活性、脱腺苷酸化活性、SUMO化活性、脱SUMO化活性、核糖基化活性、脱核糖基化活性、豆蔻酰化活性、脱豆蔻酰化活性、糖基化活性(例如来自O-GlcNAc转移酶)和脱糖基化活性。

100.如权利要求99所述的方法,其中所述异源多肽表现出选自以下的一种或多种酶活性:甲基转移酶活性、脱甲基酶活性、乙酰转移酶活性和脱乙酰酶活性。

101.如权利要求80-93中任一项所述的方法,其中所述异源多肽是增加或减少转录的蛋白质。

102.如权利要求101所述的方法,其中所述异源多肽是转录阻遏子结构域。

103.如权利要求101所述的方法,其中所述异源多肽是转录激活结构域。

104.如权利要求80-93中任一项所述的方法,其中所述异源多肽是蛋白质结合结构域。

105.一种转基因多细胞非人生物体,其基因组包含有包含编码以下一种或多种的核苷酸序列的转基因:

a)Cas12L多肽,

b)Cas12L融合多肽,和

c)Cas12L引导RNA。

106.如权利要求105所述的转基因多细胞非人生物体,其中所述Cas12L多肽包含与图2A-2Z、图2AA-2ZZ和图2AAA-2FFF中任一者中列出的氨基酸序列具有50%或更多氨基酸序列同一性的氨基酸序列。

107.如权利要求105所述的转基因多细胞非人生物体,其中所述Cas12L多肽包含与图2A-2Z、图2AA-2ZZ和图2AAA-2FFF中任一者中列出的氨基酸序列具有85%或更多氨基酸序列同一性的氨基酸序列。

108.如权利要求105-107中任一项所述的转基因多细胞非人生物体,其中所述生物体是植物、单子叶植物、双子叶植物、无脊椎动物、昆虫、节肢动物、蛛形纲动物、寄生虫、蠕虫、刺胞动物、脊椎动物、鱼、爬行动物、两栖动物、有蹄动物、鸟、猪、马、羊、啮齿动物、小鼠、大鼠或非人类灵长类动物。

109.一种系统,其包括:

a)Cas12L多肽和Cas12L引导RNA;

b)Cas12L多肽、Cas12L引导RNA和DNA供体模板；

c)Cas12L融合多肽和Cas12L引导RNA；

d)Cas12L融合多肽、Cas12L引导RNA和DNA供体模板；

e)编码Cas12L多肽的mRNA和Cas12L引导RNA；

f)编码Cas12L多肽的mRNA;Cas12L引导RNA和DNA供体模板；

g)编码Cas12L融合多肽的mRNA和Cas12L引导RNA；

h)编码Cas12L融合多肽的mRNA、Cas12L引导RNA和DNA供体模板；

i)一种或多种重组表达载体,其包含:i)编码Cas12L多肽的核苷酸序列;和ii)编码Cas12L引导RNA的核苷酸序列；

j)一种或多种重组表达载体,其包含:i)编码Cas12L多肽的核苷酸序列;ii)编码Cas12L引导RNA的核苷酸序列;和iii)DNA供体模板；

k)一种或多种重组表达载体,其包含:i)编码Cas12L融合多肽的核苷酸序列;和ii)编码Cas12L引导RNA的核苷酸序列;和

l)一种或多种重组表达载体,其包含:i)编码Cas12L融合多肽的核苷酸序列;ii)编码Cas12L引导RNA的核苷酸序列;和DNA供体模板。

110.如权利要求109所述的Cas12L系统,其中所述Cas12L多肽包含与图2A-2Z、图2AA-2ZZ和图2AAA-2FFF中任一者中描绘的氨基酸序列具有50％或更多氨基酸序列同一性的氨基酸序列。

111.如权利要求109所述的Cas12L系统,其中所述Cas12L多肽包含与图2A-2Z、图2AA-2ZZ和图2AAA-2FFF中任一者中描绘的氨基酸序列具有85％或更多氨基酸序列同一性的氨基酸序列。

112.如权利要求109-111中任一项所述的Cas12L系统,其中所述供体模板核酸具有8个核苷酸至1000个核苷酸的长度。

113.如权利要求109-111中任一项所述的Cas12L系统,其中所述供体模板核酸具有25个核苷酸至500个核苷酸的长度。

114.一种试剂盒,其包括如权利要求109-113中任一项所述的Cas12L系统。

115.如权利要求114所述的试剂盒,其中所述试剂盒的组分在同一容器中。

116.如权利要求114所述的试剂盒,其中所述试剂盒的组分在单独的容器中。

117.一种无菌容器,其包括如权利要求109-116中任一项所述的Cas12L系统。

118.如权利要求117所述的无菌容器,其中所述容器为注射器。

119.一种可植入装置,其包括如权利要求109-116中任一项所述的Cas12L系统。

120.如权利要求119所述的可植入装置,其中所述Cas12L系统在基质内。

121.如权利要求119所述的可植入装置,其中所述Cas12L系统位于储器中。

122.一种检测样品中的靶DNA的方法,所述方法包括：

(a)使所述样品与以下物质接触：

(i)Cas12L多肽；

(ii)引导RNA,所述引导RNA包含:与所述Cas12L多肽结合的区域和与所述靶DNA杂交的引导序列;和

(iii)为单链的且不与所述引导RNA的引导序列杂交的检测剂核酸;以及

(b)测量由所述Cas12L多肽切割所述单链检测剂核酸产生的可检测信号,从而检测所述靶DNA。

123.如权利要求122所述的方法,其中所述靶DNA是细菌DNA。

124.如权利要求122所述的方法,其中所述靶DNA为病毒DNA。

125.如权利要求124所述的方法,其中所述靶DNA是乳多空病毒、人乳头瘤病毒(HPV)、嗜肝DNA病毒、乙型肝炎病毒(HBV)、疱疹病毒、水痘带状疱疹病毒(VZV)、爱泼斯坦-巴尔病毒(EBV)、卡波济氏肉瘤相关疱疹病毒、腺病毒、痘病毒或细小病毒DNA。

126.如权利要求122所述的方法,其中所述靶DNA来自人细胞。

127.如权利要求122所述的方法,其中所述靶DNA是人胎儿或癌细胞DNA。

128.如权利要求122所述的方法,其中所述样品包括细胞裂解物。

129.如权利要求122所述的方法,其中所述样品包括细胞。

130.如权利要求122所述的方法,其中所述样品是血液、血清、血浆、尿液、吸出物或活检样品。

131.如权利要求122-130中任一项所述的方法,所述方法还包括确定所述样品中存在的所述靶DNA的量。

132.如权利要求131所述的方法,其中所述测量可检测信号包括以下一种或多种:基于视觉的检测、基于传感器的检测、颜色检测、基于金纳米颗粒的检测、荧光偏振、胶体相变/分散、电化学检测和基于半导体的感测。

133.如权利要求122-132中任一项所述的方法,其中所述经标记的检测剂核酸是DNA。

134.如权利要求122-132中任一项所述的方法,其中所述经标记的检测剂核酸是RNA。

135.如权利要求122-134中任一项所述的方法,其中所述经标记的检测剂核酸包含修饰的核碱基、修饰的糖部分和/或修饰的核酸键。

136.如权利要求122-135中任一项所述的方法,其中所述可检测信号在少于45分钟内是可检测的。

137.如权利要求122-135中任一项所述的方法,其中所述可检测信号在少于30分钟内是可检测的。

138.如权利要求122-137中任一项所述的方法,其还包括通过环介导等温扩增(LAMP)、解旋酶依赖性扩增(HDA)、重组酶聚合酶扩增(RPA)、链置换扩增(SDA)、基于核酸序列的扩增(NASBA)、转录介导扩增(TMA)、切口酶扩增反应(NEAR)、滚环扩增(RCA)、多置换扩增(MDA)、分枝(RAM)、环状解旋酶依赖性扩增(cHDA)、单引物等温扩增(SPIA)、信号介导RNA扩增技术(SMART)、自我持续序列复制(3SR)、基因组指数扩增反应(GEAR)或等温多置换扩增(IMDA)来扩增所述样品中的所述靶DNA。

139.如权利要求122-138中任一项所述的方法,其中所述样品中的靶DNA以小于10aM的浓度存在。

140.根据权利要求122-139中任一项所述的方法,其中所述经标记的检测剂核酸的标记包括荧光发射染料对。

141.根据权利要求140所述的方法,其中所述荧光发射染料对是荧光共振能量转移(FRET)对。

142.根据权利要求140所述的方法,其中所述荧光发射染料对是猝灭剂/荧光剂对。

143.根据权利要求142-147中任一项所述的方法,其中所述经标记的检测剂核酸包含两个或更多个荧光发射染料对。

144.根据权利要求143所述的方法,其中所述两个或更多个荧光发射染料对包括荧光共振能量转移(FRET)对和猝灭剂/荧光剂对。

CRISPR-CAS效应多肽及其使用方法

[0001] 交叉引用

[0002] 本申请要求2019年12月23日提交的美国临时专利申请号62/952,909的权益,所述申请以引用方式整体并入本文。

[0003] 引言

[0004] CRISPR-Cas系统包括Cas蛋白和一种或多种引导RNA,所述Cas蛋白参与外源DNA或RNA的获取、靶向和切割,所述引导RNA包括结合Cas蛋白的片段和结合至靶核酸的片段。例如,2类CRISPR-Cas系统包含与引导RNA结合的单一Cas蛋白,其中Cas蛋白结合至靶向的核酸并切割靶向的核酸。这些系统的可编程性质促进了它们作为用于修饰靶核酸的通用技术的用途。

发明内容

[0005] 本公开提供了RNA引导的CRISPR-Cas效应蛋白、编码所述蛋白的核酸以及包含所述蛋白的组合物。本公开提供了核糖核蛋白复合物,其包含:本公开的RNA引导的CRISPR-Cas效应蛋白;和引导RNA。本公开提供了使用本公开的RNA引导的CRISPR-Cas效应蛋白和引导RNA修饰靶核酸的方法。

附图说明

[0006] 图1呈现了Cas12超家族的最大似然种系发生树,其由具有1000次迭代的多序列比对产生。Cas12L形成与其他Cas12蛋白分离的独特分化枝。

[0007] 图2A-2Z、2AA-2ZZ和2AAA-2FFF提供了Cas12L多肽的氨基酸序列。

[0008] 图3A-3B提供了Cas12L多肽的RuvC-I、RuvC-II和RuvC-III结构域的活性位点残基的位置。

[0009] 图4描绘了CRISPR基因座结构,其中活性位点被注释。

[0010] 图5A-5P提供了图2A-2Z、2AA-2ZZ和2AAA-2FFF(从顶部到底部SEQ ID NO:102-112、112、113-126、102、127-137、157、138-156)中描绘的Cas12L多肽的氨基酸序列的比对。

[0011] 图6A-6C提供了图2A-2Z、2AA-2ZZ和2AAA-2FFF(从顶部到底部SEQ ID NO:158-161、142-144)中描绘的所选Cas12L多肽的氨基酸序列的比对。

[0012] 图7提供了Cas-λ(Cas12L)引导RNA(从顶部到底部SEQ ID NO:63、70、64、163、59、65、60、68、66、61、164、67、69)的重复区的核苷酸序列比对。提供了共有序列(SEQ ID NO:162)。

[0013] 图8A-8M描绘了各种Cas-λ引导RNA重复序列的二级结构。

[0014] 图9描绘了PAM分析的结果。

[0015] 图10提供了质粒pBAS12-L1-NT阵列的示意图。

[0016] 图11提供了质粒pBAS18的示意图。

[0017] 图12A-12C描绘的数据显示,与具有非靶向引导物(pBAS12)的Casλ阴性对照相比,具有GFP靶向引导物的Casλ(pBAS44)显示出多个数量级的集落形成单位(作为细胞活力的

代表)减少。使用来自另一个直向同源物的重复单位的具有GFP靶向导向的Casλ也显示出生长抑制和GFP消融。

[0018] 定义

[0019] 在本文中可互换使用的术语“多核苷酸”和“核酸”是指任何长度的核苷酸(核糖核苷酸或脱氧核糖核苷酸)的聚合形式。因此,该术语包括但不限于单链、双链或多链DNA或RNA、基因组DNA、cDNA、DNA-RNA杂交体、或包含嘌呤和嘧啶碱基或其他天然的、经化学或生物化学修饰的、非天然或衍生的核苷酸碱基的聚合物。

[0020] “可杂交的”或“互补的”或“大致上互补的”意指核酸(例如RNA、DNA)包含能使其在适当的温度和溶液离子强度的体外和/或体内条件下以序列特异性、反平行的方式(即,核酸特异性结合互补核酸)非共价结合(即形成沃森-克里克碱基对和/或G/U碱基对)、“退火”或“杂交”另一个核酸的核苷酸序列。标准沃森-克里克碱基配对包括:腺嘌呤(A)与胸腺嘧啶(T)配对、腺嘌呤(A)与尿嘧啶(U)配对以及鸟嘌呤(G)与胞嘧啶(C)配对[DNA、RNA]。此外,对于两个RNA分子(例如,dsRNA)之间的杂交,以及对于DNA分子与RNA分子的杂交(例如,当DNA靶核酸碱基与引导RNA配对时,等):鸟嘌呤(G)也可以与尿嘧啶(U)碱基配对。例如,G/U碱基配对为tRNA抗密码子碱基配对mRNA中的密码子的背景下的遗传密码简并(即,冗余)的至少部分原因。因此,在本公开的上下文中,鸟嘌呤(G)(例如,引导RNA分子的dsRNA双链体;与靶核酸配对的引导RNA碱基等)被认为与尿嘧啶(U)和腺嘌呤(A)互补。例如,当G/U碱基对可以在引导RNA分子的dsRNA双链体的给定核苷酸位置产生时,所述位置不被认为是非互补的,而是被认为是互补的。

[0021] 杂交和洗涤条件是熟知的,并且在下文中举例说明:Sambrook,J.,Fritsch,E.F.和Maniatis,T.Molecular Cloning:A Laboratory Manual,第二版,Cold Spring Harbor Laboratory Press,Cold Spring Harbor(1989),具体为其中的第11章和表11.1;以及Sambrook,J.和Russell,W.,Molecular Cloning:A Laboratory Manual,第三版,Cold Spring Harbor Laboratory Press,Cold Spring Harbor(2001)。温度和离子强度的条件决定杂交的“严格性”。

[0022] 杂交要求两个核酸含有互补序列,虽然碱基之间有可能存在错配。适用于两个核酸之间的杂交的条件取决于核酸的长度和互补的程度、本领域中熟知的变量。两个核苷酸序列之间的互补程度越大,对于具有那些序列的核酸的杂交体的解链温度(Tm)的值越大。对于具有短序列段互补性(例如,在35或更少、30或更少、25或更少、22或更少、20或更少或18或更少个核苷酸上的互补性)的核酸之间的杂交,错配的位置可变得重要(参见Sambrook等,同上,11.7-11.8)。通常,可杂交核酸的长度为8个核苷酸或更多(例如,10个核苷酸或更多、12个核苷酸或更多、15个核苷酸或更多、20个核苷酸或更多、22个核苷酸或更多、25个核苷酸或更多、或30个核苷酸或更多)。根据诸如互补区域的长度和互补程度的因素,可以根据需要调节温度、洗涤溶液盐浓度和其它条件。

[0023] 应当理解,多核苷酸的序列不需要与其靶核酸的序列100%互补才能特异性杂交或可杂交。此外,多核苷酸可在一个或多个区段上杂交以使得中间区段或相邻区段不涉及杂交事件(例如,凸起、环结构或发夹结构等)。多核苷酸可以与同它杂交的靶核酸序列内的靶区域具有60%或更多、65%或更多、70%或更多、75%或更多、80%或更多、85%或更多、90%或更多、95%或更多、98%或更多、99%或更多、99.5%或更多或100%序列互补性。例

如,其中反义化合物的20个核苷酸中的18个与靶区域互补并且因此将特异性杂交的反义核酸将表示90%互补性。在此实例中,剩下的非互补核苷酸可与互补核苷酸集群或穿插在互补核苷酸中并且不需要彼此邻接或与互补核苷酸邻接。可以使用任何方便的方法来确定核酸内特定核酸序列片段之间的互补百分比。示例方法包括BLAST程序(基本局部比对搜索工具)和PowerBLAST程序(Altschul等人,J.Mol.Biol.,1990,215,403-410;Zhang和Madden,Genome Res.,1997,7,649-656)、Gap程序(Wisconsin序列分析包,Unix第8版,Genetics Computer Group,University Research Park,Madison Wis.)、例如使用默认设置(其使用Smith和Waterman)(Adv Appl.Math.,1981,2,482-489)等等。

[0024] 术语"肽"、"多肽"和"蛋白质"在本文中可互换使用,并且是指任何长度的氨基酸(其可包括编码和非编码的氨基酸、经化学或生物化学修饰或衍生的氨基酸)的聚合形式,以及具有经修饰的肽骨架的多肽。

[0025] 如本文所用,"结合"(例如,关于多肽的RNA结合结构域,与靶核酸的结合等)是指大分子之间(例如,蛋白质和核酸之间;在Cas12L多肽/引导RNA复合物与靶核酸之间;等等)的非共价相互作用。当在非共价相互作用的状态下,大分子被称作"缔合"或"相互作用"或"结合"(例如,当分子X被称作与分子Y相互作用时,意指分子X以非共价方式结合分子Y)。不是所有的结合相互作用组分都需要为序列特异性的(例如,与DNA骨架中的磷酸酯残基接触),但结合相互作用的一些部分可为序列特异性的。结合相互作用的特征通常在于解离常数(K_D)小于10^{-6}M、小于10^{-7}M、小于10^{-8}M、小于10^{-9}M、小于10^{-10}M、小于10^{-11}M、小于10^{-12}M、小于10^{-13}M、小于10^{-14}M或小于10^{-15}M。"亲和力"是指结合的强度,增加的结合亲和力与较低的K_D相关。

[0026] "结合结构域"意指能够非共价结合另一分子的蛋白质结构域。结合结构域可结合例如DNA分子(DNA结合结构域)、RNA分子(RNA结合结构域)和/或蛋白质分子(蛋白质结合结构域)。就具有蛋白质结合结构域的蛋白质来说,在一些情况下,它可结合其自身(以形成同源二聚体、同源三聚体等)和/或它可结合不同蛋白质的一个或多个区。

[0027] 术语"保守氨基酸取代"是指具有类似侧链的氨基酸残基的蛋白质中的可互换性。例如,具有脂肪族侧链的一组氨基酸由甘氨酸、丙氨酸、缬氨酸、亮氨酸和异亮氨酸组成;具有脂肪族-羟基侧链的一组氨基酸由丝氨酸和苏氨酸组成;具有含酰胺侧链的一组氨基酸由天冬酰胺和谷氨酰胺组成;具有芳香族侧链的一组氨基酸由苯丙氨酸、酪氨酸和色氨酸组成;具有碱性侧链的一组氨基酸由赖氨酸、精氨酸和组氨酸组成;具有酸性侧链的一组氨基酸由谷氨酸酯和天冬氨酸酯组成;并且具有含硫侧链的一组氨基酸由半胱氨酸和甲硫氨酸组成。示例性保守氨基酸取代基团为:缬氨酸-亮氨酸-异亮氨酸、苯丙氨酸-酪氨酸、赖氨酸-精氨酸、丙氨酸-缬氨酸-甘氨酸以及天冬酰胺-谷氨酰胺。

[0028] 多核苷酸或多肽与另一多核苷酸或多肽具有一定的"序列同一性"百分比,这意味着当比对时碱基或氨基酸的百分数为相同的,并且当比较两个序列时在相同的相对位置上。可以许多不同方式确定序列同一性。为了确定序列同一性,可使用在包括ncbi.nlm.nili.gov/BL AST、ebi.ac.uk/Tools/msa/tcoffee/、ebi.ac.uk/Tools/msa/muscle/、maff t.cbrc.jp/alignment/software/的万维网网址上可获得的各种方便的方法和计算机程序(例如,BLAST、T-COFFEE、MUSCLE、MAFFT等)来比对序列。参见例如Altschul等人(1990),J.Mol.Bioi.215:403-10.

[0029] “编码”具体RNA的DNA序列为转录成RNA的DNA核苷酸序列。DNA多核苷酸可以编码翻译成蛋白质的RNA(mRNA)(因此DNA和mRNA都编码蛋白质),或者DNA多核苷酸可以编码不翻译成蛋白质的RNA(例如tRNA、rRNA、微小RNA(miRNA)、“非编码”RNA(ncRNA)、引导RNA等)。

[0030] “蛋白质编码序列”或编码具体蛋白质或多肽的序列为当置于适当调节序列的控制下时转录成mRNA(在DNA的情况下)并且在体外或体内翻译(在mRNA的情况下)成多肽的核苷酸序列。

[0031] 本文可互换使用的术语“DNA调节序列”、“控制元件”和“调节元件”是指提供和/或调节非编码序列(例如,引导RNA)或编码序列(例如,RNA引导的核酸内切酶、GeoCas9多肽、GeoCas9融合多肽等)的转录和/或调节编码多肽的翻译的转录和翻译控制序列,例如启动子、增强子、聚腺苷酸化信号、终止子、蛋白质降解信号等。

[0032] 如本文所使用,“启动子”或“启动子序列”为能够结合RNA聚合酶并且启动下游(3’方向)编码或非编码序列的转录的DNA调节区。出于本公开的目的,启动子序列在其3’末端上通过转录起始位点结合并且向上游(5’方向)序列段以包括启动高于背景的可检测水平转录所需要的最小数目的碱基或元件。在启动子序列内将发现转录起始位点以及蛋白质结合结构域为造成RNA聚合酶结合的原因。真核生物启动子将经常但不总是含有“TATA”盒和“CAT”盒。包括诱导型启动子的各种启动子可用来驱动本公开的各种载体表达。

[0033] 如本文所使用适用于核酸、多肽、细胞或生物的术语“天然存在的”或“未修饰的”或“野生型”是指存在于自然中的核酸、多肽、细胞或生物。例如,可从自然中的来源分离的存在于生物中的多肽或多核苷酸序列为天然存在的。

[0034] 如本文所使用适用于核酸或多肽的术语“融合”是指通过源自不同来源的结构定义的两种组分。例如,当在融合多肽(例如,融合Cas12L蛋白)的上下文中使用“融合”时,融合多肽包括源自不同多肽的氨基酸序列。融合多肽可包含修饰的或天然存在的多肽序列(例如,来自修饰或未修饰的Cas12L蛋白的第一氨基酸序列;和来自除Cas12L蛋白之外的修饰或未修饰蛋白的第二氨基酸序列等)。类似地,在编码融合多肽的多核苷酸的背景下的“融合”包括源自不同编码区的核苷酸序列(例如,编码修饰的或未修饰的Cas12L蛋白的第一核苷酸序列;和编码除了Cas12L蛋白以外的多肽的第二核苷酸序列)。

[0035] 术语“融合多肽”是指通常通过人干预,通过组合(即,“融合”)氨基酸序列的两个另外分开的区段而制得的多肽。

[0036] 如本文所用,“异源的”意指分别不存在于天然核酸或蛋白质中的核苷酸或多肽序列。例如,在一些情况下,在本公开的变体Cas12L蛋白中,天然存在的Cas12L多肽(或其变体)的一部分可以融合到异源多肽(即,来自除Cas12L多肽之外的蛋白质的氨基酸序列;或来自另一种生物体的氨基酸序列)。作为另一个实例,融合Cas12L多肽可包含与异源多肽融合的天然存在的Cas12L多肽(或其变体)的全部或部分,所述异源多肽即来自除Cas12L多肽之外的蛋白质的多肽或来自另一种生物体的多肽。异源多肽可表现出变体Cas12L蛋白或融合Cas12L蛋白也会表现出的活性(例如,酶促活性)(例如,生物素连接酶活性;核定位;等)。异源核酸序列可连接至天然存在的核酸序列(或其变体)(例如,通过基因工程化)以产生编码融合多肽(融合蛋白)的核苷酸序列。

[0037] 如本文所使用的“重组”意指具体核酸(DNA或RNA)为克隆、限制、聚合酶链反应(PCR)和/或连接步骤的各种组合的产物,所述步骤产生具有可与天然系统中发现的内源核

酸区别开的结构编码或非编码序列的构建体。编码多肽的DNA序列可由cDNA片段或由一系列合成寡核苷酸组装以提供能够由包含在细胞中或无细胞转录和翻译系统中的重组转录单元表达的合成核酸。包含相关序列的基因组DNA还可用于形成重组基因或转录单元。非翻译DNA的序列可存在于开放读码框的5'端或3'端，其中此类序列不干扰编码区的操纵或表达，并且实际上可通过各种机制起到调节希望的产物产生的作用(参见"DNA调节序列")。或者，未翻译的编码RNA的DNA序列(例如，引导RNA)也可被认为是重组的。因此，例如术语"重组"核酸是指非天然存在的，例如通过人干预通过人工组合序列的两个另外分开的区段而制得的核酸。这种人工组合常常通过化学合成手段或通过人工操纵核酸的分离区段(例如，通过遗传工程化技术)来完成。这通常是用编码相同氨基酸、保守氨基酸或非保守氨基酸的密码子来替代一个密码子。可替代地，执行这种操作以将具有所需功能的核酸区段连接在一起以产生所需的功能组合。这种人工组合常常通过化学合成手段或通过人工操纵核酸的分离区段(例如，通过遗传工程化技术)来完成。当重组多核苷酸编码多肽时，编码多肽的序列可为天然存在的("野生型")或可为天然存在的序列的变体(例如，突变体)。这种情况的实例是编码野生型蛋白质的DNA(重组体)，其中该DNA序列被密码子优化用于在不天然存在所述蛋白质的细胞(例如，真核细胞)中表达所述蛋白质(例如，在真核细胞中表达CRISPR/Cas RNA引导的多肽，诸如Cas12L(例如，野生型Cas 12L;变体Cas12L;融合Cas12L等等))。因此，密码子优化的DNA可以是重组的和非天然存在的，而由所述DNA编码的蛋白质可以具有野生型氨基酸序列。

[0038] 因此，术语"重组"多肽未必是指其氨基酸序列不是天然存在的多肽。相反，"重组"多肽通过重组非天然存在的DNA序列来编码，但多肽的氨基酸序列可为天然存在的("野生型")或非天然存在的(例如，变体、突变体等)。因此，"重组"多肽为人干预的结果，但可具有天然存在的氨基酸序列。

[0039] "载体"或"表达载体"为复制子，如质粒、噬菌体、病毒、人工染色体或粘粒，另一个DNA区段(即"插入物")可附着至所述复制子以便在细胞中引起所附着的区段的复制。

[0040] "表达盒"包含可操作地连接至启动子的DNA编码序列。"可操作地连接"是指并置，其中所述组分处于容许其以预期的方式起作用的关系中。例如，如果启动子影响其转录或表达，则启动子可操作地连接到编码序列(或者编码序列也可以被认为可操作地连接到启动子)。

[0041] 术语"重组表达载体"或"DNA构建体"在本文中可互换使用，是指包含载体和插入物的DNA分子。通常出于表达和/或繁殖插入物的目的或出于构建其它重组核苷酸序列而产生重组表达载体。插入物可以或可以不可操作地连接至启动子序列并且可以或可以不可操作地连接至DNA调节序列。

[0042] 当此类DNA引入到细胞内部时，细胞被外源DNA或外源RNA例如重组表达载体"遗传修饰"或"转化"或"转染"。外源DNA的存在导致永久或瞬时的遗传变化。转化DNA可以或可以不整合(共价连接)到细胞基因组中。在例如原核生物、酵母和哺乳动物细胞中，转化DNA可维持在游离元件如质粒上。相对于真核细胞，稳定转化的细胞为其中转化DNA逐渐整合到染色体中以使得通过染色体复制遗传给子细胞的一种细胞。此稳定性通过真核细胞建立包含含有转化DNA的子细胞群的细胞系或克隆的能力来展示。"克隆"为通过有丝分裂源自单个细胞或共同祖先的一群细胞。"细胞系"为能够在体外稳定生长许多代的原代细胞的克隆。

[0043] 遗传修饰(也称为“转化”)的合适的方法包括例如病毒或噬菌体感染、转染、缀合、原生质体融合、脂质体转染、电穿孔、磷酸钙沉淀、聚乙烯亚胺(PEI)介导的转染、DEAE-葡聚糖介导的转染、脂质体介导的转染、粒子枪技术、磷酸钙沉淀、直接微注射、纳米颗粒介导的核酸递送(参见例如,Panyam等人Adv Drug Deliv Rev.2012年9月13日.pii:S0169-409X(12)00283-9.doi:10.1016/j.addr.2012.09.023)等。

[0044] 遗传修饰的方法的选择通常取决于待转化的细胞类型和在其下发生转化的环境(例如,体外、离体或体内)。这些方法的一般讨论可见于Ausubel等,Short Protocols in Molecular Biology,第3版,Wiley&Sons,1995中。

[0045] 如本文所用,“靶核酸”是多核苷酸(例如DNA,诸如基因组DNA),其包括被RNA引导的核酸内切酶多肽(例如,野生型Cas12L;变体Cas12L;融合Cas12L等)靶向的位点(“靶位点”或“靶序列”)。靶序列是主题Cas12L引导RNA(例如双Cas12L引导RNA或单分子Cas12L引导RNA)的引导序列将与之杂交的序列。例如,靶核酸内的靶位点(或靶序列)5’-GAGCAUAUC-3’被序列5’-GAUAUGCUC-3’靶向(或与之结合,或与之杂交,或与之互补)。合适的杂交条件包括细胞中正常存在的生理条件。对于双链靶核酸,与引导RNA互补并杂交的靶核酸链被称为“互补链”或“靶链”;而与“靶链”互补的靶核酸链(并因此不与引导RNA互补)被称为“非靶链”或“非互补链”。

[0046] “切割”意指靶核酸分子(例如RNA、DNA)共价主链的断裂。可通过各种各样的方法来开始裂解,所述方法包括但不限于磷酸二酯键的酶水解或化学水解。单链裂解和双链裂解均是可能的,并且双链裂解可由于两个相异单链裂解事件而发生。

[0047] “核酸酶”和“核酸内切酶”在本文中可互换使用,意指对核酸切割具有催化活性的酶(例如,核糖核酸酶活性(核糖核酸切割)、脱氧核糖核酸酶活性(脱氧核糖核酸切割)等)。

[0048] 核酸酶的“裂解结构域”或“活性结构域”或“核酸酶结构域”意指具有用于核酸裂解的催化活性的核酸酶内的多肽序列或结构域。裂解结构域可包含在单个多肽链中或裂解活性可由两个(或更多个)多肽的缔合引起。单个核酸酶结构域可由给定多肽内的多于一个分离的氨基酸序列段组成。

[0049] 本文使用术语“干细胞”是指具有自我更新和产生分化的细胞类型的能力的细胞(例如,植物干细胞、脊椎动物干细胞)(参见Morrison等(1997)Cell 88:287-298)。在细胞个体发育的背景下,形容词“被分化的”或“分化的”为相对术语。“分化的细胞”为比与其比较的细胞在发育途径上更进一步向下发展的细胞。因此,多能干细胞(以下所述)可分化成谱系限制的祖细胞(例如,中胚层干细胞),所述祖细胞反过来可分化成进一步限制的细胞(例如,神经元祖细胞),所述祖细胞可分化成末期细胞(即,终止分化的细胞,例如神经元、心肌细胞等),所述祖细胞在某些组织类型中起特征性作用,并且可以或可以不保持进一步增殖的能力。干细胞可特征在于存在特异性标志物(例如,蛋白质、RNA等)和不存在特异性标志物。干细胞还可通过体外和体内的功能测定,具体地与干细胞产生多个分化子代的能力相关的测定来鉴定。

[0050] 感兴趣的干细胞包括多能干细胞(PSC)。本文使用术语“多能干细胞”或“PSC”意指能够产生生物的所有细胞类型的干细胞。因此,PSC可产生生物的所有生殖层(例如,脊椎动物的内胚层、中胚层和外胚层)的细胞。多能细胞能够形成畸胎瘤并且对活生物中的外胚层、中胚层或内胚层组织有影响。植物的多能干细胞能够产生植物的所有细胞类型(例如,

根、茎、叶等的细胞)。

[0051] 可以许多不同方式得到动物的PSC。例如,胚胎干细胞(ESC)来源于胚胎的内细胞团(Thomson等人,Science.1998年11月6日;282(5391):1145-7),而诱导多能干细胞(iPSC)来源于体细胞(Takahashi等人,Cell.2007年11月30日;131(5):861-72;Takahashi等人,Nat Protoc.2007;2(12):3081-9;Yu等人,Science.2007年12月21日;318(5858):1917-20.Epub 2007Nov 20)。因为术语PSC是指无论其来源的多能干细胞,所以术语PSC涵盖术语ESC和iPSC以及术语胚胎生殖干细胞(EGSC),所述胚胎生殖干细胞为PSC的另一个实例。PSC可以建立的细胞系形式存在,它们可直接从原代胚胎组织获得或它们可源自体细胞。PSC可为本文所述方法的靶细胞。

[0052] “胚胎干细胞(ESC)”意指从胚胎,通常从胚泡的内细胞群中分离的PSC。ESC系列出在NIH人类胚胎干细胞登记处,例如hESBGN-01、hESBGN-02、hESBGN-03、hESBGN-04(BresaGen,Inc.);HES-1、HES-2、HES-3、HES-4、HES-5、HES-6(ES Cell International);Miz-hES1(首尔国立大学米兹梅迪医院(MizMedi Hospital-Seoul National University));HSF-1、HSF-6(加州大学旧金山分校(University of California at San Francisco));以及H1、H7、H9、H13、H14(威斯康辛校友研究基金会(威赛尔研究所)(Wisconsin Alumni Research Foundation(WiCell Research Institute)))。感兴趣的干细胞还包括来自其它灵长类动物的胚胎干细胞,如恒河猴干细胞和狨猴干细胞。干细胞可获自任何哺乳动物种类,例如人、马、牛、猪、犬、猫、啮齿动物(例如,小鼠、大鼠、仓鼠)、灵长类动物等。(Thomson等(1998)Science282:1145;Thomson等(1995)Proc.Natl.Acad.Sci USA 92:7844;Thomson等人(1996)Biol.Reprod.55:254;Shamblott等人,Proc.Natl.Acad.Sci.USA 95:13726,1998)。在培养中,ESC通常呈具有高核-细胞质比、清晰边界和突出核仁的扁平集落形式生长。另外,ESC表达SSEA-3、SSEA-4、TRA-1-60、TRA-1-81和碱性磷酸酶,但不表达SSEA-1。产生和表征ESC的方法的实例可见于例如美国专利号7,029,913、美国专利号5,843,780和美国专利号6,200,806中,所述专利的公开内容以引用的方式并入本文。用于以未分化的形式增殖hESC的方法描述于WO 99/20741、WO 01/51616和WO 03/020920中。

[0053] “胚胎生殖干细胞(EGSC)”或“胚胎生殖细胞”或“EG细胞”意指源自生殖细胞和/或生殖细胞祖细胞(例如原生殖细胞,即将变为精子和卵子的原生殖细胞)的PSC。认为胚胎生殖细胞(EG细胞)具有类似于如上所述的胚胎干细胞的性质。产生和表征EG细胞的方法的实例可见于例如美国专利号7,153,684;Matsui,Y.等,(1992)Cell70:841;Shamblott,M.等(2001)Proc.Natl.Acad.Sci.USA 98:113;Shamblott,M.等(1998)Proc.Natl.Acad.Sci.USA,95:13726;以及Koshimizu,U.等(1996)Development,122:1235中,所述参考文献的公开内容以引用的方式并入本文。

[0054] “诱导多能干细胞”或“iPSC”意指源自不是PSC的细胞的PSC(即,来自相对于PSC分化的细胞)。iPSC可源自多种不同细胞类型,包括终止分化的细胞。iPSC具有呈有高核-细胞质比、清晰边界和突出核仁的扁平集落形式生长的ES细胞样形态。另外,iPSC表达本领域普通技术人员已知的一种或多种主要多能标志物,包括但不限于碱性磷酸酶、SSEA3、SSEA4、Sox2、Oct3/4、Nanog、TRA160、TRA181、TDGF 1、Dnmt3b、FoxD3、GDF3、Cyp26a1、TERT和zfp42。产生和表征iPSC的方法的实例可以在例如美国专利公开号US20090047263、

US20090068742、US20090191159、US20090227032、US20090246875和US20090304646中找到，所述专利的公开内容以引用的方式并入本文。通常，为了产生iPSC，提供具有本领域中已知的重新编程因子(例如Oct4、SOX2、KLF4、MYC、Nanog、Lin28等)以将体细胞重新编程变为多能干细胞的体细胞。

[0055] “体细胞”意指生物中的任何细胞，所述细胞在不存在实验操纵的情况下通常不产生生物中的所有类型的细胞。换言之，体细胞为具有足够分化的细胞，它们将不天然产生身体的所有三个生殖层(即外胚层、中胚层和内胚层)的细胞。例如，体细胞将包括神经元和神经祖细胞，后者的神经祖细胞可能够天然产生中枢神经系统的所有或一些细胞类型但不能产生中胚层或内胚层谱系的细胞。

[0056] “有丝分裂细胞”意指经历有丝分裂的细胞。有丝分裂为真核细胞将其细胞核中的染色体分到两个分开细胞核中的两个相同集合中的过程。通常紧接着是胞质分裂，所述胞质分裂将细胞核、细胞质、细胞器官和细胞膜分到含有粗略地相等共有这些细胞组分的两个细胞中。

[0057] “有丝分裂后细胞”意指退出有丝分裂的细胞，即其为“静止的”，即它不再经历分裂。这种静止状态可为瞬时的，即可逆的，或其可为永久的。

[0058] “减数分裂细胞”意指经历减数分裂的细胞。减数分裂为细胞出于产生配子或孢子的目的分配其核物质的过程。不像有丝分裂，在减数分裂中，染色体经历在染色体之间混乱遗传物质的重组步骤。另外，与由有丝分裂产生的两个(遗传相同的)二倍体细胞相比，减数分裂的结果为四个(遗传独特的)单倍体细胞。

[0059] 在一些情况下，组分(例如，核酸组分(例如，Cas12L引导RNA)；蛋白质组分(例如，野生型Cas12L多肽；变体Cas12L多肽；融合Cas12L多肽；等)；等)包括标记部分。如本文所用，术语“标记”、“可检测标记”或“标记部分”是指提供信号检测的任何部分，并且可以根据测定的特定性质而广泛变化。感兴趣的标记部分包括可直接检测的标记(直接标记；例如荧光标记)和间接可检测标记(间接标记；例如，结合对成员)。荧光标记可以是任何荧光标记(例如，荧光染料(例如，荧光素、德克萨斯红、罗丹明、ALEXAFLUOR®标记等)、荧光蛋白(例如，绿色荧光蛋白(GFP)、增强型GFP(EGFP)、黄色荧光蛋白(YFP)、红色荧光蛋白(RFP)、青色荧光蛋白(CFP)、樱桃色、番茄红、橘红色及其任何荧光衍生物)等)。用于所述方法的合适的可检测(直接或间接)标记部分包括可通过光谱、光化学、生物化学、免疫化学、电学、光学、化学或其他手段检测的任何部分。例如，合适的间接标记包括生物素(结合对成员)，其可以被链霉亲和素结合(其本身可以被直接或间接标记)。标记还可以包括：放射性标记(直接标记)(例如^{3}H、^{125I}、^{35}S、^{14}C或^{32}P)；酶(间接标记)(例如，过氧化物酶、碱性磷酸酶、半乳糖苷酶、荧光素酶、葡萄糖氧化酶等)；荧光蛋白(直接标记)(例如，绿色荧光蛋白、红色荧光蛋白、黄色荧光蛋白及其任何方便的衍生物)；金属标签(直接标签)；比色标签；结合对成员；等。“结合对的配偶体”或“结合对成员”是指第一部分和第二部分之一，其中第一部分和第二部分彼此具有特异性结合亲和力。合适的结合对包括但不限于：抗原/抗体(例如，地高辛配基/抗地高辛配基、二硝基苯(DNP)/抗DNP、丹磺酰-X-抗丹磺酰、荧光素/抗荧光素、路西法黄/抗路西法黄和罗丹明抗罗丹明)、生物素/抗生物素蛋白(或生物素/链霉抗生物素蛋白)和钙调素结合蛋白(CBP)/钙调素。任何结合对成员都适合用作间接可检测的标记部分。

[0060] 任何给定的组分或组分的组合可以是未标记的，或者可以用标记部分可检测地标

记。在一些情况下，当两种或多种组分被标记时，它们可以用彼此可区分的标记部分来标记。

[0061] 分子和细胞生物化学的一般方法可以在标准教科书中找到，诸如Molecular Cloning:A Laboratory Manual,第3版(Sambrook等人,HaRBor Laboratory Press 2001)；Short Protocols in Molecular Biology,第4版(Ausubel等人编,John Wiley&Sons 1999)；Protein Methods(Bollag等人,John Wiley&Sons 1996)；Nonviral Vectors for Gene Therapy(Wagner等人编,Academic Press 1999)；Viral Vectors(Kaplift和Loewy编,Academic Press 1995)；Immunology Methods Manual(I.Lefkovits编,Academic Press 1997)；以及Cell and Tissue Culture:Laboratory Procedures in Biotechnology(Doyle&Griffiths,John Wiley&Sons 1998)，其公开内容以引用的方式并入本文。

[0062] 如本文所用，术语“治疗(treatment)、治疗(treating)”等是指获得所需的药理学和/或生理学效果。就完全或部分预防疾病或其症状而言，所述效果可以是预防性的，并且/或者就部分或完全治愈疾病和/或可归因于所述疾病的副作用而言，所述效果可以是治疗性的。如本文所用，“治疗”涵盖对哺乳动物(例如，人)的疾病的任何治疗，并且包括：(a)在可能易患疾病但还未诊断患有所述疾病的受试者中预防疾病发生；(b)抑制疾病，即阻止其发展；以及(c)缓解疾病，即引起疾病消退。

[0063] 在本文中可互换使用的术语“个体”、“受试者”、“宿主”和“患者”是指个体生物体，例如哺乳动物，包括但不限于鼠类、猿、人类、非人灵长类动物、有蹄类动物、猫科动物、犬科动物、牛科动物、绵羊、哺乳类农场动物、哺乳类运动动物和哺乳动物宠物。

[0064] 在进一步描述本发明之前，应当理解，本发明不限于所述的特定实施方案，因而当然可以改变。还应当理解，本文所用的术语仅用于描述特定实施方案的目的，而不是旨在进行限制，因为本发明的范围将仅由所附权利要求书限制。

[0065] 在提供数值范围时，应该理解在所述范围的上下限之间的每个中间值(除非上下文另外清楚地指出，否则所述中间值达到下限单位的十分之一)和任何其他说明的或在所述说明范围中的中间值涵盖在本发明内。这些较小范围的上限和下限可独立地包括在较小范围内并且也涵盖在本发明内，以规定范围内任何明确排除的限值为条件。当规定范围包括一个或两个限值时，排除了那些所包括的限值中的任一个或两个的范围也包括在本发明中。

[0066] 除非另外定义，否则本文所用的所有技术和科学术语均具有与本发明所属领域的普通技术人员通常所理解的相同含义。尽管与本文所述的那些类似或等同的任何方法和材料也可用于本发明的实践或测试，但现在描述优选的方法和材料。本文提及的所有出版物以引用方式并入本文，以结合所引用的出版物公开并描述方法和/或材料。

[0067] 必须注意，如本文和所附权利要求书中所用，除非上下文另外明确指出，否则单数形式“一(a/an)”和“所述(the)”包括复数指代物。因此，例如，提及“Cas12L CRISPR-Cas效应多肽”包括多个此类多肽，并且提及“引导RNA”包括提及本领域的技术人员已知的一种或多种引导RNA及其等效物，等。还应当注意，权利要求书可被起草以排除任何任选元素。因而，这种陈述意图充当结合权利要求要素的叙述来使用诸如“仅仅”、“仅”等排他性术语或使用“否定”限制的前提基础。

[0068] 应当理解,为了清楚起见在单独的实施方案的上下文中描述的本发明的某些特征也可在单个实施方案中组合提供。相反,为了简洁起见,在单个实施方案的上下文中描述的本发明的各种特征也可单独地或以任何合适的子组合提供。本发明特别地涵盖与本发明有关的实施方案的所有组合,并且在此公开,就如同单独地和明确地公开了每个组合一样。另外,本发明也特别地涵盖各种实施方案和其元素的所有子组合,并且在此公开,就如同每个此类子组合单独地和明确地在此公开一样。

[0069] 本文所讨论的出版物仅提供其在本申请的提交日期之前的公开内容。本文的任何内容均不应解释为承认本发明无权凭借在先发明而先于此类出版物。此外,所提供的出版日期可能与实际出版日期不同,这可能需要独立地确认。

具体实施方式

[0070] 本公开提供了RNA引导的CRISPR-Cas效应蛋白(本文称为“Cas12L”多肽或“CasLambda”多肽或“Casλ”多肽)、编码所述蛋白质的核酸和包含所述蛋白质的组合物。本公开提供了核糖核蛋白复合物,其包含:本公开的Cas12L多肽;和引导RNA。本公开提供了使用本公开的Cas12L多肽和引导RNA修饰靶核酸的方法。

[0071] 本发明提供了结合Cas12L蛋白并为其提供序列特异性的引导RNA(本文称为“Cas12L引导RNA”或“CasLambda引导RNA”或“Casλ引导RNA”);编码Cas12L引导RNA的核酸;和包含Cas12L引导RNA和/或编码所述引导RNA的核酸的修饰的宿主细胞。所提供的Cas12L引导RNA可用于所提供的多种应用中。

[0072] 组合物

[0073] CRISPR/C_{AS}12L蛋白质和引导RNA

[0074] Cas12L CRISPR/Cas效应多肽(例如,Cas12L蛋白)与相应的引导RNA(例如,Cas12L引导RNA)相互作用(结合)以形成核糖核蛋白(RNP)复合物,所述复合物通过引导RNA与靶核酸分子内的靶序列之间的碱基配对被靶向至靶核酸(例如,靶DNA)中的特定位点。引导RNA包括与靶核酸的序列(靶位点)互补的核苷酸序列(引导序列)。因此,Cas12L蛋白与Cas12L引导RNA形成复合物,并且引导RNA通过引导序列为RNP复合物提供序列特异性。复合物的Cas12L蛋白提供位点特异性活性。换言之,Cas12L蛋白借助于其与引导RNA缔合而被引导至靶核酸序列(例如染色体序列或染色体外序列,例如游离型序列、微环序列、线粒体序列、叶绿体序列等)内的靶位点(例如在靶位点处稳定)。

[0075] 本公开提供了包含Cas12L多肽(和/或包含编码Cas12L多肽的核苷酸序列的核酸)的组合物(例如,其中Cas12L多肽可以是天然存在的蛋白质、切口酶Cas12L蛋白质、无催化活性的(“死的”Cas12L;本文也称为“dCas12L蛋白”)、融合Cas12L蛋白等)。本公开提供了包含Cas12L引导RNA(和/或包含编码Cas12L引导RNA的核苷酸序列的核酸)的组合物。本公开提供了组合物,其包含(a)Cas12L多肽(和/或编码Cas12L多肽的核酸)(例如,其中Cas12L多肽可以是天然存在的蛋白质、切口酶Cas12L蛋白质、dCas12L蛋白质、融合Cas12L蛋白质等)和(b)Cas12L引导RNA(和/或编码Cas12L引导RNA的核酸)。本公开提供了核酸/蛋白质复合物(RNP复合物),其包含:(a)本公开的Cas12L多肽(例如,其中Cas12L多肽可以是天然存在的蛋白质、切口酶Cas12L蛋白质、死的Cas12L蛋白质、融合Cas12L蛋白质等);和(b)Cas12L引导RNA。

[0076] Cas12L蛋白

[0077] Cas12L多肽(此术语与术语“Cas12L蛋白”可互换使用)可结合和/或修饰(例如,切割、切口、甲基化、脱甲基化等)靶核酸和/或与靶核酸相关联的多肽(例如,组蛋白尾的甲基化或乙酰化)(例如,在一些情况下,Cas12L蛋白包括具有活性的融合配偶体,并且在一些情况下,Cas12L蛋白提供核酸酶活性)。在一些情况下,Cas12L蛋白是天然存在的蛋白(例如,天然存在于噬菌体中)。在其他情况下,Cas12L蛋白不是天然存在的多肽(例如,Cas12L蛋白是变体Cas12L蛋白、融合Cas12L蛋白等)。

[0078] 用于确定给定蛋白质是否与Cas12L引导RNA相互作用的测定可以是测试蛋白质和核酸之间结合的任何方便的结合测定。合适的结合测定(例如,凝胶位移测定)将是本领域普通技术人员已知的(例如,包括向靶核酸中添加Cas12L引导RNA和蛋白质的测定)。用于确定蛋白质是否具有活性(例如,用于确定蛋白质是否具有切割靶核酸的核酸酶活性和/或一些异源活性)的测定可以是任何方便的测定(例如,测试核酸切割的任何方便的核酸切割测定)。合适的测定(例如切割测定)将是本领域普通技术人员已知的。

[0079] 天然存在的Cas12L蛋白起着核酸内切酶的作用,催化靶向双链DNA(dsDNA)中特定序列的双链断裂。序列特异性由相关联的引导RNA提供,所述引导RNA与靶DNA内的靶序列杂交。天然存在的Cas12L引导RNA是crRNA,其中crRNA包括(i)与靶DNA中的靶序列杂交的引导序列,和(ii)包含与Cas12L蛋白结合的茎环(发夹dsRNA双链体)的蛋白结合片段。

[0080] 在一些实施方案中,主题方法和/或组合物的Cas12L蛋白是(或衍生自)天然存在的(野生型)蛋白。图2A-2Z、图2AA-2ZZ和图2AAA-2FFF中描绘了天然存在的Cas12L蛋白的实例。在一些情况下,(主题组合物和/或方法的)Cas12L蛋白包含与图2(例如,图2A-2Z、图2AA-2ZZ和图2AAA-2FFF中任一者)中描绘的任一种Cas12L氨基酸序列具有20%或更多序列同一性(例如,30%或更多、40%或更多、50%或更多、60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的氨基酸序列。在一些情况下,(主题组合物和/或方法的)图2A-2Z、图2AA-2ZZ和图2AAA-2FFF蛋白质包含图2(例如,图2A-2Z、图2AA-2ZZ和图2AAA-2FFF中任一者)中描绘的氨基酸序列。

[0081] 在一些情况下,(主题组合物和/或方法的)Cas12L蛋白与图2(例如,图2A-2Z、图2AA-2ZZ和图2AAA-2FFF中任一者)中所示的氨基酸序列具有的序列同一性大于与以下任一者的序列同一性:Cas12a蛋白、Cas12b蛋白、Cas12c蛋白、Cas12d蛋白、Cas12e蛋白、Cas12g蛋白、Cas12h蛋白和Cas12i蛋白。在一些情况下,(主题组合物和/或方法的)Cas12L蛋白包含具有RuvC结构域(其包括RuvC-I、RuvC-II和RuvC-III结构域)的氨基酸序列,其与图2中描绘的氨基酸序列的RuvC结构域(例如,图2中描绘的任何Cas12L氨基酸序列的RuvC结构域)的序列同一性高于与以下任一者的RuvC结构域的序列同一性:Cas12a蛋白、Cas12b蛋白、Cas12c蛋白、Cas12d蛋白、Cas12e蛋白、Cas12 g蛋白、Cas12h蛋白和Cas12i蛋白。

[0082] 图3A-3B提供了存在于图2A-2Z、图2AA-2ZZ和图2AAA-2FFF中描绘的各种Cas12L多肽的RuvC结构域中的活性位点残基的位置。例如,Cas12L_1_257905508(图2A)的活性位点残基是氨基酸残基336、530和682。

[0083] 图5A-5P提供了图2A-2Z、图2AA-2ZZ和图2AAA-2FFF中描绘的58种Cas12L多肽的氨基酸序列比对。如图5A-5P所示,许多氨基酸在58种Cas12L多肽中是保守的。例如,基于

Cas12L_1_257905508(图2A)的氨基编号,K14、T15、E37、Y41、Y42、N43、S46、I49、I57、Y73、Y88、F91、N119、L121、N140、Y141、Q169、I171、E173、W185、Y191、Y205、L209、F212、Y213、F235、G236、G237、C238、R240、G281、N309、K324、I335、T346、Y357、N359、I360、Y361、V364、F370、F391、P393、L394、E395、L399、S401、R402、Q406、E417、V423、L424、Y442、K448、L449、R450、K454、A455、V459、K460、Y470、D471、E476、E477和G594是保守的。因此,例如,Cas12L多肽可包含以下一种或多种:基于Cas12L_1_257905508(图2A)的氨基酸编号K14、T15、E37、Y41、Y42、N43、S46、I49、I57、Y73、Y88、F91、N119、L121、N140、Y141、Q169、I171、E173、W185、Y191、Y205、L209、F212、Y213、F235、G236、G237、C238、R240、G281、N309、K324、I335、T346、Y357、N359、I360、Y361、V364、F370、F391、P393、L394、E395、L399、S401、R402、Q406、E417、V423、L424、Y442、K448、L449、R450、K454、A455、V459、K460、Y470、D471、E476、E477和G594,或另一仲Cas12L多肽(例如,图2A-2Z、图2AA-2ZZ和图2AAA-2FFF中描绘的Cas12L多肽)中相应位置的相同氨基酸,其中基于图5A-5P中提供的比对,本领域技术人员将理解"相应位置"将是哪些。Cas12L多肽可包含在基于Cas12L_1_257905508(图2A)的氨基酸编号K14、T15、E37、Y41、Y42、N43、S46、I49、I57、Y73、Y88、F91、N119、L121、N140、Y141、Q169、I171、E173、W185、Y191、Y205、L209、F212、Y213、F235、G236、G237、C238、R240、G281、N309、K324、I335、T346、Y357、N359、I360、Y361、V364、F370、F391、P393、L394、E395、L399、S401、R402、Q406、E417、V423、L424、Y442、K448、L449、R450、K454、A455、V459、K460、Y470、D471、E476、E477和G594中一个或多个处或在另一种Cas12L多肽(例如图2A-2Z、图2AA-2ZZ和图2AAA-2FFF中描绘的Cas12L多肽)中的相应位置处的保守性氨基酸取代。

[0084] 图6A-6C提供了Cas12L_2_196848753(图2B)、Cas12L_29_255238293(图2BB;在图2BB中称为Cas12L_28_255238293)、Cas12L_38_73472625(图2KK;在图2KK中称为Cas12L_37_73472625)、Cas12L_42_74408273_部分(图2NN;在图2NN中简称为Cas12L_41_74408273_部分)、Cas12L_54_75257103_部分(图2PP;在图2PP中称为Cas12L_43_75257103_部分)、Cas12L_52_82983331(图2XX;在图2XX中称为Cas12L_51_82983331)以及Cas12L_56_77738117(图2BBB;在图2BBB中称为Cas12L_55_77738117。方框中的氨基酸在图6A-6C中的Cas12L多肽中是保守的。Cas12L多肽可包含图6A-6C的方框中列出的氨基酸或这些氨基酸在图6A-6C所示位置处的保守氨基酸取代。

[0085] 在一些情况下,(主题组合物和/或方法的)Cas12L蛋白包含与图2(例如,图2A-2Z、图2AA-2ZZ和图2AAA-2FFF中任一者)中描绘的任一种Cas12L氨基酸序列的RuvC结构域(其包括RuvC-I、RuvC-II和RuvC-III结构域)具有20%或更多序列同一性(例如,30%或更多、40%或更多、50%或更多、60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的氨基酸序列。在一些情况下,(主题组合物和/或方法的)Cas12L蛋白包含与图2(例如,图2A-2Z、图2AA-2ZZ和图2AAA-2FFF中任一者)中描绘的任一种Cas12L氨基酸序列的RuvC结构域(其包括RuvC-I、RuvC-II和RuvC-III结构域)具有70%或更多序列同一性(例如,75%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的氨基酸序列。在一些情况下,(主题组合物和/或方法的)Cas12L蛋白包含图2(例如,图2A-2Z、图2AA-2ZZ和图2AAA-2FFF中任一者)中描绘的任一种Cas12L氨基酸序列的RuvC结构域(其包括RuvC-I、RuvC-II和RuvC-III结构域)。

[0086] 在一些情况下，结合Cas12L多肽的引导RNA包含图2所示的核苷酸序列(例如，图2A-2Z、图2AA-2ZZ和图2AAA-2FFF中任一者；其中“T”被替换为“U”)(或者在某些情况下是其反义互补序列)。在一些情况下，引导RNA包含核苷酸序列(N)nX或其反向互补序列，其中N是任何核苷酸，n是15至30的整数(例如，15至20、17至25、17至22、18至22、18至20、20至25或25至30)，并且X是图2(例如，图2A-2Z、图2AA-2ZZ和图2AAA-2FFF中任一者)中描绘的任一种核苷酸序列(或者在某些情况下是其反义互补序列)。

[0087] 各种CasLambda(Casλ;Cas12L)引导RNA的重复区的核苷酸序列的比对提供于图7中。如图7所示，CRISPR重复序列包含跨同源物的保守序列基序。在一些情况下，结合Cas12L多肽的引导RNA包含以下保守序列的核苷酸序列(重复序列；或蛋白质结合序列)：WAUUGUUGUARMWNYYWUUUURUAWGGWKURAACAAC(SEQ ID NO:162)，其中W是A或U；R是G或A；M是A或C；N是A、G、C或U；Y是U或C；并且K是G或U。例如，结合CasLambda多肽的引导RNA(Cas12L 28；图2BB)可包含有包含以下核苷酸序列的蛋白质结合区段：AUUGUUGAAAUAGUACUUUUAUAGUCUAUAUACAAC(SEQ ID NO:63)。作为另一个实例，结合CasLambda37多肽(Cas12L 37；图2KK)的引导RNA可包含有包含以下核苷酸序列的蛋白质结合区段：AUUGUUGUAACAUCUAUUUUGUAAGGUGUAAACAAC(SEQ ID NO:70)。作为另一个实例，结合CasLambda29多肽(Cas12L 29；图2CC)的引导RNA可包含有包含以下核苷酸序列的蛋白质结合区段：UAUUGUUGUAACUCUUAUUUUGUAUGGAGUAAACAAC(SEQ ID NO:64)。作为另一个实例，结合CasLambda42多肽(Cas12L 42；图OO)的引导RNA可包含有包含以下核苷酸序列的蛋白质结合区段：AAUUGUUGUAACUCUUAUUUUGUAUGGAGUAAACAAC(SEQ ID NO:163)。作为另一个实例，结合Casλ多肽1、3、4、5、7-15、18、19、21-27、29-35、38、40-46、48、49、53和54中任一者(Cas12L 1、3、4、5、7-15、18、19、21-27、29-35、38、40-46、48、49、53和54中任一者；分别为图2A、2C、2D、2E、2G-2O、2R、2S、2U-2AA、2CC-2II、2LL、2MM-2SS、2UU、2VV、2ZZ和2AAA)的引导RNA可包含有包含以下核苷酸序列的蛋白质结合区段：AUUGUUGUAACUCUUAUUUUGUAUGGAGUAAACAAC(SEQ ID NO:59)。作为另一个实例，结合CasLambda47多肽(Cas12L 47；图2TT)的引导RNA可包含有包含以下核苷酸序列的蛋白质结合区段：AUUGUUGUAACUUUUAUUUUGUAUGGAGUAAACAAC(SEQ ID NO:65)。作为另一个实例，结合CasLambda2多肽(Cas12L 2；图2B)的引导RNA可包含有包含以下核苷酸序列的蛋白质结合区段：AUUGUUGUAGACCUCUUUUUAUAAGGAUUGAACAAC(SEQ ID NO:60)。本公开的Cas12L多肽可以与包含图7中描绘的蛋白质结合区段的引导RNA形成复合物(核糖核蛋白(RNP)复合物)。在一些情况下，本公开的Cas12L多肽可以与包含图7中描绘的不同序列的蛋白质结合区段的不同引导RNA形成RNP复合物。

[0088] 作为另一个实例，结合CasLambda20多肽或CasLambda52多肽(Cas12L 20或Cas12L 52；分别为图2T和2YY)的引导RNA可包含有包含以下核苷酸序列的蛋白质结合区段：AAUGUUGUAGAUGCCUUUUUAUAAGGAUUAAACAAC(SEQ ID NO:68)。作为另一个实例，结合CasLambda50多肽(Cas12L 50；图2WW)的引导RNA可包含有包含以下核苷酸序列的蛋白质结合区段：AAUGUUGUAGAUACCUUUUUGUAAGGAUUGAACAAC(SEQ ID NO:66)。作为另一个实例，结合CasLambda16多肽(Cas12L 16；图2P)的引导RNA可包含有包含以下核苷酸序列的蛋白质结合区段：UAUUGUUGUAGAUACCUUUUUGUAAGGAUUAAACAAC(SEQ ID NO:61)。作为另一个实例，结合CasLambda6多肽(Cas12L 6；图2F)的引导RNA可包含有包含以下核苷酸序列的蛋白质结合区段：AUUGUUGUAGAUACCUUUUUGUAAGGAUUGAACAAC(SEQ ID NO:164)。作为另一个实例，结

合CasLambda51多肽(Cas12L 51;图2XX)的引导RNA可包含有保安以下核苷酸序列的蛋白质结合区段:AUUGUUGUAAUACUAUUUUUGUAAAGUAUAAACAAC(SEQ ID NO:67)。作为另一个实例,结合CasLambda55多肽(Cas12L 55;图2BBB)的引导RNA可包含有包含以下核苷酸序列的蛋白质结合区段:AUUGUUGUAAUACACUUUUUAUAAGGUAUGAACAAC(SEQ ID NO:69)。

[0089] 除了在重复(蛋白结合)区中包含保守序列基序以外,CasLambda引导RNA的重复区在同源物间共享保守的二级结构。图8A-8M描绘了各种CasLambda引导RNA的重复区的二级结构。例如,重复区可以包括可以形成茎和茎环结构的回文区。

[0090] 在一些情况下,结合Cas12L多肽的引导RNA包含与图2(例如,图2A-2Z、图2AA-2ZZ和图2AAA-2FFF中任一者)中描绘的任一种序列具有20%或更多序列同一性(例如,30%或更多、40%或更多、50%或更多、60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的氨基酸序列(或者在某些情况下是其反义互补序列)。在一些情况下,引导RNA包含核苷酸序列(N)nX或其反向互补序列,其中N是任何核苷酸,n是15至30的整数(例如,15至20、17至25、17至22、18至22、18至20、20至25或25至30),并且X是与图2(例如,图2A-2Z、图2AA-2ZZ和图2AAA-2FFF中任一者)中描绘的任一种序列具有20%或更多序列同一性的核苷酸序列(例如,30%或更多、40%或更多、50%或更多、60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的核苷酸序列。

[0091] 在一些情况下,结合Cas12L多肽的引导RNA包含与图2(例如,图2A-2Z、图2AA-2ZZ和图2AAA-2FFF中任一者)中描绘的任一种序列具有85%或更多序列同一性(例如,90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的氨基酸序列(或者在某些情况下是它的反义互补序列)。在一些情况下,引导RNA包含核苷酸序列(N)nX或其反向互补序列,其中N是任何核苷酸,n是15至30的整数(例如,15至20、17至25、17至22、18至22、18至20、20至25或25至30),并且X是与图2(例如,图2A-2Z、图2AA-2ZZ和图2AAA-2FFF中任一者)中描绘的任一种序列具有85%或更多序列同一性的核苷酸序列(例如,90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的核苷酸序列。

[0092] 在一些情况下,结合Cas12L多肽的引导RNA包含图2所示的核苷酸序列(例如,图2A-2Z、图2AA-2ZZ和图2AAA-2FFF中任一者)(或者在某些情况下是其反义互补序列)。在一些情况下,引导RNA包含核苷酸序列X(N)n,其中N是任何核苷酸,n是15至30的整数(例如,15至20、17至25、17至22、18至22、18至20、20至25或25至30),并且X是图2(例如,图2A-2Z、图2AA-2ZZ和图2AAA-2FFF中任一者)中描绘的任一种核苷酸序列(或者在某些情况下是其反义互补序列)。

[0093] 在一些情况下,结合Cas12L多肽的引导RNA包含与图2(例如,图2A-2Z、图2AA-2ZZ和图2AAA-2FFF中任一者)中描绘的任一种序列具有20%或更多序列同一性(例如,30%或更多、40%或更多、50%或更多、60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的氨基酸序列(或者在某些情况下是其反义互补序列)。在一些情况下,引导RNA包含核苷酸序列X(N)n,其中N是任何核苷酸,n是15至30的整数(例如,15至20、17至25、17至22、18至22、18至20、20

至25或25至30)，并且X是与图2(例如，图2A-2Z、图2AA-2ZZ和图2AAA-2FFF中任一者)中描绘的任一种序列具有20％或更多序列同一性的核苷酸序列(例如，30％或更多、40％或更多、50％或更多、60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的核苷酸序列。

[0094] 在一些情况下，结合Cas12L多肽的引导RNA包含与图2(例如，图2A-2Z、图2AA-2ZZ和图2AAA-2FFF中任一者)中描绘的任一种序列具有85％或更多序列同一性(例如，90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列(或者在某些情况下是它的反义互补序列)。在一些情况下，引导RNA包含核苷酸序列X(N)n，其中N是任何核苷酸，n是15至30的整数(例如，15至20、17至25、17至22、18至22、18至20、20至25或25至30)，并且X是与图2(例如，图2A-2Z、图2AA-2ZZ和图2AAA-2FFF中任一者)中描绘的任一种序列具有85％或更多序列同一性的核苷酸序列(例如，90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的核苷酸序列。

[0095] 图2A

[0096] 在一些情况下，(主题组合物和/或方法的)Cas12L蛋白包含与图2A中描绘并表示为“Cas12L_1_257905508”的Cas12L氨基酸序列具有20％或更多序列同一性(例如，30％或更多、40％或更多、50％或更多、60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。例如，在一些情况下，Cas12L蛋白包含与图2A中描绘的Cas12L氨基酸序列具有50％或更多序列同一性(例如，60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。在一些情况下，Cas12L蛋白包含与图2A中描绘的Cas12L氨基酸序列具有80％或更多序列同一性(例如，85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。在一些情况下，Cas12L蛋白包含与图2A中描绘的Cas12L氨基酸序列具有90％或更多序列同一性(例如，95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。在一些情况下，Cas12L蛋白包含具有图2A中描绘的Cas12L氨基酸序列的氨基酸序列。在一些情况下，Cas12L蛋白包含具有图2A中描绘的Cas12L蛋白序列的氨基酸序列，除了所述序列包含降低所述蛋白质的天然存在的催化活性的氨基酸取代(例如，1、2或3个氨基酸取代)。在一些情况下，Cas12L多肽具有700个氨基酸(aa)至800aa的长度，例如，700aa至725aa、725aa至746aa、746aa至775aa或775aa至800aa)。在一些情况下，Cas12L多肽具有746个氨基酸的长度。在一些情况下，结合Cas12L多肽(例如，包含与图2A中描绘的Cas12L氨基酸序列具有20％或更多、30％或更多、40％或更多、50％或更多、60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％氨基酸序列同一性的氨基酸序列的Cas12L多肽)的引导RNA包含以下核苷酸序列:ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:1)或其反向互补序列。在一些情况下，引导RNA包含核苷酸序列(N)nATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:2)或其反向互补序列，其中N是任何核苷酸，并且n是15至30的整数，例如15至20、17至25、17至22、18至22、18至20、20至25或25至30)。

[0097] 图2B

[0098] 在一些情况下，(主题组合物和/或方法的)Cas12L蛋白包含与图2B中描绘并表示

为“Cas12L_2_196848753”的Cas12L氨基酸序列具有20％或更多序列同一性(例如,30％或更多、40％或更多、50％或更多、60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。例如,在一些情况下,Cas12L蛋白包含与图2B中描绘的Cas12L氨基酸序列具有50％或更多序列同一性(例如,60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含与图2B中描绘的Cas12L氨基酸序列具有80％或更多序列同一性(例如,85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含与图2B中描绘的Cas12L氨基酸序列具有90％或更多序列同一性(例如,95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含具有图2B中描绘的Cas12L氨基酸序列的氨基酸序列。在一些情况下,Cas12L蛋白包含具有图2B中描绘的Cas12L蛋白序列的氨基酸序列,除了所述序列包含降低所述蛋白质的天然存在的催化活性的氨基酸取代(例如,1、2或3个氨基酸取代)。在一些情况下,Cas12L多肽具有675个氨基酸(aa)至800aa的长度,例如,675aa至700aa、700aa至725aa、725aa至735aa、735aa至775aa或775aa至800aa)。在一些情况下,Cas12L多肽具有735个氨基酸的长度。在一些情况下,结合Cas12L多肽(例如,包含与图2B中描绘的Cas12L氨基酸序列具有20％或更多、30％或更多、40％或更多、50％或更多、60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％氨基酸序列同一性的氨基酸序列的Cas12L多肽)的引导RNA包含以下核苷酸序列:ATTGTTGTAGACCTCTTTTTATAAGGATTGAACAAC(SEQ ID NO:3)或其反向互补序列。在一些情况下,引导RNA包含核苷酸序列(N)nATTGTTGTAGACCTCTTTTTATAAGGATTGAACAAC(SEQ ID NO:4)或其反向互补序列,其中N是任何核苷酸,并且n是15至30的整数,例如15至20、17至25、17至22、18至22、18至20、20至25或25至30)。

[0099] 图2C

[0100] 在一些情况下,(主题组合物和/或方法的)Cas12L蛋白包含与图2C中描绘并表示为“Cas12L_3_66741167”的Cas12L氨基酸序列具有20％或更多序列同一性(例如,30％或更多、40％或更多、50％或更多、60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。例如,在一些情况下,Cas12L蛋白包含与图2C中描绘的Cas12L氨基酸序列具有50％或更多序列同一性(例如,60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含与图2C中描绘的Cas12L氨基酸序列具有80％或更多序列同一性(例如,85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含与图2C中描绘的Cas12L氨基酸序列具有90％或更多序列同一性(例如,95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含具有图2C中描绘的Cas12L氨基酸序列的氨基酸序列。在一些情况下,Cas12L蛋白包含具有图2C中描绘的Cas12L蛋白序列的氨基酸序列,除了所述序列包含降低所述蛋白质的天然存在的催化活性

的氨基酸取代(例如,1、2或3个氨基酸取代)。在一些情况下,Cas12L多肽具有725个氨基酸(aa)至875aa的长度,例如,725aa至750aa、750aa至778aa、778aa至800aa、800aa至825aa、825aa至850aa或850aa至875aa)。在一些情况下,Cas12L多肽具有778个氨基酸的长度。在一些情况下,结合Cas12L多肽(例如,包含与图2C中描绘的Cas12L氨基酸序列具有20%或更多、30%或更多、40%或更多、50%或更多、60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%氨基酸序列同一性的氨基酸序列的Cas12L多肽)的引导RNA包含以下核苷酸序列:ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:1)或其反向互补序列。在一些情况下,引导RNA包含核苷酸序列(N)nATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:2)或其反向互补序列,其中N是任何核苷酸,并且n是15至30的整数,例如15至20、17至25、17至22、18至22、18至20、20至25或25至30)。

[0101] 图2D

[0102] 在一些情况下,(主题组合物和/或方法的)Cas12L蛋白包含与图2D中描绘并表示为"Cas12L_4_67031163"的Cas12L氨基酸序列具有20%或更多序列同一性(例如,30%或更多、40%或更多、50%或更多、60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的氨基酸序列。例如,在一些情况下,Cas12L蛋白包含与图2D中描绘的Cas12L氨基酸序列具有50%或更多序列同一性(例如,60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含与图2D中描绘的Cas12L氨基酸序列具有80%或更多序列同一性(例如,85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含与图2D中描绘的Cas12L氨基酸序列具有90%或更多序列同一性(例如,95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含具有图2D中描绘的Cas12L氨基酸序列的氨基酸序列。在一些情况下,Cas12L蛋白包含具有图2D中描绘的Cas12L蛋白序列的氨基酸序列,除了所述序列包含降低所述蛋白质的天然存在的催化活性的氨基酸取代(例如,1、2或3个氨基酸取代)。在一些情况下,Cas12L多肽具有725个氨基酸(aa)至875aa的长度,例如,725aa至750aa、750aa至778aa、778aa至800aa、800aa至825aa、825aa至850aa或850aa至875aa)。在一些情况下,Cas12L多肽具有778个氨基酸的长度。在一些情况下,结合Cas12L多肽(例如,包含与图2D中描绘的Cas12L氨基酸序列具有20%或更多、30%或更多、40%或更多、50%或更多、60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%氨基酸序列同一性的氨基酸序列的Cas12L多肽)的引导RNA包含以下核苷酸序列:ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:1)或其反向互补序列。在一些情况下,引导RNA包含核苷酸序列(N)nATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:2)或其反向互补序列,其中N是任何核苷酸,并且n是15至30的整数,例如15至20、17至25、17至22、18至22、18至20、20至25或25至30)。

[0103] 图2E

[0104] 在一些情况下,(主题组合物和/或方法的)Cas12L蛋白包含与图2E中描绘并表示

为“Cas12L_5_67793351”的Cas12L氨基酸序列具有20％或更多序列同一性(例如,30％或更多、40％或更多、50％或更多、60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。例如,在一些情况下,Cas12L蛋白包含与图2E中描绘的Cas12L氨基酸序列具有50％或更多序列同一性(例如,60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含与图2E中描绘的Cas12L氨基酸序列具有80％或更多序列同一性(例如,85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含与图2E中描绘的Cas12L氨基酸序列具有90％或更多序列同一性(例如,95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含具有图2E中描绘的Cas12L氨基酸序列的氨基酸序列。在一些情况下,Cas12L蛋白包含具有图2E中描绘的Cas12L蛋白序列的氨基酸序列,除了所述序列包含降低所述蛋白质的天然存在的催化活性的氨基酸取代(例如,1、2或3个氨基酸取代)。在一些情况下,Cas12L多肽具有725个氨基酸(aa)至875aa的长度,例如,725aa至750aa、750aa至778aa、778aa至800aa、800aa至825aa、825aa至850aa或850aa至875aa)。在一些情况下,Cas12L多肽具有778个氨基酸的长度。在一些情况下,结合Cas12L多肽(例如,包含与图2E中描绘的Cas12L氨基酸序列具有20％或更多、30％或更多、40％或更多、50％或更多、60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％氨基酸序列同一性的氨基酸序列的Cas12L多肽)的引导RNA包含以下核苷酸序列:ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:1)或其反向互补序列。在一些情况下,引导RNA包含核苷酸序列(N)nATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:2)或其反向互补序列,其中N是任何核苷酸,并且n是15至30的整数,例如15至20、17至25、17至22、18至22、18至20、20至25或25至30)。

[0105] 图2F

[0106] 在一些情况下,(主题组合物和/或方法的)Cas12L蛋白包含与图2F中描绘并表示为“Cas12L_6_67912869_部分”的Cas12L氨基酸序列具有20％或更多序列同一性(例如,30％或更多、40％或更多、50％或更多、60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的约350个氨基酸的连续片段。例如,在一些情况下,Cas12L蛋白包含与图2F中描绘的Cas12L氨基酸序列具有50％或更多序列同一性(例如,60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的约350个氨基酸的连续片段。在一些情况下,Cas12L蛋白包含与图2F中描绘的Cas12L氨基酸序列具有80％或更多序列同一性(例如,85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的约350个氨基酸的连续片段。在一些情况下,Cas12L蛋白包含与图2F中描绘的Cas12L氨基酸序列具有90％或更多序列同一性(例如,95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的约350个氨基酸的连续片段。在一些情况下,Cas12L蛋白包含具有图2F中描绘的Cas12L氨基酸序列的约350个氨基酸的连续片段。在一些情况下,Cas12L蛋白包含具有图2F中描绘的Cas12L蛋白序

列的约350个氨基酸的连续片段，除了所述序列包含降低所述蛋白质的天然存在的催化活性的氨基酸取代(例如，1、2或3个氨基酸取代)。在一些情况下，Cas12L多肽具有700个氨基酸(aa)至800aa的长度，例如，700aa至725aa、725aa至750aa或750aa至800aa)。在一些情况下，Cas12L多肽具有约725个氨基酸至约775个氨基酸的长度。在一些情况下，结合Cas12L多肽(例如，包含与图2F中描绘的Cas12L氨基酸序列具有20％或更多、30％或更多、40％或更多、50％或更多、60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％氨基酸序列同一性的氨基酸序列的Cas12L多肽)的引导RNA包含以下核苷酸序列：ATTGTTGTAGATACCTTTTTGTAAGGATTGAACAAC(SEQ ID NO:5)或其反向互补序列。在一些情况下，引导RNA包含核苷酸序列(N)nATTGTTGTAGATACCTTTTTGTAAGGATTGAACAAC(SEQ ID NO:6)或其反向互补序列，其中N是任何核苷酸，并且n是15至30的整数，例如15至20、17至25、17至22、18至22、18至20、20至25或25至30)。

[0107] 图2G

[0108] 在一些情况下，(主题组合物和/或方法的)Cas12L蛋白包含与图2G中描绘并表示为“Cas12L_7_68090316_部分”的Cas12L氨基酸序列具有20％或更多序列同一性(例如，30％或更多、40％或更多、50％或更多、60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。例如，在一些情况下，Cas12L蛋白包含与图2G中描绘的Cas12L氨基酸序列具有50％或更多序列同一性(例如，60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。在一些情况下，Cas12L蛋白包含与图2G中描绘的Cas12L氨基酸序列具有80％或更多序列同一性(例如，85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。在一些情况下，Cas12L蛋白包含与图2G中描绘的Cas12L氨基酸序列具有90％或更多序列同一性(例如，95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。在一些情况下，Cas12L蛋白包含具有图2G中描绘的Cas12L氨基酸序列的氨基酸序列。在一些情况下，Cas12L蛋白包含具有图2G中描绘的Cas12L蛋白序列的氨基酸序列，除了所述序列包含降低所述蛋白质的天然存在的催化活性的氨基酸取代(例如，1、2或3个氨基酸取代)。在一些情况下，Cas12L多肽具有650个氨基酸(aa)至800aa的长度，例如，650aa至700aa、700aa至717aa、717aa至725aa、725aa至750aa或750aa至800aa)。在一些情况下，Cas12L多肽具有约717个氨基酸至约775个氨基酸的长度。在一些情况下，结合Cas12L多肽(例如，包含与图2G中描绘的Cas12L氨基酸序列具有20％或更多、30％或更多、40％或更多、50％或更多、60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％氨基酸序列同一性的氨基酸序列的Cas12L多肽)的引导RNA包含以下核苷酸序列：ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:1)或其反向互补序列。在一些情况下，引导RNA包含核苷酸序列(N)nATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:2)或其反向互补序列，其中N是任何核苷酸，并且n是15至30的整数，例如15至20、17至25、17至22、18至22、18至20、20至25或25至30)。

[0109] 图2H

[0110] 在一些情况下，(主题组合物和/或方法的)Cas12L蛋白包含与图2H中描绘并表示

为“Cas12L_8_68328292_部分”的Cas12L氨基酸序列具有20％或更多序列同一性(例如，30％或更多、40％或更多、50％或更多、60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。例如，在一些情况下，Cas12L蛋白包含与图2H中描绘的Cas12L氨基酸序列具有50％或更多序列同一性(例如，60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。在一些情况下，Cas12L蛋白包含与图2H中描绘的Cas12L氨基酸序列具有80％或更多序列同一性(例如，85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。在一些情况下，Cas12L蛋白包含与图2H中描绘的Cas12L氨基酸序列具有90％或更多序列同一性(例如，95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。在一些情况下，Cas12L蛋白包含具有图2H中描绘的Cas12L氨基酸序列的氨基酸序列。在一些情况下，Cas12L蛋白包含具有图2H中描绘的Cas12L蛋白序列的氨基酸序列，除了所述序列包含降低所述蛋白质的天然存在的催化活性的氨基酸取代(例如，1、2或3个氨基酸取代)。在一些情况下，Cas12L多肽具有700个氨基酸(aa)至800aa的长度，例如，700aa至725aa、725aa至742aa、742aa至750aa、750aa至775aa或775aa至800aa)。在一些情况下，Cas12L多肽具有约742个氨基酸至约775个氨基酸的长度。在一些情况下，结合Cas12L多肽(例如，包含与图2H中描绘的Cas12L氨基酸序列具有20％或更多、30％或更多、40％或更多、50％或更多、60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％氨基酸序列同一性的氨基酸序列的Cas12L多肽)的引导RNA包含以下核苷酸序列:ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:1)或其反向互补序列。在一些情况下，引导RNA包含核苷酸序列(N)nATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:2)或其反向互补序列，其中N是任何核苷酸，并且n是15至30的整数，例如15至20、17至25、17至22、18至22、18至20、20至25或25至30)。

[0111] 图2I

[0112] 在一些情况下，(主题组合物和/或方法的)Cas12L蛋白包含与图2I中描绘并表示为“Cas12L_9_68454124”的Cas12L氨基酸序列具有20％或更多序列同一性(例如，30％或更多、40％或更多、50％或更多、60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。例如，在一些情况下，Cas12L蛋白包含与图2I中描绘的Cas12L氨基酸序列具有50％或更多序列同一性(例如，60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。在一些情况下，Cas12L蛋白包含与图2I中描绘的Cas12L氨基酸序列具有80％或更多序列同一性(例如，85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。在一些情况下，Cas12L蛋白包含与图2I中描绘的Cas12L氨基酸序列具有90％或更多序列同一性(例如，95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。在一些情况下，Cas12L蛋白包含具有图2I中描绘的Cas12L氨基酸序列的氨基酸序列。在一些情况下，Cas12L蛋白包含具有图2I中描绘的Cas12L蛋白序列的氨基酸序列，除了所述序列包含降低所述蛋白质的天然存在的催化活性

的氨基酸取代(例如,1、2或3个氨基酸取代)。在一些情况下,Cas12L多肽具有700个氨基酸(aa)至800aa的长度,例如,700aa至725aa、725aa至746aa、746aa至750aa、750aa至775aa或775aa至800aa)。在一些情况下,Cas12L多肽具有746个氨基酸的长度。在一些情况下,结合Cas12L多肽(例如,包含与图2I中描绘的Cas12L氨基酸序列具有20%或更多、30%或更多、40%或更多、50%或更多、60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%氨基酸序列同一性的氨基酸序列的Cas12L多肽)的引导RNA包含以下核苷酸序列:ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:1)或其反向互补序列。在一些情况下,引导RNA包含核苷酸序列(N)nATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:2)或其反向互补序列,其中N是任何核苷酸,并且n是15至30的整数,例如15至20、17至25、17至22、18至22、18至20、20至25或25至30)。

[0113] 图2J

[0114] 在一些情况下,(主题组合物和/或方法的)Cas12L蛋白包含与图2J中描绘并表示为"Cas12L_10_68605313"的Cas12L氨基酸序列具有20%或更多序列同一性(例如,30%或更多、40%或更多、50%或更多、60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的氨基酸序列。例如,在一些情况下,Cas12L蛋白包含与图2J中描绘的Cas12L氨基酸序列具有50%或更多序列同一性(例如,60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含与图2J中描绘的Cas12L氨基酸序列具有80%或更多序列同一性(例如,85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含与图2J中描绘的Cas12L氨基酸序列具有90%或更多序列同一性(例如,95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含具有图2J中描绘的Cas12L氨基酸序列的氨基酸序列。在一些情况下,Cas12L蛋白包含具有图2J中描绘的Cas12L蛋白序列的氨基酸序列,除了所述序列包含降低所述蛋白质的天然存在的催化活性的氨基酸取代(例如,1、2或3个氨基酸取代)。在一些情况下,Cas12L多肽具有725个氨基酸(aa)至825aa的长度,例如,725aa至750aa、750aa至775aa、775aa至782aa、782aa至800aa或800aa至825aa)。在一些情况下,Cas12L多肽具有782个氨基酸的长度。在一些情况下,结合Cas12L多肽(例如,包含与图2J中描绘的Cas12L氨基酸序列具有20%或更多、30%或更多、40%或更多、50%或更多、60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%氨基酸序列同一性的氨基酸序列的Cas12L多肽)的引导RNA包含以下核苷酸序列:ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:1)或其反向互补序列。在一些情况下,引导RNA包含核苷酸序列(N)nATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:2)或其反向互补序列,其中N是任何核苷酸,并且n是15至30的整数,例如15至20、17至25、17至22、18至22、18至20、20至25或25至30)。

[0115] 图2K

[0116] 在一些情况下,(主题组合物和/或方法的)Cas12L蛋白包含与图2K中描绘并表示

为“Cas12L_11_69266821_部分”的Cas12L氨基酸序列具有20％或更多序列同一性(例如，30％或更多、40％或更多、50％或更多、60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的约92个氨基酸的连续片段。例如，在一些情况下，Cas12L蛋白包含与图2K中描绘的Cas12L氨基酸序列具有50％或更多序列同一性(例如，60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的约92个氨基酸的连续片段。在一些情况下，Cas12L蛋白包含与图2K中描绘的Cas12L氨基酸序列具有80％或更多序列同一性(例如，85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的约92个氨基酸的连续片段。在一些情况下，Cas12L蛋白包含与图2K中描绘的Cas12L氨基酸序列具有90％或更多序列同一性(例如，95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的约92个氨基酸的连续片段。在一些情况下，Cas12L蛋白包含具有图2K中描绘的Cas12L氨基酸序列的约92个氨基酸的连续片段。在一些情况下，Cas12L蛋白包含具有图2K中描绘的Cas12L蛋白序列的约92个氨基酸的连续片段，除了所述序列包含降低所述蛋白质的天然存在的催化活性的氨基酸取代(例如，1、2或3个氨基酸取代)。在一些情况下，Cas12L多肽具有700个氨基酸(aa)至800aa的长度，例如，700aa至725aa、725aa至750aa、750aa至775aa或775aa至800aa)。在一些情况下，Cas12L多肽具有725个氨基酸至775个氨基酸的长度。在一些情况下，结合Cas12L多肽(例如，包含与图2K中描绘的Cas12L氨基酸序列具有20％或更多、30％或更多、40％或更多、50％或更多、60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％氨基酸序列同一性的氨基酸序列的Cas12L多肽)的引导RNA包含以下核苷酸序列：ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:1)或其反向互补序列。在一些情况下，引导RNA包含核苷酸序列(N)nATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:2)或其反向互补序列，其中N是任何核苷酸，并且n是15至30的整数，例如15至20、17至25、17至22、18至22、18至20、20至25或25至30)。

[0117] 图2L

[0118] 在一些情况下，(主题组合物和/或方法的)Cas12L蛋白包含与图2L中描绘并表示为“Cas12L_12_69417229_部分”的Cas12L氨基酸序列具有20％或更多序列同一性(例如，30％或更多、40％或更多、50％或更多、60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的约92个氨基酸的连续片段。例如，在一些情况下，Cas12L蛋白包含与图2L中描绘的Cas12L氨基酸序列具有50％或更多序列同一性(例如，60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的约92个氨基酸的连续片段。在一些情况下，Cas12L蛋白包含与图2L中描绘的Cas12L氨基酸序列具有80％或更多序列同一性(例如，85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的约92个氨基酸的连续片段。在一些情况下，Cas12L蛋白包含与图2L中描绘的Cas12L氨基酸序列具有90％或更多序列同一性(例如，95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的约92个氨基酸的连续片段。在一些情况下，Cas12L蛋白包含具有图2L中描绘的Cas12L氨基酸序列的约92个氨基酸的连续片段。在一些情况下，Cas12L蛋白包含具有图2L中描绘的Cas12L蛋白序列的

约92个氨基酸的连续片段，除了所述序列包含降低所述蛋白质的天然存在的催化活性的氨基酸取代(例如，1、2或3个氨基酸取代)。在一些情况下，Cas12L多肽具有700个氨基酸(aa)至800aa的长度，例如，700aa至725aa、725aa至750aa、750aa至775aa或775aa至800aa)。在一些情况下，Cas12L多肽具有725个氨基酸至775个氨基酸的长度。在一些情况下，结合Cas12L多肽(例如，包含与图2L中描绘的Cas12L氨基酸序列具有20％或更多、30％或更多、40％或更多、50％或更多、60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％氨基酸序列同一性的氨基酸序列的Cas12L多肽)的引导RNA包含以下核苷酸序列：ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:1)或其反向互补序列。在一些情况下，引导RNA包含核苷酸序列(N)nATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:2)或其反向互补序列，其中N是任何核苷酸，并且n是15至30的整数，例如15至20、17至25、17至22、18至22、18至20、20至25或25至30)。

[0119] 图2M

[0120] 在一些情况下，(主题组合物和/或方法的)Cas12L蛋白包含与图2M中描绘并表示为“Cas12L_13_69733214”的Cas12L氨基酸序列具有20％或更多序列同一性(例如，30％或更多、40％或更多、50％或更多、60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。例如，在一些情况下，Cas12L蛋白包含与图2M中描绘的Cas12L氨基酸序列具有50％或更多序列同一性(例如，60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。在一些情况下，Cas12L蛋白包含与图2M中描绘的Cas12L氨基酸序列具有80％或更多序列同一性(例如，85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。在一些情况下，Cas12L蛋白包含与图2M中描绘的Cas12L氨基酸序列具有90％或更多序列同一性(例如，95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。在一些情况下，Cas12L蛋白包含具有图2M中描绘的Cas12L氨基酸序列的氨基酸序列。在一些情况下，Cas12L蛋白包含具有图2M中描绘的Cas12L蛋白序列的氨基酸序列，除了所述序列包含降低所述蛋白质的天然存在的催化活性的氨基酸取代(例如，1、2或3个氨基酸取代)。在一些情况下，Cas12L多肽具有700个氨基酸(aa)至800aa的长度，例如，700aa至725aa、725aa至750aa、750aa至775aa或775aa至800aa)。在一些情况下，Cas12L多肽具有746个氨基酸的长度。在一些情况下，结合Cas12L多肽(例如，包含与图2M中描绘的Cas12L氨基酸序列具有20％或更多、30％或更多、40％或更多、50％或更多、60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％氨基酸序列同一性的氨基酸序列的Cas12L多肽)的引导RNA包含以下核苷酸序列：ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:1)或其反向互补序列。在一些情况下，引导RNA包含核苷酸序列(N)nATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:2)或其反向互补序列，其中N是任何核苷酸，并且n是15至30的整数，例如15至20、17至25、17至22、18至22、18至20、20至25或25至30)。

[0121] 图2N

[0122] 在一些情况下，(主题组合物和/或方法的)Cas12L蛋白包含与图2N中描绘并表示为“Cas12L_14_70235246_部分”的Cas12L氨基酸序列具有20％或更多序列同一性(例如，

30％或更多、40％或更多、50％或更多、60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的约427个氨基酸的连续片段。例如,在一些情况下,Cas12L蛋白包含与图2N中描绘的Cas12L氨基酸序列具有50％或更多序列同一性(例如,60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的约427个氨基酸的连续片段。在一些情况下,Cas12L蛋白包含与图2N中描绘的Cas12L氨基酸序列具有80％或更多序列同一性(例如,85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的约427个氨基酸的连续片段。在一些情况下,Cas12L蛋白包含与图2N中描绘的Cas12L氨基酸序列具有90％或更多序列同一性(例如,95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的约427个氨基酸的连续片段。在一些情况下,Cas12L蛋白包含具有图2N中描绘的Cas12L氨基酸序列的约427个氨基酸的连续片段。在一些情况下,Cas12L蛋白包含具有图2N中描绘的Cas12L蛋白序列的约427个氨基酸的连续片段,除了所述序列包含降低所述蛋白质的天然存在的催化活性的氨基酸取代(例如,1、2或3个氨基酸取代)。在一些情况下,Cas12L多肽具有425个氨基酸(aa)至800aa的长度,例如,425aa至500aa、500aa至600aa、600aa至700aa、700aa至725aa、725aa至750aa、750aa至775aa或775aa至800aa)。在一些情况下,Cas12L多肽具有725个氨基酸至775个氨基酸的长度。在一些情况下,结合Cas12L多肽(例如,包含与图2N中描绘的Cas12L氨基酸序列具有20％或更多、30％或更多、40％或更多、50％或更多、60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％氨基酸序列同一性的氨基酸序列的Cas12L多肽)的引导RNA包含以下核苷酸序列:ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:1)或其反向互补序列。在一些情况下,引导RNA包含核苷酸序列(N)nATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:2)或其反向互补序列,其中N是任何核苷酸,并且n是15至30的整数,例如15至20、17至25、17至22、18至22、18至20、20至25或25至30)。

[0123] 图20

[0124] 在一些情况下,(主题组合物和/或方法的)Cas12L蛋白包含与图20中描绘并表示为“Cas12L_15_70724743”的Cas12L氨基酸序列具有20％或更多序列同一性(例如,30％或更多、40％或更多、50％或更多、60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。例如,在一些情况下,Cas12L蛋白包含与图20中描绘的Cas12L氨基酸序列具有50％或更多序列同一性(例如,60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含与图20中描绘的Cas12L氨基酸序列具有80％或更多序列同一性(例如,85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含与图20中描绘的Cas12L氨基酸序列具有90％或更多序列同一性(例如,95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含具有图20中描绘的Cas12L氨基酸序列的氨基酸序列。在一些情况下,Cas12L蛋白包含具有图20中描绘的Cas12L蛋白序列的氨基酸序列,除了所述序列包含降低所述蛋白质的天然存在的催化活性

的氨基酸取代(例如,1、2或3个氨基酸取代)。在一些情况下,Cas12L多肽具有700个氨基酸(aa)至800aa的长度,例如,700aa至725aa、725aa至746aa、746aa至775aa或775aa至800aa)。在一些情况下,Cas12L多肽具有746个氨基酸的长度。在一些情况下,结合Cas12L多肽(例如,包含与图20中描绘的Cas12L氨基酸序列具有20%或更多、30%或更多、40%或更多、50%或更多、60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%氨基酸序列同一性的氨基酸序列的Cas12L多肽)的引导RNA包含以下核苷酸序列:ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:1)或其反向互补序列。在一些情况下,引导RNA包含核苷酸序列(N)nATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:2)或其反向互补序列,其中N是任何核苷酸,并且n是15至30的整数,例如15至20、17至25、17至22、18至22、18至20、20至25或25至30)。

[0125] 图2P

[0126] 在一些情况下,(主题组合物和/或方法的)Cas12L蛋白包含与图2P中描绘并表示为"Cas12L_16_70731038"的Cas12L氨基酸序列具有20%或更多序列同一性(例如,30%或更多、40%或更多、50%或更多、60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的氨基酸序列。例如,在一些情况下,Cas12L蛋白包含与图2P中描绘的Cas12L氨基酸序列具有50%或更多序列同一性(例如,60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含与图2P中描绘的Cas12L氨基酸序列具有80%或更多序列同一性(例如,85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含与图2P中描绘的Cas12L氨基酸序列具有90%或更多序列同一性(例如,95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含具有图2P中描绘的Cas12L氨基酸序列的氨基酸序列。在一些情况下,Cas12L蛋白包含具有图2P中描绘的Cas12L蛋白序列的氨基酸序列,除了所述序列包含降低所述蛋白质的天然存在的催化活性的氨基酸取代(例如,1、2或3个氨基酸取代)。在一些情况下,Cas12L多肽具有675个氨基酸(aa)至800aa的长度,例如,675aa至700aa、700aa至725aa、725aa至735aa、735aa至750aa、750aa至775aa或775aa至800aa。在一些情况下,Cas12L多肽具有735个氨基酸的长度。在一些情况下,结合Cas12L多肽(例如,包含与图2P中描绘的Cas12L氨基酸序列具有20%或更多、30%或更多、40%或更多、50%或更多、60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%氨基酸序列同一性的氨基酸序列的Cas12L多肽)的引导RNA包含以下核苷酸序列:TATTGTTGTAGATACCTTTTTGTAAGGATTAAACAAC(SEQ ID NO:7)或其反向互补序列。在一些情况下,引导RNA包含核苷酸序列(N)nTATTGTTGTAGATACCTTTTTGTAAGGATTAAACAAC(SEQ ID NO:8)或其反向互补序列,其中N是任何核苷酸,并且n是15至30的整数,例如15至20、17至25、17至22、18至22、18至20、20至25或25至30)。

[0127] 图2Q

[0128] 在一些情况下,(主题组合物和/或方法的)Cas12L蛋白包含与图2Q中描绘并表示为"Cas12L_17_70959391"的Cas12L氨基酸序列具有20%或更多序列同一性(例如,30%或

更多、40％或更多、50％或更多、60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。例如,在一些情况下,Cas12L蛋白包含与图2Q中描绘的Cas12L氨基酸序列具有50％或更多序列同一性(例如,60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含与图2Q中描绘的Cas12L氨基酸序列具有80％或更多序列同一性(例如,85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含与图2Q中描绘的Cas12L氨基酸序列具有90％或更多序列同一性(例如,95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含具有图2Q中描绘的Cas12L氨基酸序列的氨基酸序列。在一些情况下,Cas12L蛋白包含具有图2Q中描绘的Cas12L蛋白序列的氨基酸序列,除了所述序列包含降低所述蛋白质的天然存在的催化活性的氨基酸取代(例如,1、2或3个氨基酸取代)。在一些情况下,Cas12L多肽具有700个氨基酸(aa)至800aa的长度,例如,700aa至725aa、725aa至746aa、746aa至775aa或775aa至800aa)。在一些情况下,Cas12L多肽具有746个氨基酸的长度。

[0129] 图2R

[0130] 在一些情况下,(主题组合物和/或方法的)Cas12L蛋白包含与图2R中描绘并表示为“Cas12L_18_71078086”的Cas12L氨基酸序列具有20％或更多序列同一性(例如,30％或更多、40％或更多、50％或更多、60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。例如,在一些情况下,Cas12L蛋白包含与图2R中描绘的Cas12L氨基酸序列具有50％或更多序列同一性(例如,60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含与图2R中描绘的Cas12L氨基酸序列具有80％或更多序列同一性(例如,85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含与图2R中描绘的Cas12L氨基酸序列具有90％或更多序列同一性(例如,95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含具有图2R中描绘的Cas12L氨基酸序列的氨基酸序列。在一些情况下,Cas12L蛋白包含具有图2R中描绘的Cas12L蛋白序列的氨基酸序列,除了所述序列包含降低所述蛋白质的天然存在的催化活性的氨基酸取代(例如,1、2或3个氨基酸取代)。在一些情况下,Cas12L多肽具有700个氨基酸(aa)至800aa的长度,例如,700aa至725aa、725aa至746aa、746aa至775aa或775aa至800aa)。在一些情况下,Cas12L多肽具有746个氨基酸的长度。在一些情况下,结合Cas12L多肽(例如,包含与图2R中描绘的Cas12L氨基酸序列具有20％或更多、30％或更多、40％或更多、50％或更多、60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％氨基酸序列同一性的氨基酸序列的Cas12L多肽)的引导RNA包含以下核苷酸序列:ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:1)或其反向互补序列。在一些情况下,引导RNA包含核苷酸序列(N)nATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:2)或其反向互补序列,其中N是任何核苷酸,并且n是15

至30的整数，例如15至20、17至25、17至22、18至22、18至20、20至25或25至30)。

[0131] 图2S

[0132] 在一些情况下，(主题组合物和/或方法的)Cas12L蛋白包含与图2S中描绘并表示为“Cas12L_19_71193509_部分”的Cas12L氨基酸序列具有20％或更多序列同一性(例如，30％或更多、40％或更多、50％或更多、60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的约680个氨基酸的连续片段。例如，在一些情况下，Cas12L蛋白包含与图2S中描绘的Cas12L氨基酸序列具有50％或更多序列同一性(例如，60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的约680个氨基酸的连续片段。在一些情况下，Cas12L蛋白包含与图2S中描绘的Cas12L氨基酸序列具有80％或更多序列同一性(例如，85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的约680个氨基酸的连续片段。在一些情况下，Cas12L蛋白包含与图2S中描绘的Cas12L氨基酸序列具有90％或更多序列同一性(例如，95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的约680个氨基酸的连续片段。在一些情况下，Cas12L蛋白包含具有图2S中描绘的Cas12L氨基酸序列的约680个氨基酸的连续片段。在一些情况下，Cas12L蛋白包含具有图2S中描绘的Cas12L蛋白序列的约680个氨基酸的连续片段，除了所述序列包含降低所述蛋白质的天然存在的催化活性的氨基酸取代(例如，1、2或3个氨基酸取代)。在一些情况下，Cas12L多肽具有700个氨基酸(aa)至800aa的长度，例如，700aa至725aa、725aa至750aa或750aa至800aa)。在一些情况下，Cas12L多肽具有约725个氨基酸至约775个氨基酸的长度。在一些情况下，结合Cas12L多肽(例如，包含与图2S中描绘的Cas12L氨基酸序列具有20％或更多、30％或更多、40％或更多、50％或更多、60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％氨基酸序列同一性的氨基酸序列的Cas12L多肽)的引导RNA包含以下核苷酸序列：ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:1)或其反向互补序列。在一些情况下，引导RNA包含核苷酸序列(N)nATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:2)或其反向互补序列，其中N是任何核苷酸，并且n是15至30的整数，例如15至20、17至25、17至22、18至22、18至20、20至25或25至30)。

[0133] 图2T

[0134] 在一些情况下，(主题组合物和/或方法的)Cas12L蛋白包含与图2T中描绘并表示为“Cas12L_20_71210958_部分”的Cas12L氨基酸序列具有20％或更多序列同一性(例如，30％或更多、40％或更多、50％或更多、60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的约516个氨基酸的连续片段。例如，在一些情况下，Cas12L蛋白包含与图2T中描绘的Cas12L氨基酸序列具有50％或更多序列同一性(例如，60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的约516个氨基酸的连续片段。在一些情况下，Cas12L蛋白包含与图2T中描绘的Cas12L氨基酸序列具有80％或更多序列同一性(例如，85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的约516个氨基酸的连续片段。在一些情况下，Cas12L蛋白包含与图2T中描绘的Cas12L氨基酸序列具有90％或更多序列同一性(例

如,95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的约516个氨基酸的连续片段。在一些情况下,Cas12L蛋白包含具有图2T中描绘的Cas12L氨基酸序列的约516个氨基酸的连续片段。在一些情况下,Cas12L蛋白包含具有图2T中描绘的Cas12L蛋白序列的约516个氨基酸的连续片段,除了所述序列包含降低所述蛋白质的天然存在的催化活性的氨基酸取代(例如,1、2或3个氨基酸取代)。在一些情况下,Cas12L多肽具有700个氨基酸(aa)至800aa的长度,例如,700aa至725aa、725aa至750aa或750aa至800aa)。在一些情况下,Cas12L多肽具有约725个氨基酸至约775个氨基酸的长度。在一些情况下,结合Cas12L多肽(例如,包含与图2T中描绘的Cas12L氨基酸序列具有20%或更多、30%或更多、40%或更多、50%或更多、60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%氨基酸序列同一性的氨基酸序列的Cas12L多肽)的引导RNA包含以下核苷酸序列:AATGTTGTAGATGCCTTTTTATAAGGATTAAACAACTTG(SEQ ID NO:9)或其反向互补序列。在一些情况下,引导RNA包含核苷酸序列(N)nAATGTTGTAGATGCCTTTTTATAAGGATTAAACAACTTG(SEQ ID NO:10)或其反向互补序列,其中N是任何核苷酸,并且n是15至30的整数,例如15至20、17至25、17至22、18至22、18至20、20至25或25至30)。

[0135] 图2U

[0136] 在一些情况下,(主题组合物和/或方法的)Cas12L蛋白包含与图2U中描绘并表示为“Cas12L_21_71317321_部分”的Cas12L氨基酸序列具有20%或更多序列同一性(例如,30%或更多、40%或更多、50%或更多、60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的约585个氨基酸的连续片段。例如,在一些情况下,Cas12L蛋白包含与图2U中描绘的Cas12L氨基酸序列具有50%或更多序列同一性(例如,60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的约585个氨基酸的连续片段。在一些情况下,Cas12L蛋白包含与图2U中描绘的Cas12L氨基酸序列具有80%或更多序列同一性(例如,85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的约585个氨基酸的连续片段。在一些情况下,Cas12L蛋白包含与图2U中描绘的Cas12L氨基酸序列具有90%或更多序列同一性(例如,95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的约585个氨基酸的连续片段。在一些情况下,Cas12L蛋白包含具有图2U中描绘的Cas12L氨基酸序列的约585个氨基酸的连续片段。在一些情况下,Cas12L蛋白包含具有图2U中描绘的Cas12L蛋白序列的约585个氨基酸的连续片段,除了所述序列包含降低所述蛋白质的天然存在的催化活性的氨基酸取代(例如,1、2或3个氨基酸取代)。在一些情况下,Cas12L多肽具有700个氨基酸(aa)至800aa的长度,例如,700aa至725aa、725aa至750aa或750aa至800aa)。在一些情况下,Cas12L多肽具有约725个氨基酸至约775个氨基酸的长度。在一些情况下,结合Cas12L多肽(例如,包含与图2U中描绘的Cas12L氨基酸序列具有20%或更多、30%或更多、40%或更多、50%或更多、60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%氨基酸序列同一性的氨基酸序列的Cas12L多肽)的引导RNA包含以下核苷酸序列:ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:1)或其反向互补序列。在一些情况下,引导RNA包含核苷酸序列(N)nATTGTTGT

AACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:2)或其反向互补序列,其中N是任何核苷酸,并且n是15至30的整数,例如15至20、17至25、17至22、18至22、18至20、20至25或25至30)。

[0137] 图2V

[0138] 在一些情况下,(主题组合物和/或方法的)Cas12L蛋白包含与图2V中描绘并表示为"Cas12L_22_71456687"的Cas12L氨基酸序列具有20%或更多序列同一性(例如,30%或更多、40%或更多、50%或更多、60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的氨基酸序列。例如,在一些情况下,Cas12L蛋白包含与图2V中描绘的Cas12L氨基酸序列具有50%或更多序列同一性(例如,60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含与图2V中描绘的Cas12L氨基酸序列具有80%或更多序列同一性(例如,85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含与图2V中描绘的Cas12L氨基酸序列具有90%或更多序列同一性(例如,95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含具有图2V中描绘的Cas12L氨基酸序列的氨基酸序列。在一些情况下,Cas12L蛋白包含具有图2V中描绘的Cas12L蛋白序列的氨基酸序列,除了所述序列包含降低所述蛋白质的天然存在的催化活性的氨基酸取代(例如,1、2或3个氨基酸取代)。在一些情况下,Cas12L多肽具有700个氨基酸(aa)至800aa的长度,例如,700aa至725aa、725aa至746aa、746aa至775aa或775aa至800aa)。在一些情况下,Cas12L多肽具有746个氨基酸的长度。在一些情况下,结合Cas12L多肽(例如,包含与图2V中描绘的Cas12L氨基酸序列具有20%或更多、30%或更多、40%或更多、50%或更多、60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%氨基酸序列同一性的氨基酸序列的Cas12L多肽)的引导RNA包含以下核苷酸序列:ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:1)或其反向互补序列。在一些情况下,引导RNA包含核苷酸序列(N)nATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:2)或其反向互补序列,其中N是任何核苷酸,并且n是15至30的整数,例如15至20、17至25、17至22、18至22、18至20、20至25或25至30)。

[0139] 图2W

[0140] 在一些情况下,(主题组合物和/或方法的)Cas12L蛋白包含与图2W中描绘并表示为"Cas12L_23_71708971"的Cas12L氨基酸序列具有20%或更多序列同一性(例如,30%或更多、40%或更多、50%或更多、60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的氨基酸序列。例如,在一些情况下,Cas12L蛋白包含与图2W中描绘的Cas12L氨基酸序列具有50%或更多序列同一性(例如,60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含与图2W中描绘的Cas12L氨基酸序列具有80%或更多序列同一性(例如,85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含与图2W中描绘的Cas12L氨基酸序列具有90%或更多序列同一性(例如,95%或更多、97%或更多、98%或更多、99%

或更多或100％序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含具有图2W中描绘的Cas12L氨基酸序列的氨基酸序列。在一些情况下,Cas12L蛋白包含具有图2W中描绘的Cas12L蛋白序列的氨基酸序列,除了所述序列包含降低所述蛋白质的天然存在的催化活性的氨基酸取代(例如,1、2或3个氨基酸取代)。在一些情况下,Cas12L多肽具有500个氨基酸(aa)至800aa的长度,例如,500aa至550aa、550aa至585aa、585aa至600aa、600aa至650aa、650aa至700aa、700aa至725aa、725aa至750aa、750aa至775aa或775aa至800aa)。在一些情况下,Cas12L多肽具有585个氨基酸的长度。在一些情况下,结合Cas12L多肽(例如,包含与图2W中描绘的Cas12L氨基酸序列具有20％或更多、30％或更多、40％或更多、50％或更多、60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％氨基酸序列同一性的氨基酸序列的Cas12L多肽)的引导RNA包含以下核苷酸序列:ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:1)或其反向互补序列。在一些情况下,引导RNA包含核苷酸序列(N)nATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:2)或其反向互补序列,其中N是任何核苷酸,并且n是15至30的整数,例如15至20、17至25、17至22、18至22、18至20、20至25或25至30)。

[0141] 图2X

[0142] 在一些情况下,(主题组合物和/或方法的)Cas12L蛋白包含与图2X中描绘并表示为“Cas12L_24_46035167_部分”的Cas12L氨基酸序列具有20％或更多序列同一性(例如,30％或更多、40％或更多、50％或更多、60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。例如,在一些情况下,Cas12L蛋白包含与图2X中描绘的Cas12L氨基酸序列具有50％或更多序列同一性(例如,60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含与图2X中描绘的Cas12L氨基酸序列具有80％或更多序列同一性(例如,85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含与图2X中描绘的Cas12L氨基酸序列具有90％或更多序列同一性(例如,95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含具有图2X中描绘的Cas12L氨基酸序列的氨基酸序列。在一些情况下,Cas12L蛋白包含具有图2X中描绘的Cas12L蛋白序列的氨基酸序列,除了所述序列包含降低所述蛋白质的天然存在的催化活性的氨基酸取代(例如,1、2或3个氨基酸取代)。在一些情况下,Cas12L多肽具有700个氨基酸(aa)至800aa的长度,例如,700aa至725aa、725aa至758aa、758aa至775aa或775aa至800aa)。在一些情况下,Cas12L多肽具有758个氨基酸的长度。在一些情况下,结合Cas12L多肽(例如,包含与图2X中描绘的Cas12L氨基酸序列具有20％或更多、30％或更多、40％或更多、50％或更多、60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％氨基酸序列同一性的氨基酸序列的Cas12L多肽)的引导RNA包含以下核苷酸序列:ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:1)或其反向互补序列。在一些情况下,引导RNA包含核苷酸序列(N)nATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:2)或其反向互补序列,其中N是任何核苷酸,并且n是15至30的整数,例如15至20、17至25、17至22、18至22、18至20、20至25或25至30)。

[0143] 图2Y

[0144] 在一些情况下，(主题组合物和/或方法的)Cas12L蛋白包含与图2Y中描绘并表示为“Cas12L_25_46784254_部分”的Cas12L氨基酸序列具有20％或更多序列同一性(例如，30％或更多、40％或更多、50％或更多、60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的约596个氨基酸的连续片段。例如，在一些情况下，Cas12L蛋白包含与图2Y中描绘的Cas12L氨基酸序列具有50％或更多序列同一性(例如，60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的约596个氨基酸的连续片段。在一些情况下，Cas12L蛋白包含与图2Y中描绘的Cas12L氨基酸序列具有80％或更多序列同一性(例如，85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的约596个氨基酸的连续片段。在一些情况下，Cas12L蛋白包含与图2Y中描绘的Cas12L氨基酸序列具有90％或更多序列同一性(例如，95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的约596个氨基酸的连续片段。在一些情况下，Cas12L蛋白包含具有图2Y中描绘的Cas12L氨基酸序列的约596个氨基酸的连续片段。在一些情况下，Cas12L蛋白包含具有图2Y中描绘的Cas12L蛋白序列的约596个氨基酸的连续片段，除了所述序列包含降低所述蛋白质的天然存在的催化活性的氨基酸取代(例如，1、2或3个氨基酸取代)。在一些情况下，Cas12L多肽具有700个氨基酸(aa)至800aa的长度，例如，700aa至725aa、725aa至750aa或750aa至800aa)。在一些情况下，Cas12L多肽具有约725个氨基酸至约775个氨基酸的长度。在一些情况下，结合Cas12L多肽(例如，包含与图2Y中描绘的Cas12L氨基酸序列具有20％或更多、30％或更多、40％或更多、50％或更多、60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％氨基酸序列同一性的氨基酸序列的Cas12L多肽)的引导RNA包含以下核苷酸序列：ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:1)或其反向互补序列。在一些情况下，引导RNA包含核苷酸序列(N)nATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:2)或其反向互补序列，其中N是任何核苷酸，并且n是15至30的整数，例如15至20、17至25、17至22、18至22、18至20、20至25或25至30)。

[0145] 图2Z

[0146] 在一些情况下，(主题组合物和/或方法的)Cas12L蛋白包含与图2Z中描绘并表示为“Cas12L_26_46464451”的Cas12L氨基酸序列具有20％或更多序列同一性(例如，30％或更多、40％或更多、50％或更多、60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。例如，在一些情况下，Cas12L蛋白包含与图2Z中描绘的Cas12L氨基酸序列具有50％或更多序列同一性(例如，60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。在一些情况下，Cas12L蛋白包含与图2Z中描绘的Cas12L氨基酸序列具有80％或更多序列同一性(例如，85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。在一些情况下，Cas12L蛋白包含与图2Z中描绘的Cas12L氨基酸序列具有90％或更多序列同一性(例如，95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。在一些情况下，Cas12L蛋白包含具有图2Z中描

绘的Cas12L氨基酸序列的氨基酸序列。在一些情况下,Cas12L蛋白包含具有图2Z中描绘的Cas12L蛋白序列的氨基酸序列,除了所述序列包含降低所述蛋白质的天然存在的催化活性的氨基酸取代(例如,1、2或3个氨基酸取代)。在一些情况下,Cas12L多肽具有625个氨基酸(aa)至800aa的长度,例如,625aa至640aa、640aa至650aa、650aa至700aa、700aa至725aa、725aa至750aa、750aa至775aa或775aa至800aa)。在一些情况下,Cas12L多肽具有640个氨基酸的长度。在一些情况下,结合Cas12L多肽(例如,包含与图2Z中描绘的Cas12L氨基酸序列具有20%或更多、30%或更多、40%或更多、50%或更多、60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%氨基酸序列同一性的氨基酸序列的Cas12L多肽)的引导RNA包含以下核苷酸序列:ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:1)或其反向互补序列。在一些情况下,引导RNA包含核苷酸序列(N)nATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:2)或其反向互补序列,其中N是任何核苷酸,并且n是15至30的整数,例如15至20、17至25、17至22、18至22、18至20、20至25或25至30)。

[0147] 图2AA

[0148] 在一些情况下,(主题组合物和/或方法的)Cas12L蛋白包含与图2AA中描绘并表示为"Cas12L_27_254489164"的Cas12L氨基酸序列具有20%或更多序列同一性(例如,30%或更多、40%或更多、50%或更多、60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的氨基酸序列。例如,在一些情况下,Cas12L蛋白包含与图2AA中描绘的Cas12L氨基酸序列具有50%或更多序列同一性(例如,60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含与图2AA中描绘的Cas12L氨基酸序列具有80%或更多序列同一性(例如,85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含与图2AA中描绘的Cas12L氨基酸序列具有90%或更多序列同一性(例如,95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含具有图2AA中描绘的Cas12L氨基酸序列的氨基酸序列。在一些情况下,Cas12L蛋白包含具有图2AA中描绘的Cas12L蛋白序列的氨基酸序列,除了所述序列包含降低所述蛋白质的天然存在的催化活性的氨基酸取代(例如,1、2或3个氨基酸取代)。在一些情况下,Cas12L多肽具有700个氨基酸(aa)至800aa的长度,例如,700aa至725aa、725aa至746aa、746aa至775aa或775aa至800aa)。在一些情况下,Cas12L多肽具有746个氨基酸的长度。在一些情况下,结合Cas12L多肽(例如,包含与图2AA中描绘的Cas12L氨基酸序列具有20%或更多、30%或更多、40%或更多、50%或更多、60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%氨基酸序列同一性的氨基酸序列的Cas12L多肽)的引导RNA包含以下核苷酸序列:ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:1)或其反向互补序列。在一些情况下,引导RNA包含核苷酸序列(N)nATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:2)或其反向互补序列,其中N是任何核苷酸,并且n是15至30的整数,例如15至20、17至25、17至22、18至22、18至20、20至25或25至30)。

[0149] 图2BB

[0150] 在一些情况下,(主题组合物和/或方法的)Cas12L蛋白包含与图2BB中描绘并表示为“Cas12L_28_255238293”的Cas12L氨基酸序列具有20%或更多序列同一性(例如,30%或更多、40%或更多、50%或更多、60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的氨基酸序列。例如,在一些情况下,Cas12L蛋白包含与图2BB中描绘的Cas12L氨基酸序列具有50%或更多序列同一性(例如,60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含与图2BB中描绘的Cas12L氨基酸序列具有80%或更多序列同一性(例如,85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含与图2BB中描绘的Cas12L氨基酸序列具有90%或更多序列同一性(例如,95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含具有图2BB中描绘的Cas12L氨基酸序列的氨基酸序列。在一些情况下,Cas12L蛋白包含具有图2BB中描绘的Cas12L蛋白序列的氨基酸序列,除了所述序列包含降低所述蛋白质的天然存在的催化活性的氨基酸取代(例如,1、2或3个氨基酸取代)。在一些情况下,Cas12L多肽具有700个氨基酸(aa)至800aa的长度,例如,700aa至725aa、725aa至759aa、759aa至775aa或775aa至800aa)。在一些情况下,Cas12L多肽具有759个氨基酸的长度。在一些情况下,结合Cas12L多肽(例如,包含与图2BB中描绘的Cas12L氨基酸序列具有20%或更多、30%或更多、40%或更多、50%或更多、60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%氨基酸序列同一性的氨基酸序列的Cas12L多肽)的引导RNA包含以下核苷酸序列:ATTGTTGAAATAGTACTTTTATAGTCTATATACAAC(SEQ ID NO:11)或其反向互补序列。在一些情况下,引导RNA包含核苷酸序列(N)nATTGTTGAAATAGTACTTTTATAGTCTATATACAAC(SEQ ID NO:12)或其反向互补序列,其中N是任何核苷酸,并且n是15至30的整数,例如15至20、17至25、17至22、18至22、18至20、20至25或25至30)。

[0151] 图2CC

[0152] 在一些情况下,(主题组合物和/或方法的)Cas12L蛋白包含与图2CC中描绘并表示为“Cas12L_29_72167294”的Cas12L氨基酸序列具有20%或更多序列同一性(例如,30%或更多、40%或更多、50%或更多、60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的氨基酸序列。例如,在一些情况下,Cas12L蛋白包含与图2CC中描绘的Cas12L氨基酸序列具有50%或更多序列同一性(例如,60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含与图2CC中描绘的Cas12L氨基酸序列具有80%或更多序列同一性(例如,85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含与图2CC中描绘的Cas12L氨基酸序列具有90%或更多序列同一性(例如,95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含具有图2CC中描绘的Cas12L氨基酸序列的氨基酸序列。在一些情况下,Cas12L蛋白包含具有图2CC中描绘的Cas12L蛋白序列的氨基酸序列,除了所述序列包含降低所述蛋白质的天然存在的催化活性

的氨基酸取代(例如,1、2或3个氨基酸取代)。在一些情况下,Cas12L多肽具有700个氨基酸(aa)至800aa的长度,例如,700aa至725aa、725aa至746aa、746aa至775aa或775aa至800aa)。在一些情况下,Cas12L多肽具有746个氨基酸的长度。在一些情况下,结合Cas12L多肽(例如,包含与图2CC中描绘的Cas12L氨基酸序列具有20%或更多、30%或更多、40%或更多、50%或更多、60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%氨基酸序列同一性的氨基酸序列的Cas12L多肽)的引导RNA包含以下核苷酸序列:TATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:13)或其反向互补序列。在一些情况下,引导RNA包含核苷酸序列(N)nTATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:14)或其反向互补序列,其中N是任何核苷酸,并且n是15至30的整数,例如15至20、17至25、17至22、18至22、18至20、20至25或25至30)。

[0153] 图2DD

[0154] 在一些情况下,(主题组合物和/或方法的)Cas12L蛋白包含与图2DD中描绘并表示为"Cas12L_30_72369269"的Cas12L氨基酸序列具有20%或更多序列同一性(例如,30%或更多、40%或更多、50%或更多、60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的氨基酸序列。例如,在一些情况下,Cas12L蛋白包含与图2DD中描绘的Cas12L氨基酸序列具有50%或更多序列同一性(例如,60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含与图2DD中描绘的Cas12L氨基酸序列具有80%或更多序列同一性(例如,85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含与图2DD中描绘的Cas12L氨基酸序列具有90%或更多序列同一性(例如,95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含具有图2DD中描绘的Cas12L氨基酸序列的氨基酸序列。在一些情况下,Cas12L蛋白包含具有图2DD中描绘的Cas12L蛋白序列的氨基酸序列,除了所述序列包含降低所述蛋白质的天然存在的催化活性的氨基酸取代(例如,1、2或3个氨基酸取代)。在一些情况下,Cas12L多肽具有700个氨基酸(aa)至800aa的长度,例如,700aa至725aa、725aa至746aa、746aa至775aa或775aa至800aa)。在一些情况下,Cas12L多肽具有746个氨基酸的长度。在一些情况下,结合Cas12L多肽(例如,包含与图2DD中描绘的Cas12L氨基酸序列具有20%或更多、30%或更多、40%或更多、50%或更多、60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%氨基酸序列同一性的氨基酸序列的Cas12L多肽)的引导RNA包含以下核苷酸序列:ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:1)或其反向互补序列。在一些情况下,引导RNA包含核苷酸序列(N)nATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:2)或其反向互补序列,其中N是任何核苷酸,并且n是15至30的整数,例如15至20、17至25、17至22、18至22、18至20、20至25或25至30)。

[0155] 图2EE

[0156] 在一些情况下,(主题组合物和/或方法的)Cas12L蛋白包含与图2EE中描绘并表示为"Cas12L_31_72503976"的Cas12L氨基酸序列具有20%或更多序列同一性(例如,30%或更多、40%或更多、50%或更多、60%或更多、70%或更多、80%或更多、85%或更多、90%或

更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。例如,在一些情况下,Cas12L蛋白包含与图2EE中描绘的Cas12L氨基酸序列具有50％或更多序列同一性(例如,60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含与图2EE中描绘的Cas12L氨基酸序列具有80％或更多序列同一性(例如,85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含与图2EE中描绘的Cas12L氨基酸序列具有90％或更多序列同一性(例如,95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含具有图2EE中描绘的Cas12L氨基酸序列的氨基酸序列。在一些情况下,Cas12L蛋白包含具有图2EE中描绘的Cas12L蛋白序列的氨基酸序列,除了所述序列包含降低所述蛋白质的天然存在的催化活性的氨基酸取代(例如,1、2或3个氨基酸取代)。在一些情况下,Cas12L多肽具有700个氨基酸(aa)至800aa的长度,例如,700aa至725aa、725aa至746aa、746aa至775aa或775aa至800aa)。在一些情况下,Cas12L多肽具有746个氨基酸的长度。在一些情况下,结合Cas12L多肽(例如,包含与图2EE中描绘的Cas12L氨基酸序列具有20％或更多、30％或更多、40％或更多、50％或更多、60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％氨基酸序列同一性的氨基酸序列的Cas12L多肽)的引导RNA包含以下核苷酸序列:ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:1)或其反向互补序列。在一些情况下,引导RNA包含核苷酸序列(N)nATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:2)或其反向互补序列,其中N是任何核苷酸,并且n是15至30的整数,例如15至20、17至25、17至22、18至22、18至20、20至25或25至30)。

[0157] 图2FF

[0158] 在一些情况下,(主题组合物和/或方法的)Cas12L蛋白包含与图2FF中描绘并表示为“Cas12L_32_72547654_部分”的Cas12L氨基酸序列具有20％或更多序列同一性(例如,30％或更多、40％或更多、50％或更多、60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的约47个氨基酸的连续片段。例如,在一些情况下,Cas12L蛋白包含与图2FF中描绘的Cas12L氨基酸序列具有50％或更多序列同一性(例如,60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的约47个氨基酸的连续片段。在一些情况下,Cas12L蛋白包含与图2FF中描绘的Cas12L氨基酸序列具有80％或更多序列同一性(例如,85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的约47个氨基酸的连续片段。在一些情况下,Cas12L蛋白包含与图2FF中描绘的Cas12L氨基酸序列具有90％或更多序列同一性(例如,95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的约47个氨基酸的连续片段。在一些情况下,Cas12L蛋白包含具有图2FF中描绘的Cas12L氨基酸序列的约47个氨基酸的连续片段。在一些情况下,Cas12L蛋白包含具有图2FF中描绘的Cas12L蛋白序列的约47个氨基酸的连续片段,除了所述序列包含降低所述蛋白质的天然存在的催化活性的氨基酸取代(例如,1、2或3个氨基酸取代)。在一些情况下,Cas12L多肽具有700个氨基酸(aa)至800aa的长度,例如,700aa至725aa、725aa至750aa、750aa至775aa或775aa至800aa)。

在一些情况下,Cas12L多肽具有725个氨基酸至775个氨基酸的长度。在一些情况下,结合Cas12L多肽(例如,包含与图2FF中描绘的Cas12L氨基酸序列具有20%或更多、30%或更多、40%或更多、50%或更多、60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%氨基酸序列同一性的氨基酸序列的Cas12L多肽)的引导RNA包含以下核苷酸序列:ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:1)或其反向互补序列。在一些情况下,引导RNA包含核苷酸序列(N)nATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:2)或其反向互补序列,其中N是任何核苷酸,并且n是15至30的整数,例如15至20、17至25、17至22、18至22、18至20、20至25或25至30)。

[0159] 图2GG

[0160] 在一些情况下,(主题组合物和/或方法的)Cas12L蛋白包含与图2GG中描绘并表示为"Cas12L_33_72907394"的Cas12L氨基酸序列具有20%或更多序列同一性(例如,30%或更多、40%或更多、50%或更多、60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的氨基酸序列。例如,在一些情况下,Cas12L蛋白包含与图2GG中描绘的Cas12L氨基酸序列具有50%或更多序列同一性(例如,60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含与图2GG中描绘的Cas12L氨基酸序列具有80%或更多序列同一性(例如,85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含与图2GG中描绘的Cas12L氨基酸序列具有90%或更多序列同一性(例如,95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含具有图2GG中描绘的Cas12L氨基酸序列的氨基酸序列。在一些情况下,Cas12L蛋白包含具有图2GG中描绘的Cas12L蛋白序列的氨基酸序列,除了所述序列包含降低所述蛋白质的天然存在的催化活性的氨基酸取代(例如,1、2或3个氨基酸取代)。在一些情况下,Cas12L多肽具有700个氨基酸(aa)至800aa的长度,例如,700aa至725aa、725aa至746aa、746aa至775aa或775aa至800aa)。在一些情况下,Cas12L多肽具有746个氨基酸的长度。在一些情况下,结合Cas12L多肽(例如,包含与图2GG中描绘的Cas12L氨基酸序列具有20%或更多、30%或更多、40%或更多、50%或更多、60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%氨基酸序列同一性的氨基酸序列的Cas12L多肽)的引导RNA包含以下核苷酸序列:ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:1)或其反向互补序列。在一些情况下,引导RNA包含核苷酸序列(N)nATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:2)或其反向互补序列,其中N是任何核苷酸,并且n是15至30的整数,例如15至20、17至25、17至22、18至22、18至20、20至25或25至30)。

[0161] 图2HH

[0162] 在一些情况下,(主题组合物和/或方法的)Cas12L蛋白包含与图2HH中描绘并表示为"Cas12L_34_73124743_部分"的Cas12L氨基酸序列具有20%或更多序列同一性(例如,30%或更多、40%或更多、50%或更多、60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的约

245个氨基酸的连续片段。例如,在一些情况下,Cas12L蛋白包含与图2HH中描绘的Cas12L氨基酸序列具有50%或更多序列同一性(例如,60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的约245个氨基酸的连续片段。在一些情况下,Cas12L蛋白包含与图2HH中描绘的Cas12L氨基酸序列具有80%或更多序列同一性(例如,85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的约245个氨基酸的连续片段。在一些情况下,Cas12L蛋白包含与图2HH中描绘的Cas12L氨基酸序列具有90%或更多序列同一性(例如,95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的约245个氨基酸的连续片段。在一些情况下,Cas12L蛋白包含具有图2HH中描绘的Cas12L氨基酸序列的约245个氨基酸的连续片段。在一些情况下,Cas12L蛋白包含具有图2HH中描绘的Cas12L蛋白序列的约245个氨基酸的连续片段,除了所述序列包含降低所述蛋白质的天然存在的催化活性的氨基酸取代(例如,1、2或3个氨基酸取代)。在一些情况下,Cas12L多肽具有700个氨基酸(aa)至800aa的长度,例如,700aa至725aa、725aa至750aa、750aa至775aa或775aa至800aa)。在一些情况下,Cas12L多肽具有725个氨基酸至775个氨基酸的长度。在一些情况下,结合Cas12L多肽(例如,包含与图2HH中描绘的Cas12L氨基酸序列具有20%或更多、30%或更多、40%或更多、50%或更多、60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%氨基酸序列同一性的氨基酸序列的Cas12L多肽)的引导RNA包含以下核苷酸序列:ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:1)或其反向互补序列。在一些情况下,引导RNA包含核苷酸序列(N)nATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:2)或其反向互补序列,其中N是任何核苷酸,并且n是15至30的整数,例如15至20、17至25、17至22、18至22、18至20、20至25或25至30)。

[0163] 图2II

[0164] 在一些情况下,(主题组合物和/或方法的)Cas12L蛋白包含与图2II中描绘并表示为“Cas12L_35_73503649_部分”的Cas12L氨基酸序列具有20%或更多序列同一性(例如,30%或更多、40%或更多、50%或更多、60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的约178个氨基酸的连续片段。例如,在一些情况下,Cas12L蛋白包含与图2II中描绘的Cas12L氨基酸序列具有50%或更多序列同一性(例如,60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的约178个氨基酸的连续片段。在一些情况下,Cas12L蛋白包含与图2II中描绘的Cas12L氨基酸序列具有80%或更多序列同一性(例如,85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的约178个氨基酸的连续片段。在一些情况下,Cas12L蛋白包含与图2II中描绘的Cas12L氨基酸序列具有90%或更多序列同一性(例如,95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的约178个氨基酸的连续片段。在一些情况下,Cas12L蛋白包含具有图2II中描绘的Cas12L氨基酸序列的约178个氨基酸的连续片段。在一些情况下,Cas12L蛋白包含具有图2II中描绘的Cas12L蛋白序列的约178个氨基酸的连续片段,除了所述序列包含降低所述蛋白质的天然存在的催化活性的氨基酸取代(例如,1、2或3个氨基酸取代)。在一些情况下,Cas12L多肽具有700

个氨基酸(aa)至800aa的长度,例如,700aa至725aa、725aa至750aa、750aa至775aa或775aa至800aa)。在一些情况下,Cas12L多肽具有725个氨基酸至775个氨基酸的长度。在一些情况下,结合Cas12L多肽(例如,包含与图2II中描绘的Cas12L氨基酸序列具有20%或更多、30%或更多、40%或更多、50%或更多、60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%氨基酸序列同一性的氨基酸序列的Cas12L多肽)的引导RNA包含以下核苷酸序列:ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:1)或其反向互补序列。在一些情况下,引导RNA包含核苷酸序列(N)nATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:2)或其反向互补序列,其中N是任何核苷酸,并且n是15至30的整数,例如15至20、17至25、17至22、18至22、18至20、20至25或25至30)。

[0165] 图2JJ

[0166] 在一些情况下,(主题组合物和/或方法的)Cas12L蛋白包含与图2JJ中描绘并表示为"Cas12L_36_73503649_部分"的Cas12L氨基酸序列具有20%或更多序列同一性(例如,30%或更多、40%或更多、50%或更多、60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的约85个氨基酸的连续片段。例如,在一些情况下,Cas12L蛋白包含与图2JJ中描绘的Cas12L氨基酸序列具有50%或更多序列同一性(例如,60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的约85个氨基酸的连续片段。在一些情况下,Cas12L蛋白包含与图2JJ中描绘的Cas12L氨基酸序列具有80%或更多序列同一性(例如,85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的约85个氨基酸的连续片段。在一些情况下,Cas12L蛋白包含与图2JJ中描绘的Cas12L氨基酸序列具有90%或更多序列同一性(例如,95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的约85个氨基酸的连续片段。在一些情况下,Cas12L蛋白包含具有图2JJ中描绘的Cas12L氨基酸序列的约85个氨基酸的连续片段。在一些情况下,Cas12L蛋白包含具有图2JJ中描绘的Cas12L蛋白序列的约85个氨基酸的连续片段,除了所述序列包含降低所述蛋白质的天然存在的催化活性的氨基酸取代(例如,1、2或3个氨基酸取代)。在一些情况下,Cas12L多肽具有700个氨基酸(aa)至800aa的长度,例如,700aa至725aa、725aa至750aa、750aa至775aa或775aa至800aa)。在一些情况下,Cas12L多肽具有725个氨基酸至775个氨基酸的长度。

[0167] 图2KK

[0168] 在一些情况下,(主题组合物和/或方法的)Cas12L蛋白包含与图2KK中描绘并表示为"Cas12L_37_73472625"的Cas12L氨基酸序列具有20%或更多序列同一性(例如,30%或更多、40%或更多、50%或更多、60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的氨基酸序列。例如,在一些情况下,Cas12L蛋白包含与图2KK中描绘的Cas12L氨基酸序列具有50%或更多序列同一性(例如,60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含与图2KK中描绘的Cas12L氨基酸序列具有80%或更多序列同一性(例如,85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或

100％序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含与图2KK中描绘的Cas12L氨基酸序列具有90％或更多序列同一性(例如,95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含具有图2KK中描绘的Cas12L氨基酸序列的氨基酸序列。在一些情况下,Cas12L蛋白包含具有图2KK中描绘的Cas12L蛋白序列的氨基酸序列,除了所述序列包含降低所述蛋白质的天然存在的催化活性的氨基酸取代(例如,1、2或3个氨基酸取代)。在一些情况下,Cas12L多肽具有700个氨基酸(aa)至800aa的长度,例如700aa至725aa、725aa至750aa、750aa至767aa、767aa至775aa或775aa至800aa)。在一些情况下,Cas12L多肽具有767个氨基酸的长度。在一些情况下,结合Cas12L多肽(例如,包含与图2KK中描绘的Cas12L氨基酸序列具有20％或更多、30％或更多、40％或更多、50％或更多、60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％氨基酸序列同一性的氨基酸序列的Cas12L多肽)的引导RNA包含以下核苷酸序列:ATTGTTGTAACATCTATTTTGTAAGGTGTAAACAAC(SEQ ID NO:15)或其反向互补序列。在一些情况下,引导RNA包含核苷酸序列(N)nATTGTTGTAACATCTATTTTGTAAGGTGTAAACAAC(SEQ ID NO:16)或其反向互补序列,其中N是任何核苷酸,并且n是15至30的整数,例如15至20、17至25、17至22、18至22、18至20、20至25或25至30)。

[0169] 图2LL

[0170] 在一些情况下,(主题组合物和/或方法的)Cas12L蛋白包含与图2LL中描绘并表示为“Cas12L_38_73764039_部分”的Cas12L氨基酸序列具有20％或更多序列同一性(例如,30％或更多、40％或更多、50％或更多、60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的约652个氨基酸的连续片段。例如,在一些情况下,Cas12L蛋白包含与图2LL中描绘的Cas12L氨基酸序列具有50％或更多序列同一性(例如,60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的约652个氨基酸的连续片段。在一些情况下,Cas12L蛋白包含与图2LL中描绘的Cas12L氨基酸序列具有80％或更多序列同一性(例如,85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的约652个氨基酸的连续片段。在一些情况下,Cas12L蛋白包含与图2LL中描绘的Cas12L氨基酸序列具有90％或更多序列同一性(例如,95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的约652个氨基酸的连续片段。在一些情况下,Cas12L蛋白包含具有图2LL中描绘的Cas12L氨基酸序列的约652个氨基酸的连续片段。在一些情况下,Cas12L蛋白包含具有图2LL中描绘的Cas12L蛋白序列的约652个氨基酸的连续片段,除了所述序列包含降低所述蛋白质的天然存在的催化活性的氨基酸取代(例如,1、2或3个氨基酸取代)。在一些情况下,Cas12L多肽具有700个氨基酸(aa)至800aa的长度,例如,700aa至725aa、725aa至750aa、750aa至775aa或775aa至800aa)。在一些情况下,Cas12L多肽具有725个氨基酸至775个氨基酸的长度。在一些情况下,结合Cas12L多肽(例如,包含与图2LL中描绘的Cas12L氨基酸序列具有20％或更多、30％或更多、40％或更多、50％或更多、60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％氨基酸序列同一性的氨基酸序列的Cas12L多肽)的引导RNA包含以下核苷酸序列:ATTGTTGTAACTCTTATTTTGTATGGA

GTAAACAAC(SEQ ID NO:1)或其反向互补序列。在一些情况下,引导RNA包含核苷酸序列(N)nATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:2)或其反向互补序列,其中N是任何核苷酸,并且n是15至30的整数,例如15至20、17至25、17至22、18至22、18至20、20至25或25至30)。

[0171] 图2MM

[0172] 在一些情况下,(主题组合物和/或方法的)Cas12L蛋白包含与图2MM中描绘并表示为"Cas12L_40_74037305"的Cas12L氨基酸序列具有20%或更多序列同一性(例如,30%或更多、40%或更多、50%或更多、60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的氨基酸序列。例如,在一些情况下,Cas12L蛋白包含与图2MM中描绘的Cas12L氨基酸序列具有50%或更多序列同一性(例如,60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含与图2MM中描绘的Cas12L氨基酸序列具有80%或更多序列同一性(例如,85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含与图2MM中描绘的Cas12L氨基酸序列具有90%或更多序列同一性(例如,95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含具有图2MM中描绘的Cas12L氨基酸序列的氨基酸序列。在一些情况下,Cas12L蛋白包含具有图2MM中描绘的Cas12L蛋白序列的氨基酸序列,除了所述序列包含降低所述蛋白质的天然存在的催化活性的氨基酸取代(例如,1、2或3个氨基酸取代)。在一些情况下,Cas12L多肽具有700个氨基酸(aa)至800aa的长度,例如,700aa至725aa、725aa至750aa、750aa至782aa或782aa至800aa)。在一些情况下,Cas12L多肽具有782个氨基酸的长度。在一些情况下,结合Cas12L多肽(例如,包含与图2MM中描绘的Cas12L氨基酸序列具有20%或更多、30%或更多、40%或更多、50%或更多、60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%氨基酸序列同一性的氨基酸序列的Cas12L多肽)的引导RNA包含以下核苷酸序列:ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:1)或其反向互补序列。在一些情况下,引导RNA包含核苷酸序列(N)nATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:2)或其反向互补序列,其中N是任何核苷酸,并且n是15至30的整数,例如15至20、17至25、17至22、18至22、18至20、20至25或25至30)。

[0173] 图2NN

[0174] 在一些情况下,(主题组合物和/或方法的)Cas12L蛋白包含与图2NN中描绘并表示为"Cas12L_41_74408273_部分"的Cas12L氨基酸序列具有20%或更多序列同一性(例如,30%或更多、40%或更多、50%或更多、60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的氨基酸序列。例如,在一些情况下,Cas12L蛋白包含与图2NN中描绘的Cas12L氨基酸序列具有50%或更多序列同一性(例如,60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含与图2NN中描绘的Cas12L氨基酸序列具有80%或更多序列同一性(例如,85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更

多或100％序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含与图2NN中描绘的Cas12L氨基酸序列具有90％或更多序列同一性(例如,95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含具有图2NN中描绘的Cas12L氨基酸序列的氨基酸序列。在一些情况下,Cas12L蛋白包含具有图2NN中描绘的Cas12L蛋白序列的氨基酸序列,除了所述序列包含降低所述蛋白质的天然存在的催化活性的氨基酸取代(例如,1、2或3个氨基酸取代)。在一些情况下,Cas12L多肽具有850个氨基酸(aa)至950aa的长度,例如,850aa至875aa、875aa至889aa、889aa至900aa或900aa至950aa)。在一些情况下,Cas12L多肽具有889个氨基酸的长度。在一些情况下,结合Cas12L多肽(例如,包含与图2NN中描绘的Cas12L氨基酸序列具有20％或更多、30％或更多、40％或更多、50％或更多、60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％氨基酸序列同一性的氨基酸序列的Cas12L多肽)的引导RNA包含以下核苷酸序列:ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:1)或其反向互补序列。在一些情况下,引导RNA包含核苷酸序列(N)nATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:2)或其反向互补序列,其中N是任何核苷酸,并且n是15至30的整数,例如15至20、17至25、17至22、18至22、18至20、20至25或25至30)。

[0175] 图200

[0176] 在一些情况下,(主题组合物和/或方法的)Cas12L蛋白包含与图200中描绘并表示为“Cas12L_42_75186079_部分”的Cas12L氨基酸序列具有20％或更多序列同一性(例如,30％或更多、40％或更多、50％或更多、60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的约223个氨基酸的连续片段。例如,在一些情况下,Cas12L蛋白包含与图200中描绘的Cas12L氨基酸序列具有50％或更多序列同一性(例如,60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的约223个氨基酸的连续片段。在一些情况下,Cas12L蛋白包含与图200中描绘的Cas12L氨基酸序列具有80％或更多序列同一性(例如,85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的约223个氨基酸的连续片段。在一些情况下,Cas12L蛋白包含与图200中描绘的Cas12L氨基酸序列具有90％或更多序列同一性(例如,95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的约223个氨基酸的连续片段。在一些情况下,Cas12L蛋白包含具有图200中描绘的Cas12L氨基酸序列的约223个氨基酸的连续片段。在一些情况下,Cas12L蛋白包含具有图200中描绘的Cas12L蛋白序列的约223个氨基酸的连续片段,除了所述序列包含降低所述蛋白质的天然存在的催化活性的氨基酸取代(例如,1、2或3个氨基酸取代)。在一些情况下,Cas12L多肽具有700个氨基酸(aa)至800aa的长度,例如,700aa至725aa、725aa至750aa、750aa至775aa或775aa至800aa)。在一些情况下,Cas12L多肽具有725个氨基酸至775个氨基酸的长度。在一些情况下,结合Cas12L多肽(例如,包含与图200中描绘的Cas12L氨基酸序列具有20％或更多、30％或更多、40％或更多、50％或更多、60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％氨基酸序列同一性的氨基酸序列的Cas12L多肽)的引导RNA包含以下核苷酸序列:AATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:17)或其反向互补序列。在一些情况下,引导RNA包含核苷酸序列

(N)nAATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:18)或其反向互补序列,其中N是任何核苷酸,并且n是15至30的整数,例如15至20、17至25、17至22、18至22、18至20、20至25或25至30)。

[0177] 图2PP

[0178] 在一些情况下,(主题组合物和/或方法的)Cas12L蛋白包含与图2PP中描绘并表示为"Cas12L_43_75257103_部分"的Cas12L氨基酸序列具有20%或更多序列同一性(例如,30%或更多、40%或更多、50%或更多、60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的约439个氨基酸的连续片段。例如,在一些情况下,Cas12L蛋白包含与图2PP中描绘的Cas12L氨基酸序列具有50%或更多序列同一性(例如,60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的约439个氨基酸的连续片段。在一些情况下,Cas12L蛋白包含与图2PP中描绘的Cas12L氨基酸序列具有80%或更多序列同一性(例如,85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的约439个氨基酸的连续片段。在一些情况下,Cas12L蛋白包含与图2PP中描绘的Cas12L氨基酸序列具有90%或更多序列同一性(例如,95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的约439个氨基酸的连续片段。在一些情况下,Cas12L蛋白包含具有图2PP中描绘的Cas12L氨基酸序列的约439个氨基酸的连续片段。在一些情况下,Cas12L蛋白包含具有图2PP中描绘的Cas12L蛋白序列的约439个氨基酸的连续片段,除了所述序列包含降低所述蛋白质的天然存在的催化活性的氨基酸取代(例如,1、2或3个氨基酸取代)。在一些情况下,Cas12L多肽具有700个氨基酸(aa)至800aa的长度,例如,700aa至725aa、725aa至750aa、750aa至775aa或775aa至800aa)。在一些情况下,Cas12L多肽具有725个氨基酸至775个氨基酸的长度。在一些情况下,结合Cas12L多肽(例如,包含与图2PP中描绘的Cas12L氨基酸序列具有20%或更多、30%或更多、40%或更多、50%或更多、60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%氨基酸序列同一性的氨基酸序列的Cas12L多肽)的引导RNA包含以下核苷酸序列:ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:1)或其反向互补序列。在一些情况下,引导RNA包含核苷酸序列(N)nATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:2)或其反向互补序列,其中N是任何核苷酸,并且n是15至30的整数,例如15至20、17至25、17至22、18至22、18至20、20至25或25至30)。

[0179] 图2QQ

[0180] 在一些情况下,(主题组合物和/或方法的)Cas12L蛋白包含与图2QQ中描绘并表示为"Cas12L_44_75257103_部分"的Cas12L氨基酸序列具有20%或更多序列同一性(例如,30%或更多、40%或更多、50%或更多、60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的约196个氨基酸的连续片段。例如,在一些情况下,Cas12L蛋白包含与图2QQ中描绘的Cas12L氨基酸序列具有50%或更多序列同一性(例如,60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的约196个氨基酸的连续片段。在一些情况下,Cas12L蛋白包含与图2QQ中描绘的Cas12L氨

基酸序列具有80％或更多序列同一性(例如,85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的约196个氨基酸的连续片段。在一些情况下,Cas12L蛋白包含与图2QQ中描绘的Cas12L氨基酸序列具有90％或更多序列同一性(例如,95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的约196个氨基酸的连续片段。在一些情况下,Cas12L蛋白包含具有图2QQ中描绘的Cas12L氨基酸序列的约196个氨基酸的连续片段。在一些情况下,Cas12L蛋白包含具有图2QQ中描绘的Cas12L蛋白序列的约196个氨基酸的连续片段,除了所述序列包含降低所述蛋白质的天然存在的催化活性的氨基酸取代(例如,1、2或3个氨基酸取代)。在一些情况下,Cas12L多肽具有700个氨基酸(aa)至800aa的长度,例如,700aa至725aa、725aa至750aa、750aa至775aa或775aa至800aa)。在一些情况下,Cas12L多肽具有725个氨基酸至775个氨基酸的长度。在一些情况下,结合Cas12L多肽(例如,包含与图2QQ中描绘的Cas12L氨基酸序列具有20％或更多、30％或更多、40％或更多、50％或更多、60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％氨基酸序列同一性的氨基酸序列的Cas12L多肽)的引导RNA包含以下核苷酸序列:ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:1)或其反向互补序列。在一些情况下,引导RNA包含核苷酸序列(N)nATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:2)或其反向互补序列,其中N是任何核苷酸,并且n是15至30的整数,例如15至20、17至25、17至22、18至22、18至20、20至25或25至30)。

[0181] 图2RR

[0182] 在一些情况下,(主题组合物和/或方法的)Cas12L蛋白包含与图2RR中描绘并表示为“Cas12L_45_75616607_部分”的Cas12L氨基酸序列具有20％或更多序列同一性(例如,30％或更多、40％或更多、50％或更多、60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。例如,在一些情况下,Cas12L蛋白包含与图2RR中描绘的Cas12L氨基酸序列具有50％或更多序列同一性(例如,60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含与图2RR中描绘的Cas12L氨基酸序列具有80％或更多序列同一性(例如,85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含与图2RR中描绘的Cas12L氨基酸序列具有90％或更多序列同一性(例如,95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含具有图2RR中描绘的Cas12L氨基酸序列的氨基酸序列。在一些情况下,Cas12L蛋白包含具有图2RR中描绘的Cas12L蛋白序列的氨基酸序列,除了所述序列包含降低所述蛋白质的天然存在的催化活性的氨基酸取代(例如,1、2或3个氨基酸取代)。在一些情况下,Cas12L多肽具有700个氨基酸(aa)至800aa的长度,例如,700aa至725aa、725aa至748aa、748aa至775aa或775aa至800aa)。在一些情况下,Cas12L多肽具有748个氨基酸的长度。在一些情况下,结合Cas12L多肽(例如,包含与图2RR中描绘的Cas12L氨基酸序列具有20％或更多、30％或更多、40％或更多、50％或更多、60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％氨基酸序列同一性的氨基酸序列的

Cas12L多肽)的引导RNA包含以下核苷酸序列:ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:1)或其反向互补序列。在一些情况下,引导RNA包含核苷酸序列(N)nATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:2)或其反向互补序列,其中N是任何核苷酸,并且n是15至30的整数,例如15至20、17至25、17至22、18至22、18至20、20至25或25至30)。

[0183] 图2SS

[0184] 在一些情况下,(主题组合物和/或方法的)Cas12L蛋白包含与图2SS中描绘并表示为"Cas12L_46_75784289_部分"的Cas12L氨基酸序列具有20%或更多序列同一性(例如,30%或更多、40%或更多、50%或更多、60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的约481个氨基酸的连续片段。例如,在一些情况下,Cas12L)包含与图2SS中描绘的Cas12L氨基酸序列具有50%或更多序列同一性(例如,60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的约481个氨基酸的连续片段。在一些情况下,Cas12L蛋白)包含与图2SS中描绘的Cas12L氨基酸序列具有80%或更多序列同一性(例如,85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的约481个氨基酸的连续片段。在一些情况下,Cas12L蛋白)包含与图2SS中描绘的Cas12L氨基酸序列具有90%或更多序列同一性(例如,95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的约481个氨基酸的连续片段。在一些情况下,Cas12L蛋白)包含具有图2SS中描绘的Cas12L氨基酸序列的约481个氨基酸的连续片段。在一些情况下,Cas12L蛋白)包含具有图2SS中描绘的Cas12L蛋白序列的约481个氨基酸的连续片段,除了所述序列包含降低所述蛋白质的天然存在的催化活性的氨基酸取代(例如,1、2或3个氨基酸取代)。在一些情况下,Cas12L多肽具有700个氨基酸(aa)至800aa的长度,例如,700aa至725aa、725aa至750aa、750aa至775aa或775aa至800aa)。在一些情况下,Cas12L多肽具有725个氨基酸至775个氨基酸的长度。在一些情况下,结合Cas12L多肽(例如,包含与图2SS中描绘的Cas12L氨基酸序列具有20%或更多、30%或更多、40%或更多、50%或更多、60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%氨基酸序列同一性的氨基酸序列的Cas12L多肽)的引导RNA包含以下核苷酸序列:ATTGTTGTAACTCTTATTTGTATGGAGTAAACAAC(SEQ ID NO:1)或其反向互补序列。在一些情况下,引导RNA包含核苷酸序列(N)nATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:2)或其反向互补序列,其中N是任何核苷酸,并且n是15至30的整数,例如15至20、17至25、17至22、18至22、18至20、20至25或25至30)。

[0185] 图2TT

[0186] 在一些情况下,(主题组合物和/或方法的)Cas12L蛋白包含与图2TT中描绘并表示为"Cas12L_47_76512228"的Cas12L氨基酸序列具有20%或更多序列同一性(例如,30%或更多、40%或更多、50%或更多、60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的氨基酸序列。例如,在一些情况下,Cas12L蛋白包含与图2TT中描绘的Cas12L氨基酸序列具有50%或更多序列同一性(例如,60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的氨基酸序列。在

一些情况下,Cas12L蛋白包含与图2TT中描绘的Cas12L氨基酸序列具有80%或更多序列同一性(例如,85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含与图2TT中描绘的Cas12L氨基酸序列具有90%或更多序列同一性(例如,95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含具有图2TT中描绘的Cas12L氨基酸序列的氨基酸序列。在一些情况下,Cas12L蛋白包含具有图2TT中描绘的Cas12L蛋白序列的氨基酸序列,除了所述序列包含降低所述蛋白质的天然存在的催化活性的氨基酸取代(例如,1、2或3个氨基酸取代)。在一些情况下,Cas12L多肽具有700个氨基酸(aa)至800aa的长度,例如,700aa至725aa、725aa至746aa、746aa至775aa或775aa至800aa)。在一些情况下,Cas12L多肽具有746个氨基酸的长度。在一些情况下,结合Cas12L多肽(例如,包含与图2TT中描绘的Cas12L氨基酸序列具有20%或更多、30%或更多、40%或更多、50%或更多、60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%氨基酸序列同一性的氨基酸序列的Cas12L多肽)的引导RNA包括以下核苷酸序列:ATTGTTGTAACTTTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:19)或其反向互补序列。在一些情况下,引导RNA包括核苷酸序列:(N)nATTGTTGTAACTTTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:20)或其反向互补序列,其中N是任何核苷酸,并且n是15至30的整数,例如15至20、17至25、17至22、18至22、18至20、20至25或25至30)。

[0187] 图2UU

[0188] 在一些情况下,(主题组合物和/或方法的)Cas12L蛋白包含与图2UU中描绘并表示为"Cas12L_48_76600450"的Cas12L氨基酸序列具有20%或更多序列同一性(例如,30%或更多、40%或更多、50%或更多、60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的氨基酸序列。例如,在一些情况下,Cas12L蛋白包含与图2UU中描绘的Cas12L氨基酸序列具有50%或更多序列同一性(例如,60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含与图2UU中描绘的Cas12L氨基酸序列具有80%或更多序列同一性(例如,85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含与图2UU中描绘的Cas12L氨基酸序列具有90%或更多序列同一性(例如,95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含具有图2UU中描绘的Cas12L氨基酸序列的氨基酸序列。在一些情况下,Cas12L蛋白包含具有图2UU中描绘的Cas12L蛋白序列的氨基酸序列,除了所述序列包含降低所述蛋白质的天然存在的催化活性的氨基酸取代(例如,1、2或3个氨基酸取代)。在一些情况下,Cas12L多肽具有700个氨基酸(aa)至800aa的长度,例如,700aa至725aa、725aa至746aa、746aa至775aa或775aa至800aa)。在一些情况下,Cas12L多肽具有746个氨基酸的长度。在一些情况下,结合Cas12L多肽(例如,包含与图2UU中描绘的Cas12L氨基酸序列具有20%或更多、30%或更多、40%或更多、50%或更多、60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%氨基酸序列同一性的氨基酸序列的Cas12L多肽)的引导RNA包含以下核苷酸序列:ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID

NO:1)或其反向互补序列。在一些情况下,引导RNA包含核苷酸序列(N)nATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:2)或其反向互补序列,其中N是任何核苷酸,并且n是15至30的整数,例如15至20、17至25、17至22、18至22、18至20、20至25或25至30)。

[0189] 图2VV

[0190] 在一些情况下,(主题组合物和/或方法的)Cas12L蛋白包含与图2VV中描绘并表示为"Cas12L_49_44880081"的Cas12L氨基酸序列具有20%或更多序列同一性(例如,30%或更多、40%或更多、50%或更多、60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的氨基酸序列。例如,在一些情况下,Cas12L蛋白包含与图2VV中描绘的Cas12L氨基酸序列具有50%或更多序列同一性(例如,60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含与图2VV中描绘的Cas12L氨基酸序列具有80%或更多序列同一性(例如,85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含与图2VV中描绘的Cas12L氨基酸序列具有90%或更多序列同一性(例如,95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含具有图2VV中描绘的Cas12L氨基酸序列的氨基酸序列。在一些情况下,Cas12L蛋白包含具有图2VV中描绘的Cas12L蛋白序列的氨基酸序列,除了所述序列包含降低所述蛋白质的天然存在的催化活性的氨基酸取代(例如,1、2或3个氨基酸取代)。在一些情况下,Cas12L多肽具有700个氨基酸(aa)至800aa的长度,例如,700aa至725aa、725aa至746aa、746aa至775aa或775aa至800aa)。在一些情况下,Cas12L多肽具有746个氨基酸的长度。在一些情况下,结合Cas12L多肽(例如,包含与图2VV中描绘的Cas12L氨基酸序列具有20%或更多、30%或更多、40%或更多、50%或更多、60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%氨基酸序列同一性的氨基酸序列的Cas12L多肽)的引导RNA包含以下核苷酸序列:ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:1)或其反向互补序列。在一些情况下,引导RNA包含核苷酸序列(N)nATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:2)或其反向互补序列,其中N是任何核苷酸,并且n是15至30的整数,例如15至20、17至25、17至22、18至22、18至20、20至25或25至30)。

[0191] 图2WW

[0192] 在一些情况下,(主题组合物和/或方法的)Cas12L蛋白包含与图2WW中描绘并表示为"Cas12L_50_83012613"的Cas12L氨基酸序列具有20%或更多序列同一性(例如,30%或更多、40%或更多、50%或更多、60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的氨基酸序列。例如,在一些情况下,Cas12L蛋白包含与图2WW中描绘的Cas12L氨基酸序列具有50%或更多序列同一性(例如,60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含与图2WW中描绘的Cas12L氨基酸序列具有80%或更多序列同一性(例如,85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含与图2WW中描绘的Cas12L

氨基酸序列具有90％或更多序列同一性(例如,95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含具有图2WW中描绘的Cas12L氨基酸序列的氨基酸序列。在一些情况下,Cas12L蛋白包含具有图2WW中描绘的Cas12L蛋白序列的氨基酸序列,除了所述序列包含降低所述蛋白质的天然存在的催化活性的氨基酸取代(例如,1、2或3个氨基酸取代)。在一些情况下,Cas12L多肽具有700个氨基酸(aa)至800aa的长度,例如,700aa至735aa、735aa至750aa、750aa至775aa或775aa至800aa)。在一些情况下,Cas12L多肽具有735个氨基酸的长度。在一些情况下,结合Cas12L多肽(例如,包含与图2WW中描绘的Cas12L氨基酸序列具有20％或更多、30％或更多、40％或更多、50％或更多、60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％氨基酸序列同一性的氨基酸序列的Cas12L多肽)的引导RNA包含以下核苷酸序列:AATGTTGTAGATACCTTTTTGTAAGGATTGAACAAC(SEQ ID NO:21)或其反向互补序列。在一些情况下,引导RNA包含核苷酸序列(N)nAATGTTGTAGATACCTTTTTGTAAGGATTGAACAAC(SEQ ID NO:22)或其反向互补序列,其中N是任何核苷酸,并且n是15至30的整数,例如15至20、17至25、17至22、18至22、18至20、20至25或25至30)。

[0193] 图2XX

[0194] 在一些情况下,(主题组合物和/或方法的)Cas12L蛋白包含与图2XX中描绘并表示为“Cas12L_51_82983331”的Cas12L氨基酸序列具有20％或更多序列同一性(例如,30％或更多、40％或更多、50％或更多、60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。例如,在一些情况下,Cas12L蛋白包含与图2XX中描绘的Cas12L氨基酸序列具有50％或更多序列同一性(例如,60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含与图2XX中描绘的Cas12L氨基酸序列具有80％或更多序列同一性(例如,85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含与图2XX中描绘的Cas12L氨基酸序列具有90％或更多序列同一性(例如,95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。在一些情况下,Cas12L蛋白包含具有图2XX中描绘的Cas12L氨基酸序列的氨基酸序列。在一些情况下,Cas12L蛋白包含具有图2XX中描绘的Cas12L蛋白序列的氨基酸序列,除了所述序列包含降低所述蛋白质的天然存在的催化活性的氨基酸取代(例如,1、2或3个氨基酸取代)。在一些情况下,Cas12L多肽具有750个氨基酸(aa)至850aa的长度,例如,750aa至779aa、779aa至800aa或800aa至850aa)。在一些情况下,Cas12L多肽具有779个氨基酸的长度。在一些情况下,结合Cas12L多肽(例如,包含与图2XX中描绘的Cas12L氨基酸序列具有20％或更多、30％或更多、40％或更多、50％或更多、60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％氨基酸序列同一性的氨基酸序列的Cas12L多肽)的引导RNA包含以下核苷酸序列:ATTGTTGTAATACTATTTTTGTAAAGTATAAACAAC(SEQ ID NO:23)或其反向互补序列。在一些情况下,引导RNA包含核苷酸序列(N)nATTGTTGTAATACTATTTTTGTAAAGTATAAACAAC(SEQ ID NO:24)或其反向互补序列,其中N是任何核苷酸,并且n是15至30的整数,例如15至20、17至25、17至22、18至22、18至20、20至25或25至30)。

[0195] 图2YY

[0196] 在一些情况下，(主题组合物和/或方法的)Cas12L蛋白包含与图2YY中描绘并表示为“Cas12L_52_76767885”的Cas12L氨基酸序列具有20％或更多序列同一性(例如，30％或更多、40％或更多、50％或更多、60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。例如，在一些情况下，Cas12L蛋白包含与图2YY中描绘的Cas12L氨基酸序列具有50％或更多序列同一性(例如，60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。在一些情况下，Cas12L蛋白包含与图2YY中描绘的Cas12L氨基酸序列具有80％或更多序列同一性(例如，85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。在一些情况下，Cas12L蛋白包含与图2YY中描绘的Cas12L氨基酸序列具有90％或更多序列同一性(例如，95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。在一些情况下，Cas12L蛋白包含具有图2YY中描绘的Cas12L氨基酸序列的氨基酸序列。在一些情况下，Cas12L蛋白包含具有图2YY中描绘的Cas12L蛋白序列的氨基酸序列，除了所述序列包含降低所述蛋白质的天然存在的催化活性的氨基酸取代(例如，1、2或3个氨基酸取代)。在一些情况下，Cas12L多肽具有550个氨基酸(aa)至700aa的长度，例如，550aa至592aa、592aa至625aa、625aa至650aa或650aa至700aa)。在一些情况下，Cas12L多肽具有592个氨基酸的长度。在一些情况下，结合Cas12L多肽(例如，包含与图2YY中描绘的Cas12L氨基酸序列具有20％或更多、30％或更多、40％或更多、50％或更多、60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％氨基酸序列同一性的氨基酸序列的Cas12L多肽)的引导RNA包含以下核苷酸序列：AATGTTGTAGATGCCTTTTTATAAGGATTAAACAAC(SEQ ID NO:25)或其反向互补序列。在一些情况下，引导RNA包含核苷酸序列(N)nAATGTTGTAGATGCCTTTTTATAAGGATTAAACAAC(SEQ ID NO:26)或其反向互补序列，其中N是任何核苷酸，并且n是15至30的整数，例如15至20、17至25、17至22、18至22、18至20、20至25或25至30)。

[0197] 图2ZZ

[0198] 在一些情况下，(主题组合物和/或方法的)Cas12L蛋白包含与图2ZZ中描绘并表示为“Cas12L_53_77216451”的Cas12L氨基酸序列具有20％或更多序列同一性(例如，30％或更多、40％或更多、50％或更多、60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。例如，在一些情况下，Cas12L蛋白包含与图2ZZ中描绘的Cas12L氨基酸序列具有50％或更多序列同一性(例如，60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。在一些情况下，Cas12L蛋白包含与图2ZZ中描绘的Cas12L氨基酸序列具有80％或更多序列同一性(例如，85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。在一些情况下，Cas12L蛋白包含与图2ZZ中描绘的Cas12L氨基酸序列具有90％或更多序列同一性(例如，95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。在一些情况下，Cas12L蛋白包含具有图2ZZ中描绘的Cas12L氨基酸序列的氨基酸序列。在一些情况下，Cas12L蛋白包含具有图2ZZ中描绘的

Cas12L蛋白序列的氨基酸序列,除了所述序列包含降低所述蛋白质的天然存在的催化活性的氨基酸取代(例如,1、2或3个氨基酸取代)。在一些情况下,Cas12L多肽具有700个氨基酸(aa)至800aa的长度,例如,700aa至725aa、725aa至746aa、746aa至775aa或775aa至800aa)。在一些情况下,Cas12L多肽具有746个氨基酸的长度。在一些情况下,结合Cas12L多肽(例如,包含与图2ZZ中描绘的Cas12L氨基酸序列具有20%或更多、30%或更多、40%或更多、50%或更多、60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%氨基酸序列同一性的氨基酸序列的Cas12L多肽)的引导RNA包含以下核苷酸序列:ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:1)或其反向互补序列。在一些情况下,引导RNA包含核苷酸序列(N)nATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:2)或其反向互补序列,其中N是任何核苷酸,并且n是15至30的整数,例如15至20、17至25、17至22、18至22、18至20、20至25或25至30)。

[0199] 图2AAA

[0200] 在一些情况下,(主题组合物和/或方法的)Cas12L蛋白包含与图2AAA中描绘并表示为"Cas12L_54_77468912_部分"的Cas12L氨基酸序列具有20%或更多序列同一性(例如,30%或更多、40%或更多、50%或更多、60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的约29个氨基酸的连续片段。例如,在一些情况下,Cas12L蛋白包含与图2AAA中描绘的Cas12L氨基酸序列具有50%或更多序列同一性(例如,60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的约29个氨基酸的连续片段。在一些情况下,Cas12L蛋白包含与图2AAA中描绘的Cas12L氨基酸序列具有80%或更多序列同一性(例如,85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的约29个氨基酸的连续片段。在一些情况下,Cas12L蛋白包含与图2AAA中描绘的Cas12L氨基酸序列具有90%或更多序列同一性(例如,95%或更多、97%或更多、98%或更多、99%或更多或100%序列同一性)的约29个氨基酸的连续片段。在一些情况下,Cas12L蛋白包含具有图2AAA中描绘的Cas12L氨基酸序列的约29个氨基酸的连续片段。在一些情况下,Cas12L蛋白包含具有图2AAA中描绘的Cas12L蛋白序列的约29个氨基酸的连续片段,除了所述序列包含降低所述蛋白质的天然存在的催化活性的氨基酸取代(例如,1、2或3个氨基酸取代)。在一些情况下,Cas12L多肽具有700个氨基酸(aa)至800aa的长度,例如,700aa至725aa、725aa至750aa、750aa至775aa或775aa至800aa)。在一些情况下,Cas12L多肽具有725个氨基酸至775个氨基酸的长度。在一些情况下,结合Cas12L多肽(例如,包含与图2AAA中描绘的Cas12L氨基酸序列具有20%或更多、30%或更多、40%或更多、50%或更多、60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%氨基酸序列同一性的氨基酸序列的Cas12L多肽)的引导RNA包含以下核苷酸序列:ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:1)或其反向互补序列。在一些情况下,引导RNA包含核苷酸序列(N)nATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:2)或其反向互补序列,其中N是任何核苷酸,并且n是15至30的整数,例如15至20、17至25、17至22、18至22、18至20、20至25或25至30)。

[0201] 图2BBB

[0202] 在一些情况下，(主题组合物和/或方法的)Cas12L蛋白包含与图2BBB中描绘并表示为“Cas12L_55_77738117”的Cas12L氨基酸序列具有20％或更多序列同一性(例如，30％或更多、40％或更多、50％或更多、60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。例如，在一些情况下，Cas12L蛋白包含与图2BBB中描绘的Cas12L氨基酸序列具有50％或更多序列同一性(例如，60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。在一些情况下，Cas12L蛋白包含与图2BBB中描绘的Cas12L氨基酸序列具有80％或更多序列同一性(例如，85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。在一些情况下，Cas12L蛋白包含与图2BBB中描绘的Cas12L氨基酸序列具有90％或更多序列同一性(例如，95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。在一些情况下，Cas12L蛋白包含具有图2BBB中描绘的Cas12L氨基酸序列的氨基酸序列。在一些情况下，Cas12L蛋白包含具有图2BBB中描绘的Cas12L蛋白序列的氨基酸序列，除了所述序列包含降低所述蛋白质的天然存在的催化活性的氨基酸取代(例如，1、2或3个氨基酸取代)。在一些情况下，Cas12L多肽具有700个氨基酸(aa)至800aa的长度，例如，700aa至725aa、725aa至750aa、750aa至775aa或775aa至800aa)。在一些情况下，Cas12L多肽具有725个氨基酸至775个氨基酸的长度。在一些情况下，结合Cas12L多肽(例如，包含与图2BBB中描绘的Cas12L氨基酸序列具有20％或更多、30％或更多、40％或更多、50％或更多、60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％氨基酸序列同一性的氨基酸序列的Cas12L多肽)的引导RNA包含以下核苷酸序列:ATTGTTGTAATACACTTTTTATAAGGTATGAACAAC(SEQ ID NO:27)或其反向互补序列。在一些情况下，引导RNA包含核苷酸序列(N)nATTGTTGTAATACACTTTTTATAAGGTATGAACAAC(SEQ ID NO:28)或其反向互补序列，其中N是任何核苷酸，并且n是15至30的整数，例如15至20、17至25、17至22、18至22、18至20、20至25或25至30)。

[0203] 图2CCC

[0204] 在一些情况下，(主题组合物和/或方法的)Cas12L蛋白包含与图2CCC中描绘并表示为“Cas12L_56_65286425”的Cas12L氨基酸序列具有20％或更多序列同一性(例如，30％或更多、40％或更多、50％或更多、60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。例如，在一些情况下，Cas12L蛋白包含与图2CCC中描绘的Cas12L氨基酸序列具有50％或更多序列同一性(例如，60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。在一些情况下，Cas12L蛋白包含与图2CCC中描绘的Cas12L氨基酸序列具有80％或更多序列同一性(例如，85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。在一些情况下，Cas12L蛋白包含与图2CCC中描绘的Cas12L氨基酸序列具有90％或更多序列同一性(例如，95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。在一些情况下，Cas12L蛋白包含具有图2CCC中描绘的Cas12L氨基酸序列的氨基酸序列。在一些情况下，Cas12L蛋白包含具有图

2CCC中描绘的Cas12L蛋白序列的氨基酸序列，除了所述序列包含降低所述蛋白质的天然存在的催化活性的氨基酸取代(例如，1、2或3个氨基酸取代)。在一些情况下，Cas12L多肽具有650个氨基酸(aa)至750aa的长度，例如，650aa至692aa、692aa至725aa或725aa至750aa)。在一些情况下，Cas12L多肽具有692个氨基酸的长度。在一些情况下，结合Cas12L多肽(例如，包含与图2CCC中描绘的Cas12L氨基酸序列具有20％或更多、30％或更多、40％或更多、50％或更多、60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％氨基酸序列同一性的氨基酸序列的Cas12L多肽)的引导RNA包含以下核苷酸序列：ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:1)或其反向互补序列。在一些情况下，引导RNA包含核苷酸序列(N)nATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:2)或其反向互补序列，其中N是任何核苷酸，并且n是15至30的整数，例如15至20、17至25、17至22、18至22、18至20、20至25或25至30)。

[0205] 图2DDD

[0206] 在一些情况下，(主题组合物和/或方法的)Cas12L蛋白包含与图2DDD中描绘并表示为“Cas12L_57_65567118_部分”的Cas12L氨基酸序列具有20％或更多序列同一性(例如，30％或更多、40％或更多、50％或更多、60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的约441个氨基酸的连续片段。例如，在一些情况下，Cas12L蛋白)包含与图2DDD中描绘的Cas12L氨基酸序列具有50％或更多序列同一性(例如，60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的约441个氨基酸的连续片段。在一些情况下，Cas12L蛋白)包含与图2DDD中描绘的Cas12L氨基酸序列具有80％或更多序列同一性(例如，85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的约441个氨基酸的连续片段。在一些情况下，Cas12L蛋白)包含与图2DDD中描绘的Cas12L氨基酸序列具有90％或更多序列同一性(例如，95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的约441个氨基酸的连续片段。在一些情况下，Cas12L蛋白)包含具有图2DDD中描绘的Cas12L氨基酸序列的约441个氨基酸的连续片段。在一些情况下，Cas12L蛋白)包含具有图2DDD中描绘的Cas12L蛋白序列的约441个氨基酸的连续片段，除了所述序列包含降低所述蛋白质的天然存在的催化活性的氨基酸取代(例如，1、2或3个氨基酸取代)。在一些情况下，Cas12L多肽具有700个氨基酸(aa)至800aa的长度，例如，700aa至725aa、725aa至750aa、750aa至775aa或775aa至800aa)。在一些情况下，Cas12L多肽具有725个氨基酸至775个氨基酸的长度。在一些情况下，结合Cas12L多肽(例如，包含与图2DDD中描绘的Cas12L氨基酸序列具有20％或更多、30％或更多、40％或更多、50％或更多、60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％氨基酸序列同一性的氨基酸序列的Cas12L多肽)的引导RNA包含以下核苷酸序列：ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:1)或其反向互补序列。在一些情况下，引导RNA包含核苷酸序列(N)nATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:2)或其反向互补序列，其中N是任何核苷酸，并且n是15至30的整数，例如15至20、17至25、17至22、18至22、18至20、20至25或25至30)。

[0207] 图2EEE

[0208] 在一些情况下，(主题组合物和/或方法的)Cas12L蛋白包含与图2EEE中描绘并表示为“Cas12L_58_66287853_部分”的Cas12L氨基酸序列具有20％或更多序列同一性(例如，30％或更多、40％或更多、50％或更多、60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的约397个氨基酸的连续片段。例如，在一些情况下，Cas12L蛋白包含与图2EEE中描绘的Cas12L氨基酸序列具有50％或更多序列同一性(例如，60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的约397个氨基酸的连续片段。在一些情况下，Cas12L蛋白包含与图2EEE中描绘的Cas12L氨基酸序列具有80％或更多序列同一性(例如，85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的约397个氨基酸的连续片段。在一些情况下，Cas12L蛋白包含与图2EEE中描绘的Cas12L氨基酸序列具有90％或更多序列同一性(例如，95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的约397个氨基酸的连续片段。在一些情况下，Cas12L蛋白包含具有图2EEE中描绘的Cas12L氨基酸序列的约397个氨基酸的连续片段。在一些情况下，Cas12L蛋白包含具有图2EEE中描绘的Cas12L蛋白序列的约397个氨基酸的连续片段，除了所述序列包含降低所述蛋白质的天然存在的催化活性的氨基酸取代(例如，1、2或3个氨基酸取代)。在一些情况下，Cas12L多肽具有700个氨基酸(aa)至800aa的长度，例如，700aa至725aa、725aa至750aa、750aa至775aa或775aa至800aa)。在一些情况下，Cas12L多肽具有725个氨基酸至775个氨基酸的长度。在一些情况下，结合Cas12L多肽(例如，包含与图2EEE中描绘的Cas12L氨基酸序列具有20％或更多、30％或更多、40％或更多、50％或更多、60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％氨基酸序列同一性的氨基酸序列的Cas12L多肽)的引导RNA包含以下核苷酸序列：ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:1)或其反向互补序列。在一些情况下，引导RNA包含核苷酸序列(N)nATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:2)或其反向互补序列，其中N是任何核苷酸，并且n是15至30的整数，例如15至20、17至25、17至22、18至22、18至20、20至25或25至30)。

[0209] 图2FFF

[0210] 在一些情况下，(主题组合物和/或方法的)Cas12L蛋白包含与图2FFF中描绘并表示为“Cas12L_39_73877227”的Cas12L氨基酸序列具有20％或更多序列同一性(例如，30％或更多、40％或更多、50％或更多、60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。例如，在一些情况下，Cas12L蛋白包含与图2FFF中描绘的Cas12L氨基酸序列具有50％或更多序列同一性(例如，60％或更多、70％或更多、80％或更多、85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。在一些情况下，Cas12L蛋白包含与图2FFF中描绘的Cas12L氨基酸序列具有80％或更多序列同一性(例如，85％或更多、90％或更多、95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。在一些情况下，Cas12L蛋白包含与图2FFF中描绘的Cas12L氨基酸序列具有90％或更多序列同一性(例如，95％或更多、97％或更多、98％或更多、99％或更多或100％序列同一性)的氨基酸序列。在一些情况下，Cas12L蛋白包含具有图

2FFF中描绘的Cas12L氨基酸序列的氨基酸序列。在一些情况下,Cas12L蛋白包含具有图2FFF中描绘的Cas12L蛋白序列的氨基酸序列,除了所述序列包含降低所述蛋白质的天然存在的催化活性的氨基酸取代(例如,1、2或3个氨基酸取代)。在一些情况下,Cas12L多肽具有700个氨基酸(aa)至800aa的长度,例如,700aa至725aa、725aa至746aa、746aa至775aa或775aa至800aa)。在一些情况下,Cas12L多肽具有746个氨基酸的长度。在一些情况下,结合Cas12L多肽(例如,包含与图2FFF中描绘的Cas12L氨基酸序列具有20%或更多、30%或更多、40%或更多、50%或更多、60%或更多、70%或更多、80%或更多、85%或更多、90%或更多、95%或更多、97%或更多、98%或更多、99%或更多或100%氨基酸序列同一性的氨基酸序列的Cas12L多肽)的引导RNA包含以下核苷酸序列:ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:1)或其反向互补序列。在一些情况下,引导RNA包含核苷酸序列(N)nATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC(SEQ ID NO:2)或其反向互补序列,其中N是任何核苷酸,并且n是15至30的整数,例如15至20、17至25、17至22、18至22、18至20、20至25或25至30)。

[0211] Cas12L变体

[0212] 当与相应野生型Cas12L蛋白的氨基酸序列相比时,例如当与图2A-2Z、图2AA-2ZZ和图2AAA-2FFF中任一者中描绘的Cas12L氨基酸序列相比时,变体Cas12L蛋白具有相差至少一个氨基酸(例如,具有缺失、插入、取代、融合)的氨基酸序列。在一些情况下,与图2A-2Z、图2AA-2ZZ和图2AAA-2FFF中任一者中描绘的Cas12L氨基酸序列相比,Cas12L变体包含1个氨基酸取代至10个氨基酸取代。在一些情况下,与图2A-2Z、图2AA-2ZZ和图2AAA-2FFF中任一者中描绘的Cas12L氨基酸序列相比,Cas12L变体包含RuvC结构域中的1个氨基酸取代至10个氨基酸取代。

[0213] 变体-催化活性

[0214] 在一些情况下,Cas12L蛋白是变体Cas12L蛋白,例如相对于天然存在的催化活性序列突变的蛋白,并且在与对应的天然存在的序列相比时,表现出降低的切割活性(例如,表现出90%或更低、80%或更低、70%或更低、60%或更低、50%或更低、40%或更低,或30%或更低的切割活性)。在一些情况下,这种变体Cas12L蛋白是催化"死"蛋白(基本上没有切割活性)并且可被称为'dCas12L'。在一些情况下,变体Cas12L蛋白是切口酶(仅切割双链靶核酸(例如,双链靶DNA)的一条链)。如本文更详细描述的,在一些情况下,Cas12L蛋白(在一些情况下,是具有野生型切割活性的Cas12L蛋白并且在一些情况下,是具有降低的切割活性的变体Cas12L,例如dCas12L或切口酶Cas12L)与具有目标活性(例如,目标催化活性)的异源多肽融合(缀合)以形成融合蛋白(融合Cas12L蛋白)。

[0215] 在一些情况下,变体Cas12L多肽包含对图3A-3B中所示活性位点残基的一个、两个或三个氨基酸的取代,其中与不包含一个、两个或三个取代的对照Cas12L多肽相比,变体Cas12L多肽表现出降低的催化活性。

[0216] 变体-融合Cas12L多肽

[0217] 如上所示,在一些情况下,Cas12L蛋白(在一些情况下,是具有野生型切割活性的Cas12L蛋白并且在一些情况下,是具有降低的切割活性的变体Cas12L,例如dCas12L或切口酶Cas12L)与具有目标活性(例如,目标催化活性)的异源多肽融合(缀合)以形成融合蛋白。Cas12L蛋白可与之融合的异源多肽在本文中被称为"融合配偶体"。

[0218] 在一些情况下，融合配偶体可调节靶DNA的转录(例如，抑制转录、增加转录)。例如，在一些情况下，融合配偶体是抑制转录的蛋白质(或来自蛋白质的结构域)(例如，转录阻遏物，一种通过转录抑制蛋白的募集、靶DNA的修饰诸如甲基化、DNA修饰物的募集、与靶DNA相关联的组蛋白的调节、组蛋白修饰物(诸如修饰组蛋白的乙酰化和/或甲基化的那些组蛋白修饰物)的募集等起作用的蛋白质)。在一些情况下，融合配偶体是增加转录的蛋白质(或来自蛋白质的结构域)(例如，转录激活因子，一种通过转录激活蛋白的募集、靶DNA的修饰诸如甲基化、DNA修饰物的募集、与靶DNA相关联的组蛋白的调节、组蛋白修饰物(诸如修饰组蛋白的乙酰化和/或甲基化的那些组蛋白修饰物)的募集等起作用的蛋白质)。

[0219] 在某些情况下，融合配偶体(异源多肽)是逆转录酶。在某些情况下，融合配偶体是碱基编辑器。在某些情况下，融合配偶体(异源多肽)是脱氨酶。

[0220] 在一些情况下，融合Cas12L蛋白包括具有修饰靶核酸的酶促活性(例如，核酸酶活性、甲基转移酶活性、脱甲基酶活性、DNA修复活性、DNA损伤活性、脱氨基活性、歧化酶活性、烷基化活性、脱嘌呤活性、氧化活性、嘧啶二聚体形成活性、整合酶活性、转座酶活性、重组酶活性、聚合酶活性、连接酶活性、解旋酶活性、光裂合酶活性或糖基化酶活性)的异源多肽。

[0221] 在一些情况下，融合Cas12L蛋白包括具有修饰与靶核酸相关联的多肽(例如，组蛋白)的酶促活性(例如，甲基转移酶活性、脱甲基酶活性、乙酰转移酶活性、脱乙酰酶活性、激酶活性、磷酸酶活性、泛素连接酶活性、去泛素化活性、腺苷酸化活性、脱腺苷酸化活性、SUMO化活性、脱SUMO化活性、核糖基化活性、脱核糖基化活性、豆蔻酰化活性或脱豆蔻酰化活性)的异源多肽。

[0222] 可用于增加转录的蛋白质(或其片段)的实例包括但不限于：转录激活因子，诸如VP16、VP64、VP48、VP160、p65亚结构域(例如，来自NFkB)以及EDLL的激活结构域和/或TAL激活结构域(例如，针对植物中的活性)；组蛋白赖氨酸甲基转移酶，诸如SET1A、SET1B、MLL1至5、ASH1、SYMD2、NSD1等；组蛋白赖氨酸脱甲基酶，诸如JHDM2a/b、UTX、JMJD3等；组蛋白乙酰转移酶，诸如GCN5、PCAF、CBP、p300、TAF1、TIP60/PLIP、MOZ/MYST3、MORF/MYST4、SRC1、ACTR、P160、CLOCK等；以及DNA脱甲基酶，诸如10-11易位(TET)双加氧酶1(TET1CD)、TET1、DME、DML1、DML2、ROS1等。

[0223] 可用于减少转录的蛋白质(或其片段)的实例包括但不限于：转录阻遏子，诸如Krüppel相关盒(KRAB或SKD)；KOX1阻遏结构域；Mad mSIN3相互作用结构域(SID)；ERF阻遏子结构域(ERD)、SRDX阻遏结构域(例如，针对植物中的阻遏)等；组蛋白赖氨酸甲基转移酶，诸如Pr-SET7/8、SUV4-20H1、RIZ1等；组蛋白赖氨酸脱甲基酶，诸如JMJD2A/JHDM3A、JMJD2B、JMJD2C/GASC1、JMJD2D、JARID1A/RBP2、JARID1B/PLU-1、JARID1C/SMCX、JARID1D/SMCY等；组蛋白赖氨酸脱乙酰酶，诸如HDAC1、HDAC2、HDAC3、HDAC8、HDAC4、HDAC5、HDAC7、HDAC9、SIRT1、SIRT2、HDAC11等；DNA甲基化酶，诸如HhaI DNA m5c-甲基转移酶(M.HhaI)、DNA甲基转移酶1(DNMT1)、DNA甲基转移酶3a(DNMT3a)、DNA甲基转移酶3b(DNMT3b)、METI、DRM3(植物)、ZMET2、CMT1、CMT2(植物)等；以及外周募集元件，诸如核纤层蛋白A、核纤层蛋白B等。

[0224] 在一些情况下，融合配偶体具有修饰靶核酸(例如，ssRNA、dsRNA、ssDNA、dsDNA)的酶活性。可由融合配偶体提供的酶活性的实例包括但不限于：核酸酶活性，诸如由限制性酶(例如，FokI核酸酶)提供的活性；甲基转移酶活性，诸如由甲基转移酶(例如，HhaI DNA

m5c-甲基转移酶(M.HhaI)、DNA甲基转移酶1(DNMT1)、DNA甲基转移酶3a(DNMT3a)、DNA甲基转移酶3b(DNMT3b)、METI、DRM3(植物)、ZMET2、CMT1、CMT2(植物)等)提供的活性;脱甲基酶活性,诸如由脱甲基酶(例如;10-11易位(TET)双加氧酶1(TET1CD)、TET1、DME、DML1、DML2、ROS1等)提供的活性;DNA修复活性;DNA损伤活性;脱氨基活性,诸如由脱氨酶(例如,胞嘧啶脱氨酶,诸如大鼠APOBEC1)提供的活性;歧化酶活性;烷基化活性;脱嘌呤活性;氧化活性;嘧啶二聚体形成活性;整合酶活性,诸如由整合酶和/或解离酶(例如,Gin转化酶诸如Gin转化酶的过度活跃突变体GinH106Y、人免疫缺陷病毒1型整合酶(IN)、Tn3解离酶等)提供的活性;转座酶活性;重组酶活性,诸如由重组酶(例如,Gin重组酶的催化结构域)提供的活性;聚合酶活性;连接酶活性;解旋酶活性;光裂合酶活性和糖基化酶活性)。

[0225] 在一些情况下,融合配偶体具有修饰与靶核酸(例如,ssRNA、dsRNA、ssDNA、dsDNA)相关联的蛋白质(例如,组蛋白、RNA结合蛋白、DNA结合蛋白等)的酶活性。可由融合配偶体提供的酶活性(修饰与靶核酸相关联的蛋白质)的实例包括但不限于:甲基转移酶活性,诸如由组蛋白甲基转移酶(HMT)(例如,花斑抑制因子3-9同源物1(SUV39H1,也称为KMT1A)、常染色体组蛋白赖氨酸甲基转移酶2(G9A,也称为KMT1C和EHMT2)、SUV39H2、ESET/SETDB1等、SET1A、SET1B、MLL1至5、ASH1、SYMD2、NSD1、DOT1L、Pr-SET7/8、SUV4-20H1、EZH2、RIZ1)提供的活性;脱甲基酶活性,诸如由组蛋白脱甲基酶(例如,赖氨酸脱甲基酶1A(KDM1A,也称为LSD1)、JHDM2a/b、JMJD2A/JHDM3A、JMJD2B、JMJD2C/GASC1、JMJD2D、JARID1A/RBP2、JARID1B/PLU-1、JARID1C/SMCX、JARID1D/SMCY、UTX、JMJD3等)提供的活性;乙酰转移酶活性,诸如由组蛋白乙酰转移酶(例如,人乙酰转移酶p300、GCN5、PCAF、CBP、TAF1、TIP60/PLIP、MOZ/MYST3、MORF/MYST4、HBO1/MYST2、HMOF/MYST1、SRC1、ACTR、P160、CLOCK等的催化核心/片段)提供的活性;脱乙酰酶活性,诸如由组蛋白脱乙酰酶(例如,HDAC1、HDAC2、HDAC3、HDAC8、HDAC4、HDAC5、HDAC7、HDAC9、SIRT1、SIRT2、HDAC11等)提供的活性;激酶活性;磷酸酶活性;泛素连接酶活性;去泛素化活性;腺苷酸化活性;脱腺苷酸化活性;SUMO化活性;脱SUMO化活性;核糖基化活性;脱核糖基化活性;豆蔻酰化活性和脱豆蔻酰化活性。

[0226] 合适的融合配偶体的另外的实例是二氢叶酸还原酶(DHFR)去稳定化结构域(例如,以产生化学可控的融合Cas12L蛋白)和叶绿体转运肽。合适的叶绿体转运肽包括但不限于:

[0227] MASMISSSAVTTVSRASRGQSAAMAPFGGLKSMTGFPVRKVNTDITSITSNGGRVKCMQVWPPIGKKKFETLSYLPPLTRDSRA(SEQ ID NO:29);

[0228] MASMISSSAVTTVSRASRGQSAAMAPFGGLKSMTGFPVRKVNTDITSITSNGGRVKS(SEQ ID NO:30);

[0229] MASSMLSSATMVASPAQATMVAPFNGLKSSAAFPATRKANNDITSITSNGGRVNCMQVWPPIEKKKFETLSYLPDLTDSGGRVNC(SEQ ID NO:31);

[0230] MAQVSRICNGVQNPSLISNLSKSSQRKSPLSVSLKTQQHPRAYPISSSWGLKKSGMTLIGSELRPLKVMSSVSTAC(SEQ ID NO:32);

[0231] MAQVSRICNGVWNPSLISNLSKSSQRKSPLSVSLKTQQHPRAYPISSSWGLKKSGMTLIGSELRPLKVMSSVSTAC(SEQ ID NO:33);

[0232] MAQINNMAQGIQTLNPNSNFHKPQVPKSSSFLVFGSKKLKNSANSMLVLKKDSIFMQLFCSFRISASVATAC(SEQ ID NO:34);

[0233] MAALVTSQLATSGTVLSVTDRFRRPGFQGLRPRNPADAALGMRTVGASAAPKQSRKPHRFDRRCLSMVV(SEQ ID NO:35)；

[0234] MAALTTSQLATSATGFGIADRSAPSSLLRHGFQGLKPRSPAGGDATSLSVTTSARATPKQQRSVQRGSRRFPSVVVC(SEQ ID NO:36)；

[0235] MASSVLSSAAVATRSNVAQANMVAPFTGLKSAASFPVSRKQNLDITSIASNGGRVQC(SEQ ID NO:37)；

[0236] MESLAATSVFAPSRVAVPAARALVRAGTVVPTRRTSSTSGTSGVKCSAAVTPQASPVISRSAAAA(SEQ ID NO:38)；以及

[0237] MGAAATSMQSLKFSNRLVPPSRRLSPVPNNVTCNNLPKSAAPVRTVKCCASSWNSTINGAAATTNGASAASS(SEQ ID NO:39)。

[0238] 在一些情况下，本公开的Cas12L融合多肽包含：a)本公开的Cas12L多肽；和b)叶绿体转运肽。因此，例如，Cas12L多肽/引导RNA复合物可以靶向叶绿体。在一些情况下，这种靶向可通过N末端延伸的存在来实现，所述N末端延伸称为叶绿体转运肽(CTP)或质体转运肽。如果表达的多肽要在植物质体(例如，叶绿体)中区室化，则来自细菌来源的染色体转基因必须具有编码CTP序列的序列，所述CTP序列与编码表达的多肽的序列融合。因此，外源多肽到叶绿体的定位通常1通过将编码CTP序列的多核苷酸序列与编码外源多肽的多核苷酸的5'区可操作地连接来实现。在易位到质体的过程中，在加工步骤中去除CTP。然而，加工效率可能受到CTP的氨基酸序列和肽的氨基末端(NH_2末端)附近的序列的影响。已经描述的用于靶向叶绿体的其他选择是玉米cab-m7信号序列(美国专利号7,022,896、WO 97/41228)、豌豆谷胱甘肽还原酶信号序列(WO 97/41228)和US2009029861中描述的CTP。

[0239] 在一些情况下，本公开的Cas12L融合多肽可包含：a)本公开的Cas12L多肽；和b)内体逃逸肽。在一些情况下，内体逃逸多肽包含氨基酸序列GLFXALLXLLXSLWXLLLXA(SEQ ID NO:72)，其中每个X独立地选自赖氨酸、组氨酸和精氨酸。在一些情况下，内体逃逸多肽包含氨基酸序列GLFHALLHLLHSLWHLLLHA(SEQ ID NO:73)。

[0240] 对于在与Cas9蛋白、锌指蛋白和/或TALE蛋白融合的情况(用于位点特异性靶核酸修饰、转录调节和/或靶蛋白修饰，例如组蛋白修饰)中使用的一些上述融合配偶体(和更多)的实例，参见例如：Nomura等人，J Am Chem Soc.2007年7月18日；129(28):8676-7；Rivenbark等人，Epigenetics.2012年4月；7(4):350-60；Nucleic Acids Res.2016年7月8日；44(12):5615-28；Gilbert等人，Cell.2013年7月18日；154(2):442-51；Kearns等人，Nat Methods.2015年5月；12(5):401-3；Mendenhall等人，Nat Biotechnol.2013年12月；31(12):1133-6；Hilton等人，Nat Biotechnol.2015年5月；33(5):510-7；Gordley等人，Proc Natl Acad Sci U S A.2009年3月31日；106(13):5053-8；Akopian等人，Proc Natl Acad Sci U S A.2003年7月22日；100(15):8688-91；Tan等人，J Virol.2006年2月；80(4):1939-48；Tan等人，Proc Natl Acad Sci U S A.2003年10月14日；100(21):11997-2002；Papworth等人，Proc Natl Acad Sci U S A.2003年2月18日；100(4):1621-6；Sanjana等人，Nat Protoc.2012年1月5日；7(1):171-92；Beerli等人，Proc Natl Acad Sci U S A.1998年12月8日；95(25):14628-33；Snowden等人，Curr Biol.2002年12月23日；12(24):2159-66；Xu等人，Xu等人，Cell Discov.2016年5月3日；2:16009；Komor等人，Nature.2016年4月20日；533(7603):420-4；Chaikind等人，Nucleic Acids Res.2016年8月11日；

Choudhury等人,Oncotarget.2016年6月23日;Du等人,Cold Spring Harb Protoc.2016年1月4日;Pham等人,Methods Mol Biol.2016;1358:43-57;Balboa等人,Stem Cell Reports.2015年9月8日;5(3):448-59;Hara等人,Sci Rep.2015年6月9日;5:11221;Piatek等人,Plant Biotechnol J.2015年5月;13(4):578-89;Hu等人,Nucleic Acids Res.2014年4月;42(7):4375-90;Cheng等人,Cell Res.2013年10月;23(10):1163-71;以及Maeder等人,Nat Methods.2013年10月;10(10):977-9。

[0241] 另外适合的异源多肽包括但不限于直接和/或间接提供靶核酸的增加的转录和/或翻译的多肽(例如,转录激活因子或其片段、募集转录激活因子的蛋白质或其片段、小分子/药物响应性转录和/或翻译调控因子、翻译调控蛋白等)。实现增加或降低的转录的异源多肽的非限制性实例包括转录激活因子结构域和转录阻遏子结构域。在一些此类情况下,融合Cas12L多肽通过引导核酸(引导RNA)被靶向靶核酸中的特定位置(即,序列)并且发挥基因座特异性调控的作用,诸如阻断RNA聚合酶与启动子(所述启动子选择性抑制转录激活因子功能)的结合和/或修饰局部染色质状态(例如,在使用融合序列时,修饰靶核酸或修饰与靶核酸相关联的多肽)。在一些情况下,变化是瞬时的(例如,转录阻遏或激活)。在一些情况下,变化是可遗传的(例如,在对靶核酸或与靶核酸相关联的蛋白质(例如,核小体组蛋白)进行表观遗传修饰时)。

[0242] 当靶向ssRNA靶核酸时,使用的异源多肽的非限制性实例包括(但不限于):剪接因子(例如,RS结构域);蛋白质翻译组分(例如,翻译起始因子、延伸因子和/或释放因子;例如,eIF4G);RNA甲基化酶;RNA编辑酶(例如,RNA脱氨酶,例如作用于RNA的腺苷脱氨酶(ADAR),包括A至I和/或C至U编辑酶);解旋酶;RNA结合蛋白等。应理解,异源多肽可包括整个蛋白质,或者在一些情况下,可包括蛋白质的片段(例如,功能结构域)。

[0243] 主题融合Cas12L多肽的异源多肽可以是能够与ssRNA(出于本公开的目的,其包括分子内和/或分子间二级结构,例如双链RNA双链体,诸如发夹、茎环等)相互作用的任何结构域,无论是瞬时的还是不可逆的,直接的还是间接的,所述结构域包括但不限于选自由以下组成的组的效应结构域;内切核酸酶(例如RNA酶III、CRR22 DYW结构域、来自诸如SMG5和SMG6的蛋白质的Dicer和PIN(PilT N末端)结构域);负责刺激RNA切割的蛋白质和蛋白质结构域(例如CPSF、CstF、CFIm和CFIIm);外切核酸酶(例如XRN-1或外切核酸酶T);脱腺苷酶(例如HNT3);负责无义介导的RNA衰变的蛋白质和蛋白质结构域(例如UPF1、UPF2、UPF3、UPF3b、RNP S1、Y14、DEK、REF2和SRm160);负责稳定RNA的蛋白质和蛋白质结构域(例如PABP);负责阻遏翻译的蛋白质和蛋白质结构域(例如Ago2和Ago4);负责刺激翻译的蛋白质和蛋白质结构域(例如Staufen);负责(例如能够)调节翻译的蛋白质和蛋白质结构域(例如翻译因子,诸如起始因子、延伸因子、释放因子等,例如eIF4G);负责RNA的聚腺苷酸化的蛋白质和蛋白质结构域(例如PAP1、GLD-2和Star-PAP);负责RNA的聚尿苷酸化的蛋白质和蛋白质结构域(例如CI D1和末端尿苷酸转移酶);负责RNA定位的蛋白质和蛋白质结构域(例如来自IMP1、ZBP1、She2p、She3p和Bicaudal-D);负责RNA的核保留的蛋白质和蛋白质结构域(例如Rrp6);负责RNA的核输出的蛋白质和蛋白质结构域(例如TAP、NXF1、THO、TREX、REF和Aly);负责阻遏RNA剪接的蛋白质和蛋白质结构域(例如PTB、Sam68和hnRNP A1);负责刺激RNA剪接的蛋白质和蛋白质结构域(例如富含丝氨酸/精氨酸(SR)结构域);负责降低转录效率的蛋白质和蛋白质结构域(例如FUS(TLS));以及负责刺激转录的蛋白质和蛋白质结构

域(例如CDK7和HIV Tat)。可替代地,效应结构域可选自包括以下的组:内切核酸酶;能够刺激RNA切割的蛋白质和蛋白质结构域;外切核酸酶;脱腺苷酶;具有无义介导的RNA衰变活性的蛋白质和蛋白质结构域;能够稳定RNA的蛋白质和蛋白质结构域;能够阻遏翻译的蛋白质和蛋白质结构域;能够刺激翻译的蛋白质和蛋白质结构域;能够调节翻译的蛋白质和蛋白质结构域(例如,翻译因子,诸如起始因子、延伸因子、释放因子等,例如eIF4G);能够进行RNA的聚腺苷酸化的蛋白质和蛋白质结构域;能够进行RNA的聚尿苷酸化的蛋白质和蛋白质结构域;具有RNA定位活性的蛋白质和蛋白质结构域;能够进行RNA的核保留的蛋白质和蛋白质结构域;具有RNA核输出活性的蛋白质和蛋白质结构域;能够阻遏RNA剪接的蛋白质和蛋白质结构域;能够刺激RNA剪接的蛋白质和蛋白质结构域;能够降低转录效率的蛋白质和蛋白质结构域;以及能够刺激转录的蛋白质和蛋白质结构域。另一种合适的异源多肽是PUF RNA结合结构域,其在WO2012068627中更详细地描述,所述文献以引用方式整体并入本文。

[0244] 可作为融合Cas12L多肽的异源多肽(整体或作为其片段)使用的一些RNA剪接因子具有模块化结构,具有分开的序列特异性RNA结合模块和剪接效应结构域。例如,富含丝氨酸/精氨酸(SR)的蛋白质家族的成员含有N末端RNA识别基序(RRM),其结合前mRNA和C末端RS结构域中的外显子剪接增强子(ESE),所述外显子剪接增强子促进外显子包含。作为另一个实例,hnRNP蛋白hnRNP A1通过其RRM结构域与外显子剪接沉默子(ESS)结合,并通过C末端富含甘氨酸的结构域抑制外显子包含。一些剪接因子可通过结合两个替代位点之间的调控序列来调控剪接位点(ss)的替代使用。例如,ASF/SF2可识别ESE并有助于使用内含子近侧位点,而hnRNP A1可结合ESS并将剪接转到使用内含子远侧位点。此类因子的一个应用是生成调节内源基因(特别是疾病相关基因)的替代剪接的ESF。例如,Bcl-x前mRNA产生两种剪接同种型,这两种剪接同种型具有两个替代的5'剪接位点以编码具有相反功能的蛋白质。长剪接同种型Bcl-xL是在长寿命的有丝分裂后细胞中表达的有效凋亡抑制因子,并且在许多癌细胞中上调,从而保护细胞免于凋亡信号。短同种型Bcl-xS是促凋亡同种型,并且在具有高周转率的细胞(例如,发育中的淋巴细胞)中以高水平表达。两种Bcl-x剪接同种型之比由位于核心外显子区或外显子延伸区(即,两个替代5'剪接位点之间)中的多个cώ元件调控。对于更多实例,参见WO2010075303,其特此以引用方式整体并入。

[0245] 另外的合适的融合配偶体包括但不限于作为边界元件(例如,CTCF)的蛋白质(或其片段)、提供外周募集的蛋白质及其片段(例如,核纤层蛋白A、核纤层蛋白B等)、蛋白质对接元件(例如,FKBP/FRB、Pil1/Aby1等)。

[0246] 核酸酶

[0247] 在一些情况下,主题融合Cas12L多肽包含:i)本公开的Cas12L多肽;和ii)异源多肽("融合配偶体"),其中异源多肽是核酸酶。合适的核酸酶包括但不限于归巢核酸酶多肽;FokI多肽;转录激活因子样效应核酸酶(TALEN)多肽;MegaTAL多肽;大范围核酸酶多肽;锌指核酸酶(ZFN);ARCUS核酸酶;等等。大范围核酸酶可以由LADLIDADG归巢核酸内切酶(LHE)改造而来。megaTAL多肽可以包含TALE DNA结合结构域和工程化大范围核酸酶。参见例如WO2004/067736(归巢核酸内切酶);Urnov等人(2005年)Nature 435:646(ZFN);Mussolino等人(2011年)Nucle.Acids Res.39:9283(TALE核酸酶);Boissel等人(2013年)Nucl.Acids Res.42:2591(MegaTAL)。

[0248] 逆转录酶

[0249] 在一些情况下,主题融合Cas12L多肽包含:i)本公开的Cas12L多肽;和ii)异源多肽(“融合配偶体”),其中异源多肽是逆转录酶多肽。在某些情况下,Cas12L多肽没有催化活性。合适的逆转录酶包括例如鼠白血病病毒逆转录酶;劳斯肉瘤病毒逆转录酶;人免疫缺陷病毒I型逆转录酶;莫洛尼鼠白血病病毒逆转录酶;等等。

[0250] 碱基编辑器

[0251] 在一些情况下,本公开的Cas12L融合多肽包含:i)本公开的Cas12L多肽;和ii)异源多肽(“融合配偶体”),其中异源多肽是碱基编辑器。合适的碱基编辑包括例如腺苷脱氨酶;胞苷脱氨酶(例如,激活诱导的胞苷脱氨酶(AID));APOBEC3G等);等等。

[0252] 合适的腺苷脱氨酶是能够使DNA中的腺苷脱氨基的任何酶。在一些情况下,脱氨酶是TadA脱氨酶。

[0253] 在一些情况下,合适的腺苷脱氨酶包含与以下氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%氨基酸序列同一性的氨基酸序列:MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTD(SEQ ID NO:74)

[0254] 在一些情况下,合适的腺苷脱氨酶包含与以下氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%氨基酸序列同一性的氨基酸序列:MRRAFITGVFFLSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTD(SEQ ID NO:75)。

[0255] 在一些情况下,合适的腺苷脱氨酶包含与以下金黄色葡萄球菌TadA氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%氨基酸序列同一性的氨基酸序列:MGSHMTNDIYFMTLAIEEAKKAAQLGEVPIGAIITKDDEVIARAHNLRETLQQPTAHAEHIAIERAAKVLGSWRLEGCTLYVTLEPCVMCAGTIVMSRIPRVVYGADDPKGGCSGSLMNLLQQSNFNHRAIVDKGVLKEACSTLLTTFFK NLRANKKSTN:(SEQ ID NO:76)

[0256] 在一些情况下,合适的腺苷脱氨酶包含与以下枯草杆菌TadA氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%氨基酸序列同一性的氨基酸序列:MTQDELYMKEAIKEAKKAEEKGEVPIGAVLVINGEIIARAHNLRETEQRSIAHAEMLVIDEACKALGTWRLEGATLYVTLEPCPMCAGAVVLSRVEKVVFGAFDPKGGCSGTLMNLLQEERFNHQAEVVSGVLEEECGGMLSAFFRELRKKKKAARKNLSE(SEQ ID NO:77)

[0257] 在一些情况下,合适的腺苷脱氨酶包含与以下鼠伤寒沙门菌(Salmonella typhimurium)TadA具有至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%氨基酸序列同一性的氨基酸序列:MPPAFITGVTSLSDVELDHEYWMRHALTLAKRAWDEREVPVGAVLVHNHRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVLQNYRLLDTTLYVTLEPCVMCAGAMVHSRIGRVVFGARDAKTGAAGSLIDVLHHPGMNHRVEIIEGVLRDECATLLSDFFRMRRQEIKALKKADRAEGAGPAV(SEQ ID NO:78)

[0258] 在一些情况下,合适的腺苷脱氨酶包含与以下腐败希瓦菌TadA氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%氨基酸序列同一性的氨基酸序列:MDEYWMQVAMQMAEKAEAAGEVPVGAVLVKDGQQIATGYNLSISQHDPTAHAEILCLRSAGKKLE

NYRLLDATLYITLEPCAMCAGAMVHSRIARVVYGARDEKTGAAGTVVNLLQHPAFNHQVEVTSGVLAEACSAQLSR FFKRRRDEKKALKLAQRAQQGIE(SEQ ID NO:79)

[0259] 在一些情况下,合适的腺苷脱氨酶包含与以下流感嗜血杆菌F3031 TadA氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%氨基酸序列同一性的氨基酸序列:MDAAKVRSEFDEKMMRYALELADKAEALGEIPVGAVLVDDARNIIGEGWNLSIVQSD PTAHAEIIALRNGAKNIQNYRLLNSTLYVTLEPCTMCAGAILHSRIKRLVFGASDYKTGAIGSRFHFFDDYKMNHT LEITSGVLAEECSQKLS TFFQKRREEKKIEKALLKSLSDK(SEQ ID NO:80)

[0260] 在一些情况下,合适的腺苷脱氨酶包含与以下新月柄杆菌TadA氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%氨基酸序列同一性的氨基酸序列:MRTDESEDQDHRMMRLALDAARAAAEAGETPVGAVILDPSTGEVIATAGNGPIAAHDPTAHAEIA AMRAAAAKLGNYRLTDLTLVVTLEPCAMCAGAISHARIGRVVFGADDPKGGAVVHGPKFFAQPTCHWRPEVTGGVL ADESADLLRGFFRARRKAKI(SEQ ID NO:81)

[0261] 在一些情况下,合适的腺苷脱氨酶包含与以下硫还原地杆菌(Geobacter sulfurreducens)TadA氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%氨基酸序列同一性的氨基酸序列:MSSLKKTPIRDDAYWMGKAIREAAKAA ARDEVPIGAVIVRDGAVIGRGHNLREGSNDPSAHAEMIAIRQAARRSANWRLTGATLYVTLEPCLMCMGAIILARL ERVVFGCYDPKGGAAGSLYDLSADPRLNHQVRLSPGVCQEECGTMLSDFFRDLRRRKKAKATPALFIDERKVPPEP (SEQ ID NO:82)

[0262] 适合于包括在CRISPR/Cas效应子多肽融合多肽中的胞苷脱氨酶包括能够使DNA中的胞苷脱氨基的任何酶。

[0263] 在一些情况下,胞苷脱氨酶是来自脱氨酶的载脂蛋白B mRNA-编辑复合物(APOBEC)家族的脱氨酶。在一些情况下,APOBEC家族脱氨酶选自由以下构成的组:APOBEC1脱氨酶、APOBEC2脱氨酶、APOBEC3A脱氨酶、APOBEC3B脱氨酶、APOBEC3C脱氨酶、APOBEC3D脱氨酶、APOBEC3F脱氨酶、APOBEC3G脱氨酶和APOBEC3H脱氨酶。在一些情况下,胞苷脱氨酶是激活诱导的脱氨酶(AID)。

[0264] 在一些情况下,合适的胞苷脱氨酶包含与以下氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%氨基酸序列同一性的氨基酸序列:

[0265] MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGCHVELLFLRYISDW DLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLYFCEDRKAEPEGLRRLHRAGVQIAIMTFKDY FYCWNTFVENHERTFKAWEGLHENSVRLSRQLRRILLPLYEVDDLRDAFRTLGL(SEQ ID NO:83)

[0266] 在一些情况下,合适的胞苷脱氨酶是AID并且包含与以下氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%氨基酸序列同一性的氨基酸序列:MDSLLMNRRK FLYQFKNVRW AKGRRETYLC YVVKRRDSAT SFSLDFGYLR NKNGCHVELL FLRYISDWDL DPGRCYRVTW FTSWSPCYDC ARHVADFLRG NPNLSLRIFT ARLYFCEDRK AEPEGLRRLH RAGVQIAIMT FKENHERTFK AWEGLHENSV RLSRQLRRIL LPLYEVDDLR DAFRTLGL(SEQ ID NO:84)。

[0267] 在一些情况下,合适的胞苷脱氨酶是AID并且包含与以下氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%氨基酸序列同一性的氨基酸序列:MDSLLMNRRK FLYQFKNVRW AKGRRETYLC YVVKRRDSAT SFSLDFGYLR NKNGCHVELL

FLRYISDWDL DPGRCYRVTW FTSWSPCYDC ARHVADFLRG NPNLSLRIFT ARLYFCEDRK AEPEGLRRLH RAGVQIAIMT FKDYFYCWNT FVENHERTFK AWEGLHENSV RLSRQLRRIL LPLYEVDDLR DAFRTLGL (SEQ ID NO:85)。

[0268] 转录因子

[0269] 在一些情况下,本公开的Cas12L融合多肽包含:i)本公开的Cas12L多肽;和ii)异源多肽(“融合配偶体”),其中异源多肽是转录因子。转录因子可以包括:i)DNA结合结构域;和ii)转录激活因子。转录因子可以包括:i)DNA结合结构域;和ii)转录阻遏子。合适的转录因子包括包含转录激活因子或转录抑制子结构域的多肽(例如,Kruppel相关框(KRAB或SKD));Mad mSIN3相互作用结构域(SID);ERF阻遏结构域(ERD)等);基于锌指的人工转录因子(参见例如Sera(2009)Adv.Drug Deliv.61:513);基于TALE的人工转录因子(参见例如Liu等人(2013)Nat.Rev.Genetics 14:781);等等。在一些情况下,转录因子包括VP64多肽(转录激活)。在某些情况下,转录因子包含Krüppel相关盒(KRAB)多肽(转录抑制)。在一些情况下,转录因子包括Mad mSIN3相互作用结构域(SID)多肽(转录阻遏)。在一些情况下,转录因子包括ERF阻遏子结构域(ERD)多肽(转录阻遏)。例如,在一些情况下,转录因子是转录激活因子,其中转录激活因子是GAL4-VP16。

[0270] 重组酶

[0271] 在一些情况下,本公开的Cas12L融合多肽包含:i)本公开的Cas12L多肽;和ii)异源多肽(“融合配偶体”),其中异源多肽是重组酶。合适的重组酶包括例如Cre重组酶;Hin重组酶;Tre重组酶;FLP重组酶;等等。

[0272] 用于主题融合Cas12L多肽的各种另外的合适的异源多肽(或其片段)的实例包括但不限于在以下申请中描述的那些(所述公布涉及其他CRISPR内切核酸酶(诸如Cas9),但是描述的融合配偶体也可与Cas12L一起使用):PCT专利申请:WO2010075303、WO2012068627和WO2013155555,并且可见于例如以下美国专利和专利申请:8,906,616;8,895,308;8,889,418;8,889,356;8,871,445;8,865,406;8,795,965;8,771,945;8,697,359;20140068797;20140170753;20140179006;20140179770;20140186843;20140186919;20140186958;20140189896;20140227787;20140234972;20140242664;20140242699;20140242700;20140242702;20140248702;20140256046;20140273037;20140273226;20140273230;20140273231;20140273232;20140273233;20140273234;20140273235;20140287938;20140295556;20140295557;20140298547;20140304853;20140309487;20140310828;20140310830;20140315985;20140335063;20140335620;20140342456;20140342457;20140342458;20140349400;20140349405;20140356867;20140356956;20140356958;20140356959;20140357523;20140357530;20140364333;和20140377868;所述专利全部特此以引用方式整体并入。

[0273] 在一些情况下,异源多肽(融合配偶体)提供亚细胞定位,即异源多肽含有亚细胞定位序列(例如,用于靶向细胞核的核定位信号(NLS)、用于将融合蛋白保持在细胞核外的序列(例如核输出序列(NES))、用于将融合蛋白保留在细胞质中的序列、用于靶向线粒体的线粒体定位信号、用于靶向叶绿体的叶绿体定位信号、ER保留信号等)。在一些实施方案中,Cas12L融合多肽不包括NLS,使得蛋白质不靶向细胞核(这可能是有利的,例如,当靶核酸是存在于胞质溶胶中的RNA时)。在一些实施方案中,异源多肽可提供便于追踪和/或纯化的标

签(即,异源多肽是可检测标记)(例如,荧光蛋白,例如绿色荧光蛋白(GFP)、YFP、RFP、CFP、mCherry、tdTomato等;组氨酸标签,例如6XHis标签;血凝素(HA)标签;FLAG标签;Myc标签等)。

[0274] 在一些情况下,Cas12L蛋白(例如,野生型Cas12L蛋白、变体Cas12L蛋白、融合Cas12L蛋白、dCas12L蛋白等)包含(融合到)核定位信号(NLS)(例如,在一些情况下,2个或更多、3个或更多、4个或更多、或5个或更多个NLS)。因此,在一些情况下,Cas12L多肽包括一个或多个NLS(例如,2个或更多个、3个或更多个、4个或更多个或5个或更多个NLS)。在一些情况下,一个或多个NLS(2个或更多个、3个或更多个、4个或更多个或者5个或更多个NLS)定位在N末端和/或C末端处或附近(例如,在50个氨基酸内)。在一些情况下,一个或多个NLS(2个或更多个、3个或更多个、4个或更多个或者5个或更多个NLS)定位在N末端处或附近(例如,在50个氨基酸内)。在一些情况下,一个或多个NLS(2个或更多个、3个或更多个、4个或更多个或者5个或更多个NLS)定位在C末端处或附近(例如,在50个氨基酸内)。在一些情况下,一个或多个NLS(3个或更多个、4个或更多个或者5个或更多个NLS)定位在N末端和C末端二者处或附近(例如,在50个氨基酸内)。在一些情况下,NLS定位在N末端,并且NLS定位在C末端。

[0275] 在一些情况下,Cas12L蛋白(例如,野生型Cas12L蛋白、变体Cas12L蛋白、融合Cas12L蛋白、dCas12L蛋白等)包含(融合到)1至10个NLS(例如,1-9、1-8、1-7、1-6、1-5、2-10、2-9、2-8、2-7、2-6或2-5个NLS)。在一些情况下,Cas12L蛋白(例如,野生型Cas12L蛋白、变体Cas12L蛋白、融合Cas12L蛋白、dCas12L蛋白等)包含(融合到)2至5个NLS(例如,2-4个或2-3个NLS)。

[0276] NLS的非限制性实例包括衍生自以下的NLS序列:SV40病毒大T抗原的NLS,具有氨基酸序列PKKKRKV(SEQ ID NO:86);来自核质蛋白的NLS(例如,具有序列KRPAATKKAGQAKKKK(SEQ ID NO:87)的核质蛋白二分NLS);c-myc NLS,具有氨基酸序列PAAKRVKLD(SEQ ID NO:88)或RQRRNELKRSP(SEQ ID NO:89);hRNPA1 M9 NLS,具有序列NQSSNFGPMKGGNFGGRSSGPYGGGGQYFAKPRNQGGY(SEQ ID NO:90);来自输入蛋白-α的IBB结构域的序列RMRIZFKNKGKDTAELRRRRVEVSVELRKAKKDEQILKRRNV(SEQ ID NO:91);肌瘤T蛋白的序列VSRKRPRP(SEQ ID NO:92)和PPKKARED(SEQ ID NO:93);人p53的序列PQPKKKPL(SEQ ID NO:94);小鼠c-abl IV的序列SALIKKKKKMAP(SEQ ID NO:95);流感病毒NS1的序列DRLRR(SEQ ID NO:96)和PKQKKRK(SEQ ID NO:97);肝炎病毒δ抗原的序列RKLKKKIKKL(SEQ ID NO:98);小鼠Mx1蛋白的序列REKKKFLKRR(SEQ ID NO:99);人聚(ADP-核糖)聚合酶的序列KRKGDEVDGVDEVAKKKSKK(SEQ ID NO:100);以及类固醇激素受体(人)糖皮质激素的序列RKCLQAGMNLEARKTKK(SEQ ID NO:101)。一般来讲,NLS(或多个NLS)具有足够的强度来驱动Cas12L蛋白在真核细胞的细胞核中以可检测的量积累。可通过任何合适的技术执行细胞核中的积累的检测。例如,可检测标记物可与Cas12L蛋白融合,使得细胞内的位置可被可视化。也可从细胞中分离细胞核,然后可通过任何合适的检测蛋白质的方法(诸如免疫组织化学、蛋白质印迹或酶活性测定)分析细胞核的内容物。也可间接确定细胞核中的积累。

[0277] 在一些情况下,Cas12L融合多肽包含“蛋白转导结构域”或PTD(又称为CPP-细胞穿透肽),其是指促进横穿脂质双层、胶束、细胞膜、细胞器膜或囊泡膜的多肽、多核苷酸、碳水化合物或有机化合物或无机化合物。连接至另一个分子(所述分子可在小极性分子至大的

高分子和/或纳米颗粒的范围内)的PTD促进分子横穿膜,例如从细胞外空间进入细胞内空间或从胞质溶胶进入细胞器内。在一些实施方案中,PTD与多肽氨基末端共价连接(例如,与野生型Cas12L多肽连接以生成融合蛋白,或与变体Cas12L蛋白(诸如dCas12L、切口酶Cas12L或融合Cas12L蛋白)连接以生成融合蛋白)。在一些实施方案中,PTD与多肽的羧基末端共价连接(例如,与野生型Cas12L连接以生成融合蛋白,或与变体Cas12L蛋白(诸如dCas12L、切口酶Cas12L或融合Cas12L蛋白)连接以生成融合蛋白)。在一些情况下,PTD在合适的插入位点处内插在Cas12L融合多肽中(即,不在Cas12L融合多肽的N端或C端)。在一些情况下,主题Cas12L融合多肽包含(缀合至、融合至)一个或多个PTD(例如,两个或更多个、三个或更多个、四个或更多个PTD)。在一些情况下,PTD包括核定位信号(NLS)(例如,在一些情况下,2个或更多个、3个或更多个、4个或更多个或5个或更多个NLS)。因此,在一些情况下,Cas12L融合多肽包括一个或多个NLS(例如,2个或更多个、3个或更多个、4个或更多个或5个或更多个NLS)。在一些实施方案中,PTD共价连接至核酸(例如,Cas12L引导核酸、编码Cas12L引导核酸的多核苷酸、编码Cas12L融合多肽的多核苷酸、供体多核苷酸等)。PTD的实例包括但不限于最小十一氨基酸多肽蛋白转导结构域(对应于包含YGRKKRRQRRR;SEQ ID NO:40的HIV-1TAT的残基47-57);包含足以直接进入细胞中的多个精氨酸(例如,3个、4个、5个、6个、7个、8个、9个、10个或10-50个精氨酸)的聚精氨酸序列;VP22结构域(Zender等人(2002)Cancer Gene Ther.9(6):489-96);果蝇触角足基因(Antennapedia)蛋白转导结构域(Noguchi等人(2003)Diabetes 52(7):1732-1737);截短的人降钙素肽(Trehin等人(2004)Pharm.Research 21:1248-1256);聚赖氨酸(Wender等人(2000)Proc.Natl.Acad.Sci.USA 97:13003-13008);RRQRRTSKLMKR(SEQ ID NO:41);运输蛋白(Transportan)GWTLNSAGYLLGKINLKALAALAKKIL(SEQ ID NO:42);KALAWEAKLAKALAKALAKHLAKALAKALKCEA(SEQ ID NO:43);和RQIKIWFQNRRMKWKK(SEQ ID NO:44)。示例性PTD包括但不限于:YGRKKRRQRRR(SEQ ID NO:40);RKKRRQRRR(SEQ ID NO:45);具有3个精氨酸残基至50个精氨酸残基的精氨酸均聚物;示例性PTD结构域氨基酸序列包括但不限于以下序列中的任一个:YGRKKRRQRRR(SEQ ID NO:40);RKKRRQRR(SEQ ID NO:46);YARAAARQARA(SEQ ID NO:47);THRLPRRRRRR(SEQ ID NO:48);以及GGRRARRRRRR(SEQ ID NO:49)。在一些实施方案中,PTD是可激活的CPP(ACPP)(Aguilera等人(2009)Integr Biol(Camb)6月;1(5-6):371-381)。ACPP包括经由可切割接头连接至匹配聚阴离子(例如,Glu9或"E9")的聚阳离子CPP(例如,Arg9或"R9"),这使净电荷减小至接近零并由此抑制粘附和吸收到细胞中。当切割接头时,释放聚阴离子,局部暴露聚精氨酸和其固有的粘附性,从而"激活"ACPP以横穿膜。

[0278] 接头(例如,用于融合配偶体)

[0279] 在一些实施方案中,主题Cas12L蛋白可经由接头多肽(例如,一个或多个接头多肽)与融合配偶体融合。接头多肽可具有多种氨基酸序列中的任一种。蛋白质可通过间隔肽连接,间隔肽通常具有柔性性质,但不排除其他化学键。合适的接头包括长度在4个氨基酸与40个氨基酸之间或者长度在4个氨基酸与25个氨基酸之间的多肽。这些接头可通过使用合成的编码接头的寡核苷酸来产生以偶联蛋白质,或者可由编码融合蛋白的核酸序列编码。可使用具有一定程度柔性的肽接头。连接肽实际上可具有任何氨基酸序列,应记住优选的接头将具有产生总体上柔性的肽的序列。小氨基酸(诸如甘氨酸和丙氨酸)的用途用于产生柔性肽。对于本领域技术人员来说,产生此类序列是常规的。多种不同的接头是可商购获

得的并且被认为是适合使用的。

[0280] 接头多肽的实例包括甘氨酸聚合物(G)$_n$、甘氨酸-丝氨酸聚合物(包括例如(GS)$_n$、GSGGS$_n$(SEQ ID NO:50)、GGSGGS$_n$(SEQ ID NO:51)和GGGS$_n$(SEQ ID NO:52),其中n是至少为1的整数)、甘氨酸-丙氨酸聚合物、丙氨酸-丝氨酸聚合物。示例性接头可包含氨基酸序列,所述氨基酸序列包括但不限于GGSG(SEQ ID NO:53)、GGSGG(SEQ ID NO:54)、GSGSG(SEQ ID NO:55)、GSGGG(SEQ ID NO:56)、GGGSG(SEQ ID NO:57)、GSSSG(SEQ ID NO:58)等。普通技术人员将认识到,与任何所需元件缀合的肽的设计可包括全部或部分柔性的接头,使得接头可包括柔性接头以及赋予较少柔性结构的一个或多个部分。

[0281] 可检测标记

[0282] 在一些情况下,本公开的Cas12L多肽包含可检测标记。可提供可检测信号的合适的可检测标记和/或部分可包括但不限于酶、放射性同位素、特异性结合对的成员、荧光团、荧光蛋白、量子点等。

[0283] 合适的荧光蛋白包括但不限于绿色荧光蛋白(GFP)或其变体、GFP的蓝色荧光变体(BFP)、GFP的青色荧光变体(CFP)、GFP的黄色荧光变体(YFP)、增强型GFP(EGFP)、增强型CFP(ECFP)、增强型YFP(EYFP)、GFPS65T、Emerald、Topaz(TYFP)、Venus、Citrine、mCitrine、GFPuv、去稳定化EGFP(dEGFP)、去稳定化ECFP(dECFP)、去稳定化EYFP(dEYFP)、mCFPm、Cerulean、T-Sapphire、CyPet、YPet、mKO、HcRed、t-HcRed、DsRed、DsRed2、DsRed-单体、J-Red、二聚体2、t-二聚体2(12)、mRFP1、普奇洛孔蛋白(pocilloporin)、海肾GFP(Renilla GFP)、Monster GFP、paGFP、Kaede蛋白和点燃蛋白(kindling protein)、藻胆蛋白和藻胆蛋白缀合物(包括B-藻红蛋白、R-藻红蛋白和别藻蓝蛋白)。荧光蛋白的其他实例包括mHoneydew、mBanana、mOrange、dTomato、tdTomato、mTangerine、mStrawberry、mCherry、mGrape1、mRaspberry、mGrape2、mPlum(Shaner等人(2005)Nat.Methods 2:905-909)等。如在例如Matz等人(1999)Nature Biotechnol.17:969-973中所述的来自珊瑚虫物种的多种荧光蛋白和有色蛋白中的任一种是适合使用的。

[0284] 合适的酶包括但不限于辣根过氧化物酶(HRP)、碱性磷酸酶(AP)、β-半乳糖苷酶(GAL)、葡萄糖-6-磷酸脱氢酶、β-N-乙酰氨基葡糖苷酶、β-葡糖醛酸糖苷酶、转化酶、黄嘌呤氧化酶、萤火虫荧光素酶、葡萄糖氧化酶(GO)等。

[0285] 原间隔序列相邻基序(PAM)

[0286] Cas12L蛋白在由靶向DNA的RNA与靶DNA之间的互补性区域限定的靶序列处与靶DNA结合。与许多CRISPR内切核酸酶的情况一样,双链靶DNA的位点特异性结合(和/或切割)发生在由以下二者确定的位置处:(i)引导RNA与靶DNA之间的碱基配对互补性;和(ii)靶DNA中的短基序[称为原间隔序列相邻基序(PAM)]。

[0287] 在一些情况下,Cas12L蛋白的PAM紧接靶DNA的非互补链的靶序列的5’(互补链:(i)与引导RNA的引导序列杂交,而非互补链不与引导RNA直接杂交;并且(ii)是非互补链的反向互补链)。

[0288] 在一些情况下,不同的Cas12L蛋白(即,来自各种物种的Cas12L蛋白)可有利地用于各种所提供的方法中以便利用不同的Cas12L蛋白的各种酶特征(例如,用于不同PAM序列偏好;用于增加的或降低的酶促活性;用于增加的或降低的细胞毒性水平;用于改变NHEJ、同源定向修复、单链断裂、双链断裂等之间的平衡;利用短的总序列等)。来自不同物种的

Cas12L蛋白可能需要靶DNA中的不同的PAM序列。用于鉴定适当的PAM序列的各种方法(包括计算机模拟方法和/或湿实验室方法(wet lab methods))是本领域已知且常规的,并且可使用任何便利的方法。

[0289] 如实施例1中所示,本公开的Cal12L多肽可被重新编程(通过与引导RNA复合)以切割与引导RNA的靶向片段互补的靶核酸(例如靶DNA)的任何序列,其中PAM存在于靶的5'端(例如Casλ1的富含T的PAM);体内功能效应物的形成不需要额外的RNA组分。在一些情况下,PAM序列是富含T的序列(例如,TTR,其中R是嘌呤)。在某些情况下,PAM序列是TTA。在某些情况下,PAM序列是TTG。

[0290] Cas12L引导RNA

[0291] 与Cas12L蛋白结合形成核糖核蛋白复合物(RNP)并将复合物靶向靶核酸(例如,靶DNA)内的特定位置的核酸在本文中称为"Cas12L引导RNA"或者简称为"引导RNA"。应理解,在一些情况下,可制备杂交体DNA/RNA,使得Cas12L引导RNA除RNA碱基外还包含DNA碱基,但术语"Cas12L引导RNA"仍然用于涵盖本文的这种分子。

[0292] 可以说Cas12L引导RNA包含两个区段,即靶向区段和蛋白质结合区段。蛋白质结合区段在本文中也称为引导RNA的"恒定区"。Cas12L引导RNA的靶向区段包含与靶核酸(例如,靶dsDNA、靶ssRNA、靶ssDNA、双链靶DNA的互补链等)内的特定序列(靶位点)互补(并因此杂交)的核苷酸序列(引导序列)。蛋白质结合区段(或"蛋白质结合序列")与Cas12L多肽相互作用(结合)。主题Cas12L引导RNA的蛋白质结合区段可包含彼此杂交以形成双链RNA双链体(dsRNA双链体)的两段互补核苷酸。靶核酸(例如,基因组DNA、dsDNA、RNA等)的位点特异性结合和/或切割可发生在由Cas12L引导RNA(Cas12L引导RNA的引导序列)与靶核酸之间的碱基配对互补性确定的位置(例如,靶基因座的靶序列)处。

[0293] Cas12L引导RNA和Cas12L蛋白(例如,野生型Cas12L蛋白;变体Cas12L蛋白;融合Cas12L多肽;等)形成复合物(例如,通过非共价相互作用结合)。Cas12L引导RNA通过包含靶向区段为复合物提供靶特异性,所述靶向区段包含引导序列(与靶核酸序列互补的核苷酸序列)。复合物的Cas12L蛋白提供位点特异性活性(例如,由Cas12L蛋白提供的切割活性和/或在融合Cas12L蛋白的情况下由融合配偶体提供的活性)。换句话讲,Cas12L蛋白由于其与Cas12L引导RNA的缔合而被导向至靶核酸序列(例如,靶序列)。

[0294] 可修饰"引导序列",也称为Cas12L引导RNA的"靶向序列",使得Cas12L引导RNA可将Cas12L蛋白(例如,天然存在的Cas12L蛋白、融合Cas12L多肽等)靶向任何所需的靶核酸的任何所需序列,除了(例如,如本文所述)可考虑PAM序列之外。因此,例如,Cas12L引导RNA可具有与真核细胞中的核酸中的序列互补(例如,可与其杂交)的引导序列,所述核酸例如是病毒核酸、真核核酸(例如,真核染色体、染色体序列、真核RNA等)等。

[0295] Cas12L引导RNA的引导序列

[0296] 主题Cas12L引导RNA包含引导序列(即,靶向序列),其是与靶核酸中的序列(靶位点)互补的核苷酸序列。换句话讲,Cas12L引导RNA的引导序列可通过杂交(即,碱基配对)以序列特异性方式与靶核酸(例如,双链DNA(dsDNA)、单链DNA(ssDNA)、单链RNA(ssRNA)或双链RNA(dsRNA))相互作用。Cas12L引导RNA的引导序列可被修饰(例如,通过遗传工程)/设计成与靶核酸(例如,真核靶核酸,诸如基因组DNA)内的任何所需靶序列杂交(例如,当考虑PAM时,例如,当靶向dsDNA靶时)。

[0297] 在一些情况下，引导序列与靶核酸的靶位点之间的互补性百分比为60％或更高(例如，65％或更高、70％或更高、75％或更高、80％或更高、85％或更高、90％或更高、95％或更高、97％或更高、98％或更高、99％或更高或者100％)。在一些情况下，指导序列与靶核酸的靶位点之间的互补性百分比为80％或更高(例如，85％或更高、90％或更高、95％或更高、97％或更高、98％或更高、99％或更高或者100％)。在一些情况下，指导序列与靶核酸的靶位点之间的互补性百分比为90％或更高(例如，95％或更高、97％或更高、98％或更高、99％或更高或者100％)。在一些情况下，指导序列与靶核酸的靶位点之间的互补性百分比为100％。

[0298] 在一些情况下，指导序列与靶核酸的靶位点之间的互补性百分比在靶核酸的靶位点最3’端的七个连续核苷酸上为100％。

[0299] 在一些情况下，指导序列与靶核酸的靶位点之间的互补性百分比在17个或更多个(例如，18个或更多个、19个或更多个、20个或更多个、21个或更多个、22个或更多个)连续核苷酸上为60％或更高(例如，70％或更高、75％或更高、80％或更高、85％或更高、90％或更高、95％或更高、97％或更高、98％或更高、99％或更高或者100％)。在一些情况下，指导序列与靶核酸的靶位点之间的互补性百分比在17个或更多个(例如，18个或更多个、19个或更多个、20个或更多个、21个或更多个、22个或更多个)连续核苷酸上为80％或更高(例如，85％或更高、90％或更高、95％或更高、97％或更高、98％或更高、99％或更高或者100％)。在一些情况下，指导序列与靶核酸的靶位点之间的互补性百分比在17个或更多个(例如，18个或更多个、19个或更多个、20个或更多个、21个或更多个、22个或更多个)连续核苷酸上为90％或更高(例如，95％或更高、97％或更高、98％或更高、99％或更高或者100％)。在一些情况下，指导序列与靶核酸的靶位点之间的互补性百分比在17个或更多个(例如，18个或更多个、19个或更多个、20个或更多个、21个或更多个、22个或更多个)连续核苷酸上为100％。

[0300] 在一些情况下，指导序列与靶核酸的靶位点之间的互补性百分比在19个或更多个(例如，20个或更多个、21个或更多个、22个或更多个)连续核苷酸上为60％或更高(例如，70％或更高、75％或更高、80％或更高、85％或更高、90％或更高、95％或更高、97％或更高、98％或更高、99％或更高或者100％)。在一些情况下，指导序列与靶核酸的靶位点之间的互补性百分比在19个或更多个(例如，20个或更多个、21个或更多个、22个或更多个)连续核苷酸上为80％或更高(例如，85％或更高、90％或更高、95％或更高、97％或更高、98％或更高、99％或更高或者100％)。在一些情况下，指导序列与靶核酸的靶位点之间的互补性百分比在19个或更多个(例如，20个或更多个、21个或更多个、22个或更多个)连续核苷酸上为90％或更高(例如，95％或更高、97％或更高、98％或更高、99％或更高或者100％)。在一些情况下，指导序列与靶核酸的靶位点之间的互补性百分比在19个或更多个(例如，20个或更多个、21个或更多个、22个或更多个)连续核苷酸上为100％。

[0301] 在一些实施方案中，指导序列与靶核酸的靶位点之间的互补性百分比在17-25个连续核苷酸上为60％或更高(例如，70％或更高、75％或更高、80％或更高、85％或更高、90％或更高、95％或更高、97％或更高、98％或更高、99％或更高或者100％)。在一些情况下，指导序列与靶核酸的靶位点之间的互补性百分比在17-25个连续核苷酸上为80％或更高(例如，85％或更高、90％或更高、95％或更高、97％或更高、98％或更高、99％或更高或者100％)。在一些情况下，指导序列与靶核酸的靶位点之间的互补性百分比在17-25个连续核

苷酸上为90％或更高(例如,95％或更高、97％或更高、98％或更高、99％或更高或者100％)。在一些情况下,指导序列与靶核酸的靶位点之间的互补性百分比在17-25个连续核苷酸上为100％。

[0302] 在一些实施方案中,指导序列与靶核酸的靶位点之间的互补性百分比在19-25个连续核苷酸上为60％或更高(例如,70％或更高、75％或更高、80％或更高、85％或更高、90％或更高、95％或更高、97％或更高、98％或更高、99％或更高或者100％)。在一些情况下,指导序列与靶核酸的靶位点之间的互补性百分比在19-25个连续核苷酸上为80％或更高(例如,85％或更高、90％或更高、95％或更高、97％或更高、98％或更高、99％或更高或者100％)。在一些情况下,指导序列与靶核酸的靶位点之间的互补性百分比在19-25个连续核苷酸上为90％或更高(例如,95％或更高、97％或更高、98％或更高、99％或更高或者100％)。在一些情况下,指导序列与靶核酸的靶位点之间的互补性百分比在19-25个连续核苷酸上为100％。

[0303] 在一些情况下,指导序列具有在17-30个核苷酸(nt)(例如,17-25个、17-22个、17-20个、19-30个、19-25个、19-22个、19-20个、20-30个、20-25个或20-22个nt)的范围内的长度。在一些情况下,指导序列具有在17-25个核苷酸(nt)(例如,17-22个、17-20个、19-25个、19-22个、19-20个、20-25个或20-22个nt)的范围内的长度。在一些情况下,指导序列具有17或更多个nt(例如,18个或更多个、19个或更多个、20个或更多个、21个或更多个或者22个或更多个nt;19个nt、20个nt、21个nt、22个nt、23个nt、24个nt、25个nt等)的长度。在一些情况下,指导序列具有19个或更多个nt(例如,20个或更多个、21个或更多个或者22个或更多个nt;19个nt、20个nt、21个nt、22个nt、23个nt、24个nt、25个nt等)的长度。在一些情况下,指导序列具有17个nt的长度。在一些情况下,指导序列具有18个nt的长度。在一些情况下,引导序列具有19个nt的长度。在一些情况下,引导序列具有20个nt的长度。在一些情况下,引导序列具有21个nt的长度。在一些情况下,引导序列具有22个nt的长度。在一些情况下,引导序列具有23个nt的长度。

[0304] 在一些情况下,引导序列(也称为“间隔序列”)具有15至50个核苷酸的长度(例如,15个核苷酸(nt)至20nt、20nt至25nt、25nt至30nt、30nt至35nt、35nt至40nt、40nt至45nt或45nt至50nt)。

[0305] Cas12L引导RNA的蛋白质结合区段

[0306] 主题Cas12L引导RNA的蛋白结合片段(“恒定区”)与Cas12L蛋白相互作用。Cas12L引导RNA通过上文提及的引导序列将结合的Cas12L蛋白导向至靶核酸内的特定核苷酸序列。Cas12L引导RNA的蛋白质结合区段可包含两段核苷酸,它们彼此互补并杂交形成双链RNA双链体(dsRNA双链体)。因此,在一些情况下,蛋白质结合区段包括dsRNA双链体。

[0307] 在一些情况下,dsRNA双链体区包含5-25个碱基对(bp)的范围(例如,5-22个、5-20个、5-18个、5-15个、5-12个、5-10个、5-8个、8-25个、8-22个、8-18个、8-15个、8-12个、12-25个、12-22个、12-18个、12-15个、13-25个、13-22个、13-18个、13-15个、14-25个、14-22个、14-18个、14-15个、15-25个、15-22个、15-18个、17-25个、17-22个或17-18个bp,例如5个bp、6个bp、7个bp、8个bp、9个bp、10个bp等)。在一些情况下,dsRNA双链体区包含6-15个碱基对(bp)的范围(例如,6-12个、6-10个或6-8个bp,例如6个bp、7个bp、8个bp、9个bp、10个bp等)。在一些情况下,双链体区包含5个或更多个bp(例如,6个或更多个、7个或更多个或者8个或

更多个bp)。在一些情况下,双链体区包含6个或更多个bp(例如,7个或更多个或者8个或更多个bp)。在一些情况下,并非双链体区的所有核苷酸都是配对的,并且因此双链体形成区域可包含凸起。本文中的术语"凸起"用于意指一段核苷酸(其可以是一个核苷酸),这段核苷酸对双链双链体没有贡献,但是在5'端和3'端被有贡献的核苷酸围绕,并且因此凸起被认为是双链体区的一部分。在一些情况下,dsRNA包含1个或更多个凸起(例如,2个或更多个、3个或更多个、4个或更多个凸起)。在一些情况下,dsRNA双链体包含2个或更多个凸起(例如,3个或更多个、4个或更多个凸起)。在一些情况下,dsRNA双链体包含1-5个凸起(例如,1-4个、1-3个、2-5个、2-4个或2-3个凸起)。

[0308] 因此,在一些情况下,彼此杂交形成dsRNA双链体的核苷酸段彼此具有70%-100%的互补性(例如,75%-100%、80%-10%、85%-100%、90%-100%、95%-100%的互补性)。在一些情况下,彼此杂交形成dsRNA双链体的核苷酸段彼此具有70%-100%的互补性(例如,75%-100%、80%-10%、85%-100%、90%-100%、95%-100%的互补性)。在一些情况下,彼此杂交形成dsRNA双链体的核苷酸段彼此具有85%-100%的互补性(例如,90%-100%、95%-100%的互补性)。在一些情况下,彼此杂交形成dsRNA双链体的核苷酸段彼此具有70%-95%的互补性(例如,75%-95%、80%-95%、85%-95%、90%-95%的互补性)。

[0309] 换句话讲,在一些实施方案中,dsRNA双链体包含彼此具有70%-100%的互补性(例如,75%-100%、80%-10%、85%-100%、90%-100%、95%-100%的互补性)的两段核苷酸。在一些情况下,dsRNA双链体包含彼此具有85%-100%的互补性(例如,90%-100%、95%-100%的互补性)的两段核苷酸。在一些情况下,dsRNA双链体包含彼此具有70%-95%的互补性(例如,75%-95%、80%-95%、85%-95%、90%-95%的互补性)的两段核苷酸。

[0310] 相对于天然存在的双链体区,主题Cas12L引导RNA的双链体区可包含一个或多个(1个、2个、3个、4个、5个等)突变。例如,在一些情况下,可维持碱基对,同时对每个区段的碱基对有贡献的核苷酸可以是不同的。在一些情况下,与(天然存在的Cas12L引导RNA的)天然存在的双链体区相比,主题Cas12L引导RNA的双链体区包含更多配对的碱基、更少配对的碱基、更小的凸起、更大的凸起、更少的凸起、更多的凸起或它们的任何方便的组合。

[0311] 各种Cas9引导RNA的实例可在本领域中找到,并且在一些情况下,与引入Cas9引导RNA中的那些相似的变型也可引入本公开的Cas12L引导RNA中(例如,dsRNA双链体区的突变、5'或3'末端的延伸以用于增加稳定性以提供与另一种蛋白质的相互作用,等)。例如,参见Jinek等人,Science.2012年8月17日;337(6096):816-21;Chylinski等人,RNA Biol.2013年5月;10(5):726-37;Ma等人,Biomed Res Int.2013;2013:270805;Hou等人,Proc Natl Acad Sci U S A.2013年9月24日;110(39):15644-9;Jinek等人,Elife.2013;2:e00471;Pattanayak等人,Nat Biotechnol.2013年9月;31(9):839-43;Qi等人,Cell.2013年2月28日;152(5):1173-83;Wang等人,Cell.2013年5月9日;153(4):910-8;Auer等人,Genome Res.2013年10月31日;Chen等人,Nucleic Acids Res.2013年11月1日;41(20):e19;Cheng等人,Cell Res.2013年10月;23(10):1163-71;Cho等人,Genetics.2013年11月;195(3):1177-80;DiCarlo等人,Nucleic Acids Res.2013年4月;41(7):4336-43;Dickinson等人,Nat Methods.2013年10月;10(10):1028-34;Ebina等人,Sci Rep.2013;3:2510;Fujii等人,Nucleic Acids Res.2013年11月1日;41(20):e187;Hu等人,Cell Res.2013年11月;23(11):1322-5;Jiang等人,Nucleic Acids Res.2013年11月1日;41

(20):e188;Larson等人,Nat Protoc.2013年11月;8(11):2180-96;Mali等人,Nat Methods.2013年10月;10(10):957-63;Nakayama等人,Genesis.2013年12月;51(12):835-43;Ran等人,Nat Protoc.2013年11月;8(11):2281-308;Ran等人,Cell.2013年9月12日;154(6):1380-9;Upadhyay等人,G3(Bethesda).2013年12月9日;3(12):2233-8;Walsh等人,Proc Natl Acad Sci U S A.2013年9月24日;110(39):15514-5;Xie等人,Mol Plant.2013年10月9日;Yang等人,Cell.2013年9月12日;154(6):1370-9;Briner等人,Mol Cell.2014年10月23日;56(2):333-9;以及以下美国专利和专利申请:8,906,616;8,895,308;8,889,418;8,889,356;8,871,445;8,865,406;8,795,965;8,771,945;8,697,359;20140068797;20140170753;20140179006;20140179770;20140186843;20140186919;20140186958;20140189896;20140227787;20140234972;20140242664;20140242699;20140242700;20140242702;20140248702;20140256046;20140273037;20140273226;20140273230;20140273231;20140273232;20140273233;20140273234;20140273235;20140287938;20140295556;20140295557;20140298547;20140304853;20140309487;20140310828;20140310830;20140315985;20140335063;20140335620;20140342456;20140342457;20140342458;20140349400;20140349405;20140356867;20140356956;20140356958;20140356959;20140357523;20140357530;20140364333;和20140377868;所述文献全部特此以引用方式整体并入。

[0312] 图2A-2Z、图2AA-2ZZ和图2AAA-2FFF中提供了适于包含在Cas12L引导RNA中的恒定区的实例(例如,其中T被U取代)。与图2A-2Z、图2AA-2ZZ和图2AAA-2FFF中描绘的核苷酸序列中任一者相比,Cas12L引导RNA可以包含具有1至5个核苷酸取代的恒定区。

[0313] 核苷酸序列(其中T被U取代)可以与选择的间隔序列(其中间隔序列包含靶核酸结合序列(“引导序列”))组合,所述间隔序列为15至50个核苷酸(例如,长度为15个核苷酸(nt)至20nt、20nt至25nt、25nt至30nt、30nt至35nt、35nt至40nt、40nt至45nt或45nt至50nt)。在某些情况下,间隔序列的长度为35-38个核苷酸。例如,图2A-2Z、图2AA-2ZZ和图2AAA-2FFF中描绘的核苷酸序列(其中T被U取代)中任一者可以被包含在包含(N)n恒定区的引导RNA中,其中N是任何核苷酸,n是15至50的整数(例如,15至20、20至25、25至30、30至35、35至38、35至40、40至45或45至50)。

[0314] 作为一个实例,Cas12L引导RNA的恒定区可以包含核苷酸序列:AUUGUUGUAACUCUUAUUUUGUAUGGAGUAAACAAC(SEQ ID NO:59)。作为另一个实例,Cas12L引导RNA的恒定区可以包含核苷酸序列:AUUGUUGUAGACCUCUUUUUAUAAGGAUUGAACAAC(SEQ ID NO:60;参见图2B)。作为另一个实例,Cas12L引导RNA的恒定区可以包含核苷酸序列:UAUUGUUGUAGAUACCUUUUUGUAAGGAUUAAACAAC(SEQ ID NO:61;参见图2P)。作为另一个实例,Cas12L引导RNA的恒定区可以包含核苷酸序列:AAUGUUGUAGAUGCCUUUUUAUAAGGAUUAAACAACUUG(SEQ ID NO:62;参见图2T)。作为另一个实例,Cas12L引导RNA的恒定区可以包含核苷酸序列:AUUGUUGAAAUAGUACUUUUAUAGUCUAUAUACAAC(SEQ ID NO:63;参见图2BB)。作为另一个实例,Cas12L引导RNA的恒定区可以包含核苷酸序列:UAUUGUUGUAACUCUUAUUUUGUAUGGAGUAAACAAC(SEQ ID NO:64;参见图2CC)。作为另一个实例,Cas12L引导RNA的恒定区可以包含核苷酸序列:AUUGUUGUAACUUUUAUUUUGUAUGGAGUAAACAAC(SEQ ID NO:65;参见图2TT)。作为另一个实例,Cas12L引导RNA的恒定区可以包含核苷酸序列:AAUGUUGUAGAUACCUUUUUGUAAGGAUUGAACAAC(SEQ ID NO:66;

参见图2WW)。作为另一个实例,Cas12L引导RNA的恒定区可以包含核苷酸序列:AUUGUUGUAAUACUAUUUUUGUAAAGUAUAAACAAC(SEQ ID NO:67;参见图2XX)。作为另一个实例,Cas12L引导RNA的恒定区可以包含核苷酸序列:AAUGUUGUAGAUGCCUUUUUAUAAGGAUUAAACAAC(SEQ ID NO:68;参见图2YY)。作为另一个实例,Cas12L引导RNA的恒定区可以包含核苷酸序列:AUUGUUGUAAUACACUUUUUAUAAGGUAUGAACAAC(SEQ ID NO:69;参见图2BBB)。作为另一个实例,Cas12L引导RNA的恒定区可以包含核苷酸序列:AUUGUUGUAACAUCUAUUUUGUAAGGUGUAAACAAC(SEQ ID NO:70;参见图2KK)。

[0315] 图2A-2Z、图2AA-2ZZ和图2AAA-2FFF中描绘的核苷酸序列(但其中T被U取代)中任一者的反向互补序列可以被包含在包含恒定区-(N)n的引导RNA中,其中N是任何核苷酸,n是15-50的整数(例如,15至20、20至25、25至30、30至35、35至38、35至40、40至45或45至50)。作为一个实例,引导RNA可以具有以下核苷酸序列:NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNAUUGUUGUAACUCUUAUUUUGUAUGGAGUAAACAAC(SEQ ID NO:71),或者在一些情况下为反向互补序列,其中N是任何核苷酸,例如,其中N的片段包括靶核酸结合序列。

[0316] Cas12L引导多核苷酸

[0317] 在一些情况下,与Cas12L蛋白结合形成核酸/Cas12L多肽复合物并将所述复合物靶向靶核酸(例如靶DNA)内特定位置的核酸包含仅核糖核苷酸、仅脱氧核糖核苷酸或核糖核苷酸和脱氧核糖核苷酸的混合物。在某些情况下,引导多核苷酸仅包含核糖核苷酸,并且在本文中称为“引导RNA”在一些情况下,引导多核苷酸仅包含脱氧核糖核苷酸,并且在本文中称为“引导DNA”在一些情况下,引导多核苷酸包含核糖核苷酸和脱氧核糖核苷酸二者。引导多核苷酸可以包含核糖核苷酸碱基、脱氧核糖核苷酸碱基、核苷酸类似物、修饰的核苷酸等的组合;并且可以还包含天然存在的骨架残基和/或键联和/或非天然存在的骨架残基和/或键联。

[0318] C_{AS}12L系统

[0319] 本公开提供一种Cas12L系统。本公开的Cas12L系统可以包含:a)本公开的Cas12L多肽和Cas12L引导RNA;b)本公开的Cas12L多肽、Cas12L引导RNA和供体模板核酸;c)本公开的Cas12L融合多肽和Cas12L引导RNA;d)本公开的Cas12L融合多肽、Cas12L引导RNA和供体模板核酸;e)编码本公开的Cas12L多肽的mRNA和Cas12L引导RNA;f)编码本公开的Cas12L多肽的mRNA、Cas12L引导RNA和供体模板核酸;g)编码本公开的Cas12L融合多肽的mRNA和Cas12L引导RNA;h)编码本公开的Cas12L融合多肽的mRNA、Cas12L引导RNA和供体模板核酸;i)重组表达载体,其包含编码本公开的Cas12L多肽的核苷酸序列和编码Cas12L引导RNA的核苷酸序列;j)重组表达载体,其包含编码本公开的Cas12L多肽的核苷酸序列、编码Cas12L引导RNA的核苷酸序列和编码供体模板核酸的核苷酸序列;k)重组表达载体,其包含编码本公开的Cas12L融合多肽的核苷酸序列和编码Cas12L引导RNA的核苷酸序列;l)重组表达载体,其包含编码本公开的Cas12L融合多肽的核苷酸序列、编码Cas12L引导RNA的核苷酸序列和编码供体模板核酸的核苷酸序列;m)包含编码本公开的Cas12L多肽的核苷酸序列的第一重组表达载体,和包含编码Cas12L引导RNA的核苷酸序列的第二重组表达载体;n)包含编码本公开的Cas12L多肽的核苷酸序列的第一重组表达载体,和包含编码Cas12L引导RNA的核苷酸序列的第二重组表达载体;和供体模板核酸;o)包含编码本公开的Cas12L融合多肽的核苷酸序列的第一重组表达载体,和包含编码Cas12L引导RNA的核苷酸序列的第二重组表

达载体;p)包含编码本公开的Cas12L融合多肽的核苷酸序列的第一重组表达载体,和包含编码Cas12L引导RNA的核苷酸序列的第二重组表达载体;和供体模板核酸;q)重组表达载体,其包含编码本公开的Cas12L多肽的核苷酸序列、编码第一Cas12L引导RNA的核苷酸序列和编码第二Cas12L引导RNA的核苷酸序列;或者r)重组表达载体,其包含编码本公开的Cas12L融合多肽的核苷酸序列、编码第一Cas12L引导RNA的核苷酸序列和编码第二Cas12L引导RNA的核苷酸序列;或者(a)至(r)之一的某种变体。

[0320] 核酸

[0321] 本发明提供了一种或多种核酸,其包含以下一种或多种:供体多核苷酸序列、编码Cas12L多肽(例如,野生型Cas12L蛋白、切口酶Cas12L蛋白、dCas12L蛋白、融合Cas12L蛋白等)的核苷酸序列、Cas12L引导RNA和编码Cas12L引导RNA的核苷酸序列。本公开提供了一种核酸,其包含编码Cas12L融合多肽的核苷酸序列。本公开提供了一种重组表达载体,其包含编码Cas12L多肽的核苷酸序列。本公开提供了一种重组表达载体,其包含编码Cas12L融合多肽的核苷酸序列。本公开提供了一种重组表达载体,其包含:a)编码Cas12L多肽的核苷酸序列;和b)编码Cas12L引导RNA的核苷酸序列。本公开提供了一种重组表达载体,其包含:a)编码Cas12L融合多肽的核苷酸序列;和b)编码Cas12L引导RNA的核苷酸序列。在一些情况下,编码Cas12L蛋白的核苷酸序列和/或编码Cas12L引导RNA的核苷酸序列可操作地连接到启动子,所述启动子在选择的细胞类型(例如,原核细胞、真核细胞、植物细胞、动物细胞、哺乳动物细胞、灵长类动物细胞、啮齿类动物细胞、人类细胞等)中是可操作的。

[0322] 在一些情况下,编码本公开的Cas12L多肽的核苷酸序列是密码子优化的。这种类型的优化可能需要编码Cas12L的核苷酸序列的突变以模拟预期的宿主生物体或细胞同时编码相同蛋白质时的密码子偏好。因此,密码子可改变,但编码的蛋白质保持不变。例如,如果预期的靶细胞是人细胞,可使用人密码子优化的编码Cas12L的核苷酸序列。作为另一个非限制性实例,如果预期的宿主细胞是小鼠细胞,则可生成小鼠密码子优化的编码Cas12L的核苷酸序列。作为另一个非限制性实例,如果预期的宿主细胞是植物细胞,则可生成植物密码子优化的编码Cas12L的核苷酸序列。作为另一个非限制性实例,如果预期的宿主细胞是昆虫细胞,则可生成昆虫密码子优化的编码Cas12L的核苷酸序列。

[0323] 密码子选用框很容易获得,例如,在www[dot]kazusa[dot]or[dot]jp[forwardslash]codon上可获得的“密码子用法数据库”。在一些情况下,本公开的核酸包含编码Cas12L多肽的核苷酸序列,所述核苷酸序列被密码子优化以在真核细胞中表达。在一些情况下,本公开的核酸包含编码Cas12L多肽的核苷酸序列,所述核苷酸序列被密码子优化以在动物细胞中表达。在一些情况下,本公开的核酸包含编码Cas12L多肽的核苷酸序列,所述核苷酸序列被密码子优化以在真菌细胞中表达。在一些情况下,本公开的核酸包含编码Cas12L多肽的核苷酸序列,所述核苷酸序列被密码子优化以在植物细胞中表达。在一些情况下,本公开的核酸包含编码Cas12L多肽的核苷酸序列,所述核苷酸序列被密码子优化以在单子叶植物物种中表达。在一些情况下,本公开的核酸包含编码Cas12L多肽的核苷酸序列,所述核苷酸序列被密码子优化以在双子叶植物物种中表达。在一些情况下,本公开的核酸包含编码Cas12L多肽的核苷酸序列,所述核苷酸序列被密码子优化以在裸子植物物种中表达。在一些情况下,本公开的核酸包含编码Cas12L多肽的核苷酸序列,所述核苷酸序列被密码子优化以在被子植物物种中表达。在一些情况下,本公开的核酸包含编码Cas12L多

肽的核苷酸序列,所述核苷酸序列被密码子优化以在玉米细胞中表达。在一些情况下,本公开的核酸包含编码Cas12L多肽的核苷酸序列,所述核苷酸序列被密码子优化以在大豆细胞中表达。在一些情况下,本公开的核酸包含编码Cas12L多肽的核苷酸序列,所述核苷酸序列被密码子优化以在水稻细胞中表达。在一些情况下,本公开的核酸包含编码Cas12L多肽的核苷酸序列,所述核苷酸序列被密码子优化以在小麦细胞中表达。在一些情况下,本公开的核酸包含编码Cas12L多肽的核苷酸序列,所述核苷酸序列被密码子优化棉花在棉花细胞中表达。在一些情况下,本公开的核酸包含编码Cas12L多肽的核苷酸序列,所述核苷酸序列被密码子优化以在高粱细胞中表达。在一些情况下,本公开的核酸包含编码Cas12L多肽的核苷酸序列,所述核苷酸序列被密码子优化以在苜蓿细胞中表达。在一些情况下,本公开的核酸包含编码Cas12L多肽的核苷酸序列,所述核苷酸序列被密码子优化棉花在甘蔗细胞中表达。在一些情况下,本公开的核酸包含编码Cas12L多肽的核苷酸序列,所述核苷酸序列被密码子优化以在拟南芥细胞中表达。在一些情况下,本公开的核酸包含编码Cas12L多肽的核苷酸序列,所述核苷酸序列被密码子优化以在番茄细胞中表达。在一些情况下,本公开的核酸包含编码Cas12L多肽的核苷酸序列,所述核苷酸序列被密码子优化以在黄瓜细胞中表达。在一些情况下,本公开的核酸包含编码Cas12L多肽的核苷酸序列,所述核苷酸序列被密码子优化以在马铃薯细胞中表达。在一些情况下,本公开的核酸包含编码Cas12L多肽的核苷酸序列,所述核苷酸序列被密码子优化以在藻类细胞中表达。

[0324] 本公开提供了一种或多种重组表达载体,其包含(在一些情况下在不同的重组表达载体中,以及在一些情况下在相同的重组表达载体中):(i)供体模板核酸的核苷酸序列(其中供体模板包含与靶核酸(例如靶基因组)的靶序列具有同源性的核苷酸序列);(ii)编码Cas12L引导RNA的核苷酸序列,所述引导RNA与靶向基因组的靶基因座的靶序列杂交(例如,可操作地连接到在靶细胞诸如真核细胞中可操作的启动子);和(iii)编码Cas12L蛋白的核苷酸序列(例如,可操作地连接到在靶细胞诸如真核细胞中可操作的启动子)。本公开提供了一种或多种重组表达载体,其包含(在一些情况下在不同的重组表达载体中,以及在一些情况下在相同的重组表达载体中):(i)供体模板核酸的核苷酸序列(其中供体模板包含与靶核酸(例如靶基因组)的靶序列具有同源性的核苷酸序列);和(ii)编码Cas12L引导RNA的核苷酸序列,所述引导RNA与靶向基因组的靶基因座的靶序列杂交(例如,可操作地连接到在靶细胞诸如真核细胞中可操作的启动子)。本公开提供了一种或多种重组表达载体,其包含(在一些情况下在不同的重组表达载体中,以及在一些情况下在相同的重组表达载体中):(i)编码Cas12L引导RNA的核苷酸序列,所述引导RNA与靶向基因组的靶基因座的靶序列杂交(例如,可操作地连接到在靶细胞诸如真核细胞中可操作的启动子);和(ii)编码Cas12L蛋白的核苷酸序列(例如,可操作地连接到在靶细胞诸如真核细胞中可操作的启动子)。

[0325] 合适的表达载体包括病毒表达载体(例如,基于以下病毒的病毒载体:牛痘病毒;脊髓灰质炎病毒;腺病毒(参见例如Li等人,Invest Opthalmol Vis Sci 35:2543 2549,1994;Borras等人,Gene Ther 6:515524,1999;Li和Davidson,PNAS 92:7700 7704,1995;Sakamoto等人,H Gene Ther 5:1088 1097,1999;WO 94/12649、WO 93/03769;WO 93/19191;WO 94/28938;WO 95/11984和WO 95/00655);腺相关病毒(AAV)(参见例如Ali等人,Hum Gene Ther 9:81 86,1998;Flannery等人,PNAS 94:6916 6921,1997;Bennett等人,

Invest Opthalmol Vis Sci 38:2857 2863,1997;Jomary等人,Gene Ther 4:683 690,1997;Rolling等人,Hum Gene Ther 10:641 648,1999;Ali等人,Hum Mol Genet 5:591 594,1996;Srivastava的WO 93/09239;Samulski等人,J.Vir.(1989)63:3822-3828;Mendelson等人,Virol.(1988)166:154-165;以及Flotte等人,PNAS(1993)90:10613-10617);SV40;单纯疱疹病毒;人免疫缺陷病毒(参见例如,Miyoshi等人,PNAS 94:10319 23,1997;Takahashi等人,J Virol 73:7812 7816,1999);逆转录病毒载体(例如,鼠白血病病毒、脾坏死病毒和源自诸如劳斯肉瘤病毒、哈维肉瘤病毒、禽白血病病毒、慢病毒、人免疫缺陷病毒、骨髓增生肉瘤病毒以及乳腺肿瘤病毒的逆转录病毒的载体)等。在一些情况下,本公开的重组表达载体是重组腺相关病毒(AAV)载体。在一些情况下,本公开的重组表达载体是重组慢病毒载体。在一些情况下,本公开的重组表达载体是重组逆转录病毒载体。

[0326] 对于植物应用,可以使用基于烟草花叶病毒组、马铃薯x病毒组、马铃薯y病毒组、烟草脆裂病毒组、番茄丛矮病毒组、双生病毒组、雀麦花叶病毒组、香石竹斑驳病毒组、苜蓿花叶病毒组或南瓜花叶病毒组的病毒载体。参见例如Peyret和Lomonossoff(2015)Plant Biotechnol.J.13:1121。合适的烟草花叶病毒组载体包括例如番茄花叶病毒(ToMV)载体、烟草花叶病毒(TMV)载体、烟草淡绿花叶病毒(TMGMV)载体、辣椒轻型斑驳病毒(PMMoV)载体、红辣椒轻型斑驳病毒(PaMMV)载体、黄瓜绿斑驳花叶病毒(CGMMV)载体、九州绿斑驳花叶病毒(KGMMV)载体、木槿潜隐皮尔斯堡病毒(HLFPV)载体、齿舌兰环斑病毒(ORSV)载体、地黄花叶病毒(ReMV)载体、桑蒙仙人掌病毒(SOV)载体、山榆菜斑驳病毒(WMoV)载体、油菜花叶病毒(YoMV)载体、菽麻花叶病毒(SHMV)载体等。合适的马铃薯x病毒组载体包括例如马铃薯病毒X(PVX)载体、马铃薯奥古巴花叶病毒(PAMV)载体、六出花花叶病毒X(AlsVX)载体、仙人掌病毒X(CVX)载体、兰花花叶病毒(CymMV)载体、玉簪病毒X(HVX)载体、百合病毒X(LVX)载体、水仙花叶病毒(NMV)载体、石蒜病毒X(NVX)载体、车前草花叶病毒(PlAMV)载体、草莓轻型黄边病毒(SMYEV)载体、郁金香病毒X(TVX)载体、白三叶草花叶病毒(WClMV)载体、竹嵌纹病毒(BaMV)载体等。合适的马铃薯y病毒载体包括例如马铃薯病毒Y(PVY)载体、菜豆普通花叶病毒(BCMV)载体、三叶草黄脉病毒(ClYVV)载体、东亚西番莲病毒(EAPV)载体、香雪兰花叶病毒(FreMV)载体、日本山药花叶病毒(JYMV)载体、莴苣花叶病毒(LMV)载体、玉米矮小花叶病毒(MDMV)载体、洋葱黄矮病毒(OYDV)载体、番木瓜环斑病毒(PRSV)载体、辣椒斑驳病毒(PepMoV)载体、紫苏属斑驳病毒(PerMoV)载体、李痘病毒(PPV)载体、马铃薯病毒A(PVA)载体、高粱花叶病毒(SrMV)载体、大豆花叶病毒(SMV)载体、甘蔗花叶病毒(SCMV)载体、郁金香花叶病毒(TulMV)载体、芜菁花叶病毒(TuMV)载体、西瓜花叶病毒(WMV)载体、西葫芦黄色花叶病毒(ZYMV)载体、烟草蚀纹病毒(TEV)载体等。合适的烟草病毒载体包括例如烟草脆裂病毒(TRV)载体等。合适的番茄丛矮病毒组载体包括例如番茄丛矮病毒(TBSV)载体、茄斑驳皱缩病毒(EMCV)载体、葡萄藤阿尔及利亚潜伏病毒(GALV)载体等。合适的黄瓜花叶病毒载体包括例如黄瓜花叶病毒(CMV)载体、花生矮化病毒(PSV)载体、番茄不孕病毒(TAV)载体等。合适的雀麦草花叶病毒组载体包括例如雀麦花叶病毒(BMV)载体、豇豆褪绿斑点病毒(CCMV)载体等。合适的香石竹斑驳病毒组载体包括例如康乃馨斑驳病毒(CarMV)载体、甜瓜坏死斑点病毒(MNSV)载体、豌豆茎坏死病毒(PSNV)载体、芜菁皱缩病毒(TCV)载体等。合适的苜蓿花叶病毒载体包括例如苜蓿花叶病毒(AMV)载体等。

[0327] 根据所用的宿主/载体系统,可在表达载体中使用多种合适的转录和翻译控制元

件中的任一种,包括组成型启动子和诱导型启动子、转录增强子元件、转录终止子等。

[0328] 在一些实施方案中,编码Cas12L引导RNA的核苷酸序列可操作地连接至控制元件,例如转录控制元件,诸如启动子。在一些实施方案中,编码Cas12L蛋白或Cas12L融合多肽的核苷酸序列可操作地连接至控制元件,例如转录控制元件,诸如启动子。

[0329] 转录控制元件可以是启动子。在一些情况下,启动子是组成型活性启动子。在一些情况下,启动子是可调控启动子。在一些情况下,启动子是诱导型启动子。在一些情况下,启动子是组织特异性启动子。在一些情况下,启动子是细胞类型特异性启动子。在一些情况下,转录控制元件(例如,启动子)在所靶向细胞类型或所靶向细胞群中是功能性的。例如,在一些情况下,转录控制元件在真核细胞(例如,造血干细胞(例如,动员的外周血(mPB)CD34(+)细胞、骨髓(BM)CD34(+)细胞等))中可以是功能性的。

[0330] 真核启动子(在真核细胞中是功能性的启动子)的非限制性实例包括EF1α,来自巨细胞病毒(CMV)立即早期、单纯疱疹病毒(HSV)胸苷激酶、早期和晚期SV40、逆转录病毒的长末端重复序列(LTR)以及小鼠金属硫蛋白-I的那些启动子。选择适当的载体和启动子完全在本领域普通技术人员的水平之内。表达载体还可含有用于翻译起始的核糖体结合位点和转录终止子。表达载体还可包含用于扩增表达的适当序列。表达载体还可包含编码蛋白质标签(例如,6xHis标签、血凝素标签、荧光蛋白等)的核苷酸序列,所述蛋白质标签可融合至Cas12L蛋白,从而产生融合Cas12L多肽。

[0331] 在一些实施方案中,编码Cas12L引导RNA和/或Cas12L融合多肽的核苷酸序列可操作地连接至诱导型启动子。在一些实施方案中,编码Cas12L引导RNA和/或Cas12L融合蛋白的核苷酸序列可操作地连接至组成型启动子。

[0332] 启动子可以是组成型活性启动子(即,组成性地处于活性/“ON”状态的启动子),它可以是诱导型启动子(即,通过外界刺激例如特定温度、化合物或蛋白质的存在控制其状态(活性/“ON”或非活性/“OFF”)的启动子),它可以是空间限制的启动子(即,转录控制元件、增强子等)(例如,组织特异性启动子、细胞类型特异性启动子等),并且它可以是时间限制的启动子(即,启动子在胚胎发育的特定阶段过程中或在生物过程的特定阶段(例如,小鼠体内的毛囊周期)过程中处于“ON”状态或“OFF”状态)。

[0333] 合适的启动子可衍生自病毒并且可因此称为病毒启动子,或者它们可衍生自任何生物,包括原核生物或真核生物。合适的启动子可用来通过任何RNA聚合酶(例如,pol I、pol II、pol III)驱动表达。示例性启动子包括但不限于SV40早期启动子、小鼠乳腺肿瘤病毒长末端重复序列(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)启动子诸如CMV立即早期启动子区(CMVIE)、劳斯肉瘤病毒(RSV)启动子、人U6小核启动子(U6)(Miyagishi等人,Nature Biotechnology 20,497-500(2002))、增强的U6启动子(例如,Xia等人,Nucleic Acids Res.2003年9月1日;31(17))、人H1启动子(H1)等。

[0334] 在一些情况下,编码Cas12L引导RNA的核苷酸序列可操作地连接至(受控制于)在真核细胞中可操作的启动子(例如,U6启动子、增强的U6启动子、H1启动子等)。如本领域的普通技术人员所理解的,当使用U6启动子(例如,在真核细胞中)或另一种PolIII启动子由核酸(例如,表达载体)表达RNA(例如,引导RNA)时,如果连续存在若干个T(在RNA中编码U),则可能需要对RNA进行突变。这是因为DNA中的一串T(例如,5个T)可充当聚合酶III

(PolIII)的终止子。因此,为了确保引导RNA在真核细胞中的转录,有时可能需要修饰编码引导RNA的序列以消除T的作用。在一些情况下,编码Cas12L蛋白(例如,野生型Cas12L蛋白、切口酶Cas12L蛋白、dCas12L蛋白、融合Cas12L蛋白等)的核苷酸序列可操作地连接至在真核细胞中可操作的启动子(例如,CMV启动子、EF1α启动子、雌激素受体调控的启动子等)。

[0335] 诱导型启动子的实例包括但不限于T7 RNA聚合酶启动子、T3RNA聚合酶启动子、异丙基-β-D-硫代半乳糖苷(IPTG)调控的启动子、乳糖诱导的启动子、热休克启动子、四环素调控的启动子、类固醇调控的启动子、金属调控的启动子、雌激素受体调控的启动子等。因此,诱导型启动子可通过分子调控,所述分子包括但不限于强力霉素;雌激素和/或雌激素类似物;IPTG等。

[0336] 适合使用的诱导型启动子包括本文描述的或本领域普通技术人员已知的任何诱导型启动子。诱导型启动子的实例包括但不限于化学/生物化学调节的和物理调节的启动子,例如醇调节的启动子、四环素调节的启动子(例如脱水四环素(aTc)应答启动子和其它四环素应答启动子系统,包括四环素阻遏蛋白(tetR)、四环素操纵子序列(tetO)和四环素反式激活因子融合蛋白(tTA))、类固醇调节的启动子(例如基于大鼠糖皮质激素受体、人雌激素受体、蛾蜕皮激素受体和来自类固醇/类视黄醇/甲状腺受体超家族的启动子)、金属调节的启动子(例如,来源于酵母、小鼠和人的金属硫蛋白(结合和螯合金属离子的蛋白质)基因的启动子)、发病机理调节的启动子(例如,由水杨酸、乙烯或苯并噻二唑(BTH)诱导的)、温度/热诱导型启动子(例如,热休克启动子)和光调节的启动子(例如,来自植物细胞的光响应性启动子)。

[0337] 在一些情况下,启动子是空间限制的启动子(即,细胞类型特异性启动子、组织特异性启动子等),使得在多细胞生物体中,启动子在特定细胞子组中是活性的(即,“ON”)。空间限制的启动子也可称为增强子、转录控制元件、控制序列等。可使用任何方便的空间限制的启动子,只要启动子在靶向宿主细胞(例如,真核细胞;原核细胞)中是功能性的即可。

[0338] 在一些情况下,启动子是可逆启动子。合适的可逆启动子,包括可逆诱导型启动子,在本领域中是已知的。此类可逆启动子可分离自并衍生自许多生物体,例如真核生物和原核生物。用于第二生物体的衍生自第一生物体(例如,第一原核生物和第二真核生物、第一真核生物和第二原核生物等)的可逆启动子的修饰在本领域中是众所周知的。此类可逆启动子和基于此类可逆启动子但还包含另外的控制蛋白的系统包括但不限于醇调控的启动子(例如,醇脱氢酶I(alcA)基因启动子、响应于醇反式激活因子蛋白(AlcR)的启动子等)、四环素调控的启动子(例如,包括Tet激活因子、TetON、TetOFF等的启动子系统)、类固醇调控的启动子(例如,大鼠糖皮质激素受体启动子系统、人雌激素受体启动子系统、类视黄醇启动子系统、甲状腺启动子系统、蜕皮激素启动子系统、米非司酮启动子系统等)、金属调控的启动子(例如,金属硫蛋白启动子系统等)、发病原相关的调控启动子(例如,水杨酸调控的启动子、乙烯调控的启动子、苯并噻二唑调控的启动子等)、温度调控的启动子(例如,热休克诱导型启动子(例如,HSP-70、HSP-90、大豆热休克启动子等))、光调控的启动子、合成诱导型启动子等。

[0339] RNA聚合酶III(Pol III)启动子可用于驱动非蛋白质编码RNA分子(例如,引导RNA)表达。在一些情况下,合适的启动子是Pol III启动子。在一些情况下,Pol III启动子可操作地连接到编码引导RNA(gRNA)的核苷酸序列。在一些情况下,Pol III启动子可操作

地连接到编码单引导RNA(sgRNA)的核苷酸序列。在一些情况下,Pol III启动子可操作地连接到编码CRISPR RNA(crRNA)的核苷酸序列。在一些情况下,Pol III启动子可操作地连接到编码tracrRNA的核苷酸序列。

[0340] Pol III启动子的非限制性实例包括U6启动子、H1启动子、5S启动子、腺病毒2(Ad2)VAI启动子、tRNA启动子和7SK启动子。参见例如Schramm和Hernandez(2002)Genes&Development16:2593-2620。在一些情况下,Pol III启动子选自由以下组成的组:U6启动子、H1启动子、5S启动子、腺病毒2(Ad2)VAI启动子、tRNA启动子和7SK启动子。在一些情况下,引导RNA编码核苷酸序列可操作地连接到选自由U6启动子、H1启动子、5S启动子、腺病毒2(Ad2)VAI启动子、tRNA启动子和7SK启动子组成的组的启动子。在一些情况下,编码单引导RNA的核苷酸序列可操作地连接到选自由U6启动子、H1启动子、5S启动子、腺病毒2(Ad2)VAI启动子、tRNA启动子和7SK启动子组成的组的启动子。

[0341] 描述可在本文中用于在植物、植物组织和植物细胞中表达的启动子的实例包括但不限于以下中描述的启动子:美国专利号6,437,217(玉米RS81启动子)、美国专利号5,641,876(水稻肌动蛋白启动子)、美国专利号6,426,446(玉米RS324启动子)、美国专利号6,429,362(玉米PR-1启动子)、美国专利号6,232,526(玉米A3启动子)、美国专利号6,177,611(组成型玉米启动子)、美国专利号5,322,938、5,352,605、5,359,142和5,530,196(35S启动子)、美国专利号6,433,252(玉米L3油质蛋白启动子)、美国专利号6,429,357(水稻肌动蛋白2启动子以及水稻肌动蛋白2内含子)、美国专利号5,837,848(根特异性启动子)、美国专利号6,294,714(光诱导型启动子)、美国专利号6,140,078(盐诱导型启动子)、美国专利号6,252,138(病原体诱导型启动子)、美国专利号6,175,060(磷缺乏诱导型启动子)、美国专利号6,635,806(γ-胶原蛋白启动子)和美国专利申请序列号09/757,089(玉米叶绿体醛缩酶启动子)。可以使用的其它启动子包括胭脂氨酸合酶(NOS)启动子(Ebert等人,1987)、章鱼碱合酶(OCS)启动子(其携带在根癌土壤杆菌的肿瘤诱导质粒上)、花椰菜病毒启动子诸如花椰菜花叶病毒(CaMV)19S启动子(Lawton等人Plant Molecular Biology(1987)9:315-324)、CaMV 35S启动子(Odell等人,Nature(1985)313:810-812)、玄参花叶病毒35S-启动子(美国专利号6,051,753;5,378,619)、蔗糖合酶启动子(Yang和Russell,Proceedings of the National Academy of Sciences,USA(1990)87:4144-4148)、R基因复合物启动子(Chandler等人,Plant Cell(1989)1:1175-1183)和叶绿素a/b结合蛋白基因启动子PC1SV(美国专利号5,850,019)和AGRtu.nos(GenBank登录号V00087;Depicker等人,Journal of Molecular and Applied Genetics(1982)1:561-573;Bevan等人,1983)启动子。

[0342] 将核酸(例如,包含供体多核苷酸序列的核酸、一种或多种编码Cas12L蛋白和/或Cas12L引导RNA的核酸等)引入宿主细胞中的方法在本领域中是已知的,并且可使用任何方便的方法来将核酸(例如,表达构建体)引入细胞中。合适的方法包括例如病毒感染、转染、脂质体转染、电穿孔、磷酸钙沉淀、聚乙烯亚胺(PEI)介导的转染、DEAE-葡聚糖介导的转染、脂质体介导的转染、粒子枪技术、磷酸钙沉淀、直接微注射、纳米颗粒介导的核酸递送等。

[0343] 将重组表达载体引入细胞中可在促进细胞存活的任何培养基中和任何培养条件下发生。将重组表达载体引入靶细胞中可在体内或离体进行。将重组表达载体引入靶细胞中可在体外进行。

[0344] 在一些实施方案中,Cas12L蛋白可以作为RNA提供。RNA可通过直接化学合成提供,

或者可在体外从DNA(例如,编码Cas12L蛋白的DNA)转录。一旦合成,可通过用于将核酸引入细胞中的任何众所周知的技术(例如,微注射、电穿孔、转染等)将RNA引入细胞中。

[0345] 可使用开发良好的转染技术(参见例如Angel和Yanik(2010)PLoS ONE 5(7):e11756);以及可从Qiagen商购获得的TransMessenger®试剂、可从Stemgent商购获得的Stemfect™RNA转染试剂盒和可从Mirus Bio LLC商购获得的TransIT®-mRNA转染试剂盒向细胞提供核酸。还参见Beumer等人(2008)PNAS105(50):19821-19826。

[0346] 可直接向靶宿主细胞提供载体。换句话讲,使细胞与包含主题核酸的载体(例如,具有供体模板序列并编码Cas12L引导RNA的重组表达载体;编码Cas12L蛋白的重组表达载体等)接触,使得载体被细胞吸收。用于使细胞与作为质粒的核酸载体接触的方法(包括电穿孔、氯化钙转染、微注射和脂质体转染)在本领域中是众所周知的。对于病毒载体递送,可使细胞与包含主题病毒表达载体的病毒颗粒接触。

[0347] 逆转录病毒,例如慢病毒,适用于本公开的方法。通常使用的逆转录病毒载体是“缺陷型的”,即不能产生生产性感染所需要的病毒蛋白质。而且载体的复制需要在包装细胞系中生长。为了生成包含目标核酸的病毒颗粒,通过包装细胞系将包含核酸的逆转录病毒核酸包装到病毒衣壳中。不同的包装细胞系提供待并入衣壳中的不同包膜蛋白(嗜亲性、双嗜性或嗜异性),此包膜蛋白决定病毒颗粒对细胞的特异性(对鼠和大鼠的嗜亲性;对包括人、狗和小鼠的大多数哺乳动物细胞类型的双嗜性;以及对除了鼠细胞之外的大多数哺乳动物细胞类型的嗜异性)。适当的包装细胞系可用来确保细胞被包装的病毒颗粒靶向。将主题载体表达载体引入包装细胞系中以及采集由包装细胞系生成的病毒颗粒的方法在本领域中是众所周知的。还可通过直接微注射引入核酸(例如,RNA的注射)。

[0348] 用于向靶宿主细胞提供编码Cas12L引导RNA和/或Cas12L多肽的核酸的载体可包括用于驱动目标核酸的表达(即,转录激活)的合适的启动子。换句话讲,在一些情况下,目标核酸将可操作地连接至启动子。所述启动子可包括遍在活化型启动子,例如CMV-β-肌动蛋白启动子;或诱导型启动子,诸如在特定细胞群中有活性或对药物(诸如四环素的)存在有响应的启动子。通过转录激活,预期转录将在靶细胞中与基础水平相比增加10倍、100倍、更通常地1000倍。另外,用于向细胞提供编码Cas12L引导RNA和/或Cas12L蛋白的核酸的载体可包含如下核酸序列,其在靶细胞中编码可选择标记以便鉴定已经吸收Cas12L引导RNA和/或Cas12L蛋白的细胞。

[0349] 包含编码Cas12L多肽或Cas12L融合多肽的核苷酸序列的核酸在一些情况下是RNA。因此,可将Cas12L融合蛋白以RNA的形式引入细胞中。将RNA引入细胞中的方法在本领域中是已知的并且可包括例如直接注射、转染或用于引入DNA的任何其他方法。Cas12L蛋白可替代地以多肽的形式向细胞提供。这种多肽可任选地融合至增加产物溶解度的多肽结构域。所述结构域可通过限定的蛋白酶切割位点(例如,通过TEV蛋白酶切割的TEV序列)连接至多肽。接头还可包括一个或多个柔性序列,例如1至10个甘氨酸残基。在一些实施方案中,融合蛋白的切割在维持产物溶解度的缓冲液中进行,例如在0.5至2M尿素存在下、在增加溶解度的多肽和/或多核苷酸的存在下等进行。目标结构域包括核内体溶解结构域,例如流感HA结构域;和有助于产生的其他多肽,例如IF2结构域、GST结构域、GRPE结构域等。多肽可配制用于改进的稳定性。例如,肽可以是PEG化的,其中聚乙烯氧基提供在血流中的增加的寿命。

[0350] 另外或可替代地，本公开的Cas12L多肽可融合至多肽穿透结构域以促进被细胞吸收。许多穿透结构域在本领域中是已知的并且可用于本公开的非整合多肽，包括肽、肽模拟物和非肽运载体。例如，穿透肽可衍生自黑腹果蝇转录因子触角足基因(称为穿透蛋白)的第三α螺旋，所述第三α螺旋包含氨基酸序列RQIKIWFQNRRMKWKK(SEQ ID NO:44)。作为另一个实例，穿透肽包含HIV-1tat碱性区氨基酸序列，所述氨基酸序列可包括例如天然存在的tat蛋白的氨基酸49-57。其他穿透结构域包括聚精氨酸基序，例如HIV-1rev蛋白的氨基酸34-56的区域、九精氨酸、八精氨酸等。(参见例如Futaki等人(2003)Curr Protein Pept Sci.2003年4月;4(2):87-9和446;以及Wender等人(2000)Proc.Natl.Acad.Sci.U.S.A 2000年11月21日;97(24):13003-8;公布的美国专利申请20030220334;20030083256;20030032593;和20030022831，在此以引用方式明确地并入易位肽和类肽的教导内容中)。九精氨酸(R9)序列是已表征的更有效的PTD之一(Wender等人2000;Uemura等人2002)。可选择进行融合的位点以便优化多肽的生物活性、分泌或结合特征。将通过常规实验确定最佳位点。

[0351] 如上所述，在一些情况下，靶细胞是植物细胞。用重组核酸转化植物细胞中的染色体或质体的许多方法是本领域已知的，根据本申请的方法，这些方法可用于产生转基因植物细胞和/或转基因植物。可以使用本领域已知的任何合适的转化植物细胞的方法或技术。用于转化植物的有效方法包括细菌介导的转化，诸如土壤杆菌介导的或根瘤菌介导的转化和微粒轰击介导的转化。通过细菌介导的转化或微粒轰击以及随后的培养等，用转化载体转化外植体的多种方法是本领域已知的，这些外植体再生或发展转基因植物。用于植物转化的其它方法，诸如显微注射、电穿孔、真空渗透、压力、超声、碳化硅纤维搅动、PEG介导的转化等，也是本领域已知的。根据所用的方法和外植体，通过这些转化方法产生的转基因植物对于转化事件而言可以是嵌合的或非嵌合的。

[0352] 转化植物细胞的方法是本领域普通技术人员熟知的。例如，通过使用用重组DNA包被的颗粒进行微粒轰击来转化植物细胞(例如，基因枪转化)的具体说明可见于美国专利号5,550,318;5,538,8806,160,208;6,399,861;和6,153,812中，并且土壤杆菌介导的转化描述于美国专利号5,159,135;5,824,877;5,591,616;6,384,301;5,750,871;5,463,174;和5,188,958中。用于转化植物的其他方法可以在例如Compendium of Transgenic Crop Plants(2009)Blackwell Publishing中找到。本领域技术人员已知的任何合适的方法可用于用本文提供的任何核酸转化植物细胞。

[0353] 本公开的Cas12L多肽可在体外或通过真核细胞或通过原核细胞产生，并且它可通过解折叠(例如热变性、二硫苏糖醇还原等)进一步加工，并且可使用本领域已知的方法进一步再折叠。

[0354] 不改变一级序列的目标修饰包括多肽的化学衍生化，例如酰化、乙酰化、羧化、酰胺化等。还包括糖基化的修饰，例如通过在多肽的合成和加工过程中或在进一步加工步骤中修饰多肽的糖基化形式而进行的那些修饰;例如通过将多肽暴露于影响糖基化的酶(诸如哺乳动物糖基化酶或脱糖基化酶)而进行的那些修饰。还涵盖具有磷酸化氨基酸残基例如磷酸酪氨酸、磷酸丝氨酸或磷酸苏氨酸的序列。

[0355] 还适合包括在本公开的实施方案中的是核酸(例如，编码Cas12L引导RNA、编码Cas12L融合蛋白等的核酸)和蛋白质(例如，衍生自野生型蛋白质或变体蛋白质的Cas12L融

合蛋白),所述核酸和蛋白质已使用普通分子生物学技术和合成化学进行修饰,以便改进它们对蛋白水解降解的抗性,改变靶序列特异性,优化溶解特性,改变蛋白质活性(例如,转录调节活性、酶促活性等)或使它们更合适。此类多肽的类似物包括含有除了天然存在的L-氨基酸之外的残基(例如,D-氨基酸或非天然存在的合成氨基酸)的那些多肽。D-氨基酸可取代一些或所有氨基酸残基。

[0356] 可使用如本领域已知的常规方法,通过体外合成制备本公开的Cas12L多肽。可使用各种商业合成装置,例如Applied Biosystems,Inc.、Beckman等的自动合成仪。通过使用合成仪,天然存在的氨基酸可被非天然氨基酸取代。制备的具体顺序和方式将通过方便性、经济性、所需纯度等来确定。

[0357] 如果需要,可在合成过程中或在表达过程中将各种基团引入肽中,这允许连接至其他分子或表面。因此半胱氨酸可用来制备硫醚、组氨酸用于连接至金属离子络合物,羧基用于形成酰胺或酯,氨基用于形成酰胺等。

[0358] 还可根据常规重组合成方法分离和纯化本公开的Cas12L多肽。可由表达宿主制备裂解液,并且使用高效液相色谱法(HPLC)、排阻色谱法、凝胶电泳、亲和色谱法或其他纯化技术来纯化裂解液。大多数情况下,相对于与产物制备及其纯化的方法相关的污染物,所使用的组合物将占所需产物的20重量%或更多、更通常地75重量%或更多、优选地95重量%或更多,并且出于治疗目的通常为99.5重量%或更多。通常,百分数将基于总蛋白。因此,在一些情况下,本公开的Cas12L多肽或Cas12L融合多肽具有至少80%纯度、至少85%纯度、至少90%纯度、至少95%纯度、至少98%纯度或至少99%纯度(例如,不含污染物、非Cas12L蛋白质或其他大分子等)。

[0359] 为了诱导对靶核酸(例如,基因组DNA)的切割或任何所需的修饰,或对与靶核酸相关联的多肽的任何所需的修饰,向细胞提供本公开的Cas12L引导RNA和/或Cas12L多肽和/或供体模板序列(无论它们作为核酸还是多肽引入)持续约30分钟至约24小时,例如1小时、1.5小时、2小时、2.5小时、3小时、3.5小时、4小时、5小时、6小时、7小时、8小时、12小时、16小时、18小时、20小时或约30分钟至约24小时的任何其他时间段,这可以约每天至约每4天的频率来重复,例如以每1.5天、每2天、每3天或约每天至约每四天的任何其他频率来重复。可一次或多次(例如一次、两次、三次或多于三次)向主题细胞提供一种或多种剂,并且在每次接触事件之后允许将细胞与所述一种或多种剂孵育持续一定时间量,例如16-24小时,在所述时间之后用新鲜培养基替代培养基并且进一步培养细胞。

[0360] 在其中向细胞提供两种或更多种不同靶向复合物(例如,与相同或不同靶核酸内的不同序列互补的两种不同Cas12L引导RNA)的情况下,可同时提供(例如,作为两种多肽和/或核酸)或同时递送所述复合物。可替代地,可连续提供复合物,例如首先提供靶向复合物,接着提供第二靶向复合物等,或反之亦然。

[0361] 为了改进DNA载体向靶细胞的递送,可例如通过使用脂质复合物(lipoplex)和聚合复合物(polyplex)保护DNA免受损伤,并且促进DNA进入细胞中。因此,在一些情况下,本公开的核酸(例如,本公开的重组表达载体)可用有组织的结构(像胶束或脂质体)中的脂质覆盖。当有组织的结构与DNA复合时,它被称为脂质复合物。存在三种类型的脂质,阴离子脂质(带负电)、中性脂质或阳离子脂质(带正电)。利用阳离子脂质的脂质复合物已被证明可用于基因转移。阳离子脂质由于其正电荷,与带负电的DNA天然复合。同样由于它们的电荷,

它们与细胞膜相互作用。然后发生脂质复合物的内吞作用，并且将DNA释放到细胞质中。阳离子脂质还可防止细胞对DNA的降解。

[0362] 聚合物与DNA的复合物称为聚合复合物。大多数聚合复合物由阳离子聚合物组成，并且它们的产生由离子相互作用调控。聚合复合物与脂质复合物的作用方法之间的一个巨大差异是聚合复合物不能将其DNA负载释放到细胞质中，为此，必须发生与内体溶解剂(溶解内吞作用期间产生的内体)诸如灭活的腺病毒共转染。然而，并非总是如此；诸如聚乙烯亚胺的聚合物与壳聚糖和三甲基壳聚糖一样，都有自己的内体破坏方法。

[0363] 树枝状聚合物，一种球形的高度支化的大分子，也可用于遗传修饰干细胞。树枝状聚合物颗粒的表面可被官能化以改变其特性。具体地，可能构建阳离子树枝状聚合物(即，具有正表面电荷的树枝状聚合物)。当存在遗传物质(诸如DNA质粒)时，电荷互补性导致核酸与阳离子树枝状聚合物的暂时缔合。树枝状聚合物-核酸复合物在到达其目的地时，可通过内吞作用被吸收到细胞中。

[0364] 在一些情况下，本公开的核酸(例如，表达载体)包含目标引导序列的插入位点。例如，核酸可包含目标引导序列的插入位点，其中所述插入位点紧邻编码Cas12L引导RNA的部分的核苷酸序列，当引导序列被改变而与所需靶序列(例如，有助于引导RNA的Cas12L结合方面的序列，例如，有助于Cas12L引导RNA的一个或多个dsRNA双链体的序列-引导RNA的这个部分也可称为引导RNA的“支架”或“恒定区”)杂交时，Cas12L引导RNA的所述部分不会改变。因此，在一些情况下，主题核酸(例如，表达载体)包含编码Cas12L引导RNA的核苷酸序列，不同的是编码引导RNA的引导序列部分的部分是插入序列(插入位点)。插入位点是用于插入所需序列的任何核苷酸序列。用于各种技术的“插入位点”是本领域的普通技术人员已知的，并且可使用任何方便的插入位点。插入位点可用于操纵核酸序列的任何方法。例如，在一些情况下，插入位点是多克隆位点(MCS)(例如，包含一个或多个限制性酶识别序列的位点)，用于不依赖于连接的克隆的位点，用于基于重组的克隆(例如，基于att位点的重组)的位点，由基于CRISPR/Cas(例如Cas9)的技术识别的核苷酸序列等。

[0365] 插入位点可以是任何期望的长度，并且可取决于插入位点的类型(例如，可取决于位点是否包含一个或多个限制性酶识别序列(以及包含多少限制性酶识别序列)，位点是否包括CRISPR/Cas蛋白的靶位点等)。在一些情况下，主题核酸的插入位点的长度为3个或更多个核苷酸(nt)(例如，长度为5个或更多个、8个或更多个、10个或更多个、15个或更多个、17个或更多个、18个或更多个、19个或更多个、20个或更多个、或者25个或更多个、或者30个或更多个nt)。在一些情况下，主题核酸的插入位点的长度具有在2至50个核苷酸(nt)的范围内(例如，2至40个nt、2至30个nt、2至25个nt、2至20个nt、5至50个nt、5至40个nt、5至30个nt、5至25个nt、5至20个nt、10至50个nt、10至40个nt、10至30个nt、10至25个nt、10至20个nt、17至50个nt、17至40个nt、17至30个nt、17至25个nt)的长度。在一些情况下，主题核酸的插入位点的长度具有在5至40个nt的范围内的长度。

[0366] 核酸修饰

[0367] 在一些实施方案中，主题核酸(例如，Cas12L引导RNA)具有一个或多个修饰(例如，碱基修饰、骨架修饰等)以对核酸提供新的或增强的特征(例如，改进的稳定性)。核苷是碱基-糖组合。核苷的碱基部分通常是杂环碱基。此类杂环碱基的两个最常见类别是嘌呤和嘧啶。核苷酸是还包含共价连接至核苷的糖部分的磷酸酯基团的核苷。对于包含戊呋喃糖的

那些核苷，磷酸酯基团可连接至糖的2'、3'或5'羟基部分。在形成寡核苷酸中，磷酸酯基团共价连接彼此相邻的核苷以形成线性聚合化合物。继而，此线性聚合化合物的各端可进一步连接以形成环状化合物，然而，线性化合物是合适的。另外，线性化合物可具有内部核苷酸碱基互补性并且因此可以为了产生完全或部分双链化合物的方式折叠。在寡核苷酸内，磷酸酯基团通常被称为形成寡核苷酸的核苷间骨架。RNA和DNA的正常键或骨架是3'到5'的磷酸二酯键。

[0368] 合适的核酸修饰包括但不限于：2'O甲基修饰的核苷酸、2'氟修饰的核苷酸、锁核酸(LNA)修饰的核苷酸、肽核酸(PNA)修饰的核苷酸、具有硫代磷酸酯键的核苷酸和5'帽(例如，7-甲基鸟苷酸帽(m7G))。下文描述另外的细节和另外的修饰。

[0369] 2'-O-甲基修饰的核苷酸(也称为2'-O-甲基RNA)是在tRNA和其他小RNA中发现的天然存在的RNA修饰，其作为转录后修饰而出现。可直接合成含有2'-O-甲基RNA的寡核苷酸。这种修饰增加RNA:RNA双链体的Tm，但仅导致RNA:DNA稳定性的微小变化。它对于单链核糖核酸酶的攻击是稳定的，并且对DNA酶的易感性通常是DNA的5至10倍低。它通常用于反义寡核苷酸中，作为增加稳定性和对于靶信使的结合亲和力的手段。

[0370] 2'氟修饰的核苷酸(例如，2'氟碱基)具有氟修饰的核糖，其增加结合亲和力(Tm)并且与天然RNA相比还赋予一定程度的相对核酸酶抗性。这些修饰通常用于核酶和siRNA中以改进在血清或其他生物体液中的稳定性。

[0371] LNA碱基具有对核糖骨架的修饰，其将碱基锁定在C3'-内部位置，这有利于RNA A型螺旋双链体几何结构。这种修饰显著增加Tm并且还具有非常强的核酸酶抗性。可将多个LNA插入置于寡核苷酸中的除了3'末端之外的任何位置。已经描述了从反义寡核苷酸到杂交探针到SNP检测和等位基因特异性PCR的应用。由于LNA赋予Tm的大量增加，它们还可引起引物二聚体形成以及自发夹的形成的增加。在一些情况下，并入单个寡核苷酸中的LNA的数量是10个碱基或更少。

[0372] 硫代磷酸酯(PS)键(即，硫代磷酸酯键联)用硫原子取代核酸(例如，寡核苷酸)的磷酸酯骨架中的非桥接氧。这种修饰使得核苷酸间键对核酸酶降解具有抗性。可在寡核苷酸的5'或3'末端的最后3-5个核苷酸之间引入硫代磷酸酯键以抑制外切核酸酶降解。在寡核苷酸内(例如，在整个寡核苷酸中)包含硫代磷酸酯键也可帮助减少内切核酸酶的攻击。

[0373] 在一些实施方案中，主题核酸具有一个或多个核苷酸，所述一个或多个核苷酸是2'-O-甲基修饰的核苷酸。在一些实施方案中，主题核酸(例如，dsRNA、siNA等)具有一个或多个2'氟修饰的核苷酸。在一些实施方案中，主题核酸(例如，dsRNA、siNA等)具有一个或多个LNA碱基。在一些实施方案中，主题核酸(例如，dsRNA、siNA等)具有通过硫代磷酸酯键连接的一个或多个核苷酸(即，主题核酸具有一个或多个硫代磷酸酯键联)。在一些实施方案中，主题核酸(例如，dsRNA、siNA等)具有5'帽(例如，7-甲基鸟苷酸帽(m7G))。在一些实施方案中，主题核酸(例如，dsRNA、siNA等)具有修饰的核苷酸的组合。例如，除具有一个或多个具有其他修饰的核苷酸(例如，2'-O-甲基核苷酸和/或2'氟修饰的核苷酸和/或LNA碱基和/或硫代磷酸酯键联)之外，主题核酸(例如，dsRNA、siNA等)可具有5'帽(例如，7-甲基鸟苷酸帽(m7G))。

[0374] 修饰的骨架和修饰的核苷间键联

[0375] 含有修饰的合适的核酸(例如，Cas12L引导RNA)的实例包括含有修饰的骨架或非

天然的核苷间键联的核酸。具有修饰的骨架的核酸包括在骨架中保留磷原子的那些核酸和在骨架中不具有磷原子的那些核酸。

[0376] 其中含有磷原子的合适的修饰的寡核苷酸骨架包括例如，硫代磷酸酯、手性硫代磷酸酯、二硫代磷酸酯、磷酸三酯、氨基烷基磷酸三酯、甲基和其他烷基磷酸酯(包括3'-亚烷基磷酸酯、5'-亚烷基磷酸酯和手性磷酸酯)、次膦酸酯、氨基磷酸酯(包括3'-氨基氨基磷酸酯和氨基烷基氨基磷酸酯)、二氨基磷酸酯、硫羰氨基磷酸酯、硫羰烷基磷酸酯、硫羰烷基磷酸三酯，具有正常3'-5'键联的硒代磷酸酯和硼代磷酸酯、这些物质的2'-5'连接类似物以及具有反极性的那些寡核苷酸骨架，其中一个或多个核苷酸间键联为3'至3'、5'至5'或2'至2'键联。具有反极性的合适的寡核苷酸在最3'核苷酸间键处包含单个3'至3'键联，即可为碱性(核碱基丢失或其被羟基替代)的单个反核苷残基。还包括各种盐(例如像钾或钠)、混合盐和游离酸形式。

[0377] 在一些实施方案中，主题核酸包含一个或多个硫代磷酸酯和/或杂原子核苷间键联，具体地是-CH_2-NH-O-CH_2-、-CH_2-N(CH_3)-O-CH_2-(称为亚甲基(甲基亚氨基)或MMI骨架)、-CH_2-O-N(CH_3)-CH_2-、-CH_2-N(CH_3)-N(CH_3)-CH_2-和-O-N(CH_3)-CH_2-CH_2-(其中天然磷酸二酯核苷酸间键联表示为-O-P(=O)(OH)-O-CH_2-)。MMI型核苷间键联公开于上文提及的美国专利号5,489,677中，所述专利的公开内容以引用方式整体并入本文。合适的酰胺核苷间键联公开于美国专利号5,602,240中，所述专利的公开内容以引用方式整体并入本文。

[0378] 还合适的是具有吗啉代骨架结构的核酸，如例如美国专利号5,034,506中所述。例如，在一些实施方案中，主题核酸包含替代核糖环的6元吗啉代环。在这些实施方案的一些实施方案中，二氨基磷酸酯或其他非磷酸二酯核苷间键联替代磷酸二酯键联。

[0379] 其中不包含磷原子的合适的修饰的多核苷酸骨架具有通过短链烷基或环烷基核苷间键联、混合杂原子和烷基或环烷基核苷间键联或一个或多个短链杂原子或杂环核苷间键联形成的骨架。这些包括：具有吗啉代键联(部分地由核苷的糖部分形成)的那些骨架；硅氧烷骨架；硫化物、亚砜和砜骨架；甲酰乙酰基和硫代甲酰乙酰基骨架；亚甲基甲酰乙酰基和硫代甲酰乙酰基骨架；核糖乙酰基(riboacetyl)骨架；含烯烃的骨架；氨基磺酸酯骨架；亚甲基亚胺基和亚甲基肼基骨架；磺酸酯和磺酰胺骨架；酰氨骨架；以及具有混合的N、O、S和CH_2组成部分的其他骨架。

[0380] 模拟物

[0381] 主题核酸可以是核酸模拟物。当对多核苷酸应用术语“模拟物”时意图包括其中仅呋喃糖环或呋喃糖环和核苷酸间键联两者被非呋喃糖基团替代的多核苷酸，仅呋喃糖环替代在本领域中也称为糖替代。维持杂环碱基部分或修饰的杂环碱基部分用于与适当的靶核酸的杂交。一种这样的核酸(已显示出具有优良杂交特性的多核苷酸模拟物)称为肽核酸(PNA)。在PNA中，多核苷酸的糖骨架被含酰胺的骨架替代，具体地被氨基乙基甘氨酸骨架替代。核苷酸被保留下来并且直接或间接键合至骨架的酰胺部分的氮杂氮原子。

[0382] 已报道具有优良杂交特性的一种多核苷酸模拟物是肽核酸(PNA)。PNA化合物中的骨架是给予PNA含酰胺骨架的两个或更多个连接的氨基乙基甘氨酸单元。杂环碱基部分直接或间接键合至骨架的酰胺部分的氮杂氮原子。描述PNA化合物制备的代表性美国专利包括但不限于：美国专利号5,539,082；5,714,331；和5,719,262，所述专利的公开内容以引用方式整体并入本文。

[0383] 已被研究的另一类多核苷酸模拟物基于具有附着至吗啉代环的杂环碱基的连接吗啉代单元(吗啉代核酸)。已报道连接吗啉代核酸中的吗啉代单体单元的许多连接基团。已选择一类连接基团来得到非离子型低聚化合物。基于非离子型吗啉代的低聚化合物不太可能与细胞蛋白质有不期望的相互作用。基于吗啉代的多核苷酸是不太可能与细胞蛋白质形成不期望的相互作用的寡核苷酸的非离子型模拟物(Dwaine A.Braasch和David R.Corey,Biochemistry,2002,41(14),4503-4510)。基于吗啉代的多核苷酸公开于美国专利号5,034,506中,所述专利的公开内容以引用方式整体并入本文。已制备了吗啉代类多核苷酸内的多种化合物,所述化合物具有连接单体亚单元的多种不同的连接基团。

[0384] 另一类多核苷酸模拟物称为环己烯基核酸(CeNA)。通常存在于DNA/RNA分子中的呋喃糖环被环己烯基环替代。已制备了CeNA DMT保护的亚磷酰胺单体并且用于根据经典亚磷酰胺化学性质的低聚化合物合成。已制备并且研究了完全修饰的CeNA低聚化合物和具有用CeNA修饰的特异性位置的寡核苷酸(参见Wang等人,J.Am.Chem.Soc.,2000,122,8595-8602,其公开内容以引用方式整体并入本文)。通常,CeNA单体并入到DNA链中增加了DNA/RNA杂交体的稳定性。CeNA寡腺苷酸与RNA和DNA互补序列形成具有与天然复合物相似的稳定性的复合物。通过NMR和圆二色性示出将CeNA结构并入天然核酸结构中的研究以继续进行简单的构象调整。

[0385] 另一种修饰包括锁核酸(LNA),其中2'-羟基连接至糖环的4'碳原子从而形成2'-C、4'-C-氧基亚甲基键联,从而形成双环糖部分。所述键可以是亚甲基(-CH_2-),即桥接2'氧原子和4'碳原子的基团,其中n为1或2(Singh等人,Chem.Commun.,1998,4,455-456,其公开内容以引用方式整体并入本文)。LNA和LNA类似物显现出与互补DNA和RNA具有非常高的双链体热稳定性(Tm=+3℃至+10℃)、朝向3'-核酸外切降解的稳定性和良好的溶解特性。已经描述了含有LNA的有效且无毒的反义寡核苷酸(例如Wahlestedt等人,Proc.Natl.Acad.Sci.U.S.A.,2000,97,5633-5638,其公开内容以引用方式整体并入本文)。

[0386] 已描述了LNA单体腺嘌呤、胞嘧啶、鸟嘌呤、5-甲基-胞嘧啶、胸腺嘧啶和尿嘧啶的合成和制备连同其低聚化以及核酸识别特性(例如,Koshkin等人,Tetrahedron,1998,54,3607-3630,其公开内容以引用方式整体并入本文)。LNA及其制备也描述于WO 98/39352和WO99/14226以及美国申请20120165514、20100216983、20090041809、20060117410、20040014959、20020094555和20020086998中,所述专利的公开内容以引用方式整体并入本文。

[0387] 修饰的糖部分

[0388] 主题核酸还可包含一个或多个取代的糖部分。合适的多核苷酸包含选自以下的糖取代基团:OH;F;O-、S-或N-烷基;O-、S-或N-烯基;O-、S-或N-炔基;或O-烷基-O-烷基,其中烷基、烯基和炔基可以是取代或未取代的C_1至C_{10}烷基或C_2至C_{10}烯基和炔基。特别合适的是:$O((CH_2)_nO)_mCH_3$、$O(CH_2)_nOCH_3$、$O(CH_2)_nNH_2$、$O(CH_2)_nCH_3$、$O(CH_2)_nONH_2$和$O(CH_2)_nON((CH_2)_nCH_3)_2$,其中n和m为1至约10。其他合适的多核苷酸包含选自以下的糖取代基团:C_1至C_{10}低级烷基、取代的低级烷基、烯基、炔基、烷芳基、芳烷基、O-烷芳基或O-芳烷基、SH、SCH_3、OCN、Cl、Br、CN、CF_3、OCF_3、$SOCH_3$、SO_2CH_3、ONO_2、NO_2、N_3、NH_2、杂环烷基、杂环烷芳基、氨基烷氨基、聚烷氨基、取代的硅烷基、RNA切割基团、报告基团、嵌入剂、具有改进寡核苷酸的药物代谢动力学

特性的基团、或具有改进寡核苷酸的药效动力学特性的基团，以及其他具有相似特性的取代基。适合的修饰包括2'-甲氧基乙氧基(2'-O-$CH_2CH_2OCH_3$，也称为2'-O-(2-甲氧基乙基)或2'-MOE)(Martin等人，Helv.Chim.Acta,1995,78,486-504，其公开内容以引入方式整体并入本文)，即烷氧基烷氧基。另外合适的修饰包括2'-二甲基氨基氧基乙氧基，即O$(CH_2)_2$ON$(CH_3)_2$基团，又称为2'-DMAOE，如在下文的实施例中所述；和2'-二甲基氨基乙氧基乙氧基(在本领域中又称为2'-O-二甲基-氨基-乙氧基-乙基或2'-DMAEOE)，即2'-O-CH_2-O-CH_2-N$(CH_3)_2$。

[0389] 其他合适的糖取代基团包括甲氧基(-O-CH_3)、氨基丙氧基(--O$CH_2CH_2CH_2NH_2$)、烯丙基(-CH_2-CH＝CH_2)、-O-烯丙基(--O--CH_2—CH＝CH_2)和氟(F)。2'-糖取代基团可处于阿拉伯糖(上)位或核糖(下)位。合适的2'-阿拉伯糖修饰是2'-F。还可在低聚化合物上的其他位置上做出相似的修饰，具体地在糖的3'末端核苷上或在2'-5'连接的寡核苷酸中的3'位置以及5'末端核苷酸的5'位置。低聚化合物还可具有替代呋喃戊糖的糖模拟物，诸如环丁基部分。

[0390] 碱基修饰和取代

[0391] 主题核酸还可包括核碱基(在本领域中常常简称为“碱基”)修饰或取代。如本文所用，“未修饰的”或“天然”核碱基包括嘌呤碱基腺嘌呤(A)和鸟嘌呤(G)以及嘧啶碱基胸腺嘧啶(T)、胞嘧啶(C)和尿嘧啶(U)。修饰的核碱基包括其他合成和天然的核碱基，诸如5-甲基胞嘧啶(5-me-C)、5-羟甲基胞嘧啶、黄嘌呤、次黄嘌呤、2-氨基腺嘌呤、腺嘌呤和鸟嘌呤的6-甲基衍生物和其他烷基衍生物、腺嘌呤和鸟嘌呤的2-丙基衍生物和其他烷基衍生物、2-硫尿嘧啶、2-硫胸腺嘧啶和2-硫胞嘧啶、5-卤代尿嘧啶和胞嘧啶、5-丙炔基(-C＝C-CH_3)尿嘧啶和胞嘧啶以及嘧啶碱基的其他炔基衍生物、6-偶氮基尿嘧啶、胞嘧啶和胸腺嘧啶、5-尿嘧啶(假尿嘧啶)、4-硫尿嘧啶、8-卤代基、8-氨基、8-巯基、8-硫烷基、8-羟基和其他8-取代的腺嘌呤和鸟嘌呤、5-卤代基(具体为5-溴代基)、5-三氟甲基和其他5-取代的尿嘧啶和胞嘧啶、7-甲基鸟嘌呤和7-甲基腺嘌呤、2-F-腺嘌呤、2-氨基-腺嘌呤、8-氮杂鸟嘌呤和8-氮杂腺嘌呤、7-脱氮鸟嘌呤和7-脱氮腺嘌呤以及3-脱氮鸟嘌呤和3-脱氮腺嘌呤。另外的修饰的核碱基包括三环嘧啶，诸如吩噁嗪胞苷(1H-嘧啶并(5,4-b)(1,4)苯并噁嗪-2(3H)-酮)、吩噻嗪胞苷(1H-嘧啶并(5,4-b)(1,4)苯并噻嗪-2(3H)-酮)、G-夹诸如取代的吩噁嗪胞苷(例如9-(2-氨基乙氧基)-H-嘧啶并(5,4-(b)(1,4)苯并噁嗪-2(3H)-酮)、咔唑胞苷(2H-嘧啶并(4,5-b)吲哚-2-酮)、吡啶并吲哚胞苷(H-吡啶并(3',2':4,5)吡咯并(2,3-d)嘧啶-2-酮)。

[0392] 杂环碱基部分还可包括其中嘌呤或嘧啶碱基被其他杂环替代的那些碱基，例如7-脱氮腺嘌呤、7-脱氮鸟苷、2-氨基吡啶和2-吡啶酮。另外的核碱基包括公开于美国专利号3,687,808中的那些、公开于The Concise Encyclopedia Of Polymer Science And Engineering,第858-859页,Kroschwitz,J.I.编John Wiley&Sons,1990中的那些、由Englisch等人,Angewandte Chemie,International Edition,1991,30,613公开的那些以及由Sanghvi,Y.S.,第15章,Antisense Research and Applications,第289-302页,Crooke,S.T.和Lebleu,B.编,CRC Press,1993公开的那些，这些文献的公开内容以引用方式整体并入本文。这些核碱基中的某些可用于增加低聚化合物的结合亲和力。这些包括5-取代的嘧啶，6-氮杂嘧啶以及N-2、N-6和O-6取代的嘌呤，包括2-氨基丙基腺嘌呤、5-丙炔基尿嘧啶和5-丙炔基胞嘧啶。5-甲基胞嘧啶取代已显示出使核酸双链体稳定性增加0.6℃-

1.2℃(Sanghvi等人编Antisense Research and Applications,CRC Press,Boca Raton,1993,第276-278页;其公开内容以引用方式整体并入本文)并且例如当与2'-O-甲氧基乙基糖修饰组合时是适合的碱基取代。

[0393] 缀合物

[0394] 主题核酸的另一种可能的修饰涉及将增强寡核苷酸的活性、细胞分布或细胞吸收的一个或多个部分或缀合物化学连接至多核苷酸。这些部分或缀合物可包括共价键合至诸如伯羟基或仲羟基的官能团的缀合物基团。缀合物基团包括但不限于嵌入剂、报告分子、多胺、聚酰胺、聚乙二醇、聚醚、增强低聚物的药效动力学特性的基团以及增强低聚物的药物代谢动力学特性的基团。合适的缀合物基团包括但不限于胆固醇、脂质、磷脂、生物素、吩嗪、叶酸酯、菲啶、蒽醌、吖啶、荧光素、罗丹明、香豆素以及染料。增强药效动力学特性的基团包括改进吸收、增强对降解的抗性和/或加强与靶核酸的序列特异性杂交的基团。增强药物代谢动力学特性的基团包括改进主题核酸的吸收、分布、代谢或排泄的基团。

[0395] 缀合物部分包括但不限于脂质部分,诸如胆固醇部分(Letsinger等人,Proc.Natl.Acad.Sci.USA,1989,86,6553-6556)、胆酸(Manoharan等人,Bioorg.Med.Chem.Let.,1994,4,1053-1060)、硫醚例如己基-S-三苯甲基硫醇(Manoharan等人,Ann.N.Y.Acad.Sci.,1992,660,306-309;Manoharan等人,Bioorg.Med.Chem.Let.,1993,3,2765-2770)、巯基胆固醇(Oberhauser等人,Nucl.Acids Res.,1992,20,533-538)、脂族链例如十二烷二醇或十一烷基残基(Saison-Behmoaras等人,EMBO J.,1991,10,1111-1118;Kabanov等人,FEBS Lett.,1990,259,327-330;Svinarchuk等人,Biochimie,1993,75,49-54)、磷脂例如二-十六烷基-外消旋-甘油或三乙铵1,2-二-O-十六烷基-外消旋-甘油-3-H-磷酸酯(Manoharan等人,Tetrahedron Lett.,1995,36,3651-3654;Shea等人,Nucl.Acids Res.,1990,18,3777-3783)、多胺或聚乙二醇链(Manoharan等人,Nucleosides&Nucleotides,1995,14,969-973),或金刚烷乙酸(Manoharan等人,Tetrahedron Lett.,1995,36,3651-3654)、棕榈基部分(Mishra等人,Biochim.Biophys.Acta,1995,1264,229-237),或十八烷基胺或己基氨基-羰基-氧基胆固醇部分(Crooke等人,J.Pharmacol.Exp.Ther.,1996,277,923-937)。

[0396] 缀合物可包括"蛋白转导结构域"或PTD(又称为CPP-细胞穿透肽),其可指促进横穿脂质双层、胶束、细胞膜、细胞器膜或囊泡膜的多肽、多核苷酸、碳水化合物或有机化合物或无机化合物。连接至另一个分子(所述分子可在小极性分子至大的高分子和/或纳米颗粒的范围内)的PTD促进分子横穿膜,例如从细胞外空间进入细胞内空间或从胞质溶胶进入细胞器(例如,细胞核)内。在一些实施方案中,PTD与外源多核苷酸的3'末端共价连接。在一些实施方案中,PTD与外源多核苷酸的5'末端共价连接。示例性PTD包括但不限于最小十一氨基酸多肽蛋白转导结构域(对应于包含YGRKKRRQRRR;SEQ ID NO:40的HIV-1TAT的残基47-57);包含足以直接进入细胞中的多个精氨酸(例如,3个、4个、5个、6个、7个、8个、9个、10个或10-50个精氨酸)的聚精氨酸序列;VP22结构域(Zender等人(2002)Cancer Gene Ther.9(6):489-96);果蝇触角足基因(Antennapedia)蛋白转导结构域(Noguchi等人(2003)Diabetes 52(7):1732-1737);截短的人降钙素肽(Trehin等人(2004)Pharm.Research 21:1248-1256);聚赖氨酸(Wender等人(2000)Proc.Natl.Acad.Sci.USA 97:13003-13008);RRQRRTSKLMKR(SEQ ID NO:41);运输蛋白(Transportan)GWTLNSAGYLLGKINLKALAALAKKIL

(SEQ ID NO:42);KALAWEAKLAKALAKALAKHLAKALAKALKCEA(SEQ ID NO:43);和RQIKIWFQNRRMKWKK(SEQ ID NO:44)。示例性PTD包括但不限于YGRKKRRQRRR(SEQ ID NO:40)、RKKRRQRRR(SEQ ID NO:45);具有3个精氨酸残基至50个精氨酸残基的精氨酸均聚物;示例性PTD结构域氨基酸序列包括但不限于以下的任何序列:YGRKKRRQRRR SEQ ID NO:40);RKKRRQRR(SEQ ID NO:46);YARAAARQARA SEQ ID NO:47);THRLPRRRRRR(SEQ ID NO:48);以及GGRRARRRRRR(SEQ ID NO:49)。在一些实施方案中,PTD是可激活的CPP(ACPP)(Aguilera等人(2009)Integr Biol(Camb)6月;1(5-6):371-381)。ACPP包括经由可切割接头连接至匹配聚阴离子(例如,Glu9或"E9")的聚阳离子CPP(例如,Arg9或"R9"),这使净电荷减小至接近零并由此抑制粘附和吸收到细胞中。当切割接头时,释放聚阴离子,局部暴露聚精氨酸和其固有的粘附性,从而"激活"ACPP以横穿膜。

[0397] 将组分引入靶细胞中

[0398] 可通过多种熟知的方法中的任一种将Cas12L引导RNA(或包含编码所述引导RNA的核苷酸序列的核酸)和/或本公开的Cas12L多肽(或包含编码所述多肽的核苷酸序列的核酸)和/或本公开的Cas12L融合多肽(或包含编码本公开的Cas12L融合多肽的核苷酸序列的核酸)和/或供体多核苷酸(供体模板)引入到宿主细胞中。

[0399] 多种化合物和方法中的任一种可用于将本公开的Cas12L系统递送至靶细胞(例如,其中Cas12L系统包含:a)本公开的Cas12L多肽和Cas12L引导RNA;b)本公开的Cas12L多肽、Cas12L引导RNA和供体模板核酸;c)本公开的Cas12L融合多肽和Cas12L引导RNA;d)本公开的Cas12L融合多肽、Cas12L引导RNA和供体模板核酸;e)编码本公开的Cas12L多肽的mRNA;和Cas12L引导RNA;f)编码本公开的Cas12L多肽的mRNA、Cas12L引导RNA和供体模板核酸;g)编码本公开的Cas12L融合多肽的mRNA和Cas12L引导RNA;h)编码本公开的Cas12L融合多肽的mRNA、Cas12L引导RNA和供体模板核酸;i)重组表达载体,其包含编码本公开的Cas12L多肽的核苷酸序列和编码Cas12L引导RNA的核苷酸序列;j)重组表达载体,其包含编码本公开的Cas12L多肽的核苷酸序列、编码Cas12L引导RNA的核苷酸序列和编码供体模板核酸的核苷酸序列;k)重组表达载体,其包含编码本公开的Cas12L融合多肽的核苷酸序列和编码Cas12L引导RNA的核苷酸序列;l)重组表达载体,其包含编码本公开的Cas12L融合多肽的核苷酸序列、编码Cas12L引导RNA的核苷酸序列和编码供体模板核酸的核苷酸序列;m)包含编码本公开的Cas12L多肽的核苷酸序列的第一重组表达载体,和包含编码Cas12L引导RNA的核苷酸序列的第二重组表达载体;n)包含编码本公开的Cas12L多肽的核苷酸序列的第一重组表达载体,和包含编码Cas12L引导RNA的核苷酸序列的第二重组表达载体;和供体模板核酸;o)包含编码本公开的Cas12L融合多肽的核苷酸序列的第一重组表达载体,和包含编码Cas12L引导RNA的核苷酸序列的第二重组表达载体;p)包含编码本公开的Cas12L融合多肽的核苷酸序列的第一重组表达载体,和包含编码Cas12L引导RNA的核苷酸序列的第二重组表达载体;和供体模板核酸;q)重组表达载体,其包含编码本公开的Cas12L多肽的核苷酸序列、编码第一Cas12L引导RNA的核苷酸序列和编码第二Cas12L引导RNA的核苷酸序列;或r)重组表达载体,其包含编码本公开的Cas12L融合多肽的核苷酸序列、编码第一Cas12L引导RNA的核苷酸序列和编码第二Cas12L引导RNA的核苷酸序列;或者是(a)到(r)之一的某种变体。作为非限制性实例,本公开的Cas12L系统可与脂质组合。作为另一个非限制性实例,本公开的Cas12L系统可与颗粒组合或配制成颗粒。

[0400] 将核酸引入宿主细胞中的方法在本领域中是已知的,并且可使用任何方便的方法来将主题核酸(例如,表达构建体/载体)引入靶细胞(例如,原核细胞、真核细胞、植物细胞、动物细胞、哺乳动物细胞、人细胞等)中。合适的方法包括例如病毒感染、转染、缀合、原生质体融合、脂质体转染、电穿孔、磷酸钙沉淀、聚乙烯亚胺(PEI)介导的转染、DEAE-葡聚糖介导的转染、脂质体介导的转染、粒子枪技术、磷酸钙沉淀、直接微注射、纳米颗粒介导的核酸递送(参见例如,Panyam等人Adv Drug Deliv Rev.2012年9月13日.pii:S0169-409X(12)00283-9.doi:10.1016/j.addr.2012.09.023)等。

[0401] 在一些情况下,本公开的Cas12L多肽作为编码Cas12L多肽的核酸(例如,mRNA、DNA、质粒、表达载体、病毒载体等)提供。在一些情况下,本公开的Cas12L多肽直接作为蛋白质(例如,不与相关联的引导RNA一起或与相关联的引导RNA一起,即作为核糖核蛋白复合物)提供。可通过任何方便的方法将本公开的Cas12L多肽引入细胞中(提供至细胞);此类方法是本领域的普通技术人员已知的。作为说明性实例,可将本公开的Cas12L多肽直接注射到细胞中(例如,与或不与Cas12L引导RNA或编码Cas12L引导RNA的核酸一起,并且与或不与供体多核苷酸一起)。作为另一个实例,可将本公开的Cas12L多肽和Cas12L引导RNA的预先形成的复合物(RNP)引入细胞(例如,真核细胞)中(例如,通过注射、通过核转染;通过缀合至一种或多种组分的蛋白转导结构域(PTD),例如缀合至Cas12L蛋白、缀合至引导RNA、缀合至本公开的Cas12L多肽以及引导RNA;等)。

[0402] 在一些情况下,Cas12L融合多肽(例如,融合到融合配偶体的dCas12L、融合到融合配偶体的切口酶Cas12L等)作为编码Cas12L融合多肽的核酸(例如,mRNA、DNA、质粒、表达载体、病毒载体等)提供。在一些情况下,本公开的Cas12L融合多肽直接作为蛋白质(例如,不与相关联的引导RNA一起或与相关联的引导RNA一起,即作为核糖核蛋白复合物)提供。本公开的Cas12L融合多肽可通过任何方便的方法引入到细胞中(提供给细胞);此类方法是本领域普通技术人员已知的。作为说明性实例,本公开的Cas12L融合多肽可以直接注射到细胞中(例如,与或不与编码Cas12L引导RNA的核酸一起,并且与或不与供体多核苷酸一起)。作为另一个实例,可将本公开的Cas12L融合多肽和Cas12L引导RNA(RNP)的预形成复合物引入到细胞中(例如,通过注射;通过核转染;通过与一种或多种组分缀合的蛋白质转导结构域(PTD),例如与Cas12L融合蛋白缀合,与引导RNA缀合,与本公开的Cas12L融合多肽和引导RNA缀合;等)。

[0403] 在一些情况下,核酸(例如,Cas12L引导RNA;包含编码本公开的Cas12L多肽的核苷酸序列的核酸;等)被递送至细胞(例如,靶宿主细胞)和/或在颗粒中或与颗粒结合的多肽(例如,Cas12L多肽;Cas12L融合多肽)。在一些情况下,将本公开的Cas12L系统递送至颗粒中的或与颗粒缔合的细胞。术语“颗粒”和“纳米颗粒”可适当地互换使用。包含编码本公开的Cas12L多肽的核苷酸序列和/或Cas12L引导RNA的重组表达载体、包含编码本公开的Cas12L多肽的核苷酸序列的mRNA以及引导RNA可使用颗粒或脂质包膜同时递送;例如,Cas12L多肽和Cas12L引导RNA,例如作为复合物(例如,核糖核蛋白(RNP)复合物)可通过颗粒递送,例如通过包含脂质或类脂质以及亲水性聚合物(例如,阳离子脂质和亲水聚合物)的递送颗粒递送,例如,其中阳离子脂质包括1,2-二油酰基-3-三甲基铵-丙烷(DOTAP)或1,2-二十四烷酰基-sn-甘油基-3-磷酸胆碱(DMPC)并且/或者其中亲水性聚合物包括乙二醇或聚乙二醇(PEG);并且/或者其中颗粒还包含胆固醇(例如,来自制剂1的颗粒＝DOTAP

100、DMPC 0、PEG 0、胆固醇0;制剂编号2＝DOTAP 90、DMPC 0、PEG 10、胆固醇0;制剂编号3＝DOTAP 90、DMPC 0、PEG 5、胆固醇5)。例如,可使用多步骤方法形成颗粒,其中将Cas12L多肽和Cas12L引导RNA例如以1:1的摩尔比、例如在室温下、例如持续30分钟、例如在无菌无核酸酶的1x磷酸盐缓冲盐水(PBS)中混合在一起;并且将适用于制剂的DOTAP、DMPC、PEG和胆固醇单独地溶于醇(例如,100％乙醇),并且将两种溶液混合在一起以形成含有复合物的颗粒)。

[0404] 本公开的Cas12L多肽(或包含编码本公开的Cas12L多肽的核苷酸序列的mRNA;或包含编码本公开的Cas12L多肽的核苷酸序列的重组表达载体)和/或Cas12L引导RNA(或核酸,诸如一种或多种编码Cas12L引导RNA的表达载体)可使用颗粒或脂质包膜同时递送。例如,可使用具有由磷脂双层壳包封的聚(β-氨基酯)(PBAE)核的可生物降解的核壳结构的纳米颗粒。在一些情况下,使用基于自组装生物粘附聚合物的颗粒/纳米颗粒;此类颗粒/纳米颗粒可应用于肽的口服递送、肽的静脉内递送和肽的鼻内递送,例如递送至脑。还考虑了其他实施方案,诸如疏水性药物的口服吸收和眼部递送。可使用分子包膜技术,其涉及受保护并递送至疾病部位的工程化聚合物包膜。可以单剂量或多剂量使用约5mg/kg的剂量,这取决于各种因素,例如靶组织。

[0405] 脂质类化合物(例如,如美国专利申请20110293703中所述)也可用于施用多核苷酸,并可用于递送本公开的Cas12L多肽、本公开的Cas12L融合多肽、本公开的RNP、本公开的核酸或本公开的Cas12L系统(例如,其中Cas12L系统包含:a)本公开的Cas12L多肽和Cas12L引导RNA;b)本公开的Cas12L多肽、Cas12L引导RNA和供体模板核酸;c)本公开的Cas12L融合多肽和Cas12L引导RNA;d)本公开的Cas12L融合多肽、Cas12L引导RNA和供体模板核酸;e)编码本公开的Cas12L多肽的mRNA;和Cas12L引导RNA;f)编码本公开的Cas12L多肽的mRNA、Cas12L引导RNA和供体模板核酸;g)编码本公开的Cas12L融合多肽的mRNA和Cas12L引导RNA;h)编码本公开的Cas12L融合多肽的mRNA、Cas12L引导RNA和供体模板核酸;i)重组表达载体,其包含编码本公开的Cas12L多肽的核苷酸序列和编码Cas12L引导RNA的核苷酸序列;j)重组表达载体,其包含编码本公开的Cas12L多肽的核苷酸序列、编码Cas12L引导RNA的核苷酸序列和编码供体模板核酸的核苷酸序列;k)重组表达载体,其包含编码本公开的Cas12L融合多肽的核苷酸序列和编码Cas12L引导RNA的核苷酸序列;l)重组表达载体,其包含编码本公开的Cas12L融合多肽的核苷酸序列、编码Cas12L引导RNA的核苷酸序列和编码供体模板核酸的核苷酸序列;m)包含编码本公开的Cas12L多肽的核苷酸序列的第一重组表达载体,和包含编码Cas12L引导RNA的核苷酸序列的第二重组表达载体;n)包含编码本公开的Cas12L多肽的核苷酸序列的第一重组表达载体,和包含编码Cas12L引导RNA的核苷酸序列的第二重组表达载体;和供体模板核酸;o)包含编码本公开的Cas12L融合多肽的核苷酸序列的第一重组表达载体,和包含编码Cas12L引导RNA的核苷酸序列的第二重组表达载体;p)包含编码本公开的Cas12L融合多肽的核苷酸序列的第一重组表达载体,和包含编码Cas12L引导RNA的核苷酸序列的第二重组表达载体;和供体模板核酸;q)重组表达载体,其包含编码本公开的Cas12L多肽的核苷酸序列、编码第一Cas12L引导RNA的核苷酸序列和编码第二Cas12L引导RNA的核苷酸序列;或r)重组表达载体,其包含编码本公开的Cas12L融合多肽的核苷酸序列、编码第一Cas12L引导RNA的核苷酸序列和编码第二Cas12L引导RNA的核苷酸序列;或者是(a)到(r)之一的某种变体。在一方面,将氨基醇类脂质化合物与待递送至

细胞或受试者的剂组合以形成微颗粒、纳米颗粒、脂质体或胶束。氨基醇类脂质化合物可以与其他氨基醇类脂质化合物、聚合物(合成的或天然的)、表面活性剂、胆固醇、碳水化合物、蛋白质、脂质等组合以形成颗粒。然后可任选地将这些颗粒与药物赋形剂组合以形成药物组合物。

[0406] 聚(β-氨基醇)(PBAA)可用于将本公开的Cas12L多肽、本公开的Cas12L融合多肽、本公开的RNP、本公开的核酸或本公开的Cas12L系统递送至靶细胞。美国专利公开号20130302401涉及使用组合聚合制备的一类聚(β-氨基醇)(PBAA)。

[0407] 可使用基于糖的颗粒,例如,如参考WO2014118272(以引用方式并入本文)和Nair,J K等人,2014,Journal of the American Chemical Society 136(49),16958-16961)所述的GalNAc可用于将本公开的Cas12L多肽、本公开的Cas12L融合多肽、本公开的RNP、本公开的核酸或本公开的Cas12L系统递送至靶细胞。

[0408] 在一些情况下,脂质纳米颗粒(LNP)用于将本公开的Cas12L多肽、本公开的Cas12L融合多肽、本公开的RNP、本公开的核酸或本公开的Cas12L系统递送至靶细胞。带负电的聚合物(诸如RNA)可在低pH值(例如,pH 4)下装载到LNP中,其中可电离的脂质显示正电荷。然而,在生理pH值下,LNP表现出与较长的循环时间相容的低表面电荷。已经关注了四种可电离的阳离子脂质,即1,2-二亚油酰基-3-二甲基铵-丙烷(DLinDAP)、1,2-二亚油基氧基-3-N,N-二甲基氨基丙烷(DLinDMA)、1,2-二亚油基氧基-酮基-N,N-二甲基-3-氨基丙烷(DLinKDMA)和1,2-二亚油基-4-(2-二甲基氨基乙基)-[1,3]-二氧戊环(DLinKC2-DMA)。LNP的制备描述于例如Rosin等人(2011)Molecular Therapy 19:1286-2200)中。可使用阳离子脂质1,2-二亚油酰基-3-二甲基铵-丙烷(DLinDAP)、1,2-二亚油基氧基-3-N,N-二甲基氨基丙烷(DLinDMA)、1,2-二亚油基氧基酮基-N,N-二甲基-3-氨基丙烷(DLinK-DMA)、1,2-二亚油基-4-(2-二甲基氨基乙基)-[1,3]-二氧戊环(DLinKC2-DMA)、(3-o-[2''-(甲氧基聚乙二醇2000)琥珀酰基]-1,2-二肉豆蔻酰基-sn-乙二醇(PEG-S-DMG),以及R-3-[(ω-甲氧基-聚(乙二醇)2000)氨甲酰基]-1,2-二肉豆蔻酰氧基丙基-3-胺(PEG-C-DOMG)。核酸(例如,Cas12L引导RNA;本公开的核酸等)可包封在含有DLinDAP、DLinDMA、DLinK-DMA和DLinKC2-DMA(阳离子脂质:DSPC:CHOL:PEGS-DMG或PEG-C-DOMG的摩尔比为40:10:40:10)的LNP中。在一些情况下,并入0.2%SP-DiOC18。

[0409] 球形核酸(SNA™)构建体和其它纳米颗粒(特别是金纳米颗粒)可用于将本公开的Cas12L多肽、本公开的Cas12L融合多肽、本公开的RNP、本公开的核酸或本公开的Cas12L系统递送至靶细胞。参见例如Cutler等人,J.Am.Chem.Soc.2011 133:9254-9257;Hao等人,Small.2011 7:3158-3162;Zhang等人,ACS Nano.2011 5:6962-6970;Cutler等人,J.Am.Chem.Soc.2012 134:1376-1391;Young等人,Nano Lett.2012 12:3867-71;Zheng等人,Proc.Natl.Acad.Sci.USA.2012109:11975-80;Mirkin,Nanomedicine 2012 7:635-638;Zhang等人,J.Am.Chem.Soc.2012 134:16488-1691;Weintraub,Nature 2013495:S14-S16;Choi等人,Proc.Natl.Acad.Sci.USA.2013 110(19):7625-7630;Jensen等人,Sci.Transl.Med.5,209ra152(2013)以及Mirkin等人,Small,10:186-192。

[0410] 具有RNA的自组装纳米颗粒可以用聚乙烯亚胺(PEI)来构建,所述聚乙烯亚胺用连接在聚乙二醇(PEG)远端的Arg-Gly-Asp(RGD)肽配体聚乙二醇化。

[0411] 一般来讲,“纳米颗粒”是指具有小于1000nm的直径的任何颗粒。在一些情况下,适

用于将本公开的Cas12L多肽、本公开的Cas12L融合多肽、本公开的RNP、本公开的核酸或本公开的Cas12L系统递送至靶细胞的纳米颗粒具有500nm或更小，例如，25nm至35nm、35nm至50nm、50nm至75nm、75nm至100nm、100nm至150nm、150nm至200nm、200nm至300nm、300nm至400nm或400nm至500nm的直径。在一些情况下，适用于将本公开的Cas12L多肽、本公开的Cas12L融合多肽、本公开的RNP、本公开的核酸或本公开的Cas12L系统递送至靶细胞的纳米颗粒具有25nm至200nm的直径。在一些情况下，适用于将本公开的Cas12L多肽、本公开的Cas12L融合多肽、本公开的RNP、本公开的核酸或本公开的Cas12L系统递送至靶细胞的纳米颗粒具有100nm或更小的直径。在一些情况下，适用于将本公开的Cas12L多肽、本公开的Cas12L融合多肽、本公开的RNP、本公开的核酸或本公开的Cas12L系统递送至靶细胞的纳米颗粒具有35nm至60nm的直径。

[0412] 适用于将本公开的Cas12L多肽、本公开的Cas12L融合多肽、本公开的RNP、本公开的核酸或本公开的Cas12L系统递送至靶细胞的纳米颗粒可以不同的形式提供，例如，作为固体纳米颗粒(例如，金属(诸如银、金、铁、钛)、非金属、基于脂质的固体、聚合物)、纳米颗粒的悬浮液或它们的组合提供。可制备金属、介电和半导体纳米颗粒，以及混合结构(例如，核壳纳米颗粒)。如果由半导体材料制成的纳米颗粒足够小(通常低于10nm)以致发生电子能级的量子化，则也可将它们标记量子点。此类纳米级颗粒在生物医学应用中用作药物运载体或成像剂，并且可适用于本公开中的相似目的。

[0413] 半固体和软纳米颗粒也适用于将本公开的Cas12L多肽、本公开的Cas12L融合多肽、本公开的RNP、本公开的核酸或本公开的Cas12L系统递送至靶细胞。具有半固体性质的原型纳米颗粒是脂质体。

[0414] 在一些情况下，外来体用于将本公开的Cas12L多肽、本公开的Cas12L融合多肽、本公开的RNP、本公开的核酸或本公开的Cas12L系统递送至靶细胞。外泌体是内源性纳米囊泡，其运输RNA和蛋白质，并且可将RNA递送至脑和其他靶器官。

[0415] 在一些情况下，脂质体用于将本公开的Cas12L多肽、本公开的Cas12L融合多肽、本公开的RNP、本公开的核酸或本公开的Cas12L系统递送至靶细胞。脂质体是球形囊泡结构，其由围绕内部水性隔室的单层或多层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。尽管当脂质膜与水性溶液混合时，脂质体形成是自发的，但是也可通过使用匀化器、超声波破碎仪或挤出装置以摇动的形式施加力来加速脂质体的形成。可将若干种其他添加剂添加到脂质体中以便改变它们的结构和特性。例如，可将胆固醇或鞘磷脂添加到脂质体混合物中，以便帮助稳定脂质体结构并防止脂质体内容物(inner cargo)的泄漏。脂质体制剂可主要由以下组成：天然磷脂和脂质，诸如1,2-二硬脂酰基-sn-甘油基-3-磷脂酰胆碱(DSPC)、鞘磷脂、卵磷脂酰胆碱和单唾液酸神经节苷脂。

[0416] 稳定的核酸-脂质颗粒(SNALP)可用于将本公开的Cas12L多肽、本公开的Cas12L融合多肽、本公开的RNP、本公开的核酸或本公开的Cas12L系统递送至靶细胞。SNALP制剂可含有2:40:10:48摩尔百分比的脂质3-N-[(甲氧基聚(乙二醇)2000)氨基甲酰基]-1,2-二肉豆蔻酰氧基-丙胺(PEG-C-DMA)、1,2-二亚油基氧基-N,N-二甲基-3-氨基丙烷(DLinDMA)、1,2-二硬脂酰基-sn-甘油基-3-磷酸胆碱(DSPC)和胆固醇。可通过使用25:1的脂质/siRNA比和48/40/10/2摩尔比的胆固醇/D-Lin-DMA/DSPC/PEG-C-DMA来配制D-Lin-DMA和PEG-C-DMA以

及二硬脂酰基磷脂酰胆碱(DSPC)、胆固醇和siRNA来制备SNALP脂质体。所得的SNALP脂质体的尺寸可以是约80-100nm。SNALP可包含合成胆固醇(Sigma-Aldrich,St Louis,Mo.,USA)、二棕榈酰磷脂酰胆碱(Avanti Polar Lipids,Alabaster,Ala.,USA)、3-N-[(w-甲氧基聚(乙二醇)2000)氨基甲酰基]-1,2-二肉豆蔻酰氧基丙胺和阳离子1,2-二亚油基氧基-3-N,N二甲基氨基丙烷。SNALP可以包含合成胆固醇(Sigma-Aldrich)、1,2-二硬脂酰基-sn-甘油基-3-磷酸胆碱(DSPC;Avanti Polar Lipids Inc.)、PEG-cDMA和1,2-二亚油基氧基-3-(N;N-二甲基)氨基丙烷(DLinDMA)。

[0417] 其它阳离子脂质,例如氨基脂质2,2-二油烯基-4-二甲氨基乙基-[1,3]-二氧戊环(DLin-KC2-DMA)可用于将本公开的Cas12L多肽、本公开的Cas12L融合多肽、本公开的RNP、本公开的核酸或本公开的Cas12L系统递送至靶细胞。可考虑具有以下脂质组成的预成形的囊泡:摩尔比分别为40/10/40/10的并且FVII siRNA/总脂质比为大约0.05(w/w)的氨基脂质、二硬脂酰磷脂酰胆碱(DSPC)、胆固醇和(R)-2,3-双(十八烷氧基)丙基-1-(甲氧基聚(乙二醇)2000)丙基氨基甲酸酯(PEG-脂质)。为了确保在70-90nm范围内的窄粒径分布和0.11.+-.0.04(n＝56)的低多分散指数,可在添加指导RNA之前将颗粒通过80nm膜挤出最高达三次。可使用含有高效氨基脂质16的颗粒,其中四种脂质组分16、DSPC、胆固醇和PEG-脂质的摩尔比(50/10/38.5/1.5)可进一步优化以增强体内活性。

[0418] 脂质可用本公开的Cas12L系统或其一种或多种组分或编码其的核酸配制以形成脂质纳米颗粒(LNP)。合适的脂质包括但不限于DLin-KC2-DMA4、C12-200和辅脂质(colipid)二硬脂酰磷脂酰胆碱、胆固醇和PEG-DMG可用本公开的Cas12L系统或其组分使用自发的囊泡形成程序配制。组分摩尔比可以是约50/10/38.5/1.5(DLin-KC2-DMA或C12-200/二硬脂酰磷脂酰胆碱/胆固醇/PEG-DMG)。

[0419] 本公开的Cas12L系统或其组分可包封在PLGA微球中递送,所述微球诸如在美国公布申请20130252281和20130245107和20130244279中进一步描述的微球。

[0420] 增压蛋白可用于将本公开的Cas12L多肽、本公开的Cas12L融合多肽、本公开的RNP、本公开的核酸或本公开的Cas12L系统递送至靶细胞。超电荷蛋白是一类工程化或天然存在的蛋白质,其具有异常高的正或负净理论电荷。超负电荷蛋白和超正电荷蛋白均表现出耐受热或化学诱导的聚集的能力。超正电荷蛋白也能够穿透哺乳动物细胞。使货物与这些蛋白质(诸如质粒DNA、RNA或其他蛋白质)缔合可实现这些大分子在体外和体内向哺乳动物细胞的功能性递送。

[0421] 细胞穿透肽(CPPs)可用于将本公开的Cas12L多肽、本公开的Cas12L融合多肽、本公开的RNP、本公开的核酸或本公开的Cas12L系统递送至靶细胞。CPP通常具有以下氨基酸组成,其含有高相对丰度的带正电荷的氨基酸(诸如赖氨酸或精氨酸),或者具有含有极性/带电荷氨基酸和非极性疏水氨基酸的交替模式的序列。

[0422] 可使用可植入装置将本公开的Cas12L多肽、本公开的Cas12L融合多肽、本公开的RNP、本公开的核酸(例如,Cas12L引导RNA、编码Cas12L引导RNA的核酸、编码Cas12L多肽的核酸、供体模板等)或本公开的Cas12L系统递送至靶细胞(例如,体内靶细胞,其中靶细胞是循环中的靶细胞、组织中的靶细胞、器官中的靶细胞等)。适用于将本公开的Cas12L多肽、本公开的Cas12L融合多肽、本公开的RNP、本公开的核酸或本公开的Cas12L系统递送至靶细胞(例如,体内靶细胞,其中靶细胞是循环中的靶细胞、组织中的靶细胞、器官中的靶细胞等)

的可植入装置可包括容器(例如,储库、基质等),所述容器包含Cas12L多肽、Cas12L融合多肽、RNP或Cas12L系统(或其组分,例如本公开的核酸)。

[0423] 合适的可植入装置可包括例如用作装置主体的聚合物基底(诸如基质),并且在一些情况下包括另外的支架材料(诸如金属或另外的聚合物),以及增强可见性和成像的材料。可植入递送装置可以有利地提供局部和长时间的释放,其中待递送的多肽和/或核酸被直接释放到靶位点,例如细胞外基质(ECM)、肿瘤周围的血管系统、患病组织等。合适的可植入递送装置包括适用于递送至诸如腹腔的腔和/或其中药物递送系统未被锚定或附着的任何其他类型的施用的装置,包括生物稳定和/或可降解和/或可生物吸收的聚合物基质,其可以例如任选地为基质。在一些情况下,合适的可植入药物递送装置包含可降解聚合物,其中主要释放机制是整体侵蚀(bulk erosion)。在一些情况下,合适的可植入药物递送装置包含不可降解或缓慢降解的聚合物,其中主要释放机制是扩散而不是整体侵蚀,使得外部部分用作膜并且其内部部分用作药物储库,实际上,所述药物储库长时间内(例如约一周至约几个月)不会受到周围环境的影响。也可任选地使用具有不同释放机制的不同聚合物的组合。在总释放期的有效期内,浓度梯度可保持有效恒定,并且因此扩散速率是有效恒定的(称为“零模式”扩散)。术语“恒定”意指扩散速率维持高于治疗有效性的下阈值,但其仍然任选地以初始突发为特征并且/或者可波动,例如增加和降低到某一程度。扩散速率可长时间这样维持,并且可认为扩散速率恒定到某一水平以优化治疗有效期,例如有效的沉默期。

[0424] 在一些情况下,可植入递送系统被设计成保护基于核苷酸的治疗剂免于降解,无论是化学性质还是由于受试者体内酶和其他因素的攻击而引起的降解。

[0425] 可选择装置的植入位点或靶位点,用于获得最大的治疗功效。例如,递送装置可植入在肿瘤环境内或附近,或者与肿瘤相关联的血液供给内或附近。靶位置可以是,例如:1)大脑退化位点,如在帕金森病或阿尔茨海默病中在基底神经节、白质和灰质处;2)脊柱,如就肌萎缩侧索硬化症(ALS)而言;3)子宫颈;4)活动性和慢性炎症关节;5)真皮,如就牛皮癣而言;7)交感神经和感觉神经位点,用于镇痛作用;7)骨;8)急性或慢性感染位点;9)阴道内;10)内耳-听觉系统、内耳迷路、前庭系统;11)气管内;12)心内;冠状动脉、心外膜;13)泌尿道或膀胱;14)胆系统;15)实质组织,包括但不限于肾、肝、脾;16)淋巴结;17)唾液腺;18)牙龈;19)关节内(到关节中);20)眼内;21)脑组织;22)脑室;23)腔,包括腹腔(例如但不限于卵巢癌);24)食管内;和25)直肠内;和26)到脉管系统中。

[0426] 插入方法(诸如植入)可任选地已经用于其他类型的组织植入和/或用于插入和/或用于组织取样,任选地无需修改,或者可替代地仅在此类方法中任选地进行非主要修改。此类方法任选地包括但不限于近距离放射治疗方法、活组织检查、使用和/或不使用超声的内窥镜检查(诸如进入脑组织的立体定位方法)、腹腔镜检查(包括用腹腔镜植入关节、腹部器官、膀胱壁和体腔中)。

[0427] 修饰的宿主细胞

[0428] 本公开提供一种修饰的细胞,所述修饰的细胞包含本公开的Cas12L多肽和/或包含编码本公开的Cas12L多肽的核苷酸序列的核酸。本公开提供一种修饰的细胞,所述修饰的细胞包含本公开的Cas12L多肽,其中所述修饰的细胞是通常不包含本公开的Cas12L多肽的细胞。本公开提供一种修饰的细胞(例如,遗传修饰的细胞),所述修饰的细胞包含核酸,所述核酸包含编码本公开的Cas12L多肽的核苷酸序列。本公开提供一种用mRNA遗传修饰的

遗传修饰的细胞，所述mRNA包含编码本公开的Cas12L多肽的核苷酸序列。本公开提供一种用重组表达载体遗传修饰的遗传修饰的细胞，所述重组表达载体包含编码本公开的Cas12L多肽的核苷酸序列。本公开提供一种用重组表达载体遗传修饰的遗传修饰细胞，所述重组表达载体包含：a)编码本公开的Cas12L多肽的核苷酸序列；和b)编码本公开的Cas12L引导RNA的核苷酸序列。本公开提供一种用重组表达载体遗传修饰的遗传修饰的细胞，所述重组表达载体包含：a)编码本公开的Cas12L多肽的核苷酸序列；b)编码本公开的Cas12L引导RNA的核苷酸序列；和c)编码供体模板的核苷酸序列。

[0429] 用作本公开的Cas12L多肽和/或包含编码本公开的Cas12L多肽和/或本公开的Cas12L引导RNA的核苷酸序列的核酸的受体的细胞可以是多种细胞中的任一种，这些细胞包括例如体外细胞；体内细胞；离体细胞；原代细胞；癌细胞；动物细胞；植物细胞；藻类细胞；真菌细胞等。用作本公开的Cas12L多肽和/或包含编码本公开的Cas12L多肽和/或本公开的Cas12L引导RNA的核苷酸序列的核酸的受体的细胞被称为“宿主细胞”或“靶细胞”。宿主细胞或靶细胞可以是本公开的Cas12L系统的受体。宿主细胞或靶细胞可以是本公开的Cas12L RNP的受体。宿主细胞或靶细胞可以是本公开的Cas12L系统的单一组分的受体。

[0430] 细胞(靶细胞)的非限制性实例包括：原核细胞、真核细胞、细菌细胞、古细菌细胞、单细胞真核生物体的细胞、原生动物细胞、来自植物的细胞(例如，来自植物作物、水果、蔬菜、谷物、大豆、玉米(corn)、玉米(maize)、小麦、种子、番茄、大米、木薯、甘蔗、南瓜、干草、马铃薯、棉花、大麻、烟草、开花植物、针叶树、裸子植物、被子植物、蕨类植物、石松类、角苔类、苔类、苔藓、双子叶植物、单子叶植物等的细胞)、藻类细胞(例如，布朗葡萄藻(Botryococcus braunii)、莱茵衣藻(Chlamydomonas reinhardtii)、海洋富油微拟球藻(Nannochloropsis gaditana)、蛋白核小球藻(Chlorella pyrenoidosa)、展枝马尾藻(Sargassum patens)、羽藻(C.agardh)等)、海藻(例如巨藻(kelp))、真菌细胞(例如，酵母细胞、来自蘑菇的细胞)、动物细胞、来自无脊椎动物(例如，果蝇、刺胞动物、棘皮动物、线虫等)的细胞、来自脊椎动物(例如，鱼类、两栖动物、爬行动物、鸟类、哺乳动物)的细胞、来自哺乳动物(例如，有蹄类动物(例如，猪、牛、山羊、绵羊)；啮齿动物(例如，大鼠、小鼠)；非人灵长类动物；人；猫科动物(例如，猫)；犬(例如，狗)等)的细胞等。在一些情况下，细胞是不源自天然生物体的细胞(例如，细胞可以是合成制得的细胞；也称为人造细胞)。

[0431] 细胞可以是体外细胞(例如，建立的培养细胞系)。细胞可以是离体细胞(来自个体的培养细胞)。细胞可以是体内细胞(例如，个体中的细胞)。细胞可以是分离的细胞。细胞可以是生物体内部的细胞。细胞可以是生物体。细胞可以是细胞培养物(例如，体外细胞培养物)中的细胞。细胞可以是细胞集合中的一者。细胞可以是原核细胞或衍生自原核细胞。细胞可以是细菌细胞或可衍生自细菌细胞。细胞可以是古细菌细胞或衍生自古细菌细胞。细胞可以是真核细胞或衍生自真核细胞。细胞可以是植物细胞或衍生自植物细胞。细胞可以是动物细胞或衍生自动物细胞。细胞可以是无脊椎动物细胞或衍生自无脊椎动物细胞。细胞可以是脊椎动物细胞或衍生自脊椎动物细胞。细胞可以是哺乳动物细胞或衍生自哺乳动物细胞。细胞可以是啮齿动物细胞或衍生自啮齿动物细胞。细胞可以是人细胞或衍生自人细胞。细胞可以是微生物细胞或衍生自微生物细胞。细胞可以是真菌细胞或衍生自真菌细胞。细胞可以是昆虫细胞。细胞可以是节肢动物细胞。细胞可以是原生动物细胞。细胞可以是蠕虫细胞。

[0432] 合适的细胞包括干细胞(例如胚胎干(ES)细胞、诱导多能干(iPS)细胞;生殖细胞(例如,卵母细胞、精子、卵原细胞、精原细胞等);体细胞,例如成纤维细胞、少突胶质细胞、神经胶质细胞、造血细胞、神经元、肌细胞、骨细胞、肝细胞、胰腺细胞等。

[0433] 合适的细胞包括人胚胎干细胞、胚胎心肌细胞、肌成纤维细胞、间充质干细胞、自体移植的扩增的心肌细胞、脂肪细胞、全能细胞、多能细胞、血液干细胞、成肌细胞、成体干细胞、骨髓细胞、间充质细胞、胚胎干细胞、实质细胞、上皮细胞、内皮细胞、间皮细胞、成纤维细胞、成骨细胞、软骨细胞、外源细胞、内源细胞、干细胞、造血干细胞、骨髓衍生祖细胞、心肌细胞、骨骼细胞、胎儿细胞、未分化细胞、多能祖细胞、单能祖细胞、单核细胞、心脏成肌细胞、骨骼成肌细胞、巨噬细胞、毛细血管内皮细胞、异种细胞、同种异体细胞和产后干细胞。

[0434] 在一些情况下,细胞是免疫细胞、神经元、上皮细胞和内皮细胞或干细胞。在一些情况下,免疫细胞是T细胞、B细胞、单核细胞、天然杀伤细胞、树突状细胞或巨噬细胞。在一些情况下,免疫细胞是细胞毒性T细胞。在一些情况下,免疫细胞是辅助性T细胞。在一些情况下,免疫细胞是调节性T细胞(Treg)。

[0435] 在一些情况下,细胞是干细胞。干细胞包括成体干细胞。成体干细胞也称为体细胞干细胞。

[0436] 成体干细胞驻留在分化组织中,但保留自我更新的特性和产生多种细胞类型的能力,通常是干细胞所存在于的组织中的典型细胞类型。体细胞干细胞的许多实例是本领域的技术人员已知的,包括肌肉干细胞;造血干细胞;上皮干细胞;神经干细胞;间充质干细胞;乳腺干细胞;肠干细胞;中胚层干细胞;内皮干细胞;嗅干细胞;神经嵴干细胞等。

[0437] 目标干细胞包括哺乳动物干细胞,其中术语“哺乳动物”是指被分类为哺乳动物的任何动物,包括人;非人灵长类动物;家畜和农场动物;以及动物园、实验室、运动或宠物动物,诸如狗、马、猫、牛、小鼠、大鼠、兔等。在一些情况下,干细胞是人干细胞。在一些情况下,干细胞是啮齿动物(例如,小鼠;大鼠)干细胞。在一些情况下,干细胞是非人灵长类动物干细胞。

[0438] 干细胞可以表达一种或多种干细胞标记物,例如SOX9、KRT19、KRT7、LGR5、CA9、FXYD2、CDH6、CLDN18、TSPAN8、BPIFB1、OLFM4、CDH17和PPARGC1A。

[0439] 在一些实施方案中,干细胞是造血干细胞(HSC)。HSC是中胚层衍生的细胞,其可从骨髓、血液、脐带血、胎儿肝脏和卵黄囊中分离。HSC的特征在于$CD34^+$和$CD3^-$。HSC可在体内重新生成红系细胞、中性粒细胞-巨噬细胞、巨核细胞和淋巴样造血细胞谱系。在体外,可诱导HSC经历至少一些自我更新的细胞分裂,并且可诱导HSC分化成与体内所见相同的谱系。因此,可诱导HSC分化成红系细胞、巨核细胞、中性粒细胞、巨噬细胞和淋巴细胞中的一种或多种。

[0440] 在其他实施方案中,干细胞是神经干细胞(NSC)。神经干细胞(NSC)能够分化成神经元和神经胶质细胞(包括少突胶质细胞和星形胶质细胞)。神经干细胞是能够进行多次分裂的多能干细胞,并且在特定条件下可产生作为神经干细胞的子细胞,或可作为成神经细胞或成胶质细胞的神经祖细胞,例如,分别致力于成为一种或多种类型的神经元和神经胶质细胞的细胞。获得NSC的方法在本领域中是已知的。

[0441] 在其他实施方案中,干细胞是间充质干细胞(MSC)。MSC最初衍生自胚胎中胚层并

从成人骨髓中分离，可分化形成肌肉、骨、软骨、脂肪、骨髓基质和肌腱。分离MSC的方法在本领域中是已知的；并且可使用任何已知的方法来获得MSC。参见例如美国专利号5,736,396，其描述了人MSC的分离。

[0442] 在一些情况下，细胞是植物细胞。植物细胞可以是单子叶植物的细胞。细胞可以是双子叶植物的细胞。

[0443] 在一些情况下，细胞是植物细胞。例如，细胞可以是主要农业植物的细胞，例如大麦、豆类(干食用)、油菜、玉米、棉花(皮玛棉)、棉花(陆地棉)、亚麻籽、干草(苜蓿)、干草(非苜蓿)、燕麦、花生、大米、高粱、大豆、甜菜、甘蔗、向日葵(油)、向日葵(非油)、甘薯、烟草(白肋烟)、烟草(烤烟)、番茄、小麦(硬质小麦)、小麦(春小麦)、小麦(冬小麦)等。作为另一个实例，细胞是蔬菜作物的细胞，所述蔬菜作物包括但不限于例如，苜蓿芽、芦荟叶、葛根、慈菇、朝鲜蓟、芦笋、竹笋、香蕉花、豆芽、豆类、甜菜叶、甜菜、苦瓜、白菜、西兰花、球花甘蓝(芜菁)、球芽甘蓝、卷心菜、卷心菜芽、仙人掌叶(仙人掌果)、笋瓜、刺棘蓟、胡萝卜、花椰菜、芹菜、佛手瓜、中国洋蓟、大白菜、中国芹菜、中国韭菜、菜心、菊花叶(茼蒿)、羽衣甘蓝、玉米秸秆、甜玉米、黄瓜、白萝卜、蒲公英嫩叶、芋头、dau mue(豌豆尖)、donqua(冬瓜)、茄子、菊苣、莴苣、琴头蕨、田地水芹、苦苣、盖菜(芥菜)、gailon、良姜(暹罗、泰国姜)、大蒜、姜根、牛蒡、嫩叶、汉诺威沙拉用绿叶、huauzontle、洋姜、豆薯、羽衣甘蓝嫩叶、大头菜、羊腿藜、生菜(贝比生菜)、生菜(波士顿生菜)、生菜(波士顿红生菜)、生菜(绿叶)、生菜(冰山生菜)、生菜(红毛菜)、生菜(绿橡树叶)、生菜(红橡树叶)、生菜(加工生菜)、生菜(红叶)、生菜(罗马生菜)、生菜(红罗马生菜)、生菜(俄罗斯红芥末)、linkok、白萝卜、长豆、莲藕、野苣、龙舌兰(龙舌兰)叶、黄肉芋、混和生菜、京水菜、moap(光滑丝瓜)、moo、moqua(有绒毛的南瓜)、蘑菇、芥末、山药、秋葵、通菜、洋葱嫩叶、opo(长南瓜)、观赏玉米、观赏葫芦、欧芹、欧洲防风草、豌豆、辣椒(铃铛型)、辣椒、南瓜、菊苣、萝卜芽、萝卜、青芸苔、青芸苔、大黄、罗马生菜、芜菁甘蓝、盐角草(海豆)、丝瓜(角形/脊状丝瓜)、菠菜、南瓜、稻草捆、甘蔗、甘薯、唐莴苣、罗望子、芋艿、芋艿叶、芋艿芽、塌棵菜、tepeguaje(葫芦)、红瓜、粘果酸浆、番茄、番茄(樱桃型)、番茄(葡萄型)、番茄(李子型)、姜黄、芜菁茎嫩叶、芜菁、荸荠、薯蓣、山药(名称)、油菜、木薯(木薯)等。

[0444] 在一些情况下，植物细胞是植物组分的细胞，例如叶、茎、根、种子、花、花粉、花药、胚珠、花梗、果实、分生组织、子叶、下胚轴、荚果、胚、胚乳、外植体、愈伤组织或芽。

[0445] 在一些情况下，细胞是节肢动物细胞。例如，细胞可以是以下的亚目、家族、亚家族、群体、亚群或物种的细胞：例如，有螯肢亚门(Chelicerata)、多足亚门(Myriapodia)、Hexipodia、蛛形纲(Arachnida)、昆虫纲(Insecta)、石蛃目(Archaeognatha)、缨尾目(Thysanura)、古翅下纲(Palaeoptera)、蜉蝣目(Ephemeroptera)、蜻蜓目(Odonata)、差翅亚目(Anisoptera)、束翅亚目(Zygoptera)、新翅亚纲(Neoptera)、外翅总目(Exopterygota)、襀翅目(Plecoptera)、纺足目(Embioptera)、直翅目(Orthoptera)、缺翅目(Zoraptera)、革翅目(Dermaptera)、网翅目(Dictyoptera)、蛩蠊目(Notoptera)、蛩蠊科(Grylloblattidae)、螳科(Mantophasmatidae)、竹节虫目(Phasmatodea)、蜚蠊目(Blattaria)、等翅目(Isoptera)、螳螂目(Mantodea)、Parapneuroptera、啮虫目(Psocoptera)、缨翅目(Thysanoptera)、虱毛目(Phthiraptera)、半翅目(Hemiptera)、内翅类(Endopterygota)或全变态类(Holometabola)、膜翅目(Hymenoptera)、鞘翅目

(Coleoptera)、捻翅目(Strepsiptera)、蛇蛉目(Raphidioptera)、广翅目(Megaloptera)、脉翅目(Neuroptera)、长翅目(Mecoptera)、蚤目(Siphonaptera)、双翅目(Diptera)、毛翅目(Trichoptera)或鳞翅目(Lepidoptera)。

[0446] 在一些情况下,细胞是昆虫细胞。例如,在一些情况下,细胞是蚊子、蚱蜢、半翅目昆虫、苍蝇、跳蚤、蜜蜂、黄蜂、蚂蚁、虱子、蛾或甲虫的细胞。

[0447] 试剂盒

[0448] 本公开提供一种试剂盒,所述试剂盒包括本公开的Cas12L系统或本公开的Cas12L系统的组分。

[0449] 本公开的试剂盒可包括:a)本公开的Cas12L多肽和Cas12L引导RNA;b)本公开的Cas12L多肽、Cas12L引导RNA和供体模板核酸;c)本公开的Cas12L融合多肽和Cas12L引导RNA;d)本公开的Cas12L融合多肽、Cas12L引导RNA和供体模板核酸;e)编码本公开的Cas12L多肽的mRNA;和Cas12L引导RNA;f)编码本公开的Cas12L多肽的mRNA、Cas12L引导RNA和供体模板核酸;g)编码本公开的Cas12L融合多肽的mRNA和Cas12L引导RNA;h)编码本公开的Cas12L融合多肽的mRNA、Cas12L引导RNA和供体模板核酸;i)重组表达载体,其包含编码本公开的Cas12L多肽的核苷酸序列和编码Cas12L引导RNA的核苷酸序列;j)重组表达载体,其包含编码本公开的Cas12L多肽的核苷酸序列、编码Cas12L引导RNA的核苷酸序列和编码供体模板核酸的核苷酸序列;k)重组表达载体,其包含编码本公开的Cas12L融合多肽的核苷酸序列和编码Cas12L引导RNA的核苷酸序列;l)重组表达载体,其包含编码本公开的Cas12L融合多肽的核苷酸序列、编码Cas12L引导RNA的核苷酸序列和编码供体模板核酸的核苷酸序列;m)包含编码本公开的Cas12L多肽的核苷酸序列的第一重组表达载体,和包含编码Cas12L引导RNA的核苷酸序列的第二重组表达载体;n)包含编码本公开的Cas12L多肽的核苷酸序列的第一重组表达载体,和包含编码Cas12L引导RNA的核苷酸序列的第二重组表达载体;和供体模板核酸;o)包含编码本公开的Cas12L融合多肽的核苷酸序列的第一重组表达载体,和包含编码Cas12L引导RNA的核苷酸序列的第二重组表达载体;p)包含编码本公开的Cas12L融合多肽的核苷酸序列的第一重组表达载体,和包含编码Cas12L引导RNA的核苷酸序列的第二重组表达载体;和供体模板核酸;q)重组表达载体,其包含编码本公开的Cas12L多肽的核苷酸序列、编码第一Cas12L引导RNA的核苷酸序列和编码第二Cas12L引导RNA的核苷酸序列;或r)重组表达载体,其包含编码本公开的Cas12L融合多肽的核苷酸序列、编码第一Cas12L引导RNA的核苷酸序列和编码第二Cas12L引导RNA的核苷酸序列;或者是(a)到(r)之一的某种变体。

[0450] 本公开的试剂盒可包括:a)如上所述的本公开的Cas12L系统的组分,或者可包含本公开的Cas12L系统;和b)一种或多种另外的试剂,例如,i)缓冲剂;ii)蛋白酶抑制剂;iii)核酸酶抑制剂;iv)显影或可视化可检测标记所需的试剂;v)阳性和/或阴性对照靶DNA;vi)阳性和/或阴性对照Cas12L引导RNA等。本公开的试剂盒可包括:a)如上所述的本公开的Cas12L系统的组分,或者可包含本公开的Cas12L系统;和b)治疗剂。

[0451] 本公开的试剂盒可包括重组表达载体,所述重组表达载体包含:a)用于插入核酸的插入位点,所述核酸包含编码Cas12L引导RNA的一部分的核苷酸序列,所述Cas12L引导RNA的一部分与靶核酸中的靶核苷酸序列杂交;和b)编码Cas12L引导RNA的Cas12L结合部分的核苷酸序列。本公开的试剂盒可包括重组表达载体,所述重组表达载体包含:a)用于插入

核酸的插入位点,所述核酸包含编码Cas12L引导RNA的一部分的核苷酸序列,所述Cas12L引导RNA的一部分与靶核酸中的靶核苷酸序列杂交;b)编码Cas12L引导RNA的Cas12L结合部分的核苷酸序列;和c)编码本公开的Cas12L多肽的核苷酸序列。

[0452] 实用性

[0453] 本公开的Cas12L多肽或本公开的Cas12L融合多肽可用于多种方法中(例如,与Cas12L引导RNA组合,并且在一些情况下进一步与供体模板组合)。例如,本公开的Cas12L多肽可用于(i)修饰(例如切割,例如切口;甲基化等)靶核酸(DNA或RNA;单链或双链);(ii)调节靶核酸的转录;(iii)标记靶核酸;(iv)结合靶核酸(例如,用于分离、标记、成像、追踪等的目的);(v)修饰与靶核酸相关联的多肽(例如,组蛋白)等。因此,本公开提供一种修饰靶核酸的方法。在一些情况下,本公开的用于修饰靶核酸的方法包括使靶核酸与以下物质接触:a)本公开的Cas12L多肽;和b)一种或多种(例如,两种)Cas12L引导RNA。在一些情况下,本公开的用于修饰靶核酸的方法包括使靶核酸与以下物质接触:a)本公开的Cas12L多肽;b)Cas12L引导RNA;和c)供体核酸(例如,供体模板)。在一些情况下,接触步骤在体外细胞中进行。在一些情况下,接触步骤在体内细胞中进行。在一些情况下,接触步骤在离体细胞中进行。

[0454] 因为使用Cas12L多肽的方法包括将Cas12L多肽与靶核酸中的特定区域结合(通过相关联的Cas12L引导RNA靶向靶核酸中的特定区域),所述方法在本文中一般称为结合方法(例如,结合靶核酸的方法)。然而,应理解在一些情况下,虽然结合方法可能无非是导致靶核酸的结合,但在其他情况下,所述方法可具有不同的最终结果(例如,所述方法可导致靶核酸的修饰(例如,切割/甲基化等);从靶核酸转录的调节;靶核酸翻译的调节;基因组编辑;与靶核酸相关联的蛋白质的调节;靶核酸的分离等)。

[0455] 对于合适方法的实例,参见例如Jinek等人,Science.2012年8月17日;337(6096):816-21;Chylinski等人,RNA Biol.2013年5月;10(5):726-37;Ma等人,Biomed Res Int.2013;2013:270805;Hou等人,Proc Natl Acad Sci U S A.2013年9月24日;110(39):15644-9;Jinek等人,Elife.2013;2:e00471;Pattanayak等人,Nat Biotechnol.2013年9月;31(9):839-43;Qi等人,Cell.2013年2月28日;152(5):1173-83;Wang等人,Cell.2013年5月9日;153(4):910-8;Auer等人,Genome Res.2013年10月31日;Chen等人,Nucleic Acids Res.2013年11月1日;41(20):e19;Cheng等人,Cell Res.2013年10月;23(10):1163-71;Cho等人,Genetics.2013年11月;195(3):1177-80;DiCarlo等人,Nucleic Acids Res.2013年4月;41(7):4336-43;Dickinson等人,Nat Methods.2013年10月;10(10):1028-34;Ebina等人,Sci Rep.2013;3:2510;Fujii等人,Nucleic Acids Res.2013年11月1日;41(20):e187;Hu等人,Cell Res.2013年11月;23(11):1322-5;Jiang等人,Nucleic Acids Res.2013年11月1日;41(20):e188;Larson等人,Nat Protoc.2013年11月;8(11):2180-96;Mali等人,Nat Methods.2013年10月;10(10):957-63;Nakayama等人,Genesis.2013年12月;51(12):835-43;Ran等人,Nat Protoc.2013年11月;8(11):2281-308;Ran等人,Cell.2013年9月12日;154(6):1380-9;Upadhyay等人,G3(Bethesda).2013年12月9日;3(12):2233-8;Walsh等人,Proc Natl Acad Sci U S A.2013年9月24日;110(39):15514-5;Xie等人,Mol Plant.2013年10月9日;Yang等人,Cell.2013年9月12日;154(6):1370-9;以及以下美国专利和专利申请:8,906,616;8,895,308;8,889,418;8,889,356;8,871,445;8,865,406;8,795,965;8,

771,945;8,697,359;20140068797;20140170753;20140179006;20140179770;20140186843;20140186919;20140186958;20140189896;20140227787;20140234972;20140242664;20140242699;20140242700;20140242702;20140248702;20140256046;20140273037;20140273226;20140273230;20140273231;20140273232;20140273233;20140273234;20140273235;20140287938;20140295556;20140295557;20140298547;20140304853;20140309487;20140310828;20140310830;20140315985;20140335063;20140335620;20140342456;20140342457;20140342458;20140349400;20140349405;20140356867;20140356956;20140356958;20140356959;20140357523;20140357530;20140364333;和20140377868;所述文献各自特此以引用方式整体并入。

[0456] 例如,本公开提供(但不限于)切割靶核酸的方法;编辑靶核酸的方法;调节靶从核酸转录的方法;分离靶核酸的方法、结合靶核酸的方法、对靶核酸成像的方法、修饰靶核酸的方法等。

[0457] 如本文所用,术语/短语"使靶核酸,例如,与Cas12L多肽或与Cas12L融合多肽等接触",涵盖用于接触靶核酸的所有方法。例如,Cas12L多肽可以作为蛋白质、RNA(编码Cas12L多肽)或DNA(编码Cas12L多肽)提供给细胞;而Cas12L引导RNA可以作为引导RNA或编码引导RNA的核酸提供。因此,当例如在细胞中(例如,在体外细胞内部、在体内细胞内部、在离体细胞内部)执行方法时,包括接触靶核酸的方法涵盖将处于活性/最终状态的任何或所有组分(例如,呈Cas12L多肽的一种或多种蛋白质形式;呈Cas12L融合多肽的蛋白质形式;在一些情况下呈引导RNA的RNA形式)引入细胞中,并且还涵盖将编码一种或多种组分的一种或多种核酸(例如,一种或多种包含编码Cas12L多肽或Cas12L融合多肽的一种或多种核苷酸序列的核酸、一种或多种包含编码一种或多种引导RNA的一种或多种核苷酸序列的核酸、包含编码供体模板的核苷酸序列的核酸等)引入细胞中。因为所述方法也可在体外在细胞外部执行,所以包括接触靶核酸的方法(除非另外指明)涵盖在体外在细胞外部、在体外在细胞内部、在体内在细胞内部、离体在细胞内部接触等。

[0458] 在一些情况下,本公开的用于修饰靶核酸的方法包括向靶细胞中引入Cas12L基因座,例如来自包含Cas12L基因座的细胞(例如,在一些情况下,处于天然状态(天然存在的状态)包含Cas12L基因座的细胞)的核酸,所述核酸包含编码Cas12L多肽的核苷酸序列以及长度为约1千碱基(kb)至5kb的在编码Cas12L的核苷酸序列周围的核苷酸序列,其中靶细胞通常(在天然状态下)不包含Cas12L基因座。然而,可以修饰一个或多个间隔序列,一个或多个编码crRNA的编码引导序列,使得靶向一个或多个目标靶序列。因此,例如,在一些情况下,本公开的用于修饰靶核酸的方法包括向靶细胞中引入Cas12L基因座,例如,从源细胞(例如,在一些情况下,处于天然状态(天然存在的状态)包含Cas12L基因座的细胞)获得的核酸,其中核酸具有100个核苷酸(nt)至5kb(例如,100nt至500nt、500nt至1kb、1kb至1.5kb、1.5kb至2kb、2kb至2.5kb、2.5kb至3kb、3kb至3.5kb、3.5kb至4kb、或4kb至5kb)的长度并且包含编码Cas12L多肽的核苷酸序列。如上所述,在一些此类情况下,可以修饰一个或多个间隔序列,一个或多个编码crRNA的编码引导序列,使得靶向一个或多个目标靶序列。在一些情况下,所述方法包括向靶细胞中引入:i)Cas12L基因座;和ii)供体DNA模板。在一些情况下,靶核酸在体外无细胞组合物中。在一些情况下,靶核酸存在于靶细胞中。在一些情况下,靶核酸存在于靶细胞中,其中靶细胞是原核细胞。在一些情况下,靶核酸存在于靶细胞中,

其中靶细胞是真核细胞。在一些情况下，靶核酸存在于靶细胞中，其中靶细胞是哺乳动物细胞。在一些情况下，靶核酸存在于靶细胞中，其中靶细胞是植物细胞。

[0459] 在一些情况下，本公开的用于修饰靶核酸的方法包括使靶核酸与本公开的Cas12L多肽或本公开的Cas12L融合多肽接触。在一些情况下，本公开的用于修饰靶核酸的方法包括使靶核酸与Cas12L多肽和Cas12L引导RNA接触。在一些情况下，本公开的用于修饰靶核酸的方法包括使靶核酸与Cas12L多肽、第一Cas12L引导RNA和第二Cas12L引导RNA接触。在一些情况下，本公开的用于修饰靶核酸的方法包括使靶核酸与Cas12L多肽和Cas12L引导RNA和供体DNA模板接触。

[0460] 目标靶核酸和靶细胞

[0461] 当与Cas12L引导RNA结合时，本公开的Cas12L多肽或本公开的Cas12L融合多肽可结合靶核酸，并且在一些情况下，可结合并修饰靶核酸。靶核酸可以是任何核酸(例如，DNA、RNA)，可以是双链或单链的，可以是任何类型的核酸(例如，染色体(基因组DNA)、衍生自染色体、染色体DNA、质粒、病毒、细胞外、细胞内、线粒体、叶绿体、线性、环状等)并且可来自任何生物体(例如，只要Cas12L引导RNA包含与靶核酸中的靶序列杂交的核苷酸序列，使得靶核酸可被靶向即可)。

[0462] 靶核酸可以是DNA或RNA。靶核酸可以是双链的(例如，dsDNA、dsRNA)或单链的(例如，ssRNA、ssDNA)。在一些情况下，靶核酸是单链的。在一些情况下，靶核酸是单链RNA(ssRNA)。在一些情况下，靶ssRNA(例如，靶细胞ssRNA、病毒ssRNA等)选自：mRNA、rRNA、tRNA、非编码RNA(ncRNA)、长非编码RNA(lncRNA)和微小RNA(miRNA)。在一些情况下，靶核酸是单链DNA(ssDNA)(例如，病毒DNA)。如上所指出，在一些情况下，靶核酸是单链的。

[0463] 靶核酸可位于任何地方，例如，体外细胞外部、体外细胞内部、体内细胞内部、离体细胞内部。合适的靶细胞(其可包含靶核酸，诸如基因组DNA)包括但不限于：细菌细胞；古细菌细胞；单细胞真核生物体的细胞；植物细胞；藻类细胞，例如，布朗葡萄藻、莱茵衣藻、海洋富油微拟球藻、蛋白核小球藻、展枝马尾藻、羽藻等；真菌细胞(例如，酵母细胞)；动物细胞；来自无脊椎动物(例如，果蝇、刺胞动物、棘皮动物、线虫等)的细胞；昆虫(例如，蚊子；蜜蜂；农业害虫等)的细胞；蛛形纲动物(例如，蜘蛛；蜱等)的细胞；来自脊椎动物(例如，鱼类、两栖动物、爬行动物、鸟类、哺乳动物)的细胞；来自哺乳动物的细胞(例如，来自啮齿动物的细胞；来自人类的细胞；非人哺乳动物的细胞；啮齿动物(例如，小鼠、大鼠)的细胞；兔形目动物(例如，兔)的细胞；有蹄类动物(例如，牛、马、骆驼、美洲驼、骆马、绵羊、山羊等)的细胞；海洋哺乳动物(例如，鲸鱼、海豹、象海豹、海豚、海狮等)的细胞等。任何类型的细胞都可以是感兴趣的(例如干细胞，例如胚胎干(ES)细胞、诱导多能干(iPS)细胞、生殖细胞(例如，卵母细胞、精子、卵原细胞、精原细胞等)、成体干细胞、体细胞(例如，成纤维细胞)、造血细胞、神经元、肌肉细胞、骨细胞、肝细胞、胰腺细胞；在任何阶段下胚胎的体外或体内胚胎细胞(例如，1个细胞、2个细胞、4个细胞、8个细胞等阶段斑马鱼胚胎)等)。

[0464] 细胞可来自已建立的细胞系或者它们可以是原代细胞，其中“原代细胞”、“原代细胞系”和“原代培养物”在本文中可互换使用，是指衍生自受试者并且允许培养物在体外生长有限次数的传代(即，分裂)的细胞和细胞培养物。例如，原代培养物是可能已传代0次、1次、2次、4次、5次、10次或15次但不足以通过转折期的次数的培养物。通常，原代细胞系在体外维持少于10代。靶细胞可以是单细胞生物体并且/或者可在培养物中生长。如果细胞是原

代细胞，它们可通过任何方便的方法从个体收获。例如，白细胞可通过血浆分离置换法、白细胞血浆分离置换法、密度梯度分离等方便地收获，而来自组织(诸如皮肤、肌肉、骨髓、脾、肝、胰腺、肺、肠、胃等)的细胞可通过活组织检查方便地收获。

[0465] 在上述申请的一些申请中，主题方法可用于在体内和/或离体和/或体外的有丝分裂细胞或有丝分裂后细胞中诱导靶核酸切割、靶核酸修饰和/或结合靶核酸(例如，用于可视化，用于采集和/或分析等)(例如，以破坏由靶向mRNA编码的蛋白质的产生，以切割或以其他方式修饰靶DNA，以遗传修饰靶细胞等)。因为引导RNA通过与靶核酸杂交来提供特异性，所以在公开的方法中目标有丝分裂细胞和/或有丝分裂后细胞可包括来自任何生物体的细胞(例如，细菌细胞；古细菌细胞；单细胞真核生物体的细胞；植物细胞；藻类细胞，例如布朗葡萄藻、莱茵衣藻、海洋富油微拟球藻、蛋白核小球藻、展枝马尾藻、羽藻等；真菌细胞(例如，酵母细胞)；动物细胞；来自无脊椎动物(例如，果蝇、刺胞动物、棘皮动物、线虫等)的细胞；来自脊椎动物(例如，鱼类、两栖动物、爬行动物、鸟类、哺乳动物)的细胞；来自哺乳动物的细胞；来自啮齿动物的细胞；来自人的细胞等)。在一些情况下，可将主题Cas12L蛋白(和/或编码蛋白质的核酸，诸如DNA和/或RNA)和/或Cas12L引导RNA(和/或编码引导RNA的DNA)和/或供体模板和/或RNP引入个体(即，靶细胞可在体内)(例如，哺乳动物、大鼠、小鼠、猪、灵长类动物、非人灵长类动物、人等)中。在一些情况下，这种施用可例如通过编辑靶向细胞的基因组用于治疗和/或预防疾病的目的。

[0466] 植物细胞包括单子叶植物细胞和双子叶植物细胞。细胞可以是根细胞、叶细胞、木质部细胞、韧皮部细胞、形成层细胞、顶端分生组织细胞、实质细胞、厚角组织细胞、厚壁组织细胞等。植物细胞包括农作物的细胞，诸如小麦、玉米、大米、高粱、小米、大豆等的细胞。植物细胞包括农业水果和坚果植物的细胞，例如生产杏、橙子、柠檬、苹果、李子、梨、杏仁等的植物的细胞。

[0467] 靶细胞的其他实例在上文标题为“修饰的细胞”的部分中列出。细胞(靶细胞)的非限制性实例包括：原核细胞、真核细胞、细菌细胞、古细菌细胞、单细胞真核生物体的细胞、原生动物细胞、来自植物的细胞(例如，来自植物作物、水果、蔬菜、谷物、大豆、玉米(corn)、玉米(maize)、小麦、种子、番茄、大米、木薯、甘蔗、南瓜、干草、马铃薯、棉花、大麻、烟草、开花植物、针叶树、裸子植物、被子植物、蕨类植物、石松类、角苔类、苔类、苔藓、双子叶植物、单子叶植物等的细胞)、藻类细胞(例如，布朗葡萄藻、莱茵衣藻、海洋富油微拟球藻、蛋白核小球藻、展枝马尾藻、羽藻等)、海藻(例如巨藻)、真菌细胞(例如，酵母细胞、来自蘑菇的细胞)、动物细胞、来自无脊椎动物(例如，果蝇、刺胞动物、棘皮动物、线虫等)的细胞、来自脊椎动物(例如，鱼类、两栖动物、爬行动物、鸟类、哺乳动物)的细胞、来自哺乳动物(例如，有蹄类动物(例如，猪、牛、山羊、绵羊)；啮齿动物(例如，大鼠、小鼠)；非人灵长类动物；人；猫科动物(例如，猫)；犬(例如，狗)等)的细胞等。在一些情况下，细胞是不源自天然生物体的细胞(例如，细胞可以是合成制得的细胞；也称为人造细胞)。

[0468] 细胞可以是体外细胞(例如，建立的培养细胞系)。细胞可以是离体细胞(来自个体的培养细胞)。细胞可以是体内细胞(例如，个体中的细胞)。细胞可以是分离的细胞。细胞可以是生物体内部的细胞。细胞可以是生物体。细胞可以是细胞培养物(例如，体外细胞培养物)中的细胞。细胞可以是细胞集合中的一者。细胞可以是原核细胞或衍生自原核细胞。细胞可以是细菌细胞或可衍生自细菌细胞。细胞可以是古细菌细胞或衍生自古细菌细胞。细

胞可以是真核细胞或衍生自真核细胞。细胞可以是植物细胞或衍生自植物细胞。细胞可以是动物细胞或衍生自动物细胞。细胞可以是无脊椎动物细胞或衍生自无脊椎动物细胞。细胞可以是脊椎动物细胞或衍生自脊椎动物细胞。细胞可以是哺乳动物细胞或衍生自哺乳动物细胞。细胞可以是啮齿动物细胞或衍生自啮齿动物细胞。细胞可以是人细胞或衍生自人细胞。细胞可以是微生物细胞或衍生自微生物细胞。细胞可以是真菌细胞或衍生自真菌细胞。细胞可以是昆虫细胞。细胞可以是节肢动物细胞。细胞可以是原生动物细胞。细胞可以是蠕虫细胞。

[0469] 合适的细胞包括干细胞(例如胚胎干(ES)细胞、诱导多能干(iPS)细胞;生殖细胞(例如,卵母细胞、精子、卵原细胞、精原细胞等);体细胞,例如成纤维细胞、少突胶质细胞、神经胶质细胞、造血细胞、神经元、肌细胞、骨细胞、肝细胞、胰腺细胞等。

[0470] 合适的细胞包括人胚胎干细胞、胚胎心肌细胞、肌成纤维细胞、间充质干细胞、自体移植的扩增的心肌细胞、脂肪细胞、全能细胞、多能细胞、血液干细胞、成肌细胞、成体干细胞、骨髓细胞、间充质细胞、胚胎干细胞、实质细胞、上皮细胞、内皮细胞、间皮细胞、成纤维细胞、成骨细胞、软骨细胞、外源细胞、内源细胞、干细胞、造血干细胞、骨髓衍生祖细胞、心肌细胞、骨骼细胞、胎儿细胞、未分化细胞、多能祖细胞、单能祖细胞、单核细胞、心脏成肌细胞、骨骼成肌细胞、巨噬细胞、毛细血管内皮细胞、异种细胞、同种异体细胞和产后干细胞。

[0471] 在一些情况下,细胞是免疫细胞、神经元、上皮细胞和内皮细胞或干细胞。在一些情况下,免疫细胞是T细胞、B细胞、单核细胞、天然杀伤细胞、树突状细胞或巨噬细胞。在一些情况下,免疫细胞是细胞毒性T细胞。在一些情况下,免疫细胞是辅助性T细胞。在一些情况下,免疫细胞是调节性T细胞(Treg)。

[0472] 在一些情况下,细胞是干细胞。干细胞包括成体干细胞。成体干细胞也称为体细胞干细胞。

[0473] 成体干细胞驻留在分化组织中,但保留自我更新的特性和产生多种细胞类型的能力,通常是干细胞所存在于的组织中的典型细胞类型。体细胞干细胞的许多实例是本领域的技术人员已知的,包括肌肉干细胞;造血干细胞;上皮干细胞;神经干细胞;间充质干细胞;乳腺干细胞;肠干细胞;中胚层干细胞;内皮干细胞;嗅干细胞;神经嵴干细胞等。

[0474] 目标干细胞包括哺乳动物干细胞,其中术语“哺乳动物”是指被分类为哺乳动物的任何动物,包括人;非人灵长类动物;家畜和农场动物;以及动物园、实验室、运动或宠物动物,诸如狗、马、猫、牛、小鼠、大鼠、兔等。在一些情况下,干细胞是人干细胞。在一些情况下,干细胞是啮齿动物(例如,小鼠;大鼠)干细胞。在一些情况下,干细胞是非人灵长类动物干细胞。

[0475] 干细胞可以表达一种或多种干细胞标记物,例如SOX9、KRT19、KRT7、LGR5、CA9、FXYD2、CDH6、CLDN18、TSPAN8、BPIFB1、OLFM4、CDH17和PPARGC1A。

[0476] 在一些实施方案中,干细胞是造血干细胞(HSC)。HSC是中胚层衍生的细胞,其可从骨髓、血液、脐带血、胎儿肝脏和卵黄囊中分离。HSC的特征在于$CD34^+$和$CD3^-$。HSC可在体内重新生成红系细胞、中性粒细胞-巨噬细胞、巨核细胞和淋巴样造血细胞谱系。在体外,可诱导HSC经历至少一些自我更新的细胞分裂,并且可诱导HSC分化成与体内所见相同的谱系。因此,可诱导HSC分化成红系细胞、巨核细胞、中性粒细胞、巨噬细胞和淋巴细胞中的一种或多

种。

[0477] 在其他实施方案中，干细胞是神经干细胞(NSC)。神经干细胞(NSC)能够分化成神经元和神经胶质细胞(包括少突胶质细胞和星形胶质细胞)。神经干细胞是能够进行多次分裂的多能干细胞，并且在特定条件下可产生作为神经干细胞的子细胞，或可作为成神经细胞或成胶质细胞的神经祖细胞，例如，分别致力于成为一种或多种类型的神经元和神经胶质细胞的细胞。获得NSC的方法在本领域中是已知的。

[0478] 在其他实施方案中，干细胞是间充质干细胞(MSC)。MSC最初衍生自胚胎中胚层并从成人骨髓中分离，可分化形成肌肉、骨、软骨、脂肪、骨髓基质和肌腱。分离MSC的方法在本领域中是已知的；并且可使用任何已知的方法来获得MSC。参见例如美国专利号5,736,396，其描述了人MSC的分离。

[0479] 在一些情况下，细胞是植物细胞。植物细胞可以是单子叶植物的细胞。细胞可以是双子叶植物的细胞。

[0480] 在一些情况下，细胞是植物细胞。例如，细胞可以是主要农业植物的细胞，例如大麦、豆类(干食用)、油菜、玉米、棉花(皮玛棉)、棉花(陆地棉)、亚麻籽、干草(苜蓿)、干草(非苜蓿)、燕麦、花生、大米、高粱、大豆、甜菜、甘蔗、向日葵(油)、向日葵(非油)、甘薯、烟草(白肋烟)、烟草(烤烟)、番茄、小麦(硬质小麦)、小麦(春小麦)、小麦(冬小麦)等。作为另一个实例，细胞是蔬菜作物的细胞，所述蔬菜作物包括但不限于例如，苜蓿芽、芦荟叶、葛根(arrow root)、慈菇(arrowhead)、朝鲜蓟、芦笋、竹笋、香蕉花、豆芽、豆类、甜菜叶、甜菜、苦瓜、白菜、西兰花、球花甘蓝(芜菁)、球芽甘蓝、卷心菜、卷心菜芽、仙人掌叶(仙人掌果)、笋瓜、刺棘蓟、胡萝卜、花椰菜、芹菜、佛手瓜、中国洋蓟(crosne)、大白菜、中国芹菜、中国韭菜、菜心、菊花叶(茼蒿(tung ho))、羽衣甘蓝、玉米秸秆、甜玉米、黄瓜、白萝卜(daikon)、蒲公英嫩叶、芋头(dasheen)、dau mue(豌豆尖)、donqua(冬瓜)、茄子、菊苣(endive)、莴苣、琴头蕨、田地水芹、苦苣、盖菜(芥菜)、gailon、良姜(暹罗、泰国姜)、大蒜、姜根、牛蒡(gobo)、嫩叶、汉诺威沙拉用绿叶(hanover salad green)、huauzontle、洋姜(jerusalem artichoke)、豆薯、羽衣甘蓝(kale)嫩叶、大头菜(kohlrabi)、羊腿藜(quilete)、生菜(贝比生菜(bibb))、生菜(波士顿生菜(boston))、生菜(波士顿红生菜(boston red))、生菜(绿叶)、生菜(冰山生菜(iceberg))、生菜(红毛菜(lolla rossa))、生菜(绿橡树叶)、生菜(红橡树叶)、生菜(加工生菜)、生菜(红叶)、生菜(罗马生菜(romaine))、生菜(红罗马生菜(ruby romaine))、生菜(俄罗斯红芥末)、linkok、白萝卜(lobok)、长豆、莲藕、野苣(mache)、龙舌兰(龙舌兰(agave))叶、黄肉芋(malanga)、混和生菜(mesculin mix)、京水菜(mizuna)、moap(光滑丝瓜)、moo、moqua(有绒毛的南瓜)、蘑菇、芥末、山药(nagaimo)、秋葵、通菜、洋葱嫩叶、opo(长南瓜)、观赏玉米、观赏葫芦、欧芹、欧洲防风草、豌豆、辣椒(铃铛型)、辣椒、南瓜(pumpkin)、菊苣(radicchio)、萝卜芽、萝卜(radish)、青芸苔、青芸苔、大黄、罗马生菜(baby red)、芜菁甘蓝(rutabaga)、盐角草(海豆)、丝瓜(角形/脊状丝瓜)、菠菜、南瓜(squash)、稻草捆、甘蔗、甘薯、唐莴苣、罗望子、芋艿(taro)、芋艿叶、芋艿芽、塌棵菜、tepeguaje(葫芦(guaje))、红瓜(tindora)、粘果酸浆(tomatillo)、番茄、番茄(樱桃型)、番茄(葡萄型)、番茄(李子型)、姜黄、芜菁茎嫩叶、芜菁(turnip)、荸荠、薯蓣(yampi)、山药、油菜(yu choy)、木薯(yuca)(木薯)等。

[0481] 在一些情况下，细胞是节肢动物细胞。例如，细胞可以是以下的亚目、家族、亚家

族、群体、亚群或物种的细胞:例如,有螯肢亚门(Chelicerata)、多足亚门(Myriapodia)、Hexipodia、蛛形纲(Arachnida)、昆虫纲(Insecta)、石蛃目(Archaeognatha)、缨尾目(Thysanura)、古翅下纲(Palaeoptera)、蜉蝣目(Ephemeroptera)、蜻蜓目(Odonata)、差翅亚目(Anisoptera)、束翅亚目(Zygoptera)、新翅亚纲(Neoptera)、外翅总目(Exopterygota)、襀翅目(Plecoptera)、纺足目(Embioptera)、直翅目(Orthoptera)、缺翅目(Zoraptera)、革翅目(Dermaptera)、网翅目(Dictyoptera)、蛩蠊目(Notoptera)、蛩蠊科(Grylloblattidae)、螳科(Mantophasmatidae)、竹节虫目(Phasmatodea)、蜚蠊目(Blattaria)、等翅目(Isoptera)、螳螂目(Mantodea)、Parapneuroptera、啮虫目(Psocoptera)、缨翅目(Thysanoptera)、虱毛目(Phthiraptera)、半翅目(Hemiptera)、内翅类(Endopterygota)或全变态类(Holometabola)、膜翅目(Hymenoptera)、鞘翅目(Coleoptera)、捻翅目(Strepsiptera)、蛇蛉目(Raphidioptera)、广翅目(Megaloptera)、脉翅目(Neuroptera)、长翅目(Mecoptera)、蚤目(Siphonaptera)、双翅目(Diptera)、毛翅目(Trichoptera)或鳞翅目(Lepidoptera)。

[0482] 在一些情况下,细胞是昆虫细胞。例如,在一些情况下,细胞是蚊子、蚱蜢、半翅目昆虫、苍蝇、跳蚤、蜜蜂、黄蜂、蚂蚁、虱子、蛾或甲虫的细胞。

[0483] 将组分引入靶细胞中

[0484] 可通过多种众所周知的方法将Cas12L引导RNA(或包含编码所述引导RNA的核苷酸序列的核酸)和/或Cas12L融合多肽(或包含编码所述融合多肽的核苷酸序列的核酸)和/或供体多核苷酸引入到宿主细胞中。

[0485] 将核酸引入到细胞中的方法是本领域已知的,并且可以使用任何方便的方法将核酸(例如,表达构建体)引入到靶细胞(例如,真核细胞、人细胞、干细胞、祖细胞等)中。合适的方法更详细地描述于本文其他地方并且包括例如病毒或噬菌体感染、转染、缀合、原生质体融合、脂质体转染、电穿孔、磷酸钙沉淀、聚乙烯亚胺(PEI)介导的转染、DEAE-葡聚糖介导的转染、脂质体介导的转染、粒子枪技术、磷酸钙沉淀、直接微注射、纳米颗粒介导的核酸递送(参见例如,Panyam等人Adv Drug Deliv Rev.2012年9月13日.pii:S0169-409X(12)00283-9.doi:10.1016/j.addr.2012.09.023)等。可以使用已知的方法,例如诸如核转染,将任何或所有组分作为组合物(例如,包括下列的任何方便的组合:Cas12L多肽、Cas12L引导RNA、供体多核苷酸等)引入到细胞中。

[0486] 供体多核苷酸(供体模板)

[0487] 在Cas12L引导RNA的引导下,Cas12L蛋白在一些情况下在双链DNA(dsDNA)靶核酸内生成位点特异性双链断裂(DSB)或单链断裂(SSB)(例如,当Cas12L蛋白是切口酶变体时),这些断裂通过非同源末端连接(NHEJ)或同源定向重组(HDR)修复。

[0488] 在一些情况下,使靶DNA(与Cas12L蛋白和Cas12L引导RNA)接触在允许非同源末端连接或同源定向修复的条件下发生。因此,在一些情况下,主题方法包括使靶DNA与供体多核苷酸接触(例如,通过将供体多核苷酸引入细胞中),其中将供体多核苷酸、供体多核苷酸的一部分、供体多核苷酸的拷贝或供体多核苷酸的拷贝的一部分整合到靶DNA中。在一些情况下,所述方法不包括使细胞与供体多核苷酸接触,并且修饰靶DNA使得靶DNA内的核苷酸缺失。

[0489] 在一些情况下,Cas12L引导RNA(或编码Cas12L引导RNA的DNA)和Cas12L蛋白(或编

码Cas12L蛋白的核酸，诸如RNA或DNA，例如一种或多种表达载体)与供体多核苷酸序列共同施用(例如，与靶核酸接触、向细胞施用等)，所述供体多核苷酸序列包括与靶DNA序列同源的至少一个区段，主题方法可用于将核酸物质添加(即插入或替代)到靶DNA序列(例如以“敲入”核酸，例如编码蛋白质、siRNA、miRNA的核酸等)，添加标签(例如，6xHis、荧光蛋白(例如，绿色荧光蛋白;黄色荧光蛋白等)、血凝素(HA)、FLAG等)，将调控序列添加到基因(例如启动子、聚腺苷酸化信号、内部核糖体进入序列(IRES)、2A肽、起始密码子、终止密码子、剪接信号、定位信号等)，修饰核酸序列(例如，引入突变、通过引入正确的序列去除致病突变)等。因此，包含Cas12L引导RNA和Cas12L蛋白质的复合物可用于任何体外或体内应用中，在所述应用中希望以位点特异性(即“靶向的”)方式修饰DNA，例如基因敲除、基因敲入、基因编辑、基因标签等，例如在例如治疗疾病或作为抗病毒、抗病原体或抗癌治疗剂的基因疗法、出于治疗、诊断或研究目的通过细胞进行的大规模蛋白质生产、iPS细胞诱导、生物研究、用于缺失或替代的病原体基因的靶向等中所使用。

[0490] 在其中希望将多核苷酸序列插入靶序列被切割的基因组中的应用中，还可向细胞提供供体多核苷酸(包含供体序列的核酸)。“供体序列”或“供体多核苷酸”或“供体模板”意指待在Cas12L蛋白切割的位点处插入的核酸序列(例如，在dsDNA切割之后、对靶DNA进行切口之后、对靶DNA进行双切口之后等)。供体多核苷酸可含有与靶位点处的基因组序列足够的同源性(例如与侧接靶位点，例如在靶位点的约50个或更少的碱基内(例如约30个碱基内、约15个碱基内、约10个碱基内、约5个碱基内)的核苷酸序列或直接侧接靶位点的核苷酸序列，具有70％、80％、85％、90％、95％或100％的同源性)，以支持所述供体多核苷酸与和其具有同源性的基因组序列之间的同源定向修复。在供体与基因组序列之间具有序列同源性的大约25个、50个、100个或200个核苷酸或多于200个核苷酸(或10与200之间任何整数值的核苷酸或更多)可支持同源定向修复。供体多核苷酸可具有任何长度，例如10个核苷酸或更多、50个核苷酸或更多、100个核苷酸或更多、250个核苷酸或更多、500个核苷酸或更多、1000个核苷酸或更多、5000个核苷酸或更多等。

[0491] 供体序列通常不与它替代的基因组序列相同。而且，供体序列相对于基因组序列可含有至少一个或多个单个碱基变化、插入、缺失、反转或重排，只要存在足够同源性以支持同源定向修复即可(例如，用于基因校正，例如，以转化致病碱基对或非致病碱基对)。在一些实施方案中，供体序列包含侧接两个同源区的非同源序列，以使得靶DNA区域与两个侧接序列之间的同源定向修复导致在靶区域处插入非同源序列。供体序列还可包含载体骨架，所述载体骨架含有不与目标DNA区域同源并且不意图插入到目标DNA区域中的序列。通常，供体序列的一个或多个同源区将与希望与其重组的基因组序列具有至少50％的序列同一性。在某些实施方案中，存在60％、70％、80％、90％、95％、98％、99％或99.9％的序列同一性。根据供体多核苷酸的长度，可存在1％与100％之间的任何值的序列同一性。

[0492] 供体序列与基因组序列相比可包含某些序列差异，例如限制位点、核苷酸多态性、可选择标记(例如，抗药基因、荧光蛋白、酶等)等，所述序列差异可用来评估供体序列在切割位点处的成功插入或在一些情况下可用于其他目的(例如，表示靶向基因组基因座处的表达)。在一些情况下，如果位于编码区中，此类核苷酸序列差异将不会改变氨基酸序列，或将产生沉默氨基酸变化(即，不影响蛋白质结构或功能的变化)。可替代地，这些序列差异可包括侧接重组序列，诸如FLP、loxP序列等，所述侧接重组序列可在去除标记序列之后的时

间里激活。

[0493] 在一些情况下,供体序列作为单链DNA提供给细胞。在一些情况下,供体序列作为双链DNA提供给细胞。它可以线性或环状形式引入细胞中。如果以线性形式引入,供体序列的末端可通过任何方便的方法来保护(例如,免受核酸外切降解),并且此类方法是本领域的技术人员已知的。例如,可将一个或多个双脱氧核苷酸残基添加到线性分子的3'末端,并且/或者可将自身互补寡核苷酸连接至一个或两个末端。参见例如Chang等人(1987)Proc.Natl.Acad Sci USA84:4959-4963;Nehls等人(1996)Science 272:886-889。用于保护外源多核苷酸免受降解的另外方法包括但不限于添加一个或多个末端氨基以及使用修饰的核苷酸间键联,例如像硫代磷酸酯、氨基磷酸酯和O-甲基核糖或脱氧核糖残基。作为保护线性供体序列的末端的替代方案,可在同源区外部包括额外长度的序列,所述序列可在不影响重组的情况下降解。可将供体序列作为载体分子的一部分引入细胞中,所述载体分子具有另外的序列,例如像复制起点、启动子和编码抗生素耐药性的基因。此外,供体序列可作为裸核酸、作为与剂(诸如脂质体或泊洛沙姆)复合的核酸引入,或者可通过病毒(例如,腺病毒AAV)来递送,如本文其他地方对于编码Cas12L引导RNA和/或Cas12L融合多肽和/或供体多核苷酸的核酸所述。

[0494] 检测方法

[0495] 一旦通过检测靶DNA(双链或单链)而被激活,本公开的Cas12L多肽可混杂地切割非靶向核酸(例如,单链DNA(ssDNA)或RNA)。在一些情况下,一旦本公开的Cas12L多肽被引导RNA激活,这在引导RNA与靶DNA的靶序列杂交(即,样品包含靶DNA)时发生,Cas12L多肽变成混杂切割ssDNA的核酸酶(即,核酸酶切割非靶ssDNA,即不与引导RNA的引导序列杂交的ssDNA)。在一些情况下,一旦本公开的Cas12L多肽被引导RNA激活,这在引导RNA与靶DNA的靶序列杂交(即,样品包含靶DNA)时发生,Cas12L多肽变成混杂切割RNA的核酸酶(即,核酸酶切割非靶RNA,即不与引导RNA的引导序列杂交的RNA)。因此,当样品中存在靶DNA时(例如,在一些情况下超过阈值量),引起样品中ssDNA或RNA的切割,这可使用任何方便的检测方法(例如,使用标记的单链检测剂DNA或检测剂RNA)加以检测。

[0496] 用于检测靶DNA的组合物和方法

[0497] 提供了用于检测样品中的靶DNA(双链或单链)的组合物和方法。在一些情况下,使用为单链(ssDNA)并且不与引导RNA的引导序列杂交的检测剂DNA(即,检测剂ssDNA是非靶ssDNA)。此类方法可包括(a)使样品与以下物质接触:(i)本公开的Cas12L多肽;(ii)引导RNA,所述引导RNA包含与Cas12L多肽结合的区域和与靶DNA杂交的引导序列;和(iii)为单链的且不与引导RNA的引导序列杂交的检测剂DNA;以及(b)测量由Cas12L多肽切割单链检测剂DNA而产生的可检测信号,从而检测靶DNA。如上所指出,一旦本公开的Cas12L多肽被引导RNA激活(在样品中包含与引导RNA杂交的靶DNA(即,样品中包含靶向靶DNA)时发生),Cas12L多肽就会被激活并且充当内切核糖核酸酶来非特异性地切割样品中存在的ssDNA(包括非靶ssDNA)。因此,当样品中存在所靶向的靶DNA时(例如,在一些情况下超过阈值量),引起样品中ssDNA(包括非靶ssDNA)的切割,这可使用任何方便的检测方法(例如,使用标记的检测剂ssDNA)加以检测。

[0498] 还提供了用于切割单链DNA(ssDNA)(例如,非靶ssDNA)的组合物和方法。此类方法可包括使核酸群体与以下物质接触,其中所述群体包含靶DNA和多个非靶ssDNA:(i)本公开

的Cas12L多肽;和(ii)引导RNA,所述引导RNA包含与Cas12L多肽结合的区域和与靶DNA杂交的引导序列,其中Cas12L多肽切割所述多个非靶ssDNA。可使用这种方法,例如,在细胞中切割外源ssDNA(例如,病毒DNA)。

[0499] 主题方法的接触步骤可在包含二价金属离子的组合物中进行。接触步骤可在无细胞环境中,例如在细胞外部进行。接触步骤可在细胞内部进行。接触步骤可在体外细胞中进行。接触步骤可在离体细胞中进行。接触步骤可在体内细胞中进行。

[0500] 引导RNA可以RNA的形式或以编码引导RNA的核酸(例如,DNA,诸如重组表达载体)的形式提供。Cas12L多肽可以蛋白质或以编码蛋白质的核酸(例如,mRNA、DNA诸如重组表达载体)的形式提供。在一些情况下,可通过(例如,使用前体引导RNA阵列,其可被Cas12L效应蛋白切割成单个(“成熟”)引导RNA)来提供两个或更多个(例如,3个或更多个、4个或更多个、5个或更多个、或6个或更多个)引导RNA。

[0501] 在一些情况下(例如,当与本公开的引导RNA和Cas12L多肽接触时),在测量步骤之前,样品接触2小时或更少时间(例如,1.5小时或更少、1小时或更少、40分钟或更少、30分钟或更少、20分钟或更少、10分钟或更少、或5分钟或更少、或1分钟或更少)。例如,在一些情况下,在测量步骤之前使样品接触40分钟或更短时间。在一些情况下,在测量步骤之前使样品接触20分钟或更短时间。在一些情况下,在测量步骤之前使样品接触10分钟或更短时间。在一些情况下,在测量步骤之前使样品接触5分钟或更短时间。在一些情况下,在测量步骤之前使样品接触1分钟或更短时间。在一些情况下,在测量步骤之前使样品接触50秒至60秒。在一些情况下,在测量步骤之前使样品接触40秒至50秒。在一些情况下,在测量步骤之前使样品接触30秒至40秒。在一些情况下,在测量步骤之前使样品接触20秒至30秒。在一些情况下,在测量步骤之前使样品接触10秒至20秒。

[0502] 本公开的用于检测样品中的靶DNA(单链或双链)的方法可以高灵敏度检测靶DNA。在一些情况下,可使用本公开的方法检测包含多个DNA(包括靶DNA和多个非靶DNA)的样品中存在的靶DNA,其中靶DNA以每10^7个非靶DNA一个或多个拷贝(例如,每10^6个非靶DNA一个或多个拷贝、每10^5个非靶DNA一个或多个拷贝、每10^4个非靶DNA一个或多个拷贝、每10^3个非靶DNA一个或多个拷贝、每10^2个非靶DNA一个或多个拷贝、每50个非靶DNA一个或多个拷贝、每20个非靶DNA一个或多个拷贝、每10个非靶DNA一个或多个拷贝、或每5个非靶DNA一个或多个拷贝)存在。在一些情况下,可使用本公开的方法检测包含多个DNA(包括靶DNA和多个非靶DNA)的样品中存在的靶DNA,其中靶DNA以每10^{18}个非靶DNA一个或多个拷贝(例如,每10^{15}个非靶DNA一个或多个拷贝、每10^{12}个非靶DNA一个或多个拷贝、每10^9个非靶DNA一个或多个拷贝、每10^6个非靶DNA一个或多个拷贝、每10^5个非靶DNA一个或多个拷贝、每10^4个非靶DNA一个或多个拷贝、每10^3个非靶DNA一个或多个拷贝、每10^2个非靶DNA一个或多个拷贝、每50个非靶DNA一个或多个拷贝、每20个非靶DNA一个或多个拷贝、每10个非靶DNA一个或多个拷贝、或每5个非靶DNA一个或多个拷贝)存在。

[0503] 在一些情况下,本公开的方法可检测样品中存在的靶DNA,其中靶DNA以每10^7个非靶DNA一个拷贝至每10个非靶DNA一个拷贝(例如,每10^7个非靶DNA 1个拷贝至每10^2个非靶DNA 1个拷贝、每10^7个非靶DNA 1个拷贝至每10^3个非靶DNA 1个拷贝、每10^7个非靶DNA 1个拷贝至每10^4个非靶DNA 1个拷贝、每10^7个非靶DNA 1个拷贝至每10^5个非靶DNA 1个拷贝、每10^7个非靶DNA1个拷贝至每10^6个非靶DNA 1个拷贝、每10^6个非靶DNA 1个拷贝至每10个非靶

DNA 1个拷贝、每10^6个非靶DNA 1个拷贝至每10^2个非靶DNA 1个拷贝、每10^6个非靶DNA 1个拷贝至每10^3个非靶DNA 1个拷贝、每10^6个非靶DNA 1个拷贝至每10^4个非靶DNA 1个拷贝、每10^6个非靶DNA 1个拷贝至每10^5个非靶DNA 1个拷贝、每10^5个非靶DNA 1个拷贝至每10个非靶DNA 1个拷贝、每10^5个非靶DNA 1个拷贝至每10^2个非靶DNA 1个拷贝、每10^5个非靶DNA1个拷贝至每10^3个非靶DNA 1个拷贝、或每10^5个非靶DNA 1个拷贝至每10^4个非靶DNA 1个拷贝)存在。

[0504] 在一些情况下,本公开的方法可检测样品中存在的靶DNA,其中靶DNA以每10^{18}个非靶DNA一个拷贝至每10个非靶DNA一个拷贝(例如,每10^{18}个非靶DNA 1个拷贝至每10^2个非靶DNA 1个拷贝、每10^{15}个非靶DNA 1个拷贝至每10^2个非靶DNA 1个拷贝、每10^{12}个非靶DNA 1个拷贝至每10^2个非靶DNA 1个拷贝、每10^9个非靶DNA 1个拷贝至每10^2个非靶DNA 1个拷贝、每10^7个非靶DNA1个拷贝至每10^2个非靶DNA 1个拷贝、每10^7个非靶DNA 1个拷贝至每10^3个非靶DNA 1个拷贝、每10^7个非靶DNA 1个拷贝至每10^4个非靶DNA 1个拷贝、每10^7个非靶DNA 1个拷贝至每10^5个非靶DNA 1个拷贝、每10^7个非靶DNA 1个拷贝至每10^6个非靶DNA 1个拷贝、每10^6个非靶DNA 1个拷贝至每10个非靶DNA 1个拷贝、每10^6个非靶DNA 1个拷贝至每10^2个非靶DNA 1个拷贝、每10^6个非靶DNA 1个拷贝至每10^3个非靶DNA 1个拷贝、每10^6个非靶DNA 1个拷贝至每10^4个非靶DNA 1个拷贝、每10^6个非靶DNA 1个拷贝至每10^5个非靶DNA 1个拷贝、每10^5个非靶DNA 1个拷贝至每10个非靶DNA 1个拷贝、每10^5个非靶DNA 1个拷贝至每10^2个非靶DNA 1个拷贝、每10^5个非靶DNA 1个拷贝至每10^3个非靶DNA 1个拷贝、或每10^5个非靶DNA 1个拷贝至每10^4个非靶DNA 1个拷贝)存在。

[0505] 在一些情况下,本公开的方法可检测样品中存在的靶DNA,其中靶DNA以每10^7个非靶DNA一个拷贝至每100个非靶DNA一个拷贝(例如,每10^7个非靶DNA 1个拷贝至每10^2个非靶DNA 1个拷贝、每10^7个非靶DNA 1个拷贝至每10^3个非靶DNA 1个拷贝、每10^7个非靶DNA 1个拷贝至每10^4个非靶DNA 1个拷贝、每10^7个非靶DNA 1个拷贝至每10^5个非靶DNA 1个拷贝、每10^7个非靶DNA1个拷贝至每10^6个非靶DNA 1个拷贝、每10^6个非靶DNA 1个拷贝至每100个非靶DNA 1个拷贝、每10^6个非靶DNA 1个拷贝至每10^2个非靶DNA 1个拷贝、每10^6个非靶DNA 1个拷贝至每10^3个非靶DNA 1个拷贝、每10^6个非靶DNA 1个拷贝至每10^4个非靶DNA 1个拷贝、每10^6个非靶DNA 1个拷贝至每10^5个非靶DNA 1个拷贝、每10^5个非靶DNA 1个拷贝至每100个非靶DNA 1个拷贝、每10^5个非靶DNA 1个拷贝至每10^2个非靶DNA 1个拷贝、每10^5个非靶DNA 1个拷贝至每10^3个非靶DNA 1个拷贝、或每10^5个非靶DNA 1个拷贝至每10^4个非靶DNA 1个拷贝)存在。

[0506] 在一些情况下,对于检测样品中的靶DNA的主题方法,检测阈值为10nM或更小。术语“检测阈值”在本文用于描述要发生检测样品中必须存在的最小靶DNA量。因此,作为说明性实例,当检测阈值为10nM时,则当样品中存在浓度为10nM或更高的靶DNA时,可以检测到信号。在一些情况下,本公开的方法的检测阈值为5nM或更小。在一些情况下,本公开的方法的检测阈值为1nM或更小。在一些情况下,本公开的方法的检测阈值为0.5nM或更小。在一些情况下,本公开的方法的检测阈值为0.1nM或更小。在一些情况下,本公开的方法的检测阈值为0.05nM或更小。在一些情况下,本公开的方法的检测阈值为0.01nM或更小。在一些情况下,本公开的方法的检测阈值为0.005nM或更小。在一些情况下,本公开的方法的检测阈值为0.001nM或更小。在一些情况下,本公开的方法的检测阈值为0.0005nM或更小。在一些情

况下,本公开的方法的检测阈值为0.0001nM或更小。在一些情况下,本公开的方法的检测阈值为0.00005nM或更小。在一些情况下,本公开的方法的检测阈值为0.00001nM或更小。在一些情况下,本公开的方法的检测阈值为10pM或更小。在一些情况下,本公开的方法的检测阈值为1pM或更小。在一些情况下,本公开的方法的检测阈值为500fM或更小。在一些情况下,本公开的方法的检测阈值为250fM或更小。在一些情况下,本公开的方法的检测阈值为100fM或更小。在一些情况下,本公开的方法的检测阈值为50fM或更小。在一些情况下,本公开的方法的检测阈值为500aM(渺摩尔)或更小。在一些情况下,本公开的方法的检测阈值为250aM或更小。在一些情况下,本公开的方法的检测阈值为100aM或更小。在一些情况下,本公开的方法的检测阈值为50aM或更小。在一些情况下,本公开的方法的检测阈值为10aM或更小。在一些情况下,本公开的方法的检测阈值为1aM或更小。

[0507] 在一些情况下,检测阈值(用于在主题方法中检测靶DNA)的范围为500fM到1nM(例如,500fM到500pM、500fM到200pM、500fM到100pM、500fM到10pM、500fM到1pM、800fM到1nM、800fM到500pM、800fM到200pM、800fM到100pM、800fM到10pM、800fM到1pM、1pM到1nM、1pM到500pM、1pM到200pM、1pM到100pM或1pM到10pM)(其中浓度是指可以检测到靶DNA的靶DNA阈值浓度)。在一些情况下,本公开的方法的检测阈值在800fM至100pM的范围内。在一些情况下,本公开的方法的检测阈值在1pM至10pM的范围内。在一些情况下,本公开的方法的检测阈值在10fM至500fM,例如10fM至50fM、50fM至100fM、100fM至250fM、或250fM至500fM的范围内。

[0508] 在一些情况下,可以在样品中检测到靶DNA的最小浓度的范围为500fM到1nM(例如,500fM至500pM、500fM至200pM、500fM至100pM、500fM至10pM、500fM至1pM、800fM至1nM、800fM至500pM、800fM至200pM、800fM至100pM、800fM至10pM、800fM至1pM、1pM至1nM、1pM至500pM、1pM至200pM、1pM至100pM或1pM至10pM)。在一些情况下,可在样品中检测到靶DNA的最小浓度在800fM至100pM的范围内。在一些情况下,可在样品中检测到靶DNA的最小浓度在1pM至10pM的范围内。

[0509] 在一些情况下,检测阈值(在主题方法中用于检测样品中的靶DNA)的范围为1aM至1nM(例如,1aM至500pM、1aM至200pM、1aM至100pM、1aM至10pM、1aM至1pM、100aM至1nM、100aM至500pM、100aM至200pM、100aM至100pM、100aM至10pM、100aM至1pM、250aM至1nM、250aM至500pM、250aM至200pM、250aM至100pM、250aM至10pM、250aM至1pM、500aM至1nM、500aM至500pM、500aM至200pM、500aM至100pM、500aM至10pM、500aM至1pM、750aM至1nM、750aM至500pM、750aM至200pM、750aM至100pM、750aM至10pM、750aM至1pM、1fM至1nM、1fM至500pM、1fM至200pM、1fM至100pM、1fM至10pM、1fM至1pM、500fM至500pM、500fM至200pM、500fM至100pM、500fM至10pM、500fM至1pM、800fM至1nM、800fM至500pM、800fM至200pM、800fM至100pM、800fM至10pM、800fM至1pM、1pM至1nM、1pM至500pM、1pM至200pM、1pM至100pM、或1pM至10pM)(其中浓度是指可以检测到靶DNA的靶DNA阈值浓度)。在一些情况下,本公开的方法的检测阈值在1aM至800aM的范围内。在一些情况下,本公开的方法的检测阈值在50aM至1pM的范围内。在一些情况下,本公开的方法的检测阈值在50aM至500fM的范围内。

[0510] 在一些情况下,可在样品中检测到靶DNA的最小浓度在1aM至1nM(例如,1aM至500pM、1aM至200pM、1aM至100pM、1aM至10pM、1aM至1pM、100aM至1nM、100aM至500pM、100aM至200pM、100aM至100pM、100aM至10pM、100aM至1pM、250aM至1nM、250aM至500pM、250aM至

200pM、250aM至100pM、250aM至10pM、250aM至1pM、500aM至1nM、500aM至500pM、500aM至200pM、500aM至100pM、500aM至10pM、500aM至1pM、750aM至1nM、750aM至500pM、750aM至200pM、750aM至100pM、750aM至10pM、750aM至1pM、1fM至1nM、1fM至500pM、1fM至200pM、1fM至100pM、1fM至10pM、1fM至1pM、500fM至500pM、500fM至200pM、500fM至100pM、500fM至10pM、500fM至1pM、800fM至1nM、800fM至500pM、800fM至200pM、800fM至100pM、800fM至10pM、800fM至1pM、1pM至1nM、1pM至500pM、1pM至200pM、1pM至100pM、或1pM至10pM)的范围内。在一些情况下,可在样品中检测到靶DNA的最小浓度在1aM至500pM的范围内。在一些情况下,可在样品中检测到靶DNA的最小浓度在100aM至500pM的范围内。

[0511] 在一些情况下,主题组合物或方法表现出渺摩尔级(aM)的检测灵敏度。在一些情况下,主题组合物或方法表现出飞摩尔级(fM)的检测灵敏度。在一些情况下,主题组合物或方法表现出皮摩尔级(pM)的检测灵敏度。在一些情况下,主题组合物或方法表现出纳米摩尔级(nM)的检测灵敏度。

[0512] 靶DNA

[0513] 靶DNA可以是单链的(ssDNA)或双链的(dsDNA)。当靶DNA是单链的时,对靶DNA中的PAM序列没有偏好或要求。但是,当靶DNA是dsDNA时,PAM通常邻近靶DNA的靶序列存在(例如,参见本文其他地方对PAM的论述)。靶DNA的来源可与样品的来源相同,例如,如下文所述。

[0514] 靶DNA的来源可以是任何来源。在一些情况下,靶DNA是病毒DNA(例如,DNA病毒的基因组DNA)。因而,主题方法可用于检测核酸群体中(例如,样品中)病毒DNA的存在。主题方法还可用于在靶DNA存在下切割非靶ssDNA。例如,如果方法发生在细胞中,则当细胞中存在特定靶DNA时,主题方法可用于混杂地切割细胞中的非靶ssDNA(不与引导RNA的引导序列杂交的ssDNA)(例如,当细胞被病毒感染并检测到病毒靶DNA时)。

[0515] 可能的靶DNA的实例包括但不限于病毒DNA,诸如:乳多空病毒(例如,人乳头瘤病毒(HPV)、多瘤病毒属);嗜肝DNA病毒(例如,乙型肝炎病毒(HBV));疱疹病毒(例如,单纯疱疹病毒(HSV)、水痘带状疱疹病毒(VZV)、爱泼斯坦-巴尔病毒(Epstein-barr virus,EBV)、巨细胞病毒(CMV)、疱疹淋巴病毒、玫瑰糠疹、卡波西氏肉瘤相关疱疹病毒);腺病毒(例如,鸟腺病毒、禽腺病毒、鱼腺病毒(ichtadenovirus)、美洲白鲟腺病毒(mastadenovirus)、唾液酸酶腺病毒);痘病毒(例如,天花、痘苗病毒、牛痘病毒、猴痘病毒、口疮病毒、假牛痘病毒、牛丘疹性口炎病毒;特纳河痘病毒、亚巴猴肿瘤病毒;传染性软疣病毒(MCV));细小病毒(例如,腺相关病毒(AAV)、细小病毒B19、人博卡病毒、bufavirus、人parv4 G1);双生病毒科;矮化病毒科;藻类DNA病毒科(Phycodnaviridae)等。在一些情况下,靶DNA是寄生虫DNA。在一些情况下,靶DNA是细菌DNA,例如病原性细菌的DNA。

[0516] 样品

[0517] 主题样品包括核酸(例如,多个核酸)。术语“多个”在本文中用于意指两个或更多个。因此,在一些情况下,样品包含两个或更多个(例如,3个或更多个、5个或更多个、10个或更多个、20个或更多个、50个或更多个、100个或更多个、500个或更多个、1,000个或更多个、或者5,000个或更多个)核酸(例如,DNA)。主题方法可用作检测样品中(例如,诸如DNA的核酸的复杂混合物中)存在的靶DNA的非常灵敏的方法。在一些情况下,样品包含序列彼此不同的5个或更多个DNA(例如,10个或更多个、20个或更多个、50个或更多个、100个或更多个、

500个或更多个、1,000个或更多个、或者5,000个或更多个DNA)。在一些情况下,样品包含10个或更多个、20个或更多个、50个或更多个、100个或更多个、500个或更多个、10^3个或更多个、$5x10^3$个或更多个、10^4个或更多个、$5x10^4$个或更多个、10^5个或更多个、$5x10^5$个或更多个、10^6个或更多个、$5x10^6$个或更多个、或者10^7个或更多个DNA。在一些情况下,样品包含10至20个、20至50个、50至100个、100至500个、500至10^3个、10^3至$5x10^3$个、$5x10^3$至10^4个、10^4至$5x10^4$个、$5x10^4$至10^5个、10^5至$5x10^5$个、$5x10^5$至10^6个、10^6至$5x10^6$个、或$5x10^6$至10^7个、或超过10^7个DNA。在一些情况下,样品包含5至10^7个DNA(例如,序列彼此不同)(例如,5至10^6个、5至10^5个、5至50,000个、5至30,000个、10至10^6个、10至10^5个、10至50,000个、10至30,000个、20至10^6个、20至10^5个、20至50,000个、或20至30,000个DNA)。在一些情况下,样品包含20个或更多个序列彼此不同的DNA。在一些情况下,样品包含来自细胞裂解液(例如,真核细胞裂解液、哺乳动物细胞裂解液、人细胞裂解液、原核细胞裂解液、植物细胞裂解液等)的DNA。例如,在一些情况下,样品包含来自细胞诸如真核细胞,例如哺乳动物细胞诸如人细胞的DNA。

[0518] 术语“样品”在本文中用于意指包含DNA的任何样品(例如,以便确定在DNA群体中是否存在靶DNA)。样品可衍生自任何来源,例如,样品可以是纯化DNA的合成组合;样品可以是细胞裂解液、富含DNA的细胞裂解液,或从细胞裂解液中分离和/或纯化的DNA。样品可来自患者(例如,出于诊断目的)。样品可来自透化细胞。样品可来自交联细胞。样品可在组织切片中。样品可来自通过交联,之后进行脱脂和调整以形成均匀折射率而制备的组织。通过交联,之后进行脱脂和调整以形成均匀折射率的组织制备的实例描述于例如Shah等人,Development(2016)143,2862-2867doi:10.1242/dev.138560中。

[0519] “样品”可包含靶DNA和多个非靶DNA。在一些情况下,靶DNA在样品中以每10个非靶DNA一个拷贝、每20个非靶DNA一个拷贝、每25个非靶DNA一个拷贝、每50个非靶DNA一个拷贝、每100个非靶DNA一个拷贝、每500个非靶DNA一个拷贝、每10^3个非靶DNA一个拷贝、每$5x10^3$个非靶DNA一个拷贝、每10^4个非靶DNA一个拷贝、每$5x10^4$个非靶DNA一个拷贝、每10^5个非靶DNA一个拷贝、每$5x10^5$个非靶DNA一个拷贝、每10^6个非靶DNA一个拷贝、或小于每10^6个非靶DNA一个拷贝存在。在一些情况下,靶DNA在样品中以每10个非靶DNA一个拷贝至每20个非靶DNA 1个拷贝、每20个非靶DNA 1个拷贝至每50个非靶DNA 1个拷贝、每50个非靶DNA 1个拷贝至每100个非靶DNA 1个拷贝、每100个非靶DNA 1个拷贝至每500个非靶DNA 1个拷贝、每500个非靶DNA 1个拷贝至每10^3个非靶DNA 1个拷贝、每10^3个非靶DNA 1个拷贝至每$5x10^3$个非靶DNA 1个拷贝、每$5x10^3$个非靶DNA 1个拷贝至每10^4个非靶DNA 1个拷贝、每10^4个非靶DNA 1个拷贝至每10^5个非靶DNA 1个拷贝、每10^5个非靶DNA 1个拷贝至每10^6个非靶DNA 1个拷贝、或每10^6个非靶DNA 1个拷贝至每10^7个非靶DNA 1个拷贝存在。

[0520] 合适的样品包括但不限于唾液、血液、血清、血浆、尿液、抽吸物和活检样品。因此,关于患者的术语“样品”涵盖生物来源的血液和其他液体样品、实体组织样品诸如活检样本或组织培养物或来源于其的细胞及其后代。该定义还包括在获得后采用以下任何方式操作过的样品:诸如用试剂处理;洗涤;或针对某些细胞群体诸如癌细胞进行富集。该定义还包括已经富集了特定类型的分子(例如,DNA)的样品。术语“样品”涵盖生物样品,诸如临床样品,诸如血液、血浆、血清、抽吸物、脑脊髓液(CSF),并且还包括通过手术切除获得的组织、通过活检获得的组织、培养物中的细胞、细胞上清液、细胞裂解液、组织样品、器官、骨髓等。“生物样品”包括由其来源的生物流体(例如癌细胞、感染细胞等),例如从此类细胞获得的

包含DNA的样品(例如包含DNA的细胞裂解物或其他细胞提取物)。

[0521] 样品可包含或可从多种细胞、组织、器官或无细胞流体中的任一种获得。合适的样品来源包括真核细胞、细菌细胞和古细菌细胞。合适的样品来源包括单细胞生物体和多细胞生物体。合适的样品来源包括单细胞真核生物体;植物或植物细胞;藻类细胞,例如布朗葡萄藻、莱茵衣藻、海洋富油微拟球藻、蛋白核小球藻、展枝马尾藻、羽藻等;真菌细胞(例如,酵母细胞);动物细胞、组织或器官;来自无脊椎动物(例如,果蝇、刺胞动物、棘皮动物、线虫、昆虫、蛛形纲动物等)的细胞、组织或器官;来自脊椎动物(例如,鱼类、两栖动物、爬行动物、鸟类、哺乳动物)的细胞、组织、体液或器官;来自哺乳动物(例如,人;非人灵长类动物;有蹄类动物;猫科动物;牛;绵羊;山羊等)的细胞、组织、体液或器官。合适的样品来源包括线虫、原生动物等。合适的样品来源包括寄生虫,诸如蠕虫、疟疾寄生虫等。

[0522] 合适的样品来源包括例如以下六个界中任何一个界的细胞、组织或生物体:细菌界(例如,真细菌界);古细菌界;原生生物界;真菌界;植物界;和动物界。合适的样品来源包括原生生物界的植物样成员,包括但不限于藻类(例如,绿藻、红藻、灰胞藻、蓝细菌);原生生物界的真菌样成员,例如粘液菌、水霉菌等;原生生物界的动物样成员,例如鞭毛虫类(例如,眼虫藻)、变形虫类(amoeboids)(例如,变形虫)、孢子虫类(例如,顶复门(Apicomplexa)、粘体动物门(Myxozoa)、微孢子虫纲(Microsporidia))和纤毛虫类(例如,草履虫)。合适的样品来源包括包括真菌界的成员,包括但不限于以下门中的任何门的成员:担子菌门(担子菌类;例如,伞菌属、鹅膏菌属、牛肝菌属、鸡油菌属等成员);子囊菌门(子囊菌类,包括例如酵母菌);菌藻门(地衣);接合菌门(接合真菌);以及不完全菌门。合适的样品来源包括包括植物界的成员,包括但不限于以下分类中的任何分类的成员:苔藓植物门(例如,藓类)、角苔植物门(例如,角苔类)、苔类植物门(Hepaticophyta)(例如,苔类)、石松植物门(例如,石松类)、楔叶植物门(例如,木贼类)、裸蕨植物门(例如,松叶蕨类)、瓶尔小草门、蕨门(例如,蕨类)、苏铁门、银杏门、松柏门、买麻藤门和木兰门(例如,开花植物)。合适的样品来源包括包括动物界的成员,包括但不限于以下门中的任何门的成员:多孔动物门(海绵动物);扁盘动物门;直泳虫门(海洋无脊椎动物的寄生物);菱形虫门;刺胞动物门(珊瑚、海葵、海蜇、海笔、海肾、立方水母);栉水母门(栉水母类);扁虫动物门(扁虫类);纽形动物门(纽虫类);颚胃动物门(Ngathostomulida)(有颚蠕虫)p腹毛动物门;轮虫动物门;曳鳃动物门;动吻动物门;铠甲动物门;棘头动物门;内肛动物门;线虫动物门;线形动物门;环口动物门;软体动物门(软体动物);星虫动物门(方格星虫(peanut worms));环节动物门(环节蠕虫);缓步动物门(缓步动物);有爪动物门(栉蚕);节肢动物门(包括以下亚门:有螯肢亚门、多足亚门、六足亚门和甲壳亚门,其中有螯肢亚门包括例如蛛形纲、肢口纲和海蜘蛛纲,其中多足亚门包括例如唇足纲(唇足类)、倍足纲(多足类)、少足纲(Paropoda)和综合纲,其中六足亚门包括昆虫纲,并且其中甲壳亚门包括虾、磷虾、藤壶等;帚虫动物门;外肛动物门(苔藓动物);腕足动物门;棘皮动物门(例如,海星、海雏菊、毛头星、海胆、海参、海蛇尾、脆篮(brittle baskets)等);毛颚动物门(箭虫);半索动物门(玉钩虫);和脊索动物门。合适的脊索动物门成员包括以下亚门的任何成员:尾索动物亚门(海鞘纲;包括海鞘目、樽海鞘目和幼形目);头索亚门(文昌鱼);盲鳗纲(盲鳗);和脊椎动物亚门,其中脊椎动物亚门成员包括例如以下纲的成员:鳃鳗纲(七鳃类)、软骨鱼纲(软骨鱼)、辐鳍鱼纲(辐鳍鱼)、腔棘焦纲(腔棘鱼)、肺鱼纲(肺鱼)、爬行纲(爬行动物,例如蛇、短吻鳄、鳄

鱼、蜥蜴等)、鸟纲(鸟类);和哺乳纲(哺乳动物)。合适的植物包括任何单子叶植物和任何双子叶植物。

[0523] 合适的样品来源包括取自生物体;从生物体分离的特定细胞或细胞群的细胞、流体、组织或器官等。例如,在生物体是植物的情况下,合适的来源包括木质部、韧皮部、形成层、叶、根等。在生物体是动物的情况下,合适的来源包括特定组织(例如,肺、肝、心脏、肾、脑、脾、皮肤、胎儿组织等)、或特定细胞类型(例如,神经元细胞、上皮细胞、内皮细胞、星形胶质细胞、巨噬细胞、神经胶质细胞、胰岛细胞、T淋巴细胞、B淋巴细胞等)。

[0524] 在一些情况下,样品来源是病变的(或疑似病变的细胞、流体、组织或器官。在一些情况下,样品来源是正常(非病变的)细胞、流体、组织或器官。在一些情况下,样品来源是(或疑似是病原体感染的细胞、组织或器官。例如,样品来源可以是可能被感染或可能未被感染的个体-并且样品可以是从个体收集的任何生物样品(例如血液、唾液、活检物、血浆、血清、支气管肺泡灌洗液、痰液、粪便样品、脑脊液、细针抽吸物、拭子样品(例如,颊拭子、宫颈拭子、鼻拭子)、间质液、滑液、鼻分泌物、眼泪、血沉棕黄层、粘膜样品、上皮细胞样品(例如,上皮细胞刮擦物)等)。在一些情况下,样品是无细胞液体样品。在一些情况下,样品是可包含细胞的液体样品。病原体包括病毒、真菌、蠕虫、原生动物、疟疾寄生虫、疟原虫寄生虫、弓形虫寄生虫、血吸虫寄生虫等。“蠕虫”包括蛔虫、犬恶丝虫和植食性线虫(线虫纲)、吸虫(吸虫纲)、棘头虫纲和绦虫(绦虫纲)。原生动物感染包括来自贾第虫属种、毛滴虫属种、非洲锥虫病、阿米巴痢疾、巴贝虫病、小袋虫性痢疾、查加斯病(Chaga's disease)、球虫病、疟疾和弓形体病的感染。病原体(诸如寄生/原生动物病原体)的实例包括但不限于:恶性疟原虫(Plasmodium falciparum)、间日疟原虫(Plasmodium vivax)、克氏锥虫(Trypanosoma cruzi)和刚地弓形虫(Toxoplasma gondii)。真菌病原体包括但不限于:新型隐球菌(Cryptococcus neoformans)、荚膜组织胞浆菌(Histoplasma capsulatum)、粗球孢菌(Coccidioides immitis)、皮炎芽生菌(Blastomyces dermatitidis)、沙眼衣原体(Chlamydia trachomatis)和白色念珠菌(Candida albicans)。病原性病毒包括例如免疫缺陷病毒(例如,HIV);流感病毒;登革热病毒;西尼罗河病毒;疱疹病毒;黄热病毒;丙型肝炎病毒;甲型肝炎病毒;乙型肝炎病毒;乳头瘤病毒等。病原性病毒可包括DNA病毒,诸如:乳多空病毒(例如,人乳头瘤病毒(HPV)、多瘤病毒属);嗜肝DNA病毒(例如,乙型肝炎病毒(HBV));疱疹病毒(例如,单纯疱疹病毒(HSV)、水痘带状疱疹病毒(VZV)、爱泼斯坦-巴尔病毒(EBV)、巨细胞病毒(CMV)、疱疹淋巴病毒、玫瑰糠疹、卡波西氏肉瘤相关疱疹病毒);腺病毒(例如,鸟腺病毒、禽腺病毒、鱼腺病毒、美洲白鲟腺病毒、唾液酸酶腺病毒);痘病毒(例如,天花、痘苗病毒、牛痘病毒、猴痘病毒、口疮病毒、假牛痘病毒、牛丘疹性口炎病毒;特纳河痘病毒、亚巴猴肿瘤病毒;传染性软疣病毒(MCV));细小病毒(例如,腺相关病毒(AAV)、细小病毒B19、人博卡病毒、bufavirus、人parv4 G1);双生病毒科;矮化病毒科;藻科等。病原体可包括,例如DNA病毒[例如:乳多空病毒(例如,人乳头瘤病毒(HPV)、多瘤病毒属);嗜肝DNA病毒(例如,乙型肝炎病毒(HBV));疱疹病毒(例如,单纯疱疹病毒(HSV)、水痘带状疱疹病毒(VZV)、爱泼斯坦-巴尔病毒(EBV)、巨细胞病毒(CMV)、疱疹淋巴病毒、玫瑰糠疹、卡波西氏肉瘤相关疱疹病毒);腺病毒(例如,鸟腺病毒、禽腺病毒、鱼腺病毒、美洲白鲟腺病毒、唾液酸酶腺病毒);痘病毒(例如,天花、痘苗病毒、牛痘病毒、猴痘病毒、口疮病毒、假牛痘病毒、牛丘疹性口炎病毒;特纳河痘病毒、亚巴猴肿瘤病毒;传染性软疣病毒(MCV));细小病毒

(例如,腺相关病毒(AAV)、细小病毒B19、人博卡病毒、bufavirus、人parv4 G1);双生病毒科;矮化病毒科;藻科等]、结核分枝杆菌(Mycobacterium tuberculosis)、无乳链球菌(Streptococcus agalactiae)、耐甲氧西林金黄色葡萄球菌(methicillin-resistant Staphylococcus aureus)、嗜肺军团菌(Legionella pneumophila)、酿脓链球菌(Streptococcus pyogenes)、大肠杆菌(Escherichia coli)、淋病奈瑟氏菌(Neisseria gonorrhoeae)、脑膜炎奈瑟氏菌(Neisseria meningitidis)、肺炎球菌(Pneumococcus)、新型隐球菌(Cryptococcus neoformans)、荚膜组织胞浆菌(Histoplasma capsulatum)、流感嗜血杆菌B(Hemophilus influenzae B)、梅毒密螺旋体(Treponema pallidum)、莱姆病螺旋体(Lyme disease spirochetes)、铜绿假单胞菌(Pseudomonas aeruginosa)、麻风分枝杆菌(Mycobacterium leprae)、流产布鲁氏菌(Brucella abortus)、狂犬病病毒、流感病毒、巨细胞病毒、单纯疱疹病毒I、单纯疱疹病毒II、人血清细小样病毒(human serum parvo-like virus)、呼吸道合胞体病毒、水痘-带状疱疹病毒、乙型肝炎病毒、丙型肝炎病毒、麻疹病毒、腺病毒、人T细胞白血病病毒、爱泼斯坦-巴尔病毒、鼠白血病病毒、腮腺炎病毒、水疱性口炎病毒、辛德毕斯病毒(Sindbis virus)、淋巴细胞性脉络丛脑膜炎病毒、疣病毒、蓝舌病毒、仙台病毒(Sendai virus)、猫白血病病毒、呼肠孤病毒、脊髓灰质炎病毒、猿猴病毒40、小鼠乳腺肿瘤病毒、登革热病毒、风疹病毒、西尼罗河病毒、恶性疟原虫、间日疟原虫、刚地弓形虫、蓝氏锥虫(Trypanosoma rangeli)、克氏锥虫、罗氏锥虫(Trypanosoma rhodesiense)、布氏锥虫(Trypanosoma brucei)、曼氏血吸虫(Schistosoma mansoni)、日本血吸虫(Schistosoma japonicum)、巴贝斯虫(Babesia bovis)、柔嫩艾美球虫(Eimeria tenella)、盘尾丝虫(Onchocerca volvulus)、利什曼原虫(Leishmania tropica)、结核分枝杆菌、旋毛虫(Trichinella spiralis)、泰勒原虫(Theileria parva)、胞状绦虫(Taenia hydatigena)、羊绦虫(Taenia ovis)、牛肉绦虫(Taenia saginata)、细粒棘球绦虫(Echinococcus granulosus)、柯氏中殖孔绦虫(Mesocestoides corti)、关节炎支原体(Mycoplasma arthritidis)、猪鼻支原体(M.hyorhinis)、口腔支原体(M.orale)、精氨酸支原体(M.arginini)、莱氏无胆甾原体(Acholeplasma laidlawii)、唾窦支原体(M.salivarium)和肺炎支原体(M.pneumoniae)。

[0525] 测量可检测信号

[0526] 在一些情况下,主题方法包括测量步骤(例如,测量由Cas12L介导的ssDNA切割产生的可检测信号)。因为本公开的Cas12L多肽一旦被激活就切割非靶向ssDNA,这发生于当引导RNA与靶DNA在Cas12L效应蛋白存在下杂交时,可检测的信号可以是切割ssDNA时产生的任何信号。例如,在一些情况下,测量步骤可包括以下一项或多项:基于金纳米颗粒的检测(例如,参见Xu等人,Angew Chem Int Ed Engl.2007;46(19):3468-70;和Xia等人,Proc Natl Acad Sci U S A.2010年6月15日;107(24):10837-41)、荧光偏振、胶体相变/分散(例如,Baksh等人,Nature.2004年1月8日;427(6970):139-41)、电化学检测、基于半导体的感测(例如,Rothberg等人,Nature.2011年7月20日;475(7356):348-52;例如,可以在ssRNA切割反应后使用磷酸酶以通过打开2'-3'环状磷酸酯并通过将无机磷酸盐释放到溶液中而产生pH变化),以及检测经标记的检测剂ssRNA(更多细节参见本文其他地方)。此类检测方法的读出可以是任何方便的读出。可能的读出的实例包括但不限于:所测量的可检测荧光信号的量;对凝胶上条带(例如,表示切割产物对比未切割底物的条带)的视觉分析、基于视觉

或传感器的对颜色存在或不存在的检测(即,颜色检测方法),以及电信号的存在或不存在(或电信号的特定量)。

[0527] 在一些情况下,例如在检测到的信号量可用于确定样品中存在的靶DNA的量的意义上,测量可以是定量的。在一些情况下,例如在可检测信号的存在或不存在可以指示靶DNA(例如,病毒、SNP等)的存在或不存在的意义上,测量可以是定性的。在一些情况下,除非靶DNA(例如,病毒、SNP等)以高于特定阈值浓度存在,否则可检测信号将不存在(例如,高于给定阈值水平)。在一些情况下,可通过修改Cas12L效应子、引导RNA、样品体积和/或检测剂ssDNA(如果使用的话)的量来滴定检测阈值。因此,例如,如本领域的普通技术人员应当理解的,如果需要的话,可使用许多对照以建立一个或多个反应,每个反应设置用于检测靶DNA的不同阈值水平,并且因此可使用这样的一系列反应来确定样品中存在的靶DNA的量(例如,可使用这样的一系列反应来确定靶DNA '以至少X的浓度' 存在于样品中)。

[0528] 本公开的检测方法的使用实例包括例如单核苷酸多态性(SNP)检测、癌症筛查、细菌感染检测、抗生素抗性检测、病毒感染检测等。本公开的组合物和方法可用于检测任何DNA靶标。例如,可检测将核酸物质整合到基因组中的任何病毒,因为受试样品可包含细胞基因组DNA-并且引导RNA可被设计成检测整合的核苷酸序列。

[0529] 在一些情况下,可使用本公开方法确定样品(例如,包含靶DNA和多个非靶DNA的样品)中靶DNA的量。确定样品中靶DNA的量可包括将由测试样品产生的可检测信号的量与由参考样品产生的可检测信号的量进行比较。确定样品中靶DNA的量可包括:测量可检测信号以生成测试测量值;测量由参考样品产生的可检测信号以生成参考测量值;以及将测试测量值与参考测量值进行比较,以确定样品中存在的靶DNA的量。

[0530] 例如,在一些情况下,本公开的用于确定样品中靶DNA量的方法包括:a)使样品(例如,包含靶DNA和多个非靶DNA的样品)与以下物质接触:(i)与靶DNA杂交的引导RNA;(ii)切割样品中存在的RNA的本公开的Cas12L多肽;和(iii)检测剂ssDNA;b)测量由Cas12L介导的ssDNA切割(例如,检测剂ssDNA的切割)产生的可检测信号,生成测试测量值;c)测量由参考样品产生的可检测信号以生成参考测量值;以及d)将测试测量值与参考测量值进行比较,以确定样品中存在的靶DNA的量。

[0531] 作为另一个实例,在一些情况下,本公开的用于测定样品中靶DNA的量的方法包括:a)将样品(例如,包含靶DNA和多个非靶DNA的样品)与以下项接触:i)包含两个或更多个引导RNA的前体引导RNA阵列,每个引导RNA具有不同的引导序列;(ii)本公开的Cas12L多肽,其将前体引导RNA阵列切割成单个引导RNA,并且还切割样品的RNA;和(iii)检测剂ssDNA;b)测量由Cas12L介导的ssDNA切割(例如,检测剂ssDNA的切割)产生的可检测信号,产生测试测量值;c)测量由两个或更多个参考样品中的每一个产生的可检测信号,以产生两个或更多个参考测量值;以及d)将测试测量值与参考测量值进行比较,以确定样品中存在的靶DNA的量。

[0532] 样品中核酸的扩增

[0533] 在一些实施方案中,可通过将检测与核酸扩增结合来提高主题组合物和/或方法(例如,用于检测细胞基因组DNA中靶DNA(诸如病毒DNA或SNP)的存在)的灵敏度。在一些情况下,样品中的核酸在与本公开的切割ssDNA的Cas12L多肽接触之前被扩增(例如,样品中核酸的扩增可以在与本公开的Cas12L多肽接触之前开始)。在一些情况下,样品中的核酸在

与本公开的Cas12L多肽接触的同时被扩增。例如,在一些情况下,主题方法包括扩增样品的核酸(例如,通过使样品与扩增组分接触),之后使扩增的样品与本公开的Cas12L多肽接触。在一些情况下,主题方法包括在使样品与本公开的Cas12L多肽接触的同一时间(同时)使样品与扩增组分接触。如果同时添加所有组分(扩增组分和检测组分,诸如本公开的Cas12L多肽、引导RNA和检测剂DNA),则Cas12L的反式切割活性可能会在样品的核酸经历扩增的同时开始降解核酸。但是,即使是这种情况,与不进行扩增的方法相比,同时进行扩增和检测仍然可提高灵敏度。

[0534] 在一些情况下,例如使用引物从样品中扩增特定序列(例如,病毒的序列,包括目标SNP的序列)。因此,可扩增将与指导RNA杂交的序列以提高主题检测方法的灵敏度-这可实现所需序列的有偏扩增,从而增加样品中存在的目标序列相对于样品中存在的其他序列的拷贝数。作为一个说明性实例,如果使用主题方法来确定给定样品是否包含特定病毒(或特定SNP),则可扩增病毒序列(或非病毒基因组序列)的所需区域,并且如果实际上样品中存在病毒序列(或SNP),则扩增的区域将包括与引导RNA杂交的序列。

[0535] 如上所述,在一些情况下,将核酸扩增(例如,通过与扩增组分接触),之后将扩增的核酸与本公开的Cas12L多肽接触。在一些情况下,在与本公开的Cas12L多肽接触之前,扩增持续10秒或更长时间(例如,30秒或更长时间、45秒或更长时间、1分钟或更长时间、2分钟或更长时间、3分钟或更长时间、4分钟或更长时间、5分钟或更长时间、7.5分钟或更长时间、10分钟或更长时间等)。在一些情况下,在与本公开的Cas12L多肽接触之前,扩增持续2分钟或更长时间(例如,3分钟或更长时间、4分钟或更长时间、5分钟或更长时间、7.5分钟或更长时间、10分钟或更长时间等)。在一些情况下,扩增持续时间段范围为10秒至60分钟(例如,10秒至40分钟、10秒至30分钟、10秒至20分钟、10秒至15分钟、10秒至10分钟、10秒至5分钟、30秒至40分钟、30秒至30分钟、30秒至20分钟、30秒至15分钟、30秒至10分钟、30秒至5分钟、1分钟至40分钟、1分钟至30分钟、1分钟至20分钟、1分钟至15分钟、1分钟至10分钟、1分钟至5分钟、2分钟至40分钟、2分钟至30分钟、2分钟至20分钟、2分钟至15分钟、2分钟至10分钟、2分钟至5分钟、5分钟至40分钟、5分钟至30分钟、5分钟至20分钟、5分钟至15分钟、或5分钟至10分钟)。在一些情况下,扩增持续时间段范围为5分钟至15分钟。在一些情况下,扩增持续时间段范围为7分钟至12分钟。

[0536] 在一些情况下,样品在与本公开的Cas12L多肽接触的同时与扩增组分接触。在一些此类情况下,Cas12L蛋白在接触时是无活性的,并且一旦样品中的核酸被扩增就被激活。

[0537] 各种扩增方法和组分将是本领域的普通技术人员已知的,并且可使用任何方便的方法(参见例如Zanoli和Spoto,Biosensors(Basel).2013年3月;3(1):18-43;Gill和Ghaemi,Nucleosides,Nucleotides,and Nucleic Acids,2008,27:224-243;Craw和Balachandrana,Lab Chip,2012,12,2469-2486;所述文献以引用方式整体并入本文)。核酸扩增可包括聚合酶链式反应(PCR)、逆转录PCR(RT-PCR)、定量PCR(qPCR)、逆转录qPCR(RT-qPCR)、巢式PCR、多重PCR、不对称PCR、降落式PCR、随机引物PCR、半巢式PCR、聚合酶循环组装(PCA)、菌落PCR、连接酶链式反应(LCR)、数字PCR、甲基化特异性PCR(MSP)、较低变性温度下的共扩增-PCR(COLD-PCR)、等位基因特异性PCR、序列间特异性PCR(ISS-PCR)、全基因组扩增(WGA)、反向PCR和热不对称交错PCR(TAIL-PCR)。

[0538] 在一些情况下,扩增是等温扩增。术语“等温扩增”指示核酸(例如,DNA)扩增的一

种方法(例如,使用酶链式反应),该方法可使用单一温度孵育,由此不需要热循环仪。等温扩增是核酸扩增的一种形式,其在扩增反应期间不依赖于靶核酸的热变性,因此可能不需要温度的多次快速变化。因此,等温核酸扩增方法可在实验室环境内部或外部进行。通过与逆转录步骤结合,这些扩增方法可用于等温扩增RNA。

[0539] 等温扩增方法的实例包括但不限于:环介导等温扩增(LAMP)、解旋酶依赖性扩增(HDA)、重组酶聚合酶扩增(RPA)、链置换扩增(SDA)、基于核酸序列的扩增(NASBA)、转录介导扩增(TMA)、切口酶扩增反应(NEAR)、滚环扩增(RCA)、多置换扩增(MDA)、分枝(RAM)、环状解旋酶依赖性扩增(cHDA)、单引物等温扩增(SPIA)、信号介导RNA扩增技术(SMART)、自我持续序列复制(3SR)、基因组指数扩增反应(GEAR)和等温多置换扩增(IMDA)。

[0540] 在一些情况下,扩增是重组酶聚合酶扩增(RPA)(参见例如美国专利号8,030,000;8,426,134;8,945,845;9,309,502;和9,663,820,所述专利特此以引用方式整体并入)。重组酶聚合酶扩增(RPA)使用两个相对的引物(非常类似于PCR),并且采用三种酶-重组酶、单链DNA结合蛋白(SSB)和链置换聚合酶。重组酶将双链体DNA中具有同源序列的寡核苷酸引物配对,SSB结合DNA的置换链以防止引物被置换,并且链置换聚合酶开始DNA合成,其中引物已与靶DNA结合。在RPA反应中添加逆转录酶可促进RNA以及DNA的检测,而无需单独的步骤来生产cDNA。RPA反应的组分的一个实例如下(参见例如美国专利号8,030,000;8,426,134;8,945,845;9,309,502;9,663,820):50mM Tris pH 8.4、80mM乙酸钾、10mM乙酸镁、2mM DTT、5%PEG化合物(Carbowax-20M)、3mM ATP、30mM磷酸肌酸、100ng/μl肌酸激酶、420ng/μl gp32、140ng/μl UvsX、35ng/μl UvsY、2000M dNTP、300nM各寡核苷酸、35ng/μl Bsu聚合酶和含核酸样品)。

[0541] 在转录介导扩增(TMA)中,使用RNA聚合酶从引物区中工程化的启动子制备RNA,然后逆转录酶从引物合成cDNA。然后可使用第三种酶,例如RNA酶H,从cDNA降解RNA靶标,而无需进行热变性步骤。这种扩增技术类似于自我持续序列复制(3SR)和基于核酸序列的扩增(NASBA),但所用的酶有所不同。再例如,解旋酶依赖性扩增(HDA)利用热稳定解旋酶(Tte-UvrD)而非热量来解链dsDNA而产生单链,然后将单链用于通过聚合酶进行杂交和引物延伸。又例如,环介导扩增(LAMP)采用具有链置换能力的热稳定聚合酶和一组四个或更多个特异性设计的引物。每个引物均被设计成具有发夹末端,一旦移位,这些发夹末端便会卡入发夹中,以促进自动引发和进一步的聚合酶延伸。在LAMP反应中,尽管反应在等温条件下进行,但是对于双链靶标来说,需要初始的热变性步骤。另外,扩增产生各种长度产物的梯形图。又例如,链置换扩增(SDA)结合了限制性内切核酸酶对其靶DNA的未修饰链进行切口的能力和外切核酸酶缺陷型DNA聚合酶延伸切口处的3'末端并置换下游DNA链的能力。

[0542] 检测剂DNA

[0543] 在一些情况下,主题方法包括使样品(例如,包含靶DNA和多个非靶ssDNA的样品)与以下项接触:i)本公开的Cas12L多肽;ii)引导RNA(或前体引导RNA阵列);和iii)检测剂DNA,其是单链的并且不与引导RNA的引导序列杂交。例如,在一些情况下,主题方法包括使样品与标记的单链检测剂DNA(检测剂ssDNA)接触,所述检测剂DNA包含荧光发射染料对;Cas12L多肽在被激活后切割标记的检测剂ssDNA(在引导RNA与靶DNA杂交的情况下,通过与引导RNA结合);并且被测量的可检测信号由荧光发射染料对产生。例如,在一些情况下,主题方法包括使样品与经标记的检测剂ssDNA接触,所述经标记的检测剂ssDNA包含荧光共振

能量转移(FRET)对或猝灭剂/荧光剂对,或两者。在一些情况下,主题方法包括使样品与包含FRET对的经标记的检测剂ssDNA接触。在一些情况下,主题方法包括使样品与包含荧光剂/猝灭剂对的经标记的检测剂ssDNA接触。

[0544] 荧光发射染料对包括FRET对或猝灭剂/荧光剂对。在FRET对和猝灭剂/荧光对的两种情况下,染料之一的发射光谱与所述对中另一种染料的吸收光谱区域重叠。如本文所用,术语“荧光发射染料对”是用于包括“荧光共振能量转移(FRET)对”和“猝灭剂/荧光剂对”的通用术语,这两个术语将在下文中更详细地讨论。术语“荧光发射染料对”可与短语“FRET对和/或猝灭剂/荧光对”互换使用

[0545] 在一些情况下(例如,当检测剂ssDNA包括FRET对时),经标记的检测剂ssDNA在被切割之前产生一定量的可检测信号,并且当经标记的检测剂ssDNA被切割时所测量到的可检测信号的量减少。在一些情况下,经标记的检测剂ssDNA在被切割之前产生第一可检测信号(例如,来自FRET对),当经标记的检测剂ssDNA被切割时产生第二可检测信号(例如,来自猝灭剂/荧光剂对)。因此,在一些情况下,经标记的检测剂ssDNA包括FRET对和猝灭剂/荧光对。

[0546] 在一些情况下,经标记的检测剂ssDNA包含FRET对。FRET是这样的过程,通过所述过程,能量的无辐射转移发生为从激发态荧光团到紧邻的第二发色团。能量转移可以发生的范围限制在大约10纳米(100埃),并且转移效率对荧光团之间的分开距离非常敏感。因此,如本文所使用,术语“FRET”(“荧光共振能量转移”;也称为“萤光共振能量转移(Förster resonance energy transfer)”)是指这样的物理现象,所述物理现象涉及供体荧光团和匹配的受体荧光团被选择为使得供体的发射光谱与受体的激发光谱重叠,并且被进一步选择为使得当供体和受体彼此非常接近(通常为10nm或更短距离)时,供体的激发将引起来自受体的激发和发射,这是因为一些能量经由量子耦合效应从供体传递到受体。因此,FRET信号用作供体和受体的接近度尺度;只有当它们彼此非常接近时才会产生信号。FRET供体部分(例如,供体荧光团)和FRET受体部分(例如,受体荧光团)在本文中统称为“FRET对”。

[0547] 供体-受体对(FRET供体部分和FRET受体部分)在本文中被称为“FRET对”或“信号FRET对”。因此,在一些情况下,主题经标记的检测剂ssDNA包含两个信号配偶体(信号对),当一个信号配偶体是FRET供体部分时,另一个信号配偶体是FRET受体部分。因此,当信号配偶体非常接近时(例如,在相同的RNA分子上时),包含这种FRET对(FRET供体部分和FRET受体部分)的主题经标记的检测剂ssDNA将显示可检测信号(FRET信号),但是当所述配偶体分开时(例如,在本公开的Cas12L多肽切割RNA分子后),信号将减少(或不存在)。

[0548] FRET供体和受体部分(FRET对)将是本领域的普通技术人员已知的,并且可以使用任何方便的FRET对(例如,任何方便的供体和受体部分对)。合适的FRET对的实例包括但不限于表1中所示的那些。还参见:Bajar等人Sensors(Basel).2016年9月14日;16(9);和Abraham等人PLoS One.2015年8月3日;10(8):e0134436。

[0549] 表1.FRET对的实例(供体和受体FRET部分)

[0550]

供体	受体
色氨酸	丹酰
IAEDANS (1)	DDPM (2)

[0551]

供体	受体
BFP	DsRFP
丹酰	异硫氰酸荧光素(FITC)
丹酰	十八烷基罗丹明
青色荧光蛋白(CFP)	绿色荧光蛋白(GFP)
CF (3)	德克萨斯红
荧光素	四甲基罗丹明
Cy3	Cy5
GFP	黄色荧光蛋白(YFP)
BODIPY FL (4)	BODIPY FL (4)
罗丹明 110	Cy3
罗丹明 6G	孔雀石绿
FITC	曙红氨基硫脲
B-藻红蛋白	Cy5
Cy5	Cy5.5

[0552] (1)5-(2-碘乙酰基氨基乙基)氨基萘-1-磺酸

[0553] (2)N-(4-二甲基氨基-3,5-二硝基苯基)马来酰亚胺

[0554] (3)羧基荧光素琥珀酰亚胺酯

[0555] (4)4,4-二氟-4-硼杂-3a,4a-二氮杂-s-吲哒省

[0556] 在一些情况下,当经标记的检测剂ssDNA被切割时产生可检测信号(例如,在一些情况下,经标记的检测剂ssDNA包含猝灭剂/荧光剂对)。信号猝灭对的一个信号配偶体产生可检测的信号,并且另一个信号配偶体是猝灭第一信号配偶体的可检测信号的猝灭剂部分(即猝灭剂部分猝灭信号部分的信号,使得当信号配偶体彼此接近时,例如当信号对的信号配偶体非常接近时,来自信号部分的信号减少(猝灭))。

[0557] 例如,在一些情况下,当经标记的检测剂ssDNA被切割时,可检测信号的量增加。例如,在一些情况下,由一个信号配偶体(信号部分)显示的信号被另一个信号配偶体(猝灭剂信号部分)猝灭,例如当在由本公开的Cas12L多肽进行切割之前两者存在于相同的ssDNA分子上时。这种信号对在本文中称为“猝灭剂/荧光剂对”、“猝灭对”或“信号猝灭对”。例如,在一些情况下,一个信号配偶体(例如,第一信号配偶体)是产生可检测信号的信号部分,所述可检测信号由第二信号配偶体(例如,猝灭剂部分)猝灭。因此,当配偶体被分开时(例如,在由本公开的Cas12L多肽切割检测剂ssDNA之后),这种猝灭剂/荧光剂对的信号配偶体将产生可检测信号,但是当配偶体非常接近(例如,在由本公开的Cas12L多肽切割检测剂ssDNA之前)时,所述信号将被猝灭。

[0558] 猝灭剂部分可以在不同程度上猝灭来自信号部分的信号(例如,在本公开的

Cas12L多肽切割检测剂ssDNA之前)。在一些情况下,猝灭剂部分猝灭来自信号部分的信号,其中在存在猝灭剂部分的情况下(当信号配偶体彼此接近时)检测到的信号是在不存在猝灭剂部分的情况下(当信号配偶体分开时)检测到的信号的95%或更少。例如,在一些情况下,在猝灭剂部分存在下检测到的信号可以是猝灭剂部分不存在下检测到的信号的90%或更少、80%或更少、70%或更少、60%或更少、50%或更少、40%或更少、30%或更少、20%或更少、15%或更少、10%或更少或5%或更少。在一些情况下,在猝灭剂部分存在下,没有检测到信号(例如,高于背景)。

[0559] 在一些情况下,在猝灭剂部分不存在下(当信号配偶体被分离时)检测到的信号比在猝灭剂存在下检测到的信号大至少1.2倍(例如,大至少1.3倍、至少1.5倍、至少1.7倍、至少2倍、至少2.5倍、至少3倍、至少3.5倍、至少4倍、至少5倍、至少7倍、至少10倍、至少20倍或至少50倍)。

[0560] 在一些情况下,信号部分是荧光标记。在一些此类情况下,猝灭剂部分猝灭来自荧光标记的信号(光信号)(例如,通过吸收标记的发射光谱中的能量)。因此,当猝灭剂部分不靠近信号部分时,来自荧光标记的发射(信号)是可检测的,因为信号未被猝灭剂部分吸收。可以使用任何方便的供体受体对(信号部分/猝灭部分对),并且许多合适的对是本领域已知的。

[0561] 在一些情况下,猝灭剂部分从信号部分(本文也称为"可检测标记")吸收能量,然后发射信号(例如,不同波长的光)。因此,在一些情况下,猝灭剂部分本身是信号部分(例如,信号部分可以是6-羧基荧光素,而猝灭剂部分可以是6-羧基四甲基罗丹明),并且在一些此类情况下,所述对也可以是FRET对。在一些情况下,猝灭剂部分是暗猝灭剂。暗猝灭剂可以吸收激发能量,并以不同的方式(例如,作为热量)耗散能量。因此,暗猝灭剂自身的荧光很少甚至没有(不发射荧光)。暗猝灭剂的实例进一步描述于美国专利号8,822,673和8,586,718;美国专利公开20140378330、20140349295和20140194611;以及国际专利申请WO200142505和WO200186001,所有这些专利以引用方式整体并入本文。

[0562] 荧光标记的实例包括但不限于:Alexa Fluor®染料、ATTO染料(例如,ATTO 390、ATTO 425、ATTO 465、ATTO 488、ATTO 495、ATTO 514、ATTO 520、ATTO 532、ATTO Rho6G、ATTO 542、ATTO550、ATTO 565、ATTO Rho3B、ATTO Rho11、ATTO Rho12、ATTO Thio12、ATTO Rho101、ATTO 590、ATTO 594、ATTO Rho13、ATTO610、ATTO 620、ATTO Rho14、ATTO 633、ATTO 647、ATTO 647N、ATTO 655、ATTO Oxa12、ATTO 665、ATTO 680、ATTO 700、ATTO725、ATTO 740)、DyLight染料、花青染料(例如,Cy2、Cy3、Cy3.5、Cy3b、Cy5、Cy5.5、Cy7、Cy7.5)、FluoProbes染料、Sulfo Cy染料、Seta染料、IRIS染料、SeTau染料、SRfluor染料、Square染料、异硫氰酸荧光素(FITC)、四甲基罗丹明(TRITC)、德克萨斯红、俄勒冈绿、太平洋蓝、太平洋绿、太平洋橙、量子点和束缚荧光蛋白(tethered fluorescent protein)。

[0563] 在一些情况下,可检测标记是选自以下的荧光标记:Alexa Fluor®染料、ATTO染料(例如,ATTO 390、ATTO 425、ATTO 465、ATTO488、ATTO 495、ATTO 514、ATTO 520、ATTO 532、ATTO Rho6G、ATTO 542、ATTO 550、ATTO 565、ATTO Rho3B、ATTO Rho11、ATTO Rho12、ATTO Thio12、ATTO Rho101、ATTO 590、ATTO 594、ATTO Rho13、ATTO 610、ATTO 620、ATTO Rho14、ATTO 633、ATTO 647、ATTO 647N、ATTO 655、ATTO Oxa12、ATTO 665、ATTO 680、ATTO 700、ATTO 725、ATTO 740)、DyLight染料、花青染料(例如,Cy2、Cy3、Cy3.5、Cy3b、Cy5、Cy5.5、

Cy7、Cy7.5)、FluoProbes染料、Sulfo Cy染料、Seta染料、IRIS染料、SeTau染料、SRfluor染料、Square染料、荧光素(FITC)、四甲基罗丹明(TRITC)、德克萨斯红、俄勒冈绿、太平洋蓝、太平洋绿和太平洋橙。

[0564] 在一些情况下,可检测标记是选自以下的荧光标记:AlexaFluor®染料、ATTO染料(例如,ATTO 390、ATTO 425、ATTO 465、ATTO488、ATTO 495、ATTO 514、ATTO 520、ATTO 532、ATTO Rho6G、ATTO 542、ATTO 550、ATTO 565、ATTO Rho3B、ATTO Rho11、ATTO Rho12、ATTO Thio12、ATTO Rho101、ATTO 590、ATTO 594、ATTO Rho13、ATTO 610、ATTO 620、ATTO Rho14、ATTO 633、ATTO 647、ATTO 647N、ATTO 655、ATTO Oxa12、ATTO 665、ATTO 680、ATTO 700、ATTO 725、ATTO 740)、DyLight染料、花青染料(例如,Cy2、Cy3、Cy3.5、Cy3b、Cy5、Cy5.5、Cy7、Cy7.5)、FluoProbes染料、Sulfo Cy染料、Seta染料、IRIS染料、SeTau染料、SRfluor染料、Square染料、荧光素(FITC)、四甲基罗丹明(TRITC)、德克萨斯红、俄勒冈绿、太平洋蓝、太平洋绿、太平洋橙、量子点和束缚荧光蛋白。

[0565] ATTO染料的实例包括但不限于:ATTO 390、ATTO 425、ATTO465、ATTO 488、ATTO 495、ATTO 514、ATTO 520、ATTO 532、ATTO Rho6G、ATTO 542、ATTO 550、ATTO 565、ATTO Rho3B、ATTO Rho11、ATTO Rho12、ATTO Thio12、ATTO Rho101、ATTO590、ATTO 594、ATTO Rho13、ATTO 610、ATTO 620、ATTO Rho14、ATTO 633、ATTO 647、ATTO 647N、ATTO 655、ATTO Oxa12、ATTO 665、ATTO 680、ATTO 700、ATTO 725和ATTO 740。

[0566] AlexaFluor染料的实例包括但不限于:AlexaFluor®350、AlexaFluor®405、AlexaFluor®430、AlexaFluor®488、AlexaFluor®500、AlexaFluor®514、AlexaFluor®532、AlexaFluor®546、AlexaFluor®555、AlexaFluor®568、AlexaFluor®594、AlexaFluor®610、AlexaFluor®633、AlexaFluor®635、AlexaFluor®647、AlexaFluor®660、AlexaFluor®680、AlexaFluor®700、AlexaFluor®750、AlexaFluor®790等。

[0567] 猝灭剂部分的实例包括但不限于:暗猝灭剂、Black HoleQuencher® (BHQ®)(例如,BHQ-0、BHQ-1、BHQ-2、BHQ-3)、Qxl猝灭剂、ATTO猝灭剂(例如,ATTO 540Q、ATTO 580Q和ATTO612Q)、二甲氨基偶氮苯磺酸(Dabsyl)、Iowa Black RQ、Iowa Black FQ、IRDye QC-1、QSY染料(例如,QSY 7、QSY 9、QSY 21)、AbsoluteQuencher、Eclipse和金属簇诸如金纳米颗粒等。

[0568] 在一些情况下,猝灭剂部分选自:暗猝灭剂、Black HoleQuencher® (BHQ®)(例如,BHQ-0、BHQ-1、BHQ-2、BHQ-3)、Qxl猝灭剂、ATTO猝灭剂(例如,ATTO 540Q、ATTO 580Q和ATTO612Q)、二甲氨基偶氮苯磺酸(Dabsyl)、Iowa Black RQ、Iowa Black FQ、IRDye QC-1、QSY染料(例如,QSY 7、QSY 9、QSY 21)、AbsoluteQuencher、Eclipse和金属簇。

[0569] ATTO猝灭剂的实例包括但不限于:ATTO 540Q、ATTO 580Q和ATTO 612Q。Black HoleQuencher® (BHQ®)的实例包括但不限于:BHQ-0(493nm)、BHQ-1(534nm)、BHQ-2(579nm)和BHQ-3(672nm)。

[0570] 对于一些可检测标记(例如荧光染料)和/或猝灭剂部分的实例,参见例如Bao等人,Annu Rev Biomed Eng.2009;11:25-47;以及美国专利号8,822,673和8,586,718;美国专利公开20140378330、20140349295、20140194611、20130323851、20130224871、20110223677、20110190486、20110172420、20060179585和20030003486;以及国际专利申请

WO200142505和WO200186001,所有这些文献以引用方式整体并入本文中。

[0571] 在一些情况下,可以通过测量比色读出来检测对经标记的检测剂ssDNA的切割。例如,荧光团的释放(例如,从FRET对中释放、从猝灭剂/荧光剂对中释放等)可以导致可检测信号的波长偏移(并因此发生颜色偏移)。因此,在某些情况下,主题标记的检测剂ssDNA的切割可以通过颜色偏移来检测。这种偏移可以表示为一种颜色(波长)信号的量的损失、另一种颜色的量的增益、一种颜色与另一种颜色的比率的变化等。

[0572] 在一些情况下,本公开的CasL多肽可以反式切割RNA。因此,在一些情况下,经标记的检测核酸是经标记的RNA。因此,上述涉及经标记的检测剂DNA的公开内容同样适用于经标记的检测剂RNA。在一些情况下,经标记的检测剂RNA包含一个或多个修饰的糖、修饰的碱基和修饰的骨架。例如,经标记的检测剂RNA可以包含非天然核苷间键联、核酸模拟物、修饰的糖部分、修饰的核碱基、锁定核酸(LNA)、肽核酸(PNA)、吗啉代核酸和环己烯基核酸(CeNA)中的一种或多种。在某些情况下,使用两种或多种经标记的检测剂RNA。

[0573] 转基因非人生物体

[0574] 如上所述,在一些情况下,本公开的核酸(例如,重组表达载体)(例如,包含编码本公开的Cas12L多肽的核苷酸序列的核酸;包含编码本公开的Cas12L融合多肽的核苷酸序列的核酸等)用作转基因以生成转基因非人生物体,所述转基因非人生物体产生本公开的Cas12L多肽或Cas12L融合多肽。本公开提供一种转基因非人生物体,所述转基因非人生物体包含编码本公开的Cas12L多肽或Cas12L融合多肽的核苷酸序列。

[0575] 转基因非人动物

[0576] 本公开提供一种转基因非人动物,所述动物包含转基因,所述转基因包含含有编码Cas12L多肽或Cas12L融合多肽的核苷酸序列的核酸。在一些实施方案中,转基因非人动物的基因组包含编码本公开的Cas12L多肽或Cas12L融合多肽的核苷酸序列。在一些情况下,转基因非人动物对于遗传修饰是纯合的。在一些情况下,转基因非人动物对于遗传修饰是杂合的。在一些实施方案中,转基因非人动物是脊椎动物,例如鱼类(例如,鲑鱼、鳟鱼、斑马鱼、金鱼、河豚、洞穴鱼等)、两栖动物(青蛙、蝾螈、火蜥蜴等)、鸟类(例如,鸡、火鸡等)、爬行动物(例如,蛇、蜥蜴等)、非人哺乳动物(例如,有蹄类动物,例如猪、牛、山羊、绵羊等;兔形目动物(例如,兔);啮齿动物(例如,大鼠、小鼠);非人灵长类动物等)等。在一些情况下,转基因非人动物是无脊椎动物。在一些情况下,转基因非人动物是昆虫(例如,蚊子;农业害虫等)。在一些情况下,转基因非人动物是蛛形纲动物。

[0577] 编码本公开的Cas12L多肽或Cas12L融合多肽的核苷酸序列可在未知启动子(例如,当核酸随机整合到宿主细胞基因组中时)的控制之下(即,可操作地连接至未知启动子)或可在已知启动子的控制之下(即,可操作地连接至已知启动子)。合适的已知启动子可以是任何已知启动子并且包括组成型活性启动子(例如,CMV启动子)、诱导型启动子(例如,热休克启动子、四环素调控的启动子、类固醇调控的启动子、金属调控的启动子、雌激素受体调控的启动子等)、空间限制的和/或时间限制的启动子(例如,组织特异性启动子、细胞类型特异性启动子等)等。

[0578] 转基因植物

[0579] 如上所述,在一些情况下,本公开的核酸(例如,重组表达载体)(例如,包含编码本公开的Cas12L多肽的核苷酸序列的核酸;包含编码本公开的Cas12L融合多肽的核苷酸序列

的核酸等)用作转基因以生成转基因植物,所述转基因植物产生本公开的Cas12L多肽或Cas12L融合多肽。本公开提供了一种转基因植物,其包含编码本公开的Cas12L多肽或Cas12L融合多肽的核苷酸序列。在一些情况下,转基因植物的基因组包含主题核酸。在一些情况下,转基因植物对于遗传修饰为纯合的。在一些情况下,转基因植物对于遗传修饰为杂合的。

[0580] 将外源核酸引入植物细胞中的方法在本领域中是众所周知的。如上所定义,此类植物细胞被认为是"转化的"。合适的方法包括病毒感染(诸如双链DNA病毒)、转染、缀合、原生质体融合、电穿孔、粒子枪技术、磷酸钙沉淀、直接微注射、碳化硅晶须技术、土壤杆菌属介导的转化等。方法的选择一般取决于待转化的细胞类型和发生转化所在的环境(即体外、离体或体内)。

[0581] 基于土壤细菌根瘤土壤杆菌(Agrobacterium tumefaciens)的转化方法特别可用于将外源核酸分子引入维管植物中。土壤杆菌属(Agrobacterium)的野生型形式含有Ti(肿瘤诱导)质粒,该质粒引导在宿主植物上生长的致瘤冠瘿的产生。Ti质粒的肿瘤诱导T-DNA区向植物基因组的转移需要Ti质粒编码的毒力基因以及T-DNA边缘序列,所述T-DNA边缘序列是描绘待转移区域的一组正向DNA重复序列。基于土壤杆菌属的载体是Ti质粒的修饰形式,其中肿瘤诱导功能被待引入植物宿主中的目标核酸序列替代。

[0582] 土壤杆菌属介导的转化通常采用共合体载体或二元载体系统,其中Ti质粒的组分在辅助载体(所述辅助载体永久存在于土壤杆菌属宿主中并且携带毒力基因)与穿梭载体(所述穿梭载体含有被T-DNA序列限定的感兴趣的基因)之间分配。多种二元载体在本领域中是众所周知的并且可例如从Clontech(Palo Alto,Calif.)商购获得。例如用培养的植物细胞或创伤组织诸如叶组织、根外植体、下胚轴体、茎块或块茎共同培养土壤杆菌属的方法在本领域中也是众所周知的。参见例如Glick和Thompson(编),Methods in Plant Molecular Biology and Biotechnology,Boca Raton,Fla.:CRC Press(1993)。

[0583] 微粒介导的转化还可用来产生主题转基因植物。首先由Klein等人(Nature 327:70-73(1987))描述的这种方法依赖于微粒(诸如金或钨),所述微粒通过用氯化钙、亚精胺或聚乙二醇沉淀包被有所需的核酸分子。微粒颗粒使用诸如BIOLISTIC PD-1000(Biorad;Hercules Calif.)的装置在高速下被加速到被子植物组织中。

[0584] 可将本公开的核酸(例如,包含编码本公开的Cas12L多肽或Cas12L融合多肽的核苷酸序列的核酸(例如,重组表达载体))以使得核酸能够例如通过体内或离体方案进入一种或多种植物细胞的方式引入植物中。"体内"意指向植物的活体施用核酸,例如渗透。"离体"意指在植物外部修饰细胞或外植体,然后使此类细胞或器官再生为植物。已描述了适用于稳定转化植物细胞或建立转基因植物的多种载体,包括描述于Weissbach和Weissbach,(1989)Methods for Plant Molecular Biology Academic Press以及Gelvin等人,(1990)Plant Molecular Biology Manual,Kluwer Academic Publishers中的那些载体。具体实例包括衍生自根瘤土壤杆菌的Ti质粒的那些,以及由Herrera-Estrella等人(1983)Nature 303:209,Bevan(1984)Nucl Acid Res.12:8711-8721,Klee(1985)Bio/Technolo 3:637-642公开的那些。可替代地,非Ti载体可用来通过使用游离DNA递送技术将DNA转移到植物和细胞中。通过使用这些方法,可产生转基因植物,诸如小麦、大米(Christou(1991)Bio/Technology 9:957-9and 4462)和玉米(Gordon-Kamm(1990)Plant Cell 2:603-618)。未成

熟胚也可以是通过使用粒子枪的直接DNA递送技术(Weeks等人(1993)Plant Physiol102:1077-1084;Vasil(1993)Bio/Technolo 10:667-674;Wan和Lemeaux(1994)Plant Physiol 104:37-48)和土壤杆菌属介导的DNA转移(Ishida等人(1996)Nature Biotech 14:745-750)的单子叶植物的良好靶组织。用于将DNA引入叶绿体中的示例性方法是生物弹轰击、原生质体的聚乙二醇转化和微注射(Danieli等人Nat.Biotechnol16:345-348,1998;Staub等人Nat.Biotechnol 18:333-338,2000;O'Neill等人Plant J.3:729-738,1993;Knoblauch等人Nat.Biotechnol 17:906-909;美国专利号5,451,513、5,545,817、5,545,818和5,576,198;国际申请号WO 95/16783;以及Boynton等人,Methods in Enzymology217:510-536(1993);Svab等人,Proc.Natl.Acad.Sci.USA 90:913-917(1993);和McBride等人,Proc.Natl.Acad.Sci.USA 91:7301-7305(1994))。适用于生物弹轰击、原生质体的聚乙二醇转化以及微注射的方法的任何载体将适用作用于叶绿体转化的靶向载体。任何双链DNA载体可用作转化载体,尤其当引入方法没有使用土壤杆菌属时。

[0585] 可遗传修饰的植物包括谷物、饲料作物、水果、蔬菜、油籽作物、棕榈植物、林业植物和藤本植物。可修饰的植物的具体实例如下:玉米、香蕉、花生、红豌豆、向日葵、番茄、芸苔、烟草、小麦、大麦、燕麦、土豆、大豆、棉花、康乃馨、高粱、羽扇豆和大米。

[0586] 本公开提供转化的植物细胞,含有转化的植物细胞的组织、植物和产品。主题转化细胞以及包含所述转化细胞的组织和产品的特征是存在整合到基因组中的主题核酸,和通过本公开的Cas12L多肽或Cas12L融合多肽的植物细胞来产生。本发明的重组植物细胞可作为重组细胞群或作为组织、种子、全株植物、茎、果实、叶、根、花、茎、块茎、谷物、动物饲料、植田等使用。

[0587] 编码本公开的Cas12L多肽或Cas12L融合多肽的核苷酸序列可在未知启动子(例如,当核酸随机整合到宿主细胞基因组中时)的控制之下(即,可操作地连接至未知启动子)或可在已知启动子的控制之下(即,可操作地连接至已知启动子)。合适的已知启动子可以是任何已知的启动子并且包括组成型活性启动子、诱导型启动子、空间限制的和/或时间限制的启动子等。

[0588] 本公开的非限制性方面的实例

[0589] 上述本发明主题的包括各实施方案的各方面可在单独或与一个或多个其他方面或实施方案组合的情况下是有益的。在不限制前述描述的情况下,下文提供了本公开的某些非限制性方面。如本领域的技术人员在阅读本公开后将显而易见的,单独编号方面中的每一者可与先前或随后单独编号方面中的任一者一起使用或组合。这旨在为各方面的所有此类组合提供支持,并且不限于以下明确提供的各方面的组合:

[0590] 方面1.一种组合物,其包含:a)Cas12L多肽或编码所述Cas12L多肽的核酸分子;和b)Cas12L引导RNA或一种或多种编码Cas12L引导RNA的DNA分子。

[0591] 方面2.根据方面1所述的组合物,其中所述Cas12L多肽包含与图2A-2Z、图2AA-2ZZ和图2AAA-2FFF中任一者中描绘的氨基酸序列具有50%或更多氨基酸序列同一性的氨基酸序列。

[0592] 方面3.根据方面1或方面2所述的组合物,其中所述Cas12L引导RNA包含与图7、图2A-2Z、图2AA-2ZZ和图2AAA-2FFF中描绘的crRNA序列中的任一者具有80%、90%、95%、98%、99%或100%核苷酸序列同一性的核苷酸序列。

[0593] 方面4.根据方面1或方面2所述的组合物,其中所述Cas12L多肽与一个或多个核定位信号(NLS)融合。

[0594] 方面5.根据方面1-4中任一项所述的组合物,其中所述组合物包含脂质。

[0595] 方面6.根据方面1-4中任一项所述的组合物,其中a)和b)均在脂质体内。

[0596] 方面7.根据方面1-4中任一项所述的组合物,其中a)和b)均在颗粒内。

[0597] 方面8.根据方面1-7中任一项所述的组合物,其包含以下中的一种或多种:缓冲剂、核酸酶抑制剂和蛋白酶抑制剂。

[0598] 方面9.根据方面1-8中任一项所述的组合物,其中所述Cas12L多肽包含与图2A-2Z、图2AA-2ZZ和图2AAA-2FFF中任一者中描绘的氨基酸序列具有85%或更多同一性的氨基酸序列。

[0599] 方面10.根据方面1-9中任一项所述的组合物,其中所述Cas12L多肽是仅切割双链靶核酸分子的一条链的切口酶。

[0600] 方面11.根据方面1-9中任一项所述的组合物,其中所述Cas12L多肽是无催化活性的Cas12L多肽(dCas12L)。

[0601] 方面12.根据方面1-11中任一项所述的组合物,其中所述Cas12L多肽具有700个氨基酸至800个氨基酸的长度。

[0602] 方面13.根据方面1-12中任一项所述的组合物,其还包含DNA供体模板。

[0603] 方面14.一种Cas12L融合多肽,其包含:与异源多肽融合的Cas12L多肽。

[0604] 方面15.根据方面14所述的Cas12L融合多肽,其中所述Cas12L多肽包含与图2A-2Z、图2AA-2ZZ和图2AAA-2FFF中任一者中描绘的氨基酸序列具有50%或更多同一性的氨基酸序列。

[0605] 方面16.根据方面14所述的Cas12L融合多肽,其中所述Cas12L多肽包含与图2A-2Z、图2AA-2ZZ和图2AAA-2FFF中任一者中描绘的氨基酸序列具有85%或更多同一性的氨基酸序列。

[0606] 方面17.根据方面14-16中任一项所述的Cas12L融合多肽,其中所述Cas12L多肽是仅切割双链靶核酸分子的一条链的切口酶。

[0607] 方面18.根据方面14-17中任一项所述的Cas12L融合多肽,其中所述Cas12L多肽是无催化活性的Cas12L多肽(dCas12L)。

[0608] 方面19.根据方面14-18中任一项所述的Cas12L融合多肽,其中所述Cas12L多肽具有700个氨基酸至800个氨基酸的长度。

[0609] 方面20.根据方面14-19中任一项所述的Cas12L融合多肽,其中所述异源多肽融合到所述Cas12L多肽的N-末端和/或C-末端。

[0610] 方面21.根据方面14-20中任一项所述的Cas12L融合多肽,其包含核定位信号(NLS)。

[0611] 方面22.根据方面14-21中任一项所述的Cas12L融合多肽,其中所述异源多肽是提供与靶细胞或靶细胞类型上的细胞表面部分结合的靶向多肽。

[0612] 方面23.根据方面14-21中任一项所述的Cas12L融合多肽,其中所述异源多肽表现出酶促活性。

[0613] 方面24.根据方面23所述的Cas12L融合多肽,其中所述异源多肽表现出选自以下

的一种或多种酶促活性:核酸酶活性、甲基转移酶活性、脱甲基化酶活性、DNA修复活性、DNA损伤活性、脱氨基活性、歧化酶活性、烷基化活性、脱嘌呤活性、氧化活性、嘧啶二聚体形成活性、整合酶活性、转座酶活性、重组酶活性、聚合酶活性、连接酶活性、解旋酶活性、光裂合酶活性和糖基化酶活性。

[0614] 方面25.根据方面23所述的Cas12L融合多肽,其中所述异源多肽表现出选自以下的一种或多种酶活性:逆转录酶活性、核酸酶活性、甲基转移酶活性、去甲基化酶活性、脱氨基活性、脱嘌呤活性、整合酶活性、转座酶活性和重组酶活性。

[0615] 方面26.根据方面14-21中任一项所述的Cas12L融合多肽,其中所述异源多肽表现出修饰与靶核酸相关的靶多肽的酶促活性。

[0616] 方面27.根据方面26所述的Cas12L融合多肽,其中所述异源多肽表现出组蛋白修饰活性。

[0617] 方面28.根据方面26或方面27所述的Cas12L融合多肽,其中所述异源多肽表现出选自以下的一种或多种酶活性:甲基转移酶活性、脱甲基酶活性、乙酰转移酶活性、脱乙酰酶活性、激酶活性、磷酸酶活性、泛素连接酶活性、脱泛素化活性、腺苷酸化活性、脱腺苷酸化活性、SUMO化活性、脱SUMO化活性、核糖基化活性、脱核糖基化活性、豆蔻酰化活性、脱豆蔻酰化活性、糖基化活性(例如来自O-GlcNAc转移酶)和脱糖基化活性。

[0618] 方面29.根据方面28所述的Cas12L融合多肽,其中所述异源多肽表现出选自以下的一种或多种酶活性:甲基转移酶活性、脱甲基酶活性、乙酰转移酶活性和脱乙酰酶活性。

[0619] 方面30.根据方面14-21中任一项所述的Cas12L融合多肽,其中所述异源多肽是核内体逃逸多肽。

[0620] 方面31.根据方面30所述的Cas12L融合多肽,其中所述核内体逃逸多肽包含选自GLFXALLXLLXSLWXLLLXA(SEQ ID NO:72)和GLFHALLHLLHSLWHLLLHA(SEQ ID NO:73)的氨基酸序列,其中每个X独立地选自赖氨酸、组氨酸和精氨酸。

[0621] 方面32.根据方面14-21中任一项所述的Cas12L融合多肽,其中所述异源多肽是叶绿体转运肽。

[0622] 方面33.根据方面14-21中任一项所述的Cas12L融合多肽,其中所述异源多肽包含蛋白质转导结构域。

[0623] 方面34.根据方面14-21中任一项所述的Cas12L融合多肽,其中所述异源多肽是增加或减少转录的蛋白质。

[0624] 方面35.根据方面34所述的Cas12L融合多肽,其中所述异源多肽是转录阻遏子结构域。

[0625] 方面36.根据方面34所述的Cas12L融合多肽,其中所述异源多肽是转录激活结构域。

[0626] 方面37.根据方面14-21中任一项所述的Cas12L融合多肽,其中所述异源多肽是蛋白质结合结构域。

[0627] 方面38.一种核酸,其包含编码根据方面14-37中任一项所述的Cas12L融合多肽的核苷酸序列。

[0628] 方面39.根据方面38所述的核酸,其中编码所述Cas12L融合多肽的所述核苷酸序列可操作地连接至启动子。

[0629] 方面40.根据方面39所述的核酸,其中所述启动子在真核细胞中有功能。

[0630] 方面41.根据方面40所述的核酸,其中所述启动子在以下一种或多种细胞中有功能:植物细胞、真菌细胞、动物细胞、无脊椎动物细胞、蝇细胞、脊椎动物细胞、哺乳动物细胞、灵长类动物细胞、非人灵长类动物细胞和人细胞。

[0631] 方面42.根据方面39-41中任一项所述的核酸,其中所述启动子是以下一种或多种:组成型启动子、诱导型启动子、细胞类型特异性启动子和组织特异性启动子。

[0632] 方面43.根据方面38-42中任一项所述的核酸,其中所述核酸是重组表达载体。

[0633] 方面44.根据方面43所述的核酸,其中所述重组表达载体是重组腺相关病毒载体、重组逆转录病毒载体或重组慢病毒载体。

[0634] 方面45.根据方面39所述的核酸,其中所述启动子在原核细胞中有功能。

[0635] 方面46.根据方面38所述的核酸,其中所述核酸分子是mRNA。

[0636] 方面47.一种或多种核酸,其包含:(a)编码Cas12L引导RNA的核苷酸序列;和(b)编码Cas12L多肽的核苷酸序列。

[0637] 方面48.根据方面47所述的一种或多种核酸,其中所述Cas12L多肽包含与图2A-2Z、图2AA-2ZZ和图2AAA-2FFF中任一者中描绘的氨基酸序列具有50%或更多同一性的氨基酸序列。

[0638] 方面49.根据方面47所述的一种或多种核酸,其中所述Cas12L多肽包含与图2A-2Z、图2AA-2ZZ和图2AAA-2FFF中任一者中描绘的氨基酸具有85%或更多同一性的氨基酸序列。

[0639] 方面50.根据方面47-49中任一项所述的一种或多种核酸,其中所述Cas12L引导RNA包含与图7、图2A-2Z、图2AA-2ZZ和图2AAA-2FFF中列出的crRNA序列中的任一者具有80%或更多核苷酸序列同一性的核苷酸序列。

[0640] 方面51.根据方面47-50中任一项所述的一种或多种核酸,其中所述Cas12L多肽融合至一种或多种核定位信号(NLS)。

[0641] 方面52.根据方面47-51中任一项所述的一种或多种核酸,其中编码所述Cas12L引导RNA的所述核苷酸序列可操作地连接至启动子。

[0642] 方面53.根据方面47-52中任一项所述的一种或多种核酸,其中编码所述Cas12L多肽的所述核苷酸序列可操作地连接至启动子。

[0643] 方面54.根据方面52或方面53所述的一种或多种核酸,其中可操作地连接至编码所述Cas12L引导RNA的所述核苷酸序列的所述启动子和/或可操作地连接至编码所述Cas12L多肽的所述核苷酸序列的所述启动子,在真核细胞中有功能。

[0644] 方面55.根据方面54所述的一种或多种核酸,其中所述启动子在以下一种或多种细胞中有功能:植物细胞、真菌细胞、动物细胞、无脊椎动物细胞、蝇细胞、脊椎动物细胞、哺乳动物细胞、灵长类动物细胞、非人灵长类动物细胞和人细胞。

[0645] 方面56.根据方面53-55中任一项所述的一种或多种核酸,其中所述启动子是以下一种或多种:组成型启动子、诱导型启动子、细胞类型特异性启动子和组织特异性启动子。

[0646] 方面57.根据方面47-56中任一项所述的一种或多种核酸,其中所述一种或多种核酸是一种或多种重组表达载体。

[0647] 方面58.根据方面57所述的一种或多种核酸,其中所述一种或多种重组表达载体

选自:一种或多种腺相关病毒载体、一种或多种重组逆转录病毒载体或一种或多种重组慢病毒载体。

[0648] 方面59.根据方面53所述的一种或多种核酸,其中所述启动子在原核细胞中有功能。

[0649] 方面60.一种真核细胞,其包含以下一种或多种:

[0650] a)Cas12L多肽或包含编码所述Cas12L多肽的核苷酸序列的核酸,

[0651] b)Cas12L融合多肽或包含编码所述Cas12L融合多肽的核苷酸序列的核酸,和

[0652] c)Cas12L引导RNA或包含编码所述Cas12L引导RNA的核苷酸序列的核酸。

[0653] 方面61.根据方面60所述的真核细胞,其包含编码所述Cas12L多肽的所述核酸,其中所述核酸整合到所述细胞的基因组DNA中。

[0654] 方面62.根据方面60或方面61所述的真核细胞,其中所述真核细胞是植物细胞、哺乳动物细胞、昆虫细胞、蜘蛛细胞、真菌细胞、鸟类细胞、爬行动物细胞、两栖动物细胞、无脊椎动物细胞、小鼠细胞、大鼠细胞、灵长类动物细胞、非人灵长类动物细胞或人细胞。

[0655] 方面63.一种细胞,其包含包含Cas12L融合多肽或包含编码所述Cas12L融合多肽的核苷酸序列的核酸。

[0656] 方面64.根据方面63所述的细胞,其中所述细胞是原核细胞或真核细胞。

[0657] 方面65.根据方面63或方面64所述的细胞,其包含有包含编码所述Cas12L融合多肽的核苷酸序列的核酸,其中所述核酸分子整合到所述细胞的基因组DNA中。

[0658] 方面66.一种修饰靶核酸的方法,所述方法包括使靶核酸与以下项接触:a)Cas12L多肽;和b)包含与所述靶核酸的靶序列杂交的引导序列的Cas12L引导RNA,其中所述接触导致所述靶核酸被所述Cas12L多肽修饰。

[0659] 方面67.根据方面66所述的方法,其中所述修饰是所述靶核酸的切割。

[0660] 方面68.根据方面66或方面67所述的方法,其中所述靶核酸选自:双链DNA、单链DNA、RNA、基因组DNA和染色体外DNA。

[0661] 方面69.根据方面66-68中任一项所述的方法,其中所述接触在体外在细胞外部发生。

[0662] 方面70.根据方面66-68中任一项所述的方法,其中所述接触在培养的细胞内部发生。

[0663] 方面71.根据方面66-68中任一项所述的方法,其中所述接触在体内细胞内部发生。

[0664] 方面72.根据方面70或方面71所述的方法,其中所述细胞是真核细胞。

[0665] 方面73.根据方面72所述的方法,其中所述细胞选自:植物细胞、真菌细胞、哺乳动物细胞、爬行动物细胞、昆虫细胞、禽细胞、鱼细胞、寄生虫细胞、节肢动物细胞、蛛形纲动物细胞、无脊椎动物细胞、脊椎动物细胞、啮齿动物细胞、小鼠细胞、大鼠细胞、灵长类动物细胞、非人灵长类动物细胞和人细胞。

[0666] 方面74.根据方面70或方面71所述的方法,其中所述细胞是原核细胞。

[0667] 方面75.根据方面66-74中任一项所述的方法,其中所述接触导致基因组编辑。

[0668] 方面76.根据方面66-75中任一项所述的方法,其中所述接触包括:向细胞中引入:(a)所述Cas12L多肽或包含编码所述Cas12L多肽的核苷酸序列的核酸,和(b)所述Cas12L引

导RNA或包含编码所述Cas12L引导RNA的核苷酸序列的核酸。

[0669] 方面77.根据方面76所述的方法，其中所述接触还包括：将DNA供体模板引入到所述细胞中。

[0670] 方面78.根据方面66-77中任一项所述的方法，其中所述Cas12L引导RNA包含与图7、图2A-2Z、图2AA-2ZZ和图2AAA-2FFF中列出的crRNA序列中的任一者具有80％或更多核苷酸序列同一性的核苷酸序列。

[0671] 方面79.根据方面66-78中任一项所述的方法，其中所述Cas12L多肽融合至核定位信号。

[0672] 方面80.一种调节靶DNA转录、修饰靶核酸或修饰与靶核酸相关的蛋白质的方法，所述方法包括将靶核酸与以下项接触：a)包含与异源多肽融合的Cas12L多肽的Cas12L融合多肽；和b)包含与靶核酸的靶序列杂交的引导序列的Cas12L引导RNA。

[0673] 方面81.根据方面80所述的方法，其中所述Cas12L引导RNA包含与图7、图2A-2Z、图2AA-2ZZ和图2AAA-2FFF中列出的crRNA序列中的任一者具有80％或更多核苷酸序列同一性的核苷酸序列。

[0674] 方面82.根据方面80或方面81所述的方法，其中所述Cas12L融合多肽包含核定位信号。

[0675] 方面83.根据方面80-82中任一项所述的方法，其中所述修饰不是所述靶核酸的切割。

[0676] 方面84.根据方面80-83中任一项所述的方法，其中所述靶核酸选自：双链DNA、单链DNA、RNA、基因组DNA和染色体外DNA。

[0677] 方面85.根据方面80-84中任一项所述的方法，其中所述接触在体外在细胞外部发生。

[0678] 方面86.根据方面80-84中任一项所述的方法，其中所述接触发生在培养的细胞内部。

[0679] 方面87.根据方面80-84中任一项所述的方法，其中所述接触发生在体内细胞内部。

[0680] 方面88.根据方面86或方面87所述的方法，其中所述细胞是真核细胞。

[0681] 方面89.根据方面88所述的方法，其中所述细胞选自：植物细胞、真菌细胞、哺乳动物细胞、爬行动物细胞、昆虫细胞、禽细胞、鱼细胞、寄生虫细胞、节肢动物细胞、蛛形纲动物细胞、无脊椎动物细胞、脊椎动物细胞、啮齿动物细胞、小鼠细胞、大鼠细胞、灵长类动物细胞、非人灵长类动物细胞和人细胞。

[0682] 方面90.根据方面86或方面87所述的方法，其中所述细胞是原核细胞。

[0683] 方面91.根据方面80-90中任一项所述的方法，其中所述接触包括：向细胞中引入：(a)所述Cas12L融合多肽或包含编码所述Cas12L融合多肽的核苷酸序列的核酸，和(b)所述Cas12L引导RNA或包含编码所述Cas12L引导RNA的核苷酸序列的核酸。

[0684] 方面92.根据方面80-91中任一项所述的方法，其中所述Cas12L多肽是无催化活性的Cas12L多肽(dCas12L)。

[0685] 方面93.根据方面80-92中任一项所述的方法，其中所述Cas12L多肽具有700个氨基酸至800个氨基酸的长度。

[0686] 方面94.根据方面80-93中任一项所述的方法，其中所述异源多肽表现出酶促活性。

[0687] 方面95.根据方面94所述的方法，其中所述异源多肽表现出选自以下的一种或多种酶活性：核酸酶活性、甲基转移酶活性、脱甲基化酶活性、DNA修复活性、DNA损伤活性、脱氨基活性、歧化酶活性、烷基化活性、脱嘌呤活性、氧化活性、嘧啶二聚体形成活性、整合酶活性、转座酶活性、重组酶活性、聚合酶活性、连接酶活性、解旋酶活性、光裂合酶活性和糖基化酶活性。

[0688] 方面96.根据方面94所述的方法，其中所述异源多肽表现出选自以下的一种或多种酶活性：逆转录酶活性、核酸酶活性、甲基转移酶活性、去甲基化酶活性、脱氨基活性、脱嘌呤活性、整合酶活性、转座酶活性和重组酶活性。

[0689] 方面97.根据方面80-93中任一项所述的方法，其中所述异源多肽表现出修饰与靶核酸相关的靶多肽的酶促活性。

[0690] 方面98.根据方面97所述的方法，其中所述异源多肽表现出组蛋白修饰活性。

[0691] 方面99.根据方面97或方面98所述的方法，其中所述异源多肽表现出选自以下的一种或多种酶活性：甲基转移酶活性、脱甲基酶活性、乙酰转移酶活性、脱乙酰酶活性、激酶活性、磷酸酶活性、泛素连接酶活性、脱泛素化活性、腺苷酸化活性、脱腺苷酸化活性、SUMO化活性、脱SUMO化活性、核糖基化活性、脱核糖基化活性、豆蔻酰化活性、脱豆蔻酰化活性、糖基化活性(例如来自O-GlcNAc转移酶)和脱糖基化活性。

[0692] 方面100.根据方面99所述的方法，其中所述异源多肽表现出选自以下的一种或多种酶活性：甲基转移酶活性、脱甲基酶活性、乙酰转移酶活性和脱乙酰酶活性。

[0693] 方面101.根据方面80-93中任一项所述的方法，其中所述异源多肽是增加或减少转录的蛋白质。

[0694] 方面102.根据方面101所述的方法，其中所述异源多肽是转录阻遏子结构域。

[0695] 方面103.根据方面101所述的方法，其中所述异源多肽是转录激活结构域。

[0696] 方面104.根据方面80-93中任一项所述的方法，其中所述异源多肽是蛋白质结合结构域。

[0697] 方面105.一种转基因多细胞非人生物体，其基因组包含有包含编码以下一种或多种的核苷酸序列的转基因：

[0698] a)Cas12L多肽，

[0699] b)Cas12L融合多肽，和

[0700] c)Cas12L引导RNA。

[0701] 方面106.根据方面105所述的转基因多细胞非人生物体，其中所述Cas12L多肽包含与图2A-2Z、图2AA-2ZZ和图2AAA-2FFF中任一者中列出的氨基酸序列具有50％或更多氨基酸序列同一性的氨基酸序列。

[0702] 方面107.根据方面105所述的转基因多细胞非人生物体，其中所述Cas12L多肽包含与图2A-2Z、图2AA-2ZZ和图2AAA-2FFF中任一者中列出的氨基酸序列具有85％或更多氨基酸序列同一性的氨基酸序列。

[0703] 方面108.根据方面105-107中任一项所述的转基因多细胞非人生物体，其中所述生物体是植物、单子叶植物、双子叶植物、无脊椎动物、昆虫、节肢动物、蛛形纲动物、寄生

虫、蠕虫、刺胞动物、脊椎动物、鱼、爬行动物、两栖动物、有蹄动物、鸟、猪、马、羊、啮齿动物、小鼠、大鼠或非人类灵长类动物。

[0704] 方面109.一种系统,其包括:

[0705] a)Cas12L多肽和Cas12L引导RNA;

[0706] b)Cas12L多肽、Cas12L引导RNA和DNA供体模板;

[0707] c)Cas12L融合多肽和Cas12L引导RNA;

[0708] d)Cas12L融合多肽、Cas12L引导RNA和DNA供体模板;

[0709] e)编码Cas12L多肽的mRNA和Cas12L引导RNA;

[0710] f)编码Cas12L多肽的mRNA;Cas12L引导RNA和DNA供体模板;

[0711] g)编码Cas12L融合多肽的mRNA和Cas12L引导RNA;

[0712] h)编码Cas12L融合多肽的mRNA、Cas12L引导RNA和DNA供体模板;

[0713] i)一种或多种重组表达载体,其包含:i)编码Cas12L多肽的核苷酸序列;和ii)编码Cas12L引导RNA的核苷酸序列;

[0714] j)一种或多种重组表达载体,其包含:i)编码Cas12L多肽的核苷酸序列;ii)编码Cas12L引导RNA的核苷酸序列;和iii)DNA供体模板;

[0715] k)一种或多种重组表达载体,其包含:i)编码Cas12L融合多肽的核苷酸序列;和ii)编码Cas12L引导RNA的核苷酸序列;和

[0716] l)一种或多种重组表达载体,其包含:i)编码Cas12L融合多肽的核苷酸序列;ii)编码Cas12L引导RNA的核苷酸序列;和DNA供体模板。

[0717] 方面110.根据方面109所述的Cas12L系统,其中所述Cas12L多肽包含与图2A-2Z、图2AA-2ZZ和图2AAA-2FFF中任一者中描绘的氨基酸序列具有50%或更多氨基酸序列同一性的氨基酸序列。

[0718] 方面111.根据方面109所述的Cas12L系统,其中所述Cas12L多肽包含与图2A-2Z、图2AA-2ZZ和图2AAA-2FFF中任一者中描绘的氨基酸序列具有85%或更多氨基酸序列同一性的氨基酸序列。

[0719] 方面112.根据方面109-111中任一项所述的Cas12L系统,其中所述供体模板核酸具有8个核苷酸至1000个核苷酸的长度。

[0720] 方面113.根据方面109-111中任一项所述的Cas12L系统,其中所述供体模板核酸具有25个核苷酸至500个核苷酸的长度。

[0721] 方面114.一种试剂盒,其包括根据方面109-113中任一项所述的Cas12L系统。

[0722] 方面115.根据方面114所述的试剂盒,其中所述试剂盒的组分在同一容器中。

[0723] 方面116.根据方面114所述的试剂盒,其中所述试剂盒的组分在单独的容器中。

[0724] 方面117.一种无菌容器,其包括根据方面109-116中任一项所述的Cas12L系统。

[0725] 方面118.根据方面117所述的无菌容器,其中所述容器是注射器。

[0726] 方面119.一种可植入装置,其包括根据方面109-116中任一项所述的Cas12L系统。

[0727] 方面120.根据方面119所述的可植入装置,其中所述Cas12L系统在基质内。

[0728] 方面121.根据方面119所述的可植入装置,其中所述Cas12L系统位于储器中。

[0729] 方面122.一种检测样品中的靶DNA的方法,所述方法包括:(a)使所述样品与以下物质接触:(i)Cas12L多肽;(ii)引导RNA,所述引导RNA包含与所述Cas12L多肽结合的区域

和与所述靶DNA杂交的引导序列;和(iii)为单链的且不与所述引导RNA的所述引导序列杂交的检测剂核酸;以及(b)测量由所述Cas12L多肽切割所述单链检测剂核酸而产生的可检测信号,从而检测所述靶DNA。

[0730] 方面123.根据方面122所述的方法,其中所述靶DNA是细菌DNA。

[0731] 方面124.根据方面122所述的方法,其中所述靶DNA是病毒DNA。

[0732] 方面125.根据方面124所述的方法,其中所述靶DNA是乳多空病毒、人乳头瘤病毒(HPV)、嗜肝DNA病毒、乙型肝炎病毒(HBV)、疱疹病毒、水痘带状疱疹病毒(VZV)、爱泼斯坦-巴尔病毒(Epstein-Barr virus,EBV)、卡波济氏肉瘤相关疱疹病毒、腺病毒、痘病毒或细小病毒DNA。

[0733] 方面126.根据方面122所述的方法,其中所述靶DNA来自人细胞。

[0734] 方面127.根据方面122所述的方法,其中所述靶DNA是人胎儿细胞或癌细胞DNA。

[0735] 方面128.根据方面122所述的方法,其中所述样品包括细胞裂解物。

[0736] 方面129.根据方面122所述的方法,其中所述样品包括细胞。

[0737] 方面130.根据方面122所述的方法,其中所述样品是血液、血清、血浆、尿液、吸出物或活检样品。

[0738] 方面131.根据方面122-130中任一项所述的方法,所述方法还包括确定所述样品中存在的所述靶DNA的量。

[0739] 方面132.根据方面131所述的方法,其中所述测量可检测信号包括以下一种或多种:基于视觉的检测、基于传感器的检测、颜色检测、基于金纳米颗粒的检测、荧光偏振、胶体相变/分散、电化学检测和基于半导体的感测。

[0740] 方面133.根据方面122-132中任一项所述的方法,其中所述经标记的检测剂核酸是DNA。

[0741] 方面134.根据方面122-132中任一项所述的方法,其中所述经标记的检测剂核酸是RNA。

[0742] 方面135.根据方面122-134中任一项所述的方法,其中所述经标记的检测剂核酸包含修饰的核碱基、修饰的糖部分和/或修饰的核酸键。

[0743] 方面136.根据方面122-135中任一项所述的方法,其中所述可检测信号在少于45分钟内是可检测的。

[0744] 方面137.根据方面122-135中任一项所述的方法,其中所述可检测信号在少于30分钟内是可检测的。

[0745] 方面138.根据方面122-137中任一项所述的方法,其还包括通过环介导等温扩增(LAMP)、解旋酶依赖性扩增(HDA)、重组酶聚合酶扩增(RPA)、链置换扩增(SDA)、基于核酸序列的扩增(NASBA)、转录介导扩增(TMA)、切口酶扩增反应(NEAR)、滚环扩增(RCA)、多置换扩增(MDA)、分枝(RAM)、环状解旋酶依赖性扩增(cHDA)、单引物等温扩增(SPIA)、信号介导RNA扩增技术(SMART)、自我持续序列复制(3SR)、基因组指数扩增反应(GEAR)或等温多置换扩增(IMDA)来扩增所述样品中的所述靶DNA。

[0746] 方面139.根据方面122-138中任一项所述的方法,其中所述样品中的靶DNA以小于10aM的浓度存在。

[0747] 方面140.根据方面122-139中任一项所述的方法,其中所述经标记的检测剂核酸

的标记包括荧光发射染料对。

[0748] 方面141.根据方面140所述的方法,其中所述荧光发射染料对是荧光共振能量转移(FRET)对。

[0749] 方面142.根据方面140所述的方法,其中所述荧光发射染料对是猝灭剂/荧光剂对。

[0750] 方面143.根据方面142-147中任一项所述的方法,其中所述经标记的检测剂核酸包含两个或更多个荧光发射染料对。

[0751] 方面144.根据方面143所述的方法,其中所述两个或更多个荧光发射染料对包括荧光共振能量转移(FRET)对和猝灭剂/荧光剂对。

[0752] 实施例

[0753] 给出以下实施例以便向本领域普通技术人员提供对如何制备和使用本发明的完整公开和描述,并且不旨在限制发明者所视为的他们的发明的范围,它们也不旨在表示以下实验是所进行的所有或仅有实验。已努力确保关于所用数值(例如数量、温度等)的准确性,但应考虑一些实验误差和偏差。除非另有说明,否则份数是重量份,分子量是重量平均分子量,温度是摄氏度,并且压力是大气压或接近大气压。可使用标准缩写,例如,bp,碱基对;kb,千碱基;pl,皮升;s或sec,秒;min,分钟;h或hr,小时;aa,氨基酸;kb,千碱基;bp,碱基对;nt,核苷酸;i.m.,肌内;i.p.,腹膜内;s.c.,皮下;等。

[0754] 实施例1:Casλ1用TTR PAM切割双链DNA

[0755] CRISPR-Cas效应子复合物识别短的原间隔子相邻基序(PAM),所述基序在DNA干扰之前仅存在于外源DNA上,从而防止宿主DNA的CRISPR阵列的自我靶向。为了在实验上测试Casλ是否是RNA引导的DNA靶向核酸内切酶并测试PAM是否是DNA鉴定所必需的,在大肠杆菌中用Casλ家族的Casλ1蛋白同源物建立了PAM耗竭测定。在PAM耗竭测定中,Casλ靶向质粒文库中随机序列附近的DNA序列,所述序列不同于任何天然CasλCRISPR阵列中发现的任何序列。NGS测序显示Casλ和crRNA(如从pBAS18表达的;图11)在细菌中足以耗竭具有与crRNA引导互补的DNA位点的质粒,当显示的PAM序列与原间隔子相邻时(图9),相比之下,Casλ与非靶向对照引导RNA配对(如从pBAS12表达;图10),其与靶质粒没有互补性。结果表明,Casλ可以被重新编程以切割与引导RNA互补的任何DNA序列,其中PAM存在于靶标的5'末端(Casλ1的富含T的PAM),并且不需要额外的RNA组分来形成体内功能效应子。

[0756] 实施例2:可编程的DNA靶向

[0757] 在大肠杆菌中进行flp重组测定,以从含有整合到基因组中的GFP和RFP表达盒的大肠杆菌菌株中消除卡那霉素抗性盒。挑取大肠杆菌ΔKan的单个菌落接种三个5mL(LB)起始培养物,以在第二天制备电感受态细胞。从起始培养物中接种100mL(LB)主培养物,并在37℃、剧烈振荡下生长至OD600为0.6-0.7,然后通过重复冰冷的H20和10%甘油洗涤来制备电感受态细胞。将细胞重悬于10%甘油中,将50μL等分试样在液氮中快速冷冻,并储存于-80℃下。

[0758] 产生的Casλ载体含有密码子优化的Casλ1基因并且从pBAS12亚克隆包含其同源重复元件的引导序列和靶向所得大肠杆菌ΔKan菌株(pBAS41、pBAS42、pBAS43、pBAS44)中的GFP DNA的间隔子选择。也将含有Casλ1和由来自Casλ2的非同源重复单元和GFP靶向间隔子(TAGCATCACCTTCACCCTCTCCACGGACAG;SEQ ID NO:165)引导组成的Casλ载体亚克隆以形成

pBAS40。在微脉冲电穿孔仪(Bio-Rad)上的0.1mm电穿孔比色皿(Bio-Rad)中,通过电穿孔将Casλ载体和具有非靶向引导对照质粒的Casλ载体转化到具有100ng质粒的25μL电感受态细胞中,在1mL恢复培养基(Lucigen)中在37℃振荡一小时来回收细胞。然后制备10倍稀释系列,并将3.5μL各自的稀释液点涂在含有适当抗生素和IPTG诱导剂的LB-琼脂上。将平板在37℃下孵育过夜,并且第二天计数菌落以确定转化效率。

[0759] 为了评估转化效率,由每ng转化质粒的细胞形成单位计算平均值和标准偏差,以一式三份进行电穿孔。

[0760] 数据如图12A-12C中所示。所述实验显示,与非靶向对照相比,使用具有其同源引导的Casλ载体(pBAS44)的GFP大肠杆菌显著减少,这表明双链DNA在靶区域断裂。主要为RFP阳性/GFP阴性的菌落在蓝光下的生长进一步支持了赋予靶向可编程基因组编辑以产生缺乏GFP产生的菌株的能力。使用具有来自单独的Casλ直向同源物(pBAS40)的引导的Casλ载体的生长抑制,观察到的菌落主要表达RFP而不表达GFP,也表明Casλ直向同源物可以使用来自相关CRISPR-Cas系统的引导物在细胞中起到赋予编辑的功能,其中精确消除了GFP产生。这可以进一步扩展到具有整合GFP的HEK293T哺乳动物细胞,这表明在哺乳动物细胞中的活性。即使在未稀释的样品中,在两种情况下生长的大肠杆菌菌落的病态表型也表明核酸(RNA或DNA)的可能的反式切割,这可用于诊断目的,通过提供含有具有CasλRNP的靶核酸和单链DNA荧光团-猝灭剂(ssDNA-FQ)报告分子或RNA荧光团-猝灭剂(ssRNA-FQ)报告分子的样品,在靶核酸存在下产生强的荧光信号,而在靶核酸不存在下产生明显较低的荧光信号。

[0761] pBAS44和pBAS40间隔子:TAGCATCACCTTCACCCTCTCCA CGGACAG(SEQ ID NO:165)。

[0762] 图12A-12C。与具有非靶向引导物的Casλ阴性对照(pBAS12)相比,具有GFP靶向引导物的Casλ(pBAS44)显示出多个数量级的集落形成单位(作为细胞活力的代表)减少。使用来自另一个直向同源物的重复单位的具有GFP靶向引导物的Casλ也显示出生长抑制和GFP消融。

[0763] 实施例3:Casλ纯化

[0764] Casλ的纯化显示蛋白质大小与约70-85kDa的计算预测值一致。将含有His标签的Casλ过表达载体转化到化学感受态大肠杆菌BL21(DE3)-Star(QB3-Macrolab,UC Berkeley)中,并在LB-Kan琼脂平板(50μg/mL卡那霉素)上在37℃下孵育过夜。挑取单菌落以接种50mL(LB,卡那霉素50μg/mL)起始培养物,将其在37℃、剧烈摇动下孵育过夜。第二天,用40mL起始培养物接种2 750mL TB-Kan培养基(50μg/mL卡那霉素),并在37℃下生长至OD600为0.6,在冰上冷却,随后用0.5mM IPTG诱导基因表达,随后在16℃下孵育过夜。

[0765] 通过离心收获细胞,并将其重悬于低盐缓冲液中,并且然后随后通过超声处理来裂解。将可溶级分上样到在洗涤缓冲液中预平衡的5mL Ni-NTA Superflow Cartridge(Qiagen)上。用20倍柱体积(CV)的洗涤缓冲液洗涤结合的蛋白质,随后在5CV洗脱缓冲液(50mM HEPES-Na pH 7.5RT、500mM NaCl、500mM咪唑、5%甘油和0.5mM TCEP)中洗脱。将洗脱的蛋白质浓缩至1mL,然后注射到在尺寸排阻色谱缓冲液(20mM HEPES-Na pH 7.5RT、500mM NaCl、5%甘油和0.5mM TCEP)中预平衡的HiLoad 16/600 Superdex 200pg柱(GE Healthcare)中。将峰级分浓缩至1mL,并使用NanoDrop 8000分光光度计(Thermo Scientific)测定浓度。在4℃恒温下纯化蛋白质,并将浓缩的蛋白质保存在冰上以防止聚

集,在液氮中速冻,并储存在-80℃下。

[0766] 在蛋白质纯化的不同阶段,Casλ的十二烷基硫酸钠-聚丙烯酰胺凝胶电泳(SDS-PAGE)显示蛋白质大小与约70kDa的计算预测值一致。

[0767] 尽管已经参考本发明的特定实施方案描述了本发明,但本领域技术人员应当理解,在不脱离本发明的真实精神和范围的情况下,可进行各种改变并且可用等同物进行替换。此外,可进行许多修改以使特定的情况、材料、物质的组成、过程、一个或多个过程步骤适应本发明的目的、精神和范围。所有这些修改旨在落入所附权利要求的范围内。

序列表

<110> 加利福尼亚大学董事会(The Regents of the University of California)

Banfield, Jillian F.

Doudna, Jennifer A.

Al-Shayeb, Basem

<120> CRISPR-CAS效应多肽及其使用方法

<130> BERK-420WO

<150> 62/952,909

<151> 2019-12-23

<160> 165

<170> PatentIn version 3.5

<210> 1

<211> 36

<212> DNA

<213> 人工序列(Artificial sequence)

<220>

<223> 合成序列(synthetic sequence)

<400> 1

attgttgtaa ctcttatttt gtatggagta aacaac 36

<210> 2

<211> 37

<212> DNA

<213> 人工序列(Artificial sequence)

<220>

<223> 合成序列(synthetic sequence)

<220>

<221> misc_feature

<222> (1)..(1)

<223> N为15至30个核苷酸的序列,其中每个核苷酸由任何核苷酸组成

<400> 2

nattgttgta actcttattt tgtatggagt aaacaac 37

<210> 3

<211> 36

<212> DNA

<213> 人工序列(Artificial sequence)

<220>

<223> 合成序列(synthetic sequence)

<400> 3

attgttgtag acctcttttt ataaggattg aacaac 36

<210> 4

<211> 37

<212> DNA

<213> 人工序列(Artificial sequence)

<220>

<223> 合成序列(synthetic sequence)

<220>

<221> misc_feature

<222> (1)..(1)

<223> N为15至30个核苷酸的序列,其中每个核苷酸由任何核苷酸组成

<400> 4

nattgttgta gacctctttt tataaggatt gaacaac 37

<210> 5

<211> 36

<212> DNA

<213> 人工序列(Artificial sequence)

<220>

<223> 合成序列(synthetic sequence)

<400> 5

attgttgtag atacctttttt gtaaggattg aacaac 36

<210> 6

<211> 37

<212> DNA

<213> 人工序列(Artificial sequence)

<220>

<223> 合成序列(synthetic sequence)

<220>

<221> misc_feature

<222> (1)..(1)

<223> N为15至30个核苷酸的序列,其中每个核苷酸由任何核苷酸组成

<400> 6

nattgttgta gataccttttt tgtaaggatt gaacaac 37

<210> 7

<211> 37

<212> DNA

<213> 人工序列(Artificial sequence)

<220>

<223> 合成序列(synthetic sequence)

<400> 7
tattgttgta gataccttt tgtaaggatt aaacaac 37
<210> 8
<211> 38
<212> DNA
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<220>
<221> misc_feature
<222> (1)..(1)
<223> N为15至30个核苷酸的序列,其中每个核苷酸由任何核苷酸组成
<400> 8
ntattgttgt agataccttt ttgtaaggat taaacaac 38
<210> 9
<211> 39
<212> DNA
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 9
aatgttgtag atgccttttt ataaggatta aacaacttg 39
<210> 10
<211> 40
<212> DNA
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<220>
<221> misc_feature
<222> (1)..(1)
<223> N为15至30个核苷酸的序列,其中每个核苷酸由任何核苷酸组成
<400> 10
naatgttgta gatgcctttt tataaggatt aaacaacttg 40
<210> 11
<211> 36
<212> DNA
<213> 人工序列(Artificial sequence)
<220>

<223> 合成序列(synthetic sequence)
<400> 11
attgttgaaa tagtactttt atagtctata tacaac 36
<210> 12
<211> 37
<212> DNA
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<220>
<221> misc_feature
<222> (1)..(1)
<223> N为15至30个核苷酸的序列,其中每个核苷酸由任何核苷酸组成
<400> 12
nattgttgaa atagtacttt tatagtctat atacaac 37
<210> 13
<211> 37
<212> DNA
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 13
tattgttgta actcttattt tgtatggagt aaacaac 37
<210> 14
<211> 38
<212> DNA
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<220>
<221> misc_feature
<222> (1)..(1)
<223> N为15至30个核苷酸的序列,其中每个核苷酸由任何核苷酸组成
<400> 14
ntattgttgt aactcttatt ttgtatggag taaacaac 38
<210> 15
<211> 36
<212> DNA
<213> 人工序列(Artificial sequence)

<220>
<223> 合成序列(synthetic sequence)
<400> 15
attgttgtaa catctatttt gtaaggtgta aacaac 36
<210> 16
<211> 37
<212> DNA
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<220>
<221> misc_feature
<222> (1)..(1)
<223> N为15至30个核苷酸的序列,其中每个核苷酸由任何核苷酸组成
<400> 16
nattgttgta acatctattt tgtaaggtgt aaacaac 37
<210> 17
<211> 37
<212> DNA
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 17
aattgttgta actcttattt tgtatggagt aaacaac 37
<210> 18
<211> 38
<212> DNA
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<220>
<221> misc_feature
<222> (1)..(1)
<223> N为15至30个核苷酸的序列,其中每个核苷酸由任何核苷酸组成
<400> 18
naattgttgt aactcttatt ttgtatggag taaacaac 38
<210> 19
<211> 36
<212> DNA

<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 19
attgttgtaa cttttatttt gtatggagta aacaac 36
<210> 20
<211> 37
<212> DNA
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<220>
<221> misc_feature
<222> (1)..(1)
<223> N为15至30个核苷酸的序列,其中每个核苷酸由任何核苷酸组成
<400> 20
nattgttgta acttttattt tgtatggagt aaacaac 37
<210> 21
<211> 36
<212> DNA
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 21
aatgttgtag ataccttttt gtaaggattg aacaac 36
<210> 22
<211> 37
<212> DNA
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<220>
<221> misc_feature
<222> (1)..(1)
<223> N为15至30个核苷酸的序列,其中每个核苷酸由任何核苷酸组成
<400> 22
naatgttgta gataccttttt tgtaaggatt gaacaac 37
<210> 23
<211> 36

<212> DNA
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 23
attgttgtaa tactattttt gtaaagtata aacaac 36
<210> 24
<211> 37
<212> DNA
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<220>
<221> misc_feature
<222> (1)..(1)
<223> N为15至30个核苷酸的序列,其中每个核苷酸由任何核苷酸组成
<400> 24
nattgttgta atactatttt tgtaaagtat aaacaac 37
<210> 25
<211> 36
<212> DNA
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 25
aatgttgtag atgccttttt ataaggatta aacaac 36
<210> 26
<211> 37
<212> DNA
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<220>
<221> misc_feature
<222> (1)..(1)
<223> N为15至30个核苷酸的序列,其中每个核苷酸由任何核苷酸组成
<400> 26
naatgttgta gatgcctttt tataaggatt aaacaac 37
<210> 27

<211> 36
<212> DNA
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 27
attgttgtaa tacacttttt ataaggtatg aacaac 36
<210> 28
<211> 37
<212> DNA
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<220>
<221> misc_feature
<222> (1)..(1)
<223> N为15至30个核苷酸的序列,其中每个核苷酸由任何核苷酸组成
<400> 28
nattgttgta atacactttt tataaggtat gaacaac 37
<210> 29
<211> 84
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 29
Met Ala Ser Met Ile Ser Ser Ser Ala Val Thr Thr Val Ser Arg Ala
1 5 10 15
Ser Arg Gly Gln Ser Ala Ala Met Ala Pro Phe Gly Gly Leu Lys Ser
20 25 30
Met Thr Gly Phe Pro Val Arg Lys Val Asn Thr Asp Ile Thr Ser Ile
35 40 45
Thr Ser Asn Gly Gly Arg Val Lys Cys Met Gln Val Trp Pro Pro Ile
50 55 60
Gly Lys Lys Lys Phe Glu Thr Leu Ser Tyr Leu Pro Pro Leu Thr Arg
65 70 75 80
Asp Ser Arg Ala
<210> 30
<211> 57

<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 30

Met Ala Ser Met Ile Ser Ser Ser Ala Val Thr Thr Val Ser Arg Ala
1 5 10 15
Ser Arg Gly Gln Ser Ala Ala Met Ala Pro Phe Gly Gly Leu Lys Ser
20 25 30
Met Thr Gly Phe Pro Val Arg Lys Val Asn Thr Asp Ile Thr Ser Ile
35 40 45
Thr Ser Asn Gly Gly Arg Val Lys Ser
50 55

<210> 31
<211> 85
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 31

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

<210> 32
<211> 76
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 32

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

<210> 33
<211> 76
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 33

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

<210> 34
<211> 72
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 34

Met Ala Gln Ile Asn Asn Met Ala Gln Gly Ile Gln Thr Leu Asn Pro
1 5 10 15
Asn Ser Asn Phe His Lys Pro Gln Val Pro Lys Ser Ser Ser Phe Leu
20 25 30
Val Phe Gly Ser Lys Lys Leu Lys Asn Ser Ala Asn Ser Met Leu Val

35 40 45
Leu Lys Lys Asp Ser Ile Phe Met Gln Leu Phe Cys Ser Phe Arg Ile
50 55 60
Ser Ala Ser Val Ala Thr Ala Cys
65 70
<210> 35
<211> 69
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 35
Met Ala Ala Leu Val Thr Ser Gln Leu Ala Thr Ser Gly Thr Val Leu
1 5 10 15
Ser Val Thr Asp Arg Phe Arg Arg Pro Gly Phe Gln Gly Leu Arg Pro
20 25 30
Arg Asn Pro Ala Asp Ala Ala Leu Gly Met Arg Thr Val Gly Ala Ser
35 40 45
Ala Ala Pro Lys Gln Ser Arg Lys Pro His Arg Phe Asp Arg Arg Cys
50 55 60
Leu Ser Met Val Val
65
<210> 36
<211> 77
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 36
Met Ala Ala Leu Thr Thr Ser Gln Leu Ala Thr Ser Ala Thr Gly Phe
1 5 10 15
Gly Ile Ala Asp Arg Ser Ala Pro Ser Ser Leu Leu Arg His Gly Phe
20 25 30
Gln Gly Leu Lys Pro Arg Ser Pro Ala Gly Gly Asp Ala Thr Ser Leu
35 40 45
Ser Val Thr Thr Ser Ala Arg Ala Thr Pro Lys Gln Gln Arg Ser Val
50 55 60
Gln Arg Gly Ser Arg Arg Phe Pro Ser Val Val Val Cys
65 70 75

<210> 37
<211> 57
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 37

```
Met Ala Ser Ser Val Leu Ser Ser Ala Ala Val Ala Thr Arg Ser Asn
1               5                  10                  15
Val Ala Gln Ala Asn Met Val Ala Pro Phe Thr Gly Leu Lys Ser Ala
            20                  25                  30
Ala Ser Phe Pro Val Ser Arg Lys Gln Asn Leu Asp Ile Thr Ser Ile
        35                  40                  45
Ala Ser Asn Gly Gly Arg Val Gln Cys
    50                  55
```

<210> 38
<211> 65
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 38

```
Met Glu Ser Leu Ala Ala Thr Ser Val Phe Ala Pro Ser Arg Val Ala
1               5                  10                  15
Val Pro Ala Ala Arg Ala Leu Val Arg Ala Gly Thr Val Val Pro Thr
            20                  25                  30
Arg Arg Thr Ser Ser Thr Ser Gly Thr Ser Gly Val Lys Cys Ser Ala
        35                  40                  45
Ala Val Thr Pro Gln Ala Ser Pro Val Ile Ser Arg Ser Ala Ala Ala
    50                  55                  60
Ala
65
```

<210> 39
<211> 72
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 39

Met Gly Ala Ala Ala Thr Ser Met Gln Ser Leu Lys Phe Ser Asn Arg
1 5 10 15
Leu Val Pro Pro Ser Arg Arg Leu Ser Pro Val Pro Asn Asn Val Thr
20 25 30
Cys Asn Asn Leu Pro Lys Ser Ala Ala Pro Val Arg Thr Val Lys Cys
35 40 45
Cys Ala Ser Ser Trp Asn Ser Thr Ile Asn Gly Ala Ala Ala Thr Thr
50 55 60
Asn Gly Ala Ser Ala Ala Ser Ser
65 70

<210> 40
<211> 11
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 40
Tyr Gly Arg Lys Lys Arg Arg Gln Arg Arg Arg
1 5 10

<210> 41
<211> 12
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 41
Arg Arg Gln Arg Arg Thr Ser Lys Leu Met Lys Arg
1 5 10

<210> 42
<211> 27
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 42
Gly Trp Thr Leu Asn Ser Ala Gly Tyr Leu Leu Gly Lys Ile Asn Leu
1 5 10 15
Lys Ala Leu Ala Ala Leu Ala Lys Lys Ile Leu
20 25

<210> 43
<211> 33
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 43
Lys Ala Leu Ala Trp Glu Ala Lys Leu Ala Lys Ala Leu Ala Lys Ala
1 5 10 15
Leu Ala Lys His Leu Ala Lys Ala Leu Ala Lys Ala Leu Lys Cys Glu
20 25 30
Ala
<210> 44
<211> 16
<212> PRT
<213> 黑腹果蝇(Drosophila melanogaster)
<400> 44
Arg Gln Ile Lys Ile Trp Phe Gln Asn Arg Arg Met Lys Trp Lys Lys
1 5 10 15
<210> 45
<211> 9
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 45
Arg Lys Lys Arg Arg Gln Arg Arg Arg
1 5
<210> 46
<211> 8
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 46
Arg Lys Lys Arg Arg Gln Arg Arg
1 5
<210> 47
<211> 11

<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 47
Tyr Ala Arg Ala Ala Ala Arg Gln Ala Arg Ala
1 5 10
<210> 48
<211> 11
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 48
Thr His Arg Leu Pro Arg Arg Arg Arg Arg Arg
1 5 10
<210> 49
<211> 11
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 49
Gly Gly Arg Arg Ala Arg Arg Arg Arg Arg Arg
1 5 10
<210> 50
<211> 5
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 50
Gly Ser Gly Gly Ser
1 5
<210> 51
<211> 6
<212> PRT
<213> 人工序列(Artificial sequence)
<220>

<223> 合成序列(synthetic sequence)
<400> 51
Gly Gly Ser Gly Gly Ser
1 5
<210> 52
<211> 4
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 52
Gly Gly Gly Ser
1
<210> 53
<211> 4
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 53
Gly Gly Ser Gly
1
<210> 54
<211> 5
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 54
Gly Gly Ser Gly Gly
1 5
<210> 55
<211> 5
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 55
Gly Ser Gly Ser Gly

1 5

<210> 56

<211> 5

<212> PRT

<213> 人工序列(Artificial sequence)

<220>

<223> 合成序列(synthetic sequence)

<400> 56

Gly Ser Gly Gly Gly

1 5

<210> 57

<211> 5

<212> PRT

<213> 人工序列(Artificial sequence)

<220>

<223> 合成序列(synthetic sequence)

<400> 57

Gly Gly Gly Ser Gly

1 5

<210> 58

<211> 5

<212> PRT

<213> 人工序列(Artificial sequence)

<220>

<223> 合成序列(synthetic sequence)

<400> 58

Gly Ser Ser Ser Gly

1 5

<210> 59

<211> 36

<212> RNA

<213> 人工序列(Artificial sequence)

<220>

<223> 合成序列(synthetic sequence)

<400> 59

auuguuguaa cucuuauuuu guauggagua aacaac 36

<210> 60

<211> 36

<212> RNA

<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 60
auuguuguag accucuuuuu auaaggauug aacaac 36
<210> 61
<211> 37
<212> RNA
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 61
uauuguugua gauaccuuuu uguaaggauu aaacaac 37
<210> 62
<211> 39
<212> RNA
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 62
aauguuguag augccuuuuu auaaggauua aacaacuug 39
<210> 63
<211> 36
<212> RNA
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 63
auuguugaaa uaguacuuuu auagucuaua uacaac 36
<210> 64
<211> 37
<212> RNA
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 64
uauuguugua acucuuauuu uguauggagu aaacaac 37
<210> 65
<211> 36

<212> RNA
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 65
auuguuguaa cuuuuauuuu guauggagua aacaac 36
<210> 66
<211> 36
<212> RNA
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 66
aauguuguag auaccuuuuu guaaggauug aacaac 36
<210> 67
<211> 36
<212> RNA
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 67
auuguuguaa uacuauuuuu guaaaguaua aacaac 36
<210> 68
<211> 36
<212> RNA
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 68
aauguuguag augccuuuuu auaaggauua aacaac 36
<210> 69
<211> 36
<212> RNA
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 69
auuguuguaa uacacuuuuu auaagguaug aacaac 36
<210> 70

<211> 36
<212> RNA
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 70
auuguuguaa caucuauuuu guaaggugua aacaac 36
<210> 71
<211> 71
<212> RNA
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<220>
<221> misc_feature
<222> (1)..(35)
<223> n为任何核苷酸
<400> 71
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnauugu uguaacucuu auuuuguaug 60
gaguaaacaa c 71
<210> 72
<211> 20
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<220>
<221> MISC_FEATURE
<222> (4)..(4)
<223> X选自赖氨酸、组氨酸和精氨酸
<220>
<221> MISC_FEATURE
<222> (8)..(8)
<223> X选自赖氨酸、组氨酸和精氨酸
<220>
<221> MISC_FEATURE
<222> (11)..(11)
<223> X选自赖氨酸、组氨酸和精氨酸
<220>

<221> MISC_FEATURE
<222> (15)..(15)
<223> X选自赖氨酸、组氨酸和精氨酸
<220>
<221> MISC_FEATURE
<222> (19)..(19)
<223> X选自赖氨酸、组氨酸和精氨酸
<400> 72

```
Gly Leu Phe Xaa Ala Leu Leu Xaa Leu Leu Xaa Ser Leu Trp Xaa Leu
1               5                  10                  15
Leu Leu Xaa Ala
            20
```

<210> 73
<211> 20
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 73

```
Gly Leu Phe His Ala Leu Leu His Leu Leu His Ser Leu Trp His Leu
1               5                  10                  15
Leu Leu His Ala
            20
```

<210> 74
<211> 167
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 74

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

65 70 75 80
Tyr Val Thr Leu Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His
85 90 95
Ser Arg Ile Gly Arg Val Val Phe Gly Ala Arg Asp Ala Lys Thr Gly
100 105 110
Ala Ala Gly Ser Leu Met Asp Val Leu His His Pro Gly Met Asn His
115 120 125
Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu
130 135 140
Leu Ser Asp Phe Phe Arg Met Arg Arg Gln Glu Ile Lys Ala Gln Lys
145 150 155 160
Lys Ala Gln Ser Ser Thr Asp
165

<210> 75
<211> 178
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 75
Met Arg Arg Ala Phe Ile Thr Gly Val Phe Phe Leu Ser Glu Val Glu
1 5 10 15
Phe Ser His Glu Tyr Trp Met Arg His Ala Leu Thr Leu Ala Lys Arg
20 25 30
Ala Trp Asp Glu Arg Glu Val Pro Val Gly Ala Val Leu Val His Asn
35 40 45
Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Pro Ile Gly Arg His Asp
50 55 60
Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val
65 70 75 80
Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu Tyr Val Thr Leu Glu
85 90 95
Pro Cys Val Met Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg
100 105 110
Val Val Phe Gly Ala Arg Asp Ala Lys Thr Gly Ala Ala Gly Ser Leu
115 120 125
Met Asp Val Leu His His Pro Gly Met Asn His Arg Val Glu Ile Thr
130 135 140
Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu Leu Ser Asp Phe Phe

145 150 155 160
Arg Met Arg Arg Gln Glu Ile Lys Ala Gln Lys Lys Ala Gln Ser Ser
165 170 175
Thr Asp

<210> 76
<211> 160
<212> PRT
<213> 金黄色葡萄球菌(Staphylococcus aureus)
<400> 76

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

<210> 77
<211> 161
<212> PRT
<213> 枯草杆菌(Bacillus subtilis)
<400> 77

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

35 40 45

Arg Ser Ile Ala His Ala Glu Met Leu Val Ile Asp Glu Ala Cys Lys
50 55 60
Ala Leu Gly Thr Trp Arg Leu Glu Gly Ala Thr Leu Tyr Val Thr Leu
65 70 75 80
Glu Pro Cys Pro Met Cys Ala Gly Ala Val Val Leu Ser Arg Val Glu
85 90 95
Lys Val Val Phe Gly Ala Phe Asp Pro Lys Gly Gly Cys Ser Gly Thr
100 105 110
Leu Met Asn Leu Leu Gln Glu Glu Arg Phe Asn His Gln Ala Glu Val
115 120 125
Val Ser Gly Val Leu Glu Glu Glu Cys Gly Gly Met Leu Ser Ala Phe
130 135 140
Phe Arg Glu Leu Arg Lys Lys Lys Lys Ala Ala Arg Lys Asn Leu Ser
145 150 155 160
Glu

<210> 78
<211> 183
<212> PRT
<213> 鼠伤寒沙门菌(Salmonella typhimurium)
<400> 78

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

Glu Gly Val Leu Arg Asp Glu Cys Ala Thr Leu Leu Ser Asp Phe Phe
145 150 155 160
Arg Met Arg Arg Gln Glu Ile Lys Ala Leu Lys Lys Ala Asp Arg Ala
165 170 175
Glu Gly Ala Gly Pro Ala Val
180

<210> 79
<211> 164
<212> PRT
<213> 腐败希瓦菌(Shewanella putrefaciens)
<400> 79

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

<210> 80
<211> 173
<212> PRT
<213> 流感嗜血杆菌(Haemophilus influenzae)
<400> 80

Met Asp Ala Ala Lys Val Arg Ser Glu Phe Asp Glu Lys Met Met Arg
1 5 10 15

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

<210> 81
<211> 161
<212> PRT
<213> 新月柄杆菌(Caulobacter crescentus)
<400> 81

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

Pro Lys Gly Gly Ala Val Val His Gly Pro Lys Phe Phe Ala Gln Pro
115 120 125
Thr Cys His Trp Arg Pro Glu Val Thr Gly Gly Val Leu Ala Asp Glu
130 135 140
Ser Ala Asp Leu Leu Arg Gly Phe Phe Arg Ala Arg Arg Lys Ala Lys
145 150 155 160
Ile

<210> 82
<211> 179
<212> PRT
<213> 硫还原地杆菌(Geobacter sulfurreducens)
<400> 82

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

<210> 83
<211> 198
<212> PRT
<213> 人工序列(Artificial sequence)

<220>
<223> 合成序列(synthetic sequence)
<400> 83

Met Asp Ser Leu Leu Met Asn Arg Arg Lys Phe Leu Tyr Gln Phe Lys
1 5 10 15
Asn Val Arg Trp Ala Lys Gly Arg Arg Glu Thr Tyr Leu Cys Tyr Val
20 25 30
Val Lys Arg Arg Asp Ser Ala Thr Ser Phe Ser Leu Asp Phe Gly Tyr
35 40 45
Leu Arg Asn Lys Asn Gly Cys His Val Glu Leu Leu Phe Leu Arg Tyr
50 55 60
Ile Ser Asp Trp Asp Leu Asp Pro Gly Arg Cys Tyr Arg Val Thr Trp
65 70 75 80
Phe Thr Ser Trp Ser Pro Cys Tyr Asp Cys Ala Arg His Val Ala Asp
85 90 95
Phe Leu Arg Gly Asn Pro Asn Leu Ser Leu Arg Ile Phe Thr Ala Arg
100 105 110
Leu Tyr Phe Cys Glu Asp Arg Lys Ala Glu Pro Glu Gly Leu Arg Arg
115 120 125
Leu His Arg Ala Gly Val Gln Ile Ala Ile Met Thr Phe Lys Asp Tyr
130 135 140
Phe Tyr Cys Trp Asn Thr Phe Val Glu Asn His Glu Arg Thr Phe Lys
145 150 155 160
Ala Trp Glu Gly Leu His Glu Asn Ser Val Arg Leu Ser Arg Gln Leu
165 170 175
Arg Arg Ile Leu Leu Pro Leu Tyr Glu Val Asp Asp Leu Arg Asp Ala
180 185 190
Phe Arg Thr Leu Gly Leu
195

<210> 84
<211> 178
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 84

Phe Leu Tyr Gln Phe Lys Asn Val Arg Trp Ala Lys Gly Arg Arg Glu
1 5 10 15
Thr Tyr Leu Cys Tyr Val Val Lys Arg Arg Asp Ser Ala Thr Ser Phe

20 25 30

Ser Leu Asp Phe Gly Tyr Leu Arg Asn Lys Asn Gly Cys His Val Glu
35 40 45

Leu Leu Phe Leu Arg Tyr Ile Ser Asp Trp Asp Leu Asp Pro Gly Arg
50 55 60

Cys Tyr Arg Val Thr Trp Phe Thr Ser Trp Ser Pro Cys Tyr Asp Cys
65 70 75 80

Ala Arg His Val Ala Asp Phe Leu Arg Gly Asn Pro Asn Leu Ser Leu
85 90 95

Arg Ile Phe Thr Ala Arg Leu Tyr Phe Cys Glu Asp Arg Lys Ala Glu
100 105 110

Pro Glu Gly Leu Arg Arg Leu His Arg Ala Gly Val Gln Ile Ala Ile
115 120 125

Met Thr Phe Lys Glu Asn His Glu Arg Thr Phe Lys Ala Trp Glu Gly
130 135 140

Leu His Glu Asn Ser Val Arg Leu Ser Arg Gln Leu Arg Arg Ile Leu
145 150 155 160

Leu Pro Leu Tyr Glu Val Asp Asp Leu Arg Asp Ala Phe Arg Thr Leu
165 170 175

Gly Leu

<210> 85

<211> 188

<212> PRT

<213> 人工序列(Artificial sequence)

<220>

<223> 合成序列(synthetic sequence)

<400> 85

Phe Leu Tyr Gln Phe Lys Asn Val Arg Trp Ala Lys Gly Arg Arg Glu
1 5 10 15

Thr Tyr Leu Cys Tyr Val Val Lys Arg Arg Asp Ser Ala Thr Ser Phe
20 25 30

Ser Leu Asp Phe Gly Tyr Leu Arg Asn Lys Asn Gly Cys His Val Glu
35 40 45

Leu Leu Phe Leu Arg Tyr Ile Ser Asp Trp Asp Leu Asp Pro Gly Arg
50 55 60

Cys Tyr Arg Val Thr Trp Phe Thr Ser Trp Ser Pro Cys Tyr Asp Cys
65 70 75 80

Ala Arg His Val Ala Asp Phe Leu Arg Gly Asn Pro Asn Leu Ser Leu
85 90 95

Arg Ile Phe Thr Ala Arg Leu Tyr Phe Cys Glu Asp Arg Lys Ala Glu
100 105 110
Pro Glu Gly Leu Arg Arg Leu His Arg Ala Gly Val Gln Ile Ala Ile
115 120 125
Met Thr Phe Lys Asp Tyr Phe Tyr Cys Trp Asn Thr Phe Val Glu Asn
130 135 140
His Glu Arg Thr Phe Lys Ala Trp Glu Gly Leu His Glu Asn Ser Val
145 150 155 160
Arg Leu Ser Arg Gln Leu Arg Arg Ile Leu Leu Pro Leu Tyr Glu Val
165 170 175
Asp Asp Leu Arg Asp Ala Phe Arg Thr Leu Gly Leu
180 185

<210> 86
<211> 7
<212> PRT
<213> SV40病毒(SV40 virus)
<400> 86
Pro Lys Lys Lys Arg Lys Val
1 5

<210> 87
<211> 16
<212> PRT
<213> 未知(Unknown)
<220>
<223> 核质蛋白二分NLS
<400> 87
Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys Lys Lys Lys
1 5 10 15

<210> 88
<211> 9
<212> PRT
<213> 未知(Unknown)
<220>
<223> c-myc NLS
<400> 88
Pro Ala Ala Lys Arg Val Lys Leu Asp
1 5

<210> 89
<211> 11

<212> PRT
<213> 未知(Unknown)
<220>
<223> c-myc NLS
<400> 89
Arg Gln Arg Arg Asn Glu Leu Lys Arg Ser Pro
1 5 10
<210> 90
<211> 38
<212> PRT
<213> 未知(Unknown)
<220>
<223> hRNPA1 M9 NLS
<400> 90
Asn Gln Ser Ser Asn Phe Gly Pro Met Lys Gly Gly Asn Phe Gly Gly
1 5 10 15
Arg Ser Ser Gly Pro Tyr Gly Gly Gly Gly Gln Tyr Phe Ala Lys Pro
20 25 30
Arg Asn Gln Gly Gly Tyr
35
<210> 91
<211> 42
<212> PRT
<213> 未知(Unknown)
<220>
<223> 来自输入蛋白-α的IBB结构域
<400> 91
Arg Met Arg Ile Glx Phe Lys Asn Lys Gly Lys Asp Thr Ala Glu Leu
1 5 10 15
Arg Arg Arg Arg Val Glu Val Ser Val Glu Leu Arg Lys Ala Lys Lys
20 25 30
Asp Glu Gln Ile Leu Lys Arg Arg Asn Val
35 40
<210> 92
<211> 8
<212> PRT
<213> 未知(Unknown)
<220>
<223> 肌瘤T蛋白

<400> 92
Val Ser Arg Lys Arg Pro Arg Pro
1 5
<210> 93
<211> 8
<212> PRT
<213> 未知(Unknown)
<220>
<223> 肌瘤T蛋白
<400> 93
Pro Pro Lys Lys Ala Arg Glu Asp
1 5
<210> 94
<211> 8
<212> PRT
<213> 智人(Homo sapiens)
<400> 94
Pro Gln Pro Lys Lys Lys Pro Leu
1 5
<210> 95
<211> 12
<212> PRT
<213> 小家鼠(Mus musculus)
<400> 95
Ser Ala Leu Ile Lys Lys Lys Lys Lys Met Ala Pro
1 5 10
<210> 96
<211> 5
<212> PRT
<213> 流感病毒(Influenza virus)
<400> 96
Asp Arg Leu Arg Arg
1 5
<210> 97
<211> 7
<212> PRT
<213> 流感病毒(Influenza virus)
<400> 97
Pro Lys Gln Lys Lys Arg Lys

1 5

<210> 98

<211> 10

<212> PRT

<213> 肝炎病毒(Hepatitus virus)

<400> 98

Arg Lys Leu Lys Lys Lys Ile Lys Lys Leu

1 5 10

<210> 99

<211> 10

<212> PRT

<213> 小家鼠(Mus musculus)

<400> 99

Arg Glu Lys Lys Lys Phe Leu Lys Arg Arg

1 5 10

<210> 100

<211> 20

<212> PRT

<213> 智人(Homo sapiens)

<400> 100

Lys Arg Lys Gly Asp Glu Val Asp Gly Val Asp Glu Val Ala Lys Lys

1 5 10 15

Lys Ser Lys Lys

20

<210> 101

<211> 17

<212> PRT

<213> 智人(Homo sapiens)

<400> 101

Arg Lys Cys Leu Gln Ala Gly Met Asn Leu Glu Ala Arg Lys Thr Lys

1 5 10 15

Lys

<210> 102

<211> 746

<212> PRT

<213> 人工序列(Artificial sequence)

<220>

<223> 合成序列(synthetic sequence)

<400> 102

Met Ala Ser His Lys Lys Thr Glu Ser Asn Gln Ile Ile Lys Thr Phe
1 5 10 15
Ser Phe Lys Ile Lys Asn Ala Asn Gly Leu Ser Leu Asp Val Leu Asn
20 25 30
Asp Ala Ile Thr Glu Tyr Gln Asn Tyr Tyr Asn Ile Cys Ser Asp Trp
35 40 45
Ile Lys Asp His Leu Thr Met Lys Ile Ser Glu Leu Tyr Lys Tyr Ile
50 55 60
Pro Asn Glu Lys Lys Asn Ser Gly Tyr Ala Leu Thr Leu Ile Ser Asp
65 70 75 80
Glu Trp Lys Asp Lys Pro Met Tyr Met Met Phe Lys Lys Gly Tyr Pro
85 90 95
Ala Asn Asn Arg Asp Asn Ala Ile Tyr Glu Thr Leu Asn Thr Cys Asn
100 105 110
Thr Glu His Tyr Thr Gly Asn Ile Leu Asn Phe Ser Asp Thr Tyr Tyr
115 120 125
Arg Arg Phe Gly Tyr Val Ala Ser Ala Ile Ser Asn Tyr Val Thr Lys
130 135 140
Ile Ser Lys Met Ser Thr Gly Ser Arg Ser Lys Asn Ile Ser Asn Asp
145 150 155 160
Ser Asp Val Asp Thr Ile Met Glu Gln Val Ile Tyr Glu Met Glu His
165 170 175
Asn Gly Trp Thr Ser Val Lys Asp Trp Glu Asn Gln Met Glu Tyr Leu
180 185 190
Glu Ser Lys Thr Asp Ser Asn Pro Asn Phe Val Tyr Arg Met Thr Thr
195 200 205
Leu Tyr Glu Phe Tyr Lys Ser His Ile Asp Glu Val Asn Ser Lys Met
210 215 220
Glu Thr Met Ser Ile Asp Ser Leu Ile Lys Phe Gly Gly Cys Arg Arg
225 230 235 240
Lys Asp Ser Lys Lys Ser Met Tyr Ile Met Gly Gly Ser Asn Thr Pro
245 250 255
Phe Asp Ile Thr Gln Ile Gly Gly Asn Ser Leu Asn Ile Lys Phe Ser
260 265 270
Lys Asn Leu Asn Val Asp Val Phe Gly Arg Tyr Asp Val Ile Lys Asp
275 280 285
Asn Thr Leu Leu Val Asp Ile Ile Asn Gly His Gly Ala Ser Phe Val
290 295 300
Leu Lys Ile Ile Asn Asp Glu Ile Tyr Ile Asp Ile Asn Val Ser Val

305 310 315 320
Pro Phe Asp Lys Lys Ile Ala Thr Thr Asn Lys Val Val Gly Ile Asp
325 330 335
Val Asn Ile Lys His Met Leu Leu Ala Thr Asn Ile Leu Asp Asp Gly
340 345 350
Asn Val Lys Gly Tyr Val Asn Ile Tyr Lys Glu Val Ile Asn Asp Ser
355 360 365
Asp Phe Lys Lys Val Cys Asn Ser Thr Val Met Gln Tyr Phe Thr Asp
370 375 380
Phe Ser Lys Phe Val Thr Phe Cys Pro Leu Glu Phe Asp Phe Leu Phe
385 390 395 400
Ser Arg Val Cys Asn Gln Lys Gly Ile Tyr Asn Asp Asn Ser Ala Met
405 410 415
Glu Lys Ser Phe Ser Asp Val Leu Asn Lys Leu Lys Trp Asn Phe Ile
420 425 430
Glu Thr Gly Asp Asn Thr Lys Arg Ile Tyr Ile Glu Asn Val Met Lys
435 440 445
Leu Arg Ser Gln Met Lys Ala Tyr Ala Ile Val Lys Asn Ala Tyr Tyr
450 455 460
Lys Gln Gln Ser Glu Tyr Asp Phe Gly Lys Ser Glu Glu Phe Ile Gln
465 470 475 480
Glu His Pro Phe Ser Asn Thr Asp Lys Gly Ile Glu Ile Leu Asn Lys
485 490 495
Leu Asp Asn Ile Ser Lys Lys Ile Leu Gly Cys Arg Asn Asn Ile Ile
500 505 510
Gln Tyr Ser Tyr Asn Leu Phe Glu Ile Asn Gly Tyr Asp Met Val Ser
515 520 525
Leu Glu Lys Leu Thr Ser Ser Gln Phe Lys Lys Lys Pro Phe Pro Thr
530 535 540
Val Asn Ser Leu Leu Lys Tyr His Lys Ile Leu Gly Cys Thr Gln Glu
545 550 555 560
Glu Met Glu Lys Lys Asp Ile Tyr Ser Val Ile Lys Lys Gly Tyr Tyr
565 570 575
Asp Ile Ile Phe Asp Asn Asp Val Val Thr Asp Ala Lys Leu Ser Ala
580 585 590
Lys Gly Glu Leu Ser Lys Phe Lys Asp Asp Phe Phe Asn Leu Met Ile
595 600 605
Lys Ser Ile His Phe Ala Asp Ile Lys Asp Tyr Phe Ile Thr Leu Ser
610 615 620

Asn Asn Gly Thr Ala Gly Val Ser Leu Val Pro Ser Tyr Phe Thr Ser
625 630 635 640
Gln Met Asp Ser Ile Asp His Lys Ile Tyr Phe Val Gln Asp Asn Lys
645 650 655
Ser Gly Lys Leu Lys Leu Ala Asn Lys His Lys Val Arg Ser Ser Gln
660 665 670
Glu Lys His Ile Asn Gly Leu Asn Ala Asp Tyr Asn Ala Ala Arg Asn
675 680 685
Ile Ala Tyr Ile Met Glu Asn Thr Asp Cys Arg Asn Met Phe Met Lys
690 695 700
Gln Ser Arg Thr Asp Lys Ser Leu Tyr Asn Lys Pro Ser Tyr Glu Thr
705 710 715 720
Phe Ile Lys Thr Gln Gly Ser Ala Val Ala Lys Leu Lys Lys Glu Gly
725 730 735
Phe Val Lys Ile Leu Asp Glu Ala Ser Val
740 745

<210> 103
<211> 735
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 103
Met Ala His Lys Lys Asn Val Gly Ala Glu Ile Val Lys Thr Tyr Ser
1 5 10 15
Phe Lys Val Lys Asn Thr Asn Gly Ile Thr Met Glu Lys Leu Met Asn
20 25 30
Ala Ile Asp Glu Phe Gln Ser Tyr Tyr Asn Leu Cys Ser Asp Trp Ile
35 40 45
Cys Lys Asn Leu Thr Thr Met Thr Ile Gly Asp Leu Asp Gln Tyr Ile
50 55 60
Pro Glu Lys Ala Lys Gly Asn Thr Tyr Ala Thr Val Leu Leu Asp Glu
65 70 75 80
Ala Trp Lys Asn Gln Pro Leu Tyr Lys Ile Phe Gly Lys Lys Tyr Ser
85 90 95
Ser Asn Asn Arg Asn Asn Ala Leu Tyr Cys Ala Leu Ser Ser Val Ile
100 105 110
Asp Met Thr Lys Glu Asn Val Leu Gly Phe Ser Lys Thr His Tyr Ile
115 120 125

Arg Asn Asp Tyr Ile Leu Asn Val Ile Ser Asn Tyr Ala Ser Lys Leu
130 135 140
Ser Lys Leu Asn Thr Gly Val Lys Ser Arg Ala Ile Lys Glu Thr Ser
145 150 155 160
Asp Glu Ala Thr Ile Ile Glu Gln Val Ile Tyr Glu Met Glu His Asn
165 170 175
Lys Trp Glu Ser Ile Glu Asp Trp Lys Asn Gln Ile Glu Tyr Leu Asn
180 185 190
Ser Lys Thr Asp Tyr Asn Pro Thr Tyr Met Glu Arg Met Lys Thr Leu
195 200 205
Ser Ala Tyr Tyr Ser Thr His Lys Ser Glu Val Asp Ala Lys Met Gln
210 215 220
Glu Met Ala Val Glu Asn Leu Val Lys Phe Gly Gly Cys Arg Arg Asn
225 230 235 240
Asn Ser Lys Lys Ser Met Phe Ile Met Gly Ser Asn Thr Thr Asn Tyr
245 250 255
Thr Ile Ser Tyr Ile Gly Gly Asn Ser Phe Asn Ile Asn Phe Ala Asn
260 265 270
Ile Leu Asn Phe Asp Val Tyr Gly Arg Arg Asp Val Val Lys Asn Gly
275 280 285
Glu Val Leu Val Asp Ile Met Ala Asn His Gly Asp Ser Ile Val Leu
290 295 300
Lys Ile Val Asn Gly Glu Leu Tyr Ala Asp Val Pro Cys Ser Val Thr
305 310 315 320
Leu Asn Lys Val Glu Ser Asn Phe Asp Lys Val Val Gly Ile Asp Val
325 330 335
Asn Met Lys His Met Leu Leu Ser Thr Ser Ile Thr Asp Asn Gly Ser
340 345 350
Ser Asp Phe Leu Asn Ile Tyr Lys Glu Met Ser Asn Asn Ala Glu Phe
355 360 365
Met Ala Leu Cys Pro Glu Glu Asp Arg Lys Tyr Tyr Lys Asp Ile Ser
370 375 380
Lys Tyr Val Thr Phe Ala Pro Leu Glu Leu Asp Leu Leu Phe Ser Arg
385 390 395 400
Ile Ser Lys Gln Gly Lys Val Lys Met Glu Lys Val Tyr Ser Glu Ile
405 410 415
Leu Glu Ala Leu Lys Trp Lys Phe Phe Ala Asn Gly Asp Asn Lys Asn
420 425 430
Arg Ile Tyr Val Glu Ser Ile Gln Lys Ile Arg Gln Gln Ile Lys Ala

435 440 445

Leu Cys Val Ile Lys Asn Ala Tyr Tyr Glu Gln Gln Ser Ala Tyr Asp
450 455 460

Ile Asp Lys Thr Gln Glu Tyr Ile Glu Thr His Pro Phe Ser Leu Thr
465 470 475 480

Glu Lys Gly Met Ser Ile Lys Ser Lys Met Asp Lys Ile Cys Gln Thr
485 490 495

Ile Ile Gly Cys Arg Asn Asn Ile Ile Asp Tyr Ala Tyr Ser Phe Phe
500 505 510

Glu Arg Asn Gly Tyr Ser Ile Ile Gly Leu Glu Lys Leu Thr Ser Ser
515 520 525

Gln Phe Glu Lys Thr Lys Ser Met Pro Thr Cys Lys Ser Leu Leu Asn
530 535 540

Phe His Lys Val Leu Gly His Thr Leu Ser Glu Leu Glu Thr Leu Pro
545 550 555 560

Ile Asn Asp Val Val Lys Lys Gly Tyr Tyr Thr Phe Thr Thr Asp Asn
565 570 575

Glu Gly Lys Ile Thr Asp Ala Ser Leu Ser Glu Lys Gly Lys Val Arg
580 585 590

Lys Met Lys Asp Asp Phe Phe Asn Gln Ala Ile Lys Ala Ile His Phe
595 600 605

Ala Asp Val Lys Asp Tyr Phe Ala Thr Leu Ser Asn Asn Gly Gln Thr
610 615 620

Gly Ile Phe Phe Val Pro Ser Gln Phe Thr Ser Gln Met Asp Ser Asn
625 630 635 640

Thr His Asn Leu Tyr Phe Glu Asn Ala Lys Asn Gly Gly Leu Lys Leu
645 650 655

Ala Pro Lys Tyr Lys Val Arg Gln Thr Gln Glu Tyr His Leu Asn Gly
660 665 670

Leu Pro Ala Asp Tyr Asn Ala Ala Arg Asn Ile Ala Tyr Ile Gly Leu
675 680 685

Asp Glu Thr Met Arg Asn Thr Phe Leu Lys Lys Ala Asn Ser Asn Lys
690 695 700

Ser Leu Tyr Asn Gln Pro Ile Tyr Asp Thr Gly Ile Lys Lys Thr Ala
705 710 715 720

Gly Val Phe Ser Arg Met Lys Lys Leu Lys Arg Tyr Glu Ile Ile
725 730 735

<210> 104

<211> 778

<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 104

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

Lys Asp Ser Lys Lys Ser Met Tyr Ile Met Gly Gly Ser Asn Thr Pro
275 280 285
Phe Asp Ile Thr Gln Ile Gly Gly Asn Ser Leu Asn Ile Lys Phe Ser
290 295 300
Lys Asn Leu Asn Val Asp Val Phe Gly Arg Tyr Asp Val Ile Lys Asp
305 310 315 320
Asn Thr Leu Leu Val Asp Ile Ile Asn Gly His Gly Ala Ser Phe Val
325 330 335
Leu Lys Ile Ile Asn Gly Glu Ile Tyr Ile Asp Ile Asn Val Ser Val
340 345 350
Pro Phe Asp Lys Lys Ile Ala Thr Thr Asn Lys Val Val Gly Ile Asp
355 360 365
Val Asn Ile Lys His Met Leu Leu Ala Thr Asn Ile Leu Asp Asp Gly
370 375 380
Asn Val Asn Gly Tyr Val Asn Ile Tyr Lys Glu Val Ile Asn Asp Ser
385 390 395 400
Asp Phe Lys Lys Val Cys Asn Ser Thr Val Met Lys Tyr Phe Thr Asp
405 410 415
Phe Ser Lys Phe Val Thr Phe Cys Pro Leu Glu Phe Asp Phe Leu Phe
420 425 430
Ser Arg Val Cys Asn Gln Lys Gly Ile Tyr Asn Asp Asn Ser Ala Met
435 440 445
Glu Lys Ser Phe Ser Asp Val Leu Asn Lys Leu Lys Trp Asn Phe Ile
450 455 460
Glu Thr Gly Asp Asn Thr Lys Arg Ile Tyr Ile Glu Asn Val Met Lys
465 470 475 480
Leu Arg Ser Gln Met Lys Ala Tyr Ala Ile Val Lys Asn Ala Tyr Tyr
485 490 495
Lys Gln Gln Ser Glu Tyr Asp Phe Gly Lys Ser Glu Glu Phe Ile Gln
500 505 510
Glu His Pro Phe Ser Asn Thr Asp Lys Gly Ile Glu Ile Leu Asn Lys
515 520 525
Leu Asp Asn Ile Ser Lys Lys Ile Leu Gly Cys Arg Asn Asn Ile Ile
530 535 540
Gln Tyr Ser Tyr Asn Leu Phe Glu Ile Asn Gly Tyr Asp Met Ile Ser
545 550 555 560
Leu Glu Lys Leu Thr Ser Ser Gln Phe Lys Lys Lys Pro Phe Pro Thr
565 570 575
Val Asn Ser Leu Leu Lys Tyr His Lys Ile Leu Gly Cys Thr Gln Glu

580 585 590
Glu Met Glu Lys Lys Asp Ile Tyr Ser Val Ile Lys Lys Gly Tyr Tyr
595 600 605
Asp Ile Ile Phe Asp Asn Asp Val Val Thr Asp Ala Lys Leu Ser Thr
610 615 620
Lys Gly Glu Leu Ser Lys Phe Lys Asp Asp Phe Phe Asn Leu Met Ile
625 630 635 640
Lys Ser Ile His Phe Ala Asp Ile Lys Asp Tyr Phe Ile Thr Leu Ser
645 650 655
Asn Asn Gly Thr Ala Gly Val Ser Leu Val Pro Ser Tyr Phe Thr Ser
660 665 670
Gln Met Asp Ser Ile Asp His Lys Ile Tyr Phe Val Gln Asp Asn Lys
675 680 685
Ser Gly Lys Leu Lys Leu Ala Asn Lys His Lys Val Arg Ser Ser Gln
690 695 700
Glu Lys His Ile Asn Gly Leu Asn Ala Asp Tyr Asn Ala Ala Arg Asn
705 710 715 720
Ile Ala Tyr Ile Met Glu Asn Thr Glu Cys Arg Asn Met Phe Met Lys
725 730 735
Gln Ser Arg Thr Asp Lys Ser Leu Tyr Asn Lys Pro Ser Tyr Glu Thr
740 745 750
Phe Ile Lys Thr Gln Gly Ser Ala Val Ala Lys Leu Lys Lys Glu Gly
755 760 765
Phe Val Lys Ile Leu Asp Glu Ala Ser Val
770 775

<210> 105
<211> 778
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 105

Met Arg Ile Ser Pro His Leu Phe Tyr Ile Phe Phe Lys Lys Ile Trp
1 5 10 15
Lys Ser His Phe Phe Val Leu Ser Leu Tyr Gln Leu Asn Gln Tyr Ile
20 25 30
Met Ala Ser His Lys Lys Thr Glu Ser Asn Gln Ile Ile Lys Thr Phe
35 40 45
Ser Phe Lys Ile Lys Asn Ala Asn Gly Leu Ser Leu Asp Val Leu Asn

50 55 60

Asp Ala Ile Thr Glu Tyr Gln Asn Tyr Tyr Asn Ile Cys Ser Asp Trp
65 70 75 80
Ile Lys Asp His Leu Thr Met Lys Ile Gly Glu Leu Tyr Lys Tyr Ile
85 90 95
Pro Asp Glu Lys Lys Asn Ser Gly Tyr Ala Leu Thr Leu Ile Ser Asp
100 105 110
Glu Trp Lys Asp Lys Pro Met Tyr Met Met Phe Lys Lys Gly Tyr Pro
115 120 125
Ala Asn Asn Arg Asp Asn Ala Ile Tyr Glu Thr Leu Asn Thr Cys Asn
130 135 140
Thr Glu His Tyr Thr Gly Asn Ile Leu Asn Phe Pro Asp Thr Tyr Tyr
145 150 155 160
Arg Arg Phe Gly Tyr Val Ala Ser Thr Ile Ser Asn Tyr Val Thr Lys
165 170 175
Ile Ser Lys Met Ser Thr Gly Ser Arg Ser Lys Asn Ile Ser Asn Asp
180 185 190
Ser Asp Val Asp Thr Ile Met Glu Gln Val Ile Tyr Glu Met Glu His
195 200 205
Asn Gly Trp Thr Ser Val Lys Asp Trp Glu Asn Gln Met Glu Tyr Leu
210 215 220
Glu Ser Lys Thr Asp Ser Asn Pro Asn Phe Val Tyr Arg Met Thr Thr
225 230 235 240
Leu Tyr Glu Phe Tyr Lys Ser His Ile Asp Glu Val Asn Ser Lys Met
245 250 255
Glu Thr Met Ser Ile Asp Ser Leu Ile Lys Phe Gly Gly Cys Arg Arg
260 265 270
Lys Asp Ser Lys Lys Ser Met Tyr Ile Met Gly Gly Ser Asn Thr Pro
275 280 285
Phe Asp Ile Thr Gln Ile Gly Gly Asn Ser Leu Asn Ile Lys Phe Ser
290 295 300
Lys Asn Leu Asn Val Asp Val Phe Gly Arg Tyr Asp Val Ile Lys Asp
305 310 315 320
Asn Thr Leu Leu Val Asp Ile Ile Asn Gly His Gly Ala Ser Phe Val
325 330 335
Leu Lys Ile Ile Asn Gly Glu Ile Tyr Ile Asp Ile Asn Val Ser Val
340 345 350
Pro Phe Asp Lys Lys Ile Ala Thr Thr Asn Lys Val Val Gly Ile Asp
355 360 365

Val Asn Ile Lys His Met Leu Leu Ala Thr Asn Ile Leu Asp Asp Gly
370 375 380
Asn Val Asn Gly Tyr Val Asn Ile Tyr Lys Glu Val Ile Asn Asp Ser
385 390 395 400
Asp Phe Lys Lys Val Cys Asn Ser Thr Val Met Lys Tyr Phe Thr Asp
405 410 415
Phe Ser Lys Phe Val Thr Phe Cys Pro Leu Glu Phe Asp Phe Leu Phe
420 425 430
Ser Arg Val Cys Asn Gln Lys Gly Ile Tyr Asn Asp Asn Ser Ala Met
435 440 445
Glu Lys Ser Phe Ser Asp Val Leu Asn Lys Leu Lys Trp Asn Phe Ile
450 455 460
Glu Thr Gly Asp Asn Thr Lys Arg Ile Tyr Ile Glu Asn Val Met Lys
465 470 475 480
Leu Arg Ser Gln Met Lys Ala Tyr Ala Ile Val Lys Asn Ala Tyr Tyr
485 490 495
Lys Gln Gln Ser Glu Tyr Asp Phe Gly Lys Ser Glu Glu Phe Ile Gln
500 505 510
Glu His Pro Phe Ser Asn Thr Asp Lys Gly Ile Glu Ile Leu Asn Lys
515 520 525
Leu Asp Asn Ile Ser Lys Lys Ile Leu Gly Cys Arg Asn Asn Ile Ile
530 535 540
Gln Tyr Ser Tyr Asn Leu Phe Glu Ile Asn Gly Tyr Asp Met Ile Ser
545 550 555 560
Leu Glu Lys Leu Thr Ser Ser Gln Phe Lys Lys Asn Pro Phe Pro Thr
565 570 575
Val Asn Ser Leu Leu Lys Tyr His Lys Ile Leu Gly Cys Thr Gln Glu
580 585 590
Glu Met Glu Lys Lys Asp Ile Tyr Ser Val Ile Lys Lys Gly Tyr Tyr
595 600 605
Asp Ile Ile Phe Asp Asn Asp Val Val Thr Asp Ala Lys Leu Ser Thr
610 615 620
Lys Gly Glu Leu Ser Lys Phe Lys Asp Asp Phe Phe Asn Leu Met Ile
625 630 635 640
Lys Ser Ile His Phe Ala Asp Ile Lys Asp Tyr Phe Ile Thr Leu Ser
645 650 655
Asn Asn Gly Thr Ala Gly Val Ser Leu Val Pro Ser Tyr Phe Thr Ser
660 665 670
Gln Met Asp Ser Ile Asp His Lys Ile Tyr Phe Val Gln Asp Asn Lys

675 680 685

Ser Gly Lys Leu Lys Leu Ala Asn Lys His Lys Val Arg Ser Ser Gln

690 695 700

Glu Lys His Ile Asn Gly Leu Asn Ala Asp Tyr Asn Ala Ala Arg Asn

705 710 715 720

Ile Ala Tyr Ile Met Glu Asn Thr Glu Cys Arg Asn Met Phe Met Lys

725 730 735

Gln Ser Arg Thr Asp Lys Ser Leu Tyr Asn Lys Pro Ser Tyr Glu Thr

740 745 750

Phe Ile Lys Thr Gln Gly Ser Ala Val Ala Lys Leu Lys Lys Glu Gly

755 760 765

Phe Val Lys Ile Leu Asp Glu Ala Ser Val

770 775

<210> 106

<211> 778

<212> PRT

<213> 人工序列(Artificial sequence)

<220>

<223> 合成序列(synthetic sequence)

<400> 106

Met Arg Ile Ser Pro His Leu Phe Tyr Ile Phe Phe Lys Lys Ile Trp

1 5 10 15

Lys Cys His Ile Phe Val Leu Ser Leu Tyr Gln Leu Asn Gln Tyr Ile

20 25 30

Met Ala Ser His Lys Lys Thr Glu Ser Asn Gln Ile Ile Lys Thr Phe

35 40 45

Ser Phe Lys Ile Lys Asn Ala Asn Gly Leu Ser Leu Asp Val Leu Asn

50 55 60

Asp Ala Ile Thr Glu Tyr Gln Asn Tyr Tyr Asn Ile Cys Ser Asp Trp

65 70 75 80

Ile Lys Asp His Leu Thr Met Lys Ile Gly Glu Leu Tyr Lys Tyr Ile

85 90 95

Pro Asp Glu Lys Lys Asn Ser Gly Tyr Ala Leu Thr Leu Ile Ser Asp

100 105 110

Glu Trp Lys Asp Lys Pro Met Tyr Met Met Phe Lys Lys Gly Tyr Pro

115 120 125

Ala Asn Asn Arg Asp Asn Ala Ile Tyr Glu Thr Leu Asn Thr Cys Asn

130 135 140

Thr Glu His Tyr Thr Gly Asn Ile Leu Asn Phe Ser Asp Thr Tyr Tyr

145 150 155 160
Arg Arg Phe Gly Tyr Val Ala Ser Thr Ile Ser Asn Tyr Val Thr Lys
165 170 175
Ile Ser Lys Met Ser Thr Gly Ser Arg Ser Lys Asn Ile Ser Asn Asp
180 185 190
Ser Asp Val Asp Thr Ile Met Glu Gln Val Ile Tyr Glu Met Glu His
195 200 205
Asn Gly Trp Thr Ser Val Lys Asp Trp Glu Asn Gln Met Glu Tyr Leu
210 215 220
Glu Ser Lys Thr Asp Ser Asn Pro Asn Phe Val Tyr Arg Met Thr Thr
225 230 235 240
Leu Tyr Glu Phe Tyr Lys Ser His Ile Asp Glu Val Asn Ser Lys Met
245 250 255
Glu Thr Met Ser Ile Asp Ser Leu Ile Lys Phe Gly Gly Cys Arg Arg
260 265 270
Lys Asp Ser Lys Lys Ser Met Tyr Ile Met Gly Gly Ser Asn Thr Pro
275 280 285
Phe Asp Ile Thr Gln Ile Gly Gly Asn Ser Leu Asn Ile Lys Phe Ser
290 295 300
Lys Asn Leu Asn Val Asp Val Phe Gly Arg Tyr Asp Val Ile Lys Asp
305 310 315 320
Asn Thr Leu Leu Val Asp Ile Ile Asn Gly His Gly Ala Ser Phe Val
325 330 335
Leu Lys Ile Ile Asn Gly Glu Ile Tyr Ile Asp Ile Asn Val Ser Val
340 345 350
Pro Phe Asp Lys Lys Ile Ala Thr Thr Asn Lys Val Val Gly Val Asp
355 360 365
Val Asn Ile Lys His Met Leu Leu Ala Thr Asn Ile Leu Asp Asp Gly
370 375 380
Asn Val Asn Gly Tyr Val Asn Ile Tyr Lys Glu Val Ile Asn Asp Ser
385 390 395 400
Asp Phe Lys Lys Val Cys Asn Ser Thr Val Met Lys Tyr Phe Thr Asp
405 410 415
Phe Ser Lys Phe Val Thr Phe Cys Pro Leu Glu Phe Asp Phe Leu Phe
420 425 430
Ser Arg Val Cys Asn Gln Lys Gly Ile Tyr Asn Asp Asn Ser Ala Met
435 440 445
Glu Lys Ser Phe Ser Asp Val Leu Asn Lys Leu Lys Trp Asn Phe Ile
450 455 460

Glu Thr Gly Asp Asn Thr Lys Arg Ile Tyr Ile Glu Asn Val Met Lys
465 470 475 480
Leu Arg Ser Gln Met Lys Ala Tyr Ala Ile Val Lys Asn Ala Tyr Tyr
485 490 495
Lys Gln Gln Ser Glu Tyr Asp Phe Gly Lys Ser Glu Glu Phe Ile Gln
500 505 510
Glu His Pro Phe Ser Asn Thr Asp Lys Gly Ile Glu Ile Leu Asn Lys
515 520 525
Leu Asp Asn Ile Ser Lys Lys Ile Leu Gly Cys Arg Asn Asn Ile Ile
530 535 540
Gln Tyr Ser Tyr Asn Leu Phe Glu Ile Asn Gly Tyr Asp Met Ile Ser
545 550 555 560
Leu Glu Lys Leu Thr Ser Ser Gln Phe Lys Lys Lys Pro Phe Pro Thr
565 570 575
Val Asn Ser Leu Leu Lys Tyr His Lys Ile Leu Gly Cys Thr Gln Glu
580 585 590
Glu Met Glu Lys Lys Asp Ile Tyr Ser Val Ile Lys Lys Gly Tyr Tyr
595 600 605
Asp Ile Ile Phe Asp Asn Asp Val Val Thr Asp Ala Lys Leu Ser Thr
610 615 620
Lys Gly Glu Leu Ser Lys Phe Lys Asp Asp Phe Phe Asn Leu Met Ile
625 630 635 640
Lys Ser Ile His Phe Ala Asp Ile Lys Asp Tyr Phe Ile Thr Leu Ser
645 650 655
Asn Asn Gly Thr Ala Gly Val Ser Leu Val Pro Ser Tyr Phe Thr Ser
660 665 670
Gln Met Asp Ser Ile Asp His Lys Ile Tyr Phe Val Gln Asp Asn Lys
675 680 685
Ser Gly Lys Leu Lys Leu Ala Asn Lys His Lys Val Arg Ser Ser Gln
690 695 700
Glu Lys His Ile Asn Gly Leu Asn Ala Asp Tyr Asn Ala Ala Arg Asn
705 710 715 720
Ile Ala Tyr Ile Met Glu Asn Thr Asp Cys Arg Asn Met Phe Met Lys
725 730 735
Gln Ser Arg Thr Asp Lys Ser Leu Tyr Asn Lys Pro Ser Tyr Glu Thr
740 745 750
Phe Ile Lys Thr Gln Gly Ser Ala Val Ala Lys Leu Lys Lys Glu Gly
755 760 765
Phe Val Lys Ile Leu Asp Glu Ala Ser Val

770 775

<210> 107

<211> 351

<212> PRT

<213> 人工序列(Artificial sequence)

<220>

<223> 合成序列(synthetic sequence)

<400> 107

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

245 250 255
Thr His Asn Leu Tyr Phe Glu Asn Ala Lys Asn Gly Gly Leu Lys Leu
260 265 270
Ala Ser Lys Ser Lys Val Arg Lys Ser Gln Glu Tyr His Leu Asn Gly
275 280 285
Leu Pro Ala Asp Tyr Asn Ala Ala Arg Asn Ile Ala Tyr Ile Gly Leu
290 295 300
Asp Glu Ile Met Arg Asn Thr Phe Leu Lys Lys Ala Asn Ser Asn Lys
305 310 315 320
Ser Leu Tyr Asn Gln Pro Ile Tyr Asp Thr Gly Ile Lys Lys Thr Ala
325 330 335
Gly Val Phe Ser Arg Met Lys Lys Leu Lys Lys Tyr Lys Val Ile
340 345 350

<210> 108
<211> 717
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 108
Val Leu Asn Asp Ala Ile Thr Glu Tyr Gln Asn Tyr Tyr Asn Ile Cys
1 5 10 15
Ser Asp Trp Ile Lys Asp His Leu Thr Met Lys Ile Gly Glu Leu Tyr
20 25 30
Lys Tyr Ile Pro Asp Glu Lys Lys Asn Ser Gly Tyr Ala Leu Thr Leu
35 40 45
Ile Ser Asp Glu Trp Lys Asp Lys Pro Met Tyr Met Met Phe Lys Lys
50 55 60
Gly Tyr Pro Ala Asn Ser Arg Asp Asn Ala Ile Tyr Glu Ala Leu Asn
65 70 75 80
Thr Cys Asn Thr Glu His Tyr Thr Gly Asn Ile Leu Asn Phe Ser Asp
85 90 95
Thr Tyr Tyr Arg Arg Phe Gly Tyr Val Ala Ser Ala Ile Ser Asn Tyr
100 105 110
Val Thr Lys Ile Ser Lys Met Ser Thr Gly Ser Arg Ser Lys Asn Ile
115 120 125
Ser Asn Asp Ser Asp Val Asp Thr Ile Met Glu Gln Val Ile Tyr Glu
130 135 140
Met Glu His Asn Gly Trp Thr Ser Val Lys Asp Trp Glu Asn Gln Met

145 150 155 160
Glu Tyr Leu Glu Ser Lys Thr Asp Ser Asn Pro Asn Phe Val Tyr Arg
165 170 175
Met Thr Thr Leu Tyr Glu Phe Tyr Lys Ser His Ile Asp Glu Val Asn
180 185 190
Ser Lys Met Glu Thr Met Ser Ile Asp Ser Leu Ile Lys Phe Gly Gly
195 200 205
Cys Arg Arg Lys Asp Ser Lys Lys Ser Met Tyr Ile Met Gly Gly Ser
210 215 220
Asn Thr Pro Phe Asp Ile Thr Gln Ile Gly Gly Asn Ser Leu Asn Ile
225 230 235 240
Lys Phe Ser Lys Asn Leu Asn Val Asp Val Phe Gly Arg Tyr Asp Val
245 250 255
Ile Lys Asp Asn Thr Leu Leu Val Asp Ile Ile Asn Gly His Gly Ala
260 265 270
Ser Phe Val Leu Lys Ile Ile Asn Asp Glu Ile Tyr Ile Asp Ile Asn
275 280 285
Val Ser Val Pro Phe Asp Lys Lys Ile Ala Thr Thr Asn Lys Val Val
290 295 300
Gly Ile Asp Val Asn Ile Lys His Met Leu Leu Ala Thr Asn Ile Leu
305 310 315 320
Asp Asp Gly Asn Val Asn Gly Tyr Val Asn Ile Tyr Lys Glu Val Ile
325 330 335
Asn Asp Ser Asp Phe Lys Lys Val Cys Asn Ser Thr Val Met Lys Tyr
340 345 350
Phe Thr Asp Phe Ser Lys Phe Val Thr Phe Cys Pro Leu Glu Phe Asp
355 360 365
Phe Leu Phe Ser Arg Val Cys Asn Gln Lys Gly Ile Tyr Asn Asp Asn
370 375 380
Ser Ala Met Glu Lys Ser Phe Ser Asn Val Leu Asn Lys Leu Lys Trp
385 390 395 400
Asn Phe Ile Glu Thr Gly Asp Asn Thr Lys Arg Ile Tyr Ile Glu Asn
405 410 415
Val Met Lys Leu Arg Ser Gln Met Lys Ala Tyr Ala Ile Val Lys Asn
420 425 430
Ala Tyr Tyr Lys Gln Gln Ser Glu Tyr Asp Phe Gly Lys Ser Glu Glu
435 440 445
Phe Ile Gln Glu His Pro Phe Ser Asn Thr Asp Lys Gly Ile Glu Ile
450 455 460

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

<210> 109
<211> 742
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 109

Ile Phe Phe Lys Lys Ile Trp Lys Cys His Ile Phe Val Leu Ser Leu
1 5 10 15
Tyr Gln Leu Asn Gln Tyr Ile Met Ala Ser His Lys Lys Thr Glu Ser
20 25 30
Asn Gln Ile Ile Lys Thr Phe Ser Phe Lys Ile Lys Asn Ala Asn Gly
35 40 45
Leu Ser Leu Asp Val Leu Asn Asp Ala Ile Thr Glu Tyr Gln Asn Tyr
50 55 60
Tyr Asn Ile Cys Ser Asp Trp Ile Lys Asp His Leu Thr Met Lys Ile
65 70 75 80
Gly Glu Leu Tyr Lys Tyr Ile Pro Asp Glu Lys Lys Asn Ser Gly Tyr
85 90 95
Ala Leu Thr Leu Ile Ser Asp Glu Trp Lys Asp Lys Pro Met Tyr Met
100 105 110
Met Phe Lys Lys Gly Tyr Pro Ala Asn Ser Arg Asp Asn Ala Ile Tyr
115 120 125
Glu Ala Leu Asn Thr Cys Asn Thr Glu His Tyr Thr Gly Asn Ile Leu
130 135 140
Asn Phe Ser Asp Thr Tyr Tyr Arg Arg Phe Gly Tyr Val Ala Ser Ala
145 150 155 160
Ile Ser Asn Tyr Val Thr Lys Ile Ser Lys Met Ser Thr Gly Ser Arg
165 170 175
Ser Lys Asn Ile Ser Asn Asp Ser Asp Val Asp Thr Ile Met Glu Gln
180 185 190
Val Ile Tyr Glu Met Glu His Asn Gly Trp Thr Ser Val Lys Asp Trp
195 200 205
Glu Asn Gln Met Glu Tyr Leu Glu Ser Lys Thr Asp Ser Asn Pro Asn
210 215 220
Phe Val Tyr Arg Met Thr Thr Leu Tyr Glu Phe Tyr Lys Ser His Ile
225 230 235 240
Asp Glu Val Asn Ser Lys Met Glu Thr Met Ser Ile Asp Ser Leu Ile
245 250 255
Lys Phe Gly Gly Cys Arg Arg Lys Asp Ser Lys Lys Ser Met Tyr Ile
260 265 270
Met Gly Gly Ser Asn Thr Pro Phe Asp Ile Thr Gln Ile Asp Gly Asn
275 280 285
Ser Leu Asn Ile Lys Phe Ser Lys Asn Leu Asn Val Asp Val Phe Gly
290 295 300
Arg Tyr Asp Val Ile Lys Asp Asn Thr Leu Leu Val Asp Ile Ile Asn

305 310 315 320
Gly His Gly Ala Ser Phe Val Leu Lys Ile Ile Asn Asp Glu Ile Tyr
325 330 335
Ile Asp Ile Asn Val Ser Val Pro Phe Asp Lys Lys Ile Ala Thr Thr
340 345 350
Asn Lys Val Val Gly Val Asp Val Asn Ile Lys His Met Leu Leu Ala
355 360 365
Thr Asn Ile Leu Asp Asp Gly Asn Val Asn Gly Tyr Val Asn Ile Tyr
370 375 380
Lys Glu Val Ile Asn Asp Ser Asp Phe Lys Lys Val Cys Asn Ser Thr
385 390 395 400
Val Met Lys Tyr Phe Thr Asp Phe Ser Lys Phe Val Thr Phe Cys Pro
405 410 415
Leu Glu Phe Asp Phe Leu Phe Ser Arg Val Cys Asn Gln Lys Gly Ile
420 425 430
Tyr Asn Asp Asn Ser Ala Met Glu Lys Ser Phe Ser Asp Val Leu Asn
435 440 445
Lys Leu Lys Trp Asn Phe Ile Glu Thr Gly Asp Asn Thr Lys Arg Ile
450 455 460
Tyr Ile Glu Asn Val Met Lys Leu Arg Ser Gln Met Lys Ala Tyr Ala
465 470 475 480
Ile Val Lys Asn Ala Tyr Tyr Lys Gln Gln Ser Glu Tyr Asp Phe Gly
485 490 495
Lys Ser Glu Glu Phe Ile Gln Glu His Pro Phe Ser Asn Thr Asp Lys
500 505 510
Gly Ile Glu Ile Leu His Lys Leu Asp Asn Ile Ser Lys Lys Ile Leu
515 520 525
Gly Cys Arg Asn Asn Ile Ile Gln Tyr Ser Tyr Asn Leu Phe Glu Ile
530 535 540
Asn Gly Tyr Asp Met Ile Ser Leu Glu Lys Leu Thr Ser Ser Gln Phe
545 550 555 560
Lys Lys Lys Pro Phe Pro Thr Val Asn Ser Leu Leu Lys Tyr His Lys
565 570 575
Ile Leu Gly Cys Thr Gln Glu Glu Met Glu Lys Lys Asp Ile Tyr Ser
580 585 590
Val Ile Lys Lys Gly Tyr Tyr Asp Ile Ile Phe Asp Asn Asp Val Val
595 600 605
Thr Asp Ala Lys Leu Ser Thr Lys Gly Glu Leu Ser Lys Phe Lys Asp
610 615 620

Asp Phe Phe Asn Leu Met Ile Lys Ser Ile His Phe Ala Asp Ile Lys
625 630 635 640
Asp Tyr Phe Ile Thr Leu Ser Asn Asn Gly Thr Ala Gly Val Ser Leu
645 650 655
Val Pro Ser Tyr Phe Thr Ser Gln Met Asp Ser Ile Asp His Lys Ile
660 665 670
Tyr Phe Val Gln Asp Asn Lys Ser Gly Lys Leu Lys Leu Ala Asn Lys
675 680 685
His Lys Val Arg Ser Ser Gln Glu Lys His Ile Asn Gly Leu Asn Ala
690 695 700
Asp Tyr Asn Ala Ala Arg Asn Ile Ala Tyr Ile Met Glu Asn Thr Glu
705 710 715 720
Gln Gly Ser Ala Val Ala Lys Leu Lys Lys Glu Gly Phe Val Lys Ile
725 730 735
Leu Asp Glu Ala Ser Val
740

<210> 110
<211> 746
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 110

Met Ala Ser His Lys Lys Thr Glu Ser Asn Gln Ile Ile Lys Thr Phe
1 5 10 15
Ser Phe Lys Ile Lys Asn Ala Asn Gly Leu Ser Leu Asp Val Leu Asn
20 25 30
Asp Ala Ile Thr Glu Tyr Gln Asn Tyr Tyr Asn Ile Cys Ser Asp Trp
35 40 45
Ile Lys Asp His Leu Thr Met Lys Ile Gly Glu Leu Tyr Lys Tyr Ile
50 55 60
Pro Asp Glu Lys Lys Asn Ser Gly Tyr Ala Leu Thr Leu Ile Ser Asp
65 70 75 80
Glu Trp Lys Asp Lys Pro Met Tyr Met Met Phe Lys Lys Gly Tyr Pro
85 90 95
Ala Asn Asn Arg Asp Asn Ala Ile Tyr Glu Thr Leu Asn Thr Cys Asn
100 105 110
Thr Glu His Tyr Thr Gly Asn Ile Leu Asn Phe Ser Asp Thr Tyr Tyr
115 120 125

Arg Arg Phe Gly Tyr Val Ala Ser Ala Ile Ser Asn Tyr Val Thr Lys
130 135 140
Ile Ser Lys Met Ser Thr Gly Ser Arg Ser Lys Asn Ile Ser Asn Asp
145 150 155 160
Ser Asp Val Asp Thr Ile Met Glu Gln Val Ile Tyr Glu Met Glu His
165 170 175
Asn Gly Trp Thr Ser Val Lys Asp Trp Glu Asn Gln Met Glu Tyr Leu
180 185 190
Glu Ser Lys Thr Asp Ser Asn Pro Asn Phe Val Tyr Arg Met Thr Thr
195 200 205
Leu Tyr Glu Phe Tyr Lys Ser His Ile Asp Glu Val Asn Ser Lys Met
210 215 220
Glu Thr Met Ser Ile Asp Ser Leu Ile Lys Phe Gly Gly Cys Arg Arg
225 230 235 240
Lys Asp Ser Lys Lys Ser Met Tyr Ile Met Gly Gly Ser Asn Thr Pro
245 250 255
Phe Asp Ile Thr Gln Ile Gly Gly Asn Ser Leu Asn Ile Lys Phe Ser
260 265 270
Lys Asn Leu Asn Val Asp Val Phe Gly Arg Tyr Asp Val Ile Lys Asp
275 280 285
Asn Thr Leu Leu Val Asp Ile Ile Asn Gly His Gly Ala Ser Phe Val
290 295 300
Leu Lys Ile Ile Asn Gly Glu Ile Tyr Ile Asp Ile Asn Val Ser Val
305 310 315 320
Pro Phe Asp Lys Lys Ile Ala Thr Thr Asn Lys Val Val Gly Val Asp
325 330 335
Val Asn Ile Lys His Met Leu Leu Ala Thr Asn Ile Leu Asp Asp Gly
340 345 350
Asn Val Asn Gly Tyr Val Asn Ile Tyr Lys Glu Val Ile Asn Asp Ser
355 360 365
Asp Phe Lys Lys Val Cys Asn Ser Thr Val Met Lys Tyr Phe Thr Asp
370 375 380
Phe Ser Lys Phe Val Thr Phe Cys Pro Leu Glu Phe Asp Phe Leu Phe
385 390 395 400
Ser Arg Val Cys Asn Gln Lys Gly Ile Tyr Asn Asp Asn Ser Ala Met
405 410 415
Glu Lys Ser Phe Ser Asp Val Leu Asn Lys Leu Lys Trp Asn Phe Ile
420 425 430
Glu Thr Gly Asp Asn Thr Lys Arg Ile Tyr Ile Glu Asn Val Met Lys

435 440 445

Leu Arg Ser Gln Met Lys Ala Tyr Ala Ile Val Lys Asn Ala Tyr Tyr

450 455 460

Lys Gln Gln Ser Glu Tyr Asp Phe Gly Lys Ser Glu Glu Phe Ile Gln

465 470 475 480

Glu His Pro Phe Ser Asn Thr Asp Lys Gly Ile Glu Ile Leu Asn Lys

485 490 495

Leu Asp Asn Ile Ser Lys Lys Ile Leu Gly Cys Arg Asn Asn Ile Ile

500 505 510

Gln Tyr Ser Tyr Asn Leu Phe Glu Ile Asn Gly Tyr Asp Met Ile Ser

515 520 525

Leu Glu Lys Leu Thr Ser Ser Gln Phe Lys Lys Lys Pro Phe Pro Thr

530 535 540

Val Asn Ser Leu Leu Lys Tyr His Lys Ile Leu Gly Cys Thr Gln Glu

545 550 555 560

Glu Met Glu Lys Lys Asp Ile Tyr Ser Val Ile Lys Lys Gly Tyr Tyr

565 570 575

Asp Ile Ile Phe Asp Asn Asp Val Val Thr Asp Ala Lys Leu Ser Ala

580 585 590

Lys Gly Glu Leu Ser Lys Phe Lys Asp Asp Phe Phe Asn Leu Met Ile

595 600 605

Lys Ser Ile His Phe Ala Asp Ile Lys Asp Tyr Phe Ile Thr Leu Ser

610 615 620

Asn Asn Gly Thr Ala Gly Val Ser Leu Val Pro Ser Tyr Phe Thr Ser

625 630 635 640

Gln Met Asp Ser Ile Asp His Lys Ile Tyr Phe Val Gln Asp Asn Lys

645 650 655

Ser Gly Lys Leu Lys Leu Ala Asn Lys His Lys Val Arg Ser Ser Gln

660 665 670

Glu Lys His Ile Asn Gly Leu Asn Ala Asp Tyr Asn Ala Ala Arg Asn

675 680 685

Ile Ala Tyr Ile Met Glu Asn Thr Glu Cys Arg Asn Met Phe Met Lys

690 695 700

Gln Ser Arg Thr Asp Lys Ser Leu Tyr Asn Lys Pro Ser Tyr Glu Thr

705 710 715 720

Phe Ile Lys Thr Gln Gly Ser Ala Val Ser Lys Leu Lys Lys Asp Gly

725 730 735

Phe Val Lys Ile Leu Asp Glu Ala Ser Val

740 745

<210> 111
<211> 782
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 111

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

Asn Ser Lys Met Glu Thr Met Ser Ile Asp Ser Leu Ile Lys Phe Gly
260 265 270
Gly Cys Arg Arg Lys Asp Ser Lys Lys Ser Met Tyr Ile Met Gly Gly
275 280 285
Ser Asn Thr Pro Phe Asp Ile Thr Gln Ile Gly Gly Asn Ser Leu Asn
290 295 300
Ile Lys Phe Ser Lys Asn Leu Asn Val Asp Val Phe Gly Arg Tyr Asp
305 310 315 320
Val Ile Lys Asp Asn Thr Leu Leu Val Asp Ile Ile Asn Glu His Gly
325 330 335
Ala Ser Phe Val Leu Lys Ile Ile Asn Asp Glu Ile Tyr Ile Asp Ile
340 345 350
Asn Val Ser Val Pro Phe Asp Lys Lys Ile Ala Thr Thr Asn Lys Val
355 360 365
Val Gly Val Asp Val Asn Ile Lys His Met Leu Leu Ala Thr Asn Ile
370 375 380
Leu Asp Asp Gly Asn Val Asn Gly Tyr Val Asn Ile Tyr Lys Glu Val
385 390 395 400
Ile Asn Asp Ser Asp Phe Lys Lys Val Cys Asn Ser Thr Val Met Lys
405 410 415
Tyr Phe Thr Asp Phe Ser Lys Phe Val Thr Phe Cys Pro Leu Glu Phe
420 425 430
Asp Phe Leu Phe Ser Arg Val Cys Asn Gln Lys Gly Ile Tyr Asn Asp
435 440 445
Asn Ser Ala Met Glu Lys Ser Phe Ser Asp Val Leu Asn Lys Leu Lys
450 455 460
Trp Asn Phe Ile Glu Thr Gly Asp Asn Thr Lys Arg Ile Tyr Ile Glu
465 470 475 480
Asn Val Met Lys Leu Arg Ser Gln Met Lys Ala Tyr Ala Ile Val Lys
485 490 495
Asn Ala Tyr Tyr Lys Gln Gln Ser Glu Tyr Asp Phe Gly Lys Ser Glu
500 505 510
Glu Phe Ile Gln Glu His Pro Phe Ser Asn Thr Asp Lys Gly Ile Glu
515 520 525
Ile Leu Asn Lys Leu Asp Asn Ile Ser Lys Lys Ile Leu Gly Cys Arg
530 535 540
Asn Asn Ile Ile Gln Tyr Ser Tyr Asn Leu Phe Glu Ile Asn Gly Tyr
545 550 555 560
Asp Met Ile Ser Leu Glu Lys Leu Thr Ser Ser Gln Phe Lys Lys Lys

565 570 575
Pro Phe Pro Thr Val Asn Ser Leu Leu Lys Tyr His Lys Ile Leu Gly
580 585 590
Cys Thr Gln Glu Glu Met Glu Lys Lys Asp Ile Tyr Ser Val Ile Lys
595 600 605
Lys Gly Tyr Tyr Asp Ile Ile Phe Asp Asn Asp Val Val Thr Asp Ala
610 615 620
Lys Leu Ser Thr Lys Gly Glu Leu Ser Lys Phe Lys Asp Asp Phe Phe
625 630 635 640
Asn Leu Met Ile Lys Ser Ile His Phe Ala Asp Ile Lys Asp Tyr Phe
645 650 655
Ile Thr Leu Ser Asn Asn Gly Thr Ala Gly Val Ser Leu Val Pro Ser
660 665 670
Tyr Phe Thr Ser Gln Met Asp Ser Ile Asp His Lys Ile Tyr Phe Val
675 680 685
Gln Asp Asn Lys Ser Gly Lys Leu Lys Leu Ala Asn Lys His Lys Val
690 695 700
Arg Ser Ser Gln Glu Lys His Ile Asn Gly Leu Asn Ala Asp Tyr Asn
705 710 715 720
Ala Ala Arg Asn Ile Ala Tyr Ile Met Glu Asn Thr Glu Cys Arg Asn
725 730 735
Met Phe Met Lys Gln Ser Arg Thr Asp Lys Ser Leu Tyr Asn Lys Pro
740 745 750
Ser Tyr Glu Thr Phe Ile Lys Thr Gln Gly Ser Ala Val Ala Lys Leu
755 760 765
Lys Lys Glu Gly Phe Val Lys Ile Leu Asp Glu Ala Ser Ala
770 775 780

<210> 112
<211> 92
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 112
Asn Lys Ser Gly Lys Leu Lys Leu Ala Asn Lys His Lys Val Arg Ser
1 5 10 15
Ser Gln Glu Lys His Ile Asn Gly Leu Asn Ala Asp Tyr Asn Ala Ala
20 25 30
Arg Asn Ile Ala Tyr Ile Met Glu Asn Thr Glu Cys Arg Asn Met Phe

35 40 45

Met Lys Gln Ser Arg Thr Asp Lys Ser Leu Tyr Asn Lys Pro Ser Tyr
50 55 60

Glu Thr Phe Ile Lys Thr Gln Gly Ser Ala Val Ala Lys Leu Lys Lys
65 70 75 80

Glu Gly Phe Val Lys Ile Leu Asp Glu Ala Ser Val
85 90

<210> 113

<211> 746

<212> PRT

<213> 人工序列(Artificial sequence)

<220>

<223> 合成序列(synthetic sequence)

<400> 113

Met Ala Ser His Lys Lys Thr Glu Ser Asn Gln Ile Ile Lys Thr Phe
1 5 10 15

Ser Phe Lys Ile Lys Asn Ala Asn Gly Leu Ser Leu Asp Val Leu Asn
20 25 30

Asp Ala Ile Thr Glu Tyr Gln Asn Tyr Tyr Asn Ile Cys Ser Asp Trp
35 40 45

Ile Lys Asp His Leu Thr Met Lys Ile Gly Glu Leu Tyr Lys Tyr Ile
50 55 60

Pro Asp Glu Lys Lys Asn Ser Gly Tyr Ala Leu Thr Leu Ile Ser Asp
65 70 75 80

Glu Trp Lys Asp Lys Pro Met Tyr Met Met Phe Lys Lys Gly Tyr Pro
85 90 95

Ala Asn Ser Arg Asp Asn Ala Ile Tyr Glu Ala Leu Asn Thr Cys Asn
100 105 110

Thr Glu His Tyr Thr Gly Asn Ile Leu Asn Phe Ser Asp Thr Tyr Tyr
115 120 125

Arg Arg Phe Gly Tyr Val Ala Ser Thr Ile Ser Asn Tyr Val Thr Lys
130 135 140

Ile Ser Lys Met Ser Thr Gly Ser Arg Ser Lys Asn Ile Ser Asn Asp
145 150 155 160

Ser Asp Val Asp Thr Ile Met Glu Gln Val Ile Tyr Glu Met Glu His
165 170 175

Asn Gly Trp Thr Ser Val Lys Asp Trp Glu Asn Gln Met Glu Tyr Leu
180 185 190

Glu Ser Lys Thr Asp Ser Asn Pro Asn Phe Val Tyr Arg Met Thr Thr

195 200 205

Leu Tyr Glu Phe Tyr Lys Ser His Ile Asp Glu Val Asn Ser Lys Met

210 215 220

Glu Thr Met Ser Ile Asp Ser Leu Ile Lys Phe Gly Gly Cys Arg Arg

225 230 235 240

Lys Asp Ser Lys Lys Ser Met Tyr Ile Met Gly Gly Ser Asn Thr Pro

245 250 255

Phe Asp Ile Thr Gln Ile Gly Gly Asn Ser Leu Asn Ile Lys Phe Ser

260 265 270

Lys Asn Leu Asn Val Asp Val Phe Gly Arg Tyr Asp Val Ile Lys Asp

275 280 285

Asn Thr Leu Leu Val Asp Ile Ile Asn Gly His Gly Ala Ser Phe Val

290 295 300

Leu Lys Ile Ile Asn Gly Glu Ile Tyr Ile Asp Ile Asn Val Ser Val

305 310 315 320

Pro Phe Asp Lys Lys Ile Ala Thr Thr Asn Lys Val Val Gly Val Asp

325 330 335

Val Asn Ile Lys His Met Leu Leu Ala Thr Asn Ile Leu Asp Asp Gly

340 345 350

Asn Val Asn Gly Tyr Val Asn Ile Tyr Lys Glu Val Ile Asn Asp Ser

355 360 365

Asp Phe Lys Lys Val Cys Asn Ser Thr Val Met Lys Tyr Phe Thr Asp

370 375 380

Phe Ser Lys Phe Val Thr Phe Cys Pro Leu Glu Phe Asp Phe Leu Phe

385 390 395 400

Ser Arg Val Cys Asn Gln Lys Gly Ile Tyr Asn Asp Asn Ser Ala Met

405 410 415

Glu Lys Ser Phe Ser Asp Val Leu Asn Lys Leu Lys Trp Asn Phe Ile

420 425 430

Glu Thr Gly Asp Asn Thr Lys Arg Ile Tyr Ile Glu Asn Val Met Lys

435 440 445

Leu Arg Ser Gln Met Lys Ala Tyr Ala Ile Val Lys Asn Ala Tyr Tyr

450 455 460

Lys Gln Gln Ser Glu Tyr Asp Phe Gly Lys Ser Glu Glu Phe Ile Gln

465 470 475 480

Glu His Pro Phe Ser Asn Thr Asp Lys Gly Ile Glu Ile Leu Asn Lys

485 490 495

Leu Asp Asn Ile Ser Lys Lys Ile Leu Gly Cys Arg Asn Asn Ile Ile

500 505 510

Gln Tyr Ser Tyr Asn Leu Phe Glu Ile Asn Gly Tyr Asp Met Ile Ser
515 520 525
Leu Glu Lys Leu Thr Ser Ser Gln Phe Glu Lys Lys Pro Phe Pro Thr
530 535 540
Val Asn Ser Leu Leu Lys Tyr His Lys Ile Leu Gly Cys Thr Gln Glu
545 550 555 560
Glu Met Glu Lys Lys Asp Ile Tyr Ser Val Ile Lys Lys Gly Tyr Tyr
565 570 575
Asp Ile Ile Phe Asp Asn Asp Val Val Thr Asp Ala Lys Leu Ser Thr
580 585 590
Lys Gly Glu Leu Ser Lys Phe Lys Asp Asp Phe Phe Asn Leu Met Ile
595 600 605
Lys Ser Ile His Phe Ala Asp Ile Lys Asp Tyr Phe Ile Thr Leu Ser
610 615 620
Asn Asn Gly Thr Ala Gly Val Ser Leu Val Pro Ser Tyr Phe Thr Ser
625 630 635 640
Gln Met Asp Ser Ile Asp His Lys Ile Tyr Phe Val Gln Asp Asn Lys
645 650 655
Ser Gly Lys Leu Lys Leu Ala Asn Lys His Lys Val Arg Ser Ser Gln
660 665 670
Glu Lys His Ile Asn Gly Leu Asn Ala Asp Tyr Asn Ala Ala Arg Asn
675 680 685
Ile Ala Tyr Ile Met Glu Asn Thr Asp Cys Arg Asn Met Phe Met Lys
690 695 700
Gln Ser Arg Thr Asp Lys Ser Leu Tyr Asn Lys Pro Ser Tyr Glu Thr
705 710 715 720
Phe Ile Lys Thr Gln Gly Ser Ala Val Ala Lys Leu Lys Lys Glu Gly
725 730 735
Phe Val Lys Ile Leu Asp Glu Ala Ser Val
740 745

<210> 114
<211> 427
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 114
Val Pro Phe Asp Lys Lys Ile Ala Thr Thr Asn Lys Val Val Gly Val
1 5 10 15

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

325 330 335
Lys Ser Gly Lys Leu Lys Leu Ala Asn Lys His Lys Val Arg Ser Ser
340 345 350
Gln Glu Lys His Ile Asn Gly Leu Asn Ala Asp Tyr Asn Ala Ala Arg
355 360 365
Asn Ile Ala Tyr Ile Met Glu Asn Thr Asp Cys Arg Asn Met Phe Met
370 375 380
Lys Gln Ser Arg Thr Asp Lys Ser Leu Tyr Asn Lys Pro Ser Tyr Glu
385 390 395 400
Thr Phe Ile Lys Thr Gln Gly Ser Ala Val Ala Lys Leu Lys Lys Glu
405 410 415
Gly Phe Val Lys Ile Leu Asp Glu Ala Ser Val
420 425

<210> 115
<211> 746
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 115
Met Ala Ser His Lys Lys Thr Glu Ser Asn Gln Ile Ile Lys Thr Phe
1 5 10 15
Ser Phe Lys Ile Lys Asn Ala Asn Gly Leu Ser Leu Asp Val Leu Asn
20 25 30
Asp Ala Ile Thr Glu Tyr Gln Asn Tyr Tyr Asn Ile Cys Ser Asp Trp
35 40 45
Ile Lys Asp His Leu Thr Met Lys Ile Gly Glu Leu Tyr Lys Tyr Ile
50 55 60
Pro Asp Glu Lys Lys Asn Ser Gly Tyr Ala Leu Thr Leu Ile Ser Asp
65 70 75 80
Glu Trp Lys Asp Lys Pro Met Tyr Met Met Phe Lys Lys Gly Tyr Pro
85 90 95
Ala Asn Asn Arg Asp Asn Ala Ile Tyr Glu Thr Leu Asn Thr Cys Asn
100 105 110
Thr Glu His Tyr Thr Gly Asn Ile Leu Asn Phe Ser Asp Thr Tyr Tyr
115 120 125
Arg Arg Phe Gly Tyr Val Ala Ser Ala Ile Ser Asn Tyr Val Thr Lys
130 135 140
Ile Ser Lys Met Ser Thr Gly Ser Arg Ser Lys Asn Ile Ser Asn Asp

145 150 155 160
Ser Asp Val Asp Thr Ile Met Glu Gln Val Ile Tyr Glu Met Glu His
165 170 175
Asn Gly Trp Thr Ser Val Lys Asp Trp Glu Asn Gln Met Glu Tyr Leu
180 185 190
Glu Ser Lys Thr Asp Ser Asn Pro Asn Phe Val Tyr Arg Met Thr Thr
195 200 205
Leu Tyr Glu Phe Tyr Lys Ser His Ile Asp Glu Val Asn Ser Lys Met
210 215 220
Glu Thr Met Ser Ile Asp Ser Leu Ile Lys Phe Gly Gly Cys Arg Arg
225 230 235 240
Lys Asp Ser Lys Lys Ser Met Tyr Ile Met Gly Gly Ser Asn Thr Pro
245 250 255
Phe Asp Ile Thr Gln Ile Gly Gly Asn Ser Leu Asn Ile Lys Phe Ser
260 265 270
Lys Asn Leu Asn Val Asp Val Phe Gly Arg Tyr Asp Val Ile Lys Asp
275 280 285
Asn Thr Leu Leu Val Asp Ile Ile Asn Gly His Gly Ala Ser Phe Val
290 295 300
Leu Lys Ile Ile Asn Gly Glu Ile Tyr Ile Asp Ile Asn Val Ser Val
305 310 315 320
Pro Phe Asp Lys Lys Ile Ala Thr Thr Asn Lys Val Val Gly Val Asp
325 330 335
Val Asn Ile Lys His Met Leu Leu Ala Thr Asn Ile Leu Asp Asp Gly
340 345 350
Asn Val Asn Gly Tyr Val Asn Ile Tyr Lys Glu Val Ile Asn Asp Ser
355 360 365
Asp Phe Lys Lys Val Cys Asn Ser Thr Val Met Lys Tyr Phe Thr Asp
370 375 380
Phe Ser Lys Phe Val Thr Phe Cys Pro Leu Glu Phe Asp Phe Leu Phe
385 390 395 400
Ser Arg Val Cys Asn Gln Lys Gly Ile Tyr Asn Asp Asn Ser Ala Met
405 410 415
Glu Lys Ser Phe Ser Asp Val Leu Asn Lys Leu Lys Trp Asn Phe Ile
420 425 430
Glu Thr Gly Asp Asn Thr Lys Arg Ile Tyr Ile Glu Asn Val Met Lys
435 440 445
Leu Arg Ser Gln Met Lys Ala Tyr Ala Ile Val Lys Asn Ala Tyr Tyr
450 455 460

Lys Gln Gln Ser Glu Tyr Asp Phe Gly Lys Ser Glu Glu Phe Ile Gln
465 470 475 480
Glu His Pro Phe Ser Asn Thr Asp Lys Gly Ile Glu Ile Leu Asn Lys
485 490 495
Leu Asp Asn Ile Ser Lys Lys Ile Leu Gly Cys Arg Asn Asn Ile Ile
500 505 510
Gln Tyr Ser Tyr Asn Leu Phe Glu Ile Asn Gly Tyr Asp Met Ile Ser
515 520 525
Leu Glu Lys Leu Thr Ser Ser Gln Phe Lys Lys Lys Pro Phe Pro Thr
530 535 540
Val Asn Ser Leu Leu Lys Tyr His Lys Ile Leu Gly Cys Thr Gln Glu
545 550 555 560
Glu Met Glu Lys Lys Asp Ile Tyr Ser Val Ile Lys Lys Gly Tyr Tyr
565 570 575
Asp Ile Ile Phe Asp Asn Asp Val Val Thr Asp Ala Lys Leu Ser Thr
580 585 590
Lys Gly Glu Leu Ser Lys Phe Lys Asp Asp Phe Phe Asn Leu Met Ile
595 600 605
Lys Ser Ile His Phe Ala Asp Ile Lys Asp Tyr Phe Ile Thr Leu Ser
610 615 620
Asn Asn Gly Thr Ala Gly Val Ser Leu Val Pro Ser Tyr Phe Thr Ser
625 630 635 640
Gln Met Asp Ser Ile Asp His Lys Ile Tyr Phe Val Gln Asp Asn Lys
645 650 655
Ser Gly Lys Leu Lys Leu Ala Asn Lys His Lys Val Arg Ser Ser Gln
660 665 670
Glu Lys His Ile Asn Gly Leu Asn Ala Asp Tyr Asn Ala Ala Arg Asn
675 680 685
Ile Ala Tyr Ile Met Glu Asn Thr Glu Cys Arg Asn Met Phe Met Lys
690 695 700
Gln Ser Arg Thr Asp Lys Ser Leu Tyr Asn Lys Pro Ser Tyr Glu Thr
705 710 715 720
Phe Ile Lys Thr Gln Gly Ser Ala Val Ala Lys Leu Lys Lys Glu Gly
725 730 735
Phe Val Lys Ile Leu Asp Glu Ala Ser Val
740 745

<210> 116
<211> 735
<212> PRT

<213> 人工序列(Artificial sequence)

<220>

<223> 合成序列(synthetic sequence)

<400> 116

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

275 280 285

Glu Val Leu Val Asp Ile Met Ala Asn His Gly Asp Ser Ile Val Leu

290 295 300

Lys Ile Val Asn Gly Glu Leu Tyr Ala Asp Val Pro Cys Ser Thr Thr

305 310 315 320

Leu Asn Lys Val Glu Ser Thr Phe Asp Lys Val Ala Gly Ile Asp Val

325 330 335

Asn Met Lys His Met Leu Leu Ser Thr Ser Val Thr Asp Asn Gly Asn

340 345 350

Ser Asp Phe Val Asn Ile Tyr Lys Glu Met Ser Asn Asn Ala Glu Phe

355 360 365

Met Ala Leu Cys Pro Glu Glu Asp Arg Lys Tyr Tyr Lys Asp Ile Ser

370 375 380

Gln Tyr Val Thr Phe Ala Pro Leu Glu Leu Asp Leu Leu Phe Ser Arg

385 390 395 400

Ile Ser Lys Gln Gly Lys Val Lys Met Glu Lys Ala Tyr Ser Glu Ile

405 410 415

Leu Glu Ala Leu Lys Trp Lys Phe Phe Ala Asn Gly Asp Asn Lys Asn

420 425 430

Arg Ile Tyr Val Glu Asn Ile Gln Lys Ile Arg Gln Gln Ile Lys Ala

435 440 445

Leu Cys Val Ile Lys Asn Ala Tyr Tyr Glu Gln Gln Ser Ala Tyr Asp

450 455 460

Ile Asp Lys Thr Gln Glu Tyr Ile Glu Ala His Pro Phe Ser Leu Thr

465 470 475 480

Glu Lys Gly Met Ser Ile Lys Ser Lys Met Asp Asn Ile Cys Arg Thr

485 490 495

Ile Ile Gly Cys Arg Asn Asn Ile Ile Asp Tyr Ala Tyr Ser Phe Phe

500 505 510

Glu Arg Asn Asp Tyr Ser Ile Ile Gly Leu Glu Lys Leu Thr Ser Ser

515 520 525

Gln Phe Glu Lys Thr Lys Ser Leu Pro Thr Cys Lys Ser Leu Leu Asn

530 535 540

Phe His Lys Val Leu Gly His Thr Leu Ser Glu Leu Glu Thr Leu Pro

545 550 555 560

Ile Asn Asp Val Val Lys Lys Gly Tyr Tyr Thr Phe Thr Thr Asp Asn

565 570 575

Glu Gly Arg Ile Thr Asp Ala Ser Leu Ser Glu Lys Gly Lys Val Arg

580 585 590

Lys Met Lys Asp Asp Phe Phe Asn Gln Ala Ile Lys Ala Ile His Phe
595 600 605
Ala Asp Val Lys Asp Tyr Phe Ala Thr Leu Ser Asn Asn Gly Gln Thr
610 615 620
Gly Ile Phe Phe Val Pro Ser Gln Phe Thr Ser Gln Met Asp Ser Asn
625 630 635 640
Thr His Thr Leu Tyr Phe Glu Asn Ala Lys Asn Gly Gly Leu Lys Leu
645 650 655
Ala Ser Lys Tyr Lys Val Arg Lys Ser Gln Glu Tyr His Leu Asn Gly
660 665 670
Leu Pro Ala Asp Tyr Asn Ala Ala Arg Asn Ile Ala Tyr Ile Gly Leu
675 680 685
Asp Glu Ile Met Arg Asn Thr Phe Leu Lys Lys Ala Asn Ser Asn Lys
690 695 700
Ser Leu Tyr Asn Gln Pro Ile Tyr Asp Thr Gly Ile Lys Lys Thr Ala
705 710 715 720
Gly Val Phe Ser Arg Met Lys Lys Leu Lys Lys Tyr Lys Val Ile
725 730 735

<210> 117
<211> 746
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 117

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

Thr Glu His Tyr Thr Gly Asn Ile Leu Asn Phe Ser Asp Thr Tyr Tyr
115 120 125
Arg Arg Phe Gly Tyr Val Ala Ser Thr Ile Ser Asn Tyr Val Thr Lys
130 135 140
Ile Ser Lys Met Ser Thr Gly Ser Arg Ser Lys Asn Ile Ser Asn Asp
145 150 155 160
Ser Asp Val Asp Thr Ile Met Glu Gln Val Ile Tyr Glu Met Glu His
165 170 175
Asn Gly Trp Thr Ser Val Lys Asp Trp Glu Asn Gln Met Glu Tyr Leu
180 185 190
Glu Ser Lys Thr Asp Ser Asn Pro Asn Phe Val Tyr Arg Met Thr Thr
195 200 205
Leu Tyr Glu Phe Tyr Lys Ser His Ile Asp Glu Val Asn Ser Lys Met
210 215 220
Glu Thr Met Ser Ile Asp Ser Leu Ile Lys Phe Gly Gly Cys Arg Arg
225 230 235 240
Lys Asp Ser Lys Lys Ser Met Tyr Ile Met Gly Gly Ser Asn Thr Pro
245 250 255
Phe Asp Ile Thr Gln Ile Gly Gly Asn Ser Leu Asn Ile Lys Phe Ser
260 265 270
Lys Asn Leu Asn Val Asp Val Phe Gly Arg Tyr Asp Val Ile Lys Asp
275 280 285
Asn Thr Leu Leu Val Asp Ile Ile Asn Gly His Gly Ala Ser Phe Val
290 295 300
Leu Lys Ile Ile Asn Gly Glu Ile Tyr Ile Asp Ile Asn Val Ser Val
305 310 315 320
Pro Phe Asp Lys Lys Ile Ala Thr Thr Asn Lys Val Val Gly Val Asp
325 330 335
Val Asn Ile Lys His Met Leu Leu Ala Thr Asn Ile Leu Asp Asp Gly
340 345 350
Asn Val Asn Gly Tyr Val Asn Ile Tyr Lys Glu Val Ile Asn Asp Ser
355 360 365
Asp Phe Lys Lys Val Cys Asn Ser Thr Val Met Gln Tyr Phe Thr Asp
370 375 380
Phe Ser Lys Phe Val Thr Phe Cys Pro Leu Glu Phe Asp Phe Leu Phe
385 390 395 400
Ser Arg Val Cys Asn Gln Lys Gly Ile Tyr Asn Asp Asn Ser Ala Met
405 410 415
Glu Lys Ser Phe Ser Asp Val Leu Asn Lys Leu Lys Trp Asn Phe Ile

420 425 430

Glu Thr Gly Asp Asn Thr Lys Arg Ile Tyr Ile Glu Asn Val Met Lys

435 440 445

Leu Arg Ser Gln Met Lys Ala Tyr Ala Ile Val Lys Asn Ala Tyr Tyr

450 455 460

Lys Gln Gln Ser Glu Tyr Asp Phe Gly Lys Ser Glu Glu Phe Ile Gln

465 470 475 480

Glu His Pro Phe Ser Asn Thr Asp Lys Gly Ile Glu Ile Leu Asn Lys

485 490 495

Leu Asp Asn Ile Ser Lys Lys Ile Leu Gly Cys Arg Asn Asn Ile Ile

500 505 510

Gln Tyr Ser Tyr Asn Leu Phe Glu Ile Asn Gly Tyr Asp Met Ile Ser

515 520 525

Leu Glu Lys Leu Thr Ser Ser Gln Phe Lys Lys Lys Pro Phe Pro Thr

530 535 540

Val Asn Ser Leu Leu Lys Tyr His Lys Ile Leu Gly Cys Thr Gln Glu

545 550 555 560

Glu Met Glu Lys Lys Asp Ile Tyr Ser Val Ile Lys Lys Gly Tyr Tyr

565 570 575

Asp Ile Ile Phe Asp Asn Asp Val Val Thr Asp Ala Lys Leu Ser Thr

580 585 590

Lys Gly Glu Leu Ser Lys Phe Lys Asp Asp Phe Phe Asn Leu Met Ile

595 600 605

Lys Ser Ile His Phe Ala Asp Ile Lys Asp Tyr Phe Ile Thr Leu Ser

610 615 620

Asn Asn Gly Thr Ala Gly Val Ser Leu Val Pro Ser Tyr Phe Thr Ser

625 630 635 640

Gln Met Asp Ser Ile Asp His Lys Ile Tyr Phe Val Gln Asp Asn Lys

645 650 655

Ser Gly Lys Leu Lys Leu Ala Asn Lys His Lys Val Arg Ser Ser Gln

660 665 670

Glu Lys His Ile Asn Gly Leu Asn Ala Asp Tyr Asn Ala Ala Arg Asn

675 680 685

Ile Ala Tyr Ile Met Glu Asn Thr Glu Cys Arg Asn Met Phe Met Lys

690 695 700

Gln Ser Arg Thr Asp Lys Ser Leu Tyr Asn Lys Pro Ser Tyr Glu Thr

705 710 715 720

Phe Ile Lys Thr Gln Gly Ser Ala Val Ala Lys Leu Lys Lys Glu Gly

725 730 735

Phe Val Lys Ile Ile Asp Glu Ala Ser Val
740 745

<210> 118
<211> 746
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 118

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

Lys Asp Ser Lys Lys Ser Met Tyr Ile Met Gly Gly Ser Asn Thr Pro
245 250 255
Phe Asp Ile Thr Gln Ile Gly Gly Asn Ser Leu Asn Ile Lys Phe Ser
260 265 270
Lys Asn Leu Asn Val Asp Val Phe Gly Arg Tyr Asp Val Ile Lys Asp
275 280 285
Asn Thr Leu Leu Val Asp Ile Ile Asn Gly His Gly Ala Ser Phe Val
290 295 300
Leu Lys Ile Ile Asn Gly Glu Ile Tyr Ile Asp Ile Asn Val Ser Val
305 310 315 320
Pro Phe Asp Lys Lys Ile Ala Thr Thr Asn Lys Val Val Gly Val Asp
325 330 335
Val Asn Ile Lys His Met Leu Leu Ala Thr Asn Ile Leu Asp Asp Gly
340 345 350
Asn Val Asn Gly Tyr Val Asn Ile Tyr Lys Glu Val Ile Asn Asp Ser
355 360 365
Asp Phe Lys Lys Val Cys Asn Ser Thr Val Met Lys Tyr Phe Thr Asp
370 375 380
Phe Ser Lys Phe Val Thr Phe Cys Pro Leu Glu Phe Asp Phe Leu Phe
385 390 395 400
Ser Arg Val Cys Asn Gln Lys Gly Ile Tyr Asn Asp Asn Ser Ala Met
405 410 415
Glu Lys Ser Phe Ser Asp Val Leu Asn Lys Leu Lys Trp Asn Phe Ile
420 425 430
Glu Thr Gly Asp Asn Thr Lys Arg Ile Tyr Ile Glu Asn Val Met Lys
435 440 445
Leu Arg Ser Gln Met Lys Ala Tyr Ala Ile Val Lys Asn Ala Tyr Tyr
450 455 460
Lys Gln Gln Ser Glu Tyr Asp Phe Gly Lys Ser Glu Glu Phe Ile Gln
465 470 475 480
Glu His Pro Phe Ser Asn Thr Asp Lys Gly Ile Glu Ile Leu Asn Lys
485 490 495
Leu Asp Asn Ile Ser Lys Lys Ile Leu Gly Cys Arg Asn Asn Ile Ile
500 505 510
Gln Tyr Ser Tyr Asn Leu Phe Glu Ile Asn Gly Tyr Asp Met Ile Ser
515 520 525
Leu Glu Lys Leu Thr Ser Ser Gln Phe Lys Lys Lys Pro Phe Pro Thr
530 535 540
Val Asn Ser Leu Leu Lys Tyr His Lys Ile Leu Gly Cys Thr Gln Glu

545 550 555 560
Glu Met Glu Lys Lys Asp Ile Tyr Ser Val Ile Lys Lys Gly Tyr Tyr
565 570 575
Asp Ile Ile Phe Asp Asn Gly Val Val Thr Asp Ala Lys Leu Ser Thr
580 585 590
Lys Gly Glu Leu Ser Lys Phe Lys Asp Asp Phe Phe Asn Leu Met Ile
595 600 605
Lys Ser Ile His Phe Ala Asp Ile Lys Asp Tyr Phe Ile Thr Leu Ser
610 615 620
Asn Asn Gly Thr Ala Gly Val Ser Leu Val Pro Ser Tyr Phe Thr Ser
625 630 635 640
Gln Met Asp Ser Ile Asp His Lys Ile Tyr Phe Val Gln Asp Asn Lys
645 650 655
Ser Gly Lys Leu Lys Leu Ala Asn Lys His Lys Val Arg Ser Ser Gln
660 665 670
Glu Lys His Ile Asn Gly Leu Asn Ala Asp Tyr Asn Ala Ala Arg Asn
675 680 685
Ile Ala Tyr Ile Met Glu Asn Thr Glu Cys Arg Asn Met Phe Met Lys
690 695 700
Gln Ser Arg Thr Asp Lys Ser Leu Tyr Asn Lys Pro Ser Tyr Glu Thr
705 710 715 720
Phe Ile Lys Thr Gln Gly Ser Ala Val Ala Lys Leu Lys Lys Glu Gly
725 730 735
Phe Val Lys Ile Leu Asp Glu Ala Ser Val
740 745

<210> 119
<211> 680
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 119
Glu Lys Lys Asn Ser Gly Tyr Ala Leu Thr Leu Ile Ser Asp Glu Trp
1 5 10 15
Lys Asp Lys Pro Met Tyr Met Met Phe Lys Lys Gly Tyr Pro Ala Asn
20 25 30
Asn Arg Asp Asn Ala Ile Tyr Glu Thr Leu Asn Thr Cys Asn Thr Glu
35 40 45
His Tyr Thr Gly Asn Ile Leu Asn Phe Ser Asp Thr Tyr Tyr Arg Arg

50 55 60

Phe Gly Tyr Val Ala Ser Ala Ile Ser Asn Tyr Val Thr Lys Ile Ser
65 70 75 80

Lys Met Ser Thr Gly Ser Arg Ser Lys Asn Ile Ser Asn Asp Ser Asp
85 90 95

Val Asp Thr Ile Met Glu Gln Val Ile Tyr Glu Met Glu His Asn Gly
100 105 110

Trp Thr Ser Val Lys Asp Trp Glu Asn Gln Met Glu Tyr Leu Glu Ser
115 120 125

Lys Thr Asp Ser Asn Pro Asn Phe Val Tyr Arg Met Thr Thr Leu Tyr
130 135 140

Glu Phe Tyr Lys Ser His Ile Asp Glu Val Asn Ser Lys Met Glu Thr
145 150 155 160

Met Ser Ile Asp Ser Leu Ile Lys Phe Gly Gly Cys Arg Arg Lys Asp
165 170 175

Ser Lys Lys Ser Met Tyr Ile Met Gly Gly Ser Asn Thr Pro Phe Asp
180 185 190

Ile Thr Gln Ile Gly Gly Asn Ser Leu Asn Ile Lys Phe Ser Lys Asn
195 200 205

Leu Asn Val Asp Val Phe Gly Arg Tyr Asp Val Ile Lys Asp Asn Thr
210 215 220

Leu Leu Val Asp Ile Ile Asn Glu His Gly Ala Ser Phe Val Leu Lys
225 230 235 240

Ile Ile Asn Asp Glu Ile Tyr Ile Asp Ile Asn Val Ser Val Pro Phe
245 250 255

Asp Lys Lys Ile Ala Thr Thr Asn Lys Val Val Gly Val Asp Val Asn
260 265 270

Ile Lys His Met Leu Leu Ala Thr Asn Ile Leu Asp Asp Gly Asn Val
275 280 285

Asn Gly Tyr Val Asn Ile Tyr Lys Glu Val Ile Asn Asp Ser Asp Phe
290 295 300

Lys Lys Val Cys Asn Ser Thr Val Met Lys Tyr Phe Thr Asp Phe Ser
305 310 315 320

Lys Phe Val Thr Phe Cys Pro Leu Glu Phe Asp Phe Leu Phe Ser Arg
325 330 335

Val Cys Asn Gln Lys Gly Ile Tyr Asn Asp Asn Ser Ala Met Glu Lys
340 345 350

Ser Phe Ser Asp Val Leu Asn Lys Leu Lys Trp Asn Phe Ile Glu Thr
355 360 365

Gly Asp Asn Thr Lys Arg Ile Tyr Ile Glu Asn Val Met Lys Leu Arg
370 375 380
Ser Gln Met Lys Ala Tyr Ala Ile Val Lys Asn Ala Tyr Tyr Lys Gln
385 390 395 400
Gln Ser Glu Tyr Asp Phe Gly Lys Ser Glu Glu Phe Ile Gln Glu His
405 410 415
Pro Phe Ser Asn Thr Asp Lys Gly Ile Glu Ile Leu Asn Lys Leu Asp
420 425 430
Asn Ile Ser Lys Lys Ile Leu Gly Cys Arg Asn Asn Ile Ile Gln Tyr
435 440 445
Ser Tyr Asn Leu Phe Glu Ile Asn Gly Tyr Asp Met Ile Ser Leu Glu
450 455 460
Lys Leu Thr Ser Ser Gln Phe Lys Lys Lys Pro Phe Pro Thr Val Asn
465 470 475 480
Ser Leu Leu Lys Tyr His Lys Ile Leu Gly Cys Thr Gln Glu Glu Met
485 490 495
Glu Lys Lys Asp Ile Tyr Ser Val Ile Lys Lys Gly Tyr Tyr Asp Ile
500 505 510
Ile Phe Asp Asn Asp Val Val Thr Asp Ala Lys Leu Ser Thr Lys Gly
515 520 525
Glu Leu Ser Lys Phe Lys Asp Asp Phe Phe Asn Leu Met Ile Lys Ser
530 535 540
Ile His Phe Ala Asp Ile Lys Asp Tyr Phe Ile Thr Leu Ser Asn Asn
545 550 555 560
Gly Thr Val Gly Val Ser Leu Val Pro Ser Tyr Phe Thr Ser Gln Met
565 570 575
Asp Ser Ile Asp His Lys Ile Tyr Phe Val Gln Asp Asn Lys Ser Gly
580 585 590
Lys Leu Lys Leu Ala Asn Lys His Lys Val Arg Ser Ser Gln Glu Lys
595 600 605
His Ile Asn Gly Leu Asn Ala Asp Tyr Asn Ala Ala Arg Asn Ile Ala
610 615 620
Tyr Ile Met Glu Asn Asn Glu Cys Arg Asn Met Phe Met Lys Gln Ser
625 630 635 640
Arg Thr Asp Lys Ser Leu Tyr Asn Lys Pro Ser Tyr Glu Thr Phe Ile
645 650 655
Lys Thr Gln Gly Ser Ala Val Ala Lys Leu Lys Lys Glu Gly Phe Val
660 665 670
Lys Ile Leu Asp Glu Ala Ser Val

675 680

<210> 120
<211> 516
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 120

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

245 250 255
Pro Phe Ser Leu Thr Glu Lys Gly Met Ser Ile Lys Ser Lys Met Asp
260 265 270
Asn Ile Cys Arg Thr Ile Ile Gly Cys Arg Asn Asn Ile Ile Asp Tyr
275 280 285
Ala Tyr Ser Phe Phe Glu Arg Asn Asp Tyr Ser Ile Ile Gly Leu Glu
290 295 300
Lys Leu Thr Ser Ser Gln Phe Glu Lys Thr Lys Ser Leu Pro Thr Cys
305 310 315 320
Lys Ser Leu Leu Asn Phe His Lys Val Leu Gly His Thr Leu Ser Glu
325 330 335
Leu Glu Thr Leu Pro Ile Asn Asp Val Val Lys Lys Gly Tyr Tyr Thr
340 345 350
Phe Thr Thr Asp Asn Glu Gly Arg Ile Thr Asp Ala Ser Leu Ser Glu
355 360 365
Lys Gly Lys Val Arg Lys Met Lys Asp Asp Phe Phe Asn Gln Ala Ile
370 375 380
Lys Ala Ile His Phe Ala Asp Val Lys Asp Tyr Phe Ala Thr Leu Ser
385 390 395 400
Asn Asn Gly Gln Thr Gly Ile Phe Phe Val Pro Ser Gln Phe Thr Ser
405 410 415
Gln Met Asp Ser Asn Thr His Thr Leu Tyr Phe Glu Asn Ala Lys Asn
420 425 430
Gly Gly Leu Lys Leu Ala Ser Lys Tyr Gln Val Arg Gln Thr Gln Glu
435 440 445
Tyr His Leu Asn Gly Leu Pro Ala Asp Tyr Asn Ala Ala Arg Asn Ile
450 455 460
Ala Tyr Ile Gly Leu Asp Glu Thr Met Arg Asn Thr Phe Leu Lys Lys
465 470 475 480
Ala Asn Ser Asn Lys Ser Leu Tyr Asn Gln Pro Ile Tyr Asp Thr Gly
485 490 495
Ile Lys Lys Thr Ala Gly Val Phe Ser Arg Met Lys Lys Leu Lys Arg
500 505 510
Tyr Glu Ile Ile
515

<210> 121
<211> 585
<212> PRT
<213> 人工序列(Artificial sequence)

<220>

<223> 合成序列(synthetic sequence)

<400> 121

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

Arg Ser Gln Met Lys Ala Tyr Ala Ile Val Lys Asn Ala Tyr Tyr Lys
290 295 300
Gln Gln Ser Glu Tyr Asp Phe Gly Lys Ser Glu Glu Phe Ile Gln Glu
305 310 315 320
His Pro Phe Ser Asn Thr Asp Lys Gly Ile Glu Ile Leu Asn Lys Leu
325 330 335
Asp Asn Ile Ser Lys Lys Ile Leu Gly Cys Arg Asn Asn Ile Ile Gln
340 345 350
Tyr Ser Tyr Asn Leu Phe Glu Ile Asn Gly Tyr Asp Met Ile Ser Leu
355 360 365
Glu Lys Leu Thr Ser Ser Gln Phe Lys Lys Lys Pro Phe Pro Thr Val
370 375 380
Asn Ser Leu Leu Lys Tyr His Lys Ile Leu Gly Cys Thr Gln Glu Glu
385 390 395 400
Met Glu Lys Lys Asp Ile Tyr Ser Val Ile Lys Lys Gly Tyr Tyr Asp
405 410 415
Ile Ile Phe Asp Asn Asp Val Val Thr Asp Ala Lys Leu Ser Thr Lys
420 425 430
Gly Glu Leu Ser Lys Phe Lys Asp Asp Phe Phe Asn Leu Met Ile Lys
435 440 445
Ser Ile His Phe Ala Asp Ile Lys Asp Tyr Phe Ile Thr Leu Ser Asn
450 455 460
Asn Gly Thr Ala Gly Val Ser Leu Val Pro Ser Tyr Phe Thr Ser Gln
465 470 475 480
Met Asp Ser Ile Asp His Lys Ile Tyr Phe Val Gln Asp Asn Lys Ser
485 490 495
Gly Lys Leu Lys Leu Ala Asn Lys His Lys Val Arg Ser Ser Gln Glu
500 505 510
Lys His Ile Asn Gly Leu Asn Ala Asp Tyr Asn Ala Ala Arg Asn Ile
515 520 525
Ala Tyr Ile Met Glu Asn Thr Glu Cys Arg Asn Met Phe Met Lys Gln
530 535 540
Ser Arg Thr Asp Lys Ser Leu Tyr Asn Lys Pro Ser Tyr Glu Thr Phe
545 550 555 560
Ile Lys Thr Gln Gly Ser Ala Val Ala Lys Leu Lys Lys Glu Gly Phe
565 570 575
Val Lys Ile Ile Asp Glu Ala Ser Val
580 585

<210> 122

<211> 746
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 122

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

260 265 270

Lys Asn Leu Asn Val Asp Val Phe Gly Arg Tyr Asp Val Ile Lys Asp

275 280 285

Asn Thr Leu Leu Val Asp Ile Ile Asn Glu His Gly Ala Ser Phe Val

290 295 300

Leu Lys Ile Ile Asn Asp Glu Ile Tyr Ile Asp Ile Asn Val Ser Val

305 310 315 320

Pro Phe Asp Lys Lys Ile Ala Thr Thr Asn Lys Val Val Gly Val Asp

325 330 335

Val Asn Ile Lys His Met Leu Leu Ala Thr Asn Ile Leu Asp Asp Gly

340 345 350

Asn Val Asn Gly Tyr Val Asn Ile Tyr Lys Glu Val Ile Asn Asp Ser

355 360 365

Asp Phe Lys Lys Val Cys Asn Ser Thr Val Met Lys Tyr Phe Thr Asp

370 375 380

Phe Ser Lys Phe Val Thr Phe Cys Pro Leu Glu Phe Asp Phe Leu Phe

385 390 395 400

Ser Arg Val Cys Asn Gln Lys Gly Ile Tyr Asn Asp Asn Ser Ala Met

405 410 415

Glu Lys Ser Phe Ser Asp Val Leu Asn Lys Leu Lys Trp Asn Phe Ile

420 425 430

Glu Thr Gly Asp Asn Thr Lys Arg Ile Tyr Ile Glu Asn Val Met Lys

435 440 445

Leu Arg Ser Gln Met Lys Ala Tyr Ala Ile Val Lys Asn Ala Tyr Tyr

450 455 460

Lys Gln Gln Ser Glu Tyr Asp Phe Gly Lys Ser Glu Glu Phe Ile Gln

465 470 475 480

Glu His Pro Phe Ser Asn Thr Asp Lys Gly Ile Glu Ile Leu Asn Lys

485 490 495

Leu Asp Asn Ile Ser Lys Lys Ile Leu Gly Cys Arg Asn Asn Ile Ile

500 505 510

Gln Tyr Ser Tyr Asn Leu Phe Glu Ile Asn Gly Tyr Asp Met Ile Ser

515 520 525

Leu Glu Lys Leu Thr Ser Ser Gln Phe Glu Lys Lys Pro Phe Pro Thr

530 535 540

Val Asn Ser Leu Leu Lys Tyr His Lys Ile Leu Gly Cys Thr Gln Glu

545 550 555 560

Glu Met Glu Lys Lys Asp Ile Tyr Ser Val Ile Lys Lys Gly Tyr Tyr

565 570 575

Asp Ile Ile Phe Asp Asn Asp Val Val Thr Asp Ala Lys Leu Ser Thr
580 585 590
Lys Gly Glu Leu Ser Lys Phe Lys Asp Asp Phe Phe Asn Leu Met Ile
595 600 605
Lys Ser Ile His Phe Ala Asp Ile Lys Asp Tyr Phe Ile Thr Leu Ser
610 615 620
Asn Asn Gly Thr Ala Gly Val Ser Leu Val Pro Ser Tyr Phe Thr Ser
625 630 635 640
Gln Met Asp Ser Ile Asp His Lys Ile Tyr Phe Val Gln Asp Asn Lys
645 650 655
Ser Gly Lys Leu Lys Leu Ala Asn Lys His Lys Val Arg Ser Ser Gln
660 665 670
Glu Lys His Ile Asn Gly Leu Asn Ala Asp Tyr Asn Ala Ala Arg Asn
675 680 685
Ile Ala Tyr Ile Met Glu Asn Thr Glu Cys Arg Asn Met Phe Met Lys
690 695 700
Gln Ser Arg Thr Asp Lys Ser Leu Tyr Asn Lys Pro Ser Tyr Glu Thr
705 710 715 720
Phe Ile Lys Thr Gln Gly Ser Ala Val Ala Lys Leu Lys Lys Glu Gly
725 730 735
Phe Val Lys Ile Leu Asp Glu Ala Ser Val
740 745

<210> 123
<211> 585
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 123
Asp Val Asp Thr Ile Met Glu Gln Val Ile Tyr Glu Met Glu His Asn
1 5 10 15
Gly Trp Thr Ser Val Lys Asp Trp Glu Asn Gln Met Glu Tyr Leu Glu
20 25 30
Ser Lys Thr Asp Ser Asn Pro Asn Phe Val Tyr Arg Met Thr Thr Leu
35 40 45
Tyr Glu Phe Tyr Lys Ser His Ile Asp Glu Val Asn Ser Lys Met Glu
50 55 60
Thr Met Ser Ile Asp Ser Leu Ile Lys Phe Gly Gly Cys Arg Arg Lys
65 70 75 80

Asp Ser Lys Lys Ser Met Tyr Ile Met Gly Gly Ser Asn Thr Pro Phe
85 90 95
Asp Ile Thr Gln Ile Gly Gly Asn Ser Leu Asn Ile Lys Phe Ser Lys
100 105 110
Asn Leu Asn Val Asp Val Phe Gly Arg Tyr Asp Val Ile Lys Asp Asn
115 120 125
Thr Leu Leu Val Asp Ile Ile Asn Glu His Gly Ala Ser Phe Val Leu
130 135 140
Lys Ile Ile Asn Asp Glu Ile Tyr Ile Asp Ile Asn Val Ser Val Pro
145 150 155 160
Phe Asp Lys Lys Ile Ala Thr Thr Asn Lys Val Val Gly Val Asp Val
165 170 175
Asn Ile Lys His Met Leu Leu Ala Thr Asn Ile Leu Asp Asp Gly Asn
180 185 190
Val Asn Gly Tyr Val Asn Ile Tyr Lys Glu Val Ile Asn Asp Ser Asp
195 200 205
Phe Lys Lys Val Cys Asn Ser Thr Val Met Lys Tyr Phe Thr Asp Phe
210 215 220
Ser Lys Phe Val Thr Phe Cys Pro Leu Glu Phe Asp Phe Leu Phe Ser
225 230 235 240
Arg Val Cys Asn Gln Lys Gly Ile Tyr Asn Asp Asn Ser Ala Met Glu
245 250 255
Lys Ser Phe Ser Asp Val Leu Asn Lys Leu Lys Trp Asn Phe Ile Glu
260 265 270
Thr Gly Asp Asn Thr Lys Arg Ile Tyr Ile Glu Asn Val Met Lys Leu
275 280 285
Arg Ser Gln Met Lys Ala Tyr Ala Ile Val Lys Asn Ala Tyr Tyr Lys
290 295 300
Gln Gln Ser Glu Tyr Asp Phe Gly Lys Ser Glu Glu Phe Ile Gln Glu
305 310 315 320
His Pro Phe Ser Asn Thr Asp Lys Gly Ile Glu Ile Leu Asn Lys Leu
325 330 335
Asp Asn Ile Ser Lys Lys Ile Leu Gly Cys Arg Asn Asn Ile Ile Gln
340 345 350
Tyr Ser Tyr Asn Leu Phe Glu Ile Asn Gly Tyr Asp Met Ile Ser Leu
355 360 365
Glu Lys Leu Thr Ser Ser Gln Phe Lys Lys Lys Pro Phe Pro Thr Val
370 375 380
Asp Ser Leu Leu Lys Tyr His Lys Ile Leu Gly Cys Thr Gln Glu Glu

385 390 395 400
Met Glu Lys Lys Asp Ile Tyr Ser Val Ile Lys Lys Gly Tyr Tyr Asp
405 410 415
Ile Ile Phe Asp Asn Asp Val Val Thr Asp Ala Lys Leu Ser Thr Lys
420 425 430
Gly Glu Leu Ser Lys Phe Lys Asp Asp Phe Phe Asn Leu Met Ile Lys
435 440 445
Ser Ile His Phe Ala Asp Ile Lys Asp Tyr Phe Ile Thr Leu Ser Asn
450 455 460
Asn Gly Thr Ala Gly Val Ser Leu Val Pro Ser Tyr Phe Thr Ser Gln
465 470 475 480
Met Asp Ser Ile Asp His Lys Ile Tyr Phe Val Gln Asp Asn Lys Ser
485 490 495
Gly Lys Leu Lys Leu Ala Asn Lys His Lys Val Arg Ser Ser Gln Glu
500 505 510
Lys His Ile Asn Gly Leu Asn Ala Asp Tyr Asn Ala Ala Arg Asn Ile
515 520 525
Ala Tyr Ile Met Glu Asn Thr Asp Cys Arg Asn Met Phe Met Lys Gln
530 535 540
Ser Arg Thr Asp Lys Ser Leu Tyr Asn Lys Pro Ser Tyr Glu Thr Phe
545 550 555 560
Ile Lys Thr Gln Gly Ser Ala Val Ala Lys Leu Lys Lys Glu Gly Phe
565 570 575
Val Lys Ile Leu Asp Glu Ala Ser Val
580 585

<210> 124
<211> 758
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 124
Phe Val Leu Ser Leu Tyr Gln Leu Asn Gln Tyr Ile Met Ala Ser His
1 5 10 15
Lys Lys Thr Glu Ser Asn Gln Ile Ile Lys Thr Phe Ser Phe Lys Ile
20 25 30
Lys Asn Ala Asn Gly Leu Ser Leu Asp Val Leu Asn Asp Ala Ile Thr
35 40 45
Glu Tyr Gln Asn Tyr Tyr Asn Ile Cys Ser Asp Trp Ile Lys Asp His

50 55 60

Leu Thr Met Lys Ile Gly Glu Leu Tyr Lys Tyr Ile Pro Asp Glu Lys
65 70 75 80
Lys Asn Ser Gly Tyr Ala Leu Thr Leu Ile Ser Asp Glu Trp Lys Asp
85 90 95
Lys Pro Met Tyr Met Met Phe Lys Lys Gly Tyr Pro Ala Asn Asn Arg
100 105 110
Asp Asn Ala Ile Tyr Glu Thr Leu Asn Thr Cys Asn Thr Glu His Tyr
115 120 125
Thr Gly Asn Ile Leu Asn Phe Ser Asp Thr Tyr Tyr Arg Arg Phe Gly
130 135 140
Tyr Val Ala Ser Thr Ile Ser Asn Tyr Val Thr Lys Ile Ser Lys Met
145 150 155 160
Ser Thr Gly Ser Arg Ser Lys Asn Ile Ser Asn Asp Ser Asp Val Asp
165 170 175
Thr Ile Met Glu Gln Val Ile Tyr Glu Met Glu His Asn Gly Trp Thr
180 185 190
Ser Val Lys Asp Trp Glu Asn Gln Met Glu Tyr Leu Glu Ser Lys Thr
195 200 205
Asp Ser Asn Pro Asn Phe Val Tyr Arg Met Thr Thr Leu Tyr Glu Phe
210 215 220
Tyr Lys Ser His Ile Asp Glu Val Asn Ser Lys Met Glu Thr Met Ser
225 230 235 240
Ile Asp Ser Leu Ile Lys Phe Gly Gly Cys Arg Arg Lys Asp Ser Lys
245 250 255
Lys Ser Met Tyr Ile Met Gly Gly Ser Asn Thr Pro Phe Asp Ile Thr
260 265 270
Gln Ile Gly Gly Asn Ser Leu Asn Ile Lys Phe Ser Lys Asn Leu Asn
275 280 285
Val Asp Val Phe Gly Arg Tyr Asp Val Ile Lys Asp Asn Thr Leu Leu
290 295 300
Val Asp Ile Ile Asn Gly His Gly Ala Ser Phe Val Leu Lys Ile Ile
305 310 315 320
Asn Gly Glu Ile Tyr Ile Asp Ile Asn Val Ser Val Pro Phe Asp Lys
325 330 335
Lys Ile Ala Thr Thr Asn Lys Val Val Gly Val Asp Val Asn Ile Lys
340 345 350
His Met Leu Leu Ala Thr Asn Ile Leu Asp Asp Gly Asn Val Asn Gly
355 360 365

Tyr Val Asn Ile Tyr Lys Glu Val Ile Asn Asp Ser Asp Phe Lys Lys
370 375 380
Val Cys Asn Ser Thr Val Met Lys Tyr Phe Thr Asp Phe Ser Lys Phe
385 390 395 400
Val Thr Phe Cys Pro Leu Glu Phe Asp Phe Leu Phe Ser Arg Val Cys
405 410 415
Asn Gln Lys Gly Ile Tyr Asn Asp Asn Ser Ala Met Glu Lys Ser Phe
420 425 430
Ser Asp Val Leu Asn Lys Leu Lys Trp Asn Phe Ile Glu Thr Gly Asp
435 440 445
Asn Thr Lys Arg Ile Tyr Ile Glu Asn Val Met Lys Leu Arg Ser Gln
450 455 460
Met Lys Ala Tyr Ala Ile Val Lys Asn Ala Tyr Tyr Lys Gln Gln Ser
465 470 475 480
Glu Tyr Asp Phe Gly Lys Ser Glu Glu Phe Ile Gln Glu His Pro Phe
485 490 495
Ser Asn Thr Asp Lys Gly Ile Glu Ile Leu Asn Lys Leu Asp Asn Ile
500 505 510
Ser Lys Lys Ile Leu Gly Cys Arg Asn Asn Ile Ile Gln Tyr Ser Tyr
515 520 525
Asn Leu Phe Glu Ile Asn Gly Tyr Asp Met Ile Ser Leu Glu Lys Leu
530 535 540
Thr Ser Ser Gln Phe Lys Lys Lys Ser Phe Pro Thr Val Asp Ser Leu
545 550 555 560
Leu Lys Tyr His Lys Ile Leu Gly Cys Thr Gln Glu Glu Met Glu Lys
565 570 575
Lys Asp Ile Tyr Ser Val Ile Lys Lys Gly Tyr Tyr Asp Ile Ile Phe
580 585 590
Asp Asn Asp Val Val Thr Asp Ala Lys Leu Ser Thr Lys Gly Glu Leu
595 600 605
Ser Lys Phe Lys Asp Asp Phe Phe Asn Leu Met Ile Lys Ser Ile His
610 615 620
Phe Ala Asp Ile Lys Asp Tyr Phe Ile Thr Leu Ser Asn Asn Gly Thr
625 630 635 640
Ala Gly Val Ser Leu Val Pro Ser Tyr Phe Thr Ser Gln Met Asp Ser
645 650 655
Ile Asp His Lys Ile Tyr Phe Val Gln Asp Asn Lys Ser Gly Lys Leu
660 665 670
Lys Leu Ala Asn Lys His Lys Val Arg Ser Ser Gln Glu Lys His Ile

675 680 685

Asn Gly Leu Asn Ala Asp Tyr Asn Ala Ala Arg Asn Ile Ala Tyr Ile
690 695 700

Met Glu Asn Asn Glu Cys Arg Asn Met Phe Met Lys Gln Ser Arg Thr
705 710 715 720

Asp Lys Ser Leu Tyr Asn Lys Pro Ser Tyr Glu Thr Phe Ile Lys Thr
725 730 735

Gln Gly Ser Ala Val Ala Lys Leu Lys Lys Glu Gly Phe Val Lys Ile
740 745 750

Leu Asp Glu Ala Ser Val
755

<210> 125
<211> 596
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 125

Gly Ser Arg Ser Lys Asn Ile Ser Asn Asp Ser Asp Val Asp Thr Ile
1 5 10 15

Met Glu Gln Val Ile Tyr Glu Met Glu His Asn Gly Trp Thr Ser Val
20 25 30

Lys Asp Trp Glu Asn Gln Met Glu Tyr Leu Glu Ser Lys Thr Asp Ser
35 40 45

Asn Pro Asn Phe Val Tyr Arg Met Thr Thr Leu Tyr Glu Phe Tyr Lys
50 55 60

Ser His Ile Asp Glu Val Asn Ser Lys Met Glu Thr Met Ser Ile Asp
65 70 75 80

Ser Leu Ile Lys Phe Gly Gly Cys Arg Arg Lys Asp Ser Lys Lys Ser
85 90 95

Met Tyr Ile Met Gly Gly Ser Asn Thr Pro Phe Asp Ile Thr Gln Ile
100 105 110

Gly Gly Asn Ser Leu Asn Ile Lys Phe Ser Lys Asn Leu Asn Val Asp
115 120 125

Val Phe Gly Arg Tyr Asp Val Ile Lys Asp Asn Thr Leu Leu Val Asp
130 135 140

Ile Ile Asn Gly His Gly Ala Ser Phe Val Leu Lys Ile Ile Asn Gly
145 150 155 160

Glu Ile Tyr Ile Asp Ile Asn Val Ser Val Pro Phe Asp Lys Lys Ile

165 170 175
Ala Thr Thr Asn Lys Val Val Gly Val Asp Val Asn Ile Lys His Met
180 185 190
Leu Leu Ala Thr Asn Ile Leu Asp Asp Gly Asn Val Asn Gly Tyr Val
195 200 205
Asn Ile Tyr Lys Glu Val Ile Asn Asp Ser Asp Phe Lys Lys Val Cys
210 215 220
Asn Ser Thr Val Met Lys Tyr Phe Thr Asp Phe Ser Lys Phe Val Thr
225 230 235 240
Phe Cys Pro Leu Glu Phe Asp Phe Leu Phe Ser Arg Val Cys Asn Gln
245 250 255
Lys Gly Ile Tyr Asn Asp Asn Ser Ala Met Glu Lys Ser Phe Ser Asp
260 265 270
Val Leu Asn Lys Leu Lys Trp Asn Phe Ile Glu Thr Gly Asp Asn Thr
275 280 285
Lys Arg Ile Tyr Ile Glu Asn Val Met Lys Leu Arg Ser Gln Met Lys
290 295 300
Ala Tyr Ala Ile Val Lys Asn Ala Tyr Tyr Lys Gln Gln Ser Glu Tyr
305 310 315 320
Asp Phe Gly Lys Ser Glu Glu Phe Ile Gln Glu His Pro Phe Ser Asn
325 330 335
Thr Asp Lys Gly Ile Glu Ile Leu Asn Lys Leu Asp Asn Ile Ser Lys
340 345 350
Lys Ile Leu Gly Cys Arg Asn Asn Ile Ile Gln Tyr Ser Tyr Asn Leu
355 360 365
Phe Glu Ile Asn Gly Tyr Asp Met Ile Ser Leu Glu Lys Leu Thr Ser
370 375 380
Ser Gln Phe Lys Lys Lys Ser Phe Pro Thr Val Asp Ser Leu Leu Lys
385 390 395 400
Tyr His Lys Ile Leu Gly Cys Thr Gln Glu Glu Met Glu Lys Lys Asp
405 410 415
Ile Tyr Ser Val Ile Lys Lys Gly Tyr Tyr Asp Ile Ile Phe Asp Asn
420 425 430
Asp Val Val Thr Asp Ala Lys Leu Ser Thr Lys Gly Glu Leu Ser Lys
435 440 445
Phe Lys Asp Asp Phe Phe Asn Leu Met Ile Lys Ser Ile His Phe Ala
450 455 460
Asp Ile Lys Asp Tyr Phe Ile Thr Leu Ser Asn Asn Gly Thr Ala Gly
465 470 475 480

Val Ser Leu Val Pro Ser Tyr Phe Thr Ser Gln Met Asp Ser Ile Asp
485 490 495
His Lys Ile Tyr Phe Val Gln Asp Asn Lys Ser Gly Lys Leu Lys Leu
500 505 510
Ala Asn Lys His Lys Val Arg Ser Ser Gln Glu Lys His Ile Asn Gly
515 520 525
Leu Asn Ala Asp Tyr Asn Ala Ala Arg Asn Ile Ala Tyr Ile Met Glu
530 535 540
Asn Asn Glu Cys Arg Asn Met Phe Met Lys Gln Ser Arg Thr Asp Lys
545 550 555 560
Ser Leu Tyr Asn Lys Pro Ser Tyr Glu Thr Phe Ile Lys Thr Gln Gly
565 570 575
Ser Ala Val Ala Lys Leu Lys Lys Glu Gly Phe Val Lys Ile Leu Asp
580 585 590
Glu Ala Ser Val
595

<210> 126
<211> 640
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 126

Met Ala Ser His Lys Lys Thr Glu Ser Asn Gln Ile Ile Lys Thr Phe
1 5 10 15
Ser Phe Lys Ile Lys Asn Ala Asn Gly Leu Ser Leu Asp Val Leu Asn
20 25 30
Asp Ala Ile Thr Glu Tyr Gln Asn Tyr Tyr Asn Ile Cys Ser Asp Trp
35 40 45
Ile Lys Asp His Leu Thr Met Lys Ile Gly Glu Leu Tyr Lys Tyr Ile
50 55 60
Pro Asp Glu Lys Lys Asn Ser Gly Tyr Ala Leu Thr Leu Ile Ser Asp
65 70 75 80
Glu Trp Lys Asp Lys Pro Met Tyr Met Met Phe Lys Lys Gly Tyr Pro
85 90 95
Ala Asn Asn Arg Asp Asn Ala Ile Tyr Glu Thr Leu Asn Thr Cys Asn
100 105 110
Thr Glu His Tyr Thr Gly Asn Ile Leu Asn Phe Ser Asp Thr Tyr Tyr
115 120 125

Arg Arg Phe Gly Tyr Val Ala Ser Thr Ile Ser Asn Tyr Val Thr Lys
130 135 140
Ile Ser Lys Met Ser Thr Gly Ser Arg Ser Lys Asn Ile Ser Asn Asp
145 150 155 160
Ser Asp Val Asp Thr Ile Met Glu Gln Val Ile Tyr Glu Met Glu His
165 170 175
Asn Gly Trp Thr Ser Val Lys Asp Trp Glu Asn Gln Met Glu Tyr Leu
180 185 190
Glu Ser Lys Thr Asp Ser Asn Pro Asn Phe Val Tyr Arg Met Thr Thr
195 200 205
Leu Tyr Glu Phe Tyr Lys Ser His Ile Asp Glu Val Asn Ser Lys Met
210 215 220
Glu Thr Met Ser Ile Asp Ser Leu Ile Lys Phe Gly Gly Cys Arg Arg
225 230 235 240
Lys Asp Ser Lys Lys Ser Met Tyr Ile Met Gly Gly Ser Asn Thr Pro
245 250 255
Phe Asp Ile Thr Gln Ile Asp Gly Asn Ser Leu Asn Ile Lys Phe Ser
260 265 270
Lys Asn Leu Asn Val Asp Val Phe Gly Arg Tyr Asp Val Ile Lys Asp
275 280 285
Asn Thr Leu Leu Val Asp Ile Ile Asn Gly His Gly Ala Ser Phe Val
290 295 300
Leu Lys Ile Ile Asn Gly Glu Ile Tyr Ile Asp Ile Asn Val Ser Val
305 310 315 320
Pro Phe Asp Lys Lys Ile Ala Thr Thr Asn Lys Val Val Gly Val Asp
325 330 335
Val Asn Ile Lys His Met Leu Leu Ala Thr Asn Ile Leu Asp Asp Gly
340 345 350
Asn Val Asn Gly Tyr Val Asn Ile Tyr Lys Glu Val Ile Asn Asp Ser
355 360 365
Asp Phe Lys Lys Val Cys Asn Ser Thr Val Met Lys Tyr Phe Thr Asp
370 375 380
Phe Ser Lys Phe Val Thr Phe Cys Pro Leu Glu Phe Asp Phe Leu Phe
385 390 395 400
Ser Arg Val Cys Asn Gln Lys Gly Ile Tyr Asn Asp Asn Ser Ala Met
405 410 415
Glu Lys Ser Phe Ser Asp Val Leu Asn Lys Leu Lys Trp Asn Phe Ile
420 425 430
Glu Thr Gly Asp Asn Thr Lys Arg Ile Tyr Ile Glu Asn Val Met Lys

435 440 445

Leu Arg Ser Gln Met Lys Ala Tyr Ala Ile Val Lys Asn Ala Tyr Tyr
450 455 460

Lys Gln Gln Ser Glu Tyr Asp Phe Gly Lys Ser Glu Glu Phe Ile Gln
465 470 475 480

Glu His Pro Phe Ser Asn Thr Asp Lys Gly Ile Glu Ile Leu Asn Lys
485 490 495

Leu Asp Asn Ile Ser Lys Lys Ile Leu Gly Cys Arg Asn Asn Ile Ile
500 505 510

Gln Tyr Ser Tyr Asn Leu Phe Glu Ile Asn Gly Tyr Asp Met Ile Ser
515 520 525

Leu Glu Lys Leu Thr Ser Ser Gln Phe Lys Lys Lys Pro Phe Pro Thr
530 535 540

Val Asn Ser Leu Leu Lys Tyr His Lys Ile Leu Gly Cys Thr Gln Glu
545 550 555 560

Glu Met Glu Lys Lys Asp Ile Tyr Ser Val Ile Lys Lys Gly Tyr Tyr
565 570 575

Asp Ile Ile Phe Asp Asn Gly Val Val Ile Asp Ala Lys Leu Ser Ala
580 585 590

Lys Gly Glu Leu Ser Lys Phe Lys Asp Asp Phe Phe Asn Leu Met Ile
595 600 605

Lys Ser Ile His Phe Ala Asp Ile Lys Asp Tyr Phe Ile Thr Leu Ser
610 615 620

Asn Asn Gly Thr Ala Gly Val Ser Leu Val Asp Leu Phe Cys Pro Arg
625 630 635 640

<210> 127
<211> 759
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 127

Met Ala His Lys Lys Asn Leu Glu Gly Glu Asn Leu Gln Val Lys Thr
1 5 10 15

Ile Cys Leu Lys Ala Asn Leu Ser Lys Glu Glu Val Lys Glu Lys Trp
20 25 30

Leu Pro Val Ile Asn Glu Tyr Asn Val Tyr Tyr Asn Arg Met Ser Asp
35 40 45

Tyr Ile Cys Ser Leu Leu Gly Thr Asn Ile Thr Ile Gly Glu Phe Ala

50 55 60

Glu Gln Leu Ser Ile Glu Lys Arg Lys Asn Gly Tyr Phe Thr Ile Cys
65 70 75 80

Gln Asp Asp Lys Phe Lys Asn Glu Ser Leu Tyr Lys Ile Phe His Lys
85 90 95

Ser Phe Pro Ile Asn His Gly Thr Asn Ile Ile Asn Asn Ile Ile Ser
100 105 110

Glu Lys Asn Ile Asp Gln Tyr Asp Gly Asn Thr Leu Gly Phe Arg Pro
115 120 125

Thr Met Tyr Arg Leu Arg Gly Tyr Val Asp Ser Val Ile Gly Asn Tyr
130 135 140

Arg Thr Thr Ile Arg Thr Ile Lys Pro Ser Val Lys Arg His Lys Ile
145 150 155 160

Ser Val Asp Ser Ser Phe Asp Glu Lys Met Glu Gln Cys Ile Tyr Glu
165 170 175

Ile Gln Lys Gly Asn Leu Lys Thr Val Ser Glu Trp Asn Asn Lys Ile
180 185 190

Asp Tyr Leu Leu Ser Lys Ser Asp Met Asn Pro Leu Thr Ile Asp Arg
195 200 205

Phe Asn Leu Leu Arg Asp Phe Tyr Val Asp Asn Glu Thr Glu Val Asn
210 215 220

Glu Lys Ser Asn Asn Ser Ser Ile Glu Gln Leu Val Lys Phe Gly Gly
225 230 235 240

Cys His Arg Lys Gly Asp Asn Met Thr Leu Ser Leu Thr Glu Ala Asn
245 250 255

Phe Ser Ile Glu Glu Ile Asp Asp Ser Tyr Gly Tyr Leu Leu Thr Leu
260 265 270

His Thr Asp Ser Gly Asp Tyr Lys Ile Pro Leu Met Gly Ser Lys Met
275 280 285

Leu Lys Lys Gly Asp Lys Cys Leu Ile Asp Phe Val Asn Cys Lys Lys
290 295 300

Gly Lys Ser Leu Thr Ala Lys Ile Asp Asn Asp Tyr Asn Leu Tyr Phe
305 310 315 320

His Phe Val Val Tyr Ser Asn Phe Glu Lys Ile Glu Asp Asp Asn Ile
325 330 335

Asn Asn Val Val Gly Val Asp Val Asn Ser Lys His Met Leu Leu Met
340 345 350

Thr Asn Val Ile Asp Asp Asn Ile Asp Gly Tyr Val Asn Ile Tyr Lys
355 360 365

Ala Leu Val Asn Asp Asp Glu Phe Lys Ser Leu Val Thr Lys Ser Glu
370 375 380
Tyr Asp Asp Tyr Val Thr Met Ser Lys Tyr Val Thr Phe Cys Pro Ile
385 390 395 400
Glu Leu Lys Tyr Leu Tyr Ala Arg Tyr Cys Val Gln Lys Asp Tyr Pro
405 410 415
Ile Ser Asn Lys Asp Val Ala Ile Glu Gln Cys Ile Ser Arg Val Ile
420 425 430
Asp Lys Leu Cys Lys Glu Thr Leu Asp Ser Arg Ala Asn Asn Tyr Leu
435 440 445
Cys Met Val Arg Arg Ile Arg His Tyr Tyr Lys Ser Tyr Tyr Val Leu
450 455 460
Lys Met Thr Tyr Tyr Asp Lys Met Ser Glu Tyr Asp Thr Asn Met Glu
465 470 475 480
Tyr Asn Asp Ile Ser Thr Thr Ser Lys Glu Thr Met Asp Gln Arg Arg
485 490 495
Phe Glu Asn Ser Phe Arg Glu Thr Asp Cys Ala Lys Glu Ile Leu Ser
500 505 510
Lys Leu Asp Lys Ile Gly Asn Asp Ile Leu Gly Cys Arg Asn Asn Ile
515 520 525
Leu Thr Tyr Ala Tyr Lys Leu Phe Glu Glu Leu Gly Tyr Asp Thr Ile
530 535 540
Ala Leu Glu Asn Leu Glu Ser Ser Gln Phe Asp Lys Met Lys Ser Leu
545 550 555 560
Pro Ser Cys Gln Ser Met Leu Lys Tyr His Lys Leu Glu Gly Lys Thr
565 570 575
Met Glu Glu Val Met Ser Asn Thr Ser Val Lys Ser Leu Ile Glu Asn
580 585 590
Glu Tyr Tyr Asp Phe Ser Leu Asn Asp Asn Lys Thr Val Glu Asn Ile
595 600 605
Thr Tyr Thr Lys Asn Gly Leu Met Lys Lys Gly Phe Asp Glu Phe Phe
610 615 620
Asn Leu Phe Met Lys Ile Ile His Phe Ala Asp Ile Lys Asp Lys Phe
625 630 635 640
Leu Gln Leu Tyr Asn Asn Gly Ser Val Lys Val Ile Leu Val Pro Ser
645 650 655
Tyr Phe Thr Ser Gln Met Asp Ser Ser Asn His Ser Ile Tyr Met Glu
660 665 670
Lys Ser Lys Asn Asp Lys Leu Val Phe Ala Ser Lys His Lys Val Arg

675 680 685

Lys Thr Gln Glu Thr His Leu Asn Gly Leu Asn Ala Asp Tyr Asn Ala
690 695 700

Ala Cys Asn Ile Ala Tyr Ile Ala Lys Asp Ile Lys Trp Arg Glu Lys
705 710 715 720

Phe Cys Lys Lys Thr Ser Asn Asn Gly Tyr Ser Thr Pro Phe Tyr Asp
725 730 735

Cys Ala Thr Lys Asn Gln Ile Glu Met Val Lys Arg Ile Lys Gln Leu
740 745 750

Asn Ala Ile Lys Met Leu Ala
755

<210> 128
<211> 746
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 128

Met Ala Ser His Lys Lys Thr Glu Ser Asn Gln Ile Ile Lys Thr Phe
1 5 10 15

Ser Phe Lys Ile Lys Asn Ala Asn Gly Leu Ser Leu Asp Val Leu Asn
20 25 30

Asp Ala Ile Thr Glu Tyr Gln Asn Tyr Tyr Asn Ile Cys Ser Asp Trp
35 40 45

Ile Lys Asp His Leu Thr Met Lys Ile Gly Glu Leu Tyr Lys Tyr Ile
50 55 60

Pro Asp Glu Lys Lys Asn Ser Gly Tyr Ala Leu Thr Leu Ile Ser Asp
65 70 75 80

Glu Trp Lys Asp Lys Pro Met Tyr Met Met Phe Lys Lys Gly Tyr Pro
85 90 95

Ala Asn Ser Arg Asp Asn Ala Ile Tyr Glu Ala Leu Asn Thr Cys Asn
100 105 110

Thr Glu His Tyr Thr Gly Asn Ile Leu Asn Phe Ser Asp Thr Tyr Tyr
115 120 125

Arg Arg Phe Gly Tyr Val Ala Ser Thr Ile Ser Asn Tyr Val Thr Lys
130 135 140

Ile Ser Lys Met Ser Thr Gly Ser Arg Ser Lys Asn Ile Ser Asn Asp
145 150 155 160

Ser Asp Val Asp Thr Ile Met Glu Gln Val Ile Tyr Glu Met Glu His

165 170 175
Asn Gly Trp Thr Ser Val Lys Asp Trp Glu Asn Gln Met Glu Tyr Leu
180 185 190
Glu Ser Lys Thr Asp Ser Asn Pro Asn Phe Val Tyr Arg Met Thr Thr
195 200 205
Leu Tyr Glu Phe Tyr Lys Ser His Ile Asp Glu Val Asn Ser Lys Met
210 215 220
Glu Thr Met Ser Ile Asp Ser Leu Ile Lys Phe Gly Gly Cys Arg Arg
225 230 235 240
Lys Asp Ser Lys Lys Ser Met Tyr Ile Met Gly Gly Ser Asn Thr Pro
245 250 255
Phe Asp Ile Thr Gln Ile Gly Gly Asn Ser Leu Asn Ile Lys Phe Ser
260 265 270
Lys Asn Leu Asn Val Asp Val Phe Gly Arg Tyr Asp Val Ile Lys Asp
275 280 285
Asn Thr Leu Leu Val Asp Ile Ile Asn Gly His Gly Ala Ser Phe Val
290 295 300
Leu Lys Ile Ile Asn Asp Glu Ile Tyr Ile Asp Ile Asn Val Ser Val
305 310 315 320
Pro Phe Asp Lys Lys Ile Ala Thr Thr Asn Lys Val Val Gly Val Asp
325 330 335
Val Asn Ile Lys His Met Leu Leu Ala Thr Asn Ile Leu Asp Asp Gly
340 345 350
Asn Val Asn Gly Tyr Val Asn Ile Tyr Lys Glu Val Ile Asn Asp Ser
355 360 365
Asp Phe Lys Lys Val Cys Asn Ser Thr Val Met Lys Tyr Phe Thr Asp
370 375 380
Phe Ser Lys Phe Val Thr Phe Cys Pro Leu Glu Phe Asp Phe Leu Phe
385 390 395 400
Ser Arg Val Cys Asn Gln Lys Gly Ile Tyr Asn Asp Asn Ser Ala Met
405 410 415
Glu Lys Ser Phe Ser Asp Val Leu Asn Lys Leu Lys Trp Asn Phe Ile
420 425 430
Glu Thr Gly Asp Asn Thr Lys Arg Ile Tyr Ile Glu Asn Val Met Lys
435 440 445
Leu Arg Ser Gln Met Lys Ala Tyr Ala Ile Val Lys Asn Ala Tyr Tyr
450 455 460
Lys Gln Gln Ser Glu Tyr Asp Phe Gly Lys Ser Glu Glu Phe Ile Gln
465 470 475 480

Glu His Pro Phe Ser Asn Thr Asp Lys Gly Ile Glu Ile Leu Asn Lys
485 490 495
Leu Asp Asn Ile Ser Lys Lys Ile Leu Gly Cys Arg Asn Asn Ile Ile
500 505 510
Gln Tyr Ser Tyr Asn Leu Phe Glu Ile Asn Gly Tyr Asp Met Ile Ser
515 520 525
Leu Glu Lys Leu Thr Ser Ser Gln Phe Lys Lys Lys Ser Phe Pro Thr
530 535 540
Val Asp Ser Leu Leu Lys Tyr His Lys Ile Leu Gly Cys Thr Gln Glu
545 550 555 560
Glu Met Glu Lys Lys Asp Ile Tyr Ser Val Ile Lys Lys Gly Tyr Tyr
565 570 575
Asp Ile Ile Phe Asp Asn Asp Val Val Thr Asp Ala Lys Leu Ser Thr
580 585 590
Lys Gly Glu Leu Ser Lys Phe Lys Asp Asp Phe Phe Asn Leu Met Ile
595 600 605
Lys Ser Ile His Phe Ala Asp Ile Lys Asp Tyr Phe Ile Thr Leu Ser
610 615 620
Asn Asn Gly Thr Ala Gly Val Ser Leu Val Pro Ser Tyr Phe Thr Ser
625 630 635 640
Gln Met Asp Ser Ile Asp His Lys Ile Tyr Phe Val Gln Asp Asn Lys
645 650 655
Ser Gly Lys Leu Lys Leu Ala Asn Lys His Lys Val Arg Ser Ser Gln
660 665 670
Glu Lys His Ile Asn Gly Leu Asn Ala Asp Tyr Asn Ala Ala Arg Asn
675 680 685
Ile Ala Tyr Ile Met Glu Asn Thr Glu Cys Arg Asn Met Phe Met Lys
690 695 700
Gln Ser Arg Thr Asp Lys Ser Leu Tyr Asn Lys Pro Ser Tyr Glu Thr
705 710 715 720
Phe Ile Lys Thr Gln Gly Ser Ala Val Ala Lys Leu Lys Lys Glu Gly
725 730 735
Phe Val Lys Ile Leu Asp Glu Ala Ser Val
740 745

<210> 129
<211> 746
<212> PRT
<213> 人工序列(Artificial sequence)
<220>

<223> 合成序列(synthetic sequence)

<400> 129

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

290 295 300
Leu Lys Ile Ile Asn Asp Glu Ile Tyr Ile Asp Ile Asn Val Ser Val
305 310 315 320
Pro Phe Asp Lys Lys Ile Ala Thr Thr Asn Lys Val Val Gly Ile Asp
325 330 335
Val Asn Ile Lys His Met Leu Leu Ala Thr Asn Ile Leu Asp Asp Gly
340 345 350
Asn Val Asn Gly Tyr Val Asn Ile Tyr Lys Glu Val Ile Asn Asp Ser
355 360 365
Asp Phe Lys Lys Val Cys Asn Ser Thr Val Met Lys Tyr Phe Thr Asp
370 375 380
Phe Ser Lys Phe Val Thr Phe Cys Pro Leu Glu Phe Asp Phe Leu Phe
385 390 395 400
Ser Arg Val Cys Asn Gln Lys Gly Ile Tyr Asn Asp Asn Ser Ala Met
405 410 415
Glu Lys Ser Phe Ser Asn Val Leu Asn Lys Leu Lys Trp Asn Phe Ile
420 425 430
Glu Thr Gly Asp Asn Thr Lys Arg Ile Tyr Ile Glu Asn Val Met Lys
435 440 445
Leu Arg Ser Gln Met Lys Ala Tyr Ala Ile Val Lys Asn Ala Tyr Tyr
450 455 460
Lys Gln Gln Ser Glu Tyr Asp Phe Gly Lys Ser Glu Glu Phe Ile Gln
465 470 475 480
Glu His Pro Phe Ser Asn Thr Asp Lys Gly Ile Glu Ile Leu Asn Lys
485 490 495
Leu Asp Asn Ile Ser Lys Lys Ile Leu Gly Cys Arg Asn Asn Ile Ile
500 505 510
Gln Tyr Ser Tyr Asn Leu Phe Glu Ile Asn Gly Tyr Asp Met Ile Ser
515 520 525
Leu Glu Lys Leu Thr Ser Ser Gln Phe Lys Lys Lys Pro Phe Pro Thr
530 535 540
Val Asn Ser Leu Leu Lys Tyr His Lys Ile Leu Gly Cys Thr Gln Glu
545 550 555 560
Glu Met Glu Lys Lys Asp Ile Tyr Ser Val Ile Lys Lys Gly Tyr Tyr
565 570 575
Asp Ile Ile Phe Asp Asn Asp Val Val Thr Asp Ala Lys Leu Ser Ala
580 585 590
Lys Gly Glu Leu Ser Lys Phe Lys Asp Asp Phe Phe Asn Leu Met Ile
595 600 605

Lys Ser Ile His Phe Ala Asp Ile Lys Asp Tyr Phe Ile Thr Leu Ser
610 615 620
Asn Asn Gly Thr Ala Gly Val Ser Leu Val Pro Ser Tyr Phe Thr Ser
625 630 635 640
Gln Met Asp Ser Ile Asp His Lys Ile Tyr Phe Val Gln Asp Asn Lys
645 650 655
Ser Gly Lys Leu Lys Leu Ala Asn Lys His Lys Val Arg Ser Ser Gln
660 665 670
Glu Lys His Ile Asn Gly Leu Asn Ala Asp Tyr Asn Ala Ala Arg Asn
675 680 685
Ile Ala Tyr Ile Met Glu Asn Thr Asp Cys Arg Asn Met Phe Met Lys
690 695 700
Gln Ser Arg Thr Asp Lys Ser Leu Tyr Asn Lys Pro Ser Tyr Glu Thr
705 710 715 720
Phe Ile Lys Thr Gln Gly Ser Ala Val Ala Lys Leu Lys Lys Glu Gly
725 730 735
Phe Val Lys Ile Leu Asp Glu Ala Ser Val
740 745

<210> 130
<211> 746
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 130

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

Thr Glu His Tyr Thr Gly Asn Ile Leu Asn Phe Ser Asp Thr Tyr Tyr
115 120 125
Arg Arg Phe Gly Tyr Val Ala Ser Thr Ile Ser Asn Tyr Val Thr Lys
130 135 140
Ile Ser Lys Met Ser Thr Arg Ser Arg Ser Lys Asn Ile Ser Asn Asp
145 150 155 160
Ser Asp Val Asp Thr Ile Met Glu Gln Val Ile Tyr Glu Met Glu His
165 170 175
Asn Gly Trp Thr Ser Val Lys Asp Trp Glu Asn Gln Met Glu Tyr Leu
180 185 190
Glu Ser Lys Thr Asp Ser Asn Pro Asn Phe Val Tyr Arg Met Thr Thr
195 200 205
Leu Tyr Glu Phe Tyr Lys Ser His Ile Asp Glu Val Asn Ser Lys Met
210 215 220
Glu Thr Met Ser Ile Asp Ser Leu Ile Lys Phe Gly Gly Cys Arg Arg
225 230 235 240
Lys Asp Ser Lys Lys Ser Met Tyr Ile Met Gly Gly Ser Asn Thr Pro
245 250 255
Phe Asp Ile Thr Gln Ile Gly Gly Asn Ser Leu Asn Ile Lys Phe Ser
260 265 270
Lys Asn Leu Asn Val Asp Val Phe Gly Arg Tyr Asp Val Ile Lys Asp
275 280 285
Asn Thr Leu Leu Val Asp Ile Ile Asn Gly His Gly Ala Ser Phe Val
290 295 300
Leu Lys Ile Ile Asn Gly Glu Ile Tyr Ile Asp Ile Asn Val Ser Val
305 310 315 320
Pro Phe Asp Lys Lys Ile Ala Thr Thr Asn Lys Val Val Gly Ile Asp
325 330 335
Val Asn Ile Lys His Met Leu Leu Ala Thr Asn Ile Leu Asp Asp Gly
340 345 350
Asn Val Asn Gly Tyr Val Asn Ile Tyr Lys Glu Val Ile Asn Asp Ser
355 360 365
Asp Phe Lys Lys Val Cys Asn Ser Thr Val Met Lys Tyr Phe Thr Asp
370 375 380
Phe Ser Lys Phe Val Thr Phe Cys Pro Leu Glu Phe Asp Phe Leu Phe
385 390 395 400
Ser Arg Val Cys Asn Gln Lys Gly Ile Tyr Asn Asp Asn Ser Ala Met
405 410 415
Glu Lys Ser Phe Ser Asp Val Leu Asn Lys Leu Lys Trp Asn Phe Ile

420 425 430

Glu Thr Gly Asp Asn Thr Lys Arg Ile Tyr Ile Glu Asn Val Met Lys

435 440 445

Leu Arg Ser Gln Met Lys Ala Tyr Ala Ile Val Lys Asn Ala Tyr Tyr

450 455 460

Lys Gln Gln Ser Glu Tyr Asp Phe Gly Lys Ser Glu Glu Phe Ile Gln

465 470 475 480

Glu His Pro Phe Ser Asn Thr Asp Lys Gly Ile Glu Ile Leu Asn Lys

485 490 495

Leu Asp Asn Ile Ser Lys Lys Ile Leu Gly Cys Arg Asn Asn Ile Ile

500 505 510

Gln Tyr Ser Tyr Asn Leu Phe Glu Ile Asn Gly Tyr Asp Met Ile Ser

515 520 525

Leu Glu Lys Leu Thr Ser Ser Gln Phe Lys Lys Lys Pro Phe Pro Thr

530 535 540

Val Asn Ser Leu Leu Lys Tyr His Lys Ile Leu Gly Cys Thr Gln Glu

545 550 555 560

Glu Met Glu Lys Lys Asp Ile Tyr Ser Val Ile Lys Lys Gly Tyr Tyr

565 570 575

Asp Ile Ile Phe Asp Asn Asp Val Val Thr Asp Ala Lys Leu Ser Thr

580 585 590

Lys Gly Glu Leu Ser Lys Phe Lys Asp Asp Phe Phe Asn Leu Met Ile

595 600 605

Lys Ser Ile His Phe Ala Asp Ile Lys Asp Tyr Phe Ile Thr Leu Ser

610 615 620

Asn Asn Gly Thr Ala Gly Val Ser Leu Val Pro Ser Tyr Phe Thr Ser

625 630 635 640

Gln Met Asp Ser Ile Asp His Lys Ile Tyr Phe Val Gln Asp Asn Lys

645 650 655

Ser Gly Lys Leu Lys Leu Ala Asn Lys His Lys Val Arg Ser Ser Gln

660 665 670

Glu Lys His Ile Asn Gly Leu Asn Ala Asp Tyr Asn Ala Ala Arg Asn

675 680 685

Ile Ala Tyr Ile Met Glu Asn Thr Asp Cys Arg Asn Met Phe Met Lys

690 695 700

Gln Ser Arg Thr Asp Lys Ser Leu Tyr Asn Lys Pro Ser Tyr Glu Thr

705 710 715 720

Phe Ile Lys Thr Gln Gly Ser Ala Val Ala Lys Leu Lys Lys Glu Gly

725 730 735

Phe Val Lys Ile Leu Asp Glu Ala Ser Val
740 745

<210> 131
<211> 47
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 131

Asn Met Phe Met Lys Gln Ser Arg Thr Asp Lys Ser Leu Tyr Asn Lys
1 5 10 15
Pro Ser Tyr Glu Thr Phe Ile Lys Thr Gln Gly Ser Ala Val Ala Lys
20 25 30
Leu Lys Lys Glu Gly Phe Val Lys Ile Leu Asp Glu Ala Ser Val
35 40 45

<210> 132
<211> 746
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 132

Met Ala Ser His Lys Lys Thr Glu Ser Asn Gln Ile Ile Lys Thr Phe
1 5 10 15
Ser Phe Lys Ile Lys Asn Ala Asn Gly Leu Ser Leu Asp Val Leu Asn
20 25 30
Asp Ala Ile Thr Glu Tyr Gln Asn Tyr Tyr Asn Ile Cys Ser Asp Trp
35 40 45
Ile Lys Asp His Leu Thr Met Lys Ile Gly Glu Leu Tyr Lys Tyr Ile
50 55 60
Pro Asp Glu Lys Lys Asn Ser Gly Tyr Ala Leu Thr Leu Ile Ser Asp
65 70 75 80
Glu Trp Lys Asp Lys Pro Met Tyr Met Met Phe Lys Lys Gly Tyr Pro
85 90 95
Ala Asn Asn Arg Asp Asn Ala Ile Tyr Glu Thr Leu Asn Thr Cys Asn
100 105 110
Thr Glu His Tyr Thr Gly Asn Ile Leu Asn Phe Ser Asp Thr Tyr Tyr
115 120 125
Arg Arg Phe Gly Tyr Val Ala Ser Thr Ile Ser Asn Tyr Val Thr Lys

130 135 140
Ile Ser Lys Met Ser Thr Gly Ser Arg Ser Lys Asn Ile Ser Asn Asp
145 150 155 160
Ser Asp Val Asp Thr Ile Met Glu Gln Val Ile Tyr Glu Met Glu His
165 170 175
Asn Gly Trp Thr Ser Val Lys Asp Trp Glu Asn Gln Met Glu Tyr Leu
180 185 190
Glu Ser Lys Thr Asp Ser Asn Pro Asn Phe Val Tyr Arg Met Thr Thr
195 200 205
Leu Tyr Glu Phe Tyr Lys Ser His Ile Asp Glu Val Asn Ser Lys Met
210 215 220
Glu Thr Met Ser Ile Asp Ser Leu Ile Lys Phe Gly Gly Cys Arg Arg
225 230 235 240
Lys Asp Ser Lys Lys Ser Met Tyr Ile Met Gly Gly Ser Asn Thr Pro
245 250 255
Phe Asp Ile Thr Gln Ile Gly Gly Asn Phe Leu Asn Ile Lys Phe Ser
260 265 270
Lys Asn Leu Asn Val Asp Val Phe Gly Arg Tyr Asp Val Ile Lys Asp
275 280 285
Asn Thr Leu Leu Val Asp Ile Ile Asn Glu His Gly Ala Ser Phe Val
290 295 300
Leu Lys Ile Ile Asn Asp Glu Ile Tyr Ile Asp Ile Asn Val Ser Val
305 310 315 320
Pro Phe Asp Lys Lys Ile Ala Thr Thr Asn Lys Val Val Gly Val Asp
325 330 335
Val Asn Ile Lys His Met Leu Leu Ala Thr Asn Ile Leu Asp Asp Gly
340 345 350
Asn Val Asn Gly Tyr Val Asn Ile Tyr Lys Glu Val Ile Asn Asp Ser
355 360 365
Asp Phe Lys Lys Val Cys Asn Ser Thr Val Met Lys Tyr Phe Thr Asp
370 375 380
Phe Ser Lys Phe Val Thr Phe Cys Pro Leu Glu Phe Asp Phe Leu Phe
385 390 395 400
Ser Arg Val Cys Asn Gln Lys Gly Ile Tyr Asn Asp Asn Ser Ala Met
405 410 415
Glu Lys Ser Phe Ser Asp Val Leu Asn Lys Leu Lys Trp Asn Phe Ile
420 425 430
Glu Thr Gly Asp Asn Thr Lys Arg Ile Tyr Ile Glu Asn Val Met Lys
435 440 445

Leu Arg Ser Gln Met Lys Ala Tyr Ala Ile Val Lys Asn Ala Tyr Tyr
450 455 460
Lys Gln Gln Ser Glu Tyr Asp Phe Gly Lys Ser Glu Glu Phe Ile Gln
465 470 475 480
Glu His Pro Phe Ser Asn Thr Asp Lys Gly Ile Glu Ile Leu His Lys
485 490 495
Leu Asp Asn Ile Ser Lys Lys Ile Leu Gly Cys Arg Asn Asn Ile Ile
500 505 510
Gln Tyr Ser Tyr Asn Leu Phe Glu Ile Asn Gly Tyr Asp Met Ile Ser
515 520 525
Leu Glu Lys Leu Thr Ser Ser Gln Phe Lys Lys Lys Pro Phe Pro Thr
530 535 540
Val Asn Ser Leu Leu Lys Tyr His Lys Ile Leu Gly Cys Thr Gln Glu
545 550 555 560
Glu Met Glu Lys Lys Asp Ile Tyr Ser Val Ile Lys Lys Gly Tyr Tyr
565 570 575
Asp Ile Ile Phe Asp Asn Asp Val Val Thr Asp Ala Lys Leu Ser Thr
580 585 590
Lys Gly Glu Leu Ser Lys Phe Lys Asp Asp Phe Phe Asn Leu Met Ile
595 600 605
Lys Ser Ile His Phe Ala Asp Ile Lys Asp Tyr Phe Ile Thr Leu Ser
610 615 620
Asn Asn Gly Thr Ala Gly Val Ser Leu Val Pro Ser Tyr Phe Thr Ser
625 630 635 640
Gln Met Asp Ser Ile Asp His Lys Ile Tyr Phe Val Gln Asp Asn Lys
645 650 655
Ser Gly Lys Leu Lys Leu Ala Asn Lys His Lys Val Arg Ser Ser Gln
660 665 670
Glu Lys His Ile Asn Gly Leu Asn Ala Asp Tyr Asn Ala Ala Arg Asn
675 680 685
Ile Ala Tyr Ile Met Glu Asn Thr Glu Cys Arg Asn Met Phe Met Lys
690 695 700
Gln Ser Arg Thr Asp Lys Ser Leu Tyr Asn Lys Pro Ser Tyr Glu Thr
705 710 715 720
Phe Ile Lys Thr Gln Gly Ser Ala Val Ala Lys Leu Lys Lys Glu Gly
725 730 735
Phe Val Lys Ile Leu Asp Glu Ala Ser Val
740 745

<210> 133

<211> 245
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 133

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

<210> 134

<211> 178
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 134

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

<210> 135
<211> 85
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 135

Met Ala Asn Lys His Lys Val Arg Ser Ser Gln Glu Lys His Ile Asn
1 5 10 15
Gly Leu Asn Ala Asp Tyr Asn Ala Ala Arg Asn Ile Ala Tyr Ile Ile

20 25 30

Glu Asn Thr Glu Cys Arg Asn Met Phe Met Lys Gln Ser Arg Thr Asp
35 40 45

Lys Ser Leu Tyr Asn Lys Pro Ser Tyr Glu Thr Phe Ile Lys Thr Gln
50 55 60

Gly Ser Thr Val Ala Lys Leu Lys Lys Glu Gly Phe Val Lys Ile Leu
65 70 75 80

Asp Glu Ala Ser Val
85

<210> 136
<211> 767
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 136

Met Ala His Lys Lys Gln Lys Glu Glu Asn Glu Ile Ile Lys Thr Ile
1 5 10 15

Ser Leu Lys Val Lys Asp Tyr Ala Gly Tyr Pro Ile Val Glu Ala Met
20 25 30

Arg Glu Tyr Thr Lys Tyr Tyr Asn Lys Ile Ser Gln Trp Ile Asn Ser
35 40 45

Asn Leu Leu Thr Ile Lys Ile Gly Glu Leu Ser Ala Phe Met Pro Asp
50 55 60

Glu Cys Lys Thr His Asn Tyr Tyr Thr Tyr Met Met Ser Pro Asp Trp
65 70 75 80

Val Asn Glu Pro Leu Tyr Lys Met Phe Met Lys Gly Phe His Ala Gln
85 90 95

His Cys Asp Asn Ile Leu Phe Asn Val Val Lys Thr Leu Asn Ile Asp
100 105 110

Glu Tyr Ala Gly Asn Ser Leu Gly Leu Ser Ala Ser Cys Phe Arg Arg
115 120 125

Ser Gly Tyr Phe Gln Asn Val Val Ser Asn Tyr Lys Ser Lys Phe Ala
130 135 140

Asn Pro His Ile Ser Ile Arg Arg Lys Asn Leu Ser Asp Leu Pro Thr
145 150 155 160

Glu Asp Glu Leu Val Glu Gln Cys Ile Tyr Glu Ile Gln Asn Gly Leu
165 170 175

Ser Ser Lys Thr Lys Trp Glu Glu Gln Ile Glu Tyr Leu Lys Glu Arg

180 185 190

Asp Asp Ser Lys Gln Ile Tyr Leu Thr Arg Leu Asn Thr Leu Phe Met

195 200 205

Tyr Tyr Lys Ala Asn Lys Asp Phe Val Asp Glu Gln Ile Gln Ile Lys

210 215 220

Ser Val Glu Ser Leu Ala Asn Phe Gly Gly Cys Val Arg Lys Asp Asp

225 230 235 240

Lys Leu Ser Met Asn Leu Val Phe Ser Ser Asn Ser Pro Tyr Lys Val

245 250 255

Val Leu Asn Glu Lys Arg Asn Gly Tyr Ile Leu Ser Tyr Ser Asn Asn

260 265 270

Phe Ser Ile Glu Leu Tyr Gly Asn Arg Met Gly Leu Leu Asn Gly Val

275 280 285

Glu Val Phe Asn Val Gly Asp Lys His Ser Asn Asn Ile Thr Phe Lys

290 295 300

Met Asp Asn Asp Glu Leu Phe Val Asn Ile Pro Val Ser Val Asn Phe

305 310 315 320

Val Lys Lys Ala Asn Glu Thr Asn Lys Val Val Gly Val Asp Val Asn

325 330 335

Leu Lys His Ser Ile Phe Ala Thr Asn Ile Ile Asp Asp Gly Lys Leu

340 345 350

Asp Gly Phe Val Asn Ile Tyr Arg Glu Leu Leu Asn Asp Val Asp Phe

355 360 365

Val Lys Ser Cys Pro Asn Glu Leu Leu Asn Phe Ile Leu Asp Val Glu

370 375 380

Lys Tyr Ala Phe Phe Met Pro Leu Glu Leu Gly Leu Leu Ser Ser Arg

385 390 395 400

Val Met Asn Gln Cys Gly Tyr Ser Thr Ile Gly Lys Tyr Glu Lys Leu

405 410 415

Phe Thr Thr Glu Glu His Phe Phe Arg Val Leu Arg Gln Leu Glu Lys

420 425 430

Arg Phe Gln Glu Ser Gly Glu Asn Gln Lys Arg Ile Tyr Ile Glu Asn

435 440 445

Val Ile Lys Met Arg Ala Gln Val Lys Ala Tyr Phe Thr Leu Lys Tyr

450 455 460

Ala Tyr Asn Lys Ala Asn Lys Asp Tyr Asp Leu Lys Met Gly Phe Val

465 470 475 480

Asp Glu Ser Thr Ala Asn Lys Glu Thr Met Asp Gln Arg Arg Phe Glu

485 490 495

Asn Gln Phe Val Asn Thr Tyr Thr Ala Lys Glu Ile Leu Gly Lys Met
500 505 510
Arg Arg Ile Ala Asn Val Ile Thr Ser Cys Arg Asn Asn Ile Ile Cys
515 520 525
Tyr Met Tyr Lys Ile Phe Glu Asn Asn Gly Tyr Gly Val Val Ala Leu
530 535 540
Glu Lys Leu Gln Ser Ser Gln Met Lys Lys Glu Lys Arg Ile Pro Ser
545 550 555 560
Leu Leu Ser Leu Leu Lys Lys Gln Lys Val Glu Gly Tyr Thr Ile Asn
565 570 575
Glu Leu Lys Asp Lys Ser Val Phe Lys Phe Ile Glu Arg Gly Tyr Tyr
580 585 590
Thr Phe Asp Phe Asp Asp Asp Asn Lys Ile Thr Gly Val Gln Phe Ser
595 600 605
Asp Ala Gly Glu Val Val Asn Met Glu Thr Glu Leu His Asn Leu Ala
610 615 620
Leu Lys Thr Ile His Phe Ala Asp Ala Lys Asp Tyr Phe Val Thr Leu
625 630 635 640
Ser Asn Asn Gly Ser Val Ser Val Ala Leu Val Pro Ser Gln Phe Thr
645 650 655
Ser Gln Met Asp Ser Thr Lys His Val Leu Tyr Ala Lys Lys Asn Asn
660 665 670
Lys Gly Lys Leu Gly Ile Val Ser Glu His Glu Val Arg Pro Lys Gln
675 680 685
Glu Cys His Ile Asn Gly Leu Asn Gly Asp Tyr Asn Ala Ala Cys Asn
690 695 700
Ile Ala Tyr Ile Phe Glu Asn Asp Glu Trp Arg Asn Ala Phe Met Lys
705 710 715 720
Met Asn Pro Asn Glu Tyr Gly Lys Ala Leu Phe Glu Thr Asn Met Glu
725 730 735
Ser Thr Ser Thr Ile Ile Asn Thr Leu Lys Lys Ile Asn Pro Asp Asn
740 745 750
Ile Ile Ser Phe Asp Glu Tyr Glu Lys Thr Lys Lys Val Ala Ala
755 760 765

<210> 137
<211> 652
<212> PRT
<213> 人工序列(Artificial sequence)
<220>

<223> 合成序列(synthetic sequence)

<400> 137

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

290 295 300
Leu Phe Ser Arg Val Cys Asn Gln Lys Gly Ile Tyr Asn Asp Asn Ser
305 310 315 320
Ala Met Glu Lys Ser Phe Ser Asp Val Leu Asn Lys Leu Lys Trp Asn
325 330 335
Phe Ile Glu Thr Gly Asp Asn Thr Lys Arg Ile Tyr Ile Glu Asn Val
340 345 350
Met Lys Leu Arg Ser Gln Met Lys Ala Tyr Ala Ile Val Lys Asn Ala
355 360 365
Tyr Tyr Lys Gln Gln Ser Glu Tyr Asp Phe Gly Lys Ser Glu Glu Phe
370 375 380
Ile Gln Glu His Pro Phe Ser Asn Thr Asp Lys Gly Ile Glu Ile Leu
385 390 395 400
Asn Lys Leu Asp Asn Ile Ser Lys Lys Ile Leu Gly Cys Arg Asn Asn
405 410 415
Ile Ile Gln Tyr Ser Tyr Asn Leu Phe Glu Ile Asn Gly Tyr Asp Met
420 425 430
Ile Ser Leu Glu Lys Leu Thr Ser Ser Gln Phe Lys Lys Lys Ser Phe
435 440 445
Pro Thr Val Asn Ser Leu Leu Lys Tyr His Lys Ile Leu Gly Cys Thr
450 455 460
Gln Glu Glu Met Glu Lys Lys Asp Ile Tyr Ser Val Ile Lys Lys Gly
465 470 475 480
Tyr Tyr Asp Ile Ile Phe Asp Asn Asp Val Val Thr Asp Ala Lys Leu
485 490 495
Ser Thr Lys Gly Glu Leu Ser Lys Phe Lys Asp Asp Phe Phe Asn Leu
500 505 510
Met Ile Lys Ser Ile His Phe Ala Asp Ile Lys Asp Tyr Phe Ile Thr
515 520 525
Leu Ser Asn Asn Gly Thr Ala Gly Val Ser Leu Val Pro Ser Tyr Phe
530 535 540
Thr Ser Gln Met Asp Ser Ile Asp His Lys Ile Tyr Phe Val Gln Asp
545 550 555 560
Asn Lys Ser Gly Lys Leu Lys Leu Ala Asn Lys His Lys Val Arg Ser
565 570 575
Ser Gln Glu Lys His Ile Asn Gly Leu Asn Ala Asp Tyr Asn Ala Ala
580 585 590
Arg Asn Ile Ala Tyr Ile Met Glu Asn Thr Glu Cys Arg Asn Met Phe
595 600 605

Met Lys Gln Ser Arg Thr Asp Lys Ser Leu Tyr Asn Lys Pro Ser Tyr
610 615 620
Glu Thr Phe Ile Lys Thr Gln Gly Ser Ala Val Ala Lys Leu Lys Lys
625 630 635 640
Glu Gly Phe Val Lys Ile Leu Asp Glu Ala Ser Val
645 650

<210> 138
<211> 782
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 138

Met Met Lys Lys Met Arg Ile Ser Pro His Leu Phe Tyr Ile Phe Phe
1 5 10 15
Lys Lys Ile Trp Lys Cys His Phe Phe Val Leu Ser Leu Tyr Gln Leu
20 25 30
Asn Gln Tyr Ile Met Ala Ser His Lys Lys Thr Glu Ser Asn Gln Ile
35 40 45
Ile Lys Thr Phe Ser Phe Lys Ile Lys Asn Ala Asn Gly Leu Ser Leu
50 55 60
Asp Val Leu Asn Asp Ala Ile Thr Glu Tyr Gln Asn Tyr Tyr Asn Ile
65 70 75 80
Cys Ser Asp Trp Ile Lys Asp His Leu Thr Met Lys Ile Gly Glu Leu
85 90 95
Tyr Lys Tyr Ile Pro Asp Glu Lys Lys Asn Ser Gly Tyr Ala Leu Thr
100 105 110
Leu Ile Ser Asp Glu Trp Lys Asp Lys Pro Met Tyr Met Met Phe Lys
115 120 125
Lys Gly Tyr Pro Ala Asn Asn Arg Asp Asn Ala Ile Tyr Glu Thr Leu
130 135 140
Asn Thr Cys Asn Thr Glu His Tyr Thr Gly Asn Ile Leu Asn Phe Ser
145 150 155 160
Asp Thr Tyr Tyr Arg Arg Phe Gly Tyr Val Ala Ser Ala Ile Ser Asn
165 170 175
Tyr Val Thr Lys Ile Ser Lys Met Ser Thr Gly Ser Arg Ser Lys Asn
180 185 190
Ile Ser Asn Asp Ser Asp Val Asp Thr Ile Met Glu Gln Val Ile Tyr
195 200 205

Glu Met Glu His Asn Gly Trp Thr Ser Val Lys Asp Trp Glu Asn Gln
210 215 220
Met Glu Tyr Leu Glu Ser Lys Thr Asp Ser Asn Pro Asn Phe Val Tyr
225 230 235 240
Arg Met Thr Thr Leu Tyr Glu Phe Tyr Lys Ser His Ile Asp Glu Val
245 250 255
Asn Ser Lys Met Glu Thr Met Ser Ile Asp Ser Leu Ile Lys Phe Gly
260 265 270
Gly Cys Arg Arg Lys Asp Ser Lys Lys Ser Met Tyr Ile Met Gly Gly
275 280 285
Ser Asn Thr Pro Phe Asp Ile Thr Gln Ile Gly Gly Asn Ser Leu Asn
290 295 300
Ile Lys Phe Ser Lys Asn Leu Asn Val Asp Val Phe Gly Arg Tyr Asp
305 310 315 320
Val Ile Lys Asp Asn Thr Leu Leu Val Asp Ile Ile Asn Glu His Gly
325 330 335
Ala Ser Phe Val Leu Lys Ile Ile Asn Asp Glu Ile Tyr Ile Asp Ile
340 345 350
Asn Val Ser Val Pro Phe Asp Lys Lys Ile Ala Thr Thr Asn Lys Val
355 360 365
Val Gly Val Asp Val Asn Ile Lys His Met Leu Leu Ala Thr Asn Ile
370 375 380
Leu Asp Asp Gly Asn Val Asn Gly Tyr Val Asn Ile Tyr Lys Glu Val
385 390 395 400
Ile Asn Asp Ser Asp Phe Lys Lys Val Cys Asn Ser Thr Val Met Lys
405 410 415
Tyr Phe Thr Asp Phe Ser Lys Phe Val Thr Phe Cys Pro Leu Glu Phe
420 425 430
Asp Phe Leu Phe Ser Arg Val Cys Asn Gln Lys Gly Ile Tyr Asn Asp
435 440 445
Asn Ser Ala Met Glu Lys Ser Phe Ser Asn Val Leu Asn Lys Leu Lys
450 455 460
Trp Asn Phe Ile Glu Thr Gly Asp Asn Thr Lys Arg Ile Tyr Ile Glu
465 470 475 480
Asn Val Met Lys Leu Arg Ser Gln Met Lys Ala Tyr Ala Ile Val Lys
485 490 495
Asn Ala Tyr Tyr Lys Gln Gln Ser Glu Tyr Asp Phe Gly Lys Ser Glu
500 505 510
Glu Phe Ile Gln Glu His Pro Phe Ser Asn Thr Asp Lys Gly Ile Glu

515 520 525

Ile Leu Asn Lys Leu Asp Asn Ile Ser Lys Lys Ile Leu Gly Cys Arg
530 535 540
Asn Asn Ile Ile Gln Tyr Ser Tyr Asn Leu Phe Glu Ile Asn Gly Tyr
545 550 555 560
Asp Met Ile Ser Leu Glu Lys Leu Thr Ser Ser Gln Phe Lys Lys Lys
565 570 575
Pro Phe Pro Thr Val Asn Ser Leu Leu Lys Tyr His Lys Ile Leu Gly
580 585 590
Cys Thr Gln Glu Glu Met Glu Lys Lys Asp Ile Tyr Ser Val Ile Lys
595 600 605
Lys Gly Tyr Tyr Asp Ile Ile Phe Asp Asn Asp Val Val Thr Asp Ala
610 615 620
Lys Leu Ser Thr Lys Gly Glu Leu Ser Lys Phe Lys Asp Asp Phe Phe
625 630 635 640
Asn Leu Met Ile Lys Ser Ile His Phe Ala Asp Ile Lys Asp Tyr Phe
645 650 655
Ile Thr Leu Ser Asn Asn Gly Thr Ala Gly Val Ser Leu Val Pro Ser
660 665 670
Tyr Phe Thr Ser Gln Met Asp Ser Ile Asp His Lys Ile Tyr Phe Val
675 680 685
Gln Asp Asn Lys Ser Gly Lys Leu Lys Leu Ala Asn Lys His Lys Val
690 695 700
Arg Ser Ser Gln Glu Lys His Ile Asn Gly Leu Asn Ala Asp Tyr Asn
705 710 715 720
Ala Ala Arg Asn Ile Ala Tyr Ile Met Glu Asn Thr Asp Cys Arg Asn
725 730 735
Met Phe Met Lys Gln Ser Arg Thr Asp Lys Ser Leu Tyr Asn Lys Pro
740 745 750
Ser Tyr Glu Thr Phe Ile Lys Thr Gln Gly Ser Ala Val Ala Lys Leu
755 760 765
Lys Lys Glu Gly Phe Val Lys Ile Leu Asp Glu Ala Ser Val
770 775 780

<210> 139
<211> 889
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)

<400> 139

Trp Thr Asp Glu Asp Tyr Cys Lys Phe Phe Ala Lys Tyr Gly Met Ser
1 5 10 15
Glu Glu Cys Gln Lys Trp Met Cys Arg Asp Val Tyr Asp Tyr Arg Ile
20 25 30
Lys Asp Phe Val Asp Tyr Glu Lys Phe Asp Asp Ser Lys Ile Glu Glu
35 40 45
Glu Gln Glu Glu Tyr Ser Glu Ile Pro Val Glu Ser Glu Asp Ser Ser
50 55 60
Pro Ser Thr Ser Glu Thr Leu Pro Lys Ser Ile Tyr Asp Leu Asp Leu
65 70 75 80
Arg Arg Leu Asp Pro Asn Pro Glu Ile Pro Glu Asp Glu Lys Tyr Asn
85 90 95
Glu Glu Asp Leu Lys Asn Ala Tyr Pro Asp Lys Tyr Glu Arg Phe Glu
100 105 110
Lys Asp Gly Glu Asp Tyr Ser Tyr Ile Phe Phe Lys Lys Ile Trp Lys
115 120 125
Cys His Phe Phe Val Leu Ser Leu Tyr Gln Leu Asn Gln Tyr Ile Met
130 135 140
Ala Ser His Lys Lys Thr Glu Ser Asn Gln Ile Ile Lys Thr Phe Ser
145 150 155 160
Phe Lys Ile Lys Asn Ala Asn Gly Leu Ser Leu Asp Val Leu Asn Asp
165 170 175
Ala Ile Thr Glu Tyr Gln Asn Tyr Tyr Asn Ile Cys Ser Asp Trp Ile
180 185 190
Lys Asp His Leu Thr Met Lys Ile Gly Glu Leu Tyr Lys Tyr Ile Pro
195 200 205
Asp Glu Lys Lys Asn Ser Gly Tyr Ala Leu Thr Leu Ile Ser Asp Glu
210 215 220
Trp Lys Asp Lys Pro Met Tyr Met Met Phe Lys Lys Gly Tyr Pro Ala
225 230 235 240
Asn Asn Arg Asp Asn Ala Ile Tyr Glu Ala Leu Asn Thr Cys Asn Thr
245 250 255
Glu His Tyr Thr Gly Asn Ile Leu Asn Phe Ser Asp Thr Tyr Tyr Arg
260 265 270
Arg Phe Gly Tyr Val Ala Ser Thr Ile Ser Asn Tyr Val Thr Lys Ile
275 280 285
Ser Lys Met Ser Thr Gly Ser Arg Ser Lys Asn Ile Ser Asn Asp Ser
290 295 300

Asp Val Asp Thr Ile Met Glu Gln Val Ile Tyr Glu Met Glu His Asn
305 310 315 320
Gly Trp Thr Ser Val Lys Asp Trp Glu Asn Gln Met Glu Tyr Leu Glu
325 330 335
Ser Lys Thr Asp Ser Asn Pro Asn Phe Val Tyr Arg Met Thr Thr Leu
340 345 350
Tyr Glu Phe Tyr Lys Ser His Ile Asp Glu Val Asn Ser Lys Met Glu
355 360 365
Thr Met Ser Ile Asp Ser Leu Ile Lys Phe Gly Gly Cys Arg Arg Lys
370 375 380
Asp Ser Lys Lys Ser Met Tyr Ile Met Gly Gly Ser Asn Thr Pro Phe
385 390 395 400
Asp Ile Thr Gln Ile Asp Gly Asn Ser Leu Asn Ile Lys Phe Ser Lys
405 410 415
Asn Leu Asn Val Asp Val Phe Gly Arg Tyr Asp Val Ile Lys Asp Asn
420 425 430
Thr Leu Leu Val Asp Ile Ile Asn Gly His Gly Ala Ser Phe Val Leu
435 440 445
Lys Ile Ile Asn Asp Glu Ile Tyr Ile Asp Ile Asn Val Ser Val Pro
450 455 460
Phe Asp Lys Lys Ile Ala Thr Thr Asn Lys Val Val Gly Val Asp Val
465 470 475 480
Asn Ile Lys His Met Leu Leu Ala Thr Asn Ile Leu Asp Asp Gly Asn
485 490 495
Val Asn Gly Tyr Val Asn Ile Tyr Lys Glu Val Ile Asn Asp Ser Asp
500 505 510
Phe Lys Lys Val Cys Asn Ser Thr Val Met Lys Tyr Phe Thr Asp Phe
515 520 525
Ser Lys Phe Val Thr Phe Cys Pro Leu Glu Phe Asp Phe Leu Phe Ser
530 535 540
Arg Val Cys Asn Gln Lys Gly Ile Tyr Asn Asp Asn Ser Ala Met Glu
545 550 555 560
Lys Ser Phe Ser Asp Val Leu Asn Lys Leu Lys Trp Asn Phe Ile Glu
565 570 575
Thr Gly Asp Asn Thr Lys Arg Ile Tyr Ile Glu Asn Val Met Lys Leu
580 585 590
Arg Ser Gln Met Lys Ala Tyr Ala Ile Val Lys Asn Ala Tyr Tyr Lys
595 600 605
Gln Gln Ser Glu Tyr Asp Phe Gly Lys Ser Glu Glu Phe Ile Gln Glu

610 615 620
His Pro Phe Ser Asn Thr Asp Lys Gly Ile Glu Ile Leu Asn Lys Leu
625 630 635 640
Asp Asn Ile Ser Lys Lys Ile Leu Gly Cys Arg Asn Asn Ile Ile Gln
645 650 655
Tyr Ser Tyr Asn Leu Phe Glu Ile Asn Gly Tyr Asp Met Ile Ser Leu
660 665 670
Glu Lys Leu Thr Ser Ser Gln Phe Glu Lys Lys Pro Phe Pro Thr Val
675 680 685
Asn Ser Leu Leu Lys Tyr His Lys Ile Leu Gly Cys Thr Gln Glu Glu
690 695 700
Met Glu Lys Lys Asp Ile Tyr Ser Val Ile Lys Lys Gly Tyr Tyr Asp
705 710 715 720
Ile Ile Phe Asp Asn Gly Val Val Thr Asp Ala Lys Leu Ser Thr Lys
725 730 735
Gly Glu Leu Ser Lys Phe Lys Asp Asp Phe Phe Asn Leu Met Ile Lys
740 745 750
Ser Ile His Phe Ala Asp Ile Lys Asp Tyr Phe Ile Thr Leu Ser Asn
755 760 765
Asn Gly Thr Ala Gly Val Ser Leu Val Pro Ser Tyr Phe Thr Ser Gln
770 775 780
Met Asp Ser Ile Asp His Lys Ile Tyr Phe Val Gln Asp Asn Lys Ser
785 790 795 800
Gly Lys Leu Lys Leu Ala Asn Lys His Lys Val Arg Ser Ser Gln Glu
805 810 815
Lys His Ile Asn Gly Leu Asn Ala Asp Tyr Asn Ala Ala Arg Asn Ile
820 825 830
Ala Tyr Ile Met Glu Asn Thr Glu Cys Arg Asn Met Phe Met Lys Gln
835 840 845
Ser Arg Thr Asp Lys Ser Leu Tyr Asn Lys Pro Ser Tyr Glu Thr Phe
850 855 860
Ile Lys Thr Gln Gly Ser Ala Val Ala Lys Leu Lys Lys Glu Gly Phe
865 870 875 880
Val Lys Ile Leu Asp Glu Ala Ser Val
885

<210> 140
<211> 223
<212> PRT
<213> 人工序列(Artificial sequence)

<220>
<223> 合成序列(synthetic sequence)
<400> 140

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

<210> 141
<211> 439
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 141

Met Lys Ile Gly Glu Leu Tyr Lys Tyr Ile Pro Asp Glu Lys Lys Asn

1 5 10 15
Ser Gly Tyr Ala Leu Thr Leu Ile Ser Asp Glu Trp Lys Asp Lys Pro
20 25 30
Met Tyr Met Met Phe Lys Lys Gly Tyr Pro Ala Asn Asn Arg Asp Asn
35 40 45
Ala Ile Tyr Glu Thr Leu Asn Thr Cys Asn Thr Glu His Tyr Thr Gly
50 55 60
Asn Ile Leu Asn Phe Ser Asp Thr Tyr Tyr Arg Arg Phe Gly Tyr Val
65 70 75 80
Ala Ser Ala Ile Ser Asn Tyr Val Thr Lys Ile Ser Lys Met Ser Thr
85 90 95
Gly Ser Arg Ser Lys Asn Ile Ser Asn Asp Ser Asp Val Asp Thr Ile
100 105 110
Met Glu Gln Val Ile Tyr Glu Met Glu His Asn Gly Trp Thr Ser Val
115 120 125
Lys Asp Trp Glu Asn Gln Met Glu Tyr Leu Glu Ser Lys Thr Asp Ser
130 135 140
Asn Pro Asn Phe Val Tyr Arg Met Thr Thr Leu Tyr Glu Phe Tyr Lys
145 150 155 160
Ser His Ile Asp Glu Val Asn Ser Lys Met Glu Thr Met Ser Ile Asp
165 170 175
Ser Leu Ile Lys Phe Gly Gly Cys Arg Arg Lys Asp Ser Lys Lys Ser
180 185 190
Met Tyr Ile Met Gly Gly Ser Asn Thr Pro Phe Asp Ile Thr Gln Ile
195 200 205
Asp Gly Asn Ser Leu Asn Ile Lys Phe Ser Lys Asn Leu Asn Val Asp
210 215 220
Val Phe Gly Arg Tyr Asp Val Ile Lys Asp Asn Thr Leu Leu Val Asp
225 230 235 240
Ile Ile Asn Gly His Gly Ala Ser Phe Val Leu Lys Ile Ile Asn Gly
245 250 255
Glu Ile Tyr Ile Asp Ile Asn Val Ser Val Pro Phe Asp Lys Lys Ile
260 265 270
Ala Thr Thr Asn Lys Val Val Gly Val Asp Val Asn Ile Lys His Met
275 280 285
Leu Leu Ala Thr Asn Ile Leu Asp Asp Gly Asn Val Asn Gly Tyr Val
290 295 300
Asn Ile Tyr Lys Glu Val Ile Asn Asp Ser Asp Phe Lys Lys Val Cys
305 310 315 320

Asn Ser Thr Val Met Lys Tyr Phe Thr Asp Phe Ser Lys Phe Val Thr
325 330 335
Phe Cys Pro Leu Glu Phe Asp Phe Leu Phe Ser Arg Val Cys Asn Gln
340 345 350
Lys Gly Ile Tyr Asn Asp Asn Ser Ala Met Glu Lys Ser Phe Ser Asp
355 360 365
Val Leu Asn Lys Leu Lys Trp Asn Phe Ile Glu Thr Gly Asp Asn Thr
370 375 380
Lys Arg Ile Tyr Ile Glu Asn Val Met Lys Leu Arg Ser Gln Met Lys
385 390 395 400
Ala Tyr Ala Ile Val Lys Asn Ala Tyr Tyr Lys Gln Gln Ser Glu Tyr
405 410 415
Asp Phe Gly Lys Ser Glu Glu Phe Ile Gln Glu Tyr Asp Phe Phe Gly
420 425 430
Glu Thr Tyr Gln Phe Ser Ile
435

<210> 142
<211> 196
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 142

Met Ile Ser Leu Glu Lys Leu Thr Ser Ser Gln Phe Lys Lys Lys Pro
1 5 10 15
Phe Pro Thr Val Asn Ser Leu Leu Lys Tyr His Lys Ile Leu Gly Cys
20 25 30
Thr Gln Glu Glu Met Glu Lys Lys Lys Gly Tyr Tyr Asp Ile Ile Phe
35 40 45
Asp Asn Asp Val Val Thr Asp Ala Lys Leu Ser Thr Lys Gly Glu Leu
50 55 60
Ser Lys Phe Lys Asp Asp Phe Phe Asn Leu Met Ile Lys Ser Ile His
65 70 75 80
Phe Ala Asp Ile Lys Asp Tyr Phe Ile Thr Leu Ser Asn Asn Gly Thr
85 90 95
Ala Gly Val Ser Leu Val Pro Ser Tyr Phe Thr Ser Gln Met Asp Arg
100 105 110
Ala Asn Lys His Lys Val Arg Ser Ser Gln Glu Lys His Ile Asn Gly
115 120 125

Leu Asn Ala Asp Tyr Asn Ala Ala Arg Asn Ile Ala Tyr Ile Met Glu
130 135 140
Asn Thr Glu Cys Arg Asn Met Phe Met Lys Gln Ser Arg Thr Asp Lys
145 150 155 160
Ser Leu Tyr Asn Lys Pro Ser Tyr Glu Thr Phe Ile Lys Thr Gln Gly
165 170 175
Ser Ala Val Ala Lys Leu Lys Lys Glu Gly Phe Val Lys Ile Leu Asp
180 185 190
Glu Ala Ser Val
195

<210> 143
<211> 748
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 143
Tyr Ile Met Ala Ser His Lys Lys Thr Glu Ser Asn Gln Ile Ile Lys
1 5 10 15
Thr Phe Ser Phe Lys Ile Lys Asn Ala Asn Gly Leu Ser Leu Asp Val
20 25 30
Leu Asn Asp Ala Ile Thr Glu Tyr Gln Asn Tyr Tyr Asn Ile Cys Ser
35 40 45
Asp Trp Ile Lys Asp His Leu Thr Met Lys Ile Gly Glu Leu Tyr Lys
50 55 60
Tyr Ile Pro Asp Glu Lys Lys Asn Ser Gly Tyr Ala Leu Thr Leu Ile
65 70 75 80
Ser Asp Glu Trp Lys Asp Lys Pro Met Tyr Met Met Phe Lys Lys Gly
85 90 95
Tyr Pro Ala Asn Ser Arg Asp Asn Ala Ile Tyr Glu Thr Leu Asn Thr
100 105 110
Cys Asn Thr Glu His Tyr Thr Gly Asn Ile Leu Asn Phe Ser Asp Thr
115 120 125
Tyr Tyr Arg Arg Phe Gly Tyr Val Ala Ser Thr Ile Ser Asn Tyr Val
130 135 140
Thr Lys Ile Ser Lys Met Ser Thr Gly Ser Arg Ser Lys Asn Ile Ser
145 150 155 160
Asn Asp Ser Asp Val Asp Thr Ile Met Glu Gln Val Ile Tyr Glu Met
165 170 175

Glu His Asn Gly Trp Thr Ser Val Lys Asp Trp Glu Asn Gln Met Glu
180 185 190
Tyr Leu Glu Ser Lys Thr Asp Ser Asn Pro Asn Phe Val Tyr Arg Met
195 200 205
Thr Thr Leu Tyr Glu Phe Tyr Lys Ser His Ile Asp Glu Val Asn Ser
210 215 220
Lys Met Glu Thr Met Ser Ile Asp Ser Leu Ile Lys Phe Gly Gly Cys
225 230 235 240
Arg Arg Lys Asp Ser Lys Lys Ser Met Tyr Ile Met Gly Gly Ser Asn
245 250 255
Thr Pro Phe Asp Ile Thr Gln Ile Gly Gly Asn Ser Leu Asn Ile Lys
260 265 270
Phe Ser Lys Asn Leu Asn Val Asp Val Phe Gly Arg Tyr Asp Val Ile
275 280 285
Lys Asp Asn Thr Leu Leu Val Asp Ile Ile Asn Gly His Gly Ala Ser
290 295 300
Phe Val Leu Lys Ile Ile Asn Gly Glu Ile Tyr Ile Asp Ile Asn Val
305 310 315 320
Ser Val Pro Phe Asp Lys Lys Ile Ala Thr Thr Asn Lys Val Val Gly
325 330 335
Val Asp Val Asn Ile Lys His Met Leu Leu Ala Thr Asn Ile Leu Asp
340 345 350
Asp Gly Asn Val Asn Gly Tyr Val Asn Ile Tyr Lys Glu Val Ile Asn
355 360 365
Asp Ser Asp Phe Lys Lys Val Cys Asn Ser Thr Val Met Lys Tyr Phe
370 375 380
Thr Asp Phe Ser Lys Phe Val Thr Phe Cys Pro Leu Glu Phe Asp Phe
385 390 395 400
Leu Phe Ser Arg Val Cys Asn Gln Lys Gly Ile Tyr Asn Asp Asn Ser
405 410 415
Ala Met Glu Lys Ser Phe Ser Asn Val Leu Asn Lys Leu Lys Trp Asn
420 425 430
Phe Ile Glu Thr Gly Asp Asn Thr Lys Arg Ile Tyr Ile Glu Asn Val
435 440 445
Met Lys Leu Arg Ser Gln Met Lys Ala Tyr Ala Ile Val Lys Asn Ala
450 455 460
Tyr Tyr Lys Gln Gln Ser Glu Tyr Asp Phe Gly Lys Ser Glu Glu Phe
465 470 475 480
Ile Gln Glu His Pro Phe Ser Asn Thr Asp Lys Gly Ile Glu Ile Leu

485 490 495

Asn Lys Leu Asp Asn Ile Ser Lys Lys Ile Leu Gly Cys Arg Asn Asn

500 505 510

Ile Ile Gln Tyr Ser Tyr Asn Leu Phe Glu Ile Asn Gly Tyr Asp Met

515 520 525

Ile Ser Leu Glu Lys Leu Thr Ser Ser Gln Phe Glu Lys Lys Pro Phe

530 535 540

Pro Thr Val Asn Ser Leu Leu Lys Tyr His Lys Ile Leu Gly Cys Thr

545 550 555 560

Gln Glu Glu Met Glu Lys Lys Asp Ile Tyr Ser Val Ile Lys Lys Gly

565 570 575

Tyr Tyr Asp Ile Ile Phe Asp Asn Gly Val Val Thr Asp Ala Lys Leu

580 585 590

Ser Thr Lys Gly Glu Leu Ser Lys Phe Lys Asp Asp Phe Phe Asn Leu

595 600 605

Met Ile Lys Ser Ile His Phe Ala Asp Ile Lys Asp Tyr Phe Ile Thr

610 615 620

Leu Ser Asn Asn Gly Thr Ala Gly Val Ser Leu Val Pro Ser Tyr Phe

625 630 635 640

Thr Ser Gln Met Asp Ser Ile Asp His Lys Ile Tyr Phe Val Gln Asp

645 650 655

Asn Lys Ser Gly Lys Leu Lys Leu Ala Asn Lys His Lys Val Arg Ser

660 665 670

Ser Gln Glu Lys His Ile Asn Gly Leu Asn Ala Asp Tyr Asn Ala Ala

675 680 685

Arg Asn Ile Ala Tyr Ile Met Glu Asn Thr Glu Cys Arg Asn Met Phe

690 695 700

Met Lys Gln Ser Arg Thr Asp Lys Ser Leu Tyr Asn Lys Pro Ser Tyr

705 710 715 720

Glu Thr Phe Ile Lys Thr Gln Gly Ser Ala Val Ala Lys Leu Lys Lys

725 730 735

Glu Gly Phe Val Lys Ile Leu Asp Glu Ala Ser Val

740 745

<210> 144

<211> 481

<212> PRT

<213> 人工序列(Artificial sequence)

<220>

<223> 合成序列(synthetic sequence)

<400> 144

Ser Leu Asn Ile Lys Phe Ser Lys Asn Leu Asn Val Asp Val Phe Gly
1 5 10 15
Arg Tyr Asp Val Ile Lys Asp Asn Thr Leu Leu Val Asp Ile Ile Asn
20 25 30
Gly His Gly Ala Ser Phe Val Leu Lys Ile Ile Asn Gly Glu Ile Tyr
35 40 45
Ile Asp Ile Asn Val Ser Val Pro Phe Asp Lys Lys Ile Ala Thr Thr
50 55 60
Asn Lys Val Val Gly Val Asp Val Asn Ile Lys His Met Leu Leu Ala
65 70 75 80
Thr Asn Ile Leu Asp Asp Gly Asn Val Asn Gly Tyr Val Asn Ile Tyr
85 90 95
Lys Glu Val Ile Asn Asp Ser Asp Phe Lys Lys Val Cys Asn Ser Thr
100 105 110
Val Met Lys Tyr Phe Thr Asp Phe Ser Lys Phe Val Thr Phe Cys Pro
115 120 125
Leu Glu Phe Asp Phe Leu Phe Ser Arg Val Cys Asn Gln Lys Gly Ile
130 135 140
Tyr Asn Asp Asn Ser Ala Met Glu Lys Ser Phe Ser Asp Val Leu Asn
145 150 155 160
Lys Leu Lys Trp Asn Phe Ile Glu Thr Gly Asp Asn Thr Lys Arg Ile
165 170 175
Tyr Ile Glu Asn Val Met Lys Leu Arg Ser Gln Met Lys Ala Tyr Ala
180 185 190
Ile Val Lys Asn Ala Tyr Tyr Lys Gln Gln Ser Glu Tyr Asp Phe Gly
195 200 205
Lys Ser Glu Glu Phe Ile Gln Glu His Pro Phe Ser Asn Thr Asp Lys
210 215 220
Gly Ile Glu Ile Leu Asn Lys Leu Asp Asn Ile Ser Lys Lys Ile Leu
225 230 235 240
Gly Cys Arg Asn Asn Ile Ile Gln Tyr Ser Tyr Asn Leu Phe Glu Ile
245 250 255
Asn Gly Tyr Asp Met Ile Ser Leu Glu Lys Leu Thr Ser Ser Gln Phe
260 265 270
Lys Lys Lys Pro Phe Pro Thr Val Asn Ser Leu Leu Lys Tyr His Lys
275 280 285
Ile Leu Gly Cys Thr Gln Glu Glu Met Glu Lys Lys Asp Ile Tyr Ser
290 295 300

Val Ile Lys Lys Gly Tyr Tyr Asp Ile Ile Phe Asp Asn Gly Val Val
305 310 315 320
Ile Asp Ala Lys Leu Ser Ala Lys Gly Glu Leu Ser Lys Phe Lys Asp
325 330 335
Asp Phe Phe Asn Leu Met Ile Lys Ser Ile His Phe Ala Asp Ile Lys
340 345 350
Asp Tyr Phe Ile Thr Leu Ser Asn Asn Gly Thr Ala Gly Val Ser Leu
355 360 365
Val Pro Ser Tyr Phe Thr Ser Gln Met Asp Ser Ile Asp His Lys Ile
370 375 380
Tyr Phe Val Gln Asp Asn Lys Ser Gly Lys Leu Lys Leu Ala Asn Lys
385 390 395 400
His Lys Val Arg Ser Ser Gln Glu Lys His Ile Asn Gly Leu Asn Ala
405 410 415
Asp Tyr Asn Ala Ala Arg Asn Ile Ala Tyr Ile Met Glu Asn Thr Glu
420 425 430
Cys Arg Asn Met Phe Met Lys Gln Ser Arg Thr Asp Lys Ser Leu Tyr
435 440 445
Asn Lys Pro Ser Tyr Glu Thr Phe Ile Lys Thr Gln Gly Ser Ala Val
450 455 460
Ala Lys Leu Lys Lys Glu Gly Phe Val Lys Ile Leu Asp Glu Ala Ser
465 470 475 480
Val

<210> 145
<211> 746
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 145
Met Ala Ser His Lys Lys Thr Glu Ser Asn Gln Ile Ile Lys Thr Phe
1 5 10 15
Ser Phe Lys Ile Lys Asn Ala Asn Gly Leu Ser Leu Asp Val Leu Asn
20 25 30
Asp Ala Ile Thr Glu Tyr Gln Asn Tyr Tyr Asn Ile Cys Ser Asp Trp
35 40 45
Ile Lys Asp His Leu Thr Met Lys Ile Gly Glu Leu Tyr Lys Tyr Ile
50 55 60
Pro Asp Glu Lys Lys Asn Ser Gly Tyr Ala Leu Thr Leu Ile Ser Asp

65 70 75 80
Glu Trp Lys Asp Lys Pro Met Tyr Met Met Phe Lys Lys Gly Tyr Pro
85 90 95
Ala Asn Asn Arg Asp Asn Ala Ile Tyr Glu Thr Leu Asn Thr Cys Asn
100 105 110
Thr Glu His Tyr Thr Gly Asn Ile Leu Asn Phe Ser Asp Thr Tyr Tyr
115 120 125
Arg Arg Phe Gly Tyr Val Ala Ser Ala Ile Ser Asn Tyr Val Thr Lys
130 135 140
Ile Ser Lys Met Ser Thr Gly Ser Arg Ser Lys Asn Ile Ser Asn Asp
145 150 155 160
Ser Asp Val Asp Thr Ile Met Glu Gln Val Ile Tyr Glu Met Glu His
165 170 175
Asn Gly Trp Thr Ser Val Lys Asp Trp Glu Asn Gln Met Glu Tyr Leu
180 185 190
Glu Ser Lys Thr Asp Ser Asn Pro Asn Phe Val Tyr Arg Met Thr Thr
195 200 205
Leu Tyr Glu Phe Tyr Lys Ser His Ile Asp Glu Val Asn Ser Lys Met
210 215 220
Glu Thr Met Ser Ile Asp Ser Leu Ile Lys Phe Gly Gly Cys Arg Arg
225 230 235 240
Lys Asp Ser Lys Lys Ser Met Tyr Ile Met Gly Gly Ser Asn Thr Pro
245 250 255
Phe Asp Ile Thr Gln Ile Gly Gly Asn Ser Leu Asn Ile Lys Phe Ser
260 265 270
Lys Asn Leu Asn Val Asp Val Phe Gly Arg Tyr Asp Val Ile Lys Asp
275 280 285
Asn Thr Leu Leu Val Asp Ile Ile Asn Glu His Gly Ala Ser Phe Val
290 295 300
Leu Lys Ile Ile Asn Asp Glu Ile Tyr Ile Asp Ile Asn Val Ser Val
305 310 315 320
Pro Phe Asp Lys Lys Ile Ala Thr Thr Asn Lys Val Val Gly Ile Asp
325 330 335
Val Asn Ile Lys His Met Leu Leu Ala Thr Asn Ile Leu Asp Asp Gly
340 345 350
Asn Val Asn Gly Tyr Val Asn Ile Tyr Lys Glu Val Ile Asn Asp Ser
355 360 365
Asp Phe Lys Lys Val Cys Asn Ser Thr Val Met Lys Tyr Phe Thr Asp
370 375 380

Phe Ser Lys Phe Val Thr Phe Cys Pro Leu Glu Phe Asp Phe Leu Phe
385 390 395 400
Ser Arg Val Cys Asn Gln Lys Gly Ile Tyr Asn Asp Asn Ser Ala Met
405 410 415
Glu Lys Ser Phe Ser Asp Val Leu Asn Lys Leu Lys Trp Asn Phe Ile
420 425 430
Glu Thr Gly Asp Asn Thr Lys Arg Ile Tyr Ile Glu Asn Val Met Lys
435 440 445
Leu Arg Ser Gln Met Lys Ala Tyr Ala Ile Val Lys Asn Ala Tyr Tyr
450 455 460
Lys Gln Gln Ser Glu Tyr Asp Phe Gly Lys Ser Glu Glu Phe Ile Gln
465 470 475 480
Glu His Pro Phe Ser Asn Thr Asp Lys Gly Ile Glu Ile Leu Asn Lys
485 490 495
Leu Asp Asn Ile Ser Lys Lys Ile Leu Gly Cys Arg Asn Asn Ile Ile
500 505 510
Gln Tyr Ser Tyr Asn Leu Phe Glu Ile Asn Gly Tyr Asp Met Ile Ser
515 520 525
Leu Glu Lys Leu Thr Ser Ser Gln Phe Lys Lys Lys Ser Phe Pro Thr
530 535 540
Val Asp Ser Leu Leu Lys Tyr His Lys Ile Leu Gly Cys Thr Gln Glu
545 550 555 560
Glu Met Glu Lys Lys Asp Ile Tyr Ser Val Ile Lys Lys Gly Tyr Tyr
565 570 575
Asp Ile Ile Phe Asp Asn Asp Val Val Thr Asp Ala Lys Leu Ser Thr
580 585 590
Lys Gly Glu Leu Ser Lys Phe Lys Asp Asp Phe Phe Asn Leu Met Ile
595 600 605
Lys Ser Ile His Phe Ala Asp Ile Lys Asp Tyr Phe Ile Thr Leu Ser
610 615 620
Asn Asn Gly Thr Ala Gly Val Ser Leu Val Pro Ser Tyr Phe Thr Ser
625 630 635 640
Gln Met Asp Ser Ile Asp His Lys Ile Tyr Phe Val Gln Asp Asn Lys
645 650 655
Ser Gly Lys Leu Lys Leu Ala Asn Lys His Lys Val Arg Ser Ser Gln
660 665 670
Glu Lys His Ile Asn Gly Leu Asn Ala Asp Tyr Asn Ala Ala Arg Asn
675 680 685
Ile Ala Tyr Ile Met Glu Asn Thr Glu Cys Arg Asn Met Phe Met Lys

690 695 700

Gln Ser Arg Thr Asp Lys Ser Leu Tyr Asn Lys Pro Ser Tyr Glu Thr
705 710 715 720

Phe Ile Lys Thr Gln Gly Ser Ala Val Ala Lys Leu Lys Lys Glu Gly
725 730 735

Phe Val Lys Ile Leu Asp Glu Ala Ser Val
740 745

<210> 146

<211> 746

<212> PRT

<213> 人工序列(Artificial sequence)

<220>

<223> 合成序列(synthetic sequence)

<400> 146

Met Ala Ser His Lys Lys Thr Glu Ser Asn Gln Ile Ile Lys Thr Phe
1 5 10 15

Ser Phe Lys Ile Lys Asn Ala Asn Gly Leu Ser Leu Asp Val Leu Asn
20 25 30

Asp Ala Ile Thr Glu Tyr Gln Asn Tyr Tyr Asn Ile Cys Ser Asp Trp
35 40 45

Ile Lys Asp His Leu Thr Met Lys Ile Gly Glu Leu Tyr Lys Tyr Ile
50 55 60

Pro Asp Glu Lys Lys Asn Ser Gly Tyr Ala Leu Thr Leu Ile Ser Asp
65 70 75 80

Glu Trp Lys Asp Lys Pro Met Tyr Met Met Phe Lys Lys Gly Tyr Pro
85 90 95

Ala Asn Asn Arg Asp Asn Ala Ile Tyr Glu Thr Leu Asn Thr Cys Asn
100 105 110

Thr Glu His Tyr Thr Gly Asn Ile Leu Asn Phe Ser Asp Thr Tyr Tyr
115 120 125

Arg Arg Phe Gly Tyr Val Ala Ser Ala Ile Ser Asn Tyr Val Thr Lys
130 135 140

Ile Ser Lys Met Ser Thr Gly Ser Arg Ser Lys Asn Ile Ser Asn Asp
145 150 155 160

Ser Asp Val Asp Thr Ile Met Glu Gln Val Ile Tyr Glu Met Glu His
165 170 175

Asn Gly Trp Thr Ser Val Lys Asp Trp Glu Asn Gln Met Glu Tyr Leu
180 185 190

Glu Ser Lys Thr Asp Ser Asn Pro Asn Phe Val Tyr Arg Met Thr Thr

195 200 205

Leu Tyr Glu Phe Tyr Lys Ser His Ile Asp Glu Val Asn Ser Lys Met

210 215 220

Glu Thr Met Ser Ile Asp Ser Leu Ile Lys Phe Gly Gly Cys Arg Arg

225 230 235 240

Lys Asp Ser Lys Lys Ser Met Tyr Ile Met Gly Gly Ser Asn Thr Pro

245 250 255

Phe Asp Ile Thr Gln Ile Asp Gly Asn Ser Leu Asn Ile Lys Phe Ser

260 265 270

Lys Asn Leu Asn Val Asp Val Phe Gly Arg Tyr Asp Val Ile Lys Asp

275 280 285

Asn Thr Leu Leu Val Asp Ile Ile Asn Gly His Gly Ala Ser Phe Val

290 295 300

Leu Lys Ile Ile Asn Asp Glu Ile Tyr Ile Asp Ile Asn Val Ser Val

305 310 315 320

Pro Phe Asp Lys Lys Ile Ala Thr Thr Asn Lys Val Val Gly Val Asp

325 330 335

Val Asn Ile Lys His Met Leu Leu Ala Thr Asn Ile Leu Asp Asp Gly

340 345 350

Asn Val Asn Gly Tyr Val Asn Ile Tyr Lys Glu Val Ile Asn Asp Ser

355 360 365

Asp Phe Lys Lys Val Cys Asn Ser Thr Val Met Lys Tyr Phe Thr Asp

370 375 380

Phe Ser Lys Phe Val Thr Phe Cys Pro Leu Glu Phe Asp Phe Leu Phe

385 390 395 400

Ser Arg Val Cys Asn Gln Lys Gly Ile Tyr Asn Asp Asn Ser Ala Met

405 410 415

Glu Lys Ser Phe Ser Asp Val Leu Asn Lys Leu Lys Trp Asn Phe Ile

420 425 430

Glu Thr Gly Asp Asn Thr Lys Arg Ile Tyr Ile Glu Asn Val Met Lys

435 440 445

Leu Arg Ser Gln Met Lys Ala Tyr Ala Ile Val Lys Asn Ala Tyr Tyr

450 455 460

Lys Gln Gln Ser Glu Tyr Asp Phe Gly Lys Ser Glu Glu Phe Ile Gln

465 470 475 480

Glu His Pro Phe Ser Asn Thr Asp Lys Gly Ile Glu Ile Leu Asn Lys

485 490 495

Leu Asp Asn Ile Ser Lys Lys Ile Leu Gly Cys Arg Asn Asn Ile Ile

500 505 510

Gln Tyr Ser Tyr Asn Leu Phe Glu Ile Asn Gly Tyr Asp Met Ile Ser
515 520 525
Leu Glu Lys Leu Thr Ser Ser Gln Phe Lys Lys Lys Pro Phe Pro Thr
530 535 540
Val Asn Ser Leu Leu Lys Tyr His Lys Ile Leu Gly Cys Thr Gln Glu
545 550 555 560
Glu Met Glu Lys Lys Asp Ile Tyr Ser Val Ile Lys Lys Gly Tyr Tyr
565 570 575
Asp Ile Ile Phe Asp Asn Asp Val Val Thr Asp Ala Lys Leu Ser Thr
580 585 590
Lys Gly Glu Leu Ser Lys Phe Lys Asp Asp Phe Phe Asn Leu Met Ile
595 600 605
Lys Ser Ile His Phe Ala Asp Ile Lys Asp Tyr Phe Ile Thr Leu Ser
610 615 620
Asn Asn Gly Thr Ala Gly Val Ser Leu Val Pro Ser Tyr Phe Thr Ser
625 630 635 640
Gln Met Asp Ser Ile Asp His Lys Ile Tyr Phe Val Gln Asp Asn Lys
645 650 655
Ser Gly Lys Leu Lys Leu Ala Asn Lys His Lys Val Arg Ser Ser Gln
660 665 670
Glu Lys His Ile Asn Gly Leu Asn Ala Asp Tyr Asn Ala Ala Arg Asn
675 680 685
Ile Ala Tyr Ile Met Glu Asn Thr Glu Cys Arg Asn Met Phe Met Lys
690 695 700
Gln Ser Arg Thr Asp Lys Ser Leu Tyr Asn Lys Pro Ser Tyr Glu Thr
705 710 715 720
Phe Ile Lys Thr Gln Gly Ser Ala Val Ala Lys Leu Lys Lys Glu Gly
725 730 735
Phe Met Lys Ile Leu Asp Glu Ala Ser Val
740 745

<210> 147
<211> 746
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 147
Met Ala Ser His Lys Lys Thr Glu Ser Asn Gln Ile Ile Lys Thr Phe
1 5 10 15

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

325 330 335
Val Asn Ile Lys His Met Leu Leu Ala Thr Asn Ile Leu Asp Asp Gly
340 345 350
Asn Val Asn Gly Tyr Val Asn Ile Tyr Lys Glu Val Ile Asn Asp Ser
355 360 365
Asp Phe Lys Lys Val Cys Asn Ser Thr Val Met Lys Tyr Phe Thr Asp
370 375 380
Phe Ser Lys Phe Val Thr Phe Cys Pro Leu Glu Phe Asp Phe Leu Phe
385 390 395 400
Ser Arg Val Cys Asn Gln Lys Gly Ile Tyr Asn Asp Asn Ser Ala Met
405 410 415
Glu Lys Ser Phe Ser Asp Val Leu Asn Lys Leu Lys Trp Asn Phe Ile
420 425 430
Glu Thr Gly Asp Asn Thr Lys Arg Ile Tyr Ile Glu Asn Val Met Lys
435 440 445
Leu Arg Ser Gln Met Lys Ala Tyr Ala Ile Val Lys Asn Ala Tyr Tyr
450 455 460
Lys Gln Gln Ser Glu Tyr Asp Phe Gly Lys Ser Glu Glu Phe Ile Gln
465 470 475 480
Glu His Pro Phe Ser Asn Thr Asp Lys Gly Ile Glu Ile Leu His Lys
485 490 495
Leu Asp Asn Ile Ser Lys Lys Ile Leu Gly Cys Arg Asn Asn Ile Ile
500 505 510
Gln Tyr Ser Tyr Asn Leu Phe Glu Ile Asn Gly Tyr Asp Met Ile Ser
515 520 525
Leu Glu Lys Leu Thr Ser Ser Gln Phe Lys Lys Lys Pro Phe Pro Thr
530 535 540
Val Asn Ser Leu Leu Lys Tyr His Lys Ile Leu Gly Cys Thr Gln Glu
545 550 555 560
Glu Met Glu Lys Lys Asp Ile Tyr Ser Val Ile Lys Lys Gly Tyr Tyr
565 570 575
Asp Ile Ile Phe Asp Asn Asp Val Val Thr Asp Ala Lys Leu Ser Thr
580 585 590
Lys Gly Glu Leu Ser Lys Phe Lys Asp Asp Phe Phe Asn Leu Met Ile
595 600 605
Lys Ser Ile His Phe Ala Asp Ile Lys Asp Tyr Phe Ile Thr Leu Ser
610 615 620
Asn Asn Gly Thr Ala Gly Val Ser Leu Val Pro Ser Phe Phe Thr Ser
625 630 635 640

Gln Met Asp Ser Ile Asp His Lys Ile Tyr Phe Val Gln Asp Asn Lys
645 650 655
Ser Gly Lys Leu Lys Leu Ala Asn Lys His Lys Val Arg Ser Ser Gln
660 665 670
Glu Lys His Ile Asn Gly Leu Asn Ala Asp Tyr Asn Ala Ala Arg Asn
675 680 685
Ile Ala Tyr Ile Met Glu Asn Thr Glu Cys Arg Asn Met Phe Met Lys
690 695 700
Gln Ser Arg Thr Asp Lys Ser Leu Tyr Asn Lys Pro Ser Tyr Glu Thr
705 710 715 720
Phe Ile Lys Thr Gln Gly Ser Ala Val Ala Lys Leu Lys Lys Glu Gly
725 730 735
Phe Val Lys Ile Leu Asp Glu Ala Ser Val
740 745

<210> 148
<211> 735
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 148

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

Ser Lys Leu Asn Thr Gly Val Lys Ser Arg Ala Ile Lys Glu Thr Ser
145 150 155 160
Asp Glu Glu Thr Ile Ile Glu Gln Val Ile Tyr Glu Met Glu His Asn
165 170 175
Lys Trp Glu Ser Ile Glu Asp Trp Lys Asn Gln Ile Glu Tyr Leu Asn
180 185 190
Ser Lys Thr Asp Tyr Asn Pro Thr Tyr Met Glu Arg Met Lys Thr Leu
195 200 205
Ser Ala Tyr Tyr Ser Glu His Lys Ser Glu Val Asp Ala Lys Met Gln
210 215 220
Glu Met Ala Val Glu Asn Leu Val Lys Phe Gly Gly Cys Arg Arg Asn
225 230 235 240
Asn Ser Lys Lys Ser Met Phe Ile Met Gly Thr Thr Asn Thr Asn Tyr
245 250 255
Thr Ile Ser Tyr Ile Gly Gly Asn Ser Phe Asn Ile Asn Phe Ala Asn
260 265 270
Ile Leu Asn Phe Asp Val Tyr Gly Arg Arg Asp Val Val Lys Asn Gly
275 280 285
Glu Val Leu Val Asp Ile Met Ala Asn His Gly Asp Ser Ile Val Leu
290 295 300
Lys Ile Val Asn Gly Glu Leu Tyr Ala Asp Val Pro Cys Ser Val Thr
305 310 315 320
Leu Asn Lys Val Glu Ser Asn Phe Asp Lys Val Val Gly Ile Asp Val
325 330 335
Asn Met Lys His Met Leu Leu Ser Thr Ser Val Thr Asp Asn Gly Ser
340 345 350
Leu Asp Phe Leu Asn Ile Tyr Lys Glu Met Ser Asn Asn Ala Glu Phe
355 360 365
Met Ala Leu Cys Ser Glu Asp Asp Arg Lys Tyr Tyr Lys Asp Ile Ser
370 375 380
Gln Tyr Val Thr Phe Ala Pro Leu Glu Leu Asp Leu Leu Phe Ser Arg
385 390 395 400
Ile Ser Lys Gln Gly Lys Val Lys Met Glu Lys Ala Tyr Ser Glu Ile
405 410 415
Leu Glu Ala Leu Lys Trp Lys Phe Phe Ala Asn Gly Asp Asn Lys Asn
420 425 430
Arg Ile Tyr Val Glu Ser Ile Gln Lys Ile Arg Gln Gln Ile Lys Ala
435 440 445
Leu Cys Ile Ile Lys Asn Ala Tyr Tyr Glu Gln Gln Ser Ala Tyr Asp

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

<210> 149
<211> 779
<212> PRT
<213> 人工序列(Artificial sequence)

<220>

<223> 合成序列(synthetic sequence)

<400> 149

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

Gly Leu Leu Gly Gly Glu Glu Ile Leu Asn Ile Pro Glu Lys His Ser
290 295 300
Thr Ser Ile Thr Phe Val Met Arg Asn Asn Ser Leu Tyr Val Asp Ile
305 310 315 320
Pro Val Ala Val Pro Phe Ser Lys Val Ile Asn Asp Cys Asp Gly Lys
325 330 335
Thr Val Gly Ile Asp Val Asn Leu Lys His Ala Leu Phe Ala Thr Ser
340 345 350
Glu Val Asp Asn Gly Gln Phe Tyr Asp Tyr Val Asn Val Tyr Ala Glu
355 360 365
Leu Leu Lys Asp Glu Asn Phe Val Lys Val Cys His Lys Glu Leu Leu
370 375 380
Asp Tyr Ile Lys Asp Val Ser Lys Tyr Val Phe Phe Ala Pro Ile Glu
385 390 395 400
Leu Asn Leu Leu Leu Ser Arg Val Met Lys Gln Lys Gly Tyr Glu Asn
405 410 415
Ile Asp Asn Tyr Lys Lys Leu Tyr Lys Val Glu Glu Ala Tyr Leu Cys
420 425 430
Val Leu Asp Lys Leu Gln Lys Arg Phe Ile Asp Glu Gly Asn Asn Thr
435 440 445
Lys Arg Ile Tyr Ile Glu Asn Leu Lys Lys Met Arg Ala Gln Met Lys
450 455 460
Ala Tyr Tyr Ile Leu Lys Asp Thr Tyr Ser Lys Tyr Gln Lys Asp Tyr
465 470 475 480
Asp Ile Glu Met Gly Phe Val Asp Glu Ser Thr Glu Ser Lys Glu Thr
485 490 495
Met Asp Ala Arg Arg Ser Glu Asn Pro Phe Arg Ser Thr Asp Ile Ala
500 505 510
Gln Asp Ile Leu Lys Lys Met Asn Asn Val Gly Lys Thr Val Glu Ala
515 520 525
Cys Arg Asn Asn Ile Ile Ala Tyr Ile Tyr Lys Val Phe Glu Asn Ser
530 535 540
Asp Phe Ala Thr Ile Val Leu Glu Lys Leu Gln Ser Ser Gln Met Lys
545 550 555 560
Lys His Lys Arg Ile Pro Thr Val Asn Ser Leu Leu Lys Tyr His His
565 570 575
Val Glu Gly His Thr Ile Glu Glu Thr Lys Glu Met Lys Ile Tyr Asp
580 585 590
Val Val Glu Lys Gly Tyr Tyr Asn Phe Ile Val Asn Glu Lys Asn Glu

595 600 605
Ile Ile Asp Ala Thr Leu Thr Asp Lys Gly Lys Val Ile Met Ile Glu
610 615 620
Ala Glu Phe Tyr Asn Phe Ala Leu Lys Ser Ile His Phe Ala Asp Ala
625 630 635 640
Lys Asp Tyr Phe Ile Thr Leu Ser Asn Asn Gly Ser Val Asn Ile Ala
645 650 655
Leu Val Pro Ser Gln Phe Thr Ser Gln Met Asp Ser Ile Arg His Ala
660 665 670
Ile Phe Val Thr Lys Gly Lys Lys Gly Lys Lys Val Ile Val Asp Lys
675 680 685
Lys Tyr Val Arg Pro Lys Gln Glu Lys His Ile Asn Gly Leu Asn Gly
690 695 700
Asp Tyr Asn Ala Ser Arg Asn Ile Ala Tyr Ile Phe Glu Asn Glu Glu
705 710 715 720
Leu Arg Glu Glu Leu Leu Lys Lys Glu Glu Glu Tyr Asn Lys Tyr Gly
725 730 735
Lys Val Leu Tyr Asp Thr Leu Ile Lys Phe Pro Ser Gly Val Ile Asn
740 745 750
Lys Leu Lys Lys Phe Gly Asp Lys Tyr Met Thr Thr Ile Glu Asn Leu
755 760 765
Asp Glu Ile Gln Val Glu Asp Val Ala Tyr Val
770 775

<210> 150
<211> 592
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 150
Met Ser Lys Leu Asn Thr Gly Val Lys Arg His Thr Ile Lys Glu Thr
1 5 10 15
Ser Asp Glu Ala Thr Ile Val Glu Gln Val Ile Tyr Glu Met Glu His
20 25 30
Asn Lys Trp Glu Ser Ile Glu Asp Trp Lys Asn Gln Ile Glu Tyr Leu
35 40 45
Asn Ser Lys Thr Asp Tyr Asn Pro Thr Tyr Met Glu Arg Met Lys Thr
50 55 60
Leu Ser Ala Tyr Tyr Ser Glu His Lys Ser Glu Val Asp Ala Lys Met

65 70 75 80
Gln Glu Met Ala Val Glu Asn Leu Val Lys Phe Gly Gly Cys Arg Arg
85 90 95
Asn Asn Ser Lys Lys Ser Met Phe Ile Met Gly Ser Ser Lys Thr Thr
100 105 110
Tyr Thr Ile Ser Tyr Ile Gly Asp Asn Cys Phe Asn Ile Asn Phe Ala
115 120 125
Asn Ile Leu Asn Phe Asp Val Tyr Gly Arg Arg Asp Val Val Lys Asn
130 135 140
Gly Glu Val Leu Val Asp Ile Met Ala Asn His Gly Asp Ser Ile Val
145 150 155 160
Leu Lys Ile Val Asn Gly Glu Leu Tyr Ala Asp Val Pro Cys Ser Thr
165 170 175
Thr Leu Asn Lys Val Glu Ser Thr Phe Asp Lys Val Ala Gly Ile Asp
180 185 190
Val Asn Met Lys His Met Leu Leu Ser Thr Ser Val Thr Asp Asn Gly
195 200 205
Asn Ser Asp Phe Val Asn Ile Tyr Lys Glu Met Ser Asn Asn Ala Glu
210 215 220
Phe Met Ala Leu Cys Pro Glu Glu Asp Arg Lys Tyr Tyr Lys Asp Ile
225 230 235 240
Ser Gln Tyr Val Thr Phe Ala Pro Leu Glu Leu Asp Leu Leu Phe Ser
245 250 255
Arg Ile Ser Lys Gln Gly Lys Val Lys Met Glu Lys Ala Tyr Ser Glu
260 265 270
Ile Leu Glu Thr Leu Lys Trp Lys Phe Phe Ala Asn Gly Asp Asn Lys
275 280 285
Asn Arg Ile Tyr Val Glu Ser Ile Gln Lys Ile Arg Gln Gln Ile Lys
290 295 300
Ala Leu Cys Val Ile Lys Asn Ala Tyr Tyr Glu Gln Gln Ser Ala Tyr
305 310 315 320
Asp Ile Asp Lys Thr Gln Glu Tyr Ile Glu Ala His Pro Phe Ser Leu
325 330 335
Thr Glu Lys Gly Met Ser Ile Lys Ser Lys Met Asp Asn Ile Cys Arg
340 345 350
Thr Ile Ile Gly Cys Arg Asn Asn Ile Ile Asp Tyr Ala Tyr Ser Phe
355 360 365
Phe Glu Arg Asn Asp Tyr Ser Ile Ile Gly Leu Glu Lys Leu Thr Ser
370 375 380

Ser Gln Phe Glu Lys Thr Lys Ser Leu Pro Thr Cys Lys Ser Leu Leu
385 390 395 400
Asn Phe His Lys Val Leu Gly His Thr Leu Ser Glu Leu Glu Thr Leu
405 410 415
Pro Ile Asn Asp Val Val Lys Lys Gly Tyr Tyr Thr Phe Thr Thr Asp
420 425 430
Asn Glu Gly Arg Ile Thr Asp Ala Ser Leu Ser Glu Lys Gly Lys Val
435 440 445
Arg Lys Met Lys Asp Asp Phe Phe Asn Gln Ala Ile Lys Ala Ile His
450 455 460
Phe Ala Asp Val Lys Asp Tyr Phe Ala Thr Leu Ser Asn Asn Gly Gln
465 470 475 480
Thr Gly Ile Phe Phe Val Pro Ser Gln Phe Thr Ser Gln Met Asp Ser
485 490 495
Asn Thr His Thr Leu Tyr Phe Glu Asn Ala Lys Asn Gly Gly Leu Lys
500 505 510
Leu Ala Ser Lys Tyr Lys Val Arg Gln Thr Gln Glu Tyr His Leu Asn
515 520 525
Gly Leu Pro Ala Asp Tyr Asn Ala Ala Arg Asn Ile Ala Tyr Ile Gly
530 535 540
Leu Asp Glu Thr Met Arg Asn Thr Phe Leu Lys Lys Ala Asn Ser Asn
545 550 555 560
Lys Ser Leu Tyr Asn Gln Pro Ile Tyr Asp Thr Gly Ile Lys Lys Thr
565 570 575
Ala Gly Val Phe Ser Arg Met Lys Lys Leu Lys Arg Tyr Glu Ile Ile
580 585 590

<210> 151
<211> 746
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 151

Met Ala Ser His Lys Lys Thr Glu Ser Asn Gln Ile Ile Lys Thr Phe
1 5 10 15
Ser Phe Lys Ile Lys Asn Ala Asn Gly Leu Ser Leu Asp Val Leu Asn
20 25 30
Asp Ala Ile Thr Glu Tyr Gln Asn Tyr Tyr Asn Ile Cys Ser Asp Trp
35 40 45

Ile Lys Asp His Leu Thr Met Lys Ile Gly Glu Leu Tyr Lys Tyr Ile
50 55 60
Pro Asp Glu Lys Lys Asn Ser Gly Tyr Ala Leu Thr Leu Ile Ser Asp
65 70 75 80
Glu Trp Lys Asp Lys Pro Met Tyr Met Met Phe Lys Lys Gly Tyr Pro
85 90 95
Ala Asn Asn Arg Asp Asn Ala Ile Tyr Glu Ala Leu Asn Thr Cys Asn
100 105 110
Thr Glu His Tyr Thr Gly Asn Ile Leu Asn Phe Ser Asp Thr Tyr Tyr
115 120 125
Arg Arg Phe Gly Tyr Val Ala Ser Ala Ile Ser Asn Tyr Val Thr Lys
130 135 140
Ile Ser Lys Met Ser Thr Gly Ser Arg Ser Lys Asn Ile Ser Asn Asp
145 150 155 160
Ser Asp Val Asp Thr Ile Met Glu Gln Val Ile Tyr Glu Met Glu His
165 170 175
Asn Gly Trp Thr Ser Val Lys Asp Trp Glu Asn Gln Met Glu Tyr Leu
180 185 190
Glu Ser Lys Thr Asp Ser Asn Pro Asn Phe Val Tyr Arg Met Thr Thr
195 200 205
Leu Tyr Glu Phe Tyr Lys Ser His Ile Asp Glu Val Asn Ser Lys Met
210 215 220
Glu Thr Met Ser Ile Asp Ser Leu Ile Lys Phe Gly Gly Cys Arg Arg
225 230 235 240
Lys Asp Ser Lys Lys Ser Met Tyr Ile Met Gly Gly Ser Asn Thr Pro
245 250 255
Phe Asp Ile Thr Gln Ile Gly Gly Asn Ser Leu Asn Ile Lys Phe Ser
260 265 270
Lys Asn Leu Asn Val Asp Val Phe Gly Arg Tyr Asp Val Ile Lys Asp
275 280 285
Asn Thr Leu Leu Val Asp Ile Ile Asn Gly His Gly Ala Ser Phe Val
290 295 300
Leu Lys Ile Ile Asn Asp Glu Ile Tyr Ile Asp Ile Asn Val Ser Val
305 310 315 320
Pro Phe Asp Lys Lys Ile Ala Thr Thr Asn Lys Val Val Gly Ile Asp
325 330 335
Val Asn Ile Lys His Met Leu Leu Ala Thr Asn Ile Leu Asp Asp Gly
340 345 350
Asn Val Asn Gly Tyr Val Asn Ile Tyr Lys Glu Val Ile Asn Asp Ser

355 360 365
Asp Phe Lys Lys Val Cys Asn Ser Thr Val Met Lys Tyr Phe Thr Asp
370 375 380
Phe Ser Lys Phe Val Thr Phe Cys Pro Leu Glu Phe Asp Phe Leu Phe
385 390 395 400
Ser Arg Val Cys Asn Gln Lys Gly Ile Tyr Asn Asp Asn Ser Ala Met
405 410 415
Glu Lys Ser Phe Ser Asp Val Leu Asn Lys Leu Lys Trp Asn Phe Ile
420 425 430
Glu Thr Gly Asp Asn Thr Lys Arg Ile Tyr Ile Glu Asn Val Met Lys
435 440 445
Leu Arg Ser Gln Met Lys Ala Tyr Ala Ile Val Lys Asn Ala Tyr Tyr
450 455 460
Lys Gln Gln Ser Glu Tyr Asp Phe Gly Lys Ser Glu Glu Phe Ile Gln
465 470 475 480
Glu His Pro Phe Ser Asn Thr Asp Lys Gly Ile Glu Ile Leu Asn Lys
485 490 495
Leu Asp Asn Ile Ser Lys Lys Ile Leu Gly Cys Arg Asn Asn Ile Ile
500 505 510
Gln Tyr Ser Tyr Asn Leu Phe Glu Ile Asn Gly Tyr Asp Met Ile Ser
515 520 525
Leu Glu Lys Leu Thr Ser Ser Gln Phe Lys Lys Lys Pro Phe Pro Thr
530 535 540
Val Asn Ser Leu Leu Lys Tyr His Lys Ile Leu Gly Cys Thr Gln Glu
545 550 555 560
Glu Met Glu Lys Lys Asp Ile Tyr Ser Val Ile Lys Lys Gly Tyr Tyr
565 570 575
Asp Ile Ile Phe Asp Asn Asp Val Val Thr Asp Ala Lys Leu Ser Thr
580 585 590
Lys Gly Glu Leu Ser Lys Phe Lys Asp Asp Phe Phe Asn Leu Met Ile
595 600 605
Lys Ser Ile His Phe Ala Asp Ile Lys Asp Tyr Phe Ile Thr Leu Ser
610 615 620
Asn Asn Gly Thr Ala Gly Val Ser Leu Val Pro Ser Tyr Phe Thr Ser
625 630 635 640
Gln Met Asp Ser Ile Asp His Lys Ile Tyr Phe Val Gln Asp Asn Lys
645 650 655
Ser Gly Lys Leu Lys Leu Ala Asn Lys His Lys Val Arg Ser Ser Gln
660 665 670

Glu Lys His Ile Asn Gly Leu Asn Ala Asp Tyr Asn Ala Ala Arg Asn
675 680 685
Ile Ala Tyr Ile Met Glu Asn Thr Asp Cys Arg Asn Met Phe Met Lys
690 695 700
Gln Ser Arg Thr Asp Lys Ser Leu Tyr Asn Lys Pro Ser Tyr Glu Thr
705 710 715 720
Phe Ile Lys Thr Gln Gly Ser Ala Val Ala Lys Leu Lys Lys Glu Gly
725 730 735
Phe Val Lys Ile Leu Asp Glu Ala Ser Val
740 745

<210> 152
<211> 29
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 152
Tyr Glu Thr Phe Ile Lys Thr Gln Gly Ser Ala Val Ala Lys Leu Lys
1 5 10 15
Lys Glu Gly Phe Val Lys Ile Leu Asp Glu Ala Ser Val
20 25

<210> 153
<211> 738
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 153
Met Ala Ser His Ser Ser Leu Ser Asn Asn Gln Ile Phe Lys Thr Phe
1 5 10 15
Ser Phe Lys Val Lys Ser Ser Asn Leu Ser Lys Asp Phe Phe Asp Val
20 25 30
Ile Lys Glu Tyr Gln Glu Tyr Tyr Asn Lys Cys Ser Asp Val Ile Leu
35 40 45
Glu Asn Leu Thr Cys Ile Lys Ile Gly Asp Ile Phe Asp Met Ile Pro
50 55 60
Glu Lys Ser Lys Lys Ser Asp Tyr Ala Gln Tyr Ala Ile Ser Asp Glu
65 70 75 80
Trp Lys Asn Val Pro Leu Tyr Asn Ile Phe Ser Lys Ala Phe Ala Pro

85 90 95

Met His Arg Asp Asn Leu Leu Tyr Ile Trp Leu Thr Lys Leu Val Pro
100 105 110

Tyr Thr Gly Asn Leu Leu Lys Ile Ser Asp Thr Phe Tyr Lys Arg Asn
115 120 125

Gly Phe Ile Lys Ser Val Ile Ser Asn Tyr Thr Thr Ser Phe Thr Asn
130 135 140

Ile Lys Pro Lys Val Lys Phe Gln Lys Leu Thr Gly Asp Asp Ser His
145 150 155 160

Glu Met Leu Leu Thr Gln Thr Ile Cys Asp Met Val Lys Phe Asn Leu
165 170 175

Tyr Asp Ile Lys Ser Trp Glu Glu Met Val Ser Tyr Phe Glu Met Lys
180 185 190

Ser Glu Thr Ser Glu Asp Thr Leu Asn Arg Ile His Thr Leu Phe Asp
195 200 205

Phe Tyr Lys Asn Asn Thr Pro Glu Val Glu Asp Lys Tyr Asn Glu Leu
210 215 220

Val Thr Glu Ser Leu Ser Lys Phe Gly Gly Cys Arg Arg Asp Met Ser
225 230 235 240

Lys Leu Thr Met Ser Ile Gln Leu Ser Lys Lys Val Ile Lys Val Thr
245 250 255

His Gly Tyr Asn Thr Leu Asn Tyr Lys Tyr Gly Lys Leu Ile Asp Leu
260 265 270

Glu Leu Trp Gly Arg Lys Asp Ile Ile Asn Asn Asp Glu Leu Leu Ile
275 280 285

Asn Leu Glu Asn Val Cys Glu Met Ile Val Phe Lys Ile Lys Asn Gly
290 295 300

Glu Ile Tyr Val Asp Ile Pro Phe Lys Val Asp Phe Ile Lys Lys Asp
305 310 315 320

Gln Thr Ile Asp Lys Ile Ala Gly Val Asp Ala Asn Ile Lys His Met
325 330 335

Leu Leu Ser Thr Ser Val Lys Asp Glu Asn Leu Ile Gly Tyr Thr Asn
340 345 350

Ile Tyr Lys Glu Val Ile Asn Asp Ser Asp Phe Gln Lys Val Cys Asp
355 360 365

Asn Lys Thr Met Lys Ile Leu Gln Glu Ile Ser Asn Tyr Val Thr Phe
370 375 380

Ala Pro Ile Glu Phe Asp Met Leu Phe Ser Arg Ile Ser Lys Gln Arg
385 390 395 400

Glu Met Lys Asp Lys Tyr Ile Asn Met Glu Ile Ala Phe Thr Asn Val
405 410 415
Leu Asn Lys Leu Lys Gln Lys Phe Ile Val Asn Ser Asp Asn Lys Asn
420 425 430
Arg Ile Tyr Ile Glu Ser Ile Leu Lys Ile Arg Ser Gln Leu Lys Ser
435 440 445
Tyr Ala Ile Leu Glu Glu Val Lys Tyr Lys Lys Ala Ser Glu Tyr Asp
450 455 460
Ser Ala Ile Val Glu Glu Phe Gly Val Glu Tyr Leu Glu Thr His Pro
465 470 475 480
Phe Lys Asp Thr Glu Thr Tyr Lys Glu Ile Asn Lys Lys Ile Leu Asn
485 490 495
Ile Ser Glu Asn Ile Ile Gly Cys Arg Asn Asn Ile Ile Gln Tyr Ala
500 505 510
Tyr Lys Ile Phe Glu Ser Asn Gly Phe Asp Met Ile Ser Leu Glu Asn
515 520 525
Leu Thr Asn Ser Asn Phe Lys Lys Glu Lys Asn Met Pro Thr Ile Lys
530 535 540
Ser Leu Leu Ser Tyr His His Val Leu Gly Lys Thr Asn Glu Glu Ile
545 550 555 560
Glu Lys Leu Asp Val Tyr Ser Val Ile Lys Lys Gly Tyr Tyr Thr Phe
565 570 575
Glu Tyr Lys Asp Gly Lys Val Val Asn Ala Lys Leu Ser Glu Ile Gly
580 585 590
Glu Met Ile Lys Ile Lys Thr Thr Met Phe Asn Met Met Ile Lys Ser
595 600 605
Ile His Phe Ala Glu Ile Lys Asp Tyr Phe Ile Thr Leu Ala Asn Asn
610 615 620
Gly Glu Val Gly Val Ser Leu Val Pro Ser Tyr Tyr Thr Ser Gln Met
625 630 635 640
Asp Ser Thr Asp His Lys Val Phe Gly Leu Leu Ser Lys Lys Gly Lys
645 650 655
Trp Thr Leu Val Asp Lys Arg Lys Val Arg Lys Asn Gln Glu Thr His
660 665 670
Ile Asn Gly Leu Asn Ala Asp Tyr Asn Ala Ala Lys Asn Ile Ala Phe
675 680 685
Ile Leu Ser Asp Glu Val Trp Arg Asn Lys Phe Thr Lys Lys Thr Lys
690 695 700
Thr Pro Lys Tyr Asn Thr Pro Ser Tyr Tyr Thr Ser Ile Asn Ser Gln

705 710 715 720
Gly Lys Met Leu Arg Ala Leu Lys Ser Leu Lys Ala Phe Lys Glu Phe
725 730 735
Lys Ile

<210> 154
<211> 692
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 154

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

Val Phe Gly Arg Tyr Asp Val Ile Lys Asp Asn Thr Leu Leu Val Asp
225 230 235 240
Ile Ile Asn Glu His Gly Ala Ser Phe Val Leu Lys Ile Ile Asn Asp
245 250 255
Glu Ile Tyr Ile Asp Ile Asn Val Ser Val Pro Phe Asp Lys Lys Ile
260 265 270
Ala Thr Thr Asn Lys Val Val Gly Ile Asp Val Asn Ile Lys His Met
275 280 285
Leu Leu Ala Thr Asn Ile Leu Asp Asp Gly Asn Val Asn Gly Tyr Val
290 295 300
Asn Ile Tyr Lys Glu Val Ile Asn Asp Ser Asp Phe Lys Lys Val Cys
305 310 315 320
Asn Ser Thr Val Met Gln Tyr Phe Thr Asp Phe Ser Lys Phe Val Thr
325 330 335
Phe Cys Pro Leu Glu Phe Asp Phe Leu Phe Ser Arg Val Cys Asn Gln
340 345 350
Lys Gly Ile Tyr Asn Asp Asn Ser Ala Met Glu Lys Ser Phe Ser Asp
355 360 365
Val Leu Asn Lys Leu Lys Trp Asn Phe Ile Glu Thr Gly Asp Asn Thr
370 375 380
Lys Arg Ile Tyr Ile Glu Asn Val Met Lys Leu Arg Ser Gln Met Lys
385 390 395 400
Ala Tyr Ala Ile Val Lys Asn Ala Tyr Tyr Lys Gln Gln Ser Glu Tyr
405 410 415
Asp Phe Gly Lys Ser Glu Glu Phe Ile Gln Glu His Pro Phe Ser Asn
420 425 430
Thr Asp Lys Gly Ile Glu Ile Leu Asn Lys Leu Asp Asn Ile Ser Lys
435 440 445
Lys Ile Leu Gly Cys Arg Asn Asn Ile Ile Gln Tyr Ser Tyr Asn Leu
450 455 460
Phe Glu Ile Asn Gly Tyr Asp Met Ile Ser Leu Glu Lys Leu Thr Ser
465 470 475 480
Ser Gln Phe Lys Lys Lys Pro Phe Pro Thr Val Asn Ser Leu Leu Lys
485 490 495
Tyr His Lys Ile Leu Gly Cys Thr Gln Glu Glu Met Glu Lys Lys Asp
500 505 510
Ile Tyr Ser Val Ile Lys Lys Gly Tyr Tyr Asp Ile Ile Phe Asp Asn
515 520 525
Asp Val Val Thr Asp Ala Lys Leu Ser Thr Lys Gly Glu Leu Ser Lys

530 535 540
Phe Lys Asp Asp Phe Phe Asn Leu Met Ile Lys Ser Ile His Phe Ala
545 550 555 560
Asp Ile Lys Asp Tyr Phe Ile Thr Leu Ser Asn Asn Gly Thr Ala Gly
565 570 575
Val Ser Leu Val Pro Ser Tyr Phe Thr Ser Gln Met Asp Ser Ile Asp
580 585 590
His Lys Ile Tyr Phe Val Gln Asp Asn Lys Ser Gly Lys Leu Lys Leu
595 600 605
Ala Asn Lys His Lys Val Arg Ser Ser Gln Glu Lys His Ile Asn Gly
610 615 620
Leu Asn Ala Asp Tyr Asn Ala Ala Arg Asn Ile Ala Tyr Ile Met Glu
625 630 635 640
Asn Thr Glu Cys Arg Asn Met Phe Met Lys Gln Ser Arg Thr Asp Lys
645 650 655
Ser Leu Tyr Asn Lys Pro Ser Tyr Glu Thr Phe Ile Lys Thr Gln Gly
660 665 670
Ser Ala Val Ala Lys Leu Lys Lys Glu Gly Phe Val Lys Ile Leu Asp
675 680 685
Glu Ala Ser Val
690

<210> 155
<211> 441
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 155
Lys Ile Ile Asn Asp Glu Ile Tyr Ile Asp Ile Asn Val Ser Val Pro
1 5 10 15
Phe Asp Lys Lys Ile Ala Thr Thr Asn Lys Val Val Gly Ile Asp Val
20 25 30
Asn Ile Lys His Met Leu Leu Ala Thr Asn Ile Leu Asp Asp Gly Asn
35 40 45
Val Asn Gly Tyr Val Asn Ile Tyr Lys Glu Val Ile Asn Asp Ser Asp
50 55 60
Phe Lys Lys Val Cys Asn Ser Thr Val Met Lys Tyr Phe Thr Asp Phe
65 70 75 80
Ser Lys Phe Val Thr Phe Cys Pro Leu Glu Phe Asp Phe Leu Phe Ser

85 90 95

Arg Val Cys Asn Gln Lys Gly Ile Tyr Asn Asp Asn Ser Ala Met Glu
100 105 110

Lys Ser Phe Ser Asp Val Leu Asn Arg Leu Lys Trp Asn Phe Ile Glu
115 120 125

Thr Gly Asp Asn Thr Lys Arg Ile Tyr Ile Glu Asn Val Met Lys Leu
130 135 140

Arg Ser Gln Met Lys Ala Tyr Ala Ile Val Lys Asn Ala Tyr Tyr Lys
145 150 155 160

Gln Gln Ser Glu Tyr Asp Phe Gly Lys Ser Glu Glu Phe Ile Gln Glu
165 170 175

His Pro Phe Ser Asn Thr Asp Lys Gly Ile Glu Ile Leu Asn Lys Leu
180 185 190

Asp Asn Ile Ser Lys Lys Ile Leu Gly Cys Arg Asn Asn Ile Ile Gln
195 200 205

Tyr Ser Tyr Asn Leu Phe Glu Ile Asn Gly Tyr Asp Met Ile Ser Leu
210 215 220

Glu Lys Leu Thr Ser Ser Gln Phe Lys Lys Lys Pro Phe Pro Thr Val
225 230 235 240

Asn Ser Leu Leu Lys Tyr His Lys Ile Arg Gly Gly Thr Gln Glu Glu
245 250 255

Met Glu Lys Lys Asp Ile Tyr Ser Val Ile Lys Lys Gly Tyr Tyr Asp
260 265 270

Ile Ile Phe Asp Asn Asp Val Val Thr Asp Ala Lys Leu Ser Thr Lys
275 280 285

Gly Glu Leu Ser Lys Phe Lys Asp Asp Phe Phe Asn Leu Met Ile Lys
290 295 300

Ser Ile His Phe Ala Asp Ile Lys Asp Tyr Phe Ile Thr Leu Ser Asn
305 310 315 320

Asn Gly Thr Ala Gly Val Ser Leu Val Pro Ser Tyr Phe Thr Ser Gln
325 330 335

Met Asp Ser Ile Asp His Lys Ile Tyr Phe Val Gln Asp Asn Lys Ser
340 345 350

Gly Lys Leu Lys Leu Ala Asn Lys His Lys Val Arg Ser Ser Gln Glu
355 360 365

Lys His Ile Asn Gly Leu Asn Ala Asp Tyr Asn Ala Ala Arg Asn Ile
370 375 380

Ala Tyr Ile Met Glu Asn Asn Glu Cys Arg Asn Met Phe Met Lys Gln
385 390 395 400

Ser Arg Thr Asp Lys Ser Leu Tyr Asn Lys Pro Ser Tyr Glu Thr Phe
405 410 415
Ile Lys Thr Gln Gly Ser Ala Val Ala Lys Leu Lys Lys Glu Gly Phe
420 425 430
Val Lys Ile Leu Asp Glu Ala Ser Val
435 440

<210> 156
<211> 397
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 156

Asp Asp Gly Asn Val Asn Gly Tyr Val Asn Ile Tyr Lys Glu Val Ile
1 5 10 15
Asn Asp Ser Asp Phe Lys Lys Val Cys Asn Ser Thr Val Met Lys Tyr
20 25 30
Phe Thr Asp Phe Ser Lys Phe Val Thr Phe Cys Pro Leu Glu Phe Asp
35 40 45
Phe Leu Phe Ser Arg Val Cys Asn Gln Lys Gly Ile Tyr Asn Asp Asn
50 55 60
Ser Ala Met Glu Lys Ser Phe Ser Asp Val Leu Asn Lys Leu Lys Trp
65 70 75 80
Asn Phe Ile Glu Thr Gly Asp Asn Thr Lys Arg Ile Tyr Ile Glu Asn
85 90 95
Val Met Lys Leu Arg Ser Gln Met Lys Ala Tyr Ala Ile Val Lys Asn
100 105 110
Ala Tyr Tyr Lys Gln Gln Ser Glu Tyr Asp Phe Gly Lys Ser Glu Glu
115 120 125
Phe Ile Gln Glu His Pro Phe Ser Asn Thr Asp Lys Gly Ile Glu Ile
130 135 140
Leu Asn Lys Leu Asp Asn Ile Ser Lys Lys Ile Leu Gly Cys Arg Asn
145 150 155 160
Asn Ile Ile Gln Tyr Ser Tyr Asn Leu Phe Glu Ile Asn Gly Tyr Asp
165 170 175
Met Ile Ser Leu Glu Lys Leu Thr Ser Ser Gln Phe Lys Lys Lys Pro
180 185 190
Phe Pro Thr Val Asn Ser Leu Leu Lys Tyr His Lys Ile Leu Gly Cys
195 200 205

Thr Gln Glu Glu Met Glu Lys Lys Asp Ile Tyr Ser Val Ile Lys Lys
210 215 220
Gly Tyr Tyr Asp Ile Ile Phe Asp Asn Asp Val Val Thr Asp Ala Lys
225 230 235 240
Leu Ser Thr Lys Gly Glu Leu Ser Lys Phe Lys Asp Asp Phe Phe Asn
245 250 255
Leu Met Ile Lys Ser Ile His Phe Ala Asp Ile Lys Asp Tyr Phe Ile
260 265 270
Thr Leu Ser Asn Asn Gly Thr Ala Gly Val Ser Leu Val Pro Ser Tyr
275 280 285
Phe Thr Ser Gln Met Asp Ser Ile Asp His Lys Ile Tyr Phe Val Gln
290 295 300
Asp Asn Lys Ser Gly Lys Leu Lys Leu Ala Asn Lys His Lys Val Arg
305 310 315 320
Ser Ser Gln Glu Lys His Ile Asn Gly Leu Asn Ala Asp Tyr Asn Ala
325 330 335
Ala Arg Asn Ile Ala Tyr Ile Met Glu Asn Thr Glu Cys Arg Asn Met
340 345 350
Phe Met Lys Gln Ser Arg Thr Asp Lys Ser Leu Tyr Asn Lys Pro Ser
355 360 365
Tyr Glu Thr Phe Ile Lys Thr Gln Gly Ser Ala Val Ala Lys Leu Lys
370 375 380
Lys Asp Gly Phe Val Lys Ile Leu Asp Glu Ala Ser Ala
385 390 395

<210> 157
<211> 746
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 157

Met Ala Ser His Lys Lys Thr Glu Ser Asn Gln Ile Ile Lys Thr Phe
1 5 10 15
Ser Phe Lys Ile Lys Asn Ala Asn Gly Leu Ser Leu Asp Val Leu Asn
20 25 30
Asp Ala Ile Thr Glu Tyr Gln Asn Tyr Tyr Asn Ile Cys Ser Asp Trp
35 40 45
Ile Lys Asp His Leu Thr Met Lys Ile Gly Glu Leu Tyr Lys Tyr Ile
50 55 60

Pro Asp Glu Lys Lys Asn Ser Gly Tyr Ala Leu Thr Leu Ile Ser Asp
65 70 75 80
Glu Trp Lys Asp Lys Pro Met Tyr Met Met Phe Lys Lys Gly Tyr Pro
85 90 95
Ala Asn Ser Arg Asp Asn Ala Ile Tyr Glu Ala Leu Asn Thr Cys Asn
100 105 110
Thr Glu His Tyr Thr Gly Asn Ile Leu Asn Phe Ser Asp Thr Tyr Tyr
115 120 125
Arg Arg Phe Gly Tyr Val Ala Ser Ala Ile Ser Asn Tyr Val Thr Lys
130 135 140
Ile Ser Lys Met Ser Thr Gly Ser Arg Ser Lys Asn Ile Ser Asn Asp
145 150 155 160
Ser Asp Val Asp Thr Ile Met Glu Gln Val Ile Tyr Glu Met Glu His
165 170 175
Asn Gly Trp Thr Ser Val Lys Asp Trp Glu Asn Gln Met Glu Tyr Leu
180 185 190
Glu Ser Lys Thr Asp Ser Asn Pro Asn Phe Val Tyr Arg Met Thr Thr
195 200 205
Leu Tyr Glu Phe Tyr Lys Ser His Ile Asp Glu Val Asn Ser Lys Met
210 215 220
Glu Thr Met Ser Ile Asp Ser Leu Ile Lys Phe Gly Gly Cys Arg Arg
225 230 235 240
Lys Asp Ser Lys Lys Ser Met Tyr Ile Met Gly Gly Ser Asn Thr Pro
245 250 255
Phe Asp Ile Thr Gln Ile Asp Gly Asn Ser Leu Asn Ile Lys Phe Ser
260 265 270
Lys Asn Leu Asn Val Asp Val Phe Gly Arg Tyr Asp Val Ile Lys Asp
275 280 285
Asn Thr Leu Leu Val Asp Ile Ile Asn Gly His Gly Ala Ser Phe Val
290 295 300
Leu Lys Ile Ile Asn Gly Glu Ile Tyr Ile Asp Ile Asn Val Ser Val
305 310 315 320
Pro Phe Asp Lys Lys Ile Ala Thr Thr Asn Lys Val Val Gly Ile Asp
325 330 335
Val Asn Ile Lys His Met Leu Leu Ala Thr Asn Ile Leu Asp Asp Gly
340 345 350
Asn Val Asn Gly Tyr Val Asn Ile Tyr Lys Glu Val Ile Asn Asp Ser
355 360 365
Asp Phe Lys Lys Val Cys Asn Ser Thr Val Met Lys Tyr Phe Thr Asp

370 375 380
Phe Ser Lys Phe Val Thr Phe Cys Pro Leu Glu Phe Asp Phe Leu Phe
385 390 395 400
Ser Arg Val Cys Asn Gln Lys Gly Ile Tyr Asn Asp Asn Ser Ala Met
405 410 415
Glu Lys Ser Phe Ser Asp Val Leu Asn Lys Leu Lys Trp Asn Phe Ile
420 425 430
Glu Thr Gly Asp Asn Thr Lys Arg Ile Tyr Ile Glu Asn Val Met Lys
435 440 445
Leu Arg Ser Gln Met Lys Ala Tyr Ala Ile Val Lys Asn Ala Tyr Tyr
450 455 460
Lys Gln Gln Ser Glu Tyr Asp Phe Gly Lys Ser Glu Glu Phe Ile Gln
465 470 475 480
Glu His Pro Phe Ser Asn Thr Asp Lys Gly Ile Glu Ile Leu Asn Lys
485 490 495
Leu Asp Asn Ile Ser Lys Lys Ile Leu Gly Cys Arg Asn Asn Ile Ile
500 505 510
Gln Tyr Ser Tyr Asn Leu Phe Glu Ile Asn Gly Tyr Asp Met Ile Ser
515 520 525
Leu Glu Lys Leu Thr Ser Ser Gln Phe Lys Lys Lys Pro Phe Pro Thr
530 535 540
Val Asn Ser Leu Leu Lys Tyr His Lys Ile Leu Gly Cys Thr Gln Glu
545 550 555 560
Glu Met Glu Lys Lys Asp Ile Tyr Ser Val Ile Lys Lys Gly Tyr Tyr
565 570 575
Asp Ile Ile Phe Asp Asn Asp Val Val Thr Asp Ala Lys Leu Ser Thr
580 585 590
Lys Gly Glu Leu Ser Lys Phe Lys Asp Asp Phe Phe Asn Leu Met Ile
595 600 605
Lys Ser Ile His Phe Ala Asp Ile Lys Asp Tyr Phe Ile Thr Leu Ser
610 615 620
Asn Asn Gly Thr Ala Gly Val Ser Leu Val Pro Ser Tyr Phe Thr Ser
625 630 635 640
Gln Met Asp Ser Ile Asp His Lys Ile Tyr Phe Val Gln Asp Asn Lys
645 650 655
Ser Gly Lys Leu Lys Leu Ala Asn Lys His Lys Val Arg Ser Ser Gln
660 665 670
Glu Lys His Ile Asn Gly Leu Asn Ala Asp Tyr Asn Ala Ala Arg Asn
675 680 685

Ile Ala Tyr Ile Met Glu Asn Thr Glu Cys Arg Asn Met Phe Met Lys
690 695 700
Gln Ser Arg Thr Asp Lys Ser Leu Tyr Asn Lys Pro Ser Tyr Glu Thr
705 710 715 720
Phe Ile Lys Thr Gln Gly Ser Ala Val Ala Lys Leu Lys Lys Glu Gly
725 730 735
Phe Val Lys Ile Leu Asp Glu Ala Ser Val
740 745

<210> 158
<211> 736
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 158

Met Ala His Lys Lys Asn Val Gly Ala Glu Ile Val Lys Thr Tyr Ser
1 5 10 15
Phe Lys Val Lys Asn Thr Asn Gly Ile Thr Met Glu Lys Leu Met Asn
20 25 30
Ala Ile Asp Glu Phe Gln Ser Tyr Tyr Asn Leu Cys Ser Asp Trp Ile
35 40 45
Cys Lys Asn Leu Thr Thr Met Thr Ile Gly Asp Leu Asp Gln Tyr Ile
50 55 60
Pro Glu Lys Ala Lys Gly Asn Thr Tyr Ala Thr Val Leu Leu Asp Glu
65 70 75 80
Ala Trp Lys Asn Gln Pro Leu Tyr Lys Ile Phe Gly Lys Lys Tyr Ser
85 90 95
Ser Asn Asn Arg Asn Asn Ala Leu Tyr Cys Ala Leu Ser Ser Val Ile
100 105 110
Asp Met Thr Lys Glu Asn Val Leu Gly Phe Ser Lys Thr His Tyr Ile
115 120 125
Arg Asn Asp Tyr Ile Leu Asn Val Ile Ser Asn Tyr Ala Ser Lys Leu
130 135 140
Ser Lys Leu Asn Thr Gly Val Lys Ser Arg Ala Ile Lys Glu Thr Ser
145 150 155 160
Asp Glu Ala Thr Ile Ile Glu Gln Val Ile Tyr Glu Met Glu His Asn
165 170 175
Lys Trp Glu Ser Ile Glu Asp Trp Lys Asn Gln Ile Glu Tyr Leu Asn
180 185 190

Ser Lys Thr Asp Tyr Asn Pro Thr Tyr Met Glu Arg Met Lys Thr Leu
195 200 205
Ser Ala Tyr Tyr Ser Thr His Lys Ser Glu Val Asp Ala Lys Met Gln
210 215 220
Glu Met Ala Val Glu Asn Leu Val Lys Phe Gly Gly Cys Arg Arg Asn
225 230 235 240
Asn Ser Lys Lys Ser Met Phe Ile Met Gly Ser Asn Thr Thr Asn Tyr
245 250 255
Thr Ile Ser Tyr Ile Gly Gly Asn Ser Phe Asn Ile Asn Phe Ala Asn
260 265 270
Ile Leu Asn Phe Asp Val Tyr Gly Arg Arg Asp Val Val Lys Asn Gly
275 280 285
Glu Val Leu Val Asp Ile Met Ala Asn His Gly Asp Ser Ile Val Leu
290 295 300
Lys Ile Val Asn Gly Glu Leu Tyr Ala Asp Val Pro Cys Ser Val Thr
305 310 315 320
Leu Asn Lys Val Glu Ser Asn Phe Asp Lys Val Val Gly Ile Asp Val
325 330 335
Asn Met Lys His Met Leu Leu Ser Thr Ser Ile Thr Asp Asn Gly Ser
340 345 350
Ser Asp Phe Leu Asn Ile Tyr Lys Glu Met Ser Asn Asn Ala Glu Phe
355 360 365
Met Ala Leu Cys Pro Glu Glu Asp Arg Lys Tyr Tyr Lys Asp Ile Ser
370 375 380
Lys Tyr Val Thr Phe Ala Pro Leu Glu Leu Asp Leu Leu Phe Ser Arg
385 390 395 400
Ile Ser Lys Gln Gly Lys Val Lys Met Glu Lys Val Tyr Ser Glu Ile
405 410 415
Leu Glu Ala Leu Lys Trp Lys Phe Phe Ala Asn Gly Asp Asn Lys Asn
420 425 430
Arg Ile Tyr Val Glu Ser Ile Gln Lys Ile Arg Gln Gln Ile Lys Ala
435 440 445
Leu Cys Val Ile Lys Asn Ala Tyr Tyr Glu Gln Gln Ser Ala Tyr Asp
450 455 460
Ile Asp Lys Thr Gln Glu Tyr Ile Glu Thr His Pro Phe Ser Leu Thr
465 470 475 480
Glu Lys Gly Met Ser Ile Lys Ser Lys Met Asp Lys Ile Cys Gln Thr
485 490 495
Ile Ile Gly Cys Arg Asn Asn Ile Ile Asp Tyr Ala Tyr Ser Phe Phe

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

<210> 159
<211> 757
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 159
Met Ala His Lys Lys Asn Leu Glu Gly Glu Asn Leu Gln Val Lys Thr
1 5 10 15
Ile Cys Leu Lys Ala Asn Leu Ser Lys Glu Glu Val Lys Glu Lys Trp

20 25 30

Leu Pro Val Ile Asn Glu Tyr Asn Val Tyr Tyr Asn Arg Met Ser Asp
35 40 45

Tyr Ile Cys Ser Leu Leu Gly Thr Asn Ile Thr Ile Gly Glu Phe Ala
50 55 60

Glu Gln Leu Ser Ile Glu Lys Arg Lys Asn Gly Tyr Phe Thr Ile Cys
65 70 75 80

Gln Asp Asp Lys Phe Lys Asn Glu Ser Leu Tyr Lys Ile Phe His Lys
85 90 95

Ser Phe Pro Ile Asn His Gly Thr Asn Ile Ile Asn Asn Ile Ile Ser
100 105 110

Glu Lys Asn Ile Asp Gln Tyr Asp Gly Asn Thr Leu Gly Phe Arg Pro
115 120 125

Thr Met Tyr Arg Leu Arg Gly Tyr Val Asp Ser Val Ile Gly Asn Tyr
130 135 140

Arg Thr Thr Ile Arg Thr Ile Lys Pro Ser Val Lys Arg His Lys Ile
145 150 155 160

Ser Val Asp Ser Ser Phe Asp Glu Lys Met Glu Gln Cys Ile Tyr Glu
165 170 175

Ile Gln Lys Gly Asn Leu Lys Thr Val Ser Glu Trp Asn Asn Lys Ile
180 185 190

Asp Tyr Leu Leu Ser Lys Ser Asp Met Asn Pro Leu Thr Ile Asp Arg
195 200 205

Phe Asn Leu Leu Arg Asp Phe Tyr Val Asp Asn Glu Thr Glu Val Asn
210 215 220

Glu Lys Ser Asn Asn Ser Ser Ile Glu Gln Leu Val Lys Phe Gly Gly
225 230 235 240

Cys His Arg Lys Gly Asp Asn Met Thr Leu Ser Leu Thr Glu Ala Asn
245 250 255

Phe Ser Ile Glu Glu Ile Asp Asp Ser Tyr Gly Tyr Leu Leu Thr Leu
260 265 270

His Thr Asp Ser Gly Asp Tyr Lys Ile Pro Leu Met Gly Ser Lys Met
275 280 285

Leu Lys Lys Gly Asp Lys Cys Leu Ile Asp Phe Val Asn Asn Cys Lys
290 295 300

Lys Gly Lys Ser Leu Thr Ala Lys Ile Asp Asn Asp Tyr Asn Leu Tyr
305 310 315 320

Phe His Phe Val Val Tyr Ser Asn Phe Glu Lys Ile Glu Asp Asp Asn
325 330 335

Ile Asn Asn Val Val Gly Val Asp Val Asn Ser Lys His Met Leu Leu
340 345 350
Met Thr Asn Val Ile Asp Asp Asn Ile Asp Gly Tyr Val Asn Ile Tyr
355 360 365
Lys Ala Leu Val Asn Asp Asp Glu Phe Lys Ser Leu Val Thr Lys Ser
370 375 380
Glu Tyr Asp Asp Tyr Val Thr Met Ser Lys Tyr Val Thr Phe Cys Pro
385 390 395 400
Ile Glu Leu Lys Tyr Leu Tyr Ala Arg Tyr Cys Val Gln Lys Asp Tyr
405 410 415
Pro Ile Ser Asn Lys Asp Val Ala Ile Glu Gln Cys Ile Ser Arg Val
420 425 430
Ile Asp Lys Leu Cys Lys Glu Thr Leu Asp Ser Arg Ala Asn Asn Tyr
435 440 445
Leu Cys Met Val Arg Arg Ile Arg His Tyr Tyr Lys Ser Tyr Tyr Val
450 455 460
Leu Lys Met Thr Tyr Tyr Asp Lys Met Ser Glu Tyr Asp Thr Asn Met
465 470 475 480
Glu Tyr Asn Asp Ile Ser Thr Thr Ser Lys Glu Thr Met Asp Gln Arg
485 490 495
Arg Phe Glu Asn Ser Phe Arg Glu Thr Asp Cys Ala Lys Glu Ile Leu
500 505 510
Ser Lys Leu Asp Lys Ile Gly Asn Asp Ile Leu Gly Cys Arg Asn Asn
515 520 525
Ile Leu Thr Tyr Ala Tyr Lys Leu Phe Glu Glu Leu Gly Tyr Asp Thr
530 535 540
Ile Ala Leu Glu Asn Leu Glu Ser Ser Gln Phe Asp Lys Met Lys Ser
545 550 555 560
Leu Pro Ser Cys Gln Ser Met Leu Lys Tyr His Lys Leu Glu Gly Lys
565 570 575
Thr Met Glu Glu Val Met Ser Asn Thr Ser Val Lys Ser Leu Ile Glu
580 585 590
Asn Glu Tyr Tyr Asp Phe Ser Leu Asn Asp Asn Lys Thr Val Glu Asn
595 600 605
Ile Thr Tyr Thr Lys Asn Gly Leu Met Lys Lys Gly Phe Asp Glu Phe
610 615 620
Phe Asn Leu Phe Met Lys Ile His Phe Ala Asp Ile Lys Asp Lys Phe
625 630 635 640
Leu Gln Leu Tyr Asn Asn Gly Ser Val Lys Val Ile Leu Val Pro Ser

645 650 655
Tyr Phe Thr Ser Gln Met Asp Ser Asn His Ser Ile Tyr Met Glu Lys
660 665 670
Ser Lys Asn Asp Lys Leu Val Phe Ala Ser Lys His Lys Val Arg Lys
675 680 685
Thr Gln Glu Thr His Leu Asn Gly Leu Asn Ala Asp Tyr Asn Ala Cys
690 695 700
Asn Ile Ala Tyr Ile Ala Lys Asp Ile Lys Trp Arg Glu Lys Phe Cys
705 710 715 720
Lys Lys Thr Ser Asn Asn Gly Tyr Ser Thr Pro Phe Tyr Asp Cys Ala
725 730 735
Thr Lys Asn Gln Ile Glu Met Val Lys Arg Ile Lys Gln Leu Asn Ala
740 745 750
Ile Lys Met Leu Ala
755

<210> 160
<211> 767
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 160
Met Ala His Lys Lys Gln Lys Glu Glu Asn Glu Ile Ile Lys Thr Ile
1 5 10 15
Ser Leu Lys Val Lys Asp Tyr Ala Gly Tyr Pro Ile Val Glu Ala Met
20 25 30
Arg Glu Tyr Thr Lys Tyr Tyr Asn Lys Ile Ser Gln Trp Ile Asn Ser
35 40 45
Asn Leu Leu Thr Ile Lys Ile Gly Glu Leu Ser Ala Phe Met Pro Asp
50 55 60
Glu Cys Lys Thr His Asn Tyr Tyr Thr Tyr Met Met Ser Pro Asp Trp
65 70 75 80
Val Asn Glu Pro Leu Tyr Lys Met Phe Met Lys Gly Phe His Ala Gln
85 90 95
His Cys Asp Asn Ile Leu Phe Asn Val Val Lys Thr Leu Asn Ile Asp
100 105 110
Glu Tyr Ala Gly Asn Ser Leu Gly Leu Ser Ala Ser Cys Phe Arg Arg
115 120 125
Ser Gly Tyr Phe Gln Asn Val Val Ser Asn Tyr Lys Ser Lys Phe Ala

130 135 140

Asn Pro His Ile Ser Ile Arg Arg Lys Asn Leu Ser Asp Leu Pro Thr
145 150 155 160
Glu Asp Glu Leu Val Glu Gln Cys Ile Tyr Glu Ile Gln Asn Gly Leu
165 170 175
Ser Ser Lys Thr Lys Trp Glu Glu Gln Ile Glu Tyr Leu Lys Glu Arg
180 185 190
Asp Asp Ser Lys Gln Ile Tyr Leu Thr Arg Leu Asn Thr Leu Phe Met
195 200 205
Tyr Tyr Lys Ala Asn Lys Asp Phe Val Asp Glu Gln Ile Gln Ile Lys
210 215 220
Ser Val Glu Ser Leu Ala Asn Phe Gly Gly Cys Val Arg Lys Asp Asp
225 230 235 240
Lys Leu Ser Met Asn Leu Val Phe Ser Ser Asn Ser Pro Tyr Lys Val
245 250 255
Val Leu Asn Glu Lys Arg Asn Gly Tyr Ile Leu Ser Tyr Ser Asn Asn
260 265 270
Phe Ser Ile Glu Leu Tyr Gly Asn Arg Met Gly Leu Leu Asn Gly Val
275 280 285
Glu Val Phe Asn Val Gly Asp Lys His Ser Asn Asn Ile Thr Phe Lys
290 295 300
Met Asp Asn Asp Glu Leu Phe Val Asn Ile Pro Val Ser Val Asn Phe
305 310 315 320
Val Lys Lys Ala Asn Glu Thr Asn Lys Val Val Gly Val Asp Val Asn
325 330 335
Leu Lys His Ser Ile Phe Ala Thr Asn Ile Ile Asp Asp Gly Lys Leu
340 345 350
Asp Gly Phe Val Asn Ile Tyr Arg Glu Leu Leu Asn Asp Val Asp Phe
355 360 365
Val Lys Ser Cys Pro Asn Glu Leu Leu Asn Phe Ile Leu Asp Val Glu
370 375 380
Lys Tyr Ala Phe Phe Met Pro Leu Glu Leu Gly Leu Leu Ser Ser Arg
385 390 395 400
Val Met Asn Gln Cys Gly Tyr Ser Thr Ile Gly Lys Tyr Glu Lys Leu
405 410 415
Phe Thr Thr Glu Glu His Phe Phe Arg Val Leu Arg Gln Leu Glu Lys
420 425 430
Arg Phe Gln Glu Ser Gly Glu Asn Gln Lys Arg Ile Tyr Ile Glu Asn
435 440 445

Val Ile Lys Met Arg Ala Gln Val Lys Ala Tyr Phe Thr Leu Lys Tyr
450 455 460
Ala Thr Asn Lys Ala Asn Lys Asp Tyr Asp Leu Lys Met Gly Phe Val
465 470 475 480
Asp Glu Ser Thr Ala Asn Lys Glu Thr Met Asp Gln Arg Arg Phe Glu
485 490 495
Asn Gln Phe Val Asn Thr Tyr Thr Ala Lys Glu Ile Leu Gly Lys Met
500 505 510
Arg Arg Ile Ala Asn Val Ile Thr Ser Cys Arg Asn Asn Ile Ile Cys
515 520 525
Tyr Met Tyr Lys Ile Phe Glu Asn Asn Gly Tyr Gly Val Val Ala Leu
530 535 540
Glu Lys Leu Gln Ser Ser Gln Met Lys Lys Glu Lys Arg Ile Pro Ser
545 550 555 560
Leu Leu Ser Leu Leu Lys Lys Gln Lys Val Glu Gly Tyr Thr Ile Asn
565 570 575
Glu Leu Lys Asp Lys Ser Val Phe Lys Phe Ile Glu Arg Gly Tyr Tyr
580 585 590
Thr Phe Asp Phe Asp Asp Asp Asn Lys Ile Thr Gly Val Gln Phe Ser
595 600 605
Asp Ala Gly Glu Val Val Asn Met Glu Thr Glu Leu His Asn Leu Ala
610 615 620
Leu Lys Thr Ile His Phe Ala Asp Ala Lys Asp Tyr Phe Val Thr Leu
625 630 635 640
Ser Asn Asn Gly Ser Val Ser Val Ala Leu Val Pro Ser Gln Phe Thr
645 650 655
Ser Gln Met Asp Ser Thr Lys His Val Leu Tyr Ala Lys Lys Asn Asn
660 665 670
Lys Gly Lys Leu Gly Ile Val Ser Glu His Glu Val Arg Pro Lys Gln
675 680 685
Glu Cys His Ile Asn Gly Leu Asn Gly Asp Tyr Asn Ala Ala Cys Asn
690 695 700
Ile Ala Tyr Ile Phe Glu Asn Asp Glu Trp Arg Asn Ala Phe Met Lys
705 710 715 720
Met Asn Pro Asn Glu Tyr Gly Lys Ala Leu Phe Glu Thr Asn Met Glu
725 730 735
Ser Thr Ser Thr Ile Ile Asn Thr Leu Lys Lys Ile Asn Pro Asp Asn
740 745 750
Ile Ile Ser Phe Asp Glu Tyr Glu Lys Thr Lys Lys Val Ala Ala

755 760 765

<210> 161
<211> 889
<212> PRT
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 161

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

245 250 255
Glu His Tyr Thr Gly Asn Ile Leu Asn Phe Ser Asp Thr Tyr Tyr Arg
260 265 270
Arg Phe Gly Tyr Val Ala Ser Thr Ile Ser Asn Tyr Val Thr Lys Ile
275 280 285
Ser Lys Met Ser Thr Gly Ser Arg Ser Lys Asn Ile Ser Asn Asp Ser
290 295 300
Asp Val Asp Thr Ile Met Glu Gln Val Ile Tyr Glu Met Glu His Asn
305 310 315 320
Gly Trp Thr Ser Val Lys Asp Trp Glu Asn Gln Met Glu Tyr Leu Glu
325 330 335
Ser Lys Thr Asp Ser Asn Pro Asn Phe Val Tyr Arg Met Thr Thr Leu
340 345 350
Tyr Glu Phe Tyr Lys Ser His Ile Asp Glu Val Asn Ser Lys Met Glu
355 360 365
Thr Met Ser Ile Asp Ser Leu Ile Lys Phe Gly Gly Cys Arg Arg Lys
370 375 380
Asp Ser Lys Lys Ser Met Tyr Ile Met Gly Gly Ser Asn Thr Pro Phe
385 390 395 400
Asp Ile Thr Gln Ile Asp Gly Asn Ser Leu Asn Ile Lys Phe Ser Lys
405 410 415
Asn Leu Asn Val Asp Val Phe Gly Arg Tyr Asp Val Ile Lys Asp Asn
420 425 430
Thr Leu Leu Val Asp Ile Ile Asn Gly His Gly Ala Ser Phe Val Leu
435 440 445
Lys Ile Ile Asn Asp Glu Ile Tyr Ile Asp Ile Asn Val Ser Val Pro
450 455 460
Phe Asp Lys Lys Ile Ala Thr Thr Asn Lys Val Val Gly Val Asp Val
465 470 475 480
Asn Ile Lys His Met Leu Leu Ala Thr Asn Ile Leu Asp Asp Gly Asn
485 490 495
Val Asn Gly Tyr Val Asn Ile Tyr Lys Glu Val Ile Asn Asp Ser Asp
500 505 510
Phe Lys Lys Val Cys Asn Ser Thr Val Met Lys Tyr Phe Thr Asp Phe
515 520 525
Ser Lys Phe Val Thr Phe Cys Pro Leu Glu Phe Asp Phe Leu Phe Ser
530 535 540
Arg Val Cys Asn Gln Lys Gly Ile Tyr Asn Asp Asn Ser Ala Met Glu
545 550 555 560

Lys Ser Phe Ser Asp Val Leu Asn Lys Leu Lys Trp Asn Phe Ile Glu
565 570 575
Thr Gly Asp Asn Thr Lys Arg Ile Tyr Ile Glu Asn Val Met Lys Leu
580 585 590
Arg Ser Gln Met Lys Ala Tyr Ala Ile Val Lys Asn Ala Tyr Tyr Lys
595 600 605
Gln Gln Ser Glu Tyr Asp Phe Gly Lys Ser Glu Glu Phe Ile Gln Glu
610 615 620
His Pro Phe Ser Asn Thr Asp Lys Gly Ile Glu Ile Leu Asn Lys Leu
625 630 635 640
Asp Asn Ile Ser Lys Lys Ile Leu Gly Cys Arg Asn Asn Ile Ile Gln
645 650 655
Tyr Ser Tyr Asn Leu Phe Glu Ile Asn Gly Tyr Asp Met Ile Ser Leu
660 665 670
Glu Lys Leu Thr Ser Ser Gln Phe Glu Lys Lys Pro Phe Pro Thr Val
675 680 685
Asn Ser Leu Leu Lys Tyr His Lys Ile Leu Gly Cys Thr Gln Glu Glu
690 695 700
Met Glu Lys Lys Asp Ile Tyr Ser Val Ile Lys Lys Gly Tyr Tyr Asp
705 710 715 720
Ile Ile Phe Asp Asn Gly Val Val Thr Asp Ala Lys Leu Ser Thr Lys
725 730 735
Gly Glu Leu Ser Lys Phe Lys Asp Asp Phe Phe Asn Leu Met Ile Lys
740 745 750
Ser Ile His Phe Ala Asp Ile Lys Asp Tyr Phe Ile Thr Leu Ser Asn
755 760 765
Asn Gly Thr Ala Gly Val Ser Leu Val Pro Ser Tyr Phe Thr Ser Gln
770 775 780
Met Asp Ser Ile Asp His Lys Ile Tyr Phe Val Gln Asp Asn Lys Ser
785 790 795 800
Gly Lys Leu Lys Leu Ala Asn Lys His Lys Val Arg Ser Ser Gln Glu
805 810 815
Lys His Ile Asn Gly Leu Asn Ala Asp Tyr Asn Ala Ala Arg Asn Ile
820 825 830
Ala Tyr Ile Met Glu Asn Thr Glu Cys Arg Asn Met Phe Leu Lys Gln
835 840 845
Ser Arg Thr Asp Lys Ser Leu Tyr Asn Lys Pro Ser Tyr Glu Thr Phe
850 855 860
Ile Lys Thr Gln Gly Ser Ala Val Ala Lys Leu Lys Lys Glu Gly Phe

865 870 875 880
Val Lys Ile Leu Asp Glu Ala Ser Val
885

<210> 162
<211> 37
<212> RNA
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<220>
<221> misc_feature
<222> (14)..(14)
<223> n为a、c、g或u
<400> 162
wauuguugua rmwnyywuuu uruawggwku raacaac 37
<210> 163
<211> 37
<212> RNA
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 163
aauuguugua acucuuauuu uguauggagu aaacaac 37
<210> 164
<211> 36
<212> RNA
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 164
auuguuguag auaccuuuuu guaaggauug aacaac 36
<210> 165
<211> 30
<212> DNA
<213> 人工序列(Artificial sequence)
<220>
<223> 合成序列(synthetic sequence)
<400> 165
tagcatcacc ttcaccctct ccacggacag 30

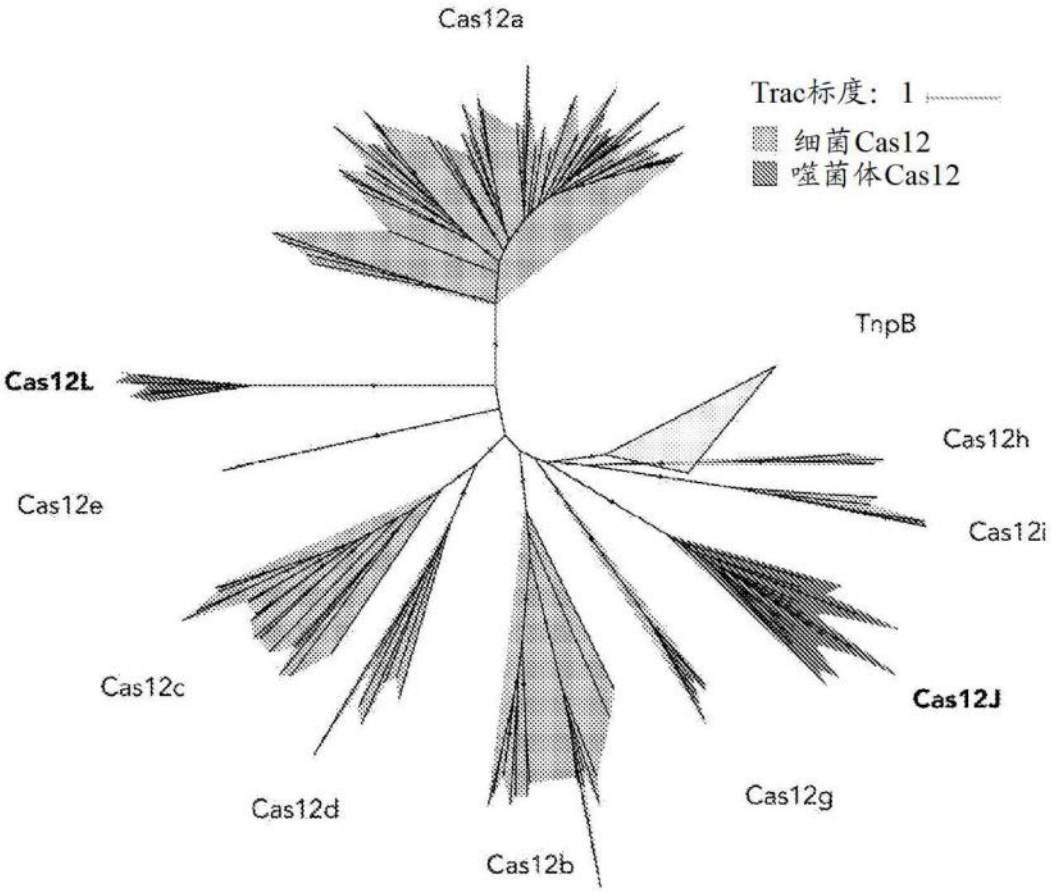


图1

>Cas12L_1_257905508 (SEQ ID NO:102)

MASHKKTESNQIIKTFSFKIKNANGLSLDVLNDAITEYQNYYNICSDWIKDHLTMKISELYKYIPNEKKN
SGYALTLISDEWKDKPMYMMFKKGYPANNRDNAIYETLNTCNTEHYTGNILNFSDTYYRRFGYVASAISN
YVTKISKMSTGSRSKNISNDSDVDTIMEQVIYEMEHNGWTSVKDWENQMEYLESKTDSNPNFVYRMTTLY
EFYKSHIDEVNSKMETMSIDSLIKFGGCRRKDSKKSMYIMGGSNTPFDITQIGGNSLNIKFSKNLNVDVF
GRYDVIKDNTLLVDIINGHGASFVLKIINDEIYIDINVSVPFDKKIATTNKVVGIDVNIKHMLLATNILD
DGNVKGYVNIYKEVINDSDFKKVCNSTVMQYFTDFSKFVTFCPLEFDFLFSRVCNQKGIYNDNSAMEKSF
SDVLNKLKWNFIETGDNTKRIYIENVMKLRSQMKAYAIVKNAYYKQQSEYDFGKSEEFIQEHPFSNTDKG
IEILNKLDNISKKILGCRNNIIQYSYNLFEINGYDMVSLEKLTSSQFKKKPFPTVNSLLKYHKILGCTQE
EMEKKDIYSVIKKGYYDIIFDNDVVTDAKLSAKGELSKFKDDFFNLMIKSIHFADIKDYFITLSNNGTAG
VSLVPSYFTSQMDSIDHKIYFVQDNKSGKLKLANKHKVRSSQEKHINGLNADYNAARNIAYIMENTDCRN
MFMKQSRTDKSLYNKPSYETFIKTQGSAVAKLKKEGFVKILDEASV*

重复: ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC (SEQ ID NO:1)

间隔子长度: 30

图2A

>Cas12L_2_196848753 (SEQ ID NO:103)

MAHKKNVGAEIVKTYSFKVKNTNGITMEKLMNAIDEFQSYYNLCSDWICKNLTTMTIGDLDQYIP
EKAKGNTYATVLLDEAWKNQPLYKIFGKKYSSNNRNNALYCALSSVIDMTKENVLGFSKTHYIR
NDYILNVISNYASKLSKLNTGVKSRAIKETSDEATIIEQVIYEMEHNKWESIEDWKNQIEYLNSKT
DYNPTYMERMKTLSAYYSTHKSEVDAKMQEMAVENLVKFGGCRRNNSKKSMFIMGSNTTNYT
ISYIGGNSFNINFANILNFDVYGRRDVVKNGEVLVDIMANHGDSIVLKIVNGELYADVPCSVTLNK
VESNFDKVVGIDVNMKHMLLSTSITDNGSSDFLNIYKEMSNNAEFMALCPEEDRKYYKDISKYV
TFAPLELDLLFSRISKQGKVKMEKVYSEILEALKWKFFANGDNKNRIYVESIQKIRQQIKALCVIK
NAYYEQQSAYDIDKTQEYIETHPFSLTEKGMSIKSKMDKICQTIIGCRNNIIDYAYSFFERNGYSII
GLEKLTSSQFEKTKSMPTCKSLLNFHKVLGHTLSELETLPINDVVKKGYYTFTTDNEGKITDASL
SEKGKVRKMKDDFFNQAIKAIHFADVKDYFATLSNNGQTGIFFVPSQFTSQMDSNTHNLYFENA
KNGGLKLAPKYKVRQTQEYHLNGLPADYNAARNIAYIGLDETMRNTFLKKANSNKSLYNQPIYD
TGIKKTAGVFSRMKKLKRYEII*

重复：ATTGTTGTAGACCTCTTTTTATAAGGATTGAACAAC (SEQ ID NO:3)

间隔子长度：30

图2B

>Cas12L_3_66741167 (SEQ ID NO:104)

MRISPHLFYIFFKKIWKCHFFVLSLYQLNQYIMASHEKTESNQIIKTFSFKIKNANGLSLDVLNDAI
TEYQNYYNICSDWIKDHLTMKIGELYKYIPDEKKNSGYALTLISDEWKDKPMYMMFKKGYPANN
RDNAIYETLNTCNTEHYTGNILNFSDTYYRRFGYVASAISNYVTKISKMSTGSRSKNISNDSDVD
TIMEQVIYEMEHNGWTSVKDWENQMEYLESKTDSNPNFVYRMTTLYEFYKSHIDEVNSKMET
MSIDSLIKFGGCRRKDSKKSMYIMGGSNTPFDITQIGGNSLNIKFSKNLNVDVFGRYDVIKDNTL
LVDIINGHGASFVLKIINGEIYIDINVSVPFDKKIATTNKVVGIDVNIKHMLLATNILDDGNVNGYVNI
YKEVINDSDFKKVCNSTVMKYFTDFSKFVTFCPLEFDFLFSRVCNQKGIYNDNSAMEKSFSDVL
NKLKWNFIETGDNTKRIYIENVMKLRSQMKAYAIVKNAYYKQQSEYDFGKSEEFIQEHPFSNTD
KGIEILNKLDNISKKILGCRNNIIQYSYNLFEINGYDMISLEKLTSSQFKKKPFPTVNSLLKYHKILG
CTQEEMEKKDIYSVIKKGYYDIIFDNDVVTDAKLSTKGELSKFKDDFFNLMIKSIHFADIKDYFITL
SNNGTAGVSLVPSYFTSQMDSIDHKIYFVQDNKSGKLKLANKHKVRSSQEKHINGLNADYNAA
RNIAYIMENTECRNMFMKQSRTDKSLYNKPSYETFIKTQGSAVAKLKKEGFVKILDEASV*

重复：ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC (SEQ ID NO:1)

间隔子长度：30

图2C

>Cas12L_4_67031163 (SEQ ID NO:105)

MRISPHLFYIFFKKIWKSHFFVLSLYQLNQYIMASHKKTESNQIIKTFSFKIKNANGLSLDVLNDAI
TEYQNYYNICSDWIKDHLTMKIGELYKYIPDEKKNSGYALTLISDEWKDKPMYMMFKKGYPANN
RDNAIYETLNTCNTEHYTGNILNFPDTYYRRFGYVASTISNYVTKISKMSTGSRSKNISNDSDVD
TIMEQVIYEMEHNGWTSVKDWENQMEYLESKTDSNPNFVYRMTTLYEFYKSHIDEVNSKMET
MSIDSLIKFGGCRRKDSKKSMYIMGGSNTPFDITQIGGNSLNIKFSKNLNVDVFGRYDVIKDNTL
LVDIINGHGASFVLKIINGEIYIDINVSVPFDKKIATTNKVVGIDVNIKHMLLATNILDDGNVNGYVNI
YKEVINDSDFKKVCNSTVMKYFTDFSKFVTFCPLEFDFLFSRVCNQKGIYNDNSAMEKSFSDVL
NKLKWNFIETGDNTKRIYIENVMKLRSQMKAYAIVKNAYYKQQSEYDFGKSEEFIQEHPFSNTD
KGIEILNKLDNISKKILGCRNNIIQYSYNLFEINGYDMISLEKLTSSQFKKNPFPTVNSLLKYHKILG
CTQEEMEKKDIYSVIKKGYYDIIFDNDVVTDAKLSTKGELSKFKDDFFNLMIKSIHFADIKDYFITL
SNNGTAGVSLVPSYFTSQMDSIDHKIYFVQDNKSGKLKLANKHKVRSSQEKHINGLNADYNAA
RNIAYIMENTECRNMFMKQSRTDKSLYNKPSYETFIKTQGSAVAKLKKEGFVKILDEASV*

重复: ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC (SEQ ID NO:1)

间隔子长度: 30

图2D

>Cas12L_5_67793351 (SEQ ID NO:106)

MRISPHLFYIFFKKIWKCHIFVLSLYQLNQYIMASHKKTESNQIIKTFSFKIKNANGLSLDVLNDAIT
EYQNYYNICSDWIKDHLTMKIGELYKYIPDEKKNSGYALTLISDEWKDKPMYMMFKKGYPANNR
DNAIYETLNTCNTEHYTGNILNFSDTYYRRFGYVASTISNYVTKISKMSTGSRSKNISNDSDVDTI
MEQVIYEMEHNGWTSVKDWENQMEYLESKTDSNPNFVYRMTTLYEFYKSHIDEVNSKMETMS
IDSLIKFGGCRRKDSKKSMYIMGGSNTPFDITQIGGNSLNIKFSKNLNVDVFGRYDVIKDNTLLVD
IINGHGASFVLKIINGEIYIDINVSVPFDKKIATTNKVVGVDVNIKHMLLATNILDDGNVNGYVNIYK
EVINDSDFKKVCNSTVMKYFTDFSKFVTFCPLEFDFLFSRVCNQKGIYNDNSAMEKSFSDVLNK
LKWNFIETGDNTKRIYIENVMKLRSQMKAYAIVKNAYYKQQSEYDFGKSEEFIQEHPFSNTDKGI
EILNKLDNISKKILGCRNNIIQYSYNLFEINGYDMISLEKLTSSQFKKKPFPTVNSLLKYHKILGCTQ
EEMEKKDIYSVIKKGYYDIIFDNDVVTDAKLSTKGELSKFKDDFFNLMIKSIHFADIKDYFITLSNN
GTAGVSLVPSYFTSQMDSIDHKIYFVQDNKSGKLKLANKHKVRSSQEKHINGLNADYNAARNIA
YIMENTDCRNMFMKQSRTDKSLYNKPSYETFIKTQGSAVAKLKKEGFVKILDEASV*

重复: ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC (SEQ ID NO:1)

间隔子长度: 30

图2E

>Cas12L_6_67912869_ 部分(SEQ ID NO:107)

QYVTFAPLELDLLFSRISKQGKVKMEKAYSEILEALKWKFFANGDNKNRIYVESIQKIRQQIKALC
VIKNAYYEQQSAYDIDKTQEYIETHPFSLTEKGMSIKSKMDKICQTIIGCRNNIIDYAYSFFERNGY
TIIGLEKLTSSQFEKTKSMPTCKSLLNFHKVLGHTLSELETLPINDVVKKGYYAFTTDNEGRITDA
SLSEKGKVRKMKDDFFNQAIKAIHFADVKDYFATLSNNGQTGIFFVPSQFTSQMDSNTHNLYFE
NAKNGGLKLASKSKVRKSQEYHLNGLPADYNAARNIAYIGLDEIMRNTFLKKANSNKSLYNQPI
YDTGIKKTAGVFSRMKKLKKYKVI*

重复:ATTGTTGTAGATACCTTTTTGTAAGGATTGAACAAC (SEQ ID NO:5)

间隔子长度: 30

图2F

>Cas12L_7_68090316_ 部分 (SEQ ID NO:108)

VLNDAITEYQNYYNICSDWIKDHLTMKIGELYKYIPDEKKNSGYALTLISDEWKDKPMYMMFKK
GYPANSRDNAIYEALNTCNTEHYTGNILNFSDTYYRRFGYVASAISNYVTKISKMSTGSRSKNIS
NDSDVDTIMEQVIYEMEHNGWTSVKDWENQMEYLESKTDSNPNFVYRMTTLYEFYKSHIDEV
NSKMETMSIDSLIKFGGCRRKDSKKSMYIMGGSNTPFDITQIGGNSLNIKFSKNLNVDVFGRYD
VIKDNTLLVDIINGHGASFVLKIINDEIYIDINVSVPFDKKIATTNKVVGIDVNIKHMLLATNILDDGN
VNGYVNIYKEVINDSDFKKVCNSTVMKYFTDFSKFVTFCPLEFDFLFSRVCNQKGIYNDNSAME
KSFSNVLNKLKWNFIETGDNTKRIYIENVMKLRSQMKAYAIVKNAYYKQQSEYDFGKSEEFIQE
HPFSNTDKGIEILHKLDNISKKILGCRNNIIQYSYNLFEINGYDMISLEKLTSSQFKKKPFPTVNSLL
KYHKILGCTQEEMEKKDIYSVIKKGYYDIIFDNDVVTDAKLSTKGELSKFKDDFFNLMIKSIHFADI
KDYFITLSNNGTAGVSLVPSYFTSQMDSIDHKIYFVQDNKSGKLKLANKHKVRSSQEKHINGLN
ADYNAARNIAYIMENTDCRNMFMKQSRTDKSLYNKPSYETFIKTQGSAVAKLKKEGFMKILDEA
SV*

重复: ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC (SEQ ID NO:1)

间隔子长度: 30

图2G

>Cas12L_8_68328292_ 部分 (SEQ ID NO:109)

IFFKKIWKCHIFVLSLYQLNQYIMASHKKTESNQIIKTFSFKIKNANGLSLDVLNDAITEYQNYYNIC
SDWIKDHLTMKIGELYKYIPDEKKNSGYALTLISDEWKDKPMYMMFKKGYPANSRDNAIYEALN
TCNTEHYTGNILNFSDTYYRRFGYVASAISNYVTKISKMSTGSRSKNISNDSDVDTIMEQVIYEM
EHNGWTSVKDWENQMEYLESKTDSNPNFVYRMTTLYEFYKSHIDEVNSKMETMSIDSLIKFGG
CRRKDSKKSMYIMGGSNTPFDITQIDGNSLNIKFSKNLNVDVFGRYDVIKDNTLLVDIINGHGAS
FVLKIINDEIYIDINVSVPFDKKIATTNKVVGVDVNIKHMLLATNILDDGNVNGYVNIYKEVINDSDF
KKVCNSTVMKYFTDFSKFVTFCPLEFDFLFSRVCNQKGIYNDNSAMEKSFSDVLNKLKWNFIET
GDNTKRIYIENVMKLRSQMKAYAIVKNAYYKQQSEYDFGKSEEFIQEHPFSNTDKGIEILHKLDNI
SKKILGCRNNIIQYSYNLFEINGYDMISLEKLTSSQFKKKPFPTVNSLLKYHKILGCTQEEMEKKDI
YSVIKKGYYDIIFDNDVVTDAKLSTKGELSKFKDDFFNLMIKSIHFADIKDYFITLSNNGTAGVSLV
PSYFTSQMDSIDHKIYFVQDNKSGKLKLANKHKVRSSQEKHINGLNADYNAARNIAYIMENTEQ
GSAVAKLKKEGFVKILDEASV*

重复: ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC (SEQ ID NO:1)
间隔子长度: 30

图2H

>Cas12L_9_68454124 (SEQ ID NO:110)

MASHKKTESNQIIKTFSFKIKNANGLSLDVLNDAITEYQNYYNICSDWIKDHLTMKIGELYKYIPDE
KKNSGYALTLISDEWKDKPMYMMFKKGYPANNRDNAIYETLNTCNTEHYTGNILNFSDTYYRR
FGYVASAISNYVTKISKMSTGSRSKNISNDSDVDTIMEQVIYEMEHNGWTSVKDWENQMEYLE
SKTDSNPNFVYRMTTLYEFYKSHIDEVNSKMETMSIDSLIKFGGCRRKDSKKSMYIMGGSNTPF
DITQIGGNSLNIKFSKNLNVDVFGRYDVIKDNTLLVDIINGHGASFVLKIINGEIYIDINVSVPFDKKI
ATTNKVVGVDVNIKHMLLATNILDDGNVNGYVNIYKEVINDSDFKKVCNSTVMKYFTDFSKFVTF
CPLEFDFLFSRVCNQKGIYNDNSAMEKSFSDVLNKLKWNFIETGDNTKRIYIENVMKLRSQMKA
YAIVKNAYYKQQSEYDFGKSEEFIQEHPFSNTDKGIEILNKLDNISKKILGCRNNIIQYSYNLFEIN
GYDMISLEKLTSSQFKKKPFPTVNSLLKYHKILGCTQEEMEKKDIYSVIKKGYYDIIFDNDVVTDA
KLSAKGELSKFKDDFFNLMIKSIHFADIKDYFITLSNNGTAGVSLVPSYFTSQMDSIDHKIYFVQD
NKSGKLKLANKHKVRSSQEKHINGLNADYNAARNIAYIMENTECRNMFMKQSRTDKSLYNKPS
YETFIKTQGSAVSKLKKDGFVKILDEASV*

重复: ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC (SEQ ID NO:1)
间隔子长度: 30

图2I

>Cas12L_10_68605313 (SEQ ID NO:111)

MMKKMRTNPHLFYICFKKIWKCHFFALSLYQLNQYIMASHKKTESNQIIKTFSFKIKNANGLSLD
VLNDAITEYQNYYNICSDWIKDHLTMKIGELYKYIPDEKKNSGYALTLISDEWKDKPMYMMFKK
GYPANSRDNAIYEALNTCNTEHYTGNILNFSDTYYRRFGYVASTISNYVTKISKMSTGSRSKNIS
NDSDVDTIMEQVIYEMEHNGWTSVKDWENQMEYLESKTDSNPNFVYRMTTLYEFYKSHIDEV
NSKMETMSIDSLIKFGGCRRKDSKKSMYIMGGSNTPFDITQIGGNSLNIKFSKNLNVDVFGRYD
VIKDNTLLVDIINEHGASFVLKIINDEIYIDINVSVPFDKKIATTNKVVGVDVNIKHMLLATNILDDGN
VNGYVNIYKEVINDSDFKKVCNSTVMKYFTDFSKFVTFCPLEFDFLFSRVCNQKGIYNDNSAME
KSFSDVLNKLKWNFIETGDNTKRIYIENVMKLRSQMKAYAIVKNAYYKQQSEYDFGKSEEFIQE
HPFSNTDKGIEILNKLDNISKKILGCRNNIIQYSYNLFEINGYDMISLEKLTSSQFKKKPFPTVNSLL
KYHKILGCTQEEMEKKDIYSVIKKGYYDIIFDNDVVTDAKLSTKGELSKFKDDFFNLMIKSIHFADI
KDYFITLSNNGTAGVSLVPSYFTSQMDSIDHKIYFVQDNKSGKLKLANKHKVRSSQEKHINGLN
ADYNAARNIAYIMENTECRNMFMKQSRTDKSLYNKPSYETFIKTQGSAVAKLKKEGFVKILDEA
SA*

重复: ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC (SEQ ID NO:1)
间隔子长度: 30

图2J

>Cas12L_11_69266821_ 部分 (SEQ ID NO:112)

NKSGKLKLANKHKVRSSQEKHINGLNADYNAARNIAYIMENTECRNMFMKQSRTDKSLYNKPS
YETFIKTQGSAVAKLKKEGFVKILDEASV*

重复: ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC (SEQ ID NO:1)
间隔子长度: 30

图2K

>Cas12L_12_69417229_ 部分 (SEQ ID NO:112)

NKSGKLKLANKHKVRSSQEKHINGLNADYNAARNIAYIMENTECRNMFMKQSRTDKSLYNKPS
YETFIKTQGSAVAKLKKEGFVKILDEASV*

重复: ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC (SEQ ID NO:1)
间隔子长度: 30

图2L

>Cas12L_13_69733214 (SEQ ID NO:113)

MASHKKTESNQIIKTFSFKIKNANGLSLDVLNDAITEYQNYYNICSDWIKDHLTMKIGELYKYIPDE
KKNSGYALTLISDEWKDKPMYMMFKKGYPANSRDNAIYEALNTCNTEHYTGNILNFSDTYYRR
FGYVASTISNYVTKISKMSTGSRSKNISNDSDVDTIMEQVIYEMEHNGWTSVKDWENQMEYLE
SKTDSNPNFVYRMTTLYEFYKSHIDEVNSKMETMSIDSLIKFGGCRRKDSKKSMYIMGGSNTPF
DITQIGGNSLNIKFSKNLNVDVFGRYDVIKDNTLLVDIINGHGASFVLKIINGEIYIDINVSVPFDKKI
ATTNKVVGVDVNIKHMLLATNILDDGNVNGYVNIYKEVINDSDFKKVCNSTVMKYFTDFSKFVTF
CPLEFDFLFSRVCNQKGIYNDNSAMEKSFSDVLNKLKWNFIETGDNTKRIYIENVMKLRSQMKA
YAIVKNAYYKQQSEYDFGKSEEFIQEHPFSNTDKGIEILNKLDNISKKILGCRNNIIQYSYNLFEIN
GYDMISLEKLTSSQFEKKPFPTVNSLLKYHKILGCTQEEMEKKDIYSVIKKGYYDIIFDNDNVVTDA
KLSTKGELSKFKDDFFNLMIKSIHFADIKDYFITLSNNGTAGVSLVPSYFTSQMDSIDHKIYFVQD
NKSGKLKLANKHKVRSSQEKHINGLNADYNAARNIAYIMENTDCRNMFMKQSRTDKSLYNKPS
YETFIKTQGSAVAKLKKEGFVKILDEASV*

重复: ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC (SEQ ID NO:1)

间隔子长度: 30

图2M

>Cas12L_14_70235246_ 部分 (SEQ ID NO:114)

VPFDKKIATTNKVVGVDVNIKHMLLATNILDDGNVNGYVNIYKEVINDSDFKKVCNSTVMKYFTD
FSKFVTFCPLEFDFLFSRVCNQKGIYNDNSAMEKSFSDVLNKLKWNFIETGDNTKRIYIENVMKL
RSQMKAYAIVKNAYYKQQSEYDFGKSEEFIQEHPFSNTDKGIEILNKLDNISKKILGCRNNIIQYS
YNLFEINGYDMISLEKLTSSQFKKKSFPTVDSLLKYHKILGCTQEEMEKKDIYSVIKKGYYDIIFDN
DVVTDAKLSTKGELSKFKDDFFNLMIKSIHFADIKDYFITLSNNGTAGVSLVPSYFTSQMDSIDHK
IYFVQDNKSGKLKLANKHKVRSSQEKHINGLNADYNAARNIAYIMENTDCRNMFMKQSRTDKS
LYNKPSYETFIKTQGSAVAKLKKEGFVKILDEASV*

重复: ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC (SEQ ID NO:1)

间隔子长度: 30

图2N

>Cas12L_15_70724743 (SEQ ID NO:115)

MASHKKTESNQIIKTFSFKIKNANGLSLDVLNDAITEYQNYYNICSDWIKDHLTMKIGELYKYIPDE
KKNSGYALTLISDEWKDKPMYMMFKKGYPANNRDNAIYETLNTCNTEHYTGNILNFSDTYYRR
FGYVASAISNYVTKISKMSTGSRSKNISNDSDVDTIMEQVIYEMEHNGWTSVKDWENQMEYLE
SKTDSNPNFVYRMTTLYEFYKSHIDEVNSKMETMSIDSLIKFGGCRRKDSKKSMYIMGGSNTPF
DITQIGGNSLNIKFSKNLNVDVFGRYDVIKDNTLLVDIINGHGASFVLKIINGEIYIDINVSVPFDKKI
ATTNKVVGVDVNIKHMLLATNILDDGNVNGYVNIYKEVINDSDFKKVCNSTVMKYFTDFSKFVTF
CPLEFDFLFSRVCNQKGIYNDNSAMEKSFSDVLNKLKWNFIETGDNTKRIYIENVMKLRSQMKA
YAIVKNAYYKQQSEYDFGKSEEFIQEHPFSNTDKGIEILNKLDNISKKILGCRNNIIQYSYNLFEIN
GYDMISLEKLTSSQFKKKPFPTVNSLLKYHKILGCTQEEMEKKDIYSVIKKGYYDIIFDNDVVTDA
KLSTKGELSKFKDDFFNLMIKSIHFADIKDYFITLSNNGTAGVSLVPSYFTSQMDSIDHKIYFVQD
NKSGKLKLANKHKVRSSQEKHINGLNADYNAARNIAYIMENTECRNMFMKQSRTDKSLYNKPS
YETFIKTQGSAVAKLKKEGFVKILDEASV*

重复:ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC (SEQ ID NO:1)

间隔子长度: 30

图20

>Cas12L_16_70731038 (SEQ ID NO:116)

MAHKKNIGAEIVKTYSFKVKNTNGITMEKLMNAIDEYQSYYNLCSDWICKNLTTMTIGDLDRYIP
EKAKDNIYATVLLDEVWKNQPLYKIFGKKYSSNNRINALYCTLSSVIDINKKNILGLSQTYYARNG
YILNVISNYASKLSKLNTGVKRHTIKETSDEATIVEQVIYEMEHNKWESIEDWKNQIEYLNSKTDY
NPTYMERMKTLSAYYSEHKSEVDAKMQEMAVENLVKFGGCRRNNSKKSMFIMGSSKTTYTIS
YIGDNCFNINFANILNFDVYGRRDVVKNGEVLVDIMANHGDSIVLKIVNGELYADVPCSTTLNKV
ESTFDKVAGIDVNMKHMLLSTSVTDNGNSDFVNIYKEMSNNAEFMALCPEEDRKYYKDISQYV
TFAPLELDLLFSRISKQGKVKMEKAYSEILEALKWKFFANGDNKNRIYVENIQKIRQQIKALCVIK
NAYYEQQSAYDIDKTQEYIEAHPFSLTEKGMSIKSKMDNICRTIIGCRNNIIDYAYSFFERNDYSII
GLEKLTSSQFEKTKSLPTCKSLLNFHKVLGHTLSELETLPINDVVKKGYYTFTTDNEGRITDASL
SEKGKVRKMKDDFFNQAIKAIHFADVKDYFATLSNNGQTGIFFVPSQFTSQMDSNTHTLYFENA
KNGGLKLASKYKVRKSQEYHLNGLPADYNAARNIAYIGLDEIMRNTFLKKANSNKSLYNQPIYDT
GIKKTAGVFSRMKKLKKYKVI*

重复:TATTGTTGTAGATACCTTTTTGTAAGGATTAAACAAC (SEQ ID NO:7)

间隔子长度: 30

图2P

>Cas12L_17_70959391 (SEQ ID NO:117)

MASHKKTESNQIIKTFSFKIKNANGLSLDVLNDAITEYQNYYNICSDWIKDHLTMKIGELYKYIPDE
KKNSGYALTLISDEWKDKPMYMMFKKGYPANSRDNAIYEALNTCNTEHYTGNILNFSDTYYRR
FGYVASTISNYVTKISKMSTGSRSKNISNDSDVDTIMEQVIYEMEHNGWTSVKDWENQMEYLE
SKTDSNPNFVYRMTTLYEFYKSHIDEVNSKMETMSIDSLIKFGGCRRKDSKKSMYIMGGSNTPF
DITQIGGNSLNIKFSKNLNVDVFGRYDVIKDNTLLVDIINGHGASFVLKIINGEIYIDINVSVPFDKKI
ATTNKVVGVDVNIKHMLLATNILDDGNVNGYVNIYKEVINDSDFKKVCNSTVMQYFTDFSKFVTF
CPLEFDFLFSRVCNQKGIYNDNSAMEKSFSDVLNKLKWNFIETGDNTKRIYIENVMKLRSQMKA
YAIVKNAYYKQQSEYDFGKSEEFIQEHPFSNTDKGIEILNKLDNISKKILGCRNNIIQYSYNLFEIN
GYDMISLEKLTSSQFKKKPFPTVNSLLKYHKILGCTQEEMEKKDIYSVIKKGYYDIIFDNDVVTDA
KLSTKGELSKFKDDFFNLMIKSIHFADIKDYFITLSNNGTAGVSLVPSYFTSQMDSIDHKIYFVQD
NKSGKLKLANKHKVRSSQEKHINGLNADYNAARNIAYIMENTECRNMFMKQSRTDKSLYNKPS
YETFIKTQGSAVAKLKKEGFVKIIDEASV*

图2Q

>Cas12L_18_71078086 (SEQ ID NO:118)

MASHKKTESNQIIKTFSFKIKNANGLSLDVLNDAITEYQNYYNICSDWIKDHLTMKIGELYKYIPDE
KKNSGYALTLISDEWKDKPMYMMFKKGYPANNRDNAIYEALNTCNTEHYTGNILNFSDTYYRR
FGYVASTISNYVTKISKMSTGSRSKNISNDSDVDTIMEQVIYEMEHNGWTSVKDWENQMEYLE
SKTDSNPNFVYRMTTLYEFYKSHIDEVNSKMETMSIDSLIKFGGCRRKDSKKSMYIMGGSNTPF
DITQIGGNSLNIKFSKNLNVDVFGRYDVIKDNTLLVDIINGHGASFVLKIINGEIYIDINVSVPFDKKI
ATTNKVVGVDVNIKHMLLATNILDDGNVNGYVNIYKEVINDSDFKKVCNSTVMKYFTDFSKFVTF
CPLEFDFLFSRVCNQKGIYNDNSAMEKSFSDVLNKLKWNFIETGDNTKRIYIENVMKLRSQMKA
YAIVKNAYYKQQSEYDFGKSEEFIQEHPFSNTDKGIEILNKLDNISKKILGCRNNIIQYSYNLFEIN
GYDMISLEKLTSSQFKKKPFPTVNSLLKYHKILGCTQEEMEKKDIYSVIKKGYYDIIFDNGVVTDA
KLSTKGELSKFKDDFFNLMIKSIHFADIKDYFITLSNNGTAGVSLVPSYFTSQMDSIDHKIYFVQD
NKSGKLKLANKHKVRSSQEKHINGLNADYNAARNIAYIMENTECRNMFMKQSRTDKSLYNKPS
YETFIKTQGSAVAKLKKEGFVKILDEASV*

重复:ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC (SEQ ID NO:1)

间隔子长度: 30

图2R

>Cas12L_19_71193509_ 部分 (SEQ ID NO:119)

EKKNSGYALTLISDEWKDKPMYMMFKKGYPANNRDNAIYETLNTCNTEHYTGNILNFSDTYYR
RFGYVASAISNYVTKISKMSTGSRSKNISNDSDVDTIMEQVIYEMEHNGWTSVKDWENQMEYL
ESKTDSNPNFVYRMTTLYEFYKSHIDEVNSKMETMSIDSLIKFGGCRRKDSKKSMYIMGGSNTP
FDITQIGGNSLNIKFSKNLNVDVFGRYDVIKDNTLLVDIINEHGASFVLKIINDEIYIDINVSVPFDKK
IATTNKVVGVDVNIKHMLLATNILDDGNVNGYVNIYKEVINDSDFKKVCNSTVMKYFTDFSKFVT
FCPLEFDFLFSRVCNQKGIYNDNSAMEKSFSDVLNKLKWNFIETGDNTKRIYIENVMKLRSQMK
AYAIVKNAYYKQQSEYDFGKSEEFIQEHPFSNTDKGIEILNKLDNISKKILGCRNNIIQYSYNLFEI
NGYDMISLEKLTSSQFKKKPFPTVNSLLKYHKILGCTQEEMEKKDIYSVIKKGYYDIIFDNDVVTD
AKLSTKGELSKFKDDFFNLMIKSIHFADIKDYFITLSNNGTVGVSLVPSYFTSQMDSIDHKIYFVQ
DNKSGKLKLANKHKVRSSQEKHINGLNADYNAARNIAYIMENNECRNMFMKQSRTDKSLYNKP
SYETFIKTQGSAVAKLKKEGFVKILDEASV*

重复: ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC (SEQ ID NO:1)

间隔子长度: 30

图2S

>Cas12L_20_71210958_ 部分 (SEQ ID NO:120)

DAKMQEMAVENLVKFGGCRRNNSKKSMFIMGSSKTTYTISYIGDNCFNINFANILNFDVYGRRD
VVKNGEVLVDIMANHGDSIVLKIVNGELYADVPCSTTLNKVESTFDKVAGIDVNMKHMLLSTSVT
DNGNSDFVNIYKEISNNAEFMALCPEEDRKYYKDISQYVTFAPLELDLLFSRISKQGKVKMEKAY
SEILETLKWKFFANGDNKNRIYVESIQKIRQQIKALCVIKNAYYEQQSAYDIDKTQEYIEAHPFSLT
EKGMSIKSKMDNICRTIIGCRNNIIDYAYSFFERNDYSIIGLEKLTSSQFEKTKSLPTCKSLLNFHK
VLGHTLSELETLPINDVVKKGYYTFTTDNEGRITDASLSEKGKVRKMKDDFFNQAIKAIHFADVK
DYFATLSNNGQTGIFFVPSQFTSQMDSNTHTLYFENAKNGGLKLASKYQVRQTQEYHLNGLPA
DYNAARNIAYIGLDETMRNTFLKKANSNKSLYNQPIYDTGIKKTAGVFSRMKKLKRYEII*

重复: AATGTTGTAGATGCCTTTTTATAAGGATTAAACAACTTG (SEQ ID NO:9)

间隔子长度: 27

图2T

>Cas12L_21_71317321_ 部分 (SEQ ID NO:121)

DVDTIMEQVIYEMEHNGWTSVKDWENQMEYLESKTDSNPNFVYRMTTLYEFYKSHIDEVNSK
METMSIDSLIKFGGCRRKDSKKSMYIMGGSNTPFDITQIGGNSLNIKFSKNLNVDVFGRYDVIKD
NTLLVDIINGHGASFVLKIINGEIYIDINVSVPFDKKIATTNKVVGVDVNIKHMLLATNILDDGNVNG
YVNIYKEVINDSDFKKVCNSTVMKYFTDFSKFVTFCPLEFDFLFSRVCNQKGIYNDNSAMEKSF
SDVLNKLKWNFIETGDNTKRIYIENVMKLRSQMKAYAIVKNAYYKQQSEYDFGKSEEFIQEHPF
SNTDKGIEILNKLDNISKKILGCRNNIIQYSYNLFEINGYDMISLEKLTSSQFKKKPFPTVNSLLKYH
KILGCTQEEMEKKDIYSVIKKGYYDIIFDNDVVTDAKLSTKGELSKFKDDFFNLMIKSIHFADIKDY
FITLSNNGTAGVSLVPSYFTSQMDSIDHKIYFVQDNKSGKLKLANKHKVRSSQEKHINGLNADY
NAARNIAYIMENTECRNMFMKQSRTDKSLYNKPSYETFIKTQGSAVAKLKKEGFVKIIDEASV*

重复: ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC (SEQ ID NO:1)

间隔子长度: 30

图2U

>Cas12L_22_71456687 (SEQ ID NO:122)

MASHKKTESNQIIKTFSFKIKNANGLSLDVLNDAITEYQNYYNICSDWIKDHLTMKIGELYKYIPDE
KKNSGYALTLISDEWKDKPMYMMFKKGYPANSRDNAIYEALNTCNTEHYTGNILNFSDTYYRR
FGYVASAISNYVTKISKMSTGSRSKNISNDSDVDTIMEQVIYEMEHNGWTSVKDWENQMEYLE
SKTDSNPNFVYRMTTLYEFYKSHIDEVNSKMETMSIDSLIKFGGCRRKDSKKSMYIMGGSNTPF
DITQIGGNSLNIKFSKNLNVDVFGRYDVIKDNTLLVDIINEHGASFVLKIINDEIYIDINVSVPFDKKI
ATTNKVVGVDVNIKHMLLATNILDDGNVNGYVNIYKEVINDSDFKKVCNSTVMKYFTDFSKFVTF
CPLEFDFLFSRVCNQKGIYNDNSAMEKSFSDVLNKLKWNFIETGDNTKRIYIENVMKLRSQMKA
YAIVKNAYYKQQSEYDFGKSEEFIQEHPFSNTDKGIEILNKLDNISKKILGCRNNIIQYSYNLFEIN
GYDMISLEKLTSSQFEKKPFPTVNSLLKYHKILGCTQEEMEKKDIYSVIKKGYYDIIFDNDVVTDA
KLSTKGELSKFKDDFFNLMIKSIHFADIKDYFITLSNNGTAGVSLVPSYFTSQMDSIDHKIYFVQD
NKSGKLKLANKHKVRSSQEKHINGLNADYNAARNIAYIMENTECRNMFMKQSRTDKSLYNKPS
YETFIKTQGSAVAKLKKEGFVKILDEASV*

重复:ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC (SEQ ID NO:1)

间隔子长度: 30

图2V

>Cas12L_23_71708971 (SEQ ID NO:123)

DVDTIMEQVIYEMEHNGWTSVKDWENQMEYLESKTDSNPNFVYRMTTLYEFYKSHIDEVNSK
METMSIDSLIKFGGCRRKDSKKSMYIMGGSNTPFDITQIGGNSLNIKFSKNLNVDVFGRYDVIKD
NTLLVDIINEHGASFVLKIINDEIYIDINVSVPFDKKIATTNKVVGVDVNIKHMLLATNILDDGNVNG
YVNIYKEVINDSDFKKVCNSTVMKYFTDFSKFVTFCPLEFDFLFSRVCNQKGIYNDNSAMEKSF
SDVLNKLKWNFIETGDNTKRIYIENVMKLRSQMKAYAIVKNAYYKQQSEYDFGKSEEFIQEHPF
SNTDKGIEILNKLDNISKKILGCRNNIIQYSYNLFEINGYDMISLEKLTSSQFKKKPFPTVDSLLKYH
KILGCTQEEMEKKDIYSVIKKGYYDIIFDNDVVTDAKLSTKGELSKFKDDFFNLMIKSIHFADIKDY
FITLSNNGTAGVSLVPSYFTSQMDSIDHKIYFVQDNKSGKLKLANKHKVRSSQEKHINGLNADY
NAARNIAYIMENTDCRNMFMKQSRTDKSLYNKPSYETFIKTQGSAVAKLKKEGFVKILDEASV*

重复:ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC (SEQ ID NO:1)

间隔子长度: 30

图2W

>Cas12L_24_46035167_ 部分 (SEQ ID NO:124)

FVLSLYQLNQYIMASHKKTESNQIIKTFSFKIKNANGLSLDVLNDAITEYQNYYNICSDWIKDHLT
MKIGELYKYIPDEKKNSGYALTLISDEWKDKPMYMMFKKGYPANNRDNAIYETLNTCNTEHYTG
NILNFSDTYYRRFGYVASTISNYVTKISKMSTGSRSKNISNDSDVDTIMEQVIYEMEHNGWTSVK
DWENQMEYLESKTDSNPNFVYRMTTLYEFYKSHIDEVNSKMETMSIDSLIKFGGCRRKDSKKS
MYIMGGSNTPFDITQIGGNSLNIKFSKNLNVDVFGRYDVIKDNTLLVDIINGHGASFVLKIINGEIYI
DINVSVPFDKKIATTNKVVGVDVNIKHMLLATNILDDGNVNGYVNIYKEVINDSDFKKVCNSTVM
KYFTDFSKFVTFCPLEFDFLFSRVCNQKGIYNDNSAMEKSFSDVLNKLKWNFIETGDNTKRIYIE
NVMKLRSQMKAYAIVKNAYYKQQSEYDFGKSEEFIQEHPFSNTDKGIEILNKLDNISKKILGCRN
NIIQYSYNLFEINGYDMISLEKLTSSQFKKKSFPTVDSLLKYHKILGCTQEEMEKKDIYSVIKKGYY
DIIFDNDVVTDAKLSTKGELSKFKDDFFNLMIKSIHFADIKDYFITLSNNGTAGVSLVPSYFTSQM
DSIDHKIYFVQDNKSGKLKLANKHKVRSSQEKHINGLNADYNAARNIAYIMENNECRNMFMKQS
RTDKSLYNKPSYETFIKTQGSAVAKLKKEGFVKILDEASV*

重复: ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC (SEQ ID NO:1)
间隔子长度: 30

图2X

>Cas12L_25_46784254_ 部分 (SEQ ID NO:125)

GSRSKNISNDSDVDTIMEQVIYEMEHNGWTSVKDWENQMEYLESKTDSNPNFVYRMTTLYEF
YKSHIDEVNSKMETMSIDSLIKFGGCRRKDSKKSMYIMGGSNTPFDITQIGGNSLNIKFSKNLNV
DVFGRYDVIKDNTLLVDIINGHGASFVLKIINGEIYIDINVSVPFDKKIATTNKVVGVDVNIKHMLLA
TNILDDGNVNGYVNIYKEVINDSDFKKVCNSTVMKYFTDFSKFVTFCPLEFDFLFSRVCNQKGIY
NDNSAMEKSFSDVLNKLKWNFIETGDNTKRIYIENVMKLRSQMKAYAIVKNAYYKQQSEYDFG
KSEEFIQEHPFSNTDKGIEILNKLDNISKKILGCRNNIIQYSYNLFEINGYDMISLEKLTSSQFKKKS
FPTVDSLLKYHKILGCTQEEMEKKDIYSVIKKGYYDIIFDNDVVTDAKLSTKGELSKFKDDFFNLM
IKSIHFADIKDYFITLSNNGTAGVSLVPSYFTSQMDSIDHKIYFVQDNKSGKLKLANKHKVRSSQE
KHINGLNADYNAARNIAYIMENNECRNMFMKQSRTDKSLYNKPSYETFIKTQGSAVAKLKKEGF
VKILDEASV*

重复: ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC (SEQ ID NO:1)
间隔子长度: 30

图2Y

>Cas12L_26_46464451 (SEQ ID NO:126)

MASHKKTESNQIIKTFSFKIKNANGLSLDVLNDAITEYQNYYNICSDWIKDHLTMKIGELYKYIPDE
KKNSGYALTLISDEWKDKPMYMMFKKGYPANNRDNAIYETLNTCNTEHYTGNILNFSDTYYRR
FGYVASTISNYVTKISKMSTGSRSKNISNDSDVDTIMEQVIYEMEHNGWTSVKDWENQMEYLE
SKTDSNPNFVYRMTTLYEFYKSHIDEVNSKMETMSIDSLIKFGGCRRKDSKKSMYIMGGSNTPF
DITQIDGNSLNIKFSKNLNVDVFGRYDVIKDNTLLVDIINGHGASFVLKIINGEIYIDINVSVPFDKKI
ATTNKVVGVDVNIKHMLLATNILDDGNVNGYVNIYKEVINDSDFKKVCNSTVMKYFTDFSKFVTF
CPLEFDFLFSRVCNQKGIYNDNSAMEKSFSDVLNKLKWNFIETGDNTKRIYIENVMKLRSQMKA
YAIVKNAYYKQQSEYDFGKSEEFIQEHPFSNTDKGIEILNKLDNISKKILGCRNNIIQYSYNLFEIN
GYDMISLEKLTSSQFKKKPFPTVNSLLKYHKILGCTQEEMEKKDIYSVIKKGYYDIIFDNGVVIDA
KLSAKGELSKFKDDFFNLMIKSIHFADIKDYFITLSNNGTAGVSLVDLFCPR*

重复:ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC (SEQ ID NO:1)

间隔子长度: 30

图2Z

>Cas12L_27_254489164 (SEQ ID NO:102)

MASHKKTESNQIIKTFSFKIKNANGLSLDVLNDAITEYQNYYNICSDWIKDHLTMKISELYKYIPNE
KKNSGYALTLISDEWKDKPMYMMFKKGYPANNRDNAIYETLNTCNTEHYTGNILNFSDTYYRR
FGYVASAISNYVTKISKMSTGSRSKNISNDSDVDTIMEQVIYEMEHNGWTSVKDWENQMEYLE
SKTDSNPNFVYRMTTLYEFYKSHIDEVNSKMETMSIDSLIKFGGCRRKDSKKSMYIMGGSNTPF
DITQIGGNSLNIKFSKNLNVDVFGRYDVIKDNTLLVDIINGHGASFVLKIINDEIYIDINVSVPFDKKI
ATTNKVVGIDVNIKHMLLATNILDDGNVKGYVNIYKEVINDSDFKKVCNSTVMQYFTDFSKFVTF
CPLEFDFLFSRVCNQKGIYNDNSAMEKSFSDVLNKLKWNFIETGDNTKRIYIENVMKLRSQMKA
YAIVKNAYYKQQSEYDFGKSEEFIQEHPFSNTDKGIEILNKLDNISKKILGCRNNIIQYSYNLFEIN
GYDMVSLEKLTSSQFKKKPFPTVNSLLKYHKILGCTQEEMEKKDIYSVIKKGYYDIIFDNDVVTD
AKLSAKGELSKFKDDFFNLMIKSIHFADIKDYFITLSNNGTAGVSLVPSYFTSQMDSIDHKIYFVQ
DNKSGKLKLANKHKVRSSQEKHINGLNADYNAARNIAYIMENTDCRNMFMKQSRTDKSLYNKP
SYETFIKTQGSAVAKLKKEGFVKILDEASV*

重复:ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC (SEQ ID NO:1)

间隔子长度: 30

图2AA

>Cas12L_28_255238293 (SEQ ID NO:127)

MAHKKNLEGENLQVKTICLKANLSKEEVKEKWLPVINEYNVYYNRMSDYICSLLGTNITIGEFAE
QLSIEKRKNGYFTICQDDKFKNESLYKIFHKSFPINHGTNIINNIISEKNIDQYDGNTLGFRPTMYR
LRGYVDSVIGNYRTTIRTIKPSVKRHKISVDSSFDEKMEQCIYEIQKGNLKTVSEWNNKIDYLLSK
SDMNPLTIDRFNLLRDFYVDNETEVNEKSNNSSIEQLVKFGGCHRKGDNMTLSLTEANFSIEEID
DSYGYLLTLHTDSGDYKIPLMGSKMLKKGDKCLIDFVNCKKGKSLTAKIDNDYNLYFHFVVYSN
FEKIEDDNINNVVGVDVNSKHMLLMTNVIDDNIDGYVNIYKALVNDDEFKSLVTKSEYDDYVTMS
KYVTFCPIELKYLYARYCVQKDYPISNKDVAIEQCISRVIDKLCKETLDSRANNYLCMVRRIRHYY
KSYYVLKMTYYDKMSEYDTNMEYNDISTTSKETMDQRRFENSFRETDCAKEILSKLDKIGNDIL
GCRNNILTYAYKLFEELGYDTIALENLESSQFDKMKSLPSCQSMLKYHKLEGKTMEEVMSNTSV
KSLIENEYYDFSLNDNKTVENITYTKNGLMKKGFDEFFNLFMKIIHFADIKDKFLQLYNNGSVKVI
LVPSYFTSQMDSSNHSIYMEKSKNDKLVFASKHKVRKTQETHLNGLNADYNAACNIAYIAKDIK
WREKFCKKTSNNGYSTPFYDCATKNQIEMVKRIKQLNAIKMLA*

重复: ATTGTTGAAATAGTACTTTTATAGTCTATATACAAC (SEQ ID NO:11)

间隔子长度: 30

图2BB

>Cas12L_29_72167294 (SEQ ID NO:128)

MASHKKTESNQIIKTFSFKIKNANGLSLDVLNDAITEYQNYYNICSDWIKDHLTMKIGELYKYIPDE
KKNSGYALTLISDEWKDKPMYMMFKKGYPANSRDNAIYEALNTCNTEHYTGNILNFSDTYYRR
FGYVASTISNYVTKISKMSTGSRSKNISNDSDVDTIMEQVIYEMEHNGWTSVKDWENQMEYLE
SKTDSNPNFVYRMTTLYEFYKSHIDEVNSKMETMSIDSLIKFGGCRRKDSKKSMYIMGGSNTPF
DITQIGGNSLNIKFSKNLNVDVFGRYDVIKDNTLLVDIINGHGASFVLKIINDEIYIDINVSVPFDKKI
ATTNKVVGVDVNIKHMLLATNILDDGNVNGYVNIYKEVINDSDFKKVCNSTVMKYFTDFSKFVTF
CPLEFDFLFSRVCNQKGIYNDNSAMEKSFSDVLNKLKWNFIETGDNTKRIYIENVMKLRSQMKA
YAIVKNAYYKQQSEYDFGKSEEFIQEHPFSNTDKGIEILNKLDNISKKILGCRNNIIQYSYNLFEIN
GYDMISLEKLTSSQFKKKSFPTVDSLLKYHKILGCTQEEMEKKDIYSVIKKGYYDIIFDNDVVTDA
KLSTKGELSKFKDDFFNLMIKSIHFADIKDYFITLSNNGTAGVSLVPSYFTSQMDSIDHKIYFVQD
NKSGKLKLANKHKVRSSQEKHINGLNADYNAARNIAYIMENTECRNMFMKQSRTDKSLYNKPS
YETFIKTQGSAVAKLKKEGFVKILDEASV*

重复: TATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC (SEQ ID NO:13)

间隔子长度: 29

图2CC

>Cas12L_30_72369269 (SEQ ID NO:129)

MASHKKTESNQIIKTFSFKIKNANGLSLDVLNDAITEYQNYYNICSDWIKDHLTMKIGELYKYIPDE
KKNSGYALTLISDEWKDKPMYMMFKKGYPANSRDNAIYEALNTCNTEHYTGNILNFSDTYYRR
FGYVASAISNYVTKISKMSTGSRSKNISNDSDVDTIMEQVIYEMEHNGWTSVKDWENQMEYLE
SKTDSNPNFVYRMTTLYEFYKSHIDEVNSKMETMSIDSLIKFGGCRRKDSKKSMYIMGGSNTPF
DITQIDGNSLNIKFSKNLNVDVFGRYDVIKDNTLLVDIINGHGASFVLKIINDEIYIDINVSVPFDKKI
ATTNKVVGIDVNIKHMLLATNILDDGNVNGYVNIYKEVINDSDFKKVCNSTVMKYFTDFSKFVTF
CPLEFDFLFSRVCNQKGIYNDNSAMEKSFSNVLNKLKWNFIETGDNTKRIYIENVMKLRSQMKA
YAIVKNAYYKQQSEYDFGKSEEFIQEHPFSNTDKGIEILNKLDNISKKILGCRNNIIQYSYNLFEIN
GYDMISLEKLTSSQFKKKPFPTVNSLLKYHKILGCTQEEMEKKDIYSVIKKGYYDIIFDNDVVTDA
KLSAKGELSKFKDDFFNLMIKSIHFADIKDYFITLSNNGTAGVSLVPSYFTSQMDSIDHKIYFVQD
NKSGKLKLANKHKVRSSQEKHINGLNADYNAARNIAYIMENTDCRNMFMKQSRTDKSLYNKPS
YETFIKTQGSAVAKLKKEGFVKILDEASV*

重复: ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC (SEQ ID NO:1)
间隔子长度: 30

图2DD

>Cas12L_31_72503976 (SEQ ID NO:130)

MASHKKTESNQIIKTFSFKIKNANGLSLDVLNDAITEYQNYYNICSDWIKDHLTMKIGELYKYIPDE
KKNSGYALTLISDEWKDKPMYMMFKKGYPANNRDNAIYETLNTCNTEHYTGNILNFSDTYYRR
FGYVASTISNYVTKISKMSTRSRSKNISNDSDVDTIMEQVIYEMEHNGWTSVKDWENQMEYLE
SKTDSNPNFVYRMTTLYEFYKSHIDEVNSKMETMSIDSLIKFGGCRRKDSKKSMYIMGGSNTPF
DITQIGGNSLNIKFSKNLNVDVFGRYDVIKDNTLLVDIINGHGASFVLKIINGEIYIDINVSVPFDKKI
ATTNKVVGIDVNIKHMLLATNILDDGNVNGYVNIYKEVINDSDFKKVCNSTVMKYFTDFSKFVTF
CPLEFDFLFSRVCNQKGIYNDNSAMEKSFSDVLNKLKWNFIETGDNTKRIYIENVMKLRSQMKA
YAIVKNAYYKQQSEYDFGKSEEFIQEHPFSNTDKGIEILNKLDNISKKILGCRNNIIQYSYNLFEIN
GYDMISLEKLTSSQFKKKPFPTVNSLLKYHKILGCTQEEMEKKDIYSVIKKGYYDIIFDNDVVTDA
KLSTKGELSKFKDDFFNLMIKSIHFADIKDYFITLSNNGTAGVSLVPSYFTSQMDSIDHKIYFVQD
NKSGKLKLANKHKVRSSQEKHINGLNADYNAARNIAYIMENTDCRNMFMKQSRTDKSLYNKPS
YETFIKTQGSAVAKLKKEGFVKILDEASV*

重复: ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC (SEQ ID NO:1)
间隔子长度: 30

图2EE

>Cas12L_32_72547654_ 部分 (SEQ ID NO:131)

NMFMKQSRTDKSLYNKPSYETFIKTQGSAVAKLKKEGFVKILDEASV*

重复: ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC (SEQ ID NO:1)
间隔子长度: 30

图2FF

>Cas12L_33_72907394 (SEQ ID NO:132)

MASHKKTESNQIIKTFSFKIKNANGLSLDVLNDAITEYQNYYNICSDWIKDHLTMKIGELYKYIPDE
KKNSGYALTLISDEWKDKPMYMMFKKGYPANNRDNAIYETLNTCNTEHYTGNILNFSDTYYRR
FGYVASTISNYVTKISKMSTGSRSKNISNDSDVDTIMEQVIYEMEHNGWTSVKDWENQMEYLE
SKTDSNPNFVYRMTTLYEFYKSHIDEVNSKMETMSIDSLIKFGGCRRKDSKKSMYIMGGSNTPF
DITQIGGNFLNIKFSKNLNVDVFGRYDVIKDNTLLVDIINEHGASFVLKIINDEIYIDINVSVPFDKKI
ATTNKVVGVDVNIKHMLLATNILDDGNVNGYVNIYKEVINDSDFKKVCNSTVMKYFTDFSKFVTF
CPLEFDFLFSRVCNQKGIYNDNSAMEKSFSDVLNKLKWNFIETGDNTKRIYIENVMKLRSQMKA
YAIVKNAYYKQQSEYDFGKSEEFIQEHPFSNTDKGIEILHKLDNISKKILGCRNNIIQYSYNLFEIN
GYDMISLEKLTSSQFKKKPFPTVNSLLKYHKILGCTQEEMEKKDIYSVIKKGYYDIIFDNDVVTDA
KLSTKGELSKFKDDFFNLMIKSIHFADIKDYFITLSNNGTAGVSLVPSYFTSQMDSIDHKIYFVQD
NKSGKLKLANKHKVRSSQEKHINGLNADYNAARNIAYIMENTECRNMFMKQSRTDKSLYNKPS
YETFIKTQGSAVAKLKKEGFVKILDEASV*

重复:ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC (SEQ ID NO:1)

间隔子长度: 30

图2GG

>Cas12L_34_73124743_ 部分 (SEQ ID NO:133)

KKILGCRNNIIQYSYNLFEINGYDMISLEKLTSSQFKKKPFPTVNSLLKYHKILGCTQEEMEKKDIY
SVIKKGYYDIIFDNDVVTDAKLSTKGELSKFKDDFFNLMIKSIHFADIKDYFITLSNNGTAGVSLVP
SYFTSQMDSIDHKIYFVQDNKSGKLKLANKHKVRSSQEKHINGLNADYNAARNIAYIMENTECR
NMFMKQSRTDKSLYNKPSYETFIKTQGSAVAKLKKEGFVKILDEASV*

重复:ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC (SEQ ID NO:1)

间隔子长度: 30

图2HH

>Cas12L_35_73503649_ 部分 (SEQ ID NO:134)

NAYYKQQSEYDFGKSEEFIQEHPFSNTDKGIEILHKLDNISKKILGCRNNIIQYSYNLFEINGYDMI
SLEKLTSSQFEKKPFPTVNSLLKYHKILGCTQEEMEKKDIYSVIKKGYYDIIFDNDVVTDAKLSTK
GELSKFKDDFFNLMIKSIHFADIKDYFITLSNNGTVGVSLVPSTV*

重复:ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC (SEQ ID NO:1)

间隔子长度: 30

图2II

>Cas12L_36_73503649_ 部分 (SEQ ID NO:135)

MANKHKVRSSQEKHINGLNADYNAARNIAYIIENTECRNMFMKQSRTDKSLYNKPSYETFIKTQ
GSTVAKLKKEGFVKILDEASV*

图2JJ

>Cas12L_37_73472625 (SEQ ID NO:136)

MAHKKQKEENEIIKTISLKVKDYAGYPIVEAMREYTKYYNKISQWINSNLLTIKIGELSAFMPDEC
KTHNYYTYMMSPDWVNEPLYKMFMKGFHAQHCDNILFNVVKTLNIDEYAGNSLGLSASCFRR
SGYFQNVVSNYKSKFANPHISIRRKNLSDLPTEDELVEQCIYEIQNGLSSKTKWEEQIEYLKERD
DSKQIYLTRLNTLFMYYKANKDFVDEQIQIKSVESLANFGGCVRKDDKLSMNLVFSSNSPYKVV
LNEKRNGYILSYSNNFSIELYGNRMGLLNGVEVFNVGDKHSNNITFKMDNDELFVNIPVSVNFV
KKANETNKVVGVDVNLKHSIFATNIIDDGKLDGFVNIYRELLNDVDFVKSCPNELLNFILDVEKYA
FFMPLELGLLSSRVMNQCGYSTIGKYEKLFTTEEHFFRVLRQLEKRFQESGENQKRIYIENVIKM
RAQVKAYFTLKYAYNKANKDYDLKMGFVDESTANKETMDQRRFENQFVNTYTAKEILGKMRRI
ANVITSCRNNIICYMYKIFENNGYGVVALEKLQSSQMKKEKRIPSLLSLLKKQKVEGYTINELKDK
SVFKFIERGYYTFDFDDDNKITGVQFSDAGEVVNMETELHNLALKTIHFADAKDYFVTLSNNGS
VSVALVPSQFTSQMDSTKHVLYAKKNNKGKLGIVSEHEVRPKQECHINGLNGDYNAACNIAYIF
ENDEWRNAFMKMNPNEYGKALFETNMESTSTIINTLKKINPDNIISFDEYEKTKKVAA*

重复: ATTGTTGTAACATCTATTTTGTAAGGTGTAAACAAC (SEQ ID NO:15)
间隔子长度: 30

图2KK

>Cas12L_38_73764039_ 部分 (SEQ ID NO:137)

YPANSRDNAIYEALNTCNTEHYTGNILNFSDTYYRRFGYVASAISNYVTKISKMSTGSRSKNISN
DSDVDTIMEQVIYEMEHNGWTSVKDWENQMEYLESKTDSNPNFVYRMTTLYEFYKSHIDEVN
SKMETMSIDSLIKFGGCRRKDSKKSMYIMGGSNTPFDITQIGGNSLNIKFSKNLNVDVFGRYDVI
KDNTLLVDIINGHGASFVLKIINGEIYIDINVSVPFDKKIATTNKVVGVDVNIKHMLLATNILDDGNV
KGYVNIYKEVINDSDFKKVCNSTVMKYFTDFSKFVTFCPLEFDFLFSRVCNQKGIYNDNSAMEK
SFSDVLNKLKWNFIETGDNTKRIYIENVMKLRSQMKAYAIVKNAYYKQQSEYDFGKSEEFIQEH
PFSNTDKGIEILNKLDNISKKILGCRNNIIQYSYNLFEINGYDMISLEKLTSSQFKKKSFPTVNSLLK
YHKILGCTQEEMEKKDIYSVIKKGYYDIIFDNDVVTDAKLSTKGELSKFKDDFFNLMIKSIHFADIK
DYFITLSNNGTAGVSLVPSYFTSQMDSIDHKIYFVQDNKSGKLKLANKHKVRSSQEKHINGLNA
DYNAARNIAYIMENTECRNMFMKQSRTDKSLYNKPSYETFIKTQGSAVAKLKKEGFVKILDEAS
V*

重复: ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC (SEQ ID NO:1)
间隔子长度: 30

图2LL

>Cas12L_40_74037305 (SEQ ID NO:138)

MMKKMRISPHLFYIFFKKIWKCHFFVLSLYQLNQYIMASHKKTESNQIIKTFSFKIKNANGLSLDV
LNDAITEYQNYYNICSDWIKDHLTMKIGELYKYIPDEKKNSGYALTLISDEWKDKPMYMMFKKG
YPANNRDNAIYETLNTCNTEHYTGNILNFSDTYYRRFGYVASAISNYVTKISKMSTGSRSKNISN
DSDVDTIMEQVIYEMEHNGWTSVKDWENQMEYLESKTDSNPNFVYRMTTLYEFYKSHIDEVN
SKMETMSIDSLIKFGGCRRKDSKKSMYIMGGSNTPFDITQIGGNSLNIKFSKNLNVDVFGRYDVI
KDNTLLVDIINEHGASFVLKIINDEIYIDINVSVPFDKKIATTNKVVGVDVNIKHMLLATNILDDGNV
NGYVNIYKEVINDSDFKKVCNSTVMKYFTDFSKFVTFCPLEFDFLFSRVCNQKGIYNDNSAMEK
SFSNVLNKLKWNFIETGDNTKRIYIENVMKLRSQMKAYAIVKNAYYKQQSEYDFGKSEEFIQEH
PFSNTDKGIEILNKLDNISKKILGCRNNIIQYSYNLFEINGYDMISLEKLTSSQFKKKPFPTVNSLLK
YHKILGCTQEEMEKKDIYSVIKKGYYDIIFDNDVVTDAKLSTKGELSKFKDDFFNLMIKSIHFADIK
DYFITLSNNGTAGVSLVPSYFTSQMDSIDHKIYFVQDNKSGKLKLANKHKVRSSQEKHINGLNA
DYNAARNIAYIMENTDCRNMFMKQSRTDKSLYNKPSYETFIKTQGSAVAKLKKEGFVKILDEAS
V*

重复:ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC (SEQ ID NO:1)

间隔子长度: 30

图2MM

>Cas12L_41_74408273_ 部分 (SEQ ID NO:139)

WTDEDYCKFFAKYGMSEECQKWMCRDVYDYRIKDFVDYEKFDDSKIEEEQEEYSEIPVESED
SSPSTSETLPKSIYDLDLRRLDPNPEIPEDEKYNEEDLKNAYPDKYERFEKDGEDYSYIFFKKIW
KCHFFVLSLYQLNQYIMASHKKTESNQIIKTFSFKIKNANGLSLDVLNDAITEYQNYYNICSDWIK
DHLTMKIGELYKYIPDEKKNSGYALTLISDEWKDKPMYMMFKKGYPANNRDNAIYEALNTCNTE
HYTGNILNFSDTYYRRFGYVASTISNYVTKISKMSTGSRSKNISNDSDVDTIMEQVIYEMEHNG
WTSVKDWENQMEYLESKTDSNPNFVYRMTTLYEFYKSHIDEVNSKMETMSIDSLIKFGGCRRK
DSKKSMYIMGGSNTPFDITQIDGNSLNIKFSKNLNVDVFGRYDVIKDNTLLVDIINGHGASFVLKII
NDEIYIDINVSVPFDKKIATTNKVVGVDVNIKHMLLATNILDDGNVNGYVNIYKEVINDSDFKKVC
NSTVMKYFTDFSKFVTFCPLEFDFLFSRVCNQKGIYNDNSAMEKSFSDVLNKLKWNFIETGDNT
KRIYIENVMKLRSQMKAYAIVKNAYYKQQSEYDFGKSEEFIQEHPFSNTDKGIEILNKLDNISKKIL
GCRNNIIQYSYNLFEINGYDMISLEKLTSSQFEKKPFPTVNSLLKYHKILGCTQEEMEKKDIYSVI
KKGYYDIIFDNGVVTDAKLSTKGELSKFKDDFFNLMIKSIHFADIKDYFITLSNNGTAGVSLVPSY
FTSQMDSIDHKIYFVQDNKSGKLKLANKHKVRSSQEKHINGLNADYNAARNIAYIMENTECRNM
FMKQSRTDKSLYNKPSYETFIKTQGSAVAKLKKEGFVKILDEASV*

重复:ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC (SEQ ID NO:1)

间隔子长度: 30

图2NN

>Cas12L_42_75186079_ 部分 (SEQ ID NO:140)
YDMISLEKLTSSQFKKKPFPTVNSLLKYHKILGCTEEEMEKKDIYSVIKKGYYDIIFDNDVVTDAKL
STKGELSKFKDDFFNLMIKSIHFADIKDYFITLSNNGTAGVSLVPSYFTSQMDSIDHKIYFVQDNK
SGKLKLANKHKVRSSQEKHINGLNADYNAARNIAYIMENTECRNMFMKQSRTDKSLYNKPSYE
TFIKTQGSAVAKLKKEGFVKILDEASA*

重复:AATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC (SEQ ID NO:17)
间隔子长度: 29

图200

>Cas12L_43_75257103_ 部分 (SEQ ID NO:141)
MKIGELYKYIPDEKKNSGYALTLISDEWKDKPMYMMFKKGYPANNRDNAIYETLNTCNTEHYTG
NILNFSDTYYRRFGYVASAISNYVTKISKMSTGSRSKNISNDSDVDTIMEQVIYEMEHNGWTSVK
DWENQMEYLESKTDSNPNFVYRMTTLYEFYKSHIDEVNSKMETMSIDSLIKFGGCRRKDSKKS
MYIMGGSNTPFDITQIDGNSLNIKFSKNLNVDVFGRYDVIKDNTLLVDIINGHGASFVLKIINGEIYI
DINVSVPFDKKIATTNKVVGVDVNIKHMLLATNILDDGNVNGYVNIYKEVINDSDFKKVCNSTVM
KYFTDFSKFVTFCPLEFDFLFSRVCNQKGIYNDNSAMEKSFSDVLNKLKWNFIETGDNTKRIYIE
NVMKLRSQMKAYAIVKNAYYKQQSEYDFGKSEEFIQEYDFFGETYQFSI*

重复:ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC (SEQ ID NO:1)
间隔子长度: 30

图2PP

>Cas12L_44_75257103_ 部分 (SEQ ID NO:142)
MISLEKLTSSQFKKKPFPTVNSLLKYHKILGCTQEEMEKKKGYYDIIFDNDVVTDAKLSTKGELS
KFKDDFFNLMIKSIHFADIKDYFITLSNNGTAGVSLVPSYFTSQMDRANKHKVRSSQEKHINGLN
ADYNAARNIAYIMENTECRNMFMKQSRTDKSLYNKPSYETFIKTQGSAVAKLKKEGFVKILDEA
SV*

重复:ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC (SEQ ID NO:1)
间隔子长度: 30

图2QQ

>Cas12L_45_75616607_ 部分 (SEQ ID NO:143)

YIMASHKKTESNQIIKTFSFKIKNANGLSLDVLNDAITEYQNYYNICSDWIKDHLTMKIGELYKYIP
DEKKNSGYALTLISDEWKDKPMYMMFKKGYPANSRDNAIYETLNTCNTEHYTGNILNFSDTYY
RRFGYVASTISNYVTKISKMSTGSRSKNISNDSDVDTIMEQVIYEMEHNGWTSVKDWENQMEY
LESKTDSNPNFVYRMTTLYEFYKSHIDEVNSKMETMSIDSLIKFGGCRRKDSKKSMYIMGGSNT
PFDITQIGGNSLNIKFSKNLNVDVFGRYDVIKDNTLLVDIINGHGASFVLKIINGEIYIDINVSVPFDK
KIATTNKVVGVDVNIKHMLLATNILDDGNVNGYVNIYKEVINDSDFKKVCNSTVMKYFTDFSKFV
TFCPLEFDFLFSRVCNQKGIYNDNSAMEKSFSNVLNKLKWNFIETGDNTKRIYIENVMKLRSQM
KAYAIVKNAYYKQQSEYDFGKSEEFIQEHPFSNTDKGIEILNKLDNISKKILGCRNNIIQYSYNLFE
INGYDMISLEKLTSSQFEKKPFPTVNSLLKYHKILGCTQEEMEKKDIYSVIKKGYYDIIFDNGVVT
DAKLSTKGELSKFKDDFFNLMIKSIHFADIKDYFITLSNNGTAGVSLVPSYFTSQMDSIDHKIYFV
QDNKSGKLKLANKHKVRSSQEKHINGLNADYNAARNIAYIMENTECRNMFMKQSRTDKSLYNK
PSYETFIKTQGSAVAKLKKEGFVKILDEASV*

重复: ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC (SEQ ID NO:1)
间隔子长度: 30

图2RR

>Cas12L_46_75784289_ 部分 (SEQ ID NO:144)

SLNIKFSKNLNVDVFGRYDVIKDNTLLVDIINGHGASFVLKIINGEIYIDINVSVPFDKKIATTNKVV
GVDVNIKHMLLATNILDDGNVNGYVNIYKEVINDSDFKKVCNSTVMKYFTDFSKFVTFCPLEFDF
LFSRVCNQKGIYNDNSAMEKSFSDVLNKLKWNFIETGDNTKRIYIENVMKLRSQMKAYAIVKNA
YYKQQSEYDFGKSEEFIQEHPFSNTDKGIEILNKLDNISKKILGCRNNIIQYSYNLFEINGYDMISL
EKLTSSQFKKKPFPTVNSLLKYHKILGCTQEEMEKKDIYSVIKKGYYDIIFDNGVVIDAKLSAKGE
LSKFKDDFFNLMIKSIHFADIKDYFITLSNNGTAGVSLVPSYFTSQMDSIDHKIYFVQDNKSGKLK
LANKHKVRSSQEKHINGLNADYNAARNIAYIMENTECRNMFMKQSRTDKSLYNKPSYETFIKTQ
GSAVAKLKKEGFVKILDEASV*

重复: ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC (SEQ ID NO:1)
间隔子长度: 30

图2SS

>Cas12L_47_76512228 (SEQ ID NO:145)

MASHKKTESNQIIKTFSFKIKNANGLSLDVLNDAITEYQNYYNICSDWIKDHLTMKIGELYKYIPDE
KKNSGYALTLISDEWKDKPMYMMFKKGYPANNRDNAIYETLNTCNTEHYTGNILNFSDTYYRR
FGYVASAISNYVTKISKMSTGSRSKNISNDSDVDTIMEQVIYEMEHNGWTSVKDWENQMEYLE
SKTDSNPNFVYRMTTLYEFYKSHIDEVNSKMETMSIDSLIKFGGCRRKDSKKSMYIMGGSNTPF
DITQIGGNSLNIKFSKNLNVDVFGRYDVIKDNTLLVDIINEHGASFVLKIINDEIYIDINVSVPFDKKI
ATTNKVVGIDVNIKHMLLATNILDDGNVNGYVNIYKEVINDSDFKKVCNSTVMKYFTDFSKFVTF
CPLEFDFLFSRVCNQKGIYNDNSAMEKSFSDVLNKLKWNFIETGDNTKRIYIENVMKLRSQMKA
YAIVKNAYYKQQSEYDFGKSEEFIQEHPFSNTDKGIEILNKLDNISKKILGCRNNIIQYSYNLFEIN
GYDMISLEKLTSSQFKKKSFPTVDSLLKYHKILGCTQEEMEKKDIYSVIKKGYYDIIFDNDVVTDA
KLSTKGELSKFKDDFFNLMIKSIHFADIKDYFITLSNNGTAGVSLVPSYFTSQMDSIDHKIYFVQD
NKSGKLKLANKHKVRSSQEKHINGLNADYNAARNIAYIMENTECRNMFMKQSRTDKSLYNKPS
YETFIKTQGSAVAKLKKEGFVKILDEASV*

重复: ATTGTTGTAACTTTTATTTTGTATGGAGTAAACAAC (SEQ ID NO:19)
间隔子长度: 30

图2TT

>Cas12L_48_76600450 (SEQ ID NO:146)

MASHKKTESNQIIKTFSFKIKNANGLSLDVLNDAITEYQNYYNICSDWIKDHLTMKIGELYKYIPDE
KKNSGYALTLISDEWKDKPMYMMFKKGYPANNRDNAIYETLNTCNTEHYTGNILNFSDTYYRR
FGYVASAISNYVTKISKMSTGSRSKNISNDSDVDTIMEQVIYEMEHNGWTSVKDWENQMEYLE
SKTDSNPNFVYRMTTLYEFYKSHIDEVNSKMETMSIDSLIKFGGCRRKDSKKSMYIMGGSNTPF
DITQIDGNSLNIKFSKNLNVDVFGRYDVIKDNTLLVDIINGHGASFVLKIINDEIYIDINVSVPFDKKI
ATTNKVVGVDVNIKHMLLATNILDDGNVNGYVNIYKEVINDSDFKKVCNSTVMKYFTDFSKFVTF
CPLEFDFLFSRVCNQKGIYNDNSAMEKSFSDVLNKLKWNFIETGDNTKRIYIENVMKLRSQMKA
YAIVKNAYYKQQSEYDFGKSEEFIQEHPFSNTDKGIEILNKLDNISKKILGCRNNIIQYSYNLFEIN
GYDMISLEKLTSSQFKKKPFPTVNSLLKYHKILGCTQEEMEKKDIYSVIKKGYYDIIFDNDVVTDA
KLSTKGELSKFKDDFFNLMIKSIHFADIKDYFITLSNNGTAGVSLVPSYFTSQMDSIDHKIYFVQD
NKSGKLKLANKHKVRSSQEKHINGLNADYNAARNIAYIMENTECRNMFMKQSRTDKSLYNKPS
YETFIKTQGSAVAKLKKEGFMKILDEASV*

重复: ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC (SEQ ID NO:1)
间隔子长度: 30

图2UU

>Cas12L_49_44880081 (SEQ ID NO:147)

MASHKKTESNQIIKTFSFKLKNANGLSLDVLNDAITEYQNYYNICSDWIKDHLTMKIGELYKYIPD
EKKNSGYALTLISDEWKDKPMYMMFKKGYPANNRDNAIYEALNTCNTEHYTGNILNFSDTYYR
RFGYVASTISNYVTKISKMSTGSRSKNISNDSDVDTIMEQVIYEMEHNGWTSVKDWENQMEYL
ESKTDSNPNFVYRMTTLYEFYKSHIDEVNSKMETMSIDSLIKFGGCRRKDSKKSMYIMGGSNTP
FDITQIGSNSLNIKFSKNLNVDVFGRYDVIKDNTLLVDIIKGHGASFALKIINDEIYIDINVSVPFDKK
IATTNKVVGIDVNIKHMLLATNILDDGNVNGYVNIYKEVINDSDFKKVCNSTVMKYFTDFSKFVTF
CPLEFDFLFSRVCNQKGIYNDNSAMEKSFSDVLNKLKWNFIETGDNTKRIYIENVMKLRSQMKA
YAIVKNAYYKQQSEYDFGKSEEFIQEHPFSNTDKGIEILHKLDNISKKILGCRNNIIQYSYNLFEIN
GYDMISLEKLTSSQFKKKPFPTVNSLLKYHKILGCTQEEMEKKDIYSVIKKGYYDIIFDNDVVTDA
KLSTKGELSKFKDDFFNLMIKSIHFADIKDYFITLSNNGTAGVSLVPSFFTSQMDSIDHKIYFVQD
NKSGKLKLANKHKVRSSQEKHINGLNADYNAARNIAYIMENTECRNMFMKQSRTDKSLYNKPS
YETFIKTQGSAVAKLKKEGFVKILDEASV*

重复:ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC (SEQ ID NO:1)

间隔子长度: 30

图2VV

>Cas12L_50_83012613 (SEQ ID NO:148)

MAHKKNIGAEIVKTYSFKVKNTNGITMEKLMNAIDEYQSYYNLCSDWICKNLTTMTIGDLDQYIP
EKAKDNTYATVLLDEAWKNQPLYKIFGKKYSSNNRDAVLYHALSSIVNASEKNILGISKTYYARK
GYVLNVASNYASKLSKLNTGVKSRAIKETSDEETIIEQVIYEMEHNKWESIEDWKNQIEYLNSKT
DYNPTYMERMKTLSAYYSEHKSEVDAKMQEMAVENLVKFGGCRRNNSKKSMFIMGTTNTNYT
ISYIGGNSFNINFANILNFDVYGRRDVVKNGEVLVDIMANHGDSIVLKIVNGELYADVPCSVTLNK
VESNFDKVVGIDVNMKHMLLSTSVTDNGSLDFLNIYKEMSNNAEFMALCSEDDRKYYKDISQY
VTFAPLELDLLFSRISKQGKVKMEKAYSEILEALKWKFFANGDNKNRIYVESIQKIRQQIKALCIIK
NAYYEQQSAYDIDKTQEYIEAHPFSLTEKGMSIKSKMDNICRTIIGCRNNIIDYAYSFFERNDYSII
GLEKLTSAQFEKTKSLPTCKSLLNFHKVLGHTLSELGTLPINDVVKKGYYTFTTDNEGRITDASL
SEKGKVRKMKDDFFNQTIKAIHFADVKDYFATLSNNGQTGIFFVPSQFTSQMDSSTHNLYFENA
KNGGLKLAPKYKVRQMQEYHLNGLPADYNAARNIAYIGLDETMRNTFLKKANSNKSLYNQPIYD
TGIKKTAGVFYRMKKLKRYEII*

重复:AATGTTGTAGATACCTTTTTGTAAGGATTGAACAAC (SEQ ID NO:21)

间隔子长度: 30

图2WW

>Cas12L_51_82983331 (SEQ ID NO:149)

MKRNQKHIKNIEESETIKTISFKVKDYAGIPIVEAMHEYRNYYNRLSRFINSKLLTMTIGELASLLP
ERCKSKGYYLYMTSDEWVNEYVYKMFMESFNSQSCDNIWFNYIKMNNPEEYNGNILGISDSYY
RRNGYFINVISNYKTKFKSPQINVKSKKLSESPTEEELKEQCVYEYVKHNLHSKKDWEEQIKYLD
ERGESKINILERIRTLYQYYKENIPTIKEYIELKSIESIEKFGGCVRKEDKLSMSLQYVSTHNYEIKL
NDTRNGYIISGISKDLSFEVYGNRMGLLGGEEILNIPEKHSTSITFVMRNNSLYVDIPVAVPFSKVI
NDCDGKTVGIDVNLKHALFATSEVDNGQFYDYVNVYAELLKDENFVKVCHKELLDYIKDVSKYV
FFAPIELNLLLSRVMKQKGYENIDNYKKLYKVEEAYLCVLDKLQKRFIDEGNNTKRIYIENLKKMR
AQMKAYYILKDTYSKYQKDYDIEMGFVDESTESKETMDARRSENPFRSTDIAQDILKKMNNVGK
TVEACRNNIIAYIYKVFENSDFATIVLEKLQSSQMKKHKRIPTVNSLLKYHHVEGHTIEETKEMKIY
DVVEKGYYNFIVNEKNEIIDATLTDKGKVIMIEAEFYNFALKSIHFADAKDYFITLSNNGSVNIALV
PSQFTSQMDSIRHAIFVTKGKKGKKVIVDKKYVRPKQEKHINGLNGDYNASRNIAYIFENEELRE
ELLKKEEEYNKYGKVLYDTLIKFPSGVINKLKKFGDKYMTTIENLDEIQVEDVAYV*

重复: ATTGTTGTAATACTATTTTTGTAAAGTATAAACAAC (SEQ ID NO:23)

间隔子长度: 30

图2XX

>Cas12L_52_76767885 (SEQ ID NO:150)

MSKLNTGVKRHTIKETSDEATIVEQVIYEMEHNKWESIEDWKNQIEYLNSKTDYNPTYMERMKT
LSAYYSEHKSEVDAKMQEMAVENLVKFGGCRRNNSKKSMFIMGSSKTTYTISYIGDNCFNINFA
NILNFDVYGRRDVVKNGEVLVDIMANHGDSIVLKIVNGELYADVPCSTTLNKVESTFDKVAGIDV
NMKHMLLSTSVTDNGNSDFVNIYKEMSNNAEFMALCPEEDRKYYKDISQYVTFAPLELDLLFSR
ISKQGKVKMEKAYSEILETLKWKFFANGDNKNRIYVESIQKIRQQIKALCVIKNAYYEQQSAYDID
KTQEYIEAHPFSLTEKGMSIKSKMDNICRTIIGCRNNIIDYAYSFFERNDYSIIGLEKLTSSQFEKT
KSLPTCKSLLNFHKVLGHTLSELETLPINDVVKKGYYTFTTDNEGRITDASLSEKGKVRKMKDDF
FNQAIKAIHFADVKDYFATLSNNGQTGIFFVPSQFTSQMDSNTHTLYFENAKNGGLKLASKYKV
RQTQEYHLNGLPADYNAARNIAYIGLDETMRNTFLKKANSNKSLYNQPIYDTGIKKTAGVFSRM
KKLKRYEII*

重复: AATGTTGTAGATGCCTTTTTATAAGGATTAAACAAC (SEQ ID NO:25)

间隔子长度: 30

图2YY

>Cas12L_53_77216451 (SEQ ID NO:151)
MASHKKTESNQIIKTFSFKIKNANGLSLDVLNDAITEYQNYYNICSDWIKDHLTMKIGELYKYIPDE
KKNSGYALTLISDEWKDKPMYMMFKKGYPANNRDNAIYEALNTCNTEHYTGNILNFSDTYYRR
FGYVASAISNYVTKISKMSTGSRSKNISNDSDVDTIMEQVIYEMEHNGWTSVKDWENQMEYLE
SKTDSNPNFVYRMTTLYEFYKSHIDEVNSKMETMSIDSLIKFGGCRRKDSKKSMYIMGGSNTPF
DITQIGGNSLNIKFSKNLNVDVFGRYDVIKDNTLLVDIINGHGASFVLKIINDEIYIDINVSVPFDKKI
ATTNKVVGIDVNIKHMLLATNILDDGNVNGYVNIYKEVINDSDFKKVCNSTVMKYFTDFSKFVTF
CPLEFDFLFSRVCNQKGIYNDNSAMEKSFSDVLNKLKWNFIETGDNTKRIYIENVMKLRSQMKA
YAIVKNAYYKQQSEYDFGKSEEFIQEHPFSNTDKGIEILNKLDNISKKILGCRNNIIQYSYNLFEIN
GYDMISLEKLTSSQFKKKPFPTVNSLLKYHKILGCTQEEMEKKDIYSVIKKGYYDIIFDNDVVTDA
KLSTKGELSKFKDDFFNLMIKSIHFADIKDYFITLSNNGTAGVSLVPSYFTSQMDSIDHKIYFVQD
NKSGKLKLANKHKVRSSQEKHINGLNADYNAARNIAYIMENTDCRNMFMKQSRTDKSLYNKPS
YETFIKTQGSAVAKLKKEGFVKILDEASV*

重复:ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC (SEQ ID NO:1)
间隔子长度: 30

图2ZZ

>Cas12L_54_77468912_ 部分 (SEQ ID NO:152)
YETFIKTQGSAVAKLKKEGFVKILDEASV*

重复:ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC (SEQ ID NO:1)
间隔子长度: 30

图2AAA

>Cas12L_55_77738117 (SEQ ID NO:153)
MASHSSLSNNQIFKTFSFKVKSSNLSKDFFDVIKEYQEYYNKCSDVILENLTCIKIGDIFDMIPEKS
KKSDYAQYAISDEWKNVPLYNIFSKAFAPMHRDNLLYIWLTKLVPYTGNLLKISDTFYKRNGFIK
SVISNYTTSFTNIKPKVKFQKLTGDDSHEMLLTQTICDMVKFNLYDIKSWEEMVSYFEMKSETSE
DTLNRIHTLFDFYKNNTPEVEDKYNELVTESLSKFGGCRRDMSKLTMSIQLSKKVIKVTHGYNTL
NYKYGKLIDLELWGRKDIINNDELLINLENVCEMIVFKIKNGEIYVDIPFKVDFIKKDQTIDKIAGVD
ANIKHMLLSTSVKDENLIGYTNIYKEVINDSDFQKVCDNKTMKILQEISNYVTFAPIEFDMLFSRIS
KQREMKDKYINMEIAFTNVLNKLKQKFIVNSDNKNRIYIESILKIRSQLKSYAILEEVKYKKASEYD
SAIVEEFGVEYLETHPFKDTETYKEINKKILNISENIIGCRNNIIQYAYKIFESNGFDMISLENLTNS
NFKKEKNMPTIKSLLSYHHVLGKTNEEIEKLDVYSVIKKGYYTFEYKDGKVVNAKLSEIGEMIKIK
TTMFNMMIKSIHFAEIKDYFITLANNGEVGVSLVPSYYTSQMDSTDHKVFGLLSKKGKWTLVDK
RKVRKNQETHINGLNADYNAAKNIAFILSDEVWRNKFTKKTKTPKYNTPSYYTSINSQGKMLRA
LKSLKAFKEFKI*

重复: ATTGTTGTAATACACTTTTTATAAGGTATGAACAAC (SEQ ID NO:27)
间隔子长度: 30

图2BBB

>Cas12L_56_65286425 (SEQ ID NO:154)

MKiGELYKYIPDEKKNSGYALTLISDEWKDKPMYMMFKKGYPANNRDNAIYEALNTCNTEHYTG
NILNFSDTYYRRFGYVASAISNYVTKISKMSTGSRSKNISNDSDVDTIMEQVIYEMEHNGWTSVK
DWENQMEYLESKTDSNPNFVYRMTTLYEFYKSHIDEVNSKMETMSIDSLIKFGGCRRKDSKKS
MYIMGGSNTPFDITQIGGNSLNIKFSKNLNVDVFGRYDVIKDNTLLVDIINEHGASFVLKIINDEIYI
DINVSVPFDKKIATTNKVVGIDVNIKHMLLATNILDDGNVNGYVNIYKEVINDSDFKKVCNSTVMQ
YFTDFSKFVTFCPLEFDFLFSRVCNQKGIYNDNSAMEKSFSDVLNKLKWNFIETGDNTKRIYIEN
VMKLRSQMKAYAIVKNAYYKQQSEYDFGKSEEFIQEHPFSNTDKGIEILNKLDNISKKILGCRNNI
IQYSYNLFEINGYDMISLEKLTSSQFKKKPFPTVNSLLKYHKILGCTQEEMEKKDIYSVIKKGYYDI
IFDNDVVTDAKLSTKGELSKFKDDFFNLMIKSIHFADIKDYFITLSNNGTAGVSLVPSYFTSQMDS
IDHKIYFVQDNKSGKLKLANKHKVRSSQEKHINGLNADYNAARNIAYIMENTECRNMFMKQSRT
DKSLYNKPSYETFIKTQGSAVAKLKKEGFVKILDEASV*

重复: ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC (SEQ ID NO:1)
间隔子长度: 30

图2CCC

>Cas12L_57_65567118_ 部分 (SEQ ID NO:155)

KIINDEIYIDINVSVPFDKKIATTNKVVGIDVNIKHMLLATNILDDGNVNGYVNIYKEVINDSDFKKV
CNSTVMKYFTDFSKFVTFCPLEFDFLFSRVCNQKGIYNDNSAMEKSFSDVLNRLKWNFIETGD
NTKRIYIENVMKLRSQMKAYAIVKNAYYKQQSEYDFGKSEEFIQEHPFSNTDKGIEILNKLDNISK
KILGCRNNIIQYSYNLFEINGYDMISLEKLTSSQFKKKPFPTVNSLLKYHKIRGGTQEEMEKKDIY
SVIKKGYYDIIFDNDVVTDAKLSTKGELSKFKDDFFNLMIKSIHFADIKDYFITLSNNGTAGVSLVP
SYFTSQMDSIDHKIYFVQDNKSGKLKLANKHKVRSSQEKHINGLNADYNAARNIAYIMENNECR
NMFMKQSRTDKSLYNKPSYETFIKTQGSAVAKLKKEGFVKILDEASV*

重复: ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC (SEQ ID NO:1)
间隔子长度: 30

图2DDD

>Cas12L_58_66287853_ 部分 (SEQ ID NO:156)

DDGNVNGYVNIYKEVINDSDFKKVCNSTVMKYFTDFSKFVTFCPLEFDFLFSRVCNQKGIYNDN
SAMEKSFSDVLNKLKWNFIETGDNTKRIYIENVMKLRSQMKAYAIVKNAYYKQQSEYDFGKSEE
FIQEHPFSNTDKGIEILNKLDNISKKILGCRNNIIQYSYNLFEINGYDMISLEKLTSSQFKKKPFPTV
NSLLKYHKILGCTQEEMEKKDIYSVIKKGYYDIIFDNDVVTDAKLSTKGELSKFKDDFFNLMIKSIH
FADIKDYFITLSNNGTAGVSLVPSYFTSQMDSIDHKIYFVQDNKSGKLKLANKHKVRSSQEKHIN
GLNADYNAARNIAYIMENTECRNMFMKQSRTDKSLYNKPSYETFIKTQGSAVAKLKKDGFVKIL
DEASA*

重复: ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC (SEQ ID NO:1)
间隔子长度: 30

图2EEE

>Cas12L_39_73877227 (SEQ ID NO:157)

MASHKKTESNQIIKTFSFKIKNANGLSLDVLNDAITEYQNYYNICSDWIKDHLTMKIGELYKYIPDE
KKNSGYALTLISDEWKDKPMYMMFKKGYPANSRDNAIYEALNTCNTEHYTGNILNFSDTYYRR
FGYVASAISNYVTKISKMSTGSRSKNISNDSDVDTIMEQVIYEMEHNGWTSVKDWENQMEYLE
SKTDSNPNFVYRMTTLYEFYKSHIDEVNSKMETMSIDSLIKFGGCRRKDSKKSMYIMGGSNTPF
DITQIDGNSLNIKFSKNLNVDVFGRYDVIKDNTLLVDIINGHGASFVLKIINGEIYIDINVSVPFDKKI
ATTNKVVGIDVNIKHMLLATNILDDGNVNGYVNIYKEVINDSDFKKVCNSTVMKYFTDFSKFVTF
CPLEFDFLFSRVCNQKGIYNDNSAMEKSFSDVLNKLKWNFIETGDNTKRIYIENVMKLRSQMKA
YAIVKNAYYKQQSEYDFGKSEEFIQEHPFSNTDKGIEILNKLDNISKKILGCRNNIIQYSYNLFEIN
GYDMISLEKLTSSQFKKKPFPTVNSLLKYHKILGCTQEEMEKKDIYSVIKKGYYDIIFDNDVVTDA
KLSTKGELSKFKDDFFNLMIKSIHFADIKDYFITLSNNGTAGVSLVPSYFTSQMDSIDHKIYFVQD
NKSGKLKLANKHKVRSSQEKHINGLNADYNAARNIAYIMENTECRNMFMKQSRTDKSLYNKPS
YETFIKTQGSAVAKLKKEGFVKILDEASV*

重复: ATTGTTGTAACTCTTATTTTGTATGGAGTAAACAAC (SEQ ID NO:1)

间隔子长度: 30

图2FFF

	RuvC-I	RuvC-II	RuvC-III
Cas12L_1_257905508	336	530	682
Cas12L_2_196848753	335	523	676
Cas12L_3_66741167	368	562	714
Cas12L_4_67031163	368	562	714
Cas12L_5_67793351	368	562	714
Cas12L_6_67912869_部分		139	292
Cas12L_7_68090316_部分	302	501	653
Cas12L_8_68328292_部分	359	553	705
Cas12L_9_68454124	336	530	682
Cas12L_10_68605313	372	566	718
Cas12L_11_69266821_部分			28
Cas12L_12_69417229_部分			28
Cas12L_13_69733214	336	530	682
Cas12L_14_70235246_部分	17	211	363
Cas12L_15_70724743	336	530	682
Cas12L_16_70731038	335	523	676
Cas12L_17_70959391	336	530	682
Cas12L_18_71078086	336	530	682
Cas12L_19_71193509_部分	270	464	616
Cas12L_20_71210958_部分	116	304	457
Cas12L_21_71317321_部分	175	369	521
Cas12L_22_71456687	336	530	682
Cas12L_23_71708971	175	369	521
Cas12L_24_46035167_部分	348	542	594
Cas12L_25_46784254_部分	186	380	532
Cas12L_26_46464451	336	530	
Cas12L_27_254489164	336	530	682
Cas12L_28_255238293	343	542	701
Cas12L_29_72167294	336	530	682
Cas12L_30_72369269	336	530	682
Cas12L_31_72503976	336	530	682
Cas12L_33_72907394	336	530	682
Cas12L_34_73124743_部分		29	181
Cas12L_35_73503649_部分		70	
Cas12L_36_73503649_部分			21
Cas12L_37_73472625	334	545	698
Cas12L_38_73764039_部分	242	436	588
Cas12L_39_73877227	336	530	682
Cas12L_40_74037305	372	566	718

图3A

Cas12L_41_74408273_部分	479	673	825
Cas12L_42_75186079_部分		7	159
Cas12L_43_75257103_部分	282		
Cas12L_44_75257103_部分		5	132
Cas12L_45_75616607_部分	338	532	684
Cas12L_46_75784289_部分	71	265	417
Cas12L_47_76512228	336	530	682
Cas12L_48_76600450	336	530	682
Cas12L_49_44880081	336	530	682
Cas12L_50_83012613	335	523	676
Cas12L_51_82983331	341	552	705
Cas12L_52_76767885	192	380	533
Cas12L_53_77216451	336	530	682
Cas12L_55_77738117	330	527	679
Cas12L_56_65286425	282	476	628
Cas12L_57_65567118_部分	31	225	377
Cas12L_58_66287853_部分		181	333

图3B

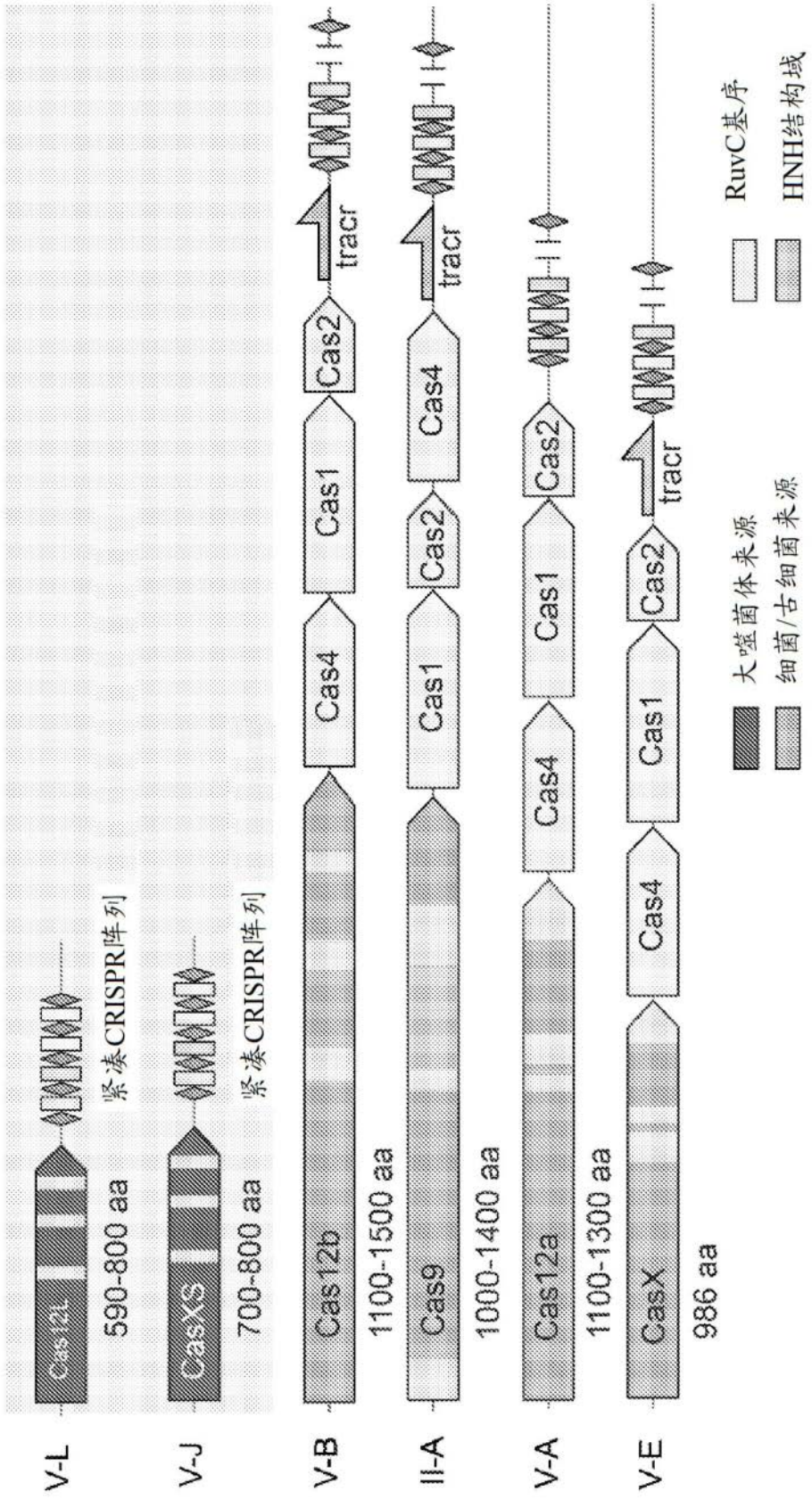

V-L
Cas12L
590-800 aa
紧凑CRISPR阵列
V-J
CasXS
700-800 aa
紧凑CRISPR阵列
V-B
Cas12b
1100-1500 aa
Cas4
Cas1
Cas2
tracr
II-A
Cas9
1000-1400 aa
Cas1
Cas2
Cas4
tracr
V-A
Cas12a
1100-1300 aa
Cas4
Cas1
Cas2
V-E
CasX
986 aa
Cas4
Cas1
Cas2
tracr
大噬菌体来源
细菌/古细菌来源
RuvC基序
HNH结构域

图4

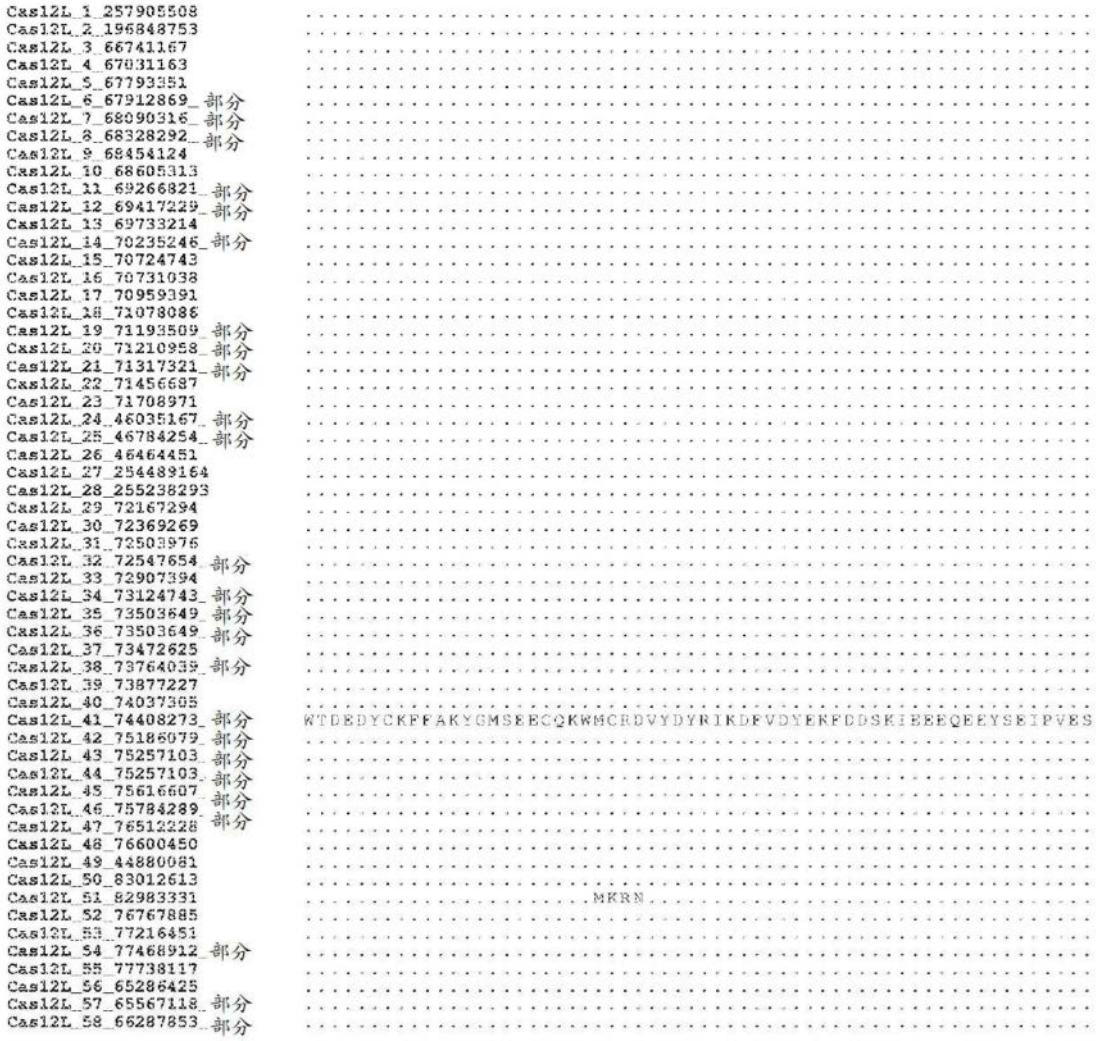

Cas12L_1_257905508
Cas12L_2_196848753
Cas12L_3_66741167
Cas12L_4_67031163
Cas12L_5_67793351
Cas12L_6_67912869_部分
Cas12L_7_68090316_部分
Cas12L_8_68328292_部分
Cas12L_9_68454124
Cas12L_10_68605313
Cas12L_11_69266821_部分
Cas12L_12_69417229_部分
Cas12L_13_69733214
Cas12L_14_70235246_部分
Cas12L_15_70724743
Cas12L_16_70731038
Cas12L_17_70959391
Cas12L_18_71078086
Cas12L_19_71193509_部分
Cas12L_20_71210958_部分
Cas12L_21_71317321_部分
Cas12L_22_71456687
Cas12L_23_71708971
Cas12L_24_46035167_部分
Cas12L_25_46784254_部分
Cas12L_26_46464451
Cas12L_27_254489164
Cas12L_28_255238293
Cas12L_29_72167294
Cas12L_30_72369269
Cas12L_31_72503976
Cas12L_32_72547654_部分
Cas12L_33_72907394
Cas12L_34_73124743_部分
Cas12L_35_73503649_部分
Cas12L_36_73503649_部分
Cas12L_37_73472625
Cas12L_38_73764039_部分
Cas12L_39_73877227
Cas12L_40_74037305
Cas12L_41_74408273_部分
Cas12L_42_75186079_部分
Cas12L_43_75257103_部分
Cas12L_44_75257103_部分
Cas12L_45_75616607_部分
Cas12L_46_75784289_部分
Cas12L_47_76512228
Cas12L_48_76600450
Cas12L_49_44880081
Cas12L_50_83012613
Cas12L_51_82983331
Cas12L_52_76767885
Cas12L_53_77216451
Cas12L_54_77468912_部分
Cas12L_55_77738117
Cas12L_56_65286425
Cas12L_57_65567118_部分
Cas12L_58_66287853_部分
MKRN
WTDEDYCKPFAKYGMSEECQKWMCRDVYDYRIKDFVDYEKFDDSKIEEEQEEYSEIPVES

图5A

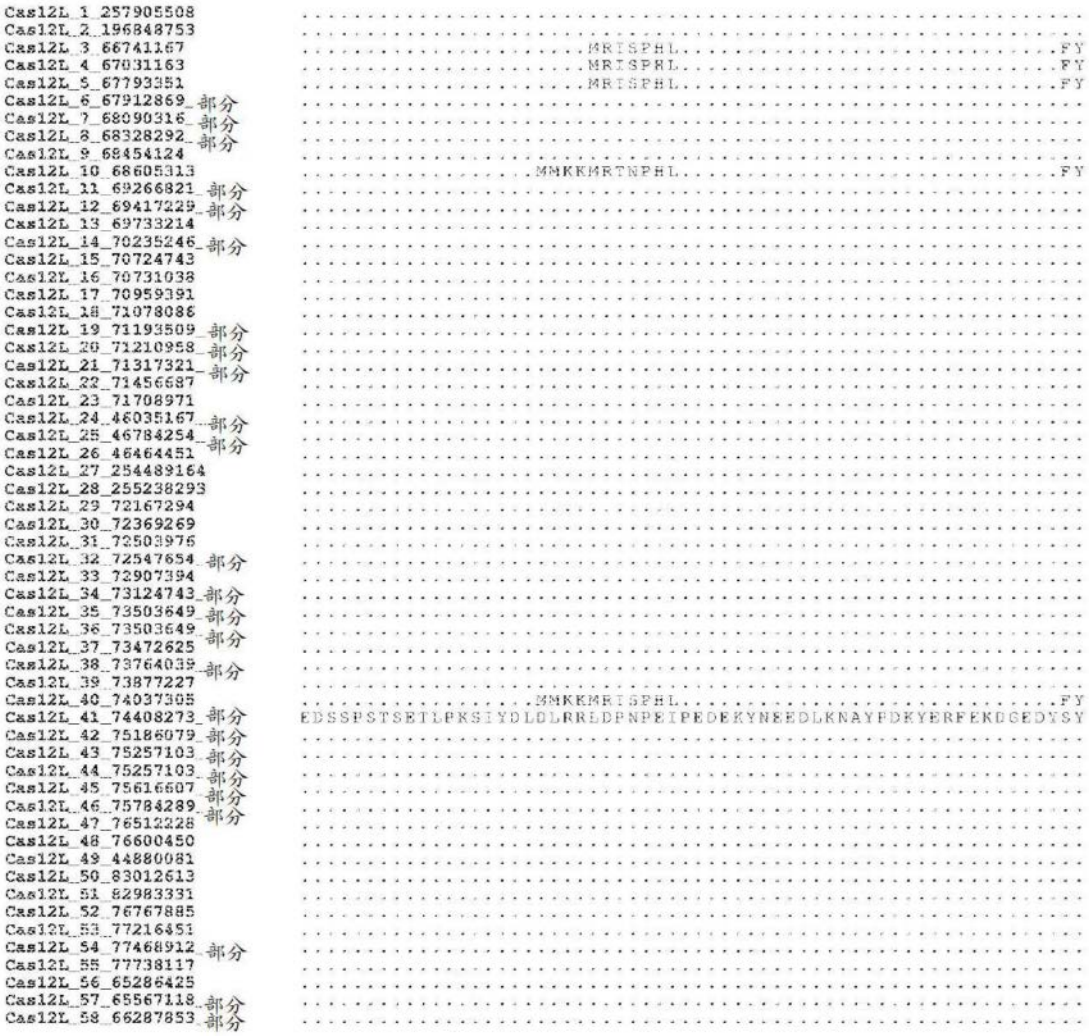

Cas12L_1_257905508
Cas12L_2_196848753
Cas12L_3_66741167
Cas12L_4_67031163
Cas12L_5_67793351
Cas12L_6_67912869_部分
Cas12L_7_68090316_部分
Cas12L_8_68328292_部分
Cas12L_9_68454124
Cas12L_10_68605313
Cas12L_11_69266821_部分
Cas12L_12_69417229_部分
Cas12L_13_69733214
Cas12L_14_70235246_部分
Cas12L_15_70724743
Cas12L_16_70731038
Cas12L_17_70959391
Cas12L_18_71078086
Cas12L_19_71193509_部分
Cas12L_20_71210958_部分
Cas12L_21_71317321_部分
Cas12L_22_71456687
Cas12L_23_71708971
Cas12L_24_46035167_部分
Cas12L_25_46784254_部分
Cas12L_26_46464451
Cas12L_27_254489164
Cas12L_28_255238293
Cas12L_29_72167294
Cas12L_30_72369269
Cas12L_31_72503976
Cas12L_32_72547654_部分
Cas12L_33_72907394
Cas12L_34_73124743_部分
Cas12L_35_73503649_部分
Cas12L_36_73503649_部分
Cas12L_37_73472625_部分
Cas12L_38_73764039_部分
Cas12L_39_73877227
Cas12L_40_74037305
Cas12L_41_74408273_部分
Cas12L_42_75186079_部分
Cas12L_43_75257103_部分
Cas12L_44_75257103_部分
Cas12L_45_75616607_部分
Cas12L_46_75784289_部分
Cas12L_47_76512228
Cas12L_48_76600450
Cas12L_49_44880081
Cas12L_50_83012613
Cas12L_51_82983331
Cas12L_52_76767885
Cas12L_53_77216451
Cas12L_54_77468912_部分
Cas12L_55_77738117
Cas12L_56_65286425
Cas12L_57_65567118_部分
Cas12L_58_66287853_部分

图5B

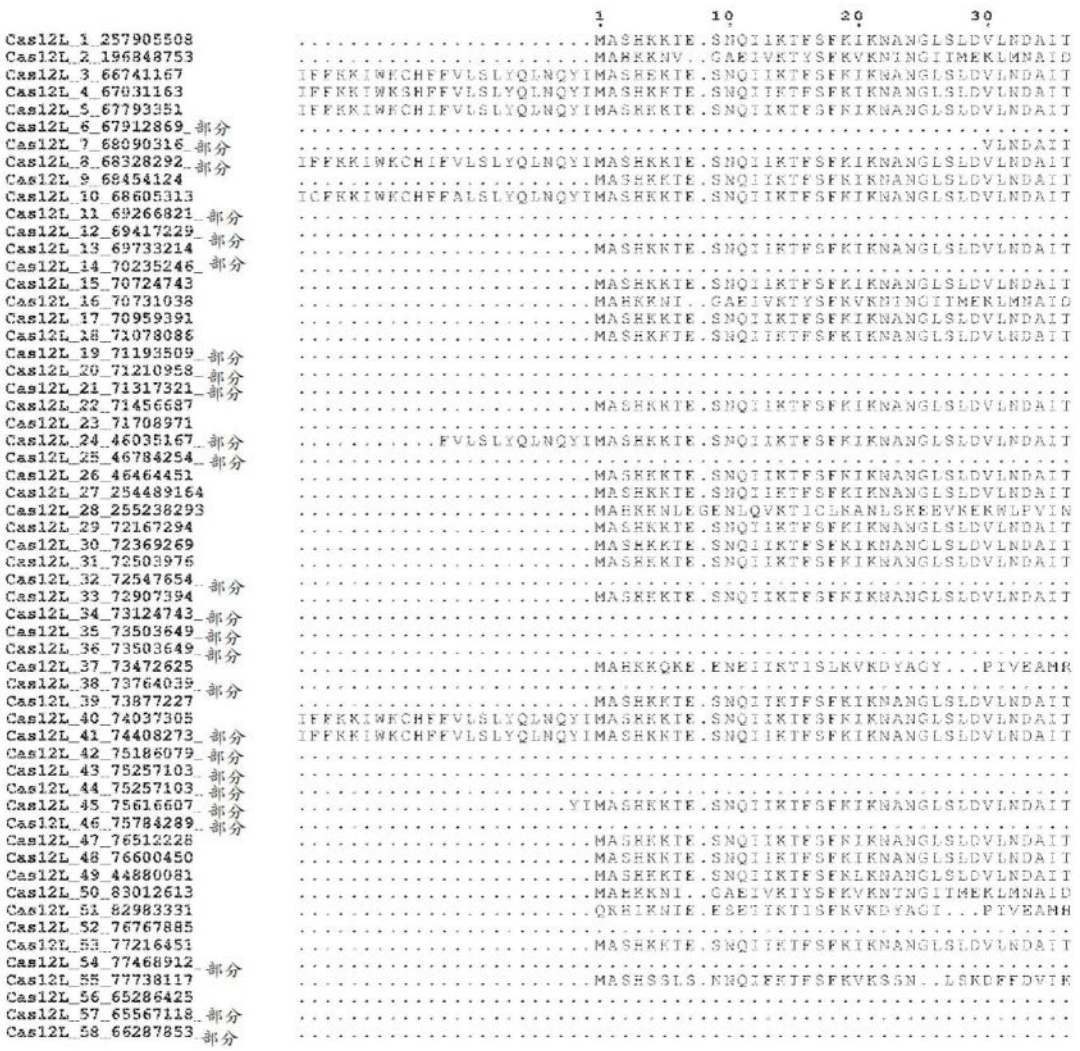

1 10 20 30
Cas12L_1_257905508
Cas12L_2_196848753
Cas12L_3_66741167
Cas12L_4_67031163
Cas12L_5_67793351
Cas12L_6_67912869_部分
Cas12L_7_68090316_部分
Cas12L_8_68328292_部分
Cas12L_9_68454124
Cas12L_10_68605313
Cas12L_11_69266821_部分
Cas12L_12_69417229_部分
Cas12L_13_69733214
Cas12L_14_70235246_部分
Cas12L_15_70724743
Cas12L_16_70731038
Cas12L_17_70959391
Cas12L_18_71078086
Cas12L_19_71193509_部分
Cas12L_20_71210958_部分
Cas12L_21_71317321_部分
Cas12L_22_71456687
Cas12L_23_71708971
Cas12L_24_46035167_部分
Cas12L_25_46784254_部分
Cas12L_26_46464451
Cas12L_27_254489164
Cas12L_28_255238293
Cas12L_29_72167294
Cas12L_30_72369269
Cas12L_31_72503976
Cas12L_32_72547654_部分
Cas12L_33_72907394
Cas12L_34_73124743_部分
Cas12L_35_73503649_部分
Cas12L_36_73503649_部分
Cas12L_37_73472625
Cas12L_38_73764039_部分
Cas12L_39_73877227
Cas12L_40_74037305
Cas12L_41_74408273_部分
Cas12L_42_75186079_部分
Cas12L_43_75257103_部分
Cas12L_44_75257103_部分
Cas12L_45_75616607_部分
Cas12L_46_75784289_部分
Cas12L_47_76512228
Cas12L_48_76600450
Cas12L_49_44880081
Cas12L_50_83012613
Cas12L_51_82983331
Cas12L_52_76767885
Cas12L_53_77216451
Cas12L_54_77468912_部分
Cas12L_55_77738117
Cas12L_56_65286425
Cas12L_57_65567118_部分
Cas12L_58_66287853_部分

图5C

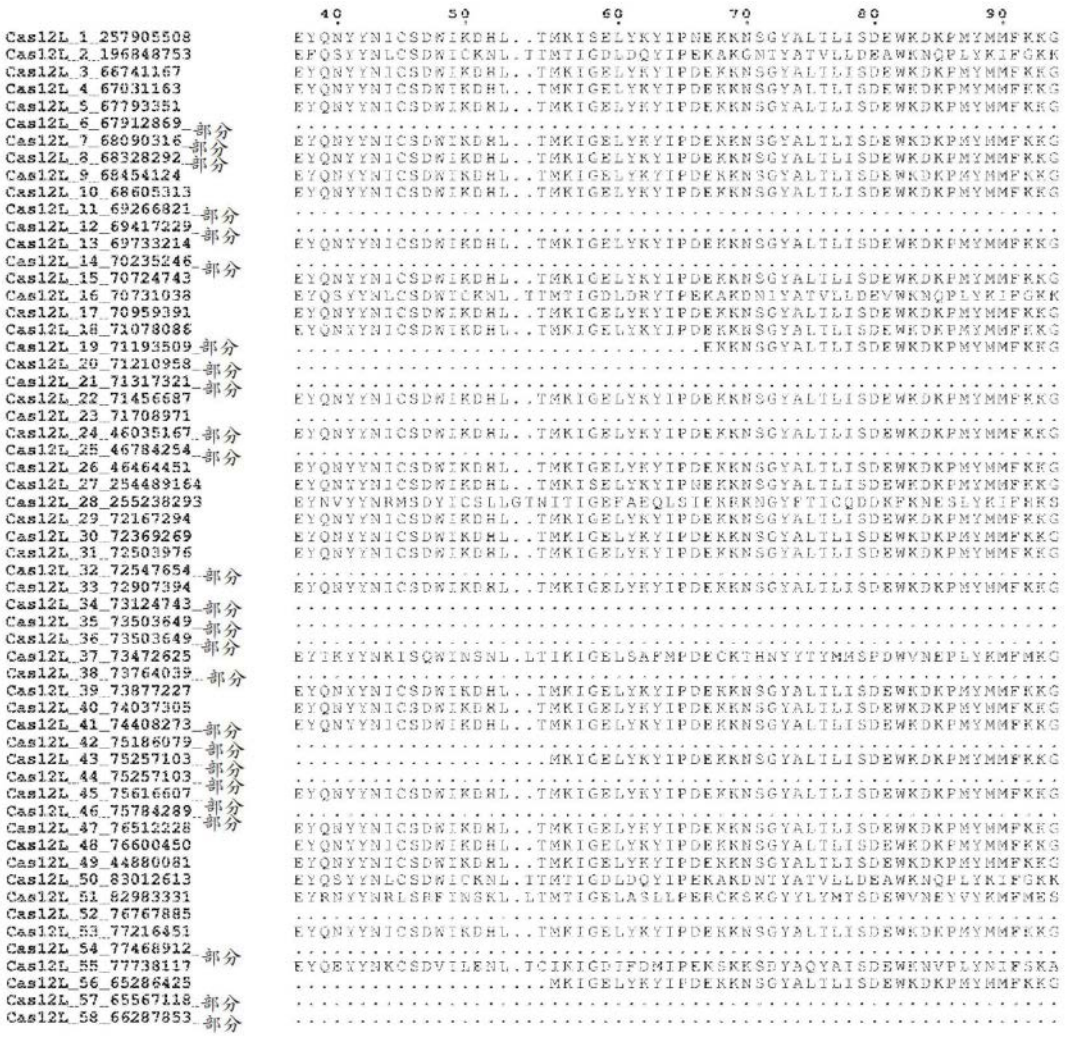


图5D

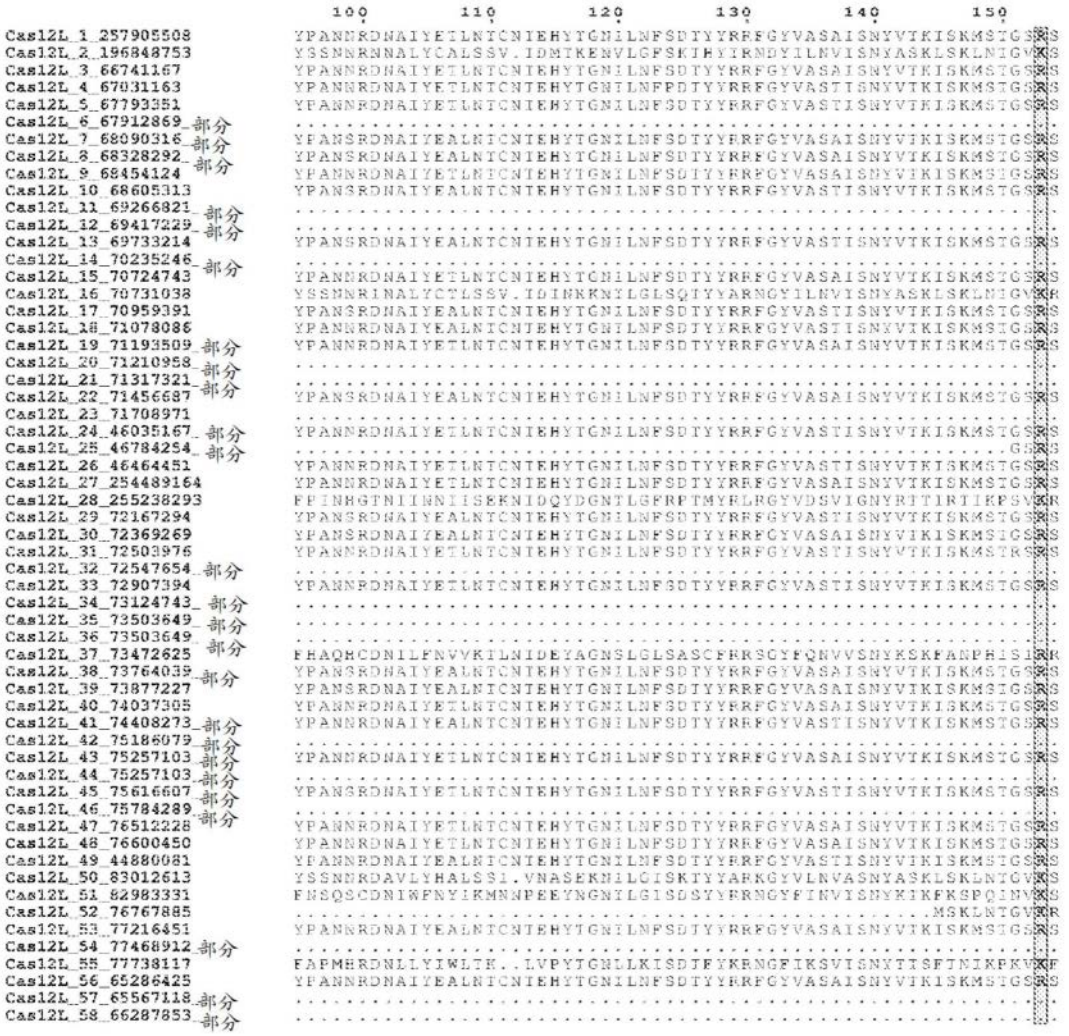


图5E

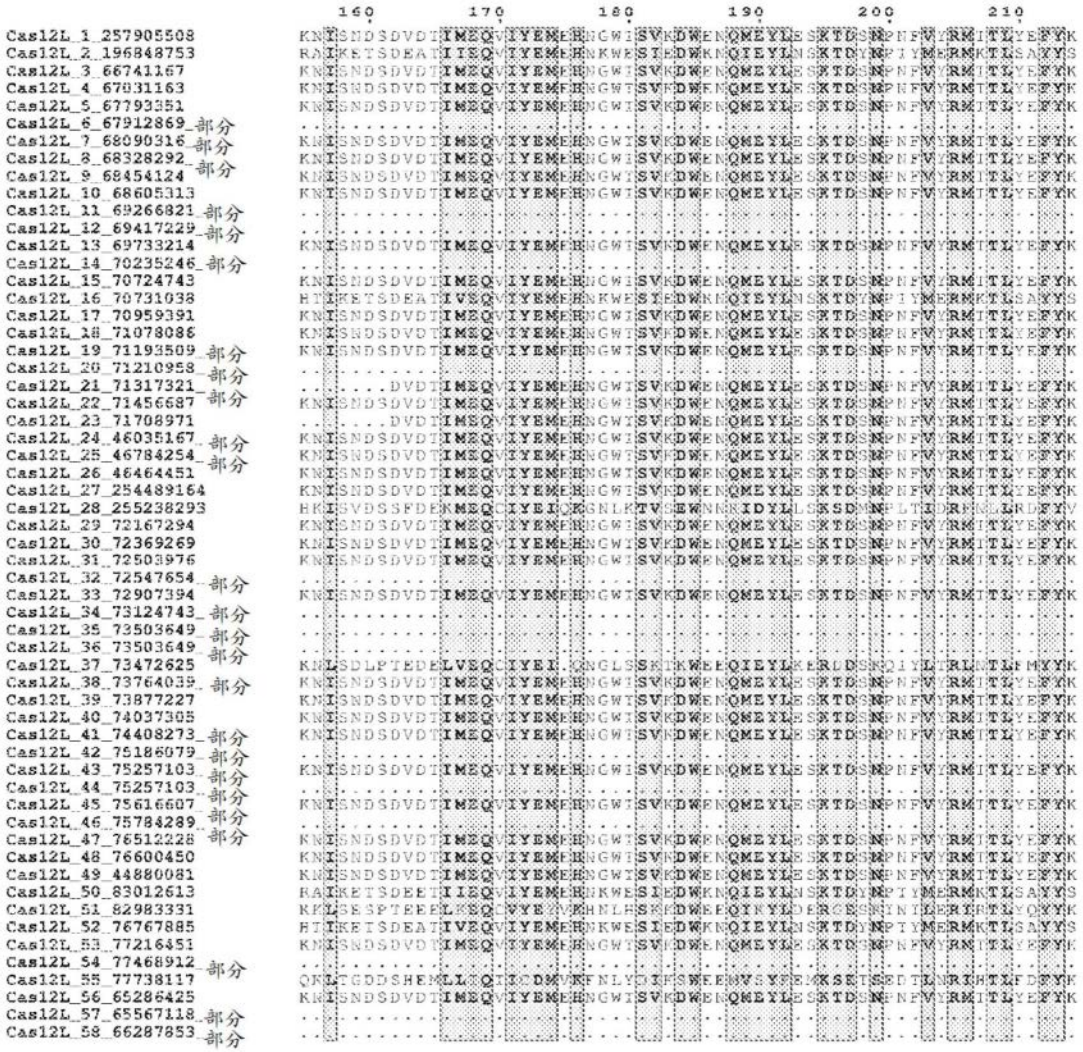

160
170
180
190
200
210
Cas12L_1_257905508
Cas12L_2_196848753
Cas12L_3_66741167
Cas12L_4_67831163
Cas12L_5_67793351
Cas12L_6_67912869_部分
Cas12L_7_68090316_部分
Cas12L_8_68328292_部分
Cas12L_9_68454124
Cas12L_10_68605313
Cas12L_11_69266821_部分
Cas12L_12_69417229_部分
Cas12L_13_69733214
Cas12L_14_70235246_部分
Cas12L_15_70724743
Cas12L_16_70731038
Cas12L_17_70959391
Cas12L_18_71078086
Cas12L_19_71193509_部分
Cas12L_20_71210958_部分
Cas12L_21_71317321_部分
Cas12L_22_71456687
Cas12L_23_71708971
Cas12L_24_46035167_部分
Cas12L_25_46784254_部分
Cas12L_26_46464451
Cas12L_27_254489164
Cas12L_28_255238293
Cas12L_29_72167294
Cas12L_30_72369269
Cas12L_31_72503976
Cas12L_32_72547654_部分
Cas12L_33_72907394
Cas12L_34_73124743_部分
Cas12L_35_73503649_部分
Cas12L_36_73503649_部分
Cas12L_37_73472625
Cas12L_38_73764039_部分
Cas12L_39_73877227
Cas12L_40_74037305
Cas12L_41_74408273_部分
Cas12L_42_75186079_部分
Cas12L_43_75257103_部分
Cas12L_44_75257103_部分
Cas12L_45_75616607_部分
Cas12L_46_75784289_部分
Cas12L_47_76512228_部分
Cas12L_48_76600450
Cas12L_49_44880081
Cas12L_50_83012613
Cas12L_51_82983331
Cas12L_52_76767885
Cas12L_53_77216451
Cas12L_54_77468912_部分
Cas12L_55_77738117
Cas12L_56_65286425
Cas12L_57_65567118_部分
Cas12L_58_66287853_部分

图5F

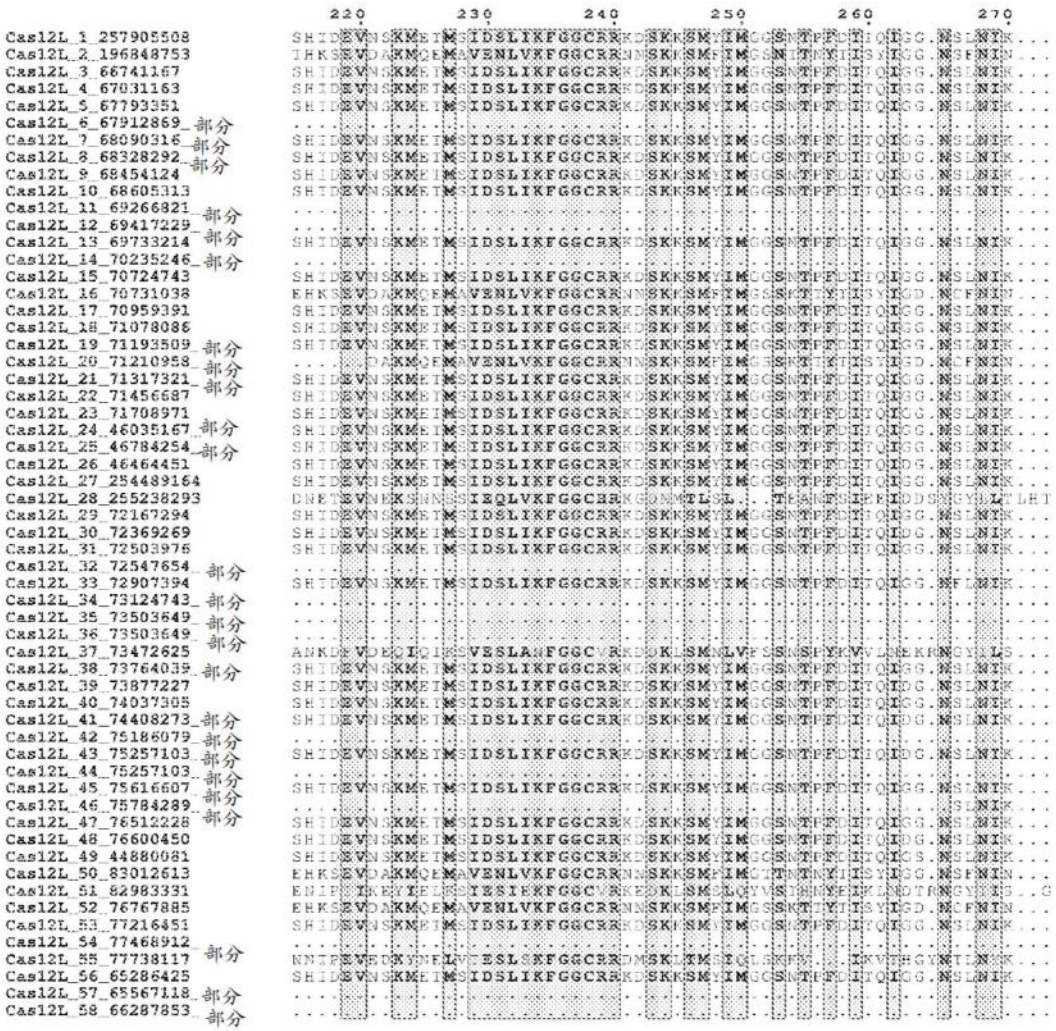

220
230
240
250
260
270
Cas12L_1_257905508
Cas12L_2_196848753
Cas12L_3_66741167
Cas12L_4_67031163
Cas12L_5_67793351
Cas12L_6_67912869_部分
Cas12L_7_68090316_部分
Cas12L_8_68328292_部分
Cas12L_9_68454124
Cas12L_10_68605313
Cas12L_11_69266821_部分
Cas12L_12_69417229_部分
Cas12L_13_69733214_部分
Cas12L_14_70235246_部分
Cas12L_15_70724743
Cas12L_16_70731038
Cas12L_17_70959391
Cas12L_18_71078086
Cas12L_19_71193509_部分
Cas12L_20_71210958_部分
Cas12L_21_71317321_部分
Cas12L_22_71456687
Cas12L_23_71708971
Cas12L_24_46035167_部分
Cas12L_25_46784254_部分
Cas12L_26_46464451
Cas12L_27_254489164
Cas12L_28_255238293
Cas12L_29_72167294
Cas12L_30_72369269
Cas12L_31_72503976
Cas12L_32_72547654_部分
Cas12L_33_72907394
Cas12L_34_73124743_部分
Cas12L_35_73503649_部分
Cas12L_36_73503649_部分
Cas12L_37_73472625
Cas12L_38_73764039_部分
Cas12L_39_73877227
Cas12L_40_74037305
Cas12L_41_74408273_部分
Cas12L_42_75186079_部分
Cas12L_43_75257103_部分
Cas12L_44_75257103_部分
Cas12L_45_75616607_部分
Cas12L_46_75784289_部分
Cas12L_47_76512228
Cas12L_48_76600450
Cas12L_49_44880081
Cas12L_50_83012613
Cas12L_51_82983331
Cas12L_52_76767885
Cas12L_53_77216451
Cas12L_54_77468912_部分
Cas12L_55_77738117
Cas12L_56_65286425
Cas12L_57_65567118_部分
Cas12L_58_66287853_部分

图5G

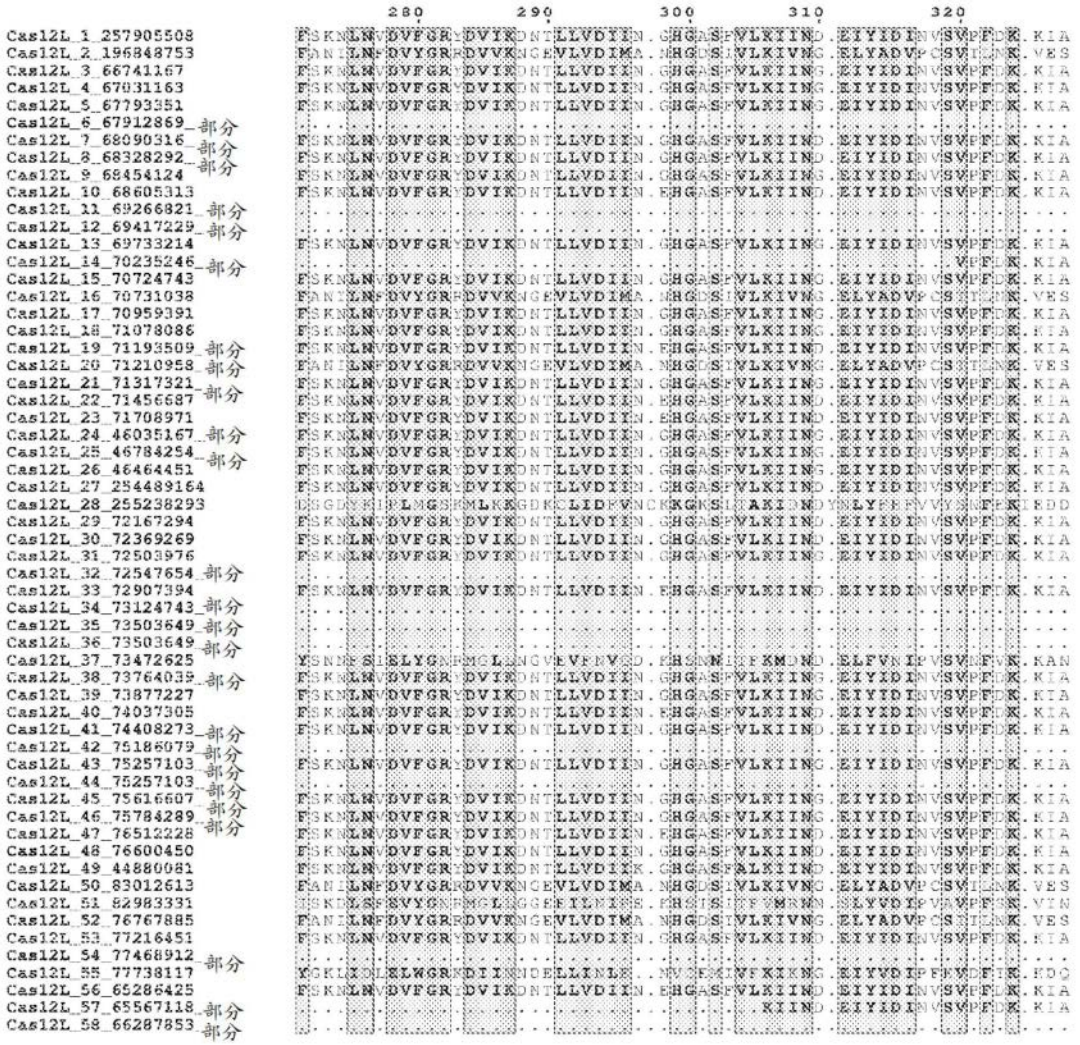

280
290
300
310
320
Cas12L_1_257905508
Cas12L_2_196848753
Cas12L_3_66741167
Cas12L_4_67031163
Cas12L_5_67793351
Cas12L_6_67912869_部分
Cas12L_7_68090316_部分
Cas12L_8_68328292_部分
Cas12L_9_68454124
Cas12L_10_68605313
Cas12L_11_69266821_部分
Cas12L_12_69417229_部分
Cas12L_13_69733214
Cas12L_14_70235246_部分
Cas12L_15_70724743
Cas12L_16_70731038
Cas12L_17_70959391
Cas12L_18_71078086
Cas12L_19_71193509_部分
Cas12L_20_71210958_部分
Cas12L_21_71317321_部分
Cas12L_22_71456687_部分
Cas12L_23_71708971
Cas12L_24_46035167_部分
Cas12L_25_46784254_部分
Cas12L_26_46464451
Cas12L_27_254489164
Cas12L_28_255238293
Cas12L_29_72167294
Cas12L_30_72369269
Cas12L_31_72503976
Cas12L_32_72547654_部分
Cas12L_33_72907394
Cas12L_34_73124743_部分
Cas12L_35_73503649_部分
Cas12L_36_73503649_部分
Cas12L_37_73472625
Cas12L_38_73764039_部分
Cas12L_39_73877227
Cas12L_40_74037305
Cas12L_41_74408273_部分
Cas12L_42_75186079_部分
Cas12L_43_75257103_部分
Cas12L_44_75257103_部分
Cas12L_45_75616607_部分
Cas12L_46_75784289_部分
Cas12L_47_76512228
Cas12L_48_76600450
Cas12L_49_44880081
Cas12L_50_83012613
Cas12L_51_82983331
Cas12L_52_76767885
Cas12L_53_77216451
Cas12L_54_77468912_部分
Cas12L_55_77738117
Cas12L_56_65286425
Cas12L_57_65567118_部分
Cas12L_58_66287853_部分

图5H

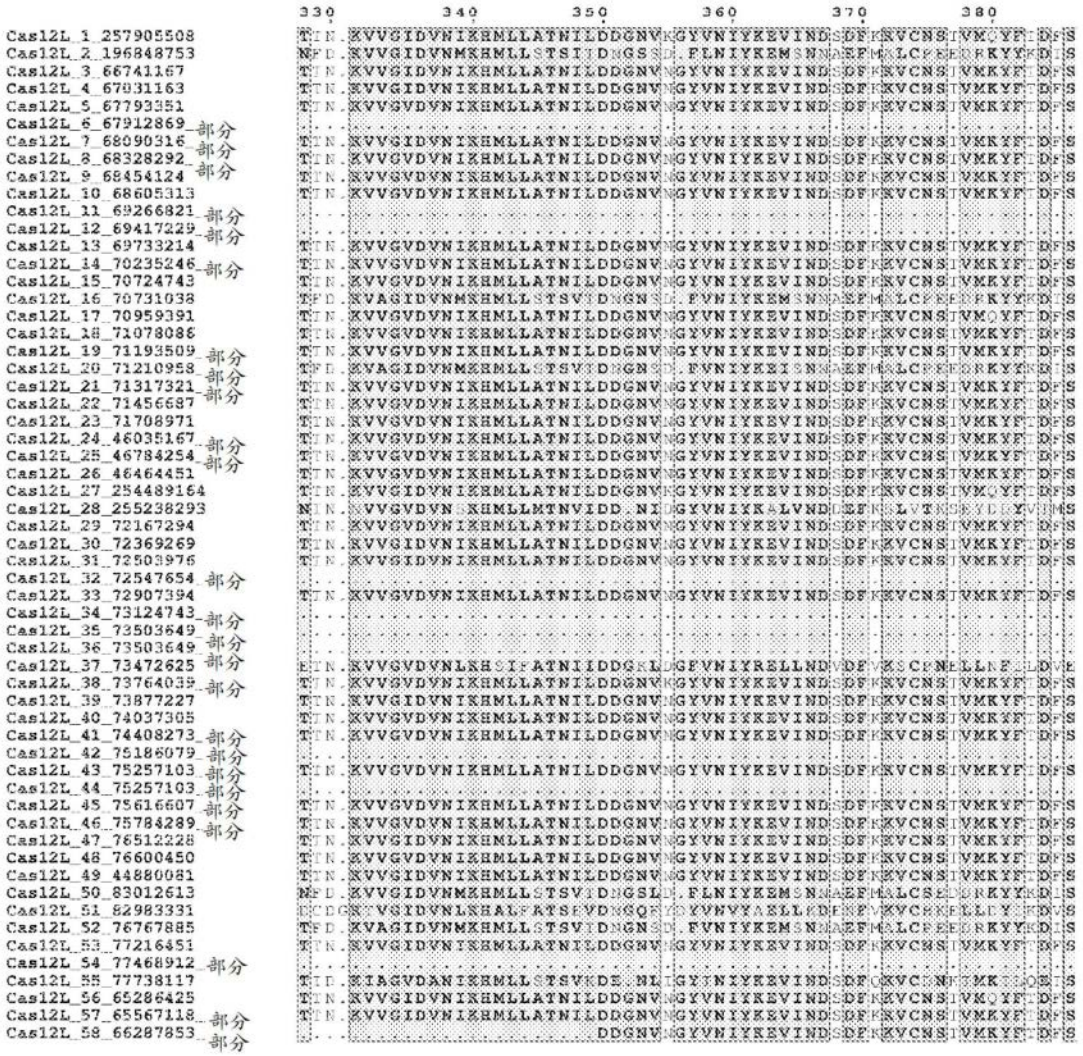

330
340
350
360
370
380
Cas12L_1_257905508
Cas12L_2_196848753
Cas12L_3_66741167
Cas12L_4_67031163
Cas12L_5_67793351
Cas12L_6_67912869_部分
Cas12L_7_68090316_部分
Cas12L_8_68328292_部分
Cas12L_9_68454124
Cas12L_10_68605313
Cas12L_11_69266821_部分
Cas12L_12_69417229_部分
Cas12L_13_69733214
Cas12L_14_70235246_部分
Cas12L_15_70724743
Cas12L_16_70731038
Cas12L_17_70959391
Cas12L_18_71078086
Cas12L_19_71193509_部分
Cas12L_20_71210958_部分
Cas12L_21_71317321_部分
Cas12L_22_71456687
Cas12L_23_71708971
Cas12L_24_46035167_部分
Cas12L_25_46784254_部分
Cas12L_26_46464451
Cas12L_27_254489164
Cas12L_28_255238293
Cas12L_29_72167294
Cas12L_30_72369269
Cas12L_31_72503976
Cas12L_32_72547654_部分
Cas12L_33_72907394
Cas12L_34_73124743_部分
Cas12L_35_73503649_部分
Cas12L_36_73503649_部分
Cas12L_37_73472625_部分
Cas12L_38_73764039_部分
Cas12L_39_73877227
Cas12L_40_74037305
Cas12L_41_74408273_部分
Cas12L_42_75186079_部分
Cas12L_43_75257103_部分
Cas12L_44_75257103_部分
Cas12L_45_75616607_部分
Cas12L_46_75784289_部分
Cas12L_47_76512228
Cas12L_48_76600450
Cas12L_49_44880081
Cas12L_50_83012613
Cas12L_51_82983331
Cas12L_52_76767885
Cas12L_53_77216451
Cas12L_54_77468912_部分
Cas12L_55_77738117
Cas12L_56_65286425
Cas12L_57_65567118_部分
Cas12L_58_66287853_部分

图5I

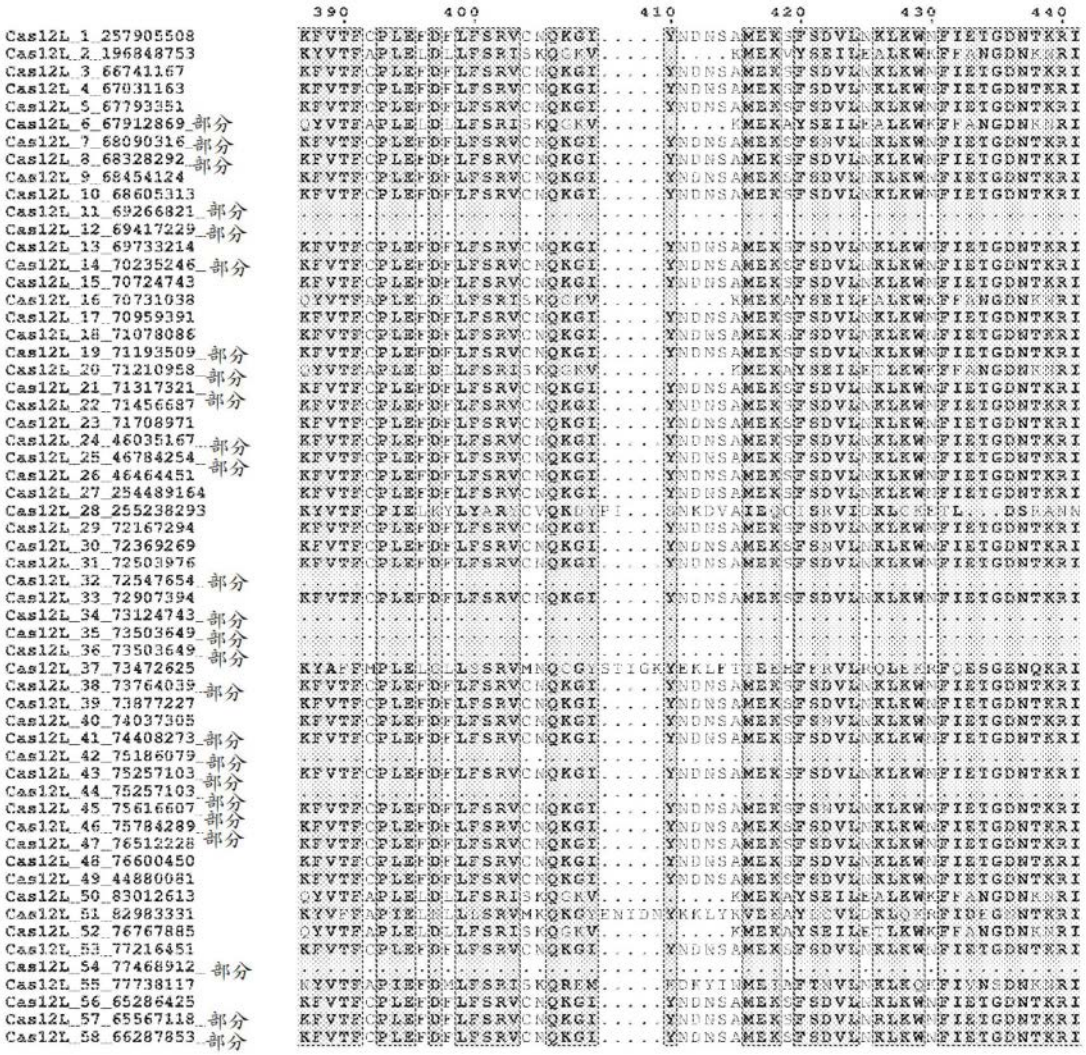

390
400
410
420
430
440
Cas12L_1_257905508
Cas12L_2_196848753
Cas12L_3_66741167
Cas12L_4_67031163
Cas12L_5_67793351
Cas12L_6_67912869_部分
Cas12L_7_68090316_部分
Cas12L_8_68328292_部分
Cas12L_9_68454124
Cas12L_10_68605313
Cas12L_11_69266821_部分
Cas12L_12_69417229_部分
Cas12L_13_69733214
Cas12L_14_70235246_部分
Cas12L_15_70724743
Cas12L_16_70731038
Cas12L_17_70959391
Cas12L_18_71078086
Cas12L_19_71193509_部分
Cas12L_20_71210958_部分
Cas12L_21_71317321_部分
Cas12L_22_71456687
Cas12L_23_71708971
Cas12L_24_46035167_部分
Cas12L_25_46784254_部分
Cas12L_26_46464451
Cas12L_27_254489164
Cas12L_28_255238293
Cas12L_29_72167294
Cas12L_30_72369269
Cas12L_31_72503976
Cas12L_32_72547654_部分
Cas12L_33_72907394
Cas12L_34_73124743_部分
Cas12L_35_73503649_部分
Cas12L_36_73503649_部分
Cas12L_37_73472625
Cas12L_38_73764039_部分
Cas12L_39_73877227
Cas12L_40_74037305
Cas12L_41_74408273_部分
Cas12L_42_75186079_部分
Cas12L_43_75257103_部分
Cas12L_44_75257103_部分
Cas12L_45_75616607_部分
Cas12L_46_75784289_部分
Cas12L_47_76512228
Cas12L_48_76600450
Cas12L_49_44880081
Cas12L_50_83012613
Cas12L_51_82983331
Cas12L_52_76767885
Cas12L_53_77216451
Cas12L_54_77468912_部分
Cas12L_55_77738117
Cas12L_56_65286425
Cas12L_57_65567118_部分
Cas12L_58_66287853_部分

图5J

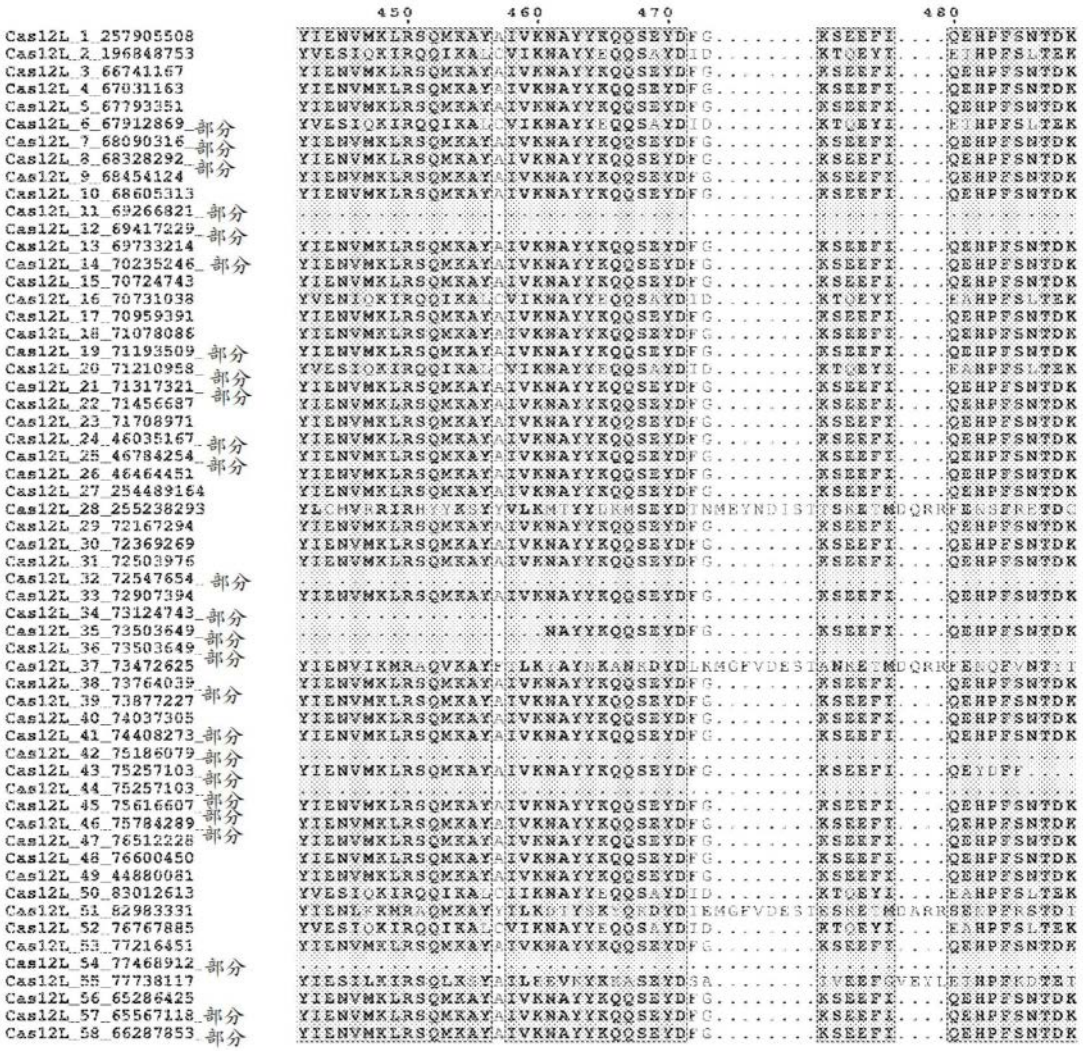


图5K

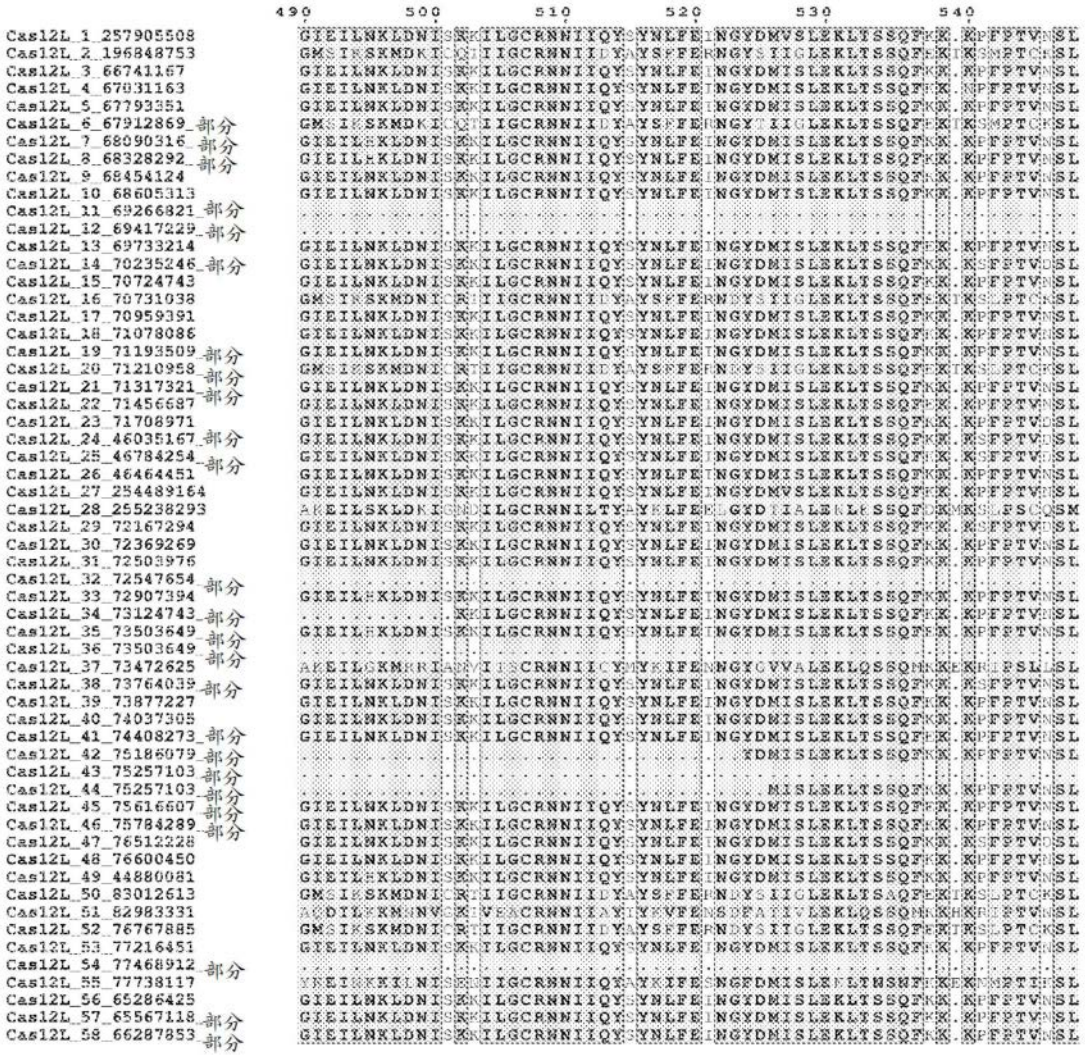


图5L

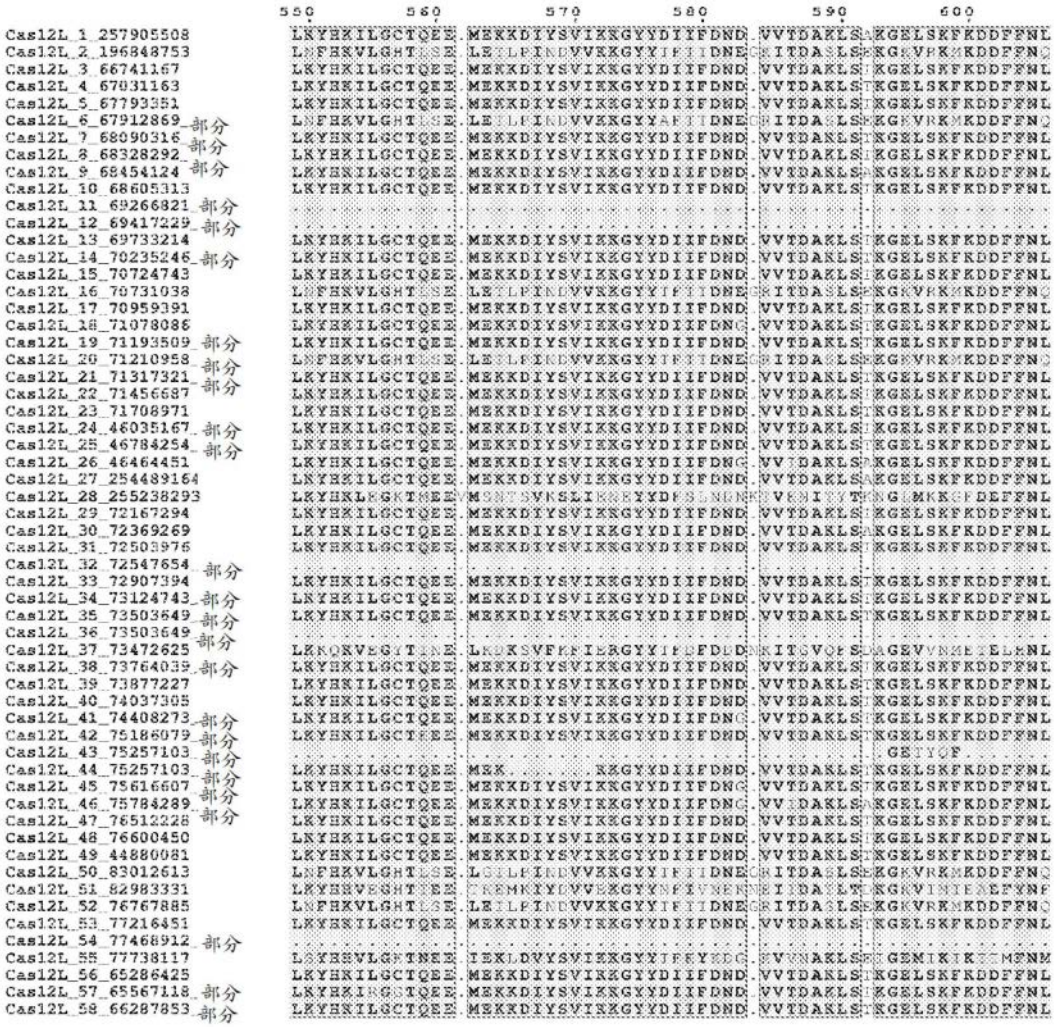


图5M

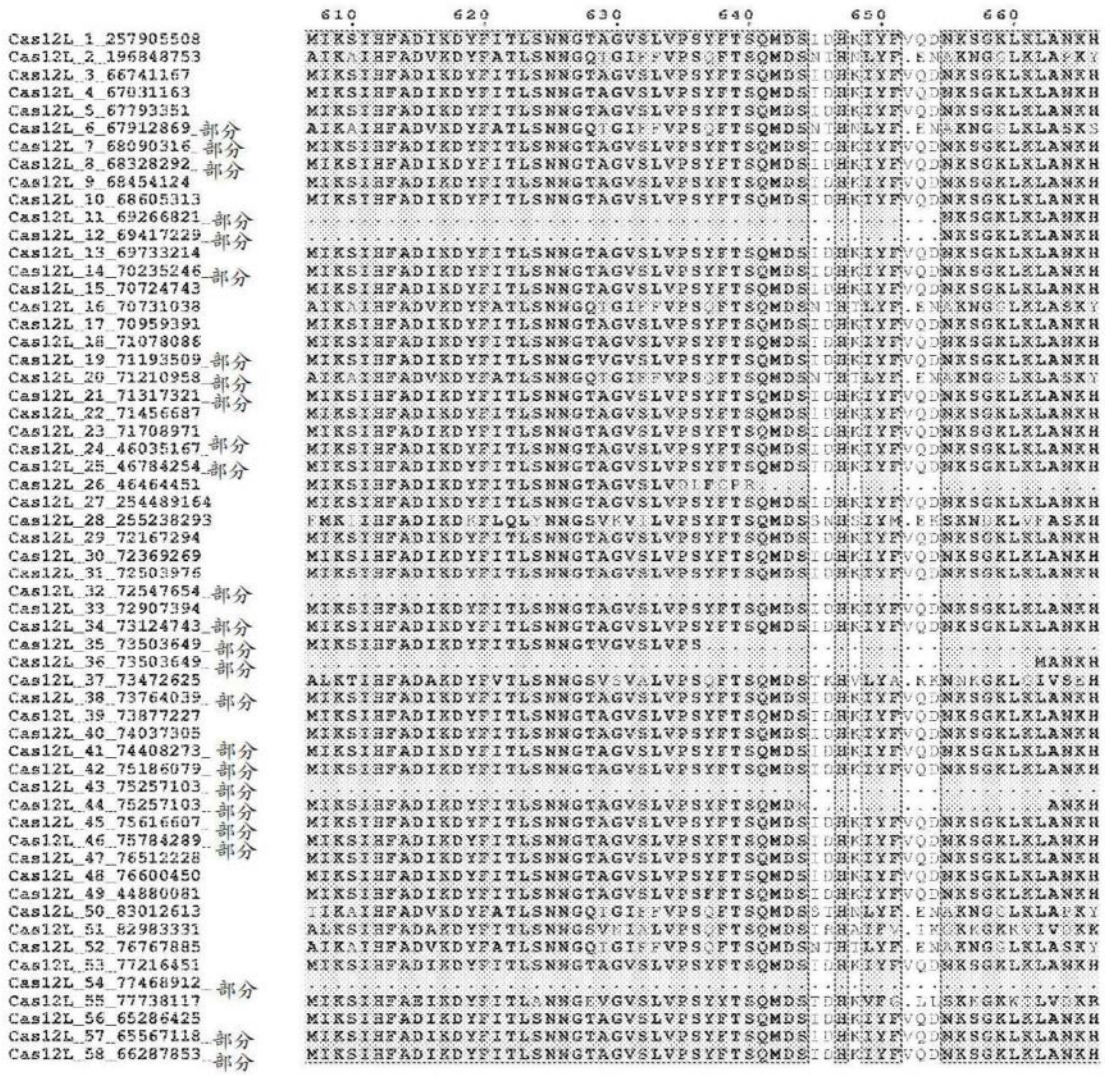

610
620
630
640
650
660
Cas12L_1_257905508
Cas12L_2_196848753
Cas12L_3_66741167
Cas12L_4_67031163
Cas12L_5_67793351
Cas12L_6_67912869_部分
Cas12L_7_68090316_部分
Cas12L_8_68328292_部分
Cas12L_9_68454124
Cas12L_10_68605313
Cas12L_11_69266821_部分
Cas12L_12_69417229_部分
Cas12L_13_69733214
Cas12L_14_70235246_部分
Cas12L_15_70724743
Cas12L_16_70731038
Cas12L_17_70959391
Cas12L_18_71078086
Cas12L_19_71193509_部分
Cas12L_20_71210958_部分
Cas12L_21_71317321_部分
Cas12L_22_71456687
Cas12L_23_71708971
Cas12L_24_46035167_部分
Cas12L_25_46784254_部分
Cas12L_26_46464451
Cas12L_27_254489164
Cas12L_28_255238293
Cas12L_29_72167294
Cas12L_30_72369269
Cas12L_31_72503976
Cas12L_32_72547654_部分
Cas12L_33_72907394
Cas12L_34_73124743_部分
Cas12L_35_73503649_部分
Cas12L_36_73503649_部分
Cas12L_37_73472625
Cas12L_38_73764039_部分
Cas12L_39_73877227
Cas12L_40_74037305
Cas12L_41_74408273_部分
Cas12L_42_75186079_部分
Cas12L_43_75257103_部分
Cas12L_44_75257103_部分
Cas12L_45_75616607_部分
Cas12L_46_75784289_部分
Cas12L_47_76512228
Cas12L_48_76600450
Cas12L_49_44880081
Cas12L_50_83012613
Cas12L_51_82983331
Cas12L_52_76767885
Cas12L_53_77216451
Cas12L_54_77468912_部分
Cas12L_55_77738117
Cas12L_56_65286425
Cas12L_57_65567118_部分
Cas12L_58_66287853_部分

图5N

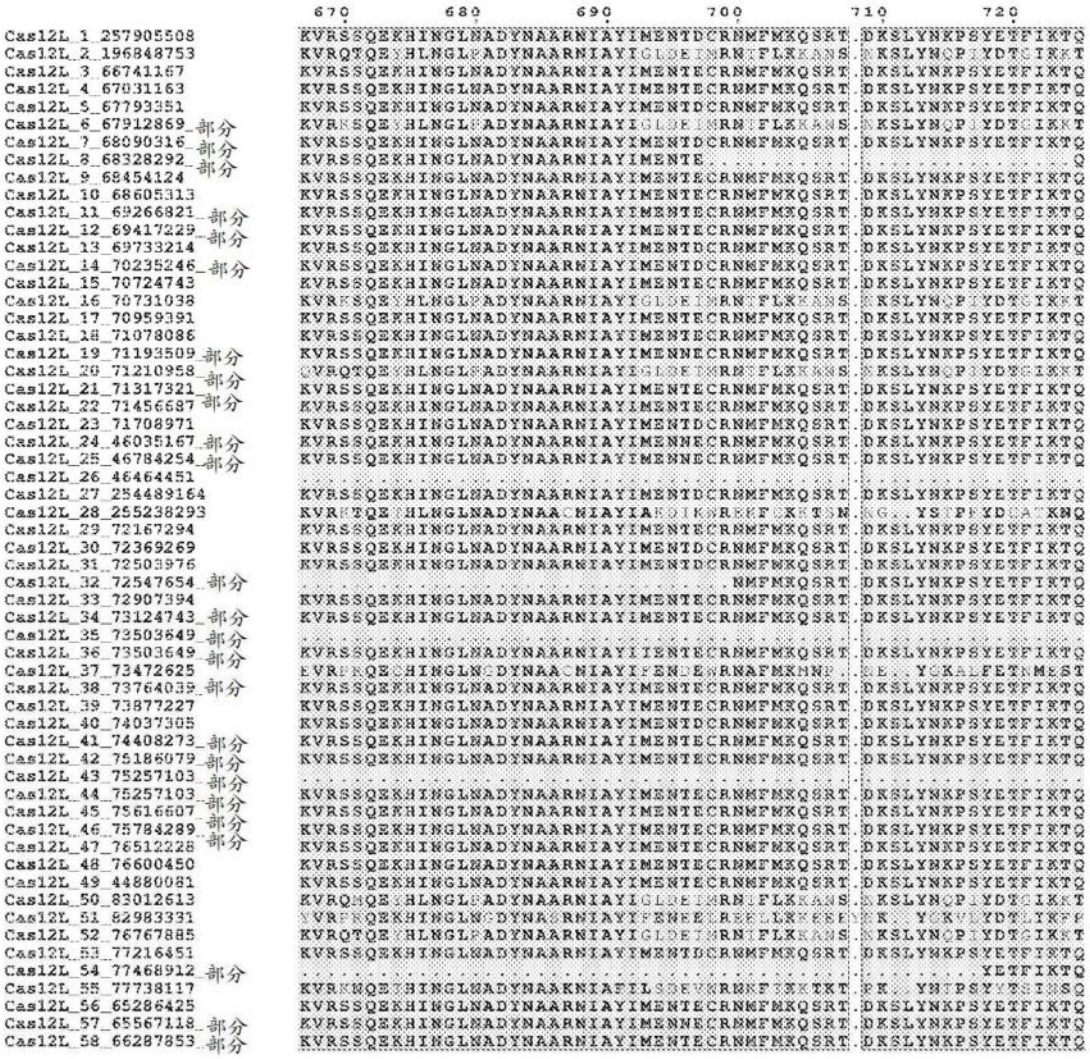


图50

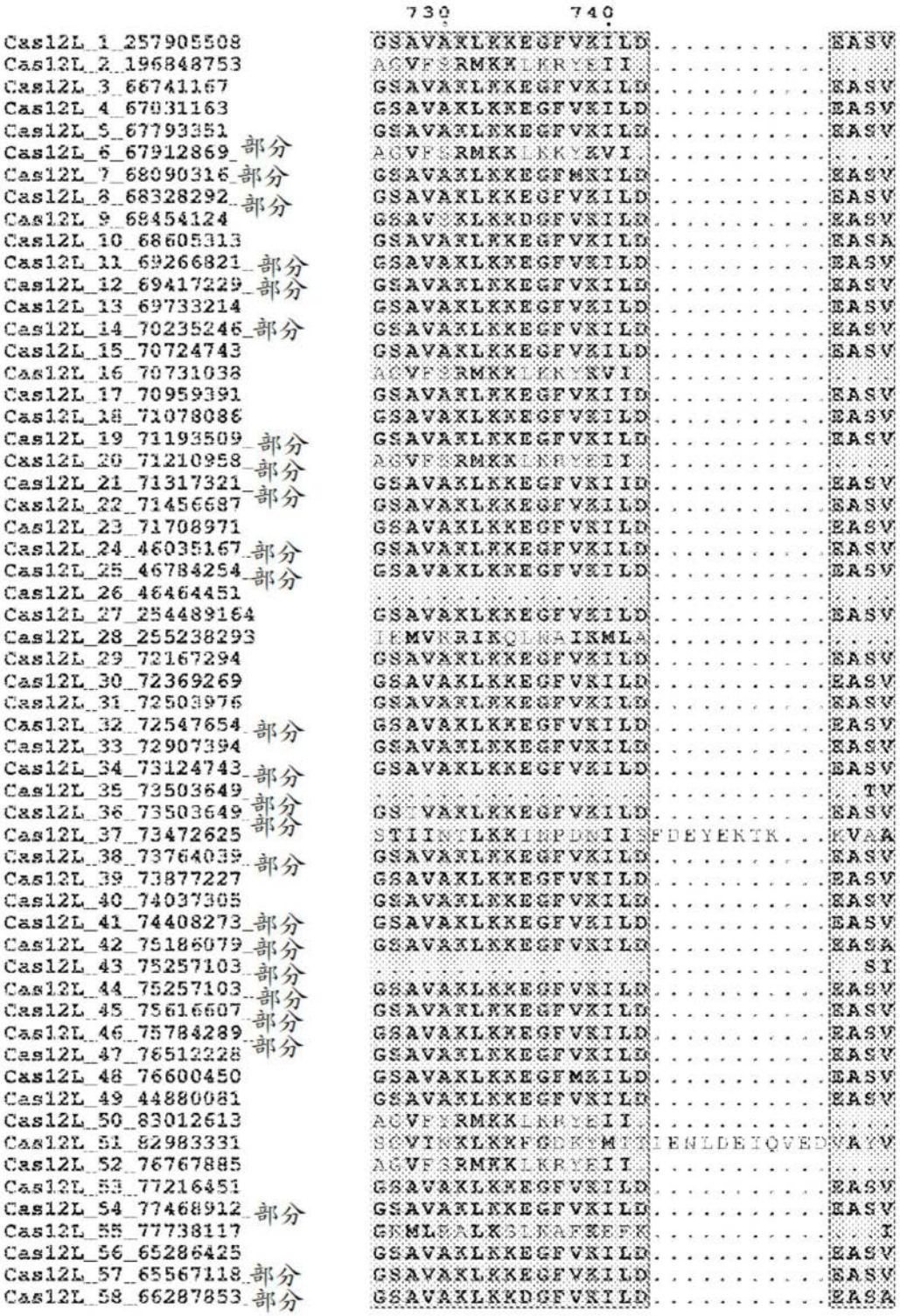

730
740
Cas12L_1_257905508
Cas12L_2_196848753
Cas12L_3_66741167
Cas12L_4_67031163
Cas12L_5_67793351
Cas12L_6_67912869_部分
Cas12L_7_68090316_部分
Cas12L_8_68328292_部分
Cas12L_9_68454124
Cas12L_10_68605313
Cas12L_11_69266821_部分
Cas12L_12_69417229_部分
Cas12L_13_69733214
Cas12L_14_70235246_部分
Cas12L_15_70724743
Cas12L_16_70731038
Cas12L_17_70959391
Cas12L_18_71078086
Cas12L_19_71193509_部分
Cas12L_20_71210958_部分
Cas12L_21_71317321_部分
Cas12L_22_71456687_部分
Cas12L_23_71708971
Cas12L_24_46035167_部分
Cas12L_25_46784254_部分
Cas12L_26_46464451
Cas12L_27_254489164
Cas12L_28_255238293
Cas12L_29_72167294
Cas12L_30_72369269
Cas12L_31_72503976
Cas12L_32_72547654_部分
Cas12L_33_72907394
Cas12L_34_73124743_部分
Cas12L_35_73503649_部分
Cas12L_36_73503649_部分
Cas12L_37_73472625_部分
Cas12L_38_73764039_部分
Cas12L_39_73877227
Cas12L_40_74037305
Cas12L_41_74408273_部分
Cas12L_42_75186079_部分
Cas12L_43_75257103_部分
Cas12L_44_75257103_部分
Cas12L_45_75616607_部分
Cas12L_46_75784289_部分
Cas12L_47_76512228
Cas12L_48_76600450
Cas12L_49_44880081
Cas12L_50_83012613
Cas12L_51_82983331
Cas12L_52_76767885
Cas12L_53_77216451
Cas12L_54_77468912_部分
Cas12L_55_77738117
Cas12L_56_65286425
Cas12L_57_65567118_部分
Cas12L_58_66287853_部分

图5P

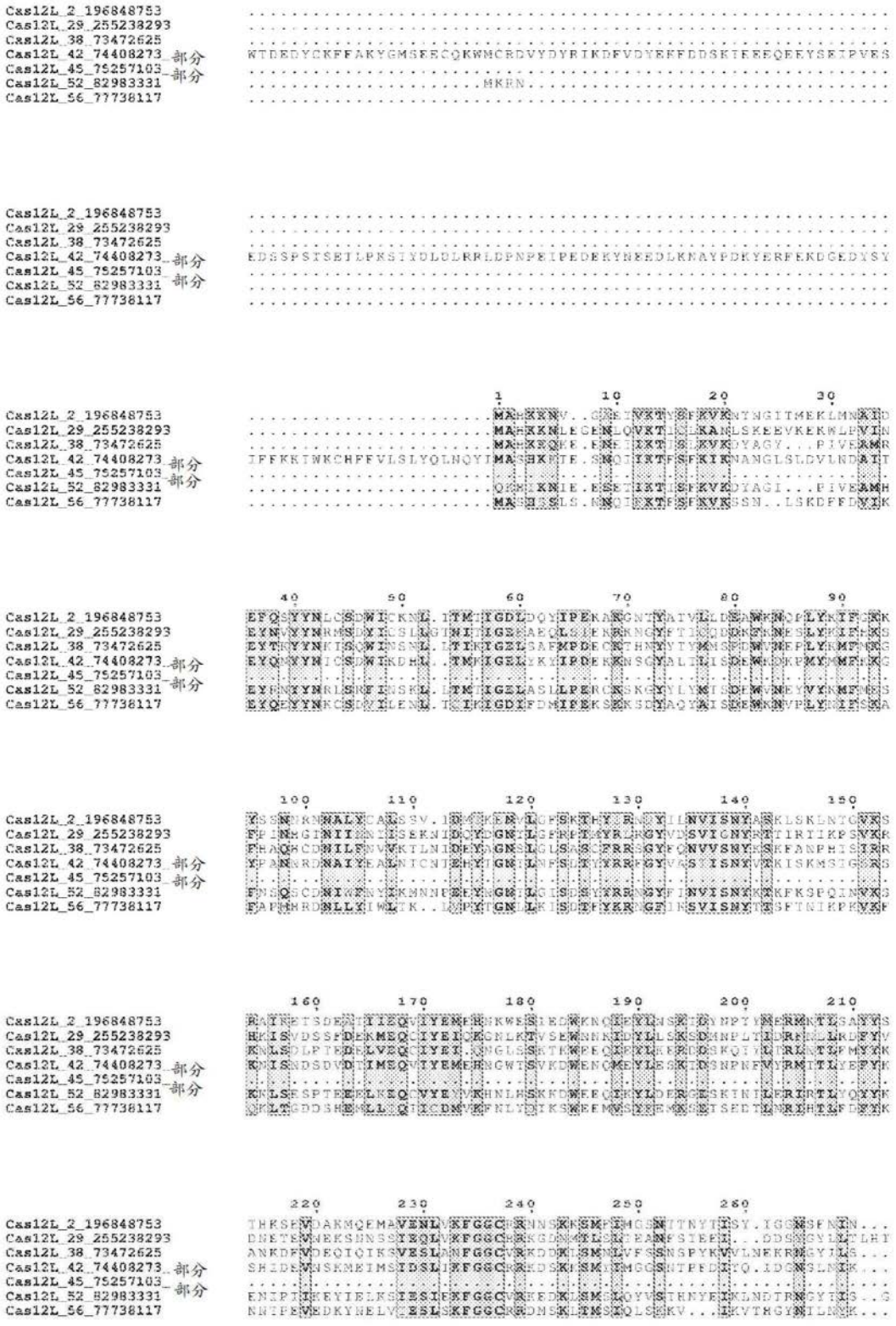

Cas12L_2_196848753
Cas12L_29_255238293
Cas12L_38_73472625
Cas12L_42_74408273 部分
Cas12L_45_75257103 部分
Cas12L_52_82983331
Cas12L_56_77738117

图6A

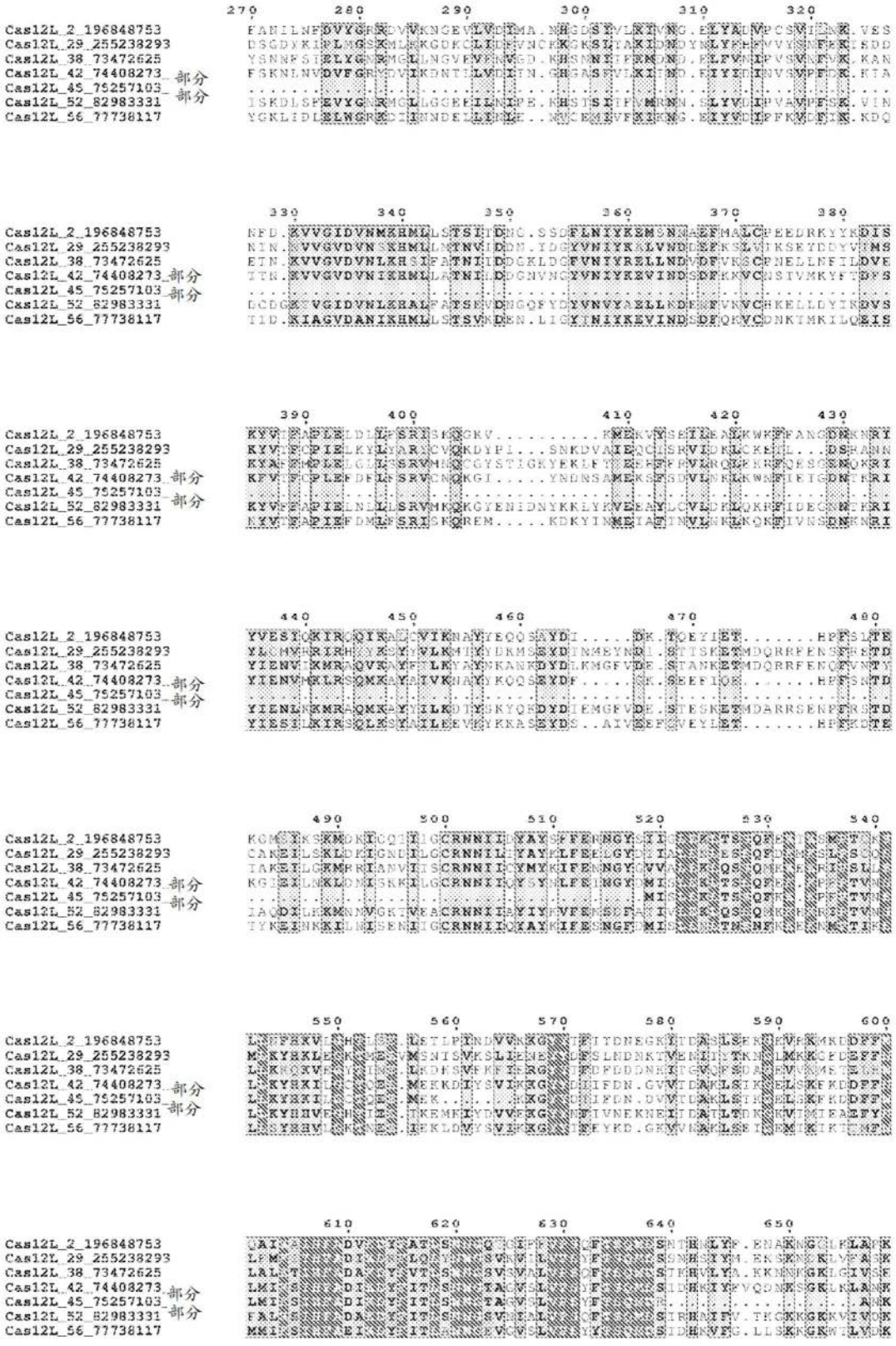

Cas12L_2_196848753
Cas12L_29_255238293
Cas12L_38_73472625
Cas12L_42_74408273_部分
Cas12L_45_75257103_部分
Cas12L_52_82983331_部分
Cas12L_56_77738117

图6B

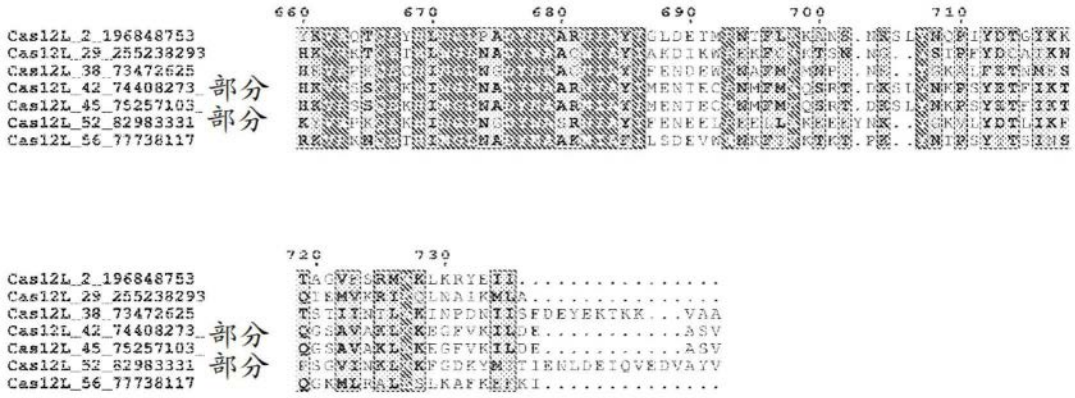

660
670
680
690
700
710
Cas12L_2_196848753
Cas12L_29_255238293
Cas12L_38_73472625
Cas12L_42_74408273_部分
Cas12L_45_75257103_部分
Cas12L_52_82983331
Cas12L_56_77738117
720
730

图6C

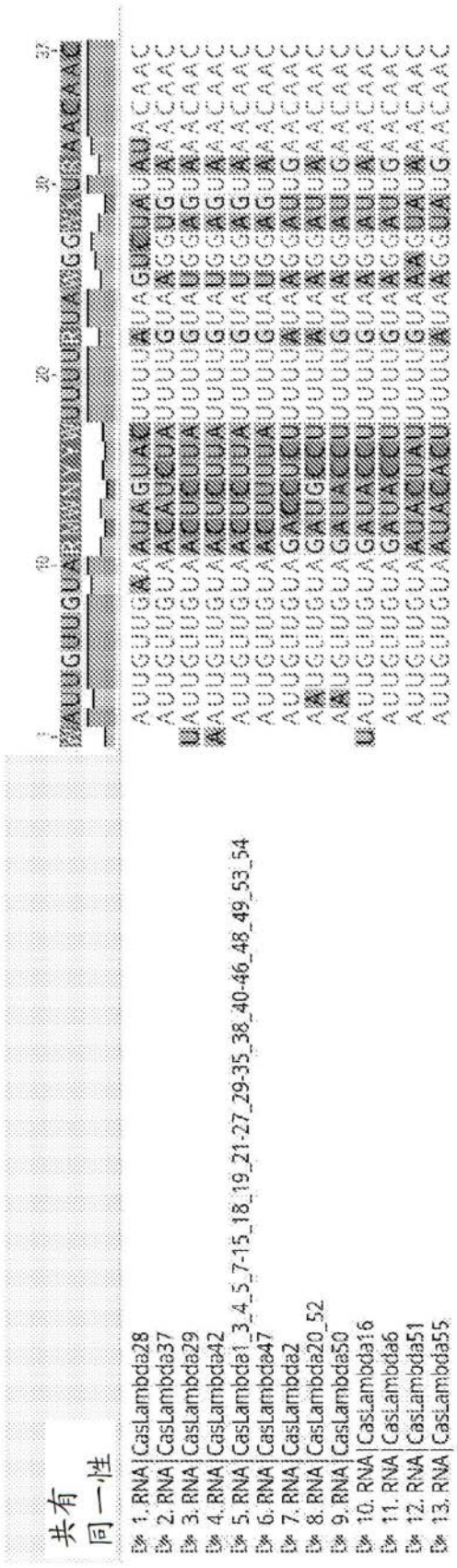

共有
同一性
1. RNA|CasLambda28
2. RNA|CasLambda37
3. RNA|CasLambda29
4. RNA|CasLambda42
5. RNA|CasLambda1_3_4_5_7-15_18_19_21-27_29-35_38_40-46_48_49_53_54
6. RNA|CasLambda47
7. RNA|CasLambda2
8. RNA|CasLambda20_52
9. RNA|CasLambda50
10. RNA|CasLambda16
11. RNA|CasLambda6
12. RNA|CasLambda51
13. RNA|CasLambda55

图7

Casλ1 (SEQ ID NO:59)

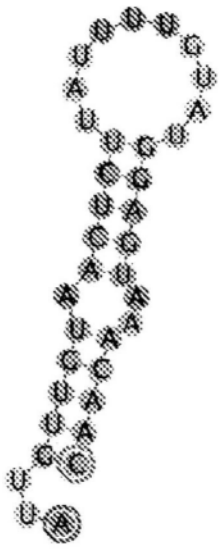


图8A

Casλ2 (SEQ ID NO:60)

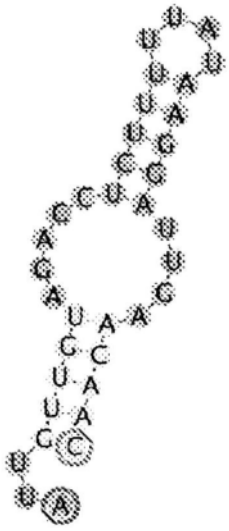


图8B

Casλ6 (SEQ ID NO:164)

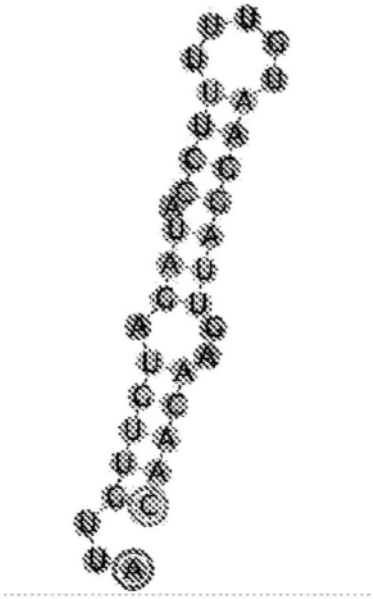


图8C

Casλ16 (SEQ ID NO:61)

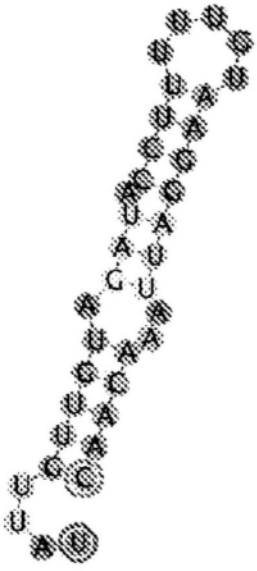


图8D

Casλ20_52 (SEQ ID NO:68)

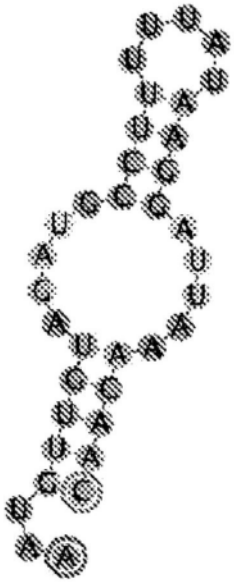


图8E

Casλ28 (SEQ ID NO:63)

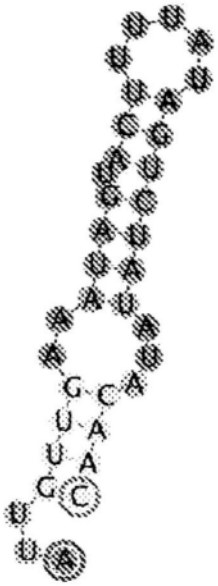


图8F

Casλ29 (SEQ ID NO:64)

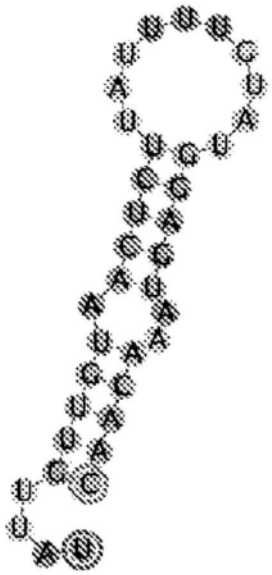


图8G

Casλ37 (SEQ ID NO:70)

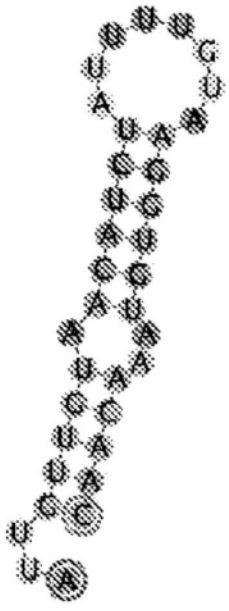


图8H

Casλ42 (SEQ ID NO:163)

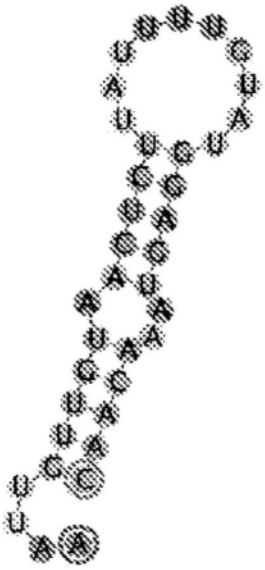


图8I

Casλ47 (SEQ ID NO:65)

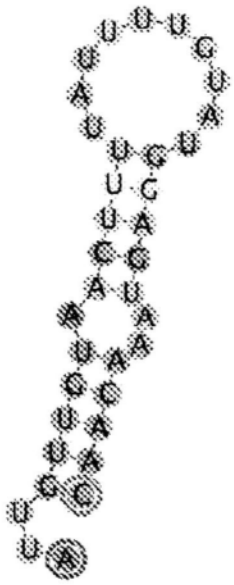


图8J

Casλ50 (SEQ ID NO:66)

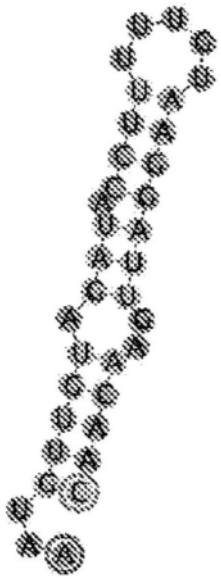


图8K

Casλ51 (SEQ ID NO:67)

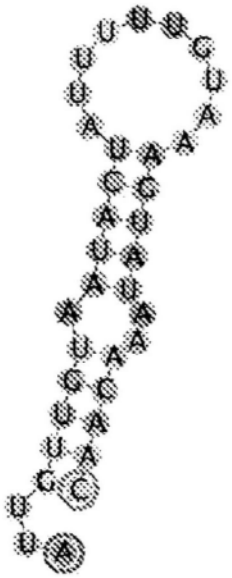


图8L

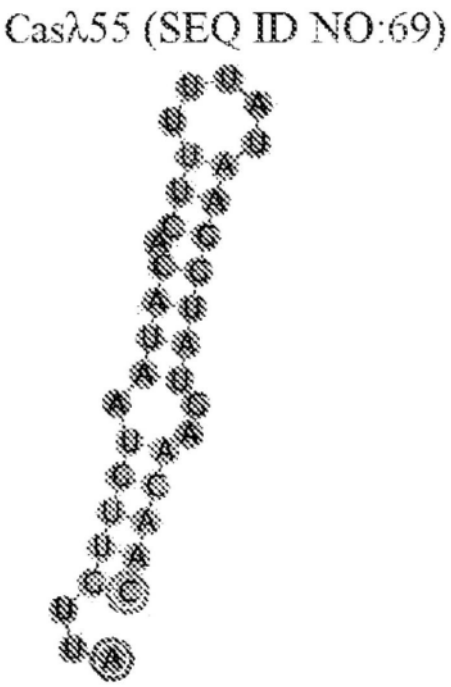


图8M

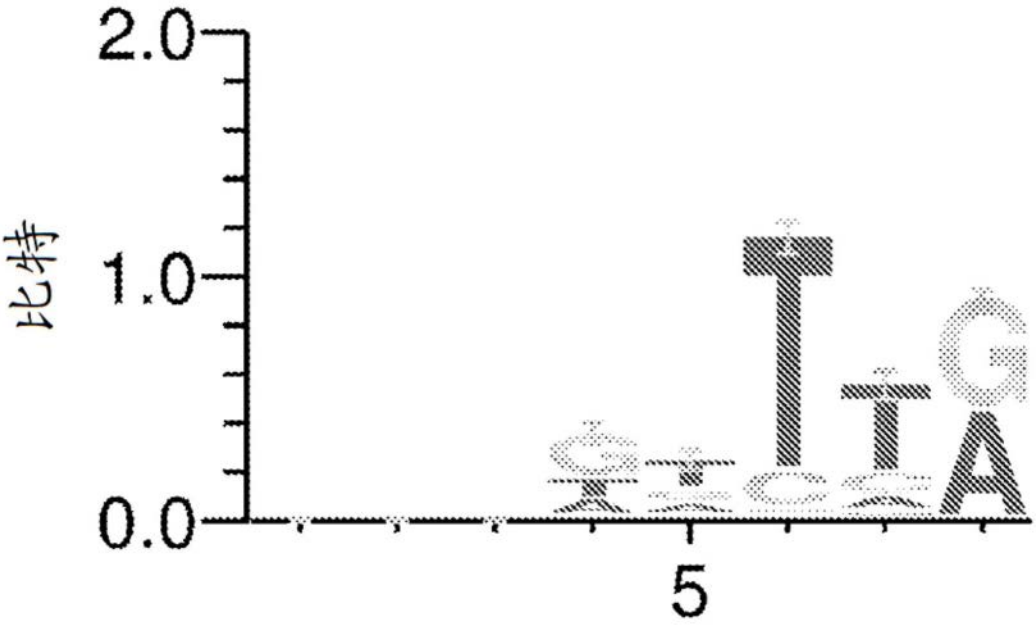


图9

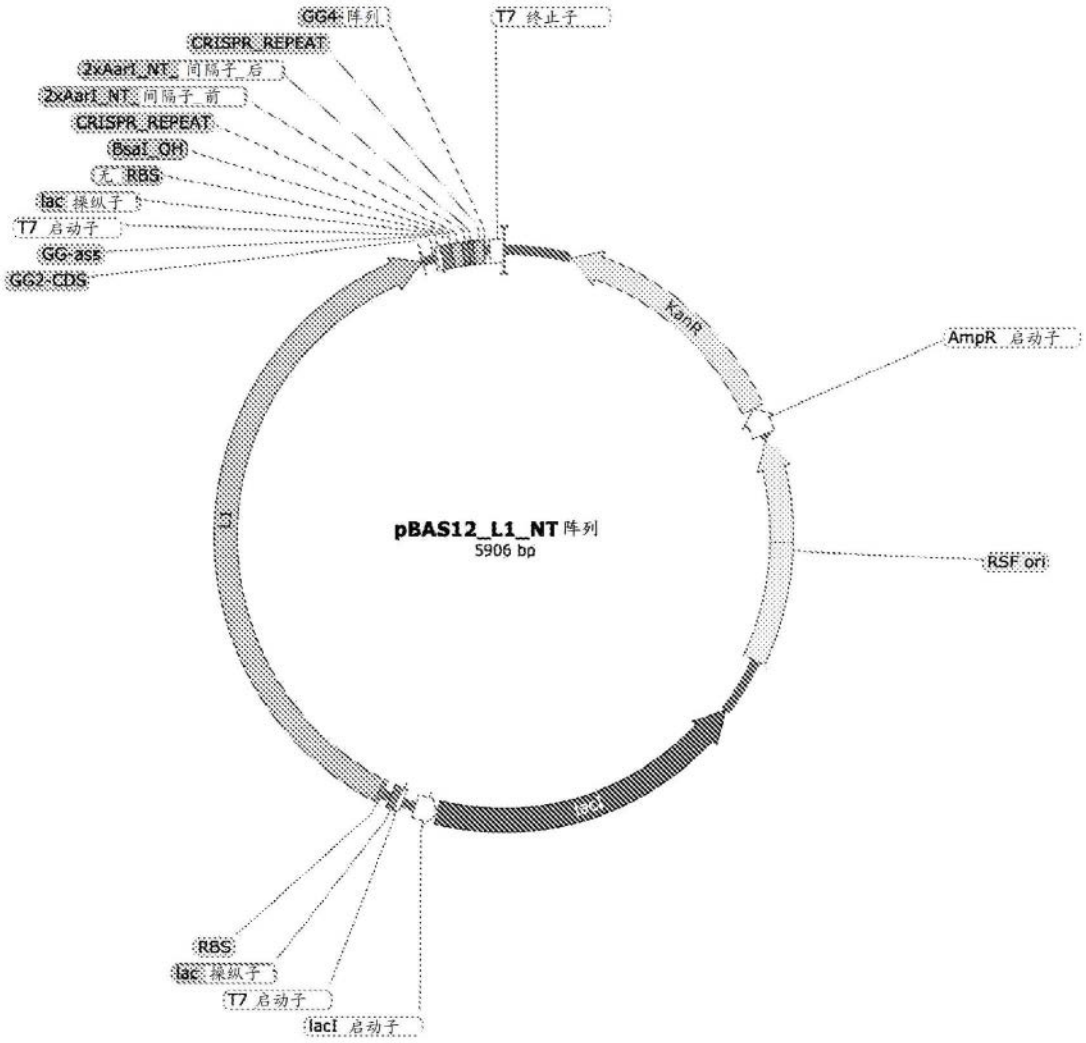

GG4 阵列
T7 终止子
CRISPR_REPEAT
2xAarI_NT 间隔子 后
2xAarI_NT 间隔子 前
CRISPR_REPEAT
BsaI_OH
RBS
lac 操纵子
T7 启动子
GG-ass
GG2-CDS
KanR
AmpR 启动子
pBAS12_L1_NT 阵列
5906 bp
RSF ori
RBS
lac 操纵子
T7 启动子
lacI 启动子

图10

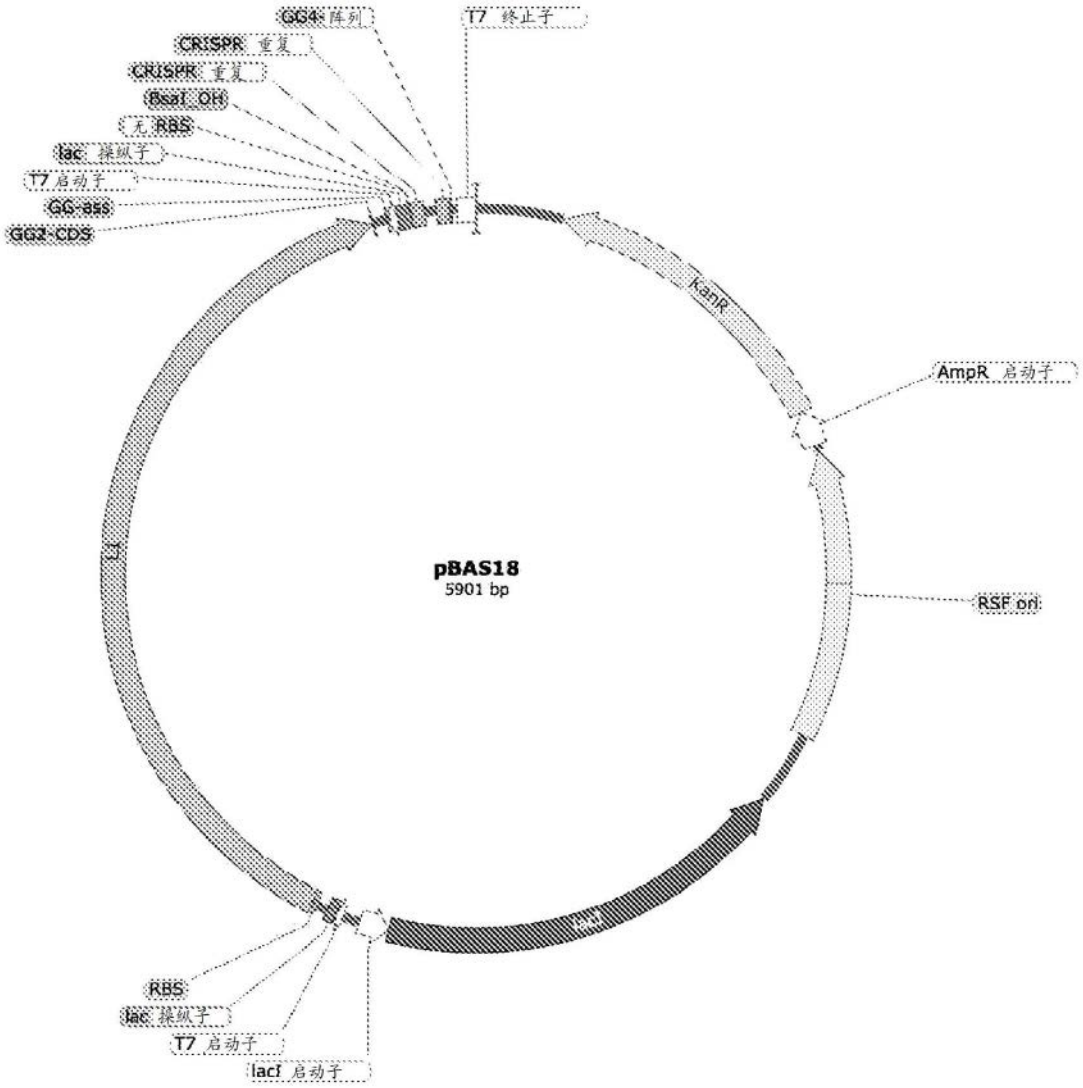

GG4 阵列
T7 终止子
CRISPR 重复
CRISPR 重复
BsaI OH
无 RBS
lac 操纵子
T7 启动子
GG-ass
GG2-CDS
KanR
AmpR 启动子
pBAS18
5901 bp
RSF ori
RBS
lac 操纵子
T7 启动子
lacI 启动子

图11

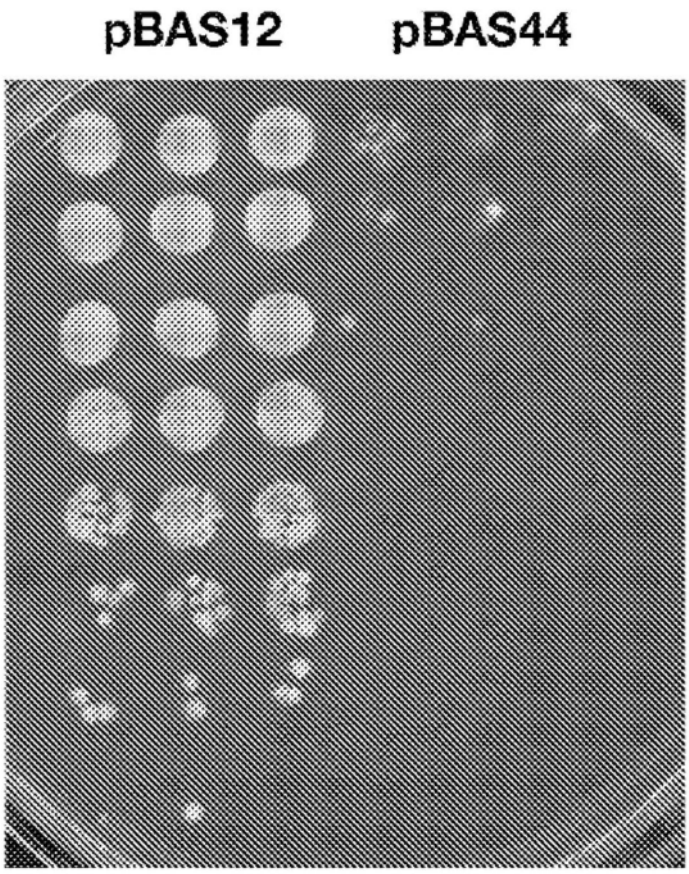


图12A

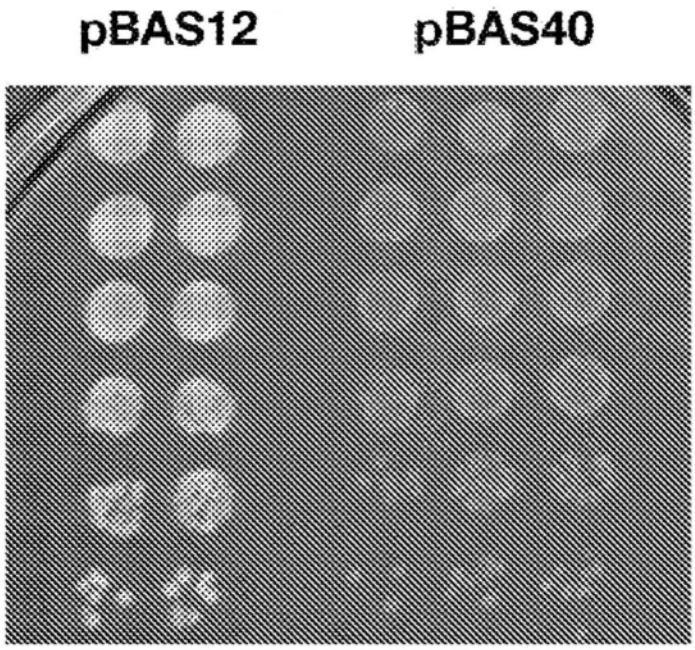


图12B

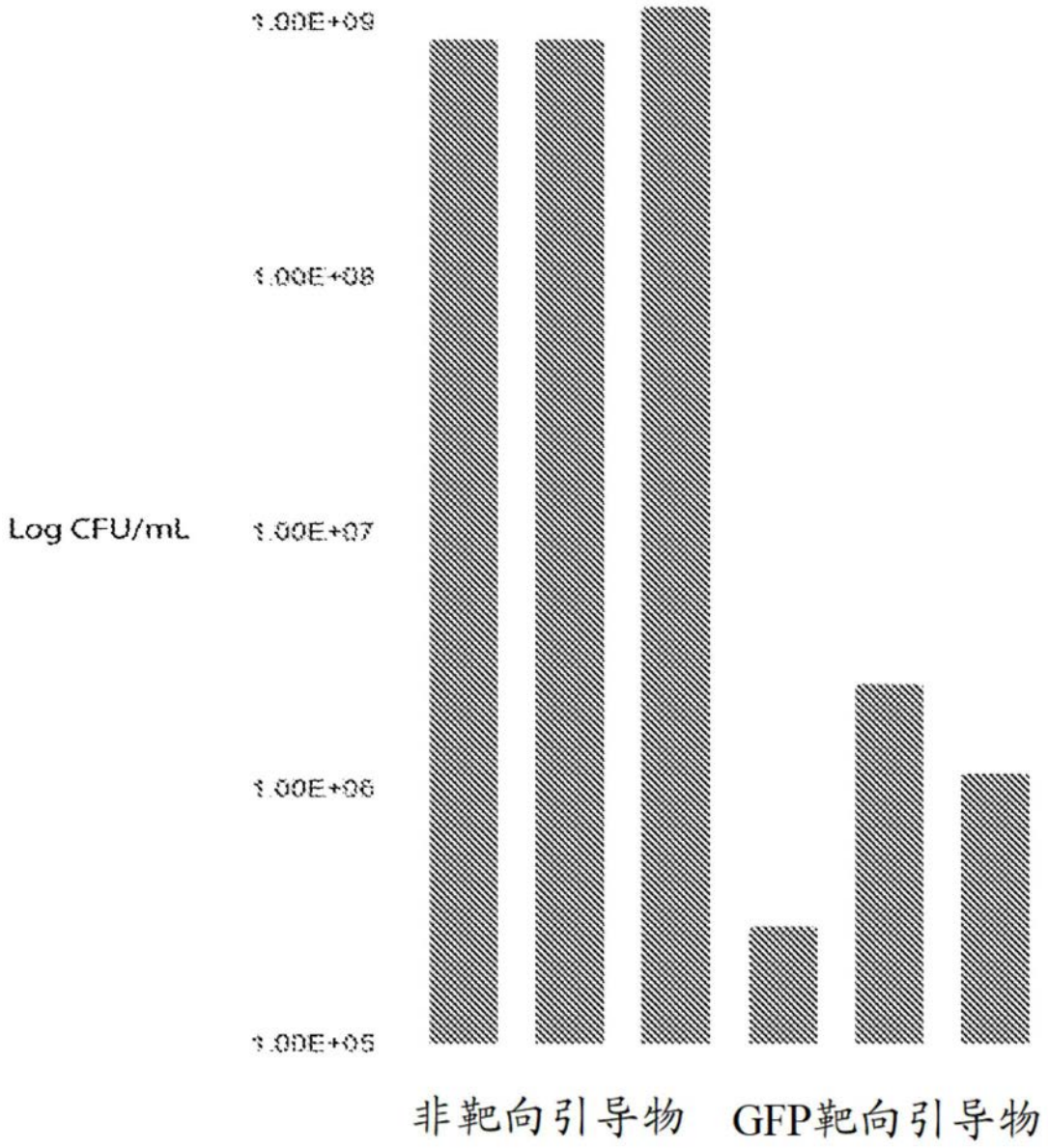

1.00E+09
1.00E+08
Log CFU/mL
1.00E+07
1.00E+06
1.00E+05
非靶向引导物
GFP靶向引导物

图12C