(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 112424348 B
(45)授权公告日 2024.12.31

(21)申请号 201980033565.0
(22)申请日 2019.03.19
(65)同一申请的已公布的文献号
申请公布号 CN 112424348 A
(43)申请公布日 2021.02.26
(30)优先权数据
18162683.9 2018.03.19 EP
18162681.3 2018.03.19 EP
18172625.8 2018.05.16 EP
18174707.2 2018.05.29 EP
18181680.2 2018.07.04 EP
62/745246 2018.10.12 US
62/745240 2018.10.12 US
62/745238 2018.10.12 US
62/745239 2018.10.12 US
(85)PCT国际申请进入国家阶段日
2020.11.18
(86)PCT国际申请的申请数据
PCT/US2019/023044 2019.03.19
(87)PCT国际申请的公布数据
WO2019/183150 EN 2019.09.26
(73)专利权人 克里斯珀医疗股份公司
地址 瑞士楚格
专利权人 拜尔健康护理有限责任公司
(72)发明人 A·科亨 M·施密特 W·科科
M·B·加马林达 A·古普塔
C·皮茨勒 F·里希特 J·特贝
C·程 R·竹内 C·W·赖斯
(74)专利代理机构 中国专利代理(香港)有限公司 72001
专利代理师 甘霖 彭昶
(51)Int.Cl.
C12N 9/22(2006.01)
C12N 15/10(2006.01)
(56)对比文件
WO 2015077318 A1,2015.05.28
审查员 颜泉梅

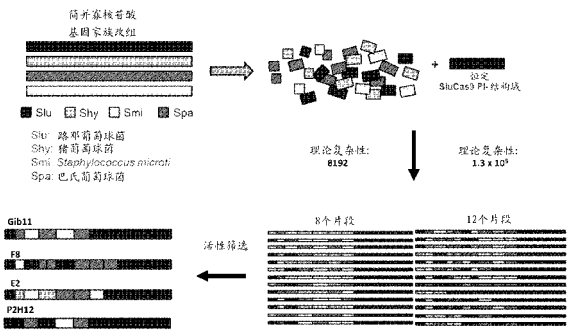
权利要求书3页 说明书145页
序列表290页 附图9页

(54)发明名称

新颖的RNA-可编程的内切核酸酶系统及其用途

(57)摘要

本发明的方面尤其涉及使用新颖的合成的RNA-指导的核酸酶(sRGN)靶向、编辑或操纵细胞中的DNA的新颖系统。所述sRGN衍生自野生型或亲本小II型CRISPR Cas9内切核酸酶。



CN 112424348 B

1.合成的RNA-指导的核酸酶(sRGN)多肽,其从N-端至C-端包含:

1)小结构域1,其包含SEQ ID NO:34-37中的任一个的氨基酸序列,

2)小结构域2,其包含SEQ ID NO:38-41中的任一个的氨基酸序列,

3)小结构域3,其包含SEQ ID NO:42-45中的任一个的氨基酸序列,

4)小结构域4,其包含SEQ ID NO:46-49中的任一个的氨基酸序列,

5)小结构域5,其包含SEQ ID NO:50-53中的任一个的氨基酸序列,

6)小结构域6,其包含SEQ ID NO:54-57中的任一个的氨基酸序列,

7)小结构域7,其包含SEQ ID NO:58-61中的任一个的氨基酸序列,和

8)小结构域8,其包含SEQ ID NO:62-65中的任一个的氨基酸序列,

其中所述小结构域中的至少2个衍生自选自以下的不同亲本Cas9内切核酸酶:如SEQ ID NO:30中所示的路邓葡萄球菌Cas9、如SEQ ID NO:31中所示的巴氏葡萄球菌Cas9、如SEQ ID NO:33中所示的Staphylococcus microti Cas9和如SEQ ID NO:32中所示的猪葡萄球菌Cas9;并且

其中所述sRGN多肽由选自SEQ ID NO:1、2、3、4、5、6、7、8、9、10、11和12的氨基酸序列组成。

2.权利要求1的sRGN多肽,其中所述小结构域中的至少3个衍生自选自以下的不同亲本Cas9内切核酸酶:如SEQ ID NO:30中所示的路邓葡萄球菌Cas9、如SEQ ID NO:31中所示的巴氏葡萄球菌Cas9、如SEQ ID NO:33中所示的Staphylococcus microti Cas9和如SEQ ID NO:32中所示的猪葡萄球菌Cas9。

3.权利要求1或2的sRGN多肽,其中小结构域8包含SEQ ID NO:62的氨基酸序列。

4.权利要求1或2的sRGN多肽,其中小结构域8包含SEQ ID NO:63的氨基酸序列。

5.权利要求1或2的sRGN多肽,其中小结构域1包含SEQ ID NO:34的氨基酸序列。

6.权利要求1的sRGN多肽,其选自:

a)由SEQ ID NO:1的氨基酸序列组成的Gib11多肽,

b)由SEQ ID NO:2的氨基酸序列组成的Gib11Spa-1多肽,和

c)由SEQ ID NO:4的氨基酸序列组成的Gib11Spa-3多肽。

7.一种核酸,其编码权利要求1-6中的任一项的sRGN多肽。

8.权利要求7的核酸,其中所述核酸进行密码子优化以用于在宿主细胞中的表达。

9.权利要求7或8的核酸,其包含SEQ ID NO:13-29中的任一个的核苷酸序列。

10.一种系统,其包含:

(a)根据权利要求1-6中的任一项的sRGN多肽或编码所述sRGN多肽的核酸;和

(b)指导RNA(gRNA)或编码所述gRNA的核酸,其中所述gRNA能够将所述sRGN多肽或其变体指导至靶多核苷酸序列。

11.权利要求10的系统,所述系统进一步包括包含异源多核苷酸序列的供体模板,其中所述异源多核苷酸序列能够被插入所述靶多核苷酸序列中。

12.权利要求10或11的系统,其中所述编码sRGN多肽的核酸进行密码子优化以用于在宿主细胞中的表达。

13.权利要求11的系统,其中所述异源多核苷酸序列进行密码子优化以用于在宿主细胞中的表达。

14.权利要求10或11的系统,其中所述编码sRGN多肽的核酸是脱氧核糖核酸(DNA)。

15.权利要求10或11的系统,其中所述编码sRGN多肽的核酸是核糖核酸(RNA)。

16.权利要求15的系统,其中所述编码sRGN多肽的RNA是mRNA。

17.权利要求11和16中的任一项的系统,其中在腺相关病毒(AAV)载体中编码所述供体模板。

18.权利要求10-11和16中的任一项的系统,其中将所述sRGN多肽或编码所述sRGN多肽的核酸配制在脂质体或脂质纳米颗粒中。

19.权利要求18的系统,其中所述脂质体或脂质纳米颗粒也包含gRNA或编码所述gRNA的核酸。

20.权利要求10-11和16中的任一项的系统,所述系统包含与gRNA预络合的sRGN多肽,从而形成核糖核蛋白(RNP)复合物。

21.一种靶向、编辑、修饰或操纵靶基因座处的靶DNA的方法,所述方法包括将以下物质提供给靶DNA:

(a)根据权利要求1-6中的任一项的sRGN多肽或编码所述sRGN多肽的核酸;和

(b)指导RNA(gRNA)或编码所述gRNA的核酸,其中所述gRNA能够将所述sRGN多肽指导至所述靶基因座。

22.根据权利要求21所述的方法,所述方法进一步包括给所述靶DNA提供包含异源多核苷酸序列的供体模板,其中所述异源多核苷酸序列能够被插入所述靶基因座中。

23.根据权利要求21或22所述的方法,其中所述编码sRGN多肽的核酸进行密码子优化以用于在宿主细胞中的表达。

24.根据权利要求22所述的方法,其中所述异源多核苷酸序列进行密码子优化以用于在宿主细胞中的表达。

25.根据权利要求21或22所述的方法,其中所述编码sRGN多肽的核酸是脱氧核糖核酸(DNA)。

26.根据权利要求21或22所述的方法,其中所述编码sRGN多肽的核酸是核糖核酸(RNA)。

27.根据权利要求26所述的方法,其中所述编码sRGN多肽的RNA是mRNA。

28.根据权利要求21-22和27中的任一项所述的方法,其中在腺相关病毒(AAV)载体中编码所述供体模板。

29.根据权利要求21-22和27中的任一项所述的方法,其中将所述sRGN多肽或编码所述sRGN多肽的核酸配制在脂质体或脂质纳米颗粒中。

30.根据权利要求29所述的方法,其中所述脂质体或脂质纳米颗粒也包含gRNA或编码所述gRNA的核酸。

31.根据权利要求21-22和27中的任一项所述的方法,所述方法包括给所述靶DNA提供sRGN多肽,所述sRGN多肽与所述gRNA预络合为RNP复合物。

32.一种经修饰的细胞,其包含:

(a)根据权利要求1-6中的任一项的sRGN多肽或编码所述sRGN多肽的核酸;和

(b)指导RNA(gRNA)或编码所述gRNA的核酸,其中所述gRNA能够将所述sRGN多肽或其变体指导至靶多核苷酸序列;

其中所述经修饰的细胞是细菌细胞、真菌细胞、古细菌细胞、原生生物或动物细胞。

33.权利要求32的经修饰的细胞,所述细胞进一步含有包含异源多核苷酸序列的供体模板,其中所述异源多核苷酸序列能够被插入所述靶多核苷酸序列中。

34.一种遗传修饰的细胞,其中通过根据权利要求21-31中的任一项所述的方法编辑所述细胞的基因组,其中所述遗传修饰的细胞是细菌细胞、真菌细胞、古细菌细胞、原生生物或动物细胞。

35.一种试剂盒,其包含:

(a)根据权利要求1-6中的任一项的sRGN多肽或编码所述sRGN多肽的核酸;和

(b)指导RNA(gRNA)或编码所述gRNA的核酸,其中所述gRNA能够将所述sRGN多肽或其变体指导至靶多核苷酸序列。

36.根据权利要求35所述的试剂盒,所述试剂盒进一步含有包含异源多核苷酸序列的供体模板,其中所述异源多核苷酸序列能够被插入所述靶多核苷酸序列中。

新颖的RNA-可编程的内切核酸酶系统及其用途

[0001] 相关申请的交叉引用

[0002] 本申请要求2018年3月19日提交的欧洲专利申请号18162681.3、2018年3月19日提交的欧洲专利申请号18162683.9、2018年5月16日提交的欧洲专利申请号18172625.8、2018年5月16日提交的欧洲专利申请号18174707.2、2018年7月4日提交的欧洲专利申请号18181680.2、2018年10月12日提交的美国专利申请号62/745,238、2018年10月12日提交的美国专利申请号62/745,239、2018年10月12日提交的美国专利申请号62/745,240和2018年10月12日提交的美国专利申请号62/745,246的优先权,它们中的每一篇的内容明确地通过引用整体并入本文,包括任何附图。

[0003] 领域

[0004] 本公开内容总体上涉及分子生物学领域,包括组合物和方法,其涉及包括RNA-可编程的内切核酸酶、相关的指导RNA和/或靶序列的新颖系统,以及在各种应用中生产和使用它们的方法,包括调节转录的方法,以及在细胞中靶向、编辑和/或操纵DNA的方法。

[0005] 背景

[0006] 内切核酸酶如锌指内切核酸酶(ZFN)、转录激活因子样效应物核酸酶(TALEN)和核糖核酸酶已被作为位点特异性核酸酶用于基因组靶向、基因组编辑、基因沉默、转录调节、促进重组和其它分子生物学技术。成簇的规律间隔的短回文重复序列(CRISPR)-CRISPR相关蛋白(Cas)系统提供了新颖核酸酶和内切核酸酶的来源,包括CRISPR-Cas9。

[0007] CRISPR系统是为原核生物提供对外源基因元件诸如噬菌体的抗性的原核系统。如在酿脓链球菌(*Streptococcus pyogenes*)中所证明的,由成熟的激活tracr RNA和靶向crRNA之间形成的双链体指导的Cas9在侵入关联DNA中引入位点特异性双链DNA(dsDNA)断裂。Cas9是一种多结构域酶,其使用HNH核酸酶结构域切割靶链(定义为与crRNA的间隔区序列互补),并使用RuvC-样结构域切割非靶链,从而使得能够通过选择性基序失活将dsDNA切割性Cas9转化成切口酶。DNA切割特异性由两个参数决定:靶向原间隔区(protospacer)序列的crRNA的可变的、间隔区衍生的序列(不与crRNA的间隔区互补的DNA靶标上的序列)和紧接在非靶DNA链上的原间隔区的3’(下游)定位的短序列,原间隔区相邻基序(Protospacer Adjacent Motif)(PAM)。

[0008] 已在过去几年中广泛利用使用CRISPR-Cas的RNA-指导的DNA靶向原理编辑基因组,如在WO2013/176722中所述。研究已经证实,RNA-指导的Cas9可用作人细胞、小鼠、斑马鱼、果蝇、蠕虫、植物、酵母和细菌以及各种其它物种中的基因组编辑工具。所述系统是通用的,从而使得能够通过对Cas9进行编程以使用多种指导RNA同时编辑基因组中的几个位点来进行多重基因组工程改造。Cas9向切口酶的转化被证明会促进在诱变活性降低的情况下在哺乳动物基因组中的同源性指导的修复。此外,例如,已经利用Cas9催化失活突变体的DNA结合活性来工程改造RNA-可编程的转录沉默和激活装置或表现遗传修饰因子(epigenetic modifiers)。

[0009] 尽管CRISPR-Cas 9系统有望用于基因编辑,但该系统的使用仍存在许多问题。例如,它们具有一个或多个以下缺点:

[0010] a)它们的大小太大,以致于无法携带到确立的治疗上合适的病毒转染系统(如腺相关病毒(AAV))的基因组内。

[0011] b)它们在异源环境中的活性通常对于在这些环境中使用来说太低了,例如对于在真核生物且特别在哺乳动物环境中有效使用来说太低了。

[0012] c)它们的核酸酶活性缺乏保真度,从而导致不希望的脱靶(off target)效应,这将例如使它们不适合于基因治疗用途或需要高精度的其它应用。

[0013] d)它们可能触发免疫应答,这可限制其在哺乳动物体内应用的用途。

[0014] e)它们要求复杂和/或长的PAM,这限制了对于DNA靶向区段的靶选择。

[0015] f)它们表现出从质粒或病毒载体的差表达。

发明内容

[0016] 这个部分提供了本公开内容的总概述,而不包含其全部范围或所有其特征。

[0017] 在某些实施方案中,本文提供了与新颖的RNA-可编程的内切核酸酶有关的组合物和方法,以及使用这样的核酸酶的系统。本发明还涉及其它组分,包括指导RNA和/或靶序列,以及在各种应用中产生和使用它们的方法。这样的应用的例子包括用于调节转录的方法,以及使用新颖的核酸酶和其它组分诸如核酸和/或多肽来靶向、编辑和/或操纵DNA的方法。本公开内容的一些实施方案还涉及包含本文公开的一种或多种系统元件的重组细胞和试剂盒。

[0018] 在一个方面,本文提供了一种合成的RNA-指导的核酸酶(sRGN)多肽,其从N-端至C-端包含:1)小结构域1,其包含SEQ ID NO: 34-37中的任一个的氨基酸序列或其与SEQ ID NO: 34-37中的任一个具有至少约90%序列同一性的变体,2)小结构域2,其包含SEQ ID NO: 38-41中的任一个的氨基酸序列或其与SEQ ID NO: 38-41中的任一个具有至少约90%序列同一性的变体,3)小结构域3,其包含SEQ ID NO: 42-45中的任一个的氨基酸序列或其与SEQ ID NO: 42-45中的任一个具有至少约90%序列同一性的变体,4)小结构域4,其包含SEQ ID NO: 46-49中的任一个的氨基酸序列或其与SEQ ID NO: 46-49中的任一个具有至少约90%序列同一性的变体,5)小结构域5,其包含SEQ ID NO: 50-53中的任一个的氨基酸序列或其与SEQ ID NO: 50-53中的任一个具有至少约90%序列同一性的变体,6)小结构域6,其包含SEQ ID NO: 54-57中的任一个的氨基酸序列或其与SEQ ID NO: 54-57中的任一个具有至少约90%序列同一性的变体,7)小结构域7,其包含SEQ ID NO: 58-61中的任一个的氨基酸序列或其与SEQ ID NO: 58-61中的任一个具有至少约90%序列同一性的变体,和8)小结构域8,其包含SEQ ID NO: 62-65中的任一个的氨基酸序列或其与SEQ ID NO: 62-65中的任一个具有至少约90%序列同一性的变体,其中所述小结构域中的至少2个衍生自选自以下的不同亲本Cas9内切核酸酶:路邓葡萄球菌(*Staphylococcus lugdunensis*)Cas9 (SEQ ID NO: 30), 巴氏葡萄球菌(*Staphylococcus pasteuri*)Cas9 (SEQ ID NO: 31), *Staphylococcus microti* Cas9 (SEQ ID NO: 33),和猪葡萄球菌(*Staphylococcus hyicus*)Cas9 (SEQ ID NO: 32)。在某些实施方案中,所述小结构域中的至少3个衍生自选自以下的不同亲本Cas9内切核酸酶:路邓葡萄球菌Cas9 (SEQ ID NO: 30)、巴氏葡萄球菌Cas9 (SEQ ID NO: 31)、*Staphylococcus microti* Cas9 (SEQ ID NO: 33)和猪葡萄球菌Cas9 (SEQ ID NO: 32)。在某些实施方案中,小结构域8包含SEQ ID NO: 62的氨基酸序列

或其与SEQ ID NO: 62具有至少约90%序列同一性的变体。在某些实施方案中,小结构域8包含SEQ ID NO: 63的氨基酸序列或其与SEQ ID NO: 63具有至少约90%序列同一性的变体。在某些实施方案中,小结构域1包含SEQ ID NO: 34的氨基酸序列或其与SEQ ID NO: 34具有至少约90%序列同一性的变体。

[0019] 在另一个方面,本文提供了一种sRGN多肽,其从N-端至C-端包含:1)小结构域1,其包含SEQ ID NO: 66-69中的任一个的氨基酸序列或其与SEQ ID NO: 66-69中的任一个具有至少约90%序列同一性的变体,2)小结构域2,其包含SEQ ID NO: 70-73中的任一个的氨基酸序列或其与SEQ ID NO: 70-73中的任一个具有至少约90%序列同一性的变体,3)小结构域3,其包含SEQ ID NO: 74-77中的任一个的氨基酸序列或其与SEQ ID NO: 74-77中的任一个具有至少约90%序列同一性的变体,4)小结构域4,其包含SEQ ID NO: 78-81中的任一个的氨基酸序列或其与SEQ ID NO: 78-81中的任一个具有至少约90%序列同一性的变体,5)小结构域5,其包含SEQ ID NO: 82-85中的任一个的氨基酸序列或其与SEQ ID NO: 82-85中的任一个具有至少约90%序列同一性的变体,6)小结构域6,其包含SEQ ID NO: 86-89中的任一个的氨基酸序列或其与SEQ ID NO: 86-89中的任一个具有至少约90%序列同一性的变体,7)小结构域7,其包含SEQ ID NO: 90-93中的任一个的氨基酸序列或其与SEQ ID NO: 90-93中的任一个具有至少约90%序列同一性的变体,8)小结构域8,其包含SEQ ID NO: 94-97中的任一个的氨基酸序列或其与SEQ ID NO: 94-97中的任一个具有至少约90%序列同一性的变体,9)小结构域9,其包含SEQ ID NO: 98-101中的任一个的氨基酸序列或其与SEQ ID NO: 98-101中的任一个具有至少约90%序列同一性的变体,10)小结构域10,其包含SEQ ID NO: 102-105中的任一个的氨基酸序列或其与SEQ ID NO: 102-105中的任一个具有至少约90%序列同一性的变体,11)小结构域11,其包含SEQ ID NO: 106-109中的任一个的氨基酸序列或其与SEQ ID NO: 106-109中的任一个具有至少约90%序列同一性的变体,和12)小结构域12,其包含SEQ ID NO: 110-113中的任一个的氨基酸序列或其与SEQ ID NO: 110-113中的任一个具有至少约90%序列同一性的变体,其中所述小结构域中的至少2个衍生自选自以下的不同亲本Cas9内切核酸酶:路邓葡萄球菌Cas9 (SEQ ID NO: 30)、巴氏葡萄球菌Cas9 (SEQ ID NO: 31)、*Staphylococcus microti* Cas9 (SEQ ID NO: 33)和猪葡萄球菌Cas9 (SEQ ID NO: 32)。在某些实施方案中,所述小结构域中的至少3个衍生自选自以下的不同亲本Cas9内切核酸酶:路邓葡萄球菌Cas9 (SEQ ID NO: 30)、巴氏葡萄球菌Cas9 (SEQ ID NO: 31)、*Staphylococcus microti* Cas9 (SEQ ID NO: 33)和猪葡萄球菌Cas9 (SEQ ID NO: 32)。在某些实施方案中,小结构域12包含SEQ ID NO: 110的氨基酸序列或其与SEQ ID NO: 110具有至少约90%序列同一性的变体。在某些实施方案中,小结构域12包含SEQ ID NO: 111的氨基酸序列或其与SEQ ID NO: 111具有至少约90%序列同一性的变体。在某些实施方案中,(i)小结构域1包含SEQ ID NO: 66的氨基酸序列或其与SEQ ID NO: 66具有至少约90%序列同一性的变体,和(ii)小结构域2包含SEQ ID NO: 70的氨基酸序列或其与SEQ ID NO: 70具有至少约90%序列同一性的变体。

[0020] 在另一个方面,本文提供了一种sRGN多肽,其选自:a) Gib11多肽,其包含SEQ ID NO: 1的氨基酸序列或其与SEQ ID NO: 1具有至少约90%序列同一性的变体,b) Gib11Spa-1多肽,其包含SEQ ID NO: 2的氨基酸序列或其与SEQ ID NO: 2具有至少约90%序列同一性的变体,c) Gib11Spa-2多肽,其包含SEQ ID NO: 3的氨基酸序列或其与SEQ ID NO: 3具有

至少约90%序列同一性的变体,d) Gib11Spa-3多肽,其包含SEQ ID NO: 4的氨基酸序列或其与SEQ ID NO: 4具有至少约90%序列同一性的变体,e) P2H12多肽,其包含SEQ ID NO: 5的氨基酸序列或其与SEQ ID NO: 5具有至少约90%序列同一性的变体,f) E2多肽,其包含SEQ ID NO: 6的氨基酸序列或其与SEQ ID NO: 6具有至少约90%序列同一性的变体,g) E2+K741D+L743K多肽,其包含SEQ ID NO: 7的氨基酸序列或其与SEQ ID NO: 7具有至少约90%序列同一性的变体,h) E2+S670T+N675D多肽,其包含SEQ ID NO: 8的氨基酸序列或其与SEQ ID NO: 8具有至少约90%序列同一性的变体,i) E2+K741N+L743N多肽,其包含SEQ ID NO: 9的氨基酸序列或其与SEQ ID NO: 9具有至少约90%序列同一性的变体,j) F8多肽,其包含SEQ ID NO: 10的氨基酸序列或其与SEQ ID NO: 10具有至少约90%序列同一性的变体;k) F8+K737D+L739K多肽,其包含SEQ ID NO: 11的氨基酸序列或其与SEQ ID NO: 11具有至少约90%序列同一性的变体;和l) F8+K737N+L739N多肽,其包含SEQ ID NO: 12的氨基酸序列或其与SEQ ID NO: 12具有至少约90%序列同一性的变体。

[0021] 在某些实施方案中,所述sRGN多肽选自:a) Gib11多肽,其包含SEQ ID NO: 1的氨基酸序列或其与SEQ ID NO: 1具有至少约90%序列同一性的变体,b) Gib11Spa-1多肽,其包含SEQ ID NO: 2的氨基酸序列或其与SEQ ID NO: 2具有至少约90%序列同一性的变体,和c) Gib11Spa-3多肽,其包含SEQ ID NO: 4的氨基酸序列或其与SEQ ID NO: 4具有至少约90%序列同一性的变体。在某些实施方案中,所述sRGN多肽选自:a)包含SEQ ID NO: 1的氨基酸序列的Gib11多肽, b)包含SEQ ID NO: 2的氨基酸序列的Gib11Spa-1多肽,和c)包含SEQ ID NO: 4的氨基酸序列的Gib11Spa-3多肽。

[0022] 在另一个方面,本文提供了一种核酸,其编码根据上述实施方案中的任一个的sRGN多肽。在某些实施方案中,所述核酸为在宿主细胞中的表达经过密码子优化。在某些实施方案中,所述核酸包含SEQ ID NO: 13-29中的任一个的核苷酸序列或其与SEQ ID NO: 13-29中的任一个具有至少约90%序列同一性的变体。

[0023] 在另一个方面,本文提供了一种系统,其包含:(a)根据权利要求1-11中的任一项的sRGN多肽或编码sRGN多肽的核酸;和(b)指导RNA (gRNA)或编码所述gRNA的核酸,其中所述gRNA能够将所述sRGN多肽或其变体指导至靶多核苷酸序列。在某些实施方案中,所述系统进一步包括包含异源多核苷酸序列的供体模板,其中所述异源多核苷酸序列能够被插入所述靶多核苷酸序列中。

[0024] 在某些实施方案中,根据上述系统中的任一个,编码所述sRGN多肽的核酸为在宿主细胞中的表达经过密码子优化和/或所述异源多核苷酸序列为在宿主细胞中的表达经过密码子优化。

[0025] 在某些实施方案中,根据上述系统中的任一个,编码所述sRGN多肽的核酸是脱氧核糖核酸(DNA)。

[0026] 在某些实施方案中,根据上述系统中的任一个,编码所述sRGN多肽的核酸是核糖核酸(RNA)。在某些实施方案中,编码所述sRGN多肽的RNA是mRNA。

[0027] 在某些实施方案中,根据上述系统中的任一个,在腺相关病毒(AAV)载体中编码供体模板。

[0028] 在某些实施方案中,根据上述系统中的任一个,将所述sRGN多肽或编码sRGN多肽的核酸配制在脂质体或脂质纳米颗粒中。在某些实施方案中,所述脂质体或脂质纳米颗粒

也包含所述gRNA或编码所述gRNA的核酸。

[0029] 在某些实施方案中,根据上述系统中的任一个,所述系统包含与gRNA预络合的sRGN多肽,从而形成核糖核蛋白(RNP)复合物。

[0030] 在另一个方面,本文提供了一种靶向、编辑、修饰或操纵靶基因座处的靶DNA的方法,所述方法包括将以下物质提供给靶DNA: (a)根据上述实施方案中的任一个的sRGN多肽或编码所述sRGN多肽的核酸;和(b)指导RNA (gRNA)或编码所述gRNA的核酸,其中所述gRNA能够将所述sRGN多肽指导至所述靶基因座。在某些实施方案中,所述方法还包括给所述靶DNA提供包含异源多核苷酸序列的供体模板,其中所述异源多核苷酸序列能够被插入所述靶基因座中。

[0031] 在某些实施方案中,根据上述方法中的任一个,编码所述sRGN多肽的核酸为在宿主细胞中的表达经过密码子优化和/或所述异源多核苷酸序列为在宿主细胞中的表达经过密码子优化。

[0032] 在某些实施方案中,根据上述方法中的任一个,编码所述sRGN多肽的核酸是脱氧核糖核酸(DNA)。

[0033] 在某些实施方案中,根据上述方法中的任一个,编码所述sRGN多肽的核酸是核糖核酸(RNA)。在某些实施方案中,编码所述sRGN多肽的RNA是mRNA。

[0034] 在某些实施方案中,根据上述方法中的任一个,在腺相关病毒(AAV)载体中编码供体模板。

[0035] 在某些实施方案中,根据上述方法中的任一个,将所述sRGN多肽或编码sRGN多肽的核酸配制在脂质体或脂质纳米颗粒中。在某些实施方案中,所述脂质体或脂质纳米颗粒也包含所述gRNA或编码所述gRNA的核酸。在某些实施方案中,所述方法包括给所述靶DNA提供sRGN多肽,所述sRGN多肽与所述gRNA预络合为RNP复合物。

[0036] 在另一个方面,本文提供了一种经修饰的细胞,其包含: (a)根据上述实施方案中的任一个的sRGN多肽或编码所述sRGN多肽的核酸;和(b)指导RNA (gRNA)或编码所述gRNA的核酸,其中所述gRNA能够将所述sRGN多肽或其变体指导至靶多核苷酸序列。在某些实施方案中,所述经修饰的细胞进一步含有包含异源多核苷酸序列的供体模板,其中所述异源多核苷酸序列能够被插入所述靶多核苷酸序列中。

[0037] 在另一个方面,本文提供了一种遗传修饰的细胞,其中通过根据上述实施方案中的任一个的方法编辑所述细胞的基因组。

[0038] 在另一个方面,本文提供了一种试剂盒,其包含:(a)根据权利要求1-11中的任一项的sRGN多肽或编码sRGN多肽的核酸;和(b)指导RNA (gRNA)或编码所述gRNA的核酸,其中所述gRNA能够将所述sRGN多肽或其变体指导至靶多核苷酸序列。在某些实施方案中,所述试剂盒进一步含有包含异源多核苷酸序列的供体模板,其中所述异源多核苷酸序列能够被插入所述靶多核苷酸序列中。

附图说明

[0039] 图1显示了用于产生Gib11、P2H12、E2和F8的基因家族改组技术(gene family shuffling technique)的图示。

[0040] 图2A显示了在破坏测定中基于荧光偏振的生化活性的结果(在某些情况下,活性

超过了野生型Cas9诸如SluCas9的活性),针对Gib11的荧光偏振测定。菱形描绘了具有sgRNA的情况,方框描绘了没有sgRNA的对照情况。

[0041] 图2B显示了关于SluCas9的基于荧光偏振的生化切割测定的结果。菱形描绘了具有sgRNA的情况,方框描绘了没有sgRNA的对照情况。

[0042] 图2C显示了Gib11和SluCas9在切割测定中的相对活性,如通过分别如图2A和2B所示的具有sgRNA的情况的曲线的初始斜率所确定的。

[0043] 图3显示了Gib11识别"NNGG"PAM基序的结果,如通过液体培养中的"活-死"测定法所确定的。

[0044] 图4显示了关于P2H12的基于荧光偏振的生化切割测定的结果。菱形描绘了具有sgRNA的情况,方框描绘了没有sgRNA的对照情况。

[0045] 图5显示了关于E2的基于BFP破坏的切割测定的结果。

[0046] 图6显示了基于金黄色葡萄球菌(*Staphylococcus aureus*)蛋白的结构域组构的Gib11Spa-1、Gib11Spa-2和Gib11Spa-3的预测结构域结构。名为PI的括号表示基础Gib11序列的PI结构域。对应的cRGN-A部分显示为阴影框(在该情况下,关于Gib11),而cRGN-B巴氏葡萄球菌部分显示为白框。氨基酸位置参考Gib11蛋白(SEQ ID NO: 1)。

[0047] 图7显示了关于Gib11Spa-1、Gib11Spa-2和Gib11Spa-3的基于BFP破坏的切割测定的结果。

[0048] 图8显示了关于F8的基于BFP破坏的切割测定的结果。

[0049] 图9显示了Gib11、Gib11Spa-1、Gib11Spa-3、E2和F8的特异性谱,如使用生化切割测定所评估的。

[0050] 图10显示了基于它们的特异性谱的Gib11、Gib11Spa-1、Gib11Spa-3、E2和F8的总特异性。

[0051] 图11显示了Gib11、Gib11Spa-1、Gib11Spa-3、E2和F8的活性谱的分布,如使用生化切割测定所评估的。在x轴上,从左到右是靶DNA底物,按每种指示的核酸酶的活性从最高至最低排序。

具体实施方式

[0052] 大多数现有的II型CRISPR Cas系统是基于来自酿脓链球菌的酶,其具有对于包装入病毒载体如AAV来说太大的特殊缺点(1638个氨基酸)。存在备选的基于来自金黄色葡萄球菌的核酸酶的II型CRISPR Cas系统(EP 2 898 075),其大小显著更小。但是,这种核酸酶需要复杂的PAM,这极大地限制了其用于基因编辑用途的应用。此外,小核酸酶对于基因编辑方法而言可以更有用,但是,已经鉴定出有限数目的这样的小核酸酶,并且它们用于基因编辑的最佳特征至少在大多数情况下未充分定义以用于基因编辑。

[0053] 本文提供了新颖的合成的CRISPR-Cas多肽(在本文中也被称作"sRGN多肽"),其衍生自四个不同葡萄球菌属种(路邓葡萄球菌、巴氏葡萄球菌、*Staphylococcus microti*和猪葡萄球菌)的CRISPR-Cas内切核酸酶的人工定义小结构域,及其具有不同和有利特征和功能性的变体。这些sRGN提供了以前不存在的基因组编辑的其它机会。Cas核酸酶可以具有不同的活性,取决于在其中使用它们的系统。本发明的一个特征是提供另外的经工程改造的核酸酶,其增加了可用于基因编辑和相关方法的核酸酶的工具箱。

[0054] 在许多情况下,天然存在的已知的Cas核酸酶靶向较长的PAM(诸如NNAAAA),这可能限制那些核酸酶的效用。本发明提供了衍生自天然存在的小核酸酶但可识别NNGG PAM的合成核酸酶(例如,小核酸酶)。

[0055] 本发明进一步提供了用于用在原核、真核和体外环境中的合适的PAM序列和合适的向导,包括单个指导RNA (sgRNA)。

[0056] 本文提供的sRGN多肽相对于已报道的CRISPR-Cas内切核酸酶可以表现出有利的特性,例如在原核、真核和/或体外环境中的更高活性,和/或sRGN多肽在真核环境(例如,人宿主细胞)中从核酸的更大表达。在某些情况下,sRGN组合了以下一种或多种:小尺寸,高编辑活性,和仅需要短PAM序列。关于RNA-指导的核酸酶的小尺寸是指长度不大于约1100个氨基酸的核酸酶。

[0057] 还提供了用于靶向、编辑或操纵细胞中DNA的系统,其包含sRGN多肽或编码sRGN多肽的核酸,和一种或多种指导RNA (gRNA),例如,一种或多种单个指导RNA (sgRNA),或编码一种或多种gRNA的核酸。

[0058] 在以下详细描述中,参考形成其一部分的附图。在附图中,除非上下文另外指出,否则相似的符号通常标识相似的部件。在详细描述、附图和权利要求中描述的示例性备选方案不是用来进行限制的。在不背离此处呈现的主题的精神或范围的情况下,可以使用其它备选方案,并且可以做出其它改变。应当容易地理解,可以以多种不同的构型来布置、置换、组合和设计如本文一般地描述的以及在附图中举例说明的各方面,所有这些都被明确地预期并且构成本申请的一部分。

[0059] 除非另有定义,否则本文中使用的所有技术术语、符号和其它科学术语或专门名词意图具有本申请所属领域的技术人员通常理解的含义。在一些情况下,为了清楚和/或为了易于参考,在本文定义了具有通常理解的含义的术语,并且在本文中对这种定义的包括不一定被解释为表示与本领域的通常理解相比相当大的差异。本领域技术人员使用常规方法较好地理解并通常采用本文描述或参考的许多技术和程序。

[0060] 定义

[0061] 在本文中互换使用的术语“多核苷酸”和“核酸”表示任何长度的核苷酸的聚合形式,不是核糖核苷酸就是脱氧核糖核苷酸。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/三螺旋,或包含嘌呤和嘧啶碱基或其它天然的、化学或生物化学修饰的、非天然的或衍生的核苷酸碱基的聚合物。

[0062] “寡核苷酸”通常表示单链或双链DNA的约5至约100个核苷酸的多核苷酸。但是,对于本公开内容的目的,寡核苷酸的长度没有上限。寡核苷酸也被称为“寡聚体”或“寡物(oligos)”,并且可以从基因中分离,或通过本领域已知的方法化学合成。在适用于所描述的实施方案的情况下,应当将术语“多核苷酸”和“核酸”理解为包括单链(诸如有义或反义)和双链多核苷酸。

[0063] “基因组DNA”表示生物的基因组的DNA,包括、但不限于,细菌、真菌、古细菌、原生生物、病毒、植物或动物的基因组的DNA。

[0064] “操纵”DNA包括结合、在一条链产生切口、或切割(即切断)DNA的两条链,或包括修饰或编辑DNA或与DNA结合的多肽。操纵DNA可以沉默、激活或调节(增加或减少)由所述DNA编码的RNA或多肽的表达,或防止或增强多肽与DNA的结合。

[0065] “茎-环结构”表示具有二级结构的核酸,所述二级结构包括已知或预测形成双链的核苷酸区域(茎部分),该区域在一侧通过主要为单链的核苷酸区域(环部分)连接。术语“发夹”和“折回”结构在本文中也用于表示茎-环结构。这样的结构是本领域众所周知的,并且这些术语与其在本领域中的已知含义一致地使用。如本领域中已知的,茎-环结构不需要精确的碱基配对。因此,茎可以包括一个或多个碱基错配。可替换地,碱基配对可以是精确的,即,不包括任何错配。

[0066] “可杂交的”或“互补的”或“基本上互补的”是指,核酸(例如RNA)包含核苷酸序列,该序列使得其能够在适当的体外和/或体内温度和溶液离子强度条件下以序列特异性的、反平行的方式(即,核酸特异性结合互补核酸)与另一核酸非共价地结合,即,形成沃森-克里克碱基对和/或G/U碱基对、“退火”或“杂交”。如本领域已知的,标准的沃森-克里克碱基配对包括:腺嘌呤(A)与胸苷(T)配对、腺嘌呤(A)与尿嘧啶(U)配对和鸟嘌呤(G)与胞嘧啶(C)配对[DNA,RNA]。此外,在本领域中还已知,对于两个RNA分子(例如dsRNA)之间的杂交,鸟嘌呤(G)与尿嘧啶(U)碱基配对。例如,在与mRNA中的密码子进行碱基配对的tRNA反密码子的情况下,G/U碱基配对部分地负责遗传密码的简并性(即,冗余)。在本公开内容的上下文中,指导RNA分子的蛋白结合区段(dsRNA双链体)的鸟嘌呤(G)被认为与尿嘧啶(U)互补,且反之亦然。因此,当可以在指导RNA分子的蛋白结合区段(dsRNA双链体)的给定核苷酸位置产生G/U碱基对时,所述位置不被认为是非互补的,而是被认为是互补的。

[0067] 杂交和洗涤条件是众所周知的,并在Sambrook, J. 和Russell, W., Molecular Cloning: A Laboratory Manual, 第三版, Cold Spring Harbor Laboratory Press, Cold Spring Harbor (2001);和Green, M. R., 和Sambrook, J., Molecular cloning: A Laboratory Manual, 第四版, Cold Spring Harbor Laboratory Press, Cold Spring Harbor (2012)中例示。温度和离子强度的条件决定了杂交的“严格性”。

[0068] 杂交需要两个核酸包含互补序列,尽管碱基之间的错配是可能的。适合于两个核酸之间杂交的条件取决于核酸的长度和互补程度,本领域众所周知的变量。两个核苷酸序列之间的互补程度越大,具有那些序列的核酸的杂交体的解链温度(Tm)的值越大。对于具有短互补序列段(例如在35个或更少、30个或更少、25个或更少、22个或更少、20个或更少或18个或更少的核苷酸范围内的互补性)的核酸之间的杂交,错配的位置变得重要(参见Sambrook等人, 出处同上,11.7-11.8)。通常,可杂交的核酸的长度为至少10个核苷酸。可杂交的核酸的示例性最小长度为:至少15个核苷酸;至少20个核苷酸;至少22个核苷酸;至少25个核苷酸;和至少30个核苷酸)。此外,熟练的技术人员会认识到,根据诸如互补区域的长度和互补程度的因素,可以在必要时调节温度和洗涤溶液盐浓度。

[0069] 在本领域中理解,多核苷酸的序列不需要与其可特异性杂交的靶核酸的序列100%互补。此外,多核苷酸可在一个或多个区段上杂交,从而使得在杂交事件中不涉及间插或相邻区段(例如,环结构或发夹结构)。多核苷酸可包含与其所靶向的靶核酸序列内的靶区域的至少70%、至少80%、至少90%、至少95%、至少99%或100%的序列互补性。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代表90%的互补性。在这个实例中,剩余的非互补核苷酸可以与互补核苷酸成簇或散布,并且不需要彼此或与互补核苷酸邻接。使用本领域已知的BLAST程序(基本局部比对搜索工具(basic local比对search tools))和PowerBLAST程序(Altschul等人, J.Mol.Biol.1990,

215, 403-410;Zhang和Madden, Genome Res., 1997,7, 649-656)或通过使用Gap程序(Wisconsin Sequence Analysis Package, Version 8 for Unix, Genetics Computer Group, University Research Park, Madison Wis.),使用默认设置,其使用Smith和Waterman (Adv. Appl. Math. 1981(2) 482-489)的算法,可以常规地确定核酸内特定核酸序列段之间的百分比互补性。

[0070] 术语“肽”、“多肽”和“蛋白”在本文中可互换地使用,且表示任何长度的氨基酸的聚合形式,其可以包括编码和非编码的氨基酸,化学或生物化学修饰或衍生的氨基酸,和具有修饰的肽主链的多肽。

[0071] 本文中使用的“结合”(例如,关于多肽的RNA-结合结构域)表示大分子之间(例如,蛋白与核酸之间)的非共价相互作用。当处于非共价相互作用状态时,大分子被认为是“缔合的”或“相互作用”或“结合”(例如,当分子X被认为与分子Y相互作用时,其是指分子X以非共价方式结合分子Y)。并非结合相互作用的所有组分都需要是序列特异性的(例如与DNA主链中的磷酸残基接触),但是结合相互作用的一些部分可能是序列特异性的。结合相互作用的特征通常在于小于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的解离常数(Kd)。“亲和力”表示结合强度,增加的结合亲和力与较低的Kd有关。

[0072] “结合结构域”是指能够非共价地结合另一分子的蛋白结构域。结合结构域可以结合至例如DNA分子(DNA-结合蛋白)、RNA分子(RNA-结合蛋白)和/或蛋白分子(蛋白-结合蛋白)。在蛋白结构域-结合蛋白的情况下,它可以结合它本身(以形成同二聚体、同三聚体等)和/或它可以结合一种或多种不同蛋白质的一个或多个分子。

[0073] 术语“保守的氨基酸置换”表示蛋白中具有相似侧链的氨基酸残基的可互换性。例如,具有脂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;具有含酰胺的侧链的一组氨基酸由天冬酰胺和谷氨酰胺组成;具有芳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和天冬氨酸组成;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。示例性的保守氨基酸置换组是: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-缬氨酸和天冬酰胺谷氨酰胺。

[0074] 多核苷酸或多肽与另一多核苷酸或多肽具有一定百分比的“序列同一性”,这意味着,当对比两个序列时,比对时所述百分比的碱基或氨基酸是相同的且在相同的相对位置。可以以多种不同方式确定序列同一性。为了确定序列同一性,可以使用在包括ncbi.nlm.nili.gov/BLAST、ebi.ac.uk/Tools/msa/tcoffee、ebi.Ac.Uk!Tools/msa/muscle、mafft.cbrc/alignment/software的网站在环球网上可获得的各种方法和计算机程序(例如BLAST、T-COFFEE、MUSCLE、MAFFT等)比对序列。参见,例如,Altschul等人(1990), J. Mol. Biol. 215:403-10。根据本发明,使用本领域中标准的序列比对来确定在一个sRGN变体中“对应于”另一个sRGN变体中的氨基酸残基的氨基酸残基。与其它sRGN变体的氨基酸残基相对应的sRGN变体的氨基酸残基出现在序列比对中的相同位置。

[0075] “编码”特定RNA的DNA序列是转录为RNA的DNA核酸序列。DNA多核苷酸可以编码被翻译成蛋白的RNA(mRNA),或者DNA多核苷酸可以编码不被翻译成蛋白质的RNA(例如,tRNA、

rRNA或指导RNA;也称为“非编码”RNA或“ncRNA”)。“蛋白编码序列”或编码特定蛋白或多肽的序列是当置于适当的调节序列的控制下时在体外或体外转录成mRNA(在DNA的情况下)并翻译(在mRNA的情况下)成多肽的核酸序列。编码序列的边界由在5’末端(N-末端)的起始密码子和在3’末端(C-末端)的翻译终止无义密码子确定。编码序列可以包括、但不限于,来自原核或真核mRNA的cDNA、来自原核或真核DNA的基因组DNA序列以及合成的核酸。转录终止序列通常位于编码序列的3’。

[0076] 本文中使用的“启动子序列”或“启动子”是能够结合RNA聚合酶并起始下游(3’方向)编码或非编码序列的转录的DNA调节区。为了限定本发明的目的,启动子序列在其3’末端处由转录起始位点结合,且向上游(5’方向)延伸以包括在背景以上可检测的水平起始转录所需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,以及负责RNA聚合酶结合的蛋白结合结构域。真核启动子经常但不总是含有“TATA”框和“CAAT”框。各种启动子,包括诱导型启动子,可用于驱动本发明的各种载体。启动子可以是组成活性的启动子(即,组成性地处于活性“打开”状态的启动子),其可以是诱导型启动子(即,其状态活性/“打开”或无活性/“关闭”由外界刺激例如特定温度、化合物或蛋白的存在控制的启动子),其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),并且其可以是时间限制的启动子(即,在胚胎发育的特定阶段过程中或生物过程的特定阶段(例如小鼠中的毛囊循环)过程中处于“打开”状态或“关闭”状态的启动子)。合适的启动子可以衍生自病毒,且因此可以被称为病毒启动子,或者它们可以衍生自任何生物,包括原核或真核生物。合适的启动子可用于通过任何RNA聚合酶(例如,pol I、pol II、pol Ill)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)启动子例如CMV立即早期启动子区域(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 Sep 1; 31(17)),人H1启动子(H1)等。诱导型启动子的例子包括、但不限于T7 RNA聚合酶启动子、T3 RNA聚合酶启动子、异丙基-β-D-硫代吡喃半乳糖苷(IPTG)调节的启动子、乳糖诱导的启动子、热激启动子、四环素调节的启动子、类固醇调节的启动子、金属调节的启动子、雌激素受体调节的启动子等。因此,诱导型启动子可以被包括、但不限于下述的分子调节:多西环素;RNA聚合酶,例如T7 RNA聚合酶;雌激素受体;雌激素受体融合体;等。

[0077] 在某些实施方案中,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),从而使得在多细胞生物中,所述启动子在特定细胞亚群中是有活性的(即,“打开”)。空间限制的启动子也可以称为增强子、转录控制元件、控制序列等。可以使用任何方便的空间限制的启动子,且合适的启动子(例如脑特异性启动子、驱动神经元亚群中表达的启动子、驱动种系中表达的启动子、驱动肺中表达的启动子、驱动肌肉中表达的启动子、驱动胰腺的胰岛细胞中表达的启动子等)的选择将取决于生物。例如,已知对于植物、蝇、蠕虫、哺乳动物、小鼠等的各种空间限制的启动子。因此,取决于生物,空间限制的启动子可用于调节在多种不同组织和细胞类型中编码sRGN多肽的核酸的表达。一些空间限制的启动子也是时间限制的,从而使得所述启动子在胚胎发育的特定阶段过程中或生物过程的特定阶段(例如小鼠中的毛囊循环)过程中处于“打开”状态或“关闭”状态。为了解释目的,

空间限制的启动子的例子包括、但不限于,神经元特异性启动子、脂肪细胞特异性启动子、心肌细胞特异性启动子、平滑肌特异性启动子、光感受器特异性启动子等。神经元特异性的空间限制的启动子包括、但不限于,神经元特异性烯醇化酶(NSE)启动子(例如,EMBL HSENO2, X51956);芳族氨基酸脱羧酶(AADC)启动子;神经丝启动子(例如,GenBank HUMNFL, L04147);突触蛋白启动子(例如,GenBank HUMSYNIB, M55301);thy-1启动子(例如,Chen等人(1987) Cell 51:7-19;和Llewellyn, 等人(2010) Nat. Med. 16(10):1161-1166);血清素受体启动子(例如,GenBank S62283);酪氨酸羟化酶启动子(TH) (例如,Oh等人(2009) Gene Ther. 16:437;Sasaoka等人(1992) Mol. Brain Res. 16:274;Boundy等人(1998) J. Neurosci. 18:9989;和Kaneda等人(1991) Neuron 6:583-594);GnRH启动子(例如,Radovick等人(1991) Proc. Natl. Acad. Sci. USA 88:3402-3406);L7启动子(例如,Oberdick等人(1990) Science 248:223-226);DNMT启动子(例如,Bartge等人(1988) Proc. Natl. Acad. Sci. USA 85:3648-3652);脑啡肽启动子(例如,Comb等人(1988) EMBO J. 17:3793-3805);髓磷脂碱蛋白(MBP)启动子;Ca2+-钙调蛋白依赖性蛋白激酶11-α(CamKIM)启动子(例如,Mayford等人(1996) Proc.Natl. Acad.Sci. USA 93:13250;和Casanova等人(2001) Genesis 31:37);CMV增强子/血小板-衍生的生长因子-p启动子(例如,Liu等人(2004) Gene Therapy 11:52-60);等。

[0078] 脂肪细胞特异性的空间限制的启动子包括、但不限于aP2基因启动子/增强子,例如人aP2基因从-5.4 kb到+21 bp的区域(例如,Tozzo等人(1997) Endocrinol. 138:1604;Ross等人(1990) Proc. Natl. Acad. Sci. USA 87:9590;和Pavjani等人(2005) Nat. Med. 11:797);葡萄糖转运蛋白-4 (GLUT4)启动子(例如,Knight等人(2003) Proc. Natl. Acad. Sci. USA 100:14725);脂肪酸移位酶(FAT/CD36)启动子(例如,Kuriki等人(2002) Biol. Pharm. Bull. 25:1476;和Sato等人(2002) J. Biol. Chem. 277:15703);硬脂酰辅酶A去饱和酶-1(SCD1)启动子(Tabor等人(1999) J. Biol. Chem. 274:20603);瘦素启动子(例如,Mason等人(1998) Endocrinol. 139:1013;和Chen等人(1999) Biochem. Biophys. Res. Comm. 262:187);脂联素启动子(例如,Kita等人(2005) Biochem. Biophys. Res. Comm. 331:484;和Chakrabarti (2010) Endocrinol. 151:2408);降脂蛋白启动子(例如,Platt等人(1989) Proc. Natl. Acad. Sci. USA 86:7490);抵抗素启动子(例如,Seo等人(2003) Molec. Endocrinol. 17:1522);等。

[0079] 心肌细胞特异性的空间限制的启动子包括、但不限于衍生自下述基因的控制序列:肌球蛋白轻链-2、a-肌球蛋白重链、AE3、心肌肌钙蛋白C、心肌动蛋白等。Franz等人(1997) Cardiovasc. Res. 35:560-566;Robbins等人(1995) Ann. N.Y. Acad. Sci. 752:492-505;Linn等人(1995) Circ. Res. 76:584591;Parmacek等人(1994) Mol. Cell. Biol. 14:1870-1885;Hunter等人(1993) Hypertension 22:608-617;和Sartorelli等人(1992) Proc. Natl. Acad. Sci. USA 89:4047-4051。

[0080] 平滑肌特异性的空间限制的启动子包括、但不限于SM22a启动子(例如,Akyiirek等人(2000) Mol. Med. 6:983;和美国专利号7,169,874);平滑肌特异性蛋白启动子(例如,WO 2001/018048);a-平滑肌肌动蛋白启动子;等。例如,已经证实两个CArG元件位于其中的SM22a启动子的0.4 kb区域介导血管平滑肌细胞特异性表达(例如,Kim, 等人(1997) Mol. Cell. Biol. 17, 2266-2278;Li, 等人, (1996) J. Cell Biol. 132, 849-859;和

Moessler, 等人(1996) Development 122, 2415-2425)。

[0081] 光感受器特异性的空间限制的启动子包括、但不限于视紫质启动子;视紫质激酶启动子(Young等人(2003) Ophthalmol. Vis. Sci. 44:4076);β磷酸二酯酶基因启动子(Nicoud等人(2007) J. Gene Med. 9:1015);色素性视网膜炎基因启动子(Nicoud等人(2007)出处同上);光感受器间类视色素-结合蛋白(IRBP)基因增强子(Nicoud等人(2007)出处同上);IRBP基因启动子(Yokoyama等人(1992) Exp Eye Res. 55:225);等。

[0082] 在本文中互换使用的术语"DNA调节序列"、"控制元件"和"调节元件"表示转录和翻译控制序列,诸如启动子、增强子、聚腺苷酸化信号、终止子、蛋白降解信号等,其提供和/或调节非编码序列(例如,指导RNA)或编码序列(例如,sRGN多肽)的转录和/或调节编码的多肽的翻译。

[0083] 本文中使用的术语"天然存在的"或"未修饰的"当应用于核酸、多肽、细胞或生物时,表示在自然界中发现的核酸、多肽、细胞或生物。例如,可从自然界中的来源分离且尚未在实验室中由人有意修饰的生物(包括病毒)中存在的多肽或多核苷酸序列是天然存在的。

[0084] 本文中使用的术语"嵌合的"当应用于核酸或多肽时表示由衍生自不同来源的结构组成的一个实体。例如,在嵌合多肽(例如,嵌合sRGN多肽)的上下文中使用"嵌合"的情况下,所述嵌合多肽包含衍生自不同多肽的氨基酸序列,诸如具有除Cas核酸酶以外的功能的多肽,包括例如具有DNA修饰活性、转录因子活性和/或DNA相关多肽修饰活性的多肽。嵌合多肽可包含修饰的或天然存在的多肽序列(例如,来自修饰的或未修饰的sRGN蛋白的第一氨基酸序列;和不同于sRGN蛋白的第二氨基酸序列)。类似地,在编码嵌合多肽的核酸的上下文中的"嵌合"包括衍生自不同编码区的核苷酸序列(例如,编码修饰的或未修饰的sRGN蛋白的第一核苷酸序列;以及编码除sRGN蛋白以外的多肽的第二核苷酸序列)。

[0085] 术语"嵌合多肽"表示非天然存在的多肽,例如通过人工干预将两个或更多个否则分离的氨基序列区段人工组合(即"融合")而制成。包含嵌合氨基酸序列的多肽是嵌合多肽。一些嵌合多肽可以被称作"融合变体"。

[0086] 本文中使用的"异源的"是指分别在天然核酸或蛋白质中未发现的核苷酸或肽。例如,在嵌合sRGN蛋白中,天然存在的细菌sRGN多肽(或其变体)的RNA结合结构域可以与异源多肽序列(即,来自除sRGN以外的蛋白的多肽序列,或来自另一生物的多肽序列)融合。异源多肽可以表现出也由嵌合sRGN蛋白表现出的活性(例如,酶活性)(例如,甲基转移酶活性、乙酰基转移酶活性、激酶活性、泛素化活性等)。异源核酸可以与天然存在的核酸(或其变体)连接(例如,通过基因工程)以产生编码嵌合多肽的嵌合核酸。作为另一个例子,在融合变体sRGN定位多肽中,可以将变体sRGN定位多肽与异源多肽(即,除sRGN以外的多肽)融合,该异源多肽表现出也将由融合变体sRGN定位多肽表现出的活性。异源核酸可与变体sRGN定位多肽连接(例如,通过基因工程)以产生编码融合变体sRGN定位多肽的核酸。本文中使用的"异源的"另外是指在并非其天然细胞的细胞中的核苷酸或多肽。

[0087] 术语"同族(cognate)"表示通常在自然界中相互作用或共存的两个生物分子。

[0088] 本文中使用的"重组"意指,特定的核酸(DNA或RNA)或载体是克隆、限制、聚合酶链式反应(PCR)和/或连接步骤的各种组合的产物,从而产生具有与天然系统中发现的内源核酸可辨别的结构编码或非编码序列的构建体。可以从cDNA片段或从一系列合成的寡核苷酸装配编码多肽的DNA序列,以提供能够从细胞或无细胞转录和翻译系统中含有的重组转录

单位表达的合成核酸。包含相关序列的基因组DNA也可以用于重组基因或转录单位的形成。非翻译DNA的序列可以存在于开放读码框的5'或3',在那里这种序列不干扰编码区的操作或表达,并且确实可以起作用以通过各种机制来调节所期望的产物的产生(参见下面的"DNA调节序列")。可替换地,编码未被翻译的RNA(例如,指导RNA)的DNA序列也可以被认为是重组的。因而,例如,术语"重组"核酸表示非天然存在的核酸,例如,通过人干预将两个否则分开的序列区段人工组合而制备。通过化学合成手段或通过人工操作分离的核酸区段,例如通过基因工程技术,可以实现这种人工组合。通常这样做以用编码相同氨基酸、保守氨基酸或非保守氨基酸的密码子置换密码子。可替换地,执行它以将具有所期望的功能的核酸区段连接在一起以产生所期望的功能的组合。通过化学合成手段或通过人工操作分离的核酸区段,例如通过基因工程技术,可以实现这种人工组合。当重组多核苷酸编码多肽时,编码的多肽的序列可以是天然存在的("野生型"),或可以是天然存在的序列的变体(例如,突变体)。因此,术语"重组"多肽不一定表示其序列不天然存在的多肽。相反,"重组"多肽由重组DNA序列编码,但是所述多肽的序列可以是天然存在的("野生型")或非天然存在的(例如,变体、突变体等)。因此,"重组"多肽是人干预的结果,但可以是天然存在的氨基酸序列。术语"非天然存在的"包括与它们天然存在的对应物明显不同的分子,包括化学修饰的或突变的分子。

[0089] "载体"或"表达载体"是复制子,诸如质粒、噬菌体、病毒或粘粒,可以在其上附着另一种DNA区段,即,"插入片段",以便实现附着的区段在细胞中的复制。

[0090] "表达盒"包含可操作地连接至启动子的DNA编码序列。"可操作地连接的"表示并列连接,其中如此描述的组分处于允许它们以其预期方式起作用的关系中。例如,如果启动子影响编码序列的转录或表达,则所述启动子可操作地连接至所述编码序列。术语"重组表达载体"或"DNA构建体"在本文中互换地用于表示包含载体和至少一个插入片段的DNA分子。重组表达载体通常是为了表达和/或扩增插入片段的目的或为了构建其它重组核苷酸序列而产生。核酸可以与或不与启动子序列可操作地连接,并且可以与或不与DNA调节序列可操作地连接。

[0091] 当外源DNA例如重组表达载体已经被引入细胞内时,细胞已被所述DNA"遗传上修饰"或"转化"或"转染"。外源DNA的存在导致永久或短暂的遗传改变。转化DNA可以或可以不整合(共价连接)到细胞的基因组中。

[0092] 例如,在原核生物、酵母和哺乳动物细胞中,转化DNA可以维持在附加型元件例如质粒上。就真核细胞而言,稳定转化的细胞是这样的细胞:其中转化DNA已经变成整合到染色体内,使得所述转化DNA通过染色体复制由子代细胞继承。通过真核细胞确立包含含有转化DNA的子细胞群体的细胞系或克隆的能力证明了这种稳定性。"克隆"是通过有丝分裂从单个细胞或共同祖先衍生出的细胞群体。"细胞系"是能够在体外稳定生长多代的原代细胞克隆。

[0093] 合适的遗传修饰方法(也称为"转化")包括、但不限于,例如病毒或噬菌体感染、转染、接合、原生质体融合、脂转染、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、基因枪技术、磷酸钙沉淀、直接显微注射、纳米颗粒介导的核酸递送(例如,Panyam等人, Adv Drug Deliv Rev. 2012年9月13日. pp: S0169-409X(12)00283-9. doi:10.1016/j.addr.2012.09.023)等。

[0094] 遗传修饰方法的选择通常取决于被转化的细胞的类型和发生转化的环境(例如,体外、离体或体内)。这些方法的一般讨论可以在Ausubel,等人, Short Protocols in Molecular Biology, 第3版, Wiley & Sons, 1995中找到。

[0095] 本文中使用的"宿主细胞"表示体内或体外的真核细胞、原核细胞(例如细菌或古细菌细胞)或来自作为单细胞实体培养的多细胞生物(例如细胞系)的细胞,所述真核或原核细胞可以或已经被用作核酸的接受者,并包括已经被核酸转化的原始细胞的后代。应当理解,由于自然、偶然或故意的突变,单个细胞的后代在形态学或基因组或总DNA互补体中不必定与原始亲本完全相同。"重组宿主细胞"(也被称作"遗传修饰的宿主细胞")是已在其中引入了异源核酸(例如表达载体)的宿主细胞。例如,细菌宿主细胞是借助于外源核酸(例如,质粒或重组表达载体)向合适的细菌宿主细胞中的引入而遗传修饰的细菌宿主细胞,而真核宿主细胞是借助于外源核酸向合适的真核宿主细胞中的引入而遗传修饰的真核宿主细胞(例如哺乳动物生殖细胞)。

[0096] 本文中使用的"靶DNA"是包含"靶位点"或"靶序列"的DNA多核苷酸。术语"靶位点"、"靶序列"、"靶原间隔区DNA"或"原间隔区样序列"在本文中互换地用于表示存在于靶DNA中的核酸序列,如果存在足够的用于结合的条件,则指导RNA的DNA-靶向区段将与其结合。例如,靶DNA内的靶位点(或靶序列)5'GAGCATATC-3'被RNA序列5'-GAUAUGCUC-3'靶向(或由其结合,或与其杂交,或与其互补)。合适的DNA/RNA结合条件包括在细胞中通常存在的生理条件。其它合适的DNA/RNA结合条件(例如,无细胞系统中的条件)是本领域已知的;例如,Sambrook, 出处同上。与指导RNA互补并与其杂交的靶DNA的链被称为"互补链",并且与"互补链"互补(且因此与指导RNA不互补)的靶DNA的链被称为"非互补链"或"非互补链"。"定位修饰多肽"或"RNA-结合位点指导的多肽"或"RNA-结合位点指导的修饰多肽"或"定位多肽"是指结合RNA并靶向至特定DNA序列的多肽。如本文所述的定位修饰多肽被与其结合的RNA分子靶向至特定DNA序列。RNA分子包含与靶DNA内的靶序列结合、杂交或互补的序列,从而将结合的多肽靶向至靶DNA内的特定位置(靶序列)。"切割"是指DNA分子的共价主链的断裂。切割可通过多种方法起始,包括、但不限于磷酸二酯键的酶促或化学水解。单链切割和双链切割都是可能的,并且双链切割可以作为两个不同的单链切割事件的结果发生。DNA切割可导致平头末端或交错末端的产生。在某些实施方案中,将包含指导RNA和定位修饰多肽的复合物用于靶向的双链DNA切割。

[0097] "核酸酶"和"内切核酸酶"在本文中互换地用于指具有用于多核苷酸切割的内切核酸降解(endonucleolytic)催化活性的酶。

[0098] 核酸酶的"切割结构域"或"活性结构域"或"核酸酶结构域"是指在核酸酶内具有用于DNA切割的催化活性的多肽序列或结构域。切割结构域可以包含在单个多肽链中,或者切割活性可以由两个(或更多个)多肽的缔合产生。单个核酸酶结构域可以由给定多肽内的超过一个分离的氨基酸序列段组成。

[0099] "定位多肽"或"RNA-结合位点指导的多肽"或"RNA-结合位点指导的多肽"是指结合RNA并靶向至特定DNA序列的多肽。如本文所述的定位多肽被与其结合的RNA分子靶向至特定DNA序列。RNA分子包含与靶DNA内的靶序列互补的序列,从而将结合的多肽靶向至靶DNA内的特定位置(靶序列)。

[0100] "指导序列"或DNA-靶向区段(或"DNA-靶向序列")包含与在本文中称为"原间隔区

样”序列的靶DNA内的特定序列互补的核苷酸序列(靶DNA的互补链)。蛋白结合区段(或“蛋白结合序列”)与定位修饰多肽相互作用。当定位修饰多肽是sRGN或sRGN相关多肽(在下面更详细地描述)时,靶DNA的位点特异性切割发生在由(i)指导RNA和靶DNA之间的碱基配对互补性;和(ii)靶DNA中的短基序(称为原间隔区相邻基序(PAM))两者确定的位置。

[0101] 指导RNA的蛋白结合区段部分地包含两个互补的核苷酸序列段,其彼此杂交以形成双链RNA双链体(dsRNA双链体)。

[0102] 在某些实施方案中,核酸(例如,指导RNA,包含编码指导RNA的核苷酸序列的核酸;编码定位多肽的核酸;等等)包含提供额外的所期望的特征(例如,修饰或调节的稳定性;亚细胞靶向;追踪,例如荧光标记;蛋白或蛋白复合物的结合位点;等)的修饰或序列。非限制性例子包括:5’帽(例如7-甲基鸟苷酸帽(m7G));3’聚腺苷酸化尾(即,3’聚腺苷酸尾);核糖开关序列(例如以允许蛋白和/或蛋白复合物调节的稳定性和/或调节的可接近性);稳定性控制序列;形成dsRNA双链体(即发夹)的序列);将RNA靶向至亚细胞位置(例如细胞核、线粒体、叶绿体等)的修饰或序列;提供追踪的修饰或序列(例如直接缀合至荧光分子,缀合至促进荧光检测的部分,允许荧光检测的序列等);提供蛋白的结合位点的修饰或序列(例如作用于DNA的蛋白,包括转录激活因子、转录阻遏蛋白、DNA甲基转移酶、DNA脱甲基酶、组蛋白乙酰转移酶、组蛋白脱乙酰酶等);及其组合。

[0103] 在某些实施方案中,指导RNA在5’或3’端包含额外的区段,其提供上述特征的任一种。例如,合适的第三区段可以包括5’帽(例如7-甲基鸟苷酸帽(m7G));3’聚腺苷酸化尾(即,3’聚腺苷酸尾);核糖开关序列(例如以允许蛋白和蛋白复合物调节的稳定性和/或调节的可接近性);稳定性控制序列;形成dsRNA双链体(即发夹)的序列);将RNA靶向亚细胞位置(例如细胞核、线粒体、叶绿体等)的序列;提供追踪的修饰或序列(例如直接缀合至荧光分子,缀合至促进荧光检测的部分,允许荧光检测的序列等);提供蛋白的结合位点的修饰或序列(例如作用于DNA的蛋白,包括转录激活因子、转录阻遏蛋白、DNA甲基转移酶、DNA脱甲基酶、组蛋白乙酰转移酶、组蛋白脱乙酰酶等);及其组合。

[0104] 指导RNA和sRGN多肽形成复合物(即,通过非共价相互作用结合)。指导RNA通过包含与靶DNA的序列互补的核苷酸序列,为复合物提供靶特异性。所述复合物的sRGN多肽提供位点特异性活性。换句话说,sRGN多肽借助于其与指导RNA的蛋白结合区段的缔合被指导至靶DNA序列(例如,染色体核酸中的靶序列;染色体外核酸中的靶序列,例如附加型核酸、小环等;线粒体核酸中的靶序列;叶绿体核酸中的靶序列;质粒中的靶序列;等等)。RNA适体是本领域已知的,并且通常是核糖开关的合成形式。术语“RNA适体”和“核糖开关”在本文中互换地用于包括合成的和天然的核酸序列,其提供它们是其一部分的RNA分子的结构的可诱导的调节(以及因此特定序列的可用性)。RNA适体通常包含折叠成特定结构(例如发夹)的序列,其特异性地结合特定药物(例如小分子)。药物的结合导致RNA折叠中的结构变化,这改变了适体是其一部分的核酸的特征。作为非限制性例子:(i)具有适体的活化子-RNA可能不能与相应靶向子-RNA结合,除非适体被适当的药物结合;(ii)具有适体的靶向子-RNA可能不能与相应活化子-RNA结合,除非适体被适当的药物结合;和(iii)各自包含结合不同药物的不同适体的靶向子-RNA和活化子-RNA可能不能彼此结合,除非同时存在两种药物。如这些实例所举例说明的,可以对两分子指导RNA进行设计以为可诱导的。

[0105] 适体和核糖开关的例子可在例如:Nakamura等人, Genes Cells. 2012 May;17

(5):344-64;Vavalle等人, Future Cardiol. 2012年5月;8(3):371-82;Citartan等人, Biosens Bioelectron. 2012年4月15日;34(1):1-11;和Liberman等人, Wiley lnterdiscip Rev RNA. 2012年5月-6月;3(3):369-84中找到;它们都通过引用整体并入本文。

[0106] 术语“干细胞”在本文中用于表示既具有自我更新能力又具有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等人(1997) Cell 88:287-298)。在细胞个体发育的上下文中,形容词“分化的”或“分化”是相对术语。“分化的细胞”是比与其进行比较的细胞已进一步向发育途径下游发展的细胞。因此,多潜能干细胞(如下所述)可以分化为谱系限制的祖先细胞(例如中胚层干细胞),后者又可以分化为进一步限制的细胞(例如神经元祖先细胞),所述细胞可以分化为终末阶段细胞(即,终末分化的细胞,例如神经元、心肌细胞等),它们在某种组织类型中起特有的作用,并且可能或可能不保留进一步增殖的能力。干细胞可以通过存在特定标志物(例如蛋白、RNA等)和不存在特定标志物来表征。干细胞还可以通过体外和体内功能测定来鉴定,特别是与干细胞产生多种分化的后代的能力有关的测定。

[0107] 感兴趣的干细胞包括多潜能干细胞(PSC)。术语“多潜能干细胞”或“PSC”在本文中用于表示能够产生生物的所有细胞类型的干细胞。因此,PSC可以产生生物的所有胚层的细胞(例如脊椎动物的内胚层、中胚层和外胚层)。多潜能细胞能够形成畸胎瘤并能够影响活生物中的外胚层、中胚层或内胚层组织。植物的多潜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)。

[0108] 动物的PSC可以以多种不同方式获得。例如,胚胎干细胞(ESC)衍生自胚胎的内细胞团(Thomson等人, Science.1998年11月6日;282(5391):1145-7),而诱导的多潜能干细胞(iPSC)衍生自体细胞(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 2007年11月20日)。

[0109] 因为术语PSC表示多潜能干细胞,而不管它们的起源,所以术语PSC包括术语ESC和iPSC,以及术语胚胎生殖干细胞(EGSC),其是PSC的另一个实例。PSC可以是确立的细胞系的形式,它们可以直接从原代胚胎组织获得,或它们可以从体细胞获得。PSC可以是本文所述方法的靶细胞。

[0110] “胚胎干细胞”(ESC)是指从胚胎分离的PSC,通常从胚泡的内细胞团分离。ESC在NIH Human Embryonic Stem Cell Registry中列举,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 (MizMedi Hospital-Seoul National University);HSF-1、HSF-6 (University of California at San Francisco);和H1、H7、H9、H13、H14(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,诸如恒河猴干细胞和绒猴干细胞。干细胞可获得自任何哺乳动物物种,例如人,马科动物,牛科动物,猪,犬科动物,猫科动物,啮齿类动物,例如,小鼠、大鼠、仓鼠、灵长类动物等(Thomson等人(1998) Science 282:1145;Thomson等人(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通常作为具有大的核质比例、明确的边缘和突出的核仁的扁平集落生长。此外,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的方法的例子可以在例如美国专利号7,029,913、美国专利号5,843,780和美国专利号6,200,806中找到,其公开内容通过引用并入本文。在WO 99/20741、WO 01/51616和WO 03/020920中描述了用于以未分化的形式增殖hESC的方法。"胚胎生殖干细胞"(EGSC)或"胚胎生殖细胞"或"EG细胞"是指PSC,其衍生自生殖细胞和/或生殖细胞祖先,例如原始生殖细胞,即,将变为精子和卵的那些。胚胎生殖细胞(EG细胞)被认为具有与如上所述的胚胎干细胞相似的性质。产生和表征EG细胞的方法的例子可以在,例如,美国专利号7,153,684;Matsui, Y., 等人, (1992) Cell 70:841;Shamblott, M., 等人(2001) Proc. Natl. Acad. Sci. USA 98: 113;Shamblott, M., 等人(1998) Proc. Natl. Acad. Sci. USA, 95:13726;和Koshimizu, U., 等人(1996) Development, 122:1235中找到,其公开内容通过引用并入本文。

[0111] "诱导的多潜能干细胞"或"iPSC"是指从非PSC的细胞(即,从相对于PSC分化的细胞)衍生的PSC。iPSC可以衍生自多种不同的细胞类型,包括终末分化的细胞。iPSC具有ES细胞样形态学,作为具有大的核质比例、明确的边缘和突出的核仁的扁平集落生长。此外,iPSC表达一种或多种本领域普通技术人员已知的关键多潜能性标记,包括、但不限于碱性磷酸酶、SSEA3、SSEA4、Sox2、Oct3/4、Nanog、TRA160、TRA181、TDGF 1、Dnmt3b、Fox03、GDF3、Cyp26al、TERT和zfp42。

[0112] 产生和表征iPSC的方法的例子可以在,例如,美国专利公开号US20090047263、US20090068742、US20090191159、US20090227032、US20090246875和US20090304646中找到,其公开内容通过引用并入本文。通常,为了产生iPSC,向体细胞提供本领域已知的重新程序化因子(例如,Oct4、SOX2、KLF4、MYC、Nanog、Lin28等)以将体细胞重新程序化以变为多潜能干细胞。

[0113] "体细胞"是指在生物中在没有实验操作的情况下通常不在生物中产生所有类型的细胞的任何细胞。换句话说,体细胞是已经充分分化的细胞,它们将不自然地产生身体的所有三个胚层(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先,后者可能能够自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。

[0114] "有丝分裂细胞"是指经历有丝分裂的细胞。有丝分裂是真核细胞将其细胞核中的染色体分裂成两个独立的细胞核中的两个相同组的过程。其后面通常紧接着胞质分裂,其将细胞核、细胞质、细胞器和细胞膜分裂为两个含有大致相等份额的这些细胞组分的细胞。

[0115] "有丝分裂后细胞"是指已经从有丝分裂中退出的细胞,即,它是"休眠的",即它不再经历分裂。这个休眠状态可以是暂时的,即,可逆的,或其可以是永久的。

[0116] "减数分裂细胞"是指经历减数分裂的细胞。减数分裂是细胞为了产生配子或孢子而分裂其核物质的过程。与有丝分裂不同,在减数分裂中,染色体经历重组步骤,其在染色体之间改组遗传物质。此外,与从有丝分裂产生的两个(遗传上相同的)二倍体细胞相比,减数分裂的结果是四个(遗传上独特的)单倍体细胞。

[0117] "重组"是指在两个多核苷酸之间交换遗传信息的过程。本文中使用的"同源性指导的修复(HDR)"表示例如在细胞中双链断裂修复期间发生的特殊化形式的DNA修复。这个

过程需要核苷酸序列同源性,使用“供体”分子为“靶”分子(即,经历双链断裂的分子)的修复提供模板,并导致遗传信息从供体转移至靶。如果供体多核苷酸不同于靶分子,并且供体多核苷酸的部分或全部序列被并入到靶DNA中,则同源性指导的修复可能导致靶分子序列的改变(例如插入、缺失、突变)。在某些实施方案中,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。

[0118] “非同源末端连接(NHEJ)”是指在不需要同源模板的情况下通过将断裂末端彼此直接连接来修复DNA中的双链断裂(与需要同源序列以指导修复的同源性指导的修复形成对照)。NHEJ经常导致双链断裂位点附近的核苷酸序列丧失(缺失)。

[0119] 术语“治疗(treatment)”、“治疗(treating)”等在本文中通常用于指获得所期望的药理和/或生理效果。所述效果就完全或部分预防疾病或其症状而论,可以是预防性的,和/或就对于疾病和/或可归因于所述疾病的副作用的部分或完全治愈而论,可以是治疗性的。本文中使用的“治疗”包括哺乳动物中疾病或症状的任何治疗,并且包括:(a)防止所述疾病或症状在可能倾向于获得所述疾病或症状但尚未被诊断为具有所述疾病或症状的受试者中发生;(b)抑制疾病或症状,即,阻止其发展;或(c)减轻疾病,即,导致疾病消退。可以在疾病或损伤发作之前、期间或之后施用治疗剂。特别重要的事是正在进行的疾病的治疗,其中所述治疗稳定或减轻了患者的不期望的临床症状。期望在受影响的组织中完全丧失功能之前进行这种治疗。将期望在疾病的有症状阶段期间以及在某些情况下在疾病的有症状阶段之后施用治疗。

[0120] 术语“个体”、“受试者”、“宿主”和“患者”在本文中可互换地使用,并且表示期望对其进行诊断、治疗或疗法的任何哺乳动物受试者,特别是人。

[0121] 分子和细胞生物化学中的一般方法可以在像下述那样的这种标准教科书中找到:Molecular Cloning: A Laboratory Manual, 第3版(Sambrook等人, Harboor 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 & Loewy编., Academic Press 1995);Immunology Methods Manual (1. Lefkovits编., Academic Press 1997);和Cell and Tissue Culture: Laboratory Procedures in Biotechnology (Doyle & Griffiths, John Wiley & Sons 1998),其公开内容通过引用并入本文。

[0122] 在提供了值的范围的情况下,应当理解,介于该范围的上限与下限之间的每个居间值(直至下限单位的十分之一,除非上下文另外清楚地指明)以及该所述范围内的任何其它所述值或居间值均被包含在本发明内。这些较小范围的上限和下限可以独立地被包含在所述较小范围内,并且也被包含在本发明内,受所述范围内任何具体排除的限值的约束。在所述范围包括一个或两个限值的情况下,排除那些被包括的限值中的任一个或两个限值的范围也被包括在本发明中。

[0123] 短语“基本上由……组成”在本文中是指排除不是系统的一种或多种指定活性组分或不是分子的一种或多种指定活性部分的一切。

[0124] 在本文呈现的某些范围的数值前面存在术语“约”。术语“约”在本文中用于为其后面的确切数字以及在该术语后面的数字附近或靠近的数字提供字面支持。在确定数字是否

接近或靠近特别提及的数字时,接近或者靠近的未提及的数字可能是这样的数字:该数字在呈现它的内容中提供了特别提及的数字的实质等同数字。

[0125] 应当理解,为了清楚而在单独的实施方案的上下文中描述的本发明的某些特征还可被组合提供在单个实施方案中。相反,为了简洁而在单个实施方案的上下文中描述的本发明的多种特征也可以单独地或以任何合适的子组合(sub-combination)来提供。与本发明有关的实施方案的所有组合都被本发明特别地包括,并且在本文中公开,正如每个和每种组合被个别地和明确地公开一样。此外,各种实施方案及其要素的所有子组合也都被本发明特别地包括,并且在本文中公开,正如每个和每种这样的子组合被在本文中个别地和明确地公开一样。

[0126] 合成的RNA-指导的核酸酶(sRGN)

[0127] 可以产生合成的RNA-指导的核酸酶(sRGN),例如,使用不同物种的CRISPR-Cas9内切核酸酶的人工定义的小结构域的基于同源性的基因家族改组,例如,四个不同葡萄球菌属种(路邓葡萄球菌、巴氏葡萄球菌、*Staphylococcus microti*和猪葡萄球菌)。在图1中示意性地描绘了一种这样的基因家族改组方案。1.简而言之,比较不同Cas9内切核酸酶的序列以鉴定具有高度同源性的区域作为锚点,将每个Cas9内切核酸酶分解成多个(诸如5、6、7、8、9、10、11、12、13、14、15或更多)相应的小结构域。通过随机分配要衍生自原始Cas9内切核酸酶之一的每个小结构域,制备基因家族改组的sRGN的文库。在某些情况下,可能合乎需要的是,保持PAM-相互作用(PI)结构域固定,以使单个PAM序列被每个文库构建体识别。例如,通过将每个Cas9内切核酸酶分解为8个小结构域可以制备文库,保持C-端小结构域固定,从而导致8192的理论复杂性。在另一个实施例中,通过将每个Cas9内切核酸酶分成12个小结构域可以制备文库,保持C-端小结构域固定,从而导致1.3 x 10^5的理论复杂性。使用已知的针对RNA-指导的内切核酸酶活性的测定法(例如细菌活/死测定法和报道基因破坏测定法)可以筛选文库,以鉴定具有高活性的sRGN。

[0128] 在一个方面,本文提供了根据基于同源性的基因家族改组方案设计的候选sRGN。尚未证明候选sRGN具有任何活性。还提供了已证明具有至少一种活性的sRGN。在某些实施方案中,至少一种活性包括例如在切割测定(诸如基于荧光偏振的生化切割测定)、基因编辑测定(诸如从GFP至BFP的转化)(Glaser等人, 2016,见下文)、基因插入测定或其它合适参数的测定中的活性。在某些实施方案中,测定靶上切割,测定脱靶切割,和/或测定靶上编辑(例如,突变的校正或基因或基因的一部分的插入),其可以是与脱靶编辑相比。在某些情况下,将活性与参考或对照核酸酶(诸如野生型Cas9)进行比较。测定是本领域中描述的并且可从商业资源获得。可以在生物体中例如具有有害突变的动物中进行合适的测定,并且通过与施用了sRGN系统的动物中与突变相关的一种或多种症状的改善来测定sRGN的功效。在某些实施方案中,至少一种活性大于从其衍生出sRGN的一种或多种亲本Cas9内切核酸酶中的相应活性。在某些实施方案中,至少一种活性大于另一种野生型Cas9诸如SluCas9中的相应活性。

[0129] 在某些实施方案中,本文提供了包含8个小结构域的候选sRGN,其中所述小结构域从N-端至C-端为1)小结构域1,其包含SEQ ID NO: 34-37中的任一个的氨基酸序列或其与SEQ ID NO: 34-37中的任一个具有至少约90%(诸如至少约91%、92%、93%、94%、95%、96%、97%、98%、99%或更多中的任一种)序列同一性的变体,2)小结构域2,其包含SEQ ID NO: 38-

41中的任一个的氨基酸序列或其与SEQ ID NO: 38-41中的任一个具有至少约90%(诸如至少约91%、92%、93%、94%、95%、96%、97%、98%、99%或更多中的任一种)序列同一性的变体，3)小结构域3,其包含SEQ ID NO: 42-45中的任一个的氨基酸序列或其与SEQ ID NO: 42-45中的任一个具有至少约90%(诸如至少约91%、92%、93%、94%、95%、96%、97%、98%、99%或更多中的任一种)序列同一性的变体，4)小结构域4,其包含SEQ ID NO: 46-49中的任一个的氨基酸序列或其与SEQ ID NO: 46-49中的任一个具有至少约90%(诸如至少约91%、92%、93%、94%、95%、96%、97%、98%、99%或更多中的任一种)序列同一性的变体，5)小结构域5,其包含SEQ ID NO: 50-53中的任一个的氨基酸序列或其与SEQ ID NO: 50-53中的任一个具有至少约90%(诸如至少约91%、92%、93%、94%、95%、96%、97%、98%、99%或更多中的任一种)序列同一性的变体，6)小结构域6,其包含SEQ ID NO: 54-57中的任一个的氨基酸序列或其与SEQ ID NO: 54-57中的任一个具有至少约90%(诸如至少约91%、92%、93%、94%、95%、96%、97%、98%、99%或更多中的任一种)序列同一性的变体，7)小结构域7,其包含SEQ ID NO: 58-61中的任一个的氨基酸序列或其与SEQ ID NO: 58-61中的任一个具有至少约90%(诸如至少约91%、92%、93%、94%、95%、96%、97%、98%、99%或更多中的任一种)序列同一性的变体,和8)小结构域8,其包含SEQ ID NO: 62-65中的任一个的氨基酸序列或其与SEQ ID NO: 62-65中的任一个具有至少约90%(诸如至少约91%、92%、93%、94%、95%、96%、97%、98%、99%或更多中的任一种)序列同一性的变体。在某些实施方案中,所述候选sRGN包含来自路邓葡萄球菌、巴氏葡萄球菌、*Staphylococcus microti*和猪葡萄球菌Cas9的小结构域。在某些实施方案中,所述候选sRGN包含来自路邓葡萄球菌、巴氏葡萄球菌、*Staphylococcus microti*和猪葡萄球菌中的小于所有4种(诸如2或3种) Cas9的小结构域。在某些实施方案中,小结构域8包含SEQ ID NO: 62的氨基酸序列或其与SEQ ID NO: 62具有至少约90%(诸如至少约91%、92%、93%、94%、95%、96%、97%、98%、99%或更多中的任一种)序列同一性的变体。在某些实施方案中,小结构域8包含SEQ ID NO: 63的氨基酸序列或其与SEQ ID NO: 63具有至少约90%(诸如至少约91%、92%、93%、94%、95%、96%、97%、98%、99%或更多中的任一种)序列同一性的变体。在某些实施方案中,小结构域1包含SEQ ID NO: 34的氨基酸序列或其与SEQ ID NO: 34具有至少约90%(诸如至少约91%、92%、93%、94%、95%、96%、97%、98%、99%或更多中的任一种)序列同一性的变体。

[0130] 在某些实施方案中,本文提供了包含12个小结构域的候选sRGN,其中所述小结构域从N-端至C-端为1)小结构域1,其包含SEQ ID NO: 66-69中的任一个的氨基酸序列或其与SEQ ID NO: 66-69中的任一个具有至少约90%(诸如至少约91%、92%、93%、94%、95%、96%、97%、98%、99%或更多中的任一种)序列同一性的变体，2)小结构域2,其包含SEQ ID NO: 70-73中的任一个的氨基酸序列或其与SEQ ID NO: 70-73中的任一个具有至少约90%(诸如至少约91%、92%、93%、94%、95%、96%、97%、98%、99%或更多中的任一种)序列同一性的变体，3)小结构域3,其包含SEQ ID NO: 74-77中的任一个的氨基酸序列或其与SEQ ID NO: 74-77中的任一个具有至少约90%(诸如至少约91%、92%、93%、94%、95%、96%、97%、98%、99%或更多中的任一种)序列同一性的变体，4)小结构域4,其包含SEQ ID NO: 78-81中的任一个的氨基酸序列或其与SEQ ID NO: 78-81中的任一个具有至少约90%(诸如至少约91%、92%、93%、94%、95%、96%、97%、98%、99%或更多中的任一种)序列同一性的变体，5)小结构域5,其包含SEQ ID NO: 82-85中的任一个的氨基酸序列或其与SEQ ID NO: 82-85中的任一个具有至

少约90%(诸如至少约91%、92%、93%、94%、95%、96%、97%、98%、99%或更多中的任一种)序列同一性的变体，6)小结构域6,其包含SEQ ID NO: 86-89中的任一个的氨基酸序列或其与SEQ ID NO: 86-89中的任一个具有至少约90%(诸如至少约91%、92%、93%、94%、95%、96%、97%、98%、99%或更多中的任一种)序列同一性的变体，7)小结构域7,其包含SEQ ID NO: 90-93中的任一个的氨基酸序列或其与SEQ ID NO: 90-93中的任一个具有至少约90%(诸如至少约91%、92%、93%、94%、95%、96%、97%、98%、99%或更多中的任一种)序列同一性的变体，8)小结构域8,其包含SEQ ID NO: 94-97中的任一个的氨基酸序列或其与SEQ ID NO: 94-97中的任一个具有至少约90%(诸如至少约91%、92%、93%、94%、95%、96%、97%、98%、99%或更多中的任一种)序列同一性的变体，9)小结构域9,其包含SEQ ID NO: 98-101中的任一个的氨基酸序列或其与SEQ ID NO: 98-101中的任一个具有至少约90%(诸如至少约91%、92%、93%、94%、95%、96%、97%、98%、99%或更多中的任一种)序列同一性的变体，10)小结构域10,其包含SEQ ID NO: 102-105中的任一个的氨基酸序列或其与SEQ ID NO: 102-105中的任一个具有至少约90%(诸如至少约91%、92%、93%、94%、95%、96%、97%、98%、99%或更多中的任一种)序列同一性的变体，11)小结构域11,其包含SEQ ID NO: 106-109中的任一个的氨基酸序列或其与SEQ ID NO: 106-109中的任一个具有至少约90%(诸如至少约91%、92%、93%、94%、95%、96%、97%、98%、99%或更多中的任一种)序列同一性的变体,和12)小结构域12,其包含SEQ ID NO: 110-113中的任一个的氨基酸序列或其与SEQ ID NO: 110-113中的任一个具有至少约90%(诸如至少约91%、92%、93%、94%、95%、96%、97%、98%、99%或更多中的任一种)序列同一性的变体。在某些实施方案中,所述候选sRGN包含来自路邓葡萄球菌、巴氏葡萄球菌、*Staphylococcus microti*和猪葡萄球菌Cas9的小结构域。在某些实施方案中,所述候选sRGN包含来自路邓葡萄球菌、巴氏葡萄球菌、*Staphylococcus microti*和猪葡萄球菌的小于所有4种(诸如2或3种) Cas9的小结构域。在某些实施方案中,小结构域12包含SEQ ID NO: 110的氨基酸序列或其与SEQ ID NO: 110具有至少约90%(诸如至少约91%、92%、93%、94%、95%、96%、97%、98%、99%或更多中的任一种)序列同一性的变体。在某些实施方案中,小结构域12包含SEQ ID NO: 111的氨基酸序列或其与SEQ ID NO: 111具有至少约90%(诸如至少约91%、92%、93%、94%、95%、96%、97%、98%、99%或更多中的任一种)序列同一性的变体。在某些实施方案中,(i)小结构域1包含SEQ ID NO: 66的氨基酸序列或其与SEQ ID NO: 66具有至少约90%(诸如至少约91%、92%、93%、94%、95%、96%、97%、98%、99%或更多中的任一种)序列同一性的变体和(ii)小结构域2包含SEQ ID NO: 70的氨基酸序列或其与SEQ ID NO: 70具有至少约90%(诸如至少约91%、92%、93%、94%、95%、96%、97%、98%、99%或更多中的任一种)序列同一性的变体。

[0131] 使用这样的基因家族改组方案鉴定了本文中公开的sRGN多肽Gib11 (SEQ ID NO: 1)、P2H12 (SEQ ID NO: 5)、E2 (SEQ ID NO: 6)和F8 (SEQ ID NO: 10),并通过用巴氏葡萄球菌Cas9的相应部分替换Gib11的不同C末端部分,产生了Gib11Spa变体(SEQ ID NO: 2-4)。

[0132] 本文公开的sRGN多肽与来自酿脓链球菌的Cas9(在本文中也被称作“SpCas9”或“SpyCas9”)相比更小,并且与NNGG PAM序列相互作用。

[0133] 在一个方面,本文提供了一种sRGN多肽,其包含SEQ ID NO: 1-12中的任一个的氨基酸序列或其与SEQ ID NO: 1-12中的任一个具有至少约90%(诸如至少约91%、92%、93%、

94%、95%、96%、97%、98%、99%或更多中的任一种)序列同一性的变体。在某些实施方案中,所述sRGN多肽包含SEQ ID NO: 1-12中的任一个的氨基酸序列。

[0134] 在某些实施方案中,本文提供了一种sRGN多肽,其包含与SEQ ID NO: 1-12中的任一个的子序列具有至少约90%(诸如至少约91%、92%、93%、94%、95%、96%、97%、98%、99%或更多中的任一种)序列同一性的氨基酸序列,其中所述子序列包含一个或多个功能性sRGN结构域。在某些实施方案中,所述sRGN多肽包含SEQ ID NO: 1-12中的任一个的子序列(或由其组成)。在某些实施方案中,所述一个或多个sRGN结构域包括核酸酶结构域(例如,HNH或RuvC结构域)。在某些实施方案中,所述一个或多个sRGN结构域包括RNA-结合结构域(例如,Rec I结构域)。在某些实施方案中,所述一个或多个sRGN结构域包括PAM-相互作用结构域。

[0135] 在某些实施方案中,本文提供了一种Gib11变体,其包括:

[0136] (I)在其整个长度上或在SEQ ID NO: 1的从位置787至位置1057的序列上与SEQ ID NO: 1的氨基酸序列具有至少90%(诸如至少约91%、92%、93%、94%、95%、96%、97%、98%、99%或更多中的任一种)序列同一性的变体;

[0137] (II)根据(I)的变体,其进一步包含组分例如核定位信号,以获得Gib11 CRISPR系统不仅在无细胞反应或原核细胞中而且在真核细胞环境(例如,活生物如植物或动物)中的适当活性;

[0138] (III)编码Gib11和根据(I)和(II)的变体的对应多核苷酸序列的密码子优化变体;

[0139] (IV)根据(I)至(III)中的任一项的变体,其进一步包含:

[0140] (a)一种或多种修饰或突变,其产生具有显著减小的或不可检测的核酸酶活性的sRGN,例如,在关于SEQ ID NO: 1的位置10 (例如,D10A)、位置562 (例如,H562A)和/或位置585 (例如,N585A)具有丙氨酸,或导致显著减小的核酸酶活性的任意其它交换;和

[0141] (b)具有核苷酸碱基编辑活性或脱氨酶活性的C-或N-端连接的多肽,诸如经由接头肽。

[0142] 在某些实施方案中,本文提供了一种P2H12变体,其包括:

[0143] (I)在其整个长度上或在SEQ ID NO: 5的从位置784至位置1055的序列上与SEQ ID NO: 5的氨基酸序列具有至少90%(诸如至少约91%、92%、93%、94%、95%、96%、97%、98%、99%或更多中的任一种)序列同一性的变体;

[0144] (II)根据(I)的变体,其进一步包含组分例如核定位信号,以获得P2H12 CRISPR系统不仅在无细胞反应或原核细胞中而且在真核细胞环境(例如,活生物如植物或动物)中的适当活性;

[0145] (III)编码P2H12和根据(I)和(II)的变体的对应多核苷酸序列的密码子优化变体;

[0146] (IV)根据(I)至(III)中的任一项的变体,其进一步包含:

[0147] (a)一种或多种修饰或突变,其产生具有显著减小的或不可检测的核酸酶活性的sRGN,例如,在关于SEQ ID NO: 5的位置10 (例如,D10A)、位置559 (例如,H559A)和/或位置582 (例如,N582A)具有丙氨酸,或导致显著减小的核酸酶活性的任意其它交换;和

[0148] (b)具有核苷酸碱基编辑活性或脱氨酶活性的C-或N-端连接的多肽,诸如经由接头肽。

[0149] 在某些实施方案中,本文提供了一种E2变体,其包括:

[0150] (I)在其整个长度上或在SEQ ID NO: 6的从位置1至位置789的序列上与SEQ ID NO: 6的氨基酸序列具有至少90%(诸如至少约91%、92%、93%、94%、95%、96%、97%、98%、99%或更多中的任一种)序列同一性的变体;

[0151] (II)根据(I)的变体,其进一步包含组分例如核定位信号,以获得E2 CRISPR系统不仅在无细胞反应或原核细胞中而且在真核细胞环境(例如,活生物如植物或动物)中的适当活性;

[0152] (III)编码E2和根据(I)和(II)的变体的对应多核苷酸序列的密码子优化变体;

[0153] (IV)根据(I)至(III)中的任一项的变体,其进一步包含:

[0154] (a)一种或多种修饰或突变,其产生具有显著减小的或不可检测的核酸酶活性的sRGN,例如,在关于SEQ ID NO: 6的位置10 (例如,D10A)、位置562 (例如,H562A)和/或位置585 (例如,N585A)具有丙氨酸,或导致显著减小的核酸酶活性的任意其它交换;和

[0155] (b)具有核苷酸碱基编辑活性或脱氨酶活性的C-或N-端连接的多肽,诸如经由接头肽。

[0156] 在某些实施方案中,本文提供了一种F8变体,其包括:

[0157] (I)在其整个长度上或在SEQ ID NO: 10的从位置1至位置789的序列上与SEQ ID NO: 10的氨基酸序列具有至少90%(诸如至少约91%、92%、93%、94%、95%、96%、97%、98%、99%或更多中的任一种)序列同一性的变体;

[0158] (II)根据(I)的变体,其进一步包含组分例如核定位信号,以获得F8 CRISPR系统不仅在无细胞反应或原核细胞中而且在真核细胞环境(例如,活生物如植物或动物)中的适当活性;

[0159] (III)编码F8和根据(I)和(II)的变体的对应多核苷酸序列的密码子优化变体;

[0160] (IV)根据(I)至(III)中的任一项的变体,其进一步包含:

[0161] (a)一种或多种修饰或突变,其产生具有显著减小的或不可检测的核酸酶活性的sRGN,例如,在关于SEQ ID NO: 10的位置10 (例如,D10A)、位置562 (例如,H562A)和/或位置585 (例如,N585A)具有丙氨酸,或导致显著减小的核酸酶活性的任意其它交换;和

[0162] (b)具有核苷酸碱基编辑活性或脱氨酶活性的C-或N-端连接的多肽,诸如经由接头肽。

[0163] 在某些实施方案中,本文提供了一种Gib11Spa-1变体,其包括:

[0164] (I)与SEQ ID NO: 2的氨基酸序列具有至少约90%(诸如至少约91%、92%、93%、94%、95%、96%、97%、98%、99%或更多中的任一种)序列同一性的变体;

[0165] (II)根据(I)的变体,其进一步包含组分例如核定位信号,以获得Gib11Spa-1 CRISPR系统不仅在无细胞反应或原核细胞中而且在真核细胞环境(例如,活生物如植物或动物)中的适当活性;

[0166] (III)编码Gib11Spa和根据(I)和(II)的变体的对应多核苷酸序列的密码子优化变体;

[0167] (IV)根据(I)至(III)中的任一项的变体,其进一步包含:

[0168] (a)一种或多种修饰或突变,其产生具有显著减小的或不可检测的核酸酶活性的sRGN,例如,在关于SEQ ID NO: 2的位置10 (例如,D10A)、位置562 (例如,H562A)和/或位

置585 (例如,N585A)具有丙氨酸,或导致显著减小的核酸酶活性的任意其它交换;和

[0169] (b)具有核苷酸碱基编辑活性或脱氨酶活性的C-或N-端连接的多肽,诸如经由接头肽。

[0170] 在某些实施方案中,本文提供了一种Gib11Spa-2变体,其包括:

[0171] (I)与SEQ ID NO: 3的氨基酸序列具有至少约90%(诸如至少约91%、92%、93%、94%、95%、96%、97%、98%、99%或更多中的任一种)序列同一性的变体;

[0172] (II)根据(I)的变体,其进一步包含组分例如核定位信号,以获得Gib11Spa-2 CRISPR系统不仅在无细胞反应或原核细胞中而且在真核细胞环境(例如,活生物如植物或动物)中的适当活性;

[0173] (III)编码Gib11Spa和根据(I)和(II)的变体的对应多核苷酸序列的密码子优化变体;

[0174] (IV)根据(I)至(III)中的任一项的变体,其进一步包含:

[0175] (a)一种或多种修饰或突变,其产生具有显著减小的或不可检测的核酸酶活性的sRGN,例如,在关于SEQ ID NO: 3的位置10 (例如,D10A)、位置562 (例如,H562A)和/或位置585 (例如,N585A)具有丙氨酸,或导致显著减小的核酸酶活性的任意其它交换;和

[0176] (b)具有核苷酸碱基编辑活性或脱氨酶活性的C-或N-端连接的多肽,诸如经由接头肽。

[0177] 在某些实施方案中,本文提供了一种Gib11Spa-3变体,其包括:

[0178] (I)与SEQ ID NO: 4的氨基酸序列具有至少约90%(诸如至少约91%、92%、93%、94%、95%、96%、97%、98%、99%或更多中的任一种)序列同一性的变体;

[0179] (II)根据(I)的变体,其进一步包含组分例如核定位信号,以获得Gib11Spa-3 CRISPR系统不仅在无细胞反应或原核细胞中而且在真核细胞环境(例如,活生物如植物或动物)中的适当活性;

[0180] (III)编码Gib11Spa和根据(I)和(II)的变体的对应多核苷酸序列的密码子优化变体;

[0181] (IV)根据(I)至(III)中的任一项的变体,其进一步包含:

[0182] (a)一种或多种修饰或突变,其产生具有显著减小的或不可检测的核酸酶活性的sRGN,例如,在关于SEQ ID NO: 4的位置10 (例如,D10A)、位置562 (例如,H562A)和/或位置585 (例如,N585A)具有丙氨酸,或导致显著减小的核酸酶活性的任意其它交换;和

[0183] (b)具有核苷酸碱基编辑活性或脱氨酶活性的C-或N-端连接的多肽,诸如经由接头肽。

[0184] 例如在酶活性的上下文中使用的术语“显著减小的”是指,这样的酶活性低于参照蛋白(例如,SEQ ID NO: 1-12中的任一个)的活性的约10%,例如,低于这样的参照酶活性的5%,低于2%,低于1%,或低于0.1%。

[0185] 在(IV) (b)下的合适多肽和它们与在(I)、(II)和(III)下的sRGN的C-端结构域的连接,描述在例如WO2017/070632, Gaudelli等人, Nature 551(2017年11月23日), 464-471;Komor等人, Sci. Adv. 2017;3eaao4774;Kim等人, Nature Biotechnol. 2017年4月;35(4) 371-376;Komor等人, Nature 533(7603);420-424。

[0186] 在(IV)下的sRGN变体可以将任何碱基对转化为任何可能的其它碱基对,而不是在

靶DNA序列中引入双链断裂。一个非限制性例子是脱氨酶活性,其可作用于胞嘧啶、鸟嘌呤或腺嘌呤碱基,并随后通过脱氨基位点复制并在细胞内修复,分别产生鸟嘌呤、胸腺嘧啶和鸟嘌呤。

[0187] 如果没有另外说明,术语sRGN包括SEQ ID NO: 1-12中的任一个的sRGN多肽以及本文所述的其所有变体,例如在(I)、(II)、(III)和(IV)下指定的那些。

[0188] 在某些实施方案中,本文提供了E2的突变体,其相对于SEQ ID NO: 6而言包含:a)在位置741的天冬氨酸残基,和在位置743的赖氨酸残基(E2+K741D+L743K, SEQ ID NO: 7);或b)在位置670的苏氨酸残基,和在位置675的天冬氨酸残基(E2+S670T+N675D, SEQ ID NO: 8);或c)在位置741的天冬酰胺残基,和在位置743的天冬酰胺残基(E2+K741N+L743N, SEQ ID NO: 9);或d)以上(a)至(c)的任意组合。

[0189] 在某些实施方案中,本文提供了E2的突变体,其包含i)在其整个长度上与根据SEQ ID NO: 6的序列或与SEQ ID NO: 6的从位置1至位置789的序列具有至少约90%(诸如至少约91%、92%、93%、94%、95%、96%、97%、98%、99%或更多中的任一种)氨基酸同一性;和ii)当与从在表1的a1前面的5个位置开始且在a1后面的5个位置结束的连续段比对时,表现出如在表1中所示的在位置a1的氨基酸残基;和iii)当与从在表1的a2前面的5个位置开始且在a2后面的5个位置结束的连续段比对时,表现出如在表1中所示的在位置a2的氨基酸残基;其中比对是指使用标准的局部比对工具或算法将11个氨基酸残基的段与根据SEQ ID NO: 6的序列比对,所述工具或算法包括例如Protein BLAST (https://blast.ncbi.nlm.nih.gov, the National Center for Biotechnology [2018年5月17日引用] Information;Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990)"Basic local alignment search tool."J. Mol. Biol. 215:403-410.),其使用标准参数,且其中a1或a2的5个在前氨基酸中的至少4个以及在a1或a2后面的5个氨基酸中的4个是相同的。

[0190] 表1

[0191]

根据SEQ ID NO: 6的位置	在位置a1的氨基酸残基	在位置a2的氨基酸残基
a1=670, a2=675	苏氨酸	天冬氨酸
a1=741, a2=743	天冬氨酸	赖氨酸
a1=741, a2=743	天冬酰胺	天冬酰胺

[0192] 如果没有另外说明,术语E2包含E2突变体,包括E2+K741D+L743K;E2+S670T+N675D;和E2+K741N+L743N。

[0193] 在某些实施方案中,本文提供了F8的突变体,其相对于SEQ ID NO: 10包含a)在位置737的天冬氨酸残基,和在位置739的赖氨酸残基(F8+K737D+L739K, SEQ ID NO: 11);或b)在位置737的天冬酰胺残基,和在位置739的天冬酰胺残基(F8+K737N+L739N, SEQ ID NO: 12)。

[0194] 在某些实施方案中,本文提供了F8的突变体,其包含i)在其整个长度上与根据SEQ ID NO: 10的序列或与SEQ ID NO: 10的从位置1至位置787的序列具有至少约90%(诸如至少约91%、92%、93%、94%、95%、96%、97%、98%、99%或更多中的任一种)氨基酸同一性;和ii)当与从在表2的a1前面的5个位置开始且在a1后面的5个位置结束的连续段比对时,表现出如在表2中所示的在位置a1的氨基酸残基;和iii)当与从在表2的a2前面的5个位置开始且在

a2后面的5个位置结束的连续段比对时,表现出如在表2中所示的在位置a2的氨基酸残基;其中比对是指使用标准的局部比对工具或算法将11个氨基酸残基的段与根据SEQ ID NO:10的序列比对,所述工具或算法包括例如Protein BLAST (https://blast.ncbi.nlm.nih.gov, the National Center for Biotechnology [在2018年5月17日引用] Information;Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990)"Basic local alignment search tool."J. Mol. Biol. 215:403-410.),其使用标准参数,且其中a1或a2的5个在前氨基酸中的至少4个以及在a1或a2后面的5个氨基酸中的4个是相同的。

[0195] 表2

[0196]

根据SEQ ID NO: 10的位置	在位置a1的氨基酸残基	在位置a2的氨基酸残基
a1=737, a2=739	天冬氨酸	赖氨酸
a1=737, a2=739	天冬酰胺	天冬酰胺

[0197] 如果没有另外说明,术语F8包含F8突变体,包括F8+K737D+L739K;和F8+K737N+L739N。

[0198] 可以如下产生其它sRGN:使用给定文库鉴定在通过基因家族改组产生的高活性sRGN中的小结构域的成对组合,并在合理设计的sRGN中包括这些成对组合中的至少一个,其中剩余的sRGN小结构域选自所述文库。剩余的小结构域可以选自起源于与成对组合中的小结构域相同的亲本Cas9的那些。合理设计的sRGN可以进一步包含在高活性sRGN中的小结构域的一个或多个其它成对组合。例如,合理设计的sRGN可以包含在高活性sRGN中存在的小结构域的2个、3个、4个、5个或更多个成对组合。

[0199] 嵌合的RNA-指导的核酸酶(cRGN)

[0200] 本文还提供了嵌合的RNA-指导的核酸酶(cRGN),其包含:A) N-端部分(cRGN-A),其取自葡萄球菌属的II型CRISPR RNA-指导的内切核酸酶Cas9,或通过两种或更多种葡萄球菌属II型CRISPR RNA-指导的内切核酸酶Cas蛋白的DNA改组序列产生的合成的RNA-指导的核酸酶,诸如本文描述的sRGN中的任一种;和b) C-端部分(cRGN-B),其取自来自巴氏葡萄球菌的II型CRISPR RNA-指导的内切核酸酶Cas9 (NCBI Refseq WP_023374365.1);其中cRGN-B包含i)来自巴氏葡萄球菌的II型CRISPR RNA-指导的内切核酸酶Cas9 (NCBI Refseq WP_023374365.1)的整个PI结构域,这是从根据NCBI Refseq WP_023374365.1的序列的氨基酸位置905开始且在氨基酸位置1054结束的序列;或ii)来自巴氏葡萄球菌的II型CRISPR RNA-指导的内切核酸酶Cas9 (NCBI Refseq WP_023374365.1)的整个PI结构域和整个WED结构域,后者位于根据NCBI Refseq WP_023374365.1的序列的氨基酸位置784至氨基酸位置904;cRGN-A包含相应II型CRISPR RNA-指导的内切核酸酶Cas9序列的前700个氨基酸,诸如前739个氨基酸或前783个氨基酸,例外是,在氨基酸水平与来自巴氏葡萄球菌的NCBI Refseq WP_023374365.1的各个部分具有超过99.9%同一性、在某些情况下超过99.5%同一性的序列;和cRGN-A直接地或通过接头肽序列连接至cRGN-B。在某些实施方案中,cRGN-A包含各个II型CRISPR RNA-指导的内切核酸酶Cas9序列的超过前700个氨基酸,诸如前739个氨基酸或前783个氨基酸。

[0201] 金黄色葡萄球菌的II型CRISPR RNA-指导的内切核酸酶Cas9的PI结构域和WED结构域描述在Nishimasu等人, Cell. 2015年8月27日;162(5): 1113-1126。在图6中进一步

示出两个结构域的潜在定位。

[0202] 在本专利申请中对NCBI Refseq数据库的任何提及均表示以下提及:国家生物技术信息中心(NCBI)[因特网]. Bethesda (MD): 国立医学图书馆(US), 国家生物技术信息中心;[1988] - [2018年5月11日引用]。可得自: https://www.ncbi.nlm.nih.gov/protein。例如,关于NCBI Refseq WP_048803085,这是https://www.ncbi.nlm.nih.gov/protein/WP_048803085。

[0203] 对单独葡萄球菌属的任何提及是指葡萄球菌属。

[0204] 在本发明的上下文中,DNA改组是指在不同的(例如,来自不同的葡萄球菌属种)II型CRISPR Cas9内切核酸酶DNA序列之间交换序列片段,以产生编码具有RNA-指导的内切核酸酶活性(诸如II型CRISPR Cas9内切核酸酶的活性)的合成蛋白的嵌合DNA序列。这样的技术尤其描述在Joern J.M. (2003) DNA Shuffling. 见: Arnold F.H., Georgiou G. (编) Directed Evolution Library Creation. Methods in Molecular Biology™, vol 231. Humana Press, 以及Gibbs等人, Gene. 2001年6月13日; 271(1):13-20。

[0205] 在某些实施方案中,通过NCBI Refseq数据库的下述入口之一鉴定来自葡萄球菌属的II型CRISPR RNA-指导的内切核酸酶Cas9:

[0206]

WP_053019794,	WP_071859985,	WP_049437627,	WP_081329738,	WP_081330634,
WP_088922804,	WP_061854099,	WP_096792116,	WP_107588422,	WP_107543406,
WP_105966910,	WP_083326931,	WP_105977729,	WP_107578657,	WP_107533955,
WP_107597066,	WP_105980293,	WP_105978400,	WP_107539784,	WP_083326835,
WP_096754380,	WP_107580550,	WP_107530431,	WP_107597643,	WP_107596301,
WP_082732265,	WP_083310250,	WP_082709447,	WP_105994700,	WP_107568091,

[0207] WP_107544007, WP_103361957, WP_001573634, WP_107576310, WP_107571611, WP_107593728, WP_103167028, WP_107532850, WP_002460848, WP_107506206, WP_104681501, WP_107546539, WP_058710220, WP_075777761, WP_050331073, WP_050345681, WP_053017934, WP_049415449, WP_060829977, WP_070855141, WP_107641154, WP_107392933, WP_107389582, WP_107390356, WP_107378401, WP_107378676, WP_060552032, WP_107376447, WP_039643679, WP_101457364, WP_104039168, WP_104052030, WP_085237539, WP_103357343, WP_044361501, WP_107366415, WP_107393309, WP_107633689, WP_107397003, WP_107386954, WP_107371508, WP_107538271, WP_107642914, WP_107536061, WP_107605852, WP_107547877, WP_107634675, WP_107532082, WP_107604007, WP_103356745, WP_107642811, WP_103298901, WP_107611983, WP_107623815, WP_107522281, WP_107560076, WP_107368542, WP_081502240, WP_107604002, WP_107377516, WP_016424980, WP_009382362, WP_001573633, WP_016424981, WP_103298687, WP_103361956, WP_107637979, WP_107612621, WP_107642817, WP_103356723, WP_107593719, WP_070592992, WP_107536245, WP_103209613, WP_063284667, WP_101457463, WP_105503156, WP_096548249, WP_096536567, WP_014613259, WP_063278948, WP_096601671, WP_103863320, WP_096559644, WP_019167918, WP_086428210, WP_096665615, WP_105976295, WP_107576302, WP_105977863, WP_107544006, WP_107538330, WP_070848771, WP_096605716, WP_095105824, WP_105966809, WP_070469119, WP_096589032, WP_105996442, WP_107552556, WP_060803934, WP_107549437, WP_107579080, WP_107547813, WP_107591747, WP_105978348, WP_096598476, WP_107587102, WP_107533825, WP_107588308, WP_107567989, WP_096544347, WP_023015764, WP_103322053, WP_060803559, WP_048803085, WP_107530433, WP_107580731, WP_107368971, WP_107559912, WP_107371387, WP_107571609, WP_107559911, WP_107571610, WP_107530436, WP_107580472, WP_017000597, WP_023374365, WP_062197343, WP_035512507, WP_079708828, WP_101040307, WP_053216997, WP_106588293, WP_107585021, WP_082099322, WP_017185731, WP_101662761, WP_084780162, WP_016837331, WP_084781893, WP_106494556, WP_088825434, WP_088816271, WP_059140148, WP_007547525, WP_084347835, WP_070710475, WP_077907981, WP_035723552, WP_058095017, WP_006446566, WP_074693676, WP_050980543, WP_001271093, WP_001271092, WP_087971021, WP_061668060, WP_016119566, WP_016106885, WP_088031364, WP_088038716, WP_007476473, WP_003343632, WP_088049424, WP_003354196, WP_066148467, WP_082806588, WP_087094968, WP_048723014,

WP_033844707, WP_055358891, WP_081209836, WP_033018780, WP_023633350, WP_046580811, WP_081157433, WP_087959824, WP_041267823, WP_018922791, WP_063577905, WP_066227285, WP_044736072, WP_033016936, WP_100664518, WP_064213580, WP_053532223, WP_042224718, WP_084138993, WP_010632729, WP_018130201, WP_088027793, WP_065212529, WP_048539452, WP_001105083, WP_001105082, WP_086397116, WP_086390158, WP_070006567, WP_003739838, WP_070034634, WP_075702521, WP_029090905, WP_038409211, WP_069125601, WP_046323366, WP_069887401, WP_069134523, WP_061665472, WP_070294293, WP_003749665, WP_070784981, WP_031669209, WP_106787163, WP_077287021, WP_085392451, WP_023548323, WP_061108493, WP_101152964, WP_101143843, WP_101143453, WP_101140817, WP_066465432, WP_033920898, WP_069001072, WP_085400884, WP_101057368, WP_088838826, WP_072240946, WP_033838504,

[0208] WP_010991369, WP_072218760, WP_072238933, WP_070227966, WP_061663015, WP_003733029, WP_101142252, WP_070031693, WP_070785826, WP_003730785, WP_070274575, WP_003727705, WP_069890501, WP_070233243, WP_070039312, WP_031665337, WP_070222802, WP_070264592, WP_103682188, WP_070299153, WP_070307355, WP_061395959, WP_070764199, WP_061128889, WP_070283519, WP_069009724, WP_070228842, WP_014601172, WP_103757671, WP_070293394, WP_060587936, WP_070238603, WP_070215465, WP_003723650, WP_070214481, WP_058876445, WP_060567941, WP_029499861, WP_027129613, WP_003145379, WP_080151624, WP_021752441, WP_069132012, WP_084038511, WP_004632196, WP_066230975, WP_072520207, WP_082729137, WP_080149038, WP_072233091, WP_072240445, WP_072218465, WP_072239624, WP_080151670, WP_077914265, WP_080151712, WP_052025682, WP_077914264, WP_100353964, WP_046517046, WP_008341324,

[0209] 在某些实施方案中,cRGN-A是来自Gib11 (SEQ ID NO: 1)。在某些实施方案中,所述cRGN是Gib11Spa-1 (SEQ ID NO: 2)、Gib11Spa-2 (SEQ ID NO: 3)或Gib11Spa-3 (SEQ ID NO: 4)。

[0210] 其它小Cas核酸酶

[0211] 在一个方面,本文提供了来自巴氏葡萄球菌、Staphylococcus microti或猪葡萄球菌的Cas9。在某些实施方案中,所述Cas9是来自巴氏葡萄球菌,且包含SEQ ID NO: 31的氨基酸序列或其与SEQ ID NO: 31具有至少约90%(诸如至少约91%、92%、93%、94%、95%、96%、97%、98%、99%或更多中的任一种)序列同一性的变体。巴氏葡萄球菌Cas9或“SpaCas9”会识别PAM序列NNGG。在某些实施方案中,所述Cas9是来自Staphylococcus microti,且包含SEQ ID NO: 33的氨基酸序列或其与SEQ ID NO: 33具有至少约90%(诸如至少约91%、92%、93%、94%、95%、96%、97%、98%、99%或更多中的任一种)序列同一性的变体。Staphylococcus microti Cas9或“SmiCas9”会识别PAM序列NNGG。在某些实施方案中,所述Cas9是来自猪葡萄球菌,且包含SEQ ID NO: 32的氨基酸序列或其与SEQ ID NO: 32具有至少约90%(诸如至

少约91%、92%、93%、94%、95%、96%、97%、98%、99%或更多中的任一种)序列同一性的变体。猪葡萄球菌Cas9或“ShyCas9”会识别PAM序列NNAAAA。在某些情况下,这些小Cas9可用作本文所述的任何系统、组合物和方法中的sRGN的替代物。

[0212] 基于sRGN多肽的CRISPR-Cas系统

[0213] 在某些实施方案中,本文提供了一种系统,其包含: (a) sRGN多肽或其与SEQ ID NO: 1-12中的任一个具有至少约90%(诸如至少约91%、92%、93%、94%、95%、96%、97%、98%、99%或更多中的任一种)序列同一性的变体,或编码所述sRGN多肽或其变体的核酸;和(b)指导RNA (gRNA)或编码所述gRNA的核酸,其中所述gRNA能够将所述sRGN多肽或其变体指导至靶多核苷酸序列。在某些实施方案中,所述系统包含sRGN多肽或其变体。在某些实施方案中,所述sRGN多肽包含SEQ ID NO: 1-12中的任一个的氨基酸序列(或由其组成)。在某些实施方案中,所述系统包含编码sRGN的核酸。在某些实施方案中,所述核酸与SEQ ID NO: 13-29具有至少约90%(诸如至少约91%、92%、93%、94%、95%、96%、97%、98%、99%或更多中的任一种)序列同一性。在某些实施方案中,所述核酸包含SEQ ID NO: 13-29的序列(或由其组成)。在某些实施方案中,所述系统包含gRNA。在某些实施方案中,所述系统包含编码gRNA的核酸。在某些实施方案中,所述gRNA是单个指导RNA (sgRNA)。在某些实施方案中,所述系统进一步包括一种或多种另外gRNA或编码一种或多种另外gRNA的核酸。

[0214] 在某些实施方案中,根据包含编码本文描述的sRGN多肽或其变体的多核苷酸序列的任何核酸,所述多核苷酸序列可操作地连接至(i)用于在细胞或体外环境中表达的合适启动子;和/或(ii)合适的核定位信号。

[0215] 在某些实施方案中,根据本文所述的编码gRNA的任何核酸,所述核酸包含编码gRNA的多核苷酸序列,所述多核苷酸序列可操作地连接至(i)用于在细胞或体外环境中表达的合适启动子;和/或(ii)合适的核定位信号。

[0216] 在某些实施方案中,本文提供了一种系统,其包含:(a) sRGN多肽,其包含SEQ ID NO: 1-12中的任一个的氨基酸序列或其与SEQ ID NO: 1-12中的任一个具有至少90%序列同一性的变体;和(b) gRNA,其中所述gRNA能够将所述sRGN多肽或其变体指导至靶多核苷酸序列。在某些实施方案中,所述gRNA是sgRNA。

[0217] 在某些实施方案中,本文提供了一种系统,其包含:(a) sRGN多肽,其包含SEQ ID NO: 1-12中的任一个的氨基酸序列或其与SEQ ID NO: 1-12中的任一个具有至少90%序列同一性的变体;和(b)编码gRNA的核酸,其中所述gRNA能够将所述sRGN多肽或其变体指导至靶多核苷酸序列。在某些实施方案中,所述gRNA是sgRNA。

[0218] 在某些实施方案中,本文提供了一种系统,其包含:(a)编码sRGN多肽的核酸,所述sRGN多肽包含SEQ ID NO: 1-12中的任一个的氨基酸序列或其与SEQ ID NO: 1-12中的任一个具有至少90%序列同一性的变体;和(b) gRNA,其中所述gRNA能够将所述sRGN多肽或其变体指导至靶多核苷酸序列。在某些实施方案中,所述核酸包含SEQ ID NO: 13-29中的任一个的多核苷酸序列或其与SEQ ID NO: 13-29中的任一个的多核苷酸序列具有至少90%序列同一性的变体。在某些实施方案中,所述gRNA是sgRNA。

[0219] 在某些实施方案中,本文提供了一种系统,其包含:(a)编码sRGN多肽的核酸,所述sRGN多肽包含SEQ ID NO: 1-12中的任一个的氨基酸序列或其与SEQ ID NO: 1-12中的任一个具有至少90%序列同一性的变体;和(b)编码gRNA的核酸,其中所述gRNA能够将所述

sRGN多肽或其变体指导至靶多核苷酸序列。在某些实施方案中,所述核酸包含SEQ ID NO: 13-29中的任一个的多核苷酸序列或其与SEQ ID NO: 13-29中的任一个的多核苷酸序列具有至少90%序列同一性的变体。在某些实施方案中,所述gRNA是sgRNA。在某些实施方案中,所述系统包含核酸,所述核酸包含(i)编码sRGN多肽或其变体的核酸;和(ii)编码gRNA的核酸。

[0220] 在某些实施方案中,根据本文所述的任何系统,所述系统进一步包括供体模板。在某些实施方案中,所述供体模板包含用于插入靶多核苷酸序列的供体序列。在某些实施方案中,所述供体序列是外源序列,例如用于在靶标中表达的转基因。在某些实施方案中,所述供体模板包含用于修饰靶多核苷酸序列中一个或多个碱基的供体序列。在某些实施方案中,所述系统进一步包括一种或多种另外的供体模板。在某些实施方案中,供体模板与系统中的gRNA物理连接。

[0221] 根据本发明的一个实施方案提供了组合物,其包含:

[0222] (a)sRGN多肽或编码这样的sRGN的核酸;

[0223] (b)单一异源指导RNA(sgRNA)或允许原位产生这种sgRNA的核酸(例如,DNA)(例如,编码所述sgRNA的核酸),其中所述sgRNA包含:

[0224] i.包含RNA并且能够与多核苷酸基因座中的靶序列杂交的DNA-靶向区段,

[0225] ii.包含RNA的tracr配对物序列,和

[0226] iii.包含RNA的tracr RNA序列,

[0227] 其中所述tracr配对物序列能够与所述tracr序列杂交,且其中(i)、(ii)和(iii)以5'至3'的取向排列。

[0228] 在sgRNA内,tracr配对物序列和tracr序列通常通过合适的环序列连接并形成茎-环结构。

[0229] 进一步,在某些实施方案中,编码sRGN多肽的核酸和/或所述sgRNA含有用于在细胞或体外环境中表达的合适启动子和/或合适的核定位信号。

[0230] 根据本发明的另一个实施方案提供了在细胞中或体外的一个或多个位置处靶向、编辑、修饰或操纵靶DNA的方法,包括:

[0231] (a)将异源sRGN多肽或编码sRGN多肽的核酸引入细胞或体外环境中;和

[0232] (b)引入单一异源指导RNA(sgRNA)或适合于原位产生这种sgRNA的核酸(例如,DNA)(例如,编码所述sgRNA的核酸),其中所述sgRNA包含:

[0233] i.包含RNA并且能够与多核苷酸基因座中的靶序列杂交的DNA-靶向区段,

[0234] ii.包含RNA的tracr配对物序列,和

[0235] iii.包含RNA的tracr RNA序列,

[0236] 其中所述tracr配对物序列能够与所述tracr序列杂交,且其中(i)、(ii)和(iii)以5'至3'的取向排列;和

[0237] (c)在所述靶DNA中产生一个或多个切割、切口或编辑,其中所述sRGN多肽通过其加工的或未加工的形式的gRNA被指导至所述靶DNA。

[0238] 根据本发明的另一个实施方案是包含下述的组合物用于在细胞(例如,体外或体内)或无细胞系统中一个或多个位置处靶向、编辑、修饰或操纵靶DNA的用途:

[0239] (a)sRGN多肽或编码这样的sRGN的核酸;

[0240] (b)单一异源指导RNA(sgRNA)或适合于原位产生这种sgRNA的核酸(例如,DNA)(例如,编码所述sgRNA的核酸),其中所述sgRNA包含:

[0241] i.包含RNA并且能够与多核苷酸基因座中的这种靶序列杂交的DNA-靶向区段,

[0242] ii.包含RNA的tracr配对物序列,和

[0243] iii.包含RNA的tracr RNA序列,

[0244] 其中所述tracr配对物序列能够与所述tracr序列杂交,且其中(i)、(ii)和(iii)以5'至3'的取向排列。

[0245] 根据本发明的另一个实施方案是离体或体外细胞,其包含:

[0246] (a)异源sRGN多肽或编码它的核酸;

[0247] (b)单一异源指导RNA(sgRNA)或适合于原位产生这种sgRNA的核酸(例如,DNA)(例如,编码所述sgRNA的核酸),其中所述sgRNA包含:

[0248] i.包含RNA并且能够与多核苷酸基因座中的这种靶序列杂交的DNA-靶向区段,

[0249] ii.包含RNA的tracr配对物序列,和

[0250] iii.包含RNA的tracr RNA序列,

[0251] 其中所述tracr配对物序列能够与所述tracr序列杂交,且其中(i)、(ii)和(iii)以5'至3'的取向排列;

[0252] 或已使用上述(a)和(b)对其基因组进行靶向、编辑、修饰或操纵的这样的细胞。

[0253] 根据本发明的额外的实施方案包括试剂盒,其包括:

[0254] (a)编码sRGN多肽的核酸,其中所述核酸可操作地连接至启动子或核糖体结合位点;和

[0255] (b)单一异源指导RNA(sgRNA)或适合于原位产生这种sgRNA的核酸(例如,DNA)(例如,编码所述sgRNA的核酸),其中所述sgRNA包含:

[0256] i.包含RNA并且能够与多核苷酸基因座中的这种靶序列杂交的DNA-靶向区段,

[0257] ii.包含RNA的tracr配对物序列,和

[0258] iii.包含RNA的tracr RNA序列,

[0259] 其中所述tracr配对物序列能够与所述tracr序列杂交,且其中(i)、(ii)和(iii)以5'至3'的取向排列。

[0260] 或者

[0261] (a)sRGN多肽;和

[0262] (b)一种或多种单一异源指导RNA(sgRNA),其每一种都包含:

[0263] iv.包含RNA并且能够与多核苷酸基因座中的这种靶序列杂交的DNA-靶向区段,

[0264] v.包含RNA的tracr配对物序列,和

[0265] vi.包含RNA的tracr RNA序列,

[0266] 其中所述tracr配对物序列能够与所述tracr序列杂交,且其中(iv)、(v)和(vi)以5'至3'的取向排列。

[0267] 根据本发明的再另一个实施方案包含用于在细胞(例如,体外或体内)或无细胞系统中一个或多个位置处靶向、编辑、修饰或操纵一种或多种靶DNA的组合物和方法,其包含:

[0268] (a)sRGN多肽或编码sRGN多肽的核酸;和

[0269] (b)指导RNA(gRNA)或适合于原位产生这种gRNA的核酸(例如,DNA)(例如,编码所

述gRNA的核酸),其中所述gRNA包含:

[0270] i.包含RNA并且能够与多核苷酸基因座中的这种靶序列杂交的DNA-靶向区段,

[0271] ii.包含RNA的tracr RNA序列;

[0272] 其中(i)和(ii)在单个RNA分子上,且(iii)在单独的RNA分子上。

[0273] *突变体*

[0274] 在某些实施方案中,sRGN多肽是具有改变的核糖核酸内切酶活性或前-crRNA、中间crRNA或成熟crRNA的相关半衰期的突变多肽。在某些实施方案中,sRGN多肽是在没有相当大地减少或增强的内切核糖核酸酶活性或对DNA的结合亲和力的情况下的具有改变的或取消的DNA内切核酸酶活性的突变多肽。这种修饰可便于为了转录调节、激活或阻抑起见的sRGN多肽的序列特异性DNA靶向;通过甲基化作用、脱甲基作用、乙酰化作用或脱乙酰作用的外遗传修饰或染色质修饰,或本领域已知的DNA结合和/或DNA修饰蛋白的任何其它修饰。

[0275] 在某些实施方案中,sRGN多肽是不具有DNA内切核酸酶活性的突变多肽。

[0276] 在某些实施方案中,细胞是细菌细胞、真菌细胞、古细菌细胞、原生生物、植物细胞或动物细胞。

[0277] 在某些实施方案中,通过编码多肽和核酸的相同或不同重组载体将sRGN多肽和一种或多种指导RNA(gRNA)如单一异源指导RNA(sgRNA)引入细胞中。

[0278] 在某些实施方案中,编码多肽、核酸或多肽和核酸两者的核酸被修饰。

[0279] 在某些实施方案中,方法或系统进一步包括添加供体DNA序列,并且其中靶DNA序列通过同源性指导的修复进行编辑。在某些实施方案中,多核苷酸供体模板与crRNA或指导RNA物理连接。

[0280] 在另一方面,本文提供了用于在不具有对单链RNA(ssRNA)、双链RNA(dsRNA)或RNA和DNA的异源双链体的活性的情况下修饰或编辑双链DNA或单链靶DNA的方法。

[0281] *多重化*

[0282] 在另一方面,本文提供了一种用于在细胞中多个位置处编辑或修饰DNA的方法,其基本上由下述组成:i)将sRGN多肽或编码sRGN多肽的核酸引入细胞;和ii)将包含两种或更多种作为RNA或编码为DNA并在一个启动子控制下的前-CRISPR RNA(前-crRNA)的单一异源核酸引入细胞,每种前-crRNA包含重复序列间隔区阵列或重复序列间隔区,其中所述间隔区包含与DNA中靶序列互补的核酸序列,且所述重复序列包含茎-环结构,其中所述sRGN多肽在茎-环结构上游切割所述两种或更多种前-crRNA,以产生两种或更多种中间crRNA,其中所述两种或更多种中间crRNA被加工为两种或更多种成熟crRNA,并且其中两种或更多种成熟crRNA的每一种指导所述sRGN多肽使得向DNA中产生两个或更多个双链断裂(DSB)。例如,sRGN多肽的一个优点是可能仅引入一种包含几种重复序列间隔区单位的前-crRNA,其在引入时,由sRGN多肽加工成靶向DNA上几种不同序列的活性重复序列间隔区单位。

[0283] 在某些实施方案中,单一异源核酸中的前-crRNA序列在特定的位置、取向、顺序或以特定的化学键连接在一起,以在由不同的crRNA序列所指定的每个位点指导或有差异地调节sRGN多肽的内切核酸酶活性。

[0284] 在另一方面,本文提供了用于在细胞中多个位置处编辑或修饰DNA的结构或功能的总的方法的例子,其基本上由下述组成:i)将RNA-指导的内切核酸酶如sRGN作为多肽或编码所述RNA-指导的内切核酸酶的核酸引入细胞;和ii)引入包含或编码两种或更多种作

为RNA或编码为DNA并在一个或多个启动子控制下的指导RNA的单一异源核酸,其中所述RNA-指导的内切核酸酶的活性或功能由所述单一异源核酸中的指导RNA序列指导。

[0285] *sRGN多肽的密码子优化的DNA序列*

[0286] 在方法的一些实施方案中,编码sRGN多肽的核酸序列是修饰的核酸,例如密码子优化的。在方法的一些实施方案中,单一异源核酸是修饰的核酸。在方法的一些实施方案中,方法进一步包括将多核苷酸供体模板引入细胞中。在某些实施方案中,多核苷酸供体模板与crRNA或指导RNA物理连接。在方法的一些实施方案中,通过同源性指导的修复、非同源末端连接或微同源性介导的末端连接(microhomology-mediated end joining)在DSBs处修复DNA。

[0287] 在方法的一些实施方案中,sRGN多肽更容易与局部环境中的成熟crRNA复合,且因此由于crRNA被相同sRGN多肽自局部环境中的前-crRNA加工,所以更易于用于指导DNA内切核酸酶活性。

[0288] 在方法的一些实施方案中,sRGN多肽用于从修饰的前-crRNA寡核苷酸序列中切割、分离或纯化一种或多种成熟crRNA序列,在所述修饰的前-crRNA寡核苷酸序列中,异源序列以5'或3'并入RNA寡核苷酸或DNA表达构建体中的一种或多种crRNA序列。可并入异源序列以修饰稳定性、半衰期、表达水平或时间控制、与sRGN多肽或靶DNA序列的相互作用或本领域已知的任何其它物理或生化特征。

[0289] 在方法的一些实施方案中,对前-crRNA序列进行修饰以提供对前-crRNA序列内两种或更多种成熟crRNA序列的有差异的调节,以有差异地修饰稳定性、半衰期、表达水平或时间控制、与sRGN多肽或靶DNA序列的相互作用或本领域已知的任何其它物理或生化特征。

[0290] 在某些实施方案中,对sRGN多肽(或其核酸编码的变体)进行修饰以改善所期望的特征,例如功能、活性、动力学、半衰期等。这种修饰的一个这种非限制性例子是用来自不同核酸酶的同源或异源切割结构域(例如来自类型II CRISPR-相关核酸酶Cas9的RuvC或HNH结构域)置换sRGN多肽的'切割结构域'。

[0291] 在一方面,本文提供了用于靶向、编辑或操纵细胞中DNA的方法,包括将完整或部分或完全缺陷的sRGN多肽或前-crRNA或crRNA部分连接至二聚体FOK1核酸酶,以在通过一种或多种crRNA分子指导到一个或多个特定DNA靶位点时指导内切核酸酶切割。在另一个实施方案中,FOK1核酸酶系统是切口酶或温度敏感突变体或本领域已知的任何其它变体。

[0292] 在某些实施方案中,将与二聚体FOK1核酸酶连接的sRGN多肽与进入细胞的包含两种或更多种作为RNA或编码为DNA并在一个启动子控制下的前-crISPR RNA(前-crRNA)的单一异源核酸一起引入细胞,每种前-crRNA包含重复序列间隔区阵列,其中所述间隔区包含与DNA中靶序列互补的核酸序列,且所述重复序列包含茎-环结构,其中所述sRGN多肽在茎-环结构上游切割所述两种或更多种前-crRNA,以产生两种或更多种中间crRNA。

[0293] 在一方面,本文提供了用于靶向、编辑或操纵细胞中DNA的方法,包括将完整或部分或完全缺陷的sRGN多肽或前-crRNA、中间crRNA、成熟crRNA部分或gRNA(统称为crRNA)与供体单链或双链DNA供体模板连接,以在通过一种或多种指导RNA或crRNA分子指导到一个或多个特定DNA靶位点时促进外源DNA序列的同源重组。

[0294] 在再另一方面,本文提供了用于将DNA模板为了进行同源重组或同源性指导的修复指导至基因编辑的特定位点的方法。在这方面,单链或双链DNA模板用化学方法或通过本

领域已知的其它方式连接至crRNA或指导RNA。在某些实施方案中,DNA模板保持与crRNA或指导RNA连接;在再其它实例中,sRGN多肽切割crRNA或指导RNA,从而释放DNA模板以使得能够或促进同源重组。

[0295] 在再另一方面,本文提供了用于靶向、编辑或操纵细胞中DNA的方法,包括将完整或部分或完全缺陷的sRGN多肽或前-crRNA或crRNA部分连接至转录激活因子或阻遏蛋白,或外遗传修饰因子,例如甲基化酶、脱甲基酶、乙酰基转移酶或脱乙酰酶,或信号传导或检测,在通过一种或多种crRNA分子指导到一个或多个特定DNA靶位点时,其所有方面以前已关于Cas9系统进行了描述。

[0296] 在另一方面,本文提供了组合物,其包含连接至crRNA或指导RNA的多核苷酸供体模板。一种靶向、编辑或操纵细胞中DNA的方法,包括将前-crRNA或crRNA或指导RNA连接至供体单链或双链多核苷酸供体模板,以使供体模板通过sRGN多肽从前-crRNA或crRNA或指导RNA上切割下来,从而在通过一种或多种指导RNA或crRNA分子指导到一个或多个特定DNA靶位点时促进通过供体模板的同源性指导的修复。

[0297] sRGN多肽也可用于形成嵌合结合蛋白,其中引入了其它结构域和活性。作为举例说明,Fok1结构域可以与sRGN蛋白融合,其可以含有催化活性的内切核酸酶结构域,或者Fok1结构域可以与sRGN蛋白融合,其已经被修饰以使sRGN多肽内切核酸酶结构域无活性。可以与sRGN多肽融合以制备嵌合蛋白的其它结构域包括转录调节剂、外遗传修饰因子、标签和其它标记或成像剂、组蛋白和/或调节或修饰基因序列的结构或活性的本领域中已知的其它形态。

[0298] 核酸和/或氨基酸修饰

[0299] 在某些实施方案中,引入细胞的多核苷酸包含一种或多种修饰,其可用于例如增强活性、稳定性或特异性,改变递送,减少宿主细胞中的先天免疫应答,进一步减小蛋白大小或用于其它增强,如本文进一步描述和本领域已知。

[0300] 在某些实施方案中,在基于sRGN的CRISPR-Cas系统中使用修饰的多核苷酸,在这种情况下,可以修饰引入细胞中的指导RNA和/或编码sRGN多肽的DNA或RNA,如下文所述和所举例说明的。这种修饰的多核苷酸可用于CRISPR-Cas系统中以编辑任何一个或多个基因组基因座。

[0301] CRISPR-Cas系统的组分

[0302] A. 指导RNA/sgRNA

[0303] 在本发明的方面,术语“嵌合RNA”、“嵌合指导RNA”、“指导RNA”、“单一指导RNA”和“合成指导RNA”可互换使用,并且是指包含DNA-靶向区段、tracr序列和tracr配对物序列的多核苷酸序列。术语“指导序列”或“DNA-靶向序列”是指指定靶位点的指导RNA内约20bp的序列,并且可以与术语“指导”或“间隔区”互换使用。术语“tracr配对物序列”也可以与术语“一种或多种同向重复序列”互换使用。

[0304] 在一些方面,根据本发明的单一指导RNA(sgRNA)包含:

[0305] i.包含RNA并且能够与多核苷酸基因座中的靶序列杂交的DNA-靶向区段或DNA-靶向序列,

[0306] ii.包含RNA的tracr配对物序列,和

[0307] iii.包含RNA的tracr RNA序列,

[0308] 其中所述tracr配对物序列能够与所述tracr序列杂交,且其中(i)、(ii)和(iii)以5'至3'的取向排列。在某些实施方案中,组分(i)、(ii)和(iii)在单个嵌合RNA上。

[0309] 在sgRNA内,tracr配对物序列和tracr序列可以通过合适的环序列连接并形成茎-环结构。

[0310] 通常,tracr配对物序列包括与tracr序列具有足够互补性的任何序列以促进下述一种或多种:(1)在含有相应tracr序列的细胞中侧接tracr配对物序列的DNA-靶向区段的切除;和(2)在靶序列处CRISPR复合物的形成,其中所述CRISPR复合物包含与tracr序列杂交的tracr配对物序列。通常,互补性程度是以沿着两个序列中较短者的长度,tracr配对物序列和tracr序列的最佳比对为基准的。最佳比对可以通过任何合适的比对算法来确定,并且可以进一步计及二级结构,例如tracr序列或tracr配对物序列内的自身互补性。在某些实施方案中,当最佳比对时,沿着两者中较短者的30个核苷酸长度,tracr序列和tracr配对物序列之间的互补性程度为约或超过25%、30%、40%、50%、60%、70%、80%、90%、95%、97.5%、99%或更高。在某些实施方案中,tracr序列长度为约或超过5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、25、30、40、50个或更多个核苷酸。在某些实施方案中,tracr序列和tracr配对物序列包含在单个转录物中,以使两者之间的杂交产生具有二级结构的转录物,例如发夹。在本发明的实施方案中,转录物或转录的多核苷酸序列具有至少两个或更多个发夹。

[0311] 在某些实施方案中,转录物具有两个、三个、四个或五个发夹。在本发明的进一步的实施方案中,转录物具有至多五个发夹。在发夹结构中,最后"N"的5'和环的上游的序列部分对应于tracr配对物序列,而环的3'的序列部分对应于tracr序列。包含DNA-靶向区段、tracr配对物序列和tracr序列的单一多核苷酸的进一步的非限制性例子如下(5'至3'地列举),其中"N"代表DNA-靶向区段的碱基,第一组小写字体字母代表tracr配对物序列,且第二组小写字体字母代表tracr序列,且最后的聚胸苷酸(poly-T)序列代表转录终止子:

[0312] 指导RNA的DNA-靶向区段包含与靶DNA中的序列互补的核苷酸序列。换句话说,指导RNA的DNA-靶向区段通过杂交(即,碱基配对)以序列特异性方式与靶DNA相互作用。因此,DNA-靶向区段的核苷酸序列可以变化并确定指导RNA和靶DNA将相互作用的靶DNA内的位置。可以修饰(例如通过遗传工程)指导RNA的DNA-靶向区段,以与靶DNA内的任何所期望的序列杂交。

[0313] DNA-靶向区段可以具有10个核苷酸至30个核苷酸的长度。

[0314] 在某些实施方案中,DNA-靶向区段具有13个核苷酸至25个核苷酸的长度。

[0315] 在某些实施方案中,DNA-靶向区段具有15个核苷酸至23个核苷酸的长度。

[0316] 在某些实施方案中,DNA-靶向区段具有18个核苷酸至22个核苷酸,例如20至22个核苷酸的长度。

[0317] DNA-靶向区段的DNA-靶向序列与靶DNA的靶序列之间的百分比互补性在20-22个核苷酸范围内可以为至少60%(例如,至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少97%、至少98%、至少99%或100%)。

[0318] 与靶DNA中的靶序列互补的原间隔区序列可以在其3'端与对于sRGN多肽的合适的PAM序列相邻,或者这种PAM序列可以是DNA-靶向区段的3'部分的一部分。

[0319] 为了非限制性地举例说明这种用途起见使用CRISPR-Cas系统,可以使用指导RNA的修饰来增强CRISPR-Cas基因组编辑复合物的形成或稳定性,所述复合物包含指导RNA和

Cas内切核酸酶,例如sRGN。指导RNA的修饰可以也或替换地用于增强基因组编辑复合物与基因组中靶序列之间相互作用的起始、稳定性或动力学,这可以例如用于增强中靶(on-target)活性。指导RNA的修饰可以也或替换地用于增强特异性,例如,与在其它(脱靶)位点的作用相比,在中靶位点的基因组编辑的相对比率。

[0320] 修饰可以也或替换地用于增加指导RNA的稳定性,例如,通过增加其对细胞中存在的核糖核酸酶(RNA酶)降解的抗性,从而导致其在细胞中的半衰期增加。增强指导RNA半衰期的修饰在其中经由需要翻译以产生sRGN多肽的RNA将Cas内切核酸酶例如sRGN多肽引入到待编辑的细胞中的实施方案中是特别有用的,这是因为增加在与编码内切核酸酶的RNA同时引入的指导RNA的半衰期可用于增加指导RNA和编码的Cas内切核酸酶在细胞中共存的时间。

[0321] B. 原间隔区-相邻基序(PAM或PAM序列)

[0322] 对于sRGN多肽的合适原间隔区相邻基序(PAM)序列包括“ANGG”和“NNGG”。

[0323] 切割位点通常存在于PAM序列上游的一到三个碱基对内,例如PAM序列上游的一或三个碱基对内、PAM序列的三个碱基对内或PAM序列“NNGG”上游的三个碱基对内。其它PAM是本领域已知的,且sRGN可以工程改造成靶向这样的PAM。

[0324] C. 在真核细胞中的用途

[0325] 在根据本发明的一个实施方案中,本文所述的sRGN系统可以用于真核细胞,例如哺乳动物细胞中。在某些实施方案中,所述细胞不是人胎儿细胞或不源自人胎儿细胞。在某些实施方案中,所述细胞不是人胚胎细胞或不源自人胚胎细胞。在某些实施方案中,本文中公开的组合物和方法不涉及人胎儿或人胚胎的破坏。在某些实施方案中,所述受试者(或个体,不论在说明书中使用哪个)不是人胎儿或胚胎。

[0326] 本发明的其它特征

[0327] 修饰可以也或替换地用于降低引入细胞中的RNA引出先天免疫应答的可能或程度。如下文和本领域中所述的,在RNA干扰(RNAi)包括小干扰RNA(siRNA)的背景中已被充分表征的这种应答倾向于与RNA的降低的半衰期和/或细胞因子或其它与免疫应答有关的因子的引出有关。

[0328] 还可以对引入细胞中的编码内切核酸酶(例如sRGN多肽)的RNA进行一种或多种类型的修饰,包括、但不限于增强RNA稳定性的修饰(例如通过减少其被细胞中存在的RNA酶的降解)、增强所得产物(即,内切核酸酶)翻译的修饰和/或降低引入细胞中的RNA引出先天免疫应答的可能或程度的修饰。同样可以使用修饰的组合,例如前述和其它的。例如,在CRISPR-Cas的情况下,可以对指导RNA进行一种或多种类型的修饰(包括以上例示的那些),和/或可以对编码CAS内切核酸酶的RNA进行一种或多种类型的修饰(包括以上例示的那些)。

[0329] 作为举例说明,可以通过化学方法容易地合成在CRISPR-Cas系统中使用的指导RNA或其它较小的RNA,从而使得能够容易地并入许多修饰,如下文所举例说明和本领域中所述的。尽管化学合成程序不断地发展,但随着多核苷酸长度显著增加超过一百个左右核苷酸,通过诸如高效液相层析(HPLC,其避免使用凝胶如PAGE)的程序纯化这种RNA倾向于变得更具挑战性。用于产生更大长度的化学修饰的RNA的一种方法是产生两个或多个连接在一起的分子。更长的RNA(例如编码sRGN多肽的那些)更容易酶促产生。尽管通常可用于供酶

促产生的RNA之用的修饰类型较少,但仍存在可用于例如增强稳定性、降低先天免疫应答的可能或程度和/或增强其它属性的修饰,如下文和本领域中进一步描述的;并且正经常开发新颖修饰类型。作为各种类型的修饰的举例说明,尤其是常常与较小的化学合成的RNA一起使用的那些,修饰可以包含在糖的2'位置修饰的一种或多种核苷酸,在某些实施方案中为2'-O-烷基、2'-O-烷基-O-烷基或2'-氟修饰的核苷酸。在某些实施方案中,RNA修饰包括在嘧啶的核糖、碱基残基(basic residues)或RNA的3'端的反向碱基上的2'-氟、2'-氨基和2'O-甲基修饰。这种修饰常规地并入寡核苷酸中,并且已显示这些寡核苷酸具有比2'-脱氧寡核苷酸对给定靶的更高的Tm(即,更高的靶结合亲和力)。

[0330] 已经证实许多核苷酸和核苷修饰使它们所并入的寡核苷酸比天然寡核苷酸对核酸酶消化更具有抗性;这些修饰的寡核苷酸比未修饰的寡核苷酸完整地存活更长的时间。修饰的寡核苷酸的具体实例包括包含修饰的主链的那些,例如,硫代磷酸酯、磷酸三酯、甲基膦酸酯、短链烷基或环烷基糖间键或短链杂原子或杂环糖间键。一些寡核苷酸是具有硫代磷酸酯主链的寡核苷酸,以及具有杂原子主链的那些,特别地是CH_2 -NH-O-CH_2、CH,-N(CH_3)-O-CH_2(称为亚甲基(甲基亚氨基)或MMI主链)、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主链;酰胺主链[参见De Mesmaeker等人, Ace.Chem.Res., 28:366-374 (1995)];吗啉代主链结构(参见Summerton和Weller,美国专利号5,034,506);肽核酸(PNA)主链(其中寡核苷酸的磷酸二酯主链被聚酰胺主链置换,核苷酸直接或间接与聚酰胺主链的氮杂氮原子结合,参见Nielsen等人, Science 1991, 254, 1497)。含磷的键包括、但不限于,硫代磷酸酯,手性硫代磷酸酯,二硫代磷酸酯,磷酸三酯,氨烷基磷酸三酯,甲基和其它烷基膦酸酯包括3'亚烷基膦酸酯和手性膦酸酯,次膦酸酯,氨基磷酸酯包括3'-氨基氨基磷酸酯和氨烷基氨基磷酸酯,硫羰氨基磷酸酯,硫羰烷基膦酸酯,硫羰烷基磷酸三酯,和具有正常3'-5'键的硼烷磷酸酯(boranophosphates),这些的2'-5'连接的类似物,以及具有反向极性的那些,其中相邻的核苷单位对3'-5'至5'-3'或2'-5'至5'-2'连接;参见美国专利号3,687,808;4,469,863;4,476,301;5,023,243;5,177,196;5,188,897;5,264,423;5,276,019;5,278,302;5,286,717;5,321,131;5,399,676;5,405,939;5,453,496;5,455, 233;5,466,677;5,476,925;5,519,126;5,536,821;5,541,306;5,550,111;5,563, 253;5,571,799;5,587,361;和5,625,050。

[0331] 基于吗啉代的寡聚化合物描述于Braasch和David Corey, Biochemistry, 41(14): 4503-4510 (2002);Genesis, 第30卷, 第3期, (2001);Heasman, Dev. Biol., 243:209-214 (2002);Nasevicius等人, Nat.Genet., 26:216-220 (2000);Lacenra等人, Proc.Nat/.Acad.Sci., 97: 9591-9596 (2000);和1991年7月23日授权的美国专利号5,034,506。环己烯基核酸寡核苷酸模拟物描述于Wang等人, J. Am. Chem. Soc., 122: 8595-8602 (2000)。

[0332] 其中不包含磷原子的修饰的寡核苷酸主链具有由短链烷基或环烷基核苷间键、混合的杂原子和烷基或环烷基核苷间键或一种或多种短链杂原子或杂环核苷间键形成的主链。这些包括具有吗啉代键(部分由核苷的糖部分形成)的那些;硅氧烷主链;硫化物、亚砜和砜主链;甲乙酰基(formacetyl)和硫代甲乙酰基(thioformacetyl)主链;亚甲基甲乙酰基和硫代甲乙酰基主链;含烯的主链;氨基磺酸酯主链;亚甲基亚氨基和亚甲基肼基主链;磺酸酯和氨磺酰主链;酰胺主链;和其它具有混合的N、O、S和CH2组分部分的;参见美国专利

号5,034,506;5,166,315;5,185,444;5,214,134;5,216,141;5,235,033;5,264, 562;5,264,564;5,405,938;5,434,257;5,466,677;5,470,967;5,489,677;5,541,307;5,561,225;5,596, 086;5,602,240;5,610,289;5,602,240;5,608,046;5,610,289;5,618,704;5,623, 070;5,663,312;5,633,360;5,677,437;和5,677,439, 它们中的每一篇通过引用并入本文。

[0333] 也可以包括一种或多种取代的糖部分,例如,在2'位置的下述之一:OH、SH、SCH_3、F、OCN、OCH_3、OCH_3 O(CH_2)n CH_3、O(CH_2)n NH_2或O$(CH_2)_n$ CH_3,其中n为1至10;C_1至C_{10}低级烷基,烷氧基烷氧基,取代的低级烷基、烷芳基或芳烷基;Cl;Br;CN;CF3;OCF3;O-、S-或N-烷基;O-、S-或N-链烯基: $SOCH_3$;SO_2 CH_3;ONO_2;NO_2;N_3;NH_2;杂环烷基;杂环烷芳基;氨烷基氨基;聚烷基氨基;取代的甲硅烷基;RNA切割基团;报道基团;嵌入剂;用于改善寡核苷酸的药物代谢动力学性质的基团;或用于改善寡核苷酸的药物动力学性质的基团和具有类似性质的其它取代基。在某些实施方案中,修饰包括2'-甲氧基乙氧基(2'-O-CH2CH2OCH3,也称为2'-O-(2-甲氧基乙基))(Martin等人, Helv. Chim. Acta, 1995, 78, 486)。其它修饰包括2'-甲氧基(2'-O-CH3)、2'-丙氧基(2'-OCH2 CH2CH3)和2'-氟(2'-F)。还可以在寡核苷酸的其它位置进行类似的修饰,特别是3'末端核苷酸上糖的3'位置和5'末端核苷酸的5'位置。寡核苷酸也可以具有糖模拟物,例如环丁基代替呋喃戊糖基。在某些实施方案中,核苷酸单位的糖和核苷间键即主链两者都被新颖基团置换。维持碱基单位用于与适当的核酸靶化合物杂交。一种这种寡聚化合物,已经证实出具有优良的杂交性质的寡核苷酸模拟物,被称为肽核酸(PNA)。在PNA化合物中,寡核苷酸的糖-主链被含有酰胺的主链例如氨乙基甘氨酸主链置换。核苷碱基被保留并直接或间接结合到主链的酰胺部分的氮杂氮原子上。教导PNA化合物的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。PNA化合物的进一步教导可以在Nielsen等人, Science, 254: 1497-1500(1991)中找到。

[0334] 指导RNA还可以额外地或替换地包括核苷碱基(可替代地被称作"碱基")修饰或取代。本文中使用的"未修饰的"或"天然的"核苷碱基包括腺嘌呤(A)、鸟嘌呤(G)、胸腺嘧啶(T)、胞嘧啶(C)和尿嘧啶(U)。修饰的核苷碱基包括仅在天然核酸中偶尔或短暂地发现的核苷碱基,例如次黄嘌呤,6-甲基腺嘌呤,5-Me嘧啶,特别是5-甲基胞嘧啶(也称为5 -甲基-2'脱氧胞嘧啶,并且在本领域中可以被称为5-Me-C),5-羟甲基胞嘧啶(HMC),糖基HMC和龙胆二糖基(gentobiosyl)HMC,以及合成的核苷碱基,例如2-氨基腺嘌呤,2-(甲氨基)腺嘌呤,2-(咪唑基烷基)腺嘌呤,2-(氨烷基氨基)腺嘌呤或其它杂取代的烷基腺嘌呤,2-硫尿嘧啶,2-硫胸腺嘧啶,5-溴尿嘧啶,5-羟甲基尿嘧啶,8-氮鸟嘌呤,7-脱氮鸟嘌呤,N6 (6-氨基己基)腺嘌呤和2.6-二氨基嘌呤。Kornberg, A, DNA Replication, W.H.Freeman & Co., San Francisco, 第75-77页(1980);Gebeyehu等人, Nucl. Acids Res. 15:4513 (1997)。也可以包括本领域已知的"通用"碱基,例如,肌苷。5-Me-C取代已显示将核酸双链体稳定性提高0.6-1.2摄氏度(Sanghvi, Y.S., 在Crooke, S.T.和Lebleu, B., 编, Antisense Research and Applications, CRC Press, Boca Raton, 1993, 第276-278页中)且是碱基取代的实施方案。

[0335] 修饰的核苷碱基包括其它合成的和天然的核苷碱基,例如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-甲基和其它烷

基衍生物,腺嘌呤和鸟嘌呤的2-丙基和其它烷基衍生物,2-硫尿嘧啶,2-硫胸腺嘧啶和2-硫胞嘧啶,5-卤尿嘧啶和胞嘧啶,5-丙炔基尿嘧啶和胞嘧啶,6-偶氮尿嘧啶、胞嘧啶和胸腺嘧啶,5-尿嘧啶(假尿嘧啶),4-硫尿嘧啶,8-卤、8-氨基、8-硫羟、8-硫代烷基、8-羟基和其它a-取代的腺嘌呤和鸟嘌呤,5-卤特别是5-溴、5-三氟甲基和其它5-取代的尿嘧啶和胞嘧啶,7-甲基鸟嘌呤和7-甲基腺嘌呤,8-氮鸟嘌呤和8-氮腺嘌呤,7-脱氮鸟嘌呤和7-脱氮腺嘌呤和3-脱氮鸟嘌呤和3-脱氮腺嘌呤。

[0336] 进一步地,核苷碱基包括在美国专利号3,687,808中公开的那些,在'The Concise Encyclopedia of Polymer Science And Engineering', 第858-859页, Kroschwitz, J.l., 编John Wiley & Sons, 1990中公开的那些,由Englisch等人, Angewandte Chemie, International Edition', 1991, 30, 第613页公开的那些,以及由Sanghvi, Y.S., 第15章, Antisense Research and Applications', 第289- 302页, Crooke, S.T.和Lebleu, B.ea., CRC Press, 1993公开的那些。这些核苷碱基中的某些对于增加本发明的寡聚化合物的结合亲和力特别有用。这些包括5-取代的嘧啶,6-氮嘧啶和N-2、N-6和-O-6取代的嘌呤,包括2-氨丙基腺嘌呤,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将核酸双链体稳定性提高0.6-1.2℃(Sanghvi, Y.S., Crooke, S.T.和Lebleu, B., 编, 'Antisense Research and Applications', CRC Press, Boca Raton, 1993, 第276-278页)且是碱基取代的实施方案,甚至更特别是当与2'-O-甲氧基乙基糖修饰组合时。修饰的核苷碱基描述于美国专利号3,687,808, 以及4,845,205;5,130,302;5,134,066;5,175, 273;5, 367,066;5,432,272;5,457,187;5,459,255;5,484,908;5,502,177;5,525,711;5,552,540;5,587,469;5,596,091;5,614,617;5,681,941;5,750,692;5,763,588;5,830,653;6,005,096;和美国专利申请公开20030158403中。

[0337] 给定寡核苷酸中的所有位置被均匀地修饰不是必需的,并且实际上可以将超过一种上述修饰并入单个寡核苷酸中,或甚至在寡核苷酸内的单个核苷中。

[0338] 在某些实施方案中,将指导RNA和/或编码内切核酸酶诸如sRGN多肽的mRNA(或DNA)化学连接至增强所述寡核苷酸的活性、细胞分布或细胞摄取的一种或多种部分或缀合物。这种部分包括、但不限于,脂质部分,例如胆固醇部分[Letsinger等人, Proc.Nat/.Acad.Sci.USA, 86: 6553-6556 (1989)];胆酸[Manoharan等人, Bioorg.Med.Chem.Let., 4: 1053-1060 (1994)];硫醚,例如己基-S-三苯甲基硫醇[Manoharan等人, Ann.N.Y Acad.Sci., 660: 306-309 (1992)和Manoharan等人, Bioorg.Med.Chem.Let., 3.•2765-2770 (1993)];硫胆固醇(thiocholesterol)[Oberhauser等人,Nucl.Acids Res., 20: 533-538 (1992)];脂族链,例如十二烷二醇或十一烷基残基[Kabanov等人, FEBS Lett., 259: 327-330 (1990)和Svinarchuk等人, Biochimie, 75: 49- 54 (1993)];磷脂,例如二-十六烷基-外消旋-甘油或1,2-二-O-十六烷基-外消旋-甘油-3-H-膦酸三乙铵[Manoharan等人, Tetrahedron Lett., 36:3651-3654 (1995)和Shea等人, Nucl.Acids Res., 18: 3777-3783 (1990)];多胺或聚乙二醇链[Mancharan等人, Nucfeosides & Nucleotides, 14: 969-973 (1995)];金刚烷乙酸[Manoharan等人, Tetrahedron Lett., 36: 3651-3654 (1995)];棕榈基部分[(Mishra等人, Biochim.Biophys.Acta, 1264: 229-237 (1995)];或十八胺或己基氨基-羰基-t羟胆固醇部分[Crooke等人, J.Pharmacal.Exp.Ther., 277: 923-937 (1996)]。也参见美国专

利号4,828,979;4,948,882;5,218,105;5,525,465;5,541,313;5,545,730;5,552, 538;5,578,717, 5,580,731;5,580,731;5,591,584;5,109,124;5,118,802;5,138,045;5,414,077;5,486, 603;5,512,439;5,578,718;5,608,046;4,587,044;4,605,735;4,667,025;4,762, 779;4,789,737;4,824,941;4,835,263;4,876,335;4,904,582;4,958,013;5,082,830;5,112,963;5,214,136;5,082,830;5,112,963;5,214,136;5,245,022;5,254,469;5,258,506;5,262,536;5,272,250;5,292,873;5,317,098;5,371,241, 5,391, 723;5,416,203, 5,451,463;5,510,475;5,512,667;5,514,785;5,565,552;5,567,810;5,574,142;5,585,481;5,587,371;5,595,726;5,597,696;5,599,923;5,599, 928和5,688,941。

[0339] 糖和其它部分可用于将包含核苷酸的蛋白质和复合物(例如阳离子多核糖体和脂质体)靶向特定位点。例如,肝细胞定向转移可以通过脱唾液酸糖蛋白受体(ASGPRs)介导;例如,Hu, 等人,Protein Pept Lett.21(1 0):1025-30 (2014)。可以使用本领域已知的和经常开发的其它系统将本情况下使用的生物分子和/或其复合物靶向特定的感兴趣的靶细胞。

[0340] 这些靶向部分或缀合物可以包括与官能团例如伯或仲羟基共价结合的缀合物基团。本发明的缀合物基团包括嵌入剂、报道分子、多胺、聚酰胺、聚乙二醇、聚醚、增强寡聚体的药物动力学性质的基团和增强寡聚体的药物代谢动力学性质的基团。典型的缀合物基团包括胆固醇、脂质、磷脂、生物素、吩嗪、叶酸、菲啶、蒽醌、吖啶、荧光素、罗丹明、香豆素和染料。在本发明的上下文中,增强药物动力学性质的基团包括改善摄取、增强对降解的抗性和/或增强与靶核酸的序列特异性杂交的基团。在本发明的上下文中,增强药物代谢动力学性质的基团包括改善本发明化合物的摄取、分布、代谢或排泄的基团。在1992年10月23日提交的国际专利申请No.PCT/US92/09196和美国专利号6,287,860中公开了代表性的缀合物基团,其通过引用并入本文。缀合物部分包括、但不限于,脂质部分,例如胆固醇部分,胆酸,硫醚,例如己基-5-三苯甲基硫醇,硫胆固醇,脂族链,例如十二烷二醇或十一烷基残基,磷脂,例如二-十六烷基-外消旋-甘油或1,2-二-O-十六烷基-外消旋-甘油-3-H-膦酸三乙铵,多胺或聚乙二醇链,或金刚烷乙酸,棕榈基部分,或十八胺或己基氨基-羰基-羟胆固醇部分。参见,例如,美国专利号4,828,979;4,948,882;5,218,105;5,525,465;5,541,313;5,545,730;5,552,538;5,578,717, 5,580,731;5,580,731;5,591,584;5,109,124;5,118,802;5,138,045;5,414,077;5,486,603;5,512,439;5,578,718;5,608,046;4,587,044;4,605,735;4,667,025;4,762,779;4,789,737;4,824,941;4,835,263;4,876,335;4,904,582;4,958,013;5,082,830;5,112,963;5,214,136;5,082,830;5,112,963;5,214,136;5,245,022;5,254,469;5,258,506;5,262,536;5,272,250;5,292,873;5,317,098;5,371,241, 5,391,723;5,416,203, 5,451,463;5,510,475;5,512,667;5,514,785;5,565,552;5,567,810;5,574,142;5,585,481;5,587,371;5,595,726;5,597,696;5,599,923;5,599,928和5,688,941。

[0341] 不太适合化学合成并且通常通过酶促合成生产的较长的多核苷酸可以通过各种方式修饰。这种修饰可以包括,例如,某些核苷酸类似物的引入,在分子的5'或3'端并入特定序列或其它部分,以及其它修饰。作为举例说明,编码sRGN多肽的mRNA具有大约3 kb至3.3 kb的长度,并且可以通过体外转录合成。对mRNA的修饰可以应用于例如增加其翻译或稳定性(例如通过增加其对细胞降解的抗性),或降低RNA引出先天免疫应答的倾向,这种可

能在引入外源RNA(特别是较长的RNA如编码sRGN的那些)后发生。

[0342] 本领域中已经描述了许多这种修饰,例如聚腺苷酸尾,5'帽类似物(例如抗反向帽类似物(Anti Reverse Cap Analog)(ARCA)或m7G(5')ppp(5')G (mCAP)),修饰的5'或3'非翻译区(UTRs),修饰的碱基的使用(例如假-UTP、2-硫-UTP、5-甲基胞苷-5'-三磷酸(5-甲基-CTP)或N6-甲基-ATP),或用磷酸酶处理以去除5'末端磷酸。这些和其它修饰是本领域已知的,并且正在开发RNA的新修饰。

[0343] 可以得到经修饰的RNA,例如,从商业供应商,包括例如Trilink Biotech、Axolabs、Bio-Synthesis Inc.、Dharmacon和许多其它。如Trilink所述的,例如,5-甲基-CTP可用于赋予所期望的特征,例如增加的核酸酶稳定性,增加的翻译或降低的先天免疫受体与体外转录的RNA的相互作用。5'-甲基胞苷-5'-三磷酸(5-甲基-CTP)、N6-甲基-ATP以及假-UTP和2-硫-UTP也已显示在增强翻译的同时减少培养中和体内的先天免疫刺激,如下文参考的Konmann等人和Warren等人的出版物所举例说明的。

[0344] 已经证实体内递送的化学修饰的mRNA可以用于实现改善的治疗效果;参见,例如,Kormann等人, Nature Biotechnology 29, 154-157 (2011)。这种修饰可用于例如增加RNA分子的稳定性和/或降低其免疫原性。使用化学修饰,例如假-U、N6-甲基-A、2-硫-U和5-甲基-C,发现分别用2-硫-U和5-甲基-C仅取代四分之一的尿苷和胞苷残基就导致小鼠中mRNA的toll-样受体(TLR)介导的识别显著降低。通过减少先天免疫系统的激活,这些修饰可因此用于有效地提高体内mRNA的稳定性和寿命;例如,Konmann等人, 出处同上。

[0345] 还已显示重复施用并入了设计用于绕过先天抗病毒反应的修饰的合成信使RNA可以将分化的人细胞重编程序为多潜能性。参见,例如Warren等人, Cell Stem Cell, 7(5): 618-30 (2010)。充当初级重编程序蛋白的这种修饰的mRNA可以是对多种人细胞类型进行重编程序的有效手段。这种细胞称为诱导的多潜能性干细胞(iPSC),且发现并入了5-甲基-CTP、假-UTP和抗反向帽类似物(ARCA)的酶促合成的RNA可用于有效逃避细胞的抗病毒反应;例如,Warren等人, 出处同上。本领域中描述的多核苷酸的其它修饰包括,例如,聚腺苷酸尾的使用,5'帽类似物(例如m7G(5')ppp(5')G (mCAP))的添加,5'或3'非翻译区(UTR)的修饰或用磷酸酶处理以去除5'末端磷酸-且正经常开发新方法。

[0346] 已经连同RNA干扰(RNAi)的修饰,包括小干扰RNA(siRNA),开发了适用于产生供本文之用的修饰的RNA的许多组合物和技术。siRNA在体内引起特殊的挑战,这是因为它们通过mRNA干扰对基因沉默的作用通常是瞬时的,其可能需要重复施用。此外,siRNA是双链RNA(dsRNA),且哺乳动物细胞具有已进行进化以检测和中和dsRNA的免疫应答,所述dsRNA经常是病毒感染的副产物。因此,存在哺乳动物酶如PKR(dsRNA-应答激酶),和可能的视黄酸诱导型基因I(RIG-I),其可以介导对dsRNA的细胞应答,以及Toll-样受体(例如TLR3、TLR7和TLR8),其可以响应于这种分子而触发细胞因子的诱导;例如,Angart等人, Pharmaceuticals (Basel) 6(4): 440-468 (2013);Kanasty等人, Molecular Therapy 20(3): 513-524 (2012);Burnett等人, Biotechnol J.6(9):1130-46 (2011);Judge和Maclachlan, Hum Gene Ther 19(2):111-24 (2008)的综述;以及其中引用的参考文献。

[0347] 本领域描述的修饰已经开发并将应用于增强RNA稳定性、减少先天免疫应答和/或获得其它可以在如本文所述的将多核苷酸引入人细胞中的方面有用的益处;例如Whitehead KA等人, Annual Review of Chemical and Biomolecular Engineering, 2:

77-96 (2011);Gaglione和Messere, Mini Rev Med Chem, 10(7):578-95 (2010);Chernolovskaya等人, Curr Opin Mol Ther., 12(2):158-67 (2010);Deleavey等人, Curr Protoc Nucleic Acid Chem第16章:Unit 16.3 (2009);Behlke, Oligonucleotides 18(4):305-19 (2008): Fucini等人, Nucleic Acid Ther 22(3): 205-210 (2012);Bremsen等人, Front Genet 3:154 (2012)的综述。

[0348] 如上文所指出的,有许多修饰的RNA的商业供应商,其中许多专门做设计用于提高siRNA的有效性的修饰。基于文献中报道的各种发现,提供了多种方法。例如,Dharmacon指出,用硫置换非桥联氧(硫代磷酸酯,PS)已被广泛用于改善siRNA的核酸酶抗性,如Kale, Nature Reviews Drug Discovery 11:125-140 (2012)所报道的。已经报道核糖2'-位置的修饰在增加双链体稳定性(Tm)的同时改善核苷酸间磷酸键的核酸酶抗性,其也已显示提供保护免于免疫激活。中等PS主链修饰与小的、良好耐受的2'-取代(2'-O-甲基、2'-氟、2'-氢(Hydro))的组合已与用于体内应用的高度稳定的siRNA有关联,如Soutschek等人Nature 432:173-178 (2004)所报道的;且已报道2'-O-甲基修饰在改善稳定性中有效,如Volkov, Oligonucleotides 19:191-202 (2009)所报道的。关于减少先天免疫应答的诱导,已经报道用2'-O-甲基、2'-氟、2'-氢修饰具体序列在通常保持沉默活性的同时减少TLR7/TLR8相互作用;例如Judge等人, Mol.Ther.13:494-505 (2006);和Cekaite等人, J.Mol.Biol.365:90-108 (2007)。还已显示额外的修饰,例如2-硫尿嘧啶、假尿嘧啶、5-甲基胞嘧啶、5-甲基尿嘧啶和N6-甲基腺苷,使由TLR3、TLR7和TLR8介导的免疫作用减至最小;例如Kariko, K.等人, Immunity 23:165-175 (2005)。

[0349] 如本领域中也已知的并且可商购获得的,许多缀合物可以应用于多核苷酸,例如供本文之用的RNA,其可以增强它们的递送和/或被细胞摄取,包括例如胆固醇、生育酚和叶酸、脂质、肽、聚合物接头和适体;例如Winkler, Ther Deliv. 4:791-809 (2013)的综述,以及其中引用的参考文献。

[0350] *模拟物*

[0351] 核酸可以是核酸模拟物。当应用于多核苷酸时术语"模拟物"意图包括其中仅呋喃糖环或呋喃糖环和核苷酸间键两者均被非呋喃糖基团置换的多核苷酸,仅呋喃糖环的置换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杂交。一种这种核酸,已显示具有优良的杂交性质的多核苷酸模拟物,被称为肽核酸(PNA)。在PNA中,多核苷酸的糖-主链被含酰胺的主链特别是氨乙基甘氨酸主链置换。核苷酸被保留并直接或间接结合到主链的酰胺部分的氮杂氮原子上。

[0352] 已报道具有优良杂交性质的一种多核苷酸模拟物是肽核酸(PNA)。PNA化合物中的主链是两个或更多个连接的氨乙基甘氨酸单位,其施用PNA含酰胺的主链。杂环碱基部分直接或间接结合到主链的酰胺部分的氮杂氮原子上。描述PNA化合物的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。

[0353] 已研究的另一类多核苷酸模拟物基于具有附着到吗啉代环的杂环碱基的连接的吗啉代单位(吗啉代核酸)。已经报道了连接吗啉代核酸中的吗啉代单体单位的许多连接基团。已选择了一类连接基团以产生非离子型寡聚化合物。非离子型基于吗啉代的寡聚化合物不太可能与细胞蛋白具有不期望的相互作用。基于吗啉代的多核苷酸是寡核苷酸的非离子型模拟物,其不太可能与细胞蛋白质形成不期望的相互作用(Braasch, D. A., 和

Corey, D. R., Biochemistry, 2002, 41(14), 4503-4510)。基于吗啉代的多核苷酸在美国专利号5,034,506中公开。已经制备了多核苷酸的吗啉代类内的多种化合物,其具有连接单体亚单位的多种不同的连接基团。

[0354] 另一类多核苷酸模拟物被称为环己烯基核酸(GeNA)。通常存在于DNA/RNA分子中的呋喃糖环被环己烯基环置换。已经按照经典的亚磷酰胺化学原理制备了GeNA DMT保护的亚磷酰胺单体,并将其用于寡聚化合物的合成。已经制备并研究了完全修饰的GeNA寡聚化合物和使具体位置被GeNA修饰的寡核苷酸(参见Wang等人, J. Am. Chem. Soc., 2000, 122, 8595-8602)。通常,将GeNA单体并入DNA链增加了DNA/RNA杂交体的其稳定性。GeNA寡腺苷酸与RNA和DNA互补体(complements)形成具有与天然复合物相似稳定性的复合物。NMR和圆二色性表明,将GeNA结构并入天然核酸结构中的研究从容易的构象适应开始。

[0355] 进一步的修饰包括锁定核酸(LNA),其中2'-羟基与糖环的4'碳原子连接,从而形成2'-C,4'-C-氧亚甲基键,从而形成双环糖部分。键可以是亚甲基(-CHz-),桥连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)。LNA和LNA类似物显示与互补的DNA和RNA非常高的双链体热稳定性(T_m = +3至+10℃)、对3'-核酸外切降解的稳定性和良好的溶解度性质。已经描述了含有LNA的有效且无毒性的反义寡核苷酸(Wahlestedt等人, Proc. Natl. Acad. Sci. U.S.A, 2000, 97, 5633-5638)。

[0356] 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,以及它们的寡聚化和核酸识别性质(Koshkin等人, Tetrahedron, 1998, 54, 3607-3630)。LNA及其制备也在WO 98/39352和WO 99/14226中描述。

[0357] *修饰的糖部分*

[0358] 核酸还可包含一种或多种取代的糖部分。合适的多核苷酸包含选自下述的糖取代基基团:OH;F;O-、S-或N-烷基;O-、S-或N-链烯基;O-、S-或N-炔基;或O-烷基-O-烷基,其中烷基、链烯基和炔基可以是取代或未取代的C_1 -C_{10}烷基或C_2 -C_{10}链烯基和炔基。特别合适的是O((CH2)nO)mCH3、O(CHz)nOCH3、O(CHz)nNH2、O(CH2)CH3、O(CH2)nONH2和O(CH2)nON((CH2)nCH3)2,其中n和m为1至约10。其它合适的多核苷酸包含选自下述的糖取代基:C_1至C_{10}低级烷基、取代的低级烷基、链烯基、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_z、NO_2、N_3、NH2、杂环烷基、杂环烷芳基、氨烷基氨基、聚烷基氨基、取代的甲硅烷基、RNA切割基团、报道基团、嵌入剂、用于改善寡核苷酸的药物代谢动力学性质的基团、或用于改善寡核苷酸的药物动力学性质的基团和具有类似性质的其它取代基。合适的修饰包括2'-甲氧基乙氧基2'-O-CH2 CHzOCH3,也称为-2'-O-(2-甲氧基乙基)或2'-MOE)(Martin等人, Hely.Chim.Acta, 1995, 78, 486-504),即,烷氧基烷氧基基团。进一步的合适的修饰包括2'-二甲氨基氧乙氧基,即,O(CHz)zON(CH3)z基团,也称为2'-DMAOE,如在下文实施例中所述的,和2'-二甲氨基乙氧基乙氧基(本领域中也称为2'-O-二甲基-氨基-乙氧基-乙基或2'- DMAEOE),即,2'-O-CHz-O-CHz-N(CH3)z。

[0359] 其它合适的糖取代基基团包括甲氧基(-O-CH_3)、氨基丙氧基(-O- $CH_zCH_zCH_zNH_z$)、烯丙基(-CH_z-CH=CH_z)、-O-烯丙基(-O-CH_z-CH=CH_z)和氟(F)。2'-糖取代基基团可以在阿拉伯糖(上)位置或核糖(下)位置。合适的2'-阿拉伯糖修饰是2'-F。还可以在寡聚化合物上的其它位置进行类似的修饰,特别是3'末端核苷上或2'-5'链接的寡核苷酸中糖的3'位置和5'末端核苷酸的5'位置。寡聚化合物也可以具有糖模拟物,例如环丁基部分代替呋喃戊糖

基糖。

[0360] *碱基修饰和取代*

[0361] 核酸还可以包括核苷碱基(在本领域中经常简称为"碱基")修饰或取代。本文中使用的"未修饰的"或"天然的"核苷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)。修饰的核苷碱基包括其它合成的和天然的核苷碱基,例如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-甲基和其它烷基衍生物,腺嘌呤和鸟嘌呤的2-丙基和其它烷基衍生物,2-硫尿嘧啶,2-硫胸腺嘧啶和2-硫胞嘧啶,5-卤尿嘧啶和胞嘧啶,5-丙炔基(-C=C-CH_3)尿嘧啶和胞嘧啶和嘧啶碱基的其它炔基衍生物,6-偶氮尿嘧啶、胞嘧啶和胸腺嘧啶,5-尿嘧啶(假尿嘧啶),4-硫尿嘧啶,8-卤、8-氨基、8-硫羟、8-硫代烷基、8-羟基和其它8-取代的腺嘌呤和鸟嘌呤,5-卤特别是5-溴、5-三氟甲基和其它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-氮鸟嘌呤和8-氮腺嘌呤,7-脱氮鸟嘌呤和7-脱氮腺嘌呤和3-脱氮鸟嘌呤和3-脱氮腺嘌呤。进一步的修饰的核苷碱基包括三环嘧啶,例如吩噁嗪胞苷(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-封条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)。

[0362] 杂环碱基部分还可以包括其中嘌呤或嘧啶碱基被其它杂环置换的那些,例如7-脱氮-腺嘌呤、7-脱氮鸟苷、2-氨基吡啶和2-吡啶酮。进一步地,核苷碱基包括在美国专利号3,687,808中公开的那些,在The Concise Encyclopedia Of Polymer Science And Engineering, 第858-859页, Kroschwitz, J.1., 编John Wiley & Sons, 1990中公开的那些,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。这些核苷碱基中的某些对于增加寡聚化合物的结合亲和力有用。这些包括5-取代的嘧啶,6-氮嘧啶和N-2、N-6和-O-6取代的嘌呤,包括2-氨丙基腺嘌呤,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将核酸双链体稳定性提高0.6-1.2 oc(Sanghvi等人, 编, Antisense Research and Applications, CRC Press, Boca Raton, 1993, 第276-278页)且是合适的碱基取代,例如,当与2'-O-甲氧基乙基糖修饰组合时。

[0363] "互补的"是指通过碱基堆积和特异性氢键合在包含天然或非天然存在的(例如如上文所述的修饰的)碱基(核苷)或其类似物的两个序列之间配对的能力。例如,如果核酸的一个位置处的碱基能够与靶的相应位置处的碱基氢键合,则认为所述碱基在那个位置处彼此互补。核酸可包含通用碱基,或惰性无碱基间隔区,其对氢键合不提供正或负贡献。碱基配对可能包括典型的沃森-克里克碱基配对和非沃森-克里克碱基配对(例如wobble碱基配对和Hoogsteen碱基配对)。

[0364] 应当理解,对于互补碱基配对,腺苷型碱基(A)与胸苷型碱基(T)或尿嘧啶型碱基(U)互补,胞嘧啶型碱基(C)与鸟苷型碱基(G)互补,并且诸如诸如3-硝基吡咯或5-硝基吲哚的通用碱基可以与任何A、C、U或T杂交并被认为与其互补。Nichols等人, Nature, 1994;369:492-493和Loakes等人, Nucleic Acids Res., 1994;22:4039-4043。肌苷(I)在本领

域中也已被认为是通用碱基,并且被认为与任何A、C、U或T互补。参见Watkins和Santalucia, Nucl.Acids Research, 2005;33 (19): 6258-6267。

[0365] *缀合物*

[0366] 核酸的另一种可能的修饰包括将一种或多种增强寡核苷酸的活性、细胞分布或细胞摄取的部分或缀合物化学连接至多核苷酸。这些部分或缀合物可以包括与官能团例如伯或仲羟基共价结合的缀合物基团。缀合物基团包括、但不限于,嵌入剂、报道分子、多胺、聚酰胺、聚乙二醇、聚醚、增强寡聚体的药物动力学性质的基团和增强寡聚体的药物代谢动力学性质的基团。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、菲啶、蒽醌、吖啶、荧光素、罗丹明、香豆素和染料。增强药物动力学性质的基团包括改善摄取、增强对降解的抗性和/或增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性质的基团包括改善核酸的摄取、分布、代谢或排泄的基团。

[0367] 缀合物部分包括、但不限于脂质部分,例如胆固醇部分(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),硫胆固醇(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1,2-二-O-十六烷基-外消旋-甘油-3-H-膦酸三乙铵(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, 969-97),或金刚烷乙酸(Manoharan等人, Tetrahedron Lett., 1995, 36, 36513654),棕榈基部分(Mishra等人, Biochim.Biophys.Acta, 1995, 1264, 229-237),或十八胺或己基氨基-羰基-羟胆固醇部分(Crooke等人, J.Pharmacal.Exp.Ther., 1996,277, 923-937)。

[0368] 缀合物可以包括“蛋白转导结构域”或PTD(也称为CPP-细胞穿透肽),其可以指促进通过脂双层、微团、细胞膜、细胞器膜或小泡膜的多肽、多核苷酸、碳水化合物或有机或无机化合物。与从小的极性分子一直到大的大分子和/或纳米颗粒的另一种分子附着的PTD促进分子通过膜,例如从细胞外空间到细胞内空间,或从细胞溶胶到细胞器内。在某些实施方案中,PTD共价连接至外源多肽(例如sRGN多肽)的氨基末端。在某些实施方案中,PTD共价连接至外源多肽(例如sRGN多肽)的C-末端或N-末端。在某些实施方案中,PTD共价连接至核酸(例如,指导RNA、编码指导RNA的核酸、编码sRGN多肽的核酸等)。示例性的PTD包括、但不限于最小的十一肽蛋白转导结构域(例如,对应于包含YGRKKRRQRRR (SEQ ID NO: 126)基序的HIV-1 TAT的残基47-57(用于鉴定但不是为了在本申请中公开起见);包含足以指导进入细胞的许多精氨酸的聚精氨酸序列(例如3、4、5、6、7、8、9、10或10-50个精氨酸);VP22结构域(Zender等人(2002) Cancer Gene Ther.9(6):489-96);果蝇触角足蛋白转导结构域(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. USA97:13003-13008);在某些实施方案中,PTD是能被活化的CPP(ACPP)

(Aguilera等人(2009) lntegr Biol (Camb) June;1(5-6): 371-381)。ACPP包含通过可切割的接头连接至匹配的聚阴离子(例如Glu9或"E9")的聚阳离子CPP(例如Arg9或"R9"),其将净电荷降低至几乎为零,且从而抑制黏附和摄取到细胞中。在切割接头后,聚阴离子被释放,从而局部地暴露了聚精氨酸及其固有的黏着性,从而"活化"ACPP通过膜。在某些实施方案中,对PTD进行化学修饰以增加PTD的生物利用率。示例性修饰公开于Expert Opin Drug Deliv. 2009 6(11):1195-205。

[0369] *编码指导RNA和/或sRGN多肽的核酸*

[0370] 本公开内容提供了包含编码指导RNA和/或sRGN多肽的核苷酸序列的核酸。在某些实施方案中,编码指导RNA的核酸是表达载体,例如重组表达载体。

[0371] 在某些实施方案中,方法包括接触靶DNA或将一种或多种包含编码指导RNA和/或sRGN多肽的核苷酸序列的核酸引入细胞(或细胞群体)。在某些实施方案中,包含靶DNA的细胞是体外的。在某些实施方案中,包含靶DNA的细胞是体内的。包含编码指导RNA和/或sRGN多肽的核苷酸序列的合适核酸包括表达载体,其中包含编码指导RNA和/或sRGN多肽的核苷酸序列的表达载体是"重组表达载体"。

[0372] 在某些实施方案中,重组表达载体是病毒构建体,例如重组腺相关病毒构建体(例如美国专利号7,078,387)、重组腺病毒构建体、重组慢病毒构建体、重组反转录病毒构建体等。合适的表达载体包括、但不限于,病毒载体(例如基于痘苗病毒的病毒载体;脊髓灰质炎病毒;腺病毒(例如Li等人, Invest Opthalmol Vis Sci 35:2543 2549, 1994;Borras等人, Gene Ther 6:515 524, 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(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 Ther4:683 690, 1997, Rolling等人, Hum Gene Ther 10:641 648, 1999;Ali等人, Hum Mol Genet 5:591 594, 1996;Srivastava在WO 93/09239中,Samulski等人, J.Vir.(1989) 63:3822-3828;Mendelson等人, Viral.(1988) 166:154-165;和Flotte等人, PNAS (1993) 90:10613-10617);SV40;单纯疱疹病毒;人免疫缺陷病毒(例如Miyoshi等人, PNAS 94:10319 23, 1997;Takahashi等人, J Viral 73:7812 7816, 1999);反转录病毒载体(例如鼠白血病病毒,脾脏坏死病毒,以及衍生自反转录病毒的载体,例如劳斯肉瘤病毒,Harvey肉瘤病毒,禽类白血病病毒,慢病毒,人免疫缺陷病毒,骨髓增生肉瘤病毒和乳腺肿瘤病毒);等等。

[0373] 许多合适的表达载体是本领域技术人员已知的,并且许多是可商购获得的。作为实例提供下述载体;对于真核宿主细胞:pXT1、pSGS(Stratagene)、pSVK3、pBPV、pMSG和pSVLSV40(Phanmacia)。但是,可以使用任何其它载体,只要它与宿主细胞相容即可。取决于所使用的宿主/载体系统,可以在表达载体中使用许多合适的转录和翻译控制元件中的任何一种,包括组成型和诱导型启动子、转录增强子元件、转录终止子等(例如Bitter等人(1987) Methods in Enzymology, 153:516-544)。

[0374] 在某些实施方案中,编码指导RNA和/或sRGN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,例如启动子。转录控制元件可以在真核细胞,例如哺乳动物细胞;或原核细胞(例如细菌或古细菌细胞)中起作用。在某些实施方案中,编码指导RNA和/或

sRGN多肽的核苷酸序列可操作地连接至多个控制元件,其允许在原核和真核细胞两者中表达编码指导RNA和/或sRGN多肽的核苷酸序列。

[0375] 合适的真核启动子(在真核细胞中起作用的启动子)的非限制性例子包括来自巨细胞病毒(CMV)立即早期、单纯疱疹病毒(HSV)胸苷激酶、早期和晚期SV40、来自反转录病毒的长末端重复序列(LTRs)和小鼠金属硫蛋白-1的那些。合适的载体和启动子的选择是在本领域普通技术水平内。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。表达载体还可以包括用于扩增表达的合适序列。表达载体还可以包括编码与sRGN多肽融合的蛋白标签(例如6xHis标签、血凝素标签、绿色荧光蛋白等)的核苷酸序列,从而结果产生嵌合多肽。

[0376] 在某些实施方案中,编码指导RNA和/或sRGN多肽的核苷酸序列可操作地连接至诱导型启动子。在某些实施方案中,编码指导RNA和/或sRGN多肽的核苷酸序列可操作地连接至组成型启动子。

[0377] 将核酸引入宿主细胞的方法是本领域已知的,并且可以使用这样的方法将核酸(例如表达构建体)引入细胞。合适的方法包括,例如病毒或噬菌体感染、转染、接合、原生质体融合、脂质转染法、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、基因枪技术、磷酸钙沉淀、直接显微注射、纳米颗粒介导的核酸递送(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。

[0378] *嵌合多肽*

[0379] 本公开内容提供了嵌合定位修饰多肽,其包含衍生自sRGN多肽的序列。嵌合定位修饰多肽与指导RNA(上文所述的)相互作用(例如结合)。指导RNA将嵌合定位修饰多肽指导至靶DNA内的靶序列(例如染色体序列或染色体外序列,例如附加型序列、小环序列、线粒体序列、叶绿体序列、无细胞多核苷酸等)。嵌合定位修饰多肽修饰靶DNA(例如靶DNA的切割或甲基化作用)和/或与靶DNA有关的多肽(例如组蛋白尾的甲基化作用或乙酰化作用)。嵌合定位修饰多肽在本文中也称为“嵌合定位多肽”或“嵌合RNA结合位点指导的修饰多肽”。

[0380] 嵌合定位修饰多肽包含两个部分,RNA-结合部分和活性部分。嵌合定位修饰多肽包含衍生自至少两种不同多肽的氨基酸序列。嵌合定位修饰多肽可以包含修饰的和/或天然存在的多肽序列(例如,来自修饰或未修饰的sRGN蛋白的第一氨基酸序列;和除sRGN蛋白之外的第二氨基酸序列)。

[0381] *RNA-结合部分*

[0382] 在某些实施方案中,嵌合定位修饰多肽的RNA-结合部分包含与SEQ ID NO: 1-12中的任一个所示的多肽的RNA-结合部分具有至少75%、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。

[0383] *活性部分*

[0384] 除RNA-结合部分外,嵌合定位修饰多肽包含“活性部分”。在某些实施方案中,嵌合定位修饰多肽的活性部分包含定位修饰多肽(例如,Cas9内切核酸酶或sRGN)的活性部分。在其它实施方案中,主题嵌合定位修饰多肽的活性部分包含定位修饰多肽的天然存在的活性部分的修饰的氨基酸序列(例如,取代、缺失、插入)。感兴趣的天然存在的活性部分衍生自本领域已知的定位修饰多肽。嵌合定位修饰多肽的活性部分是可变的,并且可以包含可

用于本文公开的方法的任何异源多肽序列。在某些实施方案中,嵌合定位修饰多肽的活性部分包含与SEQ ID NO: 1-12中的任一个的活性部分具有至少90%同一性的sRGN多肽的一部分。在某些实施方案中,嵌合定位修饰多肽包含:(i)与指导RNA相互作用的RNA-结合部分,其中所述指导RNA包含与靶DNA中的序列互补的核苷酸序列;(ii)表现出定位酶活性(例如,RNA切割活性)的活性部分,其中酶活性的位点由前-crRNA的重复序列形成的回文发夹结构确定并在发夹的上游4个核苷酸切割前-crRNA,从而产生由重复序列间隔区(5'-3')组成的crRNA的中间形式;和(iii)表现出定位酶活性(例如,DNA切割活性)的活性部分,其中酶活性的位点由指导RNA确定。

[0385] *示例性的嵌合定位修饰多肽*

[0386] 在某些实施方案中,嵌合定位修饰多肽的活性部分包含sRGN蛋白的修饰形式,包括任何sRGN直系同源物的修饰形式。在某些情况下,sRGN蛋白的修饰形式包含减少或增加sRGN蛋白的天然存在的核酸酶活性的氨基酸变化(例如,缺失、插入或取代)。例如,在某些情况下,sRGN蛋白的修饰形式具有小于50%、小于40%、小于30%、小于20%、小于10%、小于5%或小于1%的相应野生型sRGN多肽的核酸酶活性。在某些情况下,sRGN多肽的修饰形式没有显著的核酸酶活性。在其它情况下,它可能具有50%、2-倍、4-倍或最多到10-倍以上的核酸酶活性。

[0387] 在某些情况下,嵌合定位修饰多肽包含与SEQ ID NO: 1-12中的任一个具有至少75%、至少80%、至少85%、至少90%、至少95%、至少99%或100%氨基酸序列同一性的氨基酸序列。在某些实施方案中,定位修饰多肽的活性部分包含具有DNA-修饰活性和/或转录因子活性和/或DNA-相关多肽修饰活性的异源多肽。在某些情况下,异源多肽置换提供核酸酶活性的sRGN多肽的部分。在其它实施方案中,定位修饰多肽既包含正常提供核酸酶活性的sRGN多肽的部分(且所述部分可以是完全活性的,或者可以代之以被修饰为具有小于100%的相应野生型活性),又包含异源多肽。换句话说,在某些情况下,嵌合定位修饰多肽是既包含正常提供核酸酶活性的sRGN多肽部分又包含异源多肽的融合多肽。在其它情况下,嵌合定位修饰多肽是包含sRGN多肽的活性部分的修饰的变体(例如氨基酸改变、缺失、插入)和异源多肽的融合多肽。在再其它情况下,嵌合定位修饰多肽是融合多肽,其包含异源多肽和天然存在的或修饰的定位修饰多肽的RNA-结合部分。

[0388] 例如,在嵌合sRGN蛋白中,可以将天然存在的(或修饰的,例如突变、缺失、插入)sRGN多肽融合至异源多肽序列(即,来自除sRGN之外的蛋白的多肽序列或来自另一种生物的多肽序列)。异源多肽序列可以表现出也将由嵌合sRGN蛋白表现出的活性(例如,酶活性)(例如,甲基转移酶活性、乙酰转移酶活性、激酶活性、泛素化活性等)。异源核酸序列可以与另一核酸序列连接(例如通过遗传工程)以产生编码嵌合多肽的嵌合核苷酸序列。在某些实施方案中,通过将sRGN多肽(例如野生型sRGN或sRGN变体,例如具有降低或失活的核酸酶活性的sRGN)与提供亚细胞定位的异源序列(例如用于靶向细胞核的核定位信号(NLS);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)融合来产生嵌合sRGN多肽。在某些实施方案中,异源序列可以提供用于易于追踪或纯化的标签(例如荧光蛋白,例如绿色荧光蛋白(GFP)、YFP、RFP、CFP、mCherry、tdTomato等;HIS标签,例如6XHis标签;血凝素(HA)标签;FLAG标签;Myc标签;等)。在某些实施方案中,异源序列可提供增加或降低的稳定性。在某些实施方案中,异源序列可提供结合结构域(例如,以提供嵌

合的sRGN多肽结合另一种感兴趣的蛋白的能力,例如DNA或组蛋白修饰蛋白、转录因子或转录阻遏蛋白、募集蛋白等)或感兴趣的核苷酸(例如,适体或核苷酸结合蛋白的靶位点)。

[0389] *编码嵌合定位修饰多肽的核酸*

[0390] 本公开内容提供了包含编码嵌合定位修饰多肽的核苷酸序列的核酸。在某些实施方案中,包含编码嵌合定位修饰多肽的核苷酸序列的核酸是表达载体,例如重组表达载体。

[0391] 在某些实施方案中,方法包括接触靶DNA或将一种或多种包含嵌合定位修饰多肽的核酸引入细胞(或细胞群体)。包含编码嵌合定位修饰多肽的核苷酸序列的合适核酸包括表达载体,其中包含编码嵌合定位修饰多肽的核苷酸序列的表达载体是"重组表达载体"。

[0392] 在某些实施方案中,重组表达载体是病毒构建体,例如重组腺相关病毒构建体(例如美国专利号7,078,387)、重组腺病毒构建体、重组慢病毒构建体等。合适的表达载体包括、但不限于,病毒载体(例如基于痘苗病毒的病毒载体;脊髓灰质炎病毒;腺病毒(例如Li等人, Invest Opthalmol Vis Sci 35:2543 2549, 1994;Borras等人, Gene Ther 6:515 524, 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(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在WO 93/09239中,Samulski等人, J.Vir.(1989) 63:3822-3828;Mendelson等人, Viral.(1988) 166:154-165;和Flotte等人, PNAS (1993) 90:10613-10617);SV40;单纯疱疹病毒;人免疫缺陷病毒(例如Miyoshi等人, PNAS 94:10319 23, 1997;Takahashi等人, J Viral73:7812 7816, 1999);反转录病毒载体(例如鼠白血病病毒,脾脏坏死病毒,以及衍生自反转录病毒的载体,例如劳斯肉瘤病毒,Harvey肉瘤病毒,禽类白血病病毒,慢病毒,人免疫缺陷病毒,骨髓增生肉瘤病毒和乳腺肿瘤病毒);等等。

[0393] 许多合适的表达载体是本领域技术人员已知的,并且许多是可商购获得的。作为实例提供下述载体;对于真核宿主细胞:pXT1、pSGS(Stratagene)、pSVK3、pBPV、pMSG和pSVLSV40(Pharmacia)。但是,可以使用任何其它载体,只要它与宿主细胞相容即可。

[0394] 取决于所使用的宿主/载体系统,可以在表达载体中使用许多合适的转录和翻译控制元件中的任何一种,包括组成型和诱导型启动子、转录增强子元件、转录终止子等(参见,例如Bitter等人(1987) Methods in Enzymology, 153:516-544)。

[0395] 在某些实施方案中,编码嵌合定位修饰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,例如启动子。转录控制元件可以在真核细胞,例如哺乳动物细胞;或原核细胞(例如细菌或古细菌细胞)中起作用。在某些实施方案中,编码嵌合定位修饰多肽的核苷酸序列可操作地连接至多个控制元件,其允许在原核和真核细胞两者中表达编码嵌合定位修饰多肽的核苷酸序列。

[0396] 合适的真核启动子(在真核细胞中起作用的启动子)的非限制性例子包括来自巨细胞病毒(CMV)立即早期、单纯疱疹病毒(HSV)胸苷激酶、早期和晚期SV40、来自反转录病毒的长末端重复序列(LTRs)和小鼠金属硫蛋白-1的那些。合适的载体和启动子的选择完全在本领域普通技术水平内。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。表达载体还可以包括用于扩增表达的合适序列。表达载体还可以包括编码与嵌合定位修饰多肽融合的蛋白标签(例如6xHis标签、血凝素(HA)标签、荧光蛋白(例如绿色荧光蛋白;黄色荧光蛋白等)等)的核苷酸序列。

[0397] 在某些实施方案中,编码嵌合定位修饰多肽的核苷酸序列可操作地连接至诱导型启动子(例如热激启动子、四环素调节的启动子、类固醇调节的启动子、金属调节的启动子、雌激素受体调节的启动子等)。在某些实施方案中,编码嵌合定位修饰多肽的核苷酸序列可操作地连接至空间限制和/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)。在某些实施方案中,编码嵌合定位修饰多肽的核苷酸序列可操作地连接至组成型启动子。

[0398] 将核酸引入宿主细胞的方法是本领域已知的,并且可以使用任何已知方法将核酸(例如表达构建体)引入干细胞或祖先细胞。合适的方法包括例如病毒或噬菌体感染、转染、接合、原生质体融合、脂质转染法、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、基因枪技术、磷酸钙沉淀、直接显微注射、纳米颗粒介导的核酸递送(例如Panyam等人Adv Drug Deliv Rev.2012年9月13日.pp:50169-409X(12)00283-9.doi: 10.1016/j.addr.2012.09.023)等。

[0399] *方法*

[0400] 本公开内容提供了用于修饰靶DNA和/或靶DNA-相关多肽的方法。通常,方法包括使靶DNA与复合物(“靶向复合物”)接触,所述复合物包含指导RNA和sRGN多肽。

[0401] 如上文所讨论的,gRNA或sgRNA和sRGN多肽形成复合物。指导RNA通过包含与靶DNA的序列互补的核苷酸序列,为复合物提供靶特异性。复合物的sRGN多肽提供位点特异性活性。在某些实施方案中,复合物修饰靶DNA,从而导致例如DNA切割、DNA甲基化作用、DNA损伤、DNA修复等。在其它实施方案中,复合物修饰与靶DNA相关的靶多肽(例如,组蛋白、DNA-结合蛋白等),从而导致例如组蛋白甲基化作用、组蛋白乙酰化作用、组蛋白泛素化等。靶DNA可以是例如体外裸(即未由DNA相关的蛋白结合)DNA、体外细胞中的染色体DNA、体内细胞中的染色体DNA等。

[0402] 来自各种物种的sRGN蛋白可能在靶DNA中需要不同的PAM序列。因此,对于选择的特定sRGN蛋白,PAM序列要求可能与上述PAM序列不同。

[0403] 提供利用sRGN直系同源物的特征的示例性方法包括下述。

[0404] 核酸酶活性切割靶DNA以产生双链断裂。这些断裂然后由细胞以两种方式之一修复:非同源末端连接和同源性指导的修复。在非同源末端连接(NHEJ)中,通过将断裂末端直接相互连接来修复双链断裂。在所述过程中,可以在切割位点插入或缺失少数碱基对。在同源性指导的修复中,与切割的靶DNA序列具有同源性的供体多核苷酸被用作用于修复切割的靶DNA序列的模板,从而导致遗传信息从供体多核苷酸转移至靶DNA。因此,可以将新核酸材料插入/复制到位点中。在某些情况下,靶DNA与供体多核苷酸接触。在某些情况下,将供体多核苷酸引入细胞中。由于NHEJ和/或同源性指导的修复对靶DNA的修饰导致例如基因矫正、基因置换、基因标记(gene tagging)、转基因插入、核苷酸缺失、核苷酸插入、基因破坏、基因突变、序列置换等。因此,sRGN多肽对DNA的切割可用于通过切割靶DNA序列并允许细胞在没有外源提供的供体多核苷酸的情况下修复所述序列来从靶DNA序列删除核酸材料(例如以破坏使细胞对感染易感的基因(例如CCR5或CXCR4基因,其使T细胞对HIV感染易感,以

去除神经元中引起疾病的三核苷酸重复序列,以产生作为研究中的疾病模型的基因敲除和突变等)。因此,所述方法可用于敲除基因(从而导致完全缺乏转录或改变的转录)或将遗传物质敲入靶DNA中选择的基因座。

[0405] 可替换地,如果将指导RNA和sRGN多肽共施用至具有至少包括与靶DNA序列具有同源性的区段的供体多核苷酸序列的细胞,则可以使用主题方法添加,即插入或置换,核酸材料至靶DNA序列(例如以"敲入"编码蛋白、siRNA、miRNA等的核酸),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),向基因添加调节序列(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)等。因此,包含指导RNA和sRGN多肽的复合物可用于其中期望以位点特异性(即,"靶向的")方式修饰DNA的任何体外或体内应用,例如基因敲除、基因敲入、基因编辑、基因标记、序列置换等,如在例如下述中使用的,基因治疗,例如以治疗疾病或作为抗病毒、抗病原物质或抗癌治疗,农业中遗传修饰的生物的生产,用于治疗、诊断或研究目的的由细胞大规模生产蛋白,诱导iPS细胞,生物学研究,用于缺失或置换的病原体基因的靶向等。

[0406] 在某些实施方案中,sRGN多肽包含sRGN蛋白的修饰形式。在某些情况下,sRGN蛋白的修饰形式包含降低sRGN蛋白的天然存在的核酸酶活性的氨基酸变化(例如,缺失、插入或取代)。例如,在某些情况下,sRGN蛋白的修饰形式具有小于50%、小于40%、小于30%、小于20%、小于10%、小于5%或小于1%的相应野生型sRGN多肽的核酸酶活性。在某些情况下,sRGN多肽的修饰形式没有显著的核酸酶活性。当sRGN多肽是没有显著的核酸酶活性的sRGN蛋白的修饰形式时,它可以被称为"dsRGN"。

[0407] 在某些实施方案中,sRGN多肽包含异源序列(例如融合体)。在某些实施方案中,异源序列可以提供sRGN多肽的亚细胞定位(例如,用于靶向细胞核的核定位信号(NLS);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)。在某些实施方案中,异源序列可以提供用于易于追踪或纯化的标签(例如荧光蛋白,例如绿色荧光蛋白(GFP)、YFP、RFP、CFP、mCherry、tdTomato等;His标签,例如6XHis标签;血凝素(HA)标签;FLAG标签;Myc标签;等)。在某些实施方案中,异源序列可提供增加或降低的稳定性。

[0408] 在某些实施方案中,将修饰的多核苷酸用于本文所述的CRISPR-sRGN系统中,其中可以修饰指导RNA和/或DNA或包含编码sRGN多肽或其变体的多核苷酸序列的RNA,如下文所述和所示。此类修饰的多核苷酸可用于CRISPR-sRGN系统,例如,以编辑任何一个或多个基因组基因座。在某些实施方案中,本公开内容的多核苷酸中的此类修饰是通过密码子优化来实现,例如基于要在其中表达编码的多肽的特定宿主细胞的密码子优化。熟练的技术人员将理解,可以对本公开内容的任何核苷酸序列和/或重组核酸进行密码子优化以在任何目标物种中表达。密码子优化是本领域众所周知的,并且涉及使用物种特异性密码子选择表来修改核苷酸序列的密码子选择偏向。基于对目标物种的最高表达基因的序列分析而生成密码子选择表。在一个非限制性例子中,当核苷酸序列要在宿主细胞中翻译时,基于对目标物种的高表达核基因的序列分析而产生密码子选择表。通过将物种特异性密码子选择表与天然多核苷酸序列中存在的密码子进行对比,确定核苷酸序列的修饰。例如,如果预期的靶细胞是人细胞,则编码sRGN(或sRGN变体,例如,酶无活性变体)的人密码子优化的多核苷酸序列将是合适的。作为另一个非限制性例子,如果预期的宿主细胞是小鼠细胞,则编码

sRGN(或sRGN变体,例如,酶无活性变体)的小鼠密码子优化的多核苷酸序列将是合适的。

[0409] 用于密码子优化的策略和方法是本领域已知的,并且已经针对各种系统进行了描述,这些系统包括、但不限于酵母(Outchkourov等人,Protein Expr Purif, 24(1):18-24(2002))和大肠杆菌(Feng等人,Biochemistry, 39(50):15399-15409 (2000))。在某些实施方案中,通过使用GeneGPS ® 表达优化技术(ATUM)并使用制造商推荐的表达优化算法来进行密码子优化。在某些实施方案中,本公开内容的多核苷酸经密码子优化以增加在人细胞中的表达。在某些实施方案中,本公开内容的多核苷酸经密码子优化以增加在大肠杆菌细胞中的表达。在某些实施方案中,本公开内容的多核苷酸经密码子优化以增加在昆虫细胞中的表达。在某些实施方案中,定义在密码子优化程序中使用的表达优化算法以避免假定的聚-A信号(例如,AATAAA和ATTAAA)以及长(大于4个)的A段,其可能导致聚合酶滑移。还可以选择聚腺苷酸化信号以优化在预期宿主中的表达。还可以选择聚腺苷酸化信号以优化在预期宿主中的表达。

[0410] 如本领域中充分理解的,核苷酸序列的密码子优化产生这样的核苷酸序列:其与天然核苷酸序列具有小于100%同一性(例如,小于70%、71%. 72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%),但其仍编码具有与原始天然核苷酸序列编码的多肽相同的功能的多肽。因此,在本公开内容的代表性实施方案中,可以为在特定目标物种中的表达对本公开内容的核苷酸序列和/或重组核酸进行密码子优化。

[0411] 在某些实施方案中,编码sRGN的密码子优化的多核苷酸序列与SEQ ID NO: 15、18、20、22和27中的任一个具有至少90%、92%、93%、94%、95%、96%、97%、98%、99%、99.2%、99.5%、99.8%、99.9%或100%序列同一性。在某些实施方案中,本公开内容的多核苷酸经密码子优化以增加编码的sRGN多肽在靶细胞中的表达。在某些实施方案中,本公开内容的多核苷酸经密码子优化以增加在人细胞中的表达。在某些实施方案中,本公开内容的多核苷酸经密码子优化以增加在*大肠杆菌*细胞中的表达。在某些实施方案中,本公开内容的多核苷酸经密码子优化以增加在昆虫细胞中的表达。

[0412] 在某些实施方案中,可以对sRGN多肽进行密码子优化。这种类型的优化是本领域已知的,并且需要外源DNA的突变以在编码相同蛋白的同时模拟预期宿主生物或细胞的密码子偏好。因此,密码子被改变,但是编码的蛋白保持不变。例如,如果预期的靶细胞是人细胞,则人密码子优化的sRGN (或变体,例如,酶无活性变体)将是合适的。可以对任何合适的sRGN多肽(例如,任何sRGN,例如SEQ ID NO: 1-12中的任一个所示的序列)进行密码子优化。作为另一个非限制性例子,如果预期的宿主细胞是小鼠细胞,则小鼠密码子优化的sRGN(或变体,例如,酶无活性变体)将是合适的。

[0413] 在一个方面,本文提供了一种核酸,其包含编码sRGN多肽(例如,包含SEQ ID NO: 1-12中的任一个的氨基酸序列的多肽)或其与SEQ ID NO: 1-12中的任一个具有至少约90%(诸如至少约91%、92%、93%、94%、95%、96%、97%、98%、99%或更多中的任一种)序列同一性的变体的密码子优化的多核苷酸序列。这种编码sRGN或其变体的密码子优化的核酸在本文中也称为“密码子优化的sRGN多肽”或“密码子优化的sRGN”。在某些实施方案中,密码子优化的sRGN编码SEQ ID NO: 1-12中的任一个的sRGN多肽。在某些实施方案中,密码子优化的sRGN为在原核细胞如细菌细胞中表达进行密码子优化。在某些实施方案中,密码子优化的sRGN

为在真核细胞如昆虫或哺乳动物细胞中表达进行密码子优化。在某些实施方案中,密码子优化的sRGN为在人细胞中表达进行密码子优化。在某些实施方案中,密码子优化根据GeneGPS ® 优化(ATUM)进行。在某些实施方案中,密码子优化根据GeneOptimizer方法(ThermoFisher Scientific)进行。在某些实施方案中,密码子优化的sRGN与SEQ ID NO: 15、18、20、22和27中的任一个具有至少约90%(诸如至少约91%、92%、93%、94%、95%、96%、97%、98%、99%或更多中的任一种)序列同一性。在某些实施方案中,密码子优化的sRGN包含SEQ ID NO: 15、18、20、22和27中的任一个的序列(或由其组成)。

[0414] 在根据本文所述的包含编码sRGN多肽或其变体的密码子优化的多核苷酸序列的任何核酸的一些实施方案中,所述核酸进一步包括编码gRNA的核苷酸序列。在某些实施方案中,核酸是表达载体,例如重组表达载体。在某些实施方案中,编码sRGN多肽或其变体的核苷酸序列包含SEQ ID NO: 15、18、20、22和27中的任一个的密码子优化的多核苷酸序列或其与SEQ ID NO: 15、18、20、22和27中的任一个的密码子优化的多核苷酸序列具有至少90％序列同一性的变体。

[0415] 在某些实施方案中,本文提供了一种核酸,其包含与SEQ ID NO: 15、18、20、22和27中的任一个具有至少90%、91%、92%、93%、94%、95%、96%、97%、98%或99%序列同一性的多核苷酸序列,其中所述多核苷酸序列编码如本文所述的sRGN多肽或其变体。在某些实施方案中,多核苷酸序列包含SEQ ID NO: 15、18、20、22和27中的任一个的密码子优化的多核苷酸序列(或由其组成)。

[0416] 在某些实施方案中,本文提供了一种核酸,其包含与SEQ ID NO: 15、18、20、22和27中的任一个的子序列具有至少90%、91%、92%、93%、94%、95%、96%、97%、98%或99%序列同一性的多核苷酸序列,其中所述多核苷酸序列编码一个或多个功能性sRGN结构域。在某些实施方案中,所述多核苷酸序列包含SEQ ID NO: 15、18、20、22和27中的任一个的子序列(或由其组成)。在某些实施方案中,所述一个或多个sRGN结构域包括核酸酶结构域(例如,HNH或RuvC结构域)。在某些实施方案中,所述一个或多个sRGN结构域包括RNA-结合结构域(例如,Rec I结构域)。在某些实施方案中,所述一个或多个sRGN结构域包括PAM-相互作用结构域。

[0417] 在根据本文描述的包含密码子优化的多核苷酸序列的任何核酸的一些实施方案中,与在宿主细胞中sRGN多肽或其变体从包含参考多核苷酸序列(从其衍生出密码子优化的多核苷酸序列)的相应核酸的表达相比,在宿主细胞中sRGN多肽或其变体从所述核酸的表达增加。在某些实施方案中,密码子优化的多核苷酸序列包含SEQ ID NO: 15、18、20、22和27中的任一个的密码子优化的多核苷酸序列,或其与SEQ ID NO: 15、18、20、22和27中的任一个的密码子优化的多核苷酸序列具有至少90％序列同一性的变体。在某些实施方案中,参考多核苷酸序列包含来自SEQ ID NO: 13、14、16、17、19、21、23-26、28和29中的任一个的编码sRGN的多核苷酸序列,或其与来自SEQ ID NO: 13、14、16、17、19、21、23-26、28和29中的任一个的编码sRGN的多核苷酸序列具有至少90%序列同一性的变体。在某些实施方案中,与在宿主细胞中sRGN多肽或其变体从参考多核苷酸序列的表达相比,在宿主细胞中sRGN多肽或其变体从密码子优化的多核苷酸序列的的表达增加至少约10％(例如至少约15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、100%或更多的任一种)。

[0418] 在某些实施方案中,指导RNA和sRGN多肽用作用于关闭细菌细胞中基因表达的诱导系统。在某些情况下,将编码合适的指导RNA和/或合适的sRGN多肽的核酸并入靶细胞的染色体中,并在诱导型启动子的控制下。当指导RNA和/或sRGN多肽被诱导时,当存在指导RNA和sRGN多肽两者并形成复合物时,靶DNA在感兴趣的位置(例如,在独立的质粒上的靶基因)被切割(或以其它方式被修饰)。因此,在某些情况下,细菌表达菌株被工程改造以包含细菌基因组中编码适当sRGN多肽的核酸序列和/或质粒上适当的指导RNA(例如在诱导型启动子的控制下),从而允许其中通过诱导指导RNA和sRGN多肽的表达可以控制任何靶向的基因(由引入菌株中的独立质粒表达)的表达的实验。在某些情况下,sRGN多肽具有酶活性,其以除引入双链断裂外的方式修饰靶DNA。可用于修饰靶DNA的感兴趣的酶活性(例如,通过将具有酶活性的异源多肽融合到sRGN多肽,从而产生嵌合sRGN多肽)包括、但不限于甲基转移酶活性、脱甲基酶活性、DNA修复活性、DNA损伤活性、脱氨作用活性、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、光解酶活性或糖基化酶活性)。甲基化作用和脱甲基作用在本领域中被认为是外遗传基因调节的重要模式,而DNA损伤和修复活性对于细胞存活和对于响应环境应激的正确基因组维持是必不可少的。因此,本文中的方法应用于靶DNA的外遗传修饰中,并可通过将所期望的互补核酸序列基因工程改造到指导RNA的DNA-靶向区段中而用于控制在靶DNA中任何位置处的靶DNA的外遗传修饰。本文的方法还应用于在靶DNA内任何所需位置处的DNA的有意和受控的损伤中。本文的方法还应用于在靶DNA内任何所需位置处的DNA的序列特异性和受控修复。将DNA-修饰酶活性靶向靶DNA中特定位置的方法应用于研究和临床应用两者中。

[0419] 在某些情况下,sRGN多肽具有调节靶DNA的转录的活性(例如,在嵌合sRGN多肽的情况下,等等)。在某些情况下,包含表现出增加或减少转录的能力的异源多肽(例如转录激活因子或转录阻遏蛋白多肽)的嵌合sRGN多肽用于增加或减少靶DNA中特定位置处的靶DNA转录,其由指导RNA的DNA-靶向区段指导。用于提供具有转录调节活性的嵌合sRGN多肽的来源多肽的例子包括、但不限于光诱导的转录调节物、小分子/药物应答转录调节物、转录因子、转录阻遏蛋白等。在某些情况下,所述方法用于控制靶向的编码-RNA(蛋白编码基因)和/或靶向的非编码RNA(例如tRNA、rRNA、snoRNA、siRNA、miRNA、长ncRNA等)的表达。在某些情况下,sRGN多肽具有修饰与DNA相关的多肽(例如组蛋白)的酶活性。在某些实施方案中,酶活性是甲基转移酶活性、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(即,泛素化活性)、脱泛素活性、腺苷化(adenylation)活性、脱腺苷化活性、苏素化(SUMOylating)活性、脱苏素化(deSUMOylating)活性、核糖基化活性、脱核糖基化(deribosylation)活性、豆蔻酰化活性、脱豆蔻酰化(demyristoylation)活性糖基化活性(例如来自GlcNAc转移酶)或去糖基化活性。本文列举的酶活性催化对蛋白的共价修饰。这种修饰在本领域中已知改变靶蛋白的稳定性或活性(例如,由于激酶活性的磷酸化可以取决于靶蛋白刺激或沉默蛋白活性)。作为蛋白靶特别令人感兴趣的是组蛋白。组蛋白在本领域中已知结合DNA并形成称为核小体的复合物。可以修饰组蛋白(例如通过甲基化作用、乙酰化作用、泛素化、磷酸化)以引出周围DNA的结构变化,从而控制DNA的潜在大部分对相互作用因子如转录因子、聚合酶等的可接近性。单个组蛋白可以许多不同方式和以许多不同组合进行修饰(例如,组蛋白3的赖氨酸27,H3K27的三甲基化与受阻遏的转录的DNA区

域相关,而组蛋白3的赖氨酸4,H3K4的三甲基化与活跃转录的DNA区域相关)。因此,具有组蛋白修饰活性的定位修饰多肽应用于DNA结构的位点特异性控制,并可用于改变靶DNA的选定区域中的组蛋白修饰模式。这种方法应用于研究和临床应用两者中。

[0420] 在某些实施方案中,同时使用多种指导RNA以同时修饰相同靶DNA或不同靶DNA上的不同位置。在某些实施方案中,两种或更多种指导RNA靶向相同的基因或转录物或基因座。在某些实施方案中,两种或更多种指导RNA靶向不同的不相关基因座。在某些实施方案中,两种或更多种指导RNA靶向不同但相关的基因座。

[0421] 在某些情况下,sRGN多肽直接作为蛋白提供。作为一个非限制性例子,可以使用原生质球转化用外源蛋白和/或核酸转化真菌(例如酵母)(参见Kawai等人, Bioeng Bugs.2010年11月-12月;1(6):395-403;“Transformation of Saccharomyces cerevisiae and other fungi: methods and possible underlying mechanism”;和Tanka等人, Nature.2004 Mar 18;428(6980):323-8:“Conformational variations in an infectious protein determine prion strain differences”;两者均整体通过引用并入本文)。因此,可以将sRGN多肽并入原生质球(具有或不具有编码指导RNA的核酸以及具有或不具有供体多核苷酸),并且所述原生质球可用于将内容物引入酵母细胞。可以通过任何方便的方法将sRGN多肽引入细胞(提供给细胞);这种方法是本领域普通技术人员已知的。作为另一个非限制性例子,可以将sRGN多肽直接注射入细胞(例如,具有或不具有编码指导RNA的核酸以及具有或不具有供体多核苷酸),例如斑马鱼胚胎的细胞、受精的小鼠卵母细胞的前核等。

[0422] *感兴趣的靶细胞*

[0423] 在一些上述应用中,所述方法可用于在体内和/或离体和/或体外在有丝分裂或有丝分裂后细胞中诱导DNA切割、DNA修饰和/或转录调节(例如以产生可以再引入个体的遗传修饰的细胞)。因为指导RNA通过与靶DNA杂交提供特异性,所以在公开的方法中,感兴趣的有丝分裂和/或有丝分裂后细胞可能包括来自任何生物的细胞(例如细菌细胞,古细菌细胞,单细胞真核生物的细胞,植物细胞,藻类细胞,例如丛粒藻(Botryococcus braunii)、雷氏衣藻(Chlamydomonas reinhardtii)、微拟球藻(Nannochloropsis gaditana)、蛋白核小球藻(Chlorella pyrenoidosa)、展枝马尾藻(Sargassum patens C.Agardh)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、两栖动物、爬行动物、鸟、哺乳动物)的细胞,来自哺乳动物的细胞,来自啮齿类动物的细胞,来自灵长类动物的细胞,来自人的细胞,等)。

[0424] 任何类型的细胞都可能是有意义的(例如干细胞,例如胚胎干(ES)细胞、诱导的多潜能干(iPS)细胞、生殖细胞;体细胞,例如成纤维细胞、造血细胞、神经元、肌细胞、骨细胞、肝细胞、胰细胞;处于任何阶段的胚胎的体外或体内胚胎细胞,例如1-细胞、2-细胞、4-细胞、8-细胞等阶段斑马鱼胚胎;等)。细胞可以来自确立的细胞系,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,以指已经衍生自生物并被允许在体外生长有限数目的培养物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经传代0次、1次、2次、4次、5次、10次或15次但未传代通过转捩阶段的足够次数的培养物。通常,本发明的原代细胞系在体外维持少于10次传代。靶细胞在许多实施方案中是单细胞生物,或在培养物中生长。

[0425] 如果细胞是原代细胞,则可以通过任何方便的方法从个体中收获它们。例如,白细胞可通过单采血液成分术、白细胞除去法(leukocytapheresis)、密度梯度分离等方便地收获,而来自诸如皮肤、肌肉、骨髓、脾、肝、胰、肺、肠、胃等的组织的细胞最方便通过活组织检查收获。可以使用适当的溶液分散或悬浮收获的细胞。这种溶液通常将是平衡盐溶液,例如生理盐水、磷酸缓冲盐水(PBS)、Hank氏平衡盐溶液等,方便地补加了胎牛血清或其它天然存在的因子,连带着可接受的低浓度缓冲液,通常为5-25 mM。方便的缓冲液包括HEPES、磷酸盐缓冲液、乳酸盐缓冲液等。细胞可以立即使用,或者它们可以冷冻贮藏长的时间段,被解冻并能够再使用。在这种情况下,通常将细胞在10%DMSO,50%血清,40%缓冲的培养基或如本领域中所通常使用的一些其它这种溶液中冷冻,以在这种冷冻温度保存细胞,并以如本领域中所通常已知的方式解冻,以用于解冻冷冻的培养的细胞。

[0426] *编码指导RNA和/或sRGN多肽的核酸*

[0427] 在某些实施方案中,方法包括接触靶DNA或将一种或多种包含编码指导RNA和/或sRGN多肽的核苷酸序列的核酸和/或供体多核苷酸引入细胞(或细胞群体)。包含编码指导RNA和/或sRGN多肽的核苷酸序列的合适核酸包括表达载体,其中包含编码指导RNA和/或sRGN多肽的核苷酸序列的表达载体是“重组表达载体”。

[0428] 在某些实施方案中,重组表达载体是病毒构建体,例如重组腺相关病毒构建体(例如美国专利号7,078,387)、重组腺病毒构建体、重组慢病毒构建体等。

[0429] 合适的表达载体包括、但不限于,病毒载体(例如基于痘苗病毒的病毒载体;脊髓灰质炎病毒;腺病毒(例如Li等人, Invest Opthalmol Vis Sci 35:2543 2549, 1994; Borras等人, Gene Ther 6:515 524, 1999;Li和Davidson, PNAS 92:7700 7704, 1995; Sakamoto等人, H Gene Ther 5:1088 1097, 1999;WO 94112649、WO 93/03769;WO 93/19191;WO 94/28938;WO 95/11984和WO 95/00655);腺相关病毒(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在WO 93/09239中,Samulski等人, J.Vir.(1989) 63:3822-3828; Mendelson等人, Viral.(1988) 166:154-165;和Flotte等人, PNAS (1993) 90:10613-10617);SV40;单纯疱疹病毒;人免疫缺陷病毒(例如Miyoshi等人, PNAS 94:10319 23, 1997;Takahashi等人, J Viral73:7812 7816, 1999);反转录病毒载体(例如鼠白血病病毒,脾脏坏死病毒,以及衍生自反转录病毒的载体,例如劳斯肉瘤病毒,Harvey肉瘤病毒,禽类白血病病毒,慢病毒,人免疫缺陷病毒,骨髓增生肉瘤病毒和乳腺肿瘤病毒);等等。

[0430] 许多合适的表达载体是本领域技术人员已知的,并且许多是可商购获得的。作为实例提供下述载体;对于真核宿主细胞:pXT1、pSG5(Stratagene)、pSVK3、pBPV、pMSG和pSVLSV40(Pharmacia)。但是,可以使用任何其它载体,只要它与宿主细胞相容即可。

[0431] 在某些实施方案中,编码指导RNA和/或sRGN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,例如启动子。转录控制元件可以在真核细胞,例如哺乳动物细胞,或原核细胞(例如细菌或古细菌细胞)中起作用。在某些实施方案中,编码指导RNA和/或sRGN多肽的核苷酸序列可操作地连接至多个控制元件,其允许在原核和真核细胞两者中表达编码指导RNA和/或sRGN多肽的核苷酸序列。

[0432] 取决于所使用的宿主/载体系统,可以在表达载体中使用许多合适的转录和翻译控制元件中的任何一种,包括组成型和诱导型启动子、转录增强子元件、转录终止子等(例如U6启动子、H1启动子等;参见上文)(参见,例如Bitter等人(1987) Methods in Enzymology, 153:516-544)。

[0433] 在某些实施方案中,指导RNA和/或sRGN多肽可以作为RNA提供。在这种情况下,指导RNA和/或编码sRGN多肽的RNA可以通过直接化学合成产生,或者可以从编码指导RNA的DNA体外转录。从DNA模板合成RNA的方法是本领域众所周知的。在某些情况下,指导RNA和/或编码sRGN多肽的RNA将使用RNA聚合酶(例如T7聚合酶、T3聚合酶、SP6聚合酶等)体外合成。一旦合成,RNA就可以直接接触靶DNA,或者可以通过用于将核酸引入细胞的任何众所周知的技术(例如显微注射、电穿孔、转染等)引入细胞。

[0434] 可以使用十分发达的转染技术将编码指导RNA(作为DNA或RNA引入)和/或sRGN多肽(作为DNA或RNA引入)的核苷酸和/或供体多核苷酸提供给细胞;例如Angel和Yanik(2010) PLoS ONE 5(7): e 11756,和来自Qiagen的可商购获得的TransMessenger®试剂,来自Stemgent的Stemfect™RNA转染试剂盒,和来自Mims Bio的TransiT®-mRNA转染试剂盒。也参见Beumer等人(2008) Efficient gene targeting in Drosophila by direct embryo injection with zinc-finger nucleases.PNAS 105(50):19821-19826。另一方面,可以在DNA载体上提供编码指导RNA和/或sRGN多肽和/或嵌合sRGN多肽的核酸和/或供体多核苷酸。用于将核酸转移到靶细胞中的许多载体是可获得的,例如质粒、粘粒、小环、噬菌体、病毒等。包含一种或多种核酸的载体可以附加型地维持,例如作为质粒,小环DNA,病毒,例如巨细胞病毒,腺病毒等,或者它们可以通过同源重组或随机整合整合到靶细胞基因组中,例如反转录病毒衍生的载体,例如MMLV、HIV-1、ALV等。

[0435] 载体可以直接提供给细胞。换句话说,使细胞与包含编码指导RNA和/或sRGN多肽和/或嵌合sRGN多肽的核酸和/或供体多核苷酸的载体接触,以使载体被细胞摄取。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、显微注射和脂质转染法是本领域众所周知的。对于病毒载体递送,使细胞与包含编码指导RNA和/或sRGN多肽和/或嵌合sRGN多肽的核酸和/或供体多核苷酸的病毒颗粒接触。反转录病毒,例如慢病毒,特别适合于本发明的方法。通常使用的反转录病毒载体是“缺陷的”,即,不能产生生产性感染所需的病毒蛋白。相反,载体的复制需要在包装细胞系中生长。为了产生包含感兴趣的核酸的病毒颗粒,通过包装细胞系将包含所述核酸的反转录病毒核酸包装到病毒壳体中。不同的包装细胞系提供了不同的要并入壳体中的包膜蛋白(亲嗜性、双嗜性或异嗜性),这种包膜蛋白决定了病毒颗粒对细胞的特异性(对于鼠和大鼠是亲嗜性的;对于大多数哺乳动物细胞类型包括人、狗和小鼠是双嗜性的;且对于除鼠细胞外的大多数哺乳动物细胞类型是异嗜性的)。适当的包装细胞系可用于确保细胞被包装的病毒颗粒靶向。将包含编码重编程序因子的核酸的反转录病毒载体引入包装细胞系并收集由包装系产生的病毒颗粒的方法是本领域众所周知的。核酸也可以通过直接显微注射引入(例如将RNA注射到斑马鱼胚胎中)。

[0436] 用于向细胞提供编码指导RNA和/或sRGN多肽和/或嵌合sRGN多肽的核酸和/或供体多核苷酸的载体通常将包含合适的启动子,以用于驱动感兴趣的核酸的表达,即,转录激活。换句话说,感兴趣的核酸将可操作地连接至启动子。这可以包括遍在起作用的启动子,例如,CMV-13-肌动蛋白启动子,或诱导型启动子,例如在特定细胞群体中有活性或对诸如

四环素的药物的存在起反应的启动子。通过转录激活,意图是转录将在靶细胞中增加至基础水平的至少10倍、至少100倍、更通常至少1000倍。此外,用于向细胞提供指导RNA和/或sRGN多肽和/或嵌合sRGN多肽和/或供体多核苷酸的载体可以包括编码靶细胞中选择标记的核酸序列,以便鉴定已摄取指导RNA和/或sRGN多肽和/或嵌合sRGN多肽和/或供体多核苷酸的细胞。

[0437] 指导RNA和/或sRGN多肽和/或嵌合sRGN多肽可代之以用于接触DNA或作为RNA引入细胞中。将RNA引入细胞的方法是本领域已知的,并且可以包括,例如,直接注射、转染或用于引入DNA的任何其它方法。

[0438] sRGN多肽可以作为多肽提供给细胞。这种多肽可以任选地与增加产物溶解度的多肽结构域融合。所述结构域可以通过确定的蛋白酶切割位点连接至多肽,例如TEV序列,其可被TEV蛋白酶切割。接头还可以包括一种或多种柔性序列,例如,1至10个甘氨酸残基。在某些实施方案中,融合蛋白的切割在维持产物的溶解度的缓冲液中进行,例如在有0.5至2M尿素的情况下、在有增加溶解度的多肽和/或多核苷酸的情况下等。感兴趣的结构域包括内体裂解(endosomolytic)结构域,例如,流感HA域;和其它帮助生产的多肽,例如IF2结构域、GST结构域、GRPE结构域等。可以配制多肽用于提高的稳定性。例如,肽可以被聚乙二醇化(PEGylated),其中聚氧乙烯(polyethyleneoxy)基团提供了血流中增加的生存期。

[0439] 额外地或可替换地,sRGN多肽可以与多肽透膜体(permeant)结构域融合以促进由细胞摄取。许多透膜体结构域是本领域已知的,并且可以用于本发明的非整合多肽,包括肽、肽模拟物(peptidomimetics)和非肽载体。例如,透膜体肽可以衍生自黑尾果蝇(Drosophila melanogaster)转录因子触角足(Antennapaedia)的第三α螺旋,称为penetratin,其包含也在J. Biol. Chem., 271 (1996), 第18188-18193页中公开的氨基酸序列RQIKIWFQNRRMKWKK (SEQ ID NO: 127)基序(用于鉴定但不是为了在本申请中公开起见)。作为另一个实例,透膜体肽包含HIV-1 tat碱性区氨基酸序列,其可以包括例如天然存在的tat蛋白的氨基酸49-57。

[0440] 其它透膜体结构域包括聚精氨酸基序,例如HIV-1 rev蛋白的氨基酸34-56的区域、九精氨酸、acta-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;和Wender等人(2000) Proc.Natl.Acad.Sci.U.S.A.2000年11月21日;97(24):13003-8;公开的美国专利申请公开号20030220334;20030083256;20030032593;和20030022831,关于易位肽和拟肽的教导在本文特别引入作为参考。九精氨酸(R9)序列是已被表征的更有效的PTDs之一(Wender等人,2000;Uemura等人,2002)。可以选择进行融合的位点,以最优化多肽的生物学活性、分泌或结合特征。最佳位点将通过常规实验确定。在某些实施方案中,多肽透膜体结构域被化学修饰以增加PTD的生物利用率。示例性修饰公开于Expert Opin Drug Deliv.2009年11月;6(11):1195-205。

[0441] sRGN多肽可以在体外或通过真核细胞、通过原核细胞或通过体外转录和翻译(IVTT)来产生,并且其可以通过解折叠例如热变性、DTT还原等进一步加工,并且可以使用本领域已知的方法进一步重折叠。

[0442] 不改变一级序列的感兴趣的修饰包括多肽的化学衍生,例如酰化作用、乙酰化作用、羧化作用、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

进一步的加工步骤中修饰多肽的糖基化模式而产生的那些;例如通过使多肽暴露于影响糖基化的酶,例如哺乳动物的糖基化或去糖基化酶。还包括具有磷酸化的氨基酸残基的序列,例如,磷酸酪氨酸,磷酸丝氨酸或磷酸苏氨酸。

[0443] 本发明还包括已使用普通分子生物学技术和合成化学修饰的指导RNA和sRGN多肽,以便提高其对蛋白水解降解的抗性、改变靶序列特异性、最优化溶解度性质、改变蛋白活性(例如转录调节活性、酶活性等)或使其更适合作为治疗剂。这种多肽的类似物包括含有除天然存在的L-氨基酸以外的残基的那些,例如O-氨基酸或非天然存在的合成氨基酸。D-氨基酸可以取代一些或所有氨基酸残基。sRGN多肽可以使用本领域已知的常规方法通过体外合成来制备。各种商业合成装置是可获得的,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以被非天然氨基酸置换。特定的序列和制备方式将由便利、经济、所需的纯度等确定。

[0444] 如果期望,可以在合成过程中或表达过程中将各种基团引入肽中,其允许与其它分子或表面连接。因此,半胱氨酸可用于制备硫醚,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。

[0445] sRGN多肽也可以根据重组合成的常规方法分离和纯化。可以制备表达宿主的裂解物,并使用HPLC、排阻层析、凝胶电泳、亲和层析或其它纯化技术纯化裂解物。通常,相对于与产物的制备及其纯化方法相关的污染物,所使用的组合物将包含至少20重量%的所期望的产物,例如至少约75重量%或至少约95重量%,并且对于治疗目的,至少99.5重量%。通常,百分比将基于总蛋白。为了诱导DNA切割和重组,或对靶DNA的任何期望的修饰,或对与靶DNA相关的多肽的任何期望的修饰,指导RNA和/或sRGN多肽和/或供体多核苷酸,不管它们作为核酸还是作为多肽引入,被提供给细胞达约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小时或从约30分钟到约24小时的任何其它时间段,其可以以约每天至约每4天的频率重复,例如每1.5天、每2天、每3天或从约每天至约每四天的任何其它频率。可将一种或多种试剂一次或多次地提供给细胞,例如一次、两次、三次或超过三次,并且在每次接触事件之后,允许细胞与一种或多种试剂温育一些量的时间,如16-24小时,所述时间之后将培养基置换为新鲜培养基,并进一步培养细胞。在向细胞提供两种或更多种不同的靶向复合物(例如,与相同或不同靶DNA内的不同序列互补的两种不同的指导RNA)的情况下,复合物可以同时提供(例如,作为两种多肽和/或核酸),或同时递送。另一方面,它们可以被连续地提供,例如首先提供靶向复合物,后面是第二靶向复合物等,或反之亦然。

[0446] 通常,将有效量的指导RNA和/或sRGN多肽和/或供体多核苷酸提供给靶DNA或细胞以诱导靶修饰。指导RNA和/或sRGN多肽和/或供体多核苷酸的有效量是相对于阴性对照(例如与空载体或不相干多肽接触的细胞),诱导在两个同源序列之间观察到的靶修饰量的2-倍增加或更多的量。也就是说,有效量或剂量的指导RNA和/或sRGN多肽和/或供体多核苷酸将诱导在靶DNA区域观察到的靶修饰量的2-倍增加、3-倍增加、4-倍增加或更多,在一些情况下观察到的重组量的5-倍增加、6-倍增加或更多,有时7-倍或8-倍增加或更多,例如观察到的重组量的10-倍、50-倍或100-倍增加或更多,在一些情况下200-倍、500-倍、700-倍或1000-倍增加或更多,例如5000-倍或10,000-倍增加。靶修饰的量可以通过任何方便的方法来测量。例如,可以将沉默报道基因构建体共转染到细胞中,所述沉默报道基因构建体包含

与侧接重复序列的指导RNA的靶向区段(靶向序列)互补的序列,其在重组时将重建编码活性报告基因的核酸,且在与指导RNA和/或sRGN多肽和/或供体多核苷酸接触后评估报道蛋白的量,例如与指导RNA和/或sRGN多肽和/或供体多核苷酸接触后2小时、4小时、8小时、12小时、24小时、36小时、48小时、72小时或更多。作为另一种更灵敏性的测定,例如,在与指导RNA和/或sRGN多肽和/或供体多核苷酸接触后,可通过所述区域的PCR或DNA杂交来评估在包含靶DNA序列的感兴趣的基因组DNA区域的重组程度,例如与指导RNA和/或sRGN多肽和/或供体多核苷酸接触后2小时、4小时、8小时、12小时、24小时、36小时、48小时、72小时或更多。

[0447] 使细胞与指导RNA和/或sRGN多肽和/或供体多核苷酸接触可以在任何培养基中和在任何促进细胞存活的培养条件下发生。例如,可以将细胞悬浮在任何合适的方便的营养培养基中,例如Iscove氏改良DMEM或RPMI1640,补加了胎牛血清或热灭活的山羊血清(约5-10%),L-谷氨酰胺,硫醇,特别是2-巯基乙醇,和抗生素,例如青霉素和链霉素。培养物可以含有细胞对其响应性的生长因子。如本文所定义的,生长因子是能够通过对跨膜受体的特异性作用而在培养物中或在完整组织中促进细胞存活、生长和/或分化的分子。生长因子包括多肽和非多肽因子。促进细胞存活的条件通常容许非同源末端连接和同源性指导的修复。在期望将多核苷酸序列插入靶DNA序列的应用中,还向细胞提供了包含要插入的供体序列的多核苷酸。“供体序列”或“供体多核苷酸”是指在由sRGN多肽诱导的切割位点处插入的核酸序列。供体多核苷酸将含有与切割位点处的基因组序列足够的同源性,例如与侧接切割位点的核苷酸序列70%、80%、85%、90%、95%或100%的同源性,例如在切割位点约50个碱基或更少之内,例如在约30个碱基之内、在约15个碱基之内、在约10个碱基之内、在约5个碱基之内或紧接地侧接切割位点,以支持其与与其具有同源性的基因组序列之间的同源性指导的修复。供体和基因组序列之间具有序列同源性的约25、50、100或200个核苷酸或超过200个核苷酸(或10至200个核苷酸之间的任何整数值,或更多)将支持同源性指导的修复。供体序列可以具有任何长度,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。

[0448] 供体序列通常与其置换的基因组序列不同。相反,供体序列可以含有相对于基因组序列的至少一个或多个单碱基改变、插入、缺失、倒位或重排,只要存在足够的同源性以支持同源性指导的修复即可。在某些实施方案中,供体序列包含侧接两个同源性区域的非同源序列,从而使得靶DNA区域和两个侧翼序列之间的同源性指导的修复导致非同源序列插入靶区域。供体序列也可以包含载体主链,该载体主链含有与感兴趣的DNA区域不同源并且不打算插入感兴趣的DNA区域的序列。通常,供体序列的一个或多个同源区域将与期望与其进行重组的基因组序列具有至少50%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%序列同一性。取决于供体多核苷酸的长度,可以存在1%至100%序列同一性之间的任何值。与基因组序列相比,供体序列可包含某些序列差异,例如限制位点、核苷酸多态性、选择标记(例如抗药性基因、荧光蛋白、酶等)等,其可用于评估供体序列在切割位点的成功插入,或在某些情况下可用于其它目的(例如,表示在靶向的基因组基因座的表达)。在某些情况下,如果位于编码区中,则这种核苷酸序列差异将不改变氨基酸序列,或将产生沉默氨基酸改变(即,不影响蛋白结构或功能的改变)。另一方面,

这些序列差异可以包括侧翼重组序列,例如FLPs、loxP序列等,其可以在较后的时间激活以去除标记序列。

[0449] 供体序列可以作为单链DNA、单链RNA、双链DNA或双链RNA提供给细胞。可以将其以线性或环状形式引入细胞中。如果以线性形式引入,则可以通过本领域技术人员已知的方法保护供体序列的末端(例如,免于核酸外切降解)。例如,将一个或多个双脱氧核苷酸残基添加至线性分子的3'末端和/或将自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.USA 84:4959-4963;Nehls等人(1996) Science 272:886-889。保护外源多核苷酸免于降解的额外方法包括、但不限于,添加一个或多个末端氨基和使用修饰的核苷酸间键,如例如硫代磷酸酯,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末端的备选方案,可以在具有同源性的区域之外包括可以在不影响重组的情况下降解的额外长度的序列。可以将供体序列作为载体分子的部分引入细胞中,所述载体分子具有额外的序列,如例如复制起点、启动子和编码抗生素抗性的基因。此外,供体序列可以作为裸(即未修饰的)核酸,作为与诸如脂质体或泊洛沙姆的试剂复合的核酸引入,或者可以通过病毒(例如腺病毒、AAV)递送,如上文对于编码指导RNA和/或sRGN多肽的核酸和/或供体多核苷酸所述的。

[0450] 按照上述方法,感兴趣的DNA区域可以离体切割和修饰,即,"遗传修饰的"。在某些实施方案中,当已将选择标记物插入感兴趣的DNA区域中时,可通过将遗传修饰的细胞与剩余群体分开来对于包含遗传修饰的那些富集细胞群体。在富集之前,"遗传修饰的"细胞可能仅占细胞群体的约1%或更多(例如2%或更多、3%或更多、4%或更多、5%或更多、6%或更多、7%或更多、8%或更多、9%或更多、10%或更多、15%或更多或20%或更多)。可以通过适合于所用的选择标记的任何方便的分离技术来实现"遗传修饰的"细胞的分离。例如,如果已经插入了荧光标记,则可以通过荧光激活的细胞分选术来分离细胞,而如果已经插入了细胞表面标记,则可以通过亲和分离技术将细胞与异质群体分离,例如,磁性分离、亲和层析、用附着于固体基质的亲和试剂"淘选"或其它方便的技术。提供精确分离的技术包括荧光激活细胞分选仪,其可以具有不同程度的复杂化。例如多颜色通道、低角度和钝角光散射检测通道、阻抗通道等。通过使用与死细胞相关的染料(例如碘化丙锭),可以相对于死细胞选择细胞。可以采用对遗传修饰的细胞的生存力不过度有害的任何技术。以这种方式获得了对于包含修饰的DNA的细胞高度富集的细胞组合物。"高度富集的"是指遗传修饰的细胞将为细胞组合物的70%或更多、75%或更多、80%或更多、85%或更多、90%或更多,例如,细胞组合物的约95%或更多或98%或更多。换句话说,所述组合物可以是遗传修饰的细胞的基本上纯的组合物。

[0451] 通过本文描述的方法产生的遗传修饰的细胞可以立即使用。另一方面,可将细胞于液氮温度冷冻并贮藏长的时间段,被解冻并能够再使用。在这种情况下,通常将细胞在10%二甲基亚砜(DMSO),50%血清,40%缓冲的培养基或如本领域中所通常使用的一些其它这种溶液中冷冻,以在这种冷冻温度保存细胞,并以如本领域中所通常已知的方式解冻,以用于解冻冷冻的培养的细胞。

[0452] 使用本领域已知的方法可以在体外培养遗传修饰的细胞。可以在培养中扩增细胞,即,在促进其增殖的条件下生长。培养基可以是液体或半固体,例如含有琼脂、甲基纤维素等。可以将细胞群体悬浮在合适的营养培养基中,例如lscove氏改良DMEM或RPMI 1640,

正常补加了胎牛血清(约5-10％),L-谷氨酰胺,硫醇,特别是2-巯基乙醇,和抗生素,例如青霉素和链霉素。培养物可以含有各种细胞对其响应性的生长因子。如本文所定义的,生长因子是能够通过对跨膜受体的特异性作用而在培养物中或在完整组织中促进细胞存活、生长和/或分化的分子。生长因子包括多肽和非多肽因子。可以将已经以这种方式遗传修饰的的细胞移植到受试者中,以用于诸如基因治疗的目的,例如以治疗疾病或作为抗病毒、抗病原物质或抗癌治疗,用于农业中遗传修饰的生物的生产,或用于生物学研究。在某些情况下,在无Xeno培养基中培养用于人类的细胞。受试者可以是新生儿、幼年或成体。特别感兴趣的是哺乳动物受试者。可以用本方法治疗的哺乳动物物种包括犬科动物和猫科动物;马科动物;牛科动物;羊;等和灵长类动物,特别是人。动物模型,特别是小哺乳动物(例如小鼠、大鼠、豚鼠、仓鼠、兔形目(lagomorpha)(例如兔)等)可用于实验研究。

[0453] 可以将细胞单独地或与合适的基底或基质一起提供给受试者,例如以支持它们在它们要移植到的组织中的生长和/或组织。通常,将施用至少1x10^3的细胞,例如$5x10^3$的细胞、$1x10^4$的细胞、$5x10^4$的细胞、$1x10^5$的细胞、1×10^6的细胞或更多。可以通过下述途径的任一种将细胞引入受试者:胃肠外、皮下、静脉内、颅内、脊柱内、眼内或脊髓液中。可以通过注射、导管等引入细胞。用于局部递送(即,递送至损伤部位)的方法的例子包括,例如通过Ommaya贮液囊,例如用于鞘内递送(例如美国专利号5,222,982和5,385,582,通过引用并入本文);通过快速浓注,例如通过注射器,例如进入关节中;通过持续输注,例如通过套管插入术,例如具有对流(例如,美国申请No.20070254842,通过引用并入本文);或通过植入在其上已可逆附着细胞的设备(例如美国申请号20080081064和20090196903,通过引用并入本文)。为了产生转基因动物(例如转基因小鼠)起见,也可以将细胞引入胚胎(例如胚泡)中。

[0454] *sRGN系统的递送*

[0455] 指导RNA多核苷酸(RNA或DNA)和/或sRGN多肽(RNA或DNA)可以通过本领域已知的病毒或非病毒递送运载体来递送。

[0456] 多核苷酸可以通过非病毒递送运载体递送,包括、但不限于,纳米颗粒、脂质体、核糖核蛋白、带阳电荷的肽、小分子RNA-缀合物、适体-RNA嵌合体和RNA-融合蛋白复合物。Peer和Lieberman, Gene Therapy, 18: 1127-1133 (2011)中描述了一些示例性非病毒递送运载体(其集中于用于siRNA的非病毒递送运载体,其也可用于递送其它多核苷酸)。

[0457] 重组腺相关病毒(AAV)载体可用于递送。产生rAAV颗粒的技术在本领域中是标准的,其中将包括待递送的多核苷酸的待包装的AAV基因组,rep和cap基因以及辅助病毒功能提供给细胞。生产rAAV需要在单个细胞(在本文中表示为包装细胞)内存在下述组分:rAAV基因组,与rAAV基因组分开(即不在rAAV基因组中)的AAV rep和cap基因,以及辅助病毒功能。AAV rep和cap基因可能来自可衍生重组病毒的任何AAV血清型,并且可能来自与rAAV基因组ITRs不同的AAV血清型,包括、但不限于,AAV血清型AAV-1、AAV-2、AAV-3、AAV-4、AAV-5、AAV-6、AAV-7、AAV-8、AAV-9、AAV-10、AAV-11、AAV-12、AAV-13和AAV rh.74。假型rAAV的产生公开于例如WO 01/83692。

[0458] 表3

[0459]

AAV 血清型	Genbank 登记号
AAV-1	NC_002077.1
AAV-2	NC_001401.2
AAV-3	NC_001729.1
AAV-38	AF028705.1
AAV-4	NC_001829.1
AAV-5	NC_006152.1
AAV-6	AF028704.1
AAV-7	NC_006260.1
AAV-8	NC_006261.1
AAV-9	AX753250.1
AAV-10	AY631965.1
AAV-11	AY631966.1
AAV-12	00813647.1
AAV-13	EU285562.1

[0460] 产生包装细胞的方法是产生稳定表达用于AAV颗粒生产的所有必需组分的细胞系。例如,将包含缺乏AAV rep和cap基因的rAAV基因组,与rAAV基因组分开的AAV rep和cap基因以及选择标记(例如新霉素抗性基因)的质粒(或多个质粒)整合到细胞的基因组中。已通过诸如GC加尾(Samulski等人, 1982, Proc.Natl.Acad.S6.USA, 79:2077-2081)、添加含有限制性内切核酸酶切割位点的合成接头(Laughlin等人, 1983, Gene, 23:65-73)或通过直接平端连接(Senapathy & Carter, 1984, J. Biol. Chem., 259:4661-4666)的程序将AAV基因组引入细菌质粒中。然后用辅助病毒如腺病毒感染包装细胞系。这个方法的优点是细胞是可选择的并且适合于大规模生产rAAV。合适方法的其它例子采用腺病毒或杆状病毒而不是质粒,以将rAAV基因组和/或rep和cap基因引入包装细胞。

[0461] rAAV生产的一般原理在例如Carter, 1992, Current Opinions in Biotechnology, 1533-539;和Muzyczka, 1992, Curr. Topics in Micro Biol. and Immunol., 158:97-129中综述。各种方法描述于Ratschin等人, Mol. Cell. Bio. 4:2072 (1984); Hermonat等人, Proc. Natl. Acad. Sci. USA, 81:6466 (1984); Tratschin等人, Mol. Cell. Biol. 5:3251 (1985); Mclaughlin等人, J. Virol., 62:1963 (1988);和Lebkowski等人, 1988 Mol. Cell. Biol., 7:349 (1988). Samulski等人 (1989, J. Virol., 63:3822-3828);美国专利号5,173,414;WO 95/13365和对应的美国专利号5,658.776;WO 95/13392;WO 96/17947;PCT/US98/18600;WO 97/09441 (PCT/US96/14423);WO 97/08298 (PCT/US96/13872);WO 97/21825 (PCT/US96/20777);WO 97/06243 (PCT/FR96/01064);WO 99/11764;Perrin等人(1995) Vaccine 13:1244-1250; Paul等人 (1993) Human Gene Therapy 4:609-615; Clark等人(1996) Gene Therapy 3:1124-1132;美国专利号5,786,211;美国专利号5,871,982;和美国专利号6,258,595。

[0462] AAV载体血清型可以与靶细胞类型匹配。例如,由除了别的以外的指示的AAV血清型转导的下述示例性细胞类型。

[0463] 表4

[0464]

组织/细胞类型	血清型
肝	AAV8, AAV9
骨骼肌	AAV1, AAV7, AAV6, AAV8, AAV9
中枢神经系统	AAV5, AAV1, AAV4
RPE	AAV5, AAV4
感光细胞	AAV5
肺	AAV9
心脏	AAV8
胰腺	AAV8
肾	AAV2

[0465] 对受试者施用治疗的数目可能有变化。将遗传修饰的细胞引入受试者可能是一次事件;但在某些情况下,这种治疗可能引出改善达有限的时间段,并且需要正在进行的一系列的重复治疗。在其它情况下,在观察到效果之前,可能需要多次施用遗传修饰的细胞。确切的规程取决于疾病或状况、疾病的病期以及正被治疗的个别受试者的参数。

[0466] 在本公开内容的其它方面,采用指导RNA和/或sRGN多肽和/或供体多核苷酸体内修饰细胞DNA,例如用于诸如基因治疗的目的,例如以治疗疾病或作为抗病毒、抗病原物质或抗癌治疗,用于农业中遗传修饰的生物的生产,或用于生物学研究。在这些体内实施方案中,将指导RNA和/或sRGN多肽和/或供体多核苷酸直接施用于个体。可以通过本领域中已知的许多用于向受试者施用肽、小分子和核酸的方法中的任一种来施用指导RNA和/或sRGN多肽和/或供体多核苷酸。可以将指导RNA和/或sRGN多肽和/或供体多核苷酸并入多种制剂中。更特别地,本发明的指导RNA和/或sRGN多肽和/或供体多核苷酸可以通过与合适的药学上可接受的载体或稀释剂组合而配制成药物组合物。在某些情况下,将指导RNA和sRGN配制成用于离体递送。

[0467] 药物制剂是包含存在于药学上可接受的运载体中的一种或多种指导RNA和/或sRGN多肽和/或供体多核苷酸的组合物。“药学上可接受的运载体”可以是由联邦或州政府的管理机构批准或在美国药典(US Pharmacopeia)或其它公认的供哺乳动物例如人之用的药典中列举的运载体。术语“运载体”是指稀释剂、佐剂、赋形剂或载体,本发明的化合物与其一起配制用于向哺乳动物施用。这种药物运载体可以是脂质,例如脂质体,例如脂质体树状聚体;液体,例如水和油,包括石油、动物、植物或合成起源的那些,例如花生油、大豆油、矿物油、芝麻油等,盐水;阿拉伯树胶,明胶,淀粉糊,滑石,角蛋白,胶态二氧化硅,尿素等。此外,可以使用辅助剂、稳定剂、增稠剂、润滑剂和着色剂。可以将药物组合物配制成固体、半固体、液体或气体形式的制剂,例如片剂、胶囊、粉末、颗粒、软膏、溶液、栓剂、注射液、吸入剂、凝胶、小球体和气溶胶。因此,可以以多种方式实现指导RNA和/或sRGN多肽和/或供体多核苷酸的施用,包括口服、口腔含化(buccal)、直肠、胃肠外、腹膜内、皮内、经皮、气管内、眼内等施用。活性剂在施用后可以是全身性的,或者可以通过使用区域施用、壁内施用或使用其作用以将活性剂量保持在植入部位的植入物来局部化。可以将活性剂进行配制用于立

即活性,或者可以将其进行配制用于持续释放。

[0468] 对于一些状况,特别是中枢神经系统状况,可能必须配制试剂以穿越血脑屏障(BBB)。用于通过BBB进行药物递送的一种策略需要通过渗透手段(例如甘露糖醇或白三烯)或通过使用血管活性物质例如缓激肽生化地破坏BBB。使用BBB开孔(opening)以将特定试剂靶向脑肿瘤的潜力也是一种选择。当通过血管内注射施用组合物时,可以将BBB破坏剂与本发明的治疗用组合物共施用。穿过BBB递送的其它策略可能需要使用内源性转运系统,包括窖蛋白(Caveolin)-1介导的胞吞转运作用,载体介导的转运蛋白(例如葡萄糖和氨基酸载体),用于胰岛素或运铁蛋白的受体介导的胞吞转运作用以及主动流出转运蛋白(例如p-糖蛋白)。主动转运部分也可以与供本发明之用的治疗用化合物缀合,以促进穿过血管内皮壁的转运。另一方面,在BBB后面的治疗剂的药物递送可以为通过局部递送,例如通过鞘内递送,例如通过Ommaya贮液囊(参见例如美国专利号5,222,982和5385582,通过引用并入本文);通过快速浓注,例如通过注射器,例如玻璃体内或颅内地;通过持续输注,例如通过套管插入术,例如具有对流(参见例如,美国申请号20070254842,通过引用并入本文);或通过植入在其上已可逆附着试剂的设备(参见例如美国申请号20080081064和20090196903,通过引用并入本文)。

[0469] 通常,提供有效量的指导RNA和/或sRGN多肽和/或供体多核苷酸。如上文关于离体方法所讨论的,在体内的指导RNA和/或sRGN多肽和/或供体多核苷酸的有效量或有效剂量是相对于阴性对照(例如与空载体或不相干多肽接触的细胞),诱导在两个同源序列之间观察到的重组量的2倍增加或更多的量。重组的量可以通过任何方便的方法来测量,例如如上文所述和本领域已知的。指导RNA和/或sRGN多肽和/或供体多核苷酸的有效量或有效剂量的计算在本领域普通技术人员的技术内,并且对于本领域技术人员而言将是常规的。要施用的最终量将取决于施用途径以及要治疗的病症或状况的特性。

[0470] 施用特定患者的有效量将取决于多种因素,其中的几种将从一个患者到另一个患者不同。胜任的临床医生将能够确定治疗剂的有效量,以向患者施用以停止或逆转疾病状况的进展,正如所需要的。利用LD50动物数据以及对于所述试剂可获得的其它信息,临床医生可以取决于施用途径确定对于个体的最大安全剂量。例如,考虑到正向其中施用治疗用组合物的更大的流体实体,静脉内施用的剂量可以具有比鞘内施用的剂量更大量的治疗剂。类似地,可以以较高剂量或以重复剂量施用从身体快速清除的组合物,以维持治疗浓度。利用普通技术,胜任的临床医生将能够在常规临床试验过程中最优化特定治疗剂的剂量。

[0471] 对于在药物中的包含,可以从合适的商业来源获得指导RNA和/或sRGN多肽和/或供体多核苷酸。作为一般建议,每剂胃肠外施用的指导RNA和/或sRGN多肽和/或供体多核苷酸的总药学有效量将在可以通过剂量反应曲线测量的范围内。

[0472] 基于指导RNA和/或sRGN多肽和/或供体多核苷酸的疗法,即,要用于治疗性施用的指导RNA和/或sRGN多肽和/或供体多核苷酸的制剂必须是无菌的。无菌可由通过无菌滤过膜(例如0.2 1-1m膜)过滤达到。通常将治疗用组合物放置入具有无菌通道口的容器中,例如,具有可通过皮下注射针刺穿的塞子的静脉内溶液袋或小瓶。基于指导RNA和/或sRGN多肽和/或供体多核苷酸的疗法可以作为水溶液或作为用于重构的冻干制剂贮藏在单位或多剂容器中,例如,密封的安瓿或小瓶。作为冻干制剂的例子,将10- ml小瓶装填5 ml无菌过

滤的1％(w/v)化合物水溶液,并将所得混合物冻干。通过使用抑制细菌注射用水(Water-for-Injection)重构冻干的化合物来制备输注溶液。

[0473] 取决于所期望的制剂,药物组合物可以包括稀释剂的药学上可接受的无毒性载体,其被定义为通常用于配制用于动物或人施用的药物组合物的运载体。选择稀释剂以便不影响组合的生物活性。这种稀释剂的例子是蒸馏水、缓冲的水、生理盐水、PBS、林格液、葡萄糖溶液和Hank氏溶液(Hank's solution)。此外,药物组合物或制剂可包含其它载体,佐剂或无毒性、非治疗性、非免疫原性的稳定剂,赋形剂等。组合物还可包含额外的物质以近似生理条件,例如pH调节和缓冲剂、毒性调节剂、湿润剂和去污剂。

[0474] 组合物还可以包含多种稳定剂中的任何一种,如例如抗氧化剂。当药物组合物包含多肽时,可以将多肽与各种众所周知的化合物复合,所述化合物增强多肽的体内稳定性,或者以其它方式增强其药理性质(例如,增加多肽的半衰期、降低其毒性、提高溶解度或摄取)。这种修饰或复合剂的例子包括硫酸盐、葡糖酸盐、柠檬酸盐和磷酸盐。组合物的核酸或多肽也可以与增强其体内属性的分子复合。这种分子包括,例如,碳水化合物、多胺、氨基酸、其它肽、离子(例如钠、钾、钙、镁、锰)和脂质。

[0475] 关于适合于各种类型施用的制剂的进一步指导是本领域已知的,例如,Remington's Pharmaceutical Sciences, Mace Publishing Company, Philadelphia, Pa., 第21版(2006)。关于药物递送的方法的简要综述,参见, Langer, Science 249: 1527-1533 (1990)。

[0476] 可以施用药物组合物用于预防性和/或治疗性治疗。可以根据细胞培养和/或实验动物中的标准药学程序确定活性成分的毒性和治疗功效,包括,例如,确定LD50(对群体的50％致死的剂量)和ED50(在群体的50％治疗有效的剂量)。毒性和治疗效果之间的剂量比是治疗指数,且其可以表示为比率LD50/ED50。表现出大的治疗指数的治疗可能是优选的。

[0477] 从细胞培养和/或动物研究获得的数据可用于配制用于人的剂量范围。活性成分的剂量通常在包括具有低毒性的ED50的循环浓度范围内。取决于所采用的剂量形式和所使用的施用途径,剂量可以在这个范围内变化。用于配制药物组合物的组分在某些实施方案中具有高纯度并且基本上不含潜在有害的污染物(例如,至少国家食品(National Food)(NF)级,通常至少分析级,包括至少药物级)。此外,计划用于体内使用的组合物通常是无菌的。就必须在使用前合成给定化合物来说,所得产物通常基本上不含任何潜在的毒性剂,特别是任何内毒素,其在合成或纯化过程中可能存在。用于胃肠外(parental)施用的组合物也是无菌的、基本上是等渗的并在GMP条件下制备。

[0478] 要施用特定患者的治疗用组合物的有效量将取决于多种因素,其中的几种将从一个患者到另一个患者不同。胜任的临床医生将能够确定治疗剂的有效量,以向患者施用以停止或逆转疾病状况的进展,正如所需要的。利用LD50动物数据以及对于所述试剂可获得的其它信息,临床医生可以取决于施用途径确定对于个体的最大安全剂量。例如,考虑到正向其中施用治疗用组合物的更大的流体实体,静脉内施用的剂量可以大于鞘内施用的剂量。类似地,可以以较高剂量或以重复剂量施用从身体快速清除的组合物,以维持治疗浓度。利用普通技术,胜任的临床医生将能够在常规临床试验过程中最优化特定治疗剂的剂量。

[0479] *遗传修饰的宿主细胞*

[0480] 本公开内容提供了遗传修饰的宿主细胞,包括分离的遗传修饰的宿主细胞,其中遗传修饰的宿主细胞包含(已经用下述遗传修饰:

[0481] 1)外源指导RNA;2)外源核酸,其包含编码指导RNA的核苷酸序列;3)外源sRGN多肽(例如天然存在的sRGN;修饰的,即,突变的或变体,sRGN;嵌合的sRGN;等);4)外源核酸,其包含编码sRGN多肽的核苷酸序列;或5)以上的任何组合。遗传修饰的细胞是通过用例如下述遗传修饰的宿主细胞而产生的:1)外源指导RNA;2)外源核酸,其包含编码指导RNA的核苷酸序列;3)外源sRGN多肽;4)外源核酸,其包含编码sRGN多肽的核苷酸序列;或5)以上的任何组合)。

[0482] 适合作为靶细胞的所有细胞也适合作为遗传修饰的宿主细胞。例如,感兴趣的遗传修饰的宿主细胞可以是来自任何生物的细胞(例如细菌细胞,古细菌细胞,单细胞真核生物的细胞,植物细胞,藻类细胞,例如丛粒藻(Botryococcus braunii)、雷氏衣藻(Chlamydomonas reinhardtii)、微拟球藻(Nannochloropsis gaditana)、蛋白核小球藻(Chlorel/a pyrenoidosa)、展枝马尾藻(Sargassum patens C.Agardh)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、两栖动物、爬行动物、鸟、哺乳动物)的细胞,来自哺乳动物(例如猪、牛、山羊、绵羊、啮齿类动物、大鼠、小鼠、非人灵长类动物、人等)的细胞,等。

[0483] 在某些实施方案中,已用包含编码sRGN多肽(例如天然存在的sRGN;修饰的,即,突变的或变体,sRGN;嵌合的sRGN;等)的核苷酸序列的外源核酸对遗传修饰的宿主细胞进行了遗传修饰。通过将指导RNA(或编码指导RNA的DNA,其确定待修饰的基因组位置/序列)和任选地共同核酸引入细胞中,可以靶向遗传修饰的宿主细胞的DNA用于修饰。在某些实施方案中,编码sRGN多肽的核苷酸序列可操作地连接至诱导型启动子(例如热激启动子、四环素调节的启动子、类固醇调节的启动子、金属调节的启动子、雌激素受体调节的启动子等)。在某些实施方案中,编码sRGN多肽的核苷酸序列可操作地连接至空间限制和/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、细胞周期特异性启动子)。在某些实施方案中,编码sRGN多肽的核苷酸序列可操作地连接至组成型启动子。

[0484] 在某些实施方案中,遗传修饰的宿主细胞是体外的。在某些实施方案中,遗传修饰的宿主细胞是体内的。在某些实施方案中,遗传修饰的宿主细胞是原核细胞或衍生自原核细胞。在某些实施方案中,遗传修饰的宿主细胞是细菌细胞或衍生自细菌细胞。在某些实施方案中,遗传修饰的宿主细胞是古细菌细胞或衍生自古细菌细胞。在某些实施方案中,遗传修饰的宿主细胞是真核细胞或衍生自真核细胞。在某些实施方案中,遗传修饰的宿主细胞是植物细胞或衍生自植物细胞。在某些实施方案中,遗传修饰的宿主细胞是动物细胞或衍生自动物细胞。在某些实施方案中,遗传修饰的宿主细胞是无脊椎动物细胞或衍生自无脊椎动物细胞。在某些实施方案中,遗传修饰的宿主细胞是脊椎动物细胞或衍生自脊椎动物细胞。在某些实施方案中,遗传修饰的宿主细胞是哺乳动物细胞或衍生自哺乳动物细胞。在某些实施方案中,遗传修饰的宿主细胞是啮齿类动物细胞或衍生自啮齿类动物细胞。在某些实施方案中,遗传修饰的宿主细胞是人细胞或衍生自人细胞。

[0485] 本公开内容进一步提供了遗传修饰的细胞的后代,其中所述后代可以包含与衍生其的遗传修饰的细胞相同的外源核酸或多肽。本公开内容进一步提供了包含遗传修饰的宿主细胞的组合物。

[0486] *遗传修饰的干细胞和遗传修饰的祖先细胞*

[0487] 在某些实施方案中,遗传修饰的宿主细胞是遗传修饰的干细胞或祖先细胞。合适的宿主细胞包括,例如干细胞(成体干细胞、胚胎干细胞、iPS细胞等)和祖先细胞(例如心脏祖先细胞、神经祖先细胞等)。合适的宿主细胞包括哺乳动物干细胞和祖先细胞,包括,例如啮齿类动物干细胞、啮齿类动物祖先细胞、人干细胞、人祖先细胞等。合适的宿主细胞包括体外宿主细胞,例如分离的宿主细胞。

[0488] 在某些实施方案中,遗传修饰的宿主细胞包含外源指导RNA核酸。在某些实施方案中,遗传修饰的宿主细胞包含外源核酸,其包含编码指导RNA的核苷酸序列。在某些实施方案中,遗传修饰的宿主细胞包含外源sRGN多肽(例如天然存在的sRGN;修饰的,即,突变的或变体,sRGN;嵌合的sRGN;等)。在某些实施方案中,遗传修饰的宿主细胞包含外源核酸,其包含编码sRGN多肽的核苷酸序列。在某些实施方案中,遗传修饰的宿主细胞包含外源核酸,其包含编码1)指导RNA和2)sRGN多肽的核苷酸序列。

[0489] 在某些情况下,sRGN多肽包含与SEQ ID NO: 1-12中的任何序列或其长度为至少100、150、200、300、350、400或500个氨基酸的活性部分具有至少75%、至少80%、至少85%、至少90%、至少95%、至少99%或100%氨基酸序列同一性的氨基酸序列。在某些实施方案中,所述活性部分包含RNA酶结构域。在其它实施方案中,所述活性部分包含DNA酶结构域。

[0490] *组合物*

[0491] 本公开内容提供了包含指导RNA和/或定位修饰多肽的组合物。在某些情况下,sRGN多肽是嵌合多肽。组合物可用于执行本公开内容的方法,例如用于靶DNA的位点特异性修饰的方法;用于与靶DNA相关的多肽的位点特异性修饰的方法;等等。

[0492] *包含指导RNA的组合物*

[0493] 本公开内容提供了包含指导RNA的组合物。除指导RNA外,组合物可包含下述的一种或多种:盐,例如NaCl、MgCl、KCl、MgSO4等;缓冲剂,例如Tris缓冲液、N-(2-羟基乙基)哌嗪-N'-(2-乙磺酸) (HEPES)、2-(N-吗啉代)乙磺酸(MES)、MES钠盐、3-(N- 吗啉代)丙烷磺酸(MOPS)、N-三[羟基甲基]甲基-3-氨基丙烷磺酸(TAPS)等;增溶剂;去污剂,例如非离子型去污剂诸如吐温®20等;核酸酶抑制剂;等。例如,在某些情况下,组合物包含指导RNA和用于稳定核酸的缓冲液。

[0494] 在某些实施方案中,存在于组合物中的指导RNA是纯的,例如,至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或超过99%纯的,其中"%纯度"是指,指导RNA为不含在指导RNA生产过程中可能存在的其它大分子或污染物的所述百分比。

[0495] *包含嵌合多肽的组合物*

[0496] 本公开内容提供了嵌合多肽的组合物。除指导RNA外,组合物可包含下述的一种或多种:盐,例如,NaCl、MgCl、KCl、MgSO4等;缓冲剂,例如,Tris缓冲液、HEPES、MES、MES钠盐、MOPS、TAPS等;增溶剂;去污剂,例如,非离子型去污剂诸如吐温®20等;蛋白酶抑制剂;还原剂(例如,二硫苏糖醇);等。

[0497] 在某些实施方案中,存在于组合物中的嵌合多肽是纯的,例如,至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或超过99%纯的,其中"%纯度"是指,定位修饰多肽为不含在嵌合多肽生产过程中可能存在的其它蛋白、其它大分子或污

染物的所述百分比。

[0498] *包含指导RNA和sRGN多肽的组合物*

[0499] 本公开内容提供了组合物,其包含:(i)指导RNA或编码它的DNA核酸;和ii)sRGN多肽或编码它的核酸。在某些情况下,sRGN多肽是嵌合的sRGN多肽。在其它情况下,sRGN多肽是天然存在的sRGN多肽。在某些情况下,sRGN多肽表现出修饰靶DNA的酶活性。在其它情况下,sRGN多肽表现出修饰与靶DNA相关的多肽的酶活性。在再其它情况下,sRGN多肽调节靶DNA的转录。

[0500] 本公开内容提供了组合物,其包含:(i)如上所述的指导RNA,或编码它的DNA核酸,所述指导RNA包含:(a)包含与靶DNA中的序列互补的核苷酸序列的第一区段;和(b)与sRGN多肽相互作用的第二区段;和(ii)sRGN多肽,或编码它的核酸,所述sRGN多肽包含:(a)与所述指导RNA相互作用的RNA-结合部分;和(b)表现出定位酶活性的活性部分,其中酶活性的位点由所述指导RNA确定。

[0501] 在某些情况下,组合物包含:(i)指导RNA,所述指导RNA包含:(a)包含与靶DNA中的序列互补的核苷酸序列的第一区段;和(b)与sRGN多肽相互作用的第二区段;和(ii)sRGN多肽,所述sRGN多肽包含:(a)与所述指导RNA相互作用的RNA-结合部分;和(b)表现出定位酶活性的活性部分,其中酶活性的位点由所述指导RNA确定。

[0502] 在其它实施方案中,组合物包含:(i)编码指导RNA的核酸,所述指导RNA包含:(a)包含与靶DNA中的序列互补的核苷酸序列的第一区段;和(b)与定位修饰多肽相互作用的第二区段;和(ii)编码sRGN多肽的核酸,所述sRGN多肽包含:(a)与所述指导RNA相互作用的RNA-结合部分;和(b)表现出定位酶活性的活性部分,其中酶活性的位点由所述指导RNA确定。

[0503] 本公开内容提供了组合物,其包含:(i)指导RNA或编码它的DNA核酸,所述指导RNA包含:(a)包含与靶DNA中的序列互补的核苷酸序列的第一区段;和(b)与sRGN多肽相互作用的第二区段;和(ii)sRGN多肽,或编码它的核酸,所述sRGN多肽包含:(a)与所述指导RNA相互作用的RNA-结合部分;和(b)调节靶DNA内转录的活性部分,其中靶DNA内调节的转录的位点由所述指导RNA确定。

[0504] 例如,在某些情况下,组合物包含:(i)指导RNA,所述指导RNA包含:(a)包含与靶DNA中的序列互补的核苷酸序列的第一区段;和(b)与sRGN多肽相互作用的第二区段;和(ii)sRGN多肽,所述sRGN多肽包含:(a)与所述指导RNA相互作用的RNA-结合部分;和(b)调节靶DNA内转录的活性部分,其中靶DNA内调节的转录的位点由所述指导RNA确定。

[0505] 作为另一个实例,在某些情况下,组合物包含:(i)编码指导RNA的DNA核酸,所述指导RNA包含:(a)包含与靶DNA中的序列互补的核苷酸序列的第一区段;和(b)与sRGN多肽相互作用的第二区段;和(ii)编码定位修饰多肽的多核苷酸,所述sRGN多肽包含:(a)与所述指导RNA相互作用的RNA-结合部分;和(b)调节靶DNA内转录的活性部分,其中靶DNA内调节的转录的位点由所述指导RNA确定。除i)指导RNA或编码它的DNA核酸;和ii)sRGN多肽或编码它的核酸外,组合物可以包含下述的一种或多种:盐,例如NaCl、MgCl、KCl、MgSO4等;缓冲剂,例如Tris缓冲液、HEPES、MES、MES钠盐、MOPS、TAPS等;增溶剂;去污剂,例如非离子型去污剂诸如吐温-20等;蛋白酶抑制剂;还原剂(例如二硫苏糖醇);等。

[0506] 在某些情况下,组合物的组分各自是纯的,例如每种组分为至少75%、至少80%、

至少90％、至少95％、至少98％、至少99％或至少99％纯的。在某些情况下,组合物的各种组分在加入到组合物中之前是纯的。

[0507] 例如,在某些实施方案中,存在于组合物中的sRGN多肽是纯的,例如,至少75％、至少80％、至少85％、至少90％、至少95％、至少98％、至少99％或超过99％纯的,其中“％纯度”是指sRGN多肽为不含在sRGN多肽生产过程中可能存在的其它蛋白(例如,除sRGN多肽以外的蛋白)、其它大分子或污染物的所述百分比。

[0508] *试剂盒*

[0509] 本公开内容提供了用于执行方法的试剂盒。试剂盒可以包括下述的一种或多种:sRGN多肽;包含编码定位修饰多肽的核苷酸的核酸;指导RNA;包含编码指导RNA的核苷酸序列的核酸。试剂盒可以包括复合物,所述复合物包含下述的两种或更多种:sRGN多肽;包含编码sRGN多肽的核苷酸的核酸;指导RNA;包含编码指导RNA的核苷酸序列的核酸。在某些实施方案中,试剂盒包括sRGN多肽或编码它的核酸。在某些实施方案中,sRGN多肽包含:(a)与指导RNA相互作用的RNA-结合部分;和(b)调节靶DNA内转录的活性部分,其中所述指导RNA确定靶DNA内调节的转录的位点。在某些情况下,sRGN多肽的活性部分表现出降低或失活的核酸酶活性。在某些情况下,sRGN多肽是嵌合的sRGN多肽。

[0510] 在某些实施方案中,试剂盒包括:sRGN多肽或编码它的核酸,以及用于重构和/或稀释sRGN多肽的试剂。在其它实施方案中,试剂盒包括核酸(例如DNA、RNA),其包含编码sRGN多肽的核苷酸。在某些实施方案中,试剂盒包括:核酸(例如DNA、RNA),其包含编码sRGN多肽的核苷酸;以及用于重构和/或稀释sRGN多肽的试剂。

[0511] 包含sRGN多肽或编码它的核酸的试剂盒可以进一步包括一种或多种额外的试剂,其中这种额外的试剂可以选自:用于将sRGN多肽引入细胞的缓冲液;洗涤缓冲液;对照试剂;对照表达载体或RNA多核苷酸;用于从DNA体外生产sRGN多肽的试剂等。在某些情况下,试剂盒中包括的定位修饰多肽是如上所述的嵌合sRGN多肽。

[0512] 在某些实施方案中,试剂盒包括指导RNA或编码它的DNA核酸,所述指导RNA包含:(a)包含与靶DNA中的序列互补的核苷酸序列的第一区段;和(b)与定位修饰多肽相互作用的第二区段。在某些实施方案中,试剂盒包括:(i)指导RNA或编码它的DNA核酸,所述指导RNA包含:(a)包含与靶DNA中的序列互补的核苷酸序列的第一区段;和(b)与sRGN多肽相互作用的第二区段;和(ii)sRGN多肽或编码它的核酸,所述sRGN多肽包含:(a)与所述指导RNA相互作用的RNA-结合部分;和(b)表现出定位酶活性的活性部分,其中酶活性的位点由所述指导RNA确定。在某些实施方案中,sRGN多肽的活性部分不表现出酶活性(包括失活的核酸酶,例如通过突变)。在某些情况下,试剂盒包括指导RNA和sRGN多肽。在其它情况下,试剂盒包括:(i)包含编码指导RNA的核苷酸序列的核酸;和(ii)包含编码sRGN多肽的核苷酸序列的核酸。作为另一个例子,试剂盒可以包括:(i)指导RNA或编码它的DNA核酸,包含:

[0513] (a)包含与靶DNA中的序列互补的核苷酸序列的第一区段;和(b)与sRGN多肽相互作用的第二区段;和(ii)sRGN多肽或编码它的核酸,包含:(a)与所述指导RNA相互作用的RNA-结合部分;和(b)调节靶DNA内转录的活性部分,其中靶DNA内调节的转录的位点由所述指导RNA确定。在某些情况下,试剂盒包括:(i)指导RNA;和sRGN多肽。在其它情况下,试剂盒包括:(i)包含编码指导RNA的核苷酸序列的核酸;和(ii)包含编码sRGN多肽的核苷酸序列的核酸。本公开内容提供了试剂盒,其包括:(1)重组表达载体,其包含(i)编码指导RNA的核

苷酸序列,其中所述指导RNA包含:(a)包含与靶DNA中的序列互补的核苷酸序列的第一区段;和(b)与sRGN多肽相互作用的第二区段;和(ii)编码sRGN多肽的核苷酸序列,其中所述sRGN多肽包含:(a)与所述指导RNA相互作用的RNA-结合部分;和(b)表现出定位酶活性的活性部分,其中酶活性的位点由所述指导RNA确定;和(2)用于重构和/或稀释所述表达载体的试剂。

[0514] 本公开内容提供了试剂盒,其包括:(1)重组表达载体,其包含(i)编码指导RNA的核苷酸序列,其中所述指导RNA包含:(a)包含与靶DNA中的序列互补的核苷酸序列的第一区段;和(b)与sRGN多肽相互作用的第二区段;和(ii)编码sRGN多肽的核苷酸序列,其中所述sRGN多肽包含:(a)与所述指导RNA相互作用的RNA-结合部分;和(b)调节靶DNA内转录的活性部分,其中靶DNA内调节的转录的位点由所述指导RNA确定;和(2)用于重构和/或稀释所述重组表达载体的试剂。

[0515] 本公开内容提供了试剂盒,其包括:(1)包含核酸的重组表达载体,所述核酸包含编码DNA靶向RNA的核苷酸序列,所述DNA靶向RNA包含:(i)包含与靶DNA中的序列互补的核苷酸序列的第一区段;和(ii)与sRGN多肽相互作用的第二区段;和(2)用于重构和/或稀释所述重组表达载体的试剂。在这个试剂盒的一些实施方案中,试剂盒包括:重组表达载体,其包含编码sRGN多肽的核苷酸序列,其中所述sRGN多肽包含:(a)与指导RNA相互作用的RNA-结合部分;和(b)表现出定位酶活性的活性部分,其中酶活性的位点由所述指导RNA确定。在这个试剂盒的其它实施方案中,试剂盒包括:重组表达载体,其包含编码sRGN多肽的核苷酸序列,其中所述sRGN多肽包含:(a)与指导RNA相互作用的RNA-结合部分;和(b)调节靶DNA内转录的活性部分,其中靶DNA内调节的转录的位点由所述指导RNA确定。

[0516] 在任何上述试剂盒的一些实施方案中,试剂盒包括单分子指导RNA。在任何上述试剂盒的一些实施方案中,试剂盒包括两种或更多种单分子指导RNA。在任何上述试剂盒的一些实施方案中,可以作为阵列(例如,RNA分子的阵列,编码一种或多种指导RNA的DNA分子的阵列等)提供指导RNA(例如,包括两种或更多种指导RNA)。这种试剂盒可用于例如与包含sRGN多肽的上述遗传修饰的宿主细胞一起使用。在任何上述试剂盒的一些实施方案中,试剂盒进一步包括供体多核苷酸以实现所期望的遗传修饰。试剂盒的组分可以在单独的容器中;或可以组合在单个容器中。

[0517] 在某些情况下,试剂盒进一步包括一种或多种相对于野生型sRGN表现出降低的内切脱氧核糖核酸酶(endodeoxyribonuclease)活性的变体sRGN定位多肽。

[0518] 在某些情况下,试剂盒进一步包括一种或多种包含编码变体sRGN定位多肽的核苷酸序列的核酸,所述变体sRGN定位多肽相对于野生型sRGN表现出降低的内切脱氧核糖核酸酶活性。

[0519] 任何上述试剂盒都可进一步包括一种或多种额外的试剂,其中这种额外的试剂可选自:稀释缓冲液;重构溶液;洗涤缓冲液;对照试剂;对照表达载体或RNA多核苷酸;用于从DNA体外生产sRGN多肽的试剂,等。

[0520] 除上述组分外,试剂盒可进一步包括使用试剂盒的组分以实施方法的说明书。通常将用于实施方法的说明书记录在合适的记录介质上。例如,说明书可以印刷在诸如纸或塑料等的基底上。因此,说明书可以作为包装插页存在于试剂盒中,存在于试剂盒或其组分的容器标签中(即,与包装或分包装(subpackaging)相关的)等。在其它实施方案中,说明书

作为存在于适当的计算机可读存储介质(例如CD-ROM、软磁盘、闪存驱动器(flash drive)等)上的电子存储数据文件呈现。在再其它的实施方案中,试剂盒中不存在实际的说明书,但是提供了用于从远程来源(例如,经由因特网)获得说明书的手段。这个实施方案的示例是包括网址的试剂盒,在所述网址可以查看说明书和/或可以从其下载说明书。如同说明书一样,用于获得说明书的这种手段被记录在合适的基底上。

[0521] *非人遗传修饰的生物*

[0522] 在某些实施方案中,已用包含编码sRGN多肽(例如天然存在的sRGN;修饰的,即,突变的或变体,sRGN;嵌合的sRGN;等)的核苷酸序列的外源核酸对遗传修饰的宿主细胞进行了遗传修饰。如果这种细胞是真核单细胞生物,那么修饰的细胞可以被认为是遗传修饰的生物。在某些实施方案中,非人遗传修饰的生物是sRGN转基因多细胞生物。

[0523] 在某些实施方案中,遗传修饰的非人宿主细胞(例如已用包含编码定位修饰多肽(例如天然存在的sRGN;修饰的,即,突变的或变体,sRGN;嵌合的sRGN;等)的核苷酸序列的外源核酸遗传修饰的的细胞)可以产生遗传修饰的非人生物(例如,小鼠、鱼、蛙、蝇、蠕虫等)。例如,如果遗传修饰的宿主细胞是多潜能干细胞(即,PSC)或生殖细胞(例如,精子、卵母细胞等),则整个遗传修饰的生物都可以衍生自遗传修饰的宿主细胞。在某些实施方案中,遗传修饰的宿主细胞是可产生遗传修饰的生物的体内或体外的多潜能干细胞(例如ESC、iPSC、多潜能植物干细胞等)或生殖细胞(例如精子细胞、卵母细胞等)。在某些实施方案中,遗传修饰的宿主细胞是脊椎动物PSC(例如ESC、iPSC等),并用于产生遗传修饰的生物(例如,通过将PSC注射入胚泡中以产生嵌合/镶嵌型动物,其然后可以交配以产生非嵌合/非镶嵌型遗传修饰的生物;在植物的情况下嫁接;等)。用于产生遗传修饰的生物的任何方便的方法/规程,包括本文所述的方法,都适合于产生包含外源核酸的遗传修饰的宿主细胞,所述外源核酸包含编码sRGN多肽(例如天然存在的sRGN;修饰的,即,突变的或变体,sRGN;嵌合的sRGN;等)的核苷酸序列。产生遗传修饰的生物的方法是本领域已知的。例如,参见Cho等人, Curr Protoc Cell Biol.2009年3月;第19章:Unit 19.11: Generation of transgenic mice;Gama等人, Brain Struct Funct.2010年3月;214(2-3):91-109.Epub 2009年11月25日: Animal transgenesis: an overview;Husaini等人, GM Crops.2011年6月-12月;2(3):150-62.Epub 2011年6月1日: Approaches for gene targeting and targeted gene expression in plants。

[0524] 在某些实施方案中,遗传修饰的生物包含用于本发明方法的靶细胞,且因此可以被认为是靶细胞的来源。例如,如果包含外源核酸的遗传修饰的细胞被用于产生遗传修饰的生物,所述外源核酸包含编码定位修饰多肽(例如天然存在的sRGN;修饰的,即,突变的或变体,sRGN;嵌合的sRGN;等)的核苷酸序列,那么遗传修饰的生物的细胞包含所述外源核酸,其包含编码sRGN多肽(例如天然存在的sRGN;修饰的,即,突变的或变体,sRGN;嵌合的sRGN;等)的核苷酸序列。在一些这种实施方案中,可以通过将指导RNA(或编码指导RNA的DNA)和任选地供体核酸引入一个或多个细胞来靶向遗传修饰的生物的一个或多个细胞的DNA用于修饰。例如,将指导RNA(或编码指导RNA的DNA)引入遗传修饰的生物的细胞亚群(例如脑细胞、肠细胞、肾细胞、肺细胞、血细胞等)可以靶向这种细胞的DNA用于修饰,其基因组位置将取决于引入的指导RNA的DNA-靶向序列。

[0525] 在某些实施方案中,遗传修饰的生物是用于本发明方法的靶细胞的来源。例如,包

含用外源核酸遗传修饰的的细胞的遗传修饰的生物可以提供遗传修饰的细胞的来源,例如PSC(例如ESC、iPSC、精子、卵母细胞等)、神经元、祖先细胞、心肌细胞等,所述外源核酸包含编码sRGN多肽(例如天然存在的sRGN;修饰的,即,突变的或变体,sRGN;嵌合的sRGN;等)的核苷酸序列。

[0526] 在某些实施方案中,遗传修饰的细胞是包含外源核酸的PSC,所述外源核酸包含编码sRGN多肽(例如天然存在的sRGN;修饰的,即,突变的或变体,sRGN;嵌合的sRGN;等)的核苷酸序列。因此,PSC可以是靶细胞,从而使得可以通过向PSC中引入指导RNA(或编码指导RNA的DNA)和任选地供体核酸来靶向PSC的DNA用于修饰,且修饰的基因组位置将取决于引入的指导RNA的DNA-靶向序列。因此,在某些实施方案中,本文描述的方法可用于修饰衍生自遗传修饰的生物的PSC的DNA(例如,删除和/或置换任何期望的基因组位置)。然后可以将这种修饰的PSC用于产生具有下述两者的生物:(i)包含编码sRGN多肽(例如天然存在的sRGN;修饰的,即,突变的或变体,sRGN;嵌合的sRGN;等)的核苷酸序列的外源核酸和(ii)引入PSC中的DNA修饰。

[0527] 包含编码sRGN多肽(例如天然存在的sRGN;修饰的,即,突变的或变体,sRGN;嵌合的sRGN;等)的核苷酸序列的外源核酸可以在未知启动子的控制下(即,与其可操作地连接)(例如,当核酸随机整合到宿主细胞基因组中时),或可以在已知启动子的控制下(即,与其可操作地连接)。合适的已知启动子可以是任何已知的启动子,且包括组成型活性启动子(例如CMV启动子)、诱导型启动子(例如热激启动子、四环素调节的启动子、类固醇调节的启动子、金属调节的启动子、雌激素受体调节的启动子等)、空间限制和/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。

[0528] 遗传修饰的生物(例如,其细胞包含编码sRGN多肽(例如天然存在的sRGN;修饰的,即,突变的或变体,sRGN;嵌合的sRGN;等)的核苷酸序列的生物)可以是任何生物,包括例如,植物;藻类;无脊椎动物(例如刺胞动物、棘皮动物、蠕虫、蝇等);脊椎动物(例如鱼(例如斑马鱼、河豚、金鱼等)、两栖动物(例如蝾螈、蛙等)、爬行动物、鸟、哺乳动物等);有蹄类动物(ungulate)(例如山羊、猪、绵羊、牛等);啮齿类动物(例如小鼠、大鼠、仓鼠、豚鼠);兔形目(例如兔);等。

[0529] 在某些情况下,sRGN多肽包含与SEQ ID NO: 1-12具有至少75%、至少80%、至少85%、至少90%、至少95%、至少99%或100%氨基酸序列同一性的氨基酸序列。

[0530] *转基因的非人动物*

[0531] 如上所述,在某些实施方案中,核酸(例如编码sRGN多肽(例如天然存在的sRGN;修饰的,即,突变的或变体,sRGN;嵌合的sRGN;等)的核苷酸序列)或重组表达载体被用作转基因以产生生产sRGN多肽的转基因动物。因此,本公开内容进一步提供了转基因非人动物,所述动物包含转基因,所述转基因包含核酸,所述核酸包含编码定位修饰多肽(例如天然存在的sRGN;修饰的,即,突变的或变体,sRGN;嵌合的sRGN;等)的核苷酸序列,如上所述。在某些实施方案中,转基因非人动物的基因组包含编码sRGN多肽的核苷酸序列。在某些实施方案中,转基因非人动物对于遗传修饰是纯合的。在某些实施方案中,转基因非人动物对于遗传修饰是杂合的。在某些实施方案中,转基因非人动物是脊椎动物,例如,鱼(例如斑马鱼、金鱼、河豚、洞穴鱼(cave fish)等)、两栖动物(蛙、蝾螈等)、鸟(例如鸡、火鸡等)、爬行动物(例如蛇、蜥蜴等)、哺乳动物(例如有蹄类动物,例如猪、牛、山羊、绵羊等;兔形目(例如兔);

啮齿类动物(例如大鼠、小鼠);非人灵长类动物;等)等。

[0532] 包含编码sRGN多肽(例如天然存在的sRGN;修饰的,即,突变的或变体,sRGN;嵌合的sRGN;等)的核苷酸序列的外源核酸可以在未知启动子的控制下(即,与其可操作地连接)(例如,当核酸随机整合到宿主细胞基因组中时),或可以在已知启动子的控制下(即,与其可操作地连接)。合适的已知启动子可以是任何已知的启动子,且包括组成型活性启动子(例如CMV启动子)、诱导型启动子(例如热激启动子、四环素调节的启动子、类固醇调节的启动子、金属调节的启动子、雌激素受体调节的启动子等)、空间限制和/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。

[0533] *转基因植物*

[0534] 如上所述,在某些实施方案中,核酸(例如编码sRGN多肽(例如天然存在的sRGN;修饰的,即,突变的或变体,sRGN;嵌合的sRGN;等)的核苷酸序列)或重组表达载体被用作转基因以产生生产sRGN多肽的转基因植物。因此,本公开内容进一步提供了转基因植物,所述植物包含转基因,所述转基因包含核酸,所述核酸包含编码sRGN多肽(例如天然存在的sRGN;修饰的,即,突变的或变体,sRGN;嵌合的sRGN;等)的核苷酸序列,如上所述。在某些实施方案中,转基因植物的基因组包含核酸。在某些实施方案中,转基因植物对于遗传修饰是纯合的。在某些实施方案中,转基因植物对于遗传修饰是杂合的。

[0535] 将外源核酸引入植物细胞的方法是本领域众所周知的。如上文所定义,这种植物细胞被认为是“转化的”。合适的方法包括病毒感染(例如双链DNA病毒)、转染、接合、原生质体融合、电穿孔、基因枪技术、磷酸钙沉淀、直接显微注射、碳化硅须晶(silicon carbide whiskers)技术、土壤杆菌属(Agrobacterium)介导的转化等。方法的选择通常取决于正被转化的细胞的类型和发生转化的环境(即,体外、离体或体内)。基于土壤细菌根癌土壤杆菌(Agrobacterium tumefaciens)的转化方法对于将外源核酸分子引入维管植物中特别有用。土壤杆菌属的野生型形式含有Ti(根瘤诱导)质粒,其指导在宿主植物上产生根瘤的根癌生长的产生。Ti质粒的根瘤诱导T-DNA区域转移到植物基因组需要Ti质粒编码的毒力基因以及T-DNA边界,其是一组画出要转移的区域的轮廓的同向DNA重复序列。基于土壤杆菌属的载体是Ti质粒的修饰形式,其中根瘤诱导功能被引入植物宿主的感兴趣的核酸序列置换。

[0536] 土壤杆菌属介导的转化通常使用共整合载体或双元载体(binary vector)系统,其中Ti质粒的组分被分配给永久存在于土壤杆菌属宿主中并携带毒力基因的辅助载体和含有以T-DNA序列为界的感兴趣的基因的穿梭载体。多种双元载体在本领域中是众所周知的,并且可以例如从Clontech (Palo Alto, Calif.)商购获得。例如,将土壤杆菌属与培养的植物细胞或受伤的组织如叶组织、根外植体、下胚轴(hypocotyledons)、茎断片或块茎共培养的方法也是本领域众所周知的。参见,例如Glick和Thompson, (编), Methods in Plant Molecular Biology and Biotechnology, Boca Raton, Fla.: CRC Press (1993)。

[0537] 微粒介导的转化也可以用于产生转基因植物。首先由Klein等人(Nature 327:70-73 (1987))描述的这种方法依赖于通过用氯化钙、亚精胺或聚乙二醇沉淀而用所期望的核酸分子包被的微粒,例如金或钨。使用诸如BIOLISTIC PD-1000 (Biorad;Hercules Calif.)的设备将微粒颗粒高速地加速到被子植物组织中。

[0538] 可以以使得核酸能够进入一种或多种植物细胞的方式将核酸引入到植物中,例如通过体内或离体规程。“体内”是指将核酸施用于活的植物体,例如渗入。“离体”是指在植物外部修饰细胞或外植体,并然后将这种细胞或器官再生为植物。已经描述了许多适合于植物细胞稳定转化或适合于确立转基因植物的载体,包括在Weissbach和Weissbach, (1989) Methods for Plant Molecular Biology Academic Press和Gelvin等人, (1990) Plant Molecular Biology Manual, Kluwer Academic Publishers中描述的那些。具体的例子包括衍生自根癌土壤杆菌的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。另一方面,非Ti载体可以用于通过使用游离DNA递送技术将DNA转移到植物和细胞中。通过使用这些方法,可以产生转基因植物,例如小麦、稻(Christou (1991) Bio/Technology 9:957-9和4462)和玉米(Gordon-Kamm (1990) Plant Cell 2: 603-618)。未成熟胚胎也可以是用于通过使用基因枪进行直接DNA递送技术(Weeks等人(1993) Plant Physiol 102: 1077-1084;Vasil (1993) Bio/Technolo 10: 667-674;Wan和Lemeaux (1994) Plant Physiol104: 37-48和用于土壤杆菌属介导的DNA转移(Ishida等人(1996) Nature Biotech 14: 745-750)的单子叶植物的良好靶组织。将DNA引入叶绿体的示例性方法是生物射弹轰击、原生质体的聚乙二醇转化和显微注射(Daniell等人Nat.Biotechnol 16:345-348, 1998;Staub等人Nat.Biotechnol 18: 333-338, 2000;O’Neill等人Plant J.3:729-738, 1993;Knoblauch等人Nat.Biotechnol17: 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;和Boynton等人, Methods in Enzymology 217: 510-536 (1993)、Svab等人, Proc.Natl.Acad.Sci.USA 90: 913-917 (1993)和McBride等人, Proc.Nati.Acad.Sci.USA 91: 7301-7305 (1994))。任何适合于生物射弹轰击、原生质体的聚乙二醇转化和显微注射的方法的载体都将适合作为用于叶绿体转化的靶向载体。任何双链DNA载体都可以用作转化载体,尤其是当引入方法不利用土壤杆菌属时。

[0539] 可以进行遗传修饰的的植物包括谷物、饲料作物、水果、蔬菜、油料种子作物、棕榈、森林和藤本植物。可以被修饰的植物的具体实例如下:玉米、香蕉、花生、紫花豌豆、向日葵、番茄、油菜、烟草、小麦、大麦、燕麦、马铃薯、大豆、棉花、康乃馨、高粱、羽扇豆和稻。

[0540] 本公开内容还提供了转化的植物细胞,含有转化的植物细胞的组织、植物和产物。转化的细胞以及包含其的组织和产物的特征是整合到基因组中的核酸的存在,以及由植物细胞产生sRGN多肽,例如天然存在的sRGN;修饰的,即,突变的或变体,sRGN;嵌合的sRGN;等。本发明的重组植物细胞作为重组细胞群体,或作为组织、种子、完整植物、茎、果实、叶、根、花、茎、块茎、谷粒、动物饲料、植物田(a field of plants)等是有用的。

[0541] 包含编码sRGN多肽(例如天然存在的sRGN;修饰的,即,突变的或变体,sRGN;嵌合的sRGN;等)的核苷酸序列的核酸可以在未知启动子的控制下(即,与其可操作地连接)(例如,当核酸随机整合到宿主细胞基因组中时),或可以在已知启动子的控制下(即,与其可操作地连接)。合适的已知启动子可以是任何已知的启动子,且包括组成型活性启动子、诱导型启动子、空间限制和/或时间限制的启动子等。

[0542] 本公开内容提供了调节宿主细胞中靶核酸的转录的方法。所述方法通常包括使靶核酸与酶促无活性的sRGN多肽和指导RNA接触。所述方法在也被提供的多种应用中有用。

[0543] 本公开内容的转录调节方法克服了涉及RNAi的方法的一些缺点。本公开内容的转录调节方法应用于多种应用中,包括研究应用、药物发现(例如高通量筛选)、靶标确认、工业应用(例如作物工程;微生物工程等)、诊断应用、治疗应用和成像技术。

[0544] *调节转录的方法*

[0545] 本公开内容提供了选择性调节宿主细胞中靶DNA的转录的方法。所述方法通常包括:a)向宿主细胞引入:i)指导RNA,或包含编码所述指导RNA的核苷酸序列的核酸;和ii)变体sRGN定位多肽("变体sRGN多肽"),或包含编码所述变体sRGN多肽的核苷酸序列的核酸,其中所述变体sRGN多肽表现出降低的内切脱氧核糖核酸酶活性。

[0546] 指导RNA(本文也称为"指导RNA";或"gRNA")包含:i)包含与靶DNA中的靶序列互补的核苷酸序列的第一区段;ii)与sRGN多肽相互作用的第二区段;和iii)转录终止子。包含与靶DNA中的靶序列互补的核苷酸序列的第一区段在本文中被称为"靶向区段"。与sRGN多肽相互作用的第二区段在本文中也称为"蛋白结合序列"或"dsRGN-结合发夹"或"dsRGN柄(handle)"。"区段"是指分子的区段/部分/区域,例如RNA中核苷酸的邻接序列段。除非在特定上下文中另外明确定义,否则"区段"的定义不限于特定数目的总碱基对,并且可以包括任何总长度的RNA分子区域,并且可以包括或可以不包括与其它分子具有互补性的区域。变体sRGN定位多肽包含:i)与指导RNA相互作用的RNA-结合部分;以及表现出降低的内切脱氧核糖核酸酶活性的活性部分。

[0547] 指导RNA和变体sRGN多肽在宿主细胞中形成复合物;所述复合物选择性地调节宿主细胞中靶DNA的转录。

[0548] 在某些情况下,本公开内容的转录调节方法提供宿主细胞中靶核酸的选择性调节(例如减少或增加)。例如,与在没有指导RNA/变体sRGN多肽复合物的情况下靶核酸的转录水平相比,靶核酸的转录的"选择性"减少使靶核酸的转录减少至少10%、至少20%、至少30%、至少40%、至少50%、至少60%、至少70%、至少80%、至少90%或大于90%。靶核酸的转录的选择性减少减少了靶核酸的转录,但不显著地减少非靶核酸的转录,例如,与在没有指导RNA/变体sRGN多肽复合物的情况下的非靶核酸的转录水平相比,非靶核酸的转录减少(即使有的话)小于10%。

[0549] *增加的转录*

[0550] 与在没有指导RNA/变体sRGN多肽复合物的情况下靶DNA的转录水平相比,靶DNA的"选择性"增加的转录可以使靶DNA的转录增加到至少1.1倍(例如至少1.2倍、至少1.3倍、至少1.4倍、至少1.5倍、至少1.6倍、至少1.7倍、至少1.8倍、至少1.9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4.5倍、至少5倍、至少6倍、至少7倍、至少8倍、至少9倍、至少10倍、至少12倍、至少15倍或至少20倍)。靶DNA的转录的选择性增加增加了靶DNA的转录,但是不显著地增加非靶DNA的转录,例如,与在没有指导RNA/变体sRGN多肽复合物的情况下的非靶向的DNA的转录水平相比,非靶DNA的转录增加到(即使有的话)小于约5-倍(例如,小于约4-倍、小于约3-倍、小于约2-倍、小于约1.8-倍、小于约1.6-倍、小于约1.4-倍、小于约1.2-倍或小于约1.1-倍)。

[0551] 作为非限制性例子,可以通过将dsRGN融合至异源序列来实现增加的转录。合适的融合配偶体包括、但不限于,提供通过直接作用于靶DNA或作用于与靶DNA相关的多肽(例如组蛋白或其它DNA-结合蛋白)而间接增加转录的活性的多肽。合适的融合配偶体包括、但不

限于提供甲基转移酶活性、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、脱泛素活性、腺苷化活性、脱腺苷化活性、苏素化活性、脱苏素化活性、核糖基化活性、脱核糖基化活性、巴豆酰化(crotonylation)、脱巴豆酰化(decrotonylation)、丙酰化(propionylation)、脱丙酰化(depropionylation)、豆蔻酰化活性或脱豆蔻酰化活性的多肽。

[0552] 额外的合适的融合配偶体包括、但不限于,直接提供靶核酸增加的转录的多肽(例如转录激活因子或其片段、募集转录激活因子的蛋白或其片段、小分子/药物应答转录调节物等)。

[0553] 使用dsRGN融合蛋白以增加原核生物中转录的方法的非限制性例子包括细菌单杂交(B1H)或双杂交(B2H)系统的修饰。在B1H系统中,将DNA结合结构域(BD)与细菌转录激活结构域(AD,例如,大肠杆菌(Escherichia coli)RNA聚合酶的α亚基(RNAPa))融合。因此,可以将dsRGN与包含AD的异源序列融合。当dsRGN融合蛋白到达启动子的上游区域(被指导RNA靶向到那里)时,dsRGN融合蛋白的AD(例如RNAPa)募集RNAP全酶,从而导致转录激活。在B2H系统中,BD不直接与AD融合;而是它们的相互作用由蛋白-蛋白相互作用(例如GAL11P-GAL4相互作用)介导。为了修饰供所述方法之用的这种系统,可以将dsRGN与提供蛋白-蛋白相互作用的第一蛋白序列融合(例如酵母GAL11P和/或GAL4蛋白质),并将RNAa与完成蛋白-蛋白相互作用的第二蛋白序列融合(例如如果GAL11P与dsRGN融合,则为GAL4,如果GAL4与dsRGN融合,则为GAL11P,等等)。GAL11P和GAL4之间的结合亲和力提高了结合效率和转录启动速率(firing rate)。

[0554] 使用dsRGN融合蛋白以增加真核生物中转录的方法的非限制性例子包括dsRGN与激活结构域(AD)的融合(例如GAL4、疱疹病毒激活蛋白VP16或VP64、人核因子NF-KB p65亚基等)。为了使系统可诱导,dsRGN融合蛋白的表达可以通过诱导型启动子控制(例如Tet-ON、Tet-OFF等)。指导RNA可以被进行设计以靶向已知的转录应答元件(例如启动子、增强子等)、已知的上游激活序列(UAS)、被怀疑能够控制靶DNA的表达的功能未知或已知的序列等。

[0555] *额外的融合配偶体*

[0556] 用于达到增加或减少的转录的融合配偶体的非限制性例子包括、但不限于,转录激活因子和转录阻遏蛋白结构域(例如Krüppel相关框(KRAB或SKD);Mad mSIN3相互作用结构域(SID);ERF阻遏蛋白结构域(ERD)等)。在一些这样的情况下,dsRGN融合蛋白被指导RNA靶向靶DNA中的特定位置(即序列),并发挥基因座特异性调节,例如阻断RNA聚合酶与启动子结合(其选择性地抑制转录激活因子功能)和/或修饰局部染色质状态(例如,当使用修饰靶DNA或修饰与靶DNA相关的多肽的融合序列时)。在某些情况下,所述变化是瞬时的(例如转录阻遏或激活)。在某些情况下,所述变化是可遗传的(例如,当对靶DNA或与靶DNA相关的蛋白如核小体组蛋白进行外遗传修饰时)。在某些实施方案中,异源序列可以与dsRGN多肽的C-末端融合。在某些实施方案中,异源序列可以与dsRGN多肽的N-末端融合。在某些实施方案中,异源序列可以与dsRGN多肽的内部部分(即,除了N-或C-末端以外的部分)融合。使用dsRGN融合蛋白的方法的生物学作用可以通过任何方便的方法进行检测(例如基因表达测定;基于染色质的测定,例如染色质免疫沉淀(ChIP)、染色质体内测定(CiA)等)。

[0557] 在某些情况下,方法包括使用两种或更多种不同的指导RNA。例如,可以在单个宿

主细胞中使用两种不同的指导RNA,其中所述两种不同的指导RNA靶向相同靶核酸中的两种不同的靶序列。因此,例如,转录调节方法可以进一步包括将第二指导RNA或包含编码所述第二指导RNA的核苷酸序列的核酸引入宿主细胞,其中所述第二指导RNA包含:i)包含与靶DNA中第二靶序列互补的核苷酸序列的第一区段;ii)与定位多肽相互作用的第二区段;和iii)转录终止子。在某些情况下,使用靶向相同靶核酸中的两种不同靶向序列的两种不同的指导RNA提供了靶核酸转录的增加的调节(例如减少或增加)。

[0558] 作为另一个例子,可以在单个宿主细胞中使用两种不同的指导RNA,其中所述两种不同的指导RNA靶向两种不同的靶核酸。因此,例如,转录调节方法可以进一步包括将第二指导RNA或包含编码所述第二指导RNA的核苷酸序列的核酸引入宿主细胞,其中所述第二指导RNA包含:i)包含与至少第二靶DNA中的靶序列互补的核苷酸序列的第一区段;ii)与定位多肽相互作用的第二区段;和iii)转录终止子。

[0559] 在某些实施方案中,核酸(例如,指导RNA,例如单分子指导RNA;供体多核苷酸;编码sRGN多肽的核酸;等等)包含提供额外的所期望的特征(例如,修饰或调节的稳定性;亚细胞靶向;追踪,例如荧光标记;蛋白或蛋白复合物的结合位点;等)的修饰或序列。非限制性例子包括:5'帽(例如7-甲基鸟苷酸帽(m 7G));3'聚腺苷酸化尾(即,3'聚腺苷酸尾);核糖开关(riboswitch)序列或适体序列(例如以允许蛋白和/或蛋白复合物调节的稳定性和/或调节的可接近性);终止子序列;形成dsRNA双链体(即发夹)的序列);将RNA靶向亚细胞位置(例如细胞核、线粒体、叶绿体等)的修饰或序列;提供追踪的修饰或序列(例如直接缀合至荧光分子,缀合至促进荧光检测的部分,允许荧光检测的序列等);提供蛋白的结合位点的修饰或序列(例如作用于DNA的蛋白,包括转录激活因子、转录阻遏蛋白、DNA甲基转移酶、DNA脱甲基酶、组蛋白乙酰转移酶、组蛋白脱乙酰酶等);RNA的修饰,其改变了这种RNA的结构,从而改变了sRGN核糖核蛋白;及其组合。

[0560] *DNA-靶向区段*

[0561] 指导RNA的DNA-靶向区段(或"DNA-靶向序列")包含与靶DNA内的特定序列互补的核苷酸序列(靶DNA的互补链)。

[0562] 换句话说,指导RNA的DNA-靶向区段通过杂交(即,碱基配对)以序列特异性方式与靶DNA相互作用。因此,DNA-靶向区段的核苷酸序列可以变化并确定指导RNA和靶DNA将相互作用的靶DNA内的位置。可以修饰(例如通过遗传工程)指导RNA的DNA-靶向区段以与靶DNA内的任何所期望的序列杂交。

[0563] *稳定性控制序列(例如转录终止子区段)*

[0564] 稳定性控制序列影响RNA(例如指导RNA)的稳定性。合适的稳定性控制序列的一个例子是转录终止子区段(即,转录终止序列)。指导RNA的转录终止子区段可具有约10个核苷酸至约100个核苷酸的总长度,例如约10个核苷酸(nt)至约20 nt、约20 nt至约30 nt、约30 nt至约40 nt、约40 nt至约50 nt、约50 nt至约60 nt、约60 nt至约70 nt、约70 nt至约80 nt、约80 nt至约90 nt或约90 nt至约100 nt。例如,转录终止子区段可以具有约15个核苷酸(nt)至约80 nt、约15 nt至约50 nt、约15 nt至约40 nt、约15 nt至约30 nt或约15 nt至约25 nt的长度。

[0565] 在某些情况下,转录终止序列是在真核细胞中起作用的序列。在某些情况下,转录终止序列是在原核细胞中起作用的序列。

[0566] 可以包括在稳定性控制序列(例如转录终止区段,或在指导RNA的任何区段中以提供增加的稳定性)中的核苷酸序列包括,例如,不依赖ρ因子的trp终止位点。

[0567] *额外的序列*

[0568] 在某些实施方案中,指导RNA在5'或3'端包含至少一个额外的区段。例如,合适的额外的区段可以包括5'帽(例如7-甲基鸟苷酸帽(m7G));3'聚腺苷酸化尾(即,3'聚腺苷酸尾);核糖开关序列(例如以允许蛋白和蛋白复合物调节的稳定性和/或调节的可接近性);形成dsRNA双链体(即发夹)的序列);将RNA靶向亚细胞位置(例如细胞核、线粒体、叶绿体等)的序列;提供追踪的修饰或序列(例如直接缀合至荧光分子,缀合至促进荧光检测的部分,允许荧光检测的序列等);提供蛋白的结合位点的修饰或序列(例如作用于DNA的蛋白,包括转录激活因子、转录阻遏蛋白、DNA甲基转移酶、DNA脱甲基酶、组蛋白乙酰转移酶、组蛋白脱乙酰酶等)提供增加的、降低的和/或可控制的稳定性的修饰或序列;及其组合。

[0569] *多种同时的指导RNA*

[0570] 在某些实施方案中,在相同细胞中同时使用多种指导RNA以同时调节相同靶DNA或不同靶DNA上的不同位置的转录。在某些实施方案中,两种或更多种指导RNA靶向相同的基因或转录物或基因座。在某些实施方案中,两种或更多种指导RNA靶向不同的不相关基因座。在某些实施方案中,两种或更多种指导RNA靶向不同但相关的基因座。

[0571] 因为指导RNA小且坚固,所以它们可以同时存在于相同的表达载体上,且甚至可以在相同的转录控制之下,如果期望这样的话。在某些实施方案中,两种或更多种(例如3种或更多种、4种或更多种、5种或更多种、10种或更多种、15种或更多种、20种或更多种、25种或更多种、30种或更多种、35种或更多种、40种或更多种、45种或更多种或50种或更多种)指导RNA在靶细胞中同时表达(从相同或不同的载体/从相同或不同的启动子)。在某些情况下,可以在模拟靶向子(targeter)RNA的天然存在的CRISPR阵列的阵列中编码多种指导RNA。靶向区段作为约30个核苷酸长的序列编码(可以为约16到约100 nt),并被CRISPR重复序列隔开。可以通过编码RNA的DNA或作为RNA将阵列引入细胞。

[0572] 为了表达多种指导RNA,可以使用由Csy4内切核糖核酸酶介导的人工RNA加工系统。例如,多种指导RNA可以在前体转录物(例如,从U6启动子表达)上连接起来成为串联阵列,并被Csy4-特异性RNA序列隔开。共表达的Csy4蛋白将前体转录物切割成多种指导RNA。使用RNA加工系统的优点包括:首先,无需使用多种启动子;其次,因为所有的指导RNA都是从前体转录物中加工的,所以它们的浓度针对相似的dsRGN-结合进行了归一化。

[0573] Csy4是衍生自细菌铜绿假单胞菌(Pseudomonas aeruginosa)的小内切核糖核酸酶(RNA酶)蛋白。Csy4特异性识别最小的17-bp RNA发夹,并显示快速(<1分钟)和高效(>99.9%)的RNA切割。与大多数RNA酶不同,切割的RNA片段保持稳定并具有功能活性。基于Csy4的RNA切割可重新应用于人工RNA加工系统。在这个系统中,将17-bp RNA发夹插入作为前体转录物从单个启动子转录的多个RNA片段之间。Csy4的共表达在产生个别RNA片段中有效。

[0574] *变体sRGN定位多肽*

[0575] 如上所述,指导RNA和变体sRGN定位多肽形成复合物。指导RNA通过包含与靶DNA的序列互补的核苷酸序列,为复合物提供靶特异性。变体sRGN定位多肽具有降低的内切脱氧核糖核酸酶活性。例如,适合于供本公开内容的转录调节方法之用的变体sRGN定位多肽表

现出小于约20%、小于约15%、小于约10%、小于约5%、小于约1%或小于约0.1%的参照sRGN多肽(如包含SEQ ID NO: 1-12中的任一个所示的氨基酸序列的参照sRGN多肽)的内切脱氧核糖核酸酶活性。在某些实施方案中,变体sRGN定位多肽基本上没有可检测的内切脱氧核糖核酸酶活性(dsRGN)。在某些实施方案中,当变体sRGN定位多肽具有降低的催化活性时,所述多肽仍可以以位点特异性方式结合靶DNA(因为它仍被指导RNA指导至靶DNA序列),只要它保留与指导RNA相互作用的能力。在某些情况下,合适的变体sRGN定位多肽包含与SEQ ID NO: 1-12中的任一个具有至少75%、至少80%、至少85%、至少90%、至少95%、至少99%或更大氨基酸序列同一性的氨基酸序列。

[0576] 在某些情况下,变体sRGN定位多肽是切口酶,其可以切割靶DNA的互补链,但具有降低的切割靶DNA的非互补链的能力。

[0577] 在某些情况下,在切口酶中的变体sRGN定位多肽可以切割靶DNA的非互补链,但具有降低的切割靶DNA的互补链的能力。

[0578] 在某些情况下,变体sRGN定位多肽具有降低的切割靶DNA的互补链和非互补链两者的能力。例如,预期了丙氨酸取代。

[0579] 在某些情况下,变体sRGN定位多肽是融合多肽("变体sRGN融合多肽"),即,融合多肽,其包含:i)变体sRGN定位多肽;和ii)共价连接的异源多肽(也称为"融合配偶体")。

[0580] 异源多肽可以表现出也将由变体sRGN融合多肽表现出的活性(例如,酶活性)(例如,甲基转移酶活性、乙酰转移酶活性、激酶活性、泛素化活性等)。异源核酸序列可以与另一核酸序列连接(例如通过遗传工程)以产生编码嵌合多肽的嵌合核苷酸序列。在某些实施方案中,通过将变体sRGN多肽与提供亚细胞定位的异源序列融合来产生变体sRGN融合多肽(即,异源序列是亚细胞定位序列,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)。在某些实施方案中,异源序列可以提供用于易于追踪和/或纯化的标签(即,异源序列是可检测的标记)(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)。在某些实施方案中,异源序列可提供增加或降低的稳定性(即,异源序列是稳定性控制肽,例如降解决定子,其在某些情况下是可控制的(例如温度敏感或药物可控制的降解决定子序列,参见下文)。在某些实施方案中,异源序列可以提供从靶DNA的增加或减少的转录(即,异源序列是转录调节序列,例如转录因子/激活因子或其片段、募集转录因子/激活因子的蛋白或其片段、转录阻遏蛋白或其片段、募集转录阻遏蛋白的蛋白或其片段、小分子/药物应答转录调节物等)。在某些实施方案中,异源序列可以提供结合结构域(即,异源序列是蛋白结合序列,例如以提供嵌合的dsRGN多肽结合另一种感兴趣的蛋白的能力,例如DNA或组蛋白修饰蛋白、转录因子或转录阻遏蛋白、募集蛋白等)。

[0581] 提供增加或降低的稳定性的合适融合配偶体包括、但不限于降解决定子序列。本领域普通技术人员理解降解决定子为控制它们是其一部分的蛋白的稳定性的氨基酸序列。例如,包含降解决定子序列的蛋白的稳定性至少部分地由降解决定子序列控制。在某些情况下,合适的降解决定子是组成型的,以使降解决定子可以不依赖于实验控制而对蛋白稳定性发挥其影响(即,降解决定子不是药物可诱导的、温度可诱导的等)。在某些情况下,降解决定子为变体sRGN多肽提供了可控制的稳定性,从而使得变体sRGN多肽可以依赖于所期

望的条件“打开”(即,稳定的)或“关闭”(即,不稳定的,降解的)。例如,如果降解决定子是温度敏感的降解决定子,则变体sRGN多肽可以在低于阈值温度(例如42℃、41℃、40℃、39℃、38℃、37℃、36℃、35℃、34℃、33℃、32℃、31℃、30℃等)时有功能(即,“打开”,稳定的),但在高于阈值温度时无功能(即,“关闭”,降解的)。作为另一个例子,如果降解决定子是药物可诱导的降解决定子,则药物的存在或不存在可将蛋白从“关闭”(即,不稳定的)状态切换到“打开”(即,稳定的)状态,或反之亦然。示例性的药物可诱导的降解决定子衍生自FKBP12蛋白。降解决定子的稳定性由与降解决定子结合的小分子的存在或不存在来控制。

[0582] 合适的降解决定子的例子包括、但不限于由Shield-1、DFHR、生长素和/或温度控制的那些降解决定子。合适的降解决定子的非限制性例子是本领域已知的(例如Dohmen等人, Science, 1994.263(5151): 第1273-1276页: Heat-inducible degron: a method for constructing temperature-sensitive mutants;Schoeber等人, Am J Physiol Renal Physiol.2009年1月;296(1):F204-11;Conditional fast expression and function of multimeric TRPV5 channels using Shield-1;Chu等人, Bioorg Med Chem Lett.2008年11月15日;18(22):5941-4: Recent progress with FKBP-derived destabilizing domains;Kanemaki, Pflugers Arch.2012年12月28日: Frontiers of protein expression control with conditional degrons;Yang等人, Mol Cell.2012年11月30日;48(4):487-8: Titivated for destruction: the methyl degron;Barbour等人, Biosci Rep.2013年1月18日;33(1).: Characterization of the bipartite degron that regulates ubiquitin-independent degradation ofthymidylate synthase;和Greussing等人, J Vis Exp.2012年11月10日;(69): Monitoring of ubiquitin-proteasome activity in living cells using a Degron (dgn)-destabilized green fluorescent protein (GFP)-based reporter protein;其全部特此整体引入作为参考)。

[0583] 示例性的降解决定子序列已经在细胞和动物两者中得到充分表征和测试。因此,将sRGN融合至降解决定子序列产生“可调的”和“可诱导的”sRGN多肽。本文所述的任何融合配偶体都可以以任何期望的组合使用。作为举例说明这一点的一个非限制性例子,sRGN融合蛋白可包含用于检测的YFP序列、用于稳定性的降解决定子序列和增加靶DNA转录的转录激活因子序列。此外,可以在sRGN融合蛋白中使用的融合配偶体的数目是没有限制的。在某些情况下,sRGN融合蛋白包含一个或多个(例如两个或更多个、三个或更多个、四个或更多个或五个或更多个)异源序列。

[0584] 合适的融合配偶体包括、但不限于,提供甲基转移酶活性、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、脱泛素活性、腺苷化活性、脱腺苷化活性、苏素化活性、脱苏素化活性、核糖基化活性、脱核糖基化活性、巴豆酰化活性、脱巴豆酰化活性、丙酰化活性、脱丙酰化活性、豆蔻酰化活性或脱豆蔻酰化活性的多肽,它们中的任何一种都可以旨在直接修饰DNA(例如DNA的甲基化作用)或修饰DNA-相关多肽(例如组蛋白或DNA结合蛋白)。进一步合适的融合配偶体包括、但不限于边界元件(例如CTCF)、提供外周募集的蛋白及其片段(例如核纤层蛋白A、核纤层蛋白B等)和蛋白质停靠元件(例如FKBP/FRB、Pil 1/Aby 1等)。

[0585] 在某些实施方案中,可以对sRGN多肽进行密码子优化。这种类型的优化是本领域已知的,并且需要外源DNA的突变以在编码相同蛋白的同时模拟预期宿主生物或细胞的密

码子偏好。因此,密码子被改变,但是编码的蛋白保持不变。例如,如果预期的靶细胞是人细胞,则人密码子优化的dsRGN(或dsRGN变体)将是合适的定位修饰多肽。作为另一个非限制性例子,如果预期的宿主细胞是小鼠细胞,则小鼠密码子优化的sRGN(或变体,例如酶无活性变体)将是合适的sRGN定位多肽。尽管不需要密码子优化,但它是可以接受的,且在某些情况下可能是优选的。

[0586] 还可以选择聚腺苷酸化信号以最优化在预期宿主中的表达。

[0587] *宿主细胞*

[0588] 本公开内容的调节转录的方法可以用于在体内和/或离体和/或体外在有丝分裂或有丝分裂后细胞中诱导转录调节。因为指导RNA通过与靶DNA杂交提供特异性,所以有丝分裂和/或有丝分裂后细胞可以是多种宿主细胞中的任何一种,其中合适的宿主细胞包括、但不限于,细菌细胞;古细菌细胞;单细胞真核生物;植物细胞;藻类细胞,例如丛粒藻(Botryococcus braunii)、雷氏衣藻(Chlamydomonas reinhardtii)、微拟球藻(Nannochloropsis gaditana)、蛋白核小球藻(Chlorella pyrenoidosa)、展枝马尾藻(Sargassum patens,C. Agardh)等;真菌细胞;动物细胞;来自无脊椎动物(例如昆虫、刺胞动物、棘皮动物、线虫等)的细胞;真核寄生物(例如疟原虫,例如恶性疟虫(Plasmodium fakiparum);蠕虫等);来自脊椎动物(例如鱼、两栖动物、爬行动物、鸟、哺乳动物)的细胞;哺乳动物细胞,例如啮齿类动物细胞,人细胞,非人灵长类动物细胞等。合适的宿主细胞包括天然存在的细胞;遗传修饰的细胞(例如在实验室中例如通过“人的手”进行遗传修饰的的细胞);以及以任何方式在体外操作的细胞。在某些情况下,宿主细胞是分离的。

[0589] 任何类型的细胞都可能是有意义的(例如干细胞,例如胚胎干(ES)细胞、诱导的多潜能干(iPS)细胞、生殖细胞;体细胞,例如成纤维细胞、造血细胞、神经元、肌细胞、骨细胞、肝细胞、胰细胞;处于任何阶段的胚胎的体外或体内胚胎细胞,例如1-细胞、2-细胞、4-细胞、8-细胞等阶段斑马鱼胚胎;等)。细胞可以来自确立的细胞系,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,以指已经衍生自受试者并被允许在体外生长有限数目的培养物传代即分裂的细胞和细胞培养物。例如,原代培养物包括可能已经传代0次、1次、2次、4次、5次、10次或15次但未传代通过转捩阶段的足够次数的培养物。原代细胞系可在体外维持少于10次传代。靶细胞在许多实施方案中是单细胞生物,或在培养物中生长。

[0590] 如果细胞是原代细胞,则可以通过任何方便的方法从个体中收获这种细胞。例如,白细胞可通过单采血液成分术、白细胞除去法、密度梯度分离等方便地收获,而来自诸如皮肤、肌肉、骨髓、脾、肝、胰、肺、肠、胃等的组织的细胞最方便通过活组织检查收获。可以使用适当的溶液分散或悬浮收获的细胞。这种溶液通常将是平衡盐溶液,例如生理盐水、磷酸缓冲盐水(PBS)、Hank氏平衡盐溶液等,方便地补加了胎牛血清或其它天然存在的因子,连带着可接受的低浓度缓冲液,例如5-25 mM。方便的缓冲液包括HEPES、磷酸盐缓冲液、乳酸盐缓冲液等。细胞可以立即使用,或者它们可以冷冻贮藏长的时间段,被解冻并能够再使用。在这种情况下,通常将细胞在10%二甲基亚砜(DMSO),50%血清,40%缓冲的培养基或如本领域中所通常使用的一些其它这种溶液中冷冻,以在这种冷冻温度保存细胞,并以如本领域中所通常已知的方式解冻,以用于解冻冷冻的培养的细胞。

[0591] *将核酸引入宿主细胞*

[0592] 可以通过多种众所周知的方法中的任一种将指导RNA或包含编码其的核苷酸序列的核酸引入宿主细胞。类似地,在方法包括将包含编码变体sRGN定位多肽的核苷酸序列的核酸引入宿主细胞的场合,可以通过多种众所周知的方法中的任一种将这种核酸引入宿主细胞。

[0593] 将核酸引入宿主细胞的方法是本领域已知的,并且可以使用任何已知方法将核酸(例如表达构建体)引入干细胞或祖先细胞。合适的方法包括,例如病毒或噬菌体感染、转染、接合、原生质体融合、脂质转染法、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、基因枪技术、磷酸钙沉淀、直接显微注射、纳米颗粒介导的核酸递送(例如Panyam等人, Adv Drug Deliv Rev.2012年9月13日.pii: 50169-409X(12)00283-9.doi: 10.1016/j.addr.2012.09.023)等,包括、但不限于外来体递送。

[0594] *核酸*

[0595] 本公开内容提供了分离的核酸,其包含编码指导RNA的核苷酸序列。在某些情况下,核酸还包含编码变体sRGN定位多肽的核苷酸序列。

[0596] 在某些实施方案中,方法包括将一种或多种包含编码指导RNA和/或变体sRGN定位多肽的核苷酸序列的核酸引入宿主细胞(或宿主细胞群体)。在某些实施方案中,包含靶DNA的细胞是体外的。在某些实施方案中,包含靶DNA的细胞是体内的。包含编码指导RNA和/或变体sRGN定位多肽的核苷酸序列的合适核酸包括表达载体,其中表达载体包含编码指导的核苷酸序列。

[0597] *表达载体*

[0598] 在某些实施方案中,重组表达载体是病毒构建体,例如重组腺相关病毒构建体(例如美国专利号7,078,387)、重组腺病毒构建体、重组慢病毒构建体、重组反转录病毒构建体等。合适的表达载体包括、但不限于,病毒载体(例如基于痘苗病毒的病毒载体;脊髓灰质炎病毒;腺病毒(例如Li等人, Invest Opthalmol Vis Sci 35:2543 2549, 1994;Borras等人, Gene Ther 6:515 524, 1999;Li和Davidson, PNAS 92:7700 7704, 1995;Sakamoto等人, H Gene Ther 5:10881097, 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(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在WO 93/09239中,Samulski等人, J.Vir.(1989) 63:3822-3828;Mendelson等人, Viral.(1988) 166:154-165;和Flotte等人, PNAS (1993) 90:10613-10617);SV40;单纯疱疹病毒;人免疫缺陷病毒(例如Miyoshi等人, PNAS 94:10319 23, 1997;Takahashi等人, J Virol 73:7812 7816, 1999);反转录病毒载体(例如鼠白血病病毒,脾脏坏死病毒,以及衍生自反转录病毒的载体,例如劳斯肉瘤病毒,Harvey肉瘤病毒,禽类白血病病毒,慢病毒,人免疫缺陷病毒,骨髓增生肉瘤病毒和乳腺肿瘤病毒);等等。

[0599] 许多合适的表达载体是本领域技术人员已知的,并且许多是可商购获得的。作为实例提供下述载体;对于真核宿主细胞:pXT1、pSG5(Stratagene)、pSVK3、pBPV、pMSG和pSVLSV40(Pharmacia)。但是,可以使用任何其它载体,只要它与宿主细胞相容即可。

[0600] 取决于所使用的宿主/载体系统,可以在表达载体中使用许多合适的转录和翻译控制元件中的任何一种,包括组成型和诱导型启动子、转录增强子元件、转录终止子等(参见,例如Bitter等人(1987) Methods in Enzymology, 153:516-544)。

[0601] 在某些实施方案中,编码指导RNA和/或变体sRGN定位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,例如启动子。转录控制元件可以在真核细胞,例如哺乳动物细胞;或原核细胞(例如细菌或古细菌细胞)中起作用。在某些实施方案中,编码指导RNA和/或变体sRGN定位多肽的核苷酸序列可操作地连接至多个控制元件,其允许在原核和真核细胞两者中表达编码指导RNA和/或变体sRGN定位多肽的核苷酸序列。

[0602] 启动子可以是组成型活性启动子(即,组成型地处于活性“打开”状态的启动子),其也可以是诱导型启动子(即,其状态活性/“打开”或无活性/“关闭”由外界刺激例如特定温度、化合物或蛋白的存在控制的启动子),其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),并且其可以是时间限制的启动子(即,在胚胎发育的特定阶段过程中或生物过程的特定阶段(例如小鼠中的毛囊循环)过程中处于“打开”状态或“关闭”状态的启动子)。

[0603] 合适的启动子可以衍生自病毒,且因此可以被称为病毒启动子,或者它们可以衍生自任何生物,包括原核或真核生物。合适的启动子可用于通过任何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)启动子例如CMV立即早期启动子区域(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 Sep 1;31(17)),人H1启动子(H1)等。

[0604] 诱导型启动子的例子包括、但不限于T7 RNA聚合酶启动子、T3 RNA聚合酶启动子、异丙基-β-D-硫代吡喃半乳糖苷(IPTG)调节的启动子、乳糖诱导的启动子、热激启动子、四环素调节的启动子(例如Tet-ON、Tet-OFF等)、类固醇调节的启动子、金属调节的启动子、雌激素受体调节的启动子等。因此,诱导型启动子可以被包括、但不限于下述的分子调节:多西环素;RNA聚合酶,例如T7 RNA聚合酶;雌激素受体;雌激素受体融合体;等。

[0605] 在某些实施方案中,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),从而使得在多细胞生物中,所述启动子在特定细胞亚群中是有活性的(即,“打开”)。空间限制的启动子也可以称为增强子、转录控制元件、控制序列等。可以使用任何方便的空间限制的启动子,且合适的启动子(例如脑特异性启动子、驱动神经元亚群中表达的启动子、驱动种系中表达的启动子、驱动肺中表达的启动子、驱动肌肉中表达的启动子、驱动胰的胰岛细胞中表达的启动子等)的选择将取决于生物。例如,已知对于植物、蝇、蠕虫、哺乳动物、小鼠等的各种空间限制的启动子。因此,取决于生物,空间限制的启动子可用于调节在各种各样的不同组织和细胞类型中编码变体sRGN定位多肽的核酸的表达。一些空间限制的启动子也是时间限制的,从而使得所述启动子在胚胎发育的特定阶段过程中或生物过程的特定阶段(例如小鼠中的毛囊循环)过程中处于“打开”状态或“关闭”状态。

[0606] 为了举例说明起见,空间限制的启动子的例子包括、但不限于,神经元特异性启动子、脂肪细胞特异性启动子、心肌细胞特异性启动子、平滑肌特异性启动子、光感受器特异

性启动子等。神经元特异性的空间限制的启动子包括、但不限于,神经元特异性烯醇化酶(NSE)启动子(例如EMBL HSENO2,X51956);芳族氨基酸脱羧酶(AADC)启动子;神经丝启动子(例如GenBank HUMNFL,L04147);突触蛋白启动子(例如GenBank HUMSYNIB,M55301);thy-1启动子(例如Chen等人(1987) Ce/151:7-19;和Llewellyn等人(2010) Nat.Med.16(10):1161-1166);5-羟色胺受体启动子(例如GenBank S62283);酪氨酸羟化酶启动子(TH)(例如Oh等人(2009) Gene Ther 16:437;Sasaoka等人(1992) Mol.Brain Res.16:274;Boundy等人(1998) J.Neurosci.18:9989;和Kaneda等人(1991) Neuron 6:583-594);GnRH启动子(例如Radovick等人(1991) Proc.Nat!.Acad.Sci.USA 88:3402-3406);L7启动子(例如Oberdick等人(1990) Science 248:223-226);DNMT启动子(例如Bartge等人(1988) Proc.Nat/.Acad.Sci.USA 85:3648- 3652);脑啡肽启动子(例如Comb等人(1988) EMBO J.17:3793-3805);髓鞘碱性蛋白(MBP)启动子;Ca2+-钙调蛋白依赖性蛋白激酶11-α(CamKIIa)启动子(例如Mayford等人(1996) Proc.Nat/.Acad.Sci.USA 93:13250;和Casanova等人(2001) Genesis 31:37);CMV增强子/血小板衍生生长因子-0启动子(例如Liu等人(2004) Gene Therapy 11:52-60);等等。

[0607] 脂肪细胞特异性的空间限制的启动子包括、但不限于aP2基因启动子/增强子,例如人aP2基因从-5.4 kb到+21 bp的区域(例如,Tozzo等人(1997) Endocrinol. 138:1604;Ross等人(1990) Proc. Nat!. Acad. Sci. USA 87:9590;和Pavjani等人(2005) Nat. Med. 11:797);葡萄糖转运蛋白-4(GLUT4)启动子(例如,Knight等人(2003) Proc. Nat/. Acad. Sci. USA 100:14725);脂肪酸移位酶(FAT/CD36)启动子(例如,Kuriki等人(2002) Biol. Pharm. Bull. 25:1476;和Sato等人(2002) J. Biol. Chem. 277:15703);硬脂酰辅酶A去饱和酶-1 (SCD1)启动子(Tabor等人(1999) J. Biol. Chem. 274:20603);瘦蛋白启动子(例如,Mason等人(1998) Endocrinol. 139:1013;和Chen等人(1999) Biochem. Biophys. Res. Comm. 262:187);脂联素启动子(例如,Kita等人(2005) Biochem. Biophys. Res. Comm. 331:484;和Chakrabarti (2010) Endocrinol. 151:2408);降脂蛋白启动子(例如,Platt等人(1989) Proc. Natl. Acad. Sci. USA 86:7490);抵抗素启动子(例如,Seo等人(2003) Malec. Endocrinol. 17:1522);等等。

[0608] 心肌细胞特异性的空间限制的启动子包括、但不限于衍生自下述基因的控制序列:肌球蛋白轻链-2、a-肌球蛋白重链、AE3、心肌钙蛋白C、心肌动蛋白等。Franz等人(1997) Cardiovasc.Res.35:560-566;Robbins等人(1995) Ann.N.Y.Acad.Sci.752:492-505;Linn等人(1995) Circ.Res.76:584591;Parmacek等人(1994) Mol.Cell.Biol.14:1870-1885;Hunter等人(1993) Hypertension 22:608-617;和Sartorelli等人(1992) Proc.Natl.Acad.Sci.USA 89:4047-4051。

[0609] 平滑肌特异性的空间限制的启动子包括、但不限于SM22a启动子(例如Akyilrek等人(2000) Mol.Med.6:983;和美国专利号7,169,874);平滑肌特异性蛋白启动子(例如WO 2001/018048);a-平滑肌肌动蛋白启动子;等等。例如,已经证实两个CArG元件位于其中的SM22a启动子的0.4 kb区域介导血管平滑肌细胞特异性表达(例如Kim等人(1997) Mol.Cell.Biol.17, 2266-2278;Li等人, (1996) J.Cell Biol.132, 849-859;和Moessler等人(1996) Development 122, 2415-2425)。

[0610] 光感受器特异性的空间限制的启动子包括、但不限于,视紫红质启动子;视紫红质

激酶启动子(Young等人(2003) Ophthalmol.Vis.Sci.44:4076);β磷酸二酯酶基因启动子(Nicoud等人(2007) J.Gene Med.9:1015);色素性视网膜炎基因启动子(Nicoud等人(2007)上文);光感受器间类视黄醇结合蛋白(IRBP)基因增强子(Nicoud等人(2007)上文);IRBP基因启动子(Yokoyama等人(1992) Exp Eye Res.55:225);等等。

[0611] *用途*

[0612] 根据本公开内容的用于调节转录的方法应用于也被提供的多种应用中。应用包括研究应用;诊断应用;工业应用;和治疗应用。

[0613] 研究应用包括,例如确定减少或增加靶核酸的转录对例如发育、代谢、下游基因的表达等的作用。可以使用转录调节方法进行高通量的基因组分析,其中仅指导RNA的DNA-靶向区段需要改变,而蛋白结合区段和转录终止区段可以(在某些情况下)保持不变。包含在基因组分析中使用的许多核酸的文库(例如文库)将包括:与编码指导RNA的核苷酸序列可操作地连接的启动子,其中每种核酸将包括共同的蛋白结合区段,不同的DNA-靶向区段和共同的转录终止区段。芯片可含有超过5 x 10^4个独特的指导RNA。应用将包括大规模的表型分析、基因到功能作图以及宏基因组分析(meta-genomic analysis)。

[0614] 本文公开的方法应用于代谢工程领域。如本文所公开的,因为可以通过设计合适的指导RNA来有效地和可预测地控制转录水平,所以可以通过控制感兴趣的代谢途径中特定酶的水平(例如通过增加或减少的转录)来精确地控制和调节代谢途径(例如生物合成途径)的活性。感兴趣的代谢途径包括用于化学(精细化学品、燃料、抗生素、毒素、激动剂、拮抗剂等)和/或药物生产的那些。

[0615] 感兴趣的生物合成途径包括、但不限于(1)甲羟戊酸途径(例如β-羟基-β-甲基戊二酸单酰辅酶A还原酶途径)(将乙酰辅酶A转化为二甲基烯丙基焦磷酸(DMAPP)和异戊烯焦磷酸(IPP),它们用于包括类萜/类异戊二烯在内的各种各样生物分子的生物合成),(2)非甲羟戊酸途径(即,"2-C-甲基-D-赤藓糖醇4-磷酸/1-脱氧-D-木酮糖5-磷酸途径"或"MEP/DOXP途径"或"DXP途径")(也产生DMAPP和IPP,而通过经由甲羟戊酸途径的替换途径将丙酮酸和甘油醛3-磷酸转化为DMAPP和IPP),(3)聚酮化合物合成途径(通过多种聚酮化合物合酶生产多种聚酮化合物。聚酮化合物包括用于化学疗法的天然存在的小分子(例如四环素和大环内酯),且工业上重要的聚酮化合物包括雷帕霉素(免疫抑制剂)、红霉素(抗生素)、抑甲羟酶素(抗胆固醇药)和埃博霉素B(抗癌药)),(4)脂肪酸合成途径,(5)DAHP(3-脱氧-D-阿拉伯庚酮糖酸7-磷酸)合成途径,(6)产生潜在生物燃料(例如短链醇和链烷,脂肪酸甲酯和脂肪醇,类异戊二烯等)的途径,等。

[0616] *网络和级联*

[0617] 本文公开的方法可以用于设计控制的综合网络(即,一种或多种级联)。例如,指导RNA I变体sRGN定位多肽可用于控制(即,调节,例如增加、减少)另一种DNA-靶向RNA或另一种变体sRGN定位多肽的表达。例如,第一指导RNA可以被进行设计以靶向具有与第一变体sRGN定位多肽不同的功能(例如甲基转移酶活性、脱甲基酶活性、乙酰转移酶活性、脱乙酰酶活性等)的第二嵌合dsRGN多肽的转录的调节。此外,因为不同的dsRGN蛋白(例如衍生自不同物种)可能需要不同的sRGN柄(即,蛋白结合区段),所以第二嵌合dsRGN多肽可以衍生自与上述第一dsRGN多肽不同的物种。因此,在某些情况下,可以选择第二嵌合dsRGN多肽,从而使得其可以不与第一指导RNA相互作用。在其它情况下,可以选择第二嵌合dsRGN多肽,

从而使得其与第一指导RNA相互作用。在一些这样的情况下,两种(或更多种)dsRGN蛋白的活性可以竞争(例如,如果多肽具有相反的活性)或可以协同(例如,如果多肽具有相似或协同的活性)。同样,如上文所指出的,可以对网络中的任何复合物(即,指导RNA I dsRGN多肽)进行设计以控制其它指导RNA或dsRGN多肽。因为指导RNA和变体sRGN定位多肽可以靶向任何所期望的DNA序列,所以本文所述的方法可以用于控制和调节任何所期望的靶的表达。可以进行设计的综合网络(即,相互作用的级联)从非常简单的一直到非常复杂的,并且没有限制。

[0618] 在其中两种或更多种组分(例如指导RNA或dsRGN多肽)各自处于另一指导RNA/dsRGN多肽复合物的调节控制下的网络中,网络的一种组分的表达水平可能影响网络的另一种组分的表达水平(例如可能增加或减少表达)。通过这种机制,一种组分的表达可能影响相同网络中不同组分的表达,并且所述网络可能包括增加其它组分表达的组分以及减少其它组分表达的组分的混合物。如本领域的技术人员将理解的,上述一种组分的表达水平可能影响一种或多种不同组分的表达水平的例子是为了举例说明起见的,而不是限制性的。当将一种或多种组分进行修饰(如上所述)以使其可操作时(即,在实验控制下,例如温度控制;药物控制,即,药物可诱导的控制;光控制;等等),可以任选地将额外层次的复杂性引入网络。

[0619] 作为一个非限制性例子,第一指导RNA可以结合第二指导RNA的启动子,其控制靶治疗/代谢基因的表达。在这种情况下,第一指导RNA的条件性表达间接激活治疗/代谢基因。这种类型的RNA级联可用于例如将阻遏蛋白转化为激活物,并可用于控制靶基因表达的逻辑或动力学。

[0620] 转录调节方法也可用于药物发现和靶标确认。

[0621] 本发明的各个方面利用下述材料和方法,并通过下述非限制性实施例进行举例说明。

[0622] 基因组编辑

[0623] 基因组编辑通常是指修饰基因组的核苷酸序列的过程,例如以精确或预定的方式。本文描述的基因组编辑方法的例子包括使用定位核酸酶以在基因组中精确的靶位置切割DNA,从而在基因组内特定位置产生双链或单链DNA断裂的方法。这种断裂可以并经常通过自然的、内源性细胞过程例如同源性指导的修复(HDR)和非同源末端连接(NHEJ)进行修复,如最近在Cox等人, Nature Medicine 21(2), 121-31 (2015)中所综述的。NHEJ有时在丧失或添加核苷酸序列的情况下直接连接由双链断裂产生的DNA末端,这可能破坏或增强基因表达。HDR利用同源序列或供体序列作为模板,用于在断裂点处插入确定的DNA序列。同源序列可以在内源基因组中,例如姐妹染色单体。另一方面,供体可以是外源核酸,例如质粒、单链寡核苷酸、双链体寡核苷酸或病毒,其具有与核酸酶切割的基因座具有高同源性的区域,但其也可以含有额外的序列或序列变化,包括可以并入切割的靶基因座的缺失。第三种修复机制是微同源性介导的末端连接(MMEJ),也称为“备选NHEJ,其中遗传结果与NHEJ相似,这是因为在切割位点可发生小的缺失和插入。MMEJ利用侧接DNA断裂位点的几个碱基对的同源序列来驱动更有利的DNA末端连接修复结果,且最近的报道进一步阐明了这个过程的分子机制;参见,例如Cho和Greenberg, Nature 518, 174-76 (2015);Kent等人, Nature Structural and Molecular Biology, Adv.Online doi:10.1038/nsmb.2961

(2015);Mateos-Gomez等人, Nature 518,254-57 (2015);Ceccaldi等人, Nature 528, 258-62 (2015)。在某些情况下,可以基于对在DNA断裂位点的潜在的微同源性的分析来预测可能的修复结果。

[0624] 这些基因组编辑机制的每一种均可用于产生所期望的基因组改变。基因组编辑过程通常在尽可能接近预期突变位点的靶基因座中产生一个或两个DNA断裂。如本文所述和所举例说明的,这可以通过使用定位多肽来实现。

[0625] 定位多肽可以在核酸(例如基因组DNA)中引入双链断裂或单链断裂。双链断裂可刺激细胞的内源性DNA-修复途径(例如同源性依赖性修复(HDR)和非同源末端连接(NHEJ)或备选的非同源末端连接(A-NHEJ)或微同源性介导的末端连接(MMEJ))。NHEJ可在不需要同源模板的情况下修复切割的靶核酸。这有时可导致靶核酸在切割位点的小缺失或插入(插入/缺失),并可导致基因表达的破坏或改变。当同源修复模板或供体可获得时,可发生HDR。同源供体模板包含与侧接靶核酸切割位点的序列同源的序列。姐妹染色单体通常被细胞用作修复模板。但是,为了基因组编辑起见,修复模板经常作为外源核酸提供,例如质粒、双链体寡核苷酸、单链寡核苷酸或病毒核酸。在外源供体模板的情况下,通常在同源性的侧翼区之间引入额外的核酸序列(例如转基因)或修饰(例如单碱基改变或缺失),从而额外的或改变的核酸序列也变得并入靶基因座中。MMEJ导致与NHEJ相似的遗传结果,这是因为在切割位点可发生小的缺失和插入。MMEJ利用侧接切割位点的几个碱基对的同源序列来驱动有利的末端连接DNA修复结果。在某些情况下,有可能基于对核酸酶靶区域中潜在的微同源性的分析来预测可能的修复结果。因此,在某些情况下,同源重组用于将外源多核苷酸序列插入靶核酸切割位点。外源多核苷酸序列在本文中称为供体多核苷酸。在某些实施方案中,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插入靶核酸切割位点。在某些实施方案中,供体多核苷酸是外源多核苷酸序列,即,在靶核酸切割位点不天然存在的序列。

[0626] 归因于NHEJ和/或HDR的靶DNA的修饰可导致,例如,突变、缺失、改变、整合、基因矫正、基因置换、基因标记、转基因插入、核苷酸缺失、基因破坏、易位和/或基因突变。删除基因组DNA和将非天然核酸整合到基因组DNA中的方法是基因组编辑的例子。

实施例

[0627] 实施例1A:sRGN多肽Gib11的活性的证实

[0628] 在基于荧光偏振的生化切割测定中测试了Gib11和SluCas9。

[0629] *荧光偏振测定方案:*

[0630] 在10 mM Tris (pH = 7.8) 50 mM NaCl中作为10μM溶液(来自100μM原液)制备寡核苷酸双链体,并在循环变温器中于95℃退火5分钟,且然后缓慢冷却(每分钟6℃)。随后将原液在10 mM Tris (pH = 7.8) 50 mM NaCl + 0.05%普朗尼克(pluronic)中稀释。将20 nM寡物(20μL)固定在抗生蛋白链菌素包被的平板上,在5分钟后洗涤两次,且然后与20μL样品一起温育60分钟的动力学(激发波长:635 nm;发射波长:670 nm)。在切割之前,形成反应RNP。向817.6μl 1xPBS $MgCl_2$中添加从pCas606 (SEQ ID NO: 118)转录的核酸酶和$sgRNA_{VEGFA}$ (SEQ ID NO: 114),并于37℃温育10分钟。向每个孔中添加20μl RNP,并于37℃测量偏振达60分钟。*参见*,例如,Methods Enzymol. 2014;546:1-20。

[0631] 针对恒定水平的sgRNA滴定蛋白质的量。在温育靶向VEGFA寡核苷酸序列的Gib11-或SluCas9-RNP复合物后,通过荧光偏振和荧光强度信号的变化(成功切割后随时间推移偏振值降低和荧光强度增加)可以观察切割。作为切割反应的定量估计,分析了图的初始斜率。结果(在图2A和2B中表示)表明,Gib11和SluCas9成功切割寡核苷酸底物。切割动力学在图2C中示出。

[0632] 实施例1B:Gib11“NNGG”PAM基序的证实

[0633] 在液体培养物中和在琼脂平板上证实了Gib11的活性。

[0634] *在液体培养物中:*

[0635] 在IPTG-可诱导的Trc启动子下用编码嵌合构建体1 (交换了PI/WEDGE结构域,根据SEQ ID NO: 122的pCas889)或嵌合构建体2 (交换了PI结构域,根据SEQ ID NO: 121的pCas888)的质粒转化BL21DE3 origami细胞,该细胞携带毒性报道质粒(pCas634, SEQ ID NO: 119,编码具有VEGFA靶位点和“NNGG”PAM的阿拉伯糖-可诱导的ccdB毒性基因)和编码靶向VEGFA的sgRNA的质粒(pCas606, SEQ ID NO: 118)。转化后,将细胞悬浮液用1 ml SOC提取(rescue),并于37℃在1000 rpm摇动下温育(Eppendorf Thermomix)。将200µl用于接种LB培养基(补加了卡那霉素、氯霉素和氨苄西林),并将液体培养物在190 rpm摇动下于37℃温育过夜。将200µl这种过夜培养物用于接种5 ml LB培养基,其补加了卡那霉素和氨苄西林且具有/不具有0.5%L-阿拉伯糖,具有/不具有0.05 mM IPTG。在190 rpm摇动下于37℃温育过夜(18小时)后,测定OD_{600}。

[0636] 在图3的条形图中显示了OD_{600}测量结果,为Gib11、SluCas9和阴性对照获得OD_{600}值。

[0637] *在平板上:*

[0638] 在IPTG-可诱导的Trc启动子下用编码嵌合构建体1 (交换了PI/WEDGE结构域,根据SEQ ID NO: 122的pCas889)或嵌合构建体2 (交换了PI结构域,根据SEQ ID NO: 121的pCas888)的质粒(pCas81, SEQ ID NO: 120)转化BL21DE3 origami细胞,该细胞携带毒性报道质粒(pCas634, SEQ ID NO: 119,编码具有VEGFA靶位点和“NNGG”PAM的阿拉伯糖-可诱导的ccdB毒性基因)和编码靶向VEGFA的sgRNA的质粒(pCas606, SEQ ID NO: 118)。转化后,将细胞悬浮液用1 ml SOC培养基提取,并于37℃在1000 rpm摇动下温育(Eppendorf Thermomix)。将200µl铺板在补加了卡那霉素、氨苄西林、0.05 mM IPTG和0.5%L-阿拉伯糖的LB-琼脂平板中。于37℃温育18小时后,对集落进行计数。

[0639] 携带编码Gib11的质粒的BL21DE3 origami细胞成功地切割毒性报道质粒,并在诱导毒性报道蛋白ccdB后在温育中存活。通过在液体培养物中增加的OD_{600}值和通过在琼脂平板上的集落形成来测量存活。

[0640] 表5

[0641]

在琼脂平板上的集落形成	Gib11	SluCas9 WT	阴性对照
向SluCas9野生型的归一化生长[%]	94	100	37

[0642] 实施例1C: 在哺乳动物细胞中Gib11介导的基因组编辑的证实

[0643] 在以下测定中证明了Gib11在哺乳动物细胞中的活性:

[0644] *使用靶向BFP基因的指导的BFP破坏测定:*

[0645] 用表达Gib11和sgRNA (靶向BFP的指导RNA [SEQ ID NO: 117]和SluCas9 tracr

RNA [SEQ ID NO: 123])的质粒转染了在AAVS1基因座中携带编码蓝色荧光蛋白(BFP)的基因的HEK293T细胞,所述sgRNA靶向所述的BFP基因组序列的反向链。按照供应商的描述使用Lipofectamine 3000 (Thermo Fischer Scientific, 目录号L3000015)执行质粒的转染。在转染后7天进行BFP信号破坏的流式细胞计数分析。如下准备细胞:用1X PBS洗涤,胰蛋白酶消化,并重新悬浮在200µl FACS缓冲液(补充了2%FCS的1X PBS)中。使用BD FACS Canto® II中的V450过滤器组(Glaser等人, 2016),测定来自10,000个细胞的BFP信号。

[0646] 用编码Gib11和靶向BFP基因的sgRNA的质粒转染HEK293T细胞。观察到蓝色荧光的显著下降(定量为%BFP破坏)。与Cas9蛋白的功能相一致,其可能原因是Gib11辅助的DNA切割和随后在BFP开放读码框(ORF)处的插入/缺失从而导致BFP ORF的破坏。Gib11导致34.35%±7.15%BFP破坏。

[0647] 表6

[0648]

样品	%BFP破坏	标准偏差
阴性对照	4,1	0
SluCas9	6,35	0,65
Gib11	34,35	7,15

[0649] 实施例2A: sRGN多肽P2H12的活性的证实

[0650] *荧光偏振测定方案:*

[0651] 在10 mM Tris (pH = 7.8) 50 mM NaCl中作为10µM溶液(来自100µM原液)制备寡核苷酸双链体,并在循环变温器中于95℃退火5分钟,且然后缓慢冷却(每分钟6℃)。随后将原液在10 mM Tris (pH = 7.8) 50 mM NaCl + 0.05%普朗尼克中稀释。将20 nM寡物(20µL)固定在抗生蛋白链菌素包被的平板上,在5分钟后洗涤两次,且然后与20µL样品一起温育60分钟的动力学(激发波长: 635 nm;发射波长: 670 nm)。在切割之前,形成反应RNP。向817.6µl 1xPBS $MgCl_2$中添加从pCas606 (SEQ ID NO: 118)转录的核酸酶和$sgRNA_{VEGFA}$ (SEQ ID NO: 114),并于37℃温育10分钟。向每个孔中添加20µl RNP,并于37℃测量偏振达60分钟。*参见*,例如,Methods Enzymol. 2014;546:1-20。

[0652] 温育靶向VEGFA寡核苷酸序列的P2H12-RNP复合物后,通过图4所示的荧光偏振和荧光强度信号的显著变化(成功切割后随时间推移偏振值降低和荧光强度增加)观察切割。

[0653] 实施例2B: 在哺乳动物细胞中P2H12介导的基因组编辑的证实

[0654] 在以下测定中证明了P2H12在哺乳动物细胞中的活性:

[0655] *使用靶向BFP基因的指导的BFP破坏测定:*

[0656] 用表达P2H12和sgRNA (靶向BFP的指导RNA [SEQ ID NO: 117]和SluCas9 tracr RNA [SEQ ID NO: 123])的质粒转染了在AAVS1基因座中携带编码BFP的基因的HEK293T细胞,所述sgRNA靶向所述的BFP基因组序列的反向链。按照供应商的描述使用Lipofectamine 3000 (Thermo Fischer Scientific, 目录号L3000015)执行质粒的转染。在转染后7天进行BFP信号破坏的流式细胞计数分析。如下准备细胞:用1X PBS洗涤,胰蛋白酶消化,并重新悬浮在200µl FACS缓冲液(补充了2%FCS的1X PBS)中。使用BD FACS Canto II中的V450过滤器组(Glaser等人, 2016),测定来自10,000个细胞的BFP信号。

[0657] 用编码P2H12和靶向BFP基因的sgRNA的质粒转染HEK293T细胞。观察到蓝色荧光的显著下降(定量为%BFP破坏)。与Cas9蛋白的功能相一致,其可能原因是P2H12辅助的DNA切

割和随后在BFP开放读码框(ORF)处的插入/缺失从而导致BFP ORF的破坏。P2H12导致22.9%±4.4%BFP破坏。

[0658] 表7

[0659]

样品	%BFP破坏
阴性对照	4.1
SluCas9	6.35 ± 0.65
P2H12	22.9 ± 4.4

[0660] 实施例3: 在哺乳动物细胞中E2介导的基因组编辑的证实

[0661] 在以下测定中证明了E2在哺乳动物细胞中的活性:

[0662] *使用靶向BFP基因的指导的BFP破坏测定:*

[0663] 用表达E2和sgRNA (靶向BFP的指导RNA [SEQ ID NO: 117]和SluCas9 tracr RNA [SEQ ID NO: 123])的质粒转染了在AAVS1基因座中携带编码BFP的基因的HEK293T细胞,所述sgRNA靶向所述的BFP基因组序列的反向链。按照供应商的描述使用Lipofectamine 3000 (Thermo Fischer Scientific, 目录号L3000015)执行质粒的转染。在转染后7天进行BFP信号破坏的流式细胞计数分析。如下准备细胞:用1X PBS洗涤,胰蛋白酶消化,并重新悬浮在200μl FACS缓冲液(补充了2%FCS的1X PBS)中。使用BD FACS Canto II中的V450过滤器组(Glaser等人, 2016),测定来自10,000个细胞的BFP信号。

[0664] 在表8中示出并在图5中描绘该实验的三个重复的结果(E2、三个E2突变体、阴性对照和作为参考的SluCas9 (SEQ ID NO: 125))。这些结果表明,E2和三个突变体能够介导靶向的DNA切割。

[0665] 表8

[0666]

Cas9蛋白	%BFP破坏	标准偏差
阴性对照	4.05	0.55
SluCas9	7.4	1.9
E2	35.0	5.3
E2 +K741D+L743K	49.5	14
E2 +S670T+N675D	49.5	12
E2 +K741N+L743N	45.1	12

[0667] 实施例4: 在哺乳动物细胞中Gib11Spa介导的基因组编辑的证实

[0668] 在以下测定中证明了Gib11Spa-1、Gib11Spa-2和Gib11Spa-3在哺乳动物细胞中的活性:

[0669] *使用靶向BFP基因的指导的BFP破坏测定:*

[0670] 用表达Gib11Spa-1、Gib11Spa-2或Gib11Spa-3和sgRNA (靶向BFP的指导RNA (SEQ ID NO: 117)和SluCas9 tracr RNA (SEQ ID NO: 123))的质粒转染了在AAVS1基因座中携带编码BFP的基因的HEK293T细胞,所述sgRNA靶向所述的BFP基因组序列的反向链。按照供应商的描述使用Lipofectamine 3000 (Thermo Fischer Scientific, 目录号L3000015)

执行质粒的转染。在转染后7天进行BFP信号破坏的流式细胞计数分析。如下准备细胞:用1X PBS洗涤,胰蛋白酶消化,并重新悬浮在200μl FACS缓冲液(补充了2%FCS的1X PBS)中。使用BD FACS Canto II中的V450过滤器组(Glaser等人, 2016, Molecular Therapy Nucleic Acids 5, e334),测定来自10,000个细胞的BFP信号。

[0671] *使用Gib11Spa-1、Gib11Spa-2和Gib11Spa-3在哺乳动物细胞中的编辑*

[0672] 在用编码Gib11Spa-1、Gib11Spa-2或Gib11Spa-3和靶向BFP基因的sgRNA的质粒转染HEK293T细胞后,观察到蓝色荧光的显著下降(定量为%BFP破坏)。与Cas9蛋白的功能相一致,其可能原因是Gib11Spa辅助的DNA切割和随后在BFP开放读码框(ORF)处的插入/缺失和从而导致的BFP ORF破坏。结果显示在表9中和图7的条形图中。

[0673] 表9

[0674]

Cas9	%BFP破坏	标准偏差.
Gib11	19.7	3.4
Gib11+Spa PID 3	51.7	-
Gib11+Spa PID 2	44.0	5.4
Gib11+Spa PID 1	54.2	0.5
阴性对照	4.4	1.1
SluCas9	6.3	0.7

[0675] 实施例5: 在哺乳动物细胞中F8介导的基因组编辑的证实

[0676] 在以下测定中证明了F8在哺乳动物细胞中的活性:

[0677] *使用靶向BFP基因的指导的BFP破坏测定:*

[0678] 用表达F8和sgRNA (靶向BFP的指导RNA [SEQ ID NO: 117]和SluCas9 tracr RNA [SEQ ID NO: 123])的质粒转染了在AAVS1基因座中携带编码BFP的基因的HEK293T细胞,所述sgRNA靶向所述的BFP基因组序列的反向链。按照供应商的描述使用Lipofectamine 3000 (Thermo Fischer Scientific, 目录号L3000015)执行质粒的转染。在转染后7天进行BFP信号破坏的流式细胞计数分析。如下准备细胞:用1X PBS洗涤,胰蛋白酶消化,并重新悬浮在200μl FACS缓冲液(补充了2%FCS的1X PBS)中。使用BD FACS Canto II中的V450过滤器组(Glaser等人, 2016),测定来自10,000个细胞的BFP信号。

[0679] 在表10中示出并在图8中描绘该实验的三个重复的结果(F8、两个F8突变体、阴性对照和作为参考的SluCas9 (SEQ ID NO: 125))。这些结果表明,F8和两个突变体能够介导靶向的DNA切割。

[0680] 表10

[0681]

Cas9	%BFP破坏
阴性对照	5.2
SluCas9	2.6
F8	64
F8 +K737D+L739K	61.9
F8 +K737N+L739N	61.3

[0682] 实施例6: sRGN的生化特异性谱的确定

[0683] 为了评估Gib11、Gib11Spa-1、Gib11Spa-3、E2和F8的特异性谱,确定了中靶序列和

所有60-衍生的单核苷酸错配的dsDNA底物(根据SEQ ID NO: 129-250成对使用,例如,分别具有SEQ ID NO: 129和130的R01-1-A和R01-1-B代表一个双链体对)的生化切割。在10 mM Tris (pH 7.8) 50 mM NaCl中作为10μM溶液(来自100μM原液)制备寡核苷酸双链体,并在循环变温器中于95℃退火5分钟,且然后缓慢冷却(每分钟6℃)。随后将原液在10 mM Tris (pH 7.8)、50 mM NaCl和0.05%Pluronic中稀释。将20μl每种寡核苷酸(20 nM)固定在抗生蛋白链菌素包被的平板上,在10分钟后洗涤两次,且然后与20μL样品一起温育60分钟的动力学(激发波长: 635 nm;发射波长: 670 nm)。在切割之前,形成反应RNP。通过在有和没有sgRNA (SEQ ID NO: 128;60 nM终浓度)存在下在3200μl 1xPBS + 5mM MgCl2中混合12.03μl各种sRGN蛋白(1.21μg/μl)来组装RNP,在反应之前将RNP在37℃温育5 min。将20μl RNP加入每个孔,并在37℃测量偏振60分钟。

[0684] 通过计算寡核苷酸切割反应的初始斜率来分析切割动力学。为61种底物中的每一种计算斜率,并将其归一化为中靶底物的值(定义为1)。然后将所有60种脱靶底物的归一化切割值根据其位置分组。对于靶序列中20个核苷酸位置中的每一个,将三个单核苷酸错配的归一化切割值进行印迹,以说明所测试核酸酶的位置特异性核苷酸耐受性。

[0685] 结果显示在图9中。使用此指导/靶标组合,E2的特异性谱与SluCas9的特异性谱相当。Gib11、Gib11Spa-1和Gib11Spa-3具有与SpaCas9相当的特异性谱。

[0686] *sRGN的总特异性:*

[0687] 通过计算寡核苷酸切割反应的初始斜率,分析了以下对象的切割动力学:a) Gib11;b) Gib11Spa-1;c) Gib11Spa-2, d) E2, e) F8, f) SluCas9, g)*化脓性葡萄球菌*野生型(New England Biolabs),和h)金黄色葡萄球菌。为61种底物中的每一种计算斜率,并将其归一化为中靶底物的值(定义为1)。将跨60种脱靶底物的整个组的所有归一化值相加以得出总特异性值(如图10中的条形图所示)。使用此指南/目标组合,sRGN比SpyCas9更有特异性。

[0688] 实施例7:sRGN的活性谱的确定

[0689] 为了评估Gib11、Gib11Spa-1、Gib11Spa-3、E2和F8的活性谱,确定了96种具有不同GC含量的中靶DNA底物(根据SEQ ID NO: 347-538成对使用,例如,分别具有SEQ ID NO: 347和348的Ext-A-P-1-A和Ext-A-P-1-B代表一个双链体对)的生化切割。

[0690] *荧光偏振测定方案:*

[0691] 在10 mM Tris (pH = 7.8) 50 mM NaCl中作为10μM溶液(来自100μM原液)制备寡核苷酸双链体,并在循环变温器中于95℃退火5分钟,且然后缓慢冷却(每分钟6℃)。随后将原液在10 mM Tris (pH = 7.8) 50 mM NaCl + 0.05%普朗尼克中稀释。将20 nM寡物(20μL)固定在抗生蛋白链菌素包被的平板上,在5分钟后洗涤两次,且然后与20μL样品一起温育60分钟的动力学(激发波长: 635 nm;发射波长: 670 nm)。在切割之前,形成反应RNP。对于SEQ ID NO: 251-346的每种指导序列,将核酸酶和含有指导序列的sgRNA加入817.6μl 1xPBS $MgCl_2$,并在37℃温育10分钟。将20μl RNP加入含有RNP中的sgRNA的各种DNA底物的孔中,并在37℃测量偏振60分钟。*参见*,例如,Methods Enzymol. 2014;546:1-20。

[0692] 在基于荧光偏振的生化切割测定中测试了指示的变体。针对恒定水平的sgRNA滴定蛋白质的量。在温育靶向VEGFA寡核苷酸序列的变体-RNP复合物后,通过荧光偏振和荧光强度信号的改变(成功切割后随时间推移偏振值降低和荧光强度增加)的变化可以观察切

割。作为切割反应的定量估计,分析了图的初始斜率。将数据归一化为对一种RNA-指导的核酸酶具有最高活性的底物,结果示于图11。与其它sRGN和SpyCas9相比,Gib11Spa1/3和Gib11在更宽范围的靶标中显示出更高的活性。

序列表

[0693]

SEQ ID NO	序列	描述
1	MNQKFILGLDIGITSVGYGLIDYETKNIIDAGVRLFPEANVENNEGRRSKRGS RRLKRRRIHRLERVKLLLTEYDLINKEQIPTSNNPYQIRVKGLSEILSKDELAIA LLHLAKRRGIHNVDVAADKEETASDSLSTKDQINKNAKFLESRYVCELQKER LENEGHVRGVENRFLTKDIVREAKKIIDTQMQYYPEIDETFKEKYISLVETRR EYFEGPGQGSPFGWNGDLKKWYEMLMGHCTYFPQELRSVKYAYSADLFNA LNDLNNLIIQRDNSEKLEYHEKYHIIENVFKQKKKPTLKQIAKEIGVNPEDIKG YRITKSGTPEFTSFKLFHDLKKVVKDHAILDDIDLLNQIAEILTIYQDKDSIVAE LGQLEYLMSEADKQSISELTGYTGTHSLSLKCMNMIIDELWHSSMNQMEVFT YLNMRPKKYELKGYQRIPTDMIDDAILSPVVKRTFIQSINVINKVIEKYGIPEDI IIELARENNSDDRKKFINNLQKKNEATRKRINEIIGQTGNQNAKRIVEKIRLHD QQEGKCLYSLESIPLEDLLNNPNHYEVDHIIPRSVSFDNSYHNKVLVKQSENS KKSNLTPYQYFNSGKSKLSYNQFKQHILNLSKSQDRISKKKKEYLLEERDINK FEVQKEFINRNLVDTRYATRELTNYLKAYFSANNMNVKVKTINGSFTDYLR KVWKFKKERNHGYKHHAEDALIIANADFLFKENKKLKAVNSVLEKPEIETK QLDIQVDSEDNYSEMFIIPKQVQDIKDFRNFKYSHRVDKKPNRQLINDTLYST RKKDNSTYIVQTIKDIYAKDNTTLKKQFDKSPEKFLMYQHDPRTFEKLEVIM KQYANEKNPLAKYHEETGEYLTKYSKKNNGPIVKSLKYIGNKLGSHLDVTH QFKSSTKKLVKLSIKPYRFDVYLTDKGYKFITISYLDVLKKDNYYYIPEQKYD KLKLGKAIDKNAKFIASFYKNDLIKLDGEIYKIIGVNSDTRNMIELDLPDIRYK EYCELNNIKGEPRIKKTIGKKVNSIEKLTTDVLGNVFTNTQYTKPQLLFKRGN	Gib11的多肽序列
2	MNQKFILGLDIGITSVGYGLIDYETKNIIDAGVRLFPEANVENNEGRRSKRGS RRLKRRRIHRLERVKLLLTEYDLINKEQIPTSNNPYQIRVKGLSEILSKDELAIA LLHLAKRRGIHNVDVAADKEETASDSLSTKDQINKNAKFLESRYVCELQKER LENEGHVRGVENRFLTKDIVREAKKIIDTQMQYYPEIDETFKEKYISLVETRR EYFEGPGQGSPFGWNGDLKKWYEMLMGHCTYFPQELRSVKYAYSADLFNA LNDLNNLIIQRDNSEKLEYHEKYHIIENVFKQKKKPTLKQIAKEIGVNPEDIKG YRITKSGTPEFTSFKLFHDLKKVVKDHAILDDIDLLNQIAEILTIYQDKDSIVAE LGQLEYLMSEADKQSISELTGYTGTHSLSLKCMNMIIDELWHSSMNQMEVFT YLNMRPKKYELKGYQRIPTDMIDDAILSPVVKRTFIQSINVINKVIEKYGIPEDI IIELARENNSDDRKKFINNLQKKNEATRKRINEIIGQTGNQNAKRIVEKIRLHD QQEGKCLYSLESIPLEDLLNNPNHYEVDHIIPRSVSFDNSYHNKVLVKQSENS KKSNLTPYQYFNSGKSKLSYNQFKQHILNLSKSQDRISKKKKEYLLEERDINK FEVQKEFINRNLVDTRYATRELTNYLKAYFSANNMNVKVKTINGSFTDYLR KVWKFKKERNHGYKHHAEDALIIANADFLFKENKKLKAVNSVLEKPEIETK QLDIQVDSEDNYSEMFIIPKQVQDIKDFRNFKYSHRVDKKPNRQLINDTLYST RKKDNSTYIVQTIKDIYAKDNTTLKKQFDKSPEKFLMYQHDPRTFEKLEVIM KQYANEKNPLAKYHEETGEYLTKYSKKNNGPIVKSLKYIGNKLGSHLDVTH QFKSSTKKLVKLSIKNYRFDVYLTEKGYKFVTIAYLNVFKKDNYYYIPKDKY QELKEKKKIKDTDQFIASFYKNDLIKLNGDLYKIIGVNSDDRNIIELDYYDIKY KDYCEINNIKGEPRIKKTIGKKTESIEKFTTDVLGNLYLHSTEKAPQLIFKRGL	Gib11Spa-1的多肽序列
3	MNQKFILGLDIGITSVGYGLIDYETKNIIDAGVRLFPEANVENNEGRRSKRGS RRLKRRRIHRLERVKLLLTEYDLINKEQIPTSNNPYQIRVKGLSEILSKDELAIA LLHLAKRRGIHNVDVAADKEETASDSLSTKDQINKNAKFLESRYVCELQKER LENEGHVRGVENRFLTKDIVREAKKIIDTQMQYYPEIDETFKEKYISLVETRR EYFEGPGQGSPFGWNGDLKKWYEMLMGHCTYFPQELRSVKYAYSADLFNA LNDLNNLIIQRDNSEKLEYHEKYHIIENVFKQKKKPTLKQIAKEIGVNPEDIKG YRITKSGTPEFTSFKLFHDLKKVVKDHAILDDIDLLNQIAEILTIYQDKDSIVAE LGQLEYLMSEADKQSISELTGYTGTHSLSLKCMNMIIDELWHSSMNQMEVFT YLNMRPKKYELKGYQRIPTDMIDDAILSPVVKRTFIQSINVINKVIEKYGIPEDI IIELARENNSDDRKKFINNLQKKNEATRKRINEIIGQTGNQNAKRIVEKIRLHD QQEGKCLYSLESIPLEDLLNNPNHYEVDHIIPRSVSFDNSYHNKVLVKQSENS KKSNLTPYQYFNSGKSKLSYNQFKQHILNLSKSQDRISKKKKEYLLEERDINK FEVQKEFINRNLVDTRYATRELTNYLKAYFSANNMNVKVKTINGSFTDYLR KVWKFKKERNHGYKHHAEDALIIANADFLFKENKKLKAVNSVLEKPEIETK QLDIQVDSEDNYSEMFIIPKQVQDIKDFRNFKFSHRVDKKPNRQLINDTLYST RMKDEHDYIVQTITDIYGKDNTNLKKQFNKNPEKFLMYQNDPKTFEKLSIIM	Gib11Spa-2的多肽序列

[0694]

	KQYSDEKNPLAKYYEETGEYLTKYSKKNNGPIVKKIKLLGNKVGNHLDVTN KYENSTKKLVKLSIKNYRFDVYLTEKGYKFVTIAYLNVFKKDNYYYIPKDKY QELKEKKKIKDTDQFIASFYKNDLIKLNGDLYKIIGVNSDDRNIIELDYYDIKY KDYCEINNIKGEPRIKKTIGKKTESIEKFTTDVLGNLYLHSTEKAPQLIFKRGL	
4	MNQKFILGLDIGITSVGYGLIDYETKNIIDAGVRLFPEANVENNEGRRSKRGS RRLKRRRIHRLERVKLLLTEYDLINKEQIPTSNNPYQIRVKGLSEILSKDELAIA LLHLAKRRGIHNVDVAADKEETASDSLSTKDQINKNAKFLESRYVCELQKER LENEGHVRGVENRFLTKDIVREAKKIIDTQMQYYPEIDETFKEKYISLVETRR EYFEGPGQGSPFGWNGDLKKWYEMLMGHCTYFPQELRSVKYAYSADLFNA LNDLNNLIIQRDNSEKLEYHEKYHIIENVFKQKKKPTLKQIAKEIGVNPEDIKG YRITKSGTPEFTSFKLFHDLKKVVKDHAILDDIDLLNQIAEILTIYQDKDSIVAE LGQLEYLMSEADKQSISELTGYTGTHSLSLKCMNMIIDELWHSSMNQMEVFT YLNMRPKKYELKGYQRIPTDMIDDAILSPVVKRTFIQSINVINKVIEKYGIPEDI IIELARENNSDDRKKFINNLQKKNEATRKRINEIIGQTGNQNAKRIVEKIRLHD QQEGKCLYSLESIPLEDLLNNPNHYEVDHIIPRSVSFDNSYHNKVLVKQSENS KKSNLTPYQYFNSGKSKLSYNQFKQHILNLSKSQDRISKKKKEYLLEERDINK FEVQKEFINRNLVDTRYATRELTSYLKAYFSANNMDVKVKTINGSFTNHLRK VWRFDKYRNHGYKHHAEDALIIANADFLFKENKKLQNTNKILEKPTIENNTK KVTVEKEEDYNNVFETPKLVEDIKQYRDYKFSHRVDKKPNRQLINDTLYSTR MKDEHDYIVQTITDIYGKDNTNLKKQFNKNPEKFLMYQNDPKTFEKLSIIMK QYSDEKNPLAKYYEETGEYLTKYSKKNNGPIVKKIKLLGNKVGNHLDVTNK YENSTKKLVKLSIKNYRFDVYLTEKGYKFVTIAYLNVFKKDNYYYIPKDKYQ ELKEKKKIKDTDQFIASFYKNDLIKLNGDLYKIIGVNSDDRNIIELDYYDIKYK DYCEINNIKGEPRIKKTIGKKTESIEKFTTDVLGNLYLHSTEKAPQLIFKRGL	Gib11Spa-3 的多肽序列
5	MNQKFILGLDIGITSVGYGLIDYETKNIIDAGVRLFPEANVENNEGRRSKRGS RRLKRRRIHRLERVKKLLEDYNLLDQSQIPQSTNPYAIRVKGLSEALSKDELV IALLHIAKRRGIHNINVSSEDEDASNELSTKEQINRNNKLLKDKYVCEVQLQR LKEGQIRGEKNRFKTTDILKEIDQLLKVQKDYHNLDIDFINQYKEIVETRREY FEGPGKGSPYGWEGDPKAWYETLMGHCTYFPDELRSVKYAYSADLFNALN DLNNLVIQRDGLSKLEYHEKYHIIENVFKQKKKPTLKQIANEINVNPEDIKGY RITKSGKPEFTSFKLFHDLKKVVKDHAILDDIDLLNQIAEILTIYQDKDSIVAEL GQLEYLMSEADKQSISELTGYTGTHSLSLKCMNMIIDELWHSSMNQMEVFT YLNMRPKKYELKGYQRIPTDMIDDAILSPVVKRTFIQSINVINKVIEKYGIPEDI IIELARENNSDDRKKFINNLQKKNEATRKRINEIIGQTGNQNAKRIVEKIRLHD QQEGKCLYSLESIPLEDLLNNPNHYEVDHIIPRSVSFDNSYHNKVLVKQSENS KKSNLTPYQYFNSGKSKLSYNQFKQHILNLSKSQDRISKKKKEYLLEERDINK FEVQKEFINRNLVDTRYATRELTNYLKAYFSANNMNVKVKTINGSFTDYLR KVWKFKKERNHGYKHHAEDALIIANADFLFKENKKLKAVNSVLEKPEIETK QLDIQVDSEDNYSEMFIIPKQVQDIKDFRNFKYSHRVDKKPNRQLINDTLYST RKKDNSTYIVQTIKDIYAKDNTTLKKQFDKSPEKFLMYQHDPRTFEKLEVIM KQYANEKNPLAKYHEETGEYLTKYSKKNNGPIVKSLKYIGNKLGSHLDVTH QFKSSTKKLVKLSIKPYRFDVYLTDKGYKFITISYLDVLKKDNYYYIPEQKYD KLKLGKAIDKNAKFIASFYKNDLIKLDGEIYKIIGVNSDTRNMIELDLPDIRYK EYCELNNIKGEPRIKKTIGKKVNSIEKLTTDVLGNVFTNTQYTKPQLLFKRGN	P2H12 的多肽序列
6	MNQKFILGLDIGITSVGYGLIDYETKNIIDAGVRLFPEANVENNEGRRSKRGS RRLKRRRIHRLDRVKHLLAEYDLLDLTNIPKSTNPYQTRVKGLNEKLSKDEL VIALLHIAKRRGIHNVDVAADKEETASDSLSTKDQINKNAKFLESRYVCELQ KERLENEGHVRGVENRFLTKDIVREAKKIIDTQMQYYPEIDETFKEKYISLVE TRREYFEGPGKGSPFGWEGNIKKWFEQMMGHCTYFPEELRSVKYSYSAELF NALNDLNNLVITRDEDAKLNYGEKFQIIENVFKQKKTPNLKQIAIEIGVHETEI KGYRVNKSGTPEFTEFKLYHDLKSIVFDKSILENEAILDQIAEILTIYQDEQSIK EELNKLPEILNEQDKAEIAKLIGYNGTHRLSLKCIHLINEELWQTSRNQMEIFN YLNIKPNKVDLSEQNKIPKDMVNDFILSPVVKRTFIQSINVINKVIEKYGIPEDI IIELARENNSDDRKKFINNLQKKNEATRKRINEIIGQTGNQNAKRIVEKIRLHD QQEGKCLYSLKDIPLEDLLRNPNNYDIDHIIPRSVSFDDSMHNKVLVRREQNA KKNNQTPYQYLTSGYADIKYSVFKQHVLNLAENKDRMTKKKREYLLEERDI NKFEVQKEFINRNLVDTRYATRELTNYLKAYFSANNMNVKVKTINGSFTDY LRKVWKFKKERNHGYKHHAEDALIIANADFLFKENKKLKAVNSVLEKPEIET KQLDIQVDSEDNYSEMFIIPKQVQDIKDFRNFKYSHRVDKKPNRQLINDTLYS TRKKDNSTYIVQTIKDIYAKDNTTLKKQFDKSPEKFLMYQHDPRTFEKLEVI MKQYANEKNPLAKYHEETGEYLTKYSKKNNGPIVKSLKYIGNKLGSHLDVT	E2 的多肽序列

[0695]

	HQFKSSTKKLVKLSIKPYRFDVYLTDKGYKFITISYLDVLKKDNYYYIPEQKY DKLKLGKAIDKNAKFIASFYKNDLIKLDGEIYKIIGVNSDTRNMIELDLPDIRY KEYCELNNIKGEPRIKKTIGKKVNSIEKLTTDVLGNVFTNTQYTKPQLLFKRG N	
7	MNQKFILGLDIGITSVGYGLIDYETKNIIDAGVRLFPEANVENNEGRRSKRGS RRLKRRRIHRLDRVKHLLAEYDLLDLTNIPKSTNPYQTRVKGLNEKLSKDEL VIALLHIAKRRGIHNVDVAADKEETASDSLSTKDQINKNAKFLESRYVCELQ KERLENEGHVRGVENRFLTKDIVREAKKIIDTQMQYYPEIDETFKEKYISLVE TRREYFEGPGKGSPFGWEGNIKKWFEQMMGHCTYFPEELRSVKYSYSAELF NALNDLNNLVITRDEDAKLNYGEKFQIIENVFKQKKTPNLKQIAIEIGVHETEI KGYRVNKSGTPEFTEFKLYHDLKSIVFDKSILENEAILDQIAEILTIYQDEQSIK EELNKLPEILNEQDKAEIAKLIGYNGTHRLSLKCIHLINEELWQTSRNQMEIFN YLNIKPNKVDLSEQNKIPKDMVNDFILSPVVKRTFIQSINVINKVIEKYGIPEDI IIELARENNSDDRKKFINNLQKKNEATRKRINEIIGQTGNQNAKRIVEKIRLHD QQEGKCLYSLKDIPLEDLLRNPNNYDIDHIIPRSVSFDDSMHNKVLVRREQNA KKNNQTPYQYLTSGYADIKYSVFKQHVLNLAENKDRMTKKKREYLLEERDI NKFEVQKEFINRNLVDTRYATRELTNYLKAYFSANNMNVKVKTINGSFTDY LRKVWKFKKERNHGYKHHAEDALIIANADFLFKENKKLKAVNSVLEKPEIET DQKDIQVDSEDNYSEMFIIPKQVQDIKDFRNFKYSHRVDKKPNRQLINDTLYS TRKKDNSTYIVQTIKDIYAKDNTTLKKQFDKSPEKFLMYQHDPRTFEKLEVI MKQYANEKNPLAKYHEETGEYLTKYSKKNNGPIVKSLKYIGNKLGSHLDVT HQFKSSTKKLVKLSIKPYRFDVYLTDKGYKFITISYLDVLKKDNYYYIPEQKY DKLKLGKAIDKNAKFIASFYKNDLIKLDGEIYKIIGVNSDTRNMIELDLPDIRY KEYCELNNIKGEPRIKKTIGKKVNSIEKLTTDVLGNVFTNTQYTKPQLLFKRG N	E2+K741D+L743K 的多肽序列
8	MNQKFILGLDIGITSVGYGLIDYETKNIIDAGVRLFPEANVENNEGRRSKRGS RRLKRRRIHRLDRVKHLLAEYDLLDLTNIPKSTNPYQTRVKGLNEKLSKDEL VIALLHIAKRRGIHNVDVAADKEETASDSLSTKDQINKNAKFLESRYVCELQ KERLENEGHVRGVENRFLTKDIVREAKKIIDTQMQYYPEIDETFKEKYISLVE TRREYFEGPGKGSPFGWEGNIKKWFEQMMGHCTYFPEELRSVKYSYSAELF NALNDLNNLVITRDEDAKLNYGEKFQIIENVFKQKKTPNLKQIAIEIGVHETEI KGYRVNKSGTPEFTEFKLYHDLKSIVFDKSILENEAILDQIAEILTIYQDEQSIK EELNKLPEILNEQDKAEIAKLIGYNGTHRLSLKCIHLINEELWQTSRNQMEIFN YLNIKPNKVDLSEQNKIPKDMVNDFILSPVVKRTFIQSINVINKVIEKYGIPEDI IIELARENNSDDRKKFINNLQKKNEATRKRINEIIGQTGNQNAKRIVEKIRLHD QQEGKCLYSLKDIPLEDLLRNPNNYDIDHIIPRSVSFDDSMHNKVLVRREQNA KKNNQTPYQYLTSGYADIKYSVFKQHVLNLAENKDRMTKKKREYLLEERDI NKFEVQKEFINRNLVDTRYATRELTNYLKAYFTANNMDVKVKTINGSFTDY LRKVWKFKKERNHGYKHHAEDALIIANADFLFKENKKLKAVNSVLEKPEIET KQLDIQVDSEDNYSEMFIIPKQVQDIKDFRNFKYSHRVDKKPNRQLINDTLYS TRKKDNSTYIVQTIKDIYAKDNTTLKKQFDKSPEKFLMYQHDPRTFEKLEVI MKQYANEKNPLAKYHEETGEYLTKYSKKNNGPIVKSLKYIGNKLGSHLDVT HQFKSSTKKLVKLSIKPYRFDVYLTDKGYKFITISYLDVLKKDNYYYIPEQKY DKLKLGKAIDKNAKFIASFYKNDLIKLDGEIYKIIGVNSDTRNMIELDLPDIRY KEYCELNNIKGEPRIKKTIGKKVNSIEKLTTDVLGNVFTNTQYTKPQLLFKRG N	E2+S670T+N675D 的多肽序列
9	MNQKFILGLDIGITSVGYGLIDYETKNIIDAGVRLFPEANVENNEGRRSKRGS RRLKRRRIHRLDRVKHLLAEYDLLDLTNIPKSTNPYQTRVKGLNEKLSKDEL VIALLHIAKRRGIHNVDVAADKEETASDSLSTKDQINKNAKFLESRYVCELQ KERLENEGHVRGVENRFLTKDIVREAKKIIDTQMQYYPEIDETFKEKYISLVE TRREYFEGPGKGSPFGWEGNIKKWFEQMMGHCTYFPEELRSVKYSYSAELF NALNDLNNLVITRDEDAKLNYGEKFQIIENVFKQKKTPNLKQIAIEIGVHETEI KGYRVNKSGTPEFTEFKLYHDLKSIVFDKSILENEAILDQIAEILTIYQDEQSIK EELNKLPEILNEQDKAEIAKLIGYNGTHRLSLKCIHLINEELWQTSRNQMEIFN YLNIKPNKVDLSEQNKIPKDMVNDFILSPVVKRTFIQSINVINKVIEKYGIPEDI IIELARENNSDDRKKFINNLQKKNEATRKRINEIIGQTGNQNAKRIVEKIRLHD QQEGKCLYSLKDIPLEDLLRNPNNYDIDHIIPRSVSFDDSMHNKVLVRREQNA KKNNQTPYQYLTSGYADIKYSVFKQHVLNLAENKDRMTKKKREYLLEERDI NKFEVQKEFINRNLVDTRYATRELTNYLKAYFSANNMNVKVKTINGSFTDY LRKVWKFKKERNHGYKHHAEDALIIANADFLFKENKKLKAVNSVLEKPEIET NQNDIQVDSEDNYSEMFIIPKQVQDIKDFRNFKYSHRVDKKPNRQLINDTLYS	E2+K741N+L743N 的多肽序列

[0696]

	TRKKDNSTYIVQTIKDIYAKDNTTLKKQFDKSPEKFLMYQHDPRTFEKLEVI MKQYANEKNPLAKYHEETGEYLTKYSKKNNGPIVKSLKYIGNKLGSHLDVT HQFKSSTKKLVKLSIKPYRFDVYLTDKGYKFITISYLDVLKKDNYYYIPEQKY DKLKLGKAIDKNAKFIASFYKNDLIKLDGEIYKIIGVNSDTRNMIELDLPDIRY KEYCELNNIKGEPRIKKTIGKKVNSIEKLTTDVLGNVFTNTQYTKPQLLFKRG N	
10	MNQKFILGLDIGITSVGYGLIDYETKNIIDAGVRLFPEANVENNEGRRSKRGS RRLKRRRIHRLERVKSLLSEYKIISGLAPTNNQPYNIRVKGLTEQLTKDELAV ALLHIAKRRGIHKIDVIDSNDDVGNELSTKEQLNKNSKLLKDKFVCQIQLER MNEGQVRGEKNRFKTADIIKEIIQLLNVQKNFHQLDENFINKYIELVEMRREY FEGPGQGSPFGWNGDLKKWYEMLMGHCTYFPQELRSVKYAYSADLFNALN DLNNLIIQRDNSEKLEYHEKYHIIENVFKQKKKPTLKQIAKEIGVNPEDIKGYR ITKSGTPEFTEFKLYHDLKSVLFDQSILENEDVLDQIAEILTIYQDKDSIKSKLT ELDILLNEEDKENIAQLTGYNGTHRLSLKCIRLVLEEQWYSSRNQMEIFTHLN IKPKKINLTAANKIPKAMIDEFILSPVVKRTFIQSINVINKVIEKYGIPEDIIIELA RENNSDDRKKFINNLQKKNEATRKRINEIIGQTGNQNAKRIVEKIRLHDQQEG KCLYSLESIALMDLLNNPQNYEVDHIIPRSVAFDNSIHNKVLVKQIENSKKGN RTPYQYLNSSDAKLSYNQFKQHILNLSKSKDRISKKKKDYLLEERDINKFEVQ KEFINRNLVDTRYATRELTSYLKAYFSANNMDVKVKTINGSFTNHLRKVWR FDKYRNHGYKHHAEDALIIANADFLFKENKKLKAVNSVLEKPEIETKQLDIQ VDSEDNYSEMFIIPKQVQDIKDFRNFKYSHRVDKKPNRQLINDTLYSTRKKD NSTYIVQTIKDIYAKDNTTLKKQFDKSPEKFLMYQHDPRTFEKLEVIMKQYA NEKNPLAKYHEETGEYLTKYSKKNNGPIVKSLKYIGNKLGSHLDVTHQFKSS TKKLVKLSIKPYRFDVYLTDKGYKFITISYLDVLKKDNYYYIPEQKYDKLKL GKAIDKNAKFIASFYKNDLIKLDGEIYKIIGVNSDTRNMIELDLPDIRYKEYCE LNNIKGEPRIKKTIGKKVNSIEKLTTDVLGNVFTNTQYTKPQLLFKRGN	F8 的多肽序列
11	MNQKFILGLDIGITSVGYGLIDYETKNIIDAGVRLFPEANVENNEGRRSKRGS RRLKRRRIHRLERVKSLLSEYKIISGLAPTNNQPYNIRVKGLTEQLTKDELAV ALLHIAKRRGIHKIDVIDSNDDVGNELSTKEQLNKNSKLLKDKFVCQIQLER MNEGQVRGEKNRFKTADIIKEIIQLLNVQKNFHQLDENFINKYIELVEMRREY FEGPGQGSPFGWNGDLKKWYEMLMGHCTYFPQELRSVKYAYSADLFNALN DLNNLIIQRDNSEKLEYHEKYHIIENVFKQKKKPTLKQIAKEIGVNPEDIKGYR ITKSGTPEFTEFKLYHDLKSVLFDQSILENEDVLDQIAEILTIYQDKDSIKSKLT ELDILLNEEDKENIAQLTGYNGTHRLSLKCIRLVLEEQWYSSRNQMEIFTHLN IKPKKINLTAANKIPKAMIDEFILSPVVKRTFIQSINVINKVIEKYGIPEDIIIELA RENNSDDRKKFINNLQKKNEATRKRINEIIGQTGNQNAKRIVEKIRLHDQQEG KCLYSLESIALMDLLNNPQNYEVDHIIPRSVAFDNSIHNKVLVKQIENSKKGN RTPYQYLNSSDAKLSYNQFKQHILNLSKSKDRISKKKKDYLLEERDINKFEVQ KEFINRNLVDTRYATRELTSYLKAYFSANNMDVKVKTINGSFTNHLRKVWR FDKYRNHGYKHHAEDALIIANADFLFKENKKLKAVNSVLEKPEIETDQKDIQ VDSEDNYSEMFIIPKQVQDIKDFRNFKYSHRVDKKPNRQLINDTLYSTRKKD NSTYIVQTIKDIYAKDNTTLKKQFDKSPEKFLMYQHDPRTFEKLEVIMKQYA NEKNPLAKYHEETGEYLTKYSKKNNGPIVKSLKYIGNKLGSHLDVTHQFKSS TKKLVKLSIKPYRFDVYLTDKGYKFITISYLDVLKKDNYYYIPEQKYDKLKL GKAIDKNAKFIASFYKNDLIKLDGEIYKIIGVNSDTRNMIELDLPDIRYKEYCE LNNIKGEPRIKKTIGKKVNSIEKLTTDVLGNVFTNTQYTKPQLLFKRGN	F8+K737D+L739K 的多肽序列
12	MNQKFILGLDIGITSVGYGLIDYETKNIIDAGVRLFPEANVENNEGRRSKRGS RRLKRRRIHRLERVKSLLSEYKIISGLAPTNNQPYNIRVKGLTEQLTKDELAV ALLHIAKRRGIHKIDVIDSNDDVGNELSTKEQLNKNSKLLKDKFVCQIQLER MNEGQVRGEKNRFKTADIIKEIIQLLNVQKNFHQLDENFINKYIELVEMRREY FEGPGQGSPFGWNGDLKKWYEMLMGHCTYFPQELRSVKYAYSADLFNALN DLNNLIIQRDNSEKLEYHEKYHIIENVFKQKKKPTLKQIAKEIGVNPEDIKGYR ITKSGTPEFTEFKLYHDLKSVLFDQSILENEDVLDQIAEILTIYQDKDSIKSKLT ELDILLNEEDKENIAQLTGYNGTHRLSLKCIRLVLEEQWYSSRNQMEIFTHLN IKPKKINLTAANKIPKAMIDEFILSPVVKRTFIQSINVINKVIEKYGIPEDIIIELA RENNSDDRKKFINNLQKKNEATRKRINEIIGQTGNQNAKRIVEKIRLHDQQEG KCLYSLESIALMDLLNNPQNYEVDHIIPRSVAFDNSIHNKVLVKQIENSKKGN RTPYQYLNSSDAKLSYNQFKQHILNLSKSKDRISKKKKDYLLEERDINKFEVQ KEFINRNLVDTRYATRELTSYLKAYFSANNMDVKVKTINGSFTNHLRKVWR FDKYRNHGYKHHAEDALIIANADFLFKENKKLKAVNSVLEKPEIETNQNDIQ VDSEDNYSEMFIIPKQVQDIKDFRNFKYSHRVDKKPNRQLINDTLYSTRKKD	F8+K737N+L739N 的多肽序列

[0697]

	NSTYIVQTIKDIYAKDNTTLKKQFDKSPEKFLMYQHDPRTFEKLEVIMKQYA NEKNPLAKYHEETGEYLTKYSKKNNGPIVKSLKYIGNKLGSHLDVTHQFKSS TKKLVKLSIKPYRFDVYLTDKGYKFITISYLDVLKKDNYYYIPEQKYDKLKL GKAIDKNAKFIASFYKNDLIKLDGEIYKIIGVNSDTRNMIELDLPDIRYKEYCE LNNIKGEPRIKKTIGKKVNSIEKLTTDVLGNVFTNTQYTKPQLLFKRGN	
13	ATGAACCAGAAATTTATCCTGGGTCTGGATATTGGTATTACCAGCGTTGG TTATGGCCTGATTGATTACGAAACCAAAAACATTATTGATGCCGGTGTTC GTCTGTTTCCGGAAGCAAATGTTGAAAATAATGAAGGTCGTCGTAGCAAA CGTGGTAGCCGTCGTCTGAAACGTCGTCGTATTCATCGTCTGGAACGTGT TAAACTGCTGCTGACCGAATATGATCTGATTAACAAAGAGCAGATTCCGA CCAGCAATAACCCGTATCAGATTCGTGTTAAAGGTCTGAGCGAAATCCTG AGCAAAGATGAACTGGCAATTGCACTGCTGCATCTGGCAAAACGCCGTG GCATTCATAATGTTGATGTTGCAGCAGATAAAGAAGAAACCGCAAGCGA TAGCCTGAGCACCAAAGATCAGATTAACAAAAACGCCAAATTTCTGGAA AGCCGCTATGTTTGTGAACTGCAGAAAGAACGTCTGGAAAATGAAGGTC ATGTTCGTGGTGTTGAAAATCGCTTTCTGACGAAAGATATTGTGCGTGAG GCCAAAAAAATCATCGATACCCAGATGCAGTATTACCCGGAAATTGATG AAACCTTCAAAGAGAAATATATCAGCCTGGTTGAAACCCGTCGCGAATAT TTTGAAGGTCCTGGTCAGGGTAGCCCGTTTGGTTGGAATGGTGATCTGAA AAAATGGTACGAAATGCTGATGGGTCACTGTACCTATTTTCCGCAAGAAC TGCGTAGCGTTAAATATGCCTATAGCGCAGACCTGTTTAATGCACTGAAT GATCTGAACAACCTGATTATTCAGCGCGATAATAGCGAGAAACTGGAAT ACCATGAGAAGTATCACATCATCGAGAACGTGTTCAAGCAGAAAAAAAA GCCGACGCTGAAACAAATCGCAAAAGAGATTGGCGTTAACCCGGAAGAT ATTAAAGGTTATCGTATTACCAAAAGCGGCACACCGGAGTTTACATCCTT TAAACTGTTCCACGATCTGAAAAAAGTGGTGAAAGATCATGCCATCCTGG ATGATATTGATCTGCTGAATCAGATTGCAGAAATCCTGACCATCTATCAG GATAAAGATAGCATTGTTGCAGAACTGGGTCAGCTGGAATATCTGATGAG CGAAGCCGATAAACAGAGCATTAGCGAACTGACCGGTTATACCGGTACA CATAGCCTGTCACTGAAATGCATGAACATGATTATCGATGAACTGTGGCA TAGCAGCATGAACCAGATGGAAGTTTTTACCTATCTGAATATGCGTCCGA AAAAGTATGAGCTGAAAGGTTATCAGCGTATTCCGACCGATATGATTGAT GATGCAATTCTGAGTCCGGTTGTGAAACGCACCTTTATTCAGAGCATCAA CGTGATCAACAAAGTGATCGAGAAATATGGCATCCCCGAAGATATCATTA TCGAACTGGCACGTGAAAATAACTCCGATGATCGCAAAAAGTTCATCAAC AACCTGCAGAAAAAGAATGAAGCAACCCGCAAACGCATTAACGAAATTA TTGGTCAGACCGGTAATCAGAATGCCAAACGTATTGTGGAAAAAATCCGT CTGCATGATCAGCAAGAGGGCAAATGTCTGTATAGCCTGGAAAGCATTCC TCTGGAAGATCTGCTGAACAATCCGAATCATTATGAAGTGGATCACATTA TTCCGCGTAGCGTGAGCTTTGATAATTCCTATCATAATAAAGTGCTGGTG AAACAGAGCGAAAACTCCAAAAAATCCAACCTGACACCGTATCAGTATT TCAATAGCGGCAAATCCAAACTGAGCTACAACCAGTTTAAACAGCATATT CTGAACCTGAGCAAAAGCCAGGATCGCATCAGCAAGAAGAAGAAGGAGT ACCTGCTGGAAGAACGCGACATCAACAAATTTGAAGTGCAGAAAGAATT TATCAACCGCAACCTGGTTGATACCCGTTATGCAACCCGTGAACTGACCA ATTATCTGAAAGCATATTTCAGCGCCAACAACATGAACGTGAAAGTGAA AACGATTAACGGCAGCTTTACCGATTATCTGCGTAAAGTGTGGAAATTCA AAAAAGAACGCAACCACGGCTATAAACATCATGCCGAAGATGCCCTGAT TATTGCAAATGCAGATTTCCTGTTTAAAGAAAACAAAAAACTGAAAGCCG TCAACAGCGTGCTGGAAAAACCGGAAATTGAGACAAAACAGCTGGACAT TCAGGTTGATAGCGAAGATAATTACAGCGAAATGTTTATCATCCCGAAAC AGGTGCAGGATATCAAAGATTTTCGCAACTTCAAATATAGCCACCGCGTT GACAAAAAAACCTAATCGTCAGCTGATTAACGATACCCTGTATAGCACCCG CAAAAAAGATAACAGCACCTATATTGTGCAGACCATTAAAGACATCTAC GCCAAAGATAATACCACCCTGAAAAAACAGTTCGACAAAAGCCCAGAAA AATTTCTGATGTATCAGCATGATCCGCGTACCTTCGAAAAACTGGAAGTT ATTATGAAACAGTATGCCAACGAGAAAAATCCGCTGGCCAAATATCACG AAGAAACCGGTGAATATCTGACCAAATATTCCAAGAAGAACAACGGTCC GATCGTTAAATCCCTGAAATATATCGGTAATAAACTGGGCAGCCATCTGG ATGTTACCCATCAGTTTAAAAGCTCCACAAAGAAGCTGGTTAAACTGTCC ATCAAACCGTATCGCTTTGATGTGTATCTGACCGACAAAGGCTATAAATT	编码Gib11的 DNA序列

[0698]

	CATTACCATCAGCTATCTGGACGTGCTGAAAAAAGACAACTATTATTATA TCCCGGAACAGAAATATGATAAACTGAAACTGGGTAAAGCCATCGATAA AAACGCCAAATTTATCGCCAGCTTCTACAAAAACGACCTGATTAAACTGG ATGGCGAGATCTATAAAATCATCGGTGTTAATAGCGACACCCGCAATATG ATTGAGCTGGATCTGCCGGATATTCGCTATAAAGAATATTGCGAACTGAA CAACATTAAAGGCGAACCGCGTATCAAAAAGACCATCGGCAAAAAAGTG AATAGCATCGAGAAACTGACCACCGATGTTCTGGGTAATGTGTTTACCAA TACCCAGTATACCAAACCTCAGCTGCTGTTCAAACGCGGTAAT	
14	ATGAACCAGAAATTTATCCTGGGTCTGGATATTGGTATTACCAGCGTTGG TTATGGCCTGATTGATTACGAAACCAAAAACATTATTGATGCCGGTGTTC GTCTGTTTCCGGAAGCAAATGTTGAAAATAATGAAGGTCGTCGTAGCAAA CGTGGTAGCCGTCGTCTGAAACGTCGTCGTATTCATCGTCTGGAACGTGT TAAACTGCTGCTGACCGAATATGATCTGATTAACAAAGAGCAGATTCCGA CCAGCAATAACCCGTATCAGATTCGTGTTAAAGGTCTGAGCGAAATCCTG AGCAAAGATGAACTGGCAATTGCACTGCTGCATCTGGCAAAACGCCGTG GCATTCATAATGTTGATGTTGCAGCAGATAAAGAAGAAACCGCAAGCGA TAGCCTGAGCACCAAAGATCAGATTAACAAAAACGCCAAATTTCTGGAA AGCCGCTATGTTTGTGAACTGCAGAAAGAACGTCTGGAAAATGAAGGTC ATGTTCGTGGTGTTGAAAATCGCTTTCTGACGAAAGATATTGTGCGTGAG GCCAAAAAAATCATCGATACCCAGATGCAGTATTACCCGGAAATTGATG AAACCTTCAAAGAGAAATATATCAGCCTGGTTGAAACCCGTCGCGAATAT TTTGAAGGTCCTGGTCAGGGTAGCCCGTTTGGTTGGAATGGTGATCTGAA AAAATGGTACGAAATGCTGATGGGTCACTGTACCTATTTTCCGCAAGAAC TGCGTAGCGTTAAATATGCCTATAGCGCAGACCTGTTTAATGCACTGAAT GATCTGAACAACCTGATTATTCAGCGCGATAATAGCGAGAAACTGGAAT ACCATGAGAAGTATCACATCATCGAGAACGTGTTCAAGCAGAAAAAAAA GCCGACGCTGAAACAAATCGCAAAAGAGATTGGCGTTAACCCGGAAGAT ATTAAAGGTTATCGTATTACCAAAAGCGGCACACCGGAGTTTACATCCTT TAAACTGTTCCACGATCTGAAAAAAGTGGTGAAAGATCATGCCATCCTGG ATGATATTGATCTGCTGAATCAGATTGCAGAAATCCTGACCATCTATCAG GATAAAGATAGCATTGTTGCAGAACTGGGTCAGCTGGAATATCTGATGAG CGAAGCCGATAAACAGAGCATTAGCGAACTGACCGGTTATACCGGTACA CATAGCCTGTCACTGAAATGCATGAACATGATTATCGATGAACTGTGGCA TAGCAGCATGAACCAGATGGAAGTTTTTACCTATCTGAATATGCGTCCGA AAAAGTATGAGCTGAAAGGTTATCAGCGTATTCCGACCGATATGATTGAT GATGCAATTCTGAGTCCGGTTGTGAAACGCACCTTTATTCAGAGCATCAA CGTGATCAACAAAGTGATCGAGAAATATGGCATCCCCGAAGATATCATTA TCGAACTGGCACGTGAAAATAACTCCGATGATCGCAAAAAGTTCATCAAC AACCTGCAGAAAAAGAATGAAGCAACCCGCAAACGCATTAACGAAATTA TTGGTCAGACCGGTAATCAGAATGCCAAACGTATTGTGGAAAAAATCCGT CTGCATGATCAGCAAGAGGGCAAATGTCTGTATAGCCTGGAAAGCATTCC TCTGGAAGATCTGCTGAACAATCCGAATCATTATGAAGTGGATCACATTA TTCCGCGTAGCGTGAGCTTTGATAATTCCTATCATAATAAAGTGCTGGTG AAACAGAGCGAAAACTCCAAAAAATCCAACCTGACACCGTATCAGTATT TCAATAGCGGCAAATCCAAACTGAGCTACAACCAGTTTAAACAGCATATT CTGAACCTGAGCAAAAGCCAGGATCGCATCAGCAAGAAGAAGAAGGAGT ACCTGCTGGAAGAACGCGATATTAACAAATTTGAAGTGCAGAAAGAATT TATCAACCGCAACCTGGTTGATACCCGTTATGCAACCCGTGAACTGACCA ATTATCTGAAAGCATATTTCAGCGCCAACAACATGAACGTGAAAGTGAA AACGATTAACGGCAGCTTTACCGATTATCTGCGTAAAGTGTGGAAATTCA AAAAAGAACGCAACCACGGCTATAAACATCATGCCGAAGATGCCCTGAT TATTGCAAATGCAGATTTCCTGTTTAAAGAAAACAAAAAACTGAAAGCCG TCAACAGCGTGCTGGAAAAACCGGAAATTGAGACAAAACAGCTGGACAT TCAGGTTGATAGCGAAGATAATTACAGCGAAATGTTTATCATCCCGAAAC AGGTGCAGGATATCAAAGATTTTCGCAACTTCAAATATAGCCACCGCGTT GACAAAAAACCTAATCGTCAGCTGATTAACGATACCCTGTATAGCACCCG CAAAAAAGATAACAGCACCTATATTGTGCAGACCATTAAAGACATCTAC GCCAAAGATAATACCACCCTGAAAAAACAGTTCGACAAAAGCCCAGAAA AATTTCTGATGTATCAGCATGATCCGCGTACCTTCGAAAAACTGGAAGTT ATTATGAAACAGTATGCCAACGAGAAAAATCCGCTGGCCAAATATCACG AAGAAACCGGTGAATATCTGACCAAATATTCCAAGAAGAACAACGGTCC	编码Gib11Spa-1的DNA序列

[0699]

	GATCGTTAAATCCCTGAAATATATCGGTAATAAACTGGGCAGCCATCTGG ATGTTACCCATCAGTTTAAAAGCTCCACAAAGAAGCTGGTTAAGCTGTCC ATTAAAAACTATCGCTTCGATGTGTATCTGACCGAGAAAGGTTATAAGTT TGTGACCATTGCCTACCTGAATGTGTTCAAAAAAGACAACTATTACTATA TTCCGAAAGACAAATACCAAGAACTTAAAGAGAAGAAGAAAATCAAGGA CACCGATCAGTTTATCGCCAGCTTCTATAAAAACGATCTGATCAAGCTGA ACGGCGACCTGTATAAAATCATTGGTGTGAATAGTGATGACCGCAACATC ATTGAGCTGGATTATTACGACATCAAATACAAGGATTACTGCGAGATCAA CAACATTAAAGGTGAACCGCGTATCAAAAAGACCATTGGCAAAAAAACG GAAAGCATCGAAAAGTTTACCACCGATGTTCTGGGTAATCTGTATCTGCA TAGTACCGAAAAAGCACCGCAGCTGATTTTCAAACGCGGTCTG	
15	ATGAACCAGAAGTTCATTCTCGGTCTGGACATTGGCATTACTAGCGTGGG ATACGGCTTGATTGACTACGAGACTAAGAACATCATCGATGCCGGCGTCC GCCTGTTCCCGGAAGCCAACGTGGAGAACAATGAGGGCCGGAGGTCGAA GAGAGGCTCCCGCCGCCTGAAGCGGCGGCGAATCCACCGGCTGGAGAGA GTGAAGCTGCTGCTCACCGAATACGACCTGATCAACAAAGAACAGATCC CGACCTCCAACAACCCGTACCAGATCAGAGTGAAGGGTCTGTCAGAAAT CCTGTCCAAGGACGAACTGGCAATCGCCCTGCTGCACCTGGCCAAGCGGC GCGGAATCCACAACGTGGATGTGGCTGCCGACAAGGAAGAAACCGCTTC CGACTCCCTGAGCACTAAGGACCAGATCAACAAGAACGCCAAGTTCTTG GAGTCCCGCTACGTGTGCGAGCTGCAGAAGGAACGGCTGGAAAACGAAG GTCACGTGCGCGGAGTGGAAAACCGGTTCCTGACAAAGGACATTGTGCG CGAAGCGAAGAAGATCATTGATACCCAAATGCAGTACTACCCTGAAATC GACGAGACTTTCAAGGAAAAGTACATTTCCCTGGTGGAAACCCGGCGGG AATACTTCGAAGGCCCCGGACAGGGATCGCCGTTCGGATGGAACGGGGA CCTCAAGAAGTGGTACGAGATGCTGATGGGGCACTGTACCTACTTTCCGC AAGAACTGCGCTCCGTGAAGTACGCGTACTCCGCGGATCTCTTCAACGCG TTGAATGACCTGAACAACCTGATCATTCAGAGAGACAATTCCGAAAAGCT CGAGTACCACGAGAAGTATCACATCATCGAGAATGTGTTCAAGCAGAAG AAGAAACCGACCCTCAAGCAAATCGCCAAGGAGATTGGCGTCAACCCAG AGGACATCAAGGGATATCGGATTACCAAGAGCGGCACTCCCGAGTTTAC CTCTTTCAAGCTGTTTCATGATCTGAAGAAAGTCGTGAAGGACCATGCCA TTCTCGACGACATTGATCTCCTGAATCAGATCGCAGAGATCCTGACTATC TACCAAGACAAGGACTCGATTGTGGCAGAGCTGGGTCAGCTCGAATACCT GATGTCCGAGGCCGACAAGCAGTCCATCTCCGAACTGACAGGGTACACG GGGACTCATAGCCTGTCGCTGAAGTGCATGAACATGATCATTGATGAACT GTGGCACAGCTCCATGAACCAAATGGAAGTGTTTACCTACCTCAACATGC GCCCTAAGAAGTACGAACTGAAAGGCTACCAGCGCATCCCCACCGACAT GATCGACGACGCGATCTTGTCCCCTGTGGTCAAGAGGACCTTCATTCAAT CCATCAACGTGATCAACAAGGTCATCGAAAAGTACGGAATACCAGAAGA TATCATCATTGAGCTCGCTCGGGAGAACAACTCGGATGACCGGAAGAAG TTCATCAACAATCTTCAGAAGAAGAACGAAGCGACTCGGAAACGGATCA ACGAGATCATCGGACAGACCGGAAACCAGAACGCCAAACGGATTGTCGA AAAGATTAGACTGCACGACCAGCAGGAAGGGAAGTGCCTGTACTCACTC GAGTCAATACCGCTCGAGGACCTGTTGAACAACCCTAACCACTATGAAGT GGACCACATCATCCCTCGGTCCGTGAGCTTCGACAACTCGTACCACAACA AAGTGCTCGTGAAGCAGTCCGAAAACTCCAAGAAATCCAACCTGACCCC GTACCAATACTTCAATTCGGGAAAGTCGAAGCTGTCGTACAACCAGTTCA AACAACACATACTCAACCTTAGCAAAAGCCAGGATCGCATTTCCAAGAA GAAGAAGGAATACCTCCTCGAGGAAAGGGACATCAACAAGTTCGAAGTG CAGAAAGAGTTCATCAATCGCAACTTGGTGGATACCAGATATGCCACCCG GGAACTCACCAACTATCTCAAGGCCTACTTTTCCGCCAACAACATGAACG TGAAGGTCAAGACCATCAACGGGTCCTTCACTGACTACCTGAGAAAGGTC TGGAAGTTCAAGAAGGAACGCAACCACGGATACAAGCACCACGCTGAGG ACGCTCTGATCATCGCCAATGCCGACTTCCTGTTCAAGGAAAACAAGAAG CTGAAAGCTGTCAACTCAGTGCTGGAAAAGCCTGAAATCGAGACTAAGC AGCTGGATATCCAAGTGGACTCTGAGGACAACTACAGCGAGATGTTCATC ATCCCTAAACAAGTGCAGGATATCAAGGACTTTCGCAACTTCAAGTACTC ACACCGGGTGGACAAGAAACCGAATAGACAGCTGATCAACGACACGTTG TATTCCACCCGGAAGAAGGATAACTCAACCTACATTGTGCAGACTATCAA GGATATCTACGCCAAAGATAACACTACTCTGAAGAAACAATTCGACAAG	密码子优化的 **Gib11Spa-1** 多核苷酸

[0700]

	TCCCCAGAGAAGTTCCTGATGTACCAGCACGACCCCCGAACCTTTGAGAA GCTTGAAGTGATCATGAAGCAGTACGCCAACGAGAAGAACCCGCTGGCC AAGTACCATGAAGAAACCGGAGAATACCTGACCAAGTACAGCAAGAAGA ACAACGGTCCCATTGTCAAGAGCCTGAAGTACATCGGCAACAAGCTGGG ATCCCACCTCGACGTGACACATCAGTTCAAGTCGTCGACTAAGAAGCTTG TGAAGCTGTCAATCAAGAACTATAGATTCGACGTGTACTTGACCGAAAAG GGATACAAGTTCGTGACCATAGCCTATCTGAACGTGTTCAAGAAAGATAA CTACTACTACATCCCCAAGGACAAGTACCAGGAGCTCAAAGAAAAGAAG AAGATCAAAGACACCGACCAGTTCATTGCCTCCTTCTACAAGAACGACCT GATCAAACTGAACGGCGACCTCTACAAGATCATTGGAGTGAACAGCGAT GACAGGAACATCATTGAGCTGGACTACTACGACATCAAGTACAAGGACT ACTGCGAGATCAACAACATCAAGGGCGAACCCCGGATCAAGAAAACTAT TGGAAAGAAAACCGAGTCCATTGAGAAGTTCACCACTGACGTGCTGGGA AACCTTTACCTCCACTCCACCGAGAAGGCACCACAACTGATCTTCAAGCG CGGCCTG	
16	ATGAACCAGAAATTTATCCTGGGTCTGGATATTGGTATTACCAGCGTTGG TTATGGCCTGATTGATTACGAAACCAAAAACATTATTGATGCCGGTGTTC GTCTGTTTCCGGAAGCAAATGTTGAAAATAATGAAGGTCGTCGTAGCAAA CGTGGTAGCCGTCGTCTGAAACGTCGTCGTATTCATCGTCTGGAACGTGT TAAACTGCTGCTGACCGAATATGATCTGATTAACAAAGAGCAGATTCCGA CCAGCAATAACCCGTATCAGATTCGTGTTAAAGGTCTGAGCGAAATCCTG AGCAAAGATGAACTGGCAATTGCACTGCTGCATCTGGCAAAACGCCGTG GCATTCATAATGTTGATGTTGCAGCAGATAAAGAAGAAACCGCAAGCGA TAGCCTGAGCACCAAAGATCAGATTAACAAAAACGCCAAATTTCTGGAA AGCCGCTATGTTTGTGAACTGCAGAAAGAACGTCTGGAAAATGAAGGTC ATGTTCGTGGTGTTGAAAATCGCTTTCTGACGAAAGATATTGTGCGTGAG GCCAAAAAAAATCATCGATACCCAGATGCAGTATTACCCGGAAATTGATG AAACCTTCAAAGAGAAATATATCAGCCTGGTTGAAACCCGTCGCGAATAT TTTGAAGGTCCTGGTCAGGGTAGCCCGTTTGGTTGGAATGGTGATCTGAA AAAATGGTACGAAATGCTGATGGGTCACTGTACCTATTTTCCGCAAGAAC TGCGTAGCGTTAAATATGCCTATAGCGCAGACCTGTTTAATGCACTGAAT GATCTGAACAACCTGATTATTCAGCGCGATAATAGCGAGAAACTGGAAT ACCATGAGAAGTATCACATCATCGAGAACGTGTTCAAGCAGAAAAAAAA GCCGACGCTGAAACAAATCGCAAAAGAGATTGGCGTTAACCCGGAAGAT ATTAAAGGTTATCGTATTACCAAAAGCGGCACACCGGAGTTTACATCCTT TAAACTGTTCCACGATCTGAAAAAAGTGGTGAAAGATCATGCCATCCTGG ATGATATTGATCTGCTGAATCAGATTGCAGAAATCCTGACCATCTATCAG GATAAAGATAGCATTGTTGCAGAACTGGGTCAGCTGGAATATCTGATGAG CGAAGCCGATAAACAGAGCATTAGCGAACTGACCGGTTATACCGGTACA CATAGCCTGTCACTGAAATGCATGAACATGATTATCGATGAACTGTGGCA TAGCAGCATGAACCAGATGGAAGTTTTTACCTATCTGAATATGCGTCCGA AAAAGTATGAGCTGAAAGGTTATCAGCGTATTCCGACCGATATGATTGAT GATGCAATTCTGAGTCCGGTTGTGAAACGCACCTTTATTCAGAGCATCAA CGTGATCAACAAAGTGATCGAGAAATATGGCATCCCCGAAGATATCATTA TCGAACTGGCACGTGAAAATAACTCCGATGATCGCAAAAAGTTCATCAAC AACCTGCAGAAAAAGAATGAAGCAACCCGCAAACGCATTAACGAAATTA TTGGTCAGACCGGTAATCAGAATGCCAAACGTATTGTGGAAAAAATCCGT CTGCATGATCAGCAAGAGGGCAAATGTCTGTATAGCCTGGAAAGCATTCC TCTGGAAGATCTGCTGAACAATCCGAATCATTATGAAGTGGATCACATTA TTCCGCGTAGCGTGAGCTTTGATAATTCCTATCATAATAAAGTGCTGGTG AAACAGAGCGAAAACTCCAAAAAATCCAACCTGACACCGTATCAGTATT TCAATAGCGGCAAATCCAAACTGAGCTACAACCAGTTTAAACAGCATATT CTGAACCTGAGCAAAAGCCAGGATCGCATCAGCAAGAAGAAGAAGGAGT ACCTGCTGGAAGAACGCGATATTAACAAATTTGAAGTGCAGAAAGAATT TATCAACCGCAACCTGGTTGATACCCGTTATGCAACCCGTGAACTGACCA ATTATCTGAAAGCATATTTCAGCGCCAACAACATGAACGTGAAAGTGAA AACGATTAACGGCAGCTTTACCGATTATCTGCGTAAAGTGTGGAAATTCA AAAAAGAACGCAACCACGGCTATAAACATCATGCCGAAGATGCCCTGAT TATTGCAAATGCAGATTTCCTGTTTAAAGAAAACAAAAAACTGAAAGCCG TCAACAGCGTGCTGGAAAAACCGGAAATTGAGACAAAACAGCTGGACAT TCAGGTTGATAGCGAAGATAATTACAGCGAAATGTTTATCATCCCGAAAC	编码Gib11Spa-2的DNA序列

[0701]

	AGGTGCAGGATATCAAAGATTTTCGCAACTTCAAATTCAGCCACCGCGTT GATAAAAAACCGAATCGTCAGCTGATTAACGATACCCTGTATAGCACCCG TATGAAAGATGAGCATGATTATATTGTGCAGACCATCACGGATATCTATG GCAAAGATAATACCAACCTGAAAAAACAGTTCAACAAAAACCCGGAAAA ATTTCTGATGTATCAGAACGATCCGAAAACCTTTGAGAAACTGAGCATCA TCATGAAACAGTACAGCGACGAAAAAAACCCGCTGGCCAAATATTACGA AGAAACCGGTGAATATCTGACCAAATATAGCAAGAAAAACAACGGTCCG ATCGTGAAAAAGATCAAACTGCTGGGTAATAAAGTGGGCAATCATCTGG ATGTGACCAACAAATATGAAAACTCCACGAAGAAGCTGGTTAAGCTGTC CATTAAAAACTATCGCTTCGATGTGTATCTGACCGAGAAAGGTTATAAGT TTGTGACCATTGCCTACCTGAATGTGTTCAAAAAAGACAACTATTACTAT ATTCCGAAAGACAAATACCAAGAACTTAAAGAGAAGAAGAAAATCAAGG ACACCGATCAGTTTATCGCCAGCTTCTATAAAAACGATCTGATCAAGCTG AACGGCGACCTGTATAAAATCATTGGTGTGAATAGTGATGACCGCAACAT CATTGAGCTGGATTATTACGACATCAAATACAAGGATTACTGCGAGATCA ACAACATTAAAGGTGAACCGCGTATCAAAAAGACCATTGGCAAAAAAAC GGAAAGCATCGAAAAGTTTACCACCGATGTTCTGGGTAATCTGTATCTGC ATAGTACCGAAAAAGCACCGCAGCTGATTTTCAAACGCGGTCTG	
17	ATGAACCAGAAATTTATCCTGGGTCTGGATATTGGTATTACCAGCGTTGG TTATGGCCTGATTGATTACGAAACCAAAAACATTATTGATGCCGGTGTTC GTCTGTTTCCGGAAGCAAATGTTGAAAATAATGAAGGTCGTCGTAGCAAA CGTGGTAGCCGTCGTCTGAAACGTCGTCGTATTCATCGTCTGGAACGTGT TAAACTGCTGCTGACCGAATATGATCTGATTAACAAAGAGCAGATTCCGA CCAGCAATAACCCGTATCAGATTCGTGTTAAAGGTCTGAGCGAAATCCTG AGCAAAGATGAACTGGCAATTGCACTGCTGCATCTGGCAAAACGCCGTG GCATTCATAATGTTGATGTTGCAGCAGATAAAGAAGAAACCGCAAGCGA TAGCCTGAGCACCAAAGATCAGATTAACAAAAACGCCAAATTTCTGGAA AGCCGCTATGTTTGTGAACTGCAGAAAGAACGTCTGGAAAATGAAGGTC ATGTTCGTGGTGTTGAAAATCGCTTTCTGACGAAAGATATTGTGCGTGAG GCCAAAAAAATCATCGATACCCAGATGCAGTATTACCCGGAAATTGATG AAACCTTCAAAGAGAAATATATCAGCCTGGTTGAAACCCGTCGCGAATAT TTTGAAGGTCCTGGTCAGGGTAGCCCGTTTGGTTGGAATGGTGATCTGAA AAAATGGTACGAAATGCTGATGGGTCACTGTACCTATTTTCCGCAAGAAC TGCGTAGCGTTAAATATGCCTATAGCGCAGACCTGTTTAATGCACTGAAT GATCTGAACAACCTGATTATTCAGCGCGATAATAGCGAGAAACTGGAAT ACCATGAGAAGTATCACATCATCGAGAACGTGTTCAAGCAGAAAAAAAA GCCGACGCTGAAACAAATCGCAAAAGAGATTGGCGTTAACCCGGAAGAT ATTAAAGGTTATCGTATTACCAAAAGCGGCACACCGGAGTTTACATCCTT TAAACTGTTCCACGATCTGAAAAAAGTGGTGAAAGATCATGCCATCCTGG ATGATATTGATCTGCTGAATCAGATTGCAGAAATCCTGACCATCTATCAG GATAAAGATAGCATTGTTGCAGAACTGGGTCAGCTGGAATATCTGATGAG CGAAGCCGATAAACAGAGCATTAGCGAACTGACCGGTTATACCGGTACA CATAGCCTGTCACTGAAATGCATGAACATGATTATCGATGAACTGTGGCA TAGCAGCATGAACCAGATGGAAGTTTTTACCTATCTGAATATGCGTCCGA AAAAGTATGAGCTGAAAGGTTATCAGCGTATTCCGACCGATATGATTGAT GATGCAATTCTGAGTCCGGTTGTGAAACGCACCTTTATTCAGAGCATCAA CGTGATCAACAAAGTGATCGAGAAATATGGCATCCCCGAAGATATCATTA TCGAACTGGCACGTGAAAATAACTCCGATGATCGCAAAAAGTTCATCAAC AACCTGCAGAAAAAGAATGAAGCAACCCGCAAACGCATTAACGAAATTA TTGGTCAGACCGGTAATCAGAATGCCAAACGTATTGTGGAAAAAATCCGT CTGCATGATCAGCAAGAGGGCAAATGTCTGTATAGCCTGGAAAGCATTCC TCTGGAAGATCTGCTGAACAATCCGAATCATTATGAAGTGGATCACATTA TTCCGCGTAGCGTGAGCTTTGATAATTCCTATCATAATAAAGTGCTGGTG AAACAGAGCGAAAACTCCAAAAAATCCAACCTGACACCGTATCAGTATT TCAATAGCGGCAAATCCAAACTGAGCTACAACCAGTTTAAACAGCATATT CTGAACCTGAGCAAAAGCCAGGATCGCATCAGCAAGAAGAAGAAGGAGT ACCTGCTGGAAGAACGCGATATCAACAAATTTGAAGTCCAGAAAGAGTT TATCAACCGCAATCTGGTTGATACCCGTTATGCAACCCGTGAACTGACCA GCTATCTGAAAGCATATTTCAGCGCCAATAACATGGACGTGAAAGTGAA AACAATTAACGGCAGCTTTACCAACCATCTGCGTAAAGTTTGGCGCTTTG ATAAATATCGCAACCACGGCTATAAACATCATGCCGAAGATGCACTGATT	编码Gib11Spa-3的DNA序列

[0702]

	ATTGCCAATGCAGATTTCCTGTTCAAAGAAAACAAAAAACTGCAGAACA CCAACAAGATCCTGGAAAAACCGACCATTGAAAACAACACCAAAAAAGT GACCGTCGAGAAAGAAGAGGATTACAACAACGTTTTTGAAACCCCGAAA CTGGTCGAGGATATTAAACAGTATCGCGACTATAAATTCAGCCACCGCGT TGATAAAAAACCGAATCGTCAGCTGATTAACGATACCCTGTATAGCACCC GTATGAAAGATGAGCATGATTATATTGTGCAGACCATCACGGATATCTAT GGCAAAGATAATACCAACCTGAAAAAACAGTTCAACAAAAACCCGGAAA AATTTCTGATGTATCAGAACGATCCGAAAACCTTTGAGAAACTGAGCATC ATCATGAAACAGTACAGCGACGAAAAAAACCCGCTGGCCAAATATTACG AAGAAACCGGTGAATATCTGACCAAATATAGCAAGAAAAACAACGGTCC GATCGTGAAAAAGATCAAACTGCTGGGTAATAAAGTGGGCAATCATCTG GATGTGACCAACAAATATGAAAACTCCACGAAGAAGCTGGTTAAGCTGT CCATTAAAAACTATCGCTTCGATGTGTATCTGACCGAGAAAGGTTATAAG TTTGTGACCATTGCCTACCTGAATGTGTTCAAAAAAGACAACTATTACTA TATTCCGAAAGACAAATACCAAGAACTTAAAGAGAAGAAGAAAATCAAG GACACCGATCAGTTTATCGCCAGCTTCTATAAAAACGATCTGATCAAGCT GAACGGCGACCTGTATAAAATCATTGGTGTGAATAGTGATGACCGCAAC ATCATTGAGCTGGATTATTACGACATCAAATACAAGGATTACTGCGAGAT CAACAACATTAAAGGTGAACCGCGTATCAAAAAGACCATTGGCAAAAAA ACGGAAAGCATCGAAAAGTTTACCACCGATGTTCTGGGTAATCTGTATCT GCATAGTACCGAAAAAGCACCGCAGCTGATTTTCAAACGCGGTCTG	
18	ATGAACCAGAAGTTCATTCTCGGTCTGGACATTGGCATTACTAGCGTGGG ATACGGCTTGATTGACTACGAGACTAAGAACATCATCGATGCCGGCGTCC GCCTGTTCCCGGAAGCCAACGTGGAGAACAATGAGGGCCGGAGGTCGAA GAGAGGCTCCCGCCGCCTGAAGCGGCGGCGAATCCACCGGCTGGAGAGA GTGAAGCTGCTGCTCACCGAATACGACCTGATCAACAAAGAACAGATCC CGACCTCCAACAACCCGTACCAGATCAGAGTGAAGGGTCTGTCAGAAAT CCTGTCCAAGGACGAACTGGCAATCGCCCTGCTGCACCTGGCCAAGCGGC GCGGAATCCACAACGTGGATGTGGCTGCCGACAAGGAAGAAACCGCTTC CGACTCCCTGAGCACTAAGGACCAGATCAACAAGAACGCCAAGTTCTTG GAGTCCCGCTACGTGTGCGAGCTGCAGAAGGAACGGCTGGAAAACGAAG GTCACGTGCGCGGAGTGGAAAACCGGTTCCTGACAAAGGACATTGTGCG CGAAGCGAAGAAGATCATTGATACCCAAATGCAGTACTACCCTGAAATC GACGAGACTTTCAAGGAAAAGTACATTTCCCTGGTGGAAACCCGGCGGG AATACTTCGAAGGCCCCGGACAGGGATCGCCGTTCGGATGGAACGGGGA CCTCAAGAAGTGGTACGAGATGCTGATGGGGCACTGTACCTACTTTCCGC AAGAACTGCGCTCCGTGAAGTACGCGTACTCCGCGGATCTCTTCAACGCG TTGAATGACCTGAACAACCTGATCATTCAGAGAGACAATTCCGAAAAGCT CGAGTACCACGAGAAGTATCACATCATCGAGAATGTGTTCAAGCAGAAG AAGAAACCGACCCTCAAGCAAATCGCCAAGGAGATTGGCGTCAACCCAG AGGACATCAAGGGATATCGGATTACCAAGAGCGGCACTCCCGAGTTTAC CTCTTTCAAGCTGTTTCATGATCTGAAGAAAGTCGTGAAGGACCATGCCA TTCTCGACGACATTGATCTCCTGAATCAGATCGCAGAGATCCTGACTATC TACCAAGACAAGGACTCGATTGTGGCAGAGCTGGGTCAGCTCGAATACCT GATGTCCGAGGCCGACAAGCAGTCCATCTCCGAACTGACAGGGTACACG GGGACTCATAGCCTGTCGCTGAAGTGCATGAACATGATCATTGATGAACT GTGGCACAGCTCCATGAACCAAATGGAAGTGTTTACCTACCTCAACATGC GCCCTAAGAAGTACGAACTGAAAGGCTACCAGCGCATCCCCACCGACAT GATCGACGACGCGATCTTGTCCCCTGTGGTCAAGAGGACCTTCATTCAAT CCATCAACGTGATCAACAAGGTCATCGAAAAGTACGGTATTCCAGAAGA TATCATCATTGAGCTCGCTCGGGAGAACAACTCGGATGACCGGAAGAAG TTCATCAACAATCTTCAGAAGAAGAACGAAGCGACTCGGAAACGGATCA ACGAGATCATCGGACAGACCGGAAACCAGAACGCCAAACGGATTGTCGA AAAGATTAGACTGCACGACCAGCAGGAAGGGAAGTGCCTGTACTCACTC GAGTCAATACCGCTCGAGGACCTGTTGAACAACCCTAACCACTATGAAGT GGACCACATCATCCCTCGGTCCGTGAGCTTCGACAACTCGTACCACAACA AAGTGCTCGTGAAGCAGTCCGAAAACTCCAAGAAATCCAACCTGACCCC GTACCAATACTTCAATTCGGGAAAGTCGAAGCTGTCGTACAACCAGTTCA AACAACACATACTCAACCTTAGCAAAAGCCAGGATCGCATTTCCAAGAA GAAGAAGGAATACCTCCTCGAGGAAAGGGACATCAACAAGTTCGAAGTG CAGAAAGAGTTCATCAATCGCAACTTGGTGGATACCAGATATGCCACCCG	密码子优化的 **Gib11Spa-3** 多核苷酸

[0703]

	GGAACTCACCAGCTATCTCAAGGCCTACTTTTCCGCCAACAACATGGACG TGAAGGTCAAGACCATCAACGGGTCCTTCACTAACCATCTGAGAAAGGTC TGGCGGTTTGACAAGTACCGCAACCACGGATACAAGCACCACGCTGAAG ACGCTCTGATCATCGCCAATGCCGACTTCCTGTTCAAGGAAAACAAGAAG CTGCAGAACACGAACAAGATTCTGGAAAAGCCTACCATTGAGAACAACA CTAAGAAGGTCACCGTGGAGAAGGAAGAGGACTACAACAACGTGTTCGA AACTCCTAAACTGGTGGAGGATATCAAGCAATACCGCGACTACAAGTTCT CACACCGGGTGGACAAGAAACCGAATAGACAGCTGATCAACGACACGTT GTATTCCACCCGGATGAAGGATGAGCATGACTACATTGTGCAGACTATCA CCGATATCTACGGAAAAGATAACACTAACCTGAAGAAACAATTCAACAA GAACCCAGAGAAGTTCCTGATGTACCAGAACGACCCCAAGACCTTTGAG AAGCTTTCCATCATCATGAAGCAGTACTCCGACGAGAAGAACCCGCTGGC CAAGTACTACGAAGAAACCGGAGAATACCTGACCAAGTACAGCAAGAAG AACAACGGTCCCATTGTCAAGAAGATCAAGCTGCTCGGCAACAAGGTCG GAAACCACCTCGACGTGACAAACAAGTACGAGAACTCGACTAAGAAGCT TGTGAAGCTGTCAATCAAGAACTATAGATTCGACGTGTACTTGACCGAAA AGGGATACAAGTTCGTGACCATAGCCTATCTGAACGTGTTCAAGAAAGAT AACTACTACTACATCCCCAAGGACAAGTACCAGGAGCTCAAAGAAAAGA AGAAGATCAAAGACACCGACCAGTTCATTGCCTCCTTCTACAAGAACGAC CTGATCAAACTGAACGGCGACCTCTACAAGATCATTGGAGTGAACAGCG ATGACAGGAACATCATTGAGCTGGACTACTACGACATCAAGTACAAGGA CTACTGCGAGATCAACAACATCAAGGGCGAACCCCGGATCAAGAAAACT ATTGGAAAGAAAACCGAGTCCATTGAGAAGTTCACCACTGACGTGCTGG GAAACCTTTACCTCCACTCCACCGAGAAGGCACCACAACTGATCTTCAAG CGCGGCCTG	
19	ATGAACCAGAAATTTATCCTGGGTCTGGATATTGGTATTACCAGCGTTGG TTATGGCCTGATTGATTACGAAACCAAAAACATTATTGATGCCGGTGTTC GTCTGTTTCCGGAAGCAAATGTTGAAAATAATGAAGGTCGTCGTAGCAAA CGTGGTAGCCGTCGTCTGAAACGTCGTCGTATTCATCGTCTGGAACGTGT TAAAAAACTGCTGGAAGATTATAACCTGCTGGATCAGAGCCAGATTCCGC AGAGCACCAATCCGTATGCAATTCGTGTTAAAGGTCTGAGCGAAGCACTG AGCAAAGATGAACTGGTTATTGCACTGCTGCATATTGCAAAACGCCGTGG CATTCATAATATCAATGTTAGCAGCGAAGATGAGGATGCAAGCAATGAA CTGAGCACCAAAGAACAAATTAACCGCAATAATAAGCTGCTGAAGGACA AATATGTTTGCGAAGTTCAGCTGCAGCGTCTGAAAGAAGGTCAGATTCGC GGAGAAAAAAATCGCTTTAAAACCACCGATATCCTGAAAGAAATTGATC AGCTGCTTAAAGTGCAGAAGGATTATCATAACCTGGACATCGATTTCATC AACCAGTACAAAGAAATCGTTGAAACCCGTCGCGAATATTTTGAAGGTCC GGGTAAAGGTAGCCCGTATGGTTGGGAAGGTGATCCGAAAGCATGGTAT GAAACCCTGATGGGTCATTGTACCTATTTTCCGGATGAACTGCGTAGCGT TAAATATGCCTATAGCGCAGACCTGTTTAATGCACTGAATGATCTGAATA ACCTGGTGATTCAGCGTGATGGTCTGAGCAAACTGGAATATCATGAGAAA TATCACATCATCGAAAACGTGTTCAAACAGAAGAAGAAACCGACCCTGA AACAAATCGCCAACGAAATTAATGTGAACCCGGAAGATATTAAAGGCTA CCGTATTACCAAAAGCGGCAAACCGGAATTTACATCCTTTAAACTGTTCC ACGATCTGAAAAAAGTGGTGAAAGATCATGCCATCCTGGATGATATTGAT CTGCTGAATCAGATTGCAGAAATCCTGACCATCTATCAGGATAAAGATAG CATTGTTGCAGAACTGGGTCAGCTGGAATATCTGATGAGCGAAGCCGATA AACAGAGCATTAGCGAACTGACCGGTTATACCGGTACACATAGCCTGTCA CTGAAATGCATGAACATGATTATCGATGAACTGTGGCATAGCAGCATGAA CCAGATGGAAGTTTTTACCTATCTGAATATGCGTCCGAAAAAGTATGAGC TGAAAGGTTATCAGCGTATTCCGACCGATATGATTGATGATGCAATTCTG AGTCCGGTTGTGAAACGCACCTTTATTCAGAGCATCAACGTGATCAACAA AGTGATCGAGAAATATGGCATCCCCGAAGATATCATTATCGAACTGGCAC GTGAAAATAACTCCGATGATCGCAAAAAGTTCATCAACAACCTGCAGAA AAAGAATGAAGCAACCCGCAAACGCATTAACGAAATTATTGGTCAGACC GGTAATCAGAATGCCAAACGTATTGTGGAAAAAATCCGTCTGCATGATCA GCAAGAGGGTAAATGTCTGTATAGCCTGGAAAGCATTCCTCTGGAAGATC TGCTGAACAATCCGAATCATTATGAAGTGGATCACATTATTCCGCGTAGC GTGAGCTTTGATAATTCCTATCATAATAAAGTGCTGGTGAAACAGAGCGA AAACTCCAAAAAATCCAACCTGACACCGTATCAGTATTTCAATAGCGGCA	编码P2H12的DNA序列

	AATCCAAACTGAGCTACAACCAGTTTAAACAGCATATTCTGAACCTGAGC AAAAGCCAGGATCGCATCAGCAAGAAGAAGAAGGAGTACCTGCTGGAAG AACGCGATATTAACAAATTTGAAGTGCAGAAAGAATTTATCAACCGCAA CCTGGTTGATACCCGTTATGCAACCCGTGAACTGACCAATTATCTGAAAG CATATTTCAGCGCCAACAACATGAACGTGAAAGTGAAAACGATTAACGG CAGCTTTACCGATTATCTGCGTAAAGTGTGGAAATTCAAAAAAGAACGCA ACCACGGCTATAAACATCATGCCGAAGATGCCCTGATTATTGCAAATGCA GATTTCCTGTTTAAAGAAAACAAAAAACTGAAAGCCGTCAACAGCGTGCT GGAAAAACCGGAAATTGAGACAAAACAGCTGGACATTCAGGTTGATAGC GAAGATAATTACAGCGAAATGTTTATCATCCCGAAACAGGTGCAGGATAT CAAAGATTTTCGCAACTTCAAATATAGCCACCGCGTTGACAAAAAACCTA ATCGTCAGCTGATTAACGATACCCTGTATAGCACCCGCAAAAAAGATAAC AGCACCTATATTGTGCAGACCATTAAAGACATCTACGCCAAAGATAATAC CACCCTGAAAAAACAGTTCGACAAAAGCCCAGAAAAATTTCTGATGTATC AGCATGATCCGCGTACCTTCGAAAAACTGGAAGTTATTATGAAACAGTAT GCCAACGAGAAAAATCCGCTGGCCAAATATCACGAAGAAACCGGTGAAT ATCTGACCAAATATTCCAAGAAGAACAACGGTCCGATCGTTAAATCCCTG AAATATATCGGTAATAAACTGGGCAGCCATCTGGATGTTACCCATCAGTT TAAAAGCTCCACAAAGAAGCTGGTTAAACTGTCCATCAAACCGTATCGCT TTGATGTGTATCTGACCGACAAAGGCTATAAATTCATTACCATCAGCTAT CTGGACGTGCTGAAAAAAGACAACTATTATTATATCCCGGAACAGAAAT ATGATAAACTGAAACTGGGTAAAGCCATCGATAAAAACGCCAAATTTAT CGCCAGCTTCTACAAAAACGACCTGATTAAACTGGATGGCGAGATCTATA AAATCATCGGTGTTAATAGCGACACCCGCAATATGATTGAGCTGGATCTG CCGGATATTCGCTATAAAGAATATTGCGAACTGAACAACATTAAAGGCG AACCGCGTATCAAAAAGACCATCGGCAAAAAAGTGAATAGCATCGAGAA ACTGACCACCGATGTTCTGGGTAATGTGTTTACCAATACCCAGTATACCA AACCTCAGCTGCTGTTCAAACGCGGTAATGGTG	
20	ATGAACCAAAAGTTCATTCTGGGGCTCGATATCGGCATCACCTCCGTGGG ATATGGTCTGATCGACTACGAGACTAAGAACATCATCGACGCTGGAGTGC GACTGTTCCCGGAAGCGAACGTGGAGAACAACGAAGGCCGCAGATCCAA GCGCGGGTCCAGAAGGCTCAAGAGGCGGAGGATCCATAGACTCGAGAGA GTGAAGAAGCTCCTTGAAGATTACAATCTGTTGGACCAGTCACAGATTCC CCAAAGCACCAACCCGTACGCCATCAGAGTGAAGGGCCTGTCAGAAGCA CTGTCGAAAGATGAACTGGTCATTGCCCTGCTGCATATTGCCAAACGGCG CGGAATCCATAACATCAACGTGTCGAGCGAAGATGAGGACGCGTCCAAC GAACTGTCAACCAAGGAACAGATCAACCGGAACAACAAACTGCTGAAGG ACAAATACGTCTGCGAGGTGCAGCTTCAACGGCTGAAAGAGGGACAGAT CAGGGGAGAGAAAAACCGGTTCAAGACCACCGACATCCTTAAGGAGATC GACCAACTCCTGAAAGTGCAGAAGGACTATCACAACCTCGACATTGATTT TATCAACCAGTACAAGGAGATTGTGGAAACTCGGAGGGAGTACTTCGAG GGACCTGGAAAGGGATCCCCTTATGGCTGGGAAGGGGACCCCAAGGCTT GGTACGAAACCCTGATGGGCCATTGCACTTACTTTCCGGATGAACTCCGG TCCGTGAAGTACGCTTACTCTGCCGACCTGTTCAATGCACTCAACGACCTT AACAACTTGGTGATCCAACGCGATGGTCTTTCCAAGTTGGAGTACCACGA AAAGTACCACATCATCGAGAACGTGTTCAAGCAGAAAAAGAAGCCAACT CTGAAGCAGATTGCCAACGAAATTAACGTGAACCCCGAGGATATCAAGG GATACCGGATTACCAAGTCCGGCAAACCAGAGTTCACCTCATTCAAGCTG TTTCACGATCTGAAGAAGGTCGTGAAGGACCACGCCATCCTGGATGACAT TGATCTTCTGAACCAGATTGCTGAGATCCTGACCATCTACCAGGACAAGG ACTCGATTGTGGCCGAACTGGGACAGCTCGAGTACCTGATGTCCGAAGCC GATAAGCAGTCCATCAGCGAACTCACCGGTTACACCGGTACCCACTCCTT GTCCCTTAAGTGCATGAACATGATCATTGACGAACTGTGGCACTCCAGCA TGAACCAGATGGAGGTGTTCACCTACTTGAACATGCGCCCGAAGAAGTAC GAGCTGAAGGGCTACCAGCGCATACCCACGGACATGATCGACGACGCCA TCCTCTCACCGGTGGTCAAGCGCACCTTCATTCAATCTATCAACGTGATCA ACAAGGTCATCGAGAAGTACGGCATTCCTGAGGATATCATCATCGAGCTG GCTCGGGAGAACAACTCAGACGATAGGAAGAAGTTCATTAACAACCTCC AGAAAAAGAACGAGGCCACTCGCAAGCGGATTAATGAGATCATCGGTCA GACCGGGAACCAGAACGCCAAGCGGATCGTGGAAAAGATTCGGCTCCAC GACCAACAGGAGGGAAAGTGTCTGTACTCGCTGGAGTCCATTCCCCTGGA	密码子优化的 P2H12多核苷酸

[0704]

[0705]

	GGACCTCCTGAACAACCCAAACCACTACGAGGTCGATCACATAATCCCCC GCAGCGTGTCATTCGACAACTCCTACCATAACAAGGTCCTCGTGAAGCAG TCGGAGAATAGCAAGAAGTCGAACCTGACTCCGTACCAGTACTTCAACTC CGGCAAATCCAAGCTGTCCTACAATCAGTTCAAACAGCACATTCTCAACC TGTCCAAGAGCCAGGACAGGATTTCGAAGAAGAAGAAGGAATACCTTCT CGAGGAACGGGATATCAATAAGTTCGAGGTGCAGAAGGAGTTTATCAAT AGAAACCTGGTGGACACTCGCTATGCCACCCGCGAACTGACCAACTACCT GAAGGCGTACTTCTCCGCCAACAACATGAACGTGAAGGTCAAAACTATTA ACGGCAGCTTCACCGACTATCTGCGCAAGGTCTGGAAGTTCAAGAAGGA ACGCAACCACGGTTACAAGCACCACGCGGAAGATGCGCTGATTATCGCC AACGCTGACTTCCTGTTCAAGGAAAACAAGAAGCTCAAGGCCGTGAACT CAGTGCTCGAGAAGCCTGAAATCGAGACTAAGCAGCTGGACATCCAGGT CGATTCGGAAGATAACTACTCCGAAATGTTCATCATCCCTAAGCAAGTGC AGGACATCAAGGACTTCAGGAATTTCAAGTACAGCCATCGCGTGGACAA GAAGCCAAACAGACAGCTGATCAACGATACACTGTATTCCACCCGGAAG AAGGACAACTCCACCTACATCGTCCAAACCATTAAGGACATCTACGCAAA GGACAACACCACGCTTAAGAAGCAGTTCGACAAGAGCCCCGAAAAGTTC CTCATGTACCAGCACGACCCCAGAACCTTCGAGAAGCTTGAAGTGATCAT GAAGCAGTACGCCAACGAAAAGAACCCACTGGCTAAGTACCACGAGGAA ACCGGCGAATACCTGACCAAGTACTCCAAAAAGAACAACGGACCGATCG TCAAGTCCCTGAAGTACATTGGGAACAAGCTCGGCTCGCACCTCGATGTG ACCCACCAGTTCAAGTCCTCGACCAAAAAGCTCGTGAAGCTGTCCATCAA GCCGTACCGGTTCGACGTGTACCTGACTGACAAGGGATATAAGTTCATCA CCATTTCCTACCTCGACGTGTTGAAGAAGGATAACTACTACTACATTCCG GAACAGAAGTACGACAAGCTCAAGCTCGGAAAGGCCATCGACAAAAATG CGAAGTTCATCGCGAGCTTCTACAAGAATGACTTGATCAAGCTGGATGGC GAAATCTACAAGATCATCGGGGTCAACTCCGATACCCGCAACATGATTGA GCTGGATCTGCCCGACATTCGGTACAAGGAATACTGCGAGCTGAACAAC ATCAAGGGAGAACCGCGGATCAAGAAAACCATCGGAAAGAAAGTGAAC AGCATCGAGAAACTGACTACTGACGTCCTGGGAAACGTGTTCACCAACAC ACAATACACCAAACCCCAGCTGCTGTTTAAGCGCGGGAAC	
21	ATGAACCAGAAATTTATCCTGGGTCTGGATATTGGTATTACCAGCGTTGG TTATGGCCTGATTGATTACGAAACCAAAAACATTATTGATGCCGGTGTTC GTCTGTTTCCGGAAGCAAATGTTGAAAATAATGAAGGTCGTCGTAGCAAA CGTGGTAGCCGTCGTCTGAAACGTCGTCGTATTCATCGTCTGGATCGTGTT AAACATCTGCTGGCAGAATATGATCTGCTGGATCTGACCAATATTCCGAA AAGCACCAATCCGTATCAGACCCGTGTTAAAGGTCTGAATGAAAAGCTG AGCAAAGATGAACTGGTTATTGCACTGCTGCATATTGCAAAACGCCGTGG CATTCATAACGTTGATGTTGCAGCAGATAAAGAAGAAACCGCAAGCGAT AGCCTGAGCACCAAAGATCAGATTAACAAAAACGCCAAATTTCTGGAAA GCCGCTATGTTTGTGAACTGCAGAAAGAACGTCTGGAAAATGAAGGTCAT GTTCGTGGTGTTGAAAATCGCTTTCTGACGAAAGATATTGTGCGTGAGGC CAAAAAAATCATCGATACCCAGATGCAGTATTACCCGGAAATTGATGAA ACCTTCAAAGAGAAATATATCAGCCTGGTTGAAACCCGTCGCGAATATTT TGAAGGTCCGGGTAAAGGTAGCCCGTTTGGTTGGGAAGGTAATATCAAG AAATGGTTTGAGCAGATGATGGGCCACTGTACCTATTTTCCAGAAGAACT GCGTAGCGTCAAATATAGCTATTCAGCCGAACTGTTTAACGCCCTGAATG ATCTGAATAATCTGGTGATTACCCGTGATGAAGATGCCAAACTGAATTAT GGTGAGAAATTCCAGATCATCGAAAACGTGTTCAAACAGAAGAAAACAC CGAACCTGAAACAAATCGCCATTGAAATTGGTGTGCATGAAACCGAAAT CAAAGGTTATCGTGTGAACAAAAGCGGTACACCGGAATTTACCGAATTTA AACTGTATCATGACCTGAAAAGCATCGTGTTCGATAAAAGCATTCTGGAA AATGAAGCCATCCTGGATCAGATTGCAGAAATTCTGACCATCTATCAGGA TGAGCAGAGCATTAAAGAGGAACTGAATAAACTGCCGGAAATACTGAAC GAACAGGATAAAGCAGAAATCGCCAAACTGATTGGTTATAATGGCACCC ATCGTCTGAGCCTGAAATGTATTCACCTGATTAATGAAGAACTGTGGCAG ACCAGCCGTAATCAGATGGAAATTTTCAACTACCTGAACATCAAACCGAA CAAAGTGGATCTGAGTGAGCAGAACAAAATCCCGAAAGATATGGTGAAC GACTTTATTCTGAGTCCGGTTGTGAAACGCACCTTTATTCAGAGCATCAA CGTGATCAACAAAGTGATCGAGAAATATGGCATCCCCGAAGATATCATTA TCGAACTGGCACGTGAAAATAACTCCGATGATCGCAAAAAGTTCATCAAC	编码E2的DNA序列

[0706]

	AACCTGCAGAAAAAGAATGAAGCAACCCGCAAACGCATTAACGAAATTA TTGGTCAGACCGGTAATCAGAATGCCAAACGTATTGTGGAAAAAATCCGT CTGCATGATCAGCAAGAGGGTAAATGTCTGTATAGCCTGAAAGATATCCC GCTGGAAGATCTGCTGCGCAATCCGAACAATTATGATATCGACCATATTA TTCCGCGAAGCGTGAGCTTTGATGATAGCATGCATAACAAAGTTCTGGTT CGTCGCGAACAGAATGCCAAAAAGAATAATCAGACCCCGTATCAGTATC TGACCAGTGGTTATGCAGATATCAAATACAGCGTGTTTAAGCAGCATGTT CTGAATCTGGCCGAAAATAAAGATCGCATGACCAAAAAAAAGCGCGAGT ATCTGCTGGAAGAACGCGACATTAACAAATTTGAAGTGCAGAAAGAATT TATCAACCGCAACCTGGTTGATACCCGTTATGCAACCCGTGAACTGACCA ATTATCTGAAAGCATATTTCAGCGCCAACAACATGAACGTGAAAGTGAA AACGATTAACGGCAGCTTTACCGATTATCTGCGTAAAGTGTGGAAATTCA AAAAAGAACGCAACCACGGCTATAAACATCATGCCGAAGATGCCCTGAT TATTGCAAATGCAGATTTCCTGTTTAAAGAAAACAAAAAACTGAAAGCCG TCAACAGCGTGCTGGAAAAACCGGAAATTGAGACAAAACAGCTGGACAT TCAGGTTGATAGCGAAGATAATTACAGCGAAATGTTTATCATCCCGAAAC AGGTGCAGGATATCAAAGATTTTCGCAACTTCAAATATAGCCACCGCGTT GACAAAAAACCTAATCGTCAGCTGATTAACGATACCCTGTATAGCACCCG CAAAAAAGATAACAGCACCTATATTGTGCAGACCATTAAAGACATCTAC GCCAAAGATAATACCACCCTGAAAAAACAGTTCGACAAAAGCCCAGAAA AATTTCTGATGTATCAGCATGATCCGCGTACCTTCGAAAAACTGGAAGTT ATTATGAAACAGTATGCCAACGAGAAAAATCCGCTGGCCAAATATCACG AAGAAACCGGTGAATATCTGACCAAATATTCCAAGAAGAACAACGGTCC GATCGTTAAATCCCTGAAATATATCGGTAATAAACTGGGCAGCCATCTGG ATGTTACCCATCAGTTTAAAAGCTCCACAAAGAAGCTGGTTAAACTGTCC ATCAAACCGTATCGCTTTGATGTGTATCTGACCGACAAAGGCTATAAATT CATTACCATCAGCTATCTGGACGTGCTGAAAAAAGACAACTATTATTATA TCCCGGAACAGAAATATGATAAACTGAAACTGGGTAAAGCCATCGATAA AAACGCCAAATTTATCGCCAGCTTCTACAAAAACGACCTGATTAAACTGG ATGGCGAGATCTATAAAATCATCGGTGTTAATAGCGACACCCGCAATATG ATTGAGCTGGATCTGCCGGATATTCGCTATAAAGAATATTGCGAACTGAA CAACATTAAAGGCGAACCGCGTATCAAAAAGACCATCGGCAAAAAAGTG AATAGCATCGAGAAACTGACCACCGATGTTCTGGGTAATGTGTTTACCAA TACCCAGTATACCAAACCTCAGCTGCTGTTCAAACGCGGTAAT	
22	ATGAACCAAAAGTTCATTCTGGGGCTCGATATCGGCATCACCTCCGTGGG ATATGGTCTGATCGACTACGAGACTAAGAACATCATCGACGCTGGAGTGC GACTGTTCCCGGAAGCGAACGTGGAGAACAACGAAGGCCGCAGATCCAA GCGCGGGTCCAGAAGGCTCAAGAGGCGGAGGATCCATAGACTCGACAGA GTGAAGCACCTCCTTGCCGAATACGATCTGTTGGACCTTACCAACATTCC CAAGAGCACCAACCCGTACCAAACCAGAGTGAAGGGCCTGAACGAAAAG CTGTCGAAAGATGAACTGGTCATTGCCCTGCTGCATATTGCCAAACGGCG CGGAATCCATAACGTGGACGTGGCCGCTGACAAGGAAGAGACTGCGTCC GACTCGCTGTCAACCAAGGACCAGATCAACAAGAACGCCAAATTCCTGG AAAGCCGCTACGTCTGCGAGCTTCAAAAAGAACGGCTGGAGAACGAGGG ACACGTCAGGGGAGTGGAGAACCGGTTCCTGACCAAGGACATCGTGCGG GAAGCCAAGAAGATCATCGACACCCAAATGCAGTATTATCCGGAAATTG ATGAAACTTTTAAGGAGAAGTACATTTCCCTGGTGGAAACTCGGAGGGA GTACTTCGAGGGACCTGGAAAGGGATCCCCTTTCGGCTGGGAAGGGAAC ATTAAGAAGTGGTTTGAACAGATGATGGGCCATTGCACTTACTTTCCGGA AGAACTCCGGTCCGTGAAGTACTCATACTCTGCCGAGCTGTTCAATGCAC TCAACGACCTTAACAACTTGGTGATCACCCGCGATGAAGATGCCAAGTTG AACTACGGAGAAAAGTTCCAGATCATCGAGAACGTGTTCAAGCAGAAAA AGACCCCAAATCTGAAGCAGATTGCCATCGAAATTGGCGTGCACGAGAC TGAGATCAAGGGATACCGGGTCAACAAGTCCGGCACGCCAGAGTTCACC GAGTTCAAGCTGTACCACGATCTGAAGTCGATCGTGTTTGACAAGTCCAT CCTGGAAAACGAAGCCATTCTGGACCAGATTGCTGAGATCCTGACCATCT ACCAGGACGAGCAATCGATTAAGGAAGAACTGAACAAGCTCCCCGAGAT TCTGAACGAACAGGATAAGGCCGAGATCGCGAAGCTCATTGGTTACAAT GGTACCCACCGCTTGTCCCTTAAGTGCATCCATCTGATCAATGAGGAACT GTGGCAGACCAGCCGGAACCAGATGGAGATCTTCAATTACTTGAACATCA AGCCGAACAAGGTGGACCTGTCCGAACAGAACAAGATACCCAAGGACAT	密码子优化的 E2多核苷酸

[0707]

	GGTCAACGACTTTATCCTCTCACCGGTGGTCAAGCGCACCTTCATTCAATC TATCAACGTGATCAACAAGGTCATCGAGAAGTACGGCATTCCTGAGGATA TCATCATCGAGCTGGCTCGGGAGAACAACTCAGACGATAGGAAGAAGTT CATTAACAACCTCCAGAAAAAGAACGAGGCCACTCGCAAGCGGATTAAT GAGATCATCGGTCAGACCGGGAACCAGAACGCCAAGCGGATCGTGGAAA AGATTCGGCTCCACGACCAACAGGAGGGAAAGTGTCTGTACTCGCTGAA GGACATTCCCCTGGAGGACCTCCTGAGGAACCCAAACAACTACGACATC GATCACATAATCCCCCGCAGCGTGTCATTCGACGATTCCATGCATAACAA GGTCCTCGTGCGGAGAGAGCAGAATGCCAAGAAGAACAACCAGACTCCG TACCAGTACCTGACGTCCGGCTACGCAGACATCAAGTACTCAGTGTTCAA ACAGCACGTGCTCAACCTGGCCGAGAACAAGGACAGGATGACCAAGAAG AAGCGCGAATACCTTCTCGAGGAACGGGATATCAATAAGTTCGAGGTGC AGAAGGAGTTTATCAATAGAAACCTGGTGGACACTCGCTATGCCACCCGC GAACTGACCAACTACCTGAAGGCGTACTTCTCCGCCAACAACATGAACGT GAAGGTCAAAACTATTAACGGCAGCTTCACCGACTATCTGCGCAAGGTCT GGAAGTTCAAGAAGGAACGCAACCACGGTTACAAGCACCACGCGGAAGA TGCGCTGATTATCGCCAACGCTGACTTCCTGTTCAAGGAAAACAAGAAGC TCAAGGCCGTGAACTCAGTGCTCGAGAAGCCTGAAATCGAGACTAAGCA GCTGGACATCCAGGTCGATTCGGAAGATAACTACTCCGAAATGTTCATCA TCCCTAAGCAAGTGCAGGACATCAAGGACTTCAGGAATTTCAAGTACAGC CATCGCGTGGACAAGAAGCCAAACAGACAGCTGATCAACGATACACTGT ATTCCACCCGGAAGAAGGACAACTCCACCTACATCGTCCAAACCATTAAG GACATCTACGCAAAGGACAACACCACGCTTAAGAAGCAGTTCGACAAGA GCCCCGAAAAGTTCCTCATGTACCAGCACGACCCCAGAACCTTCGAGAAG CTTGAAGTGATCATGAAGCAGTACGCCAACGAAAAGAACCCACTGGCTA AGTACCACGAGGAAACCGGCGAATACCTGACCAAGTACTCCAAAAAGAA CAACGGACCGATCGTCAAGTCCCTGAAGTACATTGGGAACAAGCTCGGCT CGCACCTCGATGTGACCCACCAGTTCAAGTCCTCGACCAAAAAGCTCGTG AAGCTGTCCATCAAGCCGTACCGGTTCGACGTGTACCTGACTGACAAGGG ATATAAGTTCATCACCATTTCCTACCTCGACGTGTTGAAGAAGGATAACT ACTACTACATTCCGGAACAGAAGTACGACAAGCTCAAGCTCGGAAAGGC CATCGACAAAAATGCGAAGTTCATCGCGAGCTTCTACAAGAATGACTTGA TCAAGCTGGATGGCGAAATCTACAAGATCATCGGGGTCAACTCCGATACC CGCAACATGATTGAGCTGGATCTGCCCGACATTCGGTACAAGGAATACTG CGAGCTGAACAACATCAAGGGAGAACCGCGGATCAAGAAAACCATCGGA AAGAAAGTGAACAGCATCGAGAAACTGACTACTGACGTCCTGGGAAACG TGTTCACCAACACACAATACACCAAACCCCAGCTGCTGTTTAAGCGCGGG AAC	
23	ATGAACCAGAAATTTATCCTGGGTCTGGATATTGGTATTACCAGCGTTGG TTATGGCCTGATTGATTACGAAACCAAAAACATTATTGATGCCGGTGTTC GTCTGTTTCCGGAAGCAAATGTTGAAAATAATGAAGGTCGTCGTAGCAAA CGTGGTAGCCGTCGTCTGAAACGTCGTCGTATTCATCGTCTGGATCGTGTT AAACATCTGCTGGCAGAATATGATCTGCTGGATCTGACCAATATTCCGAA AAGCACCAATCCGTATCAGACCCGTGTTAAAGGTCTGAATGAAAAGCTG AGCAAAGATGAACTGGTTATTGCACTGCTGCATATTGCAAAACGCCGTGG CATTCATAACGTTGATGTTGCAGCAGATAAAGAAGAAACCGCAAGCGAT AGCCTGAGCACCAAAGATCAGATTAACAAAAACGCCAAATTTCTGGAAA GCCGCTATGTTTGTGAACTGCAGAAAGAACGTCTGGAAAATGAAGGTCAT GTTCGTGGTGTTGAAAATCGCTTTCTGACGAAAGATATTGTGCGTGAGGC CAAAAAAATCATCGATACCCAGATGCAGTATTACCCGGAAATTGATGAA ACCTTCAAAGAGAAATATATCAGCCTGGTTGAAACCCGTCGCGAATATTT TGAAGGTCCGGGTAAAGGTAGCCCGTTTGGTTGGGAAGGTAATATCAAG AAATGGTTTGAGCAGATGATGGGCCACTGTACCTATTTTCCAGAAGAACT GCGTAGCGTCAAATATAGCTATTCAGCCGAACTGTTTAACGCCCTGAATG ATCTGAATAATCTGGTGATTACCCGTGATGAAGATGCCAAACTGAATTAT GGTGAGAAATTCCAGATCATCGAAAACGTGTTCAAACAGAAGAAAACAC CGAACCTGAAACAAATCGCCATTGAAATTGGTGTGCATGAAACCGAAAT CAAAGGTTATCGTGTGAACAAAAGCGGTACACCGGAATTTACCGAATTTA AACTGTATCATGACCTGAAAAGCATCGTGTTCGATAAAAGCATTCTGGAA AATGAAGCCATCCTGGATCAGATTGCAGAAATTCTGACCATCTATCAGGA TGAGCAGAGCATTAAAGAGGAACTGAATAAACTGCCGGAAATACTGAAC	编码 **E2+K741D+L743K的DNA序列**

[0708]

	GAACAGGATAAAGCAGAAATCGCCAAACTGATTGGTTATAATGGCACCC ATCGTCTGAGCCTGAAATGTATTCACCTGATTAATGAAGAACTGTGGCAG ACCAGCCGTAATCAGATGGAAATTTTCAACTACCTGAACATCAAACCGAA CAAAGTGGATCTGAGTGAGCAGAACAAAATCCCGAAAGATATGGTGAAC GACTTTATTCTGAGTCCGGTTGTGAAACGCACCTTTATTCAGAGCATCAA CGTGATCAACAAAGTGATCGAGAAATATGGCATCCCCGAAGATATCATTA TCGAACTGGCACGTGAAAATAACTCCGATGATCGCAAAAAGTTCATCAAC AACCTGCAGAAAAAGAATGAAGCAACCCGCAAACGCATTAACGAAATTA TTGGTCAGACCGGTAATCAGAATGCCAAACGTATTGTGGAAAAAATCCGT CTGCATGATCAGCAAGAGGGTAAATGTCTGTATAGCCTGAAAGATATCCC GCTGGAAGATCTGCTGCGCAATCCGAACAATTATGATATCGACCATATTA TTCCGCGAAGCGTGAGCTTTGATGATAGCATGCATAACAAAGTTCTGGTT CGTCGCGAACAGAATGCCAAAAAGAATAATCAGACCCCGTATCAGTATC TGACCAGTGGTTATGCAGATATCAAATACAGCGTGTTTAAGCAGCATGTT CTGAATCTGGCCGAAAATAAAGATCGCATGACCAAAAAAAAGCGCGAGT ATCTGCTGGAAGAACGCGACATTAACAAATTTGAAGTGCAGAAAGAATT TATCAACCGCAACCTGGTTGATACCCGTTATGCAACCCGTGAACTGACCA ATTATCTGAAAGCATATTTCAGCGCCAACAACATGAACGTGAAAGTGAA AACGATTAACGGCAGCTTTACCGATTATCTGCGTAAAGTGTGGAAATTCA AAAAAGAACGCAACCACGGCTATAAACATCATGCCGAAGATGCCCTGAT TATTGCAAATGCAGATTTCCTGTTTAAAGAAAACAAAAAACTGAAAGCCG TCAACAGCGTGCTGGAAAAACCGGAAATTGAGACAGATCAGAAAGACAT TCAGGTTGATAGCGAAGATAATTACAGCGAAATGTTTATCATCCCGAAAC AGGTGCAGGATATCAAAGATTTTCGCAACTTCAAATATAGCCACCGCGTT GACAAAAAAACCTAATCGTCAGCTGATTAACGATACCCTGTATAGCACCCG CAAAAAAGATAACAGCACCTATATTGTGCAGACCATTAAAGACATCTAC GCCAAAGATAATACCACCCTGAAAAAACAGTTCGACAAAAGCCCAGAAA AATTTCTGATGTATCAGCATGATCCGCGTACCTTCGAAAAACTGGAAGTT ATTATGAAACAGTATGCCAACGAGAAAAATCCGCTGGCCAAATATCACG AAGAAACCGGTGAATATCTGACCAAATATTCCAAGAAGAACAACGGTCC GATCGTTAAATCCCTGAAATATATCGGTAATAAACTGGGCAGCCATCTGG ATGTTACCCATCAGTTTAAAAGCTCCACAAAGAAGCTGGTTAAACTGTCC ATCAAACCGTATCGCTTTGATGTGTATCTGACCGACAAAGGCTATAAATT CATTACCATCAGCTATCTGGACGTGCTGAAAAAAGACAACTATTATTATA TCCCGGAACAGAAATATGATAAACTGAAACTGGGTAAAGCCATCGATAA AAACGCCAAATTTATCGCCAGCTTCTACAAAAACGACCTGATTAAACTGG ATGGCGAGATCTATAAAATCATCGGTGTTAATAGCGACACCCGCAATATG ATTGAGCTGGATCTGCCGGATATTCGCTATAAAGAATATTGCGAACTGAA CAACATTAAAGGCGAACCGCGTATCAAAAAGACCATCGGCAAAAAAGTG AATAGCATCGAGAAACTGACCACCGATGTTCTGGGTAATGTGTTTACCAA TACCCAGTATACCAAACCTCAGCTGCTGTTCAAACGCGGTAAT	
24	ATGAACCAGAAATTTATCCTGGGTCTGGATATTGGTATTACCAGCGTTGG TTATGGCCTGATTGATTACGAAACCAAAAACATTATTGATGCCGGTGTTC GTCTGTTTCCGGAAGCAAATGTTGAAAATAATGAAGGTCGTCGTAGCAAA CGTGGTAGCCGTCGTCTGAAACGTCGTCGTATTCATCGTCTGGATCGTGTT AAACATCTGCTGGCAGAATATGATCTGCTGGATCTGACCAATATTCCGAA AAGCACCAATCCGTATCAGACCCGTGTTAAAGGTCTGAATGAAAAGCTG AGCAAAGATGAACTGGTTATTGCACTGCTGCATATTGCAAAACGCCGTGG CATTCATAACGTTGATGTTGCAGCAGATAAAGAAGAAACCGCAAGCGAT AGCCTGAGCACCAAAGATCAGATTAACAAAAACGCCAAATTTCTGGAAA GCCGCTATGTTTGTGAACTGCAGAAAGAACGTCTGGAAAATGAAGGTCAT GTTCGTGGTGTTGAAAATCGCTTTCTGACGAAAGATATTGTGCGTGAGGC CAAAAAAATCATCGATACCCAGATGCAGTATTACCCGGAAATTGATGAA ACCTTCAAAGAGAAATATATCAGCCTGGTTGAAACCCGTCGCGAATATTT TGAAGGTCCGGGTAAAGGTAGCCCGTTTGGTTGGGAAGGTAATATCAAG AAATGGTTTGAGCAGATGATGGGCCACTGTACCTATTTTCCAGAAGAACT GCGTAGCGTCAAATATAGCTATTCAGCCGAACTGTTTAACGCCCTGAATG ATCTGAATAATCTGGTGATTACCCGTGATGAAGATGCCAAACTGAATTAT GGTGAGAAATTCCAGATCATCGAAAACGTGTTCAAACAGAAGAAAACAC CGAACCTGAAACAAATCGCCATTGAAATTGGTGTGCATGAAACCGAAAT CAAAGGTTATCGTGTGAACAAAAGCGGTACACCGGAATTTACCGAATTTA	编码 E2+S670T+N675D的DNA序列

[0709]

	AACTGTATCATGACCTGAAAAGCATCGTGTTCGATAAAAGCATTCTGGAA AATGAAGCCATCCTGGATCAGATTGCAGAAATTCTGACCATCTATCAGGA TGAGCAGAGCATTAAAGAGGAACTGAATAAACTGCCGGAAATACTGAAC GAACAGGATAAAGCAGAAATCGCCAAACTGATTGGTTATAATGGCACCC ATCGTCTGAGCCTGAAATGTATTCACCTGATTAATGAAGAACTGTGGCAG ACCAGCCGTAATCAGATGGAAATTTTCAACTACCTGAACATCAAACCGAA CAAAGTGGATCTGAGTGAGCAGAACAAAATCCCGAAAGATATGGTGAAC GACTTTATTCTGAGTCCGGTTGTGAAACGCACCTTTATTCAGAGCATCAA CGTGATCAACAAAGTGATCGAGAAATATGGCATCCCCGAAGATATCATTA TCGAACTGGCACGTGAAAATAACTCCGATGATCGCAAAAAGTTCATCAAC AACCTGCAGAAAAAGAATGAAGCAACCCGCAAACGCATTAACGAAATTA TTGGTCAGACCGGTAATCAGAATGCCAAACGTATTGTGGAAAAAATCCGT CTGCATGATCAGCAAGAGGGTAAATGTCTGTATAGCCTGAAAGATATCCC GCTGGAAGATCTGCTGCGCAATCCGAACAATTATGATATCGACCATATTA TTCCGCGAAGCGTGAGCTTTGATGATAGCATGCATAACAAAGTTCTGGTT CGTCGCGAACAGAATGCCAAAAAGAATAATCAGACCCCGTATCAGTATC TGACCAGTGGTTATGCAGATATCAAATACAGCGTGTTTAAGCAGCATGTT CTGAATCTGGCCGAAAATAAAGATCGCATGACCAAAAAAAAGCGCGAGT ATCTGCTGGAAGAACGCGACATTAACAAATTTGAAGTGCAGAAAGAATT TATCAACCGCAACCTGGTTGATACCCGTTATGCAACCCGTGAACTGACCA ATTATCTGAAAGCATATTTCACCGCCAACAACATGGATGTGAAAGTGAAA ACGATTAACGGCAGCTTTACCGATTATCTGCGTAAAGTGTGGAAATTCAA AAAAGAACGCAACCACGGCTATAAACATCATGCCGAAGATGCCCTGATT ATTGCAAATGCAGATTTCCTGTTTAAAGAAAACAAAAAACTGAAAGCCGT CAACAGCGTGCTGGAAAAACCGGAAATTGAGACAAAACAGCTGGACATT CAGGTTGATAGCGAAGATAATTACAGCGAAATGTTTATCATCCCGAAACA GGTGCAGGATATCAAAGATTTTCGCAACTTCAAATATAGCCACCGCGTTG ACAAAAAACCTAATCGTCAGCTGATTAACGATACCCTGTATAGCACCCGC AAAAAAGATAACAGCACCTATATTGTGCAGACCATTAAAGACATCTACG CCAAAGATAATACCACCCTGAAAAAACAGTTCGACAAAAGCCCAGAAAA ATTTCTGATGTATCAGCATGATCCGCGTACCTTCGAAAAACTGGAAGTTA TTATGAAACAGTATGCCAACGAGAAAAATCCGCTGGCCAAATATCACGA AGAAACCGGTGAATATCTGACCAAATATTCCAAGAAGAACAACGGTCCG ATCGTTAAATCCCTGAAATATATCGGTAATAAACTGGGCAGCCATCTGGA TGTTACCCATCAGTTTAAAAGCTCCACAAAGAAGCTGGTTAAACTGTCCA TCAAACCGTATCGCTTTGATGTGTATCTGACCGACAAAGGCTATAAATTC ATTACCATCAGCTATCTGGACGTGCTGAAAAAAGACAACTATTATTATAT CCCGGAACAGAAATATGATAAACTGAAACTGGGTAAAGCCATCGATAAA AACGCCAAATTTATCGCCAGCTTCTACAAAAACGACCTGATTAAACTGGA TGGCGAGATCTATAAAATCATCGGTGTTAATAGCGACACCCGCAATATGA TTGAGCTGGATCTGCCGGATATTCGCTATAAAGAATATTGCGAACTGAAC AACATTAAAGGCGAACCGCGTATCAAAAAGACCATCGGCAAAAAAGTGA ATAGCATCGAGAAACTGACCACCGATGTTCTGGGTAATGTGTTTACCAAT ACCCAGTATACCAAACCTCAGCTGCTGTTCAAACGCGGTAAT	
25	ATGAACCAGAAATTTATCCTGGGTCTGGATATTGGTATTACCAGCGTTGG TTATGGCCTGATTGATTACGAAACCAAAAACATTATTGATGCCGGTGTTC GTCTGTTTCCGGAAGCAAATGTTGAAAATAATGAAGGTCGTCGTAGCAAA CGTGGTAGCCGTCGTCTGAAACGTCGTCGTATTCATCGTCTGGATCGTGTT AAACATCTGCTGGCAGAATATGATCTGCTGGATCTGACCAATATTCCGAA AAGCACCAATCCGTATCAGACCCGTGTTAAAGGTCTGAATGAAAAGCTG AGCAAAGATGAACTGGTTATTGCACTGCTGCATATTGCAAAACGCCGTGG CATTCATAACGTTGATGTTGCAGCAGATAAAGAAGAAACCGCAAGCGAT AGCCTGAGCACCAAAGATCAGATTAACAAAAACGCCAAATTTCTGGAAA GCCGCTATGTTTGTGAACTGCAGAAAGAACGTCTGGAAAATGAAGGTCAT GTTCGTGGTGTTGAAAATCGCTTTCTGACGAAAGATATTGTGCGTGAGGC CAAAAAAATCATCGATACCCAGATGCAGTATTACCCGGAAATTGATGAA ACCTTCAAAGAGAAATATATCAGCCTGGTTGAAACCCGTCGCGAATATTT TGAAGGTCCGGGTAAAGGTAGCCCGTTTGGTTGGGAAGGTAATATCAAG AAATGGTTTGAGCAGATGATGGGCCACTGTACCTATTTTCCAGAAGAACT GCGTAGCGTCAAATATAGCTATTCAGCCGAACTGTTTAACGCCCTGAATG ATCTGAATAATCTGGTGATTACCCGTGATGAAGATGCCAAACTGAATTAT	编码 **E2+K741N+L743N的DNA序列**

[0710]

	GGTGAGAAATTCCAGATCATCGAAAACGTGTTCAAACAGAAGAAAACAC CGAACCTGAAACAAATCGCCATTGAAATTGGTGTGCATGAAACCGAAAT CAAAGGTTATCGTGTGAACAAAAGCGGTACACCGGAATTTACCGAATTTA AACTGTATCATGACCTGAAAAGCATCGTGTTCGATAAAAGCATTCTGGAA AATGAAGCCATCCTGGATCAGATTGCAGAAATTCTGACCATCTATCAGGA TGAGCAGAGCATTAAAGAGGAACTGAATAAACTGCCGGAAATACTGAAC GAACAGGATAAAGCAGAAATCGCCAAACTGATTGGTTATAATGGCACCC ATCGTCTGAGCCTGAAATGTATTCACCTGATTAATGAAGAACTGTGGCAG ACCAGCCGTAATCAGATGGAAATTTTCAACTACCTGAACATCAAACCGAA CAAAGTGGATCTGAGTGAGCAGAACAAAATCCCGAAAGATATGGTGAAC GACTTTATTCTGAGTCCGGTTGTGAAACGCACCTTTATTCAGAGCATCAA CGTGATCAACAAAGTGATCGAGAAATATGGCATCCCCGAAGATATCATTA TCGAACTGGCACGTGAAAATAACTCCGATGATCGCAAAAAGTTCATCAAC AACCTGCAGAAAAAGAATGAAGCAACCCGCAAACGCATTAACGAAATTA TTGGTCAGACCGGTAATCAGAATGCCAAACGTATTGTGGAAAAAATCCGT CTGCATGATCAGCAAGAGGGTAAATGTCTGTATAGCCTGAAAGATATCCC GCTGGAAGATCTGCTGCGCAATCCGAACAATTATGATATCGACCATATTA TTCCGCGAAGCGTGAGCTTTGATGATAGCATGCATAACAAAGTTCTGGTT CGTCGCGAACAGAATGCCAAAAAGAATAATCAGACCCCGTATCAGTATC TGACCAGTGGTTATGCAGATATCAAATACAGCGTGTTTAAGCAGCATGTT CTGAATCTGGCCGAAAATAAAGATCGCATGACCAAAAAAAAGCGCGAGT ATCTGCTGGAAGAACGCGACATTAACAAATTTGAAGTGCAGAAAGAATT TATCAACCGCAACCTGGTTGATACCCGTTATGCAACCCGTGAACTGACCA ATTATCTGAAAGCATATTTCAGCGCCAACAACATGAACGTGAAAGTGAA AACGATTAACGGCAGCTTTACCGATTATCTGCGTAAAGTGTGGAAATTCA AAAAAGAACGCAACCACGGCTATAAACATCATGCCGAAGATGCCCTGAT TATTGCAAATGCAGATTTCCTGTTTAAAGAAAACAAAAAACTGAAAGCCG TCAACAGCGTGCTGGAAAAACCGGAAATTGAGACAAATCAGAATGACAT TCAGGTTGATAGCGAAGATAATTACAGCGAAATGTTTATCATCCCGAAAC AGGTGCAGGATATCAAAGATTTTCGCAACTTCAAATATAGCCACCGCGTT GACAAAAAACCTAATCGTCAGCTGATTAACGATACCCTGTATAGCACCCG CAAAAAAGATAACAGCACCTATATTGTGCAGACCATTAAAGACATCTAC GCCAAAGATAATACCACCCTGAAAAAACAGTTCGACAAAAGCCCAGAAA AATTTCTGATGTATCAGCATGATCCGCGTACCTTCGAAAAACTGGAAGTT ATTATGAAACAGTATGCCAACGAGAAAAATCCGCTGGCCAAATATCACG AAGAAACCGGTGAATATCTGACCAAATATTCCAAGAAGAACAACGGTCC GATCGTTAAATCCCTGAAATATATCGGTAATAAACTGGGCAGCCATCTGG ATGTTACCCATCAGTTTAAAAGCTCCACAAAGAAGCTGGTTAAACTGTCC ATCAAACCGTATCGCTTTGATGTGTATCTGACCGACAAAGGCTATAAATT CATTACCATCAGCTATCTGGACGTGCTGAAAAAAGACAACTATTATTATA TCCCGGAACAGAAATATGATAAACTGAAACTGGGTAAAGCCATCGATAA AAACGCCAAATTTATCGCCAGCTTCTACAAAAACGACCTGATTAAACTGG ATGGCGAGATCTATAAAATCATCGGTGTTAATAGCGACACCCGCAATATG ATTGAGCTGGATCTGCCGGATATTCGCTATAAAGAATATTGCGAACTGAA CAACATTAAAGGCGAACCGCGTATCAAAAAGACCATCGGCAAAAAAGTG AATAGCATCGAGAAACTGACCACCGATGTTCTGGGTAATGTGTTTACCAA TACCCAGTATACCAAACCTCAGCTGCTGTTCAAACGCGGTAAT	
26	ATGAACCAGAAATTTATCCTGGGTCTGGATATTGGTATTACCAGCGTTGG TTATGGCCTGATTGATTACGAAACCAAAAACATTATTGATGCCGGTGTTC GTCTGTTTCCGGAAGCAAATGTTGAAAATAATGAAGGTCGTCGTAGCAAA CGTGGTAGCCGTCGTCTGAAACGTCGTCGTATTCATCGTCTGGAACGTGT TAAAAGCCTGCTGAGCGAATATAAGATTATTAGCGGTCTGGCACCGACCA ATAATCAGCCGTATAACATTCGTGTTAAAGGTCTGACCGAACAGCTGACC AAAGATGAACTGGCAGTTGCACTGCTGCATATTGCCAAACGCCGTGGCAT TCATAAAATCGATGTGATTGATAGCAATGACGATGTGGGTAATGAACTGA GCACCAAAGAACAGCTGAACAAAAATAGCAAACTGCTGAAAGACAAATT CGTGTGTCAGATTCAGCTGGAACGTATGAATGAAGGCCAGGTTCGTGGTG AAAAGAATCGCTTTAAAACCGCAGACATCATCAAAGAAATTATCCAGCT GCTGAACGTGCAGAAAAACTTCCATCAGCTGGATGAAAACTTCATCAACA AATACATCGAGCTGGTTGAAATGCGTCGCGAATATTTTGAAGGTCCTGGT CAGGGTAGCCCGTTTGGTTGGAATGGTGATCTGAAAAAATGGTACGAAAT	编码F8的DNA序列

[0711]

	GCTGATGGGTCACTGTACCTATTTTCCGCAAGAACTGCGTAGCGTTAAAT ATGCCTATAGCGCAGACCTGTTTAATGCACTGAATGATCTGAACAACCTG ATTATTCAGCGCGATAATAGCGAGAAACTGGAATACCATGAGAAGTATC ACATCATCGAGAACGTGTTCAAGCAGAAAAAAAAGCCGACGCTGAAACA AATCGCAAAAGAGATTGGCGTTAACCCGGAAGATATTAAAGGTTATCGT ATTACCAAAAGCGGTACACCGGAATTCACCGAATTTAAACTGTATCACGA TCTGAAAAGCGTGCTGTTTGATCAGAGCATTCTGGAAAATGAAGATGTGC TGGACCAGATTGCAGAAATTCTGACCATTTATCAGGACAAAGACAGCATC AAAAGCAAACTGACCGAACTGGATATTCTGCTGAATGAAGAAGATAAAG AGAACATTGCACAGCTGACCGGTTATAACGGCACACATCGCCTGAGCCTG AAATGTATTCGTCTGGTACTGGAAGAACAGTGGTATAGCAGCCGTAATCA GATGGAAATCTTTACCCATCTGAACATTAAACCGAAGAAAATCAATCTGA CCGCAGCCAACAAAATTCCGAAAGCCATGATTGATGAGTTTATTCTGAGT CCGGTTGTGAAACGCACCTTTATTCAGAGCATCAACGTGATCAACAAAGT GATCGAGAAATATGGCATCCCCGAAGATATCATTATCGAACTGGCACGTG AAAATAACTCCGATGATCGCAAAAAGTTCATCAACAACCTGCAGAAAAA GAATGAAGCAACCCGCAAACGCATTAACGAAATTATTGGTCAGACCGGT AATCAGAATGCCAAACGTATTGTGGAAAAAATCCGTCTGCATGATCAGCA AGAGGGGAAATGTCTGTATAGCCTGGAAAGCATTGCCCTGATGGATCTGC TGAATAACCCGCAGAATTATGAAGTGGATCATATTATTCCGCGTAGCGTG GCATTTGATAATTCCATTCATAACAAAGTGCTGGTGAAGCAGATCGAGAA TAGCAAAAAAGGTAATCGTACGCCGTATCAGTATCTGAATAGCAGTGATG CAAAACTGAGCTACAACCAGTTTAAACAGCATATTCTGAATCTGAGCAAA AGCAAAGATCGCATCAGCAAAAAAAAGAAGGACTACCTGCTGGAAGAAC GCGATATCAACAAATTTGAAGTCCAGAAAGAGTTTATCAACCGCAATCTG GTTGATACCCGTTATGCAACCCGTGAACTGACCAGCTATCTGAAAGCATA TTTCAGCGCCAATAACATGGACGTGAAAGTGAAAACAATTAACGGCAGC TTTACCAACCATCTGCGTAAAGTTTGGCGCTTTGATAAATATCGCAACCA CGGCTATAAACATCATGCAGAAGATGCCCTGATTATTGCAAATGCAGATT TCCTGTTTAAAGAAAACAAAAAACTGAAAGCCGTCAACAGCGTGCTGGA AAAACCGGAAATTGAGACAAAACAGCTGGACATTCAGGTTGATAGCGAA GATAATTACAGCGAAATGTTTATCATCCCGAAACAGGTGCAGGATATCAA AGATTTTCGCAACTTCAAATATAGCCACCGCGTTGACAAAAAACCTAATC GTCAGCTGATTAACGATACCCTGTATAGCACCCGCAAAAAAGATAACAG CACCTATATTGTGCAGACCATTAAAGACATCTACGCCAAAGATAATACCA CCCTGAAAAAACAGTTCGACAAAAGCCCAGAAAAATTTCTGATGTATCA GCATGATCCGCGTACCTTCGAAAAACTGGAAGTTATTATGAAACAGTATG CCAACGAGAAAAATCCGCTGGCCAAATATCACGAAGAAACCGGTGAATA TCTGACCAAATATTCCAAGAAGAACAACGGTCCGATCGTTAAATCCCTGA AATATATCGGTAATAAACTGGGCAGCCATCTGGATGTTACCCATCAGTTT AAAAGCTCCACAAAGAAGCTGGTTAAACTGTCCATCAAACCGTATCGCTT TGATGTGTATCTGACCGACAAAGGCTATAAATTCATTACCATCAGCTATC TGGACGTGCTGAAAAAAGACAACTATTATTATATCCCGGAACAGAAATAT GATAAACTGAAACTGGGTAAAGCCATCGATAAAAACGCCAAATTTATCG CCAGCTTCTACAAAAACGACCTGATTAAACTGGATGGCGAGATCTATAAA ATCATCGGTGTTAATAGCGACACCCGCAATATGATTGAGCTGGATCTGCC GGATATTCGCTATAAAGAATATTGCGAACTGAACAACATTAAAGGCGAA CCGCGTATCAAAAAGACCATCGGCAAAAAAGTGAATAGCATCGAGAAAC TGACCACCGATGTTCTGGGTAATGTGTTTACCAATACCCAGTATACCAAA CCTCAGCTGCTGTTCAAACGCGGTAAT	
27	ATGAACCAAAAGTTCATTCTGGGGCTCGATATCGGCATCACCTCCGTGGG ATATGGTCTGATCGACTACGAGACTAAGAACATCATCGACGCTGGAGTGC GACTGTTCCCGGAAGCGAACGTGGAGAACAACGAAGGCCGCAGATCCAA GCGCGGGTCCAGAAGGCTCAAGAGGCGGAGGATCCATAGACTCGAGAGA GTGAAGTCGCTCCTTTCGGAATACAAGATTATCAGCGGTCTTGCCCCCAC CAACAACCAACCGTACAACATCAGAGTGAAGGGCCTGACCGAACAGCTG ACCAAAGATGAACTGGCCGTCGCCCTGCTGCATATTGCCAAACGGCGCGG AATCCATAAGATCGACGTGATTGACAGCAACGATGACGTGGGAAACGAG CTGTCAACCAAGGAACAGCTTAACAAGAACAGCAAATTGCTGAAGGACA AGTTTGTCTGCCAAATTCAACTGGAACGGATGAACGAGGGACAAGTCAG GGGAGAGAAAAACCGGTTCAAGACCGCCGACATCATCAAGGAGATCATC	密码子优化的F8多核苷酸

[0712]

	CAACTGCTGAACGTGCAGAAGAACTTCCACCAACTGGATGAAAACTTCAT TAACAAGTACATTGAACTGGTGGAAATGCGGAGGGAGTACTTCGAGGGA CCTGGACAGGGATCCCCTTTCGGCTGGAATGGGGACCTTAAGAAGTGGTA CGAAATGTTGATGGGCCATTGCACTTACTTTCCGCAAGAACTCCGGTCCG TGAAGTACGCATACTCTGCCGACCTGTTCAATGCACTCAACGACCTTAAC AACTTGATCATCCAGCGCGATAACTCGGAAAAGTTGGAATACCACGAAA AGTATCACATCATCGAGAACGTGTTCAAGCAGAAAAAGAAGCCAACTCT GAAGCAGATTGCCAAGGAAATTGGCGTGAATCCGGAGGATATCAAGGGA TACCGGATCACTAAGTCCGGCACGCCAGAGTTCACCGAGTTCAAGCTGTA CCACGATCTGAAGTCGGTGCTCTTTGACCAGTCCATCCTGGAAAACGAAG ATGTGCTGGACCAGATTGCTGAGATCCTGACCATCTACCAGGACAAGGAC TCGATTAAGTCCAAGCTCACCGAGCTGGACATTCTGCTGAACGAAGAAGA TAAGGAGAACATCGCGCAGCTCACCGGTTACAATGGTACCCACCGCTTGT CCCTTAAGTGCATCCGCCTGGTGCTGGAGGAACAGTGGTACTCGAGCCGG AACCAGATGGAGATCTTCACTCACTTGAACATCAAGCCGAAAAAGATTA ACCTGACTGCCGCCAACAAGATACCCAAGGCCATGATCGACGAGTTTATC CTCTCACCGGTGGTCAAGCGCACCTTCATTCAATCTATCAACGTGATCAA CAAGGTCATCGAGAAGTACGGCATTCCTGAGGATATCATCATCGAGCTGG CTCGGGAGAACAACTCAGACGATAGGAAGAAGTTCATTAACAACCTCCA GAAAAAGAACGAGGCCACTCGCAAGCGGATTAATGAGATCATCGGTCAG ACCGGGAACCAGAACGCCAAGCGGATCGTGGAAAAGATTCGGCTCCACG ACCAACAGGAGGGAAAGTGTCTGTACTCGCTGGAGTCGATTGCACTGATG GACCTCCTGAACAACCCACAGAACTACGAAGTCGATCACATAATCCCCCG CAGCGTGGCATTCGACAACTCCATCCATAACAAGGTCCTCGTGAAGCAGA TCGAGAATAGCAAGAAGGGGAACCGGACTCCGTACCAGTACCTGAACTC CTCCGACGCCAAGCTGTCATACAATCAGTTCAAACAGCACATTCTCAACC TGTCCAAGTCAAAGGACAGGATCTCCAAGAAGAAGAAGGACTACCTTCT CGAGGAACGGGATATCAATAAGTTCGAGGTGCAGAAGGAGTTTATCAAT AGAAACCTGGTGGACACTCGCTATGCCACCCGCGAACTGACCAGCTACCT GAAGGCGTACTTCTCCGCCAACAACATGGACGTGAAGGTCAAAACTATTA ACGGCAGCTTCACCAACCATCTGCGCAAGGTCTGGAGGTTCGACAAGTAC CGCAACCACGGTTACAAGCACCACGCGGAAGATGCGCTGATTATCGCCA ACGCTGACTTCCTGTTCAAGGAAAACAAGAAGCTCAAGGCCGTGAACTC AGTGCTCGAGAAGCCTGAAATCGAGACTAAGCAGCTGGACATCCAGGTC GATTCGGAAGATAACTACTCCGAAATGTTCATCATCCCTAAGCAAGTGCA GGACATCAAGGACTTCAGGAATTTCAAGTACAGCCATCGCGTGGACAAG AAGCCAAACAGACAGCTGATCAACGATACACTGTATTCCACCCGGAAGA AGGACAACTCCACCTACATCGTCCAAACCATTAAGGACATCTACGCAAAG GACAACACCACGCTTAAGAAGCAGTTCGACAAGAGCCCCGAAAAGTTCC TCATGTACCAGCACGACCCCAGAACCTTCGAGAAGCTTGAAGTGATCATG AAGCAGTACGCCAACGAAAAGAACCCACTGGCTAAGTACCACGAGGAAA CCGGCGAATACCTGACCAAGTACTCCAAAAAGAACAACGGACCGATCGT CAAGTCCCTGAAGTACATTGGGAACAAGCTCGGCTCGCACCTCGATGTGA CCCACCAGTTCAAGTCCTCGACCAAAAAGCTCGTGAAGCTGTCCATCAAG CCGTACCGGTTCGACGTGTACCTGACTGACAAGGGATATAAGTTCATCAC CATTTCCTACCTCGACGTGTTGAAGAAGGATAACTACTACTACATTCCGG AACAGAAGTACGACAAGCTCAAGCTCGGAAAGGCCATCGACAAAAATGC GAAGTTCATCGCGAGCTTCTACAAGAATGACTTGATCAAGCTGGATGGCG AAATCTACAAGATCATCGGGGGTCAACTCCGATACCCGCAACATGATTGAG CTGGATCTGCCCGACATTCGGTACAAGGAATACTGCGAGCTGAACAACAT CAAGGGAGAACCGCGGATCAAGAAAACCATCGGAAAGAAAGTGAACAG CATCGAGAAACTGACTACTGACGTCCTGGGAAACGTGTTCACCAACACAC AATACACCAAACCCCAGCTGCTGTTTAAGCGCGGGAAC	
28	ATGAACCAGAAATTTATCCTGGGTCTGGATATTGGTATTACCAGCGTTGG TTATGGCCTGATTGATTACGAAACCAAAAACATTATTGATGCCGGTGTTC GTCTGTTTCCGGAAGCAAATGTTGAAAATAATGAAGGTCGTCGTAGCAAA CGTGGTAGCCGTCGTCTGAAACGTCGTCGTATTCATCGTCTGGAACGTGT TAAAAGCCTGCTGAGCGAATATAAGATTATTAGCGGTCTGGCACCGACCA ATAATCAGCCGTATAACATTCGTGTTAAAGGTCTGACCGAACAGCTGACC AAAGATGAACTGGCAGTTGCACTGCTGCATATTGCCAAACGCCGTGGCAT TCATAAAATCGATGTGATTGATAGCAATGACGATGTGGGTAATGAACTGA	编码 **F8+K737D+L739K的DNA序列**

[0713]

	GCACCAAAGAACAGCTGAACAAAAATAGCAAACTGCTGAAAGACAAATT CGTGTGTCAGATTCAGCTGGAACGTATGAATGAAGGCCAGGTTCGTGGTG AAAAGAATCGCTTTAAAACCGCAGACATCATCAAAGAAATTATCCAGCT GCTGAACGTGCAGAAAAACTTCCATCAGCTGGATGAAAACTTCATCAACA AATACATCGAGCTGGTTGAAATGCGTCGCGAATATTTTGAAGGTCCTGGT CAGGGTAGCCCGTTTGGTTGGAATGGTGATCTGAAAAAATGGTACGAAAT GCTGATGGGTCACTGTACCTATTTTCCGCAAGAACTGCGTAGCGTTAAAT ATGCCTATAGCGCAGACCTGTTTAATGCACTGAATGATCTGAACAACCTG ATTATTCAGCGCGATAATAGCGAGAAACTGGAATACCATGAGAAGTATC ACATCATCGAGAACGTGTTCAAGCAGAAAAAAAAGCCGACGCTGAAACA AATCGCAAAAGAGATTGGCGTTAACCCGGAAGATATTAAAGGTTATCGT ATTACCAAAAGCGGTACACCGGAATTCACCGAATTTAAACTGTATCACGA TCTGAAAAGCGTGCTGTTTGATCAGAGCATTCTGGAAAATGAAGATGTGC TGGACCAGATTGCAGAAATTCTGACCATTTATCAGGACAAAGACAGCATC AAAAGCAAACTGACCGAACTGGATATTCTGCTGAATGAAGAAGATAAAG AGAACATTGCACAGCTGACCGGTTATAACGGCACACATCGCCTGAGCCTG AAATGTATTCGTCTGGTACTGGAAGAACAGTGGTATAGCAGCCGTAATCA GATGGAAATCTTTACCCATCTGAACATTAAACCGAAGAAAATCAATCTGA CCGCAGCCAACAAAATTCCGAAAGCCATGATTGATGAGTTTATTCTGAGT CCGGTTGTGAAACGCACCTTTATTCAGAGCATCAACGTGATCAACAAAGT GATCGAGAAATATGGCATCCCCGAAGATATCATTATCGAACTGGCACGTG AAAATAACTCCGATGATCGCAAAAAGTTCATCAACAACCTGCAGAAAAA GAATGAAGCAACCCGCAAACGCATTAACGAAATTATTGGTCAGACCGGT AATCAGAATGCCAAACGTATTGTGGAAAAAATCCGTCTGCATGATCAGCA AGAGGGGAAATGTCTGTATAGCCTGGAAAGCATTGCCCTGATGGATCTGC TGAATAACCCGCAGAATTATGAAGTGGATCATATTATTCCGCGTAGCGTG GCATTTGATAATTCCATTCATAACAAAGTGCTGGTGAAGCAGATCGAGAA TAGCAAAAAAAGGTAATCGTACGCCGTATCAGTATCTGAATAGCAGTGATG CAAAACTGAGCTACAACCAGTTTAAACAGCATATTCTGAATCTGAGCAAA AGCAAAGATCGCATCAGCAAAAAAAAGAAGGACTACCTGCTGGAAGAAC GCGATATCAACAAATTTGAAGTCCAGAAAGAGTTTATCAACCGCAATCTG GTTGATACCCGTTATGCAACCCGTGAACTGACCAGCTATCTGAAAGCATA TTTCAGCGCCAATAACATGGACGTGAAAGTGAAAACAATTAACGGCAGC TTTACCAACCATCTGCGTAAAGTTTGGCGCTTTGATAAATATCGCAACCA CGGCTATAAACATCATGCAGAAGATGCCCTGATTATTGCAAATGCAGATT TCCTGTTTAAAGAAAACAAAAAACTGAAAGCCGTCAACAGCGTGCTGGA AAAACCGGAAATTGAGACAGATCAGAAAGACATTCAGGTTGATAGCGAA GATAATTACAGCGAAATGTTTATCATCCCGAAACAGGTGCAGGATATCAA AGATTTTCGCAACTTCAAATATAGCCACCGCGTTGACAAAAAACCTAATC GTCAGCTGATTAACGATACCCTGTATAGCACCCGCAAAAAAGATAACAG CACCTATATTGTGCAGACCATTAAAGACATCTACGCCAAAGATAATACCA CCCTGAAAAAACAGTTCGACAAAAGCCCAGAAAAATTTCTGATGTATCA GCATGATCCGCGTACCTTCGAAAAACTGGAAGTTATTATGAAACAGTATG CCAACGAGAAAAATCCGCTGGCCAAATATCACGAAGAAACCGGTGAATA TCTGACCAAATATTCCAAGAAGAACAACGGTCCGATCGTTAAATCCCTGA AATATATCGGTAATAAACTGGGCAGCCATCTGGATGTTACCCATCAGTTT AAAAGCTCCACAAAGAAGCTGGTTAAACTGTCCATCAAACCGTATCGCTT TGATGTGTATCTGACCGACAAAGGCTATAAATTCATTACCATCAGCTATC TGGACGTGCTGAAAAAAGACAACTATTATTATATCCCGGAACAGAAATAT GATAAACTGAAACTGGGTAAAGCCATCGATAAAAACGCCAAATTTATCG CCAGCTTCTACAAAAACGACCTGATTAAACTGGATGGCGAGATCTATAAA ATCATCGGTGTTAATAGCGACACCCGCAATATGATTGAGCTGGATCTGCC GGATATTCGCTATAAAGAATATTGCGAACTGAACAACATTAAAGGCGAA CCGCGTATCAAAAAGACCATCGGCAAAAAAGTGAATAGCATCGAGAAAC TGACCACCGATGTTCTGGGTAATGTGTTTACCAATACCCAGTATACCAAA CCTCAGCTGCTGTTCAAACGCGGTAAT	
29	ATGAACCAGAAATTTATCCTGGGTCTGGATATTGGTATTACCAGCGTTGG TTATGGCCTGATTGATTACGAAACCAAAAACATTATTGATGCCGGTGTTC GTCTGTTTCCGGAAGCAAATGTTGAAAATAATGAAGGTCGTCGTAGCAAA CGTGGTAGCCGTCGTCTGAAACGTCGTCGTATTCATCGTCTGGAACGTGT TAAAAGCCTGCTGAGCGAATATAAGATTATTAGCGGTCTGGCACCGACCA	编码 F8+K737N+L739 N的DNA序列

[0714]

	ATAATCAGCCGTATAACATTCGTGTTAAAGGTCTGACCGAACAGCTGACC AAAGATGAACTGGCAGTTGCACTGCTGCATATTGCCAAACGCCGTGGCAT TCATAAAATCGATGTGATTGATAGCAATGACGATGTGGGTAATGAACTGA GCACCAAAGAACAGCTGAACAAAAATAGCAAACTGCTGAAAGACAAATT CGTGTGTCAGATTCAGCTGGAACGTATGAATGAAGGCCAGGTTCGTGGTG AAAAGAATCGCTTTAAAACCGCAGACATCATCAAAGAAATTATCCAGCT GCTGAACGTGCAGAAAAACTTCCATCAGCTGGATGAAAACTTCATCAACA AATACATCGAGCTGGTTGAAATGCGTCGCGAATATTTTGAAGGTCCTGGT CAGGGTAGCCCGTTTGGTTGGAATGGTGATCTGAAAAAATGGTACGAAAT GCTGATGGGTCACTGTACCTATTTTCCGCAAGAACTGCGTAGCGTTAAAT ATGCCTATAGCGCAGACCTGTTTAATGCACTGAATGATCTGAACAACCTG ATTATTCAGCGCGATAATAGCGAGAAACTGGAATACCATGAGAAGTATC ACATCATCGAGAACGTGTTCAAGCAGAAAAAAAAGCCGACGCTGAAACA AATCGCAAAAGAGATTGGCGTTAACCCGGAAGATATTAAAGGTTATCGT ATTACCAAAAGCGGTACACCGGAATTCACCGAATTTAAACTGTATCACGA TCTGAAAAGCGTGCTGTTTGATCAGAGCATTCTGGAAAATGAAGATGTGC TGGACCAGATTGCAGAAATTCTGACCATTTATCAGGACAAAGACAGCATC AAAAGCAAACTGACCGAACTGGATATTCTGCTGAATGAAGAAGATAAAG AGAACATTGCACAGCTGACCGGTTATAACGGCACACATCGCCTGAGCCTG AAATGTATTCGTCTGGTACTGGAAGAACAGTGGTATAGCAGCCGTAATCA GATGGAAATCTTTACCCATCTGAACATTAAACCGAAGAAAATCAATCTGA CCGCAGCCAACAAAATTCCGAAAGCCATGATTGATGAGTTTATTCTGAGT CCGGTTGTGAAACGCACCTTTATTCAGAGCATCAACGTGATCAACAAAGT GATCGAGAAATATGGCATCCCCGAAGATATCATTATCGAACTGGCACGTG AAAATAACTCCGATGATCGCAAAAAGTTCATCAACAACCTGCAGAAAAA GAATGAAGCAACCCGCAAACGCATTAACGAAATTATTGGTCAGACCGGT AATCAGAATGCCAAACGTATTGTGGAAAAAATCCGTCTGCATGATCAGCA AGAGGGGAAATGTCTGTATAGCCTGGAAAGCATTGCCCTGATGGATCTGC TGAATAACCCGCAGAATTATGAAGTGGATCATATTATTCCGCGTAGCGTG GCATTTGATAATTCCATTCATAACAAAGTGCTGGTGAAGCAGATCGAGAA TAGCAAAAAAGGTAATCGTACGCCGTATCAGTATCTGAATAGCAGTGATG CAAAACTGAGCTACAACCAGTTTAAACAGCATATTCTGAATCTGAGCAAA AGCAAAGATCGCATCAGCAAAAAAAAGAAGGACTACCTGCTGGAAGAAC GCGATATCAACAAATTTGAAGTCCAGAAAGAGTTTATCAACCGCAATCTG GTTGATACCCGTTATGCAACCCGTGAACTGACCAGCTATCTGAAAGCATA TTTCAGCGCCAATAACATGGACGTGAAAGTGAAAACAATTAACGGCAGC TTTACCAACCATCTGCGTAAAGTTTGGCGCTTTGATAAATATCGCAACCA CGGCTATAAACATCATGCAGAAGATGCCCTGATTATTGCAAATGCAGATT TCCTGTTTAAAGAAAACAAAAAACTGAAAGCCGTCAACAGCGTGCTGGA AAAACCGGAAATTGAGACAAATCAGAATGACATTCAGGTTGATAGCGAA GATAATTACAGCGAAATGTTTATCATCCCGAAACAGGTGCAGGATATCAA AGATTTTCGCAACTTCAAATATAGCCACCGCGTTGACAAAAAACCTAATC GTCAGCTGATTAACGATACCCTGTATAGCACCCGCAAAAAAGATAACAG CACCTATATTGTGCAGACCATTAAAGACATCTACGCCAAAGATAATACCA CCCTGAAAAAACAGTTCGACAAAAGCCCAGAAAAATTTCTGATGTATCA GCATGATCCGCGTACCTTCGAAAAACTGGAAGTTATTATGAAACAGTATG CCAACGAGAAAAATCCGCTGGCCAAATATCACGAAGAAACCGGTGAATA TCTGACCAAATATTCCAAGAAGAACAACGGTCCGATCGTTAAATCCCTGA AATATATCGGTAATAAACTGGGCAGCCATCTGGATGTTACCCATCAGTTT AAAAGCTCCACAAAGAAGCTGGTTAAACTGTCCATCAAACCGTATCGCTT TGATGTGTATCTGACCGACAAAGGCTATAAATTCATTACCATCAGCTATC TGGACGTGCTGAAAAAAGACAACTATTATTATATCCCGGAACAGAAATAT GATAAACTGAAACTGGGTAAAGCCATCGATAAAAACGCCAAATTTATCG CCAGCTTCTACAAAAACGACCTGATTAAACTGGATGGCGAGATCTATAAA ATCATCGGTGTTAATAGCGACACCCGCAATATGATTGAGCTGGATCTGCC GGATATTCGCTATAAAGAATATTGCGAACTGAACAACATTAAAGGCGAA CCGCGTATCAAAAAGACCATCGGCAAAAAAGTGAATAGCATCGAGAAAC TGACCACCGATGTTCTGGGTAATGTGTTTACCAATACCCAGTATACCAAA CCTCAGCTGCTGTTCAAACGCGGTAAT	
30	MNQKFILGLDIGITSVGYGLIDYETKNIIDAGVRLFPEANVENNEGRRSKRGS RRLKRRRIHRLERVKKLLEDYNLLDQSQIPQSTNPYAIRVKGLSEALSKDELV	路邓葡萄球菌 **Cas9**

[0715]

	IALLHIAKRRGIHKIDVIDSNDDVGNELSTKEQLNKNSKLLKDKFVCQIQLER MNEGQVRGEKNRFKTADIIKEIIQLLNVQKNFHQLDENFINKYIELVEMRREY FEGPGKGSPYGWEGDPKAWYETLMGHCTYFPDELRSVKYAYSADLFNALN DLNNLVIQRDGLSKLEYHEKYHIIENVFKQKKKPTLKQIANEINVNPEDIKGY RITKSGKPQFTEFKLYHDLKSVLFDQSILENEDVLDQIAEILTIYQDKDSIKSKL TELDILLNEEDKENIAQLTGYTGTHRLSLKCIRLVLEEQWYSSRNQMEIFTHL NIKPKKINLTAANKIPKAMIDEFILSPVVKRTFGQAINLINKIIEKYGVPEDIIIEL ARENNSKDKQKFINEMQKKNENTRKRINEIIGKYGNQNAKRLVEKIRLHDEQ EGKCLYSLESIPLEDLLNNPNHYEVDHIIPRSVSFDNSYHNKVLVKQSENSKK SNLTPYQYFNSGKSKLSYNQFKQHILNLSKSQDRISKKKKEYLLEERDINKFE VQKEFINRNLVDTRYATRELTNYLKAYFSANNMNVKVKTINGSFTDYLRKV WKFKKERNHGYKHHAEDALIIANADFLFKENKKLKAVNSVLEKPEIETKQL DIQVDSEDNYSEMFIIPKQVQDIKDFRNFKYSHRVDKKPNRQLINDTLYSTRK KDNSTYIVQTIKDIYAKDNTTLKKQFDKSPEKFLMYQHDPRTFEKLEVIMKQ YANEKNPLAKYHEETGEYLTKYSKKNNGPIVKSLKYIGNKLGSHLDVTHQF KSSTKKLVKLSIKPYRFDVYLTDKGYKFITISYLDVLKKDNYYYIPEQKYDKL KLGKAIDKNAKFIASFYKNDLIKLDGEIYKIIGVNSDTRNMIELDLPDIRYKEY CELNNIKGEPRIKKTIGKKVNSIEKLTTDVLGNVFTNTQYTKPQLLFKRGN	
31	MKEKYILGLDLGITSVGYGIINFETKKIIDAGVRLFPEANVDNNEGRRSKRGS RRLKRRRIHRLERVKLLLTEYDLINKEQIPTSNNPYQIRVKGLSEILSKDELAIA LLHLAKRRGIHNINVSSEDEDASNELSTKEQINRNNKLLKDKYVCEVQLQRL KEGQIRGEKNRFKTTDILKEIDQLLKVQKDYHNLDIDFINQYKEIVETRREYF EGPGQGSPFGWNGDLKKWYEMLMGHCTYFPQELRSVKYAYSADLFNALND LNNLIIQRDNSEKLEYHEKYHIIENVFKQKKKPTLKQIAKEIGVNPEDIKGYRI TKSGTPQFTEFKLYHDLKSIVFDKSILENEAILDQIAEILTIYQDEQSIKEELNK LPEILNEQDKAEIAKLIGYNGTHRLSLKCIHLINEELWQTSRNQMEIFNYLNIK PNKVDLSEQNKIPKDMVNDFILSPVVKRTFIQSINVINKVIEKYGIPEDIIIELAR ENNSDDRKKFINNLQKKNEATRKRINEIIGQTGNQNAKRIVEKIRLHDQQEGK CLYSLESIALMDLLNNPQNYEVDHIIPRSVAFDNSIHNKVLVKQIENSKKGNR TPYQYLNSSDAKLSYNQFKQHILNLSKSKDRISKKKKDYLLEERDINKFEVQ KEFINRNLVDTRYATRELTSYLKAYFSANNMDVKVKTINGSFTNHLRKVWR FDKYRNHGYKHHAEDALIIANADFLFKENKKLQNTNKILEKPTIENNTKKVT VEKEEDYNNVFETPKLVEDIKQYRDYKFSHRVDKKPNRQLINDTLYSTRMK DEHDYIVQTITDIYGKDNTNLKKQFNKNPEKFLMYQNDPKTFEKLSIIMKQY SDEKNPLAKYYEETGEYLTKYSKKNNGPIVKKIKLLGNKVGNHLDVTNKYE NSTKKLVKLSIKNYRFDVYLTEKGYKFVTIAYLNVFKKDNYYYIPKDKYQEL KEKKKIKDTDQFIASFYKNDLIKLNGDLYKIIGVNSDDRNIIELDYYDIKYKD YCEINNIKGEPRIKKTIGKKTESIEKFTTDVLGNLYLHSTEKAPQLIFKRGL	巴氏葡萄球菌 Cas9
32	MNNYILGLDIGITSVGYGIVDSDTREIKDAGVRLFPEANVDNNEGRRSKRGA RRLKRRRIHRLDRVKHLLAEYDLLDLTNIPKSTNPYQTRVKGLNEKLSKDEL VIALLHIAKRRGIHNVNVMMDDNDSGNELSTKDQLKKNAKALSDKYVCELQ LERFEQDYKVRGEKNRFKTEDFVREARKLLETQSKFFEIDQTFIMRYIELIETR REYFEGPGKGSPFGWEGNIKKWFEQMMGHCTYFPEELRSVKYSYSAELFNA LNDLNNLVITRDEDAKLNYGEKFQIIENVFKQKKTPNLKQIAIEIGVHETEIKG YRVNKSGKPEFTQFKLYHDLKNIFKDPKYLNDIQLMDNIAEIITIYQDAESIIK ELNQLPELLSEREKEKISALSGYSGTHRLSLKCINLLLDDLWESSLNQMELFT KLNLKPKKIDLSQQHKIPSKLVDDFILSPVVKRAFIQSIQVVNAIIDKYGLPEDI IIELARENNSDDRRKFLNQLQKQNEETRKQVEKVLREYGNDNAKRIVQKIKL HNMQEGKCLYSLKDIPLEDLLRNPHHYEVDHIIPRSVAFDNSMHNKVLVRAD ENSKKGNRTPYQYLNSSESSLSYNEFKQHILNLSKTKDRITKKKREYLLEERD INKFDVQKEFINRNLVDTRYATRELTSLLKAYFSANNLDVKVKTINGSFTNYL RKVWKFDKDRNKGYKHHAEDALIIANADFLFKHNKKLRNINKVLDAPSKEV DKKRVTVQSEDEYNQIFEDTQKAQAIKKFEIRKFSHRVDKKPNRQLINDTLYS TRNIDGIEYVVESIKDIYSVNNDKVKTKFKKDPHRLLMYRNDPQTFEKFEKV FKQYESEKNPFAKYYEETGEKIRKFSKTGQGPYINKIKYLRERLGRHCDVTN KYINSRNKIVQLKIYSYRFDIYQYGNNYKMITISYIDLEQKSNYYYISREKYEQ KKKDKQIDDSYKFIGSFYKNDIINYNGEMYRVIGVNDSEKNKIQLDMIDISIK DYMELNNIKKTGVIYKTIGKSTTHIEKYTTDILGNLYKAAPPKKPQLIFK	猪葡萄球菌Cas9
33	MEKDYILGLDIGIGSVGYGLIDYDTKSIIDAGVRLFPEANADNNLGRRAKRGA RRLKRRRIHRLERVKSLLSEYKIISGLAPTNNQPYNIRVKGLTEQLTKDELAV ALLHIAKRRGIHNVDVAADKEETASDSLSTKDQINKNAKFLESRYVCELQKE	Staphylococcus microti Cas9

[0716]

	RLENEGHVRGVENRFLTKDIVREAKKIIDTQMQYYPEIDETFKEKYISLVETR REYYEGPGKGSPYGWDADVKKWYQLMMGHCTYFPVEFRSVKYAYTADLY NALNDLNNLTIARDDNPKLEYHEKYHIIENVFKQKRNPTLKQIAKEIGVNDIN ISGYRVTKSGKPQFTSFKLFHDLKKVVKDHAILDDIDLLNQIAEILTIYQDKDS IVAELGQLEYLMSEADKQSISELTGYTGTHSLSLKCMNMIIDELWHSSMNQM EVFTYLNMRPKKYELKGYQRIPTDMIDDAILSPVVKRSFKQAIGVVNAIIKKY GLPKDIIIELARESNSAEKSRYLRAIQKKNEKTRERIEAIIKEYGNENAKGLVQ KIKLHDAQEGKCLYSLKDIPLEDLLRNPNNYDIDHIIPRSVSFDDSMHNKVLV RREQNAKKNNQTPYQYLTSGYADIKYSVFKQHVLNLAENKDRMTKKKREY LLEERNINKYDVQKEFINRNLVDTRYTTRELTTLLKTYFTINNLDVKVKTING SFTDFLRKRWGFKKNRDEGYKHHAEDALIIANADYLFKEHKLLKEIKDVSDL AGDERNSNVKDEDQYEEVFGGYFKIEDIKKYKIKKFSHRVDKKPNRQLINDT IYSTRVKDDKRYLINTLKNLYDKSNGDLKERMQKDPESLLMYHHDPQTFEK LKIVMSQYENEKNPLAKYFEETGQYLTKYAKHDNGPAIHKIKYYGNKLVEH LDITKNYHNPQNKVVQLSQKSFRFDVYQTDKGYKFISIAYLTLKNEKNYYAI SQEKYDQLKSEKKISNNAVFIGSFYTSDIIEINNEKFRVIGVNSDKNNLIEVDRI DIRQKEFIELEEEKKNNRIKVTIGRKTTNIEKFHTDILGNMYKSKRPKAPQLVF KKG	
34	MNQKFILGLDIGITSVGYGLIDYETKNIIDAGVRLFPEANVENNEGRRSKRGS RRLKRRRIHRLERV	来自 8 个小结构域文库的路邓葡萄球菌小结构域 1
35	MKEKYILGLDLGITSVGYGIINFETKKIIDAGVRLFPEANVDNNEGRRSKRGS RRLKRRRIHRLERV	来自 8 个小结构域文库的巴氏葡萄球菌小结构域 1
36	MNNYILGLDIGITSVGYGIVDSDTREIKDAGVRLFPEANVDNNEGRRSKRGA RRLKRRRIHRLDRV	来自 8 个小结构域文库的猪葡萄球菌小结构域 1
37	MEKDYILGLDIGIGSVGYGLIDYDTKSIIDAGVRLFPEANADNNLGRRAKRGA RRLKRRRIHRLERV	来自 8 个小结构域文库的 Staphylococcus microti 小结构域 1
38	KKLLEDYNLLDQSQIPQSTNPYAIRVKGLSEALSKDELVIALLHIAKRRGIH	来自 8 个小结构域文库的路邓葡萄球菌小结构域 2
39	KLLLTEYDLINKEQIPTSNNPYQIRVKGLSEILSKDELAIALLHLAKRRGIH	来自 8 个小结构域文库的巴氏葡萄球菌小结构域 2
40	KHLLAEYDLLDLTNIPKSTNPYQTRVKGLNEKLSKDELVIALLHIAKRRGIH	来自 8 个小结构域文库的猪葡萄球菌小结构域 2
41	KSLLSEYKIISGLAPTNNQPYNIRVKGLTEQLTKDELAVALLHIAKRRGIH	来自 8 个小结构域文库的 Staphylococcus microti 小结构域 2
42	KIDVIDSNDDVGNELSTKEQLNKNSKLLKDKFVCQIQLERMNEGQVRGEKN RFKTADIIKEIIQLLNVQKNFHQLDENFINKYIELVEMRRE	来自 8 个小结构域文库的路邓葡萄球

[0717]

		菌小结构域 3
43	NINVSSEDEDASNELSTKEQINRNNKLLKDKYVCEVQLQRLKEGQIRGEKNR FKTTDILKEIDQLLKVQKDYHNLDIDFINQYKEIVETRRE	来自 8 个小结构域文库的巴氏葡萄球菌小结构域 3
44	NVNVMMDDNDSGNELSTKDQLKKNAKALSDKYVCELQLERFEQDYKVRG EKNRFKTEDFVREARKLLETQSKFFEIDQTFIMRYIELIETRRE	来自 8 个小结构域文库的猪葡萄球菌小结构域 3
45	NVDVAADKEETASDSLSTKDQINKNAKFLESRYVCELQKERLENEGHVRGV ENRFLTKDIVREAKKIIDTQMQYYPEIDETFKEKYISLVETRRE	来自 8 个小结构域文库的 Staphylococcus microti 小结构域 3
46	YFEGPGKGSPYGWEGDPKAWYETLMGHCTYFPDELRSVKYAYSADLFNAL NDLNNLVIQRDGLSKLEYHEKYHIIENVFKQKKKPTLKQIANEINVNPEDIKG YRITKSGK	来自 8 个小结构域文库的路邓葡萄球菌小结构域 4
47	YFEGPGQGSPFGWNGDLKKWYEMLMGHCTYFPQELRSVKYAYSADLFNAL NDLNNLIIQRDNSEKLEYHEKYHIIENVFKQKKKPTLKQIAKEIGVNPEDIKGY RITKSGT	来自 8 个小结构域文库的巴氏葡萄球菌小结构域 4
48	YFEGPGKGSPFGWEGNIKKWFEQMMGHCTYFPEELRSVKYSYSAELFNALN DLNNLVITRDEDAKLNYGEKFQIIENVFKQKKTPNLKQIAIEIGVHETEIKGYR VNKSGK	来自 8 个小结构域文库的猪葡萄球菌小结构域 4
49	YYEGPGKGSPYGWDADVKKWYQLMMGHCTYFPVEFRSVKYAYTADLYNA LNDLNNLTIARDDNPKLEYHEKYHIIENVFKQKRNPTLKQIAKEIGVNDINISG YRVTKSGK	来自 8 个小结构域文库的 Staphylococcus microti 小结构域 4
50	PQFTEIFKLYHDLKSVLFDQSILENEDVLDQIAEILTIYQDKDSIKSKLTELDILL NEEDKENIAQLTGYTGTHRLSLKCIRLVLEEQWYSSRNQMEIFTHLNIKPKKI NLTAANKIPKAMIDEFILSPVVK	来自 8 个小结构域文库的路邓葡萄球菌小结构域 5
51	PQFTEIFKLYHDLKSIVFDKSILENEAILDQIAEILTIYQDEQSIKEELNKLPEILN EQDKAEIAKLIGYNGTHRLSLKCIHLINEELWQTSRNQMEIFNYLNIKPNKVD LSEQNKIPKDMVNDFILSPVVK	来自 8 个小结构域文库的巴氏葡萄球菌小结构域 5
52	PEFTQFKLYHDLKNIFKDPKYLNDIQLMDNIAEIITIYQDAESIIKELNQLPELL SEREKEKISALSGYSGTHRLSLKCINLLLDDLWESSLNQMELFTKLNLKPKKI DLSQQHKIPSKLVDDFILSPVVK	来自 8 个小结构域文库的猪葡萄球菌小结构域 5
53	PQFTSFKLFHDLKKVVKDHAILDDIDLLNQIAEILTIYQDKDSIVAELGQLEYL MSEADKQSISELTGYTGTHSLSLKCMNMIIDELWHSSMNQMEVFTYLNMRP KKYELKGYQRIPTDMIDDAILSPVVK	来自 8 个小结构域文库的 Staphylococcus microti 小结构域 5
54	RTFGQAINLINKIIEKYGVPEDIIIELARENNSKDKQKFINEMQKKNENTRKRI NEIIGKYGNQNAKRLVEKIRLHDEQEGKCLYSLES	来自 8 个小结构域文库的路邓葡萄球菌小结构域 6

[0718]

55	RTFIQSINVINKVIEKYGIPEDIIIELARENNSDDRKKFINNLQKKNEATRKRINE IIGQTGNQNAKRIVEKIRLHDQQEGKCLYSLES	来自 8 个小结构域文库的巴氏葡萄球菌小结构域 6
56	RAFIQSIQVVNAIIDKYGLPEDIIIELARENNSDDRRKFLNQLQKQNEETRKQV EKVLREYGNDNAKRIVQKIKLHNMQEGKCLYSLKD	来自 8 个小结构域文库的猪葡萄球菌小结构域 6
57	RSFKQAIGVVNAIIKKYGLPKDIIIELARESNSAEKSRYLRAIQKKNEKTRERIE AIIKEYGNENAKGLVQKIKLHDAQEGKCLYSLKD	来自 8 个小结构域文库的 **Staphylococcus microti** 小结构域 6
58	IPLEDLLNNPNHYEVDHIIPRSVSFDNSYHNKVLVKQSENSKKSNLTPYQYFN SGKSKLSYNQFKQHILNLSKSQDRISKKKKEYLLEERDINKFEVQKEFINRNL VDTRYATREL	来自 8 个小结构域文库的路邓葡萄球菌小结构域 7
59	IALMDLLNNPQNYEVDHIIPRSVAFDNSIHNKVLVKQIENSKKGNRTPYQYLN SSDAKLSYNQFKQHILNLSKSKDRISKKKKDYLLEERDINKFEVQKEFINRNL VDTRYATREL	来自 8 个小结构域文库的巴氏葡萄球菌小结构域 7
60	IPLEDLLRNPHHYEVDHIIPRSVAFDNSMHNKVLVRADENSKKGNRTPYQYL NSSESSLSYNEFKQHILNLSKTKDRITKKKREYLLEERDINKFDVQKEFINRNL VDTRYATREL	来自 8 个小结构域文库的猪葡萄球菌小结构域 7
61	IPLEDLLRNPNNYDIDHIIPRSVSFDDSMHNKVLVRREQNAKKNNQTPYQYLT SGYADIKYSVFKQHVLNLAENKDRMTKKKREYLLEERNINKYDVQKEFINR NLVDTRYTTREL	来自 8 个小结构域文库的 **Staphylococcus microti** 小结构域 7
62	TNYLKAYFSANNMNVKVKTINGSFTDYLRKVWKFKKERNHGYKHHAEDAL IIANADFLFKENKKLKAVNSVLEKPEIETKQLDIQVDSEDNYSEMFIIPKQVQD IKDFRNFKYSHRVDKKPNRQLINDTLYSTRKKDNSTYIVQTIKDIYAKDNTTL KKQFDKSPEKFLMYQHDPRTFEKLEVIMKQYANEKNPLAKYHEETGEYLTK YSKKNNGPIVKSLKYIGNKLGSHLDVTHQFKSSTKKLVKLSIKPYRFDVYLT DKGYKFITISYLDVLKKDNYYYIPEQKYDKLKLGKAIDKNAKFIASFYKNDLI KLDGEIYKIIGVNSDTRNMIELDLPDIRYKEYCELNNIKGEPRIKKTIGKKVNSI EKLTTDVLGNVFTNTQYTKPQLLFKRGN	来自 8 个小结构域文库的路邓葡萄球菌小结构域 8
63	TSYLKAYFSANNMDVKVKTINGSFTNHLRKVWRFDKYRNHGYKHHAEDAL IIANADFLFKENKKLQNTNKILEKPTIENNTKKVTVEKEEDYNNVFETPKLVE DIKQYRDYKFSHRVDKKPNRQLINDTLYSTRMKDEHDYIVQTITDIYGKDNT NLKKQFNKNPEKFLMYQNDPKTFEKLSIIMKQYSDEKNPLAKYYEETGEYLT KYSKKNNGPIVKKIKLLGNKVGNHLDVTNKYENSTKKLVKLSIKNYRFDVY LTEKGYKFVTIAYLNVFKKDNYYYIPKDKYQELKEKKKIKDTDQFIASFYKN DLIKLNGDLYKIIGVNSDDRNIIELDYYDIKYKDYCEINNIKGEPRIKKTIGKKT ESIEKFTTDVLGNLYLHSTEKAPQLIFKRGL	来自 8 个小结构域文库的巴氏葡萄球菌小结构域 8
64	TSLLKAYFSANNLDVKVKTINGSFTNYLRKVWKFDKDRNKGYKHHAEDALI IANADFLFKHNKKLRNINKVLDAPSKEVDKKRVTVQSEDEYNQIFEDTQKAQ AIKKFEIRKFSHRVDKKPNRQLINDTLYSTRNIDGIEYVVESIKDIYSVNNDKV KTKFKKDPHRLLMYRNDPQTFEKFEKVFKQYESEKNPFAKYYEETGEKIRKF SKTGQGPYINKIKYLRERLGRHCDVTNKYINSRNKIVQLKIYSYRFDIYQYGN NYKMITISYIDLEQKSNYYYISREKYEQKKKDKQIDDSYKFIGSFYKNDIINYN GEMYRVIGVNDSEKNKIQLDMIDISIKDYMELNNIKKTGVIYKTIGKSTTHIEK YTTDILGNLYKAAPPKKPQLIFK	来自 8 个小结构域文库的猪葡萄球菌小结构域 8

[0719]

65	TTLLKTYFTINNLDVKVKTINGSFTDFLRKRWGFKKNRDEGYKHHAEDALII ANADYLFKEHKLLKEIKDVSDLAGDERNSNVKDEDQYEEVFGGYFKIEDIKK YKIKKFSHRVDKKPNRQLINDTIYSTRVKDDKRYLINTLKNLYDKSNGDLKE RMQKDPESLLMYHHDPQTFEKLKIVMSQYENEKNPLAKYFEETGQYLTKYA KHDNGPAIHKIKYYGNKLVEHLDITKNYHNPQNKVVQLSQKSFRFDVYQTD KGYKFISIAYLTLKNEKNYYAISQEKYDQLKSEKKISNNAVFIGSFYTSDIIEIN NEKFRVIGVNSDKNNLIEVDRIDIRQKEFIELEEEKKNNRIKVTIGRKTTNIEKF HTDILGNMYKSKRPKAPQLVFKKG	来自 8 个小结构域文库的 **Staphylococcus microti** 小结构域 8
66	MNQKFILGLDIGITSVGYGLIDYETKNIIDAGVRLFPEANV	来自 12 个小结构域文库的路邓葡萄球菌小结构域 1
67	MKEKYILGLDLGITSVGYGIINFETKKIIDAGVRLFPEANV	来自 12 个小结构域文库的巴氏葡萄球菌小结构域 1
68	MNNYILGLDIGITSVGYGIVDSDTREIKDAGVRLFPEANV	来自 12 个小结构域文库的猪葡萄球菌小结构域 1
69	MEKDYILGLDIGIGSVGYGLIDYDTKSIIDAGVRLFPEANA	来自 12 个小结构域文库的 **Staphylococcus microti** 小结构域 1
70	ENNEGRRSKRGSRRLKRRRIHRL	来自 12 个小结构域文库的路邓葡萄球菌小结构域 2
71	DNNEGRRSKRGSRRLKRRRIHRL	来自 12 个小结构域文库的巴氏葡萄球菌小结构域 2
72	DNNEGRRSKRGARRLKRRRIHRL	来自 12 个小结构域文库的猪葡萄球菌小结构域 2
73	DNNLGRRAKRGARRLKRRRIHRL	来自 12 个小结构域文库的 **Staphylococcus microti** 小结构域 2
74	ERVKKLLEDYNLLDQSQIPQSTNPYAIRVKGLSEALSKDELVIALLHIAKRRGI H	来自 12 个小结构域文库的路邓葡萄球菌小结构域 3
75	ERVKLLLTEYDLINKEQIPTSNNPYQIRVKGLSEILSKDELAIALLHLAKRRGI H	来自 12 个小结构域文库的巴氏葡萄球菌小结构域 3
76	DRVKHLLAEYDLLDLTNIPKSTNPYQTRVKGLNEKLSKDELVIALLHIAKRR GIH	来自 12 个小结构域文库的猪葡萄球菌

[0720]

		小结构域 3
77	ERVKSLLSEYKIISGLAPTNNQPYNIRVKGLTEQLTKDELAVALLHIAKRRGIH	来自 12 个小结构域文库的 Staphylococcus microti 小结构域 3
78	KIDVIDSNDDVGNELSTKEQLNKNSKLLKDKFVCQIQLERMNEGQVRGEKN RFKTADIIKEIIQLLNVQKNFHQLDENFINKYIELVEMRREY	来自 12 个小结构域文库的路邓葡萄球菌小结构域 4
79	NINVSSEDEDASNELSTKEQINRNNKLLKDKYVCEVQLQRLKEGQIRGEKNR FKTTDILKEIDQLLKVQKDYHNLDIDFINQYKEIVETRREY	来自 12 个小结构域文库的巴氏葡萄球菌小结构域 4
80	NVNVMMDDNDSGNELSTKDQLKKNAKALSDKYVCELQLERFEQDYKVRG EKNRFKTEDFVREARKLLETQSKFFEIDQTFIMRYIELIETRREY	来自 12 个小结构域文库的猪葡萄球菌小结构域 4
81	NVDVAADKEETASDSLSTKDQINKNAKFLESRYVCELQKERLENEGHVRGV ENRFLTKDIVREAKKIIDTQMQYYPEIDETFKEKYISLVETRREY	来自 12 个小结构域文库的 Staphylococcus microti 小结构域 4
82	FEGPGKGSPYGWEGDPKAWYETLMGHCTYFPDELRSVKYAYSADLFNALN DLNNLVIQRDGLSKLEYHEKYHIIENVFKQKKKPTLKQIANEINVNPEDIKGY RITKSGKPQFT	来自 12 个小结构域文库的路邓葡萄球菌小结构域 5
83	FEGPGQGSPFGWNGDLKKWYEMLMGHCTYFPQELRSVKYAYSADLFNALN DLNNLIIQRDNSEKLEYHEKYHIIENVFKQKKKPTLKQIAKEIGVNPEDIKGYR ITKSGTPQFT	来自 12 个小结构域文库的巴氏葡萄球菌小结构域 5
84	FEGPGKGSPFGWEGNIKKWFEQMMGHCTYFPEELRSVKYSYSAELFNALND LNNLVITRDEDAKLNYGEKFQIIENVFKQKKTPNLKQIAIEIGVHETEIKGYRV NKSGKPEFT	来自 12 个小结构域文库的猪葡萄球菌小结构域 5
85	YEGPGKGSPYGWDADVKKWYQLMMGHCTYFPVEFRSVKYAYTADLYNAL NDLNNLTIARDDNPKLEYHEKYHIIENVFKQKRNPTLKQIAKEIGVNDINISG YRVTKSGKPQFT	来自 12 个小结构域文库的 Staphylococcus microti 小结构域 5
86	EFKLYHDLKSVLFDQSILENEDVLDQIAEILTIYQDKDSIKSKLTELDILLNEED KENIAQLTGYTGTHRLSLKCIRLVLEEQWYSSRNQMEIFTHLNIKPKKINLTA ANKIPKAMIDEFILSPVVKR	来自 12 个小结构域文库的路邓葡萄球菌小结构域 6
87	EFKLYHDLKSIVFDKSILENEAILDQIAEILTIYQDEQSIKEELNKLPEILNEQDK AEIAKLIGYNGTHRLSLKCIHLINEELWQTSRNQMEIFNYLNIKPNKVDLSEQ NKIPKDMVNDFILSPVVKR	来自 12 个小结构域文库的巴氏葡萄球菌小结构域 6
88	QFKLYHDLKNIFKDPKYLNDIQLMDNIAEIITIYQDAESIIKELNQLPELLSERE KEKISALSGYSGTHRLSLKCINLLLDDLWESSLNQMELFTKLNLKPKKIDLSQ QHKIPSKLVDDFILSPVVKR	来自 12 个小结构域文库的猪葡萄球菌小结构域 6

[0721]

89	SFKLFHDLKKVVKDHAILDDIDLLNQIAEILTIYQDKDSIVAELGQLEYLMSE ADKQSISELTGYTGTHSLSLKCMNMIIDELWHSSMNQMEVFTYLNMRPKKY ELKGYQRIPTDMIDDAILSPVVKR	来自12个小结构域文库的Staphylococcus microti小结构域6
90	TFGQAINLINKIIEKYGVPEDIIIELARENNSKDKQKFINEMQKKNENTRKRINE IIGKYGNQNAKRLVEKIRLHDEQEGKCLYSL	来自12个小结构域文库的路邓葡萄球菌小结构域7
91	TFIQSINVINKVIEKYGIPEDIIIELARENNSDDRKKFINNLQKKNEATRKRINEII GQTGNQNAKRIVEKIRLHDQQEGKCLYSL	来自12个小结构域文库的巴氏葡萄球菌小结构域7
92	AFIQSIQVVNAIIDKYGLPEDIIIELARENNSDDRRKFLNQLQKQNEETRKQVE KVLREYGNDNAKRIVQKIKLHNMQEGKCLYSL	来自12个小结构域文库的猪葡萄球菌小结构域7
93	SFKQAIGVVNAIIKKYGLPKDIIIELARESNSAEKSRYLRAIQKKNEKTRERIEA IIKEYGNENAKGLVQKIKLHDAQEGKCLYSL	来自12个小结构域文库的Staphylococcus microti小结构域7
94	ESIPLEDLLNNPNHYEVDHIIPRSVSFDNSYHNKVLVKQSENSKKSNLTPYQY FNSGKSKLSYNQFKQHILNLSKSQDRISKKKKEYLLEER	来自12个小结构域文库的路邓葡萄球菌小结构域8
95	ESIALMDLLNNPQNYEVDHIIPRSVAFDNSIHNKVLVKQIENSKKGNRTPYQY LNSSDAKLSYNQFKQHILNLSKSKDRISKKKKDYLLEER	来自12个小结构域文库的巴氏葡萄球菌小结构域8
96	KDIPLEDLLRNPHHYEVDHIIPRSVAFDNSMHNKVLVRADENSKKGNRTPYQ YLNSSESSLSYNEFKQHILNLSKTKDRITKKKREYLLEER	来自12个小结构域文库的猪葡萄球菌小结构域8
97	KDIPLEDLLRNPNNYDIDHIIPRSVSFDDSMHNKVLVRREQNAKKNNQTPYQ YLTSGYADIKYSVFKQHVLNLAENKDRMTKKKREYLLEER	来自12个小结构域文库的Staphylococcus microti小结构域8
98	DINKFEVQKEFINRNLVDTRYATRELT	来自12个小结构域文库的路邓葡萄球菌小结构域9
99	DINKFEVQKEFINRNLVDTRYATRELT	来自12个小结构域文库的巴氏葡萄球菌小结构域9
100	DINKFDVQKEFINRNLVDTRYATRELT	来自12个小结构域文库的猪葡萄球菌小结构域9

	101	NINKYDVQKEFINRNLVDTRYTTRELT	来自12个小结构域文库的Staphylococcus microti小结构域9
	102	NYLKAYFSANNMNVKVKTINGSFTDYLRKVWKFKKERNHGYKHHAEDALIIANADFLFKENKKL	来自12个小结构域文库的路邓葡萄球菌小结构域10
	103	SYLKAYFSANNMDVKVKTINGSFTNHLRKVWRFDKYRNHGYKHHAEDALIIANADFLFKENKKL	来自12个小结构域文库的巴氏葡萄球菌小结构域10
	104	SLLKAYFSANNLDVKVKTINGSFTNYLRKVWKFDKDRNKGYKHHAEDALIIANADFLFKHNKKL	来自12个小结构域文库的猪葡萄球菌小结构域10
	105	TLLKTYFTINNLDVKVKTINGSFTDFLRKRWGFKKNRDEGYKHHAEDALIIANADYLFKEHKLL	来自12个小结构域文库的Staphylococcus microti小结构域10
	106	KAVNSVLEKPEIETKQLDIQVDSEDNYSEMFIIPKQVQDIKDFRNFKYSHRVDKKPNRQLINDTLYSTRKKDNSTYIVQTIKDIYAKDNTTLKKQFDKSPEKFLMYQHDPRTFEKLEVIMKQYANEKNPLAKYHEETGEYLTKYSKKNNGPIVKSLKYIGNKLGSHLDVTHQFKSSTKKLVKLSIK	来自12个小结构域文库的路邓葡萄球菌小结构域11
[0722]	107	QNTNKILEKPTIENNTKKVTVEKEEDYNNVFETPKLVEDIKQYRDYKFSHRVDKKPNRQLINDTLYSTRMKDEHDYIVQTITDIYGKDNTNLKKQFNKNPEKFLMYQNDPKTFEKLSIIMKQYSDEKNPLAKYYEETGEYLTKYSKKNNGPIVKKIKLLGNKVGNHLDVTNKYENSTKKLVKLSIK	来自12个小结构域文库的巴氏葡萄球菌小结构域11
	108	RNINKVLDAPSKEVDKKRVTVQSEDEYNQIFEDTQKAQAIKKFEIRKFSHRVDKKPNRQLINDTLYSTRNIDGIEYVVESIKDIYSVNNDKVKTKFKKDPHRLLMYRNDPQTFEKFEKVFKQYESEKNPFAKYYEETGEKIRKFSKTGQGPYINKIKYLRERLGRHCDVTNKYINSRNKIVQLK	来自12个小结构域文库的猪葡萄球菌小结构域11
	109	KEIKDVSDLAGDERNSNVKDEDQYEEVFGGYFKIEDIKKYKIKKFSHRVDKKPNRQLINDTIYSTRVKDDKRYLINTLKNLYDKSNGDLKERMQKDPESLLMYHHDPQTFEKLKIVMSQYENEKNPLAKYFEETGQYLTKYAKHDNGPAIHKIKYYGNKLVEHLDITKNYHNPQNKVVQLSQK	来自12个小结构域文库的Staphylococcus microti小结构域11
	110	PYRFDVYLTDKGYKFITISYLDVLKKDNYYYIPEQKYDKLKLGKAIDKNAKFIASFYKNDLIKLDGEIYKIIGVNSDTRNMIELDLPDIRYKEYCELNNIKGEPRIKKTIGKKVNSIEKLTTDVLGNVFTNTQYTKPQLLFKRGN	来自12个小结构域文库的路邓葡萄球菌小结构域12
	111	NYRFDVYLTEKGYKFVTIAYLNVFKKDNYYYIPKDKYQELKEKKKIKDTDQFIASFYKNDLIKLNGDLYKIIGVNSDDRNIIELDYYDIKYKDYCEINNIKGEPRIKKTIGKKTESIEKFTTDVLGNLYLHSTEKAPQLIFKRGL	来自12个小结构域文库的巴氏葡萄球菌小结构域12
	112	IYSYRFDIYQYGNNYKMITISYIDLEQKSNYYYISREKYEQKKKDKQIDDSYKFIGSFYKNDIINYNGEMYRVIGVNDSEKNKIQLDMIDISIKDYMELNNIKKTGVIYKTIGKSTTHIEKYTTDILGNLYKAAPPKKPQLIFK	来自12个小结构域文库的猪葡萄球菌小结构域12
	113	SFRFDVYQTDKGYKFISIAYLTLKNEKNYYAISQEKYDQLKSEKKISNNAVFI	来自12个小结构域

[0723]

	GSFYTSDIIEINNEKFRVIGVNSDKNNLIEVDRIDIRQKEFIELEEEKKNNRIKV TIGRKTTNIEKFHTDILGNMYKSKRPKAPQLVFKKG	文库的 Staphylococcus microti 小结构域 12
114	GGACCCCCUCCACCCCGCCUCGUUUUAGUACUCUGGAAACAGAAUCUAC UGAAACAAGACAAUAUGUCGUGUUUAUCCCAUCAAUUUAUUGGUGGGGA UUUUUUU	sgRNAVEGFA
115	ACCCCCTCCACCCCGCCTCCGGGCGCG	具有 "CGGGCGC"-PAM 的 ATTO647N 寡核苷酸
116	GCGCGCGCCCGGAGGCGGGGTGGAGGGGGTCGG	具有 "CGGGCGC"-PAM 的生物素化的寡核苷酸
117	GGGUCAGGGUGGUCACGAGG	靶向 BFP 的指导 RNA
118	CCACTGTAGAAGAGCAAATGCCACCTGACGTCTAAGAAATTCGCGTTAAA TTTTTGTTAAATCAGCTCATTTTTTAACCAATAGGCCGAAATCGGCAAAAT CCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGGCCGCTAC ACTAGGGCGCGTCTAATACGACTCACTATAGGACCCCCTCCACCCCGCCT CGTTTTAGTACTCTGGAAACAGAATCTACTGAAACAAGACAATATGTCGT GTTTATCCCATCAATTTATTGGTGGGATTTTTTTCTAGCATAACCCCTTGG GGCCTCTAAACGGGTCTTGAGGGGTTTTTTGTCACTGCCCGCTTTCCAGTC GGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGTTCGTCAGAACTGCT TACGCGGTTTGTCACTCGGTCGCTACGCTCCGGGCGTGAGACTGCGGCGG GCGCTGCGGACACATACAAAGTTACCCACAGATTCCGTGGATAAGCAGG GGACTAACATGTGAGGCAAAACAGCAGGGCCGCGCCGGTGGCGTTTTTC CATAGGCTCCGCCCTCCTGCCAGAGTTCACATAAACAGACGCTTTTCCGG TGCATCTGTGGGAGCCGTGAGGCTCAACCATGAATCTGACAGTACGGGCG AAACCCGACAGGACTTAAAGATCCCCACCGTTTCCGGCGGGTCGCTCCCT CTTGCGCTCTCCTGTTCCGACCCTGCCGTTTACCGGATACCTGTTCCGCCT TTCTCCCTTACGGGAAGTGTGGCGCTTTCTCATAGCTCACACACTGGTATC TCGGCTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTAAGCAAGAACTC CCCGTTCAGCCCGACTGCTGCGCCTTATCCGGTAACTGTTCACTTGAGTCC AACCCGGAAAAGCACGGTAAAACGCCACTGGCAGCAGCCATTGGTAACT GGGAGTTCGCAGAGGATTTGTTTAGCTAAACACGCGGTTGCTCTTGAAGT GTGCGCCAAAGTCCGGCTACACTGGAAGGACAGATTTGGTTGCTGTGCTC TGCGAAAGCCAGTTACCACGGTTAAGCAGTTCCCCAACTGACTTAACCTT CGATCAAACCACCTCCCCAGGTGGTTTTTTCGTTTACAGGGCAAAAGATT ACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACTGA ACCGCTCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTATTAG AAAAACTCATCGAGCATCAAATGAAACTGCAATTTATTCATATCAGGATT ATCAATACCATATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAAACT CACCGAGGCAGTTCCATAGGATGGCAAGATCCTGGTATCGGTCTGCGATT CCGACTCGTCCAACATCAATACAACCTATTAATTTCCCCTCGTCAAAAAT AAGGTTATCAAGTGAGAAATCACCATGAGTGACGACTGAATCCGGTGAG AATGGCAAAAGTTTATGCATTTCTTTCCAGACTTGTTCAACAGGCCAGCC ATTACGCTCGTCATCAAAATCACTCGCATCAACCAAACCGTTATTCATTC GTGATTGCGCCTGAGCGAGACGAAATACGCGATCGCTGTTAAAAGGACA ATTACAAACAGGAATCGAATGCAACCGGCGCAGGAACACTGCCAGCGCA TCAACAATATTTTCACCTGAATCAGGATATTCTTCTAATACCTGGAATGCT GTTTTCCCGGGGATCGCAGTGGTGAGTAACCATGCATCATCAGGAGTACG GATAAAATGCTTGATGGTCGGAAGAGGCATAAATTCCGTCAGCCAGTTTA GTCTGACCATCTCATCTGTAACATCATTGGCAACGCTACCTTTGCCATGTT TCAGAAACAACTCTGGCGCATCGGGCTTCCCATACAATCGATAGATTGTC GCACCTGATTGCCCGACATTATCGCGAGCCCATTTATACCCATATAAATC AGCATCCATGTTGGAATTTAATCGCGGCCTAGAGCAAGACGTTTCCCGTT GAATATGGCTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGG	pCas606

[0724]

	GTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAA CAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTG	
119	CCGGTGATACCACGATACTATGACTGAGAGTCAACGCCATGAGCGGCCTC ATTTCTTATTCTGAGTTACAACAGTCCGCACCGCTGCCGGTAGCTATTGAC TATCCGGCTGCACTAGCCCTGCGTCAGATGGCTCTGATCCAAGGCAAACT GCCAAAATATCTGCTGGCACCGGAAGTCAGCGCCCTGCACCATTATGTTC CGGATCTGCATCGCAGGATGCTGCTGGCTACCCTGTGGAACACCTACATC TGTATTAACGAAGCGCTGGCATTGACCCTGAGTGATTTTTCTCTGGTGCCG CCCTATCCCTTTGTGCAGCTTGCCACGCTCAAAGGGGTTTGAGGTCCAAC CGTACGAAAACGTACGGTAAGAGGAAAATTATCGTCTGAAAAATCGATT AGTAGACAAGAAAGTCCGTTAAGTGCCAATTTTCGATTAAAAAGACACC GTTTTGATGGCGTTTTCCAATGTACATTATGTTTCGATATATCAGACAGTT ACTTCACTAACGTACGTTTTCGTTCTATTGGCCTTCAGACCCCATATCCTT AATGTCCTTTATTTGCTGGGGTTATCAGATCCCCCCGACACGTTTAATTAA TGCTTTCTCCGCCGGAGATCGACGCACAGCGTTCTGTGCTCTATGATGTTA TTTCTTAATAATCATCCAGGTATTCTCTTTATCACCATACGTAGTGCGAGT GTCCACCTTAACGCAGGGCTTTCCGTCACAGCGCGATATGTCAGCCAGCG GGGCTTTCTTTTGCCAGACCGCTTCCATCCTCTGCATTTCAGCAATCTGGC TATACCCGTCATTCATAAACCACGTAAATGCCGTCACGCAGGAAGCCAGG ACGAAGAATATCGTCAGTACAAGATAAATCGCGGATTTCCACGTATAGCG TGACATCTCACGACGCATTTCATGGATCATCGCTTTCGCCGTATCGGCAG CCTGATTCAGCGCTTCTGTCGCCGGTTTCTGCTGTGCTAATCCGGCTTGTT TCAGTTCTTTCTCAACCTGAGTGAGCGCGGAACTCACCGATTTCCTGACG GTGTCAGTCATATTACCGGACGCGCTGTCCAGCTCACGAATGACCCTGCT CAGCGTTTCACTTTGCTGCTGTAATTGTGATGAGGCGGCCTGAAACTGTTC TGTCAGAGAAGTAACACGCTTTTCCAGCGCCTGATGATGCCCGATAAGGG CGGCAATTTGTTTAATTTCGTCGCTCATACAAAATCCTGCCTATCGTGAGA ATGACCAGCCTTTATCCGGCTTCTGTCGTATCTGTTCGGCGAGTCGCTGTC GTTCTTTCTCCTGCTGACGCTGTTTTTCCGCCAGACGTTCGCGCTCTCTCTG CCTTTCCATCTCCTGATGTATCCCCTGGAACTCCGCCATCGCATCGTTAAC AAGGGACTGAAGATCGATTTCTTCCTGTATATCCTTCATGGCATCACTGA CCAGTGCGTTCAGCTTGTCAGGCTCTTTTTCAAAATCAAACGTTCTGCCGG AATGGGATTCCTGCTCAGGCTCTGACTTCAGCTCCTGTTTTAGCGTCAGAG TATCCCTCTCGCTGAGGGCTTCCCGTAACGAGGTAGTCACGTCAATTACG CTGTCACGTTCATCACGGGACTGCTGCACCTGCCTTTCAGCCTCCCTGCGC TCAAGAATGGCCTGTAGCTGCTCAGTATCGAATCGCTGAACCTGACCCGC GCCCAGATGCCGCTCAGGCTCACGGTCAATGCCCTGCGCCTTCAGGGAAC GGGAATCAACCCGGTCAGCGTGCTGATACCGTTCAAGGTGCTTATTCTGG AGGTCAGCCCAGCGTTCCCTCTGGGCAACAAGGTATTCTTTGCGTTCGGT CGGTGTTTCCCCGAAACGTGCCTTTTTTGCGCCACCGCGCTCCGGCTCTTT GGTGTTAGCCCGTTTAAAATACTGCTCAGGGTCACGGTGAATACCGTCAT TAATGCGTTCAGAGAACATGATATGGGCGTGGGGCTGCTCGCCACCGGCT ATCGCTGCTTTCGGATTATGGATAGCGAACTGATAGGCATGGCGGTCGCC AATTTCCTGTTGGACAAAATCGCGGACAAGCTCAAGACGTTGTTCGGGTT TTAACTCACGCGGCAGGGCAATCTCGATTTCACGGTAGGTACAGCCGTTG GCACGTTCAGACGTGTCAGCGGCTTTCCAGAACTCGGACGGTTTATGCGC TGCCCACGCCGGCATATTGCCGGACTCCTTGTGCTCAAGGTCGGAGTCTT TTTCACGGGCATACTTTCCCTCACGCGCAATATAATCGGCATGAGGAGAG GCACTGCCTTTTCCGCCGGTTTTTACGCTGAGATGATAGGATGCCATCGTG TTTTATCCCGCTGAAGGCGCGCACCGTTTCTGAACGAAGTGAAGAAACGT CTAAGTGCGCCCTGATAAATAAAAGAGTTATCAGGGATTGTAGTGGGATT TGACCTCCTCTGCCATCACTGAGCATAATCATTCCGTTAGCATTCAGGAG GTAAACAGCATGAATAAAAGCGAAAAACAGGAACAATGGGCAGCAGAA AGAGTGCAGTATATTCGCGGCTTAAAGTCGCCGAATGAGCAACAGAAAC TTATGCTGATACTGACGGATAAAGCAGATAAAACAGCACAGGATATCAA AACGCTGTCCCTGCTGATGAAGGCTGAACAGGCAGCAGAGAAAGCGCAG GAAGCCAGAGCGAAAGTCATGAACCTGATACAGGCAGAAAAGCGAGCCG AAGCCAGAGCCGCCCGTAAAGCCCGTGACCATGCTCTGTACCAGTCTGCC GGATTGCTTATCCTGGCGGGTCTGGTTGACAGTAAGACGGGTAAGCCTGT TGATGATACCGCTGCCTTACTGGGTGCATTAGCCAGTCTGAATGACCTGT CACGGGATAATCCGAAGTGGTCAGACTGGAAAATCAGAGGGCAGGAACT	pCas634

[0725]

GCTGAACAGCAAAAAGTCAGATAGCACCACATAGCAGACCCGCCATAAA
ACGCCCTGAGAAGCCCGTGACGGGCTTTTCTTGTATTATGGGTAGTTTCCT
TGCATGAATCCATAAAAGGCGCCTGTAGTGCCATTTACCCCCATTCACTG
CCAGAGCCGTGAGCGCAGCGAACTGAATGTCACGAAAAAGACAGCGACT
CAGGTGCCTGATGGTCGGAGACAAAAGGAATATTCAGCGATTTGCCCGA
GCTTGCGAGGGTGCTACTTAAGCCTTTAGGGTTTTAAGGTCTGTTTTGTAG
AGGAGCAAACAGCGTTTGCGACATCCTTTTGTAATACTGCGGAACTGACT
AAAGTAGTGAGTTATACACAGGGCTGGGATCTATTCTTTTTATCTTTTTTT
ATTCTTTCTTTATTCTATAAATTATAACCACTTGAATATAAACAAAAAAAA
CACACAAAGGTCTAGCGGAATTTACAGAGGGTCTAGCAGAATTTACAAG
TTTTCCAGCAAAGGTCTAGCAGAATTTACAGATACCCACAACTCAAAGGA
AAAGGACTAGTAATTATCATTGACTAGCCCATCTCAATTGGTATAGTGAT
TAAAATCACCTAGACCAATTGAGATGTATGTCTGAATTAGTTGTTTTCAA
AGCAAATGAACTAGCGATTAGTCGCTATGACTTAACGGAGCATGAAACC
AAGCTAATTTTATGCTGTGTGGCACTACTCAACCCCACGATTGAAAACCC
TACAAGGAAAGAACGGACGGTATCGTTCACTTATAACCAATACGCTCAG
ATGATGAACATCAGTAGGGAAAATGCTTATGGTGTATTAGCTAAAGCAAC
CAGAGAGCTGATGACGAGAACTGTGGAAATCAGGAATCCTTTGGTTAAA
GGCTTTGAGATTTTCCAGTGGACAAACTATGCCAAGTTCTCAAGCGAAAA
ATTAGAATTAGTTTTTAGTGAAGAGATATTGCCTTATCTTTTCCAGTTAAA
AAAATTCATAAAATATAATCTGGAACATGTTAAGTCTTTTGAAAACAAAT
ACTCTATGAGGATTTATGAGTGGTTATTAAAAGAACTAACACAAAAGAA
AACTCACAAGGCAAATATAGAGATTAGCCTTGATGAATTTAAGTTCATGT
TAATGCTTGAAAATAACTACCATGAGTTTAAAAGGCTTAACCAATGGGTT
TTGAAACCAATAAGTAAAGATTTAAACACTTACAGCAATATGAAATTGGT
GGTTGATAAGCGAGGCCGCCCGACTGATACGTTGATTTTCCAAGTTGAAC
TAGATAGACAAATGGATCTCGTAACCGAACTTGAGAACAACCAGATAAA
AATGAATGGTGACAAAATACCAACAACCATTACATCAGATTCCTACCTAC
ATAACGGACTAAGAAAAACACTACACGATGCTTTAACTGCAAAAATTCA
GCTCACCAGTTTTGAGGCAAAATTTTTGAGTGACATGCAAAGTAAGTATG
ATCTCAATGGTTCGTTCTCATGGCTCACGCAAAAACAACGAACCACACTA
GAGAACATACTGGCTAAATACGGAAGGATCTGAGGTTCTTATGGCTCTTG
TATCTATCAGTGAAGCATCAAGACTAACAAACAAAAGTAGAACAACTGT
TCACCGTTACATATCAAAGGGAAAACTGTCCATATGCACAGATGAAAAC
GGTGTAAAAAAGATAGATACATCAGAGCTTTTACGAGTTTTTGGTGCATT
CAAAGCTGTTCACCATGAACAGATCGACAATGTAACAGATGAACAGCAT
GTAACACCTAATAGAACAGGTGAAACCAGTAAAACAAAGCAACTAGAAC
ATGAAATTGAACACCTGAGACAACTTGTTACAGCTCAACAGTCACACATA
GACAGCCTGAAACAGGCGATGCTGCTTATCGAATCAAAGCTGCCGACAA
CACGGGAGCCAGTGACGCCTCCCGTGGGGAAAAAATCATGGCAATTCTG
GAAGAAATAGCGCTTTCAGCCGGCAAACCGGCTGAAGCCGGATCTGCGA
TTCTGATAACAAACTAGCAACACCAGAACAGCCCGTTTGCGGGCAGCAA
AACCCGTACTTTTGGACGTTCCGGCGGTTTTTTGTGGCGAGTGGTGTTCGG
GCGGTGCGCGCAAGATCCATTATGTTAAACGGGCGAGTTTACATCTCAAA
ACCGCCCGCTTAACACCATCAGAAATCCTCAGCGCGATTTTAAGCACCAA
CCCCCCCCCGTAACACCCAAATCCATACTGAAAGTGGCTTTGTTGAATAA
ATCGAACTTTTGCTGAGTTGAAGGATCAGATCACGCATCCTCCCGACAAC
ACAGACCATTCCGTGGCAAAGCAAAAGTTCAGAATCACCAACTGGTCCA
CCTACAACAAAGCTCTCATCAACCGTGGCTCCCTCACTTTCTGGCTGGAT
GATGAGGCGATTCAGGCCTGGTATGAGTCGGCAACACCTTCATCACGAGG
AAGGCCCCAGCGCTATTCTGATCTCGCCATCACCACCGTTCTGGTGATTA
AACGCGTATTCCGGCTGACCCTGCGGGCTGCGCAGGGTTTTATTGATTCC
ATTTTTGCCCTGATGAACGTTCCGTTGCGCTGCCCGGATTACACCAGTGTC
AGTAAGCGGGCAAAGTCGGTTAATGTCAGTTTCAAAACGTCCACCCGGG
GTGAAATCGCACACCTGGTGATTGATTCCACCGGGCTGAAGGTCTTTGGT
GAAGGCGAATGGAAAGTCAGAAAGCACGGCAAAGAGCGCCGTCGTATCT
GGCGAAAGTTGCATCTTGCTGTTGACAGCAACACACATGAAGTTGTCTGT
GCAGACCTGTCGCTGAATAACGTCACGGACTCAGAAGCCTTCCCGGGCCT
TATCCGGCAGACTCACAGAAAAATCAGGGCAGCCGCGGCAGACGGGGCT
TACGATACCCGGCTCTGTCACGATGAACTGCGCCGCAAAAAAATCAGCGC
GCTTATTCCTCCCCGAAAAGGTGCGGGTTACTGGCCCGGTGAATATGCAG

[0726]

	ACCGTAACCGTGCAGTGGCTAATCAGCGAATGACCGGGAGTAATGCGCG GTGGAAATGGACAACAGATTACAACCGTCGCTCGATAGCGGAAACGGCG ATGTACCGGGTAAAACAGCTGTTCGGGGGTTCACTGACGCTGCGTGACTA CGATGGTCAGGTTGCGGAGGCTATGGCCCTGGTACGAGCGCTGAACAAA ATGACGAAAGCAGGTATGCCTGAAAGCGTGCGTATTGCCTGAAAACACA ACCCGCTACGGGGGAGACTTACCCGAAATCTGATTTATTCAACAAAGCCG GGTGTGGTGAACTACAAAGCAGACCCGTTGAGGTTATCAGTTCGATGCAC AATCAGCAGCGCATAAAATATGCACAAGAACAGGAGCACCCTTCGCATT AAGCTGTGGTGGTAACAAGTAGTGCCGGGCTACCATCAGCGAGCATGAT GCGCTCCCACAGCATTCGCCTTGGCAGTATGGAAGTTCCTCGCTCCAGTT CGGGCCGGTATCCACCTCGAGCTTTCTACGGGGTCTGACGCTCAGTGGAA CGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCT TCACCTAGATCCTTTTCGACCGAATAAATACCTGTGACGGAAGATCACTT CGCAGAATAAATAAATCCTGGTGTCCCTGTTGATACCGGGAAGCCCTGGG CCAACTTTTGGCGAAAATGAGACGTTGATCGGCACGTAAGAGGTTCCAAC TTTCACCATAATGAAATAAGATCACTACCGGGCGTATTTTTTGAGTTGTCG AGATTTTCAGGAGCTAAGGAAGCTAAAATGGAGAAAAAAATCACTGGAT ATACCACCGTTGATATATCCCAATGGCATCGTAAAGAACATTTTGAGGCA TTTCAGTCAGTTGCTCAATGTACCTATAACCAGACCGTTCAGCTGGATATT ACGGCCTTTTTAAAGACCGTAAAGAAAAATAAGCACAAGTTTTATCCGGC CTTTATTCACATTCTTGCCCGCCTGATGAATGCTCATCCGGAATTACGTAT GGCAATGAAAGACGGTGAGCTGGTGATATGGGATAGTGTTCACCCTTGTT ACACCGTTTTCCATGAGCAAACTGAAACGTTTTCATCGCTCTGGAGTGAA TACCACGACGATTTCCGGCAGTTTCTACACATATATTCGCAAGATGTGGC GTGTTACGGTGAAAACCTGGCCTATTTCCCTAAAGGGTTTATTGAGAATA TGTTTTTCGTCTCAGCCAATCCCTGGGTGAGTTTCACCAGTTTTGATTTAA ACGTGGCCAATATGGACAACTTCTTCGCCCCCGTTTTCACCTTGGGCAAA TATTATACGCAAGGCGACAAGGTGCTGATGCCGCTGGCGATTCAGGTTCA TCATGCCGTTTGTGATGGCTTCCATGTCGGCAGAATGCTTAATGAATTAC AACAGTACTGCGATGAGTGGCAGGGCGGGGCGTAATTTTTTTAAGGCAGT TATTGGTGCCCTTAAACGCCTGGTTGCTACGCCTGAATAAGTGATAATAA GCGGATGAATGGCAGAAATTCGAAAGCAAATTCGACCCGGTCGTCGGTT CAGGGCAGGGTCGTTAAATAGCCGCTTATGTCTATTGCTGGTTTACCGGA TTATTGACTACCGGAAGCAGTGTGACCGTGTGCTTCTCAAATGCCTGAGG CCAGTTTGCTCAGGCTCTCCCCGTGGAGGTAATAATTGACGATATGATCC TTTTTTTCTGATCAAAAGTGCTCCATGGAATTATGACAACTTGACGGCTAC ATCATTCACTTTTTCTTCACAACCGGCACGGAACTCGCTCGGGCTGGCCCC GGTGCATTTTTTAAATACCCGCGAGAAATAGAGTTGATCGTCAAAACCAA CATTGCGACCGACGGTGGCGATAGGCATCCGGGTGGTGCTCAAAAGCAG CTTCGCCTGGCTGATACGTTGGTCCTCGCGCCAGCTTAAGACGCTAATCC CTAACTGCTGGCGGAAAAGATGTGACAGACGCGACGGCGACAAGCAAAC ATGCTGTGCGACGCTGGCGATATCAAAATTGCTGTCTGCCAGGTGATCGC TGATGTACTGACAAGCCTCGCGTACCCGATTATCCATCGGTGGATGGAGC GACTCGTTAATCGCTTCCATGCGCCGCAGTAACAATTGCTCAAGCAGATT TATCGCCAGCAGCTCCGAATAGCGCCCTTCCCCTTGCCCGGCGTTAATGA TTTGCCCAAACAGGTCGCTGAAATGCGGCTGGTGCGCTTCATCCGGGCGA AAGAACCCCGTATTGGCAAATATTGACGGCCAGTTAAGCCATTCATGCCA GTAGGCGCGCGGACGAAAGTAAACCCACTGGTGATACCATTCGCGAGCC TCCGGATGACGACCGTAGTGATGAATCTCTCCTGGCGGGAACAGCAAAAT ATCACCCGGTCGGCAAACAAATTCTCGTCCCTGATTTTTCACCACCCCCTG ACCGCGAATGGTGAGATTGAGAATATAACCTTTCATTCCCAGCGGTCGGT CGATAAAAAAATCGAGATAACCGTTGGCCTCAATCGGCGTTAAACCCGCC ACCAGATGGGCATTAAACGAGTATCCCGGCAGCAGGGGATCATTTTGCGC TTCAGCCATACTTTTCATACTCCCGCCATTCAGAGAAGAAACCAATTGTC CATATTGCATCAGACATTGCCGTCACTGCGTCTTTTACTGGCTCTTCTCGC TAACCAAACCGGAAACCCCGCTTATTAAAAGCATTCTGTAACAAAGCGG GACCAAAGCCATGACAAAAACGCGTAACAAAAGTGTCTATAATCACGGC AGAAAAGTCCACATTGATTATTTGCACGGCGTCACACTTTGCTATGCCAT AGCATTTTTATCCATAAGATTAGCGGATCCTACCTGACGCTTTTTATCGCA ACTCTCTACTGTCGTGCGCAACGCGTGACATCGACTAGGAGGCCTTTCTA TGCAGTTCAAAGTGTATACCTATAAACGCGAAAGCCGTTATCGTCTGTTT	

[0727]

	GTTGATGTTCAGAGCGATATTATTGATACACCGGGTCGTCGTATGGTTATT CCGCTGGCAAGCGCACGTCTGCTGAGCGATAAAGTTAGCCGTGAACTGTA TCCGGTTGTTCATATTGGTGATGAAAGCTGGCGTATGATGACCACCGATA TGGCAAGCGTTCCGGTTAGCGTTATTGGTGAAGAAGTTGCAGATCTGAGC CATCGTGAAAACGATATCAAAAATGCCATCAACCTGATGTTTTGGGGCAT CTTATAGGACCCCCTCCACCCCGCCTCGCGGGCAGTTTGCCTGGCGGCAG TAGCGCGGTGGTCCCACCTGACCCCATGCCGAACTCAGAAGTGAAACGCC GTAGCGCCGATGGTAGTGTGGGCTCTCCCCATGCGAGAGTAGGGAACTGC CAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTA	
120	GGTGGAAGCGGAGGCAGCGGGGGATCAGGCCATCATCATCACCATCATT AATGAACGGGTCTTGAGGGGTTTTTTGGACTTTTGCGTATTGGGCGCTCTT CCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGA GCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGG GGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAG GAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCC CTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCC GACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGC GCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCC CTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTT CGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTT CAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCC GGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATT AGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGC CTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTG AAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACA AACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGC GCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCT GACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATT ATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTA AATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGC TTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATA GTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACC ATCTGGCCCCAGTGCTGCAATGATACCGCGAGATCCACGCTCACCGGCTC CAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAG TGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGA AGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCA TTGCTGCAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCA GCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGC AAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTT GGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTAC TGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCA AGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCG TCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCAT CATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGT TGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCA TCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAA TGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATA CTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATG AGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCC GCGCACATTTCCCCGAAAAGTGCCACTCCTCTTGCTGTGACCGTCTCCGG GAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGAGGC AGCAGATCAATTCGCGCGCGAAGGCGAAGCGGCATGCATTTACGTTGAC ACCATCGAATGGTGCAAAACCTTTCGCGGTATGGCATGATAGCGCCCGGA AGAGAGTCAATTCAGGGTGGTGAATGTGAAACCAGTAACGTTATACGAT GTCGCAGAGTATGCCGGTGTCTCTTATCAGACCGTTTCCCGCGTGGTGAA CCAGGCCAGCCACGTTTCTGCGAAAACGCGGGAAAAAGTGGAAGCGGCG ATGGCGGAGCTGAATTACATTCCCAACCGCGTGGCACAACAACTGGCGG GCAAACAGTCGTTGCTGATTGGCGTTGCCACCTCCAGTCTGGCCCTGCAC GCGCCGTCGCAAATTGTCGCGGCGATTAAATCTCGCGCCGATCAACTGGG TGCCAGCGTGGTGGTGTCGATGGTAGAACGAAGCGGCGTCGAAGCCTGT AAAGCGGCGGTGCACAATCTTCTCGCGCAACGCGTCAGTGGGCTGATCAT	pCas81

[0728]

	TAACTATCCGCTGGATGACCAGGATGCCATTGCTGTGGAAGCTGCCTGCA	
	CTAATGTTCCGGCGTTATTTCTTGATGTCTCTGACCAGACACCCATCAACA	
	GTATTATTTTCTCCCATGAAGATGGTACGCGACTGGGCGTGGAGCATCTG	
	GTCGCATTGGGTCACCAGCAAATCGCGCTGTTAGCGGGCCCATTAAGTTC	
	TGTCTCGGCGCGTCTGCGTCTGGCTGGCTGGCATAAATATCTCACTCGCA	
	ATCAAATTCAGCCGATAGCGGAACGGGAAGGCGACTGGAGTGCCATGTC	
	CGGTTTTCAACAAACCATGCAAATGCTGAATGAGGGCATCGTTCCCACTG	
	CGATGCTGGTTGCCAACGATCAGATGGCGCTGGGCGCAATGCGCGCCATT	
	ACCGAGTCCGGGCTGCGCGTTGGTGCGGATATCTCGGTAGTGGGATACGA	
	CGATACCGAAGATAGCTCATGTTATATCCCGCCGTTAACCACCATCAAAC	
	AGGATTTTCGCCTGCTGGGGCAAACCAGCGTGGACCGCTTGCTGCAACTC	
	TCTCAGGGCCAGGCGGTGAAGGGCAATCAGCTGTTGCCCGTCTCACTGGT	
	GAAAAGAAAAACCACCCTGGCGCCCAATACGCAAACCGCCTCTCCCCGC	
	GCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGA	
	AAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATTAGG	
	TCCGGCTCGTATAATGTGTGGAATTGTGAGCGGATAACAATTTTCACACA	
	GGAAACAGAATTCTAGCATTGTGAGCGGATAACAATTCCCCTCTAGAAAT	
	AATTTTGTTTAACTTTAAGAAGGAGATATACCATGAAACGTCCGGCAGCA	
	ACCAAAAAAGCAGGTCAGGCCAAGAAAAAAAAAGGTGGTGGTTCAGGTA	
	ACCAGAAATTTATCCTGGGTCTGGATATTGGTATTACCAGCGTTGGTTAT	
	GGCCTGATTGATTACGAAACCAAAAACATTATTGATGCCGGTGTTCGTCT	
	GTTTCCGGAAGCAAATGTTGAAAATAATGAAGGTCGTCGTAGCAAACGT	
	GGTAGCCGTCGTCTGAAACGTCGTCGTATTCATCGTCTGGAACGTGTTAA	
	AAAACTGCTGGAAGATTATAACCTGCTGGATCAGAGCCAGATTCCGCAG	
	AGCACCAATCCGTATGCAATTCGTGTTAAAGGTCTGAGCGAAGCACTGAG	
	CAAAGATGAACTGGTTATTGCACTGCTGCATATTGCAAAACGCCGTGGCA	
	TTCATAAAATCGATGTGATTGATAGCAATGACGATGTGGGTAATGAACTG	
	AGCACCAAAGAACAGCTGAACAAAAATAGCAAACTGCTGAAAGACAAAT	
	TCGTGTGTCAGATTCAGCTGGAACGTATGAATGAAGGCCAGGTTCGTGGT	
	GAAAAGAATCGCTTTAAAACCGCAGACATCATCAAAGAAATTATCCAGC	
	TGCTGAACGTGCAGAAAAACTTCCATCAGCTGGATGAAAACTTCATCAAC	
	AAATACATCGAGCTGGTTGAAATGCGTCGCGAATATTTTGAAGGTCCGGG	
	TAAAGGTAGCCCGTATGGTTGGGAAGGTGATCCGAAAGCATGGTATGAA	
	ACCCTGATGGGTCATTGTACCTATTTTCCGGATGAACTGCGTAGCGTTAA	
	ATATGCCTATAGCGCAGACCTGTTTAATGCACTGAATGATCTGAATAACC	
	TGGTGATTCAGCGTGATGGTCTGAGCAAACTGGAATATCATGAGAAATAT	
	CACATCATCGAAAACGTGTTCAAACAGAAGAAGAAACCGACCCTGAAAC	
	AAATCGCCAACGAAATTAATGTGAACCCGGAAGATATTAAAGGCTACCG	
	TATTACCAAAAGCGGTAAACCGCAGTTCACCGAATTTAAACTGTATCACG	
	ATCTGAAAAGCGTGCTGTTTGATCAGAGCATTCTGGAAAATGAAGATGTG	
	CTGGACCAGATTGCAGAAATTCTGACCATTTATCAGGACAAAGACAGCAT	
	CAAAAGCAAACTGACCGAACTGGATATTCTGCTGAATGAAGAAGATAAA	
	GAGAACATTGCACAGCTGACCGGTTATACCGGCACCCATCGTCTGAGCCT	
	GAAATGTATTCGTCTGGTACTGGAAGAACAGTGGTATAGCAGCCGTAATC	
	AGATGGAAATCTTTACCCATCTGAACATTAAACCGAAGAAAATCAATCTG	
	ACCGCAGCCAACAAAATTCCGAAAGCCATGATTGATGAGTTTATTCTGAG	
	TCCGGTTGTGAAACGTACCTTTGGTCAGGCAATTAACCTGATCAACAAAA	
	TCATTGAAAAATATGGCGTGCCTGAGGATATCATTATTGAACTGGCACGT	
	GAAAACAACAGCAAAGATAAACAGAAATTCATCAACGAGATGCAGAAG	
	AAGAACGAAAATACCCGCAAACGGATTAACGAGATCATTGGCAAATATG	
	GTAATCAGAATGCCAAACGCCTGGTGGAAAAAATTCGTCTGCATGATGA	
	ACAAGAGGGCAAATGTCTGTATAGCCTGGAAAGCATTCCTCTGGAAGATC	
	TGCTGAACAATCCGAATCATTATGAAGTGGATCACATTATTCCGCGTAGC	
	GTGAGCTTTGATAATTCCTATCATAATAAAGTGCTGGTGAAACAGAGCGA	
	AAACTCCAAAAAATCCAACCTGACACCGTATCAGTATTTCAATAGCGGCA	
	AATCCAAACTGAGCTACAACCAGTTTAAACAGCATATTCTGAACCTGAGC	
	AAAAGCCAGGATCGCATCAGCAAGAAGAAGAAGGAGTACCTGCTGGAAG	
	AACGCGACATCAACAAATTTGAAGTGCAGAAAGAATTTATCAACCGCAA	
	CCTGGTTGATACCCGTTATGCAACCCGTGAACTGACCAATTATCTGAAAG	
	CATATTTCAGCGCCAACAACATGAACGTGAAAGTGAAAACGATTAACGG	
	CAGCTTTACCGATTATCTGCGTAAAGTGTGGAAATTCAAAAAAGAACGCA	

[0729]

	ACCACGGCTATAAACATCATGCCGAAGATGCCCTGATTATTGCAAATGCA GATTTCCTGTTTAAAGAAAACAAAAAACTGAAAGCCGTCAACAGCGTGCT GGAAAAACCGGAAATTGAGACAAAACAGCTGGACATTCAGGTTGATAGC GAAGATAATTACAGCGAAATGTTTATCATCCCGAAACAGGTGCAGGATAT CAAAGATTTTCGCAACTTCAAATATAGCCACCGCGTTGACAAAAAACCTA ATCGTCAGCTGATTAACGATACCCTGTATAGCACCCGCAAAAAAGATAAC AGCACCTATATTGTGCAGACCATTAAAGACATCTACGCCAAAGATAATAC CACCCTGAAAAAACAGTTCGACAAAAGCCCAGAAAAATTTCTGATGTATC AGCATGATCCGCGTACCTTCGAAAAACTGGAAGTTATTATGAAACAGTAT GCCAACGAGAAAAATCCGCTGGCCAAATATCACGAAGAAACCGGTGAAT ATCTGACCAAATATTCCAAGAAGAACAACGGTCCGATCGTTAAATCCCTG AAATATATCGGTAATAAACTGGGCAGCCATCTGGATGTTACCCATCAGTT TAAAAGCTCCACAAAGAAGCTGGTTAAACTGTCCATCAAACCGTATCGCT TTGATGTGTATCTGACCGACAAAGGCTATAAATTCATTACCATCAGCTAT CTGGACGTGCTGAAAAAAGACAACTATTATTATATCCCGGAACAGAAAT ATGATAAACTGAAACTGGGTAAAGCCATCGATAAAAACGCCAAATTTAT CGCCAGCTTCTACAAAAACGACCTGATTAAACTGGATGGCGAGATCTATA AAATCATCGGTGTTAATAGCGACACCCGCAATATGATTGAGCTGGATCTG CCGGATATTCGCTATAAAGAATATTGCGAACTGAACAACATTAAAGGCG AACCGCGTATCAAAAAGACCATCGGCAAAAAAGTGAATAGCATCGAGAA ACTGACCACCGATGTTCTGGGTAATGTGTTTACCAATACCCAGTATACCA AACCTCAGCTGCTGTTCAAACGCGGTAATGGCGGAGGATCTGGCCCCCCT AAGAAAAAGCGGAAGGTG	
121	TGCCACCTCCAGTCTGGCCCTGCACGCGCCGTCGCAAATTGTCGCGGCGA TTAAATCTCGCGCCGATCAACTGGGTGCCAGCGTGGTGGTGTCGATGGTA GAACGAAGCGGCGTCGAAGCCTGTAAAGCGGCGGTGCACAATCTTCTCG CGCAACGCGTCAGTGGGCTGATCATTAACTATCCGCTGGATGACCAGGAT GCCATTGCTGTGGAAGCTGCCTGCACTAATGTTCCGGCGTTATTTCTTGAT GTCTCTGACCAGACACCCATCAACAGTATTATTTTCTCCCATGAAGATGG TACGCGACTGGGCGTGGAGCATCTGGTCGCATTGGGTCACCAGCAAATCG CGCTGTTAGCGGGCCCATTAAGTTCTGTCTCGGCGCGTCTGCGTCTGGCTG GCTGGCATAAATATCTCACTCGCAATCAAATTCAGCCGATAGCGGAACGG GAAGGCGACTGGAGTGCCATGTCCGGTTTTCAACAAACCATGCAAATGCT GAATGAGGGCATCGTTCCCACTGCGATGCTGGTTGCCAACGATCAGATGG CGCTGGGCGCAATGCGCGCCATTACCGAGTCCGGGCTGCGCGTTGGTGCG GATATCTCGGTAGTGGGATACGACGATACCGAAGATAGCTCATGTTATAT CCCGCCGTTAACCACCATCAAACAGGATTTTCGCCTGCTGGGGCAAACCA GCGTGGACCGCTTGCTGCAACTCTCTCAGGGCCAGGCGGTGAAGGGCAAT CAGCTGTTGCCCGTCTCACTGGTGAAAAGAAAAACCACCCTGGCGCCCAA TACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGG CACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACATATTCTGA AATGAGCTGTTGACAATTAATCATCCGGCTCGTATAATGTGTGGAATTGT GAGCGGATAACAATTTTCACACAGGAAACAGAATTCTAGCATTGTGAGC GGATAACAATTCCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGA TATACCATGAAACGTCCGGCAGCAACCAAAAAAGCAGGTCAGGCCAAGA AAAAAAAAGGTGGTGGTTCAGGCAACAACTATATTCTGGGTCTGGATATT GGTATTACCAGCGTTGGTTATGGTATTGTTGATAGCGATACCCGTGAAAT TAAAGATGCCGGTGTTCGTCTGTTTCCGGAAGCAAATGTTGATAATAATG AAGGTCGTCGTAGCAAACGTGGTGCACGTCGTCTGAAACGTCGTCGTATT CATCGTCTGGATCGTGTTAAACATCTGCTGGCAGAATATGATCTGCTGGA TCTGACCAATATTCCGAAAAGCACCAATCCGTATCAGACCCGTGTTAAAG GTCTGAATGAAAAGCTGAGCAAAGATGAACTGGTTATTGCACTGCTGCAT ATTGCAAAACGCCGTGGCATTCATAATGTGAATGTTATGATGGATGATAA CGATAGCGGTAATGAACTGAGCACCAAAGATCAGCTGAAGAAGAATGCA AAAGCACTGAGCGATAAATATGTTTGTGAACTGCAGCTGGAACGCTTTGA GCAGGATTATAAAGTTCGTGGTGAAAAGAACCGCTTCAAAACCGAAGAT TTTGTTCGTGAAGCACGTAAACTGCTGGAAACCCAGAGCAAATTTTTCGA AATTGATCAGACGTTCATCATGCGCTATATCGAACTGATTGAAACCCGTC GCGAATATTTTGAAGGTCCGGGTAAAGGTAGCCCGTTTGGTTGGGAAGGT AATATCAAGAAATGGTTTGAGCAGATGATGGGCCACTGTACCTATTTTCC AGAAGAACTGCGTAGCGTCAAATATAGCTATTCAGCCGAACTGTTTAACG	pCas888

[0730]

CCCTGAATGATCTGAATAATCTGGTGATTACCCGTGATGAAGATGCCAAA
CTGAATTATGGTGAGAAATTCCAGATCATCGAAAACGTGTTCAAACAGAA
GAAAACACCGAACCTGAAACAAATCGCCATTGAAATTGGTGTGCATGAA
ACCGAAATCAAAGGTTATCGTGTGAACAAAAGCGGCAAACCGGAATTTA
CCCAGTTCAAACTGTATCACGATCTGAAGAACATCTTCAAAGACCCGAAA
TACCTGAACGATATCCAGCTGATGGATAACATTGCAGAAATCATCACCAT
TTATCAGGATGCCGAAAGCATCATCAAAGAACTGAATCAGCTGCCGGAA
CTGCTGAGCGAACGTGAAAAGGAGAAGATCAGCGCACTGAGCGGTTATA
GCGGCACCCATCGTCTGAGCCTGAAATGTATTAATCTGCTGCTGGATGAT
CTGTGGGAAAGCAGCCTGAATCAGATGGAACTGTTCACAAAACTGAATCT
GAAACCGAAGAAAATCGATCTGAGCCAGCAGCATAAAATTCCGAGCAAA
CTGGTGGATGACTTTATTCTGAGTCCGGTTGTTAAACGTGCATTTATTCAG
AGTATCCAGGTGGTGAATGCCATCATTGATAAATACGGTCTGCCGGAAGA
TATCATCATTGAACTGGCACGTGAAAACAATAGTGATGATCGTCGCAAAT
TTCTGAACCAGCTGCAAAAGCAGAATGAGGAAACCCGTAAACAGGTTGA
AAAGGTGCTGCGTGAATATGGCAATGATAATGCCAAACGTATCGTGCAG
AAAATCAAACTGCATAATATGCAAGAGGGGAAATGTCTGTATAGCCTGA
AAGATATTCCGCTGGAAGATCTGCTGCGTAATCCGCATCATTATGAAGTG
GATCATATTATTCCGCGTAGCGTGGCATTTGATAATAGCATGCATAATAA
AGTTCTGGTGCGTGCCGATGAGAATAGCAAGAAAGGTAATCGTACCCCGT
ATCAGTATCTGAATAGCAGCGAAAGCAGTCTGAGCTATAACGAATTTAAA
CAGCATATCCTGAACCTGAGCAAAACCAAAGACCGCATCACCAAGAAGA
AGCGTGAATACCTGCTGGAAGAACGCGACATCAACAAATTTGATGTGCA
GAAAGAATTCATCAACCGCAATCTGGTTGATACACGTTATGCAACCCGTG
AACTGACCAGCCTGCTGAAAGCATATTTCAGCGCAAATAACCTGGACGTG
AAAGTGAAAACGATTAATGGCAGCTTTACCAACTATCTGCGTAAAGTGTG
GAAATTCGATAAAGATCGCAACAAAGGCTATAAACATCATGCAGAAGAT
GCCCTGATTATTGCCAATGCAGATTTTCTGTTCAAACACAACAAGAAACT
GCGCAACATTAACAAAGTGCTGGATGCACCGAGCAAAGAAGTTGATAAG
AAACGCGTGACAGTTCAGAGCGAAGATGAATATAACCAGATTTTTGAGG
ATACCCAGAAAGCCCAGGCAATCAAGAAATTCGAAATCCGCAAATTTAG
CCACCGCGTGGACAAGAAACCGAATCGTCAGCTGATTAATGATACCCTGT
ATAGCACCCGTAATATCGATGGTATTGAATATGTGGTCGAGAGCATCAAA
GATATCTATAGCGTGAACAACGATAAAGTTAAAACCAAATTCAAGAAAG
ACCCGCACCGTCTGCTGATGTATCGTAATGATCCGCAGACCTTTGAGAAA
TTCGAGAAAGTCTTTAAACAGTATGAGAGCGAGAAGAACCCGTTCGCCA
AATATTACGAAGAAACCGGTGAGAAAATTCGCAAATTCAGCAAAACCGG
TCAGGGTCCGTATATCAACAAAATCAAATATCTGCGTGAACGCCTGGGTC
GTCATTGTGATGTTACCAACAAATATATCAATAGCCGCAACAAAATTGTG
CAACTGTCCATCAAACCGTATCGCTTTGATGTGTATCTGACCGACAAAGG
CTATAAATTCATTACCATCAGCTATCTGGACGTGCTGAAAAAAGACAACT
ATTATTATATCCCGGAACAGAAATATGATAAACTGAAACTGGGTAAAGCC
ATCGATAAAAACGCCAAATTTATCGCCAGCTTCTACAAAAACGACCTGAT
TAAACTGGATGGCGAGATCTATAAAATCATCGGTGTTAATAGCGACACCC
GCAATATGATTGAGCTGGATCTGCCGGATATTCGCTATAAAGAATATTGC
GAACTGAACAACATTAAAGGCGAACCGCGTATCAAAAAGACCATCGGCA
AAAAAGTGAATAGCATCGAGAAACTGACCACCGATGTTCTGGGTAATGT
GTTTACCAATACCCAGTATACCAAACCTCAGCTGCTGTTCAAACGCGGTA
ATGGCGGAGGATCTGGCCCCCCTAAGAAAAAGCGGAAGGTGGGTGGAAG
CGGAGGCAGCGGGGGATCAGGCCATCATCATCACCATCATTAATGAACG
GGTCTTGAGGGGTTTTTTGGACTTTTGCGTATTGGGCGCTCTTCCGCTTCC
TCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATC
AGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAAC
GCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGT
AAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGA
GCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGA
CTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCT
GTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGA
AGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTA
GGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCG
ACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGA

[0731]

	CACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAG CGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTAC GGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGT TACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCG CTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAA AAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCA GTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAA GGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCT AAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGT GAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGA CTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCC CAGTGCTGCAATGATACCGCGAGATCCACGCTCACCGGCTCCAGATTTAT CAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGC AACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGT AAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTGCAG GCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTT CCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCG GTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGT GTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCC ATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTG AGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGG ATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAA CGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAG TTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTT CACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAA AAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTT TCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACAT ATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTC CCCGAAAAGTGCCACTCCTCTTGCTGTGACCGTCTCCGGGAGCTGCATGT GTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGAGGCAGCAGATCAAT TCGCGCGCGAAGGCGAAGCGGCATGCATTTACGTTGACACCATCGAATG GTGCAAAACCTTTCGCGGTATGGCATGATAGCGCCCGGAAGAGAGTCAA TTCAGGGTGGTGAATGTGAAACCAGTAACGTTATACGATGTCGCAGAGTA TGCCGGTGTCTCTTATCAGACCGTTTCCCGCGTGGTGAACCAGGCCAGCC ACGTTTCTGCGAAAACGCGGGAAAAAGTGGAAGCGGCGATGGCGGAGCT GAATTACATTCCCAACCGCGTGGCACAACAACTGGCGGGCAAACAGTCG TTGCTGATTGGCGT	
122	CGGAACGGGAAGGCGACTGGAGTGCCATGTCCGGTTTTCAACAAACCAT GCAAATGCTGAATGAGGGCATCGTTCCCACTGCGATGCTGGTTGCCAACG ATCAGATGGCGCTGGGCGCAATGCGCGCCATTACCGAGTCCGGGCTGCGC GTTGGTGCGGATATCTCGGTAGTGGGATACGACGATACCGAAGATAGCTC ATGTTATATCCCGCCGTTAACCACCATCAAACAGGATTTTCGCCTGCTGG GGCAAACCAGCGTGGACCGCTTGCTGCAACTCTCTCAGGGCCAGGCGGTG AAGGGCAATCAGCTGTTGCCCGTCTCACTGGTGAAAAGAAAAACCACCCT GGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAA TGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCA ATATTCTGAAATGAGCTGTTGACAATTAATCATCCGGCTCGTATAATGTG TGGAATTGTGAGCGGATAACAATTTTCACACAGGAAACAGAATTCTAGCA TTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTTTGTTTAACTTTAAG AAGGAGATATACCATGAAACGTCCGGCAGCAACCAAAAAAGCAGGTCAG GCCAAGAAAAAAAAAAGGTGGTGGTTCAGGCAACAACTATATTCTGGGTC TGGATATTGGTATTACCAGCGTTGGTTATGGTATTGTTGATAGCGATACCC GTGAAATTAAAGATGCCGGTGTTCGTCTGTTTCCGGAAGCAAATGTTGAT AATAATGAAGGTCGTCGTAGCAAACGTGGTGCACGTCGTCTGAAACGTCG TCGTATTCATCGTCTGGATCGTGTTAAACATCTGCTGGCAGAATATGATCT GCTGGATCTGACCAATATTCCGAAAAGCACCAATCCGTATCAGACCCGTG TTAAAGGTCTGAATGAAAAGCTGAGCAAAGATGAACTGGTTATTGCACTG CTGCATATTGCAAAACGCCGTGGCATTCATAATGTGAATGTTATGATGGA TGATAACGATAGCGGTAATGAACTGAGCACCAAAGATCAGCTGAAGAAG AATGCAAAAGCACTGAGCGATAAATATGTTTGTGAACTGCAGCTGGAAC GCTTTGAGCAGGATTATAAAGTTCGTGGTGAAAAGAACCGCTTCAAAACC	pCas889

[0732]

	GAAGATTTTGTTCGTGAAGCACGTAAACTGCTGGAAACCCAGAGCAAATT TTTCGAAATTGATCAGACGTTCATCATGCGCTATATCGAACTGATTGAAA CCCGTCGCGAATATTTTGAAGGTCCGGGTAAAGGTAGCCCGTTTGGTTGG GAAGGTAATATCAAGAAATGGTTTGAGCAGATGATGGGCCACTGTACCT ATTTTCCAGAAGAACTGCGTAGCGTCAAATATAGC	
123	CTTGTACTTATACCTAAAATTACAGAATCTACTGAAACAAGACAATATGT CGTGTTTATCCCATCAATTTATTGGTGGGATTTTTTTATGTTTTTAGCAAA AAGTAATACCATACTTTATATTTTTAAATTATAATAAAGATATAAATAAA GGTGG	SluCas9 tracr RNA
124	ATGAAACGTCCGGCAGCAACCAAAAAAGCAGGTCAGGCCAAGAAAAAA AAAGGTGGTGGTTCAGGTAACCAGAAATTTATCCTGGGTCTGGATATTGG TATTACCAGCGTTGGTTATGGCCTGATTGATTACGAAACCAAAAACATTA TTGATGCCGGTGTTCGTCTGTTTCCGGAAGCAAATGTTGAAAATAATGAA GGTCGTCGTAGCAAACGTGGTAGCCGTCGTCTGAAACGTCGTCGTATTCA TCGTCTGGAACGTGTTAAAAAACTGCTGGAAGATTATAACCTGCTGGATC AGAGCCAGATTCCGCAGAGCACCAATCCGTATGCAATTCGTGTTAAAGGT CTGAGCGAAGCACTGAGCAAAGATGAACTGGTTATTGCACTGCTGCATAT TGCAAAACGCCGTGGCATTCATAAAATCGATGTGATTGATAGCAATGACG ATGTGGGTAATGAACTGAGCACCAAAGAACAGCTGAACAAAAATAGCAA ACTGCTGAAAGACAAATTCGTGTGTCAGATTCAGCTGGAACGTATGAATG AAGGCCAGGTTCGTGGTGAAAAGAATCGCTTTAAAACCGCAGACATCAT CAAAGAAATTATCCAGCTGCTGAACGTGCAGAAAAACTTCCATCAGCTGG ATGAAAACTTCATCAACAAATACATCGAGCTGGTTGAAATGCGTCGCGAA TATTTTGAAGGTCCGGGTAAAGGTAGCCCGTATGGTTGGGAAGGTGATCC GAAAGCATGGTATGAAACCCTGATGGGTCATTGTACCTATTTTCCGGATG AACTGCGTAGCGTTAAATATGCCTATAGCGCAGACCTGTTTAATGCACTG AATGATCTGAATAACCTGGTGATTCAGCGTGATGGTCTGAGCAAACTGGA ATATCATGAGAAATATCACATCATCGAAAACGTGTTCAAACAGAAGAAG AAACCGACCCTGAAACAAATCGCCAACGAAATTAATGTGAACCCGGAAG ATATTAAAGGCTACCGTATTACCAAAAGCGGTAAACCGCAGTTCACCGAA TTTAAACTGTATCACGATCTGAAAAGCGTGCTGTTTGATCAGAGCATTCT GGAAAATGAAGATGTGCTGGACCAGATTGCAGAAATTCTGACCATTTATC AGGACAAAGACAGCATCAAAAGCAAACTGACCGAACTGGATATTCTGCT GAATGAAGAAGATAAAGAGAACATTGCACAGCTGACCGGTTATACCGGC ACCCATCGTCTGAGCCTGAAATGTATTCGTCTGGTACTGGAAGAACAGTG GTATAGCAGCCGTAATCAGATGGAAATCTTTACCCATCTGAACATTAAAC CGAAGAAAATCAATCTGACCGCAGCCAACAAAATTCCGAAAGCCATGAT TGATGAGTTTATTCTGAGTCCGGTTGTGAAACGTACCTTTGGTCAGGCAA TTAACCTGATCAACAAAATCATTGAAAAATATGGCGTGCCTGAGGATATC ATTATTGAACTGGCACGTGAAAACAACAGCAAAGATAAACAGAAATTCA TCAACGAGATGCAGAAGAAGAACGAAAATACCCGCAAACGGATTAACGA GATCATTGGCAAATATGGTAATCAGAATGCCAAACGCCTGGTGGAAAAA ATTCGTCTGCATGATGAACAAGAGGGCAAATGTCTGTATAGCCTGGAAAG CATTCCTCTGGAAGATCTGCTGAACAATCCGAATCATTATGAAGTGGATC ACATTATTCCGCGTAGCGTGAGCTTTGATAATTCCTATCATAATAAAGTG CTGGTGAAACAGAGCGAAAACTCCAAAAAATCCAACCTGACACCGTATC AGTATTTCAATAGCGGCAAATCCAAACTGAGCTACAACCAGTTTAAACAG CATATTCTGAACCTGAGCAAAAGCCAGGATCGCATCAGCAAGAAGAAGA AGGAGTACCTGCTGGAAGAACGCGACATCAACAAATTTGAAGTGCAGAA AGAATTTATCAACCGCAACCTGGTTGATACCCGTTATGCAACCCGTGAAC TGACCAATTATCTGAAAGCATATTTCAGCGCCAACAACATGAACGTGAAA GTGAAAACGATTAACGGCAGCTTTACCGATTATCTGCGTAAAGTGTGGAA ATTCAAAAAAGAACGCAACCACGGCTATAAACATCATGCCGAAGATGCC CTGATTATTGCAAATGCAGATTTCCTGTTTAAAGAAAACAAAAAACTGAA AGCCGTCAACAGCGTGCTGGAAAAACCGGAAATTGAGACAAAACAGCTG GACATTCAGGTTGATAGCGAAGATAATTACAGCGAAATGTTTATCATCCC GAAACAGGTGCAGGATATCAAAGATTTTCGCAACTTCAAATATAGCCACC GCGTTGACAAAAAACCTAATCGTCAGCTGATTAACGATACCCTGTATAGC ACCCGCAAAAAAGATAACAGCACCTATATTGTGCAGACCATTAAAGACA TCTACGCCAAAGATAATACCACCCTGAAAAAACAGTTCGACAAAAGCCC AGAAAAATTTCTGATGTATCAGCATGATCCGCGTACCTTCGAAAAACTGG	SluCas9 密码子优化的 DNA

[0733]

	AAGTTATTATGAAACAGTATGCCAACGAGAAAAATCCGCTGGCCAAATA TCACGAAGAAACCGGTGAATATCTGACCAAATATTCCAAGAAGAACAAC GGTCCGATCGTTAAATCCCTGAAATATATCGGTAATAAACTGGGCAGCCA TCTGGATGTTACCCATCAGTTTAAAAGCTCCACAAAGAAGCTGGTTAAAC TGTCCATCAAACCGTATCGCTTTGATGTGTATCTGACCGACAAAGGCTAT AAATTCATTACCATCAGCTATCTGGACGTGCTGAAAAAAGACAACTATTA TTATATCCCGGAACAGAAATATGATAAACTGAAACTGGGTAAAGCCATC GATAAAAACGCCAAATTTATCGCCAGCTTCTACAAAAACGACCTGATTAA ACTGGATGGCGAGATCTATAAAATCATCGGTGTTAATAGCGACACCCGCA ATATGATTGAGCTGGATCTGCCGGATATTCGCTATAAAGAATATTGCGAA CTGAACAACATTAAAGGCGAACCGCGTATCAAAAAGACCATCGGCAAAA AAGTGAATAGCATCGAGAAACTGACCACCGATGTTCTGGGTAATGTGTTT ACCAATACCCAGTATACCAAACCTCAGCTGCTGTTCAAACGCGGTAATGG CGGAGGATCTGGCCCCCCTAAGAAAAAGCGGAAGGTGGGTGGAAGCGGA GGCAGCGGGGGATCAGGCCATCATCATCACCATCATTAA	
125	MNQKFILGLDIGITSVGYGLIDYETKNIIDAGVRLFPEANVENNEGRRSKRGS RRLKRRRIHRLERVKKLLEDYNLLDQSQIPQSTNPYAIRVKGLSEALSKDELV IALLHIAKRRGIHKIDVIDSNDDVGNELSTKEQLNKNSKLLKDKFVCQIQLER MNEGQVRGEKNRFKTADIIKEIIQLLNVQKNFHQLDENFINKYIELVEMRREY FEGPGKGSPYGWEGDPKAWYETLMGHCTYFPDELRSVKYAYSADLFNALN DLNNLVIQRDGLSKLEYHEKYHIIENVFKQKKKPTLKQIANEINVNPEDIKGY RITKSGKPQFTEFKLYHDLKSVLFDQSILENEDVLDQIAEILTIYQDKDSIKSKL TELDILLNEEDKENIAQLTGYTGTHRLSLKCIRLVLEEQWYSSRNQMEIFTHL NIKPKKINLTAANKIPKAMIDEFILSPVVKRTFGQAINLINKIIEKYGVPEDIIIEL ARENNSKDKQKFINEMQKKNENTRKRINEIIGKYGNQNAKRLVEKIRLHDEQ EGKCLYSLESIPLEDLLNNPNHYEVDHIIPRSVSFDNSYHNKVLVKQSENSKK SNLTPYQYFNSGKSKLSYNQFKQHILNLSKSQDRISKKKKEYLLEERDINKFE VQKEFINRNLVDTRYATRELTNYLKAYFSANNMNVKVKTINGSFTDYLRKV WKFKKERNHGYKHHAEDALIIANADFLFKENKKLKAVNSVLEKPEIETKQL DIQVDSEDNYSEMFIIPKQVQDIKDFRNFKYSHRVDKKPNRQLINDTLYSTRK KDNSTYIVQTIKDIYAKDNTTLKKQFDKSPEKFLMYQHDPRTFEKLEVIMKQ YANEKNPLAKYHEETGEYLTKYSKKNNGPIVKSLKYIGNKLGSHLDVTHQF KSSTKKLVKLSIKPYRFDVYLTDKGYKFITISYLDVLKKDNYYYIPEQKYDKL KLGKAIDKNAKFIASFYKNDLIKLDGEIYKIIGVNSDTRNMIELDLPDIRYKEY CELNNIKGEPRIKKTIGKKVNSIEKLTTDVLGNVFTNTQYTKPQLLFKRGNGG	SluCas9 多肽
126	YGRKKRRQRRR	**HIV-1 TAT 的残基 47-57**
127	RQIKIWFQNRRMKWKK	Penetratin
128	ggugaacguggaugaaguuggauuuuaguacucuggaaacagaaucuacugaaacaagacaauaugucguguu uaucccaucaauuuauugguggggauuuuuuucuagcauaaccccuuggggccucuaaacgggucuugaggg guuuuuu	sgRNA_Ex6
129	ggtgaacgtggatgaagttgtggggtac	R01-1-A
130	gcggtaccccacaacttcatccacgttcaccgg	R01-1-B
131	agtgaacgtggatgaagttgtggggtac	R01-2-A
132	gcggtaccccacaacttcatccacgttcactgg	R01-2-B
133	cgtgaacgtggatgaagttgtggggtac	R01-3-A
134	gcggtaccccacaacttcatccacgttcacggg	R01-3-B
135	tgtgaacgtggatgaagttgtggggtac	R01-4-A
136	gcggtaccccacaacttcatccacgttcacagg	R01-4-B
137	gatgaacgtggatgaagttgtggggtac	R01-5-A
138	gcggtaccccacaacttcatccacgttcatcgg	R01-5-B
139	gctgaacgtggatgaagttgtggggtac	R01-6-A
140	gcggtaccccacaacttcatccacgttcagcgg	R01-6-B
141	gttgaacgtggatgaagttgtggggtac	R01-7-A
142	gcggtaccccacaacttcatccacgttcaacgg	R01-7-B
143	ggagaacgtggatgaagttgtggggtac	R01-8-A

144	gcggtaccccacaacttcatccacgttctccgg	R01-8-B
145	ggcgaacgtggatgaagttgtggggtac	R01-9-A
146	gcggtaccccacaacttcatccacgttcgccgg	R01-9-B
147	ggggaacgtggatgaagttgtggggtac	R01-10-A
148	gcggtaccccacaacttcatccacgttcccgg	R01-10-B
149	ggtaaacgtggatgaagttgtggggtac	R01-11-A
150	gcggtaccccacaacttcatccacgtttaccgg	R01-11-B
151	ggtcaacgtggatgaagttgtggggtac	R01-12-A
152	gcggtaccccacaacttcatccacgttgaccgg	R01-12-B
153	ggttaacgtggatgaagttgtggggtac	R01-13-A
154	gcggtaccccacaacttcatccacgttaaccgg	R01-13-B
155	ggtgcacgtggatgaagttgtggggtac	R01-14-A
156	gcggtaccccacaacttcatccacgtgcaccgg	R01-14-B
157	ggtgtacgtggatgaagttgtggggtac	R01-15-A
158	gcggtaccccacaacttcatccacgtacaccgg	R01-15-B
159	ggtggacgtggatgaagttgtggggtac	R01-16-A
160	gcggtaccccacaacttcatccacgtccaccgg	R01-16-B
161	ggtgaccgtggatgaagttgtggggtac	R01-17-A
162	gcggtaccccacaacttcatccacggtcaccgg	R01-17-B
163	ggtgatcgtggatgaagttgtggggtac	R01-18-A
164	gcggtaccccacaacttcatccacgatcaccgg	R01-18-B
165	ggtgagcgtggatgaagttgtggggtac	R01-19-A
[0734] 166	gcggtaccccacaacttcatccacgctcaccgg	R01-19-B
167	ggtgaaagtggatgaagttgtggggtac	R01-20-A
168	gcggtaccccacaacttcatccactttcaccgg	R01-20-B
169	ggtgaatgtggatgaagttgtggggtac	R01-21-A
170	gcggtaccccacaacttcatccacattcaccgg	R01-21-B
171	ggtgaaggtggatgaagttgtggggtac	R01-22-A
172	gcggtaccccacaacttcatccaccttcaccgg	R01-22-B
173	ggtgaacatggatgaagttgtggggtac	R01-23-A
174	gcggtaccccacaacttcatccatgttcaccgg	R01-23-B
175	ggtgaacctggatgaagttgtggggtac	R01-24-A
176	gcggtaccccacaacttcatccaggttcaccgg	R01-24-B
177	ggtgaacttggatgaagttgtggggtac	R01-25-A
178	gcggtaccccacaacttcatccaagttcaccgg	R01-25-B
179	ggtgaacgaggatgaagttgtggggtac	R01-26-A
180	gcggtaccccacaacttcatcctcgttcaccgg	R01-26-B
181	ggtgaacgcggatgaagttgtggggtac	R01-27-A
182	gcggtaccccacaacttcatccgcgttcaccgg	R01-27-B
183	ggtgaacggggatgaagttgtggggtac	R01-28-A
184	gcggtaccccacaacttcatccccgttcaccgg	R01-28-B
185	ggtgaacgtagatgaagttgtggggtac	R01-29-A
186	gcggtaccccacaacttcatctacgttcaccgg	R01-29-B
187	ggtgaacgtcgatgaagttgtggggtac	R01-30-A
188	gcggtaccccacaacttcatcgacgttcaccgg	R01-30-B

[0735]

189	ggtgaacgttgatgaagttgtggggtac	R01-31-A
190	gcggtaccccacaacttcatcaacgttcaccgg	R01-31-B
191	ggtgaacgtgaatgaagttgtggggtac	R01-32-A
192	gcggtaccccacaacttcattcacgttcaccgg	R01-32-B
193	ggtgaacgtgcatgaagttgtggggtac	R01-33-A
194	gcggtaccccacaacttcatgcacgttcaccgg	R01-33-B
195	ggtgaacgtgtatgaagttgtggggtac	R01-34-A
196	gcggtaccccacaacttcatacacgttcaccgg	R01-34-B
197	ggtgaacgtggctgaagttgtggggtac	R01-35-A
198	gcggtaccccacaacttcagccacgttcaccgg	R01-35-B
199	ggtgaacgtggttgaagttgtggggtac	R01-36-A
200	gcggtaccccacaacttcaaccacgttcaccgg	R01-36-B
201	ggtgaacgtgggtgaagttgtggggtac	R01-37-A
202	gcggtaccccacaacttcaccccacgttcaccgg	R01-37-B
203	ggtgaacgtggaagaagttgtggggtac	R01-38-A
204	gcggtaccccacaacttcttccacgttcaccgg	R01-38-B
205	ggtgaacgtggacgaagttgtggggtac	R01-39-A
206	gcggtaccccacaacttcgtccacgttcaccgg	R01-39-B
207	ggtgaacgtggaggaagttgtggggtac	R01-40-A
208	gcggtaccccacaacttcctccacgttcaccgg	R01-40-B
209	ggtgaacgtggataaagttgtggggtac	R01-41-A
210	gcggtaccccacaactttatccacgttcaccgg	R01-41-B
211	ggtgaacgtggatcaagttgtggggtac	R01-42-A
212	gcggtaccccacaacttgatccacgttcaccgg	R01-42-B
213	ggtgaacgtggattaagttgtggggtac	R01-43-A
214	gcggtaccccacaacttaatccacgttcaccgg	R01-43-B
215	ggtgaacgtggatgcagttgtggggtac	R01-44-A
216	gcggtaccccacaactgcatccacgttcaccgg	R01-44-B
217	ggtgaacgtggatgtagttgtggggtac	R01-45-A
218	gcggtaccccacaactacatccacgttcaccgg	R01-45-B
219	ggtgaacgtggatggagttgtggggtac	R01-46-A
220	gcggtaccccacaactccatccacgttcaccgg	R01-46-B
221	ggtgaacgtggatgacgttgtggggtac	R01-47-A
222	gcggtaccccacaacgtcatccacgttcaccgg	R01-47-B
223	ggtgaacgtggatgatgttgtggggtac	R01-48-A
224	gcggtaccccacaacatcatccacgttcaccgg	R01-48-B
225	ggtgaacgtggatgaggttgtggggtac	R01-49-A
226	gcggtaccccacaacctcatccacgttcaccgg	R01-49-B
227	ggtgaacgtggatgaaattgtggggta	R01-50-A
228	gcggtaccccacaatttcatccacgttcaccgg	R01-50-B
229	ggtgaacgtggatgaacttgtggggtac	R01-51-A
230	gcggtaccccacaagttcatccacgttcaccgg	R01-51-B
231	ggtgaacgtggatgaatttgtggggtac	R01-52-A
232	gcggtaccccacaaattcatccacgttcaccgg	R01-52-B
233	ggtgaacgtggatgaagatgtggggtac	R01-53-A

[0736]

234	gcggtaccccacatcttcatccacgttcaccgg	R01-53-B
235	ggtgaacgtggatgaagctgtggggtac	R01-54-A
236	gcggtaccccacagcttcatccacgttcaccgg	R01-54-B
237	ggtgaacgtggatgaaggtgtggggtac	R01-55-A
238	gcggtaccccacaccttcatccacgttcaccgg	R01-55-B
239	ggtgaacgtggatgaagtagtggggtac	R01-56-A
240	gcggtaccccactacttcatccacgttcaccgg	R01-56-B
241	ggtgaacgtggatgaagtcgtggggtac	R01-57-A
242	gcggtaccccacgacttcatccacgttcaccgg	R01-57-B
243	ggtgaacgtggatgaagtggtggggtac	R01-58-A
244	gcggtaccccaccacttcatccacgttcaccgg	R01-58-B
245	ggtgaacgtggatgaagttatggggtac	R01-59-A
246	gcggtaccccataacttcatccacgttcaccgg	R01-59-B
247	ggtgaacgtggatgaagttctggggtac	R01-60-A
248	gcggtaccccagaacttcatccacgttcaccgg	R01-60-B
249	ggtgaacgtggatgaagttttggggtac	R01-61-A
250	gcggtaccccaaaacttcatccacgttcaccgg	R01-61-B
251	GGAGTCCGAGCAGAAGAAGAA	sgRNA 1 靶序列
252	GGACATCGATGTCCTCCCCAT	sgRNA 2 靶序列
253	GGTCACCTCCAATGACTAGGG	sgRNA 3 靶序列
254	GGCTCCCCATTGGCCTGCTTCG	sgRNA 4 靶序列
255	GGTGCGCCACCGGTTGATGTGA	sgRNA 5 靶序列
256	GGCGCCACCGGTTGATGTGAT	sgRNA 6 靶序列
257	GGCCGTTTGTACTTTGTCCTC	sgRNA 7 靶序列
258	GGCACAGATGAGAAACTCAGG	sgRNA 8 靶序列
259	GGATTGGGTGTTCAGGGCAGAG	sgRNA 9 靶序列
260	GGTGGCGAGAGGGGCCGAGAT	sgRNA 10 靶序列
261	GGGGCCGAGATTGGGTGTTC	sgRNA 11 靶序列
262	GGTGCCATTAGCTAAATGCAT	sgRNA 12 靶序列
263	GGTACCACCCACAGGTGCCAG	sgRNA 13 靶序列
264	GGAAAGCCTCTGGGCCAGGAA	sgRNA 14 靶序列
265	GGACACCCACGGAGAGCGCTGT	sgRNA 15 靶序列
266	GGTCCAGTTGGTGACAAATAC	sgRNA 16 靶序列
267	GGTTTTAACACTTCCCTAGCCA	sgRNA 17 靶序列
268	GGCTTAATTAGCTCCTTTGGCT	sgRNA 18 靶序列
269	GGCTTAATTAGCTCCTTTGGC	sgRNA 19 靶序列

[0737]

270	GGTGCATATGATGATAGTATTA	sgRNA 20 靶序列
271	GGAATGTATGCTGGCTTTTAAG	sgRNA 21 靶序列
272	GGAAATGTATGCTGGCTTTTAA	sgRNA 22 靶序列
273	GGAAAATGTATGCTGGCTTTTA	sgRNA 23 靶序列
274	GGTAAAATGTATGCTGGCTTTT	sgRNA 24 靶序列
275	GGCTTAAAGATGATCTCTTACC	sgRNA 25 靶序列
276	GGTTAAAGATGATCTCTTACCT	sgRNA 26 靶序列
277	GGTAATAAAATTCAAACATCCT	sgRNA 27 靶序列
278	GGTAAAGCATAGTGCAATGGAT	sgRNA 28 靶序列
279	GGTTAAATAAAGCATAGTGCAA	sgRNA 29 靶序列
280	GGATTTATGAGATCAACAGCAC	sgRNA 30 靶序列
281	GGCCCATGCTCTGACCGCTCGA	sgRNA 31 靶序列
282	GGCCATCGAGCGGTCAGAGCAT	sgRNA 32 靶序列
283	GGGCAAAAGCTCATGTGATA	sgRNA 33 靶序列
284	GGACTGAGGCTTATGTTCCATG	sgRNA 34 靶序列
285	GGAGAGTGTACAAACTCACAA	sgRNA 35 靶序列
286	GGACTAGCATTATAATGCACCA	sgRNA 36 靶序列
287	GGTCAGAAGAGATTAGTTAGTA	sgRNA 37 靶序列
288	GGTGAGAGTGCCATCTCTTCCT	sgRNA 38 靶序列
289	GGTGGACCACATGGCTTTGCTC	sgRNA 39 靶序列
290	GGTGAGCAAAGCCATGTGGTCC	sgRNA 40 靶序列
291	GGATTAATGACATACGCATTT	sgRNA 41 靶序列
292	GGAGATTAATGACATACGCATT	sgRNA 42 靶序列
293	GGTTCAATCCTCTTGTCACCTG	sgRNA 43 靶序列
294	GGACATCTCTATGTCGGCCAC	sgRNA 44 靶序列
295	GGAAAACCTACCGCAAGTTGCC	sgRNA 45 靶序列
296	GGTGGACTACATAGTTGTGTGA	sgRNA 46 靶序列
297	GGCTGCCAGACTCTCTGAACCC	sgRNA 47 靶序列
298	GGATCCCTCGTATAACAATA	sgRNA 48 靶序列
299	GGTATCGCATAGTGCAAAGAAT	sgRNA 49 靶序列

[0738]

300	GGCTCGGTCCTCTCGTCACCTG	sgRNA 50 靶序列
301	GGACGCTCCCGGAGAGCGCTGG	sgRNA 51 靶序列
302	GGCCATGGCGCGGTCAGAGGGT	sgRNA 52 靶序列
303	GGCTCGGCTCCAATGACGAGGG	sgRNA 53 靶序列
304	GGCACTTCGATGACCTCCCCGT	sgRNA 54 靶序列
305	GGAGTGGGTGTTCAGGACCGCG	sgRNA 55 靶序列
306	GGCTGCTACCCACAGGTGCAAG	sgRNA 56 靶序列
307	GGAAGATGTATCCTGGAATTTA	sgRNA 57 靶序列
308	GGTTTTATCACTACCCAAGACA	sgRNA 58 靶序列
309	GGCGGCCAGGCTCCCTGAAACC	sgRNA 59 靶序列
310	GGACTGGTATTATAATACAACA	sgRNA 60 靶序列
311	GGTGAGAGTGACATTTCGTACT	sgRNA 61 靶序列
312	GGTGCGCCCCCGGGTGAGGTGC	sgRNA 62 靶序列
313	GGCAGTTTGAGCAGAAAAAGAA	sgRNA 63 靶序列
314	GGTCTGTAGAGATAAGTGAGTA	sgRNA 64 靶序列
315	GGAAGTGTGTACAAACTCAAGA	sgRNA 65 靶序列
316	GGTATCATGGATGCTGACTTTT	sgRNA 66 靶序列
317	GGCTTTACTAGCCCCTGTGGC	sgRNA 67 靶序列
318	GGCACGGAGGAGAAAAGCAGG	sgRNA 68 靶序列
319	GGATTTGTGAGATCTAAAGAAC	sgRNA 69 靶序列
320	GGCGTTGATGACATTCGCATTT	sgRNA 70 靶序列
321	GGCTTATAGATCATCTAATACC	sgRNA 71 靶序列
322	GGTGGCCTCCATAGTTGGGTGG	sgRNA 72 靶序列
323	GGTACTAGAGTTCAAACATACT	sgRNA 73 靶序列
324	GGCCCCTGCTCTGAGCACTCGG	sgRNA 74 靶序列
325	GGAGGACCACATGGCCTGCCTC	sgRNA 75 靶序列
326	GGAACAGCTGCCGCAAATTGCC	sgRNA 76 靶序列
327	GGCTCCGTATTGGCTGGCTTCG	sgRNA 77 靶序列
328	GGCCAGCCGATGACAAATAC	sgRNA 78 靶序列
329	GGACGTCTCTACGTCGGCGGC	sgRNA 79 靶序列

[0739]

330	GGCCGGCGCGAGGGGCCGAGAG	sgRNA 80 靶序列
331	GGACTGTTTGTACTTTTTCATC	sgRNA 81 靶序列
332	GGTGCTTATGATGATGACATTA	sgRNA 82 靶序列
333	GGCCGCGTCCGATTGATGTGAT	sgRNA 83 靶序列
334	GGCTTTATGAGCACCTTAGGCT	sgRNA 84 靶序列
335	GGATCCCTCCTATGACCATG	sgRNA 85 靶序列
336	GGATAAAGGTGACCTCTTACGT	sgRNA 86 靶序列
337	GGACTGTGGCTTAGGGTCCAGG	sgRNA 87 靶序列
338	GGTGAGGAAACCCAGGCGGTCC	sgRNA 88 靶序列
339	GGAAATGTATACAGGCAATTAA	sgRNA 89 靶序列
340	GGTTGAATCAAGCATAGTCCGA	sgRNA 90 靶序列
341	GGGCTACGGCTCATGAGATA	sgRNA 91 靶序列
342	GGCGGGCCGCGATTGGCAGTTC	sgRNA 92 靶序列
343	GGCTGCGATTATCTAAATCCAT	sgRNA 93 靶序列
344	GGAGCTTGATTACATACGCATG	sgRNA 94 靶序列
345	GGAATGTTTGCTGGCCTCTAGG	sgRNA 95 靶序列
346	GGAAAGCCTCTGCGCCGGGGG	sgRNA 96 靶序列
347	CCTGAGTCCGAGCAGAAGAAGAATGGGGTAC	Ext-A-P-1-A
348	GCGGTACCCCATTCTTCTTCTGCTCGGACTCAGGGG	Ext-A-P-1-B
349	GGTGACATCGATGTCCTCCCCATTGGGGTAC	Ext-A-P-2-A
350	GCGGTACCCCAATGGGGAGGACATCGATGTCACCGG	Ext-A-P-2-B
351	GATGTCACCTCCAATGACTAGGGTGGGGTAC	Ext-A-P-3-A
352	GCGGTACCCCACCCTAGTCATTGGAGGTGACATCGG	Ext-A-P-3-B
353	GTCCTCCCCATTGGCCTGCTTCGTGGGGTAC	Ext-A-P-4-A
354	GCGGTACCCCACGAAGCAGGCCAATGGGGAGGACGG	Ext-A-P-4-B
355	CAATGCGCCACCGGTTGATGTGATGGGGTAC	Ext-A-P-5-A
356	GCGGTACCCCATCACATCAACCGGTGGCGCATTGGG	Ext-A-P-5-B
357	AATGCGCCACCGGTTGATGTGATTGGGGTAC	Ext-A-P-6-A
358	GCGGTACCCCAATCACATCAACCGGTGGCGCATTGG	Ext-A-P-6-B
359	TCTGCCGTTTGTACTTTGTCCTCTGGGGTAC	Ext-A-P-7-A
360	GCGGTACCCCAGAGGACAAAGTACAAACGGCAGAGG	Ext-A-P-7-B
361	GGGGCACAGATGAGAAACTCAGGTGGGGTAC	Ext-A-P-8-A
362	GCGGTACCCCACCTGAGTTTCTCATCTGTGCCCCGG	Ext-A-P-8-B
363	GAGATTGGGTGTTCAGGGCAGAGTGGGGTAC	Ext-A-P-9-A
364	GCGGTACCCCACTCTGCCCTGAACACCCAATCTCGG	Ext-A-P-9-B
365	AGGGTGGCGAGAGGGGCCGAGATTGGGGTAC	Ext-A-P-10-A
366	GCGGTACCCCAATCTCGGCCCCTCTCGCCACCCTGG	Ext-A-P-10-B

367	AGAGGGGCCGAGATTGGGTGTTCTGGGGTAC	Ext-A-P-11-A
368	GCGGTACCCCAGAACACCCAATCTCGGCCCCTCTGG	Ext-A-P-11-B
369	TGAGTGCCATTAGCTAAATGCATTGGGGTAC	Ext-A-P-12-A
370	GCGGTACCCCAATGCATTTAGCTAATGGCACTCAGG	Ext-A-P-12-B
371	AGGGTACCACCCACAGGTGCCAGTGGGGTAC	Ext-A-P-13-A
372	GCGGTACCCCACTGGCACCTGTGGGTGGTACCCTGG	Ext-A-P-13-B
373	TGGGAAAGCCTCTGGGCCAGGAATGGGGTAC	Ext-A-P-14-A
374	GCGGTACCCCATTCCTGGCCCAGAGGCTTTCCCAGG	Ext-A-P-14-B
375	AGGACACCCACGGAGAGCGCTGTTGGGGTAC	Ext-A-P-27-A
376	GCGGTACCCCAACAGCGCTCTCCGTGGGTGTCCTGG	Ext-A-P-27-B
377	AGGGTCCAGTTGGTGACAAATACTGGGGTAC	Ext-A-P-28-A
378	GCGGTACCCCAGTATTTGTCACCAACTGGACCCTGG	Ext-A-P-28-B
379	AGGTTTTAACACTTCCCTAGCCATGGGGTAC	Ext-A-P-29-A
380	GCGGTACCCCATGGCTAGGGAAGTGTTAAAACCTGG	Ext-A-P-29-B
381	AGGCTTAATTAGCTCCTTTGGCTTGGGGTAC	Ext-A-P-30-A
382	GCGGTACCCCAAGCCAAAGGAGCTAATTAAGCCTGG	Ext-A-P-30-B
383	AGGGCTTAATTAGCTCCTTTGGCTGGGGTAC	Ext-A-P-31-A
384	GCGGTACCCCAGCCAAAGGAGCTAATTAAGCCCTGG	Ext-A-P-31-B
385	AGGTGCATATGATGATAGTATTATGGGGTAC	Ext-A-P-32-A
386	GCGGTACCCCATAATACTATCATCATATGCACCTGG	Ext-A-P-32-B
387	AGGAATGTATGCTGGCTTTTAAGTGGGGTAC	Ext-A-P-33-A
388	GCGGTACCCCACTTAAAAGCCAGCATACATTCCTGG	Ext-A-P-33-B
[0740] 389	AGGAAATGTATGCTGGCTTTTAATGGGGTAC	Ext-A-P-34-A
390	GCGGTACCCCATTAAAAGCCAGCATACATTTCCTGG	Ext-A-P-34-B
391	AGGAAAATGTATGCTGGCTTTTATGGGGTAC	Ext-A-P-35-A
392	GCGGTACCCCATAAAAGCCAGCATACATTTTCCTGG	Ext-A-P-35-B
393	AGGTAAAATGTATGCTGGCTTTTTGGGGTAC	Ext-A-P-36-A
394	GCGGTACCCCAAAAAGCCAGCATACATTTTACCTGG	Ext-A-P-36-B
395	AGGCTTAAAGATGATCTCTTACCTGGGGTAC	Ext-A-P-37-A
396	GCGGTACCCCAGGTAAGAGATCATCTTTAAGCCTGG	Ext-A-P-37-B
397	AGGTTAAAGATGATCTCTTACCTTGGGGTAC	Ext-A-P-38-A
398	GCGGTACCCCAAGGTAAGAGATCATCTTTAACCTGG	Ext-A-P-38-B
399	AGGTAATAAAATTCAAACATCCTTGGGGTAC	Ext-A-P-39-A
400	GCGGTACCCCAAGGATGTTTGAATTTTATTACCTGG	Ext-A-P-39-B
401	AGGTAAAGCATAGTGCAATGGATTGGGGTAC	Ext-A-P-40-A
402	GCGGTACCCCAATCCATTGCACTATGCTTTACCTGG	Ext-A-P-40-B
403	AGGTTAAATAAAGCATAGTGCAATGGGGTAC	Ext-A-P-41-A
404	GCGGTACCCCATTGCACTATGCTTTATTTAACCTGG	Ext-A-P-41-B
405	AGGATTTATGAGATCAACAGCACTGGGGTAC	Ext-A-P-42-A
406	GCGGTACCCCAGTGCTGTTGATCTCATAAATCCTGG	Ext-A-P-42-B
407	AGGCCCATGCTCTGACCGCTCGATGGGGTAC	Ext-A-P-43-A
408	GCGGTACCCCATCGAGCGGTCAGAGCATGGGCCTGG	Ext-A-P-43-B
409	AGGCCATCGAGCGGTCAGAGCATTGGGGTAC	Ext-A-P-44-A
410	GCGGTACCCCAATGCTCTGACCGCTCGATGGCCTGG	Ext-A-P-44-B
411	AGGGGGCAAAAGCTCATGTGATATGGGGTAC	Ext-A-P-45-A

412	GCGGTACCCCATATCACATGAGCTTTTGCCCCCTGG	Ext-A-P-45-B
413	AGGACTGAGGCTTATGTTCCATGTGGGGTAC	Ext-A-P-46-A
414	GCGGTACCCCACATGGAACATAAGCCTCAGTCCTGG	Ext-A-P-46-B
415	AGGGAGAGTGTACAAACTCACAATGGGGTAC	Ext-A-P-47-A
416	GCGGTACCCCATTGTGAGTTTGTACACTCTCCCTGG	Ext-A-P-47-B
417	AGGACTAGCATTATAATGCACCATGGGGTAC	Ext-A-P-48-A
418	GCGGTACCCCATGGTGCATTATAATGCTAGTCCTGG	Ext-A-P-48-B
419	AGGTCAGAAGAGATTAGTTAGTATGGGGTAC	Ext-A-P-49-A
420	GCGGTACCCCATACTAACTAATCTCTTCTGACCTGG	Ext-A-P-49-B
421	AGGTGAGAGTGCCATCTCTTCCTTGGGGTAC	Ext-A-P-50-A
422	GCGGTACCCCAAGGAAGAGATGGCACTCTCACCTGG	Ext-A-P-50-B
423	AGGTGGACCACATGGCTTTGCTCTGGGGTAC	Ext-A-P-51-A
424	GCGGTACCCCAGAGCAAAGCCATGTGGTCCACCTGG	Ext-A-P-51-B
425	AGGTGAGCAAAGCCATGTGGTCCTGGGGTAC	Ext-A-P-52-A
426	GCGGTACCCCAGGACCACATGGCTTTGCTCACCTGG	Ext-A-P-52-B
427	AGGGATTAATGACATACGCATTTTGGGGTAC	Ext-A-P-53-A
428	GCGGTACCCCAAAATGCGTATGTCATTAATCCCTGG	Ext-A-P-53-B
429	AGGAGATTAATGACATACGCATTTGGGGTAC	Ext-A-P-54-A
430	GCGGTACCCCAAATGCGTATGTCATTAATCTCCTGG	Ext-A-P-54-B
431	AGGTTCAATCCTCTTGTCACCTGTGGGGTAC	Ext-A-P-55-A
432	GCGGTACCCCACAGGTGACAAGAGGATTGAACCTGG	Ext-A-P-55-B
433	AGGGACATCTCTATGTCGGCCACTGGGGTAC	Ext-A-P-56-A
[0741] 434	GCGGTACCCCAGTGGCCGACATAGAGATGTCCCTGG	Ext-A-P-56-B
435	AGGAAAACCTACCGCAAGTTGCCTGGGGTAC	Ext-A-P-57-A
436	GCGGTACCCCAGGCAACTTGCGGTAGGTTTTCCTGG	Ext-A-P-57-B
437	AGGTGGACTACATAGTTGTGTGATGGGGTAC	Ext-A-P-58-A
438	GCGGTACCCCATCACACAACTATGTAGTCCACCTGG	Ext-A-P-58-B
439	AGGCTGCCAGACTCTCTGAACCCTGGGGTAC	Ext-A-P-59-A
440	GCGGTACCCCAGGGTTCAGAGAGTCTGGCAGCCTGG	Ext-A-P-59-B
441	AGGGGATCCCTCGTATAACAATATGGGGTAC	Ext-A-P-60-A
442	GCGGTACCCCATATTGTTATACGAGGGATCCCCTGG	Ext-A-P-60-B
443	AGGTATCGCATAGTGCAAAGAATTGGGGTAC	Ext-A-P-61-A
444	GCGGTACCCCAATTCTTTGCACTATGCGATACCTGG	Ext-A-P-61-B
445	AGGCTCGGTCCTCTCGTCACCTGTGGGGTAC	Ext-A-P-62-A
446	GCGGTACCCCACAGGTGACGAGAGGACCGAGCCTGG	Ext-A-P-62-B
447	AGGACGCTCCCGGAGAGCGCTGGTGGGGTAC	Ext-A-P-63-A
448	GCGGTACCCCACCAGCGCTCTCCGGGAGCGTCCTGG	Ext-A-P-63-B
449	AGGCCATGGCGCGGTCAGAGGGTTGGGGTAC	Ext-A-P-64-A
450	GCGGTACCCCAACCCTCTGACCGCGCCATGGCCTGG	Ext-A-P-64-B
451	AGGCTCGGCTCCAATGACGAGGGTGGGGTAC	Ext-A-P-65-A
452	GCGGTACCCCACCCTCGTCATTGGAGCCGAGCCTGG	Ext-A-P-65-B
453	AGGCACTTCGATGACCTCCCCGTTGGGGTAC	Ext-A-P-66-A
454	GCGGTACCCCAACGGGGAGGTCATCGAAGTGCCTGG	Ext-A-P-66-B
455	AGGAGTGGGTGTTCAGGACCGCGTGGGGTAC	Ext-A-P-67-A
456	GCGGTACCCCACGCGGTCCTGAACACCCACTCCTGG	Ext-A-P-67-B

457	AGGCTGCTACCCACAGGTGCAAGTGGGGTAC	Ext-A-P-68-A
458	GCGGTACCCCACTTGCACCTGTGGGTAGCAGCCTGG	Ext-A-P-68-B
459	AGGAAGATGTATCCTGGAATTTATGGGGTAC	Ext-A-P-69-A
460	GCGGTACCCCATAAATTCCAGGATACATCTTCCTGG	Ext-A-P-69-B
461	AGGTTTTATCACTACCCAAGACATGGGGTAC	Ext-A-P-70-A
462	GCGGTACCCCATGTCTTGGGTAGTGATAAAACCTGG	Ext-A-P-70-B
463	AGGCGGCCAGGCTCCCTGAAACCTGGGGTAC	Ext-A-P-71-A
464	GCGGTACCCCAGGTTTCAGGGAGCCTGGCCGCCTGG	Ext-A-P-71-B
465	AGGACTGGTATTATAATACAACATGGGGTAC	Ext-A-P-72-A
466	GCGGTACCCCATGTTGTATTATAATACCAGTCCTGG	Ext-A-P-72-B
467	AGGTGAGAGTGACATTTCGTACTTGGGGTAC	Ext-A-P-73-A
468	GCGGTACCCCAAGTACGAAATGTCACTCTCACCTGG	Ext-A-P-73-B
469	AGGTGCGCCCCCGGGTGAGGTGCTGGGGTAC	Ext-A-P-74-A
470	GCGGTACCCCAGCACCTCACCCGGGGGCGCACCTGG	Ext-A-P-74-B
471	AGGCAGTTTGAGCAGAAAAAGAATGGGGTAC	Ext-A-P-75-A
472	GCGGTACCCCATTCTTTTTCTGCTCAAACTGCCTGG	Ext-A-P-75-B
473	AGGTCTGTAGAGATAAGTGAGTATGGGGTAC	Ext-A-P-76-A
474	GCGGTACCCCATACTCACTTATCTCTACAGACCTGG	Ext-A-P-76-B
475	AGGAAGTGTGTACAAACTCAAGATGGGGTAC	Ext-A-P-77-A
476	GCGGTACCCCATCTTGAGTTTGTACACACTTCCTGG	Ext-A-P-77-B
477	AGGTATCATGGATGCTGACTTTTTGGGGTAC	Ext-A-P-78-A
478	GCGGTACCCCAAAAAGTCAGCATCCATGATACCTGG	Ext-A-P-78-B
479	AGGGCTTTACTAGCCCCTGTGGCTGGGGTAC	Ext-A-P-79-A
480	GCGGTACCCCAGCCACAGGGGCTAGTAAAGCCCTGG	Ext-A-P-79-B
481	AGGGCACGGAGGAGAAAAGCAGGTGGGGTAC	Ext-A-P-80-A
482	GCGGTACCCCACCTGCTTTTCTCCTCCGTGCCCTGG	Ext-A-P-80-B
483	AGGATTTGTGAGATCTAAAGAACTGGGGTAC	Ext-A-P-81-A
484	GCGGTACCCCAGTTCTTTAGATCTCACAAATCCTGG	Ext-A-P-81-B
485	AGGCGTTGATGACATTCGCATTTTGGGGTAC	Ext-A-P-82-A
486	GCGGTACCCCAAAATGCGAATGTCATCAACGCCTGG	Ext-A-P-82-B
487	AGGCTTATAGATCATCTAATACCTGGGGTAC	Ext-A-P-83-A
488	GCGGTACCCCAGGTATTAGATGATCTATAAGCCTGG	Ext-A-P-83-B
489	AGGTGGCCTCCATAGTTGGGTGGTGGGGTAC	Ext-A-P-84-A
490	GCGGTACCCCACCACCCAACTATGGAGGCCACCTGG	Ext-A-P-84-B
491	AGGTACTAGAGTTCAAACATACTTGGGGTAC	Ext-A-P-85-A
492	GCGGTACCCCAAGTATGTTTGAACTCTAGTACCTGG	Ext-A-P-85-B
493	AGGCCCCTGCTCTGAGCACTCGGTGGGGTAC	Ext-A-P-86-A
494	GCGGTACCCCACCGAGTGCTCAGAGCAGGGGCCTGG	Ext-A-P-86-B
495	AGGAGGACCACATGGCCTGCCTCTGGGGTAC	Ext-A-P-87-A
496	GCGGTACCCCAGAGGCAGGCCATGTGGTCCTCCTGG	Ext-A-P-87-B
497	AGGAACAGCTGCCGCAAATTGCCTGGGGTAC	Ext-A-P-88-A
498	GCGGTACCCCAGGCAATTTGCGGCAGCTGTTCCTGG	Ext-A-P-88-B
499	AGGCTCCGTATTGGCTGGCTTCGTGGGGTAC	Ext-A-P-89-A
500	GCGGTACCCCACGAAGCCAGCCAATACGGAGCCTGG	Ext-A-P-89-B
501	AGGGGCCAGCCGATGACAAATACTGGGGTAC	Ext-A-P-90-A

[0742]

[0743]

502	GCGGTACCCCAGTATTTGTCATCGGCTGGCCCCTGG	Ext-A-P-90-B
503	AGGGACGTCTCTACGTCGGCGGCTGGGGTAC	Ext-A-P-91-A
504	GCGGTACCCCAGCCGCCGACGTAGAGACGTCCCTGG	Ext-A-P-91-B
505	AGGCCGGCGCGAGGGGCCGAGAGTGGGGTAC	Ext-A-P-92-A
506	GCGGTACCCCACTCTCGGCCCCTCGCGCCGGCCTGG	Ext-A-P-92-B
507	AGGACTGTTTGTACTTTTTCATCTGGGGTAC	Ext-A-P-93-A
508	GCGGTACCCCAGATGAAAAAGTACAAACAGTCCTGG	Ext-A-P-93-B
509	AGGTGCTTATGATGATGACATTATGGGGTAC	Ext-A-P-94-A
510	GCGGTACCCCATAATGTCATCATCATAAGCACCTGG	Ext-A-P-94-B
511	AGGCCGCGTCCGATTGATGTGATTGGGGTAC	Ext-A-P-95-A
512	GCGGTACCCCAATCACATCAATCGGACGCGGCCTGG	Ext-A-P-95-B
513	AGGCTTTATGAGCACCTTAGGCTTGGGGTAC	Ext-A-P-96-A
514	GCGGTACCCCAAGCCTAAGGTGCTCATAAAGCCTGG	Ext-A-P-96-B
515	AGGGGATCCCTCCTATGACCATGTGGGGTAC	Ext-A-P-97-A
516	GCGGTACCCCACATGGTCATAGGAGGGATCCCCTGG	Ext-A-P-97-B
517	AGGATAAAGGTGACCTCTTACGTTGGGGTAC	Ext-A-P-98-A
518	GCGGTACCCCAACGTAAGAGGTCACCTTTATCCTGG	Ext-A-P-98-B
519	AGGACTGTGGCTTAGGGTCCAGGTGGGGTAC	Ext-A-P-99-A
520	GCGGTACCCCACCTGGACCCTAAGCCACAGTCCTGG	Ext-A-P-99-B
521	AGGTGAGGAAACCCAGGCGGTCCTGGGGTAC	Ext-A-P-100-A
522	GCGGTACCCCAGGACCGCCTGGGTTTCCTCACCTGG	Ext-A-P-100-B
523	AGGAAATGTATACAGGCAATTAATGGGGTAC	Ext-A-P-101-A
524	GCGGTACCCCATTAATTGCCTGTATACATTTCCTGG	Ext-A-P-101-B
525	AGGTTGAATCAAGCATAGTCCGATGGGGTAC	Ext-A-P-102-A
526	GCGGTACCCCATCGGACTATGCTTGATTCAACCTGG	Ext-A-P-102-B
527	AGGGGGCTACGGCTCATGAGATATGGGGTAC	Ext-A-P-103-A
528	GCGGTACCCCATATCTCATGAGCCGTAGCCCCCTGG	Ext-A-P-103-B
529	AGGCGGGCCGCGATTGGCAGTTCTGGGGTAC	Ext-A-P-104-A
530	GCGGTACCCCAGAACTGCCAATCGCGGCCCGCCTGG	Ext-A-P-104-B
531	AGGCTGCGATTATCTAAATCCATTGGGGTAC	Ext-A-P-105-A
532	GCGGTACCCCAATGGATTTAGATAATCGCAGCCTGG	Ext-A-P-105-B
533	AGGAGCTTGATTACATACGCATGTGGGGTAC	Ext-A-P-106-A
534	GCGGTACCCCACATGCGTATGTAATCAAGCTCCTGG	Ext-A-P-106-B
535	AGGAATGTTTGCTGGCCTCTAGGTGGGGTAC	Ext-A-P-107-A
536	GCGGTACCCCACCTAGAGGCCAGCAAACATTCCTGG	Ext-A-P-107-B
537	AGGGAAAGCCTCTGCGCCGGGGGTGGGGTAC	Ext-A-P-108-A
538	GCGGTACCCCACCCCCGGCGCAGAGGCTTTCCCTGG	Ext-A-P-108-B

。

序列表

<110> Casebia Therapeutics Limited Liability Partnership
Cohnen, Andre
Schmidt, Moritz
Coco, Wayne
Gamalinda, Michael Biag
Gupta, Ashish
Pitzler, Christian
Richter, Florian
Tebbe, Jan
Cheng, Christopher
Takeuchi, Ryo
Reiss, Caroline W

<120> 新颖的RNA-可编程的内切核酸酶系统及其用途

<130> 052984-523001WO

<140> PCT/US2019/023044
<141> 2019-03-19

<150> EP 18162681.3
<151> 2013-03-19

<150> EP 18162683.9
<151> 2018-03-19

<150> EP 18172625.8
<151> 2018-05-16

[0001]
<150> EP 18174707.2
<151> 2018-05-16

<150> EP 18181680.2
<151> 2018-07-04

<150> US 62/745,238
<151> 2018-10-12

<150> US 62/745,239
<151> 2018-10-12

<150> US 62/745,240
<151> 2018-10-12

<150> US 62/745,246
<151> 2018-10-12

<160> 538

<170> PatentIn 3.5版

<210> 1
<211> 1057
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature

<223> Gib11的多肽序列

<400> 1

```
Met Asn Gln Lys Phe Ile Leu Gly Leu Asp Ile Gly Ile Thr Ser Val
1               5                  10                  15

Gly Tyr Gly Leu Ile Asp Tyr Glu Thr Lys Asn Ile Ile Asp Ala Gly
            20                  25                  30

Val Arg Leu Phe Pro Glu Ala Asn Val Glu Asn Asn Glu Gly Arg Arg
        35                  40                  45

Ser Lys Arg Gly Ser Arg Arg Leu Lys Arg Arg Arg Ile His Arg Leu
    50                  55                  60

Glu Arg Val Lys Leu Leu Leu Thr Glu Tyr Asp Leu Ile Asn Lys Glu
65                  70                  75                  80

Gln Ile Pro Thr Ser Asn Asn Pro Tyr Gln Ile Arg Val Lys Gly Leu
                85                  90                  95

Ser Glu Ile Leu Ser Lys Asp Glu Leu Ala Ile Ala Leu Leu His Leu
            100                 105                 110
```

[0002]

```
Ala Lys Arg Arg Gly Ile His Asn Val Asp Val Ala Ala Asp Lys Glu
        115                 120                 125

Glu Thr Ala Ser Asp Ser Leu Ser Thr Lys Asp Gln Ile Asn Lys Asn
    130                 135                 140

Ala Lys Phe Leu Glu Ser Arg Tyr Val Cys Glu Leu Gln Lys Glu Arg
145                 150                 155                 160

Leu Glu Asn Glu Gly His Val Arg Gly Val Glu Asn Arg Phe Leu Thr
                165                 170                 175

Lys Asp Ile Val Arg Glu Ala Lys Lys Ile Ile Asp Thr Gln Met Gln
            180                 185                 190

Tyr Tyr Pro Glu Ile Asp Glu Thr Phe Lys Glu Lys Tyr Ile Ser Leu
        195                 200                 205

Val Glu Thr Arg Arg Glu Tyr Phe Glu Gly Pro Gly Gln Gly Ser Pro
    210                 215                 220

Phe Gly Trp Asn Gly Asp Leu Lys Lys Trp Tyr Glu Met Leu Met Gly
225                 230                 235                 240
```

```
His Cys Thr Tyr Phe Pro Gln Glu Leu Arg Ser Val Lys Tyr Ala Tyr
                245                 250                 255

Ser Ala Asp Leu Phe Asn Ala Leu Asn Asp Leu Asn Asn Leu Ile Ile
            260                 265                 270

Gln Arg Asp Asn Ser Glu Lys Leu Glu Tyr His Glu Lys Tyr His Ile
        275                 280                 285

Ile Glu Asn Val Phe Lys Gln Lys Lys Lys Pro Thr Leu Lys Gln Ile
    290                 295                 300

Ala Lys Glu Ile Gly Val Asn Pro Glu Asp Ile Lys Gly Tyr Arg Ile
305                 310                 315                 320

Thr Lys Ser Gly Thr Pro Glu Phe Thr Ser Phe Lys Leu Phe His Asp
                325                 330                 335

Leu Lys Lys Val Val Lys Asp His Ala Ile Leu Asp Asp Ile Asp Leu
            340                 345                 350

Leu Asn Gln Ile Ala Glu Ile Leu Thr Ile Tyr Gln Asp Lys Asp Ser
        355                 360                 365
```

[0003]

```
Ile Val Ala Glu Leu Gly Gln Leu Glu Tyr Leu Met Ser Glu Ala Asp
    370                 375                 380

Lys Gln Ser Ile Ser Glu Leu Thr Gly Tyr Thr Gly Thr His Ser Leu
385                 390                 395                 400

Ser Leu Lys Cys Met Asn Met Ile Ile Asp Glu Leu Trp His Ser Ser
                405                 410                 415

Met Asn Gln Met Glu Val Phe Thr Tyr Leu Asn Met Arg Pro Lys Lys
            420                 425                 430

Tyr Glu Leu Lys Gly Tyr Gln Arg Ile Pro Thr Asp Met Ile Asp Asp
        435                 440                 445

Ala Ile Leu Ser Pro Val Val Lys Arg Thr Phe Ile Gln Ser Ile Asn
    450                 455                 460

Val Ile Asn Lys Val Ile Glu Lys Tyr Gly Ile Pro Glu Asp Ile Ile
465                 470                 475                 480

Ile Glu Leu Ala Arg Glu Asn Asn Ser Asp Asp Arg Lys Lys Phe Ile
                485                 490                 495
```

Asn Asn Leu Gln Lys Lys Asn Glu Ala Thr Arg Lys Arg Ile Asn Glu
500 505 510

Ile Ile Gly Gln Thr Gly Asn Gln Asn Ala Lys Arg Ile Val Glu Lys
515 520 525

Ile Arg Leu His Asp Gln Gln Glu Gly Lys Cys Leu Tyr Ser Leu Glu
530 535 540

Ser Ile Pro Leu Glu Asp Leu Leu Asn Asn Pro Asn His Tyr Glu Val
545 550 555 560

Asp His Ile Ile Pro Arg Ser Val Ser Phe Asp Asn Ser Tyr His Asn
565 570 575

Lys Val Leu Val Lys Gln Ser Glu Asn Ser Lys Lys Ser Asn Leu Thr
580 585 590

Pro Tyr Gln Tyr Phe Asn Ser Gly Lys Ser Lys Leu Ser Tyr Asn Gln
595 600 605

Phe Lys Gln His Ile Leu Asn Leu Ser Lys Ser Gln Asp Arg Ile Ser
610 615 620

[0004]

Lys Lys Lys Lys Glu Tyr Leu Leu Glu Glu Arg Asp Ile Asn Lys Phe
625 630 635 640

Glu Val Gln Lys Glu Phe Ile Asn Arg Asn Leu Val Asp Thr Arg Tyr
645 650 655

Ala Thr Arg Glu Leu Thr Asn Tyr Leu Lys Ala Tyr Phe Ser Ala Asn
660 665 670

Asn Met Asn Val Lys Val Lys Thr Ile Asn Gly Ser Phe Thr Asp Tyr
675 680 685

Leu Arg Lys Val Trp Lys Phe Lys Lys Glu Arg Asn His Gly Tyr Lys
690 695 700

His His Ala Glu Asp Ala Leu Ile Ile Ala Asn Ala Asp Phe Leu Phe
705 710 715 720

Lys Glu Asn Lys Lys Leu Lys Ala Val Asn Ser Val Leu Glu Lys Pro
725 730 735

Glu Ile Glu Thr Lys Gln Leu Asp Ile Gln Val Asp Ser Glu Asp Asn
740 745 750

```
Tyr Ser Glu Met Phe Ile Ile Pro Lys Gln Val Gln Asp Ile Lys Asp
        755                 760                 765

Phe Arg Asn Phe Lys Tyr Ser His Arg Val Asp Lys Lys Pro Asn Arg
    770                 775                 780

Gln Leu Ile Asn Asp Thr Leu Tyr Ser Thr Arg Lys Lys Asp Asn Ser
785                 790                 795                 800

Thr Tyr Ile Val Gln Thr Ile Lys Asp Ile Tyr Ala Lys Asp Asn Thr
                805                 810                 815

Thr Leu Lys Lys Gln Phe Asp Lys Ser Pro Glu Lys Phe Leu Met Tyr
            820                 825                 830

Gln His Asp Pro Arg Thr Phe Glu Lys Leu Glu Val Ile Met Lys Gln
        835                 840                 845

Tyr Ala Asn Glu Lys Asn Pro Leu Ala Lys Tyr His Glu Glu Thr Gly
    850                 855                 860

Glu Tyr Leu Thr Lys Tyr Ser Lys Lys Asn Asn Gly Pro Ile Val Lys
865                 870                 875                 880
```

[0005]

```
Ser Leu Lys Tyr Ile Gly Asn Lys Leu Gly Ser His Leu Asp Val Thr
                885                 890                 895

His Gln Phe Lys Ser Ser Thr Lys Lys Leu Val Lys Leu Ser Ile Lys
            900                 905                 910

Pro Tyr Arg Phe Asp Val Tyr Leu Thr Asp Lys Gly Tyr Lys Phe Ile
        915                 920                 925

Thr Ile Ser Tyr Leu Asp Val Leu Lys Lys Asp Asn Tyr Tyr Tyr Ile
    930                 935                 940

Pro Glu Gln Lys Tyr Asp Lys Leu Lys Leu Gly Lys Ala Ile Asp Lys
945                 950                 955                 960

Asn Ala Lys Phe Ile Ala Ser Phe Tyr Lys Asn Asp Leu Ile Lys Leu
                965                 970                 975

Asp Gly Glu Ile Tyr Lys Ile Ile Gly Val Asn Ser Asp Thr Arg Asn
            980                 985                 990

Met Ile Glu Leu Asp Leu Pro Asp  Ile Arg Tyr Lys Glu  Tyr Cys Glu
        995                 1000                1005
```

Leu Asn Asn Ile Lys Gly Glu Pro Arg Ile Lys Lys Thr Ile Gly
1010 1015 1020

Lys Lys Val Asn Ser Ile Glu Lys Leu Thr Thr Asp Val Leu Gly
1025 1030 1035

Asn Val Phe Thr Asn Thr Gln Tyr Thr Lys Pro Gln Leu Leu Phe
1040 1045 1050

Lys Arg Gly Asn
1055

<210> 2
<211> 1057
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> GiblISpa-1的多肽序列

<400> 2

[0006]

Met Asn Gln Lys Phe Ile Leu Gly Leu Asp Ile Gly Ile Thr Ser Val
1 5 10 15

Gly Tyr Gly Leu Ile Asp Tyr Glu Thr Lys Asn Ile Ile Asp Ala Gly
20 25 30

Val Arg Leu Phe Pro Glu Ala Asn Val Glu Asn Asn Glu Gly Arg Arg
35 40 45

Ser Lys Arg Gly Ser Arg Arg Leu Lys Arg Arg Arg Ile His Arg Leu
50 55 60

Glu Arg Val Lys Leu Leu Leu Thr Glu Tyr Asp Leu Ile Asn Lys Glu
65 70 75 80

Gln Ile Pro Thr Ser Asn Asn Pro Tyr Gln Ile Arg Val Lys Gly Leu
85 90 95

Ser Glu Ile Leu Ser Lys Asp Glu Leu Ala Ile Ala Leu Leu His Leu
100 105 110

Ala Lys Arg Arg Gly Ile His Asn Val Asp Val Ala Ala Asp Lys Glu
115 120 125

Glu Thr Ala Ser Asp Ser Leu Ser Thr Lys Asp Gln Ile Asn Lys Asn

130 135 140

Ala Lys Phe Leu Glu Ser Arg Tyr Val Cys Glu Leu Gln Lys Glu Arg
145 150 155 160

Leu Glu Asn Glu Gly His Val Arg Gly Val Glu Asn Arg Phe Leu Thr
165 170 175

Lys Asp Ile Val Arg Glu Ala Lys Lys Ile Ile Asp Thr Gln Met Gln
180 185 190

Tyr Tyr Pro Glu Ile Asp Glu Thr Phe Lys Glu Lys Tyr Ile Ser Leu
195 200 205

Val Glu Thr Arg Arg Glu Tyr Phe Glu Gly Pro Gly Gln Gly Ser Pro
210 215 220

Phe Gly Trp Asn Gly Asp Leu Lys Lys Trp Tyr Glu Met Leu Met Gly
225 230 235 240

His Cys Thr Tyr Phe Pro Gln Glu Leu Arg Ser Val Lys Tyr Ala Tyr
245 250 255

[0007] Ser Ala Asp Leu Phe Asn Ala Leu Asn Asp Leu Asn Asn Leu Ile Ile
260 265 270

Gln Arg Asp Asn Ser Glu Lys Leu Glu Tyr His Glu Lys Tyr His Ile
275 280 285

Ile Glu Asn Val Phe Lys Gln Lys Lys Lys Pro Thr Leu Lys Gln Ile
290 295 300

Ala Lys Glu Ile Gly Val Asn Pro Glu Asp Ile Lys Gly Tyr Arg Ile
305 310 315 320

Thr Lys Ser Gly Thr Pro Glu Phe Thr Ser Phe Lys Leu Phe His Asp
325 330 335

Leu Lys Lys Val Val Lys Asp His Ala Ile Leu Asp Asp Ile Asp Leu
340 345 350

Leu Asn Gln Ile Ala Glu Ile Leu Thr Ile Tyr Gln Asp Lys Asp Ser
355 360 365

Ile Val Ala Glu Leu Gly Gln Leu Glu Tyr Leu Met Ser Glu Ala Asp
370 375 380

Lys Gln Ser Ile Ser Glu Leu Thr Gly Tyr Thr Gly Thr His Ser Leu

385 390 395 400

Ser Leu Lys Cys Met Asn Met Ile Ile Asp Glu Leu Trp His Ser Ser
405 410 415

Met Asn Gln Met Glu Val Phe Thr Tyr Leu Asn Met Arg Pro Lys Lys
420 425 430

Tyr Glu Leu Lys Gly Tyr Gln Arg Ile Pro Thr Asp Met Ile Asp Asp
435 440 445

Ala Ile Leu Ser Pro Val Val Lys Arg Thr Phe Ile Gln Ser Ile Asn
450 455 460

Val Ile Asn Lys Val Ile Glu Lys Tyr Gly Ile Pro Glu Asp Ile Ile
465 470 475 480

Ile Glu Leu Ala Arg Glu Asn Asn Ser Asp Asp Arg Lys Lys Phe Ile
485 490 495

Asn Asn Leu Gln Lys Lys Asn Glu Ala Thr Arg Lys Arg Ile Asn Glu
500 505 510

[0008] Ile Ile Gly Gln Thr Gly Asn Gln Asn Ala Lys Arg Ile Val Glu Lys
515 520 525

Ile Arg Leu His Asp Gln Gln Glu Gly Lys Cys Leu Tyr Ser Leu Glu
530 535 540

Ser Ile Pro Leu Glu Asp Leu Leu Asn Asn Pro Asn His Tyr Glu Val
545 550 555 560

Asp His Ile Ile Pro Arg Ser Val Ser Phe Asp Asn Ser Tyr His Asn
565 570 575

Lys Val Leu Val Lys Gln Ser Glu Asn Ser Lys Lys Ser Asn Leu Thr
580 585 590

Pro Tyr Gln Tyr Phe Asn Ser Gly Lys Ser Lys Leu Ser Tyr Asn Gln
595 600 605

Phe Lys Gln His Ile Leu Asn Leu Ser Lys Ser Gln Asp Arg Ile Ser
610 615 620

Lys Lys Lys Lys Glu Tyr Leu Leu Glu Glu Arg Asp Ile Asn Lys Phe
625 630 635 640

Glu Val Gln Lys Glu Phe Ile Asn Arg Asn Leu Val Asp Thr Arg Tyr

645 650 655

Ala Thr Arg Glu Leu Thr Asn Tyr Leu Lys Ala Tyr Phe Ser Ala Asn
660 665 670

Asn Met Asn Val Lys Val Lys Thr Ile Asn Gly Ser Phe Thr Asp Tyr
675 680 685

Leu Arg Lys Val Trp Lys Phe Lys Lys Glu Arg Asn His Gly Tyr Lys
690 695 700

His His Ala Glu Asp Ala Leu Ile Ile Ala Asn Ala Asp Phe Leu Phe
705 710 715 720

Lys Glu Asn Lys Lys Leu Lys Ala Val Asn Ser Val Leu Glu Lys Pro
725 730 735

Glu Ile Glu Thr Lys Gln Leu Asp Ile Gln Val Asp Ser Glu Asp Asn
740 745 750

Tyr Ser Glu Met Phe Ile Ile Pro Lys Gln Val Gln Asp Ile Lys Asp
755 760 765

[0009] Phe Arg Asn Phe Lys Tyr Ser His Arg Val Asp Lys Lys Pro Asn Arg
770 775 780

Gln Leu Ile Asn Asp Thr Leu Tyr Ser Thr Arg Lys Lys Asp Asn Ser
785 790 795 800

Thr Tyr Ile Val Gln Thr Ile Lys Asp Ile Tyr Ala Lys Asp Asn Thr
805 810 815

Thr Leu Lys Lys Gln Phe Asp Lys Ser Pro Glu Lys Phe Leu Met Tyr
820 825 830

Gln His Asp Pro Arg Thr Phe Glu Lys Leu Glu Val Ile Met Lys Gln
835 840 845

Tyr Ala Asn Glu Lys Asn Pro Leu Ala Lys Tyr His Glu Glu Thr Gly
850 855 860

Glu Tyr Leu Thr Lys Tyr Ser Lys Lys Asn Asn Gly Pro Ile Val Lys
865 870 875 880

Ser Leu Lys Tyr Ile Gly Asn Lys Leu Gly Ser His Leu Asp Val Thr
885 890 895

His Gln Phe Lys Ser Ser Thr Lys Lys Leu Val Lys Leu Ser Ile Lys

900 905 910

Asn Tyr Arg Phe Asp Val Tyr Leu Thr Glu Lys Gly Tyr Lys Phe Val
915 920 925

Thr Ile Ala Tyr Leu Asn Val Phe Lys Lys Asp Asn Tyr Tyr Tyr Ile
930 935 940

Pro Lys Asp Lys Tyr Gln Glu Leu Lys Glu Lys Lys Lys Ile Lys Asp
945 950 955 960

Thr Asp Gln Phe Ile Ala Ser Phe Tyr Lys Asn Asp Leu Ile Lys Leu
965 970 975

Asn Gly Asp Leu Tyr Lys Ile Ile Gly Val Asn Ser Asp Asp Arg Asn
980 985 990

Ile Ile Glu Leu Asp Tyr Tyr Asp Ile Lys Tyr Lys Asp Tyr Cys Glu
995 1000 1005

Ile Asn Asn Ile Lys Gly Glu Pro Arg Ile Lys Lys Thr Ile Gly
1010 1015 1020

[0010] Lys Lys Thr Glu Ser Ile Glu Lys Phe Thr Thr Asp Val Leu Gly
1025 1030 1035

Asn Leu Tyr Leu His Ser Thr Glu Lys Ala Pro Gln Leu Ile Phe
1040 1045 1050

Lys Arg Gly Leu
1055

<210> 3
<211> 1057
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> Gib11Spa-2的多肽序列

<400> 3

Met Asn Gln Lys Phe Ile Leu Gly Leu Asp Ile Gly Ile Thr Ser Val
1 5 10 15

Gly Tyr Gly Leu Ile Asp Tyr Glu Thr Lys Asn Ile Ile Asp Ala Gly
20 25 30

Val Arg Leu Phe Pro Glu Ala Asn Val Glu Asn Asn Glu Gly Arg Arg
35 40 45

Ser Lys Arg Gly Ser Arg Arg Leu Lys Arg Arg Arg Ile His Arg Leu
50 55 60

Glu Arg Val Lys Leu Leu Leu Thr Glu Tyr Asp Leu Ile Asn Lys Glu
65 70 75 80

Gln Ile Pro Thr Ser Asn Asn Pro Tyr Gln Ile Arg Val Lys Gly Leu
85 90 95

Ser Glu Ile Leu Ser Lys Asp Glu Leu Ala Ile Ala Leu Leu His Leu
100 105 110

Ala Lys Arg Arg Gly Ile His Asn Val Asp Val Ala Ala Asp Lys Glu
115 120 125

Glu Thr Ala Ser Asp Ser Leu Ser Thr Lys Asp Gln Ile Asn Lys Asn
130 135 140

Ala Lys Phe Leu Glu Ser Arg Tyr Val Cys Glu Leu Gln Lys Glu Arg
145 150 155 160

[0011]

Leu Glu Asn Glu Gly His Val Arg Gly Val Glu Asn Arg Phe Leu Thr
165 170 175

Lys Asp Ile Val Arg Glu Ala Lys Lys Ile Ile Asp Thr Gln Met Gln
180 185 190

Tyr Tyr Pro Glu Ile Asp Glu Thr Phe Lys Glu Lys Tyr Ile Ser Leu
195 200 205

Val Glu Thr Arg Arg Glu Tyr Phe Glu Gly Pro Gly Gln Gly Ser Pro
210 215 220

Phe Gly Trp Asn Gly Asp Leu Lys Lys Trp Tyr Glu Met Leu Met Gly
225 230 235 240

His Cys Thr Tyr Phe Pro Gln Glu Leu Arg Ser Val Lys Tyr Ala Tyr
245 250 255

Ser Ala Asp Leu Phe Asn Ala Leu Asn Asp Leu Asn Asn Leu Ile Ile
260 265 270

Gln Arg Asp Asn Ser Glu Lys Leu Glu Tyr His Glu Lys Tyr His Ile
275 280 285

```
Ile Glu Asn Val Phe Lys Gln Lys Lys Lys Pro Thr Leu Lys Gln Ile
    290                 295                 300

Ala Lys Glu Ile Gly Val Asn Pro Glu Asp Ile Lys Gly Tyr Arg Ile
305                 310                 315                 320

Thr Lys Ser Gly Thr Pro Glu Phe Thr Ser Phe Lys Leu Phe His Asp
                325                 330                 335

Leu Lys Lys Val Val Lys Asp His Ala Ile Leu Asp Asp Ile Asp Leu
            340                 345                 350

Leu Asn Gln Ile Ala Glu Ile Leu Thr Ile Tyr Gln Asp Lys Asp Ser
        355                 360                 365

Ile Val Ala Glu Leu Gly Gln Leu Glu Tyr Leu Met Ser Glu Ala Asp
    370                 375                 380

Lys Gln Ser Ile Ser Glu Leu Thr Gly Tyr Thr Gly Thr His Ser Leu
385                 390                 395                 400

Ser Leu Lys Cys Met Asn Met Ile Ile Asp Glu Leu Trp His Ser Ser
                405                 410                 415
```

[0012]

```
Met Asn Gln Met Glu Val Phe Thr Tyr Leu Asn Met Arg Pro Lys Lys
            420                 425                 430

Tyr Glu Leu Lys Gly Tyr Gln Arg Ile Pro Thr Asp Met Ile Asp Asp
        435                 440                 445

Ala Ile Leu Ser Pro Val Val Lys Arg Thr Phe Ile Gln Ser Ile Asn
    450                 455                 460

Val Ile Asn Lys Val Ile Glu Lys Tyr Gly Ile Pro Glu Asp Ile Ile
465                 470                 475                 480

Ile Glu Leu Ala Arg Glu Asn Asn Ser Asp Asp Arg Lys Lys Phe Ile
                485                 490                 495

Asn Asn Leu Gln Lys Lys Asn Glu Ala Thr Arg Lys Arg Ile Asn Glu
            500                 505                 510

Ile Ile Gly Gln Thr Gly Asn Gln Asn Ala Lys Arg Ile Val Glu Lys
        515                 520                 525

Ile Arg Leu His Asp Gln Gln Glu Gly Lys Cys Leu Tyr Ser Leu Glu
    530                 535                 540
```

```
Ser Ile Pro Leu Glu Asp Leu Leu Asn Asn Pro Asn His Tyr Glu Val
545                 550                 555                 560

Asp His Ile Ile Pro Arg Ser Val Ser Phe Asp Asn Ser Tyr His Asn
                565                 570                 575

Lys Val Leu Val Lys Gln Ser Glu Asn Ser Lys Lys Ser Asn Leu Thr
            580                 585                 590

Pro Tyr Gln Tyr Phe Asn Ser Gly Lys Ser Lys Leu Ser Tyr Asn Gln
        595                 600                 605

Phe Lys Gln His Ile Leu Asn Leu Ser Lys Ser Gln Asp Arg Ile Ser
    610                 615                 620

Lys Lys Lys Lys Glu Tyr Leu Leu Glu Glu Arg Asp Ile Asn Lys Phe
625                 630                 635                 640

Glu Val Gln Lys Glu Phe Ile Asn Arg Asn Leu Val Asp Thr Arg Tyr
                645                 650                 655

Ala Thr Arg Glu Leu Thr Asn Tyr Leu Lys Ala Tyr Phe Ser Ala Asn
            660                 665                 670
```

[0013]

```
Asn Met Asn Val Lys Val Lys Thr Ile Asn Gly Ser Phe Thr Asp Tyr
        675                 680                 685

Leu Arg Lys Val Trp Lys Phe Lys Lys Glu Arg Asn His Gly Tyr Lys
    690                 695                 700

His His Ala Glu Asp Ala Leu Ile Ile Ala Asn Ala Asp Phe Leu Phe
705                 710                 715                 720

Lys Glu Asn Lys Lys Leu Lys Ala Val Asn Ser Val Leu Glu Lys Pro
                725                 730                 735

Glu Ile Glu Thr Lys Gln Leu Asp Ile Gln Val Asp Ser Glu Asp Asn
            740                 745                 750

Tyr Ser Glu Met Phe Ile Ile Pro Lys Gln Val Gln Asp Ile Lys Asp
        755                 760                 765

Phe Arg Asn Phe Lys Phe Ser His Arg Val Asp Lys Lys Pro Asn Arg
    770                 775                 780

Gln Leu Ile Asn Asp Thr Leu Tyr Ser Thr Arg Met Lys Asp Glu His
785                 790                 795                 800
```

Asp Tyr Ile Val Gln Thr Ile Thr Asp Ile Tyr Gly Lys Asp Asn Thr
805 810 815

Asn Leu Lys Lys Gln Phe Asn Lys Asn Pro Glu Lys Phe Leu Met Tyr
820 825 830

Gln Asn Asp Pro Lys Thr Phe Glu Lys Leu Ser Ile Ile Met Lys Gln
835 840 845

Tyr Ser Asp Glu Lys Asn Pro Leu Ala Lys Tyr Tyr Glu Glu Thr Gly
850 855 860

Glu Tyr Leu Thr Lys Tyr Ser Lys Lys Asn Asn Gly Pro Ile Val Lys
865 870 875 880

Lys Ile Lys Leu Leu Gly Asn Lys Val Gly Asn His Leu Asp Val Thr
885 890 895

Asn Lys Tyr Glu Asn Ser Thr Lys Lys Leu Val Lys Leu Ser Ile Lys
900 905 910

Asn Tyr Arg Phe Asp Val Tyr Leu Thr Glu Lys Gly Tyr Lys Phe Val
915 920 925

[0014]

Thr Ile Ala Tyr Leu Asn Val Phe Lys Lys Asp Asn Tyr Tyr Tyr Ile
930 935 940

Pro Lys Asp Lys Tyr Gln Glu Leu Lys Glu Lys Lys Lys Ile Lys Asp
945 950 955 960

Thr Asp Gln Phe Ile Ala Ser Phe Tyr Lys Asn Asp Leu Ile Lys Leu
965 970 975

Asn Gly Asp Leu Tyr Lys Ile Ile Gly Val Asn Ser Asp Asp Arg Asn
980 985 990

Ile Ile Glu Leu Asp Tyr Tyr Asp Ile Lys Tyr Lys Asp Tyr Cys Glu
995 1000 1005

Ile Asn Asn Ile Lys Gly Glu Pro Arg Ile Lys Lys Thr Ile Gly
1010 1015 1020

Lys Lys Thr Glu Ser Ile Glu Lys Phe Thr Thr Asp Val Leu Gly
1025 1030 1035

Asn Leu Tyr Leu His Ser Thr Glu Lys Ala Pro Gln Leu Ile Phe
1040 1045 1050

Lys Arg Gly Leu
1055

<210> 4
<211> 1057
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> Gib11Spa-3的多肽序列

<400> 4

Met Asn Gln Lys Phe Ile Leu Gly Leu Asp Ile Gly Ile Thr Ser Val
1 5 10 15

Gly Tyr Gly Leu Ile Asp Tyr Glu Thr Lys Asn Ile Ile Asp Ala Gly
20 25 30

Val Arg Leu Phe Pro Glu Ala Asn Val Glu Asn Asn Glu Gly Arg Arg
35 40 45

[0015]

Ser Lys Arg Gly Ser Arg Arg Leu Lys Arg Arg Arg Ile His Arg Leu
50 55 60

Glu Arg Val Lys Leu Leu Leu Thr Glu Tyr Asp Leu Ile Asn Lys Glu
65 70 75 80

Gln Ile Pro Thr Ser Asn Asn Pro Tyr Gln Ile Arg Val Lys Gly Leu
85 90 95

Ser Glu Ile Leu Ser Lys Asp Glu Leu Ala Ile Ala Leu Leu His Leu
100 105 110

Ala Lys Arg Arg Gly Ile His Asn Val Asp Val Ala Ala Asp Lys Glu
115 120 125

Glu Thr Ala Ser Asp Ser Leu Ser Thr Lys Asp Gln Ile Asn Lys Asn
130 135 140

Ala Lys Phe Leu Glu Ser Arg Tyr Val Cys Glu Leu Gln Lys Glu Arg
145 150 155 160

Leu Glu Asn Glu Gly His Val Arg Gly Val Glu Asn Arg Phe Leu Thr
165 170 175

```
Lys Asp Ile Val Arg Glu Ala Lys Lys Ile Ile Asp Thr Gln Met Gln
            180                 185                 190

Tyr Tyr Pro Glu Ile Asp Glu Thr Phe Lys Glu Lys Tyr Ile Ser Leu
        195                 200                 205

Val Glu Thr Arg Arg Glu Tyr Phe Glu Gly Pro Gly Gln Gly Ser Pro
    210                 215                 220

Phe Gly Trp Asn Gly Asp Leu Lys Lys Trp Tyr Glu Met Leu Met Gly
225                 230                 235                 240

His Cys Thr Tyr Phe Pro Gln Glu Leu Arg Ser Val Lys Tyr Ala Tyr
                245                 250                 255

Ser Ala Asp Leu Phe Asn Ala Leu Asn Asp Leu Asn Asn Leu Ile Ile
            260                 265                 270

Gln Arg Asp Asn Ser Glu Lys Leu Glu Tyr His Glu Lys Tyr His Ile
        275                 280                 285

Ile Glu Asn Val Phe Lys Gln Lys Lys Lys Pro Thr Leu Lys Gln Ile
    290                 295                 300
```

[0016]

```
Ala Lys Glu Ile Gly Val Asn Pro Glu Asp Ile Lys Gly Tyr Arg Ile
305                 310                 315                 320

Thr Lys Ser Gly Thr Pro Glu Phe Thr Ser Phe Lys Leu Phe His Asp
                325                 330                 335

Leu Lys Lys Val Val Lys Asp His Ala Ile Leu Asp Asp Ile Asp Leu
            340                 345                 350

Leu Asn Gln Ile Ala Glu Ile Leu Thr Ile Tyr Gln Asp Lys Asp Ser
        355                 360                 365

Ile Val Ala Glu Leu Gly Gln Leu Glu Tyr Leu Met Ser Glu Ala Asp
    370                 375                 380

Lys Gln Ser Ile Ser Glu Leu Thr Gly Tyr Thr Gly Thr His Ser Leu
385                 390                 395                 400

Ser Leu Lys Cys Met Asn Met Ile Ile Asp Glu Leu Trp His Ser Ser
                405                 410                 415

Met Asn Gln Met Glu Val Phe Thr Tyr Leu Asn Met Arg Pro Lys Lys
            420                 425                 430
```

```
Tyr Glu Leu Lys Gly Tyr Gln Arg Ile Pro Thr Asp Met Ile Asp Asp
        435                 440                 445

Ala Ile Leu Ser Pro Val Val Lys Arg Thr Phe Ile Gln Ser Ile Asn
    450                 455                 460

Val Ile Asn Lys Val Ile Glu Lys Tyr Gly Ile Pro Glu Asp Ile Ile
465                 470                 475                 480

Ile Glu Leu Ala Arg Glu Asn Asn Ser Asp Asp Arg Lys Lys Phe Ile
                485                 490                 495

Asn Asn Leu Gln Lys Lys Asn Glu Ala Thr Arg Lys Arg Ile Asn Glu
            500                 505                 510

Ile Ile Gly Gln Thr Gly Asn Gln Asn Ala Lys Arg Ile Val Glu Lys
        515                 520                 525

Ile Arg Leu His Asp Gln Gln Glu Gly Lys Cys Leu Tyr Ser Leu Glu
    530                 535                 540

Ser Ile Pro Leu Glu Asp Leu Leu Asn Asn Pro Asn His Tyr Glu Val
545                 550                 555                 560
```

[0017]

```
Asp His Ile Ile Pro Arg Ser Val Ser Phe Asp Asn Ser Tyr His Asn
                565                 570                 575

Lys Val Leu Val Lys Gln Ser Glu Asn Ser Lys Lys Ser Asn Leu Thr
            580                 585                 590

Pro Tyr Gln Tyr Phe Asn Ser Gly Lys Ser Lys Leu Ser Tyr Asn Gln
        595                 600                 605

Phe Lys Gln His Ile Leu Asn Leu Ser Lys Ser Gln Asp Arg Ile Ser
    610                 615                 620

Lys Lys Lys Lys Glu Tyr Leu Leu Glu Glu Arg Asp Ile Asn Lys Phe
625                 630                 635                 640

Glu Val Gln Lys Glu Phe Ile Asn Arg Asn Leu Val Asp Thr Arg Tyr
                645                 650                 655

Ala Thr Arg Glu Leu Thr Ser Tyr Leu Lys Ala Tyr Phe Ser Ala Asn
            660                 665                 670

Asn Met Asp Val Lys Val Lys Thr Ile Asn Gly Ser Phe Thr Asn His
        675                 680                 685
```

```
Leu Arg Lys Val Trp Arg Phe Asp Lys Tyr Arg Asn His Gly Tyr Lys
    690                 695                 700

His His Ala Glu Asp Ala Leu Ile Ile Ala Asn Ala Asp Phe Leu Phe
705                 710                 715                 720

Lys Glu Asn Lys Lys Leu Gln Asn Thr Asn Lys Ile Leu Glu Lys Pro
                725                 730                 735

Thr Ile Glu Asn Asn Thr Lys Lys Val Thr Val Glu Lys Glu Glu Asp
            740                 745                 750

Tyr Asn Asn Val Phe Glu Thr Pro Lys Leu Val Glu Asp Ile Lys Gln
        755                 760                 765

Tyr Arg Asp Tyr Lys Phe Ser His Arg Val Asp Lys Lys Pro Asn Arg
    770                 775                 780

Gln Leu Ile Asn Asp Thr Leu Tyr Ser Thr Arg Met Lys Asp Glu His
785                 790                 795                 800

Asp Tyr Ile Val Gln Thr Ile Thr Asp Ile Tyr Gly Lys Asp Asn Thr
                805                 810                 815
```

[0018]

```
Asn Leu Lys Lys Gln Phe Asn Lys Asn Pro Glu Lys Phe Leu Met Tyr
            820                 825                 830

Gln Asn Asp Pro Lys Thr Phe Glu Lys Leu Ser Ile Ile Met Lys Gln
        835                 840                 845

Tyr Ser Asp Glu Lys Asn Pro Leu Ala Lys Tyr Tyr Glu Glu Thr Gly
    850                 855                 860

Glu Tyr Leu Thr Lys Tyr Ser Lys Lys Asn Asn Gly Pro Ile Val Lys
865                 870                 875                 880

Lys Ile Lys Leu Leu Gly Asn Lys Val Gly Asn His Leu Asp Val Thr
                885                 890                 895

Asn Lys Tyr Glu Asn Ser Thr Lys Lys Leu Val Lys Leu Ser Ile Lys
            900                 905                 910

Asn Tyr Arg Phe Asp Val Tyr Leu Thr Glu Lys Gly Tyr Lys Phe Val
        915                 920                 925

Thr Ile Ala Tyr Leu Asn Val Phe Lys Lys Asp Asn Tyr Tyr Tyr Ile
    930                 935                 940
```

Pro Lys Asp Lys Tyr Gln Glu Leu Lys Glu Lys Lys Lys Ile Lys Asp
945 950 955 960

Thr Asp Gln Phe Ile Ala Ser Phe Tyr Lys Asn Asp Leu Ile Lys Leu
965 970 975

Asn Gly Asp Leu Tyr Lys Ile Ile Gly Val Asn Ser Asp Asp Arg Asn
980 985 990

Ile Ile Glu Leu Asp Tyr Tyr Asp Ile Lys Tyr Lys Asp Tyr Cys Glu
995 1000 1005

Ile Asn Asn Ile Lys Gly Glu Pro Arg Ile Lys Lys Thr Ile Gly
1010 1015 1020

Lys Lys Thr Glu Ser Ile Glu Lys Phe Thr Thr Asp Val Leu Gly
1025 1030 1035

Asn Leu Tyr Leu His Ser Thr Glu Lys Ala Pro Gln Leu Ile Phe
1040 1045 1050

Lys Arg Gly Leu
1055

[0019]

<210> 5
<211> 1054
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> P2H12的多肽序列

<400> 5

Met Asn Gln Lys Phe Ile Leu Gly Leu Asp Ile Gly Ile Thr Ser Val
1 5 10 15

Gly Tyr Gly Leu Ile Asp Tyr Glu Thr Lys Asn Ile Ile Asp Ala Gly
20 25 30

Val Arg Leu Phe Pro Glu Ala Asn Val Glu Asn Asn Glu Gly Arg Arg
35 40 45

Ser Lys Arg Gly Ser Arg Arg Leu Lys Arg Arg Arg Ile His Arg Leu
50 55 60

```
Glu Arg Val Lys Lys Leu Leu Glu Asp Tyr Asn Leu Leu Asp Gln Ser
65                  70                  75                  80

Gln Ile Pro Gln Ser Thr Asn Pro Tyr Ala Ile Arg Val Lys Gly Leu
                85                  90                  95

Ser Glu Ala Leu Ser Lys Asp Glu Leu Val Ile Ala Leu Leu His Ile
            100                 105                 110

Ala Lys Arg Arg Gly Ile His Asn Ile Asn Val Ser Ser Glu Asp Glu
        115                 120                 125

Asp Ala Ser Asn Glu Leu Ser Thr Lys Glu Gln Ile Asn Arg Asn Asn
    130                 135                 140

Lys Leu Leu Lys Asp Lys Tyr Val Cys Glu Val Gln Leu Gln Arg Leu
145                 150                 155                 160

Lys Glu Gly Gln Ile Arg Gly Glu Lys Asn Arg Phe Lys Thr Thr Asp
                165                 170                 175

Ile Leu Lys Glu Ile Asp Gln Leu Leu Lys Val Gln Lys Asp Tyr His
            180                 185                 190
```

[0020]

```
Asn Leu Asp Ile Asp Phe Ile Asn Gln Tyr Lys Glu Ile Val Glu Thr
        195                 200                 205

Arg Arg Glu Tyr Phe Glu Gly Pro Gly Lys Gly Ser Pro Tyr Gly Trp
    210                 215                 220

Glu Gly Asp Pro Lys Ala Trp Tyr Glu Thr Leu Met Gly His Cys Thr
225                 230                 235                 240

Tyr Phe Pro Asp Glu Leu Arg Ser Val Lys Tyr Ala Tyr Ser Ala Asp
                245                 250                 255

Leu Phe Asn Ala Leu Asn Asp Leu Asn Asn Leu Val Ile Gln Arg Asp
            260                 265                 270

Gly Leu Ser Lys Leu Glu Tyr His Glu Lys Tyr His Ile Ile Glu Asn
        275                 280                 285

Val Phe Lys Gln Lys Lys Lys Pro Thr Leu Lys Gln Ile Ala Asn Glu
    290                 295                 300

Ile Asn Val Asn Pro Glu Asp Ile Lys Gly Tyr Arg Ile Thr Lys Ser
305                 310                 315                 320
```

```
Gly Lys Pro Glu Phe Thr Ser Phe Lys Leu Phe His Asp Leu Lys Lys
                325                 330                 335

Val Val Lys Asp His Ala Ile Leu Asp Asp Ile Asp Leu Leu Asn Gln
            340                 345                 350

Ile Ala Glu Ile Leu Thr Ile Tyr Gln Asp Lys Asp Ser Ile Val Ala
        355                 360                 365

Glu Leu Gly Gln Leu Glu Tyr Leu Met Ser Glu Ala Asp Lys Gln Ser
    370                 375                 380

Ile Ser Glu Leu Thr Gly Tyr Thr Gly Thr His Ser Leu Ser Leu Lys
385                 390                 395                 400

Cys Met Asn Met Ile Ile Asp Glu Leu Trp His Ser Ser Met Asn Gln
                405                 410                 415

Met Glu Val Phe Thr Tyr Leu Asn Met Arg Pro Lys Lys Tyr Glu Leu
            420                 425                 430

Lys Gly Tyr Gln Arg Ile Pro Thr Asp Met Ile Asp Asp Ala Ile Leu
        435                 440                 445
```

[0021]

```
Ser Pro Val Val Lys Arg Thr Phe Ile Gln Ser Ile Asn Val Ile Asn
    450                 455                 460

Lys Val Ile Glu Lys Tyr Gly Ile Pro Glu Asp Ile Ile Ile Glu Leu
465                 470                 475                 480

Ala Arg Glu Asn Asn Ser Asp Asp Arg Lys Lys Phe Ile Asn Asn Leu
                485                 490                 495

Gln Lys Lys Asn Glu Ala Thr Arg Lys Arg Ile Asn Glu Ile Ile Gly
            500                 505                 510

Gln Thr Gly Asn Gln Asn Ala Lys Arg Ile Val Glu Lys Ile Arg Leu
        515                 520                 525

His Asp Gln Gln Glu Gly Lys Cys Leu Tyr Ser Leu Glu Ser Ile Pro
    530                 535                 540

Leu Glu Asp Leu Leu Asn Asn Pro Asn His Tyr Glu Val Asp His Ile
545                 550                 555                 560

Ile Pro Arg Ser Val Ser Phe Asp Asn Ser Tyr His Asn Lys Val Leu
                565                 570                 575
```

```
Val Lys Gln Ser Glu Asn Ser Lys Lys Ser Asn Leu Thr Pro Tyr Gln
            580                 585                 590

Tyr Phe Asn Ser Gly Lys Ser Lys Leu Ser Tyr Asn Gln Phe Lys Gln
        595                 600                 605

His Ile Leu Asn Leu Ser Lys Ser Gln Asp Arg Ile Ser Lys Lys Lys
    610                 615                 620

Lys Glu Tyr Leu Leu Glu Glu Arg Asp Ile Asn Lys Phe Glu Val Gln
625                 630                 635                 640

Lys Glu Phe Ile Asn Arg Asn Leu Val Asp Thr Arg Tyr Ala Thr Arg
                645                 650                 655

Glu Leu Thr Asn Tyr Leu Lys Ala Tyr Phe Ser Ala Asn Asn Met Asn
            660                 665                 670

Val Lys Val Lys Thr Ile Asn Gly Ser Phe Thr Asp Tyr Leu Arg Lys
        675                 680                 685

Val Trp Lys Phe Lys Lys Glu Arg Asn His Gly Tyr Lys His His Ala
    690                 695                 700
```

[0022]

```
Glu Asp Ala Leu Ile Ile Ala Asn Ala Asp Phe Leu Phe Lys Glu Asn
705                 710                 715                 720

Lys Lys Leu Lys Ala Val Asn Ser Val Leu Glu Lys Pro Glu Ile Glu
                725                 730                 735

Thr Lys Gln Leu Asp Ile Gln Val Asp Ser Glu Asp Asn Tyr Ser Glu
            740                 745                 750

Met Phe Ile Ile Pro Lys Gln Val Gln Asp Ile Lys Asp Phe Arg Asn
        755                 760                 765

Phe Lys Tyr Ser His Arg Val Asp Lys Lys Pro Asn Arg Gln Leu Ile
    770                 775                 780

Asn Asp Thr Leu Tyr Ser Thr Arg Lys Lys Asp Asn Ser Thr Tyr Ile
785                 790                 795                 800

Val Gln Thr Ile Lys Asp Ile Tyr Ala Lys Asp Asn Thr Thr Leu Lys
                805                 810                 815

Lys Gln Phe Asp Lys Ser Pro Glu Lys Phe Leu Met Tyr Gln His Asp
            820                 825                 830
```

```
Pro Arg Thr Phe Glu Lys Leu Glu Val Ile Met Lys Gln Tyr Ala Asn
        835                 840                 845

Glu Lys Asn Pro Leu Ala Lys Tyr His Glu Glu Thr Gly Glu Tyr Leu
    850                 855                 860

Thr Lys Tyr Ser Lys Lys Asn Asn Gly Pro Ile Val Lys Ser Leu Lys
865                 870                 875                 880

Tyr Ile Gly Asn Lys Leu Gly Ser His Leu Asp Val Thr His Gln Phe
                885                 890                 895

Lys Ser Ser Thr Lys Lys Leu Val Lys Leu Ser Ile Lys Pro Tyr Arg
            900                 905                 910

Phe Asp Val Tyr Leu Thr Asp Lys Gly Tyr Lys Phe Ile Thr Ile Ser
        915                 920                 925

Tyr Leu Asp Val Leu Lys Lys Asp Asn Tyr Tyr Tyr Ile Pro Glu Gln
    930                 935                 940

Lys Tyr Asp Lys Leu Lys Leu Gly Lys Ala Ile Asp Lys Asn Ala Lys
945                 950                 955                 960
```

[0023]

```
Phe Ile Ala Ser Phe Tyr Lys Asn Asp Leu Ile Lys Leu Asp Gly Glu
                965                 970                 975

Ile Tyr Lys Ile Ile Gly Val Asn Ser Asp Thr Arg Asn Met Ile Glu
            980                 985                 990

Leu Asp Leu Pro Asp Ile Arg Tyr  Lys Glu Tyr Cys Glu  Leu Asn Asn
        995                 1000                1005

Ile Lys  Gly Glu Pro Arg Ile  Lys Lys Thr Ile Gly  Lys Lys Val
    1010                1015                1020

Asn Ser  Ile Glu Lys Leu Thr  Thr Asp Val Leu Gly  Asn Val Phe
    1025                1030                1035

Thr Asn  Thr Gln Tyr Thr Lys  Pro Gln Leu Leu Phe  Lys Arg Gly
    1040                1045                1050

Asn
```

<210> 6
<211> 1057
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> E2的多肽序列

<400> 6

Met Asn Gln Lys Phe Ile Leu Gly Leu Asp Ile Gly Ile Thr Ser Val
1 5 10 15

Gly Tyr Gly Leu Ile Asp Tyr Glu Thr Lys Asn Ile Ile Asp Ala Gly
20 25 30

Val Arg Leu Phe Pro Glu Ala Asn Val Glu Asn Asn Glu Gly Arg Arg
35 40 45

Ser Lys Arg Gly Ser Arg Arg Leu Lys Arg Arg Arg Ile His Arg Leu
50 55 60

Asp Arg Val Lys His Leu Leu Ala Glu Tyr Asp Leu Leu Asp Leu Thr
65 70 75 80

Asn Ile Pro Lys Ser Thr Asn Pro Tyr Gln Thr Arg Val Lys Gly Leu
[0024] 85 90 95

Asn Glu Lys Leu Ser Lys Asp Glu Leu Val Ile Ala Leu Leu His Ile
100 105 110

Ala Lys Arg Arg Gly Ile His Asn Val Asp Val Ala Ala Asp Lys Glu
115 120 125

Glu Thr Ala Ser Asp Ser Leu Ser Thr Lys Asp Gln Ile Asn Lys Asn
130 135 140

Ala Lys Phe Leu Glu Ser Arg Tyr Val Cys Glu Leu Gln Lys Glu Arg
145 150 155 160

Leu Glu Asn Glu Gly His Val Arg Gly Val Glu Asn Arg Phe Leu Thr
165 170 175

Lys Asp Ile Val Arg Glu Ala Lys Lys Ile Ile Asp Thr Gln Met Gln
180 185 190

Tyr Tyr Pro Glu Ile Asp Glu Thr Phe Lys Glu Lys Tyr Ile Ser Leu
195 200 205

Val Glu Thr Arg Arg Glu Tyr Phe Glu Gly Pro Gly Lys Gly Ser Pro

210 215 220

Phe Gly Trp Glu Gly Asn Ile Lys Lys Trp Phe Glu Gln Met Met Gly
225 230 235 240

His Cys Thr Tyr Phe Pro Glu Glu Leu Arg Ser Val Lys Tyr Ser Tyr
245 250 255

Ser Ala Glu Leu Phe Asn Ala Leu Asn Asp Leu Asn Asn Leu Val Ile
260 265 270

Thr Arg Asp Glu Asp Ala Lys Leu Asn Tyr Gly Glu Lys Phe Gln Ile
275 280 285

Ile Glu Asn Val Phe Lys Gln Lys Lys Thr Pro Asn Leu Lys Gln Ile
290 295 300

Ala Ile Glu Ile Gly Val His Glu Thr Glu Ile Lys Gly Tyr Arg Val
305 310 315 320

Asn Lys Ser Gly Thr Pro Glu Phe Thr Glu Phe Lys Leu Tyr His Asp
325 330 335

[0025] Leu Lys Ser Ile Val Phe Asp Lys Ser Ile Leu Glu Asn Glu Ala Ile
340 345 350

Leu Asp Gln Ile Ala Glu Ile Leu Thr Ile Tyr Gln Asp Glu Gln Ser
355 360 365

Ile Lys Glu Glu Leu Asn Lys Leu Pro Glu Ile Leu Asn Glu Gln Asp
370 375 380

Lys Ala Glu Ile Ala Lys Leu Ile Gly Tyr Asn Gly Thr His Arg Leu
385 390 395 400

Ser Leu Lys Cys Ile His Leu Ile Asn Glu Glu Leu Trp Gln Thr Ser
405 410 415

Arg Asn Gln Met Glu Ile Phe Asn Tyr Leu Asn Ile Lys Pro Asn Lys
420 425 430

Val Asp Leu Ser Glu Gln Asn Lys Ile Pro Lys Asp Met Val Asn Asp
435 440 445

Phe Ile Leu Ser Pro Val Val Lys Arg Thr Phe Ile Gln Ser Ile Asn
450 455 460

Val Ile Asn Lys Val Ile Glu Lys Tyr Gly Ile Pro Glu Asp Ile Ile

465 470 475 480

Ile Glu Leu Ala Arg Glu Asn Asn Ser Asp Asp Arg Lys Lys Phe Ile
485 490 495

Asn Asn Leu Gln Lys Lys Asn Glu Ala Thr Arg Lys Arg Ile Asn Glu
500 505 510

Ile Ile Gly Gln Thr Gly Asn Gln Asn Ala Lys Arg Ile Val Glu Lys
515 520 525

Ile Arg Leu His Asp Gln Gln Glu Gly Lys Cys Leu Tyr Ser Leu Lys
530 535 540

Asp Ile Pro Leu Glu Asp Leu Leu Arg Asn Pro Asn Asn Tyr Asp Ile
545 550 555 560

Asp His Ile Ile Pro Arg Ser Val Ser Phe Asp Asp Ser Met His Asn
565 570 575

Lys Val Leu Val Arg Arg Glu Gln Asn Ala Lys Lys Asn Asn Gln Thr
580 585 590

[0026] Pro Tyr Gln Tyr Leu Thr Ser Gly Tyr Ala Asp Ile Lys Tyr Ser Val
595 600 605

Phe Lys Gln His Val Leu Asn Leu Ala Glu Asn Lys Asp Arg Met Thr
610 615 620

Lys Lys Lys Arg Glu Tyr Leu Leu Glu Glu Arg Asp Ile Asn Lys Phe
625 630 635 640

Glu Val Gln Lys Glu Phe Ile Asn Arg Asn Leu Val Asp Thr Arg Tyr
645 650 655

Ala Thr Arg Glu Leu Thr Asn Tyr Leu Lys Ala Tyr Phe Ser Ala Asn
660 665 670

Asn Met Asn Val Lys Val Lys Thr Ile Asn Gly Ser Phe Thr Asp Tyr
675 680 685

Leu Arg Lys Val Trp Lys Phe Lys Lys Glu Arg Asn His Gly Tyr Lys
690 695 700

His His Ala Glu Asp Ala Leu Ile Ile Ala Asn Ala Asp Phe Leu Phe
705 710 715 720

Lys Glu Asn Lys Lys Leu Lys Ala Val Asn Ser Val Leu Glu Lys Pro

725 730 735

Glu Ile Glu Thr Lys Gln Leu Asp Ile Gln Val Asp Ser Glu Asp Asn
740 745 750

Tyr Ser Glu Met Phe Ile Ile Pro Lys Gln Val Gln Asp Ile Lys Asp
755 760 765

Phe Arg Asn Phe Lys Tyr Ser His Arg Val Asp Lys Lys Pro Asn Arg
770 775 780

Gln Leu Ile Asn Asp Thr Leu Tyr Ser Thr Arg Lys Lys Asp Asn Ser
785 790 795 800

Thr Tyr Ile Val Gln Thr Ile Lys Asp Ile Tyr Ala Lys Asp Asn Thr
805 810 815

Thr Leu Lys Lys Gln Phe Asp Lys Ser Pro Glu Lys Phe Leu Met Tyr
820 825 830

Gln His Asp Pro Arg Thr Phe Glu Lys Leu Glu Val Ile Met Lys Gln
835 840 845

[0027]

Tyr Ala Asn Glu Lys Asn Pro Leu Ala Lys Tyr His Glu Glu Thr Gly
850 855 860

Glu Tyr Leu Thr Lys Tyr Ser Lys Lys Asn Asn Gly Pro Ile Val Lys
865 870 875 880

Ser Leu Lys Tyr Ile Gly Asn Lys Leu Gly Ser His Leu Asp Val Thr
885 890 895

His Gln Phe Lys Ser Ser Thr Lys Lys Leu Val Lys Leu Ser Ile Lys
900 905 910

Pro Tyr Arg Phe Asp Val Tyr Leu Thr Asp Lys Gly Tyr Lys Phe Ile
915 920 925

Thr Ile Ser Tyr Leu Asp Val Leu Lys Lys Asp Asn Tyr Tyr Tyr Ile
930 935 940

Pro Glu Gln Lys Tyr Asp Lys Leu Lys Leu Gly Lys Ala Ile Asp Lys
945 950 955 960

Asn Ala Lys Phe Ile Ala Ser Phe Tyr Lys Asn Asp Leu Ile Lys Leu
965 970 975

Asp Gly Glu Ile Tyr Lys Ile Ile Gly Val Asn Ser Asp Thr Arg Asn

980 985 990

Met Ile Glu Leu Asp Leu Pro Asp Ile Arg Tyr Lys Glu Tyr Cys Glu
995 1000 1005

Leu Asn Asn Ile Lys Gly Glu Pro Arg Ile Lys Lys Thr Ile Gly
1010 1015 1020

Lys Lys Val Asn Ser Ile Glu Lys Leu Thr Thr Asp Val Leu Gly
1025 1030 1035

Asn Val Phe Thr Asn Thr Gln Tyr Thr Lys Pro Gln Leu Leu Phe
1040 1045 1050

Lys Arg Gly Asn
1055

<210> 7
<211> 1057
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

[0028]

<220>
<221> misc_feature
<223> E2-K741D-L743K的多肽序列

<400> 7

Met Asn Gln Lys Phe Ile Leu Gly Leu Asp Ile Gly Ile Thr Ser Val
1 5 10 15

Gly Tyr Gly Leu Ile Asp Tyr Glu Thr Lys Asn Ile Ile Asp Ala Gly
20 25 30

Val Arg Leu Phe Pro Glu Ala Asn Val Glu Asn Asn Glu Gly Arg Arg
35 40 45

Ser Lys Arg Gly Ser Arg Arg Leu Lys Arg Arg Arg Ile His Arg Leu
50 55 60

Asp Arg Val Lys His Leu Leu Ala Glu Tyr Asp Leu Leu Asp Leu Thr
65 70 75 80

Asn Ile Pro Lys Ser Thr Asn Pro Tyr Gln Thr Arg Val Lys Gly Leu
85 90 95

Asn Glu Lys Leu Ser Lys Asp Glu Leu Val Ile Ala Leu Leu His Ile
100 105 110

```
Ala Lys Arg Arg Gly Ile His Asn Val Asp Val Ala Ala Asp Lys Glu
        115                 120                 125

Glu Thr Ala Ser Asp Ser Leu Ser Thr Lys Asp Gln Ile Asn Lys Asn
    130                 135                 140

Ala Lys Phe Leu Glu Ser Arg Tyr Val Cys Glu Leu Gln Lys Glu Arg
145                 150                 155                 160

Leu Glu Asn Glu Gly His Val Arg Gly Val Glu Asn Arg Phe Leu Thr
                165                 170                 175

Lys Asp Ile Val Arg Glu Ala Lys Lys Ile Ile Asp Thr Gln Met Gln
            180                 185                 190

Tyr Tyr Pro Glu Ile Asp Glu Thr Phe Lys Glu Lys Tyr Ile Ser Leu
        195                 200                 205

Val Glu Thr Arg Arg Glu Tyr Phe Glu Gly Pro Gly Lys Gly Ser Pro
    210                 215                 220

Phe Gly Trp Glu Gly Asn Ile Lys Lys Trp Phe Glu Gln Met Met Gly
225                 230                 235                 240
```

[0029]

```
His Cys Thr Tyr Phe Pro Glu Glu Leu Arg Ser Val Lys Tyr Ser Tyr
                245                 250                 255

Ser Ala Glu Leu Phe Asn Ala Leu Asn Asp Leu Asn Asn Leu Val Ile
            260                 265                 270

Thr Arg Asp Glu Asp Ala Lys Leu Asn Tyr Gly Glu Lys Phe Gln Ile
        275                 280                 285

Ile Glu Asn Val Phe Lys Gln Lys Lys Thr Pro Asn Leu Lys Gln Ile
    290                 295                 300

Ala Ile Glu Ile Gly Val His Glu Thr Glu Ile Lys Gly Tyr Arg Val
305                 310                 315                 320

Asn Lys Ser Gly Thr Pro Glu Phe Thr Glu Phe Lys Leu Tyr His Asp
                325                 330                 335

Leu Lys Ser Ile Val Phe Asp Lys Ser Ile Leu Glu Asn Glu Ala Ile
            340                 345                 350

Leu Asp Gln Ile Ala Glu Ile Leu Thr Ile Tyr Gln Asp Glu Gln Ser
        355                 360                 365
```

```
Ile Lys Glu Glu Leu Asn Lys Leu Pro Glu Ile Leu Asn Glu Gln Asp
    370                 375                 380

Lys Ala Glu Ile Ala Lys Leu Ile Gly Tyr Asn Gly Thr His Arg Leu
385                 390                 395                 400

Ser Leu Lys Cys Ile His Leu Ile Asn Glu Glu Leu Trp Gln Thr Ser
                405                 410                 415

Arg Asn Gln Met Glu Ile Phe Asn Tyr Leu Asn Ile Lys Pro Asn Lys
            420                 425                 430

Val Asp Leu Ser Glu Gln Asn Lys Ile Pro Lys Asp Met Val Asn Asp
        435                 440                 445

Phe Ile Leu Ser Pro Val Val Lys Arg Thr Phe Ile Gln Ser Ile Asn
    450                 455                 460

Val Ile Asn Lys Val Ile Glu Lys Tyr Gly Ile Pro Glu Asp Ile Ile
465                 470                 475                 480

Ile Glu Leu Ala Arg Glu Asn Asn Ser Asp Asp Arg Lys Lys Phe Ile
                485                 490                 495
```

[0030]

```
Asn Asn Leu Gln Lys Lys Asn Glu Ala Thr Arg Lys Arg Ile Asn Glu
            500                 505                 510

Ile Ile Gly Gln Thr Gly Asn Gln Asn Ala Lys Arg Ile Val Glu Lys
        515                 520                 525

Ile Arg Leu His Asp Gln Gln Glu Gly Lys Cys Leu Tyr Ser Leu Lys
    530                 535                 540

Asp Ile Pro Leu Glu Asp Leu Leu Arg Asn Pro Asn Asn Tyr Asp Ile
545                 550                 555                 560

Asp His Ile Ile Pro Arg Ser Val Ser Phe Asp Asp Ser Met His Asn
                565                 570                 575

Lys Val Leu Val Arg Arg Glu Gln Asn Ala Lys Lys Asn Asn Gln Thr
            580                 585                 590

Pro Tyr Gln Tyr Leu Thr Ser Gly Tyr Ala Asp Ile Lys Tyr Ser Val
        595                 600                 605

Phe Lys Gln His Val Leu Asn Leu Ala Glu Asn Lys Asp Arg Met Thr
    610                 615                 620
```

Lys Lys Lys Arg Glu Tyr Leu Leu Glu Glu Arg Asp Ile Asn Lys Phe
625 630 635 640

Glu Val Gln Lys Glu Phe Ile Asn Arg Asn Leu Val Asp Thr Arg Tyr
645 650 655

Ala Thr Arg Glu Leu Thr Asn Tyr Leu Lys Ala Tyr Phe Ser Ala Asn
660 665 670

Asn Met Asn Val Lys Val Lys Thr Ile Asn Gly Ser Phe Thr Asp Tyr
675 680 685

Leu Arg Lys Val Trp Lys Phe Lys Lys Glu Arg Asn His Gly Tyr Lys
690 695 700

His His Ala Glu Asp Ala Leu Ile Ile Ala Asn Ala Asp Phe Leu Phe
705 710 715 720

Lys Glu Asn Lys Lys Leu Lys Ala Val Asn Ser Val Leu Glu Lys Pro
725 730 735

Glu Ile Glu Thr Asp Gln Lys Asp Ile Gln Val Asp Ser Glu Asp Asn
740 745 750

[0031]

Tyr Ser Glu Met Phe Ile Ile Pro Lys Gln Val Gln Asp Ile Lys Asp
755 760 765

Phe Arg Asn Phe Lys Tyr Ser His Arg Val Asp Lys Lys Pro Asn Arg
770 775 780

Gln Leu Ile Asn Asp Thr Leu Tyr Ser Thr Arg Lys Lys Asp Asn Ser
785 790 795 800

Thr Tyr Ile Val Gln Thr Ile Lys Asp Ile Tyr Ala Lys Asp Asn Thr
805 810 815

Thr Leu Lys Lys Gln Phe Asp Lys Ser Pro Glu Lys Phe Leu Met Tyr
820 825 830

Gln His Asp Pro Arg Thr Phe Glu Lys Leu Glu Val Ile Met Lys Gln
835 840 845

Tyr Ala Asn Glu Lys Asn Pro Leu Ala Lys Tyr His Glu Glu Thr Gly
850 855 860

Glu Tyr Leu Thr Lys Tyr Ser Lys Lys Asn Asn Gly Pro Ile Val Lys
865 870 875 880

Ser Leu Lys Tyr Ile Gly Asn Lys Leu Gly Ser His Leu Asp Val Thr
885 890 895

His Gln Phe Lys Ser Ser Thr Lys Lys Leu Val Lys Leu Ser Ile Lys
900 905 910

Pro Tyr Arg Phe Asp Val Tyr Leu Thr Asp Lys Gly Tyr Lys Phe Ile
915 920 925

Thr Ile Ser Tyr Leu Asp Val Leu Lys Lys Asp Asn Tyr Tyr Tyr Ile
930 935 940

Pro Glu Gln Lys Tyr Asp Lys Leu Lys Leu Gly Lys Ala Ile Asp Lys
945 950 955 960

Asn Ala Lys Phe Ile Ala Ser Phe Tyr Lys Asn Asp Leu Ile Lys Leu
965 970 975

Asp Gly Glu Ile Tyr Lys Ile Ile Gly Val Asn Ser Asp Thr Arg Asn
980 985 990

Met Ile Glu Leu Asp Leu Pro Asp Ile Arg Tyr Lys Glu Tyr Cys Glu
995 1000 1005

[0032]

Leu Asn Asn Ile Lys Gly Glu Pro Arg Ile Lys Lys Thr Ile Gly
1010 1015 1020

Lys Lys Val Asn Ser Ile Glu Lys Leu Thr Thr Asp Val Leu Gly
1025 1030 1035

Asn Val Phe Thr Asn Thr Gln Tyr Thr Lys Pro Gln Leu Leu Phe
1040 1045 1050

Lys Arg Gly Asn
1055

<210> 8
<211> 1057
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> E2-S670T-N675D的多肽序列

<400> 8

```
Met Asn Gln Lys Phe Ile Leu Gly Leu Asp Ile Gly Ile Thr Ser Val
1               5                  10                  15

Gly Tyr Gly Leu Ile Asp Tyr Glu Thr Lys Asn Ile Ile Asp Ala Gly
            20                  25                  30

Val Arg Leu Phe Pro Glu Ala Asn Val Glu Asn Asn Glu Gly Arg Arg
        35                  40                  45

Ser Lys Arg Gly Ser Arg Arg Leu Lys Arg Arg Arg Ile His Arg Leu
    50                  55                  60

Asp Arg Val Lys His Leu Leu Ala Glu Tyr Asp Leu Leu Asp Leu Thr
65                  70                  75                  80

Asn Ile Pro Lys Ser Thr Asn Pro Tyr Gln Thr Arg Val Lys Gly Leu
                85                  90                  95

Asn Glu Lys Leu Ser Lys Asp Glu Leu Val Ile Ala Leu Leu His Ile
            100                 105                 110

Ala Lys Arg Arg Gly Ile His Asn Val Asp Val Ala Ala Asp Lys Glu
        115                 120                 125
```

[0033]

```
Glu Thr Ala Ser Asp Ser Leu Ser Thr Lys Asp Gln Ile Asn Lys Asn
    130                 135                 140

Ala Lys Phe Leu Glu Ser Arg Tyr Val Cys Glu Leu Gln Lys Glu Arg
145                 150                 155                 160

Leu Glu Asn Glu Gly His Val Arg Gly Val Glu Asn Arg Phe Leu Thr
                165                 170                 175

Lys Asp Ile Val Arg Glu Ala Lys Lys Ile Ile Asp Thr Gln Met Gln
            180                 185                 190

Tyr Tyr Pro Glu Ile Asp Glu Thr Phe Lys Glu Lys Tyr Ile Ser Leu
        195                 200                 205

Val Glu Thr Arg Arg Glu Tyr Phe Glu Gly Pro Gly Lys Gly Ser Pro
    210                 215                 220

Phe Gly Trp Glu Gly Asn Ile Lys Lys Trp Phe Glu Gln Met Met Gly
225                 230                 235                 240

His Cys Thr Tyr Phe Pro Glu Glu Leu Arg Ser Val Lys Tyr Ser Tyr
                245                 250                 255
```

```
Ser Ala Glu Leu Phe Asn Ala Leu Asn Asp Leu Asn Asn Leu Val Ile
            260                 265                 270

Thr Arg Asp Glu Asp Ala Lys Leu Asn Tyr Gly Glu Lys Phe Gln Ile
        275                 280                 285

Ile Glu Asn Val Phe Lys Gln Lys Lys Thr Pro Asn Leu Lys Gln Ile
    290                 295                 300

Ala Ile Glu Ile Gly Val His Glu Thr Glu Ile Lys Gly Tyr Arg Val
305                 310                 315                 320

Asn Lys Ser Gly Thr Pro Glu Phe Thr Glu Phe Lys Leu Tyr His Asp
                325                 330                 335

Leu Lys Ser Ile Val Phe Asp Lys Ser Ile Leu Glu Asn Glu Ala Ile
            340                 345                 350

Leu Asp Gln Ile Ala Glu Ile Leu Thr Ile Tyr Gln Asp Glu Gln Ser
        355                 360                 365

Ile Lys Glu Glu Leu Asn Lys Leu Pro Glu Ile Leu Asn Glu Gln Asp
    370                 375                 380
```

[0034]

```
Lys Ala Glu Ile Ala Lys Leu Ile Gly Tyr Asn Gly Thr His Arg Leu
385                 390                 395                 400

Ser Leu Lys Cys Ile His Leu Ile Asn Glu Glu Leu Trp Gln Thr Ser
                405                 410                 415

Arg Asn Gln Met Glu Ile Phe Asn Tyr Leu Asn Ile Lys Pro Asn Lys
            420                 425                 430

Val Asp Leu Ser Glu Gln Asn Lys Ile Pro Lys Asp Met Val Asn Asp
        435                 440                 445

Phe Ile Leu Ser Pro Val Val Lys Arg Thr Phe Ile Gln Ser Ile Asn
    450                 455                 460

Val Ile Asn Lys Val Ile Glu Lys Tyr Gly Ile Pro Glu Asp Ile Ile
465                 470                 475                 480

Ile Glu Leu Ala Arg Glu Asn Asn Ser Asp Asp Arg Lys Lys Phe Ile
                485                 490                 495

Asn Asn Leu Gln Lys Lys Asn Glu Ala Thr Arg Lys Arg Ile Asn Glu
            500                 505                 510
```

Ile Ile Gly Gln Thr Gly Asn Gln Asn Ala Lys Arg Ile Val Glu Lys
515 520 525

Ile Arg Leu His Asp Gln Gln Glu Gly Lys Cys Leu Tyr Ser Leu Lys
530 535 540

Asp Ile Pro Leu Glu Asp Leu Leu Arg Asn Pro Asn Asn Tyr Asp Ile
545 550 555 560

Asp His Ile Ile Pro Arg Ser Val Ser Phe Asp Asp Ser Met His Asn
565 570 575

Lys Val Leu Val Arg Arg Glu Gln Asn Ala Lys Lys Asn Asn Gln Thr
580 585 590

Pro Tyr Gln Tyr Leu Thr Ser Gly Tyr Ala Asp Ile Lys Tyr Ser Val
595 600 605

Phe Lys Gln His Val Leu Asn Leu Ala Glu Asn Lys Asp Arg Met Thr
610 615 620

Lys Lys Lys Arg Glu Tyr Leu Leu Glu Glu Arg Asp Ile Asn Lys Phe
625 630 635 640

[0035]

Glu Val Gln Lys Glu Phe Ile Asn Arg Asn Leu Val Asp Thr Arg Tyr
645 650 655

Ala Thr Arg Glu Leu Thr Asn Tyr Leu Lys Ala Tyr Phe Thr Ala Asn
660 665 670

Asn Met Asp Val Lys Val Lys Thr Ile Asn Gly Ser Phe Thr Asp Tyr
675 680 685

Leu Arg Lys Val Trp Lys Phe Lys Lys Glu Arg Asn His Gly Tyr Lys
690 695 700

His His Ala Glu Asp Ala Leu Ile Ile Ala Asn Ala Asp Phe Leu Phe
705 710 715 720

Lys Glu Asn Lys Lys Leu Lys Ala Val Asn Ser Val Leu Glu Lys Pro
725 730 735

Glu Ile Glu Thr Lys Gln Leu Asp Ile Gln Val Asp Ser Glu Asp Asn
740 745 750

Tyr Ser Glu Met Phe Ile Ile Pro Lys Gln Val Gln Asp Ile Lys Asp
755 760 765

```
Phe Arg Asn Phe Lys Tyr Ser His Arg Val Asp Lys Lys Pro Asn Arg
    770                 775                 780

Gln Leu Ile Asn Asp Thr Leu Tyr Ser Thr Arg Lys Lys Asp Asn Ser
785                 790                 795                 800

Thr Tyr Ile Val Gln Thr Ile Lys Asp Ile Tyr Ala Lys Asp Asn Thr
                805                 810                 815

Thr Leu Lys Lys Gln Phe Asp Lys Ser Pro Glu Lys Phe Leu Met Tyr
            820                 825                 830

Gln His Asp Pro Arg Thr Phe Glu Lys Leu Glu Val Ile Met Lys Gln
        835                 840                 845

Tyr Ala Asn Glu Lys Asn Pro Leu Ala Lys Tyr His Glu Glu Thr Gly
    850                 855                 860

Glu Tyr Leu Thr Lys Tyr Ser Lys Lys Asn Asn Gly Pro Ile Val Lys
865                 870                 875                 880

Ser Leu Lys Tyr Ile Gly Asn Lys Leu Gly Ser His Leu Asp Val Thr
                885                 890                 895
```

[0036]

```
His Gln Phe Lys Ser Ser Thr Lys Lys Leu Val Lys Leu Ser Ile Lys
            900                 905                 910

Pro Tyr Arg Phe Asp Val Tyr Leu Thr Asp Lys Gly Tyr Lys Phe Ile
        915                 920                 925

Thr Ile Ser Tyr Leu Asp Val Leu Lys Lys Asp Asn Tyr Tyr Tyr Ile
    930                 935                 940

Pro Glu Gln Lys Tyr Asp Lys Leu Lys Leu Gly Lys Ala Ile Asp Lys
945                 950                 955                 960

Asn Ala Lys Phe Ile Ala Ser Phe Tyr Lys Asn Asp Leu Ile Lys Leu
                965                 970                 975

Asp Gly Glu Ile Tyr Lys Ile Ile Gly Val Asn Ser Asp Thr Arg Asn
            980                 985                 990

Met Ile Glu Leu Asp Leu Pro Asp  Ile Arg Tyr Lys Glu  Tyr Cys Glu
        995                 1000                1005

Leu Asn  Asn Ile Lys Gly Glu  Pro Arg Ile Lys Lys  Thr Ile Gly
    1010                1015                1020
```

Lys Lys Val Asn Ser Ile Glu Lys Leu Thr Thr Asp Val Leu Gly
1025 1030 1035

Asn Val Phe Thr Asn Thr Gln Tyr Thr Lys Pro Gln Leu Leu Phe
1040 1045 1050

Lys Arg Gly Asn
1055

<210> 9
<211> 1057
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> E2-K741N-L743N的多肽序列

<400> 9

Met Asn Gln Lys Phe Ile Leu Gly Leu Asp Ile Gly Ile Thr Ser Val
1 5 10 15

[0037]

Gly Tyr Gly Leu Ile Asp Tyr Glu Thr Lys Asn Ile Ile Asp Ala Gly
20 25 30

Val Arg Leu Phe Pro Glu Ala Asn Val Glu Asn Asn Glu Gly Arg Arg
35 40 45

Ser Lys Arg Gly Ser Arg Arg Leu Lys Arg Arg Arg Ile His Arg Leu
50 55 60

Asp Arg Val Lys His Leu Leu Ala Glu Tyr Asp Leu Leu Asp Leu Thr
65 70 75 80

Asn Ile Pro Lys Ser Thr Asn Pro Tyr Gln Thr Arg Val Lys Gly Leu
85 90 95

Asn Glu Lys Leu Ser Lys Asp Glu Leu Val Ile Ala Leu Leu His Ile
100 105 110

Ala Lys Arg Arg Gly Ile His Asn Val Asp Val Ala Ala Asp Lys Glu
115 120 125

Glu Thr Ala Ser Asp Ser Leu Ser Thr Lys Asp Gln Ile Asn Lys Asn
130 135 140

```
Ala Lys Phe Leu Glu Ser Arg Tyr Val Cys Glu Leu Gln Lys Glu Arg
145                 150                 155                 160

Leu Glu Asn Glu Gly His Val Arg Gly Val Glu Asn Arg Phe Leu Thr
                165                 170                 175

Lys Asp Ile Val Arg Glu Ala Lys Lys Ile Ile Asp Thr Gln Met Gln
            180                 185                 190

Tyr Tyr Pro Glu Ile Asp Glu Thr Phe Lys Glu Lys Tyr Ile Ser Leu
        195                 200                 205

Val Glu Thr Arg Arg Glu Tyr Phe Glu Gly Pro Gly Lys Gly Ser Pro
    210                 215                 220

Phe Gly Trp Glu Gly Asn Ile Lys Lys Trp Phe Glu Gln Met Met Gly
225                 230                 235                 240

His Cys Thr Tyr Phe Pro Glu Glu Leu Arg Ser Val Lys Tyr Ser Tyr
                245                 250                 255

Ser Ala Glu Leu Phe Asn Ala Leu Asn Asp Leu Asn Asn Leu Val Ile
            260                 265                 270
```

[0038]

```
Thr Arg Asp Glu Asp Ala Lys Leu Asn Tyr Gly Glu Lys Phe Gln Ile
        275                 280                 285

Ile Glu Asn Val Phe Lys Gln Lys Lys Thr Pro Asn Leu Lys Gln Ile
    290                 295                 300

Ala Ile Glu Ile Gly Val His Glu Thr Glu Ile Lys Gly Tyr Arg Val
305                 310                 315                 320

Asn Lys Ser Gly Thr Pro Glu Phe Thr Glu Phe Lys Leu Tyr His Asp
                325                 330                 335

Leu Lys Ser Ile Val Phe Asp Lys Ser Ile Leu Glu Asn Glu Ala Ile
            340                 345                 350

Leu Asp Gln Ile Ala Glu Ile Leu Thr Ile Tyr Gln Asp Glu Gln Ser
        355                 360                 365

Ile Lys Glu Glu Leu Asn Lys Leu Pro Glu Ile Leu Asn Glu Gln Asp
    370                 375                 380

Lys Ala Glu Ile Ala Lys Leu Ile Gly Tyr Asn Gly Thr His Arg Leu
385                 390                 395                 400
```

Ser Leu Lys Cys Ile His Leu Ile Asn Glu Glu Leu Trp Gln Thr Ser
405 410 415

Arg Asn Gln Met Glu Ile Phe Asn Tyr Leu Asn Ile Lys Pro Asn Lys
420 425 430

Val Asp Leu Ser Glu Gln Asn Lys Ile Pro Lys Asp Met Val Asn Asp
435 440 445

Phe Ile Leu Ser Pro Val Val Lys Arg Thr Phe Ile Gln Ser Ile Asn
450 455 460

Val Ile Asn Lys Val Ile Glu Lys Tyr Gly Ile Pro Glu Asp Ile Ile
465 470 475 480

Ile Glu Leu Ala Arg Glu Asn Asn Ser Asp Asp Arg Lys Lys Phe Ile
485 490 495

Asn Asn Leu Gln Lys Lys Asn Glu Ala Thr Arg Lys Arg Ile Asn Glu
500 505 510

Ile Ile Gly Gln Thr Gly Asn Gln Asn Ala Lys Arg Ile Val Glu Lys
515 520 525

[0039]

Ile Arg Leu His Asp Gln Gln Glu Gly Lys Cys Leu Tyr Ser Leu Lys
530 535 540

Asp Ile Pro Leu Glu Asp Leu Leu Arg Asn Pro Asn Asn Tyr Asp Ile
545 550 555 560

Asp His Ile Ile Pro Arg Ser Val Ser Phe Asp Asp Ser Met His Asn
565 570 575

Lys Val Leu Val Arg Arg Glu Gln Asn Ala Lys Lys Asn Asn Gln Thr
580 585 590

Pro Tyr Gln Tyr Leu Thr Ser Gly Tyr Ala Asp Ile Lys Tyr Ser Val
595 600 605

Phe Lys Gln His Val Leu Asn Leu Ala Glu Asn Lys Asp Arg Met Thr
610 615 620

Lys Lys Lys Arg Glu Tyr Leu Leu Glu Glu Arg Asp Ile Asn Lys Phe
625 630 635 640

Glu Val Gln Lys Glu Phe Ile Asn Arg Asn Leu Val Asp Thr Arg Tyr
645 650 655

Ala Thr Arg Glu Leu Thr Asn Tyr Leu Lys Ala Tyr Phe Ser Ala Asn
660 665 670

Asn Met Asn Val Lys Val Lys Thr Ile Asn Gly Ser Phe Thr Asp Tyr
675 680 685

Leu Arg Lys Val Trp Lys Phe Lys Lys Glu Arg Asn His Gly Tyr Lys
690 695 700

His His Ala Glu Asp Ala Leu Ile Ile Ala Asn Ala Asp Phe Leu Phe
705 710 715 720

Lys Glu Asn Lys Lys Leu Lys Ala Val Asn Ser Val Leu Glu Lys Pro
725 730 735

Glu Ile Glu Thr Asn Gln Asn Asp Ile Gln Val Asp Ser Glu Asp Asn
740 745 750

Tyr Ser Glu Met Phe Ile Ile Pro Lys Gln Val Gln Asp Ile Lys Asp
755 760 765

Phe Arg Asn Phe Lys Tyr Ser His Arg Val Asp Lys Lys Pro Asn Arg
770 775 780

[0040]

Gln Leu Ile Asn Asp Thr Leu Tyr Ser Thr Arg Lys Lys Asp Asn Ser
785 790 795 800

Thr Tyr Ile Val Gln Thr Ile Lys Asp Ile Tyr Ala Lys Asp Asn Thr
805 810 815

Thr Leu Lys Lys Gln Phe Asp Lys Ser Pro Glu Lys Phe Leu Met Tyr
820 825 830

Gln His Asp Pro Arg Thr Phe Glu Lys Leu Glu Val Ile Met Lys Gln
835 840 845

Tyr Ala Asn Glu Lys Asn Pro Leu Ala Lys Tyr His Glu Glu Thr Gly
850 855 860

Glu Tyr Leu Thr Lys Tyr Ser Lys Lys Asn Asn Gly Pro Ile Val Lys
865 870 875 880

Ser Leu Lys Tyr Ile Gly Asn Lys Leu Gly Ser His Leu Asp Val Thr
885 890 895

His Gln Phe Lys Ser Ser Thr Lys Lys Leu Val Lys Leu Ser Ile Lys
900 905 910

Pro Tyr Arg Phe Asp Val Tyr Leu Thr Asp Lys Gly Tyr Lys Phe Ile
915 920 925

Thr Ile Ser Tyr Leu Asp Val Leu Lys Lys Asp Asn Tyr Tyr Tyr Ile
930 935 940

Pro Glu Gln Lys Tyr Asp Lys Leu Lys Leu Gly Lys Ala Ile Asp Lys
945 950 955 960

Asn Ala Lys Phe Ile Ala Ser Phe Tyr Lys Asn Asp Leu Ile Lys Leu
965 970 975

Asp Gly Glu Ile Tyr Lys Ile Ile Gly Val Asn Ser Asp Thr Arg Asn
980 985 990

Met Ile Glu Leu Asp Leu Pro Asp Ile Arg Tyr Lys Glu Tyr Cys Glu
995 1000 1005

Leu Asn Asn Ile Lys Gly Glu Pro Arg Ile Lys Lys Thr Ile Gly
1010 1015 1020

Lys Lys Val Asn Ser Ile Glu Lys Leu Thr Thr Asp Val Leu Gly
1025 1030 1035

[0041]

Asn Val Phe Thr Asn Thr Gln Tyr Thr Lys Pro Gln Leu Leu Phe
1040 1045 1050

Lys Arg Gly Asn
1055

<210> 10
<211> 1053
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> F8的多肽序列

<400> 10

Met Asn Gln Lys Phe Ile Leu Gly Leu Asp Ile Gly Ile Thr Ser Val
1 5 10 15

Gly Tyr Gly Leu Ile Asp Tyr Glu Thr Lys Asn Ile Ile Asp Ala Gly
20 25 30

Val Arg Leu Phe Pro Glu Ala Asn Val Glu Asn Asn Glu Gly Arg Arg

35 40 45

Ser Lys Arg Gly Ser Arg Arg Leu Lys Arg Arg Arg Ile His Arg Leu
50 55 60

Glu Arg Val Lys Ser Leu Leu Ser Glu Tyr Lys Ile Ile Ser Gly Leu
65 70 75 80

Ala Pro Thr Asn Asn Gln Pro Tyr Asn Ile Arg Val Lys Gly Leu Thr
85 90 95

Glu Gln Leu Thr Lys Asp Glu Leu Ala Val Ala Leu Leu His Ile Ala
100 105 110

Lys Arg Arg Gly Ile His Lys Ile Asp Val Ile Asp Ser Asn Asp Asp
115 120 125

Val Gly Asn Glu Leu Ser Thr Lys Glu Gln Leu Asn Lys Asn Ser Lys
130 135 140

Leu Leu Lys Asp Lys Phe Val Cys Gln Ile Gln Leu Glu Arg Met Asn
145 150 155 160

[0042] Glu Gly Gln Val Arg Gly Glu Lys Asn Arg Phe Lys Thr Ala Asp Ile
165 170 175

Ile Lys Glu Ile Ile Gln Leu Leu Asn Val Gln Lys Asn Phe His Gln
180 185 190

Leu Asp Glu Asn Phe Ile Asn Lys Tyr Ile Glu Leu Val Glu Met Arg
195 200 205

Arg Glu Tyr Phe Glu Gly Pro Gly Gln Gly Ser Pro Phe Gly Trp Asn
210 215 220

Gly Asp Leu Lys Lys Trp Tyr Glu Met Leu Met Gly His Cys Thr Tyr
225 230 235 240

Phe Pro Gln Glu Leu Arg Ser Val Lys Tyr Ala Tyr Ser Ala Asp Leu
245 250 255

Phe Asn Ala Leu Asn Asp Leu Asn Asn Leu Ile Ile Gln Arg Asp Asn
260 265 270

Ser Glu Lys Leu Glu Tyr His Glu Lys Tyr His Ile Ile Glu Asn Val
275 280 285

Phe Lys Gln Lys Lys Lys Pro Thr Leu Lys Gln Ile Ala Lys Glu Ile

290 295 300

Gly Val Asn Pro Glu Asp Ile Lys Gly Tyr Arg Ile Thr Lys Ser Gly
305 310 315 320

Thr Pro Glu Phe Thr Glu Phe Lys Leu Tyr His Asp Leu Lys Ser Val
325 330 335

Leu Phe Asp Gln Ser Ile Leu Glu Asn Glu Asp Val Leu Asp Gln Ile
340 345 350

Ala Glu Ile Leu Thr Ile Tyr Gln Asp Lys Asp Ser Ile Lys Ser Lys
355 360 365

Leu Thr Glu Leu Asp Ile Leu Leu Asn Glu Glu Asp Lys Glu Asn Ile
370 375 380

Ala Gln Leu Thr Gly Tyr Asn Gly Thr His Arg Leu Ser Leu Lys Cys
385 390 395 400

Ile Arg Leu Val Leu Glu Glu Gln Trp Tyr Ser Ser Arg Asn Gln Met
405 410 415

[0043] Glu Ile Phe Thr His Leu Asn Ile Lys Pro Lys Lys Ile Asn Leu Thr
420 425 430

Ala Ala Asn Lys Ile Pro Lys Ala Met Ile Asp Glu Phe Ile Leu Ser
435 440 445

Pro Val Val Lys Arg Thr Phe Ile Gln Ser Ile Asn Val Ile Asn Lys
450 455 460

Val Ile Glu Lys Tyr Gly Ile Pro Glu Asp Ile Ile Ile Glu Leu Ala
465 470 475 480

Arg Glu Asn Asn Ser Asp Asp Arg Lys Lys Phe Ile Asn Asn Leu Gln
485 490 495

Lys Lys Asn Glu Ala Thr Arg Lys Arg Ile Asn Glu Ile Ile Gly Gln
500 505 510

Thr Gly Asn Gln Asn Ala Lys Arg Ile Val Glu Lys Ile Arg Leu His
515 520 525

Asp Gln Gln Glu Gly Lys Cys Leu Tyr Ser Leu Glu Ser Ile Ala Leu
530 535 540

Met Asp Leu Leu Asn Asn Pro Gln Asn Tyr Glu Val Asp His Ile Ile

545 550 555 560

Pro Arg Ser Val Ala Phe Asp Asn Ser Ile His Asn Lys Val Leu Val
565 570 575

Lys Gln Ile Glu Asn Ser Lys Lys Gly Asn Arg Thr Pro Tyr Gln Tyr
580 585 590

Leu Asn Ser Ser Asp Ala Lys Leu Ser Tyr Asn Gln Phe Lys Gln His
595 600 605

Ile Leu Asn Leu Ser Lys Ser Lys Asp Arg Ile Ser Lys Lys Lys Lys
610 615 620

Asp Tyr Leu Leu Glu Glu Arg Asp Ile Asn Lys Phe Glu Val Gln Lys
625 630 635 640

Glu Phe Ile Asn Arg Asn Leu Val Asp Thr Arg Tyr Ala Thr Arg Glu
645 650 655

Leu Thr Ser Tyr Leu Lys Ala Tyr Phe Ser Ala Asn Asn Met Asp Val
660 665 670

[0044] Lys Val Lys Thr Ile Asn Gly Ser Phe Thr Asn His Leu Arg Lys Val
675 680 685

Trp Arg Phe Asp Lys Tyr Arg Asn His Gly Tyr Lys His His Ala Glu
690 695 700

Asp Ala Leu Ile Ile Ala Asn Ala Asp Phe Leu Phe Lys Glu Asn Lys
705 710 715 720

Lys Leu Lys Ala Val Asn Ser Val Leu Glu Lys Pro Glu Ile Glu Thr
725 730 735

Lys Gln Leu Asp Ile Gln Val Asp Ser Glu Asp Asn Tyr Ser Glu Met
740 745 750

Phe Ile Ile Pro Lys Gln Val Gln Asp Ile Lys Asp Phe Arg Asn Phe
755 760 765

Lys Tyr Ser His Arg Val Asp Lys Lys Pro Asn Arg Gln Leu Ile Asn
770 775 780

Asp Thr Leu Tyr Ser Thr Arg Lys Lys Asp Asn Ser Thr Tyr Ile Val
785 790 795 800

Gln Thr Ile Lys Asp Ile Tyr Ala Lys Asp Asn Thr Thr Leu Lys Lys

805 810 815

Gln Phe Asp Lys Ser Pro Glu Lys Phe Leu Met Tyr Gln His Asp Pro
820 825 830

Arg Thr Phe Glu Lys Leu Glu Val Ile Met Lys Gln Tyr Ala Asn Glu
835 840 845

Lys Asn Pro Leu Ala Lys Tyr His Glu Glu Thr Gly Glu Tyr Leu Thr
850 855 860

Lys Tyr Ser Lys Lys Asn Asn Gly Pro Ile Val Lys Ser Leu Lys Tyr
865 870 875 880

Ile Gly Asn Lys Leu Gly Ser His Leu Asp Val Thr His Gln Phe Lys
885 890 895

Ser Ser Thr Lys Lys Leu Val Lys Leu Ser Ile Lys Pro Tyr Arg Phe
900 905 910

Asp Val Tyr Leu Thr Asp Lys Gly Tyr Lys Phe Ile Thr Ile Ser Tyr
915 920 925

[0045] Leu Asp Val Leu Lys Lys Asp Asn Tyr Tyr Tyr Ile Pro Glu Gln Lys
930 935 940

Tyr Asp Lys Leu Lys Leu Gly Lys Ala Ile Asp Lys Asn Ala Lys Phe
945 950 955 960

Ile Ala Ser Phe Tyr Lys Asn Asp Leu Ile Lys Leu Asp Gly Glu Ile
965 970 975

Tyr Lys Ile Ile Gly Val Asn Ser Asp Thr Arg Asn Met Ile Glu Leu
980 985 990

Asp Leu Pro Asp Ile Arg Tyr Lys Glu Tyr Cys Glu Leu Asn Asn Ile
995 1000 1005

Lys Gly Glu Pro Arg Ile Lys Lys Thr Ile Gly Lys Lys Val Asn
1010 1015 1020

Ser Ile Glu Lys Leu Thr Thr Asp Val Leu Gly Asn Val Phe Thr
1025 1030 1035

Asn Thr Gln Tyr Thr Lys Pro Gln Leu Leu Phe Lys Arg Gly Asn
1040 1045 1050

<210> 11

<211> 1053
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> F8-K737D-L739K的多肽序列

<400> 11

[0046]

```
Met Asn Gln Lys Phe Ile Leu Gly Leu Asp Ile Gly Ile Thr Ser Val
1               5                  10                  15

Gly Tyr Gly Leu Ile Asp Tyr Glu Thr Lys Asn Ile Ile Asp Ala Gly
            20                  25                  30

Val Arg Leu Phe Pro Glu Ala Asn Val Glu Asn Asn Glu Gly Arg Arg
        35                  40                  45

Ser Lys Arg Gly Ser Arg Arg Leu Lys Arg Arg Arg Ile His Arg Leu
    50                  55                  60

Glu Arg Val Lys Ser Leu Leu Ser Glu Tyr Lys Ile Ile Ser Gly Leu
65                  70                  75                  80

Ala Pro Thr Asn Asn Gln Pro Tyr Asn Ile Arg Val Lys Gly Leu Thr
                85                  90                  95

Glu Gln Leu Thr Lys Asp Glu Leu Ala Val Ala Leu Leu His Ile Ala
            100                 105                 110

Lys Arg Arg Gly Ile His Lys Ile Asp Val Ile Asp Ser Asn Asp Asp
        115                 120                 125

Val Gly Asn Glu Leu Ser Thr Lys Glu Gln Leu Asn Lys Asn Ser Lys
    130                 135                 140

Leu Leu Lys Asp Lys Phe Val Cys Gln Ile Gln Leu Glu Arg Met Asn
145                 150                 155                 160

Glu Gly Gln Val Arg Gly Glu Lys Asn Arg Phe Lys Thr Ala Asp Ile
                165                 170                 175

Ile Lys Glu Ile Ile Gln Leu Leu Asn Val Gln Lys Asn Phe His Gln
            180                 185                 190

Leu Asp Glu Asn Phe Ile Asn Lys Tyr Ile Glu Leu Val Glu Met Arg
        195                 200                 205
```

```
Arg Glu Tyr Phe Glu Gly Pro Gly Gln Gly Ser Pro Phe Gly Trp Asn
    210                 215                 220

Gly Asp Leu Lys Lys Trp Tyr Glu Met Leu Met Gly His Cys Thr Tyr
225                 230                 235                 240

Phe Pro Gln Glu Leu Arg Ser Val Lys Tyr Ala Tyr Ser Ala Asp Leu
                245                 250                 255

Phe Asn Ala Leu Asn Asp Leu Asn Asn Leu Ile Ile Gln Arg Asp Asn
            260                 265                 270

Ser Glu Lys Leu Glu Tyr His Glu Lys Tyr His Ile Ile Glu Asn Val
        275                 280                 285

Phe Lys Gln Lys Lys Lys Pro Thr Leu Lys Gln Ile Ala Lys Glu Ile
    290                 295                 300

Gly Val Asn Pro Glu Asp Ile Lys Gly Tyr Arg Ile Thr Lys Ser Gly
305                 310                 315                 320

Thr Pro Glu Phe Thr Glu Phe Lys Leu Tyr His Asp Leu Lys Ser Val
                325                 330                 335
```

[0047]

```
Leu Phe Asp Gln Ser Ile Leu Glu Asn Glu Asp Val Leu Asp Gln Ile
            340                 345                 350

Ala Glu Ile Leu Thr Ile Tyr Gln Asp Lys Asp Ser Ile Lys Ser Lys
        355                 360                 365

Leu Thr Glu Leu Asp Ile Leu Leu Asn Glu Glu Asp Lys Glu Asn Ile
    370                 375                 380

Ala Gln Leu Thr Gly Tyr Asn Gly Thr His Arg Leu Ser Leu Lys Cys
385                 390                 395                 400

Ile Arg Leu Val Leu Glu Glu Gln Trp Tyr Ser Ser Arg Asn Gln Met
                405                 410                 415

Glu Ile Phe Thr His Leu Asn Ile Lys Pro Lys Lys Ile Asn Leu Thr
            420                 425                 430

Ala Ala Asn Lys Ile Pro Lys Ala Met Ile Asp Glu Phe Ile Leu Ser
        435                 440                 445

Pro Val Val Lys Arg Thr Phe Ile Gln Ser Ile Asn Val Ile Asn Lys
    450                 455                 460
```

```
Val Ile Glu Lys Tyr Gly Ile Pro Glu Asp Ile Ile Ile Glu Leu Ala
465                 470                 475                 480

Arg Glu Asn Asn Ser Asp Asp Arg Lys Lys Phe Ile Asn Asn Leu Gln
                485                 490                 495

Lys Lys Asn Glu Ala Thr Arg Lys Arg Ile Asn Glu Ile Ile Gly Gln
            500                 505                 510

Thr Gly Asn Gln Asn Ala Lys Arg Ile Val Glu Lys Ile Arg Leu His
        515                 520                 525

Asp Gln Gln Glu Gly Lys Cys Leu Tyr Ser Leu Glu Ser Ile Ala Leu
    530                 535                 540

Met Asp Leu Leu Asn Asn Pro Gln Asn Tyr Glu Val Asp His Ile Ile
545                 550                 555                 560

Pro Arg Ser Val Ala Phe Asp Asn Ser Ile His Asn Lys Val Leu Val
                565                 570                 575

Lys Gln Ile Glu Asn Ser Lys Lys Gly Asn Arg Thr Pro Tyr Gln Tyr
            580                 585                 590
```

[0048]

```
Leu Asn Ser Ser Asp Ala Lys Leu Ser Tyr Asn Gln Phe Lys Gln His
        595                 600                 605

Ile Leu Asn Leu Ser Lys Ser Lys Asp Arg Ile Ser Lys Lys Lys Lys
    610                 615                 620

Asp Tyr Leu Leu Glu Glu Arg Asp Ile Asn Lys Phe Glu Val Gln Lys
625                 630                 635                 640

Glu Phe Ile Asn Arg Asn Leu Val Asp Thr Arg Tyr Ala Thr Arg Glu
                645                 650                 655

Leu Thr Ser Tyr Leu Lys Ala Tyr Phe Ser Ala Asn Asn Met Asp Val
            660                 665                 670

Lys Val Lys Thr Ile Asn Gly Ser Phe Thr Asn His Leu Arg Lys Val
        675                 680                 685

Trp Arg Phe Asp Lys Tyr Arg Asn His Gly Tyr Lys His His Ala Glu
    690                 695                 700

Asp Ala Leu Ile Ile Ala Asn Ala Asp Phe Leu Phe Lys Glu Asn Lys
705                 710                 715                 720
```

Lys Leu Lys Ala Val Asn Ser Val Leu Glu Lys Pro Glu Ile Glu Thr
725 730 735

Asp Gln Lys Asp Ile Gln Val Asp Ser Glu Asp Asn Tyr Ser Glu Met
740 745 750

Phe Ile Ile Pro Lys Gln Val Gln Asp Ile Lys Asp Phe Arg Asn Phe
755 760 765

Lys Tyr Ser His Arg Val Asp Lys Lys Pro Asn Arg Gln Leu Ile Asn
770 775 780

Asp Thr Leu Tyr Ser Thr Arg Lys Lys Asp Asn Ser Thr Tyr Ile Val
785 790 795 800

Gln Thr Ile Lys Asp Ile Tyr Ala Lys Asp Asn Thr Thr Leu Lys Lys
805 810 815

Gln Phe Asp Lys Ser Pro Glu Lys Phe Leu Met Tyr Gln His Asp Pro
820 825 830

Arg Thr Phe Glu Lys Leu Glu Val Ile Met Lys Gln Tyr Ala Asn Glu
835 840 845

[0049]

Lys Asn Pro Leu Ala Lys Tyr His Glu Glu Thr Gly Glu Tyr Leu Thr
850 855 860

Lys Tyr Ser Lys Lys Asn Asn Gly Pro Ile Val Lys Ser Leu Lys Tyr
865 870 875 880

Ile Gly Asn Lys Leu Gly Ser His Leu Asp Val Thr His Gln Phe Lys
885 890 895

Ser Ser Thr Lys Lys Leu Val Lys Leu Ser Ile Lys Pro Tyr Arg Phe
900 905 910

Asp Val Tyr Leu Thr Asp Lys Gly Tyr Lys Phe Ile Thr Ile Ser Tyr
915 920 925

Leu Asp Val Leu Lys Lys Asp Asn Tyr Tyr Tyr Ile Pro Glu Gln Lys
930 935 940

Tyr Asp Lys Leu Lys Leu Gly Lys Ala Ile Asp Lys Asn Ala Lys Phe
945 950 955 960

Ile Ala Ser Phe Tyr Lys Asn Asp Leu Ile Lys Leu Asp Gly Glu Ile
965 970 975

Tyr Lys Ile Ile Gly Val Asn Ser Asp Thr Arg Asn Met Ile Glu Leu
980 985 990

Asp Leu Pro Asp Ile Arg Tyr Lys Glu Tyr Cys Glu Leu Asn Asn Ile
995 1000 1005

Lys Gly Glu Pro Arg Ile Lys Lys Thr Ile Gly Lys Lys Val Asn
1010 1015 1020

Ser Ile Glu Lys Leu Thr Thr Asp Val Leu Gly Asn Val Phe Thr
1025 1030 1035

Asn Thr Gln Tyr Thr Lys Pro Gln Leu Leu Phe Lys Arg Gly Asn
1040 1045 1050

<210> 12
<211> 1053
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
[0050] <221> misc_feature
<223> F8-K737N-L739N的多肽序列

<400> 12

Met Asn Gln Lys Phe Ile Leu Gly Leu Asp Ile Gly Ile Thr Ser Val
1 5 10 15

Gly Tyr Gly Leu Ile Asp Tyr Glu Thr Lys Asn Ile Ile Asp Ala Gly
20 25 30

Val Arg Leu Phe Pro Glu Ala Asn Val Glu Asn Asn Glu Gly Arg Arg
35 40 45

Ser Lys Arg Gly Ser Arg Arg Leu Lys Arg Arg Arg Ile His Arg Leu
50 55 60

Glu Arg Val Lys Ser Leu Leu Ser Glu Tyr Lys Ile Ile Ser Gly Leu
65 70 75 80

Ala Pro Thr Asn Asn Gln Pro Tyr Asn Ile Arg Val Lys Gly Leu Thr
85 90 95

Glu Gln Leu Thr Lys Asp Glu Leu Ala Val Ala Leu Leu His Ile Ala
100 105 110

```
Lys Arg Arg Gly Ile His Lys Ile Asp Val Ile Asp Ser Asn Asp Asp
        115                 120                 125

Val Gly Asn Glu Leu Ser Thr Lys Glu Gln Leu Asn Lys Asn Ser Lys
    130                 135                 140

Leu Leu Lys Asp Lys Phe Val Cys Gln Ile Gln Leu Glu Arg Met Asn
145                 150                 155                 160

Glu Gly Gln Val Arg Gly Glu Lys Asn Arg Phe Lys Thr Ala Asp Ile
                165                 170                 175

Ile Lys Glu Ile Ile Gln Leu Leu Asn Val Gln Lys Asn Phe His Gln
            180                 185                 190

Leu Asp Glu Asn Phe Ile Asn Lys Tyr Ile Glu Leu Val Glu Met Arg
        195                 200                 205

Arg Glu Tyr Phe Glu Gly Pro Gly Gln Gly Ser Pro Phe Gly Trp Asn
    210                 215                 220

Gly Asp Leu Lys Lys Trp Tyr Glu Met Leu Met Gly His Cys Thr Tyr
225                 230                 235                 240
```

[0051]

```
Phe Pro Gln Glu Leu Arg Ser Val Lys Tyr Ala Tyr Ser Ala Asp Leu
                245                 250                 255

Phe Asn Ala Leu Asn Asp Leu Asn Asn Leu Ile Ile Gln Arg Asp Asn
            260                 265                 270

Ser Glu Lys Leu Glu Tyr His Glu Lys Tyr His Ile Ile Glu Asn Val
        275                 280                 285

Phe Lys Gln Lys Lys Lys Pro Thr Leu Lys Gln Ile Ala Lys Glu Ile
    290                 295                 300

Gly Val Asn Pro Glu Asp Ile Lys Gly Tyr Arg Ile Thr Lys Ser Gly
305                 310                 315                 320

Thr Pro Glu Phe Thr Glu Phe Lys Leu Tyr His Asp Leu Lys Ser Val
                325                 330                 335

Leu Phe Asp Gln Ser Ile Leu Glu Asn Glu Asp Val Leu Asp Gln Ile
            340                 345                 350

Ala Glu Ile Leu Thr Ile Tyr Gln Asp Lys Asp Ser Ile Lys Ser Lys
        355                 360                 365
```

```
Leu Thr Glu Leu Asp Ile Leu Leu Asn Glu Glu Asp Lys Glu Asn Ile
    370                 375                 380

Ala Gln Leu Thr Gly Tyr Asn Gly Thr His Arg Leu Ser Leu Lys Cys
385                 390                 395                 400

Ile Arg Leu Val Leu Glu Glu Gln Trp Tyr Ser Ser Arg Asn Gln Met
                405                 410                 415

Glu Ile Phe Thr His Leu Asn Ile Lys Pro Lys Lys Ile Asn Leu Thr
            420                 425                 430

Ala Ala Asn Lys Ile Pro Lys Ala Met Ile Asp Glu Phe Ile Leu Ser
        435                 440                 445

Pro Val Val Lys Arg Thr Phe Ile Gln Ser Ile Asn Val Ile Asn Lys
    450                 455                 460

Val Ile Glu Lys Tyr Gly Ile Pro Glu Asp Ile Ile Ile Glu Leu Ala
465                 470                 475                 480

Arg Glu Asn Asn Ser Asp Asp Arg Lys Lys Phe Ile Asn Asn Leu Gln
                485                 490                 495
```

[0052]

```
Lys Lys Asn Glu Ala Thr Arg Lys Arg Ile Asn Glu Ile Ile Gly Gln
            500                 505                 510

Thr Gly Asn Gln Asn Ala Lys Arg Ile Val Glu Lys Ile Arg Leu His
        515                 520                 525

Asp Gln Gln Glu Gly Lys Cys Leu Tyr Ser Leu Glu Ser Ile Ala Leu
    530                 535                 540

Met Asp Leu Leu Asn Asn Pro Gln Asn Tyr Glu Val Asp His Ile Ile
545                 550                 555                 560

Pro Arg Ser Val Ala Phe Asp Asn Ser Ile His Asn Lys Val Leu Val
                565                 570                 575

Lys Gln Ile Glu Asn Ser Lys Lys Gly Asn Arg Thr Pro Tyr Gln Tyr
            580                 585                 590

Leu Asn Ser Ser Asp Ala Lys Leu Ser Tyr Asn Gln Phe Lys Gln His
        595                 600                 605

Ile Leu Asn Leu Ser Lys Ser Lys Asp Arg Ile Ser Lys Lys Lys Lys
    610                 615                 620
```

```
Asp Tyr Leu Leu Glu Glu Arg Asp Ile Asn Lys Phe Glu Val Gln Lys
625                 630                 635                 640

Glu Phe Ile Asn Arg Asn Leu Val Asp Thr Arg Tyr Ala Thr Arg Glu
                645                 650                 655

Leu Thr Ser Tyr Leu Lys Ala Tyr Phe Ser Ala Asn Asn Met Asp Val
            660                 665                 670

Lys Val Lys Thr Ile Asn Gly Ser Phe Thr Asn His Leu Arg Lys Val
        675                 680                 685

Trp Arg Phe Asp Lys Tyr Arg Asn His Gly Tyr Lys His His Ala Glu
    690                 695                 700

Asp Ala Leu Ile Ile Ala Asn Ala Asp Phe Leu Phe Lys Glu Asn Lys
705                 710                 715                 720

Lys Leu Lys Ala Val Asn Ser Val Leu Glu Lys Pro Glu Ile Glu Thr
                725                 730                 735

Asn Gln Asn Asp Ile Gln Val Asp Ser Glu Asp Asn Tyr Ser Glu Met
            740                 745                 750
```

[0053]

```
Phe Ile Ile Pro Lys Gln Val Gln Asp Ile Lys Asp Phe Arg Asn Phe
        755                 760                 765

Lys Tyr Ser His Arg Val Asp Lys Lys Pro Asn Arg Gln Leu Ile Asn
    770                 775                 780

Asp Thr Leu Tyr Ser Thr Arg Lys Lys Asp Asn Ser Thr Tyr Ile Val
785                 790                 795                 800

Gln Thr Ile Lys Asp Ile Tyr Ala Lys Asp Asn Thr Thr Leu Lys Lys
                805                 810                 815

Gln Phe Asp Lys Ser Pro Glu Lys Phe Leu Met Tyr Gln His Asp Pro
            820                 825                 830

Arg Thr Phe Glu Lys Leu Glu Val Ile Met Lys Gln Tyr Ala Asn Glu
        835                 840                 845

Lys Asn Pro Leu Ala Lys Tyr His Glu Glu Thr Gly Glu Tyr Leu Thr
    850                 855                 860

Lys Tyr Ser Lys Lys Asn Asn Gly Pro Ile Val Lys Ser Leu Lys Tyr
865                 870                 875                 880
```

```
Ile Gly Asn Lys Leu Gly Ser His Leu Asp Val Thr His Gln Phe Lys
                885                 890                 895

Ser Ser Thr Lys Lys Leu Val Lys Leu Ser Ile Lys Pro Tyr Arg Phe
            900                 905                 910

Asp Val Tyr Leu Thr Asp Lys Gly Tyr Lys Phe Ile Thr Ile Ser Tyr
        915                 920                 925

Leu Asp Val Leu Lys Lys Asp Asn Tyr Tyr Tyr Ile Pro Glu Gln Lys
    930                 935                 940

Tyr Asp Lys Leu Lys Leu Gly Lys Ala Ile Asp Lys Asn Ala Lys Phe
945                 950                 955                 960

Ile Ala Ser Phe Tyr Lys Asn Asp Leu Ile Lys Leu Asp Gly Glu Ile
                965                 970                 975

Tyr Lys Ile Ile Gly Val Asn Ser Asp Thr Arg Asn Met Ile Glu Leu
            980                 985                 990

Asp Leu Pro Asp Ile Arg Tyr Lys  Glu Tyr Cys Glu Leu  Asn Asn Ile
        995                 1000                1005
```

[0054]

```
Lys Gly  Glu Pro Arg Ile Lys  Lys Thr Ile Gly Lys  Lys Val Asn
    1010                1015                1020

Ser Ile  Glu Lys Leu Thr Thr  Asp Val Leu Gly Asn  Val Phe Thr
    1025                1030                1035

Asn Thr  Gln Tyr Thr Lys Pro  Gln Leu Leu Phe Lys  Arg Gly Asn
    1040                1045                1050
```

<210> 13
<211> 3171
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> 编码Gib11的DNA序列

<400> 13

```
atgaaccaga aatttatcct gggtctggat attggtatta ccagcgttgg ttatggcctg      60

attgattacg aaaccaaaaa cattattgat gccggtgttc gtctgtttcc ggaagcaaat     120

gttgaaaata atgaaggtcg tcgtagcaaa cgtggtagcc gtcgtctgaa acgtcgtcgt     180
```

[0055]

attcatcgtc tggaacgtgt taaactgctg ctgaccgaat atgatctgat taacaaagag 240

cagattccga ccagcaataa cccgtatcag attcgtgtta aaggtctgag cgaaatcctg 300

agcaaagatg aactggcaat tgcactgctg catctggcaa aacgccgtgg cattcataat 360

gttgatgttg cagcagataa agaagaaacc gcaagcgata gcctgagcac caaagatcag 420

attaacaaaa acgccaaatt tctggaaagc cgctatgttt gtgaactgca gaaagaacgt 480

ctggaaaatg aaggtcatgt tcgtggtgtt gaaaatcgct ttctgacgaa agatattgtg 540

cgtgaggcca aaaaaatcat cgatacccag atgcagtatt acccggaaat tgatgaaacc 600

ttcaaagaga aatatatcag cctggttgaa acccgtcgcg aatattttga aggtcctggt 660

cagggtagcc cgtttggttg gaatggtgat ctgaaaaaat ggtacgaaat gctgatgggt 720

cactgtacct attttccgca agaactgcgt agcgttaaat atgcctatag cgcagacctg 780

tttaatgcac tgaatgatct gaacaacctg attattcagc gcgataatag cgagaaactg 840

gaataccatg agaagtatca catcatcgag aacgtgttca agcagaaaaa aaagccgacg 900

ctgaaacaaa tcgcaaaaga gattggcgtt aacccggaag atattaaagg ttatcgtatt 960

accaaaagcg gcacaccgga gtttacatcc tttaaactgt tccacgatct gaaaaaagtg 1020

gtgaaagatc atgccatcct ggatgatatt gatctgctga atcagattgc agaaatcctg 1080

accatctatc aggataaaga tagcattgtt gcagaactgg gtcagctgga atatctgatg 1140

agcgaagccg ataaacagag cattagcgaa ctgaccggtt ataccggtac acatagcctg 1200

tcactgaaat gcatgaacat gattatcgat gaactgtggc atagcagcat gaaccagatg 1260

gaagttttta cctatctgaa tatgcgtccg aaaaagtatg agctgaaagg ttatcagcgt 1320

attccgaccg atatgattga tgatgcaatt ctgagtccgg ttgtgaaacg caccttttatt 1380

cagagcatca acgtgatcaa caaagtgatc gagaaatatg gcatccccga agatatcatt 1440

atcgaactgg cacgtgaaaa taactccgat gatcgcaaaa agttcatcaa caacctgcag 1500

aaaaagaatg aagcaacccg caaacgcatt aacgaaatta ttggtcagac cggtaatcag 1560

aatgccaaac gtattgtgga aaaaatccgt ctgcatgatc agcaagaggg caaatgtctg 1620

tatagcctgg aaagcattcc tctggaagat ctgctgaaca atccgaatca ttatgaagtg 1680

gatcacatta ttccgcgtag cgtgagcttt gataattcct atcataataa agtgctggtg 1740

aaacagagcg aaaactccaa aaaatccaac ctgacaccgt atcagtattt caatagcggc 1800

aaatccaaac tgagctacaa ccagtttaaa cagcatattc tgaacctgag caaaagccag 1860

gatcgcatca gcaagaagaa gaaggagtac ctgctggaag aacgcgacat caacaaattt 1920

gaagtgcaga aagaatttat caaccgcaac ctggttgata cccgttatgc aacccgtgaa 1980

ctgaccaatt atctgaaagc atatttcagc gccaacaaca tgaacgtgaa agtgaaaacg 2040

attaacggca gctttaccga ttatctgcgt aaagtgtgga aattcaaaaa agaacgcaac 2100

cacggctata aacatcatgc cgaagatgcc ctgattattg caaatgcaga tttcctgttt 2160

aaagaaaaca aaaaactgaa agccgtcaac agcgtgctgg aaaaaccgga aattgagaca 2220

aaacagctgg acattcaggt tgatagcgaa gataattaca gcgaaatgtt tatcatcccg 2280

aaacaggtgc aggatatcaa agattttcgc aacttcaaat atagccaccg cgttgacaaa 2340

aaacctaatc gtcagctgat taacgatacc ctgtatagca cccgcaaaaa agataacagc 2400

acctatattg tgcagaccat taaagacatc tacgccaaag ataataccac cctgaaaaaa 2460

cagttcgaca aaagcccaga aaaatttctg atgtatcagc atgatccgcg taccttcgaa 2520

aaactggaag ttattatgaa acagtatgcc aacgagaaaa atccgctggc caaatatcac 2580

gaagaaaccg gtgaatatct gaccaaatat tccaagaaga acaacggtcc gatcgttaaa 2640

tccctgaaat atatcggtaa taaactgggc agccatctgg atgttaccca tcagtttaaa 2700

agctccacaa agaagctggt taaactgtcc atcaaaccgt atcgctttga tgtgtatctg 2760

accgacaaag gctataaatt cattaccatc agctatctgg acgtgctgaa aaaagacaac 2820

tattattata tcccggaaca gaaatatgat aaactgaaac tgggtaaagc catcgataaa 2880

aacgccaaat ttatcgccag cttctacaaa aacgacctga ttaaactgga tggcgagatc 2940

tataaaatca tcggtgttaa tagcgacacc cgcaatatga ttgagctgga tctgccggat 3000

[0056] attcgctata aagaatattg cgaactgaac aacattaaag gcgaaccgcg tatcaaaaag 3060

accatcggca aaaaagtgaa tagcatcgag aaactgacca ccgatgttct gggtaatgtg 3120

tttaccaata cccagtatac caaacctcag ctgctgttca aacgcggtaa t 3171

<210> 14
<211> 3171
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> 编码Gibl1Spa-1的DNA序列

<400> 14
atgaaccaga aatttatcct gggtctggat attggtatta ccagcgttgg ttatggcctg 60

attgattacg aaaccaaaaa cattattgat gccggtgttc gtctgtttcc ggaagcaaat 120

gttgaaaata atgaaggtcg tcgtagcaaa cgtggtagcc gtcgtctgaa acgtcgtcgt 180

attcatcgtc tggaacgtgt taaactgctg ctgaccgaat atgatctgat taacaaagag 240

cagattccga ccagcaataa cccgtatcag attcgtgtta aaggtctgag cgaaatcctg 300

agcaaagatg aactggcaat tgcactgctg catctggcaa aacgccgtgg cattcataat 360

[0057]

gttgatgttg cagcagataa agaagaaacc gcaagcgata gcctgagcac caaagatcag 420

attaacaaaa acgccaaatt tctggaaagc cgctatgttt gtgaactgca gaaagaacgt 480

ctggaaaatg aaggtcatgt tcgtggtgtt gaaaatcgct ttctgacgaa agatattgtg 540

cgtgaggcca aaaaaatcat cgatacccag atgcagtatt acccggaaat tgatgaaacc 600

ttcaaagaga aatatatcag cctggttgaa acccgtcgcg aatattttga aggtcctggt 660

cagggtagcc cgtttggttg gaatggtgat ctgaaaaaat ggtacgaaat gctgatgggt 720

cactgtacct attttccgca agaactgcgt agcgttaaat atgcctatag cgcagacctg 780

tttaatgcac tgaatgatct gaacaacctg attattcagc gcgataatag cgagaaactg 840

gaataccatg agaagtatca catcatcgag aacgtgttca agcagaaaaa aaagccgacg 900

ctgaaacaaa tcgcaaaaga gattggcgtt aacccggaag atattaaagg ttatcgtatt 960

accaaaagcg gcacaccgga gtttacatcc tttaaactgt tccacgatct gaaaaaagtg 1020

gtgaaagatc atgccatcct ggatgatatt gatctgctga atcagattgc agaaatcctg 1080

accatctatc aggataaaga tagcattgtt gcagaactgg gtcagctgga atatctgatg 1140

agcgaagccg ataaacagag cattagcgaa ctgaccggtt ataccggtac acatagcctg 1200

tcactgaaat gcatgaacat gattatcgat gaactgtggc atagcagcat gaaccagatg 1260

gaagttttta cctatctgaa tatgcgtccg aaaaagtatg agctgaaagg ttatcagcgt 1320

attccgaccg atatgattga tgatgcaatt ctgagtccgg ttgtgaaacg caccttttatt 1380

cagagcatca acgtgatcaa caaagtgatc gagaaatatg gcatccccga agatatcatt 1440

atcgaactgg cacgtgaaaa taactccgat gatcgcaaaa agttcatcaa caacctgcag 1500

aaaaagaatg aagcaacccg caaacgcatt aacgaaatta ttggtcagac cggtaatcag 1560

aatgccaaac gtattgtgga aaaaatccgt ctgcatgatc agcaagaggg caaatgtctg 1620

tatagcctgg aaagcattcc tctggaagat ctgctgaaca atccgaatca ttatgaagtg 1680

gatcacatta ttccgcgtag cgtgagcttt gataattcct atcataataa agtgctggtg 1740

aaacagagcg aaaactccaa aaaatccaac ctgacaccgt atcagtattt caatagcggc 1800

aaatccaaac tgagctacaa ccagtttaaa cagcatattc tgaacctgag caaaagccag 1860

gatcgcatca gcaagaagaa gaaggagtac ctgctggaag aacgcgatat taacaaattt 1920

gaagtgcaga aagaatttat caaccgcaac ctggttgata cccgttatgc aacccgtgaa 1980

ctgaccaatt atctgaaagc atatttcagc gccaacaaca tgaacgtgaa agtgaaaacg 2040

attaacggca gctttaccga ttatctgcgt aaagtgtgga aattcaaaaa agaacgcaac 2100

cacggctata aacatcatgc cgaagatgcc ctgattattg caaatgcaga tttcctgttt 2160

aaagaaaaca aaaaactgaa agccgtcaac agcgtgctgg aaaaaccgga aattgagaca 2220

aaacagctgg acattcaggt tgatagcgaa gataattaca gcgaaatgtt tatcatcccg 2280

aaacaggtgc aggatatcaa agattttcgc aacttcaaat atagccaccg cgttgacaaa 2340

aaacctaatc gtcagctgat taacgatacc ctgtatagca cccgcaaaaa agataacagc 2400

acctatattg tgcagaccat taaagacatc tacgccaaag ataataccac cctgaaaaaa 2460

cagttcgaca aaagcccaga aaaatttctg atgtatcagc atgatccgcg taccttcgaa 2520

aaactggaag ttattatgaa acagtatgcc aacgagaaaa atccgctggc caaatatcac 2580

gaagaaaccg gtgaatatct gaccaaatat tccaagaaga acaacggtcc gatcgttaaa 2640

tccctgaaat atatcggtaa taaactgggc agccatctgg atgttaccca tcagtttaaa 2700

agctccacaa agaagctggt taagctgtcc attaaaaact atcgcttcga tgtgtatctg 2760

accgagaaag gttataagtt tgtgaccatt gcctacctga atgtgttcaa aaaagacaac 2820

tattactata ttccgaaaga caaataccaa gaacttaaag agaagaagaa aatcaaggac 2880

accgatcagt ttatcgccag cttctataaa aacgatctga tcaagctgaa cggcgacctg 2940

tataaaatca ttggtgtgaa tagtgatgac cgcaacatca ttgagctgga ttattacgac 3000

atcaaataca aggattactg cgagatcaac aacattaaag gtgaaccgcg tatcaaaaag 3060

accattggca aaaaaacgga aagcatcgaa aagtttacca ccgatgttct gggtaatctg 3120

tatctgcata gtaccgaaaa agcaccgcag ctgattttca aacgcggtct g 3171

[0058]

<210> 15
<211> 3171
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> 密码子优化的Gib11Spa-1多核苷酸

<400> 15
atgaaccaga agttcattct cggtctggac attggcatta ctagcgtggg atacggcttg 60

attgactacg agactaagaa catcatcgat gccggcgtcc gcctgttccc ggaagccaac 120

gtggagaaca atgagggccg gaggtcgaag agaggctccc gccgcctgaa gcggcggcga 180

atccaccggc tggagagagt gaagctgctg ctcaccgaat acgacctgat caacaaagaa 240

cagatcccga cctccaacaa cccgtaccag atcagagtga agggtctgtc agaaatcctg 300

tccaaggacg aactggcaat cgccctgctg cacctggcca agcggcgcgg aatccacaac 360

gtggatgtgg ctgccgacaa ggaagaaacc gcttccgact ccctgagcac taaggaccag 420

atcaacaaga acgccaagtt cttggagtcc cgctacgtgt gcgagctgca gaaggaacgg 480

ctggaaaacg aaggtcacgt gcgcggagtg gaaaaccggt tcctgacaaa ggacattgtg 540

cgcgaagcga agaagatcat tgatacccaa atgcagtact accctgaaat cgacgagact 600

[0059]

ttcaaggaaa agtacatttc cctggtggaa acccggcggg aatacttcga aggccccgga 660

cagggatcgc cgttcggatg gaacggggac ctcaagaagt ggtacgagat gctgatgggg 720

cactgtacct actttccgca agaactgcgc tccgtgaagt acgcgtactc cgcggatctc 780

ttcaacgcgt tgaatgacct gaacaacctg atcattcaga gagacaattc cgaaaagctc 840

gagtaccacg agaagtatca catcatcgag aatgtgttca agcagaagaa gaaaccgacc 900

ctcaagcaaa tcgccaagga gattggcgtc aacccagagg acatcaaggg atatcggatt 960

accaagagcg gcactcccga gtttacctct ttcaagctgt ttcatgatct gaagaaagtc 1020

gtgaaggacc atgccattct cgacgacatt gatctcctga atcagatcgc agagatcctg 1080

actatctacc aagacaagga ctcgattgtg gcagagctgg gtcagctcga atacctgatg 1140

tccgaggccg acaagcagtc catctccgaa ctgacagggt acacggggac tcatagcctg 1200

tcgctgaagt gcatgaacat gatcattgat gaactgtggc acagctccat gaaccaaatg 1260

gaagtgttta cctacctcaa catgcgccct aagaagtacg aactgaaagg ctaccagcgc 1320

atccccaccg acatgatcga cgacgcgatc ttgtcccctg tggtcaagag gaccttcatt 1380

caatccatca acgtgatcaa caaggtcatc gaaaagtacg gaataccaga agatatcatc 1440

attgagctcg ctcgggagaa caactcggat gaccggaaga agttcatcaa caatcttcag 1500

aagaagaacg aagcgactcg gaaacggatc aacgagatca tcggacagac cggaaaccag 1560

aacgccaaac ggattgtcga aaagattaga ctgcacgacc agcaggaagg gaagtgcctg 1620

tactcactcg agtcaatacc gctcgaggac ctgttgaaca accctaacca ctatgaagtg 1680

gaccacatca tccctcggtc cgtgagcttc gacaactcgt accacaacaa agtgctcgtg 1740

aagcagtccg aaaactccaa gaaatccaac ctgaccccgt accaatactt caattcggga 1800

aagtcgaagc tgtcgtacaa ccagttcaaa caacacatac tcaaccttag caaaagccag 1860

gatcgcattt ccaagaagaa gaaggaatac ctcctcgagg aaagggacat caacaagttc 1920

gaagtgcaga aagagttcat caatcgcaac ttggtggata ccagatatgc cacccgggaa 1980

ctcaccaact atctcaaggc ctacttttcc gccaacaaca tgaacgtgaa ggtcaagacc 2040

atcaacgggt ccttcactga ctacctgaga aaggtctgga agttcaagaa ggaacgcaac 2100

cacggataca agcaccacgc tgaggacgct ctgatcatcg ccaatgccga cttcctgttc 2160

aaggaaaaca agaagctgaa agctgtcaac tcagtgctgg aaaagcctga aatcgagact 2220

aagcagctgg atatccaagt ggactctgag gacaactaca gcgagatgtt catcatccct 2280

aaacaagtgc aggatatcaa ggactttcgc aacttcaagt actcacaccg ggtggacaag 2340

aaaccgaata gacagctgat caacgacacg ttgtattcca cccggaagaa ggataactca 2400

acctacattg tgcagactat caaggatatc tacgccaaag ataacactac tctgaagaaa 2460

caattcgaca agtccccaga gaagttcctg atgtaccagc acgacccccg aacctttgag 2520

aagcttgaag tgatcatgaa gcagtacgcc aacgagaaga acccgctggc caagtaccat 2580

gaagaaaccg gagaatacct gaccaagtac agcaagaaga acaacggtcc cattgtcaag 2640

agcctgaagt acatcggcaa caagctggga tcccacctcg acgtgacaca tcagttcaag 2700

tcgtcgacta agaagcttgt gaagctgtca atcaagaact atagattcga cgtgtacttg 2760

accgaaaagg gatacaagtt cgtgaccata gcctatctga acgtgttcaa gaaagataac 2820

tactactaca tccccaagga caagtaccag gagctcaaag aaaagaagaa gatcaaagac 2880

accgaccagt tcattgcctc cttctacaag aacgacctga tcaaactgaa cggcgacctc 2940

tacaagatca ttggagtgaa cagcgatgac aggaacatca ttgagctgga ctactacgac 3000

atcaagtaca aggactactg cgagatcaac aacatcaagg gcgaaccccg gatcaagaaa 3060

actattggaa agaaaaccga gtccattgag aagttcacca ctgacgtgct gggaaacctt 3120

tacctccact ccaccgagaa ggcaccacaa ctgatcttca agcgcggcct g 3171

<210> 16
<211> 3171
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

[0060]

<220>
<221> misc_feature
<223> 编码Gib11Spa-2的DNA序列

<400> 16
atgaaccaga aatttatcct gggtctggat attggtatta ccagcgttgg ttatggcctg 60

attgattacg aaaccaaaaa cattattgat gccggtgttc gtctgtttcc ggaagcaaat 120

gttgaaaata atgaaggtcg tcgtagcaaa cgtggtagcc gtcgtctgaa acgtcgtcgt 180

attcatcgtc tggaacgtgt taaactgctg ctgaccgaat atgatctgat taacaaagag 240

cagattccga ccagcaataa cccgtatcag attcgtgtta aaggtctgag cgaaatcctg 300

agcaaagatg aactggcaat tgcactgctg catctggcaa aacgccgtgg cattcataat 360

gttgatgttg cagcagataa agaagaaacc gcaagcgata gcctgagcac caaagatcag 420

attaacaaaa acgccaaatt tctggaaagc cgctatgttt gtgaactgca gaaagaacgt 480

ctggaaaatg aaggtcatgt tcgtggtgtt gaaaatcgct ttctgacgaa agatattgtg 540

cgtgaggcca aaaaaatcat cgatacccag atgcagtatt acccggaaat tgatgaaacc 600

ttcaaagaga aatatatcag cctggttgaa acccgtcgcg aatattttga aggtcctggt 660

cagggtagcc cgtttggttg gaatggtgat ctgaaaaaat ggtacgaaat gctgatgggt 720

cactgtacct attttccgca agaactgcgt agcgttaaat atgcctatag cgcagacctg 780

[0061]

tttaatgcac tgaatgatct gaacaacctg attattcagc gcgataatag cgagaaactg 840

gaataccatg agaagtatca catcatcgag aacgtgttca agcagaaaaa aaagccgacg 900

ctgaaacaaa tcgcaaaaga gattggcgtt aacccggaag atattaaagg ttatcgtatt 960

accaaaagcg gcacaccgga gtttacatcc tttaaactgt tccacgatct gaaaaaagtg 1020

gtgaaagatc atgccatcct ggatgatatt gatctgctga atcagattgc agaaatcctg 1080

accatctatc aggataaaga tagcattgtt gcagaactgg gtcagctgga atatctgatg 1140

agcgaagccg ataaacagag cattagcgaa ctgaccggtt ataccggtac acatagcctg 1200

tcactgaaat gcatgaacat gattatcgat gaactgtggc atagcagcat gaaccagatg 1260

gaagttttta cctatctgaa tatgcgtccg aaaaagtatg agctgaaagg ttatcagcgt 1320

attccgaccg atatgattga tgatgcaatt ctgagtccgg ttgtgaaacg cacctttatt 1380

cagagcatca acgtgatcaa caaagtgatc gagaaatatg gcatccccga agatatcatt 1440

atcgaactgg cacgtgaaaa taactccgat gatcgcaaaa agttcatcaa caacctgcag 1500

aaaaagaatg aagcaacccg caaacgcatt aacgaaatta ttggtcagac cggtaatcag 1560

aatgccaaac gtattgtgga aaaaatccgt ctgcatgatc agcaagaggg caaatgtctg 1620

tatagcctgg aaagcattcc tctggaagat ctgctgaaca atccgaatca ttatgaagtg 1680

gatcacatta ttccgcgtag cgtgagcttt gataattcct atcataataa agtgctggtg 1740

aaacagagcg aaaactccaa aaaatccaac ctgacaccgt atcagtattt caatagcggc 1800

aaatccaaac tgagctacaa ccagtttaaa cagcatattc tgaacctgag caaaagccag 1860

gatcgcatca gcaagaagaa gaaggagtac ctgctggaag aacgcgatat taacaaattt 1920

gaagtgcaga aagaatttat caaccgcaac ctggttgata cccgttatgc aacccgtgaa 1980

ctgaccaatt atctgaaagc atatttcagc gccaacaaca tgaacgtgaa agtgaaaacg 2040

attaacggca gctttaccga ttatctgcgt aaagtgtgga aattcaaaaa agaacgcaac 2100

cacggctata aacatcatgc cgaagatgcc ctgattattg caaatgcaga tttcctgttt 2160

aaagaaaaca aaaaactgaa agccgtcaac agcgtgctgg aaaaaccgga attgagaca 2220

aaacagctgg acattcaggt tgatagcgaa gataattaca gcgaaatgtt tatcatcccg 2280

aaacaggtgc aggatatcaa agattttcgc aacttcaaat tcagccaccg cgttgataaa 2340

aaaccgaatc gtcagctgat taacgatacc ctgtatagca cccgtatgaa agatgagcat 2400

gattatattg tgcagaccat cacggatatc tatggcaaag ataataccaa cctgaaaaaa 2460

cagttcaaca aaaacccgga aaaatttctg atgtatcaga acgatccgaa aacctttgag 2520

aaactgagca tcatcatgaa acagtacagc gacgaaaaaa acccgctggc caaatattac 2580

gaagaaaccg gtgaatatct gaccaaatat agcaagaaaa acaacggtcc gatcgtgaaa 2640

aagatcaaac tgctgggtaa taaagtgggc aatcatctgg atgtgaccaa caaatatgaa 2700

aactccacga agaagctggt taagctgtcc attaaaaact atcgcttcga tgtgtatctg 2760

accgagaaag gttataagtt tgtgaccatt gcctacctga atgtgttcaa aaaagacaac 2820

tattactata ttccgaaaga caaataccaa gaacttaaag agaagaagaa aatcaaggac 2880

accgatcagt ttatcgccag cttctataaa aacgatctga tcaagctgaa cggcgacctg 2940

tataaaatca ttggtgtgaa tagtgatgac cgcaacatca ttgagctgga ttattacgac 3000

atcaaataca aggattactg cgagatcaac aacattaaag gtgaaccgcg tatcaaaaag 3060

accattggca aaaaaacgga aagcatcgaa aagtttacca ccgatgttct gggtaatctg 3120

tatctgcata gtaccgaaaa agcaccgcag ctgattttca aacgcggtct g 3171

〈210〉 17
〈211〉 3171
〈212〉 DNA
〈213〉 人工序列

〈220〉
〈223〉 合成的多核苷酸

〈220〉
〈221〉 misc_feature
〈223〉 编码Gib11Spa-3的DNA序列

〈400〉 17

[0062]

atgaaccaga aatttatcct gggtctggat attggtatta ccagcgttgg ttatggcctg 60

attgattacg aaaccaaaaa cattattgat gccggtgttc gtctgtttcc ggaagcaaat 120

gttgaaaata atgaaggtcg tcgtagcaaa cgtggtagcc gtcgtctgaa acgtcgtcgt 180

attcatcgtc tggaacgtgt taaactgctg ctgaccgaat atgatctgat taacaaagag 240

cagattccga ccagcaataa cccgtatcag attcgtgtta aaggtctgag cgaaatcctg 300

agcaaagatg aactggcaat tgcactgctg catctggcaa aacgccgtgg cattcataat 360

gttgatgttg cagcagataa agaagaaacc gcaagcgata gcctgagcac caaagatcag 420

attaacaaaa acgccaaatt tctggaaagc cgctatgttt gtgaactgca gaaagaacgt 480

ctggaaaatg aaggtcatgt tcgtggtgtt gaaaatcgct ttctgacgaa agatattgtg 540

cgtgaggcca aaaaaatcat cgatacccag atgcagtatt acccggaaat tgatgaaacc 600

ttcaaagaga aatatatcag cctggttgaa acccgtcgcg aatattttga aggtcctggt 660

cagggtagcc cgtttggttg gaatggtgat ctgaaaaaat ggtacgaaat gctgatgggt 720

cactgtacct attttccgca agaactgcgt agcgttaaat atgcctatag cgcagacctg 780

tttaatgcac tgaatgatct gaacaacctg attattcagc gcgataatag cgagaaactg 840

gaataccatg agaagtatca catcatcgag aacgtgttca agcagaaaaa aaagccgacg 900

ctgaaacaaa tcgcaaaaga gattggcgtt aacccggaag atattaaagg ttatcgtatt 960

accaaaagcg gcacaccgga gtttacatcc tttaaactgt tccacgatct gaaaaaagtg 1020

[0063]

gtgaaagatc atgccatcct ggatgatatt gatctgctga atcagattgc agaaatcctg 1080

accatctatc aggataaaga tagcattgtt gcagaactgg gtcagctgga atatctgatg 1140

agcgaagccg ataaacagag cattagcgaa ctgaccggtt ataccggtac acatagcctg 1200

tcactgaaat gcatgaacat gattatcgat gaactgtggc atagcagcat gaaccagatg 1260

gaagttttta cctatctgaa tatgcgtccg aaaaagtatg agctgaaagg ttatcagcgt 1320

attccgaccg atatgattga tgatgcaatt ctgagtccgg ttgtgaaacg cacctttatt 1380

cagagcatca acgtgatcaa caaagtgatc gagaaatatg gcatccccga agatatcatt 1440

atcgaactgg cacgtgaaaa taactccgat gatcgcaaaa agttcatcaa caacctgcag 1500

aaaaagaatg aagcaacccg caaacgcatt aacgaaatta ttggtcagac cggtaatcag 1560

aatgccaaac gtattgtgga aaaaatccgt ctgcatgatc agcaagaggg caaatgtctg 1620

tatagcctgg aaagcattcc tctggaagat ctgctgaaca atccgaatca ttatgaagtg 1680

gatcacatta ttccgcgtag cgtgagcttt gataattcct atcataataa agtgctggtg 1740

aaacagagcg aaaactccaa aaaatccaac ctgacaccgt atcagtattt caatagcggc 1800

aaatccaaac tgagctacaa ccagtttaaa cagcatattc tgaacctgag caaaagccag 1860

gatcgcatca gcaagaagaa gaaggagtac ctgctggaag aacgcgatat caacaaattt 1920

gaagtccaga aagagtttat caaccgcaat ctggttgata cccgttatgc aacccgtgaa 1980

ctgaccagct atctgaaagc atatttcagc gccaataaca tggacgtgaa agtgaaaaca 2040

attaacggca gctttaccaa ccatctgcgt aaagtttggc gctttgataa atatcgcaac 2100

cacggctata aacatcatgc cgaagatgca ctgattattg ccaatgcaga tttcctgttc 2160

aaagaaaaca aaaaactgca gaacaccaac aagatcctgg aaaaaccgac cattgaaaac 2220

aacaccaaaa aagtgaccgt cgagaaagaa gaggattaca acaacgtttt tgaaaccccg 2280

aaactggtcg aggatattaa acagtatcgc gactataaat tcagccaccg cgttgataaa 2340

aaaccgaatc gtcagctgat taacgatacc ctgtatagca cccgtatgaa agatgagcat 2400

gattatattg tgcagaccat cacggatatc tatggcaaag ataataccaa cctgaaaaaa 2460

cagttcaaca aaaacccgga aaaatttctg atgtatcaga acgatccgaa aacctttgag 2520

aaactgagca tcatcatgaa acagtacagc gacgaaaaaa acccgctggc caaatattac 2580

gaagaaaccg gtgaatatct gaccaaatat agcaagaaaa acaacggtcc gatcgtgaaa 2640

aagatcaaac tgctgggtaa taaagtgggc aatcatctgg atgtgaccaa caaatatgaa 2700

aactccacga agaagctggt taagctgtcc attaaaaact atcgcttcga tgtgtatctg 2760

accgagaaag gttataagtt tgtgaccatt gcctacctga atgtgttcaa aaaagacaac 2820

tattactata ttccgaaaga caaataccaa gaacttaaag agaagaagaa aatcaaggac 2880

accgatcagt ttatcgccag cttctataaa aacgatctga tcaagctgaa cggcgacctg 2940

tataaaatca ttggtgtgaa tagtgatgac cgcaacatca ttgagctgga ttattacgac 3000

atcaaataca aggattactg cgagatcaac aacattaaag gtgaaccgcg tatcaaaaag 3060

accattggca aaaaaacgga aagcatcgaa aagtttacca ccgatgttct gggtaatctg 3120

tatctgcata gtaccgaaaa agcaccgcag ctgattttca aacgcggtct g 3171

<210> 18
<211> 3171
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> 密码子优化的Gib11Spa-3多核苷酸

<400> 18
atgaaccaga agttcattct cggtctggac attggcatta ctagcgtggg atacggcttg 60

attgactacg agactaagaa catcatcgat gccggcgtcc gcctgttccc ggaagccaac 120

gtggagaaca atgagggccg gaggtcgaag agaggctccc gccgcctgaa gcggcggcga 180

atccaccggc tggagagagt gaagctgctg ctcaccgaat acgacctgat caacaaagaa 240

[0064]

cagatcccga cctccaacaa cccgtaccag atcagagtga agggtctgtc agaaatcctg 300

tccaaggacg aactggcaat cgccctgctg cacctggcca agcggcgcgg aatccacaac 360

gtggatgtgg ctgccgacaa ggaagaaacc gcttccgact ccctgagcac taaggaccag 420

atcaacaaga acgccaagtt cttggagtcc cgctacgtgt gcgagctgca gaaggaacgg 480

ctggaaaacg aaggtcacgt gcgcggagtg gaaaaccggt tcctgacaaa ggacattgtg 540

cgcgaagcga agaagatcat tgatacccaa atgcagtact accctgaaat cgacgagact 600

ttcaaggaaa agtacatttc cctggtggaa acccggcggg aatacttcga aggccccgga 660

cagggatcgc cgttcggatg gaacggggac ctcaagaagt ggtacgagat gctgatgggg 720

cactgtacct actttccgca agaactgcgc tccgtgaagt acgcgtactc cgcggatctc 780

ttcaacgcgt tgaatgacct gaacaacctg atcattcaga gagacaattc cgaaaagctc 840

gagtaccacg agaagtatca catcatcgag aatgtgttca agcagaagaa gaaaccgacc 900

ctcaagcaaa tcgccaagga gattggcgtc aacccagagg acatcaaggg atatcggatt 960

accaagagcg gcactcccga gtttacctct ttcaagctgt ttcatgatct gaagaaagtc 1020

gtgaaggacc atgccattct cgacgacatt gatctcctga atcagatcgc agagatcctg 1080

actatctacc aagacaagga ctcgattgtg gcagagctgg gtcagctcga atacctgatg 1140

tccgaggccg acaagcagtc catctccgaa ctgacagggt acacggggac tcatagcctg 1200

tcgctgaagt gcatgaacat gatcattgat gaactgtggc acagctccat gaaccaaatg 1260

gaagtgttta cctacctcaa catgcgccct aagaagtacg aactgaaagg ctaccagcgc 1320

atccccaccg acatgatcga cgacgcgatc ttgtcccctg tggtcaagag gaccttcatt 1380

caatccatca acgtgatcaa caaggtcatc gaaaagtacg gtattccaga agatatcatc 1440

attgagctcg ctcgggagaa caactcggat gaccggaaga agttcatcaa caatcttcag 1500

aagaagaacg aagcgactcg gaaacggatc aacgagatca tcggacagac cggaaaccag 1560

aacgccaaac ggattgtcga aaagattaga ctgcacgacc agcaggaagg gaagtgcctg 1620

tactcactcg agtcaatacc gctcgaggac ctgttgaaca accctaacca ctatgaagtg 1680

gaccacatca tccctcggtc cgtgagcttc gacaactcgt accacaacaa agtgctcgtg 1740

aagcagtccg aaaactccaa gaaatccaac ctgaccccgt accaatactt caattcggga 1800

aagtcgaagc tgtcgtacaa ccagttcaaa caacacatac tcaaccttag caaaagccag 1860

gatcgcattt ccaagaagaa gaaggaatac ctcctcgagg aaagggacat caacaagttc 1920

gaagtgcaga aagagttcat caatcgcaac ttggtggata ccagatatgc cacccgggaa 1980

ctcaccagct atctcaaggc ctacttttcc gccaacaaca tggacgtgaa ggtcaagacc 2040

atcaacgggt ccttcactaa ccatctgaga aaggtctggc ggtttgacaa gtaccgcaac 2100

cacggataca agcaccacgc tgaagacgct ctgatcatcg ccaatgccga cttcctgttc 2160

[0065] aaggaaaaca agaagctgca gaacacgaac aagattctgg aaaagcctac cattgagaac 2220

aacactaaga aggtcaccgt ggagaaggaa gaggactaca acaacgtgtt cgaaactcct 2280

aaactggtgg aggatatcaa gcaataccgc gactacaagt tctcacaccg ggtggacaag 2340

aaaccgaata gacagctgat caacgacacg ttgtattcca cccggatgaa ggatgagcat 2400

gactacattg tgcagactat caccgatatc tacggaaaag ataacactaa cctgaagaaa 2460

caattcaaca agaacccaga gaagttcctg atgtaccaga acgaccccaa gacctttgag 2520

aagctttcca tcatcatgaa gcagtactcc gacgagaaga acccgctggc caagtactac 2580

gaagaaaccg gagaatacct gaccaagtac agcaagaaga acaacggtcc cattgtcaag 2640

aagatcaagc tgctcggcaa caaggtcgga aaccacctcg acgtgacaaa caagtacgag 2700

aactcgacta agaagcttgt gaagctgtca atcaagaact atagattcga cgtgtacttg 2760

accgaaaagg gatacaagtt cgtgaccata gcctatctga acgtgttcaa gaaagataac 2820

tactactaca tccccaagga caagtaccag gagctcaaag aaaagaagaa gatcaaagac 2880

accgaccagt tcattgcctc cttctacaag aacgacctga tcaaactgaa cggcgacctc 2940

tacaagatca ttggagtgaa cagcgatgac aggaacatca ttgagctgga ctactacgac 3000

atcaagtaca aggactactg cgagatcaac aacatcaagg gcgaaccccg gatcaagaaa 3060

actattggaa agaaaaccga gtccattgag aagttcacca ctgacgtgct gggaaacctt 3120

tacctccact ccaccgagaa ggcaccacaa ctgatcttca agcgcggcct g 3171

<210> 19
<211> 3166
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> 编码P2H12的DNA序列

<400> 19

[0066]

atgaaccaga aatttatcct gggtctggat attggtatta ccagcgttgg ttatggcctg 60
attgattacg aaaccaaaaa cattattgat gccggtgttc gtctgtttcc ggaagcaaat 120
gttgaaaata atgaaggtcg tcgtagcaaa cgtggtagcc gtcgtctgaa acgtcgtcgt 180
attcatcgtc tggaacgtgt taaaaaactg ctggaagatt ataacctgct ggatcagagc 240
cagattccgc agagcaccaa tccgtatgca attcgtgtta aaggtctgag cgaagcactg 300
agcaaagatg aactggttat tgcactgctg catattgcaa aacgccgtgg cattcataat 360
atcaatgtta gcagcgaaga tgaggatgca agcaatgaac tgagcaccaa agaacaaatt 420
aaccgcaata ataagctgct gaaggacaaa tatgtttgcg aagttcagct gcagcgtctg 480
aaagaaggtc agattcgcgg agaaaaaaat cgctttaaaa ccaccgatat cctgaaagaa 540
attgatcagc tgcttaaagt gcagaaggat tatcataacc tggacatcga tttcatcaac 600
cagtacaaag aaatcgttga aacccgtcgc gaatattttg aaggtccggg taaaggtagc 660
ccgtatggtt gggaaggtga tccgaaagca tggtatgaaa ccctgatggg tcattgtacc 720
tattttccgg atgaactgcg tagcgttaaa tatgcctata gcgcagacct gtttaatgca 780
ctgaatgatc tgaataacct ggtgattcag cgtgatggtc tgagcaaact ggaatatcat 840
gagaaatatc acatcatcga aaacgtgttc aaacagaaga agaaaccgac cctgaaacaa 900
atcgccaacg aaattaatgt gaacccggaa gatattaaag gctaccgtat taccaaaagc 960
ggcaaaccgg aatttacatc ctttaaactg ttccacgatc tgaaaaaagt ggtgaaagat 1020
catgccatcc tggatgatat tgatctgctg aatcagattg cagaaatcct gaccatctat 1080
caggataaag atagcattgt tgcagaactg ggtcagctgg aatatctgat gagcgaagcc 1140
gataaacaga gcattagcga actgaccggt tataccggta cacatagcct gtcactgaaa 1200
tgcatgaaca tgattatcga tgaactgtgg catagcagca tgaaccagat ggaagttttt 1260
acctatctga atatgcgtcc gaaaaagtat gagctgaaag gttatcagcg tattccgacc 1320
gatatgattg atgatgcaat tctgagtccg gttgtgaaac gcacctttat tcagagcatc 1380
aacgtgatca acaaagtgat cgagaaatat ggcatccccg aagatatcat tatcgaactg 1440

[0067]

gcacgtgaaa ataactccga tgatcgcaaa aagttcatca acaacctgca gaaaaagaat 1500

gaagcaaccc gcaaacgcat taacgaaatt attggtcaga ccggtaatca gaatgccaaa 1560

cgtattgtgg aaaaaatccg tctgcatgat cagcaagagg gtaaatgtct gtatagcctg 1620

gaaagcattc ctctggaaga tctgctgaac aatccgaatc attatgaagt ggatcacatt 1680

attccgcgta gcgtgagctt tgataattcc tatcataata aagtgctggt gaaacagagc 1740

gaaaactcca aaaaatccaa cctgacaccg tatcagtatt tcaatagcgg caaatccaaa 1800

ctgagctaca accagtttaa acagcatatt ctgaacctga gcaaaagcca ggatcgcatc 1860

agcaagaaga agaaggagta cctgctggaa gaacgcgata ttaacaaatt tgaagtgcag 1920

aaagaattta tcaaccgcaa cctggttgat acccgttatg caacccgtga actgaccaat 1980

tatctgaaag catatttcag cgccaacaac atgaacgtga aagtgaaaac gattaacggc 2040

agctttaccg attatctgcg taaagtgtgg aaattcaaaa aagaacgcaa ccacggctat 2100

aaacatcatg ccgaagatgc cctgattatt gcaaatgcag atttcctgtt taaagaaaac 2160

aaaaaactga aagccgtcaa cagcgtgctg gaaaaaccgg aaattgagac aaaacagctg 2220

gacattcagg ttgatagcga agataattac agcgaaatgt ttatcatccc gaaacaggtg 2280

caggatatca aagattttcg caacttcaaa tatagccacc gcgttgacaa aaaacctaat 2340

cgtcagctga ttaacgatac cctgtatagc acccgcaaaa aagataacag cacctatatt 2400

gtgcagacca ttaaagacat ctacgccaaa gataatacca ccctgaaaaa acagttcgac 2460

aaaagcccag aaaaatttct gatgtatcag catgatccgc gtaccttcga aaaactggaa 2520

gttattatga aacagtatgc caacgagaaa aatccgctgg ccaaatatca cgaagaaacc 2580

ggtgaatatc tgaccaaata ttccaagaag aacaacggtc cgatcgttaa atccctgaaa 2640

tatatcggta ataaactggg cagccatctg gatgttaccc atcagtttaa aagctccaca 2700

aagaagctgg ttaaactgtc catcaaaccg tatcgctttg atgtgtatct gaccgacaaa 2760

ggctataaat tcattaccat cagctatctg gacgtgctga aaaaagacaa ctattattat 2820

atcccggaac agaaatatga taaactgaaa ctgggtaaag ccatcgataa aaacgccaaa 2880

tttatcgcca gcttctacaa aaacgacctg attaaactgg atggcgagat ctataaaatc 2940

atcggtgtta atagcgacac ccgcaatatg attgagctgg atctgccgga tattcgctat 3000

aaagaatatt gcgaactgaa caacattaaa ggcgaaccgc gtatcaaaaa gaccatcggc 3060

aaaaaagtga atagcatcga gaaactgacc accgatgttc tgggtaatgt gtttaccaat 3120

acccagtata ccaaacctca gctgctgttc aaacgcggta atggtg 3166

<210> 20
<211> 3162
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> 密码子优化的P2H12多核苷酸

<400> 20
atgaaccaaa agttcattct ggggctcgat atcggcatca cctccgtggg atatggtctg 60

atcgactacg agactaagaa catcatcgac gctggagtgc gactgttccc ggaagcgaac 120

gtggagaaca acgaaggccg cagatccaag cgcgggtcca gaaggctcaa gaggcggagg 180

atccatagac tcgagagagt gaagaagctc cttgaagatt acaatctgtt ggaccagtca 240

cagattcccc aaagcaccaa cccgtacgcc atcagagtga agggcctgtc agaagcactg 300

tcgaaagatg aactggtcat tgccctgctg catattgcca aacggcgcgg aatccataac 360

atcaacgtgt cgagcgaaga tgaggacgcg tccaacgaac tgtcaaccaa ggaacagatc 420

aaccggaaca acaaactgct gaaggacaaa tacgtctgcg aggtgcagct tcaacggctg 480

aaagagggac agatcagggg agagaaaaac cggttcaaga ccaccgacat ccttaaggag 540

atcgaccaac tcctgaaagt gcagaaggac tatcacaacc tcgacattga ttttatcaac 600

cagtacaagg agattgtgga aactcggagg gagtacttcg agggacctgg aaagggatcc 660

[0068] ccttatggct gggaagggga ccccaaggct tggtacgaaa ccctgatggg ccattgcact 720

tactttccgg atgaactccg gtccgtgaag tacgcttact ctgccgacct gttcaatgca 780

ctcaacgacc ttaacaactt ggtgatccaa cgcgatggtc tttccaagtt ggagtaccac 840

gaaaagtacc acatcatcga gaacgtgttc aagcagaaaa agaagccaac tctgaagcag 900

attgccaacg aaattaacgt gaaccccgag gatatcaagg ataccggat taccaagtcc 960

ggcaaaccag agttcacctc attcaagctg tttcacgatc tgaagaaggt cgtgaaggac 1020

cacgccatcc tggatgacat tgatcttctg aaccagattg ctgagatcct gaccatctac 1080

caggacaagg actcgattgt ggccgaactg ggacagctcg agtacctgat gtccgaagcc 1140

gataagcagt ccatcagcga actcaccggt tacaccggta cccactcctt gtcccttaag 1200

tgcatgaaca tgatcattga cgaactgtgg cactccagca tgaaccagat ggaggtgttc 1260

acctacttga acatgcgccc gaagaagtac gagctgaagg gctaccagcg catacccacg 1320

gacatgatcg acgacgccat cctctcaccg gtggtcaagc gcaccttcat tcaatctatc 1380

aacgtgatca acaaggtcat cgagaagtac ggcattcctg aggatatcat catcgagctg 1440

gctcgggaga acaactcaga cgataggaag aagttcatta acaacctcca gaaaaagaac 1500

gaggccactc gcaagcggat taatgagatc atcggtcaga ccgggaacca gaacgccaag 1560

cggatcgtgg aaaagattcg gctccacgac caacaggagg gaaagtgtct gtactcgctg 1620

gagtccattc ccctggagga cctcctgaac aacccaaacc actacgaggt cgatcacata 1680

atcccccgca gcgtgtcatt cgacaactcc taccataaca aggtcctcgt gaagcagtcg 1740

gagaatagca agaagtcgaa cctgactccg taccagtact tcaactccgg caaatccaag 1800

ctgtcctaca atcagttcaa acagcacatt ctcaacctgt ccaagagcca ggacaggatt 1860

tcgaagaaga agaaggaata ccttctcgag gaacgggata tcaataagtt cgaggtgcag 1920

aaggagttta tcaatagaaa cctggtggac actcgctatg ccacccgcga actgaccaac 1980

tacctgaagg cgtacttctc cgccaacaac atgaacgtga aggtcaaaac tattaacggc 2040

agcttcaccg actatctgcg caaggtctgg aagttcaaga aggaacgcaa ccacggttac 2100

aagcaccacg cggaagatgc gctgattatc gccaacgctg acttcctgtt caaggaaaac 2160

aagaagctca aggccgtgaa ctcagtgctc gagaagcctg aaatcgagac taagcagctg 2220

gacatccagg tcgattcgga agataactac tccgaaatgt tcatcatccc taagcaagtg 2280

caggacatca aggacttcag gaatttcaag tacagccatc gcgtggacaa gaagccaaac 2340

agacagctga tcaacgatac actgtattcc acccggaaga aggacaactc cacctacatc 2400

gtccaaacca ttaaggacat ctacgcaaag gacaacacca cgcttaagaa gcagttcgac 2460

aagagccccg aaaagttcct catgtaccag cacgacccca gaaccttcga gaagcttgaa 2520

gtgatcatga agcagtacgc caacgaaaag aacccactgg ctaagtacca cgaggaaacc 2580

[0069] ggcgaatacc tgaccaagta ctccaaaaag aacaacggac cgatcgtcaa gtccctgaag 2640

tacattggga acaagctcgg ctcgcacctc gatgtgaccc accagttcaa gtcctcgacc 2700

aaaaagctcg tgaagctgtc catcaagccg taccggttcg acgtgtacct gactgacaag 2760

ggatataagt tcatcaccat ttcctacctc gacgtgttga agaaggataa ctactactac 2820

attccggaac agaagtacga caagctcaag ctcggaaagg ccatcgacaa aaatgcgaag 2880

ttcatcgcga gcttctacaa gaatgacttg atcaagctgg atggcgaaat ctacaagatc 2940

atcggggtca actccgatac ccgcaacatg attgagctgg atctgcccga cattcggtac 3000

aaggaatact gcgagctgaa caacatcaag ggagaaccgc ggatcaagaa aaccatcgga 3060

aagaaagtga acagcatcga gaaactgact actgacgtcc tgggaaacgt gttcaccaac 3120

acacaataca ccaaacccca gctgctgttt aagcgcggga ac 3162

<210> 21
<211> 3171
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature

<223> 编码E2的DNA序列

<400> 21

[0070]

atgaaccaga aatttatcct gggtctggat attggtatta ccagcgttgg ttatggcctg 60

attgattacg aaaccaaaaa cattattgat gccggtgttc gtctgtttcc ggaagcaaat 120

gttgaaaata atgaaggtcg tcgtagcaaa cgtggtagcc gtcgtctgaa acgtcgtcgt 180

attcatcgtc tggatcgtgt taaacatctg ctggcagaat atgatctgct ggatctgacc 240

aatattccga aaagcaccaa tccgtatcag acccgtgtta aaggtctgaa tgaaaagctg 300

agcaaagatg aactggttat tgcactgctg catattgcaa aacgccgtgg cattcataac 360

gttgatgttg cagcagataa agaagaaacc gcaagcgata gcctgagcac caaagatcag 420

attaacaaaa acgccaaatt tctggaaagc cgctatgttt gtgaactgca gaaagaacgt 480

ctggaaaatg aaggtcatgt tcgtggtgtt gaaaatcgct ttctgacgaa agatattgtg 540

cgtgaggcca aaaaaatcat cgatacccag atgcagtatt acccggaaat tgatgaaacc 600

ttcaaagaga aatatatcag cctggttgaa acccgtcgcg aatattttga aggtccgggt 660

aaaggtagcc cgtttggttg ggaaggtaat atcaagaaat ggtttgagca gatgatgggc 720

cactgtacct attttccaga agaactgcgt agcgtcaaat atagctattc agccgaactg 780

tttaacgccc tgaatgatct gaataatctg gtgattaccc gtgatgaaga tgccaaactg 840

aattatggtg agaaattcca gatcatcgaa aacgtgttca aacagaagaa aacaccgaac 900

ctgaaacaaa tcgccattga aattggtgtg catgaaaccg aaatcaaagg ttatcgtgtg 960

aacaaaagcg gtacaccgga atttaccgaa tttaaactgt atcatgacct gaaaagcatc 1020

gtgttcgata aaagcattct ggaaaatgaa gccatcctgg atcagattgc agaaattctg 1080

accatctatc aggatgagca gagcattaaa gaggaactga ataaactgcc ggaaatactg 1140

aacgaacagg ataaagcaga aatcgccaaa ctgattggtt ataatggcac ccatcgtctg 1200

agcctgaaat gtattcacct gattaatgaa gaactgtggc agaccagccg taatcagatg 1260

gaaattttca actacctgaa catcaaaccg aacaaagtgg atctgagtga gcagaacaaa 1320

atcccgaaag atatggtgaa cgactttatt ctgagtccgg ttgtgaaacg caccttttatt 1380

cagagcatca acgtgatcaa caaagtgatc gagaaatatg gcatccccga agatatcatt 1440

atcgaactgg cacgtgaaaa taactccgat gatcgcaaaa agttcatcaa caacctgcag 1500

aaaaagaatg aagcaacccg caaacgcatt aacgaaatta ttggtcagac cggtaatcag 1560

aatgccaaac gtattgtgga aaaaatccgt ctgcatgatc agcaagaggg taaatgtctg 1620

tatagcctga aagatatccc gctggaagat ctgctgcgca atccgaacaa ttatgatatc 1680

gaccatatta ttccgcgaag cgtgagcttt gatgatagca tgcataacaa agttctggtt 1740

cgtcgcgaac agaatgccaa aaagaataat cagaccccgt atcagtatct gaccagtggt 1800

tatgcagata tcaaatacag cgtgtttaag cagcatgttc tgaatctggc cgaaaataaa 1860

gatcgcatga ccaaaaaaaa gcgcgagtat ctgctggaag aacgcgacat taacaaattt 1920

gaagtgcaga aagaatttat caaccgcaac ctggttgata cccgttatgc aacccgtgaa 1980

ctgaccaatt atctgaaagc atatttcagc gccaacaaca tgaacgtgaa agtgaaaacg 2040

attaacggca gctttaccga ttatctgcgt aaagtgtgga aattcaaaaa agaacgcaac 2100

cacggctata aacatcatgc cgaagatgcc ctgattattg caaatgcaga tttcctgttt 2160

aaagaaaaca aaaaactgaa agccgtcaac agcgtgctgg aaaaaccgga aattgagaca 2220

aaacagctgg acattcaggt tgatagcgaa gataattaca gcgaaatgtt tatcatcccg 2280

aaacaggtgc aggatatcaa agattttcgc aacttcaaat atagccaccg cgttgacaaa 2340

aaacctaatc gtcagctgat taacgatacc ctgtatagca cccgcaaaaa agataacagc 2400

acctatattg tgcagaccat taaagacatc tacgccaaag ataataccac cctgaaaaaa 2460

cagttcgaca aaagcccaga aaaatttctg atgtatcagc atgatccgcg taccttcgaa 2520

aaactggaag ttattatgaa acagtatgcc aacgagaaaa atccgctggc caaatatcac 2580

gaagaaaccg gtgaatatct gaccaaatat tccaagaaga acaacggtcc gatcgttaaa 2640

tccctgaaat atatcggtaa taaactgggc agccatctgg atgttaccca tcagtttaaa 2700

agctccacaa agaagctggt taaactgtcc atcaaaccgt atcgctttga tgtgtatctg 2760

[0071] accgacaaag gctataaatt cattaccatc agctatctgg acgtgctgaa aaaagacaac 2820

tattattata tcccggaaca gaaatatgat aaactgaaac tgggtaaagc catcgataaa 2880

aacgccaaat ttatcgccag cttctacaaa aacgacctga ttaaactgga tggcgagatc 2940

tataaaatca tcggtgttaa tagcgacacc cgcaatatga ttgagctgga tctgccggat 3000

attcgctata aagaatattg cgaactgaac aacattaaag gcgaaccgcg tatcaaaaag 3060

accatcggca aaaaagtgaa tagcatcgag aaactgacca ccgatgttct gggtaatgtg 3120

tttaccaata cccagtatac caaacctcag ctgctgttca aacgcggtaa t 3171

<210> 22
<211> 3171
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> 密码子优化的E2多核苷酸

<400> 22
atgaaccaaa agttcattct ggggctcgat atcggcatca cctccgtggg atatggtctg 60

atcgactacg agactaagaa catcatcgac gctggagtgc gactgttccc ggaagcgaac 120

[0072]

gtggagaaca acgaaggccg cagatccaag cgcgggtcca gaaggctcaa gaggcggagg 180

atccatagac tcgacagagt gaagcacctc cttgccgaat acgatctgtt ggaccttacc 240

aacattccca agagcaccaa cccgtaccaa accagagtga agggcctgaa cgaaaagctg 300

tcgaaagatg aactggtcat tgccctgctg catattgcca aacggcgcgg aatccataac 360

gtggacgtgg ccgctgacaa ggaagagact gcgtccgact cgctgtcaac caaggaccag 420

atcaacaaga acgccaaatt cctggaaagc cgctacgtct gcgagcttca aaaagaacgg 480

ctggagaacg agggacacgt caggggagtg gagaaccggt tcctgaccaa ggacatcgtg 540

cgggaagcca agaagatcat cgacacccaa atgcagtatt atccggaaat tgatgaaact 600

tttaaggaga agtacatttc cctggtggaa actcggaggg agtacttcga gggacctgga 660

aagggatccc ctttcggctg ggaagggaac attaagaagt ggtttgaaca gatgatgggc 720

cattgcactt actttccgga agaactccgg tccgtgaagt actcatactc tgccgagctg 780

ttcaatgcac tcaacgacct taacaacttg gtgatcaccc gcgatgaaga tgccaagttg 840

aactacggag aaaagttcca gatcatcgag aacgtgttca agcagaaaaa gaccccaaat 900

ctgaagcaga ttgccatcga aattggcgtg cacgagactg agatcaaggg ataccgggtc 960

aacaagtccg gcacgccaga gttcaccgag ttcaagctgt accacgatct gaagtcgatc 1020

gtgtttgaca agtccatcct ggaaaacgaa gccattctgg accagattgc tgagatcctg 1080

accatctacc aggacgagca atcgattaag gaagaactga acaagctccc cgagattctg 1140

aacgaacagg ataaggccga gatcgcgaag ctcattggtt acaatggtac ccaccgcttg 1200

tcccttaagt gcatccatct gatcaatgag gaactgtggc agaccagccg gaaccagatg 1260

gagatcttca attacttgaa catcaagccg aacaaggtgg acctgtccga acagaacaag 1320

atacccaagg acatggtcaa cgactttatc ctctcaccgg tggtcaagcg caccttcatt 1380

caatctatca acgtgatcaa caaggtcatc gagaagtacg gcattcctga ggatatcatc 1440

atcgagctgg ctcgggagaa caactcagac gataggaaga agttcattaa caacctccag 1500

aaaaagaacg aggccactcg caagcggatt aatgagatca tcggtcagac cgggaaccag 1560

aacgccaagc ggatcgtgga aaagattcgg ctccacgacc aacaggaggg aaagtgtctg 1620

tactcgctga aggacattcc cctggaggac ctcctgagga acccaaacaa ctacgacatc 1680

gatcacataa tcccccgcag cgtgtcattc gacgattcca tgcataacaa ggtcctcgtg 1740

cggagagagc agaatgccaa gaagaacaac cagactccgt accagtacct gacgtccggc 1800

tacgcagaca tcaagtactc agtgttcaaa cagcacgtgc tcaacctggc cgagaacaag 1860

gacaggatga ccaagaagaa gcgcgaatac cttctcgagg aacgggatat caataagttc 1920

gaggtgcaga aggagtttat caatagaaac ctggtggaca ctcgctatgc cacccgcgaa 1980

ctgaccaact acctgaaggc gtacttctcc gccaacaaca tgaacgtgaa ggtcaaaact 2040

attaacggca gcttcaccga ctatctgcgc aaggtctgga agttcaagaa ggaacgcaac 2100

cacggttaca agcaccacgc ggaagatgcg ctgattatcg ccaacgctga cttcctgttc 2160

aaggaaaaca agaagctcaa ggccgtgaac tcagtgctcg agaagcctga aatcgagact 2220

aagcagctgg acatccaggt cgattcggaa gataactact ccgaaatgtt catcatccct 2280

aagcaagtgc aggacatcaa ggacttcagg aatttcaagt acagccatcg cgtggacaag 2340

aagccaaaca gacagctgat caacgataca ctgtattcca cccggaagaa ggacaactcc 2400

acctacatcg tccaaaccat taaggacatc tacgcaaagg acaacaccac gcttaagaag 2460

cagttcgaca agagccccga aaagttcctc atgtaccagc acgaccccag aaccttcgag 2520

aagcttgaag tgatcatgaa gcagtacgcc aacgaaaaga acccactggc taagtaccac 2580

gaggaaaccg gcgaatacct gaccaagtac tccaaaaaga acaacggacc gatcgtcaag 2640

tccctgaagt acattgggaa caagctcggc tcgcacctcg atgtgaccca ccagttcaag 2700

tcctcgacca aaaagctcgt gaagctgtcc atcaagccgt accggttcga cgtgtacctg 2760

actgacaagg gatataagtt catcaccatt tcctacctcg acgtgttgaa gaaggataac 2820

tactactaca ttccggaaca gaagtacgac aagctcaagc tcggaaaggc catcgacaaa 2880

aatgcgaagt tcatcgcgag cttctacaag aatgacttga tcaagctgga tggcgaaatc 2940

tacaagatca tcggggtcaa ctccgatacc cgcaacatga ttgagctgga tctgcccgac 3000

[0073] attcggtaca aggaatactg cgagctgaac aacatcaagg gagaaccgcg gatcaagaaa 3060

accatcggaa agaaagtgaa cagcatcgag aaactgacta ctgacgtcct gggaaacgtg 3120

ttcaccaaca cacaatacac caaaccccag ctgctgttta agcgcgggaa c 3171

<210> 23
<211> 3171
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> 编码E2-K741D-L743K的DNA序列

<400> 23
atgaaccaga aatttatcct gggtctggat attggtatta ccagcgttgg ttatggcctg 60

attgattacg aaaccaaaaa cattattgat gccggtgttc gtctgtttcc ggaagcaaat 120

gttgaaaata atgaaggtcg tcgtagcaaa cgtggtagcc gtcgtctgaa acgtcgtcgt 180

attcatcgtc tggatcgtgt taaacatctg ctggcagaat atgatctgct ggatctgacc 240

aatattccga aaagcaccaa tccgtatcag acccgtgtta aaggtctgaa tgaaaagctg 300

agcaaagatg aactggttat tgcactgctg catattgcaa aacgccgtgg cattcataac 360

[0074]

gttgatgttg cagcagataa agaagaaacc gcaagcgata gcctgagcac caaagatcag 420

attaacaaaa acgccaaatt tctggaaagc cgctatgttt gtgaactgca gaaagaacgt 480

ctggaaaatg aaggtcatgt tcgtggtgtt gaaaatcgct ttctgacgaa agatattgtg 540

cgtgaggcca aaaaaatcat cgatacccag atgcagtatt acccggaaat tgatgaaacc 600

ttcaaagaga aatatatcag cctggttgaa acccgtcgcg aatattttga aggtccgggt 660

aaaggtagcc cgtttggttg ggaaggtaat atcaagaaat ggtttgagca gatgatgggc 720

cactgtacct attttccaga agaactgcgt agcgtcaaat atagctattc agccgaactg 780

tttaacgccc tgaatgatct gaataatctg gtgattaccc gtgatgaaga tgccaaactg 840

aattatggtg agaaattcca gatcatcgaa aacgtgttca aacagaagaa aacaccgaac 900

ctgaaacaaa tcgccattga aattggtgtg catgaaaccg aaatcaaagg ttatcgtgtg 960

aacaaaagcg gtacaccgga atttaccgaa tttaaactgt atcatgacct gaaaagcatc 1020

gtgttcgata aaagcattct ggaaaatgaa gccatcctgg atcagattgc agaaattctg 1080

accatctatc aggatgagca gagcattaaa gaggaactga ataaactgcc ggaaatactg 1140

aacgaacagg ataaagcaga aatcgccaaa ctgattggtt ataatggcac ccatcgtctg 1200

agcctgaaat gtattcacct gattaatgaa gaactgtggc agaccagccg taatcagatg 1260

gaaattttca actacctgaa catcaaaccg aacaaagtgg atctgagtga gcagaacaaa 1320

atcccgaaag atatggtgaa cgactttatt ctgagtccgg ttgtgaaacg caccttttatt 1380

cagagcatca acgtgatcaa caaagtgatc gagaaatatg gcatccccga agatatcatt 1440

atcgaactgg cacgtgaaaa taactccgat gatcgcaaaa agttcatcaa caacctgcag 1500

aaaaagaatg aagcaacccg caaacgcatt aacgaaatta ttggtcagac cggtaatcag 1560

aatgccaaac gtattgtgga aaaaatccgt ctgcatgatc agcaagaggg taaatgtctg 1620

tatagcctga aagatatccc gctggaagat ctgctgcgca atccgaacaa ttatgatatc 1680

gaccatatta ttccgcgaag cgtgagcttt gatgatagca tgcataacaa agttctggtt 1740

cgtcgcgaac agaatgccaa aaagaataat cagaccccgt atcagtatct gaccagtggt 1800

tatgcagata tcaaatacag cgtgtttaag cagcatgttc tgaatctggc cgaaaataaa 1860

gatcgcatga ccaaaaaaaa gcgcgagtat ctgctggaag aacgcgacat taacaaattt 1920

gaagtgcaga aagaatttat caaccgcaac ctggttgata cccgttatgc aacccgtgaa 1980

ctgaccaatt atctgaaagc atatttcagc gccaacaaca tgaacgtgaa agtgaaaacg 2040

attaacggca gctttaccga ttatctgcgt aaagtgtgga aattcaaaaa agaacgcaac 2100

cacggctata aacatcatgc cgaagatgcc ctgattattg caaatgcaga tttcctgttt 2160

aaagaaaaca aaaaactgaa agccgtcaac agcgtgctgg aaaaaccgga aattgagaca 2220

gatcagaaag acattcaggt tgatagcgaa gataattaca gcgaaatgtt tatcatcccg 2280

aaacaggtgc aggatatcaa agattttcgc aacttcaaat atagccaccg cgttgacaaa 2340

aaacctaatc gtcagctgat taacgatacc ctgtatagca cccgcaaaaa agataacagc 2400

acctatattg tgcagaccat taaagacatc tacgccaaag ataataccac cctgaaaaaa 2460

cagttcgaca aaagcccaga aaaatttctg atgtatcagc atgatccgcg taccttcgaa 2520

aaactggaag ttattatgaa acagtatgcc aacgagaaaa atccgctggc caaatatcac 2580

gaagaaaccg gtgaatatct gaccaaatat tccaagaaga acaacggtcc gatcgttaaa 2640

tccctgaaat atatcggtaa taaactgggc agccatctgg atgttaccca tcagtttaaa 2700

agctccacaa agaagctggt taaactgtcc atcaaaccgt atcgctttga tgtgtatctg 2760

accgacaaag gctataaatt cattaccatc agctatctgg acgtgctgaa aaaagacaac 2820

tattattata tcccggaaca gaaatatgat aaactgaaac tgggtaaagc catcgataaa 2880

aacgccaaat ttatcgccag cttctacaaa aacgacctga ttaaactgga tggcgagatc 2940

tataaaatca tcggtgttaa tagcgacacc cgcaatatga ttgagctgga tctgccggat 3000

attcgctata aagaatattg cgaactgaac aacattaaag gcgaaccgcg tatcaaaaag 3060

accatcggca aaaaagtgaa tagcatcgag aaactgacca ccgatgttct gggtaatgtg 3120

tttaccaata cccagtatac caaacctcag ctgctgttca aacgcggtaa t 3171

[0075]

<210> 24
<211> 3171
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> 编码E2-S670T-N675D的DNA序列

<400> 24
atgaaccaga aatttatcct gggtctggat attggtatta ccagcgttgg ttatggcctg 60

attgattacg aaaccaaaaa cattattgat gccggtgttc gtctgtttcc ggaagcaaat 120

gttgaaaata atgaaggtcg tcgtagcaaa cgtggtagcc gtcgtctgaa acgtcgtcgt 180

attcatcgtc tggatcgtgt taaacatctg ctggcagaat atgatctgct ggatctgacc 240

aatattccga aaagcaccaa tccgtatcag acccgtgtta aaggtctgaa tgaaaagctg 300

agcaaagatg aactggttat tgcactgctg catattgcaa aacgccgtgg cattcataac 360

gttgatgttg cagcagataa agaagaaacc gcaagcgata gcctgagcac caaagatcag 420

attaacaaaa acgccaaatt tctggaaagc cgctatgttt gtgaactgca gaaagaacgt 480

ctggaaaatg aaggtcatgt tcgtggtgtt gaaaatcgct ttctgacgaa agatattgtg 540

[0076]

cgtgaggcca aaaaaatcat cgatacccag atgcagtatt acccggaaat tgatgaaacc 600

ttcaaagaga aatatatcag cctggttgaa acccgtcgcg aatattttga aggtccgggt 660

aaaggtagcc cgtttggttg ggaaggtaat atcaagaaat ggtttgagca gatgatgggc 720

cactgtacct attttccaga agaactgcgt agcgtcaaat atagctattc agccgaactg 780

tttaacgccc tgaatgatct gaataatctg gtgattaccc gtgatgaaga tgccaaactg 840

aattatggtg agaaattcca gatcatcgaa aacgtgttca aacagaagaa aacaccgaac 900

ctgaaacaaa tcgccattga aattggtgtg catgaaaccg aaatcaaagg ttatcgtgtg 960

aacaaaagcg gtacaccgga atttaccgaa tttaaactgt atcatgacct gaaaagcatc 1020

gtgttcgata aaagcattct ggaaaatgaa gccatcctgg atcagattgc agaaattctg 1080

accatctatc aggatgagca gagcattaaa gaggaactga ataaactgcc ggaaatactg 1140

aacgaacagg ataaagcaga aatcgccaaa ctgattggtt ataatggcac ccatcgtctg 1200

agcctgaaat gtattcacct gattaatgaa gaactgtggc agaccagccg taatcagatg 1260

gaaattttca actacctgaa catcaaaccg aacaaagtgg atctgagtga gcagaacaaa 1320

atcccgaaag atatggtgaa cgactttatt ctgagtccgg ttgtgaaacg cacctttatt 1380

cagagcatca acgtgatcaa caaagtgatc gagaaatatg gcatccccga agatatcatt 1440

atcgaactgg cacgtgaaaa taactccgat gatcgcaaaa agttcatcaa caacctgcag 1500

aaaaagaatg aagcaacccg caaacgcatt aacgaaatta ttggtcagac cggtaatcag 1560

aatgccaaac gtattgtgga aaaaatccgt ctgcatgatc agcaagaggg taaatgtctg 1620

tatagcctga aagatatccc gctggaagat ctgctgcgca atccgaacaa ttatgatatc 1680

gaccatatta ttccgcgaag cgtgagcttt gatgatagca tgcataacaa agttctggtt 1740

cgtcgcgaac agaatgccaa aaagaataat cagaccccgt atcagtatct gaccagtggt 1800

tatgcagata tcaaatacag cgtgtttaag cagcatgttc tgaatctggc cgaaaataaa 1860

gatcgcatga ccaaaaaaaa gcgcgagtat ctgctggaag aacgcgacat taacaaattt 1920

gaagtgcaga aagaatttat caaccgcaac ctggttgata cccgttatgc aacccgtgaa 1980

ctgaccaatt atctgaaagc atatttcacc gccaacaaca tggatgtgaa agtgaaaacg 2040

attaacggca gctttaccga ttatctgcgt aaagtgtgga aattcaaaaa agaacgcaac 2100

cacggctata aacatcatgc cgaagatgcc ctgattattg caaatgcaga tttcctgttt 2160

aaagaaaaca aaaaactgaa agccgtcaac agcgtgctgg aaaaaccgga attgagaca 2220

aaacagctgg acattcaggt tgatagcgaa gataattaca gcgaaatgtt tatcatcccg 2280

aaacaggtgc aggatatcaa agattttcgc aacttcaaat atagccaccg cgttgacaaa 2340

aaacctaatc gtcagctgat taacgatacc ctgtatagca cccgcaaaaa agataacagc 2400

acctatattg tgcagaccat taaagacatc tacgccaaag ataataccac cctgaaaaaa 2460

cagttcgaca aaagcccaga aaaatttctg atgtatcagc atgatccgcg taccttcgaa 2520

aaactggaag ttattatgaa acagtatgcc aacgagaaaa atccgctggc caaatatcac 2580

gaagaaaccg gtgaatatct gaccaaatat tccaagaaga acaacggtcc gatcgttaaa 2640

tccctgaaat atatcggtaa taaactgggc agccatctgg atgttaccca tcagtttaaa 2700

agctccacaa agaagctggt taaactgtcc atcaaaccgt atcgctttga tgtgtatctg 2760

accgacaaag gctataaatt cattaccatc agctatctgg acgtgctgaa aaaagacaac 2820

tattattata tcccggaaca gaaatatgat aaactgaaac tgggtaaagc catcgataaa 2880

aacgccaaat ttatcgccag cttctacaaa aacgacctga ttaaactgga tggcgagatc 2940

tataaaatca tcggtgttaa tagcgacacc cgcaatatga ttgagctgga tctgccggat 3000

attcgctata aagaatattg cgaactgaac aacattaaag gcgaaccgcg tatcaaaaag 3060

accatcggca aaaaagtgaa tagcatcgag aaactgacca ccgatgttct gggtaatgtg 3120

tttaccaata cccagtatac caaacctcag ctgctgttca aacgcggtaa t 3171

<210> 25
<211> 3171
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

[0077]

<220>
<221> misc_feature
<223> 编码E2-K741N-L743N的DNA序列

<400> 25
atgaaccaga aatttatcct gggtctggat attggtatta ccagcgttgg ttatggcctg 60

attgattacg aaaccaaaaa cattattgat gccggtgttc gtctgtttcc ggaagcaaat 120

gttgaaaata atgaaggtcg tcgtagcaaa cgtggtagcc gtcgtctgaa acgtcgtcgt 180

attcatcgtc tggatcgtgt taaacatctg ctggcagaat atgatctgct ggatctgacc 240

aatattccga aaagcaccaa tccgtatcag acccgtgtta aaggtctgaa tgaaaagctg 300

agcaaagatg aactggttat tgcactgctg catattgcaa aacgccgtgg cattcataac 360

gttgatgttg cagcagataa agaagaaacc gcaagcgata gcctgagcac caaagatcag 420

attaacaaaa acgccaaatt tctggaaagc cgctatgttt gtgaactgca gaaagaacgt 480

ctggaaaatg aaggtcatgt tcgtggtgtt gaaaatcgct ttctgacgaa agatattgtg 540

cgtgaggcca aaaaaatcat cgatacccag atgcagtatt acccggaaat tgatgaaacc 600

ttcaaagaga aatatatcag cctggttgaa acccgtcgcg aatattttga aggtccgggt 660

aaaggtagcc cgtttggttg ggaaggtaat atcaagaaat ggtttgagca gatgatgggc 720

cactgtacct attttccaga agaactgcgt agcgtcaaat atagctattc agccgaactg 780

[0078]

tttaacgccc tgaatgatct gaataatctg gtgattaccc gtgatgaaga tgccaaactg 840

aattatggtg agaaattcca gatcatcgaa aacgtgttca aacagaagaa aacaccgaac 900

ctgaaacaaa tcgccattga aattggtgtg catgaaaccg aaatcaaagg ttatcgtgtg 960

aacaaaagcg gtacaccgga atttaccgaa tttaaactgt atcatgacct gaaaagcatc 1020

gtgttcgata aaagcattct ggaaaatgaa gccatcctgg atcagattgc agaaattctg 1080

accatctatc aggatgagca gagcattaaa gaggaactga ataaactgcc ggaaatactg 1140

aacgaacagg ataaagcaga aatcgccaaa ctgattggtt ataatggcac ccatcgtctg 1200

agcctgaaat gtattcacct gattaatgaa gaactgtggc agaccagccg taatcagatg 1260

gaaattttca actacctgaa catcaaaccg aacaaagtgg atctgagtga gcagaacaaa 1320

atcccgaaag atatggtgaa cgactttatt ctgagtccgg ttgtgaaacg cacctttatt 1380

cagagcatca acgtgatcaa caaagtgatc gagaaatatg gcatccccga agatatcatt 1440

atcgaactgg cacgtgaaaa taactccgat gatcgcaaaa agttcatcaa caacctgcag 1500

aaaaagaatg aagcaacccg caaacgcatt aacgaaatta ttggtcagac cggtaatcag 1560

aatgccaaac gtattgtgga aaaaatccgt ctgcatgatc agcaagaggg taaatgtctg 1620

tatagcctga aagatatccc gctggaagat ctgctgcgca atccgaacaa ttatgatatc 1680

gaccatatta ttccgcgaag cgtgagcttt gatgatagca tgcataacaa agttctggtt 1740

cgtcgcgaac agaatgccaa aaagaataat cagaccccgt atcagtatct gaccagtggt 1800

tatgcagata tcaaatacag cgtgtttaag cagcatgttc tgaatctggc cgaaaataaa 1860

gatcgcatga ccaaaaaaaa gcgcgagtat ctgctggaag aacgcgacat taacaaattt 1920

gaagtgcaga aagaatttat caaccgcaac ctggttgata cccgttatgc aacccgtgaa 1980

ctgaccaatt atctgaaagc atatttcagc gccaacaaca tgaacgtgaa agtgaaaacg 2040

attaacggca gctttaccga ttatctgcgt aaagtgtgga aattcaaaaa agaacgcaac 2100

cacggctata aacatcatgc cgaagatgcc ctgattattg caaatgcaga tttcctgttt 2160

aaagaaaaca aaaaactgaa agccgtcaac agcgtgctgg aaaaaccgga aattgagaca 2220

aatcagaatg acattcaggt tgatagcgaa gataattaca gcgaaatgtt tatcatcccg 2280

aaacaggtgc aggatatcaa agattttcgc aacttcaaat atagccaccg cgttgacaaa 2340

aaacctaatc gtcagctgat taacgatacc ctgtatagca cccgcaaaaa agataacagc 2400

acctatattg tgcagaccat taaagacatc tacgccaaag ataataccac cctgaaaaaa 2460

cagttcgaca aaagcccaga aaaatttctg atgtatcagc atgatccgcg taccttcgaa 2520

aaactggaag ttattatgaa acagtatgcc aacgagaaaa atccgctggc caaatatcac 2580

gaagaaaccg gtgaatatct gaccaaatat tccaagaaga acaacggtcc gatcgttaaa 2640

tccctgaaat atatcggtaa taaactgggc agccatctgg atgttaccca tcagtttaaa 2700

agctccacaa agaagctggt taaactgtcc atcaaaccgt atcgctttga tgtgtatctg 2760

accgacaaag gctataaatt cattaccatc agctatctgg acgtgctgaa aaaagacaac 2820

tattattata tcccggaaca gaaatatgat aaactgaaac tgggtaaagc catcgataaa 2880

aacgccaaat ttatcgccag cttctacaaa aacgacctga ttaaactgga tggcgagatc 2940

tataaaatca tcggtgttaa tagcgacacc cgcaatatga ttgagctgga tctgccggat 3000

attcgctata aagaatattg cgaactgaac aacattaaag gcgaaccgcg tatcaaaaag 3060

accatcggca aaaaagtgaa tagcatcgag aaactgacca ccgatgttct gggtaatgtg 3120

tttaccaata cccagtatac caaacctcag ctgctgttca aacgcggtaa t 3171

<210> 26
<211> 3159
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> 编码F8的DNA序列

<400> 26

[0079]

atgaaccaga aatttatcct gggtctggat attggtatta ccagcgttgg ttatggcctg 60

attgattacg aaaccaaaaa cattattgat gccggtgttc gtctgtttcc ggaagcaaat 120

gttgaaaata atgaaggtcg tcgtagcaaa cgtggtagcc gtcgtctgaa acgtcgtcgt 180

attcatcgtc tggaacgtgt taaaagcctg ctgagcgaat ataagattat tagcggtctg 240

gcaccgacca ataatcagcc gtataacatt cgtgttaaag gtctgaccga acagctgacc 300

aaagatgaac tggcagttgc actgctgcat attgccaaac gccgtggcat tcataaaatc 360

gatgtgattg atagcaatga cgatgtgggt aatgaactga gcaccaaaga acagctgaac 420

aaaaatagca aactgctgaa agacaaattc gtgtgtcaga ttcagctgga acgtatgaat 480

gaaggccagg ttcgtggtga aaagaatcgc tttaaaaccg cagacatcat caaagaaatt 540

atccagctgc tgaacgtgca gaaaaacttc catcagctgg atgaaaactt catcaacaaa 600

tacatcgagc tggttgaaat gcgtcgcgaa tattttgaag gtcctggtca gggtagcccg 660

tttggttgga atggtgatct gaaaaaatgg tacgaaatgc tgatgggtca ctgtacctat 720

tttccgcaag aactgcgtag cgttaaatat gcctatagcg cagacctgtt taatgcactg 780

aatgatctga acaacctgat tattcagcgc gataatagcg agaaactgga ataccatgag 840

aagtatcaca tcatcgagaa cgtgttcaag cagaaaaaaa agccgacgct gaaacaaatc 900

gcaaaagaga ttggcgttaa cccggaagat attaaaggtt atcgtattac caaaagcggt 960

[0080]

acaccggaat tcaccgaatt taaactgtat cacgatctga aaagcgtgct gtttgatcag 1020

agcattctgg aaaatgaaga tgtgctggac cagattgcag aaattctgac catttatcag 1080

gacaaagaca gcatcaaaag caaactgacc gaactggata ttctgctgaa tgaagaagat 1140

aaagagaaca ttgcacagct gaccggttat aacggcacac atcgcctgag cctgaaatgt 1200

attcgtctgg tactggaaga acagtggtat agcagccgta atcagatgga aatctttacc 1260

catctgaaca ttaaaccgaa gaaaatcaat ctgaccgcag ccaacaaaat tccgaaagcc 1320

atgattgatg agtttattct gagtccggtt gtgaaacgca cctttattca gagcatcaac 1380

gtgatcaaca aagtgatcga gaaatatggc atccccgaag atatcattat cgaactggca 1440

cgtgaaaata actccgatga tcgcaaaaag ttcatcaaca acctgcagaa aaagaatgaa 1500

gcaacccgca aacgcattaa cgaaattatt ggtcagaccg gtaatcagaa tgccaaacgt 1560

attgtggaaa aaatccgtct gcatgatcag caagagggga aatgtctgta tagcctggaa 1620

agcattgccc tgatggatct gctgaataac ccgcagaatt atgaagtgga tcatattatt 1680

ccgcgtagcg tggcatttga taattccatt cataacaaag tgctggtgaa gcagatcgag 1740

aatagcaaaa aaggtaatcg tacgccgtat cagtatctga atagcagtga tgcaaaactg 1800

agctacaacc agtttaaaca gcatattctg aatctgagca aaagcaaaga tcgcatcagc 1860

aaaaaaaaga aggactacct gctggaagaa cgcgatatca acaaatttga agtccagaaa 1920

gagtttatca accgcaatct ggttgatacc cgttatgcaa cccgtgaact gaccagctat 1980

ctgaaagcat atttcagcgc caataacatg gacgtgaaag tgaaaacaat taacggcagc 2040

tttaccaacc atctgcgtaa agtttggcgc tttgataaat atcgcaacca cggctataaa 2100

catcatgcag aagatgccct gattattgca aatgcagatt tcctgtttaa agaaaacaaa 2160

aaactgaaag ccgtcaacag cgtgctggaa aaaccggaaa ttgagacaaa acagctggac 2220

attcaggttg atagcgaaga taattacagc gaaatgttta tcatcccgaa acaggtgcag 2280

gatatcaaag attttcgcaa cttcaaatat agccaccgcg ttgacaaaaa acctaatcgt 2340

cagctgatta acgataccct gtatagcacc cgcaaaaaag ataacagcac ctatattgtg 2400

cagaccatta aagacatcta cgccaaagat aataccaccc tgaaaaaaca gttcgacaaa 2460

agcccagaaa aatttctgat gtatcagcat gatccgcgta ccttcgaaaa actggaagtt 2520

attatgaaac agtatgccaa cgagaaaaat ccgctggcca aatatcacga agaaaccggt 2580

gaatatctga ccaaatattc caagaagaac aacggtccga tcgttaaatc cctgaaatat 2640

atcggtaata aactgggcag ccatctggat gttacccatc agtttaaaag ctccacaaag 2700

aagctggtta aactgtccat caaaccgtat cgctttgatg tgtatctgac cgacaaaggc 2760

tataaattca ttaccatcag ctatctggac gtgctgaaaa aagacaacta ttattatatc 2820

ccggaacaga aatatgataa actgaaactg ggtaaagcca tcgataaaaa cgccaaattt 2880

atcgccagct tctacaaaaa cgacctgatt aaactggatg gcgagatcta taaaatcatc 2940

ggtgttaata gcgacacccg caatatgatt gagctggatc tgccggatat tcgctataaa 3000

gaatattgcg aactgaacaa cattaaaggc gaaccgcgta tcaaaaagac catcggcaaa 3060

aaagtgaata gcatcgagaa actgaccacc gatgttctgg gtaatgtgtt taccaatacc 3120

cagtatacca aacctcagct gctgttcaaa cgcggtaat 3159

<210> 27
<211> 3159
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> 密码子优化的F8多核苷酸

<400> 27

[0081]

atgaaccaaa agttcattct ggggctcgat atcggcatca cctccgtggg atatggtctg 60

atcgactacg agactaagaa catcatcgac gctggagtgc gactgttccc ggaagcgaac 120

gtggagaaca acgaaggccg cagatccaag cgcgggtcca gaaggctcaa gaggcggagg 180

atccatagac tcgagagagt gaagtcgctc ctttcggaat acaagattat cagcggtctt 240

gcccccacca acaaccaacc gtacaacatc agagtgaagg gcctgaccga acagctgacc 300

aaagatgaac tggccgtcgc cctgctgcat attgccaaac ggcgcggaat ccataagatc 360

gacgtgattg acagcaacga tgacgtggga aacgagctgt caaccaagga acagcttaac 420

aagaacagca aattgctgaa ggacaagttt gtctgccaaa ttcaactgga acggatgaac 480

gagggacaag tcaggggaga gaaaaaccgg ttcaagaccg ccgacatcat caaggagatc 540

atccaactgc tgaacgtgca gaagaacttc caccaactgg atgaaaactt cattaacaag 600

tacattgaac tggtggaaat gcggagggag tacttcgagg gacctggaca gggatcccct 660

ttcggctgga atggggacct taagaagtgg tacgaaatgt tgatgggcca ttgcacttac 720

tttccgcaag aactccggtc cgtgaagtac gcatactctg ccgacctgtt caatgcactc 780

aacgacctta acaacttgat catccagcgc gataactcgg aaaagttgga ataccacgaa 840

aagtatcaca tcatcgagaa cgtgttcaag cagaaaaaga agccaactct gaagcagatt 900

gccaaggaaa ttggcgtgaa tccggaggat atcaagggat accggatcac taagtccggc 960

acgccagagt tcaccgagtt caagctgtac cacgatctga agtcggtgct ctttgaccag 1020

tccatcctgg aaaacgaaga tgtgctggac cagattgctg agatcctgac catctaccag 1080

gacaaggact cgattaagtc caagctcacc gagctggaca ttctgctgaa cgaagaagat 1140

aaggagaaca tcgcgcagct caccggttac aatggtaccc accgcttgtc ccttaagtgc 1200

[0082]

atccgcctgg tgctggagga acagtggtac tcgagccgga accagatgga gatcttcact 1260

cacttgaaca tcaagccgaa aaagattaac ctgactgccg ccaacaagat acccaaggcc 1320

atgatcgacg agtttatcct ctcaccggtg gtcaagcgca ccttcattca atctatcaac 1380

gtgatcaaca aggtcatcga gaagtacggc attcctgagg atatcatcat cgagctggct 1440

cgggagaaca actcagacga taggaagaag ttcattaaca acctccagaa aaagaacgag 1500

gccactcgca agcggattaa tgagatcatc ggtcagaccg ggaaccagaa cgccaagcgg 1560

atcgtggaaa agattcggct ccacgaccaa caggagggaa agtgtctgta ctcgctggag 1620

tcgattgcac tgatggacct cctgaacaac ccacagaact acgaagtcga tcacataatc 1680

ccccgcagcg tggcattcga caactccatc cataacaagg tcctcgtgaa gcagatcgag 1740

aatagcaaga aggggaaccg gactccgtac cagtacctga actcctccga cgccaagctg 1800

tcatacaatc agttcaaaca gcacattctc aacctgtcca agtcaaagga caggatctcc 1860

aagaagaaga aggactacct tctcgaggaa cgggatatca ataagttcga ggtgcagaag 1920

gagtttatca atagaaacct ggtggacact cgctatgcca cccgcgaact gaccagctac 1980

ctgaaggcgt acttctccgc caacaacatg gacgtgaagg tcaaaactat taacggcagc 2040

ttcaccaacc atctgcgcaa ggtctggagg ttcgacaagt accgcaacca cggttacaag 2100

caccacgcgg aagatgcgct gattatcgcc aacgctgact tcctgttcaa ggaaaacaag 2160

aagctcaagg ccgtgaactc agtgctcgag aagcctgaaa tcgagactaa gcagctggac 2220

atccaggtcg attcggaaga taactactcc gaaatgttca tcatccctaa gcaagtgcag 2280

gacatcaagg acttcaggaa tttcaagtac agccatcgcg tggacaagaa gccaaacaga 2340

cagctgatca acgatacact gtattccacc cggaagaagg acaactccac ctacatcgtc 2400

caaaccatta aggacatcta cgcaaaggac aacaccacgc ttaagaagca gttcgacaag 2460

agccccgaaa agttcctcat gtaccagcac gaccccagaa ccttcgagaa gcttgaagtg 2520

atcatgaagc agtacgccaa cgaaaagaac ccactggcta agtaccacga ggaaaccggc 2580

gaatacctga ccaagtactc caaaaagaac aacggaccga tcgtcaagtc cctgaagtac 2640

attgggaaca agctcggctc gcacctcgat gtgacccacc agttcaagtc ctcgaccaaa 2700

aagctcgtga agctgtccat caagccgtac cggttcgacg tgtacctgac tgacaaggga 2760

tataagttca tcaccatttc ctacctcgac gtgttgaaga aggataacta ctactacatt 2820

ccggaacaga agtacgacaa gctcaagctc ggaaaggcca tcgacaaaaa tgcgaagttc 2880

atcgcgagct tctacaagaa tgacttgatc aagctggatg gcgaaatcta caagatcatc 2940

ggggtcaact ccgatacccg caacatgatt gagctggatc tgcccgacat tcggtacaag 3000

gaatactgcg agctgaacaa catcaaggga gaaccgcgga tcaagaaaac catcggaaag 3060

aaagtgaaca gcatcgagaa actgactact gacgtcctgg gaaacgtgtt caccaacaca 3120

caatacacca aaccccagct gctgtttaag cgcgggaac 3159

<210> 28
<211> 3159
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> 编码F8-K737D-L739K的DNA序列

<400> 28

[0083]

atgaaccaga aatttatcct gggtctggat attggtatta ccagcgttgg ttatggcctg 60

attgattacg aaaccaaaaa cattattgat gccggtgttc gtctgtttcc ggaagcaaat 120

gttgaaaata atgaaggtcg tcgtagcaaa cgtggtagcc gtcgtctgaa acgtcgtcgt 180

attcatcgtc tggaacgtgt taaaagcctg ctgagcgaat ataagattat tagcggtctg 240

gcaccgacca ataatcagcc gtataacatt cgtgttaaag gtctgaccga acagctgacc 300

aaagatgaac tggcagttgc actgctgcat attgccaaac gccgtggcat tcataaaatc 360

gatgtgattg atagcaatga cgatgtgggt aatgaactga gcaccaaaga acagctgaac 420

aaaaatagca aactgctgaa agacaaattc gtgtgtcaga ttcagctgga acgtatgaat 480

gaaggccagg ttcgtggtga aaagaatcgc tttaaaaccg cagacatcat caaagaaatt 540

atccagctgc tgaacgtgca gaaaaacttc catcagctgg atgaaaactt catcaacaaa 600

tacatcgagc tggttgaaat gcgtcgcgaa tattttgaag gtcctggtca gggtagcccg 660

tttggttgga atggtgatct gaaaaaatgg tacgaaatgc tgatgggtca ctgtacctat 720

tttccgcaag aactgcgtag cgttaaatat gcctatagcg cagacctgtt taatgcactg 780

aatgatctga acaacctgat tattcagcgc gataatagcg agaaactgga ataccatgag 840

aagtatcaca tcatcgagaa cgtgttcaag cagaaaaaaa agccgacgct gaaacaaatc 900

gcaaaagaga ttggcgttaa cccggaagat attaaaggtt atcgtattac caaaagcggt 960

acaccggaat tcaccgaatt taaactgtat cacgatctga aaagcgtgct gtttgatcag 1020

agcattctgg aaaatgaaga tgtgctggac cagattgcag aaattctgac catttatcag 1080

gacaaagaca gcatcaaaag caaactgacc gaactggata ttctgctgaa tgaagaagat 1140

aaagagaaca ttgcacagct gaccggttat aacggcacac atcgcctgag cctgaaatgt 1200

attcgtctgg tactggaaga acagtggtat agcagccgta atcagatgga aatctttacc 1260

catctgaaca ttaaaccgaa gaaaatcaat ctgaccgcag ccaacaaaat tccgaaagcc 1320

atgattgatg agtttattct gagtccggtt gtgaaacgca cctttattca gagcatcaac 1380

[0084]

gtgatcaaca aagtgatcga gaaatatggc atccccgaag atatcattat cgaactggca 1440

cgtgaaaata actccgatga tcgcaaaaag ttcatcaaca acctgcagaa aaagaatgaa 1500

gcaacccgca aacgcattaa cgaaattatt ggtcagaccg gtaatcagaa tgccaaacgt 1560

attgtggaaa aaatccgtct gcatgatcag caagagggga aatgtctgta tagcctggaa 1620

agcattgccc tgatggatct gctgaataac ccgcagaatt atgaagtgga tcatattatt 1680

ccgcgtagcg tggcatttga taattccatt cataacaaag tgctggtgaa gcagatcgag 1740

aatagcaaaa aaggtaatcg tacgccgtat cagtatctga atagcagtga tgcaaaactg 1800

agctacaacc agtttaaaca gcatattctg aatctgagca aaagcaaaga tcgcatcagc 1860

aaaaaaaaga aggactacct gctggaagaa cgcgatatca acaaatttga agtccagaaa 1920

gagtttatca accgcaatct ggttgatacc cgttatgcaa cccgtgaact gaccagctat 1980

ctgaaagcat atttcagcgc caataacatg gacgtgaaag tgaaaacaat taacggcagc 2040

tttaccaacc atctgcgtaa agtttggcgc tttgataaat atcgcaacca cggctataaa 2100

catcatgcag aagatgccct gattattgca aatgcagatt tcctgtttaa agaaaacaaa 2160

aaactgaaag ccgtcaacag cgtgctggaa aaaccggaaa ttgagacaga tcagaaagac 2220

attcaggttg atagcgaaga taattacagc gaaatgttta tcatcccgaa acaggtgcag 2280

gatatcaaag attttcgcaa cttcaaatat agccaccgcg ttgacaaaaa acctaatcgt 2340

cagctgatta acgataccct gtatagcacc cgcaaaaaag ataacagcac ctatattgtg 2400

cagaccatta aagacatcta cgccaaagat aataccaccc tgaaaaaaca gttcgacaaa 2460

agcccagaaa aatttctgat gtatcagcat gatccgcgta ccttcgaaaa actggaagtt 2520

attatgaaac agtatgccaa cgagaaaaat ccgctggcca aatatcacga agaaaccggt 2580

gaatatctga ccaaatattc caagaagaac aacggtccga tcgttaaatc cctgaaatat 2640

atcggtaata aactgggcag ccatctggat gttacccatc agtttaaaag ctccacaaag 2700

aagctggtta aactgtccat caaaccgtat cgctttgatg tgtatctgac cgacaaaggc 2760

tataaattca ttaccatcag ctatctggac gtgctgaaaa aagacaacta ttattatatc 2820

ccggaacaga aatatgataa actgaaactg ggtaaagcca tcgataaaaa cgccaaattt 2880

atcgccagct tctacaaaaa cgacctgatt aaactggatg gcgagatcta taaaatcatc 2940

ggtgttaata gcgacacccg caatatgatt gagctggatc tgccggatat tcgctataaa 3000

gaatattgcg aactgaacaa cattaaaggc gaaccgcgta tcaaaaagac catcggcaaa 3060

aaagtgaata gcatcgagaa actgaccacc gatgttctgg gtaatgtgtt taccaatacc 3120

cagtatacca aacctcagct gctgttcaaa cgcggtaat 3159

<210> 29
<211> 3159
<212> DNA

<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> 编码F8-K737N-L739N的DNA序列

<400> 29

[0085]

```
atgaaccaga aatttatcct gggtctggat attggtatta ccagcgttgg ttatggcctg      60
attgattacg aaaccaaaaa cattattgat gccggtgttc gtctgtttcc ggaagcaaat     120
gttgaaaata atgaaggtcg tcgtagcaaa cgtggtagcc gtcgtctgaa acgtcgtcgt     180
attcatcgtc tggaacgtgt taaaagcctg ctgagcgaat ataagattat tagcggtctg     240
gcaccgacca ataatcagcc gtataacatt cgtgttaaag gtctgaccga acagctgacc     300
aaagatgaac tggcagttgc actgctgcat attgccaaac gccgtggcat tcataaaatc     360
gatgtgattg atagcaatga cgatgtgggt aatgaactga gcaccaaaga acagctgaac     420
aaaaatagca aactgctgaa agacaaattc gtgtgtcaga ttcagctgga acgtatgaat     480
gaaggccagg ttcgtggtga aaagaatcgc tttaaaaccg cagacatcat caaagaaatt     540
atccagctgc tgaacgtgca gaaaaacttc catcagctgg atgaaaactt catcaacaaa     600
tacatcgagc tggttgaaat gcgtcgcgaa tattttgaag gtcctggtca gggtagcccg     660
tttggttgga atggtgatct gaaaaaatgg tacgaaatgc tgatgggtca ctgtacctat     720
tttccgcaag aactgcgtag cgttaaatat gcctatagcg cagacctgtt taatgcactg     780
aatgatctga acaacctgat tattcagcgc gataatagcg agaaactgga ataccatgag     840
aagtatcaca tcatcgagaa cgtgttcaag cagaaaaaaa agccgacgct gaaacaaatc     900
gcaaaagaga ttggcgttaa cccggaagat attaaaggtt atcgtattac caaaagcggt     960
acaccggaat tcaccgaatt taaactgtat cacgatctga aaagcgtgct gtttgatcag    1020
agcattctgg aaaatgaaga tgtgctggac cagattgcag aaattctgac catttatcag    1080
gacaaagaca gcatcaaaag caaactgacc gaactggata ttctgctgaa tgaagaagat    1140
aaagagaaca ttgcacagct gaccggttat aacggcacac atcgcctgag cctgaaatgt    1200
attcgtctgg tactggaaga acagtggtat agcagccgta atcagatgga aatctttacc    1260
catctgaaca ttaaaccgaa gaaaatcaat ctgaccgcag ccaacaaaat tccgaaagcc    1320
atgattgatg agtttattct gagtccggtt gtgaaacgca cctttattca gagcatcaac    1380
gtgatcaaca aagtgatcga gaaatatggc atccccgaag atatcattat cgaactggca    1440
cgtgaaaata actccgatga tcgcaaaaag ttcatcaaca acctgcagaa aaagaatgaa    1500
gcaacccgca aacgcattaa cgaaattatt ggtcagaccg gtaatcagaa tgccaaacgt    1560
attgtggaaa aaatccgtct gcatgatcag caagagggga aatgtctgta tagcctggaa    1620
```

[0086]

agcattgccc tgatggatct gctgaataac ccgcagaatt atgaagtgga tcatattatt 1680

ccgcgtagcg tggcatttga taattccatt cataacaaag tgctggtgaa gcagatcgag 1740

aatagcaaaa aaggtaatcg tacgccgtat cagtatctga atagcagtga tgcaaaactg 1800

agctacaacc agtttaaaca gcatattctg aatctgagca aaagcaaaga tcgcatcagc 1860

aaaaaaaaga aggactacct gctggaagaa cgcgatatca acaaatttga agtccagaaa 1920

gagtttatca accgcaatct ggttgatacc cgttatgcaa cccgtgaact gaccagctat 1980

ctgaaagcat atttcagcgc caataacatg gacgtgaaag tgaaaacaat taacggcagc 2040

tttaccaacc atctgcgtaa agtttggcgc tttgataaat atcgcaacca cggctataaa 2100

catcatgcag aagatgccct gattattgca aatgcagatt tcctgtttaa agaaaacaaa 2160

aaactgaaag ccgtcaacag cgtgctggaa aaaccggaaa ttgagacaaa tcagaatgac 2220

attcaggttg atagcgaaga taattacagc gaaatgttta tcatcccgaa acaggtgcag 2280

gatatcaaag attttcgcaa cttcaaatat agccaccgcg ttgacaaaaa acctaatcgt 2340

cagctgatta acgataccct gtatagcacc cgcaaaaaag ataacagcac ctatattgtg 2400

cagaccatta aagacatcta cgccaaagat aataccaccc tgaaaaaaca gttcgacaaa 2460

agcccagaaa aatttctgat gtatcagcat gatccgcgta ccttcgaaaa actggaagtt 2520

attatgaaac agtatgccaa cgagaaaaat ccgctggcca aatatcacga agaaaccggt 2580

gaatatctga ccaaatattc caagaagaac aacggtccga tcgttaaatc cctgaaatat 2640

atcggtaata aactgggcag ccatctggat gttacccatc agtttaaaag ctccacaaag 2700

aagctggtta aactgtccat caaaccgtat cgctttgatg tgtatctgac cgacaaaggc 2760

tataaattca ttaccatcag ctatctggac gtgctgaaaa aagacaacta ttattatatc 2820

ccggaacaga aatatgataa actgaaactg ggtaaagcca tcgataaaaa cgccaaattt 2880

atcgccagct tctacaaaaa cgacctgatt aaactggatg gcgagatcta taaaatcatc 2940

ggtgttaata gcgacacccg caatatgatt gagctggatc tgccggatat tcgctataaa 3000

gaatattgcg aactgaacaa cattaaaggc gaaccgcgta tcaaaaagac catcggcaaa 3060

aaagtgaata gcatcgagaa actgaccacc gatgttctgg gtaatgtgtt taccaatacc 3120

cagtatacca aacctcagct gctgttcaaa cgcggtaat 3159

<210> 30
<211> 1054
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>

<221> misc_feature
<223> 路邓葡萄球菌Cas9

<400> 30

```
Met Asn Gln Lys Phe Ile Leu Gly Leu Asp Ile Gly Ile Thr Ser Val
1               5                  10                  15

Gly Tyr Gly Leu Ile Asp Tyr Glu Thr Lys Asn Ile Ile Asp Ala Gly
            20                  25                  30

Val Arg Leu Phe Pro Glu Ala Asn Val Glu Asn Asn Glu Gly Arg Arg
        35                  40                  45

Ser Lys Arg Gly Ser Arg Arg Leu Lys Arg Arg Arg Ile His Arg Leu
    50                  55                  60

Glu Arg Val Lys Lys Leu Leu Glu Asp Tyr Asn Leu Leu Asp Gln Ser
65                  70                  75                  80

Gln Ile Pro Gln Ser Thr Asn Pro Tyr Ala Ile Arg Val Lys Gly Leu
                85                  90                  95

Ser Glu Ala Leu Ser Lys Asp Glu Leu Val Ile Ala Leu Leu His Ile
            100                 105                 110
```

[0087]

```
Ala Lys Arg Arg Gly Ile His Lys Ile Asp Val Ile Asp Ser Asn Asp
        115                 120                 125

Asp Val Gly Asn Glu Leu Ser Thr Lys Glu Gln Leu Asn Lys Asn Ser
    130                 135                 140

Lys Leu Leu Lys Asp Lys Phe Val Cys Gln Ile Gln Leu Glu Arg Met
145                 150                 155                 160

Asn Glu Gly Gln Val Arg Gly Glu Lys Asn Arg Phe Lys Thr Ala Asp
                165                 170                 175

Ile Ile Lys Glu Ile Ile Gln Leu Leu Asn Val Gln Lys Asn Phe His
            180                 185                 190

Gln Leu Asp Glu Asn Phe Ile Asn Lys Tyr Ile Glu Leu Val Glu Met
        195                 200                 205

Arg Arg Glu Tyr Phe Glu Gly Pro Gly Lys Gly Ser Pro Tyr Gly Trp
    210                 215                 220

Glu Gly Asp Pro Lys Ala Trp Tyr Glu Thr Leu Met Gly His Cys Thr
225                 230                 235                 240
```

Tyr Phe Pro Asp Glu Leu Arg Ser Val Lys Tyr Ala Tyr Ser Ala Asp
245 250 255

Leu Phe Asn Ala Leu Asn Asp Leu Asn Asn Leu Val Ile Gln Arg Asp
260 265 270

Gly Leu Ser Lys Leu Glu Tyr His Glu Lys Tyr His Ile Ile Glu Asn
275 280 285

Val Phe Lys Gln Lys Lys Lys Pro Thr Leu Lys Gln Ile Ala Asn Glu
290 295 300

Ile Asn Val Asn Pro Glu Asp Ile Lys Gly Tyr Arg Ile Thr Lys Ser
305 310 315 320

Gly Lys Pro Gln Phe Thr Glu Phe Lys Leu Tyr His Asp Leu Lys Ser
325 330 335

Val Leu Phe Asp Gln Ser Ile Leu Glu Asn Glu Asp Val Leu Asp Gln
340 345 350

Ile Ala Glu Ile Leu Thr Ile Tyr Gln Asp Lys Asp Ser Ile Lys Ser
355 360 365

[0088]

Lys Leu Thr Glu Leu Asp Ile Leu Leu Asn Glu Glu Asp Lys Glu Asn
370 375 380

Ile Ala Gln Leu Thr Gly Tyr Thr Gly Thr His Arg Leu Ser Leu Lys
385 390 395 400

Cys Ile Arg Leu Val Leu Glu Glu Gln Trp Tyr Ser Ser Arg Asn Gln
405 410 415

Met Glu Ile Phe Thr His Leu Asn Ile Lys Pro Lys Lys Ile Asn Leu
420 425 430

Thr Ala Ala Asn Lys Ile Pro Lys Ala Met Ile Asp Glu Phe Ile Leu
435 440 445

Ser Pro Val Val Lys Arg Thr Phe Gly Gln Ala Ile Asn Leu Ile Asn
450 455 460

Lys Ile Ile Glu Lys Tyr Gly Val Pro Glu Asp Ile Ile Ile Glu Leu
465 470 475 480

Ala Arg Glu Asn Asn Ser Lys Asp Lys Gln Lys Phe Ile Asn Glu Met
485 490 495

```
Gln Lys Lys Asn Glu Asn Thr Arg Lys Arg Ile Asn Glu Ile Ile Gly
            500                 505                 510

Lys Tyr Gly Asn Gln Asn Ala Lys Arg Leu Val Glu Lys Ile Arg Leu
        515                 520                 525

His Asp Glu Gln Glu Gly Lys Cys Leu Tyr Ser Leu Glu Ser Ile Pro
    530                 535                 540

Leu Glu Asp Leu Leu Asn Asn Pro Asn His Tyr Glu Val Asp His Ile
545                 550                 555                 560

Ile Pro Arg Ser Val Ser Phe Asp Asn Ser Tyr His Asn Lys Val Leu
                565                 570                 575

Val Lys Gln Ser Glu Asn Ser Lys Lys Ser Asn Leu Thr Pro Tyr Gln
            580                 585                 590

Tyr Phe Asn Ser Gly Lys Ser Lys Leu Ser Tyr Asn Gln Phe Lys Gln
        595                 600                 605

His Ile Leu Asn Leu Ser Lys Ser Gln Asp Arg Ile Ser Lys Lys Lys
    610                 615                 620
```

[0089]

```
Lys Glu Tyr Leu Leu Glu Glu Arg Asp Ile Asn Lys Phe Glu Val Gln
625                 630                 635                 640

Lys Glu Phe Ile Asn Arg Asn Leu Val Asp Thr Arg Tyr Ala Thr Arg
                645                 650                 655

Glu Leu Thr Asn Tyr Leu Lys Ala Tyr Phe Ser Ala Asn Asn Met Asn
            660                 665                 670

Val Lys Val Lys Thr Ile Asn Gly Ser Phe Thr Asp Tyr Leu Arg Lys
        675                 680                 685

Val Trp Lys Phe Lys Lys Glu Arg Asn His Gly Tyr Lys His His Ala
    690                 695                 700

Glu Asp Ala Leu Ile Ile Ala Asn Ala Asp Phe Leu Phe Lys Glu Asn
705                 710                 715                 720

Lys Lys Leu Lys Ala Val Asn Ser Val Leu Glu Lys Pro Glu Ile Glu
                725                 730                 735

Thr Lys Gln Leu Asp Ile Gln Val Asp Ser Glu Asp Asn Tyr Ser Glu
            740                 745                 750
```

```
Met Phe Ile Ile Pro Lys Gln Val Gln Asp Ile Lys Asp Phe Arg Asn
        755                 760                 765

Phe Lys Tyr Ser His Arg Val Asp Lys Lys Pro Asn Arg Gln Leu Ile
    770                 775                 780

Asn Asp Thr Leu Tyr Ser Thr Arg Lys Lys Asp Asn Ser Thr Tyr Ile
785                 790                 795                 800

Val Gln Thr Ile Lys Asp Ile Tyr Ala Lys Asp Asn Thr Thr Leu Lys
                805                 810                 815

Lys Gln Phe Asp Lys Ser Pro Glu Lys Phe Leu Met Tyr Gln His Asp
            820                 825                 830

Pro Arg Thr Phe Glu Lys Leu Glu Val Ile Met Lys Gln Tyr Ala Asn
        835                 840                 845

Glu Lys Asn Pro Leu Ala Lys Tyr His Glu Glu Thr Gly Glu Tyr Leu
    850                 855                 860

Thr Lys Tyr Ser Lys Lys Asn Asn Gly Pro Ile Val Lys Ser Leu Lys
865                 870                 875                 880
```

[0090]

```
Tyr Ile Gly Asn Lys Leu Gly Ser His Leu Asp Val Thr His Gln Phe
                885                 890                 895

Lys Ser Ser Thr Lys Lys Leu Val Lys Leu Ser Ile Lys Pro Tyr Arg
            900                 905                 910

Phe Asp Val Tyr Leu Thr Asp Lys Gly Tyr Lys Phe Ile Thr Ile Ser
        915                 920                 925

Tyr Leu Asp Val Leu Lys Lys Asp Asn Tyr Tyr Tyr Ile Pro Glu Gln
    930                 935                 940

Lys Tyr Asp Lys Leu Lys Leu Gly Lys Ala Ile Asp Lys Asn Ala Lys
945                 950                 955                 960

Phe Ile Ala Ser Phe Tyr Lys Asn Asp Leu Ile Lys Leu Asp Gly Glu
                965                 970                 975

Ile Tyr Lys Ile Ile Gly Val Asn Ser Asp Thr Arg Asn Met Ile Glu
            980                 985                 990

Leu Asp Leu Pro Asp Ile Arg Tyr  Lys Glu Tyr Cys Glu  Leu Asn Asn
        995                 1000                1005
```

Ile Lys Gly Glu Pro Arg Ile Lys Lys Thr Ile Gly Lys Lys Val
1010 1015 1020

Asn Ser Ile Glu Lys Leu Thr Thr Asp Val Leu Gly Asn Val Phe
1025 1030 1035

Thr Asn Thr Gln Tyr Thr Lys Pro Gln Leu Leu Phe Lys Arg Gly
1040 1045 1050

Asn

<210> 31
<211> 1054
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> 巴氏葡萄球菌Cas9

<400> 31

[0091]

Met Lys Glu Lys Tyr Ile Leu Gly Leu Asp Leu Gly Ile Thr Ser Val
1 5 10 15

Gly Tyr Gly Ile Ile Asn Phe Glu Thr Lys Lys Ile Ile Asp Ala Gly
20 25 30

Val Arg Leu Phe Pro Glu Ala Asn Val Asp Asn Asn Glu Gly Arg Arg
35 40 45

Ser Lys Arg Gly Ser Arg Arg Leu Lys Arg Arg Arg Ile His Arg Leu
50 55 60

Glu Arg Val Lys Leu Leu Leu Thr Glu Tyr Asp Leu Ile Asn Lys Glu
65 70 75 80

Gln Ile Pro Thr Ser Asn Asn Pro Tyr Gln Ile Arg Val Lys Gly Leu
85 90 95

Ser Glu Ile Leu Ser Lys Asp Glu Leu Ala Ile Ala Leu Leu His Leu
100 105 110

Ala Lys Arg Arg Gly Ile His Asn Ile Asn Val Ser Ser Glu Asp Glu
115 120 125

```
Asp Ala Ser Asn Glu Leu Ser Thr Lys Glu Gln Ile Asn Arg Asn Asn
    130                 135                 140

Lys Leu Leu Lys Asp Lys Tyr Val Cys Glu Val Gln Leu Gln Arg Leu
145                 150                 155                 160

Lys Glu Gly Gln Ile Arg Gly Glu Lys Asn Arg Phe Lys Thr Thr Asp
                165                 170                 175

Ile Leu Lys Glu Ile Asp Gln Leu Leu Lys Val Gln Lys Asp Tyr His
            180                 185                 190

Asn Leu Asp Ile Asp Phe Ile Asn Gln Tyr Lys Glu Ile Val Glu Thr
        195                 200                 205

Arg Arg Glu Tyr Phe Glu Gly Pro Gly Gln Gly Ser Pro Phe Gly Trp
    210                 215                 220

Asn Gly Asp Leu Lys Lys Trp Tyr Glu Met Leu Met Gly His Cys Thr
225                 230                 235                 240

Tyr Phe Pro Gln Glu Leu Arg Ser Val Lys Tyr Ala Tyr Ser Ala Asp
                245                 250                 255
```

[0092]

```
Leu Phe Asn Ala Leu Asn Asp Leu Asn Asn Leu Ile Ile Gln Arg Asp
            260                 265                 270

Asn Ser Glu Lys Leu Glu Tyr His Glu Lys Tyr His Ile Ile Glu Asn
        275                 280                 285

Val Phe Lys Gln Lys Lys Lys Pro Thr Leu Lys Gln Ile Ala Lys Glu
    290                 295                 300

Ile Gly Val Asn Pro Glu Asp Ile Lys Gly Tyr Arg Ile Thr Lys Ser
305                 310                 315                 320

Gly Thr Pro Gln Phe Thr Glu Phe Lys Leu Tyr His Asp Leu Lys Ser
                325                 330                 335

Ile Val Phe Asp Lys Ser Ile Leu Glu Asn Glu Ala Ile Leu Asp Gln
            340                 345                 350

Ile Ala Glu Ile Leu Thr Ile Tyr Gln Asp Glu Gln Ser Ile Lys Glu
        355                 360                 365

Glu Leu Asn Lys Leu Pro Glu Ile Leu Asn Glu Gln Asp Lys Ala Glu
    370                 375                 380
```

```
Ile Ala Lys Leu Ile Gly Tyr Asn Gly Thr His Arg Leu Ser Leu Lys
385                 390                 395                 400

Cys Ile His Leu Ile Asn Glu Glu Leu Trp Gln Thr Ser Arg Asn Gln
                405                 410                 415

Met Glu Ile Phe Asn Tyr Leu Asn Ile Lys Pro Asn Lys Val Asp Leu
            420                 425                 430

Ser Glu Gln Asn Lys Ile Pro Lys Asp Met Val Asn Asp Phe Ile Leu
        435                 440                 445

Ser Pro Val Val Lys Arg Thr Phe Ile Gln Ser Ile Asn Val Ile Asn
    450                 455                 460

Lys Val Ile Glu Lys Tyr Gly Ile Pro Glu Asp Ile Ile Ile Glu Leu
465                 470                 475                 480

Ala Arg Glu Asn Asn Ser Asp Asp Arg Lys Lys Phe Ile Asn Asn Leu
                485                 490                 495

Gln Lys Lys Asn Glu Ala Thr Arg Lys Arg Ile Asn Glu Ile Ile Gly
            500                 505                 510
```

[0093]

```
Gln Thr Gly Asn Gln Asn Ala Lys Arg Ile Val Glu Lys Ile Arg Leu
        515                 520                 525

His Asp Gln Gln Glu Gly Lys Cys Leu Tyr Ser Leu Glu Ser Ile Ala
    530                 535                 540

Leu Met Asp Leu Leu Asn Asn Pro Gln Asn Tyr Glu Val Asp His Ile
545                 550                 555                 560

Ile Pro Arg Ser Val Ala Phe Asp Asn Ser Ile His Asn Lys Val Leu
                565                 570                 575

Val Lys Gln Ile Glu Asn Ser Lys Lys Gly Asn Arg Thr Pro Tyr Gln
            580                 585                 590

Tyr Leu Asn Ser Ser Asp Ala Lys Leu Ser Tyr Asn Gln Phe Lys Gln
        595                 600                 605

His Ile Leu Asn Leu Ser Lys Ser Lys Asp Arg Ile Ser Lys Lys Lys
    610                 615                 620

Lys Asp Tyr Leu Leu Glu Glu Arg Asp Ile Asn Lys Phe Glu Val Gln
625                 630                 635                 640
```

Lys Glu Phe Ile Asn Arg Asn Leu Val Asp Thr Arg Tyr Ala Thr Arg
645 650 655

Glu Leu Thr Ser Tyr Leu Lys Ala Tyr Phe Ser Ala Asn Asn Met Asp
660 665 670

Val Lys Val Lys Thr Ile Asn Gly Ser Phe Thr Asn His Leu Arg Lys
675 680 685

Val Trp Arg Phe Asp Lys Tyr Arg Asn His Gly Tyr Lys His His Ala
690 695 700

Glu Asp Ala Leu Ile Ile Ala Asn Ala Asp Phe Leu Phe Lys Glu Asn
705 710 715 720

Lys Lys Leu Gln Asn Thr Asn Lys Ile Leu Glu Lys Pro Thr Ile Glu
725 730 735

Asn Asn Thr Lys Lys Val Thr Val Glu Lys Glu Glu Asp Tyr Asn Asn
740 745 750

Val Phe Glu Thr Pro Lys Leu Val Glu Asp Ile Lys Gln Tyr Arg Asp
755 760 765

[0094]

Tyr Lys Phe Ser His Arg Val Asp Lys Lys Pro Asn Arg Gln Leu Ile
770 775 780

Asn Asp Thr Leu Tyr Ser Thr Arg Met Lys Asp Glu His Asp Tyr Ile
785 790 795 800

Val Gln Thr Ile Thr Asp Ile Tyr Gly Lys Asp Asn Thr Asn Leu Lys
805 810 815

Lys Gln Phe Asn Lys Asn Pro Glu Lys Phe Leu Met Tyr Gln Asn Asp
820 825 830

Pro Lys Thr Phe Glu Lys Leu Ser Ile Ile Met Lys Gln Tyr Ser Asp
835 840 845

Glu Lys Asn Pro Leu Ala Lys Tyr Tyr Glu Glu Thr Gly Glu Tyr Leu
850 855 860

Thr Lys Tyr Ser Lys Lys Asn Asn Gly Pro Ile Val Lys Lys Ile Lys
865 870 875 880

Leu Leu Gly Asn Lys Val Gly Asn His Leu Asp Val Thr Asn Lys Tyr
885 890 895

Glu Asn Ser Thr Lys Lys Leu Val Lys Leu Ser Ile Lys Asn Tyr Arg
900 905 910

Phe Asp Val Tyr Leu Thr Glu Lys Gly Tyr Lys Phe Val Thr Ile Ala
915 920 925

Tyr Leu Asn Val Phe Lys Lys Asp Asn Tyr Tyr Tyr Ile Pro Lys Asp
930 935 940

Lys Tyr Gln Glu Leu Lys Glu Lys Lys Lys Ile Lys Asp Thr Asp Gln
945 950 955 960

Phe Ile Ala Ser Phe Tyr Lys Asn Asp Leu Ile Lys Leu Asn Gly Asp
965 970 975

Leu Tyr Lys Ile Ile Gly Val Asn Ser Asp Asp Arg Asn Ile Ile Glu
980 985 990

Leu Asp Tyr Tyr Asp Ile Lys Tyr Lys Asp Tyr Cys Glu Ile Asn Asn
995 1000 1005

Ile Lys Gly Glu Pro Arg Ile Lys Lys Thr Ile Gly Lys Lys Thr
1010 1015 1020

[0095]

Glu Ser Ile Glu Lys Phe Thr Thr Asp Val Leu Gly Asn Leu Tyr
1025 1030 1035

Leu His Ser Thr Glu Lys Ala Pro Gln Leu Ile Phe Lys Arg Gly
1040 1045 1050

Leu

<210> 32
<211> 1051
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> 猪葡萄球菌Cas9

<400> 32

Met Asn Asn Tyr Ile Leu Gly Leu Asp Ile Gly Ile Thr Ser Val Gly
1 5 10 15

Tyr Gly Ile Val Asp Ser Asp Thr Arg Glu Ile Lys Asp Ala Gly Val

20 25 30

Arg Leu Phe Pro Glu Ala Asn Val Asp Asn Asn Glu Gly Arg Arg Ser
35 40 45

Lys Arg Gly Ala Arg Arg Leu Lys Arg Arg Arg Ile His Arg Leu Asp
50 55 60

Arg Val Lys His Leu Leu Ala Glu Tyr Asp Leu Leu Asp Leu Thr Asn
65 70 75 80

Ile Pro Lys Ser Thr Asn Pro Tyr Gln Thr Arg Val Lys Gly Leu Asn
85 90 95

Glu Lys Leu Ser Lys Asp Glu Leu Val Ile Ala Leu Leu His Ile Ala
100 105 110

Lys Arg Arg Gly Ile His Asn Val Asn Val Met Met Asp Asp Asn Asp
115 120 125

Ser Gly Asn Glu Leu Ser Thr Lys Asp Gln Leu Lys Lys Asn Ala Lys
130 135 140

[0096] Ala Leu Ser Asp Lys Tyr Val Cys Glu Leu Gln Leu Glu Arg Phe Glu
145 150 155 160

Gln Asp Tyr Lys Val Arg Gly Glu Lys Asn Arg Phe Lys Thr Glu Asp
165 170 175

Phe Val Arg Glu Ala Arg Lys Leu Leu Glu Thr Gln Ser Lys Phe Phe
180 185 190

Glu Ile Asp Gln Thr Phe Ile Met Arg Tyr Ile Glu Leu Ile Glu Thr
195 200 205

Arg Arg Glu Tyr Phe Glu Gly Pro Gly Lys Gly Ser Pro Phe Gly Trp
210 215 220

Glu Gly Asn Ile Lys Lys Trp Phe Glu Gln Met Met Gly His Cys Thr
225 230 235 240

Tyr Phe Pro Glu Glu Leu Arg Ser Val Lys Tyr Ser Tyr Ser Ala Glu
245 250 255

Leu Phe Asn Ala Leu Asn Asp Leu Asn Asn Leu Val Ile Thr Arg Asp
260 265 270

Glu Asp Ala Lys Leu Asn Tyr Gly Glu Lys Phe Gln Ile Ile Glu Asn

275 280 285

Val Phe Lys Gln Lys Lys Thr Pro Asn Leu Lys Gln Ile Ala Ile Glu
290 295 300

Ile Gly Val His Glu Thr Glu Ile Lys Gly Tyr Arg Val Asn Lys Ser
305 310 315 320

Gly Lys Pro Glu Phe Thr Gln Phe Lys Leu Tyr His Asp Leu Lys Asn
325 330 335

Ile Phe Lys Asp Pro Lys Tyr Leu Asn Asp Ile Gln Leu Met Asp Asn
340 345 350

Ile Ala Glu Ile Ile Thr Ile Tyr Gln Asp Ala Glu Ser Ile Ile Lys
355 360 365

Glu Leu Asn Gln Leu Pro Glu Leu Leu Ser Glu Arg Glu Lys Glu Lys
370 375 380

Ile Ser Ala Leu Ser Gly Tyr Ser Gly Thr His Arg Leu Ser Leu Lys
385 390 395 400

[0097] Cys Ile Asn Leu Leu Leu Asp Asp Leu Trp Glu Ser Ser Leu Asn Gln
405 410 415

Met Glu Leu Phe Thr Lys Leu Asn Leu Lys Pro Lys Lys Ile Asp Leu
420 425 430

Ser Gln Gln His Lys Ile Pro Ser Lys Leu Val Asp Asp Phe Ile Leu
435 440 445

Ser Pro Val Val Lys Arg Ala Phe Ile Gln Ser Ile Gln Val Val Asn
450 455 460

Ala Ile Ile Asp Lys Tyr Gly Leu Pro Glu Asp Ile Ile Ile Glu Leu
465 470 475 480

Ala Arg Glu Asn Asn Ser Asp Asp Arg Arg Lys Phe Leu Asn Gln Leu
485 490 495

Gln Lys Gln Asn Glu Glu Thr Arg Lys Gln Val Glu Lys Val Leu Arg
500 505 510

Glu Tyr Gly Asn Asp Asn Ala Lys Arg Ile Val Gln Lys Ile Lys Leu
515 520 525

His Asn Met Gln Glu Gly Lys Cys Leu Tyr Ser Leu Lys Asp Ile Pro

530 535 540

Leu Glu Asp Leu Leu Arg Asn Pro His His Tyr Glu Val Asp His Ile
545 550 555 560

Ile Pro Arg Ser Val Ala Phe Asp Asn Ser Met His Asn Lys Val Leu
565 570 575

Val Arg Ala Asp Glu Asn Ser Lys Lys Gly Asn Arg Thr Pro Tyr Gln
580 585 590

Tyr Leu Asn Ser Ser Glu Ser Ser Leu Ser Tyr Asn Glu Phe Lys Gln
595 600 605

His Ile Leu Asn Leu Ser Lys Thr Lys Asp Arg Ile Thr Lys Lys Lys
610 615 620

Arg Glu Tyr Leu Leu Glu Glu Arg Asp Ile Asn Lys Phe Asp Val Gln
625 630 635 640

Lys Glu Phe Ile Asn Arg Asn Leu Val Asp Thr Arg Tyr Ala Thr Arg
645 650 655

[0098] Glu Leu Thr Ser Leu Leu Lys Ala Tyr Phe Ser Ala Asn Asn Leu Asp
660 665 670

Val Lys Val Lys Thr Ile Asn Gly Ser Phe Thr Asn Tyr Leu Arg Lys
675 680 685

Val Trp Lys Phe Asp Lys Asp Arg Asn Lys Gly Tyr Lys His His Ala
690 695 700

Glu Asp Ala Leu Ile Ile Ala Asn Ala Asp Phe Leu Phe Lys His Asn
705 710 715 720

Lys Lys Leu Arg Asn Ile Asn Lys Val Leu Asp Ala Pro Ser Lys Glu
725 730 735

Val Asp Lys Lys Arg Val Thr Val Gln Ser Glu Asp Glu Tyr Asn Gln
740 745 750

Ile Phe Glu Asp Thr Gln Lys Ala Gln Ala Ile Lys Lys Phe Glu Ile
755 760 765

Arg Lys Phe Ser His Arg Val Asp Lys Lys Pro Asn Arg Gln Leu Ile
770 775 780

Asn Asp Thr Leu Tyr Ser Thr Arg Asn Ile Asp Gly Ile Glu Tyr Val

785 790 795 800

Val Glu Ser Ile Lys Asp Ile Tyr Ser Val Asn Asn Asp Lys Val Lys
805 810 815

Thr Lys Phe Lys Lys Asp Pro His Arg Leu Leu Met Tyr Arg Asn Asp
820 825 830

Pro Gln Thr Phe Glu Lys Phe Glu Lys Val Phe Lys Gln Tyr Glu Ser
835 840 845

Glu Lys Asn Pro Phe Ala Lys Tyr Tyr Glu Glu Thr Gly Glu Lys Ile
850 855 860

Arg Lys Phe Ser Lys Thr Gly Gln Gly Pro Tyr Ile Asn Lys Ile Lys
865 870 875 880

Tyr Leu Arg Glu Arg Leu Gly Arg His Cys Asp Val Thr Asn Lys Tyr
885 890 895

Ile Asn Ser Arg Asn Lys Ile Val Gln Leu Lys Ile Tyr Ser Tyr Arg
900 905 910

[0099] Phe Asp Ile Tyr Gln Tyr Gly Asn Asn Tyr Lys Met Ile Thr Ile Ser
915 920 925

Tyr Ile Asp Leu Glu Gln Lys Ser Asn Tyr Tyr Tyr Ile Ser Arg Glu
930 935 940

Lys Tyr Glu Gln Lys Lys Lys Asp Lys Gln Ile Asp Asp Ser Tyr Lys
945 950 955 960

Phe Ile Gly Ser Phe Tyr Lys Asn Asp Ile Ile Asn Tyr Asn Gly Glu
965 970 975

Met Tyr Arg Val Ile Gly Val Asn Asp Ser Glu Lys Asn Lys Ile Gln
980 985 990

Leu Asp Met Ile Asp Ile Ser Ile Lys Asp Tyr Met Glu Leu Asn Asn
995 1000 1005

Ile Lys Lys Thr Gly Val Ile Tyr Lys Thr Ile Gly Lys Ser Thr
1010 1015 1020

Thr His Ile Glu Lys Tyr Thr Thr Asp Ile Leu Gly Asn Leu Tyr
1025 1030 1035

Lys Ala Ala Pro Pro Lys Lys Pro Gln Leu Ile Phe Lys

1040 1045 1050

<210> 33
<211> 1052
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> Staphylococcus microti Cas9

<400> 33

Met Glu Lys Asp Tyr Ile Leu Gly Leu Asp Ile Gly Ile Gly Ser Val
1 5 10 15

Gly Tyr Gly Leu Ile Asp Tyr Asp Thr Lys Ser Ile Ile Asp Ala Gly
20 25 30

Val Arg Leu Phe Pro Glu Ala Asn Ala Asp Asn Asn Leu Gly Arg Arg
35 40 45

Ala Lys Arg Gly Ala Arg Arg Leu Lys Arg Arg Arg Ile His Arg Leu
50 55 60

[0100]

Glu Arg Val Lys Ser Leu Leu Ser Glu Tyr Lys Ile Ile Ser Gly Leu
65 70 75 80

Ala Pro Thr Asn Asn Gln Pro Tyr Asn Ile Arg Val Lys Gly Leu Thr
85 90 95

Glu Gln Leu Thr Lys Asp Glu Leu Ala Val Ala Leu Leu His Ile Ala
100 105 110

Lys Arg Arg Gly Ile His Asn Val Asp Val Ala Ala Asp Lys Glu Glu
115 120 125

Thr Ala Ser Asp Ser Leu Ser Thr Lys Asp Gln Ile Asn Lys Asn Ala
130 135 140

Lys Phe Leu Glu Ser Arg Tyr Val Cys Glu Leu Gln Lys Glu Arg Leu
145 150 155 160

Glu Asn Glu Gly His Val Arg Gly Val Glu Asn Arg Phe Leu Thr Lys
165 170 175

Asp Ile Val Arg Glu Ala Lys Lys Ile Ile Asp Thr Gln Met Gln Tyr
180 185 190

```
Tyr Pro Glu Ile Asp Glu Thr Phe Lys Glu Lys Tyr Ile Ser Leu Val
        195                 200                 205

Glu Thr Arg Arg Glu Tyr Tyr Glu Gly Pro Gly Lys Gly Ser Pro Tyr
    210                 215                 220

Gly Trp Asp Ala Asp Val Lys Lys Trp Tyr Gln Leu Met Met Gly His
225                 230                 235                 240

Cys Thr Tyr Phe Pro Val Glu Phe Arg Ser Val Lys Tyr Ala Tyr Thr
                245                 250                 255

Ala Asp Leu Tyr Asn Ala Leu Asn Asp Leu Asn Asn Leu Thr Ile Ala
            260                 265                 270

Arg Asp Asp Asn Pro Lys Leu Glu Tyr His Glu Lys Tyr His Ile Ile
        275                 280                 285

Glu Asn Val Phe Lys Gln Lys Arg Asn Pro Thr Leu Lys Gln Ile Ala
    290                 295                 300

Lys Glu Ile Gly Val Asn Asp Ile Asn Ile Ser Gly Tyr Arg Val Thr
305                 310                 315                 320
```

[0101]

```
Lys Ser Gly Lys Pro Gln Phe Thr Ser Phe Lys Leu Phe His Asp Leu
                325                 330                 335

Lys Lys Val Val Lys Asp His Ala Ile Leu Asp Asp Ile Asp Leu Leu
            340                 345                 350

Asn Gln Ile Ala Glu Ile Leu Thr Ile Tyr Gln Asp Lys Asp Ser Ile
        355                 360                 365

Val Ala Glu Leu Gly Gln Leu Glu Tyr Leu Met Ser Glu Ala Asp Lys
    370                 375                 380

Gln Ser Ile Ser Glu Leu Thr Gly Tyr Thr Gly Thr His Ser Leu Ser
385                 390                 395                 400

Leu Lys Cys Met Asn Met Ile Ile Asp Glu Leu Trp His Ser Ser Met
                405                 410                 415

Asn Gln Met Glu Val Phe Thr Tyr Leu Asn Met Arg Pro Lys Lys Tyr
            420                 425                 430

Glu Leu Lys Gly Tyr Gln Arg Ile Pro Thr Asp Met Ile Asp Asp Ala
        435                 440                 445
```

```
Ile Leu Ser Pro Val Val Lys Arg Ser Phe Lys Gln Ala Ile Gly Val
    450                 455                 460

Val Asn Ala Ile Ile Lys Lys Tyr Gly Leu Pro Lys Asp Ile Ile Ile
465                 470                 475                 480

Glu Leu Ala Arg Glu Ser Asn Ser Ala Glu Lys Ser Arg Tyr Leu Arg
                485                 490                 495

Ala Ile Gln Lys Lys Asn Glu Lys Thr Arg Glu Arg Ile Glu Ala Ile
            500                 505                 510

Ile Lys Glu Tyr Gly Asn Glu Asn Ala Lys Gly Leu Val Gln Lys Ile
        515                 520                 525

Lys Leu His Asp Ala Gln Glu Gly Lys Cys Leu Tyr Ser Leu Lys Asp
    530                 535                 540

Ile Pro Leu Glu Asp Leu Leu Arg Asn Pro Asn Asn Tyr Asp Ile Asp
545                 550                 555                 560

His Ile Ile Pro Arg Ser Val Ser Phe Asp Asp Ser Met His Asn Lys
                565                 570                 575
```

[0102]

```
Val Leu Val Arg Arg Glu Gln Asn Ala Lys Lys Asn Asn Gln Thr Pro
            580                 585                 590

Tyr Gln Tyr Leu Thr Ser Gly Tyr Ala Asp Ile Lys Tyr Ser Val Phe
        595                 600                 605

Lys Gln His Val Leu Asn Leu Ala Glu Asn Lys Asp Arg Met Thr Lys
    610                 615                 620

Lys Lys Arg Glu Tyr Leu Leu Glu Glu Arg Asn Ile Asn Lys Tyr Asp
625                 630                 635                 640

Val Gln Lys Glu Phe Ile Asn Arg Asn Leu Val Asp Thr Arg Tyr Thr
                645                 650                 655

Thr Arg Glu Leu Thr Thr Leu Leu Lys Thr Tyr Phe Thr Ile Asn Asn
            660                 665                 670

Leu Asp Val Lys Val Lys Thr Ile Asn Gly Ser Phe Thr Asp Phe Leu
        675                 680                 685

Arg Lys Arg Trp Gly Phe Lys Lys Asn Arg Asp Glu Gly Tyr Lys His
    690                 695                 700
```

His Ala Glu Asp Ala Leu Ile Ile Ala Asn Ala Asp Tyr Leu Phe Lys
705 710 715 720

Glu His Lys Leu Leu Lys Glu Ile Lys Asp Val Ser Asp Leu Ala Gly
725 730 735

Asp Glu Arg Asn Ser Asn Val Lys Asp Glu Asp Gln Tyr Glu Glu Val
740 745 750

Phe Gly Gly Tyr Phe Lys Ile Glu Asp Ile Lys Lys Tyr Lys Ile Lys
755 760 765

Lys Phe Ser His Arg Val Asp Lys Lys Pro Asn Arg Gln Leu Ile Asn
770 775 780

Asp Thr Ile Tyr Ser Thr Arg Val Lys Asp Asp Lys Arg Tyr Leu Ile
785 790 795 800

Asn Thr Leu Lys Asn Leu Tyr Asp Lys Ser Asn Gly Asp Leu Lys Glu
805 810 815

Arg Met Gln Lys Asp Pro Glu Ser Leu Leu Met Tyr His His Asp Pro
820 825 830

[0103]

Gln Thr Phe Glu Lys Leu Lys Ile Val Met Ser Gln Tyr Glu Asn Glu
835 840 845

Lys Asn Pro Leu Ala Lys Tyr Phe Glu Glu Thr Gly Gln Tyr Leu Thr
850 855 860

Lys Tyr Ala Lys His Asp Asn Gly Pro Ala Ile His Lys Ile Lys Tyr
865 870 875 880

Tyr Gly Asn Lys Leu Val Glu His Leu Asp Ile Thr Lys Asn Tyr His
885 890 895

Asn Pro Gln Asn Lys Val Val Gln Leu Ser Gln Lys Ser Phe Arg Phe
900 905 910

Asp Val Tyr Gln Thr Asp Lys Gly Tyr Lys Phe Ile Ser Ile Ala Tyr
915 920 925

Leu Thr Leu Lys Asn Glu Lys Asn Tyr Tyr Ala Ile Ser Gln Glu Lys
930 935 940

Tyr Asp Gln Leu Lys Ser Glu Lys Lys Ile Ser Asn Asn Ala Val Phe
945 950 955 960

Ile Gly Ser Phe Tyr Thr Ser Asp Ile Ile Glu Ile Asn Asn Glu Lys
965 970 975

Phe Arg Val Ile Gly Val Asn Ser Asp Lys Asn Asn Leu Ile Glu Val
980 985 990

Asp Arg Ile Asp Ile Arg Gln Lys Glu Phe Ile Glu Leu Glu Glu Glu
995 1000 1005

Lys Lys Asn Asn Arg Ile Lys Val Thr Ile Gly Arg Lys Thr Thr
1010 1015 1020

Asn Ile Glu Lys Phe His Thr Asp Ile Leu Gly Asn Met Tyr Lys
1025 1030 1035

Ser Lys Arg Pro Lys Ala Pro Gln Leu Val Phe Lys Lys Gly
1040 1045 1050

<210> 34
<211> 67
<212> PRT
<213> 人工序列

[0104]

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> 来自8小结构域文库的路邓葡萄球菌小结构域1

<400> 34

Met Asn Gln Lys Phe Ile Leu Gly Leu Asp Ile Gly Ile Thr Ser Val
1 5 10 15

Gly Tyr Gly Leu Ile Asp Tyr Glu Thr Lys Asn Ile Ile Asp Ala Gly
20 25 30

Val Arg Leu Phe Pro Glu Ala Asn Val Glu Asn Asn Glu Gly Arg Arg
35 40 45

Ser Lys Arg Gly Ser Arg Arg Leu Lys Arg Arg Arg Ile His Arg Leu
50 55 60

Glu Arg Val
65

<210> 35
<211> 67
<212> PRT

<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> 来自8小结构域文库的巴氏葡萄球菌小结构域1

<400> 35

Met Lys Glu Lys Tyr Ile Leu Gly Leu Asp Leu Gly Ile Thr Ser Val
1 5 10 15

Gly Tyr Gly Ile Ile Asn Phe Glu Thr Lys Lys Ile Ile Asp Ala Gly
20 25 30

Val Arg Leu Phe Pro Glu Ala Asn Val Asp Asn Asn Glu Gly Arg Arg
35 40 45

Ser Lys Arg Gly Ser Arg Arg Leu Lys Arg Arg Arg Ile His Arg Leu
50 55 60

Glu Arg Val
65

[0105]

<210> 36
<211> 66
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> 来自8小结构域文库的猪葡萄球菌小结构域1

<400> 36

Met Asn Asn Tyr Ile Leu Gly Leu Asp Ile Gly Ile Thr Ser Val Gly
1 5 10 15

Tyr Gly Ile Val Asp Ser Asp Thr Arg Glu Ile Lys Asp Ala Gly Val
20 25 30

Arg Leu Phe Pro Glu Ala Asn Val Asp Asn Asn Glu Gly Arg Arg Ser
35 40 45

Lys Arg Gly Ala Arg Arg Leu Lys Arg Arg Arg Ile His Arg Leu Asp
50 55 60

Arg Val

65

<210> 37
<211> 67
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> 来自8小结构域文库的Staphylococcus microti小结构域1

<400> 37

Met Glu Lys Asp Tyr Ile Leu Gly Leu Asp Ile Gly Ile Gly Ser Val
1 5 10 15

Gly Tyr Gly Leu Ile Asp Tyr Asp Thr Lys Ser Ile Ile Asp Ala Gly
20 25 30

Val Arg Leu Phe Pro Glu Ala Asn Ala Asp Asn Asn Leu Gly Arg Arg
35 40 45

Ala Lys Arg Gly Ala Arg Arg Leu Lys Arg Arg Arg Ile His Arg Leu
[0106] 50 55 60

Glu Arg Val
65

<210> 38
<211> 52
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> 来自8小结构域文库的路邓葡萄球菌小结构域2

<400> 38

Lys Lys Leu Leu Glu Asp Tyr Asn Leu Leu Asp Gln Ser Gln Ile Pro
1 5 10 15

Gln Ser Thr Asn Pro Tyr Ala Ile Arg Val Lys Gly Leu Ser Glu Ala
20 25 30

Leu Ser Lys Asp Glu Leu Val Ile Ala Leu Leu His Ile Ala Lys Arg
35 40 45

Arg Gly Ile His
50

<210> 39
<211> 52
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> 来自8小结构域文库的巴氏葡萄球菌小结构域2

<400> 39

Lys Leu Leu Leu Thr Glu Tyr Asp Leu Ile Asn Lys Glu Gln Ile Pro
1 5 10 15

Thr Ser Asn Asn Pro Tyr Gln Ile Arg Val Lys Gly Leu Ser Glu Ile
20 25 30

Leu Ser Lys Asp Glu Leu Ala Ile Ala Leu Leu His Leu Ala Lys Arg
35 40 45

[0107] Arg Gly Ile His
50

<210> 40
<211> 52
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> 来自8小结构域文库的猪葡萄球菌小结构域2

<400> 40

Lys His Leu Leu Ala Glu Tyr Asp Leu Leu Asp Leu Thr Asn Ile Pro
1 5 10 15

Lys Ser Thr Asn Pro Tyr Gln Thr Arg Val Lys Gly Leu Asn Glu Lys
20 25 30

Leu Ser Lys Asp Glu Leu Val Ile Ala Leu Leu His Ile Ala Lys Arg
35 40 45

Arg Gly Ile His

50

<210> 41
<211> 51
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> 来自8小结构域文库的Staphylococcus microti小结构域2

<400> 41

Lys Ser Leu Leu Ser Glu Tyr Lys Ile Ile Ser Gly Leu Ala Pro Thr
1 5 10 15

Asn Asn Gln Pro Tyr Asn Ile Arg Val Lys Gly Leu Thr Glu Gln Leu
20 25 30

Thr Lys Asp Glu Leu Ala Val Ala Leu Leu His Ile Ala Lys Arg Arg
35 40 45

Gly Ile His
[0108] 50

<210> 42
<211> 92
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> 来自8小结构域文库的路邓葡萄球菌小结构域3

<400> 42

Lys Ile Asp Val Ile Asp Ser Asn Asp Asp Val Gly Asn Glu Leu Ser
1 5 10 15

Thr Lys Glu Gln Leu Asn Lys Asn Ser Lys Leu Leu Lys Asp Lys Phe
20 25 30

Val Cys Gln Ile Gln Leu Glu Arg Met Asn Glu Gly Gln Val Arg Gly
35 40 45

Glu Lys Asn Arg Phe Lys Thr Ala Asp Ile Ile Lys Glu Ile Ile Gln
50 55 60

Leu Leu Asn Val Gln Lys Asn Phe His Gln Leu Asp Glu Asn Phe Ile
65 70 75 80

Asn Lys Tyr Ile Glu Leu Val Glu Met Arg Arg Glu
85 90

<210> 43
<211> 92
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> 来自8小结构域文库的巴氏葡萄球菌小结构域3

<400> 43

Asn Ile Asn Val Ser Ser Glu Asp Glu Asp Ala Ser Asn Glu Leu Ser
1 5 10 15

Thr Lys Glu Gln Ile Asn Arg Asn Asn Lys Leu Leu Lys Asp Lys Tyr
20 25 30

[0109] Val Cys Glu Val Gln Leu Gln Arg Leu Lys Glu Gly Gln Ile Arg Gly
35 40 45

Glu Lys Asn Arg Phe Lys Thr Thr Asp Ile Leu Lys Glu Ile Asp Gln
50 55 60

Leu Leu Lys Val Gln Lys Asp Tyr His Asn Leu Asp Ile Asp Phe Ile
65 70 75 80

Asn Gln Tyr Lys Glu Ile Val Glu Thr Arg Arg Glu
85 90

<210> 44
<211> 93
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> 来自8小结构域文库的猪葡萄球菌小结构域3

<400> 44

Asn Val Asn Val Met Met Asp Asp Asn Asp Ser Gly Asn Glu Leu Ser

1 5 10 15

Thr Lys Asp Gln Leu Lys Lys Asn Ala Lys Ala Leu Ser Asp Lys Tyr
20 25 30

Val Cys Glu Leu Gln Leu Glu Arg Phe Glu Gln Asp Tyr Lys Val Arg
35 40 45

Gly Glu Lys Asn Arg Phe Lys Thr Glu Asp Phe Val Arg Glu Ala Arg
50 55 60

Lys Leu Leu Glu Thr Gln Ser Lys Phe Phe Glu Ile Asp Gln Thr Phe
65 70 75 80

Ile Met Arg Tyr Ile Glu Leu Ile Glu Thr Arg Arg Glu
85 90

<210> 45
<211> 95
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

[0110]

<220>
<221> misc_feature
<223> 来自8小结构域文库的Staphylococcus microti小结构域3

<400> 45

Asn Val Asp Val Ala Ala Asp Lys Glu Glu Thr Ala Ser Asp Ser Leu
1 5 10 15

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

Tyr Val Cys Glu Leu Gln Lys Glu Arg Leu Glu Asn Glu Gly His Val
35 40 45

Arg Gly Val Glu Asn Arg Phe Leu Thr Lys Asp Ile Val Arg Glu Ala
50 55 60

Lys Lys Ile Ile Asp Thr Gln Met Gln Tyr Tyr Pro Glu Ile Asp Glu
65 70 75 80

Thr Phe Lys Glu Lys Tyr Ile Ser Leu Val Glu Thr Arg Arg Glu
85 90 95

<210> 46
<211> 111

<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> 来自8小结构域文库的路邓葡萄球菌小结构域4

<400> 46

Tyr Phe Glu Gly Pro Gly Lys Gly Ser Pro Tyr Gly Trp Glu Gly Asp
1 5 10 15

Pro Lys Ala Trp Tyr Glu Thr Leu Met Gly His Cys Thr Tyr Phe Pro
20 25 30

Asp Glu Leu Arg Ser Val Lys Tyr Ala Tyr Ser Ala Asp Leu Phe Asn
35 40 45

Ala Leu Asn Asp Leu Asn Asn Leu Val Ile Gln Arg Asp Gly Leu Ser
50 55 60

Lys Leu Glu Tyr His Glu Lys Tyr His Ile Ile Glu Asn Val Phe Lys
65 70 75 80

[0111]

Gln Lys Lys Lys Pro Thr Leu Lys Gln Ile Ala Asn Glu Ile Asn Val
85 90 95

Asn Pro Glu Asp Ile Lys Gly Tyr Arg Ile Thr Lys Ser Gly Lys
100 105 110

<210> 47
<211> 111
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> 来自8小结构域文库的巴氏葡萄球菌小结构域4

<400> 47

Tyr Phe Glu Gly Pro Gly Gln Gly Ser Pro Phe Gly Trp Asn Gly Asp
1 5 10 15

Leu Lys Lys Trp Tyr Glu Met Leu Met Gly His Cys Thr Tyr Phe Pro
20 25 30

Gln Glu Leu Arg Ser Val Lys Tyr Ala Tyr Ser Ala Asp Leu Phe Asn
35 40 45

Ala Leu Asn Asp Leu Asn Asn Leu Ile Ile Gln Arg Asp Asn Ser Glu
50 55 60

Lys Leu Glu Tyr His Glu Lys Tyr His Ile Ile Glu Asn Val Phe Lys
65 70 75 80

Gln Lys Lys Lys Pro Thr Leu Lys Gln Ile Ala Lys Glu Ile Gly Val
85 90 95

Asn Pro Glu Asp Ile Lys Gly Tyr Arg Ile Thr Lys Ser Gly Thr
100 105 110

<210> 48
<211> 111
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
[0112] <223> 来自8小结构域文库的猪葡萄球菌小结构域4

<400> 48

Tyr Phe Glu Gly Pro Gly Lys Gly Ser Pro Phe Gly Trp Glu Gly Asn
1 5 10 15

Ile Lys Lys Trp Phe Glu Gln Met Met Gly His Cys Thr Tyr Phe Pro
20 25 30

Glu Glu Leu Arg Ser Val Lys Tyr Ser Tyr Ser Ala Glu Leu Phe Asn
35 40 45

Ala Leu Asn Asp Leu Asn Asn Leu Val Ile Thr Arg Asp Glu Asp Ala
50 55 60

Lys Leu Asn Tyr Gly Glu Lys Phe Gln Ile Ile Glu Asn Val Phe Lys
65 70 75 80

Gln Lys Lys Thr Pro Asn Leu Lys Gln Ile Ala Ile Glu Ile Gly Val
85 90 95

His Glu Thr Glu Ile Lys Gly Tyr Arg Val Asn Lys Ser Gly Lys
100 105 110

<210> 49

<210> 111
<211> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> 来自8小结构域文库的Staphylococcus microti小结构域4

<400> 49

Tyr Tyr Glu Gly Pro Gly Lys Gly Ser Pro Tyr Gly Trp Asp Ala Asp
1 5 10 15

Val Lys Lys Trp Tyr Gln Leu Met Met Gly His Cys Thr Tyr Phe Pro
20 25 30

Val Glu Phe Arg Ser Val Lys Tyr Ala Tyr Thr Ala Asp Leu Tyr Asn
35 40 45

Ala Leu Asn Asp Leu Asn Asn Leu Thr Ile Ala Arg Asp Asp Asn Pro
50 55 60

[0113] Lys Leu Glu Tyr His Glu Lys Tyr His Ile Ile Glu Asn Val Phe Lys
65 70 75 80

Gln Lys Arg Asn Pro Thr Leu Lys Gln Ile Ala Lys Glu Ile Gly Val
85 90 95

Asn Asp Ile Asn Ile Ser Gly Tyr Arg Val Thr Lys Ser Gly Lys
100 105 110

<210> 50
<211> 131
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> 来自8小结构域文库的路邓葡萄球菌小结构域5

<400> 50

Pro Gln Phe Thr Glu Phe Lys Leu Tyr His Asp Leu Lys Ser Val Leu
1 5 10 15

Phe Asp Gln Ser Ile Leu Glu Asn Glu Asp Val Leu Asp Gln Ile Ala
20 25 30

Glu Ile Leu Thr Ile Tyr Gln Asp Lys Asp Ser Ile Lys Ser Lys Leu
35 40 45

Thr Glu Leu Asp Ile Leu Leu Asn Glu Glu Asp Lys Glu Asn Ile Ala
50 55 60

Gln Leu Thr Gly Tyr Thr Gly Thr His Arg Leu Ser Leu Lys Cys Ile
65 70 75 80

Arg Leu Val Leu Glu Glu Gln Trp Tyr Ser Ser Arg Asn Gln Met Glu
85 90 95

Ile Phe Thr His Leu Asn Ile Lys Pro Lys Lys Ile Asn Leu Thr Ala
100 105 110

Ala Asn Lys Ile Pro Lys Ala Met Ile Asp Glu Phe Ile Leu Ser Pro
115 120 125

Val Val Lys
130

<210> 51
<211> 131
[0114] <212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> 来自8小结构域文库的巴氏葡萄球菌小结构域5

<400> 51

Pro Gln Phe Thr Glu Phe Lys Leu Tyr His Asp Leu Lys Ser Ile Val
1 5 10 15

Phe Asp Lys Ser Ile Leu Glu Asn Glu Ala Ile Leu Asp Gln Ile Ala
20 25 30

Glu Ile Leu Thr Ile Tyr Gln Asp Glu Gln Ser Ile Lys Glu Glu Leu
35 40 45

Asn Lys Leu Pro Glu Ile Leu Asn Glu Gln Asp Lys Ala Glu Ile Ala
50 55 60

Lys Leu Ile Gly Tyr Asn Gly Thr His Arg Leu Ser Leu Lys Cys Ile
65 70 75 80

```
His Leu Ile Asn Glu Glu Leu Trp Gln Thr Ser Arg Asn Gln Met Glu
                85                  90                  95

Ile Phe Asn Tyr Leu Asn Ile Lys Pro Asn Lys Val Asp Leu Ser Glu
            100                 105                 110

Gln Asn Lys Ile Pro Lys Asp Met Val Asn Asp Phe Ile Leu Ser Pro
        115                 120                 125

Val Val Lys
    130
```

<210> 52
<211> 131
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> 来自8小结构域文库的猪葡萄球菌小结构域5

<400> 52

[0115]

```
Pro Glu Phe Thr Gln Phe Lys Leu Tyr His Asp Leu Lys Asn Ile Phe
1               5                   10                  15

Lys Asp Pro Lys Tyr Leu Asn Asp Ile Gln Leu Met Asp Asn Ile Ala
            20                  25                  30

Glu Ile Ile Thr Ile Tyr Gln Asp Ala Glu Ser Ile Ile Lys Glu Leu
        35                  40                  45

Asn Gln Leu Pro Glu Leu Leu Ser Glu Arg Glu Lys Glu Lys Ile Ser
    50                  55                  60

Ala Leu Ser Gly Tyr Ser Gly Thr His Arg Leu Ser Leu Lys Cys Ile
65                  70                  75                  80

Asn Leu Leu Leu Asp Asp Leu Trp Glu Ser Ser Leu Asn Gln Met Glu
                85                  90                  95

Leu Phe Thr Lys Leu Asn Leu Lys Pro Lys Lys Ile Asp Leu Ser Gln
            100                 105                 110

Gln His Lys Ile Pro Ser Lys Leu Val Asp Asp Phe Ile Leu Ser Pro
        115                 120                 125

Val Val Lys
```

130

<210> 53
<211> 131
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> 来自8小结构域文库的Staphylococcus microti小结构域5

<400> 53

Pro Gln Phe Thr Ser Phe Lys Leu Phe His Asp Leu Lys Lys Val Val
1 5 10 15

Lys Asp His Ala Ile Leu Asp Asp Ile Asp Leu Leu Asn Gln Ile Ala
20 25 30

Glu Ile Leu Thr Ile Tyr Gln Asp Lys Asp Ser Ile Val Ala Glu Leu
35 40 45

Gly Gln Leu Glu Tyr Leu Met Ser Glu Ala Asp Lys Gln Ser Ile Ser
50 55 60

[0116]

Glu Leu Thr Gly Tyr Thr Gly Thr His Ser Leu Ser Leu Lys Cys Met
65 70 75 80

Asn Met Ile Ile Asp Glu Leu Trp His Ser Ser Met Asn Gln Met Glu
85 90 95

Val Phe Thr Tyr Leu Asn Met Arg Pro Lys Lys Tyr Glu Leu Lys Gly
100 105 110

Tyr Gln Arg Ile Pro Thr Asp Met Ile Asp Asp Ala Ile Leu Ser Pro
115 120 125

Val Val Lys
130

<210> 54
<211> 89
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>

<220>
<221> misc_feature
<223> 来自8小结构域文库的路邓葡萄球菌小结构域6

<400> 54

Arg Thr Phe Gly Gln Ala Ile Asn Leu Ile Asn Lys Ile Ile Glu Lys
1 5 10 15

Tyr Gly Val Pro Glu Asp Ile Ile Ile Glu Leu Ala Arg Glu Asn Asn
20 25 30

Ser Lys Asp Lys Gln Lys Phe Ile Asn Glu Met Gln Lys Lys Asn Glu
35 40 45

Asn Thr Arg Lys Arg Ile Asn Glu Ile Ile Gly Lys Tyr Gly Asn Gln
50 55 60

Asn Ala Lys Arg Leu Val Glu Lys Ile Arg Leu His Asp Glu Gln Glu
65 70 75 80

Gly Lys Cys Leu Tyr Ser Leu Glu Ser
85

[0117]

<210> 55
<211> 89
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> 来自8小结构域文库的巴氏葡萄球菌小结构域6

<400> 55

Arg Thr Phe Ile Gln Ser Ile Asn Val Ile Asn Lys Val Ile Glu Lys
1 5 10 15

Tyr Gly Ile Pro Glu Asp Ile Ile Ile Glu Leu Ala Arg Glu Asn Asn
20 25 30

Ser Asp Asp Arg Lys Lys Phe Ile Asn Asn Leu Gln Lys Lys Asn Glu
35 40 45

Ala Thr Arg Lys Arg Ile Asn Glu Ile Ile Gly Gln Thr Gly Asn Gln
50 55 60

Asn Ala Lys Arg Ile Val Glu Lys Ile Arg Leu His Asp Gln Gln Glu
65 70 75 80

Gly Lys Cys Leu Tyr Ser Leu Glu Ser
85

<210> 56
<211> 89
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> 来自8小结构域文库的猪葡萄球菌小结构域6

<400> 56

Arg Ala Phe Ile Gln Ser Ile Gln Val Val Asn Ala Ile Ile Asp Lys
1 5 10 15

Tyr Gly Leu Pro Glu Asp Ile Ile Ile Glu Leu Ala Arg Glu Asn Asn
20 25 30

Ser Asp Asp Arg Arg Lys Phe Leu Asn Gln Leu Gln Lys Gln Asn Glu
35 40 45

[0118]

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

Asn Ala Lys Arg Ile Val Gln Lys Ile Lys Leu His Asn Met Gln Glu
65 70 75 80

Gly Lys Cys Leu Tyr Ser Leu Lys Asp
85

<210> 57
<211> 89
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> 来自8小结构域文库的Staphylococcus microti小结构域6

<400> 57

Arg Ser Phe Lys Gln Ala Ile Gly Val Val Asn Ala Ile Ile Lys Lys
1 5 10 15

Tyr Gly Leu Pro Lys Asp Ile Ile Ile Glu Leu Ala Arg Glu Ser Asn
20 25 30

Ser Ala Glu Lys Ser Arg Tyr Leu Arg Ala Ile Gln Lys Lys Asn Glu
35 40 45

Lys Thr Arg Glu Arg Ile Glu Ala Ile Ile Lys Glu Tyr Gly Asn Glu
50 55 60

Asn Ala Lys Gly Leu Val Gln Lys Ile Lys Leu His Asp Ala Gln Glu
65 70 75 80

Gly Lys Cys Leu Tyr Ser Leu Lys Asp
85

<210> 58
<211> 116
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> 来自8小结构域文库的路邓葡萄球菌小结构域7

<400> 58

[0119]

Ile Pro Leu Glu Asp Leu Leu Asn Asn Pro Asn His Tyr Glu Val Asp
1 5 10 15

His Ile Ile Pro Arg Ser Val Ser Phe Asp Asn Ser Tyr His Asn Lys
20 25 30

Val Leu Val Lys Gln Ser Glu Asn Ser Lys Lys Ser Asn Leu Thr Pro
35 40 45

Tyr Gln Tyr Phe Asn Ser Gly Lys Ser Lys Leu Ser Tyr Asn Gln Phe
50 55 60

Lys Gln His Ile Leu Asn Leu Ser Lys Ser Gln Asp Arg Ile Ser Lys
65 70 75 80

Lys Lys Lys Glu Tyr Leu Leu Glu Glu Arg Asp Ile Asn Lys Phe Glu
85 90 95

Val Gln Lys Glu Phe Ile Asn Arg Asn Leu Val Asp Thr Arg Tyr Ala
100 105 110

Thr Arg Glu Leu
115

<210> 59
<211> 116
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> 来自8小结构域文库的巴氏葡萄球菌小结构域7

<400> 59

Ile Ala Leu Met Asp Leu Leu Asn Asn Pro Gln Asn Tyr Glu Val Asp
1 5 10 15

His Ile Ile Pro Arg Ser Val Ala Phe Asp Asn Ser Ile His Asn Lys
20 25 30

Val Leu Val Lys Gln Ile Glu Asn Ser Lys Lys Gly Asn Arg Thr Pro
35 40 45

Tyr Gln Tyr Leu Asn Ser Ser Asp Ala Lys Leu Ser Tyr Asn Gln Phe
50 55 60

[0120] Lys Gln His Ile Leu Asn Leu Ser Lys Ser Lys Asp Arg Ile Ser Lys
65 70 75 80

Lys Lys Lys Asp Tyr Leu Leu Glu Glu Arg Asp Ile Asn Lys Phe Glu
85 90 95

Val Gln Lys Glu Phe Ile Asn Arg Asn Leu Val Asp Thr Arg Tyr Ala
100 105 110

Thr Arg Glu Leu
115

<210> 60
<211> 116
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> 来自8小结构域文库的猪葡萄球菌小结构域7

<400> 60

Ile Pro Leu Glu Asp Leu Leu Arg Asn Pro His His Tyr Glu Val Asp

1 5 10 15

His Ile Ile Pro Arg Ser Val Ala Phe Asp Asn Ser Met His Asn Lys
20 25 30

Val Leu Val Arg Ala Asp Glu Asn Ser Lys Lys Gly Asn Arg Thr Pro
35 40 45

Tyr Gln Tyr Leu Asn Ser Ser Glu Ser Ser Leu Ser Tyr Asn Glu Phe
50 55 60

Lys Gln His Ile Leu Asn Leu Ser Lys Thr Lys Asp Arg Ile Thr Lys
65 70 75 80

Lys Lys Arg Glu Tyr Leu Leu Glu Glu Arg Asp Ile Asn Lys Phe Asp
85 90 95

Val Gln Lys Glu Phe Ile Asn Arg Asn Leu Val Asp Thr Arg Tyr Ala
100 105 110

Thr Arg Glu Leu
115

[0121]

<210> 61
<211> 116
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> 来自8小结构域文库的Staphylococcus microti小结构域7

<400> 61

Ile Pro Leu Glu Asp Leu Leu Arg Asn Pro Asn Asn Tyr Asp Ile Asp
1 5 10 15

His Ile Ile Pro Arg Ser Val Ser Phe Asp Asp Ser Met His Asn Lys
20 25 30

Val Leu Val Arg Arg Glu Gln Asn Ala Lys Lys Asn Asn Gln Thr Pro
35 40 45

Tyr Gln Tyr Leu Thr Ser Gly Tyr Ala Asp Ile Lys Tyr Ser Val Phe
50 55 60

Lys Gln His Val Leu Asn Leu Ala Glu Asn Lys Asp Arg Met Thr Lys
65 70 75 80

Lys Lys Arg Glu Tyr Leu Leu Glu Glu Arg Asn Ile Asn Lys Tyr Asp
85 90 95

Val Gln Lys Glu Phe Ile Asn Arg Asn Leu Val Asp Thr Arg Tyr Thr
100 105 110

Thr Arg Glu Leu
115

<210> 62
<211> 396
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> 来自8小结构域文库的路邓葡萄球菌小结构域8

<400> 62

Thr Asn Tyr Leu Lys Ala Tyr Phe Ser Ala Asn Asn Met Asn Val Lys
1 5 10 15

[0122]

Val Lys Thr Ile Asn Gly Ser Phe Thr Asp Tyr Leu Arg Lys Val Trp
20 25 30

Lys Phe Lys Lys Glu Arg Asn His Gly Tyr Lys His His Ala Glu Asp
35 40 45

Ala Leu Ile Ile Ala Asn Ala Asp Phe Leu Phe Lys Glu Asn Lys Lys
50 55 60

Leu Lys Ala Val Asn Ser Val Leu Glu Lys Pro Glu Ile Glu Thr Lys
65 70 75 80

Gln Leu Asp Ile Gln Val Asp Ser Glu Asp Asn Tyr Ser Glu Met Phe
85 90 95

Ile Ile Pro Lys Gln Val Gln Asp Ile Lys Asp Phe Arg Asn Phe Lys
100 105 110

Tyr Ser His Arg Val Asp Lys Lys Pro Asn Arg Gln Leu Ile Asn Asp
115 120 125

Thr Leu Tyr Ser Thr Arg Lys Lys Asp Asn Ser Thr Tyr Ile Val Gln
130 135 140

```
Thr Ile Lys Asp Ile Tyr Ala Lys Asp Asn Thr Thr Leu Lys Lys Gln
145                 150                 155                 160

Phe Asp Lys Ser Pro Glu Lys Phe Leu Met Tyr Gln His Asp Pro Arg
                165                 170                 175

Thr Phe Glu Lys Leu Glu Val Ile Met Lys Gln Tyr Ala Asn Glu Lys
            180                 185                 190

Asn Pro Leu Ala Lys Tyr His Glu Glu Thr Gly Glu Tyr Leu Thr Lys
        195                 200                 205

Tyr Ser Lys Lys Asn Asn Gly Pro Ile Val Lys Ser Leu Lys Tyr Ile
    210                 215                 220

Gly Asn Lys Leu Gly Ser His Leu Asp Val Thr His Gln Phe Lys Ser
225                 230                 235                 240

Ser Thr Lys Lys Leu Val Lys Leu Ser Ile Lys Pro Tyr Arg Phe Asp
                245                 250                 255

Val Tyr Leu Thr Asp Lys Gly Tyr Lys Phe Ile Thr Ile Ser Tyr Leu
            260                 265                 270
```

[0123]

```
Asp Val Leu Lys Lys Asp Asn Tyr Tyr Tyr Ile Pro Glu Gln Lys Tyr
        275                 280                 285

Asp Lys Leu Lys Leu Gly Lys Ala Ile Asp Lys Asn Ala Lys Phe Ile
    290                 295                 300

Ala Ser Phe Tyr Lys Asn Asp Leu Ile Lys Leu Asp Gly Glu Ile Tyr
305                 310                 315                 320

Lys Ile Ile Gly Val Asn Ser Asp Thr Arg Asn Met Ile Glu Leu Asp
                325                 330                 335

Leu Pro Asp Ile Arg Tyr Lys Glu Tyr Cys Glu Leu Asn Asn Ile Lys
            340                 345                 350

Gly Glu Pro Arg Ile Lys Lys Thr Ile Gly Lys Lys Val Asn Ser Ile
        355                 360                 365

Glu Lys Leu Thr Thr Asp Val Leu Gly Asn Val Phe Thr Asn Thr Gln
    370                 375                 380

Tyr Thr Lys Pro Gln Leu Leu Phe Lys Arg Gly Asn
385                 390                 395
```

<210> 63
<211> 396
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> 来自8小结构域文库的巴氏葡萄球菌小结构域8

<400> 63

Thr Ser Tyr Leu Lys Ala Tyr Phe Ser Ala Asn Asn Met Asp Val Lys
1 5 10 15

Val Lys Thr Ile Asn Gly Ser Phe Thr Asn His Leu Arg Lys Val Trp
20 25 30

Arg Phe Asp Lys Tyr Arg Asn His Gly Tyr Lys His His Ala Glu Asp
35 40 45

Ala Leu Ile Ile Ala Asn Ala Asp Phe Leu Phe Lys Glu Asn Lys Lys
50 55 60

[0124]

Leu Gln Asn Thr Asn Lys Ile Leu Glu Lys Pro Thr Ile Glu Asn Asn
65 70 75 80

Thr Lys Lys Val Thr Val Glu Lys Glu Glu Asp Tyr Asn Asn Val Phe
85 90 95

Glu Thr Pro Lys Leu Val Glu Asp Ile Lys Gln Tyr Arg Asp Tyr Lys
100 105 110

Phe Ser His Arg Val Asp Lys Lys Pro Asn Arg Gln Leu Ile Asn Asp
115 120 125

Thr Leu Tyr Ser Thr Arg Met Lys Asp Glu His Asp Tyr Ile Val Gln
130 135 140

Thr Ile Thr Asp Ile Tyr Gly Lys Asp Asn Thr Asn Leu Lys Lys Gln
145 150 155 160

Phe Asn Lys Asn Pro Glu Lys Phe Leu Met Tyr Gln Asn Asp Pro Lys
165 170 175

Thr Phe Glu Lys Leu Ser Ile Ile Met Lys Gln Tyr Ser Asp Glu Lys
180 185 190

Asn Pro Leu Ala Lys Tyr Tyr Glu Glu Thr Gly Glu Tyr Leu Thr Lys
195 200 205

Tyr Ser Lys Lys Asn Asn Gly Pro Ile Val Lys Lys Ile Lys Leu Leu
210 215 220

Gly Asn Lys Val Gly Asn His Leu Asp Val Thr Asn Lys Tyr Glu Asn
225 230 235 240

Ser Thr Lys Lys Leu Val Lys Leu Ser Ile Lys Asn Tyr Arg Phe Asp
245 250 255

Val Tyr Leu Thr Glu Lys Gly Tyr Lys Phe Val Thr Ile Ala Tyr Leu
260 265 270

Asn Val Phe Lys Lys Asp Asn Tyr Tyr Tyr Ile Pro Lys Asp Lys Tyr
275 280 285

Gln Glu Leu Lys Glu Lys Lys Lys Ile Lys Asp Thr Asp Gln Phe Ile
290 295 300

Ala Ser Phe Tyr Lys Asn Asp Leu Ile Lys Leu Asn Gly Asp Leu Tyr
305 310 315 320

[0125]

Lys Ile Ile Gly Val Asn Ser Asp Asp Arg Asn Ile Ile Glu Leu Asp
325 330 335

Tyr Tyr Asp Ile Lys Tyr Lys Asp Tyr Cys Glu Ile Asn Asn Ile Lys
340 345 350

Gly Glu Pro Arg Ile Lys Lys Thr Ile Gly Lys Lys Thr Glu Ser Ile
355 360 365

Glu Lys Phe Thr Thr Asp Val Leu Gly Asn Leu Tyr Leu His Ser Thr
370 375 380

Glu Lys Ala Pro Gln Leu Ile Phe Lys Arg Gly Leu
385 390 395

<210> 64
<211> 393
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> 来自8小结构域文库的猪葡萄球菌小结构域8

<400> 64

Thr Ser Leu Leu Lys Ala Tyr Phe Ser Ala Asn Asn Leu Asp Val Lys
1 5 10 15

Val Lys Thr Ile Asn Gly Ser Phe Thr Asn Tyr Leu Arg Lys Val Trp
20 25 30

Lys Phe Asp Lys Asp Arg Asn Lys Gly Tyr Lys His His Ala Glu Asp
35 40 45

Ala Leu Ile Ile Ala Asn Ala Asp Phe Leu Phe Lys His Asn Lys Lys
50 55 60

Leu Arg Asn Ile Asn Lys Val Leu Asp Ala Pro Ser Lys Glu Val Asp
65 70 75 80

Lys Lys Arg Val Thr Val Gln Ser Glu Asp Glu Tyr Asn Gln Ile Phe
85 90 95

Glu Asp Thr Gln Lys Ala Gln Ala Ile Lys Lys Phe Glu Ile Arg Lys
100 105 110

[0126] Phe Ser His Arg Val Asp Lys Lys Pro Asn Arg Gln Leu Ile Asn Asp
115 120 125

Thr Leu Tyr Ser Thr Arg Asn Ile Asp Gly Ile Glu Tyr Val Val Glu
130 135 140

Ser Ile Lys Asp Ile Tyr Ser Val Asn Asn Asp Lys Val Lys Thr Lys
145 150 155 160

Phe Lys Lys Asp Pro His Arg Leu Leu Met Tyr Arg Asn Asp Pro Gln
165 170 175

Thr Phe Glu Lys Phe Glu Lys Val Phe Lys Gln Tyr Glu Ser Glu Lys
180 185 190

Asn Pro Phe Ala Lys Tyr Tyr Glu Glu Thr Gly Glu Lys Ile Arg Lys
195 200 205

Phe Ser Lys Thr Gly Gln Gly Pro Tyr Ile Asn Lys Ile Lys Tyr Leu
210 215 220

Arg Glu Arg Leu Gly Arg His Cys Asp Val Thr Asn Lys Tyr Ile Asn
225 230 235 240

Ser Arg Asn Lys Ile Val Gln Leu Lys Ile Tyr Ser Tyr Arg Phe Asp

245 250 255

Ile Tyr Gln Tyr Gly Asn Asn Tyr Lys Met Ile Thr Ile Ser Tyr Ile
260 265 270

Asp Leu Glu Gln Lys Ser Asn Tyr Tyr Tyr Ile Ser Arg Glu Lys Tyr
275 280 285

Glu Gln Lys Lys Lys Asp Lys Gln Ile Asp Asp Ser Tyr Lys Phe Ile
290 295 300

Gly Ser Phe Tyr Lys Asn Asp Ile Ile Asn Tyr Asn Gly Glu Met Tyr
305 310 315 320

Arg Val Ile Gly Val Asn Asp Ser Glu Lys Asn Lys Ile Gln Leu Asp
325 330 335

Met Ile Asp Ile Ser Ile Lys Asp Tyr Met Glu Leu Asn Asn Ile Lys
340 345 350

Lys Thr Gly Val Ile Tyr Lys Thr Ile Gly Lys Ser Thr Thr His Ile
355 360 365

[0127] Glu Lys Tyr Thr Thr Asp Ile Leu Gly Asn Leu Tyr Lys Ala Ala Pro
370 375 380

Pro Lys Lys Pro Gln Leu Ile Phe Lys
385 390

<210> 65
<211> 392
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> 来自8小结构域文库的Staphylococcus microti小结构域8

<400> 65

Thr Thr Leu Leu Lys Thr Tyr Phe Thr Ile Asn Asn Leu Asp Val Lys
1 5 10 15

Val Lys Thr Ile Asn Gly Ser Phe Thr Asp Phe Leu Arg Lys Arg Trp
20 25 30

Gly Phe Lys Lys Asn Arg Asp Glu Gly Tyr Lys His His Ala Glu Asp
35 40 45

```
Ala Leu Ile Ile Ala Asn Ala Asp Tyr Leu Phe Lys Glu His Lys Leu
    50                  55                  60

Leu Lys Glu Ile Lys Asp Val Ser Asp Leu Ala Gly Asp Glu Arg Asn
65                  70                  75                  80

Ser Asn Val Lys Asp Glu Asp Gln Tyr Glu Glu Val Phe Gly Gly Tyr
                85                  90                  95

Phe Lys Ile Glu Asp Ile Lys Lys Tyr Lys Ile Lys Lys Phe Ser His
            100                 105                 110

Arg Val Asp Lys Lys Pro Asn Arg Gln Leu Ile Asn Asp Thr Ile Tyr
        115                 120                 125

Ser Thr Arg Val Lys Asp Asp Lys Arg Tyr Leu Ile Asn Thr Leu Lys
    130                 135                 140

Asn Leu Tyr Asp Lys Ser Asn Gly Asp Leu Lys Glu Arg Met Gln Lys
145                 150                 155                 160

Asp Pro Glu Ser Leu Leu Met Tyr His His Asp Pro Gln Thr Phe Glu
                165                 170                 175
```

[0128]

```
Lys Leu Lys Ile Val Met Ser Gln Tyr Glu Asn Glu Lys Asn Pro Leu
            180                 185                 190

Ala Lys Tyr Phe Glu Glu Thr Gly Gln Tyr Leu Thr Lys Tyr Ala Lys
        195                 200                 205

His Asp Asn Gly Pro Ala Ile His Lys Ile Lys Tyr Tyr Gly Asn Lys
    210                 215                 220

Leu Val Glu His Leu Asp Ile Thr Lys Asn Tyr His Asn Pro Gln Asn
225                 230                 235                 240

Lys Val Val Gln Leu Ser Gln Lys Ser Phe Arg Phe Asp Val Tyr Gln
                245                 250                 255

Thr Asp Lys Gly Tyr Lys Phe Ile Ser Ile Ala Tyr Leu Thr Leu Lys
            260                 265                 270

Asn Glu Lys Asn Tyr Tyr Ala Ile Ser Gln Glu Lys Tyr Asp Gln Leu
        275                 280                 285

Lys Ser Glu Lys Lys Ile Ser Asn Asn Ala Val Phe Ile Gly Ser Phe
    290                 295                 300
```

Tyr Thr Ser Asp Ile Ile Glu Ile Asn Asn Glu Lys Phe Arg Val Ile
305 310 315 320

Gly Val Asn Ser Asp Lys Asn Asn Leu Ile Glu Val Asp Arg Ile Asp
325 330 335

Ile Arg Gln Lys Glu Phe Ile Glu Leu Glu Glu Glu Lys Lys Asn Asn
340 345 350

Arg Ile Lys Val Thr Ile Gly Arg Lys Thr Thr Asn Ile Glu Lys Phe
355 360 365

His Thr Asp Ile Leu Gly Asn Met Tyr Lys Ser Lys Arg Pro Lys Ala
370 375 380

Pro Gln Leu Val Phe Lys Lys Gly
385 390

<210> 66
<211> 41
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

[0129]

<220>
<221> misc_feature
<223> 来自12小结构域文库的路邓葡萄球菌小结构域1

<400> 66

Met Asn Gln Lys Phe Ile Leu Gly Leu Asp Ile Gly Ile Thr Ser Val
1 5 10 15

Gly Tyr Gly Leu Ile Asp Tyr Glu Thr Lys Asn Ile Ile Asp Ala Gly
20 25 30

Val Arg Leu Phe Pro Glu Ala Asn Val
35 40

<210> 67
<211> 41
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature

<223> 来自12小结构域文库的巴氏葡萄球菌小结构域1

<400> 67

Met Lys Glu Lys Tyr Ile Leu Gly Leu Asp Leu Gly Ile Thr Ser Val
1 5 10 15

Gly Tyr Gly Ile Ile Asn Phe Glu Thr Lys Lys Ile Ile Asp Ala Gly
20 25 30

Val Arg Leu Phe Pro Glu Ala Asn Val
35 40

<210> 68
<211> 40
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> 来自12小结构域文库的猪葡萄球菌小结构域1

<400> 68

[0130]
Met Asn Asn Tyr Ile Leu Gly Leu Asp Ile Gly Ile Thr Ser Val Gly
1 5 10 15

Tyr Gly Ile Val Asp Ser Asp Thr Arg Glu Ile Lys Asp Ala Gly Val
20 25 30

Arg Leu Phe Pro Glu Ala Asn Val
35 40

<210> 69
<211> 41
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> 来自12小结构域文库的Staphylococcus microti小结构域1

<400> 69

Met Glu Lys Asp Tyr Ile Leu Gly Leu Asp Ile Gly Ile Gly Ser Val
1 5 10 15

Gly Tyr Gly Leu Ile Asp Tyr Asp Thr Lys Ser Ile Ile Asp Ala Gly
20 25 30

Val Arg Leu Phe Pro Glu Ala Asn Ala
35 40

<210> 70
<211> 23
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> 来自12小结构域文库的路邓葡萄球菌小结构域2

<400> 70

Glu Asn Asn Glu Gly Arg Arg Ser Lys Arg Gly Ser Arg Arg Leu Lys
1 5 10 15

Arg Arg Arg Ile His Arg Leu
20

[0131]

<210> 71
<211> 23
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> 来自12小结构域文库的巴氏葡萄球菌小结构域2

<400> 71

Asp Asn Asn Glu Gly Arg Arg Ser Lys Arg Gly Ser Arg Arg Leu Lys
1 5 10 15

Arg Arg Arg Ile His Arg Leu
20

<210> 72
<211> 23
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> 来自12小结构域文库的猪葡萄球菌小结构域2

<400> 72

Asp Asn Asn Glu Gly Arg Arg Ser Lys Arg Gly Ala Arg Arg Leu Lys
1 5 10 15

Arg Arg Arg Ile His Arg Leu
20

<210> 73
<211> 23
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> 来自12小结构域文库的Staphylococcus microti小结构域2

<400> 73

Asp Asn Asn Leu Gly Arg Arg Ala Lys Arg Gly Ala Arg Arg Leu Lys
1 5 10 15

Arg Arg Arg Ile His Arg Leu
20

[0132]

<210> 74
<211> 55
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> 来自12小结构域文库的路邓葡萄球菌小结构域3

<400> 74

Glu Arg Val Lys Lys Leu Leu Glu Asp Tyr Asn Leu Leu Asp Gln Ser
1 5 10 15

Gln Ile Pro Gln Ser Thr Asn Pro Tyr Ala Ile Arg Val Lys Gly Leu
20 25 30

Ser Glu Ala Leu Ser Lys Asp Glu Leu Val Ile Ala Leu Leu His Ile
35 40 45

Ala Lys Arg Arg Gly Ile His
50 55

<210> 75
<211> 55
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> 来自12小结构域文库的巴氏葡萄球菌小结构域3

<400> 75

Glu Arg Val Lys Leu Leu Leu Thr Glu Tyr Asp Leu Ile Asn Lys Glu
1 5 10 15

Gln Ile Pro Thr Ser Asn Asn Pro Tyr Gln Ile Arg Val Lys Gly Leu
20 25 30

Ser Glu Ile Leu Ser Lys Asp Glu Leu Ala Ile Ala Leu Leu His Leu
35 40 45

Ala Lys Arg Arg Gly Ile His
50 55

[0133] <210> 76
<211> 55
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> 来自12小结构域文库的猪葡萄球菌小结构域3

<400> 76

Asp Arg Val Lys His Leu Leu Ala Glu Tyr Asp Leu Leu Asp Leu Thr
1 5 10 15

Asn Ile Pro Lys Ser Thr Asn Pro Tyr Gln Thr Arg Val Lys Gly Leu
20 25 30

Asn Glu Lys Leu Ser Lys Asp Glu Leu Val Ile Ala Leu Leu His Ile
35 40 45

Ala Lys Arg Arg Gly Ile His
50 55

<210> 77

<211> 54
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> 来自12小结构域文库的Staphylococcus microti小结构域3

<400> 77

Glu Arg Val Lys Ser Leu Leu Ser Glu Tyr Lys Ile Ile Ser Gly Leu
1 5 10 15

Ala Pro Thr Asn Asn Gln Pro Tyr Asn Ile Arg Val Lys Gly Leu Thr
20 25 30

Glu Gln Leu Thr Lys Asp Glu Leu Ala Val Ala Leu Leu His Ile Ala
35 40 45

Lys Arg Arg Gly Ile His
50

<210> 78
[0134] <211> 93
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> 来自12小结构域文库的路邓葡萄球菌小结构域4

<400> 78

Lys Ile Asp Val Ile Asp Ser Asn Asp Asp Val Gly Asn Glu Leu Ser
1 5 10 15

Thr Lys Glu Gln Leu Asn Lys Asn Ser Lys Leu Leu Lys Asp Lys Phe
20 25 30

Val Cys Gln Ile Gln Leu Glu Arg Met Asn Glu Gly Gln Val Arg Gly
35 40 45

Glu Lys Asn Arg Phe Lys Thr Ala Asp Ile Ile Lys Glu Ile Ile Gln
50 55 60

Leu Leu Asn Val Gln Lys Asn Phe His Gln Leu Asp Glu Asn Phe Ile
65 70 75 80

Asn Lys Tyr Ile Glu Leu Val Glu Met Arg Arg Glu Tyr
85 90

<210> 79
<211> 93
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> 来自12小结构域文库的巴氏葡萄球菌小结构域4

<400> 79

Asn Ile Asn Val Ser Ser Glu Asp Glu Asp Ala Ser Asn Glu Leu Ser
1 5 10 15

Thr Lys Glu Gln Ile Asn Arg Asn Asn Lys Leu Leu Lys Asp Lys Tyr
20 25 30

Val Cys Glu Val Gln Leu Gln Arg Leu Lys Glu Gly Gln Ile Arg Gly
35 40 45

[0135] Glu Lys Asn Arg Phe Lys Thr Thr Asp Ile Leu Lys Glu Ile Asp Gln
50 55 60

Leu Leu Lys Val Gln Lys Asp Tyr His Asn Leu Asp Ile Asp Phe Ile
65 70 75 80

Asn Gln Tyr Lys Glu Ile Val Glu Thr Arg Arg Glu Tyr
85 90

<210> 80
<211> 94
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> 来自12小结构域文库的猪葡萄球菌小结构域4

<400> 80

Asn Val Asn Val Met Met Asp Asp Asn Asp Ser Gly Asn Glu Leu Ser
1 5 10 15

Thr Lys Asp Gln Leu Lys Lys Asn Ala Lys Ala Leu Ser Asp Lys Tyr

20 25 30

Val Cys Glu Leu Gln Leu Glu Arg Phe Glu Gln Asp Tyr Lys Val Arg
35 40 45

Gly Glu Lys Asn Arg Phe Lys Thr Glu Asp Phe Val Arg Glu Ala Arg
50 55 60

Lys Leu Leu Glu Thr Gln Ser Lys Phe Phe Glu Ile Asp Gln Thr Phe
65 70 75 80

Ile Met Arg Tyr Ile Glu Leu Ile Glu Thr Arg Arg Glu Tyr
85 90

<210> 81
<211> 96
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> 来自12小结构域文库的Staphylococcus microti小结构域4

[0136]

<400> 81

Asn Val Asp Val Ala Ala Asp Lys Glu Glu Thr Ala Ser Asp Ser Leu
1 5 10 15

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

Tyr Val Cys Glu Leu Gln Lys Glu Arg Leu Glu Asn Glu Gly His Val
35 40 45

Arg Gly Val Glu Asn Arg Phe Leu Thr Lys Asp Ile Val Arg Glu Ala
50 55 60

Lys Lys Ile Ile Asp Thr Gln Met Gln Tyr Tyr Pro Glu Ile Asp Glu
65 70 75 80

Thr Phe Lys Glu Lys Tyr Ile Ser Leu Val Glu Thr Arg Arg Glu Tyr
85 90 95

<210> 82
<211> 114
<212> PRT
<213> 人工序列

<220>

<223> 合成的多肽

<220>
<221> misc_feature
<223> 来自12小结构域文库的路邓葡萄球菌小结构域5

<400> 82

Phe Glu Gly Pro Gly Lys Gly Ser Pro Tyr Gly Trp Glu Gly Asp Pro
1 5 10 15

Lys Ala Trp Tyr Glu Thr Leu Met Gly His Cys Thr Tyr Phe Pro Asp
20 25 30

Glu Leu Arg Ser Val Lys Tyr Ala Tyr Ser Ala Asp Leu Phe Asn Ala
35 40 45

Leu Asn Asp Leu Asn Asn Leu Val Ile Gln Arg Asp Gly Leu Ser Lys
50 55 60

Leu Glu Tyr His Glu Lys Tyr His Ile Ile Glu Asn Val Phe Lys Gln
65 70 75 80

Lys Lys Lys Pro Thr Leu Lys Gln Ile Ala Asn Glu Ile Asn Val Asn
85 90 95

[0137]

Pro Glu Asp Ile Lys Gly Tyr Arg Ile Thr Lys Ser Gly Lys Pro Gln
100 105 110

Phe Thr

<210> 83
<211> 114
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> 来自12小结构域文库的巴氏葡萄球菌小结构域5

<400> 83

Phe Glu Gly Pro Gly Gln Gly Ser Pro Phe Gly Trp Asn Gly Asp Leu
1 5 10 15

Lys Lys Trp Tyr Glu Met Leu Met Gly His Cys Thr Tyr Phe Pro Gln
20 25 30

Glu Leu Arg Ser Val Lys Tyr Ala Tyr Ser Ala Asp Leu Phe Asn Ala
35 40 45

Leu Asn Asp Leu Asn Asn Leu Ile Ile Gln Arg Asp Asn Ser Glu Lys
50 55 60

Leu Glu Tyr His Glu Lys Tyr His Ile Ile Glu Asn Val Phe Lys Gln
65 70 75 80

Lys Lys Lys Pro Thr Leu Lys Gln Ile Ala Lys Glu Ile Gly Val Asn
85 90 95

Pro Glu Asp Ile Lys Gly Tyr Arg Ile Thr Lys Ser Gly Thr Pro Gln
100 105 110

Phe Thr

<210> 84
<211> 114
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

[0138]

<220>
<221> misc_feature
<223> 来自12小结构域文库的猪葡萄球菌小结构域5

<400> 84

Phe Glu Gly Pro Gly Lys Gly Ser Pro Phe Gly Trp Glu Gly Asn Ile
1 5 10 15

Lys Lys Trp Phe Glu Gln Met Met Gly His Cys Thr Tyr Phe Pro Glu
20 25 30

Glu Leu Arg Ser Val Lys Tyr Ser Tyr Ser Ala Glu Leu Phe Asn Ala
35 40 45

Leu Asn Asp Leu Asn Asn Leu Val Ile Thr Arg Asp Glu Asp Ala Lys
50 55 60

Leu Asn Tyr Gly Glu Lys Phe Gln Ile Ile Glu Asn Val Phe Lys Gln
65 70 75 80

Lys Lys Thr Pro Asn Leu Lys Gln Ile Ala Ile Glu Ile Gly Val His
85 90 95

Glu Thr Glu Ile Lys Gly Tyr Arg Val Asn Lys Ser Gly Lys Pro Glu

100 105 110

Phe Thr

<210> 85
<211> 114
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> 来自12小结构域文库的Staphylococcus microti小结构域5

<400> 85

Tyr Glu Gly Pro Gly Lys Gly Ser Pro Tyr Gly Trp Asp Ala Asp Val
1 5 10 15

Lys Lys Trp Tyr Gln Leu Met Met Gly His Cys Thr Tyr Phe Pro Val
20 25 30

Glu Phe Arg Ser Val Lys Tyr Ala Tyr Thr Ala Asp Leu Tyr Asn Ala
35 40 45

[0139]

Leu Asn Asp Leu Asn Asn Leu Thr Ile Ala Arg Asp Asp Asn Pro Lys
50 55 60

Leu Glu Tyr His Glu Lys Tyr His Ile Ile Glu Asn Val Phe Lys Gln
65 70 75 80

Lys Arg Asn Pro Thr Leu Lys Gln Ile Ala Lys Glu Ile Gly Val Asn
85 90 95

Asp Ile Asn Ile Ser Gly Tyr Arg Val Thr Lys Ser Gly Lys Pro Gln
100 105 110

Phe Thr

<210> 86
<211> 128
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>

<220>
<221> misc_feature
<223> 来自12小结构域文库的路邓葡萄球菌小结构域6

<400> 86

Glu Phe Lys Leu Tyr His Asp Leu Lys Ser Val Leu Phe Asp Gln Ser
1 5 10 15

Ile Leu Glu Asn Glu Asp Val Leu Asp Gln Ile Ala Glu Ile Leu Thr
20 25 30

Ile Tyr Gln Asp Lys Asp Ser Ile Lys Ser Lys Leu Thr Glu Leu Asp
35 40 45

Ile Leu Leu Asn Glu Glu Asp Lys Glu Asn Ile Ala Gln Leu Thr Gly
50 55 60

Tyr Thr Gly Thr His Arg Leu Ser Leu Lys Cys Ile Arg Leu Val Leu
65 70 75 80

Glu Glu Gln Trp Tyr Ser Ser Arg Asn Gln Met Glu Ile Phe Thr His
85 90 95

Leu Asn Ile Lys Pro Lys Lys Ile Asn Leu Thr Ala Ala Asn Lys Ile
100 105 110

[0140]

Pro Lys Ala Met Ile Asp Glu Phe Ile Leu Ser Pro Val Val Lys Arg
115 120 125

<210> 87
<211> 128
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> 来自12小结构域文库的巴氏葡萄球菌小结构域6

<400> 87

Glu Phe Lys Leu Tyr His Asp Leu Lys Ser Ile Val Phe Asp Lys Ser
1 5 10 15

Ile Leu Glu Asn Glu Ala Ile Leu Asp Gln Ile Ala Glu Ile Leu Thr
20 25 30

Ile Tyr Gln Asp Glu Gln Ser Ile Lys Glu Glu Leu Asn Lys Leu Pro
35 40 45

Glu Ile Leu Asn Glu Gln Asp Lys Ala Glu Ile Ala Lys Leu Ile Gly
50 55 60

Tyr Asn Gly Thr His Arg Leu Ser Leu Lys Cys Ile His Leu Ile Asn
65 70 75 80

Glu Glu Leu Trp Gln Thr Ser Arg Asn Gln Met Glu Ile Phe Asn Tyr
85 90 95

Leu Asn Ile Lys Pro Asn Lys Val Asp Leu Ser Glu Gln Asn Lys Ile
100 105 110

Pro Lys Asp Met Val Asn Asp Phe Ile Leu Ser Pro Val Val Lys Arg
115 120 125

<210> 88
<211> 128
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> 来自12小结构域文库的猪葡萄球菌小结构域6

[0141]

<400> 88

Gln Phe Lys Leu Tyr His Asp Leu Lys Asn Ile Phe Lys Asp Pro Lys
1 5 10 15

Tyr Leu Asn Asp Ile Gln Leu Met Asp Asn Ile Ala Glu Ile Ile Thr
20 25 30

Ile Tyr Gln Asp Ala Glu Ser Ile Ile Lys Glu Leu Asn Gln Leu Pro
35 40 45

Glu Leu Leu Ser Glu Arg Glu Lys Glu Lys Ile Ser Ala Leu Ser Gly
50 55 60

Tyr Ser Gly Thr His Arg Leu Ser Leu Lys Cys Ile Asn Leu Leu Leu
65 70 75 80

Asp Asp Leu Trp Glu Ser Ser Leu Asn Gln Met Glu Leu Phe Thr Lys
85 90 95

Leu Asn Leu Lys Pro Lys Lys Ile Asp Leu Ser Gln Gln His Lys Ile
100 105 110

Pro Ser Lys Leu Val Asp Asp Phe Ile Leu Ser Pro Val Val Lys Arg

115 120 125

<210> 89
<211> 128
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> 来自12小结构域文库的Staphylococcus microti小结构域6

<400> 89

Ser Phe Lys Leu Phe His Asp Leu Lys Lys Val Val Lys Asp His Ala
1 5 10 15

Ile Leu Asp Asp Ile Asp Leu Leu Asn Gln Ile Ala Glu Ile Leu Thr
20 25 30

Ile Tyr Gln Asp Lys Asp Ser Ile Val Ala Glu Leu Gly Gln Leu Glu
35 40 45

Tyr Leu Met Ser Glu Ala Asp Lys Gln Ser Ile Ser Glu Leu Thr Gly
50 55 60

[0142]

Tyr Thr Gly Thr His Ser Leu Ser Leu Lys Cys Met Asn Met Ile Ile
65 70 75 80

Asp Glu Leu Trp His Ser Ser Met Asn Gln Met Glu Val Phe Thr Tyr
85 90 95

Leu Asn Met Arg Pro Lys Lys Tyr Glu Leu Lys Gly Tyr Gln Arg Ile
100 105 110

Pro Thr Asp Met Ile Asp Asp Ala Ile Leu Ser Pro Val Val Lys Arg
115 120 125

<210> 90
<211> 86
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> 来自12小结构域文库的路邓葡萄球菌小结构域7

<400> 90

Thr Phe Gly Gln Ala Ile Asn Leu Ile Asn Lys Ile Ile Glu Lys Tyr
1 5 10 15

Gly Val Pro Glu Asp Ile Ile Ile Glu Leu Ala Arg Glu Asn Asn Ser
20 25 30

Lys Asp Lys Gln Lys Phe Ile Asn Glu Met Gln Lys Lys Asn Glu Asn
35 40 45

Thr Arg Lys Arg Ile Asn Glu Ile Ile Gly Lys Tyr Gly Asn Gln Asn
50 55 60

Ala Lys Arg Leu Val Glu Lys Ile Arg Leu His Asp Glu Gln Glu Gly
65 70 75 80

Lys Cys Leu Tyr Ser Leu
85

<210> 91
<211> 86
<212> PRT
<213> 人工序列

<220>
[0143] <223> 合成的多肽

<220>
<221> misc_feature
<223> 来自12小结构域文库的巴氏葡萄球菌小结构域7

<400> 91

Thr Phe Ile Gln Ser Ile Asn Val Ile Asn Lys Val Ile Glu Lys Tyr
1 5 10 15

Gly Ile Pro Glu Asp Ile Ile Ile Glu Leu Ala Arg Glu Asn Asn Ser
20 25 30

Asp Asp Arg Lys Lys Phe Ile Asn Asn Leu Gln Lys Lys Asn Glu Ala
35 40 45

Thr Arg Lys Arg Ile Asn Glu Ile Ile Gly Gln Thr Gly Asn Gln Asn
50 55 60

Ala Lys Arg Ile Val Glu Lys Ile Arg Leu His Asp Gln Gln Glu Gly
65 70 75 80

Lys Cys Leu Tyr Ser Leu
85

<210> 92
<211> 86
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> 来自12小结构域文库的猪葡萄球菌小结构域7

<400> 92

Ala Phe Ile Gln Ser Ile Gln Val Val Asn Ala Ile Ile Asp Lys Tyr
1 5 10 15

Gly Leu Pro Glu Asp Ile Ile Ile Glu Leu Ala Arg Glu Asn Asn Ser
20 25 30

Asp Asp Arg Arg Lys Phe Leu Asn Gln Leu Gln Lys Gln Asn Glu Glu
35 40 45

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

[0144] Ala Lys Arg Ile Val Gln Lys Ile Lys Leu His Asn Met Gln Glu Gly
65 70 75 80

Lys Cys Leu Tyr Ser Leu
85

<210> 93
<211> 86
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> 来自12小结构域文库的Staphylococcus microti小结构域7

<400> 93

Ser Phe Lys Gln Ala Ile Gly Val Val Asn Ala Ile Ile Lys Lys Tyr
1 5 10 15

Gly Leu Pro Lys Asp Ile Ile Ile Glu Leu Ala Arg Glu Ser Asn Ser
20 25 30

Ala Glu Lys Ser Arg Tyr Leu Arg Ala Ile Gln Lys Lys Asn Glu Lys
35 40 45

Thr Arg Glu Arg Ile Glu Ala Ile Ile Lys Glu Tyr Gly Asn Glu Asn
50 55 60

Ala Lys Gly Leu Val Gln Lys Ile Lys Leu His Asp Ala Gln Glu Gly
65 70 75 80

Lys Cys Leu Tyr Ser Leu
85

<210> 94
<211> 92
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> 来自12小结构域文库的路邓葡萄球菌小结构域8

<400> 94

Glu Ser Ile Pro Leu Glu Asp Leu Leu Asn Asn Pro Asn His Tyr Glu
1 5 10 15

[0145]

Val Asp His Ile Ile Pro Arg Ser Val Ser Phe Asp Asn Ser Tyr His
20 25 30

Asn Lys Val Leu Val Lys Gln Ser Glu Asn Ser Lys Lys Ser Asn Leu
35 40 45

Thr Pro Tyr Gln Tyr Phe Asn Ser Gly Lys Ser Lys Leu Ser Tyr Asn
50 55 60

Gln Phe Lys Gln His Ile Leu Asn Leu Ser Lys Ser Gln Asp Arg Ile
65 70 75 80

Ser Lys Lys Lys Lys Glu Tyr Leu Leu Glu Glu Arg
85 90

<210> 95
<211> 92
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature

<223> 来自12小结构域文库的巴氏葡萄球菌小结构域8

<400> 95

Glu Ser Ile Ala Leu Met Asp Leu Leu Asn Asn Pro Gln Asn Tyr Glu
1 5 10 15

Val Asp His Ile Ile Pro Arg Ser Val Ala Phe Asp Asn Ser Ile His
20 25 30

Asn Lys Val Leu Val Lys Gln Ile Glu Asn Ser Lys Lys Gly Asn Arg
35 40 45

Thr Pro Tyr Gln Tyr Leu Asn Ser Ser Asp Ala Lys Leu Ser Tyr Asn
50 55 60

Gln Phe Lys Gln His Ile Leu Asn Leu Ser Lys Ser Lys Asp Arg Ile
65 70 75 80

Ser Lys Lys Lys Lys Asp Tyr Leu Leu Glu Glu Arg
85 90

<210> 96
<211> 92
<212> PRT
<213> 人工序列

[0146]

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> 来自12小结构域文库的猪葡萄球菌小结构域8

<400> 96

Lys Asp Ile Pro Leu Glu Asp Leu Leu Arg Asn Pro His His Tyr Glu
1 5 10 15

Val Asp His Ile Ile Pro Arg Ser Val Ala Phe Asp Asn Ser Met His
20 25 30

Asn Lys Val Leu Val Arg Ala Asp Glu Asn Ser Lys Lys Gly Asn Arg
35 40 45

Thr Pro Tyr Gln Tyr Leu Asn Ser Ser Glu Ser Ser Leu Ser Tyr Asn
50 55 60

Glu Phe Lys Gln His Ile Leu Asn Leu Ser Lys Thr Lys Asp Arg Ile
65 70 75 80

Thr Lys Lys Lys Arg Glu Tyr Leu Leu Glu Glu Arg

85 90

<210> 97
<211> 92
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> 来自12小结构域文库的Staphylococcus microti小结构域8

<400> 97

Lys Asp Ile Pro Leu Glu Asp Leu Leu Arg Asn Pro Asn Asn Tyr Asp
1 5 10 15

Ile Asp His Ile Ile Pro Arg Ser Val Ser Phe Asp Asp Ser Met His
20 25 30

Asn Lys Val Leu Val Arg Arg Glu Gln Asn Ala Lys Lys Asn Asn Gln
35 40 45

Thr Pro Tyr Gln Tyr Leu Thr Ser Gly Tyr Ala Asp Ile Lys Tyr Ser
[0147] 50 55 60

Val Phe Lys Gln His Val Leu Asn Leu Ala Glu Asn Lys Asp Arg Met
65 70 75 80

Thr Lys Lys Lys Arg Glu Tyr Leu Leu Glu Glu Arg
85 90

<210> 98
<211> 27
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> 来自12小结构域文库的路邓葡萄球菌小结构域9

<400> 98

Asp Ile Asn Lys Phe Glu Val Gln Lys Glu Phe Ile Asn Arg Asn Leu
1 5 10 15

Val Asp Thr Arg Tyr Ala Thr Arg Glu Leu Thr
20 25

<210> 99
<211> 27
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> 来自12小结构域文库的巴氏葡萄球菌小结构域9

<400> 99

Asp Ile Asn Lys Phe Glu Val Gln Lys Glu Phe Ile Asn Arg Asn Leu
1 5 10 15

Val Asp Thr Arg Tyr Ala Thr Arg Glu Leu Thr
20 25

<210> 100
<211> 27
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

[0148]

<220>
<221> misc_feature
<223> 来自12小结构域文库的猪葡萄球菌小结构域9

<400> 100

Asp Ile Asn Lys Phe Asp Val Gln Lys Glu Phe Ile Asn Arg Asn Leu
1 5 10 15

Val Asp Thr Arg Tyr Ala Thr Arg Glu Leu Thr
20 25

<210> 101
<211> 27
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> 来自12小结构域文库的Staphylococcus microti小结构域9

<400> 101

Asn Ile Asn Lys Tyr Asp Val Gln Lys Glu Phe Ile Asn Arg Asn Leu
1 5 10 15

Val Asp Thr Arg Tyr Thr Thr Arg Glu Leu Thr
20 25

<210> 102
<211> 64
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> 来自12小结构域文库的路邓葡萄球菌小结构域10

<400> 102

Asn Tyr Leu Lys Ala Tyr Phe Ser Ala Asn Asn Met Asn Val Lys Val
1 5 10 15

Lys Thr Ile Asn Gly Ser Phe Thr Asp Tyr Leu Arg Lys Val Trp Lys
20 25 30

Phe Lys Lys Glu Arg Asn His Gly Tyr Lys His His Ala Glu Asp Ala
35 40 45

[0149]

Leu Ile Ile Ala Asn Ala Asp Phe Leu Phe Lys Glu Asn Lys Lys Leu
50 55 60

<210> 103
<211> 64
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> 来自12小结构域文库的巴氏葡萄球菌小结构域10

<400> 103

Ser Tyr Leu Lys Ala Tyr Phe Ser Ala Asn Asn Met Asp Val Lys Val
1 5 10 15

Lys Thr Ile Asn Gly Ser Phe Thr Asn His Leu Arg Lys Val Trp Arg
20 25 30

Phe Asp Lys Tyr Arg Asn His Gly Tyr Lys His His Ala Glu Asp Ala
35 40 45

Leu Ile Ile Ala Asn Ala Asp Phe Leu Phe Lys Glu Asn Lys Lys Leu
50 55 60

<210> 104
<211> 64
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> 来自12小结构域文库的猪葡萄球菌小结构域10

<400> 104

Ser Leu Leu Lys Ala Tyr Phe Ser Ala Asn Asn Leu Asp Val Lys Val
1 5 10 15

Lys Thr Ile Asn Gly Ser Phe Thr Asn Tyr Leu Arg Lys Val Trp Lys
20 25 30

Phe Asp Lys Asp Arg Asn Lys Gly Tyr Lys His His Ala Glu Asp Ala
35 40 45

[0150] Leu Ile Ile Ala Asn Ala Asp Phe Leu Phe Lys His Asn Lys Lys Leu
50 55 60

<210> 105
<211> 64
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> 来自12小结构域文库的Staphylococcus microti小结构域10

<400> 105

Thr Leu Leu Lys Thr Tyr Phe Thr Ile Asn Asn Leu Asp Val Lys Val
1 5 10 15

Lys Thr Ile Asn Gly Ser Phe Thr Asp Phe Leu Arg Lys Arg Trp Gly
20 25 30

Phe Lys Lys Asn Arg Asp Glu Gly Tyr Lys His His Ala Glu Asp Ala
35 40 45

Leu Ile Ile Ala Asn Ala Asp Tyr Leu Phe Lys Glu His Lys Leu Leu
50 55 60

<210> 106
<211> 186
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> 来自12小结构域文库的路邓葡萄球菌小结构域11

<400> 106

Lys Ala Val Asn Ser Val Leu Glu Lys Pro Glu Ile Glu Thr Lys Gln
1 5 10 15

Leu Asp Ile Gln Val Asp Ser Glu Asp Asn Tyr Ser Glu Met Phe Ile
20 25 30

Ile Pro Lys Gln Val Gln Asp Ile Lys Asp Phe Arg Asn Phe Lys Tyr
35 40 45

Ser His Arg Val Asp Lys Lys Pro Asn Arg Gln Leu Ile Asn Asp Thr
50 55 60

[0151]

Leu Tyr Ser Thr Arg Lys Lys Asp Asn Ser Thr Tyr Ile Val Gln Thr
65 70 75 80

Ile Lys Asp Ile Tyr Ala Lys Asp Asn Thr Thr Leu Lys Lys Gln Phe
85 90 95

Asp Lys Ser Pro Glu Lys Phe Leu Met Tyr Gln His Asp Pro Arg Thr
100 105 110

Phe Glu Lys Leu Glu Val Ile Met Lys Gln Tyr Ala Asn Glu Lys Asn
115 120 125

Pro Leu Ala Lys Tyr His Glu Glu Thr Gly Glu Tyr Leu Thr Lys Tyr
130 135 140

Ser Lys Lys Asn Asn Gly Pro Ile Val Lys Ser Leu Lys Tyr Ile Gly
145 150 155 160

Asn Lys Leu Gly Ser His Leu Asp Val Thr His Gln Phe Lys Ser Ser
165 170 175

Thr Lys Lys Leu Val Lys Leu Ser Ile Lys
180 185

<210> 107
<211> 186
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> 来自12小结构域文库的巴氏葡萄球菌小结构域11

<400> 107

Gln Asn Thr Asn Lys Ile Leu Glu Lys Pro Thr Ile Glu Asn Asn Thr
1 5 10 15

Lys Lys Val Thr Val Glu Lys Glu Glu Asp Tyr Asn Asn Val Phe Glu
20 25 30

Thr Pro Lys Leu Val Glu Asp Ile Lys Gln Tyr Arg Asp Tyr Lys Phe
35 40 45

Ser His Arg Val Asp Lys Lys Pro Asn Arg Gln Leu Ile Asn Asp Thr
50 55 60

[0152]

Leu Tyr Ser Thr Arg Met Lys Asp Glu His Asp Tyr Ile Val Gln Thr
65 70 75 80

Ile Thr Asp Ile Tyr Gly Lys Asp Asn Thr Asn Leu Lys Lys Gln Phe
85 90 95

Asn Lys Asn Pro Glu Lys Phe Leu Met Tyr Gln Asn Asp Pro Lys Thr
100 105 110

Phe Glu Lys Leu Ser Ile Ile Met Lys Gln Tyr Ser Asp Glu Lys Asn
115 120 125

Pro Leu Ala Lys Tyr Tyr Glu Glu Thr Gly Glu Tyr Leu Thr Lys Tyr
130 135 140

Ser Lys Lys Asn Asn Gly Pro Ile Val Lys Lys Ile Lys Leu Leu Gly
145 150 155 160

Asn Lys Val Gly Asn His Leu Asp Val Thr Asn Lys Tyr Glu Asn Ser
165 170 175

Thr Lys Lys Leu Val Lys Leu Ser Ile Lys
180 185

<210> 108
<211> 184
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> 来自12小结构域文库的猪葡萄球菌小结构域11

<400> 108

Arg Asn Ile Asn Lys Val Leu Asp Ala Pro Ser Lys Glu Val Asp Lys
1 5 10 15

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

Asp Thr Gln Lys Ala Gln Ala Ile Lys Lys Phe Glu Ile Arg Lys Phe
35 40 45

Ser His Arg Val Asp Lys Lys Pro Asn Arg Gln Leu Ile Asn Asp Thr
50 55 60

[0153] Leu Tyr Ser Thr Arg Asn Ile Asp Gly Ile Glu Tyr Val Val Glu Ser
65 70 75 80

Ile Lys Asp Ile Tyr Ser Val Asn Asn Asp Lys Val Lys Thr Lys Phe
85 90 95

Lys Lys Asp Pro His Arg Leu Leu Met Tyr Arg Asn Asp Pro Gln Thr
100 105 110

Phe Glu Lys Phe Glu Lys Val Phe Lys Gln Tyr Glu Ser Glu Lys Asn
115 120 125

Pro Phe Ala Lys Tyr Tyr Glu Glu Thr Gly Glu Lys Ile Arg Lys Phe
130 135 140

Ser Lys Thr Gly Gln Gly Pro Tyr Ile Asn Lys Ile Lys Tyr Leu Arg
145 150 155 160

Glu Arg Leu Gly Arg His Cys Asp Val Thr Asn Lys Tyr Ile Asn Ser
165 170 175

Arg Asn Lys Ile Val Gln Leu Lys
180

<210> 109

<211> 183
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> 来自12小结构域文库的Staphylococcus microti小结构域11

<400> 109

Lys Glu Ile Lys Asp Val Ser Asp Leu Ala Gly Asp Glu Arg Asn Ser
1 5 10 15

Asn Val Lys Asp Glu Asp Gln Tyr Glu Glu Val Phe Gly Gly Tyr Phe
20 25 30

Lys Ile Glu Asp Ile Lys Lys Tyr Lys Ile Lys Lys Phe Ser His Arg
35 40 45

Val Asp Lys Lys Pro Asn Arg Gln Leu Ile Asn Asp Thr Ile Tyr Ser
50 55 60

Thr Arg Val Lys Asp Asp Lys Arg Tyr Leu Ile Asn Thr Leu Lys Asn
65 70 75 80

[0154]

Leu Tyr Asp Lys Ser Asn Gly Asp Leu Lys Glu Arg Met Gln Lys Asp
85 90 95

Pro Glu Ser Leu Leu Met Tyr His His Asp Pro Gln Thr Phe Glu Lys
100 105 110

Leu Lys Ile Val Met Ser Gln Tyr Glu Asn Glu Lys Asn Pro Leu Ala
115 120 125

Lys Tyr Phe Glu Glu Thr Gly Gln Tyr Leu Thr Lys Tyr Ala Lys His
130 135 140

Asp Asn Gly Pro Ala Ile His Lys Ile Lys Tyr Tyr Gly Asn Lys Leu
145 150 155 160

Val Glu His Leu Asp Ile Thr Lys Asn Tyr His Asn Pro Gln Asn Lys
165 170 175

Val Val Gln Leu Ser Gln Lys
180

<210> 110
<211> 145

<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> 来自12小结构域文库的路邓葡萄球菌小结构域12

<400> 110

Pro Tyr Arg Phe Asp Val Tyr Leu Thr Asp Lys Gly Tyr Lys Phe Ile
1 5 10 15

Thr Ile Ser Tyr Leu Asp Val Leu Lys Lys Asp Asn Tyr Tyr Tyr Ile
20 25 30

Pro Glu Gln Lys Tyr Asp Lys Leu Lys Leu Gly Lys Ala Ile Asp Lys
35 40 45

Asn Ala Lys Phe Ile Ala Ser Phe Tyr Lys Asn Asp Leu Ile Lys Leu
50 55 60

Asp Gly Glu Ile Tyr Lys Ile Ile Gly Val Asn Ser Asp Thr Arg Asn
65 70 75 80

[0155]

Met Ile Glu Leu Asp Leu Pro Asp Ile Arg Tyr Lys Glu Tyr Cys Glu
85 90 95

Leu Asn Asn Ile Lys Gly Glu Pro Arg Ile Lys Lys Thr Ile Gly Lys
100 105 110

Lys Val Asn Ser Ile Glu Lys Leu Thr Thr Asp Val Leu Gly Asn Val
115 120 125

Phe Thr Asn Thr Gln Tyr Thr Lys Pro Gln Leu Leu Phe Lys Arg Gly
130 135 140

Asn
145

<210> 111
<211> 145
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature

<223> 来自12小结构域文库的巴氏葡萄球菌小结构域12

<400> 111

Asn Tyr Arg Phe Asp Val Tyr Leu Thr Glu Lys Gly Tyr Lys Phe Val
1 5 10 15

Thr Ile Ala Tyr Leu Asn Val Phe Lys Lys Asp Asn Tyr Tyr Tyr Ile
20 25 30

Pro Lys Asp Lys Tyr Gln Glu Leu Lys Glu Lys Lys Lys Ile Lys Asp
35 40 45

Thr Asp Gln Phe Ile Ala Ser Phe Tyr Lys Asn Asp Leu Ile Lys Leu
50 55 60

Asn Gly Asp Leu Tyr Lys Ile Ile Gly Val Asn Ser Asp Asp Arg Asn
65 70 75 80

Ile Ile Glu Leu Asp Tyr Tyr Asp Ile Lys Tyr Lys Asp Tyr Cys Glu
85 90 95

Ile Asn Asn Ile Lys Gly Glu Pro Arg Ile Lys Lys Thr Ile Gly Lys
100 105 110

[0156]

Lys Thr Glu Ser Ile Glu Lys Phe Thr Thr Asp Val Leu Gly Asn Leu
115 120 125

Tyr Leu His Ser Thr Glu Lys Ala Pro Gln Leu Ile Phe Lys Arg Gly
130 135 140

Leu
145

<210> 112
<211> 144
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> 来自12小结构域文库的猪葡萄球菌小结构域12

<400> 112

Ile Tyr Ser Tyr Arg Phe Asp Ile Tyr Gln Tyr Gly Asn Asn Tyr Lys
1 5 10 15

Met Ile Thr Ile Ser Tyr Ile Asp Leu Glu Gln Lys Ser Asn Tyr Tyr

20 25 30

Tyr Ile Ser Arg Glu Lys Tyr Glu Gln Lys Lys Lys Asp Lys Gln Ile
35 40 45

Asp Asp Ser Tyr Lys Phe Ile Gly Ser Phe Tyr Lys Asn Asp Ile Ile
50 55 60

Asn Tyr Asn Gly Glu Met Tyr Arg Val Ile Gly Val Asn Asp Ser Glu
65 70 75 80

Lys Asn Lys Ile Gln Leu Asp Met Ile Asp Ile Ser Ile Lys Asp Tyr
85 90 95

Met Glu Leu Asn Asn Ile Lys Lys Thr Gly Val Ile Tyr Lys Thr Ile
100 105 110

Gly Lys Ser Thr Thr His Ile Glu Lys Tyr Thr Thr Asp Ile Leu Gly
115 120 125

Asn Leu Tyr Lys Ala Ala Pro Pro Lys Lys Pro Gln Leu Ile Phe Lys
130 135 140

[0157]

<210> 113
<211> 144
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> 来自12小结构域文库的Staphylococcus microti小结构域12

<400> 113

Ser Phe Arg Phe Asp Val Tyr Gln Thr Asp Lys Gly Tyr Lys Phe Ile
1 5 10 15

Ser Ile Ala Tyr Leu Thr Leu Lys Asn Glu Lys Asn Tyr Tyr Ala Ile
20 25 30

Ser Gln Glu Lys Tyr Asp Gln Leu Lys Ser Glu Lys Lys Ile Ser Asn
35 40 45

Asn Ala Val Phe Ile Gly Ser Phe Tyr Thr Ser Asp Ile Ile Glu Ile
50 55 60

Asn Asn Glu Lys Phe Arg Val Ile Gly Val Asn Ser Asp Lys Asn Asn
65 70 75 80

Leu Ile Glu Val Asp Arg Ile Asp Ile Arg Gln Lys Glu Phe Ile Glu
85 90 95

Leu Glu Glu Glu Lys Lys Asn Asn Arg Ile Lys Val Thr Ile Gly Arg
100 105 110

Lys Thr Thr Asn Ile Glu Lys Phe His Thr Asp Ile Leu Gly Asn Met
115 120 125

Tyr Lys Ser Lys Arg Pro Lys Ala Pro Gln Leu Val Phe Lys Lys Gly
130 135 140

<210> 114
<211> 104
<212> RNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNAVEGFA

<400> 114
[0158] ggacccccuc caccccgccu cguuuuagua cucuggaaac agaaucuacu gaaacaagac 60

aauaugucgu guuuauccca ucaauuuauu ggugggauuu uuuu 104

<210> 115
<211> 27
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> 具有CGGGCGC-PAM的ATTO647N寡核苷酸

<400> 115
acccccctcca ccccgcctcc gggcgcg 27

<210> 116
<211> 33
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>

<220>
<221> misc_feature
<223> 具有CGGGCGC-PAM的生物素化的寡核苷酸

<400> 116
gcgcgcgccc ggaggcgggg tggaggggt cgg 33

<210> 117
<211> 20
<212> RNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> 靶向BFP的指导RNA

<400> 117
gggucagggu ggucacgagg 20

<210> 118
<211> 2188
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

[0159]

<220>
<221> misc_feature
<223> pCas606

<400> 118
ccactgtaga agagcaaatg ccacctgacg tctaagaaat tcgcgttaaa tttttgttaa 60

atcagctcat tttttaacca ataggccgaa atcggcaaaa tcccttataa atcaaaagaa 120

tagaccgaga tagggttgag tggccgctac actagggcgc gtctaatacg actcactata 180

ggaccccctc caccccgcct cgttttagta ctctggaaac agaatctact gaaacaagac 240

aatatgtcgt gtttatccca tcaatttatt ggtgggattt ttttctagca taaccccttg 300

gggcctctaa acgggtcttg aggggttttt tgtcactgcc cgctttccag tcgggaaacc 360

tgtcgtgcca gctgcattaa tgaatcgttc gtcagaactg cttacgcggt ttgtcactcg 420

gtcgctacgc tccgggcgtg agactgcggc gggcgctgcg gacacataca aagttaccca 480

cagattccgt ggataagcag gggactaaca tgtgaggcaa aacagcaggg ccgcgccggt 540

ggcgtttttc cataggctcc gccctcctgc cagagttcac ataaacagac gcttttccgg 600

tgcatctgtg ggagccgtga ggctcaacca tgaatctgac agtacgggcg aaacccgaca 660

ggacttaaag atccccaccg tttccggcgg gtcgctccct cttgcgctct cctgttccga 720

ccctgccgtt taccggatac ctgttccgcc tttctccctt acgggaagtg tggcgctttc 780

tcatagctca cacactggta tctcggctcg gtgtaggtcg ttcgctccaa gctgggctgt 840

aagcaagaac tccccgttca gcccgactgc tgcgccttat ccggtaactg ttcacttgag 900

tccaacccgg aaaagcacgg taaaacgcca ctggcagcag ccattggtaa ctgggagttc 960

gcagaggatt tgtttagcta aacacgcggt tgctcttgaa gtgtgcgcca aagtccggct 1020

acactggaag gacagatttg gttgctgtgc tctgcgaaag ccagttacca cggttaagca 1080

gttccccaac tgacttaacc ttcgatcaaa ccacctcccc aggtggtttt ttcgtttaca 1140

gggcaaaaga ttacgcgcag aaaaaaagga tctcaagaag atcctttgat cttttctact 1200

gaaccgctca atctaaagta tatatgagta aacttggtct gacagttatt agaaaaactc 1260

atcgagcatc aaatgaaact gcaatttatt catatcagga ttatcaatac catatttttg 1320

aaaaagccgt ttctgtaatg aaggagaaaa ctcaccgagg cagttccata ggatggcaag 1380

atcctggtat cggtctgcga ttccgactcg tccaacatca atacaaccta ttaatttccc 1440

ctcgtcaaaa ataaggttat caagtgagaa atcaccatga gtgacgactg aatccggtga 1500

gaatggcaaa agtttatgca tttctttcca gacttgttca acaggccagc cattacgctc 1560

gtcatcaaaa tcactcgcat caaccaaacc gttattcatt cgtgattgcg cctgagcgag 1620

acgaaatacg cgatcgctgt taaaaggaca attacaaaca ggaatcgaat gcaaccggcg 1680

caggaacact gccagcgcat caacaatatt ttcacctgaa tcaggatatt cttctaatac 1740

[0160] ctggaatgct gttttcccgg ggatcgcagt ggtgagtaac catgcatcat caggagtacg 1800

gataaaatgc ttgatggtcg gaagaggcat aaattccgtc agccagttta gtctgaccat 1860

ctcatctgta acatcattgg caacgctacc tttgccatgt ttcagaaaca actctggcgc 1920

atcgggcttc ccatacaatc gatagattgt cgcacctgat tgcccgacat tatcgcgagc 1980

ccatttatac ccatataaat cagcatccat gttggaattt aatcgcggcc tagagcaaga 2040

cgtttcccgt tgaatatggc tcatactctt cctttttcaa tattattgaa gcatttatca 2100

gggttattgt ctcatgagcg gatacatatt tgaatgtatt tagaaaaata aacaaatagg 2160

ggttccgcgc acatttcccc gaaaagtg 2188

<210> 119
<211> 9489
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> pCas634

<400> 119
ccggtgatac cacgatacta tgactgagag tcaacgccat gagcggcctc atttcttatt 60

[0161]

ctgagttaca acagtccgca ccgctgccgg tagctattga ctatccggct gcactagccc 120

tgcgtcagat ggctctgatc caaggcaaac tgccaaaata tctgctggca ccggaagtca 180

gcgccctgca ccattatgtt ccggatctgc atcgcaggat gctgctggct accctgtgga 240

acacctacat ctgtattaac gaagcgctgg cattgaccct gagtgatttt tctctggtgc 300

cgccctatcc ctttgtgcag cttgccacgc tcaaaggggt ttgaggtcca accgtacgaa 360

aacgtacggt aagaggaaaa ttatcgtctg aaaaatcgat tagtagacaa gaaagtccgt 420

taagtgccaa ttttcgatta aaaagacacc gttttgatgg cgttttccaa tgtacattat 480

gtttcgatat atcagacagt tacttcacta acgtacgttt tcgttctatt ggccttcaga 540

ccccatatcc ttaatgtcct ttatttgctg gggttatcag atccccccga cacgtttaat 600

taatgctttc tccgccggag atcgacgcac agcgttctgt gctctatgat gttatttctt 660

aataatcatc caggtattct ctttatcacc atacgtagtg cgagtgtcca ccttaacgca 720

gggctttccg tcacagcgcg atatgtcagc cagcggggct ttcttttgcc agaccgcttc 780

catcctctgc atttcagcaa tctggctata cccgtcattc ataaaccacg taaatgccgt 840

cacgcaggaa gccaggacga agaatatcgt cagtacaaga taaatcgcgg atttccacgt 900

atagcgtgac atctcacgac gcatttcatg gatcatcgct ttcgccgtat cggcagcctg 960

attcagcgct tctgtcgccg gtttctgctg tgctaatccg gcttgtttca gttctttctc 1020

aacctgagtg agcgcggaac tcaccgattt cctgacggtg tcagtcatat taccggacgc 1080

gctgtccagc tcacgaatga ccctgctcag cgtttcactt tgctgctgta attgtgatga 1140

ggcggcctga aactgttctg tcagagaagt aacacgcttt tccagcgcct gatgatgccc 1200

gataagggcg gcaatttgtt taatttcgtc gctcatacaa aatcctgcct atcgtgagaa 1260

tgaccagcct ttatccggct tctgtcgtat ctgttcggcg agtcgctgtc gttctttctc 1320

ctgctgacgc tgtttttccg ccagacgttc gcgctctctc tgcctttcca tctcctgatg 1380

tatcccctgg aactccgcca tcgcatcgtt aacaagggac tgaagatcga tttcttcctg 1440

tatatccttc atggcatcac tgaccagtgc gttcagcttg tcaggctctt tttcaaaatc 1500

aaacgttctg ccggaatggg attcctgctc aggctctgac ttcagctcct gttttagcgt 1560

cagagtatcc ctctcgctga gggcttcccg taacgaggta gtcacgtcaa ttacgctgtc 1620

acgttcatca cgggactgct gcacctgcct ttcagcctcc ctgcgctcaa gaatggcctg 1680

tagctgctca gtatcgaatc gctgaacctg acccgcgccc agatgccgct caggctcacg 1740

gtcaatgccc tgcgccttca gggaacggga atcaacccgg tcagcgtgct gataccgttc 1800

aaggtgctta ttctggaggt cagcccagcg ttccctctgg gcaacaaggt attctttgcg 1860

ttcggtcggt gtttccccga aacgtgcctt ttttgcgcca ccgcgctccg gctctttggt 1920

gttagcccgt ttaaaatact gctcagggtc acggtgaata ccgtcattaa tgcgttcaga 1980

[0162]

gaacatgata tgggcgtggg gctgctcgcc accggctatc gctgctttcg gattatggat 2040

agcgaactga taggcatggc ggtcgccaat ttcctgttgg acaaaatcgc ggacaagctc 2100

aagacgttgt tcgggtttta actcacgcgg cagggcaatc tcgatttcac ggtaggtaca 2160

gccgttggca cgttcagacg tgtcagcggc tttccagaac tcggacggtt tatgcgctgc 2220

ccacgccggc atattgccgg actccttgtg ctcaaggtcg gagtcttttt cacgggcata 2280

ctttccctca cgcgcaatat aatcggcatg aggagaggca ctgccttttc cgccggtttt 2340

tacgctgaga tgataggatg ccatcgtgtt ttatcccgct gaaggcgcgc accgtttctg 2400

aacgaagtga agaaacgtct aagtgcgccc tgataaataa aagagttatc agggattgta 2460

gtgggatttg acctcctctg ccatcactga gcataatcat tccgttagca ttcaggaggt 2520

aaacagcatg aataaaagcg aaaaacagga acaatgggca gcagaaagag tgcagtatat 2580

tcgcggctta aagtcgccga atgagcaaca gaaacttatg ctgatactga cggataaagc 2640

agataaaaca gcacaggata tcaaaacgct gtccctgctg atgaaggctg aacaggcagc 2700

agagaaagcg caggaagcca gagcgaaagt catgaacctg atacaggcag aaaagcgagc 2760

cgaagccaga gccgcccgta aagcccgtga ccatgctctg taccagtctg ccggattgct 2820

tatcctggcg ggtctggttg acagtaagac gggtaagcct gttgatgata ccgctgcctt 2880

actgggtgca ttagccagtc tgaatgacct gtcacgggat aatccgaagt ggtcagactg 2940

gaaaatcaga gggcaggaac tgctgaacag caaaaagtca gatagcacca catagcagac 3000

ccgccataaa acgccctgag aagcccgtga cgggcttttc ttgtattatg ggtagtttcc 3060

ttgcatgaat ccataaaagg cgcctgtagt gccatttacc cccattcact gccagagccg 3120

tgagcgcagc gaactgaatg tcacgaaaaa gacagcgact caggtgcctg atggtcggag 3180

acaaaaggaa tattcagcga tttgcccgag cttgcgaggg tgctacttaa gcctttaggg 3240

ttttaaggtc tgttttgtag aggagcaaac agcgtttgcg acatcctttt gtaatactgc 3300

ggaactgact aaagtagtga gttatacaca gggctgggat ctattctttt tatctttttt 3360

tattctttct ttattctata aattataacc acttgaatat aaacaaaaaa aacacacaaa 3420

ggtctagcgg aatttacaga gggtctagca gaatttacaa gttttccagc aaaggtctag 3480

cagaatttac agatacccac aactcaaagg aaaaggacta gtaattatca ttgactagcc 3540

catctcaatt ggtatagtga ttaaaatcac ctagaccaat tgagatgtat gtctgaatta 3600

gttgttttca aagcaaatga actagcgatt agtcgctatg acttaacgga gcatgaaacc 3660

aagctaattt tatgctgtgt ggcactactc aaccccacga ttgaaaaccc tacaaggaaa 3720

gaacggacgg tatcgttcac ttataaccaa tacgctcaga tgatgaacat cagtagggaa 3780

aatgcttatg gtgtattagc taaagcaacc agagagctga tgacgagaac tgtggaaatc 3840

aggaatcctt tggttaaagg ctttgagatt ttccagtgga caaactatgc caagttctca 3900

[0163]

agcgaaaaat tagaattagt ttttagtgaa gagatattgc cttatctttt ccagttaaaa 3960
aaattcataa aatataatct ggaacatgtt aagtcttttg aaaacaaata ctctatgagg 4020
atttatgagt ggttattaaa agaactaaca caaaagaaaa ctcacaaggc aaatatagag 4080
attagccttg atgaatttaa gttcatgtta atgcttgaaa ataactacca tgagtttaaa 4140
aggcttaacc aatgggtttt gaaaccaata agtaaagatt taaacactta cagcaatatg 4200
aaattggtgg ttgataagcg aggccgcccg actgatacgt tgattttcca agttgaacta 4260
gatagacaaa tggatctcgt aaccgaactt gagaacaacc agataaaaat gaatggtgac 4320
aaaataccaa caaccattac atcagattcc tacctacata acggactaag aaaaacacta 4380
cacgatgctt taactgcaaa aattcagctc accagttttg aggcaaaatt tttgagtgac 4440
atgcaaagta agtatgatct caatggttcg ttctcatggc tcacgcaaaa acaacgaacc 4500
acactagaga acatactggc taaatacgga aggatctgag gttcttatgg ctcttgtatc 4560
tatcagtgaa gcatcaagac taacaaacaa aagtagaaca actgttcacc gttacatatc 4620
aaagggaaaa ctgtccatat gcacagatga aaacggtgta aaaaagatag atacatcaga 4680
gcttttacga gtttttggtg cattcaaagc tgttcaccat gaacagatcg acaatgtaac 4740
agatgaacag catgtaacac ctaatagaac aggtgaaacc agtaaaacaa agcaactaga 4800
acatgaaatt gaacacctga gacaacttgt tacagctcaa cagtcacaca tagacagcct 4860
gaaacaggcg atgctgctta tcgaatcaaa gctgccgaca acacgggagc cagtgacgcc 4920
tcccgtgggg aaaaaatcat ggcaattctg gaagaaatag cgctttcagc cggcaaaccg 4980
gctgaagccg gatctgcgat tctgataaca aactagcaac accagaacag cccgtttgcg 5040
ggcagcaaaa cccgtacttt tggacgttcc ggcggttttt tgtggcgagt ggtgttcggg 5100
cggtgcgcgc aagatccatt atgttaaacg ggcgagttta catctcaaaa ccgcccgctt 5160
aacaccatca gaaatcctca gcgcgatttt aagcaccaac ccccccccgt aacacccaaa 5220
tccatactga aagtggcttt gttgaataaa tcgaactttt gctgagttga aggatcagat 5280
cacgcatcct cccgacaaca cagaccattc cgtggcaaag caaaagttca gaatcaccaa 5340
ctggtccacc tacaacaaag ctctcatcaa ccgtggctcc ctcactttct ggctggatga 5400
tgaggcgatt caggcctggt atgagtcggc aacaccttca tcacgaggaa ggccccagcg 5460
ctattctgat ctcgccatca ccaccgttct ggtgattaaa cgcgtattcc ggctgaccct 5520
gcgggctgcg cagggtttta ttgattccat ttttgccctg atgaacgttc cgttgcgctg 5580
cccggattac accagtgtca gtaagcgggc aaagtcggtt aatgtcagtt tcaaaacgtc 5640
cacccggggt gaaatcgcac acctggtgat tgattccacc gggctgaagg tctttggtga 5700
aggcgaatgg aaagtcagaa agcacggcaa agagcgccgt cgtatctggc gaaagttgca 5760
tcttgctgtt gacagcaaca cacatgaagt tgtctgtgca gacctgtcgc tgaataacgt 5820

[0164]

cacggactca gaagccttcc cgggccttat ccggcagact cacagaaaaa tcagggcagc 5880

cgcggcagac ggggcttacg atacccggct ctgtcacgat gaactgcgcc gcaaaaaaat 5940

cagcgcgctt attcctcccc gaaaaggtgc gggttactgg cccggtgaat atgcagaccg 6000

taaccgtgca gtggctaatc agcgaatgac cgggagtaat gcgcggtgga aatggacaac 6060

agattacaac cgtcgctcga tagcggaaac ggcgatgtac cgggtaaaac agctgttcgg 6120

gggttcactg acgctgcgtg actacgatgg tcaggttgcg gaggctatgg ccctggtacg 6180

agcgctgaac aaaatgacga aagcaggtat gcctgaaagc gtgcgtattg cctgaaaaca 6240

caacccgcta cgggggagac ttacccgaaa tctgatttat tcaacaaagc cgggtgtggt 6300

gaactacaaa gcagacccgt tgaggttatc agttcgatgc acaatcagca gcgcataaaa 6360

tatgcacaag aacaggagca cccttcgcat taagctgtgg tggtaacaag tagtgccggg 6420

ctaccatcag cgagcatgat gcgctcccac agcattcgcc ttggcagtat ggaagttcct 6480

cgctccagtt cgggccggta tccacctcga gctttctacg gggtctgacg ctcagtggaa 6540

cgaaaactca cgttaaggga ttttggtcat gagattatca aaaaggatct tcacctagat 6600

ccttttcgac cgaataaata cctgtgacgg aagatcactt cgcagaataa ataaatcctg 6660

gtgtccctgt tgataccggg aagccctggg ccaacttttg gcgaaaatga gacgttgatc 6720

ggcacgtaag aggttccaac tttcaccata atgaaataag atcactaccg ggcgtatttt 6780

ttgagttgtc gagattttca ggagctaagg aagctaaaat ggagaaaaaa atcactggat 6840

ataccaccgt tgatatatcc caatggcatc gtaaagaaca ttttgaggca tttcagtcag 6900

ttgctcaatg tacctataac cagaccgttc agctggatat tacggccttt ttaaagaccg 6960

taaagaaaaa taagcacaag ttttatccgg cctttattca cattcttgcc cgcctgatga 7020

atgctcatcc ggaattacgt atggcaatga aagacggtga gctggtgata tgggatagtg 7080

ttcacccttg ttacaccgtt ttccatgagc aaactgaaac gttttcatcg ctctggagtg 7140

aataccacga cgatttccgg cagtttctac acatatattc gcaagatgtg gcgtgttacg 7200

gtgaaaacct ggcctatttc cctaaagggt ttattgagaa tatgtttttc gtctcagcca 7260

atccctgggt gagtttcacc agttttgatt taaacgtggc caatatggac aacttcttcg 7320

cccccgtttt caccttgggc aaatattata cgcaaggcga caaggtgctg atgccgctgg 7380

cgattcaggt tcatcatgcc gtttgtgatg gcttccatgt cggcagaatg cttaatgaat 7440

tacaacagta ctgcgatgag tggcagggcg gggcgtaatt tttttaaggc agttattggt 7500

gcccttaaac gcctggttgc tacgcctgaa taagtgataa taagcggatg aatggcagaa 7560

attcgaaagc aaattcgacc cggtcgtcgg ttcagggcag ggtcgttaaa tagccgctta 7620

tgtctattgc tggtttaccg gattattgac taccggaagc agtgtgaccg tgtgcttctc 7680

aaatgcctga ggccagtttg ctcaggctct ccccgtggag gtaataattg acgatatgat 7740

[0165]

cctttttttc tgatcaaaag tgctccatgg aattatgaca acttgacggc tacatcattc 7800

acttttttctt cacaaccggc acggaactcg ctcgggctgg ccccggtgca ttttttaaat 7860

acccgcgaga aatagagttg atcgtcaaaa ccaacattgc gaccgacggt ggcgataggc 7920

atccgggtgg tgctcaaaag cagcttcgcc tggctgatac gttggtcctc gcgccagctt 7980

aagacgctaa tccctaactg ctggcggaaa agatgtgaca gacgcgacgg cgacaagcaa 8040

acatgctgtg cgacgctggc gatatcaaaa ttgctgtctg ccaggtgatc gctgatgtac 8100

tgacaagcct cgcgtacccg attatccatc ggtggatgga gcgactcgtt aatcgcttcc 8160

atgcgccgca gtaacaattg ctcaagcaga tttatcgcca gcagctccga atagcgccct 8220

tccccttgcc cggcgttaat gatttgccca aacaggtcgc tgaaatgcgg ctggtgcgct 8280

tcatccgggc gaaagaaccc cgtattggca aatattgacg gccagttaag ccattcatgc 8340

cagtaggcgc gcggacgaaa gtaaacccac tggtgatacc attcgcgagc ctccggatga 8400

cgaccgtagt gatgaatctc tcctggcggg aacagcaaaa tatcacccgg tcggcaaaca 8460

aattctcgtc cctgattttt caccaccccc tgaccgcgaa tggtgagatt gagaatataa 8520

cctttcattc ccagcggtcg gtcgataaaa aaatcgagat aaccgttggc ctcaatcggc 8580

gttaaacccg ccaccagatg ggcattaaac gagtatcccg gcagcagggg atcattttgc 8640

gcttcagcca tacttttcat actcccgcca ttcagagaag aaaccaattg tccatattgc 8700

atcagacatt gccgtcactg cgtcttttac tggctcttct cgctaaccaa accggaaacc 8760

ccgcttatta aaagcattct gtaacaaagc gggaccaaag ccatgacaaa aacgcgtaac 8820

aaaagtgtct ataatcacgg cagaaaagtc cacattgatt atttgcacgg cgtcacactt 8880

tgctatgcca tagcattttt atccataaga ttagcggatc ctacctgacg ctttttatcg 8940

caactctcta ctgtcgtgcg caacgcgtga catcgactag gaggcctttc tatgcagttc 9000

aaagtgtata cctataaacg cgaaagccgt tatcgtctgt ttgttgatgt tcagagcgat 9060

attattgata caccgggtcg tcgtatggtt attccgctgg caagcgcacg tctgctgagc 9120

gataaagtta gccgtgaact gtatccggtt gttcatattg gtgatgaaag ctggcgtatg 9180

atgaccaccg atatggcaag cgttccggtt agcgttattg gtgaagaagt tgcagatctg 9240

agccatcgtg aaaacgatat caaaaatgcc atcaacctga tgttttgggg catcttatag 9300

gacccccctcc accccgcctc gcgggcagtt tgcctggcgg cagtagcgcg gtggtcccac 9360

ctgaccccat gccgaactca gaagtgaaac gccgtagcgc cgatggtagt gtgggctctc 9420

cccatgcgag agtagggaac tgccaggcat caaataaaac gaaaggctca gtcgaaagac 9480

tgggcctta 9489

<210> 120
<211> 6740

<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> pCas81

<400> 120

[0166]

ggtggaagcg gaggcagcgg gggatcaggc catcatcatc accatcatta atgaacgggt 60

cttgaggggt tttttggact tttgcgtatt gggcgctctt ccgcttcctc gctcactgac 120

tcgctgcgct cggtcgttcg gctgcggcga gcggtatcag ctcactcaaa ggcggtaata 180

cggttatcca cagaatcagg ggataacgca ggaaagaaca tgtgagcaaa aggccagcaa 240

aaggccagga accgtaaaaa ggccgcgttg ctggcgtttt tccataggct ccgccccccct 300

gacgagcatc acaaaaatcg acgctcaagt cagaggtggc gaaacccgac aggactataa 360

agataccagg cgtttccccc tggaagctcc ctcgtgcgct ctcctgttcc gaccctgccg 420

cttaccggat acctgtccgc ctttctccct tcgggaagcg tggcgctttc tcatagctca 480

cgctgtaggt atctcagttc ggtgtaggtc gttcgctcca agctgggctg tgtgcacgaa 540

ccccccgttc agcccgaccg ctgcgcctta tccggtaact atcgtcttga gtccaacccg 600

gtaagacacg acttatcgcc actggcagca gccactggta acaggattag cagagcgagg 660

tatgtaggcg gtgctacaga gttcttgaag tggtggccta actacggcta cactagaagg 720

acagtatttg gtatctgcgc tctgctgaag ccagttacct tcggaaaaag agttggtagc 780

tcttgatccg gcaaacaaac caccgctggt agcggtggtt tttttgtttg caagcagcag 840

attacgcgca gaaaaaaagg atctcaagaa gatcctttga tcttttctac ggggtctgac 900

gctcagtgga acgaaaactc acgttaaggg attttggtca tgagattatc aaaaaggatc 960

ttcacctaga tccttttaaa ttaaaaatga agttttaaat caatctaaag tatatatgag 1020

taaacttggt ctgacagtta ccaatgctta atcagtgagg cacctatctc agcgatctgt 1080

ctatttcgtt catccatagt tgcctgactc cccgtcgtgt agataactac gatacgggag 1140

ggcttaccat ctggccccag tgctgcaatg ataccgcgag atccacgctc accggctcca 1200

gatttatcag caataaacca gccagccgga agggccgagc gcagaagtgg tcctgcaact 1260

ttatccgcct ccatccagtc tattaattgt tgccgggaag ctagagtaag tagttcgcca 1320

gttaatagtt tgcgcaacgt tgttgccatt gctgcaggca tcgtggtgtc acgctcgtcg 1380

tttggtatgg cttcattcag ctccggttcc caacgatcaa ggcgagttac atgatccccc 1440

atgttgtgca aaaaagcggt tagctccttc ggtcctccga tcgttgtcag aagtaagttg 1500

gccgcagtgt tatcactcat ggttatggca gcactgcata attctcttac tgtcatgcca 1560

[0167]

tccgtaagat gcttttctgt gactggtgag tactcaacca agtcattctg agaatagtgt 1620

atgcggcgac cgagttgctc ttgcccggcg tcaatacggg ataataccgc gccacatagc 1680

agaactttaa aagtgctcat cattggaaaa cgttcttcgg ggcgaaaact ctcaaggatc 1740

ttaccgctgt tgagatccag ttcgatgtaa cccactcgtg cacccaactg atcttcagca 1800

tcttttactt tcaccagcgt ttctgggtga gcaaaaacag gaaggcaaaa tgccgcaaaa 1860

aagggaataa gggcgacacg gaaatgttga atactcatac tcttcctttt tcaatattat 1920

tgaagcattt atcagggtta ttgtctcatg agcggataca tatttgaatg tatttagaaa 1980

aataaacaaa taggggttcc gcgcacattt ccccgaaaag tgccactcct cttgctgtga 2040

ccgtctccgg gagctgcatg tgtcagaggt tttcaccgtc atcaccgaaa cgcgcgaggc 2100

agcagatcaa ttcgcgcgcg aaggcgaagc ggcatgcatt tacgttgaca ccatcgaatg 2160

gtgcaaaacc tttcgcggta tggcatgata gcgcccggaa gagagtcaat tcagggtggt 2220

gaatgtgaaa ccagtaacgt tatacgatgt cgcagagtat gccggtgtct cttatcagac 2280

cgtttcccgc gtggtgaacc aggccagcca cgtttctgcg aaaacgcggg aaaaagtgga 2340

agcggcgatg gcggagctga attacattcc caaccgcgtg gcacaacaac tggcgggcaa 2400

acagtcgttg ctgattggcg ttgccacctc cagtctggcc ctgcacgcgc cgtcgcaaat 2460

tgtcgcggcg attaaatctc gcgccgatca actgggtgcc agcgtggtgg tgtcgatggt 2520

agaacgaagc ggcgtcgaag cctgtaaagc ggcggtgcac aatcttctcg cgcaacgcgt 2580

cagtgggctg atcattaact atccgctgga tgaccaggat gccattgctg tggaagctgc 2640

ctgcactaat gttccggcgt tatttcttga tgtctctgac cagacaccca tcaacagtat 2700

tattttctcc catgaagatg gtacgcgact gggcgtggag catctggtcg cattgggtca 2760

ccagcaaatc gcgctgttag cgggcccatt aagttctgtc tcggcgcgtc tgcgtctggc 2820

tggctggcat aaatatctca ctcgcaatca aattcagccg atagcggaac gggaaggcga 2880

ctggagtgcc atgtccggtt ttcaacaaac catgcaaatg ctgaatgagg gcatcgttcc 2940

cactgcgatg ctggttgcca acgatcagat ggcgctgggc gcaatgcgcg ccattaccga 3000

gtccgggctg cgcgttggtg cggatatctc ggtagtggga tacgacgata ccgaagatag 3060

ctcatgttat atcccgccgt taaccaccat caaacaggat tttcgcctgc tggggcaaac 3120

cagcgtggac cgcttgctgc aactctctca gggccaggcg gtgaagggca atcagctgtt 3180

gcccgtctca ctggtgaaaa gaaaaaccac cctggcgccc aatacgcaaa ccgcctctcc 3240

ccgcgcgttg gccgattcat taatgcagct ggcacgacag gtttcccgac tggaaagcgg 3300

gcagtgagcg caatattctg aaatgagctg ttgacaatta atcatccggc tcgtataatg 3360

tgtggaattg tgagcggata acaattttca cacaggaaac agaattctag cattgtgagc 3420

ggataacaat tcccctctag aaataatttt gtttaacttt aagaaggaga tataccatga 3480

[0168]

aacgtccggc agcaaccaaa aaagcaggtc aggccaagaa aaaaaaaggt ggtggttcag 3540

gtaaccagaa atttatcctg ggtctggata ttggtattac cagcgttggt tatggcctga 3600

ttgattacga aaccaaaaac attattgatg ccggtgttcg tctgtttccg gaagcaaatg 3660

ttgaaaataa tgaaggtcgt cgtagcaaac gtggtagccg tcgtctgaaa cgtcgtcgta 3720

ttcatcgtct ggaacgtgtt aaaaaactgc tggaagatta taacctgctg gatcagagcc 3780

agattccgca gagcaccaat ccgtatgcaa ttcgtgttaa aggtctgagc gaagcactga 3840

gcaaagatga actggttatt gcactgctgc atattgcaaa acgccgtggc attcataaaa 3900

tcgatgtgat tgatagcaat gacgatgtgg gtaatgaact gagcaccaaa gaacagctga 3960

acaaaaatag caaactgctg aaagacaaat tcgtgtgtca gattcagctg gaacgtatga 4020

atgaaggcca ggttcgtggt gaaaagaatc gctttaaaac cgcagacatc atcaaagaaa 4080

ttatccagct gctgaacgtg cagaaaaact tccatcagct ggatgaaaac ttcatcaaca 4140

aatacatcga gctggttgaa atgcgtcgcg aatattttga aggtccgggt aaaggtagcc 4200

cgtatggttg ggaaggtgat ccgaaagcat ggtatgaaac cctgatgggt cattgtacct 4260

attttccgga tgaactgcgt agcgttaaat atgcctatag cgcagacctg tttaatgcac 4320

tgaatgatct gaataacctg gtgattcagc gtgatggtct gagcaaactg gaatatcatg 4380

agaaatatca catcatcgaa aacgtgttca aacagaagaa gaaaccgacc ctgaaacaaa 4440

tcgccaacga aattaatgtg aacccggaag atattaaagg ctaccgtatt accaaaagcg 4500

gtaaaccgca gttcaccgaa tttaaactgt atcacgatct gaaaagcgtg ctgtttgatc 4560

agagcattct ggaaaatgaa gatgtgctgg accagattgc agaaattctg accatttatc 4620

aggacaaaga cagcatcaaa agcaaactga ccgaactgga tattctgctg aatgaagaag 4680

ataaagagaa cattgcacag ctgaccggtt ataccggcac ccatcgtctg agcctgaaat 4740

gtattcgtct ggtactggaa gaacagtggt atagcagccg taatcagatg gaaatcttta 4800

cccatctgaa cattaaaccg aagaaaatca atctgaccgc agccaacaaa attccgaaag 4860

ccatgattga tgagtttatt ctgagtccgg ttgtgaaacg tacctttggt caggcaatta 4920

acctgatcaa caaaatcatt gaaaaatatg gcgtgcctga ggatatcatt attgaactgg 4980

cacgtgaaaa caacagcaaa gataaacaga aattcatcaa cgagatgcag aagaagaacg 5040

aaaatacccg caaacggatt aacgagatca ttggcaaata tggtaatcag aatgccaaac 5100

gcctggtgga aaaaattcgt ctgcatgatg aacaagaggg caaatgtctg tatagcctgg 5160

aaagcattcc tctggaagat ctgctgaaca atccgaatca ttatgaagtg gatcacatta 5220

ttccgcgtag cgtgagcttt gataattcct atcataataa agtgctggtg aaacagagcg 5280

aaaactccaa aaaatccaac ctgacaccgt atcagtattt caatagcggc aaatccaaac 5340

tgagctacaa ccagtttaaa cagcatattc tgaacctgag caaaagccag gatcgcatca 5400

[0169]

gcaagaagaa gaaggagtac ctgctggaag aacgcgacat caacaaattt gaagtgcaga 5460

aagaatttat caaccgcaac ctggttgata cccgttatgc aacccgtgaa ctgaccaatt 5520

atctgaaagc atatttcagc gccaacaaca tgaacgtgaa agtgaaaacg attaacggca 5580

gctttaccga ttatctgcgt aaagtgtgga aattcaaaaa agaacgcaac cacggctata 5640

aacatcatgc cgaagatgcc ctgattattg caaatgcaga tttcctgttt aaagaaaaca 5700

aaaaactgaa agccgtcaac agcgtgctgg aaaaaccgga aattgagaca aaacagctgg 5760

acattcaggt tgatagcgaa gataattaca gcgaaatgtt tatcatcccg aaacaggtgc 5820

aggatatcaa agattttcgc aacttcaaat atagccaccg cgttgacaaa aaacctaatc 5880

gtcagctgat taacgatacc ctgtatagca cccgcaaaaa agataacagc acctatattg 5940

tgcagaccat taaagacatc tacgccaaag ataataccac cctgaaaaaa cagttcgaca 6000

aaagcccaga aaaatttctg atgtatcagc atgatccgcg taccttcgaa aaactggaag 6060

ttattatgaa acagtatgcc aacgagaaaa atccgctggc caaatatcac gaagaaaccg 6120

gtgaatatct gaccaaatat tccaagaaga acaacggtcc gatcgttaaa tccctgaaat 6180

atatcggtaa taaactgggc agccatctgg atgttaccca tcagtttaaa agctccacaa 6240

agaagctggt taaactgtcc atcaaaccgt atcgctttga tgtgtatctg accgacaaag 6300

gctataaatt cattaccatc agctatctgg acgtgctgaa aaaagacaac tattattata 6360

tcccggaaca gaaatatgat aaactgaaac tgggtaaagc catcgataaa aacgccaaat 6420

ttatcgccag cttctacaaa aacgacctga ttaaactgga tggcgagatc tataaaatca 6480

tcggtgttaa tagcgacacc cgcaatatga ttgagctgga tctgccggat attcgctata 6540

aagaatattg cgaactgaac aacattaaag gcgaaccgcg tatcaaaaag accatcggca 6600

aaaaagtgaa tagcatcgag aaactgacca ccgatgttct gggtaatgtg tttaccaata 6660

cccagtatac caaacctcag ctgctgttca aacgcggtaa tggcggagga tctggccccc 6720

ctaagaaaaa gcggaaggtg 6740

<210> 121
<211> 6740
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> pCas888

<400> 121
tgccacctcc agtctggccc tgcacgcgcc gtcgcaaatt gtcgcggcga ttaaatctcg 60

cgccgatcaa ctgggtgcca gcgtggtggt gtcgatggta gaacgaagcg gcgtcgaagc 120

[0170]

ctgtaaagcg gcggtgcaca atcttctcgc gcaacgcgtc agtgggctga tcattaacta 180

tccgctggat gaccaggatg ccattgctgt ggaagctgcc tgcactaatg ttccggcgtt 240

atttcttgat gtctctgacc agacacccat caacagtatt attttctccc atgaagatgg 300

tacgcgactg ggcgtggagc atctggtcgc attgggtcac cagcaaatcg cgctgttagc 360

gggcccatta agttctgtct cggcgcgtct gcgtctggct ggctggcata aatatctcac 420

tcgcaatcaa attcagccga tagcggaacg ggaaggcgac tggagtgcca tgtccggttt 480

tcaacaaacc atgcaaatgc tgaatgaggg catcgttccc actgcgatgc tggttgccaa 540

cgatcagatg gcgctgggcg caatgcgcgc cattaccgag tccgggctgc gcgttggtgc 600

ggatatctcg gtagtgggat acgacgatac cgaagatagc tcatgttata tcccgccgtt 660

aaccaccatc aaacaggatt ttcgcctgct ggggcaaacc agcgtggacc gcttgctgca 720

actctctcag ggccaggcgg tgaagggcaa tcagctgttg cccgtctcac tggtgaaaag 780

aaaaaccacc ctggcgccca atacgcaaac cgcctctccc cgcgcgttgg ccgattcatt 840

aatgcagctg gcacgacagg tttcccgact ggaaagcggg cagtgagcgc aatattctga 900

aatgagctgt tgacaattaa tcatccggct cgtataatgt gtggaattgt gagcggataa 960

caattttcac acaggaaaca gaattctagc attgtgagcg gataacaatt cccctctaga 1020

aataattttg tttaacttta agaaggagat ataccatgaa acgtccggca gcaaccaaaa 1080

aagcaggtca ggccaagaaa aaaaaaggtg gtggttcagg caacaactat attctgggtc 1140

tggatattgg tattaccagc gttggttatg gtattgttga tagcgatacc cgtgaaatta 1200

aagatgccgg tgttcgtctg tttccggaag caaatgttga taataatgaa ggtcgtcgta 1260

gcaaacgtgg tgcacgtcgt ctgaaacgtc gtcgtattca tcgtctggat cgtgttaaac 1320

atctgctggc agaatatgat ctgctggatc tgaccaatat tccgaaaagc accaatccgt 1380

atcagacccg tgttaaaggt ctgaatgaaa agctgagcaa agatgaactg gttattgcac 1440

tgctgcatat tgcaaaacgc cgtggcattc ataatgtgaa tgttatgatg gatgataacg 1500

atagcggtaa tgaactgagc accaaagatc agctgaagaa gaatgcaaaa gcactgagcg 1560

ataaatatgt ttgtgaactg cagctggaac gctttgagca ggattataaa gttcgtggtg 1620

aaaagaaccg cttcaaaacc gaagattttg ttcgtgaagc acgtaaactg ctggaaaccc 1680

agagcaaatt tttcgaaatt gatcagacgt tcatcatgcg ctatatcgaa ctgattgaaa 1740

cccgtcgcga atattttgaa ggtccgggta aaggtagccc gtttggttgg gaaggtaata 1800

tcaagaaatg gtttgagcag atgatgggcc actgtaccta ttttccagaa gaactgcgta 1860

gcgtcaaata tagctattca gccgaactgt ttaacgccct gaatgatctg aataatctgg 1920

tgattacccg tgatgaagat gccaaactga attatggtga gaaattccag atcatcgaaa 1980

acgtgttcaa acagaagaaa acaccgaacc tgaaacaaat cgccattgaa attggtgtgc 2040

[0171]

atgaaaccga aatcaaaggt tatcgtgtga acaaaagcgg caaaccggaa tttacccagt 2100
tcaaactgta tcacgatctg aagaacatct tcaaagaccc gaaatacctg aacgatatcc 2160
agctgatgga taacattgca gaaatcatca ccatttatca ggatgccgaa agcatcatca 2220
aagaactgaa tcagctgccg gaactgctga gcgaacgtga aaaggagaag atcagcgcac 2280
tgagcggtta tagcggcacc catcgtctga gcctgaaatg tattaatctg ctgctggatg 2340
atctgtggga aagcagcctg aatcagatgg aactgttcac aaaactgaat ctgaaaccga 2400
agaaaatcga tctgagccag cagcataaaa ttccgagcaa actggtggat gactttattc 2460
tgagtccggt tgttaaacgt gcatttattc agagtatcca ggtggtgaat gccatcattg 2520
ataaatacgg tctgccggaa gatatcatca ttgaactggc acgtgaaaac aatagtgatg 2580
atcgtcgcaa atttctgaac cagctgcaaa agcagaatga ggaaacccgt aaacaggttg 2640
aaaaggtgct gcgtgaatat ggcaatgata atgccaaacg tatcgtgcag aaaatcaaac 2700
tgcataatat gcaagagggg aaatgtctgt atagcctgaa agatattccg ctggaagatc 2760
tgctgcgtaa tccgcatcat tatgaagtgg atcatattat tccgcgtagc gtggcatttg 2820
ataatagcat gcataataaa gttctggtgc gtgccgatga gaatagcaag aaaggtaatc 2880
gtaccccgta tcagtatctg aatagcagcg aaagcagtct gagctataac gaatttaaac 2940
agcatatcct gaacctgagc aaaaccaaag accgcatcac caagaagaag cgtgaatacc 3000
tgctggaaga acgcgacatc aacaaatttg atgtgcagaa agaattcatc aaccgcaatc 3060
tggttgatac acgttatgca acccgtgaac tgaccagcct gctgaaagca tatttcagcg 3120
caaataacct ggacgtgaaa gtgaaaacga ttaatggcag ctttaccaac tatctgcgta 3180
aagtgtggaa attcgataaa gatcgcaaca aaggctataa acatcatgca gaagatgccc 3240
tgattattgc caatgcagat tttctgttca aacacaacaa gaaactgcgc aacattaaca 3300
aagtgctgga tgcaccgagc aaagaagttg ataagaaacg cgtgacagtt cagagcgaag 3360
atgaatataa ccagattttt gaggataccc agaaagccca ggcaatcaag aaattcgaaa 3420
tccgcaaatt tagccaccgc gtggacaaga aaccgaatcg tcagctgatt aatgataccc 3480
tgtatagcac ccgtaatatc gatggtattg aatatgtggt cgagagcatc aaagatatct 3540
atagcgtgaa caacgataaa gttaaaacca aattcaagaa agacccgcac cgtctgctga 3600
tgtatcgtaa tgatccgcag acctttgaga aattcgagaa agtctttaaa cagtatgaga 3660
gcgagaagaa cccgttcgcc aaatattacg aagaaaccgg tgagaaaatt cgcaaattca 3720
gcaaaaccgg tcagggtccg tatatcaaca aaatcaaata tctgcgtgaa cgcctgggtc 3780
gtcattgtga tgttaccaac aaatatatca atagccgcaa caaaattgtg caactgtcca 3840
tcaaaccgta tcgctttgat gtgtatctga ccgacaaagg ctataaattc attaccatca 3900
gctatctgga cgtgctgaaa aaagacaact attattatat cccggaacag aaatatgata 3960

[0172]

aactgaaact gggtaaagcc atcgataaaa acgccaaatt tatcgccagc ttctacaaaa 4020

acgacctgat taaactggat ggcgagatct ataaaatcat cggtgttaat agcgacaccc 4080

gcaatatgat tgagctggat ctgccggata ttcgctataa agaatattgc gaactgaaca 4140

acattaaagg cgaaccgcgt atcaaaaaga ccatcggcaa aaaagtgaat agcatcgaga 4200

aactgaccac cgatgttctg ggtaatgtgt ttaccaatac ccagtatacc aaacctcagc 4260

tgctgttcaa acgcggtaat ggcggaggat ctggcccccc taagaaaaag cggaaggtgg 4320

gtggaagcgg aggcagcggg ggatcaggcc atcatcatca ccatcattaa tgaacgggtc 4380

ttgaggggtt ttttggactt ttgcgtattg ggcgctcttc cgcttcctcg ctcactgact 4440

cgctgcgctc ggtcgttcgg ctgcggcgag cggtatcagc tcactcaaag gcggtaatac 4500

ggttatccac agaatcaggg gataacgcag gaaagaacat gtgagcaaaa ggccagcaaa 4560

aggccaggaa ccgtaaaaag gccgcgttgc tggcgttttt ccataggctc cgcccccctg 4620

acgagcatca caaaaatcga cgctcaagtc agaggtggcg aaacccgaca ggactataaa 4680

gataccaggc gtttcccccc ggaagctccc tcgtgcgctc tcctgttccg accctgccgc 4740

ttaccggata cctgtccgcc tttctccctt cgggaagcgt ggcgctttct catagctcac 4800

gctgtaggta tctcagttcg gtgtaggtcg ttcgctccaa gctgggctgt gtgcacgaac 4860

cccccgttca gcccgaccgc tgcgccttat ccggtaacta tcgtcttgag tccaacccgg 4920

taagacacga cttatcgcca ctggcagcag ccactggtaa caggattagc agagcgaggt 4980

atgtaggcgg tgctacagag ttcttgaagt ggtggcctaa ctacggctac actagaagga 5040

cagtatttgg tatctgcgct ctgctgaagc cagttacctt cggaaaaaga gttggtagct 5100

cttgatccgg caaacaaacc accgctggta gcggtggttt ttttgtttgc aagcagcaga 5160

ttacgcgcag aaaaaaagga tctcaagaag atcctttgat cttttctacg gggtctgacg 5220

ctcagtggaa cgaaaactca cgttaaggga ttttggtcat gagattatca aaaaggatct 5280

tcacctagat ccttttaaat taaaaatgaa gttttaaatc aatctaaagt atatatgagt 5340

aaacttggtc tgacagttac caatgcttaa tcagtgaggc acctatctca gcgatctgtc 5400

tatttcgttc atccatagtt gcctgactcc ccgtcgtgta gataactacg atacgggagg 5460

gcttaccatc tggccccagt gctgcaatga taccgcgaga tccacgctca ccggctccag 5520

atttatcagc aataaaccag ccagccggaa gggccgagcg cagaagtggt cctgcaactt 5580

tatccgcctc catccagtct attaattgtt gccgggaagc tagagtaagt agttcgccag 5640

ttaatagttt gcgcaacgtt gttgccattg ctgcaggcat cgtggtgtca cgctcgtcgt 5700

ttggtatggc ttcattcagc tccggttccc aacgatcaag gcgagttaca tgatccccca 5760

tgttgtgcaa aaaagcggtt agctccttcg gtcctccgat cgttgtcaga agtaagttgg 5820

ccgcagtgtt atcactcatg gttatggcag cactgcataa ttctcttact gtcatgccat 5880

ccgtaagatg cttttctgtg actggtgagt actcaaccaa gtcattctga gaatagtgta 5940

tgcggcgacc gagttgctct tgcccggcgt caatacggga taataccgcg ccacatagca 6000

gaactttaaa agtgctcatc attggaaaac gttcttcggg gcgaaaactc tcaaggatct 6060

taccgctgtt gagatccagt tcgatgtaac ccactcgtgc acccaactga tcttcagcat 6120

cttttacttt caccagcgtt tctgggtgag caaaaacagg aaggcaaaat gccgcaaaaa 6180

agggaataag ggcgacacgg aaatgttgaa tactcatact cttccttttt caatattatt 6240

gaagcattta tcagggttat tgtctcatga gcggatacat atttgaatgt atttagaaaa 6300

ataaacaaat aggggttccg cgcacatttc cccgaaaagt gccactcctc ttgctgtgac 6360

cgtctccggg agctgcatgt gtcagaggtt ttcaccgtca tcaccgaaac gcgcgaggca 6420

gcagatcaat tcgcgcgcga aggcgaagcg gcatgcattt acgttgacac catcgaatgg 6480

tgcaaaacct ttcgcggtat ggcatgatag cgcccggaag agagtcaatt cagggtggtg 6540

aatgtgaaac cagtaacgtt atacgatgtc gcagagtatg ccggtgtctc ttatcagacc 6600

gtttcccgcg tggtgaacca ggccagccac gtttctgcga aaacgcggga aaaagtggaa 6660

gcggcgatgg cggagctgaa ttacattccc aaccgcgtgg cacaacaact ggcgggcaaa 6720

cagtcgttgc tgattggcgt 6740

[0173]

<210> 122
<211> 1431
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> pCas889

<400> 122
cggaacggga aggcgactgg agtgccatgt ccggttttca acaaaccatg caaatgctga 60

atgagggcat cgttcccact gcgatgctgg ttgccaacga tcagatggcg ctgggcgcaa 120

tgcgcgccat taccgagtcc gggctgcgcg ttggtgcgga tatctcggta gtgggatacg 180

acgataccga agatagctca tgttatatcc cgccgttaac caccatcaaa caggattttc 240

gcctgctggg gcaaaccagc gtggaccgct tgctgcaact ctctcagggc aggcggtga 300

agggcaatca gctgttgccc gtctcactgg tgaaaagaaa aaccaccctg gcgcccaata 360

cgcaaaccgc ctctccccgc gcgttggccg attcattaat gcagctggca cgacaggttt 420

cccgactgga aagcgggcag tgagcgcaat attctgaaat gagctgttga caattaatca 480

tccggctcgt ataatgtgtg gaattgtgag cggataacaa ttttcacaca ggaaacagaa 540

ttctagcatt gtgagcggat aacaattccc ctctagaaat aattttgttt aactttaaga 600

aggagatata ccatgaaacg tccggcagca accaaaaaag caggtcaggc caagaaaaaa 660

aaaggtggtg gttcaggcaa caactatatt ctgggtctgg atattggtat taccagcgtt 720

ggttatggta ttgttgatag cgatacccgt gaaattaaag atgccggtgt tcgtctgttt 780

ccggaagcaa atgttgataa taatgaaggt cgtcgtagca aacgtggtgc acgtcgtctg 840

aaacgtcgtc gtattcatcg tctggatcgt gttaaacatc tgctggcaga atatgatctg 900

ctggatctga ccaatattcc gaaaagcacc aatccgtatc agacccgtgt taaaggtctg 960

aatgaaaagc tgagcaaaga tgaactggtt attgcactgc tgcatattgc aaaacgccgt 1020

ggcattcata atgtgaatgt tatgatggat gataacgata gcggtaatga actgagcacc 1080

aaagatcagc tgaagaagaa tgcaaaagca ctgagcgata aatatgtttg tgaactgcag 1140

ctggaacgct ttgagcagga ttataaagtt cgtggtgaaa agaaccgctt caaaaccgaa 1200

gattttgttc gtgaagcacg taaactgctg gaaacccaga gcaaattttt cgaaattgat 1260

cagacgttca tcatgcgcta tatcgaactg attgaaaccc gtcgcgaata ttttgaaggt 1320

ccgggtaaag gtagcccgtt tggttgggaa ggtaatatca agaaatggtt tgagcagatg 1380

atgggccact gtacctattt tccagaagaa ctgcgtagcg tcaaatatag c 1431

[0174]

<210> 123
<211> 156
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> SluCas9 tracr RNA

<400> 123
cttgtactta tacctaaaat tacagaatct actgaaacaa gacaatatgt cgtgtttatc 60

ccatcaattt attggtggga ttttttttatg tttttagcaa aaagtaatac catactttat 120

atttttaaat tataataaag atataaataa aggtgg 156

<210> 124
<211> 3315
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> SluCas9密码子优化的DNA

<400> 124

[0175]

atgaaacgtc cggcagcaac caaaaaagca ggtcaggcca agaaaaaaaa aggtggtggt 60

tcaggtaacc agaaatttat cctgggtctg gatattggta ttaccagcgt tggttatggc 120

ctgattgatt acgaaaccaa aaacattatt gatgccggtg ttcgtctgtt tccggaagca 180

aatgttgaaa ataatgaagg tcgtcgtagc aaacgtggta gccgtcgtct gaaacgtcgt 240

cgtattcatc gtctggaacg tgttaaaaaa ctgctggaag attataacct gctggatcag 300

agccagattc cgcagagcac caatccgtat gcaattcgtg ttaaaggtct gagcgaagca 360

ctgagcaaag atgaactggt tattgcactg ctgcatattg caaaacgccg tggcattcat 420

aaaatcgatg tgattgatag caatgacgat gtgggtaatg aactgagcac caaagaacag 480

ctgaacaaaa atagcaaact gctgaaagac aaattcgtgt gtcagattca gctggaacgt 540

atgaatgaag gccaggttcg tggtgaaaag aatcgcttta aaaccgcaga catcatcaaa 600

gaaattatcc agctgctgaa cgtgcagaaa aacttccatc agctggatga aaacttcatc 660

aacaaataca tcgagctggt tgaaatgcgt cgcgaatatt ttgaaggtcc gggtaaaggt 720

agcccgtatg gttgggaagg tgatccgaaa gcatggtatg aaaccctgat gggtcattgt 780

acctattttc cggatgaact gcgtagcgtt aaatatgcct atagcgcaga cctgtttaat 840

gcactgaatg atctgaataa cctggtgatt cagcgtgatg gtctgagcaa actggaatat 900

catgagaaat atcacatcat cgaaaacgtg ttcaaacaga agaagaaacc gaccctgaaa 960

caaatcgcca acgaaattaa tgtgaacccg gaagatatta aaggctaccg tattaccaaa 1020

agcggtaaac cgcagttcac cgaatttaaa ctgtatcacg atctgaaaag cgtgctgttt 1080

gatcagagca ttctggaaaa tgaagatgtg ctggaccaga ttgcagaaat tctgaccatt 1140

tatcaggaca aagacagcat caaaagcaaa ctgaccgaac tggatattct gctgaatgaa 1200

gaagataaag agaacattgc acagctgacc ggttataccg gcacccatcg tctgagcctg 1260

aaatgtattc gtctggtact ggaagaacag tggtatagca gccgtaatca gatggaaatc 1320

tttacccatc tgaacattaa accgaagaaa atcaatctga ccgcagccaa caaaattccg 1380

aaagccatga ttgatgagtt tattctgagt ccggttgtga aacgtacctt tggtcaggca 1440

attaacctga tcaacaaaat cattgaaaaa tatggcgtgc ctgaggatat cattattgaa 1500

ctggcacgtg aaaacaacag caaagataaa cagaaattca tcaacgagat gcagaagaag 1560

aacgaaaata cccgcaaacg gattaacgag atcattggca aatatggtaa tcagaatgcc 1620

aaacgcctgg tggaaaaaat tcgtctgcat gatgaacaag agggcaaatg tctgtatagc 1680

ctggaaagca ttcctctgga agatctgctg aacaatccga atcattatga agtggatcac 1740

attattccgc gtagcgtgag ctttgataat tcctatcata ataaagtgct ggtgaaacag 1800

agcgaaaact ccaaaaaatc caacctgaca ccgtatcagt atttcaatag cggcaaatcc 1860

[0176]

aaactgagct acaaccagtt taaacagcat attctgaacc tgagcaaaag ccaggatcgc 1920

atcagcaaga agaagaagga gtacctgctg gaagaacgcg acatcaacaa atttgaagtg 1980

cagaaagaat ttatcaaccg caacctggtt gatacccgtt atgcaacccg tgaactgacc 2040

aattatctga aagcatattt cagcgccaac aacatgaacg tgaaagtgaa aacgattaac 2100

ggcagcttta ccgattatct gcgtaaagtg tggaaattca aaaaagaacg caaccacggc 2160

tataaacatc atgccgaaga tgccctgatt attgcaaatg cagatttcct gtttaaagaa 2220

aacaaaaaac tgaaagccgt caacagcgtg ctggaaaaac cggaaattga gacaaaacag 2280

ctggacattc aggttgatag cgaagataat tacagcgaaa tgtttatcat cccgaaacag 2340

gtgcaggata tcaaagattt tcgcaacttc aaatatagcc accgcgttga caaaaaacct 2400

aatcgtcagc tgattaacga taccctgtat agcacccgca aaaaagataa cagcacctat 2460

attgtgcaga ccattaaaga catctacgcc aaagataata ccaccctgaa aaaacagttc 2520

gacaaaagcc cagaaaaatt tctgatgtat cagcatgatc cgcgtacctt cgaaaaactg 2580

gaagttatta tgaaacagta tgccaacgag aaaaatccgc tggccaaata tcacgaagaa 2640

accggtgaat atctgaccaa atattccaag aagaacaacg gtccgatcgt taaatccctg 2700

aaatatatcg gtaataaact gggcagccat ctggatgtta cccatcagtt taaaagctcc 2760

acaaagaagc tggttaaact gtccatcaaa ccgtatcgct ttgatgtgta tctgaccgac 2820

aaaggctata aattcattac catcagctat ctggacgtgc tgaaaaaaga caactattat 2880

tatatcccgg aacagaaata tgataaactg aaactgggta aagccatcga taaaaacgcc 2940

aaatttatcg ccagcttcta caaaaacgac ctgattaaac tggatggcga gatctataaa 3000

atcatcggtg ttaatagcga cacccgcaat atgattgagc tggatctgcc ggatattcgc 3060

tataaagaat attgcgaact gaacaacatt aaaggcgaac cgcgtatcaa aaagaccatc 3120

ggcaaaaaag tgaatagcat cgagaaactg accaccgatg ttctgggtaa tgtgtttacc 3180

aatacccagt ataccaaacc tcagctgctg ttcaaacgcg gtaatggcgg aggatctggc 3240

ccccctaaga aaaagcggaa ggtgggtgga agcggaggca gcgggggatc aggccatcat 3300

catcaccatc attaa 3315

<210> 125
<211> 1056
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> SluCas9多肽

<400> 125

Met Asn Gln Lys Phe Ile Leu Gly Leu Asp Ile Gly Ile Thr Ser Val
1 5 10 15

Gly Tyr Gly Leu Ile Asp Tyr Glu Thr Lys Asn Ile Ile Asp Ala Gly
20 25 30

Val Arg Leu Phe Pro Glu Ala Asn Val Glu Asn Asn Glu Gly Arg Arg
35 40 45

Ser Lys Arg Gly Ser Arg Arg Leu Lys Arg Arg Arg Ile His Arg Leu
50 55 60

Glu Arg Val Lys Lys Leu Leu Glu Asp Tyr Asn Leu Leu Asp Gln Ser
65 70 75 80

Gln Ile Pro Gln Ser Thr Asn Pro Tyr Ala Ile Arg Val Lys Gly Leu
85 90 95

Ser Glu Ala Leu Ser Lys Asp Glu Leu Val Ile Ala Leu Leu His Ile
100 105 110

Ala Lys Arg Arg Gly Ile His Lys Ile Asp Val Ile Asp Ser Asn Asp
115 120 125

[0177]

Asp Val Gly Asn Glu Leu Ser Thr Lys Glu Gln Leu Asn Lys Asn Ser
130 135 140

Lys Leu Leu Lys Asp Lys Phe Val Cys Gln Ile Gln Leu Glu Arg Met
145 150 155 160

Asn Glu Gly Gln Val Arg Gly Glu Lys Asn Arg Phe Lys Thr Ala Asp
165 170 175

Ile Ile Lys Glu Ile Ile Gln Leu Leu Asn Val Gln Lys Asn Phe His
180 185 190

Gln Leu Asp Glu Asn Phe Ile Asn Lys Tyr Ile Glu Leu Val Glu Met
195 200 205

Arg Arg Glu Tyr Phe Glu Gly Pro Gly Lys Gly Ser Pro Tyr Gly Trp
210 215 220

Glu Gly Asp Pro Lys Ala Trp Tyr Glu Thr Leu Met Gly His Cys Thr
225 230 235 240

Tyr Phe Pro Asp Glu Leu Arg Ser Val Lys Tyr Ala Tyr Ser Ala Asp
245 250 255

```
Leu Phe Asn Ala Leu Asn Asp Leu Asn Asn Leu Val Ile Gln Arg Asp
            260                 265                 270

Gly Leu Ser Lys Leu Glu Tyr His Glu Lys Tyr His Ile Ile Glu Asn
        275                 280                 285

Val Phe Lys Gln Lys Lys Lys Pro Thr Leu Lys Gln Ile Ala Asn Glu
    290                 295                 300

Ile Asn Val Asn Pro Glu Asp Ile Lys Gly Tyr Arg Ile Thr Lys Ser
305                 310                 315                 320

Gly Lys Pro Gln Phe Thr Glu Phe Lys Leu Tyr His Asp Leu Lys Ser
                325                 330                 335

Val Leu Phe Asp Gln Ser Ile Leu Glu Asn Glu Asp Val Leu Asp Gln
            340                 345                 350

Ile Ala Glu Ile Leu Thr Ile Tyr Gln Asp Lys Asp Ser Ile Lys Ser
        355                 360                 365

Lys Leu Thr Glu Leu Asp Ile Leu Leu Asn Glu Glu Asp Lys Glu Asn
    370                 375                 380
```

[0178]

```
Ile Ala Gln Leu Thr Gly Tyr Thr Gly Thr His Arg Leu Ser Leu Lys
385                 390                 395                 400

Cys Ile Arg Leu Val Leu Glu Glu Gln Trp Tyr Ser Ser Arg Asn Gln
                405                 410                 415

Met Glu Ile Phe Thr His Leu Asn Ile Lys Pro Lys Lys Ile Asn Leu
            420                 425                 430

Thr Ala Ala Asn Lys Ile Pro Lys Ala Met Ile Asp Glu Phe Ile Leu
        435                 440                 445

Ser Pro Val Val Lys Arg Thr Phe Gly Gln Ala Ile Asn Leu Ile Asn
    450                 455                 460

Lys Ile Ile Glu Lys Tyr Gly Val Pro Glu Asp Ile Ile Ile Glu Leu
465                 470                 475                 480

Ala Arg Glu Asn Asn Ser Lys Asp Lys Gln Lys Phe Ile Asn Glu Met
                485                 490                 495

Gln Lys Lys Asn Glu Asn Thr Arg Lys Arg Ile Asn Glu Ile Ile Gly
            500                 505                 510
```

Lys Tyr Gly Asn Gln Asn Ala Lys Arg Leu Val Glu Lys Ile Arg Leu
515 520 525

His Asp Glu Gln Glu Gly Lys Cys Leu Tyr Ser Leu Glu Ser Ile Pro
530 535 540

Leu Glu Asp Leu Leu Asn Asn Pro Asn His Tyr Glu Val Asp His Ile
545 550 555 560

Ile Pro Arg Ser Val Ser Phe Asp Asn Ser Tyr His Asn Lys Val Leu
565 570 575

Val Lys Gln Ser Glu Asn Ser Lys Lys Ser Asn Leu Thr Pro Tyr Gln
580 585 590

Tyr Phe Asn Ser Gly Lys Ser Lys Leu Ser Tyr Asn Gln Phe Lys Gln
595 600 605

His Ile Leu Asn Leu Ser Lys Ser Gln Asp Arg Ile Ser Lys Lys Lys
610 615 620

Lys Glu Tyr Leu Leu Glu Glu Arg Asp Ile Asn Lys Phe Glu Val Gln
625 630 635 640

[0179]

Lys Glu Phe Ile Asn Arg Asn Leu Val Asp Thr Arg Tyr Ala Thr Arg
645 650 655

Glu Leu Thr Asn Tyr Leu Lys Ala Tyr Phe Ser Ala Asn Asn Met Asn
660 665 670

Val Lys Val Lys Thr Ile Asn Gly Ser Phe Thr Asp Tyr Leu Arg Lys
675 680 685

Val Trp Lys Phe Lys Lys Glu Arg Asn His Gly Tyr Lys His His Ala
690 695 700

Glu Asp Ala Leu Ile Ile Ala Asn Ala Asp Phe Leu Phe Lys Glu Asn
705 710 715 720

Lys Lys Leu Lys Ala Val Asn Ser Val Leu Glu Lys Pro Glu Ile Glu
725 730 735

Thr Lys Gln Leu Asp Ile Gln Val Asp Ser Glu Asp Asn Tyr Ser Glu
740 745 750

Met Phe Ile Ile Pro Lys Gln Val Gln Asp Ile Lys Asp Phe Arg Asn
755 760 765

```
Phe Lys Tyr Ser His Arg Val Asp Lys Lys Pro Asn Arg Gln Leu Ile
    770                 775                 780

Asn Asp Thr Leu Tyr Ser Thr Arg Lys Lys Asp Asn Ser Thr Tyr Ile
785                 790                 795                 800

Val Gln Thr Ile Lys Asp Ile Tyr Ala Lys Asp Asn Thr Thr Leu Lys
                805                 810                 815

Lys Gln Phe Asp Lys Ser Pro Glu Lys Phe Leu Met Tyr Gln His Asp
            820                 825                 830

Pro Arg Thr Phe Glu Lys Leu Glu Val Ile Met Lys Gln Tyr Ala Asn
        835                 840                 845

Glu Lys Asn Pro Leu Ala Lys Tyr His Glu Glu Thr Gly Glu Tyr Leu
    850                 855                 860

Thr Lys Tyr Ser Lys Lys Asn Asn Gly Pro Ile Val Lys Ser Leu Lys
865                 870                 875                 880

Tyr Ile Gly Asn Lys Leu Gly Ser His Leu Asp Val Thr His Gln Phe
                885                 890                 895
```

[0180]

```
Lys Ser Ser Thr Lys Lys Leu Val Lys Leu Ser Ile Lys Pro Tyr Arg
            900                 905                 910

Phe Asp Val Tyr Leu Thr Asp Lys Gly Tyr Lys Phe Ile Thr Ile Ser
        915                 920                 925

Tyr Leu Asp Val Leu Lys Lys Asp Asn Tyr Tyr Tyr Ile Pro Glu Gln
    930                 935                 940

Lys Tyr Asp Lys Leu Lys Leu Gly Lys Ala Ile Asp Lys Asn Ala Lys
945                 950                 955                 960

Phe Ile Ala Ser Phe Tyr Lys Asn Asp Leu Ile Lys Leu Asp Gly Glu
                965                 970                 975

Ile Tyr Lys Ile Ile Gly Val Asn Ser Asp Thr Arg Asn Met Ile Glu
            980                 985                 990

Leu Asp Leu Pro Asp Ile Arg Tyr  Lys Glu Tyr Cys Glu  Leu Asn Asn
        995                 1000                1005

Ile Lys  Gly Glu Pro Arg Ile  Lys Lys Thr Ile Gly  Lys Lys Val
    1010                1015                1020
```

Asn Ser Ile Glu Lys Leu Thr Thr Asp Val Leu Gly Asn Val Phe
1025 1030 1035

Thr Asn Thr Gln Tyr Thr Lys Pro Gln Leu Leu Phe Lys Arg Gly
1040 1045 1050

Asn Gly Gly
1055

<210> 126
<211> 11
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> HIV-1 TAT的残基47-57

<400> 126

Tyr Gly Arg Lys Lys Arg Arg Gln Arg Arg Arg
1 5 10

[0181]

<210> 127
<211> 16
<212> PRT
<213> 人工序列

<220>
<223> 合成的多肽

<220>
<221> misc_feature
<223> Penetratin

<400> 127

Arg Gln Ile Lys Ile Trp Phe Gln Asn Arg Arg Met Lys Trp Lys Lys
1 5 10 15

<210> 128
<211> 150
<212> RNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA-Ex6

<400> 128
ggugaacgug gaugaaguug guuuuaguac ucuggaaaca gaaucuacug aaacaagaca 60

auaugucgug uuuaucccau caauuuauug gugggauuuu uuucuagcau aaccccuugg 120

ggccucuaaa cgggucuuga gggguuuuuu 150

<210> 129
<211> 28
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-1-A

<400> 129
ggtgaacgtg gatgaagttg tggggtac 28

<210> 130
<211> 33
<212> DNA
<213> 人工序列

<220>
[0182] <223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-1-B

<400> 130
gcggtacccc acaacttcat ccacgttcac cgg 33

<210> 131
<211> 28
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-2-A

<400> 131
agtgaacgtg gatgaagttg tggggtac 28

<210> 132
<211> 33
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-2-B

<400> 132
gcggtacccc acaacttcat ccacgttcac tgg 33

<210> 133
<211> 28
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-3-A

<400> 133
cgtgaacgtg gatgaagttg tggggtac 28

[0183]

<210> 134
<211> 33
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-3-B

<400> 134
gcggtacccc acaacttcat ccacgttcac ggg 33

<210> 135
<211> 28
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-4-A

<400> 135
tgtgaacgtg gatgaagttg tggggtac 28

<210> 136
<211> 33
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-4-B

<400> 136
gcggtacccc acaacttcat ccacgttcac agg 33

<210> 137
<211> 28
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-5-A

<400> 137
[0184] gatgaacgtg gatgaagttg tggggtac 28

<210> 138
<211> 33
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-5-B

<400> 138
gcggtacccc acaacttcat ccacgttcat cgg 33

<210> 139
<211> 28
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-6-A

<400> 139
gctgaacgtg gatgaagttg tggggtac 28

<210> 140
<211> 33
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-6-B

<400> 140
gcggtacccc acaacttcat ccacgttcag cgg 33

<210> 141
<211> 28
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

[0185]

<220>
<221> misc_feature
<223> R01-7-A

<400> 141
gttgaacgtg gatgaagttg tggggtac 28

<210> 142
<211> 33
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-7-B

<400> 142
gcggtacccc acaacttcat ccacgttcaa cgg 33

<210> 143
<211> 28
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-8-A

<400> 143
ggagaacgtg gatgaagttg tggggtac 28

<210> 144
<211> 33
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-8-B

<400> 144
gcggtacccc acaacttcat ccacgttctc cgg 33

<210> 145
<211> 28
<212> DNA
<213> 人工序列

<220>
[0186] <223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-9-A

<400> 145
ggcgaacgtg gatgaagttg tggggtac 28

<210> 146
<211> 33
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-9-B

<400> 146
gcggtacccc acaacttcat ccacgttcgc cgg 33

<210> 147
<211> 28
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-10-A

<400> 147
ggggaacgtg gatgaagttg tggggtac 28

<210> 148
<211> 33
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-10-B

<400> 148
gcggtacccc acaacttcat ccacgttccc cgg 33

[0187]

<210> 149
<211> 28
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-11-A

<400> 149
ggtaaacgtg gatgaagttg tggggtac 28

<210> 150
<211> 33
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-11-B

<400> 150
gcggtacccc acaacttcat ccacgtttac cgg 33

<210> 151

<210> 28
<211> DNA
<212> 人工序列
<213> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-12-A

<400> 151
ggtcaacgtg gatgaagttg tggggtac 28

<210> 152
<211> 33
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-12-B

<400> 152
gcggtacccc acaacttcat ccacgttgac cgg 33

[0188]

<210> 153
<211> 28
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-13-A

<400> 153
ggttaacgtg gatgaagttg tggggtac 28

<210> 154
<211> 33
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-13-B

<400> 154

gcggtacccc acaacttcat ccacgttaac cgg 33

<210> 155
<211> 28
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-14-A

<400> 155
ggtgcacgtg gatgaagttg tggggtac 28

<210> 156
<211> 33
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-14-B

[0189]

<400> 156
gcggtacccc acaacttcat ccacgtgcac cgg 33

<210> 157
<211> 28
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-15-A

<400> 157
ggtgtacgtg gatgaagttg tggggtac 28

<210> 158
<211> 33
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>

<220>
<221> misc_feature
<223> R01-15-B

<400> 158
gcggtacccc acaacttcat ccacgtacac cgg 33

<210> 159
<211> 28
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-16-A

<400> 159
ggtggacgtg gatgaagttg tggggtac 28

<210> 160
<211> 33
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

[0190]

<220>
<221> misc_feature
<223> R01-16-B

<400> 160
gcggtacccc acaacttcat ccacgtccac cgg 33

<210> 161
<211> 28
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-17-A

<400> 161
ggtgaccgtg gatgaagttg tggggtac 28

<210> 162
<211> 33
<212> DNA
<213> 人工序列

<220>

<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-17-B

<400> 162
gcggtacccc acaacttcat ccacggtcac cgg 33

<210> 163
<211> 28
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-18-A

<400> 163
ggtgatcgtg gatgaagttg tggggtac 28

<210> 164
<211> 33
<212> DNA
<213> 人工序列

[0191]

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-18-B

<400> 164
gcggtacccc acaacttcat ccacgatcac cgg 33

<210> 165
<211> 28
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-19-A

<400> 165
ggtgagcgtg gatgaagttg tggggtac 28

<210> 166
<211> 33

<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-19-B

<400> 166
gcggtacccc acaacttcat ccacgctcac cgg 33

<210> 167
<211> 28
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-20-A

<400> 167
ggtgaaagtg gatgaagttg tggggtac 28

[0192]

<210> 168
<211> 33
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-20-B

<400> 168
gcggtacccc acaacttcat ccactttcac cgg 33

<210> 169
<211> 28
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-21-A

<400> 169
ggtgaatgtg gatgaagttg tggggtac 28

<210> 170
<211> 33
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-21-B

<400> 170
gcggtacccc acaacttcat ccacattcac cgg 33

<210> 171
<211> 28
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-22-A

[0193]
<400> 171
ggtgaaggtg gatgaagttg tggggtac 28

<210> 172
<211> 33
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-22-B

<400> 172
gcggtacccc acaacttcat ccaccttcac cgg 33

<210> 173
<211> 28
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature

<223> R01-23-A

<400> 173
ggtgaacatg gatgaagttg tggggtac 28

<210> 174
<211> 33
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-23-B

<400> 174
gcggtacccc acaacttcat ccatgttcac cgg 33

<210> 175
<211> 28
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

[0194]

<220>
<221> misc_feature
<223> R01-24-A

<400> 175
ggtgaacctg gatgaagttg tggggtac 28

<210> 176
<211> 33
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-24-B

<400> 176
gcggtacccc acaacttcat ccaggttcac cgg 33

<210> 177
<211> 28
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-25-A

<400> 177
ggtgaacttg gatgaagttg tggggtac 28

<210> 178
<211> 33
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-25-B

<400> 178
gcggtacccc acaacttcat ccaagttcac cgg 33

<210> 179
<211> 28
<212> DNA
<213> 人工序列

[0195] <220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-26-A

<400> 179
ggtgaacgag gatgaagttg tggggtac 28

<210> 180
<211> 33
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-26-B

<400> 180
gcggtacccc acaacttcat cctcgttcac cgg 33

<210> 181
<211> 28
<212> DNA

<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-27-A

<400> 181
ggtgaacgcg gatgaagttg tggggtac 28

<210> 182
<211> 33
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-27-B

<400> 182
gcggtacccc acaacttcat ccgcgttcac cgg 33

[0196] <210> 183
<211> 28
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-28-A

<400> 183
ggtgaacggg gatgaagttg tggggtac 28

<210> 184
<211> 33
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-28-B

<400> 184
gcggtacccc acaacttcat ccccgttcac cgg 33

<210> 185
<211> 28
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-29-A

<400> 185
ggtgaacgta gatgaagttg tggggtac 28

<210> 186
<211> 33
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-29-B

<400> 186
[0197] gcggtacccc acaacttcat ctacgttcac cgg 33

<210> 187
<211> 28
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-30-A

<400> 187
ggtgaacgtc gatgaagttg tggggtac 28

<210> 188
<211> 33
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-30-B

<400> 188
gcggtacccc acaacttcat cgacgttcac cgg 33

<210> 189
<211> 28
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-31-A

<400> 189
ggtgaacgtt gatgaagttg tggggtac 28

<210> 190
<211> 33
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

[0198]

<220>
<221> misc_feature
<223> R01-31-B

<400> 190
gcggtacccc acaacttcat caacgttcac cgg 33

<210> 191
<211> 28
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-32-A

<400> 191
ggtgaacgtg aatgaagttg tggggtac 28

<210> 192
<211> 33
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-32-B

<400> 192
gcggtacccc acaacttcat tcacgttcac cgg 33

<210> 193
<211> 28
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-33-A

<400> 193
ggtgaacgtg catgaagttg tggggtac 28

<210> 194
<211> 33
<212> DNA
<213> 人工序列

<220>
[0199] <223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-33-B

<400> 194
gcggtacccc acaacttcat gcacgttcac cgg 33

<210> 195
<211> 28
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-34-A

<400> 195
ggtgaacgtg tatgaagttg tggggtac 28

<210> 196
<211> 33
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-34-B

<400> 196
gcggtacccc acaacttcat acacgttcac cgg 33

<210> 197
<211> 28
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-35-A

<400> 197
ggtgaacgtg gctgaagttg tggggtac 28

[0200]

<210> 198
<211> 33
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-35-B

<400> 198
gcggtacccc acaacttcag ccacgttcac cgg 33

<210> 199
<211> 28
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-36-A

<400> 199
ggtgaacgtg gttgaagttg tggggtac 28

<210> 200
<211> 33
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-36-B

<400> 200
gcggtacccc acaacttcaa ccacgttcac cgg 33

<210> 201
<211> 28
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-37-A

<400> 201
[0201] ggtgaacgtg ggtgaagttg tggggtac 28

<210> 202
<211> 33
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-37-B

<400> 202
gcggtacccc acaacttcac ccacgttcac cgg 33

<210> 203
<211> 28
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-38-A

<400> 203
ggtgaacgtg gaagaagttg tggggtac 28

<210> 204
<211> 33
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-38-B

<400> 204
gcggtacccc acaacttctt ccacgttcac cgg 33

<210> 205
<211> 28
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

[0202]

<220>
<221> misc_feature
<223> R01-39-A

<400> 205
ggtgaacgtg gacgaagttg tggggtac 28

<210> 206
<211> 33
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-39-B

<400> 206
gcggtacccc acaacttcgt ccacgttcac cgg 33

<210> 207
<211> 28
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-40-A

<400> 207
ggtgaacgtg gaggaagttg tggggtac 28

<210> 208
<211> 33
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-40-B

<400> 208
gcggtacccc acaacttcct ccacgttcac cgg 33

<210> 209
<211> 28
<212> DNA
<213> 人工序列

[0203] <220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-41-A

<400> 209
ggtgaacgtg gataaagttg tggggtac 28

<210> 210
<211> 33
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-41-B

<400> 210
gcggtacccc acaactttat ccacgttcac cgg 33

<210> 211
<211> 28
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-42-A

<400> 211
ggtgaacgtg gatcaagttg tggggtac 28

<210> 212
<211> 33
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-42-B

<400> 212
gcggtacccc acaacttgat ccacgttcac cgg 33

[0204]

<210> 213
<211> 28
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-43-A

<400> 213
ggtgaacgtg gattaagttg tggggtac 28

<210> 214
<211> 33
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-43-B

<400> 214
gcggtacccc acaacttaat ccacgttcac cgg 33

<210> 215

<211> 28
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-44-A

<400> 215
ggtgaacgtg gatgcagttg tggggtac 28

<210> 216
<211> 33
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-44-B

<400> 216
gcggtacccc acaactgcat ccacgttcac cgg 33

[0205]

<210> 217
<211> 28
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-45-A

<400> 217
ggtgaacgtg gatgtagttg tggggtac 28

<210> 218
<211> 33
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-45-B

<400> 218

gcggtacccc acaactacat ccacgttcac cgg 33

<210> 219
<211> 28
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-46-A

<400> 219
ggtgaacgtg gatggagttg tggggtac 28

<210> 220
<211> 33
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-46-B

[0206]

<400> 220
gcggtacccc acaactccat ccacgttcac cgg 33

<210> 221
<211> 28
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-47-A

<400> 221
ggtgaacgtg gatgacgttg tggggtac 28

<210> 222
<211> 33
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>

<220>
<221> misc_feature
<223> R01-47-B

<400> 222
gcggtacccc acaacgtcat ccacgttcac cgg 33

<210> 223
<211> 28
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-48-A

<400> 223
ggtgaacgtg gatgatgttg tggggtac 28

<210> 224
<211> 33
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

[0207]

<220>
<221> misc_feature
<223> R01-48-B

<400> 224
gcggtacccc acaacatcat ccacgttcac cgg 33

<210> 225
<211> 28
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-49-A

<400> 225
ggtgaacgtg gatgaggttg tggggtac 28

<210> 226
<211> 33
<212> DNA
<213> 人工序列

<220>

<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-49-B

<400> 226
gcggtacccc acaacctcat ccacgttcac cgg 33

<210> 227
<211> 27
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-50-A

<400> 227
ggtgaacgtg gatgaaattg tggggta 27

<210> 228
<211> 33
<212> DNA
<213> 人工序列

[0208]

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-50-B

<400> 228
gcggtacccc acaatttcat ccacgttcac cgg 33

<210> 229
<211> 28
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-51-A

<400> 229
ggtgaacgtg gatgaacttg tggggtac 28

<210> 230
<211> 33

<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-51-B

<400> 230
gcggtacccc acaagttcat ccacgttcac cgg 33

<210> 231
<211> 28
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-52-A

<400> 231
ggtgaacgtg gatgaatttg tggggtac 28

[0209]

<210> 232
<211> 33
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-52-B

<400> 232
gcggtacccc acaaattcat ccacgttcac cgg 33

<210> 233
<211> 28
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-53-A

<400> 233
ggtgaacgtg gatgaagatg tggggtac 28

<210> 234
<211> 33
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-53-B

<400> 234
gcggtacccc acatcttcat ccacgttcac cgg 33

<210> 235
<211> 28
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-54-A

[0210]
<400> 235
ggtgaacgtg gatgaagctg tggggtac 28

<210> 236
<211> 33
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-54-B

<400> 236
gcggtacccc acagcttcat ccacgttcac cgg 33

<210> 237
<211> 28
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature

<223> R01-55-A

<400> 237
ggtgaacgtg gatgaaggtg tggggtac 28

<210> 238
<211> 33
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-55-B

<400> 238
gcggtacccc acaccttcat ccacgttcac cgg 33

<210> 239
<211> 28
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

[0211]

<220>
<221> misc_feature
<223> R01-56-A

<400> 239
ggtgaacgtg gatgaagtag tggggtac 28

<210> 240
<211> 33
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-56-B

<400> 240
gcggtacccc actacttcat ccacgttcac cgg 33

<210> 241
<211> 28
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-57-A

<400> 241
ggtgaacgtg gatgaagtcg tggggtac 28

<210> 242
<211> 33
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-57-B

<400> 242
gcggtacccc acgacttcat ccacgttcac cgg 33

<210> 243
<211> 28
<212> DNA
<213> 人工序列

[0212]
<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-58-A

<400> 243
ggtgaacgtg gatgaagtgg tggggtac 28

<210> 244
<211> 33
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-58-B

<400> 244
gcggtacccc accacttcat ccacgttcac cgg 33

<210> 245
<211> 28
<212> DNA

<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-59-A

<400> 245
ggtgaacgtg gatgaagtta tggggtac 28

<210> 246
<211> 33
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-59-B

<400> 246
gcggtacccc ataacttcat ccacgttcac cgg 33

[0213] <210> 247
<211> 28
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-60-A

<400> 247
ggtgaacgtg gatgaagttc tggggtac 28

<210> 248
<211> 33
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-60-B

<400> 248
gcggtacccc agaacttcat ccacgttcac cgg 33

<210> 249
<211> 28
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-61-A

<400> 249
ggtgaacgtg gatgaagttt tggggtac 28

<210> 250
<211> 33
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> R01-61-B

<400> 250
[0214] gcggtacccc aaaacttcat ccacgttcac cgg 33

<210> 251
<211> 21
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 1靶序列

<400> 251
ggagtccgag cagaagaaga a 21

<210> 252
<211> 21
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 2靶序列

<400> 252
ggacatcgat gtcctcccca t 21

<210> 253
<211> 21
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 3靶序列

<400> 253
ggtcacctcc aatgactagg g 21

<210> 254
<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

[0215]

<220>
<221> misc_feature
<223> sgRNA 4靶序列

<400> 254
ggctccccat tggcctgctt cg 22

<210> 255
<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 5靶序列

<400> 255
ggtgcgccac cggttgatgt ga 22

<210> 256
<211> 21
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 6靶序列

<400> 256
ggcgccaccg gttgatgtga t 21

<210> 257
<211> 21
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 7靶序列

<400> 257
ggccgtttgt actttgtcct c 21

<210> 258
<211> 21
<212> DNA
<213> 人工序列

<220>
[0216] <223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 8靶序列

<400> 258
ggcacagatg agaaactcag g 21

<210> 259
<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 9靶序列

<400> 259
ggattgggtg ttcagggcag ag 22

<210> 260
<211> 21
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 10靶序列

<400> 260
ggtggcgaga ggggccgaga t 21

<210> 261
<211> 20
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 11靶序列

<400> 261
ggggccgaga ttgggtgttc 20

[0217]

<210> 262
<211> 21
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 12靶序列

<400> 262
ggtgccatta gctaaatgca t 21

<210> 263
<211> 21
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 13靶序列

<400> 263
ggtaccaccc acaggtgcca g 21

<210> 264
<211> 21
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 14靶序列

<400> 264
ggaaagcctc tgggccagga a 21

<210> 265
<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 15靶序列

<400> 265
[0218] ggacacccac ggagagcgct gt 22

<210> 266
<211> 21
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 16靶序列

<400> 266
ggtccagttg gtgacaaata c 21

<210> 267
<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 17靶序列

<400> 267
ggttttaaca cttccctagc ca 22

<210> 268
<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 18靶序列

<400> 268
ggcttaatta gctcctttgg ct 22

<210> 269
<211> 21
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

[0219]

<220>
<221> misc_feature
<223> sgRNA 19靶序列

<400> 269
ggcttaatta gctcctttgg c 21

<210> 270
<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 20靶序列

<400> 270
ggtgcatatg atgatagtat ta 22

<210> 271
<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 21靶序列

<400> 271
ggaatgtatg ctggctttta ag 22

<210> 272
<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 22靶序列

<400> 272
ggaaatgtat gctggctttt aa 22

<210> 273
<211> 22
<212> DNA
<213> 人工序列

<220>
[0220] <223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 23靶序列

<400> 273
ggaaaatgta tgctggcttt ta 22

<210> 274
<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 24靶序列

<400> 274
ggtaaaatgt atgctggctt tt 22

<210> 275
<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 25靶序列

<400> 275
ggcttaaaga tgatctctta cc 22

<210> 276
<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 26靶序列

<400> 276
ggttaaagat gatctcttac ct 22

[0221]

<210> 277
<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 27靶序列

<400> 277
ggtaataaaa ttcaaacatc ct 22

<210> 278
<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 28靶序列

<400> 278
ggtaaagcat agtgcaatgg at 22

<210> 279

<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 29靶序列

<400> 279
ggttaaataa agcatagtgc aa 22

<210> 280
<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 30靶序列

<400> 280
ggatttatga gatcaacagc ac 22

[0222]

<210> 281
<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 31靶序列

<400> 281
ggcccatgct ctgaccgctc ga 22

<210> 282
<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 32靶序列

<400> 282

ggccatcgag cggtcagagc at 22

<210> 283
<211> 20
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 33靶序列

<400> 283
gggcaaaagc tcatgtgata 20

<210> 284
<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 34靶序列

[0223]

<400> 284
ggactgaggc ttatgttcca tg 22

<210> 285
<211> 21
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 35靶序列

<400> 285
ggagagtgta caaactcaca a 21

<210> 286
<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>

<220>
<221> misc_feature
<223> sgRNA 36靶序列

<400> 286
ggactagcat tataatgcac ca 22

<210> 287
<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 37靶序列

<400> 287
ggtcagaaga gattagttag ta 22

<210> 288
<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

[0224]

<220>
<221> misc_feature
<223> sgRNA 38靶序列

<400> 288
ggtgagagtg ccatctcttc ct 22

<210> 289
<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 39靶序列

<400> 289
ggtggaccac atggctttgc tc 22

<210> 290
<211> 22
<212> DNA
<213> 人工序列

<220>

<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 40靶序列

<400> 290
ggtgagcaaa gccatgtggt cc 22

<210> 291
<211> 21
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 41靶序列

<400> 291
ggattaatga catacgcatt t 21

<210> 292
<211> 22
<212> DNA
<213> 人工序列

[0225]

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 42靶序列

<400> 292
ggagattaat gacatacgca tt 22

<210> 293
<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 43靶序列

<400> 293
ggttcaatcc tcttgtcacc tg 22

<210> 294
<211> 21

<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 44靶序列

<400> 294
ggacatctct atgtcggcca c 21

<210> 295
<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 45靶序列

<400> 295
ggaaaaccta ccgcaagttg cc 22

[0226]

<210> 296
<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 46靶序列

<400> 296
ggtggactac atagttgtgt ga 22

<210> 297
<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 47靶序列

<400> 297
ggctgccaga ctctctgaac cc 22

<210> 298
<211> 20
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 48靶序列

<400> 298
ggatccctcg tataacaata 20

<210> 299
<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 49靶序列

[0227]
<400> 299
ggtatcgcat agtgcaaaga at 22

<210> 300
<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 50靶序列

<400> 300
ggctcggtcc tctcgtcacc tg 22

<210> 301
<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature

<223> sgRNA 51靶序列

<400> 301
ggacgctccc ggagagcgct gg 22

<210> 302
<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 52靶序列

<400> 302
ggccatggcg cggtcagagg gt 22

<210> 303
<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

[0228]

<220>
<221> misc_feature
<223> sgRNA 53靶序列

<400> 303
ggctcggctc caatgacgag gg 22

<210> 304
<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 54靶序列

<400> 304
ggcacttcga tgacctcccc gt 22

<210> 305
<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 55靶序列

<400> 305
ggagtgggtg ttcaggaccg cg 22

<210> 306
<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 56靶序列

<400> 306
ggctgctacc cacaggtgca ag 22

<210> 307
<211> 22
<212> DNA
<213> 人工序列

[0229]
<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 57靶序列

<400> 307
ggaagatgta tcctggaatt ta 22

<210> 308
<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 58靶序列

<400> 308
ggttttatca ctacccaaga ca 22

<210> 309
<211> 22
<212> DNA

<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 59靶序列

<400> 309
ggcggccagg ctccctgaaa cc 22

<210> 310
<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 60靶序列

<400> 310
ggactggtat tataatacaa ca 22

[0230] <210> 311
<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 61靶序列

<400> 311
ggtgagagtg acatttcgta ct 22

<210> 312
<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 62靶序列

<400> 312
ggtgcgcccc cgggtgaggt gc 22

<210> 313
<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 63靶序列

<400> 313
ggcagtttga gcagaaaaag aa 22

<210> 314
<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 64靶序列

<400> 314
[0231] ggtctgtaga gataagtgag ta 22

<210> 315
<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 65靶序列

<400> 315
ggaagtgtgt acaaactcaa ga 22

<210> 316
<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 66靶序列

<400> 316
ggtatcatgg atgctgactt tt 22

<210> 317
<211> 21
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 67靶序列

<400> 317
ggctttacta gcccctgtgg c 21

<210> 318
<211> 21
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

[0232]

<220>
<221> misc_feature
<223> sgRNA 68靶序列

<400> 318
ggcacggagg agaaaagcag g 21

<210> 319
<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 69靶序列

<400> 319
ggatttgtga gatctaaaga ac 22

<210> 320
<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 70靶序列

<400> 320
ggcgttgatg acattcgcat tt 22

<210> 321
<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 71靶序列

<400> 321
ggcttataga tcatctaata cc 22

<210> 322
<211> 22
<212> DNA
<213> 人工序列

<220>
[0233] <223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 72靶序列

<400> 322
ggtggcctcc atagttgggt gg 22

<210> 323
<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 73靶序列

<400> 323
ggtactagag ttcaaacata ct 22

<210> 324
<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 74靶序列

<400> 324
ggcccctgct ctgagcactc gg 22

<210> 325
<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 75靶序列

<400> 325
ggaggaccac atggcctgcc tc 22

[0234]

<210> 326
<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 76靶序列

<400> 326
ggaacagctg ccgcaaattg cc 22

<210> 327
<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 77靶序列

<400> 327
ggctccgtat tggctggctt cg 22

<210> 328
<211> 20
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 78靶序列

<400> 328
ggccagccga tgacaaatac 20

<210> 329
<211> 21
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 79靶序列

<400> 329
[0235] ggacgtctct acgtcggcgg c 21

<210> 330
<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 80靶序列

<400> 330
ggccggcgcg aggggccgag ag 22

<210> 331
<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 81靶序列

<400> 331
ggactgtttg tacttttttca tc 22

<210> 332
<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 82靶序列

<400> 332
ggtgcttatg atgatgacat ta 22

<210> 333
<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

[0236]

<220>
<221> misc_feature
<223> sgRNA 83靶序列

<400> 333
ggccgcgtcc gattgatgtg at 22

<210> 334
<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 84靶序列

<400> 334
ggctttatga gcaccttagg ct 22

<210> 335
<211> 20
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 85靶序列

<400> 335
ggatccctcc tatgaccatg 20

<210> 336
<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 86靶序列

<400> 336
ggataaaggt gacctcttac gt 22

<210> 337
<211> 22
<212> DNA
<213> 人工序列

<220>
[0237] <223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 87靶序列

<400> 337
ggactgtggc ttagggtcca gg 22

<210> 338
<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 88靶序列

<400> 338
ggtgaggaaa cccaggcggt cc 22

<210> 339
<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 89靶序列

<400> 339
ggaaatgtat acaggcaatt aa 22

<210> 340
<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 90靶序列

<400> 340
ggttgaatca agcatagtcc ga 22

[0238]

<210> 341
<211> 20
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 91靶序列

<400> 341
gggctacggc tcatgagata 20

<210> 342
<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 92靶序列

<400> 342
ggcgggccgc gattggcagt tc 22

<210> 343

<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 93靶序列

<400> 343
ggctgcgatt atctaaatcc at 22

<210> 344
<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 94靶序列

<400> 344
ggagcttgat tacatacgca tg 22

[0239]

<210> 345
<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 95靶序列

<400> 345
ggaatgtttg ctggcctcta gg 22

<210> 346
<211> 21
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> sgRNA 96靶序列

<400> 346

ggaaagcctc tgcgccgggg g 21

<210> 347
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-1-A

<400> 347
cctgagtccg agcagaagaa gaatggggta c 31

<210> 348
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-1-B

[0240]

<400> 348
gcggtacccc attcttcttc tgctcggact cagggg 36

<210> 349
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-2-A

<400> 349
ggtgacatcg atgtcctccc cattggggta c 31

<210> 350
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>

<220>
<221> misc_feature
<223> Ext-A-P-2-B

<400> 350
gcggtacccc aatggggagg acatcgatgt caccgg 36

<210> 351
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-3-A

<400> 351
gatgtcacct ccaatgacta gggtggggta c 31

<210> 352
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

[0241]

<220>
<221> misc_feature
<223> Ext-A-P-3-B

<400> 352
gcggtacccc accctagtca ttggaggtga catcgg 36

<210> 353
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-4-A

<400> 353
gtcctcccca ttggcctgct tcgtggggta c 31

<210> 354
<211> 36
<212> DNA
<213> 人工序列

<220>

<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-4-B

<400> 354
gcggtacccc acgaagcagg ccaatgggga ggacgg 36

<210> 355
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-5-A

<400> 355
caatgcgcca ccggttgatg tgatggggta c 31

<210> 356
<211> 36
<212> DNA
<213> 人工序列

[0242]

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-5-B

<400> 356
gcggtacccc atcacatcaa ccggtggcgc attggg 36

<210> 357
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-6-A

<400> 357
aatgcgccac cggttgatgt gattggggta c 31

<210> 358
<211> 36

<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-6-B

<400> 358
gcggtacccc aatcacatca accggtggcg cattgg 36

<210> 359
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-7-A

<400> 359
tctgccgttt gtactttgtc ctctggggta c 31

[0243]

<210> 360
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-7-B

<400> 360
gcggtacccc agaggacaaa gtacaaacgg cagagg 36

<210> 361
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-8-A

<400> 361
ggggcacaga tgagaaactc aggtggggta c 31

<210> 362
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-8-B

<400> 362
gcggtacccc acctgagttt ctcatctgtg ccccgg 36

<210> 363
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-9-A

[0244]
<400> 363
gagattgggt gttcagggca gagtggggta c 31

<210> 364
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-9-B

<400> 364
gcggtacccc actctgccct gaacacccaa tctcgg 36

<210> 365
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature

<223> Ext-A-P-10-A

<400> 365
agggtggcga gaggggccga gattggggta c 31

<210> 366
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-10-B

<400> 366
gcggtacccc aatctcggcc cctctcgcca ccctgg 36

<210> 367
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

[0245]

<220>
<221> misc_feature
<223> Ext-A-P-11-A

<400> 367
agaggggccg agattgggtg ttctggggta c 31

<210> 368
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-11-B

<400> 368
gcggtacccc agaacaccca atctcggccc ctctgg 36

<210> 369
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-12-A

<400> 369
tgagtgccat tagctaaatg cattggggta c 31

<210> 370
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-12-B

<400> 370
gcggtacccc aatgcattta gctaatggca ctcagg 36

<210> 371
<211> 31
<212> DNA
<213> 人工序列

[0246]
<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-13-A

<400> 371
agggtaccac ccacaggtgc cagtggggta c 31

<210> 372
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-13-B

<400> 372
gcggtacccc actggcacct gtgggtggta ccctgg 36

<210> 373
<211> 31
<212> DNA

<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-14-A

<400> 373
tgggaaagcc tctgggccag gaatggggta c 31

<210> 374
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-14-B

<400> 374
gcggtacccc attcctggcc cagaggcttt cccagg 36

[0247] <210> 375
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-27-A

<400> 375
aggacaccca cggagagcgc tgttggggta c 31

<210> 376
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-27-B

<400> 376
gcggtacccc aacagcgctc tccgtgggtg tcctgg 36

<210> 377
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-28-A

<400> 377
agggtccagt tggtgacaaa tactggggta c 31

<210> 378
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-28-B

<400> 378
[0248] gcggtacccc agtatttgtc accaactgga ccctgg 36

<210> 379
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-29-A

<400> 379
aggttttaac acttccctag ccatggggta c 31

<210> 380
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-29-B

<400> 380
gcggtacccc atggctaggg aagtgttaaa acctgg 36

<210> 381
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-30-A

<400> 381
aggcttaatt agctcctttg gcttggggta c 31

<210> 382
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

[0249]

<220>
<221> misc_feature
<223> Ext-A-P-30-B

<400> 382
gcggtacccc aagccaaagg agctaattaa gcctgg 36

<210> 383
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-31-A

<400> 383
agggcttaat tagctccttt ggctggggta c 31

<210> 384
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-31-B

<400> 384
gcggtacccc agccaaagga gctaattaag ccctgg 36

<210> 385
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-32-A

<400> 385
aggtgcatat gatgatagta ttatggggta c 31

<210> 386
<211> 36
<212> DNA
<213> 人工序列

<220>
[0250] <223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-32-B

<400> 386
gcggtacccc ataatactat catcatatgc acctgg 36

<210> 387
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-33-A

<400> 387
aggaatgtat gctggctttt aagtggggta c 31

<210> 388
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-33-B

<400> 388
gcggtacccc acttaaaagc cagcatacat tcctgg 36

<210> 389
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-34-A

<400> 389
aggaaatgta tgctggcttt taatggggta c 31

[0251]

<210> 390
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-34-B

<400> 390
gcggtacccc attaaaagcc agcatacatt tcctgg 36

<210> 391
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-35-A

<400> 391
aggaaaatgt atgctggctt ttatggggta c 31

<210> 392
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-35-B

<400> 392
gcggtacccc ataaaagcca gcatacattt tcctgg 36

<210> 393
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-36-A

<400> 393
[0252] aggtaaaatg tatgctggct ttttggggta c 31

<210> 394
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-36-B

<400> 394
gcggtacccc aaaaagccag catacatttt acctgg 36

<210> 395
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-37-A

<400> 395
aggcttaaag atgatctctt acctggggta c 31

<210> 396
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-37-B

<400> 396
gcggtacccc aggtaagaga tcatctttaa gcctgg 36

<210> 397
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

[0253]

<220>
<221> misc_feature
<223> Ext-A-P-38-A

<400> 397
aggttaaaga tgatctctta ccttggggta c 31

<210> 398
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-38-B

<400> 398
gcggtacccc aaggtaagag atcatcttta acctgg 36

<210> 399
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-39-A

<400> 399
aggtaataaa attcaaacat ccttggggta c 31

<210> 400
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-39-B

<400> 400
gcggtacccc aaggatgttt gaattttatt acctgg 36

<210> 401
<211> 31
<212> DNA
<213> 人工序列

<220>
[0254] <223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-40-A

<400> 401
aggtaaagca tagtgcaatg gattggggta c 31

<210> 402
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-40-B

<400> 402
gcggtacccc aatccattgc actatgcttt acctgg 36

<210> 403
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-41-A

<400> 403
aggttaaata aagcatagtg caatggggta c 31

<210> 404
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-41-B

<400> 404
gcggtacccc attgcactat gctttattta acctgg 36

[0255]

<210> 405
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-42-A

<400> 405
aggatttatg agatcaacag cactggggta c 31

<210> 406
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-42-B

<400> 406
gcggtacccc agtgctgttg atctcataaa tcctgg 36

<210> 407

<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-43-A

<400> 407
aggcccatgc tctgaccgct cgatggggta c 31

<210> 408
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-43-B

<400> 408
gcggtacccc atcgagcggt cagagcatgg gcctgg 36

[0256]

<210> 409
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-44-A

<400> 409
aggccatcga gcggtcagag cattggggta c 31

<210> 410
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-44-B

<400> 410

gcggtacccc aatgctctga ccgctcgatg gcctgg 36

<210> 411
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-45-A

<400> 411
aggggggcaaa agctcatgtg atatggggta c 31

<210> 412
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-45-B

[0257]

<400> 412
gcggtacccc atatcacatg agcttttgcc ccctgg 36

<210> 413
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-46-A

<400> 413
aggactgagg cttatgttcc atgtggggta c 31

<210> 414
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>

<220>
<221> misc_feature
<223> Ext-A-P-46-B

<400> 414
gcggtacccc acatggaaca taagcctcag tcctgg 36

<210> 415
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-47-A

<400> 415
agggagagtg tacaaactca caatggggta c 31

<210> 416
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

[0258]

<220>
<221> misc_feature
<223> Ext-A-P-47-B

<400> 416
gcggtacccc attgtgagtt tgtacactct ccctgg 36

<210> 417
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-48-A

<400> 417
aggactagca ttataatgca ccatggggta c 31

<210> 418
<211> 36
<212> DNA
<213> 人工序列

<220>

<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-48-B

<400> 418
gcggtacccc atggtgcatt ataatgctag tcctgg 36

<210> 419
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-49-A

<400> 419
aggtcagaag agattagtta gtatggggta c 31

<210> 420
<211> 36
<212> DNA
<213> 人工序列

[0259]

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-49-B

<400> 420
gcggtacccc atactaacta atctcttctg acctgg 36

<210> 421
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-50-A

<400> 421
aggtgagagt gccatctctt ccttggggta c 31

<210> 422
<211> 36

<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-50-B

<400> 422
gcggtacccc aaggaagaga tggcactctc acctgg 36

<210> 423
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-51-A

<400> 423
aggtggacca catggctttg ctctggggta c 31

[0260]

<210> 424
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-51-B

<400> 424
gcggtacccc agagcaaagc catgtggtcc acctgg 36

<210> 425
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-52-A

<400> 425
aggtgagcaa agccatgtgg tcctggggta c 31

<210> 426
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-52-B

<400> 426
gcggtacccc aggaccacat ggctttgctc acctgg 36

<210> 427
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-53-A

[0261]
<400> 427
agggattaat gacatacgca ttttggggta c 31

<210> 428
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-53-B

<400> 428
gcggtacccc aaaatgcgta tgtcattaat ccctgg 36

<210> 429
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature

<223> Ext-A-P-54-A

<400> 429
aggagattaa tgacatacgc atttggggta c 31

<210> 430
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-54-B

<400> 430
gcggtacccc aaatgcgtat gtcattaatc tcctgg 36

<210> 431
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

[0262]

<220>
<221> misc_feature
<223> Ext-A-P-55-A

<400> 431
aggttcaatc ctcttgtcac ctgtggggta c 31

<210> 432
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-55-B

<400> 432
gcggtacccc acaggtgaca agaggattga acctgg 36

<210> 433
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-56-A

<400> 433
agggacatct ctatgtcggc cactggggta c 31

<210> 434
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-56-B

<400> 434
gcggtacccc agtggccgac atagagatgt ccctgg 36

<210> 435
<211> 31
<212> DNA
<213> 人工序列

[0263]
<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-57-A

<400> 435
aggaaaacct accgcaagtt gcctggggta c 31

<210> 436
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-57-B

<400> 436
gcggtacccc aggcaacttg cggtaggttt tcctgg 36

<210> 437
<211> 31
<212> DNA

<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-58-A

<400> 437
aggtggacta catagttgtg tgatggggta c 31

<210> 438
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-58-B

<400> 438
gcggtacccc atcacacaac tatgtagtcc acctgg 36

[0264] <210> 439
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-59-A

<400> 439
aggctgccag actctctgaa ccctggggta c 31

<210> 440
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-59-B

<400> 440
gcggtacccc agggttcaga gagtctggca gcctgg 36

<210> 441
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-60-A

<400> 441
aggggatccc tcgtataaca atatggggta c 31

<210> 442
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-60-B

<400> 442
[0265] gcggtacccc atattgttat acgagggatc ccctgg 36

<210> 443
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-61-A

<400> 443
aggtatcgca tagtgcaaag aattggggta c 31

<210> 444
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-61-B

<400> 444
gcggtacccc aattctttgc actatgcgat acctgg 36

<210> 445
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-62-A

<400> 445
aggctcggtc ctctcgtcac ctgtggggta c 31

<210> 446
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

[0266]

<220>
<221> misc_feature
<223> Ext-A-P-62-B

<400> 446
gcggtacccc acaggtgacg agaggaccga gcctgg 36

<210> 447
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-63-A

<400> 447
aggacgctcc cggagagcgc tggtggggta c 31

<210> 448
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-63-B

<400> 448
gcggtacccc accagcgctc tccgggagcg tcctgg 36

<210> 449
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-64-A

<400> 449
aggccatggc gcggtcagag ggttggggta c 31

<210> 450
<211> 36
<212> DNA
<213> 人工序列

<220>
[0267] <223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-64-B

<400> 450
gcggtacccc aaccctctga ccgcgccatg gcctgg 36

<210> 451
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-65-A

<400> 451
aggctcggct ccaatgacga gggtggggta c 31

<210> 452
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-65-B

<400> 452
gcggtacccc accctcgtca ttggagccga gcctgg 36

<210> 453
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-66-A

<400> 453
aggcacttcg atgacctccc cgttggggta c 31

[0268]

<210> 454
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-66-B

<400> 454
gcggtacccc aacggggagg tcatcgaagt gcctgg 36

<210> 455
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-67-A

<400> 455
aggagtgggt gttcaggacc gcgtggggta c 31

<210> 456
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-67-B

<400> 456
gcggtacccc acgcggtcct gaacacccac tcctgg 36

<210> 457
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-68-A

<400> 457
[0269] aggctgctac ccacaggtgc aagtggggta c 31

<210> 458
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-68-B

<400> 458
gcggtacccc acttgcacct gtgggtagca gcctgg 36

<210> 459
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-69-A

<400> 459
aggaagatgt atcctggaat ttatggggta c 31

<210> 460
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-69-B

<400> 460
gcggtacccc ataaattcca ggatacatct tcctgg 36

<210> 461
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

[0270]

<220>
<221> misc_feature
<223> Ext-A-P-70-A

<400> 461
aggttttatc actacccaag acatggggta c 31

<210> 462
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-70-B

<400> 462
gcggtacccc atgtcttggg tagtgataaa acctgg 36

<210> 463
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-71-A

<400> 463
aggcggccag gctccctgaa acctggggta c 31

<210> 464
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-71-B

<400> 464
gcggtacccc aggtttcagg gagcctggcc gcctgg 36

<210> 465
<211> 31
<212> DNA
<213> 人工序列

<220>
[0271] <223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-72-A

<400> 465
aggactggta ttataataca acatggggta c 31

<210> 466
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-72-B

<400> 466
gcggtacccc atgttgtatt ataataccag tcctgg 36

<210> 467
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-73-A

<400> 467
aggtgagagt gacatttcgt acttggggta c 31

<210> 468
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-73-B

<400> 468
gcggtacccc aagtacgaaa tgtcactctc acctgg 36

[0272]

<210> 469
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-74-A

<400> 469
aggtgcgccc ccgggtgagg tgctggggta c 31

<210> 470
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-74-B

<400> 470
gcggtacccc agcacctcac ccgggggcgc acctgg 36

<210> 471

<210> 31
<211> DNA
<212> 人工序列
<213> 合成的多核苷酸

<220>
<223> misc_feature
Ext-A-P-75-A

<400> 471
aggcagtttg agcagaaaaa gaatggggta c 31

<210> 472
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<223> misc_feature
Ext-A-P-75-B

<400> 472
gcggtacccc attctttttc tgctcaaact gcctgg 36

[0273]

<210> 473
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<223> misc_feature
Ext-A-P-76-A

<400> 473
aggtctgtag agataagtga gtatggggta c 31

<210> 474
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<223> misc_feature
Ext-A-P-76-B

<400> 474

gcggtacccc atactcactt atctctacag acctgg 36

<210> 475
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-77-A

<400> 475
aggaagtgtg tacaaactca agatggggta c 31

<210> 476
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-77-B

[0274]

<400> 476
gcggtacccc atcttgagtt tgtacacact tcctgg 36

<210> 477
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-78-A

<400> 477
aggtatcatg gatgctgact ttttggggta c 31

<210> 478
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>

<220>
<221> misc_feature
<223> Ext-A-P-78-B

<400> 478
gcggtacccc aaaaagtcag catccatgat acctgg 36

<210> 479
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-79-A

<400> 479
agggctttac tagcccctgt ggctggggta c 31

<210> 480
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

[0275]

<220>
<221> misc_feature
<223> Ext-A-P-79-B

<400> 480
gcggtacccc agccacaggg gctagtaaag ccctgg 36

<210> 481
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-80-A

<400> 481
agggcacgga ggagaaaagc aggtggggta c 31

<210> 482
<211> 36
<212> DNA
<213> 人工序列

<220>

<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-80-B

<400> 482
gcggtacccc acctgctttt ctcctccgtg ccctgg 36

<210> 483
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-81-A

<400> 483
aggatttgtg agatctaaag aactggggta c 31

<210> 484
<211> 36
<212> DNA
<213> 人工序列

[0276]

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-81-B

<400> 484
gcggtacccc agttctttag atctcacaaa tcctgg 36

<210> 485
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-82-A

<400> 485
aggcgttgat gacattcgca ttttggggta c 31

<210> 486
<211> 36

<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-82-B

<400> 486
gcggtacccc aaaatgcgaa tgtcatcaac gcctgg 36

<210> 487
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-83-A

<400> 487
aggcttatag atcatctaat acctggggta c 31

[0277]

<210> 488
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-83-B

<400> 488
gcggtacccc aggtattaga tgatctataa gcctgg 36

<210> 489
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-84-A

<400> 489
aggtggcctc catagttggg tggtggggta c 31

<210> 490
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-84-B

<400> 490
gcggtacccc accacccaac tatggaggcc acctgg 36

<210> 491
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-85-A

[0278]
<400> 491
aggtactaga gttcaaacat acttggggta c 31

<210> 492
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-85-B

<400> 492
gcggtacccc aagtatgttt gaactctagt acctgg 36

<210> 493
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature

<223> Ext-A-P-86-A

<400> 493
aggcccctgc tctgagcact cggtggggta c 31

<210> 494
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-86-B

<400> 494
gcggtacccc accgagtgct cagagcaggg gcctgg 36

<210> 495
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

[0279]

<220>
<221> misc_feature
<223> Ext-A-P-87-A

<400> 495
aggaggacca catggcctgc ctctggggta c 31

<210> 496
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-87-B

<400> 496
gcggtacccc agaggcaggc catgtggtcc tcctgg 36

<210> 497
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-88-A

<400> 497
aggaacagct gccgcaaatt gcctggggta c 31

<210> 498
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-88-B

<400> 498
gcggtacccc aggcaatttg cggcagctgt tcctgg 36

<210> 499
<211> 31
<212> DNA
<213> 人工序列

[0280] <220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-89-A

<400> 499
aggctccgta ttggctggct tcgtggggta c 31

<210> 500
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-89-B

<400> 500
gcggtacccc acgaagccag ccaatacgga gcctgg 36

<210> 501
<211> 31
<212> DNA

<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-90-A

<400> 501
aggggccagc cgatgacaaa tactggggta c 31

<210> 502
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-90-B

<400> 502
gcggtacccc agtatttgtc atcggctggc ccctgg 36

[0281] <210> 503
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-91-A

<400> 503
agggacgtct ctacgtcggc ggctggggta c 31

<210> 504
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-91-B

<400> 504
gcggtacccc agccgccgac gtagagacgt ccctgg 36

<210> 505
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-92-A

<400> 505
aggccggcgc gaggggccga gagtggggta c 31

<210> 506
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-92-B

<400> 506
[0282] gcggtacccc actctcggcc cctcgcgccg gcctgg 36

<210> 507
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-93-A

<400> 507
aggactgttt gtacttttc atctggggta c 31

<210> 508
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-93-B

<400> 508
gcggtacccc agatgaaaaa gtacaaacag tcctgg 36

<210> 509
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-94-A

<400> 509
aggtgcttat gatgatgaca ttatggggta c 31

<210> 510
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

[0283]

<220>
<221> misc_feature
<223> Ext-A-P-94-B

<400> 510
gcggtacccc ataatgtcat catcataagc acctgg 36

<210> 511
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-95-A

<400> 511
aggccgcgtc cgattgatgt gattggggta c 31

<210> 512
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-95-B

<400> 512
gcggtacccc aatcacatca atcggacgcg gcctgg 36

<210> 513
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-96-A

<400> 513
aggctttatg agcaccttag gcttggggta c 31

<210> 514
<211> 36
<212> DNA
<213> 人工序列

<220>
[0284] <223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-96-B

<400> 514
gcggtacccc aagcctaagg tgctcataaa gcctgg 36

<210> 515
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-97-A

<400> 515
aggggatccc tcctatgacc atgtggggta c 31

<210> 516
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-97-B

<400> 516
gcggtacccc acatggtcat aggagggatc ccctgg 36

<210> 517
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-98-A

<400> 517
aggataaagg tgacctctta cgttggggta c 31

[0285]

<210> 518
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-98-B

<400> 518
gcggtacccc aacgtaagag gtcaccttta tcctgg 36

<210> 519
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-99-A

<400> 519
aggactgtgg cttagggtcc aggtggggta c 31

<210> 520
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-99-B

<400> 520
gcggtacccc acctggaccc taagccacag tcctgg 36

<210> 521
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-100-A

<400> 521
[0286] aggtgaggaa acccaggcgg tcctggggta c 31

<210> 522
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-100-B

<400> 522
gcggtacccc aggaccgcct gggtttcctc acctgg 36

<210> 523
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-101-A

<400> 523
aggaaatgta tacaggcaat taatggggta c 31

<210> 524
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-101-B

<400> 524
gcggtacccc attaattgcc tgtatacatt tcctgg 36

<210> 525
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

[0287]

<220>
<221> misc_feature
<223> Ext-A-P-102-A

<400> 525
aggttgaatc aagcatagtc cgatggggta c 31

<210> 526
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-102-B

<400> 526
gcggtacccc atcggactat gcttgattca acctgg 36

<210> 527
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-103-A

<400> 527
aggggctac ggctcatgag atatggggta c 31

<210> 528
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-103-B

<400> 528
gcggtacccc atatctcatg agccgtagcc ccctgg 36

<210> 529
<211> 31
<212> DNA
<213> 人工序列

<220>
[0288] <223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-104-A

<400> 529
aggcgggccg cgattggcag ttctggggta c 31

<210> 530
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-104-B

<400> 530
gcggtacccc agaactgcca atcgcggccc gcctgg 36

<210> 531
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-105-A

<400> 531
aggctgcgat tatctaaatc cattggggta c 31

<210> 532
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-105-B

<400> 532
gcggtacccc aatggattta gataatcgca gcctgg 36

[0289]

<210> 533
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-106-A

<400> 533
aggagcttga ttacatacgc atgtggggta c 31

<210> 534
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-106-B

<400> 534
gcggtacccc acatgcgtat gtaatcaagc tcctgg 36

<210> 535

<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-107-A

<400> 535
aggaatgttt gctggcctct aggtggggta c 31

<210> 536
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-107-B

<400> 536
gcggtacccc acctagaggc cagcaaacat tcctgg 36

[0290]

<210> 537
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-108-A

<400> 537
agggaaagcc tctgcgccgg gggtggggta c 31

<210> 538
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成的多核苷酸

<220>
<221> misc_feature
<223> Ext-A-P-108-B

<400> 538

[0291] gcggtacccc acccccggcg cagaggcttt ccctgg 36

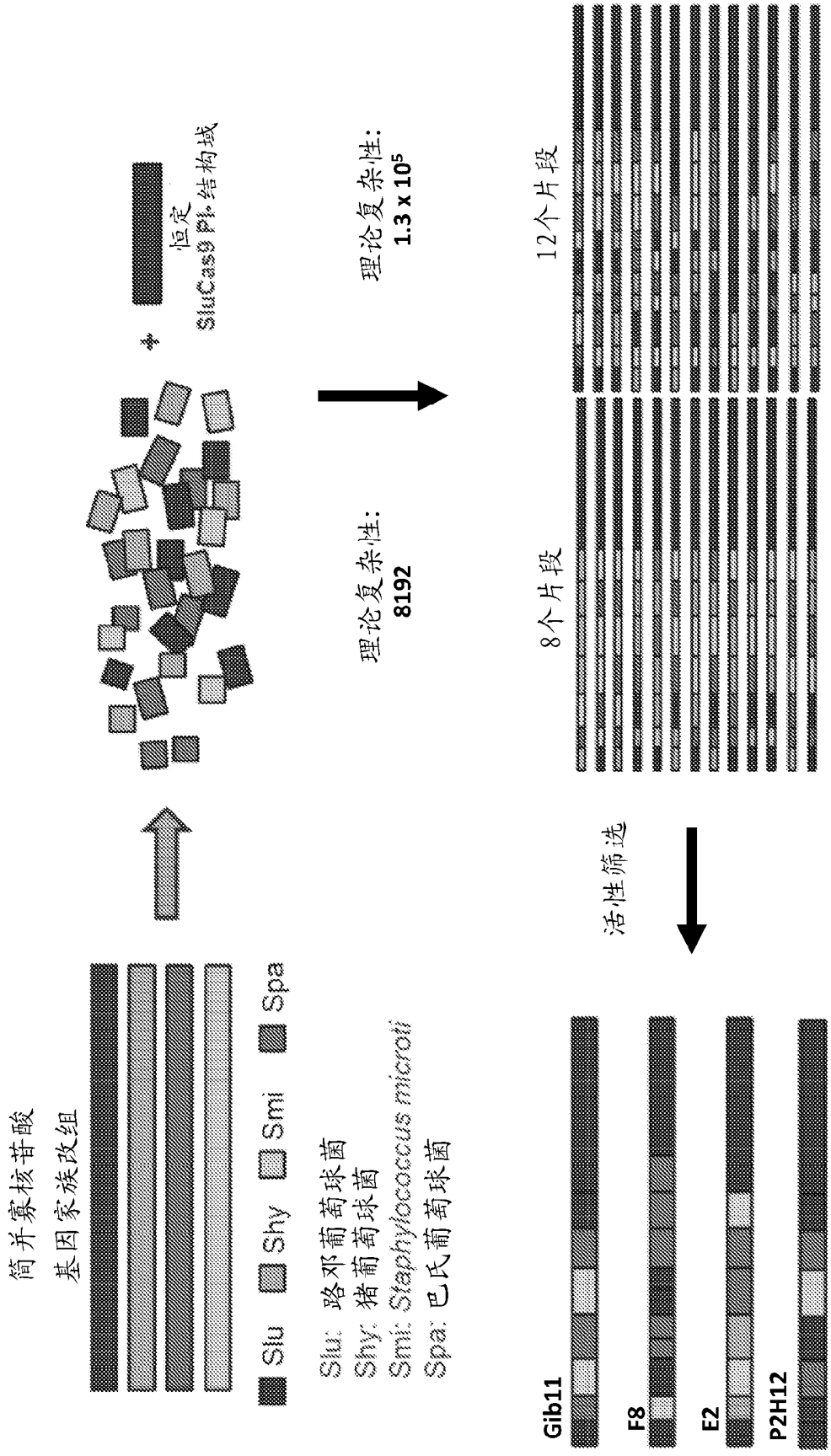

简并寡核苷酸
基因家族改组
Slu
Shy
Smi
Spa
Slu: 路邓葡萄球菌
Shy: 猪葡萄球菌
Smi: Staphylococcus microti
Spa: 巴氏葡萄球菌
+
恒定
SluCas9 PI结构域
理论复杂性:
8192
理论复杂性:
1.3 x 10^5
8个片段
12个片段
活性筛选
Gib11
F8
E2
P2H12

图 1

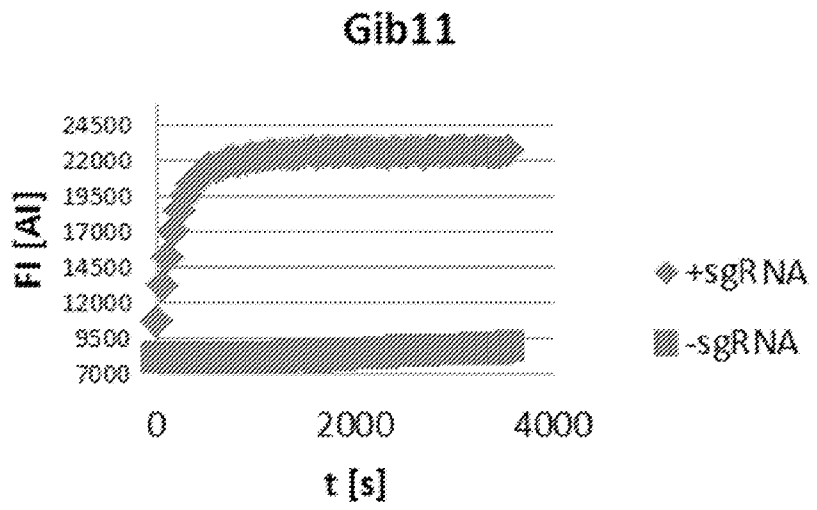

Gib11
24500
22000
19500
17000
14500
12000
9500
7000
FI [AU]
0
2000
4000
t [s]
+sgRNA
-sgRNA

图 2A

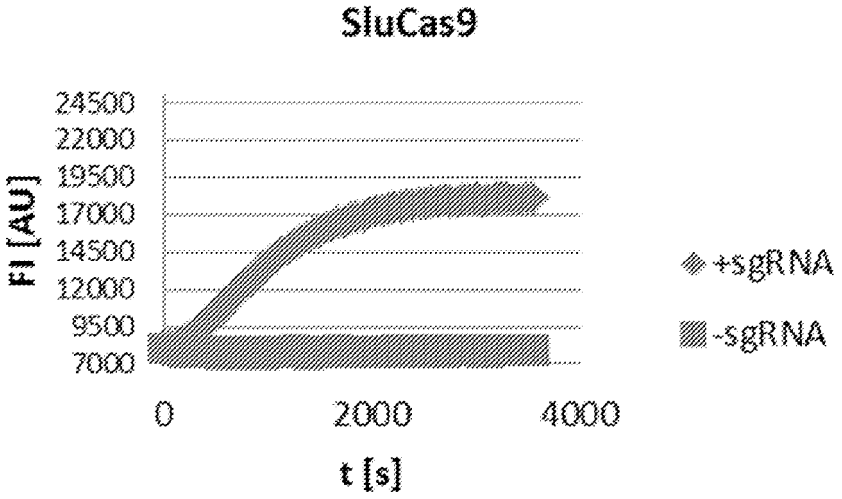

SluCas9
24500
22000
19500
17000
14500
12000
9500
7000
FI [AU]
0
2000
4000
t [s]
+sgRNA
-sgRNA

图 2B

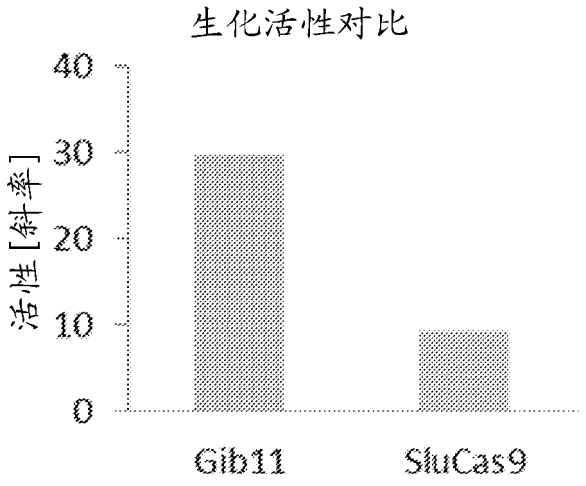

生化活性对比
活性[斜率]
40
30
20
10
0
Gib11
SluCas9

图 2C

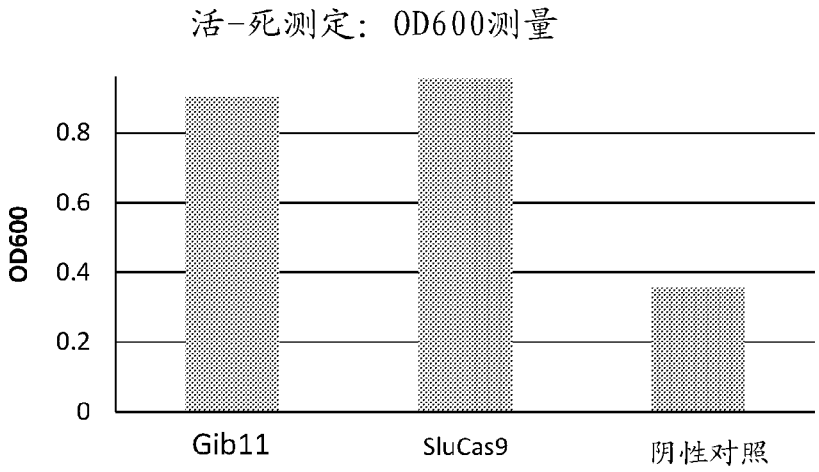

活-死测定：OD600测量
OD600
0.8
0.6
0.4
0.2
0
Gib11
SluCas9
阴性对照

图 3

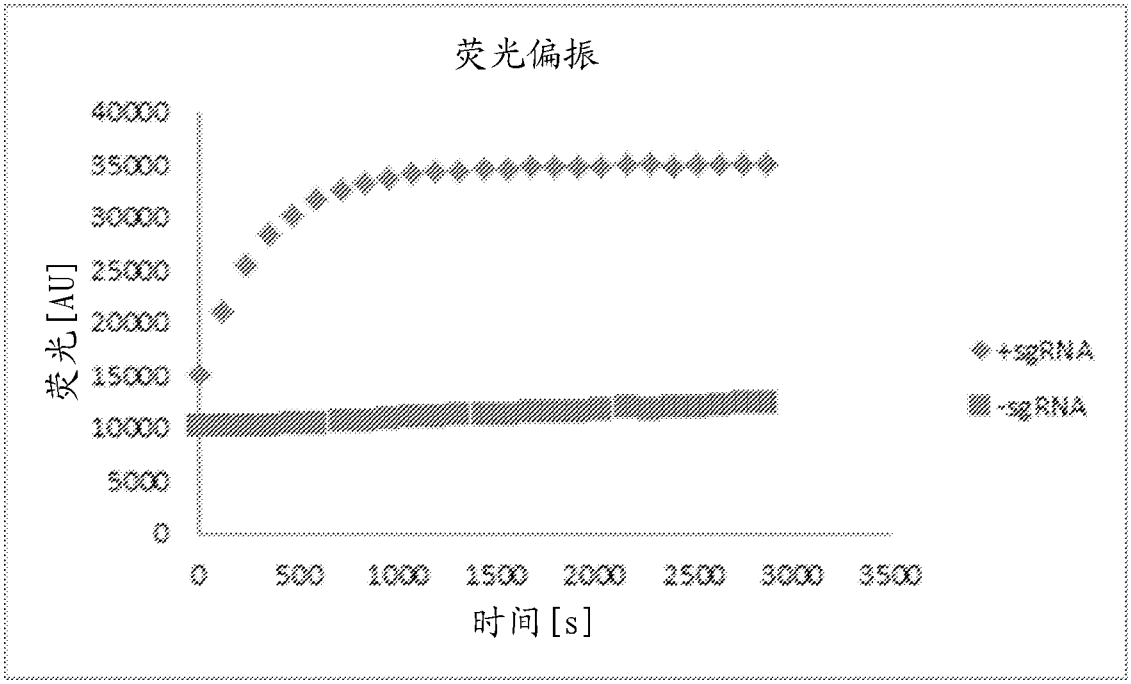

荧光偏振
荧光[AU]
40000
35000
30000
25000
20000
15000
10000
5000
0
0
500
1000
1500
2000
2500
3000
3500
时间[s]
+sgRNA
-sgRNA

图 4

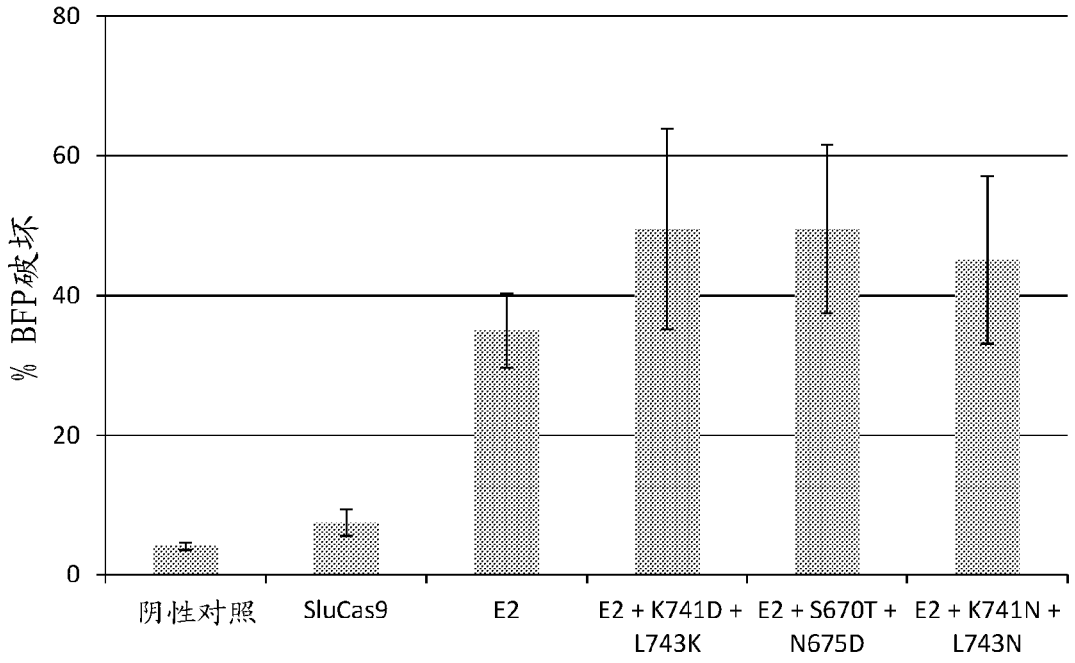

80
60
40
20
0
% BFP破坏
阴性对照
SluCas9
E2
E2 + K741D + L743K
E2 + S670T + N675D
E2 + K741N + L743N

图 5

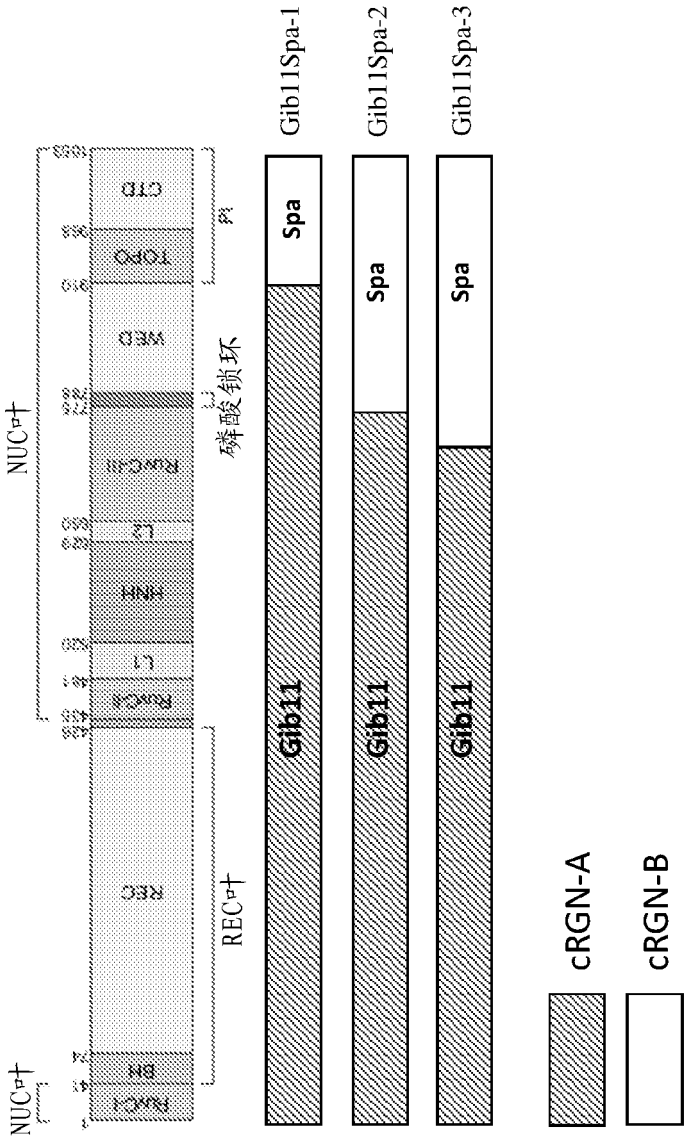

NUC叶
REC叶
磷酸锁环
Gib11Spa-1
Gib11Spa-2
Gib11Spa-3
Gib11
Spa
cRGN-A
cRGN-B
Gib11:
1) C-端交换
(交换 911-1057 AA)
2) C-端交换
(交换 789-1057 AA)
3) C-端交换
(交换 665-1057 AA)

图 6

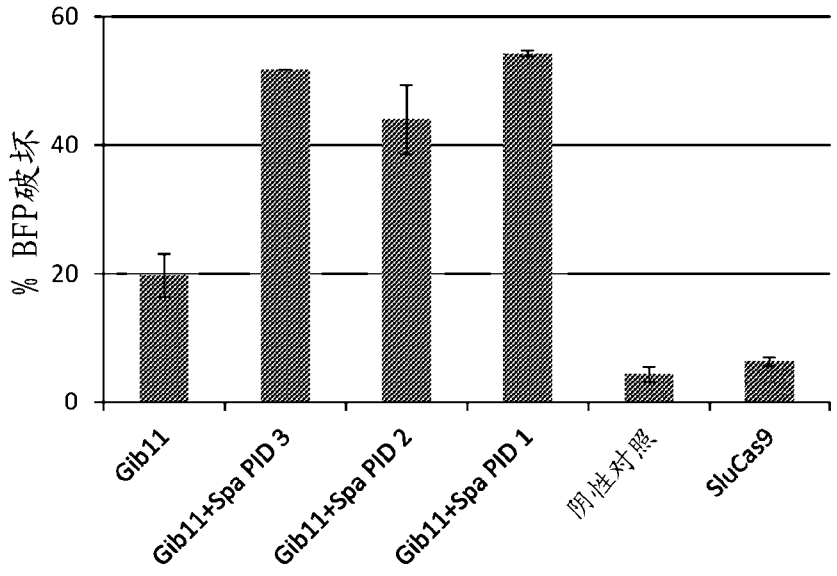

60
40
20
0
% BFP破坏
Gib11
Gib11+Spa PID 3
Gib11+Spa PID 2
Gib11+Spa PID 1
阴性对照
SluCas9

图 7

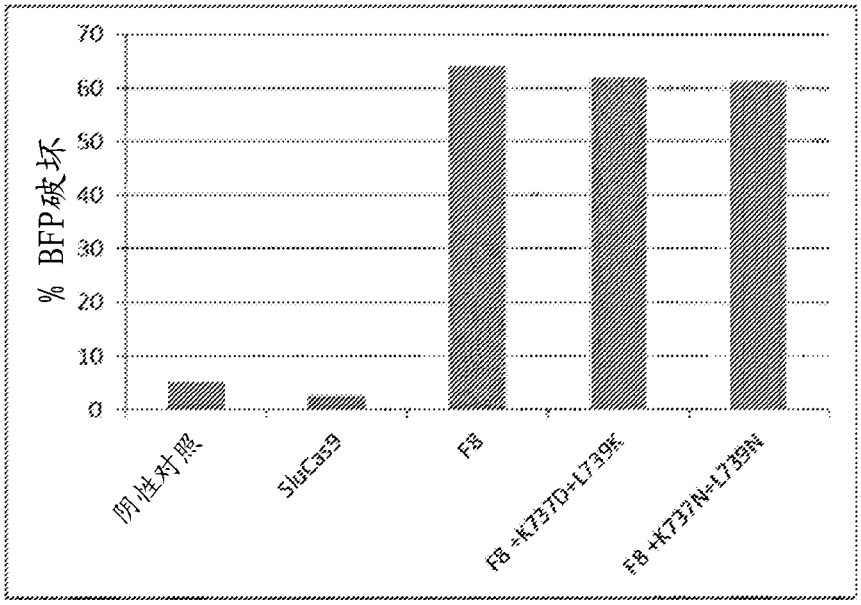

70
60
50
40
30
20
10
0
% BFP破坏
阴性对照
SluCas9
F8
F8 +K737D+L739K
F8 +K737N+L739K

图 8

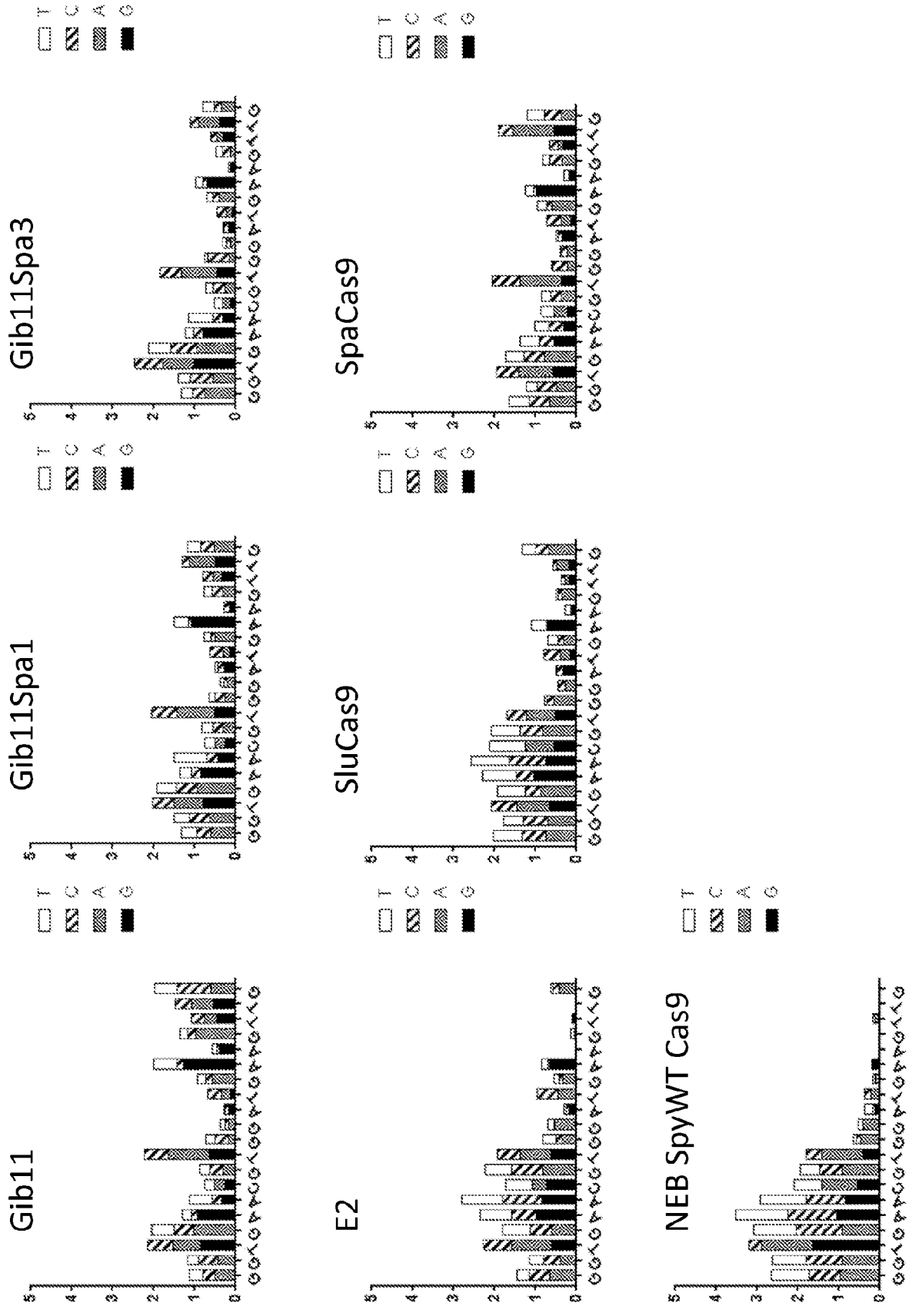

Gib11Spa3
SpaCas9
Gib11Spa1
SluCas9
Gib11
E2
NEB SpyWT Cas9
T C A G

图 9

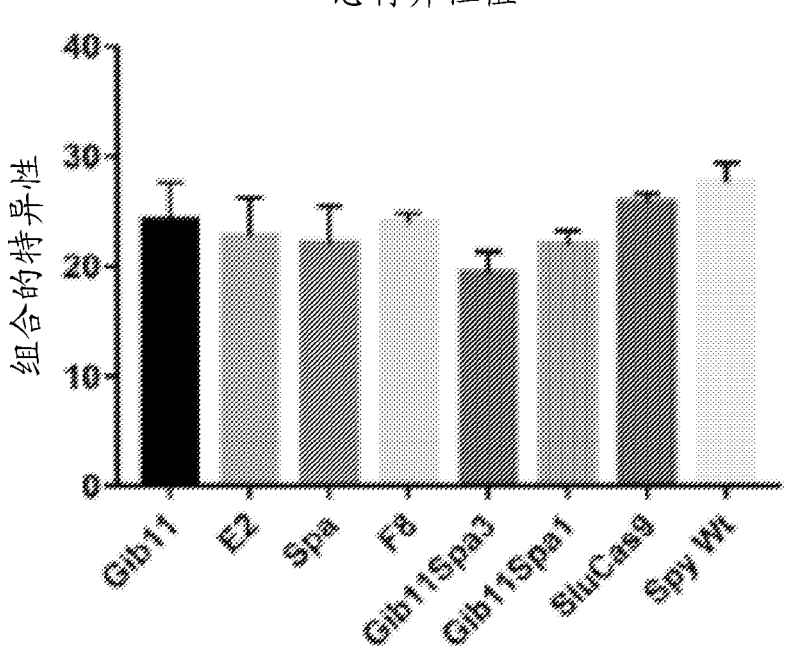

总特异性值
组合的特异性
40
30
20
10
0
Gib11
E2
Spa
Fa
Gib11Spa3
Gib11Spa1
SluCas9
Spy Wt

图 10

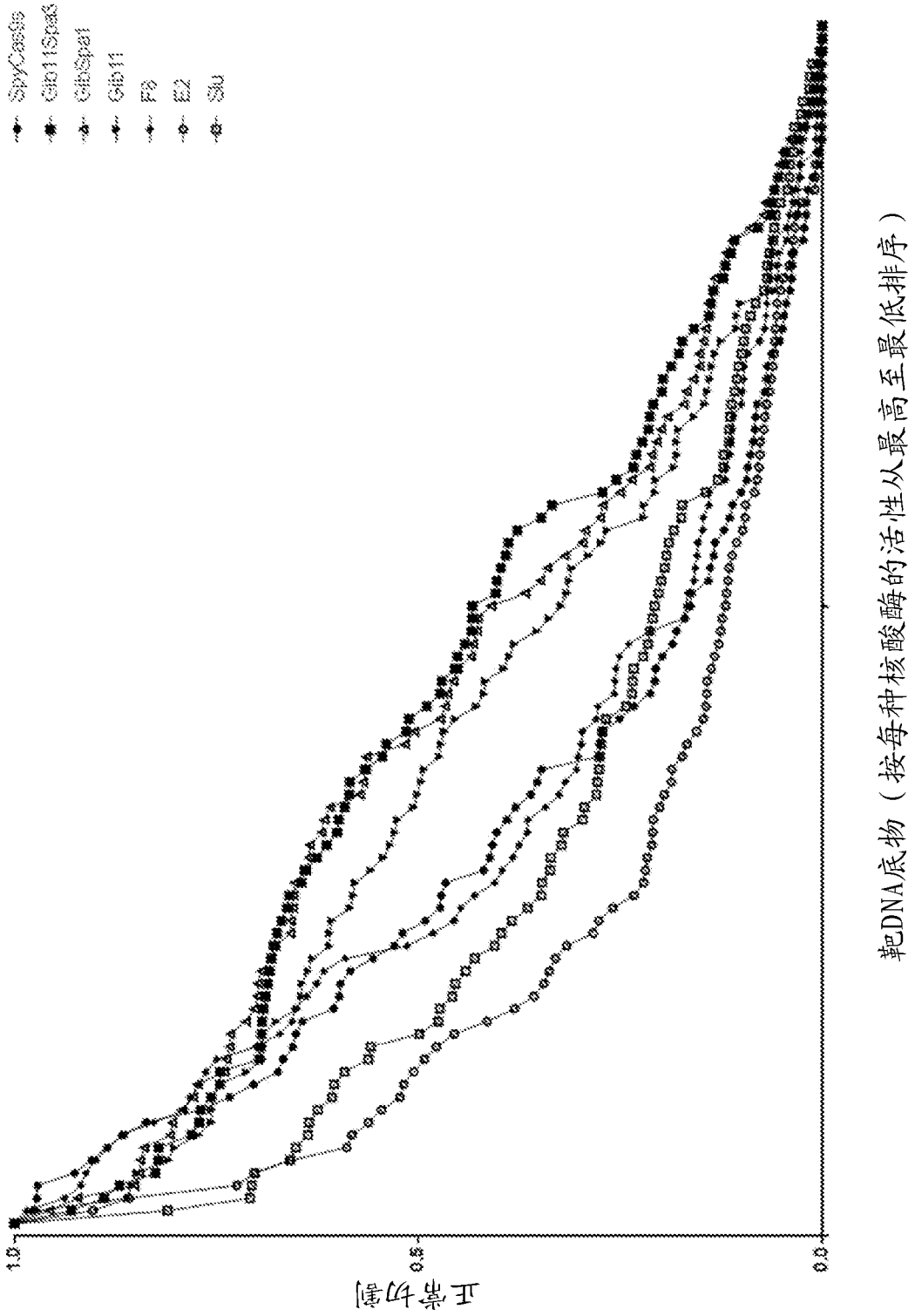

SpyCas9s
Gb11Spa3
GbSpa1
Gb11
F2
E2
SU
1.0
0.5
0.0
正常切割
靶DNA底物（按每种核酸酶的活性从最高至最低排序）

图 11