



(10) 授权公告号 CN 112654702 B

(45) 授权公告日 2025.05.13

(21) 申请号 201980058359.5

(22) 申请日 2019.09.06

(65) 同一申请的已公布的文献号
申请公布号 CN 112654702 A

(43) 申请公布日 2021.04.13

(30) 优先权数据
62/728184 2018.09.07 US

(85) PCT国际申请进入国家阶段日
2021.03.05

(86) PCT国际申请的申请数据
PCT/EP2019/073838 2019.09.06

(87) PCT国际申请的公布数据
W02020/049158 EN 2020.03.12

(73) 专利权人 阿斯利康(瑞典)有限公司
地址 瑞典南泰利耶

(72) 发明人 A-L·卡瓦洛

(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 李波 李唐

(51) Int.Cl.
C12N 9/22 (2006.01)

(56) 对比文件
WO 2018068053 A2, 2018.04.12

审查员 赵秋歌

权利要求书1页 说明书60页
序列表90页 附图25页

(54) 发明名称

改进的核酸酶的组合物和方法

(57) 摘要

本披露提供了具有比野生型Cas9更快的降解速率的重组Cas9蛋白。本披露还提供了具有减少的脱靶修饰的重组Cas9蛋白和CRISPR-Cas系统。本文还提供了利用本披露的重组Cas9蛋白的、具有减少的脱靶修饰的位点特异性修饰的方法。

1. 一种包含工程化的KFERQ基序的重组Cas9蛋白,其中该工程化的KFERQ基序在该重组Cas9蛋白的表面暴露的区域中,并且其中该工程化的KFERQ基序是VDKLN (SEQ ID NO:39),并且该重组Cas9蛋白是通过在SEQ ID NO:1的氨基酸位置185处引入F185N突变获得的。
2. 一种重组Cas9蛋白,该重组Cas9蛋白与如权利要求1所述的重组Cas9蛋白相比,进一步在SEQ ID NO:1中的位置D10、H840或其组合处包含突变,其中该突变选自D10A或D10N; H840A、H840N或H840Y;及其组合。
3. 如权利要求1或2所述的重组Cas9蛋白,其中该重组Cas9蛋白产生粘性末端。
4. 一种重组Cas9蛋白,该重组Cas9蛋白与如权利要求1或2所述的重组Cas9蛋白相比,进一步包含一个或多个核定位信号。
5. 一种编码如权利要求1至4中任一项所述的重组Cas9蛋白的多核苷酸分子。
6. 如权利要求5所述的多核苷酸分子,其中对该多核苷酸分子进行密码子优化以在真核细胞中表达。
7. 一种非天然存在的CRISPR-Cas系统,其包含:
 - (a) 如权利要求1至4中任一项所述的重组Cas9蛋白;以及
 - (b) 与该重组Cas9蛋白形成复合物并包含指导序列的指导多核苷酸。
8. 一种非天然存在的CRISPR-Cas系统,其包含:
 - (a) 如权利要求5或6所述的多核苷酸分子;以及
 - (b) 编码与该重组Cas9蛋白形成复合物并包含指导序列的指导多核苷酸的核苷酸序列。
9. 一种非天然存在的CRISPR-Cas系统,其包含:
 - (a) 可操作地连接至如权利要求5或6所述的多核苷酸分子的调节元件;以及
 - (b) 与该重组Cas9蛋白形成复合物并包含指导序列的指导多核苷酸。
10. 如权利要求7至9中任一项所述的系统,其中该指导序列与同向重复序列连接。
11. 如权利要求7至9中任一项所述的系统,其中该指导多核苷酸包含tracrRNA序列。
12. 一种包含如权利要求7至11中任一项所述的系统的递送颗粒。
13. 一种包含如权利要求7至11中任一项所述的系统的囊泡。
14. 如权利要求13所述的囊泡,其中该囊泡是外泌体或脂质体。
15. 一种包含如权利要求7至11中任一项所述的系统的病毒载体。
16. 一种在细胞基因组中的靶序列处提供位点特异性修饰的体外方法,该方法包括将如权利要求7至11中任一项所述的CRISPR-Cas系统引入该细胞中。

改进的核酸酶的组合物和方法

技术领域

[0001] 本披露提供了具有比野生型Cas9更快的降解速率的重组Cas9蛋白。本披露还提供了具有减少的脱靶修饰的重组Cas9蛋白和CRISPR-Cas系统。本文还提供了利用本披露的重组Cas9蛋白的、具有减少的脱靶修饰的位点特异性修饰的方法。

背景技术

[0002] 成簇的规律间隔的短回文重复序列 (CRISPR) 和CRISPR相关 (Cas) 系统是由Ishino在大肠杆菌中首次发现的、原核生物免疫系统 (Ishino等人, *Journal of Bacteriology* [细菌学杂志] 169 (12) : 5429-5433 (1987))。该原核生物免疫系统通过以序列特异性方式靶向病毒和质粒的核酸来提供针对病毒和质粒的免疫。还参见Soret等人, *Nature Reviews Microbiology* [自然综述微生物学] 6 (3) : 181-186 (2008)。CRISPR免疫应答涉及两个主要阶段: 第一个是采集, 第二个是干扰。采集阶段涉及切割入侵病毒和质粒的基因组, 并将入侵病毒和质粒基因组的区段整合到生物体的CRISPR基因座中。整合到生物体的CRISPR基因座中的区段被称为前间区序列, 并且有助于保护生物体免受相同病毒或质粒的后续攻击。第二阶段涉及攻击入侵的病毒或质粒。在第二阶段, 将前间区序列转录为RNA, 经过一些加工后, 该RNA与入侵病毒或质粒的DNA中的互补序列杂交, 同时还与有效裂解DNA的蛋白质或蛋白质复合物缔合。

[0003] 根据细菌物种的不同, CRISPRRNA加工的过程有所不同。例如, 最初在细菌酿脓链球菌 (*Streptococcus pyogenes*) 中描述的II型系统中, 转录的RNA与反式激活RNA (tracrRNA) 配对, 之后被RNA酶III裂解以形成单个CRISPR-RNA (crRNA)。在被Cas9核酸酶结合后, crRNA被进一步加工以产生成熟的crRNA。该crRNA/Cas9复合物随后结合包含与捕获区域互补的序列 (称为前间区序列) 的DNA。然后, 该Cas9蛋白以位点特异性的方式裂解DNA的两条链, 形成双链断裂 (DSB)。这提供了基于DNA的“记忆”, 导致病毒或质粒DNA在重复暴露和/或感染后迅速降解。关于原生CRISPR系统已经有全面的综述 (参见, 例如, Barrangou等人, *Cell* [细胞] 54 (2) : 234-244 (2014))。

[0004] 自最初发现以来, 多个研究小组围绕CRISPR系统在基因工程中的潜在应用进行了大量的研究, 包括基因编辑 (Jinek等人, *Science* [科学] 337 (6096) : 816-821 (2012); Cong等人, *Science* [科学] 339 (6121) : 819-823 (2013); 和Mali等人, *Science* [科学] 339 (6121) : 823-826 (2013))。CRISPR-Cas9基因编辑系统已经成功地用于广泛的生物体和细胞系。除基因组编辑外, CRISPR系统还具有许多其他应用, 包括调节基因表达、基因回路构建和功能基因组学等 (在Sander等人, *Nature Biotechnology* [自然生物技术] 32: 347-355 (2014) 中综述)。

[0005] CRISPR/Cas9用于治疗应用的适用性是强烈关注的话题。然而, 靶基因组的脱靶修饰 (即在非预期靶序列的基因座处的双链DNA断裂) 可能导致无法预测和不希望的结果, 引起对在临床应用中使用的CRISPR系统的关注。参见例如Hsu等人, *Nature Biotechnology* [自然生物技术] 31 (9) : 827-834 (2013); Hsu等人, *Cell* [细胞] 157 (6) : 1262-1278 (2014); 和

Schaefer等人, Nature Methods[自然方法]14(6):547-548(2017)。

[0006] CRISPR/Cas9系统的细胞毒性也受到关注。研究表明,当Cas9核酸酶效率提高时,细胞中会触发tp53依赖性毒性应答(Ihry等人, bioRxiv(2017), doi:10.1101/168443)。

[0007] 已致力于减少Cas9脱靶修饰。Fu等人描述了使用具有短靶互补区的截短的指导RNA以通过减少指导RNA-靶DNA界面的长度来减少Cas9的脱靶效应的方法(Nature Biotechnology[自然生物技术]32(3):279-284(2014))。Kleinstiver等人描述了工程化的Cas9变体,其与靶DNA序列的接触减少以最小化脱靶结合(Nature[自然]529(7587):490-495(2016))。然而,尽管Cas9脱靶活性降低,但两项研究也显示出靶上(on-target)核酸酶效率的相应降低。

[0008] 因此,在该领域中仍然需要具有降低的脱靶活性的改良CRISPR/Cas9系统,其保持靶上效率。

发明内容

[0009] 在一些实施例中,本披露提供了具有比野生型Cas9更快的降解速率的重组Cas9蛋白。在一些实施例中,本披露还提供了具有减少的脱靶修饰的重组Cas9蛋白和CRISPR-Cas系统。在一些实施例中,本披露提供了利用本文所述的重组Cas9蛋白的具有减少的脱靶修饰的位点特异性修饰的方法。

[0010] 在一些实施例中,本披露提供了包括工程化的KFERQ基序或KFERQ样基序的重组Cas9蛋白。

[0011] 在一些实施例中,工程化的KFERQ基序或KFERQ样基序选自KFERQ(SEQ ID NO:24)、RKVEQ(SEQ ID NO:25)、QDLKF(SEQ ID NO:26)、QRFFE(SEQ ID NO:27)、NRVVD(SEQ ID NO:28)、QRDKV(SEQ ID NO:29)、QKILD(SEQ ID NO:30)、QKKEL(SEQ ID NO:31)、QFREL(SEQ ID NO:32)、IKLDQ(SEQ ID NO:33)、DVVRQ(SEQ ID NO:34)、QRIVE(SEQ ID NO:35)、VKELQ(SEQ ID NO:36)、QKVFD(SEQ ID NO:37)、QELLR(SEQ ID NO:38)、VDKLN(SEQ ID NO:39)、RIKEN(SEQ ID NO:40)、NKKFE(SEQ ID NO:41)及其组合。在一些实施例中,工程化的KFERQ样基序是VDKLN(SEQ ID NO:39)。

[0012] 在一些实施例中,工程化的KFERQ基序或KFERQ样基序在重组Cas9蛋白的REC叶中。在一些实施例中,工程化的KFERQ基序或KFERQ样基序在REC叶的Rec2结构域中。在一些实施例中,工程化的KFERQ基序或KFERQ样基序在重组Cas9蛋白的HNH结构域、RuvC结构域或PI结构域中。

[0013] 在一些实施例中,工程化的KFERQ基序或KFERQ样基序在重组Cas9蛋白的表面暴露区域中。在一些实施例中,工程化的KFERQ基序或KFERQ样基序在重组Cas9蛋白的N末端或C末端处。

[0014] 在一些实施例中,本披露提供了包括野生型Cas9蛋白的一个或多个氨基酸修饰的重组Cas9蛋白,这些修饰将分子伴侣介导的自噬(CMA)靶基序或内体微自噬(eMI)靶基序引入该Cas9蛋白中,其中该重组Cas9蛋白在体内的降解比野生型Cas9蛋白或不包含CMA或eMI靶基序的Cas9蛋白快至少20%。在一些实施例中,重组Cas9蛋白在体内的降解比野生型Cas9蛋白或不包含CMA或eMI靶基序的Cas9蛋白快至少50%。在一些实施例中,重组Cas9蛋白在体内的降解比野生型Cas9蛋白或不包含CMA或eMI靶基序的Cas9蛋白快至少80%。

[0015] 在一些实施例中,本披露提供了包括野生型Cas9蛋白的一个或多个氨基酸修饰的重组Cas9蛋白,其中该重组Cas9蛋白包括CMA靶基序或eMI靶基序。

[0016] 在一些实施例中,CMA靶基序或eMI靶基序选自KFERQ (SEQ ID NO:24)、RKVEQ (SEQ ID NO:25)、QDLKF (SEQ ID NO:26)、QRFFE (SEQ ID NO:27)、NRVVD (SEQ ID NO:28)、QRDKV (SEQ ID NO:29)、QKILD (SEQ ID NO:30)、QKKEL (SEQ ID NO:31)、QFREL (SEQ ID NO:32)、IKLDQ (SEQ ID NO:33)、DVVRQ (SEQ ID NO:34)、QRIVE (SEQ ID NO:35)、VKELQ (SEQ ID NO:36)、QKVPD (SEQ ID NO:37)、QELLR (SEQ ID NO:38)、VDKLN (SEQ ID NO:39)、RIKEN (SEQ ID NO:40)、NKKFE (SEQ ID NO:41)及其组合。在一些实施例中,CMA靶基序或eMI靶基序是VDKLN (SEQ ID NO:39)。在一些实施例中,一个或多个氨基酸取代在重组Cas9蛋白的表面暴露区域中。

[0017] 在一些实施例中,本披露提供了分离自酿脓链球菌的重组Cas9蛋白 (SpCas9),该重组Cas9蛋白在SEQ ID NO:1的位置F185、A547、I548、T560、V561、D829、I830、L1087、S1088、P1199、K1200或其组合中的一个或多个处包括氨基酸修饰。

[0018] 在一些实施例中,氨基酸修饰包括以下突变中的一个或多个:F185N;A547E/I548L;T560E/V561Q;D829L/I830R;L1087E/S1088Q;或P1199D/K1200Q。在一些实施例中,氨基酸修饰是F185处的突变。在一些实施例中,突变是F185N。在一些实施例中,氨基酸修饰导致CMA靶基序或eMI靶基序。

[0019] 在一些实施例中,本披露的重组Cas9蛋白与SEQ ID NO:1具有至少90%同一性。

[0020] 在一些实施例中,本披露提供了能够与70kD的热休克同源蛋白 (HSC70) 结合的重组Cas9蛋白。

[0021] 在一些实施例中,本披露提供了分离自酿脓链球菌的重组蛋白 (SpCas9),该重组蛋白在SEQ ID NO:1的氨基酸位置185处包含工程化的KFERQ基序或KFERQ样基序。在一些实施例中,KFERQ样基序是VDKLN (SEQ ID NO:39)。

[0022] 在一些实施例中,本披露的重组Cas9蛋白进一步在SEQ ID NO:1中的位置D10、H840或其组合处包括突变。在一些实施例中,突变选自D10A或D10N;H840A、H840N或H840Y;及其组合。在一些实施例中,本披露的重组Cas9蛋白产生粘性末端。

[0023] 在一些实施例中,本披露的重组Cas9蛋白进一步包括一个或多个核定位信号。

[0024] 在一些实施例中,本披露提供了编码本披露的重组Cas9的多核苷酸序列。在一些实施例中,对多核苷酸序列进行密码子优化以在真核细胞中表达。

[0025] 在一些实施例中,本披露提供了非天然存在的CRISPR-Cas系统,该系统包括:本披露的重组Cas9蛋白;以及与重组Cas9蛋白形成复合物并包含指导序列的指导多核苷酸。

[0026] 在一些实施例中,本披露提供了非天然存在的CRISPR-Cas系统,该系统包括:编码本披露的重组Cas9蛋白的多核苷酸序列;以及与重组Cas9蛋白形成复合物并包括指导序列的指导多核苷酸。

[0027] 在一些实施例中,本披露提供了非天然存在的CRISPR-Cas系统,该系统包括:可操作地连接至编码本披露的重组Cas9蛋白的多核苷酸序列的调节元件;以及与重组Cas9蛋白形成复合物并包括指导序列的指导多核苷酸。

[0028] 在CRISPR-Cas系统的一些实施例中,指导序列与同向重复序列连接。

[0029] 在CRISPR-Cas系统的一些实施例中,指导多核苷酸包括tracrRNA序列。在一些实

施例中,CRISPR-Cas系统包括单独的多核苷酸,该多核苷酸包括tracrRNA序列。

[0030] 在CRISPR-Cas系统的一些实施例中,编码重组Cas9蛋白的多核苷酸序列和指导多核苷酸在单个载体上。在CRISPR-Cas系统的一些实施例中,编码重组Cas9蛋白的多核苷酸序列、指导多核苷酸和tracrRNA序列在单个载体上。

[0031] 在一些实施例中,递送颗粒包括本披露的CRISPR-Cas系统。在一些实施例中,囊泡包括本披露的CRISPR-Cas系统。在一些实施例中,该囊泡是外泌体或脂质体。

[0032] 在一些实施例中,病毒载体包括本披露的CRISPR-Cas系统。在一些实施例中,该病毒载体是腺病毒、慢病毒或腺相关病毒载体。

[0033] 在一些实施例中,本披露提供了在细胞基因组中的靶序列处提供位点特异性修饰的方法,该方法包括将本披露的CRISPR-Cas系统引入该细胞中。

[0034] 在该方法的一些实施例中,修饰包括靶序列的至少一部分的缺失。在该方法的一些实施例中,修饰包括靶序列的突变。在该方法的一些实施例中,修饰包括在靶序列处插入目的序列(SoI)。

[0035] 在该方法的一些实施例中,细胞基因组中的脱靶修饰小于由重组Cas9产生的基因组中修饰的约5%。在该方法的一些实施例中,细胞基因组中的脱靶修饰小于由重组Cas9产生的基因组中修饰的约2%。在该方法的一些实施例中,细胞基因组中的脱靶修饰小于由重组Cas9产生的基因组中修饰的约1%。在该方法的一些实施例中,相对于野生型CRISPR-Cas9或不包括KFERQ基序或KFERQ样基序的Cas9,细胞基因组中的脱靶修饰减少了至少约50%。

[0036] 在该方法的一些实施例中,细胞是细菌细胞、哺乳动物细胞或植物细胞。在该方法的一些实施例中,细胞是人类细胞。在该方法的一些实施例中,细胞是多能干细胞。在该方法的一些实施例中,细胞是诱导的多能干细胞。

[0037] 在该方法的一些实施例中,指导多核苷酸的指导序列能够与细胞基因组中的靶序列杂交。在该方法的一些实施例中,将CRISPR-Cas系统经由递送颗粒、囊泡或病毒载体引入细胞中。

附图说明

[0038] 图1和2涉及实例4中所述的实验。

[0039] 图1A显示在将Cas9稳定转染到人尿道上皮细胞(SVHUC-1)中后,细胞数量减少。图1B显示表达Cas9的小鼠的体重在诱导Cas9表达后降低。

[0040] 图2左小图是诱导的多能干细胞(iPSC)的显微镜图像。图2右小图是表达Cas9的iPSC的显微镜图像。

[0041] 图3显示来自酿脓链球菌的Cas9蛋白(SpCas9)的示意图。

[0042] 图4显示结合至指导RNA(Sg RNA)和DNA的SpCas9的晶体结构。

[0043] 图5至12涉及实例1中所述的实验。

[0044] 图5A显示含有野生型Cas9和包括KFERQ基序的Cas9的质粒的示意图。图5B显示含有分别用FLAG标签标记的野生型Cas9和FaDe-Cas9的质粒的示意图。

[0045] 图6A和6B显示检测Cas9或FaDe-Cas9存在的蛋白质印迹(免疫印迹)。在图6A中,通过Cas9特异性抗体可检测到野生型Cas9,但未检测到KFERQ-Cas9。在图6B中,通过对FLAG和

Cas9特异的抗体可以检测到FLAG标记的野生型Cas9,但是没有一种抗体可以检测到FLAG标记的KFERQ-Cas9。

[0046] 图7A和7B显示Cas9或FaDe-Cas9随时间的表达。图7A中的蛋白质印迹显示野生型Cas9水平随时间增加,但是通过Cas9特异性抗体未检测到FaDe-Cas9。Low exp:低暴露;high exp:高暴露;ctr:对照(无Cas9)。图7B显示Cas9和FaDe-Cas9的mRNA转录物水平在相同时间点是可比较的。

[0047] 图8显示双报告载体的示意图,其包括用于表达与GFP融合的Cas9的一个启动子和用于表达mCherry的第二个启动子。

[0048] 图9A和9B是分别显示表达Cas9-GFP和FaDe-Cas9-GFP的细胞的荧光显微镜图像。

[0049] 图10A显示通过mCherry荧光测量的Cas9和FaDe-Cas9的转染效率。图10B显示通过GFP荧光测量的Cas9和FaDe-Cas9的表达水平。

[0050] 图11A和11B分别显示Cas9-GFP和mCherry的荧光显微镜图像。图11C显示图11A和11B的合并,其表明表达GFP的相同细胞也表达mCherry。

[0051] 图12显示蛋白质印迹,其表明通过Cas9特异性抗体检测的Cas9和FaDe-Cas9在不同细胞类型中随时间的表达水平。

[0052] 图13至16涉及实例2中所述的实验。

[0053] 图13A显示通过HSC70特异性抗体检测的Cas9或FaDe-Cas9与HSC70的免疫共沉淀印迹。图13B显示Cas9或FaDe-Cas9和Lamp-2A的免疫荧光图像。

[0054] 图14显示两个质粒的示意图:第一个质粒表达与dsRed融合的Lamp-2A,并且第二个质粒表达与GFP融合的FaDe-Cas9。

[0055] 图15显示Lamp-2A-dsRed、FaDe-Cas9-GFP的荧光显微镜图像,以及显示Lamp-2A和FaDe-Cas9的共定位的合并图像。

[0056] 图16显示蛋白质印迹,其表明Cas9和FaDe-Cas9在胞质溶胶或细胞核中的定位。

[0057] 图17至19涉及实例3中所述的实验。

[0058] 图17A显示测试HEK细胞中Cas9和FaDe-Cas9核酸酶活性的Surveyor核酸酶测定(细胞测定)的结果。图17B显示表明Cas9和FaDe-Cas9的核酸酶效率的下一代测序结果。

[0059] 图18显示测试hiPSc中Cas9和FaDe-Cas9核酸酶活性的测定结果。RNP:核糖核蛋白;pl:质粒。

[0060] 图19显示在EMX和FANCF基因座处由Cas9和FaDe-Cas9进行的脱靶修饰的分析结果。左小图、中小图和右小图分别比较Cas9和FaDe-Cas9之间的靶上效率、脱靶效率和标准化的靶上效率。

[0061] 图20A表示显示FaDe-Cas9与Cas9相比具有相当的靶上效率和降低的脱靶活性的数据。图20B显示与FaDe-Cas9和未转染的细胞相比,用Cas9转染的细胞具有降低的增殖速率。

[0062] 图21显示与Cas9相比,用FaDe-Cas9编辑的细胞导致减少的染色体易位。

[0063] 图22A显示测试细胞对Cas9和FaDe-Cas9的耐受性的实验的示意图。图22B和22C显示与Cas9相比,细胞可以耐受更多拷贝的FaDe-Cas9。

[0064] 图23A和23B显示在0至100小时的时间点用Cas9或FaDe-Cas9转染的细胞的Cas9的蛋白质印迹的定量(图23A),以及0至24小时的时间点特写(图23B)。

- [0065] 图24A-24C显示通过ELISA测定法测量的Cas9和FaDe-Cas9细胞内蛋白水平。
- [0066] 图25A和25B显示通过GFP信号分析KD lamp2a效率和蛋白质积累的结果。
- [0067] 图26显示了免疫沉淀结果,其显示FaDe-Cas9与分子伴侣介导的自噬HSC70的主调节剂具有高亲和力结合。
- [0068] 图27A和27B显示体内靶上/脱靶分析图。
- [0069] 图28A显示暴露于FaDe-Cas9的小鼠肝脏由于其在肝组织中的快速降解而显示出更高的基因编辑,继而导致编辑细胞的存活率提高。图28B和28C显示与Cas9相比,FaDe-Cas9的体内快速周转导致较低的肝毒性,这通过未改变的肝糖原、几乎不可检测的有丝分裂标记Ki67、较少量的浸润物和最小的细胞坏死来证实。图28D显示病毒介导的FaDe-Cas9的体内表达导致较低的细胞毒性T淋巴细胞免疫应答。

具体实施方式

[0070] 本文描述了CRISPR-Cas系统的组分,该系统可用于基因组编辑、基因组工程和改变基因和/或遗传元件的表达。CRISPR-Cas系统可用于各种治疗应用,包括遗传疾病的治疗。本文还描述了Cas9蛋白的快速降解变体(有时称为“FaDe-Cas9”),其有助于减少CRISPR-Cas9系统的脱靶活性。本文描述了快速降解的Cas9蛋白的其他优点,这些优点包括但不限于与野生型Cas9相当的靶上效率和/或与野生型Cas9相比降低的毒性。

[0071] 定义

[0072] 如本文所用,“一个/一种(a或an)”可以意指一个/种或多个/种。如本文说明书和一项或多项权利要求中所用,当与单词“包含”结合使用时,单词“一个/种”可以意指一个/种或多于一个/种。如本文所用,“另一个/种”可以意指至少第二个/种或更多个/种。

[0073] 在整个本申请中,术语“约”用于指示值包括被采用以确定值的方法/装置的误差的固有变化,或者研究受试者之间存在的变化。典型地,该术语意指涵盖近似于或小于1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、11%、12%、13%、14%、15%、16%、17%、18%、19%或20%变化,这取决于具体情况。

[0074] 权利要求中使用术语“或”用于意指“和/或”,除非明确指示仅指替代方案或替代方案是相互排斥的,尽管本披露支持仅是指替代方案和“和/或”的定义。

[0075] 如在本说明书和一项和多项权利要求中所使用的,词语“包含”(和任何形式的包含,诸如“包含有”和“包含着”),“具有”(和任何形式的具有,诸如“具有着”和“有”),“包括”(以及任何形式的包括,诸如“包括有”和“包括着”)或“含有”(以及任何形式的含有,诸如“含有着”和“含”)是包扩性的或开放性的,并且不排除另外的未列举的元素或方法步骤。考虑到本说明书中讨论的任何实施例可以相对于本披露的任何方法、系统、宿主细胞、表达载体和/或组合物实施。此外,本披露的组合物、系统、宿主细胞和/或载体可用于实现本披露的方法和蛋白质。

[0076] 术语“例如”及其对应的简称“如”(无论是否斜体)的使用意味着所列举的特定术语是本披露的代表性实例和实施例,并不旨在限于所引用或列举的特定实例,除非另有明确说明。

[0077] “核酸”、“核酸分子”、“核苷酸”、“核苷酸序列”、“寡核苷酸”或“多核苷酸”意指包括共价连接的核苷酸的聚合化合物。术语“核酸”包括核糖核酸(RNA)和脱氧核糖核酸

(DNA),两者都可以是单链或双链的。DNA包括但不限于互补DNA(cDNA)、基因组DNA、质粒或载体DNA和合成DNA。在一些实施例中,本披露提供了编码本文披露的任何一种多肽的多核苷酸,例如,本披露涉及编码Cas蛋白或其变体的多核苷酸。

[0078] “基因”是指编码多肽的核苷酸的组装,并且包括cDNA和基因组DNA核酸分子。“基因”还指可以充当编码序列之前(5'非编码序列)和之后(3'非编码序列)的调节序列的核酸片段。

[0079] 当单链形式的核酸分子可在适合的温度和溶液离子强度条件下退火至另一核酸分子上时,核酸分子与该另一核酸分子(诸如cDNA、基因组DNA或RNA)“可杂交”或“杂交”。杂交和洗涤条件是已知的,并在Sambrook等人,Molecular Cloning:A Laboratory Manual[分子克隆:实验室手册],第二版,Cold Spring Harbor Laboratory Press[冷泉港实验室出版社],Cold Spring Harbor[冷泉港](1989),特别是其中第11章和表11.1中作了举例说明。温度和离子强度条件决定了杂交的“严格性”。可以调整严格条件以筛选适度类似的片段(如来自远缘生物体的同源序列),至高度类似的片段(如复制来自近缘生物体的功能性酶的基因)。为了初步筛选同源核酸,可以使用对应于55°C的 T_m 的低严格性杂交条件,例如,5XSSC、0.1% SDS、0.25%牛奶和不含甲酰胺;或30%甲酰胺,5XSSC,0.5% SDS。中等严格性杂交条件对应于较高的 T_m ,例如,40%甲酰胺和5X或6XSSC。高严格性杂交条件对应于最高的 T_m ,例如,50%甲酰胺,5X或6XSSC。杂交需要两个核酸包含互补序列,尽管取决于杂交的严格性,碱基之间可能存在错配。

[0080] 术语“互补的”用于描述能够彼此杂交的核苷酸碱基之间的关系。例如,对于DNA,腺苷与胸腺嘧啶互补,而胞嘧啶与鸟嘌呤互补。因此,本披露还包括与本文披露或使用的完整序列互补的分离核酸片段以及那些基本上相似的核酸序列。

[0081] DNA“编码序列”是双链DNA序列,当置于适当调节序列的控制下时,该双链DNA序列在体外细胞中或体内被转录并翻译成多肽。“合适的调节序列”是指位于编码序列上游(5'非编码序列)、内部或下游(3'非编码序列)的核苷酸序列,并且该核苷酸序列影响转录、RNA加工或稳定性或相关编码序列的翻译。调节序列可包括启动子、翻译前导序列、内含子、聚腺苷酸化识别序列、RNA加工位点、效应子结合位点和茎环结构。该编码序列的边界由5'(氨基)末端处的起始密码子和3'(羧基)末端处的翻译终止密码子确定。编码序列可以包括但不限于原核序列、来自mRNA的cDNA、基因组DNA序列,甚至是合成DNA序列。如果编码序列旨在真核细胞中表达,则多腺苷酸化信号和转录终止序列通常存在于该编码序列的3'端。

[0082] “开放阅读框”的缩写为ORF,意指包括翻译起始信号或起始密码子(诸如ATG或AUG)以及终止密码子并且可能被翻译成多肽序列的一段核酸序列(DNA、cDNA或RNA)。

[0083] 术语“同源重组”是指将外来DNA序列插入到另一个DNA分子中,例如,将载体插入染色体中。在一些情况下,该载体靶向特定的染色体位点以进行同源重组。对于特定的同源重组,该载体通常含有与染色体序列具有同源性的足够长的区域,以允许载体与染色体的互补结合和将载体掺入染色体。更长的同源性区域和更大程度的序列相似性可提高同源重组的效率。

[0084] 根据本文的披露内容,可以使用本领域已知的方法来扩增多核苷酸。一旦建立了合适的宿主系统和生长条件,就可以大量扩增和制备重组表达载体。如本文所述,可以使用的表达载体包括但不限于以下载体或其衍生物:人类或动物病毒,诸如牛痘病毒或腺病毒;

昆虫病毒,诸如杆状病毒;酵母载体;噬菌体载体(例如, λ),以及质粒和粘粒DNA载体。

[0085] 如文中所用,“可操作地连接”意指目的多核苷酸,例如编码Cas9蛋白的多核苷酸,以允许多核苷酸序列表达的方式与调节元件连接。在一些实施例中,该调节元件是启动子。在一些实施例中,目的多核苷酸可操作地连接至表达载体上的启动子。

[0086] 如本文所用,“启动子”,“启动子序列”或“启动子区”是指能够结合RNA聚合酶并涉及起始下游编码或非编码序列的转录的DNA调节区/序列。在本披露的一些实例中,启动子序列包括转录起始位点并向上游延伸以包括以高于背景可检测水平起始转录所使用的最少数目的碱基或元件。在一些实施例中,启动子序列包括转录起始位点,以及负责RNA聚合酶结合的蛋白质结合结构域。真核启动子通常但并非总是包含多个“TATA”盒和“CAT”盒。各种启动子,包括诱导型启动子,可用于驱动本披露的各种载体。

[0087] “载体”是用于将核酸克隆和/或转移至宿主细胞中的任何工具。载体可以是可能与另一个DNA区段附接的复制子,这样以便产生附接区段的复制。“复制子”是任意一种遗传因子(例如,质粒、噬菌体、粘粒、染色体、病毒),它充当DNA体内复制的自动单元,即在其自我控制下能复制。在本披露的一些实施例中,载体是附加型载体,该载体在许多个细胞世代后,通过例如不对称分配,从细胞群中去除/丢失。术语“载体”包括用于在体外、离体或体内将该核酸引入细胞中的病毒性和非病毒工具。可以使用本领域已知的大量载体,以便操作核酸,将响应元件和启动子整合至基因中等等。可能的载体包括例如质粒或经修饰的病毒,包括例如噬菌体诸如 λ 衍生物,或质粒诸如PBR322或pUC质粒衍生物,或Bluescript载体。例如,将对应于响应元件和启动子的DNA片段插入合适的载体中能伴随着将合适的DNA片段连接至具有互补结合端的选定载体。可替代地,DNA分子的末端可以被酶催化修饰或者任意位点通过将核苷酸序列(接头)连接至该DNA末端中而产生。此类载体可进行工程化处理,以包含提供对细胞进行选择的选择性标记基因,这些细胞将标记掺入到细胞基因组中。此类标记允许鉴定和/或选择宿主细胞,这些宿主细胞掺入和表达该标记所编码的蛋白质。

[0088] 病毒载体,特别是逆转录病毒载体,已经用于细胞以及活体动物的许多种基因递送应用中。可以使用的病毒载体包括但不限于逆转录病毒、腺相关病毒、痘病毒、杆状病毒、牛痘、单纯性疱疹、艾巴氏病毒、腺病毒、双生病毒和花椰菜花叶病毒载体。非病毒载体包括但不限于质粒、脂质体、带电脂质(细胞转染素)、DNA-蛋白质复合物和生物聚合物。除核酸外,载体还可包括一个或多个调节区和/或用于选择、测量和监测核酸转移结果(转移至哪个组织、表达持续时间等)的选择性标记。

[0089] 可以通过已知的方法将载体引入希望的宿主细胞中,这些方法包括但不限于转染、转导、细胞融合和脂质转染。载体可包括各种调节元件,包括启动子。在一些实施例中,载体设计可以基于由Mali等人“Cas9 as a versatile tool for engineering biology [Cas9作为工程生物学的多用途工具]”,Nature Methods[自然方法]10:957-63(2013)设计的多个构建体。在一些实施例中,本披露提供了包括本文所述的任何多核苷酸的表达载体,例如,包括编码Cas蛋白或其变体的多核苷酸的表达载体。在一些实施例中,本披露提供了包括编码Cas9蛋白或其变体的多核苷酸的表达载体。

[0090] 术语“质粒”是指额外的染色体元件,该元件通常携带不参与细胞的中央代谢的基因,并且通常呈环状双链DNA分子的形式。此类元件可以是来源于任何来源的单链或双链DNA或RNA的线性、环状或超螺旋自主复制序列、基因组整合序列、噬菌体或核苷酸序列,其

中许多核苷酸序列已连接或重组到独特结构中,该独特结构能够将针对选定基因产物的启动子片段和DNA序列连同适当的3'未翻译序列引入到细胞中。

[0091] 如本文所用,“转染”意指将外源核酸分子(包括载体)引入细胞。“转染的”细胞在细胞内部包括外源核酸分子,而“转化的”细胞是其中细胞内的外源核酸分子诱导细胞表型变化的细胞。转染的核酸分子可以整合到宿主细胞的基因组DNA中和/或可以被细胞暂时或长时间地维持在染色体外。表达外源核酸分子或片段的宿主细胞或生物体被称为“重组”、“转化”或“转基因”生物体。在一些实施例中,本披露提供了包括本文所述的任何表达载体(例如,包括编码Cas蛋白或其变体的多核苷酸的表达载体)的宿主细胞。在一些实施例中,本披露提供了包括表达载体的宿主细胞,该表达载体包括编码Cas9蛋白或其变体的多核苷酸。

[0092] 术语“宿主细胞”是指已导入重组表达载体的细胞。术语“宿主细胞”不仅指其中导入表达载体的细胞(“亲本”细胞),还指这种细胞的后代。因为修饰可能例如由于突变或环境影响而在后代中发生,所以后代可能与亲本细胞不同,但仍包括在术语“宿主细胞”的范围内。

[0093] 术语“肽”、“多肽”和“蛋白质”在本文中可互换使用,是指任何长度的氨基酸的聚合形式,这些聚合形式可以包括编码和非编码的氨基酸、经化学或生物化学修饰或衍生的氨基酸和具有经修饰的肽骨架的多肽。

[0094] 蛋白质或多肽的起点称为“N末端”(或氨基末端、NH₂末端、N末端或胺末端),是指蛋白质或多肽的第一个氨基酸残基的游离胺(-NH₂)基团。蛋白质或多肽的末端称为“C末端”(或羧基末端、羧基末端、C末端或COOH末端),是指蛋白质或肽的最后一个氨基酸残基的游离羧基(-COOH)。

[0095] 如本文所用,“氨基酸”是指包括羧基(-COOH)和氨基(-NH₂)的化合物。“氨基酸”是指天然和非天然(即合成)氨基酸。天然氨基酸及其三个字母和一个字母的缩写包括:丙氨酸(Ala;A);精氨酸(Arg,R);天冬酰胺(Asn;N);天冬氨酸(Asp;D);半胱氨酸(Cys;C);谷氨酰胺(Gln;Q);谷氨酸(Glu;E);甘氨酸(Gly;G);组氨酸(His;H);异亮氨酸(Ile;I);亮氨酸(Leu;L);赖氨酸(Lys;K);蛋氨酸(Met;M);苯丙氨酸(Phe;F);脯氨酸(Pro;P);丝氨酸(Ser;S);苏氨酸(Thr;T);色氨酸(Trp;W);酪氨酸(Tyr;Y);以及缬氨酸(Val;V)。

[0096] “氨基酸取代”是指包括一个或多个野生型或天然存在的氨基酸被相对于该野生型或天然存在的氨基酸不同的氨基酸在该氨基酸残基处取代的多肽或蛋白质。被取代的氨基酸可以是合成或天然存在的氨基酸。在一些实施例中,被取代的氨基酸是选自下组的天然存在的氨基酸,该组由以下组成:A、R、N、D、C、Q、E、G、H、I、L、K、M、F、P、S、T、W、Y和V。取代突变体可以使用缩写体系进行描述。例如,第五(第5)个氨基酸残基被取代的取代突变可以缩写为“X5Y”,其中“X”是被代替的野生型或天然存在的氨基酸,“5”是该蛋白质或多肽的氨基酸序列内氨基酸残基的位置,“Y”是被取代的或非野生型或非天然存在的氨基酸。

[0097] “分离的”多肽、蛋白质、肽或核酸是已经从其天然环境中去除的分子。还应理解,“分离的”多肽、蛋白质、肽或核酸可以与赋形剂(诸如稀释剂)或佐剂一起配制,并且仍被认为是分离的。

[0098] 当用于指核酸分子、肽、多肽或蛋白质时,术语“重组”意指未知在自然界中存在的遗传物质的新组合或由其产生。重组分子可以通过重组技术领域中的任何熟知的技术

来产生,包括但不限于聚合酶链式反应(PCR)、基因剪接(例如,使用限制性核酸内切酶)以及核酸分子、肽或蛋白质的固相合成。

[0099] 当用于指多肽或蛋白质时,术语“结构域”意指蛋白质中独特的功能和/或结构单元。结构域有时负责特定的功能或相互作用,有助于蛋白质的整体作用。在多种生物学背景下可以存在结构域。在具有不同功能的蛋白质中可以找到相似的结构域。可替代地,具有低序列同一性(即,小于约50%、小于约40%、小于约30%、小于约20%、小于约10%、小于约5%或小于约1%的序列同一性)的结构域可能具有相同的功能。在一些实施例中,Cas9结构域是RuvC结构域。在一些实施例中,Cas9结构域是HNH结构域。在一些实施例中,Cas9结构域是Rec结构域。

[0100] 当用于多肽或蛋白质时,术语“基序”通常是指一组典型地长度短于20个氨基酸的保守的氨基酸残基,其对于蛋白质功能可能是重要的。特定的序列基序可以介导多种蛋白质中的共同功能,例如蛋白质结合或靶向特定亚细胞位置。基序的实例包括但不限于核定位信号、微体靶向基序、阻止或促进分泌的基序和促进蛋白质识别和结合的基序。基序数据库和/或基序搜索工具是本领域技术人员已知的,并且包括例如PROSITE (expasy.ch/sprot/prosite.html)、Pfam (pfam.wu.stl.edu)、PRINTS (biochem.ucl.ac.uk/bsm/dbbrowser/PRINTS/PRINT_S.html)和Minimotif Miner (cse-mnm.engr.uconn.edu:8080/MNM/SMS_SearchServlet)。

[0101] 如本文所用,“工程化的”蛋白质是指在蛋白质中包括一个或多个修饰以获得所希望特性的蛋白质。示例性修饰包括但不限于插入、缺失、取代或与另一结构域或蛋白质融合。本披露的工程化的蛋白包括工程化的Cas9蛋白。

[0102] 在一些实施例中,工程化的蛋白是由野生型蛋白产生的。如本文所用,“野生型”蛋白质或核酸是天然存在的未修饰的蛋白质或核酸。例如,野生型Cas9蛋白可以从生物体酿脓链球菌中分离,并且可包括SEQ ID NO:1的氨基酸序列。野生型与“突变体”形成对比,“突变体”在蛋白质或核酸的氨基酸和/或核苷酸序列中包括一个或多个修饰。例如,酿脓链球菌Cas9的突变体可以包括SEQ ID NO:2的氨基酸序列,该变体相对于野生型酿脓链球菌Cas9(SEQ ID NO:1)具有单个氨基酸取代。

[0103] 当用于多肽或蛋白质时,术语“降解(degrade或degradation)”通常是指经由通常称为蛋白水解的过程将蛋白质分解成较小的肽片段或单个氨基酸。蛋白质的细胞内降解可以在溶酶体或蛋白酶体中实现。除了诸如本文所述的选择性分子伴侣介导的自噬途径的途径外,溶酶体降解通常是非选择性过程。在溶酶体降解中,胞质蛋白被内吞到溶酶体中进行降解。蛋白酶体降解通常是选择性的,其中要降解的蛋白用泛素标记。关于蛋白酶体蛋白降解途径的综述,参见,例如,Ciechanover,Cell[细胞]79(1):13-21(1994);Hasselgren等人,Ann Surg[外科年鉴]225(3):307-316(1997);Collins等人,Cell[细胞]169(5):792-806(2017)。通常,蛋白质的降解速率与其在细胞中的功能和生化特征有关。例如,具有富含脯氨酸、谷氨酸、丝氨酸和苏氨酸的区段的蛋白质(有时称为PEST蛋白质)具有短的半衰期(参见,例如,Voet&Voet,Biochemistry[生物化学]第二版John Wiley&Sons[约翰威利父子公司],第1010-1014页(1995),将其通过引用整体并入)。影响蛋白质降解速率的其他因素包括:谷氨酰胺和天冬酰胺的脱氨基速率;半胱氨酸、组氨酸和蛋氨酸的氧化速率;不存在稳定配体;存在连接的碳水化合物或磷酸基团;存在游离 α -氨基;蛋白质的电荷;以及蛋白

质的柔韧性和稳定性(例如,参见Creighton“Chapter 10-Degradation” in *Proteins: Structures and Molecular Properties* 2nd ed. [蛋白质:结构和分子性质第二版中“第10章-降解”] W H Freeman and Company, 第463-473页(1993),其通过引用整体并入)。测量蛋白质降解速率的方法包括,例如氨基酸同位素脉冲追踪(例如,在细胞培养物或SILAC中用氨基酸进行稳定的同位素标记)、合成后的放射性标记或报告子依赖性方法,例如整体蛋白质稳定性分析(GPSP),其利用例如GFP作为报告蛋白(参见,例如Yewdell等人, *Cell Biol Int* [国际细胞生物学] 35 (5) : 457-462 (2011))。测量蛋白质降解速率的另一种方法是通过使用例如免疫印迹的密度测定分析在不同时间点定量细胞中蛋白质的量,随时间绘制蛋白质水平,以及从蛋白质水平对时间图确定降解速率。本领域技术人员可以选择测定蛋白质降解速率的方法。

[0104] 如本文所用,术语“序列相似性”或“相似性百分比%”是指核酸序列或氨基酸序列之间的同一性或一致程度。如本文所用,“序列相似性”是指如下的核酸序列,其中一个或多个核苷酸碱基的改变导致一个或多个氨基酸的取代,但不影响该DNA序列编码的蛋白质的功能性质。“序列相似性”还指核酸的修饰,诸如基本上不影响所得转录物的功能性质的一个或多个核苷酸碱基的缺失或插入。因此,应当理解,本披露不仅仅涵盖特定的示例性序列。进行核苷酸碱基取代的方法以及确定所编码产物的生物活性保留的方法是已知的。

[0105] 此外,技术人员认识到,本披露涵盖的相似序列也由其严格条件下与本文示例的序列杂交的能力来定义。本披露的相似核酸序列是DNA序列与本文所披露的核酸的DNA序列具有至少70%、至少80%、至少90%、至少95%或至少99%相同的那些核酸。本披露的相似核酸序列是DNA序列与本文所披露的核酸的DNA序列具有约70%、至少约70%、约75%、至少约75%、约80%、至少约80%、约85%、至少约85%、约90%、至少约90%、约95%、至少约95%、约99%、至少约99%或约100%相同的那些核酸。

[0106] 如本文所用,“序列相似性”是指两个或更多个氨基酸序列,其中大于约40%的氨基酸是相同的,或者大于约60%的氨基酸是功能上相同的。功能相同或功能相似的氨基酸具有化学上相似的侧链。例如,可以根据功能相似性按照以下方式对氨基酸进行分组:

[0107] 带正电的侧链:Arg、His、Lys;

[0108] 带负电的侧链:Asn、Glu;

[0109] 极性、不带电的侧链:Ser、Thr、Asn、Gln;

[0110] 疏水性侧链:Ala、Val、Ile、Leu、Met、Phe、Tyr、Trp;

[0111] 其他:Cys、Gly、Pro。

[0112] 在一些实施例中,本披露的相似氨基酸序列具有至少40%、至少50%、至少60%、至少70%、至少80%、至少90%或至少99%相同的氨基酸。

[0113] 在一些实施例中,本披露的相似氨基酸序列具有至少60%、至少70%、至少80%、至少90%或至少95%功能相同的氨基酸。在一些实施例中,本披露的相似氨基酸序列具有约40%、至少约40%、约45%、至少约45%、约50%、至少约50%、约55%、至少约55%、约60%、至少约60%、约65%、至少约65%、约70%、至少约70%、约75%、至少约75%、约80%、至少约80%、约85%、至少约85%、约90%、至少约90%、约95%、至少约95%、约97%、至少约97%、约98%、至少约98%、约99%、至少约99%或约100%相同的氨基酸。

[0114] 在一些实施例中,本披露的相似氨基酸序列具有约60%、至少约60%、约65%、至

少约65%、约70%、至少约70%、约75%、至少约75%、约80%、至少约80%、约85%、至少约85%、约90%、至少约90%、约95%、至少约95%、约97%、至少约97%、约98%、至少约98%、约99%、至少约99%或约100%功能相同的氨基酸。

[0115] 如本文所用,术语“相同蛋白质”是指具有与参考蛋白质基本上相似的结构或氨基酸序列的蛋白质,其执行与参考蛋白质相同的生化功能,并且可包括因在氨基酸序列中的一个或多个位点处取代或缺失一个或多个氨基酸而不同于参考蛋白质的蛋白质,即至少约60%、至少约60%、约65%、至少约65%、约70%、至少约70%、约75%、至少约75%、约80%、至少约80%、约85%、至少约85%、约90%、至少约90%、约95%、至少约95%、约97%、至少约97%、约98%、至少约98%、约99%、至少约99%或约100%相同氨基酸的缺失。一方面,“相同蛋白质”是指具有与参考蛋白质相同的氨基酸序列的蛋白质。

[0116] 使用本领域常规方法,例如BLAST、MUSCLE、Clustal(包括ClustalW和ClustalX)和T-Coffee(包括变体,诸如像M-Coffee、R-Coffee和Espresso),通过序列比对可确定序列相似性。

[0117] 在核酸序列或氨基酸序列的背景下,术语“序列同一性”或“同一性百分比%”是指当该序列在指定的比较窗上比对时,比较的序列中相同的残基的百分比。在一些实施例中,仅将两个或更多个序列的特定部分进行比对以确定序列同一性。在一些实施例中,仅将两个或更多个序列的特定结构域进行比对以确定序列相似性。比较窗可以是至少10至超过1000个残基、至少20至约1000个残基或至少50至500个残基的区段,在其中可以比对和比较这些序列。用于确定序列同一性的比对方法是熟知的,并且可以使用公开可用的数据库(诸如BLAST)来进行。当指氨基酸序列时,“同一性百分比”或“同一性%”可以通过本领域已知的方法确定。例如,在一些实施例中,使用Karlin和Altschul,Proc Nat Acad Sci USA[美国国家科学院院刊]87:2264-2268(1990)的算法,按Karlin和Altschul,Proc Nat Acad Sci USA[美国国家科学院院刊]90:5873-5877(1993)所改良的,确定两个氨基酸序列的“同一性百分比”。这种算法被并入BLAST程序中,例如,Altschul等人,Journal of Molecular Biology[分子生物学杂志],215:403-410(1990)中描述的BLAST+或NBLAST和XBLAST程序。可采用诸如像XBLAST程序等的程序(分数=50、字长=3),执行BLAST蛋白质搜索,以获得与本披露的蛋白质分子同源的氨基酸序列。在两个序列之间存在空位的情况下,可以利用如Altschul等人,Nucleic Acids Research[核酸研究]25(17):3389-3402(1997)中描述的空位BLAST程序。当利用BLAST程序和空位BLAST程序时,可以使用相应程序(例如,XBLAST和NBLAST)的缺省参数。

[0118] 在一些实施例中,多肽或核酸分子与参考多肽或核酸分子(或参考多肽或核酸分子的片段)分别具有70%、至少70%、75%、至少75%、80%、至少80%、85%、至少85%、90%、至少90%、95%、至少95%、97%、至少97%、98%、至少98%、99%或至少99%或100%的序列同一性。在一些实施例中,多肽或核酸分子与参考多肽或核酸分子(或参考多肽或核酸分子的片段)分别具有约70%、至少约70%、约75%、至少约75%、约80%、至少约80%、约85%、至少约85%、约90%、至少约90%、约95%、至少约95%、约97%、至少约97%、约98%、至少约98%、约99%、至少约99%或与约100%的序列同一性。CRISPR-Cas系统的概述

[0119] CRISPR相关蛋白9(Cas9)是在细菌中发现的II型CRISPR适应性免疫系统的RNA指导的核酸内切酶,该细菌包括但不限于酿脓链球菌、嗜热链球菌、金黄色葡萄球菌和脑膜炎

奈瑟氏球菌等细菌。有关CRISPR-Cas9系统的概述,参见例如Sander等人,Nature Biotechnology[自然生物技术]32:347-355(2014)。通常,CRISPR或CRISPR-Cas系统的特征在于促进CRISPR复合物在靶序列的位点形成的元件,该CRISPR复合物包括指导多核苷酸和Cas9核酸内切酶(在本文中可互换地称为“Cas9蛋白”或“Cas9核酸酶”)。在天然存在的CRISPR-Cas系统中,将外源DNA导入CRISPR阵列中,然后产生带有与外源DNA位点互补的“前间区序列”区域的crRNA(CRISPR-RNA)。crRNA与tracrRNA(也由CRISPR系统编码)杂交,并且这对RNA与Cas9核酸酶缔合。crRNA/tracrRNA/Cas9复合物识别并裂解带有前间区序列的外源DNA。

[0120] 在一些实施例中,本披露提供了工程化的CRISPR-Cas系统。在一些实施例中,工程化的CRISPR-Cas系统包括工程化的Cas9蛋白,该蛋白包括相对于野生型Cas9的一个或多个修饰。在一些实施例中,工程化的Cas9蛋白包括野生型Cas9中不存在的一个或多个基序。导入野生型Cas9的一个或多个基序可以称为“工程化的”基序。在一些实施例中,Cas9蛋白中的一个或多个工程化的基序是分子伴侣介导的自噬(CMA)基序。

[0121] 在一些实施例中,工程化的CRISPR-Cas系统包括工程化的指导多核苷酸,该多核苷酸包括相对于野生型crRNA和/或tracrRNA的一个或多个修饰。在一些实施例中,工程化的CRISPR-Cas系统利用crRNA和tracrRNA序列的一部分(即单个指导多核苷酸)之间的融合。因此,在这种情况下,在Cas9和单个指导多核苷酸之间形成复合物。单个指导多核苷酸与Cas9形成复合物以介导靶序列的裂解,该靶序列与指导多核苷酸的前(5')20个核苷酸(即,指导多核苷酸的指导序列部分)互补,并且紧邻前间区序列邻近基序(PAM)序列。在其他实施例中,工程化的CRISPR-Cas系统包括包含tracrRNA序列的单独多核苷酸,即tracrRNA不是包括指导序列的指导多核苷酸的一部分。在这种情况下,在Cas9、指导多核苷酸和tracrRNA之间形成复合物。在一些实施例中,指导多核苷酸的tracrRNA组分激活Cas9蛋白。在一些实施例中,Cas9蛋白的激活使Cas9的核酸酶活性激活或增加。在一些实施例中,Cas9蛋白直到其与crRNA和tracrRNA形成复合物才有活性。

[0122] Cas9核酸内切酶在前间区序列邻近基序(PAM)上游在靶序列处产生双链DNA断裂。双链断裂的修复可能导致在双链断裂位点处的插入或缺失。在一些实施例中,使用细胞的内源DNA修复途径将目的序列插入靶序列。内源性DNA修复途径包括非同源末端连接(NHEJ)途径、微同源介导的末端连接(MMEJ)途径和同源定向修复(HDR)途径。NHEJ、MMEJ和HDR途径可修复双链DNA断裂。在NHEJ中,修复DNA中的断裂不需要同源模板。NHEJ修复可能容易出错,但是当DNA断裂包括相容的突出端时,错误会减少。NHEJ和MMEJ是在机理上截然不同的DNA修复途径,它们各自涉及DNA修复酶的不同子集。与在一些情况下可能精确或在一些情况下容易出错的NHEJ不同,MMEJ总是容易出错,并且会导致在修复位点处的缺失和插入。MMEJ相关的缺失归因于双链断裂两侧的微同源性(2-10个碱基对)。相比之下,HDR需要同源模板来直接进行修复,但是HDR修复典型地具有高保真度,并且不易出错。在一些实施例中,利用NHEJ和MMEJ修复的容易出错的性质在靶序列中引入非特异性核苷酸取代。

[0123] 如本文所述,一些CRISPR-Cas系统可能具有不期望的脱靶活性或脱靶基因组编辑。在基因组编辑的上下文中使用的“脱靶”是指非特异性和非预期的遗传修饰,这与“靶上”相反,后者是指在预期基因座处的修饰。当例如Cas9核酸酶未结合其预期的靶序列(即与指导多核苷酸上的指导序列互补的基因组序列)时,可能会导致脱靶修饰,这可能是由同

源序列和/或错配容许性引起。脱靶修饰可以包括但不限于非预期的点突变、缺失、插入、倒位和易位。在一些实施例中,与野生型Cas9蛋白相比,本披露的工程化的Cas9蛋白具有降低的脱靶活性。在一些实施例中,与野生型Cas9蛋白相比,本披露的工程化的Cas9蛋白的脱靶活性降低至少约50%。在一些实施例中,与野生型Cas9相比,本披露的工程化的Cas9蛋白的脱靶活性降低至少约60%、至少约70%、至少约80%、至少约90%、至少约95%或至少约100%。脱靶修饰可以使用例如靶向测序、外显子组测序、全基因组测序、BLESS(直接原位断裂标记、链霉亲和素富集和下一代测序)、GUIDE-seq(通过测序对DSB进行全基因组、无偏鉴定)、LAM-HTGTS(线性扩增介导的高通量全基因组易位测序)和Digenome-seq(体外Cas9消化的全基因组测序)。脱靶修饰的检测和定量方法描述于例如Zhang等人, *Mol Ther Nucleic Acids*[分子治疗核酸]4:e264(2014);和Zischewski等人, *Biotechnol Adv*[生物技术进展]35:95-104(2017)中。

[0124] Cas9蛋白

[0125] 在一些实施例中,Cas9蛋白衍生自以下物种:酿脓链球菌、嗜热链球菌、停乳链球菌、变形链球菌、英诺克李斯特菌、金黄色葡萄球菌或肺炎克雷伯氏菌。在一些实施例中,术语Cas9是指包括酿脓链球菌Cas9蛋白的氨基酸序列(SEQ ID NO:1)的多肽。在一些实施例中,术语Cas9是指包括嗜热链球菌Cas9蛋白的氨基酸序列(SEQ ID NO:17)的多肽。在一些实施例中,术语Cas9是指包括停乳链球菌Cas9蛋白的氨基酸序列(SEQ ID NO:18)的多肽。在一些实施例中,术语Cas9是指包括变形链球菌Cas9蛋白的氨基酸序列(SEQ ID NO:19)的多肽。在一些实施例中,术语Cas9是指包括英诺克李斯特菌Cas9蛋白的氨基酸序列(SEQ ID NO:20)的多肽。在一些实施例中,术语Cas9是指包括金黄色葡萄球菌Cas9蛋白的氨基酸序列(SEQ ID NO:21)的多肽。在一些实施例中,术语Cas9是指包括肺炎克雷伯氏菌Cas9蛋白的氨基酸序列(SEQ ID NO:22)的多肽。

[0126] 在一些实施例中,术语Cas9是指包括SEQ ID NO:1的多肽。在一些实施例中,Cas9蛋白与SEQ ID NO:1的氨基酸序列具有至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或约100%同一性。在一些实施例中,Cas9是由与SEQ ID NO:3具有至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或约100%同一性的多核苷酸序列编码的多肽。

[0127] 在一些实施例中,术语Cas9是指能够产生粘性末端的Cas9。如本文所用,术语“粘性末端(cohesive end)”、“交错末端”或“粘性末端(sticky end)”是指具有长度不等的链的核酸片段。与“平端”相反,粘性末端是通过对核酸(典型地为DNA)进行交错切割而产生的。黏性或粘性末端具有突出的单链的链(这些链具有不成对的核苷酸)或突出端,例如,3'或5'突出端。每个突出端可以与另一个互补的突出端进行退火以形成碱基对。两个互补的粘性末端可以经由相互作用诸如氢键结合一起退火。退火的粘性末端的稳定性取决于成对的突出端的解链温度。两个互补的粘性末端可以通过化学或酶促连接(例如,通过DNA连接酶)连接在一起。

[0128] 在一些实施例中,术语Cas9是指具有改变的功能的Cas9变体,例如Cas9杂合蛋白。例如,Cas9或具有无活性的DNA裂解结构域的Cas9蛋白的结合结构域可以用作经由指导多核苷酸与希望的靶序列特异性地结合的结合结构域。可以将结合结构域(即无活性的Cas9)

融合或缀合至裂解结构域(例如核酸内切酶FokI的裂解结构域),以产生工程化杂交核酸酶。Cas9-FokI杂合蛋白进一步描述于例如美国专利公开号2015/0071899和Guilinger等人,Nature Biotechnology[自然生物技术]32:577-582(2014)中。工程化的杂合核酸酶的其他实例描述于例如Wah等人,Proc Nat Acad Sci[美国国家科学院院刊]95:10564-10569(1996);Li等人,Nucl Acids Res[核酸研究]39(1):359-372(2011);和Kim等人,Proc Nat Acad Sci[美国国家科学院院刊]93:1156-1160(1996)中。

[0129] Cpf1(着丝粒和启动子因子1)也是II型CRISPR系统的RNA指导的核酸酶。Cpf1产生粘性末端。CRISPR/Cpf1系统类似于CRISPR/Cas9系统。但是,Cas9和Cpf1之间存在一些差异。与Cas9不同,Cpf1不使用tracrRNA。Cpf1蛋白识别与Cas9不同的PAM序列,并且Cpf1在与Cas9不同的位点处裂解。Cas9在与PAM相邻的序列裂解,而Cpf1在远离PAM的序列裂解。Cpf1蛋白进一步描述于例如外国专利公开GB 1506509.7、美国专利号9,580,701、美国专利公开2016/0208243和Zetsche等人,Cell[细胞]163(3):759-771(2015)中。根据本披露,可以使用在功能上类似于Cpf1的酶。因此,在一些实施例中,本披露提供了包括本文所述的氨基酸修饰的重组Cpf1蛋白。

[0130] 一些野生型或天然存在的Cas9蛋白(例如来自酿脓链球菌的Cas9蛋白)具有六个结构域:Rec1、Rec2、桥螺旋(Bridge Helix,BH)、PAM相互作用(PI)、HNH、和RuvC。Rec1结构域负责结合指导多核苷酸。当结合靶序列时,BH结构域负责引发裂解活性。PI结构域赋予PAM特异性,并负责引发与靶序列的结合。HNH和RuvC结构域是切割DNA的核酸酶结构域。对Cas9蛋白质的结构研究揭示,蛋白质具有识别叶(“REC叶”),其包括BH、Rec1和Rec2结构域;和核酸酶叶(“NUC叶”),其包括RuvC(分为RuvC I、RuvC II和RuvC III子结构域)、HNH和PI结构域。参见图3和4。可以使用基于蛋白质氨基酸序列的结构域结构预测工具来鉴定蛋白质结构域,例如SMART(Letunic等人,Nucleic Acids Research[核酸研究](2017),doi:10.1093/nar/gkx922)、PANDA(Wang等人,Scientific Reports[科学报告]8:3484(2018))或InterPro(Finn等人,Nucleic Acids Research[核酸研究](2017),doi:10.1093/nar/gkw1107)。还可以基于蛋白质结构(例如通过视觉检查)或通过使用诸如PUU(Holm等人,Proteins[蛋白质]19(3):256-268(1994))、RigidFinder(Abyzov等人,Proteins[蛋白质]78(2):309-324(2010))或PiSQRD(Aleksiev等人,Bioinformatics[生物信息学]25(20):2743-2744(2009))的算法来鉴定蛋白质结构域。基于结构表征的Cas9结构域的鉴定描述于例如Jinek等人,Science[科学]337:816-821(2012);Nishimasu等人,Cell[细胞]156(5):935-949(2014);Anders等人,Nature[自然]513:569-573(2014);和Sternberg等人,Nature[自然]507(7490):62-67(2014)中。

[0131] 在一些实施例中,本披露的Cas9蛋白包括与SEQ ID NO:5的氨基酸序列具有至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或约100%同一性的REC叶。在一些实施例中,本披露的Cas9蛋白包括与SEQ ID NO:6-7的氨基酸序列具有至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或约100%同一性的NUC叶。

[0132] 在一些实施例中,本披露的Cas9蛋白包括与SEQ ID NO:8的氨基酸序列具有至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或约100%同一性的BH结构域。在一些实施例中,本披露的Cas9蛋白包括与SEQ ID

N0:9-10的氨基酸序列具有至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或约100%同一性的Rec1结构域。在一些实施例中,本披露的Cas9蛋白包括与SEQ ID N0:11的氨基酸序列具有至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或约100%同一性的Rec2结构域。在一些实施例中,本披露的Cas9蛋白包括与SEQ ID N0:12-14的氨基酸序列具有至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或约100%同一性的RuvC结构域。在一些实施例中,本披露的Cas9蛋白包括与SEQ ID N0:15的氨基酸序列具有至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或约100%同一性的HNH结构域。在一些实施例中,本披露的Cas9蛋白包括与SEQ ID N0:16的氨基酸序列具有至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或约100%同一性的PI结构域。

[0133] Cas9蛋白的结构研究(例如晶体结构)可揭示表面暴露的蛋白质的区域。如本文所用,“表面暴露的区域”是指可接近周围环境的蛋白质区域,即,蛋白质的外“表面”上的区域。类似地,“表面暴露的残基”包括蛋白质的氨基酸残基,这些残基在表面暴露的区域中。表面暴露的残基与“埋藏”残基相反,该“埋藏”残基向内朝向蛋白质中心,并形成“埋藏区”,其不能被周围环境接近。蛋白质上的表面暴露的残基可能在与其它分子(例如其他蛋白质或细胞结构)的相互作用中起重要作用。因此,在一些实施例中,蛋白质(例如Cas9蛋白)上的某些残基以一种构象状态被表面暴露,而不以不同的构象状态被表面暴露。例如,Cas9蛋白在结合至指导RNA时可发生构象变化,使得Cas9蛋白中先前未暴露的区域在指导RNA结合时变为表面暴露的,或反之亦然(参见例如,Fagerlund等人,Proc Natl Acad Sci[美国国家科学院院刊]114(26):E5211-E5128(2017))。

[0134] 表面暴露的残基也可能决定蛋白质的物理性质并限制蛋白质的折叠结构。当在诸如PyMOL(pymol.org)或Swiss PDB Viewer(spdbv.vital-it.ch)的程序中查看蛋白质晶体结构时,可以确定表面暴露的残基。表面暴露的残基也可以使用诸如NACCESS(bioinf.manchester.ac.uk/naccess)的程序来计算。蛋白质的表面暴露的残基也可以通过计算预测来确定,例如,当晶体结构不可用时。蛋白质序列中表面暴露的残基的计算预测工具包括,例如SARpred(Garg等人,Proteins[蛋白质]61:318-24(2005))、JOY程序包中的PSA/TEM(Mizuguchi等人,Bioinformatics[生物信息学]14:617-623(1998))和RSARF(caps.ncbs.res.in/download/pugal/RSARF)。

[0135] 分子伴侣介导的自噬概述

[0136] “分子伴侣介导的自噬”或CMA是指选择性蛋白降解过程,该过程涉及对胞质蛋白质的分子伴侣依赖性选择,然后将蛋白靶向溶酶体并将它们跨过溶酶体膜易位以进行降解。CMA的示例性分子伴侣蛋白是70kD的热休克同源蛋白或HSC70。“内体微自噬”或eMI是指与CMA类似的蛋白质降解过程,不同之处在于eMI选择性地将包括KFERQ基序或KFERQ样基序的蛋白质靶向晚期内体而不是溶酶体进行降解。与CMA一样,HSC70也是eMI的分子伴侣蛋白。参见例如Kaushik等人,Trends Cell Biol[细胞生物学趋势]22(8):407-417(2012); Tekirdag等人,J Biol Chem[生物化学杂志]293:5414-5424(2018);和Pereira等人,Int J Cell Biol[国际细胞生物学杂志]2012(4):931956(2012)。

[0137] 本文所称的“KFERQ基序”是五肽序列:Lys-Phe-Glu-Arg-Gln (SEQ ID NO:24)。本文所称的“KFERQ样基序”是与KFERQ在生物化学上相似或在生物化学上相关的基序。如本文所述,在生物化学上相似或在生物化学上相关的基序可以包括在功能上等同的氨基酸残基。因此,KFERQ基序可以是具有以下参数的任何五肽:一个或两个带正电荷的残基(例如,Lys或Arg);一个或两个大体积疏水性残基(例如,Phe,Ile,Leu或Val);带负电荷的残基(例如,Asp或Glu);和位于五肽任一侧的Gln或Asn。参见例如,Dice等人,Trends Biochem Sci [生物化学科学的趋势]15(8):305-309(1990);和Kaushik等人,Trends Cell Biol [细胞生物学趋势]22(8):407-417(2012)。KFERQ样基序的实例包括但不限于表1中列出的基序。

[0138] 表1.KFERQ样基序

	基序的氨基酸序	SEQ ID NO.
	KFERQ	24
	RKVEQ	25
[0139]	QDLKF	26
	QRFFE	27
	NRVVD	28
	QRDKV	29
	QKILD	30
	QKKEL	31
	QFREL	32
	IKLDQ	33
	DVVRQ	34
[0140]	QRIVE	35
	VKELQ	36
	QKVFD	37
	QELLR	38
	VDKLN	39
	RIKEN	40
	NKKFE	41

[0141] 包括至少一个KFERQ基序或KFERQ样基序的蛋白质可以被CMA或eMI的组分识别。因此,在一些实施例中,KFERQ基序或KFERQ样基序是分子伴侣介导的自噬(CMA)靶基序。在一些实施例中,KFERQ基序或KFERQ样基序是内体微自噬(eMI)靶基序。不受特定理论的束缚,出于说明本披露的目的,在本文中描述了CMA和eMI,应理解,KFERQ基序或KFERQ样基序可以用作其他蛋白质降解途径的靶标,以及其他共有序列或基序(不同于本文所述的KFERQ基序或KFERQ样基序)可为CMA或eMI靶基序。

[0142] HSC70识别并结合至蛋白质上的CMA或eMI靶基序,例如KFERQ基序或KFERQ样基序,以形成分子伴侣蛋白复合物。然后,分子伴侣蛋白复合物与2A型溶酶体相关膜蛋白(LAMP-2A)受体结合。蛋白质展开,这触发了LAMP-2A的多聚化。随后,未折叠的蛋白质经由LAMP-2A跨过溶酶体膜易位,最后转运的蛋白被降解。参见例如Kaushik等人,Trends Cell Biol [细胞生物学趋势]22(8):407-417(2012)。

[0143] 重组Cas9蛋白

[0144] 与野生型Cas9相比,本披露的重组Cas9蛋白是功能性Cas9核酸酶,并且具有降低

的脱靶修饰。“功能性Cas9核酸酶”意指重组Cas9蛋白具有与野生型Cas9蛋白至少大约相同水平的核酸酶活性,如通过Cas9活性测定所测量的。“功能性Cas9核酸酶”也意指重组Cas9具有与野生型Cas9蛋白大约相同水平的靶上修饰(即,基因组编辑效率),如通过Cas9效率测定所测量的。

[0145] 在一些实施例中,本披露的重组Cas9蛋白具有野生型Cas9蛋白的核酸酶活性的至少约70%、至少约75%、至少约80%、至少约85%、至少约90%、至少约95%或至少约100%,如通过Cas9活性测定所测量的。在一些实施例中,本披露的重组Cas9蛋白具有比野生型Cas9蛋白更高的核酸酶活性,如通过Cas9活性测定所测量的。Cas9活性测定的非限制性实例包括T7核酸内切酶I测定和SURVEYOR测定(综述于Vouillot等人,G3 (Bethesda) 5 (3): 407-415 (2015)中)。在一些实施例中,本披露的重组Cas9蛋白具有野生型Cas9蛋白的靶上修饰的至少约70%、至少约75%、至少约80%、至少约85%、至少约90%、至少约95%或至少约100%,如通过Cas9效率测定所测量的。在一些实施例中,本披露的重组Cas9蛋白具有比野生型Cas9蛋白更高的靶上修饰,如通过Cas9效率测定所测量的。Cas9效率测定的非限制性实例包括错配检测测定和基于测序的测定(综述于Zischewski等人,Biotechnol Adv[生物技术进展]35:95-104(2017)中)。

[0146] 在一些实施例中,本披露提供了包括工程化的KFERQ基序或KFERQ样基序的重组Cas9蛋白。

[0147] 如本文所述,KFERQ基序或KFERQ样基序被CMA或eMI的组分识别。因此,在一些实施例中,包括工程化的KFERQ基序或KFERQ样基序的Cas9蛋白被CMA或eMI的组分识别。在一些实施例中,KFERQ基序或KFERQ样基序是SEQ ID NO:24-41中的任一个。因此,在一些实施例中,KFERQ基序或KFERQ样基序是KFERQ(SEQ ID NO:24)、RKVEQ(SEQ ID NO:25)、QDLKF(SEQ ID NO:26)、QRFFE(SEQ ID NO:27)、NRVVD(SEQ ID NO:28)、QRDKV(SEQ ID NO:29)、QKILD(SEQ ID NO:30)、QKKEL(SEQ ID NO:31)、QFREL(SEQ ID NO:32)、IKLDQ(SEQ ID NO:33)、DVVRQ(SEQ ID NO:34)、QRIVE(SEQ ID NO:35)、VKELQ(SEQ ID NO:36)、QKVFD(SEQ ID NO:37)、QELLR(SEQ ID NO:38)、VDKLN(SEQ ID NO:39)、RIKEN(SEQ ID NO:40)或NKKFE(SEQ ID NO:41)。在一些实施例中,工程化的KFERQ基序或KFERQ样基序是VDKLN(SEQ ID NO:39)。

[0148] 在一些实施例中,重组Cas9蛋白包括具有氨基酸序列KFERQ(SEQ ID NO:24)的工程化的KFERQ基序或KFERQ样基序。在一些实施例中,重组Cas9蛋白包括具有氨基酸序列RKVEQ(SEQ ID NO:25)的工程化的KFERQ基序或KFERQ样基序。在一些实施例中,重组Cas9蛋白包括具有氨基酸序列QDLKF(SEQ ID NO:26)的工程化的KFERQ基序或KFERQ样基序。在一些实施例中,重组Cas9蛋白包括具有氨基酸序列QRFFE(SEQ ID NO:27)的工程化的KFERQ基序或KFERQ样基序。在一些实施例中,重组Cas9蛋白包括具有氨基酸序列NRVVD(SEQ ID NO:28)的工程化的KFERQ基序或KFERQ样基序。在一些实施例中,重组Cas9蛋白包括具有氨基酸序列QRDKV(SEQ ID NO:29)的工程化的KFERQ基序或KFERQ样基序。

[0149] 在一些实施例中,重组Cas9蛋白包括具有氨基酸序列QKILD(SEQ ID NO:30)的工程化的KFERQ基序或KFERQ样基序。在一些实施例中,重组Cas9蛋白包括具有氨基酸序列QKKEL(SEQ ID NO:31)的工程化的KFERQ基序或KFERQ样基序。在一些实施例中,重组Cas9蛋白包括具有氨基酸序列QFREL(SEQ ID NO:32)的工程化的KFERQ基序或KFERQ样基序。在一些实施例中,重组Cas9蛋白包括具有氨基酸序列IKLDQ(SEQ ID NO:33)的工程化的KFERQ基

序或KFERQ样基序。在一些实施例中,重组Cas9蛋白包括具有氨基酸序列DVVRQ(SEQ ID NO:34)的工程化的KFERQ基序或KFERQ样基序。在一些实施例中,重组Cas9蛋白包括具有氨基酸序列QRIVE(SEQ ID NO:35)的工程化的KFERQ基序或KFERQ样基序。在一些实施例中,重组Cas9蛋白包括具有氨基酸序列VKELQ(SEQ ID NO:36)的工程化的KFERQ基序或KFERQ样基序。

[0150] 在一些实施例中,重组Cas9蛋白包括具有氨基酸序列QKVFD(SEQ ID NO:37)的工程化的KFERQ基序或KFERQ样基序。在一些实施例中,重组Cas9蛋白包括具有氨基酸序列QELLR(SEQ ID NO:38)的工程化的KFERQ基序或KFERQ样基序。在一些实施例中,重组Cas9蛋白包括具有氨基酸序列VDKLN(SEQ ID NO:39)的工程化的KFERQ基序或KFERQ样基序。在一些实施例中,重组Cas9蛋白包括具有氨基酸序列RIKEN(SEQ ID NO:40)的工程化的KFERQ基序或KFERQ样基序。在一些实施例中,重组Cas9蛋白包括具有氨基酸序列NKKFE(SEQ ID NO:41)的工程化的KFERQ基序或KFERQ样基序。

[0151] 在一些实施例中,工程化的KFERQ基序或KFERQ样基序在SEQ ID NO:1的第一个氨基酸残基之前。在一些实施例中,工程化的KFERQ基序或KFERQ样基序在SEQ ID NO:1的氨基酸残基1至100之间。在一些实施例中,工程化的KFERQ基序或KFERQ样基序在SEQ ID NO:1的氨基酸残基100至300之间。在一些实施例中,工程化的KFERQ基序或KFERQ样基序在SEQ ID NO:1的氨基酸残基300至700之间。在一些实施例中,工程化的KFERQ基序或KFERQ样基序在SEQ ID NO:1的氨基酸残基700至900之间。在一些实施例中,工程化的KFERQ基序或KFERQ样基序在SEQ ID NO:1的氨基酸残基900至1100之间。在一些实施例中,工程化的KFERQ基序或KFERQ样基序在SEQ ID NO:1的氨基酸残基1100至1300之间。在一些实施例中,工程化的KFERQ基序或KFERQ样基序在SEQ ID NO:1的最后一个氨基酸残基之后。

[0152] 在一些实施例中,工程化的KFERQ基序或KFERQ样基序在Cas9蛋白的REC叶中。在一些实施例中,Cas9蛋白的REC叶包括BH结构域、Rec1结构域和Rec2结构域。在一些实施例中,REC叶具有SEQ ID NO:5的氨基酸序列。在一些实施例中,工程化的KFERQ基序或KFERQ样基序在REC叶的Rec1结构域中。在一些实施例中,Rec1结构域具有SEQ ID NO:9-10的氨基酸序列。在一些实施例中,工程化的KFERQ基序或KFERQ样基序在REC叶的Rec2结构域中。在一些实施例中,Rec2结构域具有SEQ ID NO:11的氨基酸序列。在一些实施例中,工程化的KFERQ基序或KFERQ样基序在REC叶的BH结构域中。在一些实施例中,BH结构域具有SEQ ID NO:8的氨基酸序列。

[0153] 在一些实施例中,工程化的KFERQ基序或KFERQ样基序在Cas9蛋白的NUC叶中。在一些实施例中,Cas9蛋白的NUC叶包括RuvC结构域、HNH结构域和PI结构域。在一些实施例中,NUC叶具有SEQ ID NO:6-7的氨基酸序列。在一些实施例中,工程化的KFERQ基序或KFERQ样基序在Cas9蛋白的RuvC结构域、HNH结构域和PI结构域中。在一些实施例中,工程化的KFERQ基序或KFERQ样基序在RuvC结构域中。在一些实施例中,RuvC结构域具有SEQ ID NO:12-14的氨基酸序列。在一些实施例中,工程化的KFERQ基序或KFERQ样基序在HNH结构域中。在一些实施例中,HNH结构域具有SEQ ID NO:15的氨基酸序列。在一些实施例中,工程化的KFERQ基序或KFERQ样基序在PI结构域中。在一些实施例中,PI结构域具有SEQ ID NO:16的氨基酸序列。

[0154] 在一些实施例中,包括工程化的KFERQ基序或KFERQ样基序的Cas9蛋白包括与SEQ

ID NO:5的氨基酸序列具有至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或约100%同一性的叶。在一些实施例中,包括工程化的KFERQ基序或KFERQ样基序的Cas9蛋白包括与SEQ ID NO:6-7的氨基酸序列具有至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或约100%同一性的叶。

[0155] 在一些实施例中,包括工程化的KFERQ基序或KFERQ样基序的Cas9蛋白包括与SEQ ID NO:8的氨基酸序列具有至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或约100%同一性的BH结构域。在一些实施例中,包括工程化的KFERQ基序或KFERQ样基序的Cas9蛋白包括与SEQ ID NO:9-10的氨基酸序列具有至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或约100%同一性的Rec1结构域。在一些实施例中,包括工程化的KFERQ基序或KFERQ样基序的Cas9蛋白包括与SEQ ID NO:11的氨基酸序列具有至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或约100%同一性的Rec2结构域。

[0156] 在一些实施例中,包括工程化的KFERQ基序或KFERQ样基序的Cas9蛋白包括与SEQ ID NO:12-14的氨基酸序列具有至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或约100%同一性的RuvC结构域。在一些实施例中,包括工程化的KFERQ基序或KFERQ样基序的Cas9蛋白包括与SEQ ID NO:15的氨基酸序列具有至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或约100%同一性的HNH结构域。在一些实施例中,包括工程化的KFERQ基序或KFERQ样基序的Cas9蛋白包括与SEQ ID NO:16的氨基酸序列具有至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或约100%同一性的PI结构域。

[0157] 在一些实施例中,工程化的KFERQ基序或KFERQ样基序在重组Cas9蛋白的表面暴露区域中。如本文所述,表面暴露的区域是指可接近周围环境的Cas9蛋白区域,例如,可接近蛋白降解途径的组分。在一些实施例中,重组Cas9蛋白的表面暴露区域在Cas9蛋白的REC叶中。在一些实施例中,重组Cas9蛋白的表面暴露区域在Cas9蛋白的NUC叶中。在一些实施例中,重组Cas9蛋白的表面暴露区域在Cas9蛋白的Rec1结构域、Rec2结构域、BH结构域、RuvC结构域、HNH结构域或PI结构域中。在一些实施例中,重组Cas9蛋白的表面暴露区域在Cas9蛋白的氨基酸残基150与250之间。

[0158] 在一些实施例中,工程化的KFERQ基序或KFERQ样基序位于重组Cas9蛋白的N末端或C末端。如本文所述,N末端是蛋白质或多肽的“起始”,而C末端是蛋白质或多肽的“末端”。因此,在一些实施例中,KFERQ基序或KFERQ样基序位于Cas9蛋白的N末端“起始”。在一些实施例中,KFERQ基序或KFERQ样基序位于Cas9蛋白的C末端“末端”。在一些实施例中,向蛋白质的N末端或C末端添加工程化的基序不会影响蛋白质的折叠、结构或动力学。在一些实施例中,Cas9的N末端是表面暴露的。在一些实施例中,Cas9的C末端是表面暴露的。

[0159] 在一些实施例中,本披露提供了包括野生型Cas9蛋白的一个或多个氨基酸修饰的重组Cas9蛋白,这些修饰将分子伴侣介导的自噬(CMA)靶基序或内体微自噬(eMI)靶基序引入该Cas9蛋白中,其中该重组Cas9蛋白在体内的降解比野生型Cas9蛋白或不包括CMA或eMI

靶基序的Cas9蛋白快至少20%。在一些实施例中,重组Cas9蛋白质在体内的降解比野生型Cas9蛋白质或不包括CMA或eMI靶基序的Cas9蛋白质快至少50%。在一些实施例中,重组Cas9蛋白质在体内的降解比野生型Cas9蛋白质或不包括CMA或eMI靶基序的Cas9蛋白质快至少80%。

[0160] 在一些实施例中,本披露提供了包括野生型Cas9蛋白的一个或多个氨基酸修饰的重组Cas9蛋白,其中该重组Cas9蛋白包括CMA靶基序或eMI靶基序。

[0161] 如本文所述,含有CMA基序或eMI基序的蛋白质是CMA或eMI蛋白降解途径的靶标。因此,在一些实施例中,包括一个或多个引入CMA或eMI靶基序的氨基酸修饰的本披露的重组Cas9蛋白被靶向用于经由CMA或eMI的蛋白质降解。同样,在一些实施例中,包括CMA靶基序或eMI靶基序的本披露的重组Cas9蛋白被靶向用于经由CMA或eMI的蛋白质降解。

[0162] 在一些实施例中,包括CMA或eMI靶基序的重组Cas9蛋白在体内的降解比野生型Cas9蛋白质或不包括CMA或eMI靶基序的Cas9蛋白质快至少20%、快至少30%、快至少40%、快至少50%、快至少60%、快至少70%、快至少80%、快至少90%、快至少100%、快至少150%、快至少200%、快至少500%,如通过免疫印迹或GFP报告基因测定所测量的。在一些实施例中,如果同一细胞表达:(a) 包括一个或多个引入CMA或eMI靶基序的氨基酸修饰的重组Cas9,和(b) 野生型Cas9,则重组Cas9被完全降解,而至少50%的野生型Cas9仍保留在细胞内。类似地,在一些实施例中,如果同一细胞表达:(a) 包括一个或多个引入CMA或eMI靶基序的氨基酸修饰的重组Cas9,和(b) 不包括CMA或eMI靶基序的Cas9蛋白,则重组Cas9被完全降解,而至少50%的不包括CMA或eMI靶基序的Cas9蛋白仍保留在细胞内。在一些实施例中,重组Cas9被完全降解,而至少60%、至少70%、至少80%、至少90%或至少100%的野生型Cas9或不包括CMA或eMI靶基序的Cas9蛋白仍保留在细胞内。在一些实施例中,重组Cas9在引入细胞的12小时内、24小时内、36小时内、48小时内或72小时内被完全降解。如本文的实施例中所用,“完全降解”是指低于GFP报告基因测定或免疫印迹的检测水平的蛋白质。如本文所述,测量蛋白质降解速率的方法包括,例如氨基酸同位素脉冲追踪(例如,在细胞培养物或SILAC中用氨基酸进行稳定的同位素标记)、合成后的放射性标记或报告子依赖性方法,例如整体蛋白质稳定性分析(GPSP),其利用例如GFP作为报告蛋白(参见,例如Yewde11等人,Cell Biol Int [国际细胞生物学]35(5):457-462(2011))。在一些实施例中,Cas9蛋白的降解速率是通过使用例如免疫印迹的密度测定分析在不同时间点定量细胞中Cas9蛋白质的量,随时间绘制蛋白质水平,以及从Cas9蛋白质水平对时间图确定降解速率来测量。

[0163] 在一些实施例中,重组Cas9中的一个或多个氨基酸修饰包括SEQ ID NO:1的位置F185处的突变。在一些实施例中,突变是F185N。在一些实施例中,重组Cas9中的一个或多个氨基酸修饰包括在SEQ ID NO:1的位置A547和I548处的突变。在一些实施例中,突变是A547E和I548L。在一些实施例中,重组Cas9中的一个或多个氨基酸修饰包括在SEQ ID NO:1的位置T560和V561处的突变。在一些实施例中,突变是T560E和V561Q。在一些实施例中,重组Cas9中的一个或多个氨基酸修饰包括SEQ ID NO:1的位置D829和I830处的突变。在一些实施例中,突变是D829L和I830R。在一些实施例中,重组Cas9中的一个或多个氨基酸修饰包括在SEQ ID NO:1的位置L1087和S1088处的突变。在一些实施例中,突变是L1087E和S1088Q。在一些实施例中,重组Cas9中的一个或多个氨基酸修饰包括在SEQ ID NO:1的位置P1199和K1200处的突变。在一些实施例中,突变是P1199D和K1200Q。

[0164] 在一些实施例中,重组Cas9中的一个或多个氨基酸修饰包括本文所述的任何突变的组合。在一些实施例中,重组Cas9中的一个或多个氨基酸修饰包括选自F185N、A547E、I548L及其组合的突变。在一些实施例中,重组Cas9中的一个或多个氨基酸修饰包括选自F185N、T560E、V561Q及其组合的突变。在一些实施例中,重组Cas9中的一个或多个氨基酸修饰包括选自F185N、D829L、I830R及其组合的突变。在一些实施例中,重组Cas9中的一个或多个氨基酸修饰包括选自F185N、L1087E、S1088Q及其组合的突变。在一些实施例中,重组Cas9中的一个或多个氨基酸修饰包括选自F185N、P1199D、K1200Q及其组合的突变。

[0165] 在一些实施例中,重组Cas9中的一个或多个氨基酸修饰包括选自A547E、I548L、T560E、V561Q及其组合的突变。在一些实施例中,重组Cas9中的一个或多个氨基酸修饰包括选自A547E、I548L、D829L、I830R及其组合的突变。在一些实施例中,重组Cas9中的一个或多个氨基酸修饰包括选自A547E、I548L、L1087E、S1088Q及其组合的突变。在一些实施例中,重组Cas9中的一个或多个氨基酸修饰包括选自A547E、I548L、P1199D、K1200Q及其组合的突变。在一些实施例中,重组Cas9中的一个或多个氨基酸修饰包括选自T560E、V561Q、D829L、I830R及其组合的突变。在一些实施例中,重组Cas9中的一个或多个氨基酸修饰包括选自T560E、V561Q、L1087E、S1088Q及其组合的突变。在一些实施例中,重组Cas9中的一个或多个氨基酸修饰包括选自T560E、V561Q、P1199D、K1200Q及其组合的突变。在一些实施例中,重组Cas9中的一个或多个氨基酸修饰包括选自D829L、I830R、L1087E、S1088Q及其组合的突变。在一些实施例中,重组Cas9中的一个或多个氨基酸修饰包括选自D829L、I830R、P1199D、K1200Q及其组合的突变。在一些实施例中,重组Cas9中的一个或多个氨基酸修饰包括选自L1087E、S1088Q、P1199D、K1200Q及其组合的突变。在一些实施例中,如本文所述,重组Cas9蛋白中的一个或多个氨基酸修饰导致一个或多个CMA靶基序或eMI靶基序。

[0166] 在一些实施例中,重组Cas9蛋白具有与SEQ ID NO:1具有至少50%同一性的氨基酸序列,并且包括本文所述的一个或多个氨基酸修饰。在一些实施例中,重组Cas9蛋白具有与SEQ ID NO:1具有至少60%同一性的氨基酸序列,并且包括本文所述的一个或多个氨基酸修饰。在一些实施例中,重组Cas9蛋白具有与SEQ ID NO:1具有至少70%同一性的氨基酸序列,并且包括本文所述的一个或多个氨基酸修饰。在一些实施例中,重组Cas9蛋白具有与SEQ ID NO:1具有至少80%同一性的氨基酸序列,并且包括本文所述的一个或多个氨基酸修饰。在一些实施例中,重组Cas9蛋白具有与SEQ ID NO:1具有至少90%同一性的氨基酸序列,并且包括本文所述的一个或多个氨基酸修饰。在一些实施例中,重组Cas9蛋白具有与SEQ ID NO:1具有至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、或至少99%同一性的氨基酸序列,并且包括本文所述的一个或多个氨基酸修饰。

[0167] 在一些实施例中,本披露提供了分离自酿脓链球菌的重组Cas9蛋白(SpCas9),该重组Cas9蛋白在SEQ ID NO:1的位置F185、A547、I548、T560、V561、D829、I830、L1087、S1088、P1199、K1200或其组合中的一个或多个处包括氨基酸修饰。

[0168] 在一些实施例中,重组SpCas9中的一个或多个氨基酸修饰包括在SEQ ID NO:1的位置F185处的突变。在一些实施例中,突变是F185N。在一些实施例中,重组SpCas9中的一个或多个氨基酸修饰包括在SEQ ID NO:1的位置A547和I548处的突变。在一些实施例中,突变是A547E和I548L。在一些实施例中,重组SpCas9中的一个或多个氨基酸修饰包括在SEQ ID

NO:1的位置T560和V561处的突变。在一些实施例中,突变是T560E和V561Q。在一些实施例中,重组SpCas9中的一个或多个氨基酸修饰包括在SEQ ID NO:1的位置D829和I830处的突变。在一些实施例中,突变是D829L和I830R。在一些实施例中,重组SpCas9中的一个或多个氨基酸修饰包括在SEQ ID NO:1的位置L1087和S1088处的突变。在一些实施例中,突变是L1087E和S1088Q。在一些实施例中,重组SpCas9中的一个或多个氨基酸修饰包括在SEQ ID NO:1的位置P1199和K1200处的突变。在一些实施例中,突变是P1199D和K1200Q。

[0169] 在一些实施例中,重组SpCas9中的一个或多个氨基酸修饰包括本文所述的任何突变的组合。在一些实施例中,重组SpCas9中的一个或多个氨基酸修饰包括选自F185N、A547E、I548L及其组合的突变。在一些实施例中,重组SpCas9中的一个或多个氨基酸修饰包括选自F185N、T560E、V561Q及其组合的突变。在一些实施例中,重组SpCas9中的一个或多个氨基酸修饰包括选自F185N、D829L、I830R及其组合的突变。在一些实施例中,重组SpCas9中的一个或多个氨基酸修饰包括选自F185N、L1087E、S1088Q及其组合的突变。在一些实施例中,重组SpCas9中的一个或多个氨基酸修饰包括选自F185N、P1199D、K1200Q及其组合的突变。

[0170] 在一些实施例中,重组SpCas9中的一个或多个氨基酸修饰包括选自A547E、I548L、T560E、V561Q及其组合的突变。在一些实施例中,重组SpCas9中的一个或多个氨基酸修饰包括选自A547E、I548L、D829L、I830R及其组合的突变。在一些实施例中,重组SpCas9中的一个或多个氨基酸修饰包括选自A547E、I548L、L1087E、S1088Q及其组合的突变。在一些实施例中,重组SpCas9中的一个或多个氨基酸修饰包括选自A547E、I548L、P1199D、K1200Q及其组合的突变。在一些实施例中,重组SpCas9中的一个或多个氨基酸修饰包括选自T560E、V561Q、D829L、I830R及其组合的突变。在一些实施例中,重组SpCas9中的一个或多个氨基酸修饰包括选自T560E、V561Q、L1087E、S1088Q及其组合的突变。在一些实施例中,重组SpCas9中的一个或多个氨基酸修饰包括选自T560E、V561Q、P1199D、K1200Q及其组合的突变。在一些实施例中,重组SpCas9中的一个或多个氨基酸修饰包括选自D829L、I830R、L1087E、S1088Q及其组合的突变。在一些实施例中,重组SpCas9中的一个或多个氨基酸修饰包括选自D829L、I830R、P1199D、K1200Q及其组合的突变。在一些实施例中,重组SpCas9中的一个或多个氨基酸修饰包括选自L1087E、S1088Q、P1199D、K1200Q及其组合的突变。在一些实施例中,重组SpCas9蛋白中的一个或多个氨基酸修饰导致一个或多个CMA靶基序或eMI靶基序。CMA靶基序和eMI靶基序如文中所述。

[0171] 在一些实施例中,重组SpCas9蛋白具有与SEQ ID NO:1具有至少50%同一性的氨基酸序列,并且包括本文所述的一个或多个氨基酸修饰。在一些实施例中,重组SpCas9蛋白具有与SEQ ID NO:1具有至少60%同一性的氨基酸序列,并且包括本文所述的一个或多个氨基酸修饰。在一些实施例中,重组SpCas9蛋白具有与SEQ ID NO:1具有至少70%同一性的氨基酸序列,并且包括本文所述的一个或多个氨基酸修饰。在一些实施例中,重组SpCas9蛋白具有与SEQ ID NO:1具有至少80%同一性的氨基酸序列,并且包括本文所述的一个或多个氨基酸修饰。在一些实施例中,重组SpCas9蛋白具有与SEQ ID NO:1具有至少90%同一性的氨基酸序列,并且包括本文所述的一个或多个氨基酸修饰。在一些实施例中,重组SpCas9蛋白具有与SEQ ID NO:1具有至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、或至少99%同一性的氨基酸序列,并且包括本文所述的一个或

多个氨基酸修饰。

[0172] 在一些实施例中,本披露提供了能够与70kD的热休克同源蛋白(HSC70)结合的重组Cas9蛋白。

[0173] 如本文所述,HSC70是CMA和eMI蛋白降解途径中的分子伴侣蛋白。HSC70与靶向降解的蛋白结合,并将蛋白质转运至溶酶体(在CMA的情况下)或晚期内体(在eMI的情况下)进行降解。因此,在一些实施例中,与对HSC70具有较低结合亲和力的蛋白相比,对HSC70具有较高结合亲和力的蛋白降解得更快。在一些实施例中,蛋白质对HSC70的结合能力和/或亲和力通过在蛋白质上存在CMA或eMI靶基序例如KFERQ基序或KFERQ样基序来确定。

[0174] 在一些实施例中,本披露的重组Cas9蛋白能够结合HSC70。如本文所述和例示,野生型Cas9或不包括KFERQ基序或KFERQ样基序的Cas9蛋白不结合HSC70。在一些实施例中,本披露的重组Cas9蛋白能够以比野生型Cas9蛋白或不包括KFERQ基序或KFERQ样基序的Cas9蛋白高至少50%、至少60%、至少70%、至少80%、至少90%或至少100%的亲和力结合HSC70。测定蛋白质之间的结合亲和力的方法是本领域已知的,并且包括例如生化方法,例如:共免疫沉淀、双分子荧光互补、亲和电泳、下拉测定、噬菌体展示、体内交联、串联亲和纯化、交联后进行质谱和邻近连接测定;生物物理方法,例如:生物层干涉法、动态光散射、表面等离子体共振、荧光共振能量转移和等温滴定量热法;和/或遗传方法,例如:酵母双杂交筛选和细菌双杂交筛选。有关测量结合亲和力和检测蛋白质相互作用的方法的概述,参见例如Meyerkord和Fu,Protein-Protein Interactions:Methods and Applications 2ndEd. [蛋白质-蛋白质相互作用:方法和应用第二版]2015,Humana出版社。在一些实施例中,本披露的重组Cas9能够在与HSC70孵育一段时间后被HSC70抗体检测到,而野生型Cas9或不包括KFERQ基序或KFERQ样基序的Cas9蛋白在孵育同一段时间后被HSC70抗体检测不到。

[0175] 在一些实施例中,HSC70与重组Cas9之间的结合亲和力比HSC70与野生型Cas9或不包括KFERQ基序或KFERQ样基序的Cas9蛋白之间的结合亲和力高至少2倍、高至少3倍、高至少4倍、高至少5倍、高至少6倍、高至少7倍、高至少8倍、高至少9倍、高至少10倍、高至少20倍、高至少30倍、高至少40倍、高至少50倍、高至少60倍、高至少70倍、高至少80倍、高至少90倍、高至少100倍、高至少500倍、或高至少1000倍。

[0176] 在一些实施例中,本披露提供了分离自酿脓链球菌的重组Cas9蛋白(SpCas9),该重组Cas9蛋白在SEQ ID NO:1的氨基酸位置185处包括工程化的KFERQ基序或KFERQ样基序。

[0177] 如本文所述,SEQ ID NO:1包括来自酿脓链球菌的野生型Cas9蛋白(SpCas9)的氨基酸序列。SEQ ID NO:1的氨基酸位置185在对应于SpCas9的Rec2结构域的区域中。在一些实施例中,修饰SEQ ID NO:1的位置185处的氨基酸残基以产生KFERQ基序或KFERQ样基序。在一些实施例中,SEQ ID NO:1的位置185处的KFERQ样基序是VDKLN。在一些实施例中,重组SpCas9蛋白在SEQ ID NO:1的位置185处包括突变。在一些实施例中,突变是F185N。

[0178] 在一些实施例中,本披露的重组Cas9进一步在SEQ ID NO:1的位置D10和/或H840处包括突变。在野生型Cas9的位置D10和/或H840处的突变产生具有切口酶活性的Cas9,在本文中也称为“Cas9切口酶”。Cas9切口酶仅能够裂解双链DNA的一条链(即,“切口”该DNA)。Cas9切口酶描述于例如Cho等人,Genome Res[基因组研究]24:132-141(2013)中。在一些实施例中,本披露的重组Cas9蛋白进一步在SEQ ID NO:1的氨基酸位置D10处包括突变。在一些实施例中,本披露的重组Cas9蛋白进一步在SEQ ID NO:1的氨基酸位置H840处包括突变。

在一些实施例中,本披露的重组Cas9蛋白进一步包括SEQ ID NO:1的氨基酸位置D10处的突变和氨基酸位置H840处的突变。在一些实施例中,位置D10处的突变是D10A。在一些实施例中,位置D10处的突变是D10N。在一些实施例中,位置H840处的突变是H840A。在一些实施例中,位置H840处的突变是H840N。在一些实施例中,位置H840处的突变是H840Y。在一些实施例中,重组Cas9蛋白具有F185N突变和D10A突变。在一些实施例中,重组Cas9蛋白具有F185N突变和D10N突变。在一些实施例中,重组Cas9蛋白具有F185N突变和H840A突变。在一些实施例中,重组Cas9蛋白具有F185N突变和H840N突变。在一些实施例中,重组Cas9蛋白具有F185N突变和H840Y突变。

[0179] 在一些实施例中,重组Cas9蛋白具有F185N突变、D10A突变和H840A突变。在一些实施例中,重组Cas9蛋白具有F185N突变、D10A突变和H840N突变。在一些实施例中,重组Cas9蛋白具有F185N突变、D10A突变和H840Y突变。在一些实施例中,重组Cas9蛋白具有F185N突变、D10N突变和H840A突变。在一些实施例中,重组Cas9蛋白具有F185N突变、D10N突变和H840N突变。在一些实施例中,重组Cas9蛋白具有F185N突变、D10N突变和H840Y突变。在一些实施例中,重组Cas9蛋白具有F185N突变、D10A突变和H840A突变。在一些实施例中,重组Cas9蛋白具有F185N突变、D10A突变和H840N突变。在一些实施例中,重组Cas9蛋白具有F185N突变、D10A突变和H840Y突变。在一些实施例中,重组Cas9蛋白具有F185N突变、D10N突变和H840A突变。在一些实施例中,重组Cas9蛋白具有F185N突变、D10N突变和H840N突变。在一些实施例中,重组Cas9蛋白具有F185N突变、D10N突变和H840Y突变。

[0180] 在一些实施例中,本披露的重组Cas9蛋白产生粘性末端。如本文所述,粘性末端是指具有不等长度的链的核酸片段。在一些实施例中,产生粘性末端的重组Cas9蛋白是重组Cas9-FokI杂合蛋白。在一些实施例中,重组Cas9-FokI杂合蛋白在Cas9中具有F185N突变。在一些实施例中,重组Cas9-FokI杂合蛋白在Cas9中具有F185N突变和D10A突变。在一些实施例中,重组Cas9-FokI杂合蛋白在Cas9中具有F185N突变和D10N突变。在一些实施例中,重组Cas9-FokI杂合蛋白在Cas9中具有F185N突变和H840A突变。在一些实施例中,重组Cas9-FokI杂合蛋白在Cas9中具有F185N突变和H840N突变。在一些实施例中,重组Cas9-FokI杂合蛋白在Cas9中具有F185N突变和H840Y突变。

[0181] 在一些实施例中,重组Cas9-FokI杂合蛋白在Cas9中具有F185N突变、D10A突变和H840A突变。在一些实施例中,重组Cas9-FokI杂合蛋白在Cas9中具有F185N突变、D10A突变和H840N突变。在一些实施例中,重组Cas9-FokI杂合蛋白在Cas9中具有F185N突变、D10A突变和H840Y突变。在一些实施例中,重组Cas9-FokI杂合蛋白在Cas9中具有F185N突变、D10N突变和H840A突变。在一些实施例中,重组Cas9-FokI杂合蛋白在Cas9中具有F185N突变、D10N突变和H840N突变。在一些实施例中,重组Cas9-FokI杂合蛋白在Cas9中具有F185N突变、D10N突变和H840Y突变。在一些实施例中,重组Cas9-FokI杂合蛋白在Cas9中具有F185N突变、D10A突变和H840A突变。在一些实施例中,重组Cas9-FokI杂合蛋白在Cas9中具有F185N突变、D10A突变和H840N突变。在一些实施例中,重组Cas9-FokI杂合蛋白在Cas9中具有F185N突变、D10A突变和H840Y突变。在一些实施例中,重组Cas9-FokI杂合蛋白在Cas9中具有F185N突变、D10N突变和H840A突变。在一些实施例中,重组Cas9-FokI杂合蛋白在Cas9中具有F185N突变、D10N突变和H840N突变。在一些实施例中,重组Cas9-FokI杂合蛋白在Cas9中具有F185N突变、D10N突变和H840Y突变。

[0182] 在一些实施例中,产生粘性末端的重组Cas9蛋白在能够产生粘性末端的野生型Cas9蛋白上包括工程化的KFERQ基序或KFERQ样基序。在一些实施例中,能够产生粘性末端的野生型Cas9蛋白分离自新凶手弗朗西斯菌(*Francisella novicida*) (FnCas9) (SEQ ID NO:23)。在一些实施例中,重组Cas9蛋白具有与SEQ ID NO:23具有至少约90%序列同一性的氨基酸序列,并且包括本文所述的工程化的KFERQ基序或KFERQ样基序。在一些实施例中,重组Cas9蛋白具有与SEQ ID NO:23具有至少约90%序列同一性的氨基酸序列,并且包括CMA靶基序或eMI靶基序。在一些实施例中,重组Cas9具有与SEQ ID NO:23具有至少约90%序列同一性的氨基酸序列,并且能够以比野生型Cas9更高的亲和力结合HSC70。在一些实施例中,重组Cas9具有与SEQ ID NO:23具有至少约90%序列同一性的氨基酸序列,并且比野生型Cas9降解得更快。

[0183] 在一些实施例中,本披露的重组Cas9包括一个或多个核定位信号。“核定位信号”或“核定位序列”(NLS)是对蛋白质“加标签”以通过核转运导入细胞核的氨基酸序列,即具有NLS的蛋白质被转运至细胞核。典型地,NLS包括暴露在蛋白质表面的带正电荷的Lys或Arg残基。示例性核定位序列包括但不限于来自以下的NLS:SV40大T抗原、核质蛋白、EGL-13、c-Myc和TUS蛋白。在一些实施例中,该NLS包括序列PKKKRKV (SEQ ID NO:42)。在一些实施例中,该NLS包括序列AVKRPAATKKAGQAKKKKLD (SEQ ID NO:43)。在一些实施例中,该NLS包括序列PAAKRVKLD (SEQ ID NO:44)。在一些实施例中,该NLS包括序列MSRRRKNPTKLSENAKKLAKEVEN (SEQ ID NO:45)。在一些实施例中,该NLS包括序列KLKIKRPVK (SEQ ID NO:46)。其他核定位序列包括但不限于hnRNP A1的酸性M9结构域、酵母转录抑制子Mat α 2中的序列KIPIK (SEQ ID NO:47)和PY-NLS。

[0184] 核苷酸

[0185] 在一些实施例中,本披露提供了编码本文所述的重组Cas9蛋白的多核苷酸序列。在一些实施例中,本披露提供了编码包括工程化的KFERQ基序或KFERQ样基序的重组Cas9的多核苷酸序列。在一些实施例中,本披露提供了编码重组Cas9蛋白的多核苷酸序列,该重组Cas9蛋白包括野生型Cas9蛋白的一个或多个氨基酸修饰,该修饰将CMA靶基序或eMI靶基序引入Cas9蛋白中,其中该重组Cas9蛋白在体内的降解比野生型Cas9蛋白快至少20%。在一些实施例中,本披露提供了编码重组Cas9蛋白的多核苷酸序列,该重组Cas9蛋白包括野生型Cas9蛋白的一个或多个氨基酸修饰,其中该重组Cas9蛋白包括CMA靶基序或eMI靶基序。在一些实施例中,本披露提供了编码分离自酿脓链球菌的重组Cas9蛋白(SpCas9)的多核苷酸序列,该重组Cas9蛋白在SEQ ID NO:1的位置F185、A547、I548、T560、V561、D829、I830、L1087、S1088、P1199、K1200或其组合中的一个或多个处包括氨基酸修饰。在一些实施例中,本披露提供了编码能够结合HSC70的重组Cas9蛋白的多核苷酸序列。在一些实施例中,本披露提供了编码分离自酿脓链球菌的重组Cas9蛋白(SpCas9)的多核苷酸序列,该重组Cas9蛋白在SEQ ID NO:1的氨基酸位置185处包括工程化的KFERQ基序或KFERQ样基序。

[0186] 在一些实施例中,多核苷酸序列与SEQ ID NO:3具有至少50%序列同一性。在一些实施例中,多核苷酸序列与SEQ ID NO:3具有至少60%序列同一性。在一些实施例中,多核苷酸序列与SEQ ID NO:3具有至少70%序列同一性。在一些实施例中,多核苷酸序列与SEQ ID NO:3具有至少80%序列同一性。在一些实施例中,多核苷酸序列与SEQ ID NO:3具有至少90%序列同一性。在一些实施例中,多核苷酸序列与SEQ ID NO:3具有至少91%、至少

92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性。

[0187] 在一些实施例中,对编码重组Cas9的多核苷酸序列进行密码子优化以在真核细胞中表达。在一些实施例中,对编码stiCas9的多核苷酸序列进行密码子优化以在动物细胞中表达。在一些实施例中,对编码重组Cas9的多核苷酸序列进行密码子优化以在人类细胞中表达。在一些实施例中,对编码重组Cas9的多核苷酸序列进行密码子优化以在植物细胞中表达。密码子优化是对密码子进行调整来匹配表达宿主的tRNA丰度,以提高重组或异源蛋白质表达的产量和效率。密码子优化方法是本领域的常规方法,可以使用软件程序执行,诸如像集成DNA技术公司(Integrated DNA Technologies)的密码子优化工具、Entelechon的密码子使用表分析工具、GENEMAKER的Blue Heron软件、Aptagen的Gene Forge软件、DNA Builder软件、通用密码子使用分析软件、公开可用的OPTIMIZER软件以及金斯瑞的OptimumGene算法。

[0188] CRISPR-Cas系统

[0189] 在一些实施例中,本披露提供了非天然存在的CRISPR-Cas系统,该系统包括:本文提供的重组Cas9蛋白,和与重组Cas9蛋白形成复合物并包括指导序列的指导多核苷酸。

[0190] 在一些实施例中,非天然存在的CRISPR-Cas系统的重组Cas9蛋白包括工程化的KFERQ基序或KFERQ样基序。在一些实施例中,非天然存在的CRISPR-Cas系统的重组Cas9蛋白包括野生型Cas9蛋白的一个或多个氨基酸修饰,该修饰将CMA靶基序或eMI靶基序引入Cas9蛋白中,其中该重组Cas9蛋白在体内的降解比野生型Cas9蛋白或不包括CMA或eMI靶基序的Cas9蛋白快至少20%。在一些实施例中,非天然存在的CRISPR-Cas系统的重组Cas9蛋白包括野生型Cas9蛋白的一个或多个氨基酸修饰,其中该重组Cas9蛋白包括CMA靶基序或eMI靶基序。在一些实施例中,非天然存在的CRISPR-Cas系统的重组Cas9蛋白分离自酿脓链球菌(SpCas9),并且在SEQ ID NO:1的位置F185、A547、I548、T560、V561、D829、I830、L1087、S1088、P1199、K1200或其组合中的一个或多个处包括氨基酸修饰。在一些实施例中,非天然存在的CRISPR-Cas系统的重组Cas9蛋白能够结合HSC70。在一些实施例中,非天然存在的CRISPR-Cas系统的重组Cas9蛋白分离自酿脓链球菌(SpCas9),并且在SEQ ID NO:1的氨基酸位置185处包括工程化的KFERQ基序或KFERQ样基序。

[0191] 在一些实施例中,本披露提供了非天然存在的CRISPR-Cas系统,该系统包括:编码本文提供的重组Cas9蛋白的多核苷酸序列,以及与该重组Cas9蛋白形成复合物并包括指导序列的指导多核苷酸。

[0192] 在一些实施例中,本披露提供了非天然存在的CRISPR-Cas系统,该系统包括:可操作地连接至编码本文提供的重组Cas9蛋白的多核苷酸序列的调节元件,以及与该重组Cas9蛋白形成复合物并包括指导序列的指导多核苷酸。

[0193] 在一些实施例中,非天然存在的CRISPR-Cas系统的多核苷酸编码包括工程化的KFERQ基序或KFERQ样基序的重组Cas9。在一些实施例中,非天然存在的CRISPR-Cas系统的多核苷酸编码包括野生型Cas9蛋白的一个或多个氨基酸修饰的重组Cas9蛋白,该修饰将CMA靶基序或eMI靶基序引入Cas9蛋白中,其中该重组Cas9蛋白在体内的降解比野生型Cas9蛋白或不包括CMA或eMI靶基序的Cas9蛋白快至少20%。在一些实施例中,非天然存在的CRISPR-Cas系统的多核苷酸编码包括野生型Cas9蛋白的一个或多个氨基酸修饰的重组

Cas9蛋白,其中该重组Cas9蛋白包括CMA靶基序或eMI靶基序。在一些实施例中,非天然存在的CRISPR-Cas系统的多核苷酸编码分离自酿脓链球菌的重组Cas9蛋白(SpCas9),该重组Cas9蛋白在SEQ ID NO:1的位置F185、A547、I548、T560、V561、D829、I830、L1087、S1088、P1199、K1200或其组合中的一个或多个处包括氨基酸修饰。在一些实施例中,非天然存在的CRISPR-Cas系统的多核苷酸编码能够结合HSC70的重组Cas9蛋白。在一些实施例中,非天然存在的CRISPR-Cas系统的多核苷酸编码分离自酿脓链球菌的重组Cas9蛋白(SpCas9),该重组Cas9蛋白在SEQ ID NO:1的氨基酸位置185处包括工程化的KFERQ基序或KFERQ样基序。

[0194] 在一些实施例中,与编码重组Cas9蛋白的多核苷酸序列连接的调节元件是启动子。在一些实施例中,该调节元件是细菌启动子。在一些实施例中,该调节元件是病毒启动子。在一些实施例中,该调节元件是真核生物调节元件,即,真核生物启动子。在一些实施例中,该真核生物调节元件是哺乳动物启动子。

[0195] 在一些实施例中,非天然存在的CRISPR-Cas系统的指导多核苷酸是RNA分子。与CRISPR-Cas组分结合并将其靶向靶DNA内特定位置的RNA分子在本文中称为“指导RNA”、“gRNA”或“小指导RNA”,在本文中也可能称为“靶向DNA的RNA”。指导多核苷酸,例如指导RNA,包括至少两个核苷酸区段:至少一个“DNA结合区段”和至少一个“多肽结合区段”。“区段”是指分子的一个部分、节区段或区域,例如,指导多核苷酸分子的核苷酸的连续伸展段。除非另有明确定义,否则“区段”的定义不限于特定数目的总碱基对。

[0196] 在一些实施例中,指导多核苷酸的DNA结合区段(或“DNA靶向序列”)与细胞中的靶序列杂交。在一些实施例中,指导多核苷酸(例如,指导RNA)的DNA结合区段包括与靶DNA内的特定序列互补的多核苷酸序列。

[0197] 在一些实施例中,本披露的指导多核苷酸具有与细菌细胞中的靶序列杂交的指导序列。在该方法的一些实施例中,靶序列在细菌细胞中。在一些实施例中,该细菌细胞是实验室菌株。此类细胞的实例包括但不限于大肠杆菌、金黄色葡萄球菌、霍乱弧菌、肺炎链球菌、枯草芽孢杆菌、新月形柄杆菌、生殖支原体、费氏曲霉、集胞藻属、荧光假单胞菌、棕色固氮菌、天蓝色链霉菌。在一些实施例中,细菌细胞是用于制备食物和/或饮料的细菌。此类细胞的非限制性示例性属包括但不限于醋杆菌属、节杆菌属、芽孢杆菌属、双歧杆菌属、短状杆菌属、短杆菌属、肉杆菌属、棒杆菌属、肠球菌属、葡糖醋杆菌属、哈夫尼菌属、盐单胞菌属、考克氏菌属、乳杆菌属(包括耐酸乳杆菌、酸鱼乳杆菌、嗜酸乳酸杆菌、消化乳杆菌、短乳杆菌、布赫纳乳杆菌、干酪乳酸菌、弯曲乳杆菌、发酵乳杆菌、希氏乳杆菌(*L.hilgardii*)、詹氏乳杆菌(*L.jensenii*)、金氏乳杆菌(*L.kimchii*)、乳酸乳杆菌(*L.lactis*)、副干酪乳杆菌(*L.paracasei*)、植物乳杆菌和阪乳杆菌(*L.sakei*))、明串珠菌属、微杆菌属、小球菌属、丙酸杆菌属、魏斯氏菌属和发酵单胞菌属。

[0198] 在一些实施例中,本披露的指导多核苷酸具有与真核细胞中的靶序列杂交的指导序列。在一些实施例中,该真核细胞是动物或人类细胞。在一些实施例中,该真核细胞是人类或啮齿动物或牛细胞系或者细胞株。此类细胞/细胞系或细胞株的实例包括但不限于小鼠骨髓瘤(NS0)细胞系、中国仓鼠卵巢(CHO)细胞系、HT1080、H9、HepG2、MCF7、MDBK Jurkat、NIH3T3、PC12、BHK(幼仓鼠肾细胞)、VERO、SP2/0、YB2/0、Y0、C127、L细胞、COS(例如COS1和COS7)、QC1-3、HEK-293、VERO、PER.C6、HeLa、EB1、EB2、EB3、溶瘤或杂交瘤细胞系。在一些实施例中,该真核细胞是CHO细胞系。在一些实施例中,该真核细胞是CHO细胞。在一些实施例

中,该细胞是CHO-K1细胞、CHO-K1 SV细胞、DG44 CHO细胞、DUXB11 CHO细胞、CHOS、CHO GS敲除细胞、CHO FUT8 Gs敲除细胞、CHOZN或CHO衍生的细胞。CHO GS敲除细胞(例如,GSKO细胞)是例如CHO-K1 SV GS敲除细胞。CHO FUT8敲除细胞是例如POTELLIGENT CHOK1 SV(龙沙生物公司(Lonza Biologics, Inc.))。真核细胞也可以是禽细胞、细胞系或细胞株,例如EBX细胞、EB14、EB24、EB26、EB66或EBv13。

[0199] 在一些实施例中,该真核细胞是人类细胞。在一些实施例中,该人类细胞是干细胞。干细胞可以是例如多能干细胞,包括胚胎干细胞(ESC)、成年干细胞、诱导的多能干细胞(iPSC)、组织特异性干细胞(例如,造血干细胞)和间充质干细胞(MSC)。在一些实施例中,该人类细胞是本文描述的任何细胞的分化形式。在一些实施例中,该真核细胞是源自培养物中任何原代细胞的细胞。

[0200] 在一些实施例中,该真核细胞是肝细胞,诸如人肝细胞、动物肝细胞,或非实质细胞。例如,该真核细胞可以是可培养代谢合格的人肝细胞、可培养感应诱导合格的人肝细胞、可培养人肝细胞、悬浮合格的人肝细胞(包括10-供体和20-供体合并的肝细胞)、人肝库普弗细胞、人肝星状细胞、狗肝细胞(包括单个和合并的比格犬肝细胞)、小鼠肝细胞(包括CD-1和C57BI/6肝细胞)、大鼠肝细胞(包括Sprague-Dawley、Wistar Han和Wistar肝细胞)、猴肝细胞(包括食蟹猴或恒河猴肝细胞)、猫肝细胞(包括家养短毛猫肝细胞)和兔肝细胞(包括新西兰白兔肝细胞)。

[0201] 在一些实施例中,该真核细胞是植物细胞。例如,该植物细胞可以是诸如木薯、玉米、高粱、小麦或水稻的农作物的细胞。该植物细胞可以是藻类、树木或蔬菜的细胞。该植物细胞可以是单子叶植物或双子叶植物的细胞,或者可以是农作物或谷物植物、生产植物、水果或蔬菜的细胞。例如,该植物细胞可以是树木的细胞,该树木为例如柑橘属果树,诸如橘子树、葡萄柚树或柠檬树;桃树或油桃树;苹果树或梨树;坚果树,诸如杏仁树或胡桃树或开心果树;茄属植物,例如,马铃薯、芸苔属(Brassica)属植物、莴苣属(Lactuca)植物;菠菜属(Spinacia)植物;辣椒属(Capsicum)植物;棉花、烟草、芦笋、胡萝卜、卷心菜、西兰花、花椰菜、番茄、茄子、胡椒、莴苣、菠菜、草莓、蓝莓、覆盆子、黑莓、葡萄、咖啡、可可等。

[0202] 在一些实施例中,指导多核苷酸的指导序列为约5至约50个核苷酸。在一些实施例中,指导多核苷酸的指导序列为约6至约45个核苷酸。在一些实施例中,指导多核苷酸的指导序列为约7至约40个核苷酸。在一些实施例中,指导多核苷酸的指导序列为约8至约35个核苷酸。在一些实施例中,指导多核苷酸的指导序列为约9至约30个核苷酸。在一些实施例中,指导多核苷酸的指导序列为约10至约20个核苷酸。在一些实施例中,指导多核苷酸的指导序列为约12至约20个核苷酸。在一些实施例中,指导多核苷酸的指导序列为约14至约20个核苷酸。在一些实施例中,指导多核苷酸的指导序列为约16至约20个核苷酸。在一些实施例中,指导多核苷酸的指导序列为约18至约20个核苷酸。在一些实施例中,指导多核苷酸的指导序列为约5至约10个核苷酸。在一些实施例中,指导多核苷酸的指导序列为约6至约10个核苷酸。在一些实施例中,指导多核苷酸的指导序列为约7至约10个核苷酸。在一些实施例中,指导多核苷酸的指导序列为约8至约10个核苷酸。指导序列的长度可以由本领域技术人员使用指导序列设计工具来测定,例如CRISPR设计工具(Hsu等人,Nat Biotechnol[自然生物技术]31(9):827-832(2013))、ampliCan(Labun等人,bioRxiv 2018,doi:10.1101/249474)、CasFinder(Alach等人,bioRxiv 2014,doi:10.1101/005074)、CHOPCHOP(Labun等

人, *Nucleic Acids Res* [核酸研究] 2016, doi:10.1093/nar/gkw398) 等。

[0203] 在一些实施例中, 本披露的指导多核苷酸 (例如, 指导RNA) 包括多肽结合序列/区段。指导多核苷酸 (例如, 指导RNA) 的多肽结合区段 (或“蛋白结合序列”) 与本披露的Cas蛋白的多核苷酸结合结构域相互作用。此类多肽结合区段或序列是本领域技术人员已知的, 例如, 美国专利公开2014/0068797、2014/0273037、2014/0273226、2014/0295556、2014/0295557、2014/0349405、2015/0045546、2015/0071898、2015/0071899和2015/0071906中披露的那些, 这些公开的披露内容以其全文并入本文。在一些实施例中, 指导多核苷酸的多肽结合区段与Cas9结合。在一些实施例中, 指导多核苷酸的多肽结合区段与本文提供的重组Cas9蛋白结合。

[0204] 在一些实施例中, 指导多核苷酸为至少约10、15、20、25或30个核苷酸并且至多约25、30、35、40、45、50、60、70、80、90、100、110、120、130、140或150个核苷酸。在一些实施例中, 指导多核苷酸为约10至约150个核苷酸。在一些实施例中, 指导多核苷酸为约20至约120个核苷酸。在一些实施例中, 指导多核苷酸为约30至约100个核苷酸。在一些实施例中, 指导多核苷酸为约40至约80个核苷酸。在一些实施例中, 指导多核苷酸为约50至约60个核苷酸。在一些实施例中, 指导多核苷酸为约10至约35个核苷酸。在一些实施例中, 指导多核苷酸为约15至约30个核苷酸。在一些实施例中, 指导多核苷酸为约20至约25个核苷酸。

[0205] 可以将指导多核苷酸 (例如, 指导RNA) 作为分离的分子 (例如, RNA分子) 引入靶细胞中, 或使用包含编码指导多核苷酸 (例如, 指导RNA) 的DNA的表达载体将其引入细胞中。

[0206] 在一些实施例中, 将CRISPR-Cas系统的指导多核苷酸与同向重复序列连接。同向重复序列或DR序列是CRISPR基因座中的重复序列阵列, 由短伸展段的非重复序列 (间区序列) 隔开。该间区序列靶向靶序列上的前间区序列邻近基序 (PAM)。当转录CRISPR基因座的非编码部分 (即, 指导多核苷酸和tracrRNA), 转录物在DR序列上被裂解成多个短crRNA, 这些crRNA包含单个的间区序列, 这些间区序列将Cas9核酸酶引导至PAM。在一些实施例中, 该DR序列是RNA。在一些实施例中, 该DR序列由核酸编码。在一些实施例中, 该DR序列与指导多核苷酸连接。在一些实施例中, 该DR序列与指导多核苷酸的指导序列连接。在一些实施例中, 该DR序列包括二级结构。在一些实施例中, 该DR序列包括茎环结构。在一些实施例中, 该DR序列为10至20个核苷酸。在一些实施例中, 该DR序列为至少16个核苷酸。在一些实施例中, 该DR序列为至少16个核苷酸并且包括单个茎环。在一些实施例中, 该DR序列包括RNA适配体。在一些实施例中, 该DR中的二级结构或茎环被核酸酶识别以用于裂解。在一些实施例中, 该核酸酶是核糖核酸酶。在一些实施例中, 该核酸酶是RNA酶III。

[0207] 在一些实施例中, 本披露的CRISPR-Cas系统进一步包括tracrRNA。“tracrRNA”或反式激活CRISPR-RNA与前体crRNA或前体CRISPR-RNA形成RNA双链体, 然后被RNA特异性核糖核酸酶RNA酶III裂解, 形成crRNA/tracrRNA杂交体。在一些实施例中, 该指导RNA包括crRNA/tracrRNA杂交体。在一些实施例中, 该指导RNA的tracrRNA组分激活Cas9蛋白。在一些实施例中, 该CRISPR-cas系统的指导多核苷酸包括tracrRNA序列。在一些实施例中, CRISPR-Cas系统包括单独的多核苷酸, 该多核苷酸包括tracrRNA序列。

[0208] 在一些实施例中, 编码重组Cas9的多核苷酸和指导多核苷酸在单个载体上。在一些实施例中, 编码重组Cas9的多核苷酸、指导多核苷酸 (或可以转录成指导多核苷酸的核苷酸) 和tracrRNA在单个载体上。在一些实施例中, 编码重组Cas9的多核苷酸、指导多核苷酸

(或可以转录成指导多核苷酸的核苷酸)、tracrRNA和同向重复序列在单个载体上。在一些实施例中,该载体是表达载体。在一些实施例中,该载体是哺乳动物表达载体。在一些实施例中,该载体是人类表达载体。在一些实施例中,该载体是植物表达载体。

[0209] 在一些实施例中,编码重组Cas9和指导多核苷酸的多核苷酸是单个核酸分子。在一些实施例中,编码重组Cas9、指导多核苷酸和tracrRNA的多核苷酸是单个核酸分子。在一些实施例中,编码重组Cas9、指导多核苷酸、tracrRNA和同向重复序列的多核苷酸是单个核酸分子。在一些实施例中,该单个核酸分子是表达载体。在一些实施例中,该单个核酸分子是哺乳动物表达载体。在一些实施例中,该单个核酸分子是人类表达载体。在一些实施例中,该单个核酸分子是植物表达载体。

[0210] 在一些实施例中,重组Cas9和指导多核苷酸能够形成复合物。在一些实施例中,重组Cas9和指导多核苷酸的复合物在自然界中不存在。

[0211] 用于递送CRISPR-Cas系统的各种方法是本领域已知的。在一些实施例中,本披露的CRISPR-Cas系统由递送颗粒递送。递送颗粒是包括颗粒的生物递送系统或配制品。如本文所定义,“颗粒”是最大直径约100微米(μm)的实体。在一些实施例中,该颗粒的最大直径为约10 μm 。在一些实施例中,该颗粒的最大直径为约2000纳米(nm)。在一些实施例中,该颗粒的最大直径为约1000nm。在一些实施例中,该颗粒的最大直径为约900nm、约800nm、约700nm、约600nm、约500nm、约400nm、约300nm、约200nm或约100nm。在一些实施例中,该颗粒的直径为约25nm至约200nm。在一些实施例中,该颗粒的直径为约50nm至约150nm。在一些实施例中,该颗粒的直径为约75nm至约100nm。

[0212] 递送颗粒可以以任何形式提供,包括但不限于:固体、半固体、乳液或胶体颗粒。在一些实施例中,该递送颗粒是基于脂质的系统、脂质体、胶束、微囊泡、外泌体或基因枪。在一些实施例中,该递送颗粒包括CRISPR-Cas系统。在一些实施例中,该递送颗粒包括CRISPR-Cas系统,该CRISPR-Cas系统包括重组Cas9和指导多核苷酸。在一些实施例中,该递送颗粒包括CRISPR-Cas系统,该CRISPR-Cas系统包括重组Cas9和指导多核苷酸,其中重组Cas9和指导多核苷酸以复合物存在。在一些实施例中,该递送颗粒包括CRISPR-Cas系统,该CRISPR-Cas系统包括重组Cas9、指导多核苷酸和含有tracrRNA的多核苷酸。在一些实施例中,该递送颗粒包括CRISPR-Cas系统,该CRISPR-Cas系统包括重组Cas9、指导多核苷酸和tracrRNA。

[0213] 在一些实施例中,该递送颗粒进一步包括脂质、糖、金属或蛋白质。在一些实施例中,该递送颗粒是脂质包膜。例如,Su等人,Molecular Pharmacology[分子药理学]8(3):774-784(2011)中描述了使用脂质包膜或包括脂质的递送颗粒进行的mRNA递送。在一些实施例中,该递送颗粒是基于糖的颗粒,例如,GalNAc。在WO 2014/118272和Nair等人,J Am Chem Soc[美国化学学会杂志]136(49):16958-16961(2014)中描述了基于糖的颗粒。

[0214] 在一些实施例中,该递送颗粒是纳米颗粒。本披露涵盖的纳米颗粒可以以不同形式提供,例如,作为固体纳米颗粒(例如,金属,诸如银、金、铁、钛)、非金属、基于脂质的固体、聚合物、纳米颗粒或其组合。可以制备金属、电介质和半导体纳米颗粒以及混合结构(例如,核-壳纳米颗粒)。由半导体材料制成的纳米粒子如果足够小(典型地小于10nm),可以量化电子能级,则也可以将其标记为量子点。这样的纳米级颗粒在生物医学应用中用作药物运载体或显像剂,并且可以加以调整以适用于本披露中的类似用途。

[0215] 递送颗粒的制备在美国专利公开2011/0293703、2012/0251560和2013/0302401；以及美国专利号5,543,158、5,855,913、5,895,309、6,007,845和8,709,843中进一步描述。

[0216] 在一些实施例中，囊泡包括本披露的CRISPR-Cas系统。“囊泡”是细胞内具有被脂质双层包围的流体的小结构。在一些实施例中，本披露的CRISPR-Cas系统由囊泡递送。在一些实施例中，该囊泡包括重组Cas9和指导多核苷酸。在一些实施例中，该囊泡包括重组Cas9和指导多核苷酸，其中重组Cas9和指导多核苷酸以复合物存在。在一些实施例中，该囊泡包括CRISPR-Cas系统，该CRISPR-Cas系统包括重组Cas9、指导多核苷酸和含有tracrRNA的多核苷酸。在一些实施例中，该囊泡包括CRISPR-Cas系统，该CRISPR-Cas系统包括重组Cas9、指导多核苷酸和tracrRNA。

[0217] 在一些实施例中，包括重组Cas9和指导多核苷酸的囊泡是外泌体或脂质体。在一些实施例中，该囊泡是外泌体。在一些实施例中，该外泌体用于递送本披露的CRISPR-Cas系统。外泌体是内源性纳米囊泡（即，直径为约30nm至约100nm），可转运RNA和蛋白质，并且可将RNA递送至大脑和其他靶器官。例如，Alvarez-Erviti等人，*Nature Biotechnology* [自然生物学] 29:341 (2011)，El-Andaloussi等人，*Nature Protocols* [自然实验手册] 7:2112-2116 (2012)，以及Wahlgren等人，*Nucleic Acids Research* [核酸研究] 40(17):e130 (2012) 中描述了用于将内源性生物材料递送至靶器官的工程化的外泌体。

[0218] 在一些实施例中，包括stCas9和指导多核苷酸的囊泡是脂质体。在一些实施例中，该脂质体用于递送本披露的CRISPR-Cas系统。脂质体是具有至少一个脂质双层的球形囊泡结构，并且可以用作营养物和药物施用的媒介物。脂质体通常由磷脂（特别是磷脂酰胆碱）以及其他脂质（诸如蛋磷脂酰乙醇胺）组成。脂质体的类型包括但不限于多层囊泡、小单层囊泡、大单层囊泡和耳蜗囊泡。参见例如Spuch和Navarro，*Journal of Drug Delivery* [药物递送杂志]，文章号469679 (2011)。例如，Morrissey等人，*Nature Biotechnology* [自然生物技术] 23(8):1002-1007 (2005)，Zimmerman等人，*Nature Letters* [自然快报] 441:111-114 (2006)，以及Li等人，*Gene Therapy* [基因疗法] 19:775-780 (2012) 描述了用于递送诸如CRISPR-Cas组分的生物材料的脂质体。

[0219] 在一些实施例中，病毒载体包括本披露的CRISPR-Cas系统。在一些实施例中，本披露的CRISPR-Cas系统由病毒载体递送。在一些实施例中，病毒载体包括重组Cas9和指导多核苷酸。在一些实施例中，病毒载体包括重组Cas9和指导多核苷酸，其中该重组Cas9和指导多核苷酸以复合物存在。在一些实施例中，病毒载体包括CRISPR-Cas系统，该CRISPR-Cas系统包括重组Cas9、指导多核苷酸和含有tracrRNA的多核苷酸。在一些实施例中，病毒载体包括CRISPR-Cas系统，该CRISPR-Cas系统包括重组Cas9、指导多核苷酸和tracrRNA。在一些实施例中，该病毒载体是腺病毒、慢病毒或腺相关病毒载体。本文提供了病毒载体的实例。

[0220] 在一些实施例中，腺相关病毒(AAV)和/或慢病毒载体可以用作包括本文所述的CRISPR-Cas系统的元件的病毒载体。在本披露的一些实施例中，该Cas蛋白由病毒载体转导的细胞在细胞内表达。

[0221] 在一些实施例中，本披露的Cas蛋白和方法用于离体基因编辑，诸如CAR-T型疗法。这些实施例可能涉及对来自人类供体的细胞的修饰。在这些情况下，也可以使用病毒载体；但是，还有其他选择可以直接将Cas9蛋白（连同体外转录的指导RNA和供体DNA）转染到培养的细胞中。

[0222] 在一些实施例中,本披露的重组Cas9蛋白是包括一个或多个异源蛋白结构域(例如,除重组Cas9蛋白之外约或至少约1、2、3、4、5、6、7、8、9或10个或更多个结构域)的融合蛋白的一部分。Cas9融合蛋白可包括任何其他蛋白序列,以及任选地在任何两个结构域之间的接头序列。可以与重组Cas9蛋白融合的蛋白质结构域的实例包括但不限于:具有以下一项或多项活性的表位标签、报告基因序列和蛋白质结构域:甲基化酶活性、脱甲基酶活性、转录激活活性、转录抑制活性、转录释放因子活性、组蛋白修饰活性、RNA裂解活性和核酸结合活性。表位标签的非限制性实例包括:组氨酸(His)标签、V5标签、FLAG标签、流感病毒血凝素(HA)标签、Myc标签、VSV-G标签和硫氧还蛋白(Trx)标签。报告基因的实例包括但不限于谷胱甘肽-5-转移酶(GST)、辣根过氧化物酶(HRP)、氯霉素乙酰转移酶(CAT)、 β -半乳糖苷酶、 β -葡萄糖醛酸酶、萤光素酶、绿色荧光蛋白(GFP)、HcRed、DsRed、青色荧光蛋白(CFP)、黄色荧光蛋白(YFP)、自发荧光蛋白(包括蓝色荧光蛋白(BFP))和mCherry。在一些实施例中,重组Cas9蛋白与结合DNA分子或结合其他细胞分子的蛋白或蛋白片段融合,包括但不限于:麦芽糖结合蛋白(MBP)、S标签、Lex A DNA结合结构域(DBD)、GAL4 DNA结合结构域和单纯疱疹病毒(HSV)BP16蛋白。可形成包括Cas9蛋白的融合蛋白的一部分的其他结构域在美国专利公开2011/0059502中描述。在一些实施例中,加标签的重组Cas9蛋白用于鉴定靶序列的位置。

[0223] 在一些实施例中,重组Cas9蛋白可形成诱导型系统的组分。该系统的可诱导性质允许使用某种形式的能量对基因编辑或基因表达进行时空控制。这种能量形式可以包括但不限于:电磁辐射、声能、化学能和热能。诱导型系统的非限制性实例包括:四环素诱导型启动子(Tet-On或Tet-Off)、小分子双杂交转录激活系统(FKBP、ABA等)或光诱导系统(光敏色素、LOV结构域或隐色素)。在一些实施例中,该Cas9蛋白是光诱导型转录效应子(LITE)的一部分,以序列特异性方式指导转录活性的变化。光的成分可以包括Cas9蛋白、光响应性细胞色素异二聚体(例如,来自拟南芥(*Arabidopsis thaliana*))和转录激活/抑制结构域。诱导型DNA结合蛋白及其使用方法的其他实例在国际申请公开号WO 2014/018423和WO 2014/093635;美国专利号8,889,418和8,895,308;以及美国专利公开号2014/0186919、2014/0242700、2014/0273234和2014/0335620中提供;

[0224] 位点特异性修饰的方法

[0225] 在一些实施例中,本披露提供了在细胞基因组中的靶序列处提供位点特异性修饰的方法,该方法包括将本文所述的非天然存在的CRISPR-Cas系统引入细胞中。

[0226] 在该方法的一些实施例中,非天然存在的CRISPR-Cas系统的重组Cas9蛋白包括工程化的KFERQ基序或KFERQ样基序。在该方法的一些实施例中,非天然存在的CRISPR-Cas系统的重组Cas9蛋白包括野生型Cas9蛋白的一个或多个氨基酸修饰,该修饰将CMA靶基序或eMI靶基序引入Cas9蛋白中,其中该重组Cas9蛋白在体内的降解比野生型Cas9蛋白或不包括CMA或eMI靶基序的Cas9蛋白快至少20%。在该方法的一些实施例中,非天然存在的CRISPR-Cas系统的重组Cas9蛋白包括野生型Cas9蛋白的一个或多个氨基酸修饰,其中该重组Cas9蛋白包括CMA靶基序或eMI靶基序。在该方法的一些实施例中,非天然存在的CRISPR-Cas系统的重组Cas9蛋白分离自酿脓链球菌(*SpCas9*),并且在SEQ ID NO:1的位置F185、A547、I548、T560、V561、D829、I830、L1087、S1088、P1199、K1200或其组合中的一个或多个处包括氨基酸修饰。在该方法的一些实施例中,非天然存在的CRISPR-Cas系统的重组Cas9蛋

白能够结合HSC70。在该方法的一些实施例中,非天然存在的CRISPR-Cas系统的重组Cas9蛋白分离自酿脓链球菌(SpCas9),并且在SEQ ID NO:1的氨基酸位置185处包括工程化的KFERQ基序或KFERQ样基序。

[0227] 靶序列的修饰涵盖单核苷酸取代、多核苷酸取代、核酸的插入(即,敲入)和缺失(即,敲除)、移码突变和其他核酸修饰。

[0228] 在该方法的一些实施例中,该修饰是靶序列的至少一部分的缺失。可以在两个不同位点裂解靶序列并产生互补的粘性末端,并且这些互补的粘性末端可以重新连接,从而去除两个位点之间的序列部分。

[0229] 在该方法的一些实施例中,该修饰是靶序列的突变。位点特异性诱变可通过使用位点特异性核酸酶来实现的,该酶可促进含有目的突变的外源多核苷酸模板(也称为“供体多核苷酸”或“供体载体”)的同源重组。在一些实施例中,目的序列(SoI)包括目的突变。

[0230] 在该方法的一些实施例中,该修饰是将目的序列(SoI)插入靶序列。该SoI可以作为外源多核苷酸模板引入。在一些实施例中,该外源多核苷酸包括平末端。在一些实施例中,该外源多核苷酸模板包括粘性末端。在一些实施例中,该外源多核苷酸模板包括与靶序列中的粘性末端互补的粘性末端。

[0231] 外源多核苷酸模板可以具有任何合适的长度,诸如约或至少约10、15、20、25、50、75、100、150、200、250、500或1000或更多个核苷酸的长度。在一些实施例中,该外源多核苷酸模板与包括靶序列的多核苷酸的一部分互补。当进行最佳比对时,该外源多核苷酸模板与靶序列的一个或多个核苷酸(例如,约或至少约1、5、10、15、20、25、30、35、40、45、50、60、70、80、90或100或更多个核苷酸)重叠。在一些实施例中,当对外源多核苷酸模板和包括靶序列的多核苷酸进行最佳比对时,外源多核苷酸模板最接近的核苷酸在距离靶序列约1、5、10、15、20、25、50、75、100、200、300、400、500、100、1500、2000、2500、5000、10000个或更多个核苷酸内。

[0232] 在一些实施例中,该外源多核苷酸是DNA,例如,DNA质粒、细菌人工染色体(BAC)、酵母人工染色体(YAC)、病毒载体、单链或双链DNA的线性片段、寡核苷酸、PCR片段、裸核酸或与诸如脂质体的递送媒介物复合的核酸。

[0233] 在一些实施例中,使用细胞的内源DNA修复途径将该外源多核苷酸插入靶序列。内源DNA修复途径包括NHEJ、MMEJ和HDR,其各自在本文中描述。在修复过程中,可以将包括SoI的外源多核苷酸模板引入该靶序列。在一些实施例中,将包括侧接上游序列和下游序列的SoI的外源多核苷酸模板引入细胞中,其中该上游和下游序列与靶序列中整合位点的任一侧具有序列相似性。在一些实施例中,该包括SoI的外源多核苷酸包括例如突变基因。在一些实施例中,该外源多核苷酸包括对于该细胞而言是内源性或外源性的序列。在一些实施例中,该SoI包括编码蛋白质的多核苷酸,或非编码序列,诸如像微小RNA。在一些实施例中,该SoI可操作地与调节元件连接。在一些实施例中,该SoI是调节元件。在一些实施例中,该SoI包括抗性盒,例如,赋予对抗生素抗性的基因。在一些实施例中,该SoI包括野生型靶序列的突变。在一些实施例中,该SoI通过产生移码突变或核苷酸取代来破坏或校正靶序列。在一些实施例中,该SoI包括标记。将标记引入靶序列可以便于筛选靶向的整合。在一些实施例中,该标记是限制性位点、荧光蛋白或选择性标记。在一些实施例中,将该SoI作为包括SoI的载体引入。

[0234] 选择该外源多核苷酸模板中的上游和下游序列以促进靶序列和外源多核苷酸之间的同源重组。该上游序列是与用于整合的靶向位点的上游序列(靶序列)具有序列相似性的核酸序列。类似地,该下游序列是与用于整合的靶位点的下游序列具有序列相似性的核酸序列。因此,在一些实施例中,通过在上游和下游序列处的同源重组,将包括SoI的外源多核苷酸模板插入靶序列中。在一些实施例中,该外源多核苷酸模板中的上游和下游序列与靶向的基因组序列的上游和下游序列分别具有至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、分别至少97%、至少98%、至少99%或100%的序列同一性。在一些实施例中,该上游或下游序列具有至少约20、50、100、150、200、250、300、350、400或500个碱基对和至多约600、750、1000、1250、1500、1750或2000个碱基对。在一些实施例中,该上游或下游序列具有约20至2000个碱基对、或约50至1750个碱基对、或约100至1500个碱基对、或约200至1250个碱基对、或约300至1000个碱基对、或约400至约750个碱基对、或约500至600个碱基对。在一些实施例中,该上游或下游序列具有约50、约100、约250、约500、约100、约1250、约1500、约1750、约2000、约2250或约2500个碱基对。

[0235] 在该方法的一些实施例中,靶序列中的修饰是细胞中靶序列表达的失活。例如,在CRISPR-Cas复合物与靶序列结合后,靶序列即失活,使得该序列不被转录,不产生编码的蛋白质,和/或该序列不像野生型序列那样起作用。例如,蛋白质或微小RNA编码序列可能会失活,从而不产生蛋白质。

[0236] 在一些实施例中,调节序列可能会失活,使得它不再作为调节序列起作用。调节序列的实例包括启动子、转录终止子、增强子和本文所述的其他调节元件。失活的靶序列可以包括缺失突变(即,一个或多个核苷酸的缺失)、插入突变(即,一个或多个核苷酸的插入)或无义突变(即,用另一个核苷酸取代一个单核苷酸从而引入终止密码子)。在一些实施例中,靶序列的失活导致靶序列的“敲除”。

[0237] 在包括本文提供的重组Cas9的方法的一些实施例中,相对于野生型Cas9或不包括KFERQ基序或KFERQ样基序的Cas9,细胞基因组中脱靶修饰减少至少约50%。如本文所述,脱靶修饰是非特异性和非预期的基因修饰,例如非预期的点突变、缺失、插入、倒位和易位。在一些实施例中,本披露的重组Cas9蛋白由于更快的降解速率而在细胞中具有减少的脱靶。在一些实施例中,本披露的重组Cas9蛋白由于细胞对Cas9的较低可用性而在细胞中具有减少的脱靶。在一些实施例中,本披露的重组Cas9蛋白由于细胞与Cas9的较短暴露时间而在细胞中具有减少的脱靶。

[0238] 在包括本文提供的重组Cas9的方法的一些实施例中,相对于野生型Cas9,脱靶修饰减少,并且靶上修饰至少大致相同水平。在一些实施例中,重组Cas9的靶上修饰在野生型Cas9的靶上修饰的至少约20%、至少约15%、至少约10%、至少约5%、至少约4%、至少约3%、至少约2%、至少约1%、至少约0.5%内。在包括本文提供的重组Cas9的一些实施例中,相对于野生型Cas9,脱靶修饰减少,并且靶上修饰增加。在一些实施例中,重组Cas9的靶上修饰比野生型Cas9高至少约5%、高至少约10%、高至少约11%、高至少约12%、高至少约13%、高至少约14%、高至少约15%、高至少约16%、高至少约17%、高至少约18%、高至少约19%或高至少约20%。

[0239] 在一些实施例中,相对于野生型Cas9或不包括KFERQ基序或KFERQ样基序的Cas9,细胞基因组中的脱靶修饰减少至少约10%、至少约20%、至少约30%、至少约40%、至少约

50%、至少约60%、至少约70%、至少约80%、至少约90%、至少约100%、至少约150%或至少约200%。

[0240] 在包括本文提供的重组Cas9的方法的一些实施例中,相对于野生型Cas9或不包括KFERQ基序或KFERQ样基序的Cas9,细胞基因组中的脱靶修饰减少至少约2倍、至少约3倍、至少约4倍、至少约5倍、至少约6倍、至少约7倍、至少约8倍、至少约9倍、至少约10倍、至少约20倍、至少约30倍、至少约40倍、至少约50倍、至少约60倍、至少约70倍、至少约80倍、至少约90倍、至少约100倍、至少约500倍或至少约1000倍。

[0241] 在该方法的一些实施例中,细胞基因组中的脱靶修饰少于由具有KFERQ基序或KFERQ样基序的重组Cas9产生的基因组中的所有修饰的约5%。在该方法的一些实施例中,细胞基因组中的脱靶修饰少于由具有KFERQ基序或KFERQ样基序的重组Cas9产生的基因组中的所有修饰的约2%。在该方法的一些实施例中,细胞基因组中的脱靶修饰少于由具有KFERQ基序或KFERQ样基序的重组Cas9产生的基因组中的所有修饰的约1%。如本文所述,野生型Cas9的脱靶修饰可以是野生型Cas9的基因组中所有修饰的至少约2%、至少约3%、至少约4%、至少约5%、至少约6%、至少约7%、至少约8%、至少约9%或至少约10%。在一些实施例中,具有KFERQ基序或KFERQ样基序的重组Cas9的脱靶修饰少于由具有KFERQ基序或KFERQ样基序的重组Cas9产生的基因组中的所有修饰的约5%、少于约4%、少于约3%、少于约2%、少于约1.5%、少于约1%、少于约0.5%或少于约0.1%。脱靶修饰的量可以取决于指导多核苷酸的序列和靶基因组基因座而变化。通常,当使用相同的指导多核苷酸时,与野生型Cas9相比,具有KFERQ基序或KFERQ样基序的重组Cas9蛋白具有减少的脱靶修饰。

[0242] 在该方法的一些实施例中,靶序列在细菌细胞中。在一些实施例中,该细菌细胞是实验室菌株。此类细胞的实例包括但不限于大肠杆菌、金黄色葡萄球菌、霍乱弧菌、肺炎链球菌、枯草芽孢杆菌、新月形柄杆菌、生殖支原体、费氏曲霉、集胞藻属、荧光假单胞菌、棕色固氮菌、天蓝色链霉菌。在一些实施例中,细菌细胞是用于制备食物和/或饮料的细菌。此类细胞的非限制性示例性属包括但不限于醋杆菌属、节杆菌属、芽孢杆菌属、双歧杆菌属、短状杆菌属、短杆菌属、肉杆菌属、棒杆菌属、肠球菌属、葡糖醋杆菌属、哈夫尼菌属、盐单胞菌属、考克氏菌属、乳杆菌属(包括耐酸乳杆菌、酸鱼乳杆菌、嗜酸乳酸杆菌、消化乳杆菌、短乳杆菌、布赫纳乳杆菌、干酪乳酸菌、弯曲乳杆菌、发酵乳杆菌、希氏乳杆菌、詹氏乳杆菌、金氏乳杆菌、乳酸乳杆菌、副干酪乳杆菌、植物乳杆菌和阪乳杆菌)、明串珠菌属、微杆菌属、小球菌属、丙酸杆菌属、魏斯氏菌属和发酵单胞菌属。

[0243] 在该方法的一些实施例中,靶细胞在真核细胞中。在一些实施例中,该真核细胞是动物或人类细胞。在一些实施例中,该真核细胞是人类或啮齿动物或牛细胞系或者细胞株。此类细胞/细胞系或细胞株的实例包括但不限于小鼠骨髓瘤(NS0)细胞系、中国仓鼠卵巢(CHO)细胞系、HT1080、H9、HepG2、MCF7、MDBK Jurkat、NIH3T3、PC12、BHK(幼仓鼠肾细胞)、VERO、SP2/0、YB2/0、Y0、C127、L细胞、COS(例如COS1和COS7)、QC1-3、HEK-293、VERO、PER.C6、HeLA、EB1、EB2、EB3、溶瘤或杂交瘤细胞系。在一些实施例中,该真核细胞是CHO细胞系。在一些实施例中,该真核细胞是CHO细胞。在一些实施例中,该细胞是CHO-K1细胞、CHO-K1 SV细胞、DG44 CHO细胞、DUXB11 CHO细胞、CHOS、CHO GS敲除细胞、CHO FUT8 GS敲除细胞、CHOZN或CHO衍生的细胞。CHO GS敲除细胞(例如,GSKO细胞)是例如CHO-K1 SV GS敲除细胞。CHO FUT8敲除细胞是例如POTELLIGENT CHOK1 SV(龙沙生物公司)。真核细胞也可以是禽细胞、

细胞系或细胞株,例如EBX细胞、EB14、EB24、EB26、EB66或EBv13。

[0244] 在一些实施例中,该真核细胞是人类细胞。在一些实施例中,该人类细胞是干细胞。干细胞可以是例如多能干细胞,包括胚胎干细胞(ESC)、成年干细胞、诱导的多能干细胞(iPSC)、组织特异性干细胞(例如,造血干细胞)和间充质干细胞(MSC)。在一些实施例中,该细胞是多能干细胞。在一些实施例中,该细胞是诱导的多能干细胞。在一些实施例中,该人类细胞是本文描述的任何细胞的分化形式。在一些实施例中,该真核细胞是源自培养物中任何原代细胞的细胞。

[0245] 在一些实施例中,该真核细胞是肝细胞,诸如人肝细胞、动物肝细胞,或非实质细胞。例如,该真核细胞可以是可培养代谢合格的人肝细胞、可培养感应诱导合格的人肝细胞、可培养人肝细胞、悬浮合格的人肝细胞(包括10-供体和20-供体合并的肝细胞)、人肝库普弗细胞、人肝星状细胞、狗肝细胞(包括单个和合并的比格犬肝细胞)、小鼠肝细胞(包括CD-1和C57BI/6肝细胞)、大鼠肝细胞(包括Sprague-Dawley、Wistar Han和Wistar肝细胞)、猴肝细胞(包括食蟹猴或恒河猴肝细胞)、猫肝细胞(包括家养短毛猫肝细胞)和兔肝细胞(包括新西兰白兔肝细胞)。

[0246] 在一些实施例中,该真核细胞是植物细胞。例如,该植物细胞可以是诸如木薯、玉米、高粱、小麦或水稻的农作物的细胞。该植物细胞可以是藻类、树木或蔬菜的细胞。该植物细胞可以是单子叶植物或双子叶植物的细胞,或者可以是农作物或谷物植物、生产植物、水果或蔬菜的细胞。例如,该植物细胞可以是树木的细胞,该树木为例如柑橘属果树,诸如橘子树、葡萄柚树或柠檬树;桃树或油桃树;苹果树或梨树;坚果树,诸如杏仁树或胡桃树或开心果树;茄属植物,例如,马铃薯、芸苔属(*Brassica*)属植物、莴苣属(*Lactuca*)植物;菠菜属(*Spinacia*)植物;辣椒属(*Capsicum*)植物;棉花、烟草、芦笋、胡萝卜、卷心菜、西兰花、花椰菜、番茄、茄子、胡椒、莴苣、菠菜、草莓、蓝莓、覆盆子、黑莓、葡萄、咖啡、可可等。

[0247] 在该方法的一些实施例中,指导多核苷酸的指导序列能够与细胞中的靶序列杂交。在一些实施例中,指导多核苷酸的DNA结合区段与细胞中的靶序列杂交。在一些实施例中,指导多核苷酸(例如,指导RNA)的DNA结合区段包括与靶DNA内的特定序列互补的多核苷酸序列。在一些实施例中,指导多核苷酸的指导序列能够与细菌细胞中的靶序列杂交。在一些实施例中,指导多核苷酸的指导序列能够与真核细胞中的靶序列杂交。在一些实施例中,指导多核苷酸的指导序列能够与哺乳动物细胞中的靶序列杂交。在一些实施例中,指导多核苷酸的指导序列能够与人细胞中的靶序列杂交。在一些实施例中,指导多核苷酸的指导序列能够与多能干细胞中的靶序列杂交。在一些实施例中,指导多核苷酸的指导序列能够与诱导的多能干细胞中的靶序列杂交。在一些实施例中,指导多核苷酸的指导序列能够与植物细胞中的靶序列杂交。

[0248] 在一些实施例中,指导多核苷酸的指导序列为至少约5、6、7、8、9、10、12、14、16、18或20个核苷酸并且至多约20、25、30、35、40、45或50个核苷酸。在一些实施例中,指导多核苷酸的指导序列为约5至约50个核苷酸。在一些实施例中,指导多核苷酸的指导序列为约6至约45个核苷酸。在一些实施例中,指导多核苷酸的指导序列为约7至约40个核苷酸。在一些实施例中,指导多核苷酸的指导序列为约8至约35个核苷酸。在一些实施例中,指导多核苷酸的指导序列为约9至约30个核苷酸。在一些实施例中,指导多核苷酸的指导序列为约10至约20个核苷酸。在一些实施例中,指导多核苷酸的指导序列为约12至约20个核苷酸。在一些

实施例中,指导多核苷酸的指导序列为约14至约20个核苷酸。在一些实施例中,指导多核苷酸的指导序列为约16至约20个核苷酸。在一些实施例中,指导多核苷酸的指导序列为约18至约20个核苷酸。在一些实施例中,指导多核苷酸的指导序列为约5至约10个核苷酸。在一些实施例中,指导多核苷酸的指导序列为约6至约10个核苷酸。在一些实施例中,指导多核苷酸的指导序列为约7至约10个核苷酸。在一些实施例中,指导多核苷酸的指导序列为约8至约10个核苷酸。指导序列的长度可以由本领域技术人员使用文中所述的指导序列设计工具来测定。

[0249] 在该方法的一些实施例中,将CRISPR-Cas系统经由递送颗粒、囊泡或病毒载体引入细胞中。在该方法的一些实施例中,经由递送颗粒将包括重组Cas9和指导多核苷酸的CRISPR-Cas系统引入细胞中。在该方法的一些实施例中,经由囊泡将包括重组Cas9和指导多核苷酸的CRISPR-Cas系统引入细胞中。在该方法的一些实施例中,经由载体将包括重组Cas9和指导多核苷酸的CRISPR-Cas系统引入细胞中。在该方法的一些实施例中,经由病毒载体将包括重组Cas9和指导多核苷酸的CRISPR-Cas系统引入细胞中。在该方法的一些实施例中,将编码包括重组Cas9和指导多核苷酸的复合物的组分的多核苷酸引入一个或多个载体上。本文提供了递送颗粒、囊泡、载体、病毒载体和递送到细胞中的方法(例如载体的转染)的实例。

[0250] 本文引用的所有参考文献,包括专利、专利申请、论文、教科书等以及其中引用的参考文献(如同它们还未曾引用过的程度)通过引用以其全文结合在此。

[0251] 其他示例性实施例

[0252] 实施例1是包括工程化的KFERQ基序或KFERQ样基序的重组Cas9蛋白。

[0253] 实施例2包括实施例1的重组Cas9蛋白,其中该工程化的KFERQ基序或KFERQ样基序选自KFERQ(SEQ ID NO:24)、RKVEQ(SEQ ID NO:25)、QDLKF(SEQ ID NO:26)、QRFFE(SEQ ID NO:27)、NRVVD(SEQ ID NO:28)、QRDKV(SEQ ID NO:29)、QKILD(SEQ ID NO:30)、QKKEL(SEQ ID NO:31)、QFREL(SEQ ID NO:32)、IKLDQ(SEQ ID NO:33)、DVVRQ(SEQ ID NO:34)、QRIVE(SEQ ID NO:35)、VKELQ(SEQ ID NO:36)、QKVFD(SEQ ID NO:37)、QELLR(SEQ ID NO:38)、VDKLN(SEQ ID NO:39)、RIKEN(SEQ ID NO:40)、NKKFE(SEQ ID NO:41)及其组合。

[0254] 实施例3包括实施例1或2的重组Cas9蛋白,其中该工程化的KFERQ样基序是VDKLN(SEQ ID NO:39)。

[0255] 实施例4包括实施例1的重组Cas9蛋白,其中该工程化的KFERQ基序或KFERQ样基序在该Cas9蛋白的REC叶中。

[0256] 实施例5包括实施例2的重组Cas9蛋白,其中该工程化的KFERQ基序或KFERQ样基序在该REC叶的Rec2结构域中。

[0257] 实施例6包括实施例1的重组Cas9蛋白,其中该工程化的KFERQ基序或KFERQ样基序在该重组Cas9蛋白的HNH结构域、RuvC结构域或PI结构域中。

[0258] 实施例7包括实施例1至4中任一项的重组Cas9蛋白,其中该工程化的KFERQ基序或KFERQ样基序在该重组Cas9蛋白的表面暴露的区域中。

[0259] 实施例8包括实施例1至6中任一项的重组Cas9蛋白,其中该工程化的KFERQ基序或KFERQ样基序在该重组Cas9蛋白的N末端或C末端。

[0260] 实施例9是包括野生型Cas9蛋白的一个或多个氨基酸修饰的重组Cas9蛋白,这些

修饰将分子伴侣介导的自噬 (CMA) 靶基序或内体微自噬 (eMI) 靶基序引入该Cas9蛋白中,其中该重组Cas9蛋白在体内的降解比该野生型Cas9蛋白或不包括CMA或eMI靶基序的Cas9蛋白快至少20%。

[0261] 实施例10包括实施例9的重组Cas9蛋白,其中该重组Cas9蛋白在体内的降解比该野生型Cas9蛋白或不包括CMA或eMI靶基序的Cas9蛋白快至少50%。

[0262] 实施例11包括实施例9或10的重组Cas9蛋白,其中该重组Cas9蛋白在体内的降解比该野生型Cas9蛋白或不包括CMA或eMI靶基序的Cas9蛋白快至少80%。

[0263] 实施例12是包括野生型Cas9蛋白的一个或多个氨基酸修饰的重组Cas9蛋白,其中该重组Cas9蛋白包括CMA靶基序或eMI靶基序。

[0264] 实施例13包括实施例9至12中任一项的重组Cas9蛋白,其中该CMA靶基序或eMI靶基序选自KFERQ (SEQ ID NO:24)、RKVEQ (SEQ ID NO:25)、QDLKF (SEQ ID NO:26)、QRFFE (SEQ ID NO:27)、NRVVD (SEQ ID NO:28)、QRDKV (SEQ ID NO:29)、QKILD (SEQ ID NO:30)、QKKEL (SEQ ID NO:31)、QFREL (SEQ ID NO:32)、IKLDQ (SEQ ID NO:33)、DVVRQ (SEQ ID NO:34)、QRIVE (SEQ ID NO:35)、VKELQ (SEQ ID NO:36)、QKVFD (SEQ ID NO:37)、QELLR (SEQ ID NO:38)、VDKLN (SEQ ID NO:39)、RIKEN (SEQ ID NO:40)、NKKFE (SEQ ID NO:41) 及其组合。

[0265] 实施例14包括实施例13的重组Cas9蛋白,其中该CMA靶基序或eMI靶基序是VDKLN (SEQ ID NO:39)。

[0266] 实施例15包括实施例9至14中任一项的重组Cas9蛋白,其中该一个或多个氨基酸取代在该重组Cas9蛋白的表面暴露的区域中。

[0267] 实施例16包括分离自酿脓链球菌的重组Cas9蛋白 (SpCas9),该重组Cas9蛋白在SEQ ID NO:1的位置F185、A547、I548、T560、V561、D829、I830、L1087、S1088、P1199、K1200或其组合中的一个或多个处包括氨基酸修饰。

[0268] 实施例17包括实施例9至16中任一项的重组Cas9蛋白,其中该氨基酸修饰包括以下突变中的一个或多个:F185N;A547E/I548L;T560E/V561Q;D829L/I830R;L1087E/S1088Q;或P1199D/K1200Q。

[0269] 实施例18包括实施例9至17中任一项的重组Cas9蛋白,其中该氨基酸修饰是F185处的突变。

[0270] 实施例19包括实施例18的重组Cas9蛋白,其中该突变是F185N。

[0271] 实施例20包括实施例16至19中任一项的重组Cas9蛋白,其中该氨基酸修饰导致CMA靶基序或eMI靶基序。

[0272] 实施例21包括实施例9至20中任一项的重组Cas9蛋白,其中该重组Cas9蛋白与SEQ ID NO:1具有至少90%同一性。

[0273] 实施例22是能够与70kD的热休克同源蛋白 (HSC70) 结合的重组Cas9蛋白。

[0274] 实施例23是分离自酿脓链球菌的重组Cas9蛋白 (SpCas9),该重组Cas9蛋白在SEQ ID NO:1的氨基酸位置185处包括工程化的KFERQ基序或KFERQ样基序。

[0275] 实施例24包括实施例23的重组Cas9蛋白,其中该KFERQ样基序是VDKLN (SEQ ID NO:39)。

[0276] 实施例25包括实施例1至24中任一项的重组Cas9蛋白,该重组Cas9蛋白进一步在SEQ ID NO:1中的位置D10、H840或其组合处包括突变。

- [0277] 实施例26包括实施例25的重组Cas9蛋白,其中该突变选自D10A或D10N;H840A、H840N或H840Y;及其组合。
- [0278] 实施例27包括实施例1至26中任一项的重组Cas9蛋白,其中该重组Cas9蛋白产生粘性末端。
- [0279] 实施例28包括实施例1至27中任一项的重组Cas9蛋白,该重组Cas9蛋白进一步包括一个或多个核定位信号。
- [0280] 实施例29是编码实施例1至28中任一项的重组Cas9蛋白的多核苷酸序列。
- [0281] 实施例30包括实施例29的多核苷酸序列,其中该多核苷酸序列进行密码子优化以在真核细胞中表达。
- [0282] 实施例31是非天然存在的CRISPR-Cas系统,其包括:实施例1至28中任一项的重组Cas9蛋白;以及与该重组Cas9蛋白形成复合物并包括指导序列的指导多核苷酸。
- [0283] 实施例32是非天然存在的CRISPR-Cas系统,其包括:实施例29或30的多核苷酸序列;以及编码与该重组Cas9蛋白形成复合物并包括指导序列的指导多核苷酸的核苷酸序列。
- [0284] 实施例33是非天然存在的CRISPR-Cas系统,其包括:可操作地连接至实施例29或30的多核苷酸序列的调节元件;以及与该重组Cas9蛋白形成复合物并包括指导序列的指导多核苷酸。
- [0285] 实施例34包括实施例31至33中任一项的系统,其中该指导序列与同向重复序列连接。
- [0286] 实施例35包括实施例31至34中任一项的系统,其中该指导多核苷酸包括tracrRNA序列。
- [0287] 实施例36包括实施例31至34中任一项的系统,该系统进一步包括单独的多核苷酸,该多核苷酸包括tracrRNA序列。
- [0288] 实施例37包括实施例31至35中任一项的系统,其中编码该重组Cas9蛋白的多核苷酸序列和指导多核苷酸在单个载体上。
- [0289] 实施例38包括实施例36的系统,其中编码该重组Cas9蛋白的多核苷酸序列、指导多核苷酸和tracrRNA序列在单个载体上。
- [0290] 实施例39是包括实施例31至38中任一项的系统的递送颗粒。
- [0291] 实施例40是包括实施例31至38中任一项的系统的囊泡。
- [0292] 实施例41包括实施例40的囊泡,其中该囊泡是外泌体或脂质体。
- [0293] 实施例42是包括根据实施例31至38中任一项的系统的病毒载体。
- [0294] 实施例43包括实施例42的病毒载体,其中该病毒载体是腺病毒、慢病毒或腺相关病毒载体。
- [0295] 实施例44是在细胞基因组中的靶序列处提供位点特异性修饰的方法,该方法包括将实施例31至38中任一项的CRISPR-Cas系统引入该细胞中。
- [0296] 实施例45包括实施例44的方法,其中该修饰包括靶序列的至少一部分的缺失。
- [0297] 实施例46包括实施例44的方法,其中该修饰包括靶序列的突变。
- [0298] 实施例47包括实施例44的方法,其中该修饰包括在靶序列处插入目的序列(SoI)。
- [0299] 实施例48包括实施例44至47中任一项的方法,其中该细胞基因组中的脱靶修饰小

于由重组Cas9产生的基因组中修饰的约5%。

[0300] 实施例49包括实施例44至48中任一项的方法,其中该细胞基因组中的脱靶修饰小于由重组Cas9产生的基因组中修饰的约2%。

[0301] 实施例50包括实施例44至49中任一项的方法,其中该细胞基因组中的脱靶修饰小于由重组Cas9产生的基因组中修饰的约1%。

[0302] 实施例51包括实施例44至50中任一项的方法,其中相对于野生型CRISPR-Cas9或不包括KFERQ基序或KFERQ样基序的Cas9,细胞基因组中的脱靶修饰减少至少约50%。

[0303] 实施例52包括实施例44至51中任一项的方法,其中该细胞是细菌细胞、哺乳动物细胞或植物细胞。

[0304] 实施例53包括实施例52的方法,其中该细胞是人细胞。

[0305] 实施例54包括实施例53的方法,其中该细胞是多能干细胞。

[0306] 实施例55包括实施例54的方法,其中该细胞是诱导的多能干细胞。

[0307] 实施例56包括实施例44至55中任一项的方法,其中该指导多核苷酸的指导序列能够与该细胞基因组中的靶序列杂交。

[0308] 实施例57包括实施例44至56中任一项的方法,其中将CRISPR-Cas系统经由递送颗粒、囊泡或病毒载体引入细胞中。

[0309] 序列

[0310] 下表4列出了本文提供的序列。

[0311] 表4. 序列列表

[0312]

SEQ ID NO.	氨基酸 (AA) 或核苷酸 (NT)	序列
1	AA	SpCas9
2	AA	FaDe-SpCas9
3	NT	SpCas9
4	NT	FaDe-SpCas9
5	AA	SpCas9 REC 叶
6	AA	SpCas9 NUC 叶 1
7	AA	SpCas9 NUC 叶 2
8	AA	SpCas9 BH 结构域
9	AA	SpCas9 Rec1 结构域 1
10	AA	SpCas9 Rec1 结构域 2
11	AA	SpCas9 Rec2 结构域
12	AA	SpCas9 RuvC I 结构域
13	AA	SpCas9 RuvC II 结构域
14	AA	SpCas9 RuvC III 结构域
15	AA	SpCas9 HNH 结构域
16	AA	SpCas9 PI 结构域

	17	AA	嗜热链球菌 Cas9
	18	AA	停乳链球菌 Cas9
	19	AA	变形链球菌 Cas9
	20	AA	英诺克李斯特菌 Cas9
	21	AA	金黄色葡萄球菌 Cas9
[0313]	22	AA	肺炎克雷伯氏菌 Cas9
	23	AA	新凶手弗朗西斯菌 Cas9
	24-41	AA	KFERQ 基序或 KFERQ 样基序
	42-47	AA	核定位信号
	48-65	NT	本文使用的引物
	66-70	NT	本文使用的指导 RNA

[0314] SpCas9-来自酿脓链球菌的Cas9 (SEQ ID NO:1)

MDKKYSIGLDIGTNSVGVAVITDDYKVPSSKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKN
RICYLQEIFSNEMAKVDDSFHRLLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLI
YLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAIL SARLSKSRRENLIQAQPLG
EKKNGLFGNLIALSGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADFLAAKNLSDAILLSDILRV
NTEITKAPLSASMIKRYDEHHQDLTLKALVRRQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDG
TEELLVKLNREDLLRKQRTFDNGSI PHQIHLGELHAILRRQEDFY PFLKDNREKIEKILTFRIPIYVVGPLARGNSRF
AWMTRKSEETITPWNFEVVDKGASQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAF
LSGEQKKAIVDLLFKTRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKI IKDKDFLDNEENEDIL
EDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTI LDFLKSDFANRNF
[0315] MQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA IKKGI LQTVKVVDELVKVMGRHKPENIVIEMARENQTT
QKGQKNSRERMKRIE EGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFL
KDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLI TQRKFDNLTKAERGGLSELDKAGFIKRQLVET
RQITKHVAQILD SRMNTKYDENDKLIREVKVI TLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVG TALIKKY
PKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGR
DFATVRKVL SMPQVNI VKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKYGGFDSPTVAYSVLVVAKVEKGKSK
KLKSVKELLGITIMERSSEFKNPIDFLEAKGYKEVKDLI IKLPKYSLFELENGRKRMLASAGELQKGNELALPSKY
VNFLYLASHYEKLGKSPEDNEQQLFVEQHKHYLDEIIEQISEFSKRVI LADANLDKVL SAYNKHRRDKPIREQAENI
IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQSI TGLYETRIDLSQLGGD

[0316] FaDe-SpCas9-具有F185N突变的SpCas9 (SEQ ID NO:2)

MDKKYSIGLDIGTNSVGVAVITDDYKVPSSKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKN
 RICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLI
 YLALAHMIKFRGHFLIEGDLNPDNSDVDKLNIQLVQTYNQLFEENPINASGVDAKAIL SARLSKSRLENLIAQLPG
 EKKNGLFGNLIALSGLTPNFKSNFDLAEDAKLQLSKDQYADLFLAAKNLSDAILLSDI LRV
 NTEITKAPLSASMIKRYDEHHQDLTLKALVRRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDG
 TEELLVKLNREDLLRKQRTFDNGSI PHQIHLGELHAILRRQEDFY PFLKDNREKIEKILTFRI PYYVGPLARGNSRF
 AWMTRKSEETITPWNFEVVDKGSASQSFIERMTNFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAF
 LSGEQKKAIVDLLFKTRNKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKI IKDKDFLDNEENEDIL
 EDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDQSGKTI LDFLKSDGFANRNF
 MQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGI LQTVKVVDELVKVMGRHKPENIVIEMARENQTT
 QKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFL
 KDDSIDNKVLRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLI TQRKFDNLTKAERGGLSELDKAGFIKRQLVET
 RQITKHVAQILD SRMNTKYDENDKLIREVKVITLKS KLVSDFRKDFQFYK VREINNYHHAHDAYLNAVVG TALIKKY
 PKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGR
 DFATVRKVL SMPQVNI VKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGSK
 KLKSVKELLGITIMERSSEFEKNPIDFLEAKGYKEVKKDLI IKLPKYSLFELENGRKRMLASAGELQKGNELALPSKY
 VNFYLYASHYEKLGKSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVI LADANLDKVL SAYNKH RDKPIREQAENI
 IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSI TGLYETRIDLSQLGGD

[0317]

[0318] SpCas9核苷酸序列 (SEQ ID NO:3)

ATGGCTCCTAAGAAAAAGCGGAAGGTGGACAAGAAATACTCAATCGGGCTGGACATCGGAACTAACTCAGTGGGGTG
 GGCAGTCATTACTGACGAGTACAAAGTGCCAAGCAAAGAAATTTAAGGTCCTGGGCAACACCGATAGGCACTCCATCA
 AGAAAAATCTGATTGGGGCCCTGCTGTTCGACTCTGGAGAGACAGCTGAAGCAACTAGACTGAAAAGGACTGCTAGA
 AGGCGCTATACCCGGCGAAAGAATCGCATCTGCTACCTGCAGGAGATTTTCTCTAACGAAATGGCCAAGGTGGACGA
 TAGTTTCTTTTCATCGGCTGGAGGAATCATTCCTGGTCGAGGAAGATAAGAAACACGAGAGACATCCTATCTTTGGAA
 ACATTGTGGACGAGGTCTGCTTATCACGAAAAATACCCCAACCATCTATCATCTGCGCAAGAAACTGGTGGACTCTACA
 GATAAAGCAGACCTGCGGCTGATCTATCTGGCCCTGGCTCATATGATTAAGTTCAGAGGCCATTTTCTGATCGAGGG
 AGATCTGAACCCAGACAATAGCGATGTGGACAAGCTGTTTATCCAGCTGGTCCAGACATACAATCAGCTGTTTGGAG
 AAAACCTATTAATGCATCTGGCGTGGACGCAAAAAGCCATCCTGAGTGCCAGGCTGTCTAAGAGTAGAAGGCTGGAG
 AACCTGATCGCTCAGCTGCCAGGCGAAAAGAAAACGGCCTGTTTGGAAATCTGATTGCACTGTCACTGGGACTGAC
 ACCTAACTTCAAGAGCAATTTTGATCTGGCCGAGGACGCTAAACTGCAGCTGAGCAAGGACACTTATGACGATGACC
 TGGATAACCTGCTGGCTCAGATCGGAGATCAGTACGCAGACCTGTTCTCGCCGCTAAGAACTGTCTGACGCTATC
 CTGCTGAGTGATATTCTGCGGGTGAACACCGAGATTACAAAAGCCCTCTGTGCTAGCTAGCATGATCAAGAGATATGA

[0319]

CGAGCACCATCAGGATCTGACCCTGCTGAAGGCACCTGGTGCGCCAGCAGCTGCCCGAGAAGTACAAGGAAATCTTCT
TTGATCAGAGTAAGAACGGGTACGCCGTTATATTGACGGCGGAGCTTCACAGGAGGAATTCTACAAGTTTATCAA
CCTATTCTGGAGAAGATGGACGGCACCGAGGAAGTCTGGTGAAGTGAATCGCGAGGACCTGCTGCGCAAGCAGCG
GACATTTGATAACGGCTCCATCCCCACCAGATTCATCTGGGAGAGCTGCACGCAATCCTGCGACGACAGGAAGACT
TCTACCCATTTCTGAAGGATAACCGCGAGAAGATCGAAAAAATTCGACCTTCCGGATCCCTTACTATGTGGGGCC
CTGGCAAGGGTAATTCGCCCTTTGCCCTGGATGACACGGAATCTGAGGAAACAATCACTCCTTGGAATTCGAGGA
AGTGGTCGATAAGGGAGCTTCCGCACAGTCTTTCATCGAGAGAATGACAACTTCGACAAAAACCTGCCAAATGAGA
AAGTGTGCCTAAGCACAGTCTGCTGTACGAGTATTTACAGTCTATAACGAAGTACTAAGGTGAAATACGTCACC
GAGGGGATGAGGAAGCCCGCTTCTTGAGCGGTGAACAGAAGAAAGCTATCGTGGACCTGCTGTTTTAAAACCAATCG
CAAGGTGACAGTCAAGCAGCTGAAGGAGGACTACTTCAAGAAAATGAATGTTTTCGATTCTGTGGAGATCAGTGGCG
TCGAAGACAGATTTAACGCTTCTCTGGGAACCTACCACGATCTGCTGAAGATCATTAAAGGATAAAGACTTCTGGAC
AACGAGGAAAATGAGGATATCCTGGAAGACATTGTGCTGACCCTGACACTGTTTTGAGGATCGCGAAATGATCGAGGA
ACGGCTGAAAACCTTATGCCCATCTGTTTCGATGACAAGGTGATGAAACAGCTGAAGCGAAGAAGGTACACCGGCTGGG
GACGACTGAGCAGAAAGCTGATCAACGGCATTCCGGGCAAAACAGAGTGGAAAGACTATCCTGGACTTTCTGAAATCA
GATGGCTTCGCTAACAGAAATTTTATGCAGCTGATTCACGATGACAGCCTGACCTTCAAAGAGGATATCCAGAAGGC
ACAGGTGTCCGGCAGGGTGACTCTCTGCACGAGCATAATCGCAAACCTGGCCGGGTCCCCGCCATCAAGAAAGGTA
TTCTGCAGACCGTGAAGGTGGTCGATGAGCTGGTGAAGTCAATGGGCAGGCATAAGCCAGAAAACATCGTGATTGAG
ATGGCCCGGAAAATCAGACCACACAGAAAGGACAGAAAGACAGCCGCGAGCGGATGAAAAGGATCGAGGAAGGCAT
TAAGGAACTGGGATCCCAGATCCTGAAAGAGCACCTGTGGAAAACACTCAGCTGCAGAATGAGAAGCTGTATCTGT
ACTATCTGCAGAATGGGCGGGATATGTACGTGGACCAGGAGCTGGATATTAACCGACTGTCTGATTACGACGTGGAT
CATATCGTCCACAGTCATTCCTGAAAGATGACAGCATTGACAATAAGGTGCTGACCCGGAGTGACAAAAACCGAGG
AAAGAGTGATAATGTCCCTTCAGAGGAAGTGGTCAAGAAAATGAAGAACTACTGGAGACAGCTGCTGAATGCCAAAC
TGATCACACAGCGAAAGTTTGATAACCTGACTAAAGCTGAGAGAGGGGGTCTGTGAGAATGGACAAAGCAGGCTTC
ATCAAGCGACAGCTGGTGGAGACCAGACAGATCACAAAGCACGTCGCTCAGATTCTGGATAGCAGGATGAACACAAA
GTACGATGAGAATGACAACTGATCCGCGAAGTGAAGGTCATTACTCTGAAGTCAAACTTGTGAGCGACTTCAGAA
AGGATTTCCAGTTCTACAAAGTCAGGGAGATCAACAATTATACCATGCTCATGACGCATACCTGAACGCAGTGGTC
GGGACCGCCTGATTAAGAAAATACCCCAAACCTGGAGAGCGAATTCGTGTACGGTACTATAAGGTGTACGATGTCAG
AAAAATGATCGCCAAGAGTGAGCAGGAAATGGAAAAGCCACCGCTAAGTATTTCTTTTACTCAAACATCATGAATT
TCTTTAAGACTGAGATCACCTTGGCAAATGGGGAAATCCGAAAGAGACCCTGATTGAGACTAACGGCGAGACCGGA
GAAATCGTGTGGGACAAGGGTAGGGATTTTGCCACAGTGGCGAAGTCTGTCCATGCCTCAAGTGAATATTGTCAA
GAAAACAGAGGTGCAGACTGGCGGATTCAGTAAGGAATCAATTCGCCCAAACGGAACCTGATAAGCTGATCGCCC
GAAAGAAAGACTGGGATCCCAAGAAATATGGGGTTTCGACTCCCCAACAGTGGCTTACTCTGTCTGGTGGTTCGCA
AAGGTGGAGAAGGGGAAAAGCAAGAACTGAAATCCGTCAAGGAGCTGCTGGGTATCACTATTATGGAGAGGAGCTC
CTTCGAGAAGAACCCCATCGATTTTCTGGAGGCTAAAGGCTATAAGGAAGTGAAGAAAGACCTGATCATTAACTGC
CAAAGTACAGCCTGTTTGTGAGCTGGAAAACGGAAGGAAGCGAATGCTGGCATCCGCAGGAGAGCTGCAGAAGGGTAAT
GAACTGGCCCTGCCTTCTAAGTACGTGAACTTCTGTATCTGGCTAGCCACTACGAGAAGCTGAAAGGCTCCCCCGA
GGATAACGAACAGAAACAGCTGTTTGTGGAGCAGCACAAAGCATTATCTGGACGAGATCATTGAACAGATTAGCGAGT
TCTCCAAAAGAGTGATCCTGGCTGACGCAAATCTGGATAAGGTCTGAGCGCATAACAACAAACACAGAGATAAGCCA
ATCAGGGAGCAGGCCGAAAATATCATTCATCTGTTCACTCTGACCAACCTGGGAGCCCTGCAGCCTTCAAGTATTT
TGACACTACCATCGATCGGAAACGATACACATCCACTAAGGAGGTGCTGGACGCTACCTGATTACCAGAGCATTAA
CCGGCCTGTATGAAACAAGGATTGACCTGTCTCAGCTGGGGGGCGACTGA

[0320]

[0321] FaDe-SpCas9核苷酸序列 (SEQ ID NO:4)

[0322]

ATGGCTCCTAAGAAAAAGCGAAGGTGGACAAGAAATACTCAATCGGGCTGGACATCGGAACTAACTCAGTGGGGTG
 GGCAGTCATTACTGACGAGTACAAAGTGCCAAGCAAGAAATTTAAGGTCTGGGCAACACCGATAGGCACTCCATCA
 AGAAAAATCTGATTGGGGCCCTGCTGTTGACTCTGGAGAGACAGCTGAAGCAACTAGACTGAAAAGGACTGCTAGA
 AGGCGCTATACCCGGCGAAAGAATCGCATCTGCTACCTGCAGGAGATTTTCTCTAACGAAATGGCCAAGGTGGACGA
 TAGTTTCTTTCATCGGCTGGAGGAATCATTCCTGGTCGAGGAAGATAAGAAAACGAGAGACATCCTATCTTTGGAA
 ACATTGTGGACGAGGTCGCTTATCACGAAAAATACCCCACCATCTATCATCTGCGCAAGAACTGGTGGACTCTACA
 GATAAAGCAGACCTGCGGCTGATCTATCTGGCCCTGGCTCACATGATTAAGTTCAGAGGCCATTTTCTGATCGAGGG
 AGATCTGAACCCAGACAATAGCGATGTGGACAAGCTGAACATCCAGCTGGTCCAGACATACAATCAGCTGTTGAGG
 AAAACCTATTAATGCATCTGGCGTGGACGAAAAGCCATCCTGAGTGCCAGGCTGTCTAAGAGTAGAAGGCTGGAG
 AACCTGATCGCTCAGCTGCCAGGCGAAAAGAAAAACGGCCTGTTTGGAAATCTGATTGCACTGTCACTGGGACTGAC
 ACCTAACTTCAAGAGCAATTTTGATCTGGCCGAGGACGCTAAACTGCAGCTGAGCAAGGACACTTATGACGATGACC
 TGGATAACCTGCTGGCTCAGATCGGAGATCAGTACGCAGACCTGTTCCCTGGCCGCTAAGAACTGTCTGACGCTATC
 CTGCTGAGTGATATTCTGCGGGTGAACACCGAGATTACAAAAGCCCCCTGTGTCAGCTAGCATGATCAAGAGATATGA
 CGAGCACCATCAGGATCTGACCCTGCTGAAGGCCTGGTGCGCCAGCAGCTGCCCGAGAAGTACAAGGAAATCTTCT
 TTGATCAGAGTAAGAACGGGTACGCCGTTATATTGACGGCGGAGCTTCACAGGAGGAATCTACAAGTTTATCAAA
 CCTATTCTGGAGAAGATGGACGGCACCGAGGAACTGCTGGTGAACGAAATCGCGAGGACCTGCTGCGCAAGCAGCG
 GACATTTGATAACGGCTCCATCCCCACCAGATTCATCTGGGAGAGCTGCACGCAATCTGCGACGACAGGAAGACT
 TCTACCCATTTCTGAAGGATAACCGCGAGAAGATCGAAAAAATTTGACCTTCCGGATCCCTTACTATGTGGGGCC
 CTGGCAAGGGTAATTTCCCGCTTTGCCCTGGATGACACGGAAATCTGAGGAAACAATCACTCCTTGGAACTTCGAGGA
 AGTGGTCGATAAGGGAGCTTCCGCACAGTCTTTCATCGAGAGAATGACAAACTTCGACAAAAACCTGCCAAATGAGA
 AAGTGCTGCCTAAGCACAGTCTGCTGTACGAGTATTTACAGTCTATAACGAACTGACTAAGGTGAAATACGTCACC
 GAGGGGATGAGGAAGCCCGCTTCTGAGCGGTGAACAGAAGAAAGCTATCGTGGACCTGCTGTTTAAAACCAATCG
 CAAGGTGACAGTCAAGCAGCTGAAGGAGGACTACTTCAAGAAAATGAAATGTTTCGATTCTGTGGAGATCAGTGGCG
 TCGAAGACAGATTTAACGCTTCTCTGGAACTACCACGATCTGCTGAAGATCATTAAAGGATAAAGACTTCTTGGAC
 AACGAGGAAAATGAGGATATCCTGGAAGACATTGTGCTGACCCTGACACTGTTTGGAGATCGCGAAATGATCGAGGA
 ACGGCTGAAAATTTATGCCATCTGTTTCGATGACAAGGTGATGAAACAGCTGAAGCGAAGAAGGTACACCGGCTGGG
 GACGACTGAGCAGAAAGCTGATCAACGGCATTCTGGGACAAAACAGAGTGGAAAGACTATCCTGGACTTTCTGAAATCA
 GATGGCTTCGCTAACAGAAATTTTATGCAGCTGATTCACGATGACAGCCTGACCTTCAAAGAGGATATCCAGAAGGC

[0323] ACAGGTGTCCGGCAGGGTACTCTCTGCACGAGCATATCGCAAACCTGGCCGGGTCCCCGCCATCAAGAAAGGTA
 TTCTGCAGACCGTGAAGGTGGTCGATGAGCTGGTGAAAGTCATGGGCAGGCATAAGCCAGAAAACATCGTGATTGAG
 ATGGCCCGCGAAAATCAGACCACACAGAAAGGACAGAAGAACAGCCGCGAGCGGATGAAAAGGATCGAGGAAGGCAT
 TAAGGAACTGGGATCCCAGATCCTGAAAGAGCACCCCTGTGAAAACACTCAGCTGCAGAATGAGAAGCTGTATCTGT
 ACTATCTGCAGAATGGCGGGATATGTACGTGGACCAGGAGCTGGATATTAACCGACTGTCTGATTACGACGTGGAT
 CATATCGTCCCACAGTCATTCTGAAAGATGACAGCATTGACAATAAGGTGCTGACCCGGAGTGACAAAAACCGAGG
 AAAGAGTGATAATGTCCCTTCAGAGGAAGTGGTCAAGAAAATGAAGAACTACTGGAGACAGCTGCTGAATGCCAAAC
 TGATCACACAGCGAAAGTTTGATAACCTGACTAAAGCTGAGAGAGGGGGTCTGTGAGAACTGGACAAAGCAGGCTTC
 ATCAAGCGACAGCTGGTGGAGACCAGACAGATCACAAAGCACGTCGCTCAGATTCTGGATAGCAGGATGAACACAAA
 GTACGATGAGAATGACAAACTGATCCGCGAAGTGAAGGTCATTACTCTGAAGTCAAACTTGTGAGCGACTTCAGAA
 AGGATTTCCAGTTCTACAAAGTCAGGGAGATCAACAATTATCACCATGCTCATGACGCATACCTGAACGCAGTGGTC
 GGGACCGCCCTGATTAAGAAAATACCCCAAACCTGGAGAGCGAATTCGTGTACGGTGACTATAAGGTGTACGATGTCAG
 AAAATGATCGCCAAGAGTGAGCAGGAAATTGAAAAGCCACCGCTAAGTATTTCTTTTACTCAAACATCATGAATT
 TCTTTAAGACTGAGATCACCCCTGGCAAATGGGGAAATCCGAAAGAGACCCTGATTGAGACTAACGGCGAGACCGGA
 GAAATCGTGTGGGACAAGGGTAGGGATTTTGCCACAGTGCAGCAAGGTCCTGTCCATGCCTCAAGTGAATATTGTCAA
 GAAAACAGAGGTGCAGACTGGCGGATTCAGTAAGGAATCAATTCTGCCAAACGGAACCTCTGATAAGCTGATCGCCC
 GAAAGAAAGACTGGGATCCCAAGAAATATGGGGTTTCGACTCCCCAACAGTGGCTTACTCTGTCTGGTGGTTCGCA
 AAGGTGGAGAAGGGGAAAAGCAAGAACTGAAATCCGTCAAGGAGCTGCTGGGTATCACTATTATGGAGAGGAGCTC
 CTTGAGAAGAACCCATCGATTTTCTGGAGGCTAAAGGCTATAAGGAAGTGAAGAAAGACCTGATCATTAAACTGC
 CAAAGTACAGCCTGTTTGTAGCTGGAAAACGGAAGGAAGCGAATGCTGGCATCCGCAGGAGAGCTGCAGAAGGGTAAT
 GAACTGGCCCTGCCTTCTAAGTACGTGAACTTCTGTATCTGGCTAGCCACTACGAGAAGCTGAAAGGCTCCCCGA
 GGATAACGAACAGAAACAGCTGTTTGTGGAGCAGCACAAAGCATTATCTGGACGAGATCATTGAACAGATTAGCGAGT
 TCTCCAAAAGAGTGATCCTGGCTGACGCAAATCTGGATAAGGTCCTGAGCGCATAACAACACAGAGATAAGCCA
 ATCAGGGAGCAGGCCGAAAATATCATTCATCTGTTCACTCTGACCAACCTGGGAGCCCTGCAGCCTTCAAGTATTT
 TGACACTACCATCGATCGGAAACGATACACATCCACTAAGGAGGTGCTGGACGCTACCCTGATTACCAGAGCATT
 CCGGCTGTATGAAACAAGGATTGACCTGTCTCAGCTGGGGGGCGACTGA

[0324] SpCas9 REC叶:SpCas9的氨基酸61-718 (SEQ ID NO:5)

ATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLSESLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYH
 LRKKLVSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLNILQVQTYNQLFEEFNINASGVDAKAII
 RLSKSRLENLIAQLPGEKKNLFGNLIALLSLGTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFL
 AAKNLSDAIILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGAS
 QEEFYKFIKPILEKMDGTEELLVKNREDLLRQRTFDNGSIPHQIHLGELHAILRRQEDFYPFKDNREKIEKILT
 FRIPYYVGPLARGNSRFAMTRKSEETITPWNFEEVVDKGSASQSFIERMTNFDKLPNEKVLPHKSLLYEYFTVYN
 ELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGYHDLK
 IIKDKDFLDNEENEDILEDIVLTLTLFEDREMIERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG
 KTIIDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGD

[0326] SpCas9 NUC叶1:SpCas9的氨基酸1-60 (SEQ ID NO:6)

[0327] MDKKYSIGLDIGTNSVGVAVITDDYKVPKFKVLGNTDRHSIKKNLIGALLFDSGETAE

[0328] SpCas9 NUC叶2:SpCas9的氨基酸719-1368 (SEQ ID NO:7)

- SLHEHIANLAGSPAIAKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGKELGSQI
LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTNRSDKNRKGSDNVP
EEVVKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKL
IREVKVITLKSCLVSDFRKDFQFYKVRINNYHHAHDAYLNAVVGITALIKKYPKLESEFVYGDYKVDVRKMIKSE
[0329] QEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIKKTEVQTG
GFSKESILPKRNSDKLIARKKDWDPKPYGGFDSPTVAYSVLVVAKEKKGSKKLKSVKELLGITIMERSSEKFNPI
FLEAKGYKEVKKDLI IKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFYLLASHYEKLGKSPEDNEQKQ
FVEQHKHYLDEIIIEQISEFSKRVI LADANLDKVL SAYNKHRDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDR
RYTSTKEVL DATLIHQSI TGLYETRIDLSQLGGD
- [0330] SpCas9 BH结构域:SpCas9的氨基酸61-94 (SEQ ID NO:8)
- [0331] ATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVD
- [0332] SpCas9 Rec1结构域1:SpCas9的氨基酸95-180 (SEQ ID NO:9)
- [0333] DSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLV DSTDKADLRLIYLALAHMIKFRGHFLIE
GDLNPDNSD
- [0334] SpCas9 Rec1结构域2:SpCas9的氨基酸309-718 (SEQ ID NO:10)
- NTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDG
TEELLVKNREDLLRKQRTFDNGSI PHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRI PYYVGPLARGNSRF
AWMTRKSEETITPWNFEEVVDKGASQSFIERMTNFDKNLPNEKVLPHKSLLEYEFTVYNELTKVKYVTEGMRKPAF
- [0335] LSGEQKKAIVDLLFKTRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKI IKDKDFLDNEENEDIL
EDIVLTTLTFEDREMIERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTI LDFLKS DGFANRNF
MQLIHDDSLTFKEDIQKAQVSGQGD
- [0336] SpCas9 Rec2结构域:SpCas9的氨基酸181-308 (SEQ ID NO:11)
- [0337] VDKLNIQLVQTYNQLFEENPINASGVDAKAIL SARLSKSRRENLI AQLPGEKKNGLFGNLI ALSGLTPNFKSNFD
LAEDAKLQLSKD TYDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRV
- [0338] SpCas9 RuvC I结构域:SpCas9的氨基酸1-59 (SEQ ID NO:12)
- [0339] MDKYSIGLDIGTNSVGWAVITDDYKVPSPKFKVLGNTDRHS IKKNLIGALLFDSGETA
- [0340] SpCas9 RuvC II结构域:SpCas9的氨基酸718-774 (SEQ ID NO:13)
- [0341] DSLHEHIANLAGSPAIAKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQ
- [0342] SpCas9 RuvC III结构域:SpCas9的氨基酸909-1098 (SEQ ID NO:14)
- SELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVRINNYHHAH
[0343] DAYLNAVVGITALIKKYPKLESEFVYGDYKVDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPL
IETNGETGEIVWDKGRDFATVRKVLSPQVNIKKT
- [0344] SpCas9 HNH结构域:SpCas9的氨基酸775-908 (SEQ ID NO:15)
- [0345] KNSRERMKRIEEGKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
IDNKVLTNRSDKNRKGSDNVPSEEVVKMKNYWRQLLNAKLITQRKFDNLTKAERGGL
- [0346] SpCas9 PI结构域:SpCas9的氨基酸1099-1368 (SEQ ID NO:16)
- EVQTTGGFSKESILPKRNSDKLIARKKDWDPKPYGGFDSPTVAYSVLVVAKEKKGSKKLKSVKELLGITIMERSSE
KNPIDFLEAKGYKEVKKDLI IKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFYLLASHYEKLGKSPEDN
[0347] EQKQLFVEQHKHYLDEIIIEQISEFSKRVI LADANLDKVL SAYNKHRDKPIREQAENI IHLFTLTNLGAPAAFKYFDT
TIDRKRYTSTKEVL DATLIHQSI TGLYETRIDLSQLGGD
- [0348] 来自嗜热链球菌的Cas9 (SEQ ID NO:17)

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

[0349]

[0350] 来自停乳链球菌的Cas9 (SEQ ID NO:18)

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

[0351]

[0352] 来自变形链球菌的Cas9 (SEQ ID NO:19)

MKKPYSIGLDIGTNSVGWAVVTDYKVPAAKMKVLGNTDKSHIEKNLLGALLFDSGNTAEDRRLKRTARRRYTRRRN
 RILYLQEIFSEEMGVVDDSFHRLSDSFLVTEDEKRGERRHPIFGNLEEEVKYHENFPTIYHLRQYLADNPEKVDLRLV
 YLALAHIIKFRGHFLIEGKFDTRNNDVQRLFQEFLLAVYDNTFENSSLQEQNVQVEEILTDKISKSAKKDRVLKLFPN
 EKSNGRFAEFLKLIQRYNEHQMDLAQLKQFIRQKLSDKYNEVFSVSKDGYAGYIDGKTNQEAIFYKYLKGLLNKIEG
 TDVGTKAPLSASMIQRYNEHQMDLAQLKQFIRQKLSDKYNEVFSVSKDGYAGYIDGKTNQEAIFYKYLKGLLNKIEG
 SGYFLDKIEREDFLRKQRTFDNGSIHQIHLQEMRAIIRROAEFYPLADNQDRIEKLTFRIPIYVVGPLARGKSD
 AWLSRKSADKITPWNFDEIVDKESSAEAFINRMTNYDLYLPNQKVLPKHSLLYEKFTVYNELTKVKYKTEQGGKTAFF
 DANMKQEIFDGVFVKYRKYVTKDKLMDLFEKEFDEFRIVDLTGLDKENKVFNASYGTYHDLCKILDKDFLDNSKNEKI
 LEDIVLTLTLFEDREMIRKRLNYSDDLTKEQVKKLERRHYTGWGRLSAELIHGIRNKESRKTILDYLIIDDGNSNRN
 FMQLINDDALSFKEEIAKAQVIGETDNLNQVVSIDIAGSPAIKKGILOSLKIVDELVKIMGHQPENIVVEMARENQFT
 NQGRRSNQQLKGLTDSIKEFGSQILKEHPVENSQIQNDRLFLYLLQNGRDMYTGEELDIDYLSQYDIDHIIPOAFI
 KDNSIDNRVLTSSKENRGKSDDVPSKDVVRKMKSYWSKLLSAKLIHQKFDNLTKAERGGLTDDDKAGFIKRQLVET
 RQITKHVARILDERFNTETDENNKIRQVKIVTLKSNLVSFRKEFELYKREINDYHHAHDAYLNAVIGKALLGVY
 PQLEPEFVYGDYPHFHGHKKNKATAKFFYSNIMNFFKDDVVRTDKNGEIIWKKDEHISNIKKVLSYPQVNIIVKKVE
 EQTGGFSKESILPKGNSDKLIPRKTCKFYWDTKKYGGFSDPIVAYSILVIADIEKGSKLLKTVKALVGVTIMEKMT
 FERDPVAFLEKRGYRNVQENI IKLPKYSLFKLENGRKRLLASARELQKGNIEVLPNHLGLTLLYHAKNIHKVDEPKH
 LDYVDKHKHDEFKELLDVSNFSKKYTLAEGNLEKIKELYAQNNGEDLKESSFINLLTFTAIGAPATFKFFFDKNID
 RKRYTSTTEILNATLIHQSI TGLYETRIDLNKLGGD

[0353]

[0354] 来自英诺克李斯特菌的Cas9 (SEQ ID NO:20)

MKKPYTIGLDIGTNSVGWAVLTDQYDLVKKRMKIAGDSEKKQIKKNFWGVRLFDEGQTAADRRMARTARRRIERRRN
 RISYLQGI FAEEMSKTDANFFCRLSDSFYVDNEKRNRRHPFFATIEEEVEYHKNYPTIYHLREELVNSSEKADLRLV
 YLALAHIIKRYGNFLIEGALDTQNTSVDGIYKQFIQTYNQVFASGIEDGSLKKLEDNKDVAKILVEKVTRKEKLERI
 LKLYPGEKSAGMFAQFISLIVGSKGNFQKPFDLIEKSDIECAKDSYEEDESLALIGDEYAEFVAAKNAYSAVVL
 SSIITVAETETNAKLSASMIERFDTHEEDLGELKAFIKLHLPKHYYEIEFSNTEKHGYAGYIDGKTKQADFYKYMKMT
 LENIEGADYFIAKIEKENFLRKQRTFDNGAIPHQLHLEEAELHQQAKYYPFLKENYDKIKSLVTFRIPIYFVGPLA
 NGQSEFAWLTRKADGEIRPWNIEEKVDFGKSAVDIEKMTNKDTYLPKENVLPKHSLCYQKYLTVNELTKVRYINDQ
 GKTSYFSGQEKEQIFNDLQKQKRVKDKLELFLRNMSHVESPTIEGLEDSFNSSYSTYHDLKLVGKIQEILDNPVN
 TEMLENIVKILTVFEDKRMIEQLQFSDVLDGVVLLKLEERRHYTGWGRLSAKLLMGIRDKQSHLTI LDYLMNDDGL
 NRNLMLINDSNLSFKSIIIEKEQVTTADKDIQSIVADLAGSPAIKKGILOSLKIVDELVSVMGYPPQTI VVEMAREN
 QTTGKGNNSRPRYKSLKAIKEFGSQILKEHPTDNQELRNRLYLYLQNGKDMYTGGDLDIHNLNYSYDIDHIVPQ
 SFITDNSIDNLVLTSSAGNREKGDVPPLEIVRKRKVFWEKLYQGNLMSKRKFDYLTKAERGGLTEADKARFIHRQL
 VETRQITKNVANILHQRFNYEKDDHGNTMKQVRIVTLKSALVSQFRKQFQLYKVRDVNDYHHAHDAYLNGVVANTLL
 KVYPQLEPEFVYGDYHQDFWFKANKATAKQFYTNIMLFFAQKDRIIDENGEILWDKYLDTVKKVMSYRQMNIIVKK
 TEIQKGEFSKATIKPKGNSSKLI PRKTNWDPMKYGGLDSPNMAVAVVIEYAKGNKLVFEKKIIRVTIMERKAFEKD
 EKAFLEEQGYRQPKVLAKLPKYTYECEGRRRMLASANEAQKGNQVLPNHLVTLHHAANCEVSDGKSLDYIESN
 REMFAELLAHVSEFAKRYTLAEANLNKINQLFEQNKEGDIKAIQAQSFVDLMAFNAMGAPASFKFFETTIERKRYNNL
 KELLNSTIYQSI TGLYESRKRLDD

[0355]

[0356] 来自金黄色葡萄球菌的Cas9 (SEQ ID NO:21)

MKRNYYILGLDIGITSVGYGIIDYETRVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQRVKKLLFDYNLL
 TDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQ
 LERLKKDGEVRSINRFKTSYVKEAKQLLKVQKAYHQDQSFIDTYIDLLETRRTRYEGPGEPSFGWKDIKEWYE
 MLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVI TRDENEKLEYEKFQI IENVFKQKKKPTLKQIAKEILVNEE
 DIKGYRVSTGKPEFTNLKVYHDIKDITARKEI IENAELLDQIAKILTIYQSSEDIQEELTNLNSLTQEIEIQISN
 LKGYTGTHNLSLKA INLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIVIN
 AI IKKYGLPNDI I IELAREKNSKDAQMINEMQKRNRQTNERIEE I IRTTGKENAKYLIEKIKLHDMQEGKCLYSLE
 AI PLEDLLNPNFYEV DHI I PRSVSFDNSFNKVLVQEQEENS KKG NRTPFQYLS SSSDSKI SYETFKKHILNLAKGKG
 RISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKE
 RNKGYKHAEDALI IANADFI FKEWKKLDKAKKVMENQMFE EKQAESMPEIETE QEYKEIFITPHQIKHIKDFKDYK
 YSHRVDKKNRELINDTLYSTRKDDKGN TLIVNNLNGLYDKDNDKLLKLINKSPEKLLMYHHPQTYQKLKLIMEQY
 GDEKNPLYKYEETGNLYTKYSKKDNGPVIKKIKYYGNKLNALHDI TDDYPNSRNKVVKLSLKPYPFRDVYLDNGVYK
 FVTVKNL DVIKKENYEVNSKCYEEAKLKKI SNQAEFIASFYNNDLIKINGELYRIGVNNDLLNRIEVNMIDITY
 REYLENMNDKRPPRI IKTIASKTQS IKKYSTDILGNLYEVKSKKHPQI IKKG

[0357]

[0358] 来自肺炎克雷白杆菌的Cas9 (SEQ ID NO:22)

MAPKKRRKVG I HGVPAADKKYSIGLDIGTNSVGWAVITDEYKVP SKKFKVLGNTDRHS IKKNLIGALLFDSGETAEA
 TRLKRRTARRRYTRRKNR ICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHL
 RKKLV DSTDKADRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQ LFEENPINASGVDAKAILSAR
 LSKSRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAIGDQYADLFLA
 AKNLSDAI LLSDI LRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQ
 EEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFY PFLKDNREKIEKILTF
 RIPYYVGPLARGNSRFAMTRKSEETITPWNFEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNE
 LTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKI
 IKDKDFLDNEENEDI LEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGK
 TILDFLKS DGFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA I KKGILQTVKVVDELVKVMGRH
 KPENIVIAMARENQTTQKGQNSRERMKRIE EGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDIN
 RLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGL
 SELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS KLVSDFRKDFQFYK VREINNYHHAH
 DAYLNAVVG TALIKKYPKLESEFVYGDYKVYDVRKMI AKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPL
 IETNGETGEIVWDKGRDFATVRKVL SMPQVNI VKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTV
 AYSVLVVAKVEKGKSKKLSVKELLGITIMERS SFEKNPIDFLEAKGYKEVKKDLI IKLPKYSLFELENGRKRMLAS
 AGELQKGNELALPSKYVNFLYLASHYEKLGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVI LADANLDKVL SA
 YNKHDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSI TGLYETRIDLSQLGGDKR
 PAATKKAGQAKKKK

[0359]

[0360] FnCas9-来自新凶手弗朗西斯菌的Cas9 (SEQ ID NO:23)

[0361] MNFKILPIAIDLGVKNTGVFSAFYQKGTSLERLDNKNKGVYELSKDSYTLMMNNRTARRHQRRGIDRKQLVKRFLKLIWTEQLNLEWDKDTQQAISFLFNRRGFSFITDGYSPYELNIVPEQVKAILMDIFDDYNGEDDLDSYLKATEQESKISEIYNKLMQKILEFKLMLKCTDIKDDKVSTKTLKEITSYEFELLADYLANYSES�KTQKFSYTDKQGNLKELSYHHDKYNIQEFLKRHATINDRILDITLLTDDLDIWNFNFEKFDKNEEKLNQEDKDHIQAHLHHFVFAVNIKSEMASGGRHRSQYFQEITNVLDENNHQEGYLKNFCENLHNKYSNLSVKNLNLIGNLSNLELPLRKYFNDKIHAKADHWDEQKFTETECHWILGEWRVGVKDQDKKDGAKYSYKDLNELKQKVTKAGLVDFLELDEPCRTIPPYLDNMRKPPKCQSLILNPKFLDNQYPNWQQYLQELKQLSIQNYLDSFETDLKVLKSSKQPYFVEYKSSNQQIASGQRDYKDLDARILQFIFDRVKASDELLLEIYFQAKKQKASSELEKLESSKLDDEVIANSQLSQILKSQHTNGIFEQGTFLHLVCKYYKQRQRARDSRLYIMPEYRYDKKLHKYNNTRGFDDNQLLTYCNHKPRQKRYQLLNDLAGVLQVSPNFKDKIGSDDDLFISKWLVEHIRGFKKACEDSLKIQKDNRGLLNHKINIARNTKGKCEKEIFNLICKIEGSEDKKGNYKHGLAYELGVLFGEPNEASKPEFDRKIKKFNSIYSAFIQQIAFAERKGNANTCAVCSADNAHRMQQIKITEPVEDNKDKIILSAKARQLPAIPTRIVDGAVKMATILAKNIVDDNQNIKQVLSAKHQLHIPIITESNAFEFEPALADVKGKSLKDRRKALERISPENIFKDKNNRIKEFAKGISAYSANLTDGDFDGAKEELDHIIPRSHKKGTLNDEANLICVTRGDNKNKGNRIFCLRDLADNYKQFETDDLEIEKKIADTIWDANKKDFKFGNYRSFINLTPQEQAFRHALFLADENPIKQAVIRAINNRNRFTVNGTQRYFAEVLANNIYLRAKKENLNTDKISFDYFGIPTIGNRGIAEIRQLYEKVSDIQAYAKGDKPQASYSHLIDAMLAFCIAADEHRNDGSIGLEIDKNYSLYPLDKNTGEVFTKDIIFSQIKITDNEFSDKKLVKKAIEGFNTHRQMTDRGIYAENYLPILIHKELNEVRKGYTWNKSEEIKIFKGKKYDIQQLNVLVYCLKFVDKPISIDIQISTLEELRNILTTNNAATAEYYYINLKTQKLHEYYIENYNTALGYKYSKEMEFRLSLAYRSERVKIKSIDDVQVLDKDSNFIIGKITLPPFKKEWQRLYREWQNTTIKDDYEFKLSFFNVKSITKLHKKVRKDFSLPISTNEGKFLVKKRTWDNFIYQILNDSDSRADGTPFIPAFDISKNEIVEAIDSFTSKNIFWLPKNIELQKVDNKNIFAIDTSKWFEVETPSDLRDIGIATIYQKIDNNSRPKVRVKLDYVIDDDSKINYFMNHSLLKSRYPKVLEILKQSTIEFESSGFNKTIKEMLGMKLAGIYNETSNN

[0362] 表5. KFERQ或KFERQ样基序

[0363]

SEQ ID NO:24	KFERQ
SEQ ID NO:25	RKVEQ
SEQ ID NO:26	QDLKF
SEQ ID NO:27	QRFFE
SEQ ID NO:28	NRVVD
SEQ ID NO:29	QRDKV
SEQ ID NO:30	QKILD
SEQ ID NO:31	QKKEL
SEQ ID NO:32	QFREL
SEQ ID NO:33	IKLDQ
SEQ ID NO:34	DVVRQ
SEQ ID NO:35	QRIVE
SEQ ID NO:36	VKELQ
SEQ ID NO:37	QKVFD
SEQ ID NO:38	QELLR
SEQ ID NO:39	VDKLN
SEQ ID NO:40	RIKEN
SEQ ID NO:41	NKKFE

[0364] 表6. 核定位信号

[0365]	SEQ ID NO:42	PKKKRKV
	SEQ ID NO:43	AVKRPAATKKAGQAKKKLD
	SEQ ID NO:44	PAAKRVKLD
	SEQ ID NO:45	MSRRRKANPTKLENAKLAKEVEN
	SEQ ID NO:46	KLKIKRPVK
	SEQ ID NO:47	KIPIK

[0366] 表7. 引物

[0367]	SEQ ID NO.	基因	序列
	SEQ ID NO:48	EMX1-T正向	TTCCAGAACCGGAGGACAAAG
	SEQ ID NO:49	EMX1-T反向	CCACCCTAGTCATTGGAGGT
	SEQ ID NO:50	EMX1-OT1正向	TTTATTATCTGCACATGTATG
	SEQ ID NO:51	EMX1-OT1反向	CTACCTGTACATCTGCACAAG
	SEQ ID NO:52	EMX1-OT2正向	ATGTGCTTCAACCCATCACG
	SEQ ID NO:53	EMX1-OT2反向	GTTGGCTTTTACAAGGATGC
	SEQ ID NO:54	FANCF-T正向	CACGGATAAAGACGCTGGGA
	SEQ ID NO:55	FANCF-T反向	TCCCAGGTGCTGACGTAGG
	SEQ ID NO:56	FANCF-OT1正向	TAGCACTGGGTGCTTAATCCG
	SEQ ID NO:57	FANCF-OT2反向	GGGTTTGGTTGGCTGCTCAT
	SEQ ID NO:58	AAVS1T2正向	ACCGGGGCCACTAGGGACAGGAT
	SEQ ID NO:59	AAVS1T2反向	AAACATCCTGTCCCTAGTGGCCC
	SEQ ID NO:60	Cas9正向	GATAAAGCAGACCTGCGGCTGATCTATC
	SEQ ID NO:61	Cas9反向	CTGGCAGCTGAGCGATCAGGTTCTC
	SEQ ID NO:62	A1AT正向	GATGCCACCTTCCCCTCTC
	SEQ ID NO:63	A1AT反向	AGTGGTGGCCTCATTCTGGA
	SEQ ID NO:64	ABCB1正向	GGCTTCACGAGAAAAGTTGATG
	SEQ ID NO:65	ABCB1反向	GGATTCACAGGCTTACCTAC

[0368] 表8. 指导RNA序列

[0369]	SEQ ID NO.	靶	序列
	SEQ ID NO:66	EMX1	GAGTCCGAGCAGAAGAAGAA
	SEQ ID NO:67	FANCF1	GGAATCCCTTCTGCAGCACC
	SEQ ID NO:68	AAVS1-T2	CACTTTGGGAGGCCGAGGC
	SEQ ID NO:69	A1A1	GAACTTGGTGATGATATCGTGCCCTCTGGC CAGTCCTGAT
	SEQ ID NO:70	ABCB1	TTTATAGTAGGATTTACACGACCAATTCCT TCATTAATCT

[0370] 实例

[0371] 材料和方法

[0372] 将以下材料和方法用于实例中描述的实验中。寡核苷酸和指导RNA合成由西格玛公司(Sigma)和Synthego公司提供。除非另有说明,否则试剂和试剂盒均购自赛默飞世尔科技公司(ThermoFisher)。

- [0373] 试剂和试剂盒列表
- [0374] 亚细胞分级试剂盒;
- [0375] 不含EDTA的蛋白酶抑制剂混合物(西格玛公司);
- [0376] NOVEX NUPAGE蛋白凝胶,Bis-Tris 4%-12%,1.5mm厚,10孔;
- [0377] 4X Laemmli缓冲液(伯乐公司(BioRad));
- [0378] NUPAGE MOPS20X SDS运行缓冲液;
- [0379] NUPAGE 20X转移缓冲液;
- [0380] NOVEX Sharp预染蛋白标准分子量标记;
- [0381] 硝酸纤维素预切印迹膜,孔径0.45 μ m;
- [0382] 一级抗体:Cas9小鼠单克隆(艾博抗公司(Abcam));
- [0383] 单克隆抗FLAG M2抗体(西格玛公司);
- [0384] 单克隆抗 α 微管蛋白(西格玛公司);
- [0385] 二级抗体:IRDYE 680RD驴抗小鼠IgG(H+L),0.1mg(Li-COR);
- [0386] HSC70小鼠单克隆(圣克鲁斯公司(Santa Cruz));
- [0387] 蛋白A琼脂糖(艾博抗公司);
- [0388] REVERTAID RT试剂盒;
- [0389] DNA酶I,不含RNA酶;
- [0390] Phusion Flash高保真PCR主混合物;
- [0391] Q5热启动高保真2X主混合物;
- [0392] Gentra Puregene试剂盒(凯杰公司(Qiagen));
- [0393] Lipofectamine LTX Plus(英杰公司(Invitrogen));
- [0394] FUGENE HD(普洛麦格公司(Promega))
- [0395] Cas9(CRISPR相关蛋白9)ELISA试剂盒(细胞生物学实验室(Cell Biolabs))
- [0396] 放线菌酮CHX(西格玛公司)
- [0397] 亮抑酶肽(西格玛公司)
- [0398] 抗CRISPR-Cas9抗体[EPR19799](ab203933)
- [0399] 抗Ki67抗体(ab15580)
- [0400] 抗裂解胱天蛋白酶3抗体(ab2302)
- [0401] 抗 γ H2A.X(磷酸S139)抗体[9F3](ab26350)
- [0402] 抗CD8 α 抗体[144B](ab17147)
- [0403] 抗CD4抗体[EPR19514]-低内毒素,无叠氮化物(ab221775)
- [0404] 表2中列出了本文所述实验中使用的引物。
- [0405] 表2.引物

基因	引物组 (F: 正向, R: 反向)	SEQ ID NO.
EMX1-T	F: TTCCAGAACCGGAGGACAAAG	48
	R: CCACCCTAGTCATTGGAGGT	49
EMX1-OT1	F: TTTATTATCTGCACATGTATG	50
	R: CTACCTGTACATCTGCACAAG	51
EMX1-OT2	F: ATGTGCTTCAACCCATCACG	52
	R: GTTGGCTTTCACAAGGATGC	53
FANCF-T	F: CACGGATAAAGACGCTGGGA	54
	R: TCCCAGGTGCTGACGTAGG	55
FANCF-OT1	F: TAGCACTGGGTGCTTAATCCG	56
	R: GGGTTTGGTTGGCTGCTCAT	57
AAVS1 T2	F: ACCGGGGCCACTAGGGACAGGAT	58
	R: AAACATCCTGTCCCTAGTGGCCC	59
Cas9	F: GATAAAGCAGACCTGCGGCTGATCTATC	60
	R: CTGGCAGCTGAGCGATCAGGTTCTC	61
A1AT	F: GATGCCACCTTCCCCTCTC	62
	R: AGTGGTGGCCTCATTCTGGA	63
ABCB1	F: GGCTTCACGAGAAAAGTTGATG	64
	R: GGATTCACAGGCTTCACCTAC	65

[0407] 表3中列出了本文所述实验中使用的指导RNA。

[0408] 表3. 指导RNA序列

靶	gRNA	SEQ ID NO.
EMX1	GAGTCCGAGCAGAAGAAGAA	66
FANCF1	GGAATCCCTTCTGCAGCACC	67
AAVS1-T2	CACTTTGGGAGGCCGAGGC	68
A1AT	GAAGTTGGTGATGATATCGT GCCCTCTGGCCAGTCCTGAT	69
ABCB1	TTTATAGTAGGATTTACACG ACCAATTCCTTCATTAATCT	70
小鼠和人 PCSK9	CAGGTTCCATGGGATGCTCT	71
小鼠 PCSK9	AGCAGCAGCGGCGGCAACAG	72

[0410] 实验程序

[0411] 细胞培养。将SV-HUC-1细胞在补充有10% (v/v) 胎牛血清 (Gibco) 和1% 抗生素 (100U/mL青霉素和100mg/L硫酸链霉素) 的F-12K培养基 (Ham的Kaighn改良) 中培养。

[0412] 将HEK293细胞在补充有10% (v/v) 胎牛血清 (Gibco) 和1% 抗生素 (100U/mL青霉素和100mg/L硫酸链霉素) 的Dulbecco改良的Eagle培养基 (英杰公司) 中培养。使所有细胞在37°C下、具有5%CO₂的加湿气氛中生长。

[0413] 将Neuro-2a细胞在补充有10% (v/v) 胎牛血清 (Gibco) 和1% 抗生素 (100U/mL青霉素和100mg/L硫酸链霉素) 的Eagle最低必需培养基 (EMEM) 中培养。

[0414] 将HCT细胞在补充有10% (v/v) 胎牛血清 (Gibco) 和1% 抗生素 (100U/mL青霉素和100mg/L硫酸链霉素) 的RPMI-1640培养基 (ATCC 30-2001) 中培养。

[0415] 将人iPSC在补充有GF-1添加剂 (按1:333稀释)、GF-2添加剂 (按1:1000稀释) 和GF-3添加剂 (按1:1000稀释) 的CELLARTIS DEF-CS基础培养基中培养。

[0416] 质粒突变. 通过G-Blocks基因片段进行突变并合成。

[0417] 实例1. Cas9相对于FaDe-Cas9: 蛋白质周转分析

[0418] 质粒转染. 通过反向转染用编码Cas9或FaDe-Cas9 (FaDe-Cas9: 具有F185N突变的野生型Cas9; 在图5A中显示为KFERQ-Cas9) 的质粒转染永生化人或小鼠细胞系。将3 μ g质粒 (Cas9或FaDe-Cas9) 与转染剂 (Lipofectamine LTX) 在OPTIMEM培养基中混合, 并在6孔板中孵育25分钟。孵育后, 将细胞分离, 计数 (50×10^4), 并重悬于2mL完全培养基中, 并添加到含有转染试剂混合物的孔中。转染后24小时, 通过GFP表达分析细胞以评估转染效率, 并在不同时间点收获细胞用于蛋白质印迹分析。

[0419] 用于蛋白质印迹的细胞裂解和蛋白质提取。用胰蛋白酶收获转染的细胞, 并以2000rpm离心5分钟。用冷PBS1X洗涤后, 将细胞沉淀悬浮于冷“2步裂解缓冲液”和蛋白酶抑制剂 (2步裂解缓冲液: 10mM KCl, 20mM Tris HCl pH 7.4, 10mM MgCl₂, 20mM EDTA, 10% 甘油, 0.8% TRITON) 中, 涡旋, 超声处理并在冰上孵育20分钟。超声程序包括两种处理: 峰值功率20.0-占空比40.0-脉冲周期50-持续时间15秒。然后将细胞裂解物用420nM NaCl孵育5分钟以从核酸中分离蛋白质, 并通过在4°C下以15,000rpm离心30分钟进行澄清。离心后, 丢弃含有DNA和膜的沉淀, 并通过NANODROP测量上清液的蛋白质浓度。

[0420] 免疫印迹。将如上所述转染的细胞中的澄清裂解物与上样缓冲液 (10% β -巯基乙醇) 混合并煮沸8分钟。将样品加载到NUPAGE Bis-Tris SDS蛋白凝胶中, 并通过在200V (MOPS1X) 下运行40分钟进行分离。然后在35V下用NUPAGE转移缓冲液加20% 甲醇将蛋白质凝胶转移到硝酸纤维素膜上 (罐转移 (tanktransfer))。将膜在1% BSA中封闭1小时, 并在4°C下与一级抗体一起孵育。用PBS吐温0.2% 洗涤3次后, 将膜与二级抗体 (在1% BSA中1:10000) 在室温下孵育1小时, 并用PBS吐温0.2% 洗涤3次。用ODYSSEY成像系统进行检测。

[0421] 结果在图6A、6B和7中示出。Cas9与FaDe-Cas9的分析表明, FaDe-Cas9的表达低于蛋白质印迹的检测水平 (图6A)。如图6B所示, Cas9中产生FaDe-Cas9的突变不会损害Cas9抗体特异性。如图7A所示, 与Cas9相比, FaDe-Cas9在短时间窗口内显示出非常低的表达。尽管Cas9在转染后8小时至72小时以增加的水平被抗Cas9检测到, 但FaDe-Cas9甚至在转染后8小时也没有被抗Cas9检测到。

[0422] 通过GFP融合蛋白表达分析蛋白质周转。将以 7×10^3 细胞密度接种到96孔板中的HEK293细胞用100ng编码Cas9-GFP-Fused或FaDe-Cas9-GFP-Fused的双启动子驱动的报告载体转染, 并在其自身启动子下mCherry表达。用FUGENE HD转染试剂进行转染。然后将板置于INCUCYTE中, 以通过测量一段时间内的GFP荧光信号来监测Cas9和FaDe-Cas9蛋白的体内水平。分析mCherry荧光以评估转染效率。图8为转染到细胞中用于转染效率评估的双报告载体的示意图。

[0423] 结果显示在图9至11中。虽然转染效率相当 (如通过mCherry荧光测量, 图10A), 但

与Cas9-GFP-Fused相比, FaDe-Cas9-GFP-Fused显示较低的表达水平(如通过GFP荧光测量, 图10B)。图12显示了在HEK、HCT、hIPSc和Neuro-2a细胞中相似的Cas9和FaDe-Cas9蛋白质周转实验的结果, 表明FaDe-Cas9的高周转与细胞类型无关。

[0424] 通过半定量RT-PCR评估mRNA稳定性。从用编码Cas9或FaDe-Cas9的质粒转染的SVHUC1细胞中分离RNA。将在不同时间点收获的细胞沉淀悬浮在1ml TRIZOL中。在室温下5分钟后, 添加200 μ L氯仿, 然后在室温下孵育3分钟。将样品在4 $^{\circ}$ C下以15,000rpm离心15分钟, 并将水相(含有RNA)收集在单独的试管中。然后通过向水相中添加500 μ L异丙醇、室温下孵育10分钟并以15,000rpm离心来沉淀RNA。将RNA以凝胶状沉淀收集在底部, 用70%乙醇洗涤并溶于无RNA酶的水中。将RNA样品用含有1 μ g RNA、1X DNA酶缓冲液和1 μ L DNA酶加超纯水的10 μ L反应物中的DNA酶处理。将混合物在37 $^{\circ}$ C孵育30分钟, 并用NANODROP测量RNA浓度。

[0425] cDNA合成。使反应在具有500ng RNA、1 μ L随机六聚体引物、4 μ L 5X反应缓冲液、1 μ L RIBOLOCKRNA酶抑制剂(20U/ μ L)、2 μ L 10mM dNTP Mix、1 μ L REVERTAID RT(200U/ μ L)和无核酶的水的20 μ L反应中进行。将反应在42 $^{\circ}$ C孵育60分钟, 并通过在70 $^{\circ}$ C加热5分钟来终止。

[0426] PCR。使用1 μ L Q5热启动高保真2X主混合物在20 μ L反应体积(0.5 μ L正向和反向引物(10 μ M)、1 μ L cDNA和水)中进行PCR。每个反应重复三次。PCR如下进行: 在95 $^{\circ}$ C下变性1分钟, 随后将在95 $^{\circ}$ C下变性30秒、在58 $^{\circ}$ C下退火30秒、在72 $^{\circ}$ C下延伸30秒, 循环24次, 最后延伸2分钟。

[0427] 结果显示在图7B中, 表明来自Cas9和FaDe-Cas9的相似的mRNA转录水平。这些数据证实Cas9和FaDe-Cas9的转染效率是相当的。

[0428] 实例2. 评价CMA对FaDe-Cas9的蛋白质周转和亚细胞定位的作用Cas9-HSC70共免疫沉淀。用编码Cas9或FaDe-Cas9的质粒转染以70%汇合接种的SV-HUC-1细胞。转染后48小时收获细胞并用CO-IP裂解缓冲液(140mM KCl、3mM MgCl₂、0.5%NONIDET P-40、20mM HEPES pH 7.4、1mM EDTA、1.5mM EGTA、蛋白酶抑制剂(无EDTA的蛋白酶抑制剂混合物))进行裂解。将细胞沉淀悬浮于新鲜制备的冷裂解缓冲液中, 并使用1mL注射器使通过25号针头5-6次。将裂解物在冰上孵育30分钟, 并在4 $^{\circ}$ C下以15,000rpm离心20分钟。将澄清的裂解物收集在新试管中, 并用NANODROP分析以测量蛋白质浓度。对800 μ g澄清的裂解物进行Cas9免疫沉淀, 在裂解缓冲液中稀释, 并在4 $^{\circ}$ C下与Cas9一级抗体(1:100)孵育过夜。第二天, 在4 $^{\circ}$ C下将免疫复合物固定在转鼓上的50 μ L蛋白GSEPHAROSE珠上持续4小时。然后将珠用CO-IP裂解缓冲液(3X)洗涤, 重悬于30 μ L SDS样品缓冲液中, 煮沸5分钟, 并用HSC70一级抗体(1:1000)进行蛋白质印迹。

[0429] 结果显示在图13A中。共免疫沉淀显示FaDe-Cas9而不是Cas9与CMA分子伴侣HSC70相互作用。如抗HSC70印迹所示, 使用抗HSC70抗体检测到FaDe-Cas9, 但未检测到Cas9。

[0430] FaDe-Lamp2A的共定位。将以 7×10^3 细胞密度接种在96孔板中的HEK293细胞用50ng Cas9-GFP-Fused或FaDe-Cas9-GFP-Fused和30ng Lamp2A-dsRed-Fused共转染。转染后24小时, 通过INCUCYTE zoom分析细胞。参见图14为转染到细胞中的质粒的示意图。

[0431] 结果显示在图13B和15中。GFP标记的Cas9或FaDe-Cas9显示绿色荧光, 而mCherry标记的Lamp2A显示红色荧光。免疫荧光信号的可视化表明FaDe-Cas9与溶酶体蛋白和CMA调节剂Lamp2A共定位在胞质溶胶中。与Lamp2A的共定位表明活性降解。

[0432] 亚细胞蛋白分级。用胰蛋白酶-EDTA收获细胞, 然后以 $500 \times g$ 离心5分钟。用冰冷的

PBS洗涤细胞沉淀,并且干燥除去并丢弃上清液。将含有蛋白酶抑制剂的冰冷细胞提取缓冲液(CEB)添加到细胞沉淀中以进行细胞质提取,并在4°C下孵育10分钟。将裂解物以500×g离心5分钟。然后将上清液(细胞质提取物)转移到冰上的干净的预冷管中。将含有蛋白酶抑制剂的冰冷膜提取缓冲液(MEB)添加到沉淀中以提取膜,并在4°C下孵育10分钟,然后在3000×g下离心5分钟。对于核提取,将每100μL含蛋白酶抑制剂加5μL 100mM CaCl₂和3μL微球菌核酸酶(300单位)的冰冷核提取缓冲液(NEB)添加到沉淀中,并在4°C孵育10分钟,然后在室温下孵育5分钟。然后通过4°C下以15,000rpm离心10分钟来使裂解物澄清。测量来自每个细胞区室的蛋白质提取物以定量蛋白质浓度,并在NUPAGE Bis-Tris SDS蛋白质凝胶上分离。

[0433] 转染后48小时,Cas9和FaDe-Cas9的亚细胞定位结果显示在图16中。在相同的亚细胞位置发现Cas9和FaDe-Cas9,这表明FaDe-Cas9中的F185N突变不影响区室化。

[0434] 实例3.酶活性、靶上和脱靶插入缺失的分析

[0435] 基因组DNA提取。使用Gentra Puregene试剂盒进行基因组DNA提取。将细胞沉淀悬浮在300μL裂解缓冲液中,并在添加蛋白质沉淀缓冲液(100μL)后在冰上孵育5分钟。然后将裂解物以14,000rpm离心10分钟;将上清液与300μL异丙醇混合以沉淀DNA,然后以14,000rpm离心10分钟。将DNA沉淀用100μL 70%乙醇洗涤,并悬浮在30μL水中。用NANODROP测量DNA浓度。

[0436] Surveyor核酸酶测定。PCR:在20μL总体积(具有100ng gDNA和10μL 2X主混合物(PHUSION Flash高保真PCR)加1μL正向和反向引物(10μM))中进行PCR。PCR如下进行:在95°C下变性3分钟,然后将在95°C下变性5秒、在58°C下退火30秒、在72°C下延伸(1min/1,000bps),循环35次,最后延伸5分钟。

[0437] 消化:通过在99°C加热5分钟使PCR产物变性,然后通过冷却至65°C持续30分钟和冷却至23°C持续30分钟再退火以形成异源双链体。

[0438] 用Surveyor核酸酶(也称为CEL核酸酶)处理杂交的异源双链体或同源双链体,该Surveyor核酸酶识别异源双链体DNA中存在的错配并在错配失真的3'侧裂解两条链。因此,将50μL反应体积中的20μL未纯化的PCR产物(约250ng)用1μL surveyor酶+1μL增强子在42°C下消化20分钟。然后通过10%丙烯酰胺凝胶电泳40分钟分离10μL经处理的DNA。

[0439] 确定Cas9和FaDe-Cas9在HEK细胞中的核酸酶效率的测定结果示于图17A中。如显示“切割”DNA的凝胶带所示,Cas9和FaDe-Cas9核酸酶活性是相当的。图17B所示的下一代测序分析证实Cas9(12.8%)和FaDe-Cas9(16.5%)的可比核酸酶效率。图18显示在hiPSC中,FaDe-Cas9具有与Cas9相当的核酸酶活性水平。

[0440] EMX1/FANCF1脱靶分析。将HEK293细胞以 60×10^4 的密度接种到12孔板中,并用800ng编码Cas9或FaDe-Cas9的质粒和200ng编码gRNA(EMX1或FANCF1)的质粒共转染。使用FUGENE HD转染试剂进行转染。转染后72小时收获细胞,并裂解以进行基因组DNA提取。用100ng gDNA、1μl PHUSION FLASH高保真PCR主混合物以及与根据来自依诺米那公司(Illumina)的制造商的建议设计的衔接子融合的正向和反向引物进行靶上和脱靶区域的PCR扩增。PCR如下进行:在95°C下变性1分钟,然后将在95°C下变性30秒、在58°C下退火30秒、并在72°C下延伸30秒,循环30次,最后延伸2分钟。纯化PCR产物并通过下一代测序进行分析。

[0441] 结果显示在图19中。虽然FaDe-Cas9与Cas9相比具有略低的靶上修饰,但是FaDe-Cas9的脱靶修饰与Cas9相比显著降低。因此,FaDe-Cas9显示比Cas9增加的标准化靶上修饰效率。

[0442] 核糖核蛋白(RNP)电穿孔。Cas9/RNP的预复合:在1.5mL试管中,将100pmol Cas9或FaDe-Cas9与120pmol合成双gRNA在10 μ L 1X Cas9缓冲液中混合,并在室温下孵育20分钟。将20 \times 10⁴密度的HEK293细胞悬浮于20 μ L电穿孔缓冲液SF中,并与RNP复合物孵育2分钟。使用4D核转染仪(龙沙公司(Lonza))对细胞进行电穿孔。核转染后,将细胞接种于具有1mL完全培养基(DMEM)的12孔板中,并在不同时间点(3小时;6小时;10小时;24小时)收获。

[0443] RNP转染。将人iPSC以20 \times 10⁴的密度接种,并用浓度递增的Cas9或FaDe-Cas9和3 μ M双gRNA转染。用Lipofectamine CRISPRMAX转染试剂进行转染。感染后48小时,裂解细胞以提取基因组DNA,并通过PCR证实缺失。

[0444] 实例4.Cas9表达对细胞的影响

[0445] 实例4.1

[0446] 用野生型Cas9转染人尿道上皮细胞(SVHUC-1)。通过ddPCR证实含有Cas9的细胞克隆(图1A,上小图)。Cas9稳定细胞系生成后四周,细胞增殖并在24、48和72小时计数。来自野生型细胞(无Cas9)、克隆1(杂合Cas9整合)和克隆3(纯合Cas9整合)的细胞计数显示下降趋势(图1A,下小图)。

[0447] 实例4.2

[0448] 用强力霉素诱导型启动子上的野生型Cas9转染小鼠。诱导后测量小鼠的体重(“iCas dox”),并与野生型小鼠(“WT水”)、具有强力霉素的野生型小鼠(“WT dox”)以及转染但不表达Cas9的小鼠(“iCas水”)进行比较。图1B中的结果显示,与不表达Cas9的小鼠相比,转染并表达野生型Cas9的小鼠表现出体重减轻。

[0449] 实例4.3

[0450] 用野生型Cas9瞬时转染人诱导的多能干细胞(hiPSC)。在Cas9瞬时表达后5周拍摄细胞的显微镜图像。如图2所示,当瞬时表达Cas9时,hiPSC丧失其未分化表型。

[0451] 实例5.Cas9相对于FaDe-Cas9的蛋白质周转分析

[0452] 实例5.1通过ELISA测定法测量的Cas9和FaDe-Cas9细胞内蛋白水平

[0453] 将20 \times 10⁴密度的HEK293细胞悬浮在20 μ L电穿孔缓冲液SF中,并与RNP复合物一起孵育2分钟。使用4D核转染仪(4D核转染仪核心单位:龙沙公司,AAF-1002B;4D核转染仪X单位:AAF-1002X;龙沙公司)对细胞进行电穿孔。核转染后,将细胞接种于具有1mL完全培养基(DMEM)的12孔板中,并在24小时收获。将细胞裂解并按照说明手册通过商业试剂盒ELISA测定(细胞生物学实验室)分析蛋白质水平:将细胞或组织裂解液在裂解缓冲液如RIPA缓冲液(25mM Tris·HCl pH7.6、150mM NaCl、1%NP-40、1%脱氧胆酸钠、0.1%SDS)中超声处理或进行匀浆并在测定前在4 $^{\circ}$ C以10,000x g离心10分钟。

[0454] 结果如图24A所示。在RNP电穿孔7.5 μ g/10⁵个细胞时,与<0.1%FaDe-Cas9相比,在24h时回收了约5%的细胞内Cas9。在电穿孔后24h,FaDe-Cas9在细胞中的丰度降低>97%。

[0455] 实例5.2蛋白质周转和降解的测量

[0456] 用GFP融合的Cas9或FaDe-Cas9以30 \times 10⁴的密度转染HEK293细胞,并通过incucyte

随时间分析GFP表达。

[0457] 在转染后12小时,将细胞用CHX (10ug/ml) 处理以抑制蛋白质合成,并且与未处理的细胞相比,在GFP信号之后测量Cas9与FaDe-Cas9的降解蛋白质。

[0458] 通过GFP表达,FaDe-Cas9不太丰富并且从未达到用Cas9观察到的细胞内蛋白质水平(图24B)。暴露于蛋白质翻译抑制剂环己酰亚胺(CHX)的细胞显示FaDe-Cas9水平随时间的下降,相反,Cas9蛋白质水平随时间保持恒定(图24C)。

[0459] 实例6.CMA对FaDe-Cas9的高蛋白质周转和蛋白亚细胞定位的作用

[0460] 实例6.1.Lamp2A敲低

[0461] 使用RNAiMAX转染试剂和OPTIMEM,用GFP融合表达载体Cas9或FaDe-Cas9加Ds-Red lamp2a载体(2:1),加gRNA(3:1)和增加剂量的siRNA(10-20-40-60-90-100ng)共转染 7×10^3 密度的HEK293细胞。作为对照,用加扰的siRNA共转染HEK293细胞。在incucyte zoom下随时间监测转染的细胞以通过GFP信号分析KD lamp2a效率和蛋白质积累。

[0462] 为了抑制CMA,通过siRNA降低溶酶体受体Lampa2a的表达。siRNA转染导致Lamp2a受体的剂量依赖性降低(图25A/B),但Cas9的蛋白质水平未改变(图25A),FaDe-Cas9显示剂量依赖性累积,表明其降解依赖于CMA(图25B)。

[0463] 实例6.2Cas9-HSC70-共免疫沉淀。

[0464] 使用MaxCyte,用20ug的FaDe-Cas9的标记标签融合的Cas9对HEK293s细胞(5000万)进行电穿孔。电穿孔后24小时,用100uM亮抑酶肽(组织蛋白酶B抑制剂)处理细胞以暂时抑制溶酶体的降解。电穿孔后48小时收获细胞,并用CO-IP裂解缓冲液(KCl 140mM、3mM MgCl₂、0.5%Nonidet P-40、20mM Hepes pH 7.4、1mM EDTA、1.5mM EGTA、蛋白酶抑制剂(无EDTA的蛋白酶抑制剂混合物,西格玛公司))进行裂解。将细胞沉淀悬浮在冷且新鲜制备的裂解缓冲液中,并使用1ml注射器通过25号针头5-6次。将裂解物在冰上孵育30分钟,并在4°C下以15,000离心20分钟。将澄清的裂解物收集在新试管中,并用Nanodrop分析以测量蛋白质浓度。对800ug澄清裂解物进行Cas9免疫沉淀,将其在裂解缓冲液中稀释,并与标记一级抗体(抗-FLAG F7425 SIGMA 1:500)一起在4°C下孵育过夜。第二天,在4°C下将免疫复合物固定在转鼓上的蛋白质G-琼脂糖珠(50u1)上持续4小时。将珠用CO-IP裂解缓冲液(3X)洗涤,重悬于SDS样品缓冲液(30u1)中,煮沸5分钟,并用Cas9和HSC70一抗(1:1000)进行蛋白质印迹。将30ug裂解物用作INPUT以确认蛋白质的总量。

[0465] 免疫沉淀显示,FaDe-Cas9与分子伴侣介导的自噬HSC70的主调节剂具有高亲和力结合。结果显示在图26中。

[0466] 实例7.Cas9相对于FaDe-Cas9体内小鼠模型

[0467] 实例7.1腺病毒构建体

[0468] 表达Cas9或FaDe-Cas9和gRNA(Ad-Cas9-gMH和Ad-Cas9-gP和Ad-FaDe-Cas9-gMH和Ad-FaDe-Cas9-gP)的腺病毒由Vector Biolabs公司(莫尔文(Malvern))产生。Adv Cas9/FaDe和gRNA分别在复制缺陷型腺病毒血清型5(dE1/E3)骨架中由鸡 β -肌动蛋白杂合体(CBh)和U6启动子表达。也产生分别由CBh和CMV启动子表达Cas9和GFP,但不表达gRNA的阴性对照腺病毒(Ad-Cas9-GFP和Ad-FaDe-Cas9-GFP)。

[0469] 实例7.2体内靶上/脱靶分析

[0470] 对于体内脱靶编辑分析,(gP)(一种靶向小鼠PCSK9基因座混杂指导RNA)因其诱导

多个脱靶突变小鼠基因组的高可能性而被选择(图27A)。

[0471] 使9至11周龄雄性小鼠接受尾静脉注射,剂量为 1×10^9 感染单位(IFU)的腺病毒(Ad-Cas9-gp或Ad-FaDe-Cas9-gp、或Ad-Cas9-GFP或Ad-FaDe-Cas9-GFP),在200 μ l磷酸盐缓冲盐水稀释液中。在病毒施用前(基线)、病毒施用后一周和终止时(病毒施用后四天或三周)采集外周血。

[0472] 在处理第7天从注射腺病毒的小鼠的肝组织提取基因组DNA,以进行插入缺失分析。通过CIRCLE-seq,通过选择读取计数高于靶上50%的位点和各种较低排名的位点(相对于靶上,最多包含6个错配)来进行靶向深度测序,从而鉴定出脱靶。对小鼠模型中gP的靶上+9个脱靶位点的基因编辑频率的评估表明,与Cas9相比,9个中的4个不同的脱靶位点显示显著降低的FaDe-Cas9基因编辑(图27B)。

[0473] 实例7.3下一代测序(NGS)

[0474] 使用磁珠纯化PCR产物,使用QuantiFlor dsDNA System试剂盒(普洛麦格公司)进行定量,标准化至每个扩增子10ng/ μ l,并合并。使用End Prep酶混合物和来自依诺米那公司的NEBNextUltra IIDNA文库制备试剂盒中的反应缓冲液对合并的样品进行末端修复和A加尾,并使用同一试剂盒中的连接主混合物和连接增强子连接到Illumina TruSeq衔接子。然后用磁珠纯化文库样品,使用PEG/NaCl SPRI溶液(KAPA Biosystems)进行大小选择,使用液滴数字PCR(BioRad)进行定量,并加载到Illumina MiSeq上进行深度测序。

[0475] 实例7.4体内基因组编辑和生存力评估

[0476] 对于体内Pcsk9基因编辑,使携带小鼠或人PSC9的单个等位基因的9至11周龄人源化PCSK9小鼠(PCSK9KIKO)接受尾静脉注射,剂量为 1×10^9 感染单位(IFU)的腺病毒(Ad-Cas9-gMH或Ad-FaDe-Cas9-gMH和Ad-Cas9-GFP或Ad-FaDe-Cas9-GFP),在200 μ l磷酸盐缓冲盐水稀释液中。在处理第7天从注射腺病毒的小鼠的肝组织提取基因组DNA,以通过NGS分别对人和小鼠PCSK9基因座进行插入缺失分析。将肝叶包括在石蜡块中,并用苏木精和曙红(H&E)、有丝分裂标记Ki6、Cas9、裂解胱天蛋白酶3、p-H2AX和CD4/CD8进行染色。

[0477] 暴露于FaDe-Cas9的小鼠肝脏由于其在肝组织中的快速降解而显示出更高的基因编辑,继而导致编辑细胞的存活率提高,如图28A所示。与Cas9相比,FaDe-Cas9的体内快速周转导致较低的肝毒性,这通过未改变的肝糖原、几乎不可检测的有丝分裂标记Ki67、较少量的浸润物和最小的细胞坏死来证实。(图28B/C)。

[0478] 此外,发现病毒介导的FaDe-Cas9的体内表达导致较低的细胞毒性T淋巴细胞免疫应答(图28D)。AdV递送后一周,Cas9在肝细胞中仍高度表达,而FaDe-Cas9无法检测到(A,E)。FaDe-Cas9的较短潜伏期导致肝细胞中较低水平的凋亡(裂解的胱天蛋白酶3 IHC;B,F)和DNA双链断裂(磷酸H2AX IHC;C,G)。虽然在鼠肝脏中对Cas9递送的免疫反应由中等数量的CD4+(记忆)淋巴细胞和许多CD8+(细胞毒性)淋巴细胞组成,但后者在FaDe-Cas9 AdV感染的肝脏中明显减少(CD4-CD8 IHC;D,H)。

序列表

<110> 阿斯利康(瑞典) 有限公司

<120> 改进的核酸酶的组合物和方法

<130> 200777-WO-PCT

<140> PCT/EP2019/073838

<141> 2019-09-06

<150> 62/728, 184

<151> 2018-09-07

<160> 87

<170> PatentIn3.5 版

<210> 1

<211> 1368

<212> PRT

<213> 酿脓链球菌

<400> 1

Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val
1 5 10 15

[0001]

Gly Trp Ala Val Ile Thr Asp Asp Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
245 250 255

[0002]

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
 370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
 385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
 405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
 420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
 435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
 450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
 465 470 475 480

[0003] Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
 485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
 500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
 515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
 530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
 545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
 565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
 580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
 595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720

[0004]

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
 835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
 850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
 865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
 885 890 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
 900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
 915 920 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
 930 935 940

[0005] Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
 945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
 965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
 980 985 990

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe
 995 1000 1005

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala
 1010 1015 1020

Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe
 1025 1030 1035

Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala
 1040 1045 1050

Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu
 1055 1060 1065

Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val
 1070 1075 1080

Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr
 1085 1090 1095

Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys
 1100 1105 1110

Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro
 1115 1120 1125

Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val
 1130 1135 1140

Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys
 1145 1150 1155

Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser
 1160 1165 1170

[0006]

Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys
 1175 1180 1185

Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu
 1190 1195 1200

Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly
 1205 1210 1215

Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val
 1220 1225 1230

Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser
 1235 1240 1245

Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys
 1250 1255 1260

His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys
 1265 1270 1275

Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala
1280 1285 1290

Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn
1295 1300 1305

Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala
1310 1315 1320

Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser
1325 1330 1335

Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr
1340 1345 1350

Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
1355 1360 1365

<210> 2

<211> 1368

<212> PRT

<213> 人工序列

[0007]

<220>

<221> 来源

<223> /注释= “人工序列的描述: 合成多肽”

<400> 2

Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Ile Thr Asp Asp Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
 100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
 115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
 130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
 145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
 165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Asn Ile Gln Leu Val Gln Thr Tyr
 180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
 195 200 205

[0008] Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
 210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
 225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
 245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
 260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
 275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
 290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
 305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
 325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
435 440 445

[0009]

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670

[0010] Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
885 890 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
900 905 910

[0011]

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
915 920 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
930 935 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
980 985 990

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe
995 1000 1005

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala
1010 1015 1020

	Lys Ser	Glu Gln	Glu Ile	Gly	Lys Ala	Thr Ala	Lys Tyr	Phe Phe				
	1025			1030			1035					
	Tyr Ser	Asn Ile	Met Asn	Phe	Phe Lys	Thr Glu	Ile Thr	Leu Ala				
	1040			1045			1050					
	Asn Gly	Glu Ile	Arg Lys	Arg	Pro Leu	Ile Glu	Thr Asn	Gly Glu				
	1055			1060			1065					
	Thr Gly	Glu Ile	Val Trp	Asp	Lys Gly	Arg Asp	Phe Ala	Thr Val				
	1070			1075			1080					
	Arg Lys	Val Leu	Ser Met	Pro	Gln Val	Asn Ile	Val Lys	Lys Thr				
	1085			1090			1095					
	Glu Val	Gln Thr	Gly Gly	Phe	Ser Lys	Glu Ser	Ile Leu	Pro Lys				
	1100			1105			1110					
	Arg Asn	Ser Asp	Lys Leu	Ile	Ala Arg	Lys Lys	Asp Trp	Asp Pro				
	1115			1120			1125					
[0012]	Lys Lys	Tyr Gly	Gly Phe	Asp	Ser Pro	Thr Val	Ala Tyr	Ser Val				
	1130			1135			1140					
	Leu Val	Val Ala	Lys Val	Glu	Lys Gly	Lys Ser	Lys Lys	Leu Lys				
	1145			1150			1155					
	Ser Val	Lys Glu	Leu Leu	Gly	Ile Thr	Ile Met	Glu Arg	Ser Ser				
	1160			1165			1170					
	Phe Glu	Lys Asn	Pro Ile	Asp	Phe Leu	Glu Ala	Lys Gly	Tyr Lys				
	1175			1180			1185					
	Glu Val	Lys Lys	Asp Leu	Ile	Ile Lys	Leu Pro	Lys Tyr	Ser Leu				
	1190			1195			1200					
	Phe Glu	Leu Glu	Asn Gly	Arg	Lys Arg	Met Leu	Ala Ser	Ala Gly				
	1205			1210			1215					
	Glu Leu	Gln Lys	Gly Asn	Glu	Leu Ala	Leu Pro	Ser Lys	Tyr Val				
	1220			1225			1230					
	Asn Phe	Leu Tyr	Leu Ala	Ser	His Tyr	Glu Lys	Leu Lys	Gly Ser				
	1235			1240			1245					

Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys
 1250 1255 1260

His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys
 1265 1270 1275

Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala
 1280 1285 1290

Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn
 1295 1300 1305

Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala
 1310 1315 1320

Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser
 1325 1330 1335

Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr
 1340 1345 1350

[0013]

Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
 1355 1360 1365

<210> 3

<211> 4131

<212> DNA

<213> 酿脓链球菌

<400> 3

```

atggctccta agaaaagcg gaaggtggac aagaaatact caatcgggct ggacatcgga      60
actaactcag tggggtgggc agtcattact gacgagtaca aagtgcccaag caagaaatTT      120
aaggtcctgg gcaacaccga taggcactcc atcaagaaaa atctgattgg ggccctgctg      180
ttcgactctg gagagacagc tgaagcaact agactgaaaa ggactgctag aaggcgtat      240
accggcgaa agaatcgcat ctgctacctg caggagatTT tctctaacga aatggccaag      300
gtggacgata gtttctttca tcggctggag gaatcattcc tggtcgagga agataagaaa      360
cacgagagac atcctatctt tggaacatt gtggacgagg tcgcttatca cgaaaaatac      420
cccaccatct atcatctgcg caagaaactg gtggactcta cagataaagc agacctgceg      480
ctgatctatc tggccctggc tcacatgatt aagttcagag gccattttct gatcgaggga      540
    
```

	gatctgaacc cagacaatag cgatgtggac aagctgttca tccagctggt ccagacatac	600
	aatcagctgt ttgaggaaaa ccctattaat gcactctggcg tggacgcaaa agccatcctg	660
	agtgccagge tgtctaagag tagaagctg gagaacctga tegetcagct gccaggcgaa	720
	aagaaaaacg gcctgtttgg aaatctgatt gcactgtcac tgggactgac acctaacttc	780
	aagagcaatt ttgatctggc cgaggacgct aaactgcagc tgagcaagga cacttatgac	840
	gatgacctgg ataacctgct ggctcagatc ggagatcagt acgcagacct gttcctggcc	900
	gctaagaatc tgtctgacgc tatectgctg agtgatattc tgcgggtgaa caccgagatt	960
	acaaaagccc ctctgtcagc tagcatgac aagagatatg acgagacca tcaggatctg	1020
	acctgctga aggcactggt gcgccagcag ctgcccgaga agtacaagga aatcttcttt	1080
	gatcagagta agaacgggta cgccggttat attgacggcg gagcttcaca ggaggaattc	1140
	tacaagttta tcaaacctat tctggagaag atggacggca ccgaggaact gctggtgaaa	1200
	ctgaatcgcg aggacctgct gcgcaagcag cggacatttg ataacggctc catccccac	1260
	cagattcacc tgggagagct gcacgcaatc ctgcgacgac aggaagactt ctaccattt	1320
	ctgaaggata accgcgagaa gatcgaaaaa attctgacct tccggatccc ttactatgtg	1380
[0014]	gggcccctgg caaggggtaa ttcccgtttt gcctggatga cacggaaatc tgaggaaca	1440
	atcactcctt ggaacttcga ggaagtggc gataaggag cttccgcaca gtctttcacc	1500
	gagagaatga caaacttcga caaaaacctg ccaaatgaga aagtctgcc taagcacagt	1560
	ctgetgtacg agtatttcac agtctataac gaactgacta aggtgaaata cgtcaccgag	1620
	gggatgagga agccccctt cctgagcggg gaacagaaga aagctatcgt ggacctgctg	1680
	tttaaaacca atcgcaaggt gacagtcaag cagctgaagg aggactactt caagaaaatt	1740
	gaatgtttcg attctgtgga gatcagtggc gtcgaagaca gatttaacgc ttctctggga	1800
	acctaccacg atctgctgaa gatcattaag gataaagact tcctggacaa cgaggaaaat	1860
	gaggatatcc tggaagacat tgtgctgacc ctgacactgt ttgaggatcg cgaaatgac	1920
	gaggaacggc tgaaaactta tgcccactctg ttcgatgaca aggtgatgaa acagctgaag	1980
	cgaagaaggt acaccggctg gggacgactg agcagaaagc tgatcaacgg cattcgggac	2040
	aaacagagtg gaaagactat cctggacttt ctgaaatcag atggcttcgc taacagaaat	2100
	tttatgcagc tgattcacga tgacagcctg acctcaaag aggatatcca gaaggcacag	2160
	gtgtccgggc aggggtgactc tctgcacgag catatcgcaa acctggccgg gtccccgcc	2220
	atcaagaaag gtattctgca gaccgtgaag gtggctgatg agctgggtgaa agtcatgggc	2280

	aggcataagc cagaaaacat cgtgattgag atggcccgcg aaaatcagac cacacagaaa	2340
	ggacagaaga acagccgcga gcggatgaaa aggatcgagg aaggcattaa ggaactggga	2400
	tcccagatcc tgaagagca ccctgtggaa aacactcagc tgcagaatga gaagctgtat	2460
	ctgtactatc tgcagaatgg gcgggatatg tacgtggacc aggagctgga tattaaccga	2520
	ctgtctgatt acgacgtgga tcatatcgtc ccacagtcac tcctgaaaga tgacagcatt	2580
	gacaataagg tgctgacccg gagtgacaaa aaccgaggaa agagtataa tgtcccttca	2640
	gaggaagtgg tcaagaaaat gaagaactac tggagacagc tgctgaatgc caaactgatc	2700
	acacagcgaa agtttgataa cctgactaaa gctgagagag ggggtctgtc agaactggac	2760
	aaagcaggct tcatcaagcg acagctgggt gagaccagac agatcacaaa gcacgtcgt	2820
	cagattctgg atagcaggat gaacacaaag tacgatgaga atgacaaaact gatccgcgaa	2880
	gtgaaggtca ttactctgaa gtcaaaactt gtgagcgact tcagaaagga tttccagttc	2940
	tacaaagtca gggagatcaa caattatcac catgctcatg acgcatacct gaacgcagtg	3000
	gtcgggaccg ccctgattaa gaaatacccc aactggaga gcaaatcgt gtacggtgac	3060
	tataaggtgt acgatgtcag aaaaatgac gccaaagtg agcaggaaat tggaaaagcc	3120
[0015]	accgctaagt atttctttta ctcaaacatc atgaatttct ttaagactga gatcacctg	3180
	gcaaatgggg aaatccgaaa gagaccactg attgagacta acggcgagac cggagaaatc	3240
	gtgtgggaca agggtaggga ttttgccaca gtgcgcaagg tcctgtccat gcctcaagt	3300
	aatattgtca agaaaacaga ggtgcagact ggcggattca gtaaggaatc aattctgcc	3360
	aaacggaact ctgataagct gatcgcgccg aagaaagact gggatcccaa gaaatatggg	3420
	ggtttcgact ccccaacagt ggcttactct gtcttggtgg tcgcaaaggt ggagaagggg	3480
	aaaagcaaga aactgaaatc cgtcaaggag ctgctgggta tcactattat ggagaggagc	3540
	tccttcgaga agaaccceat cgattttctg gaggctaaag gctataagga agtgaagaaa	3600
	gacctgatca ttaaactgcc aaagtacagc ctgtttgagc tggaaaacgg aaggaagcga	3660
	atgetggcat ccgcaggaga gctgcagaag ggtaatgaac tggccctgcc ttctaagtac	3720
	gtgaacttcc tgtatctggc tagccactac gagaagctga aaggctcccc cgaggataac	3780
	gaacagaaac agctgtttgt ggagcagcac aagcattatc tggacgagat cattgaacag	3840
	attagcgagt tctccaaaag agtgatcctg gctgacgcaa atctggataa ggtcctgagc	3900
	gcatacaaca aacacagaga taagccaatc agggagcagg ccgaaaatat cattcatctg	3960
	ttcactctga ccaacctggg agcccctgca gccttcaagt attttgacac taccatcgat	4020

	cggaaacgat acacatccac taaggaggtg ctggacgcta ccctgattca ccagagcatt	4080
	accggcctgt atgaacaag gattgacctg tctcagctgg ggggcgactg a	4131
	<210> 4	
	<211> 4131	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<221> 来源	
	<223> /注释= “人工序列的描述：合成多核苷酸”	
	<400> 4	
	atggctccta agaaaaagcg gaaggtggac aagaaatact caatcgggct ggacatcgga	60
	actaactcag tggggtgggc agtcattact gacgagtaca aagtccaag caagaaattt	120
	aagtcctg gcaacaccga taggcactcc atcaagaaaa atctgattgg ggccctgctg	180
	ttcgactctg gagagacagc tgaagcaact agactgaaaa ggactgctag aaggcgctat	240
	accggcgaa agaatcgcat ctgctacctg caggagattt tctctaacga aatggccaag	300
	gtggacgata gtttctttca tcggetggag gaatcattcc tggtcgagga agataagaaa	360
	cacgagagac atcctatctt tggaaacatt gtggacgagg tcgcttatca cgaaaaatac	420
[0016]	cccaccatct atcatctgag caagaaactg gtggactcta cagataaagc agacctgagg	480
	ctgatctatc tggccctggc tcacatgatt aagttcagag gccattttct gatcgaggga	540
	gatctgaacc cagacaatag cgatgtggac aagctgaaca tccagctggt ccagacatac	600
	aatcagctgt ttgaggaaaa ccctattaat gcatctggcg tggacgaaa agccatcctg	660
	agtgccagc tgtctaagag tagaaggctg gagaacctga tcgctcagct gccagcgaa	720
	aagaaaaacg gcctgtttgg aaatctgatt gcactgtcac tgggactgac acctaacttc	780
	aagagcaatt ttgatctggc cgaggacgct aaactgcagc tgagcaagga cacttatgac	840
	gatgacctgg ataacctgct ggctcagatc ggagatcagt acgcagacct gttcctggcc	900
	gctaagaatc tgtctgacgc tatcctgctg agtgatattc tgcgggtgaa caccgagatt	960
	acaaaagccc ctctgtcagc tagcatgac aagagatatg acgagcacca tcaggatctg	1020
	accctgctga aggcaactgg gcgccagcag ctgcccgaga agtacaagga aatcttcttt	1080
	gatcagagta agaacgggta cgccggttat attgacggcg gagcttcaca ggaggaattc	1140
	tacaagttta tcaaacctat tctggagaag atggacggca ccgaggaact gctggtgaaa	1200
	ctgaatcgcg aggacctgct gcgcaagcag cggacatttg ataacggctc catccccac	1260
	cagattcacc tgggagagct gcacgcaatc ctgcgacgac aggaagactt ctaccattt	1320

	ctgaaggata accgcgagaa gatcgaaaaa attctgacct tccggatccc ttactatgtg	1380
	gggcccctgg caaggggtaa ttcccgcctt gcctggatga cacggaaatc tgaggaaaca	1440
	atcactcctt ggaacttcca ggaagtggtc gataaggag cttccgcaca gtctttcacc	1500
	gagagaatga caaacttcca caaaaacctg ccaaatgaga aagtgtgcc taagcacagt	1560
	ctgtgttacg agtatttcac agtctataac gaactgacta aggtgaaata cgaccaccgag	1620
	gggatgagga agcccgcctt cctgagcggg gaacagaaga aagctatcgt ggacctgctg	1680
	tttaaacca atcgcaaggt gacagtcaag cagctgaagg aggactactt caagaaaatt	1740
	gaatgtttcg attctgtgga gatcagtggc gtcgaagaca gatttaacgc ttctctggga	1800
	acctaccacg atctgtgaa gatcattaag gataaagact tcttgacaa cgaggaaaat	1860
	gaggatatcc tggaagacat tgtgtgacc ctgacactgt ttgaggatcg cgaaatgatc	1920
	gaggaacggc tgaaaactta tgcccactcg ttcgatgaca aggtgatgaa acagctgaag	1980
	cgaagaaggt acaccggctg gggacgactg agcagaaagc tgatcaacgg cattcgggac	2040
	aaacagagtg gaaagactat cctggacttt ctgaaatcag atggcttcgc taacagaaat	2100
	tttatgcagc tgattcacga tgacagcctg acctcaaag aggatatcca gaaggcacag	2160
[0017]	gtgtccgggc aggggtgactc tctgcacgag catatcgcaa acctggccgg gtccccgcc	2220
	atcaagaaag gtattctgca gaccgtgaag gtggtcgatg agctggtgaa agtcatgggc	2280
	aggcataagc cagaaaacat cgtgattgag atggcccgcg aaaatcagac cacacagaaa	2340
	ggacagaaga acagccgcga gcggatgaaa aggatcgagg aaggcattaa ggaactggga	2400
	tcccagatec tgaaagagca ccctgtggaa aacactcagc tgcagaatga gaagctgtat	2460
	ctgtactatc tgcagaatgg gcgggatatg tacgtggacc aggagctgga tattaaccga	2520
	ctgtctgatt acgacgtgga tcatatcgtc ccacagtcac tctgaaaga tgacagcatt	2580
	gacaataagg tgctgaccgg gagtgacaaa aaccgaggaa agagtgataa tgtcccttca	2640
	gaggaagtgg tcaagaaaat gaagaactac tggagacagc tgctgaatgc caaactgatc	2700
	acacagcgaa agtttgataa cctgactaaa gctgagagag ggggtctgtc agaactggac	2760
	aaagcaggct tcatcaagcg acagctgggtg gagaccagac agatcacaaa gcacgtcgt	2820
	cagattctgg atagcaggat gaacacaaag tacgatgaga atgacaaact gatcccgaa	2880
	gtgaaggtca ttactctgaa gtcaaaactt gtgagcgact tcagaaagga tttccagttc	2940
	tacaaagtca gggagatcaa caattatcac catgctcatg acgcatacct gaacgcagtg	3000
	gtcgggaccg ccctgattaa gaaatcccc aaactggaga gcaaatcgt gtacgggtgac	3060

tataaggtgt acgatgtcag aaaaatgac gccaaagtg agcaggaaat tggaaaagcc	3120
accgctaagt atttctttta ctcaaacatc atgaatttct ttaagactga gatcacccctg	3180
gcaaatgggg aaatccgaaa gagaccactg attgagacta acggcgagac cggagaaaac	3240
gtgtgggaca agggtaggga ttttgccaca gtgcgcaagg tcctgtccat gcctcaagtg	3300
aatattgtca agaaaacaga ggtgcagact ggcggattca gtaaggaatc aattctgccc	3360
aaacggaact ctgataagct gatcgcccga aagaaagact gggatcccaa gaaatatggg	3420
ggtttcgact ccccaacagt ggcttactct gtcttggtgg tcgcaaaggt ggagaagggg	3480
aaaagcaaga aactgaaatc cgtcaaggag ctgctgggta tcactattat ggagaggagc	3540
tccttcgaga agaaccocat cgattttctg gaggctaaag gctataagga agtgaagaaa	3600
gacctgatca ttaaactgcc aaagtacagc ctgtttgagc tggaaaacgg aaggaagcga	3660
atgctggcat ccgcaggaga gctgcagaag ggtaatgaac tggccctgcc ttctaagtac	3720
gtgaacttcc tgtatctggc tagccactac gagaagctga aaggctcccc cgaggataac	3780
gaacagaaac agctgtttgt ggagcagcac aagcattatc tggacgagat cattgaacag	3840
attagcgagt tctcaaaaag agtgatctg gctgacgcaa atctggataa ggtcctgagc	3900
[0018] gcatacaaca aacacagaga taagccaatc agggagcagg ccgaaaatat cattcatctg	3960
ttcactctga ccaacctggg agccccctgca gccttcaagt attttgacac taccatgat	4020
cggaaacgat acacatccac taaggagtg ctggacgcta ccctgattca ccagagcatt	4080
accggcctgt atgaaacaag gattgacctg tctcagctgg ggggcgactg a	4131

<210> 5

<211> 658

<212> PRT

<213> 人工序列

<220>

<221> 来源

<223> /注释= “人工序列的描述：合成多肽”

<400> 5

Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys
1 5 10 15

Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys
20 25 30

Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu
35 40 45

Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp
50 55 60

Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys
65 70 75 80

Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu
85 90 95

Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly
100 105 110

Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Asn Ile Gln Leu
115 120 125

Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser
130 135 140

Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg
145 150 155 160

[0019]

Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly
165 170 175

Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe
180 185 190

Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys
195 200 205

Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp
210 215 220

Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile
225 230 235 240

Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro
245 250 255

Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu
260 265 270

Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys
 275 280 285

Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp
 290 295 300

Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu
 305 310 315 320

Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu
 325 330 335

Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His
 340 345 350

Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp
 355 360 365

Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu
 370 375 380

[0020] Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser
 385 390 395 400

Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp
 405 410 415

Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile
 420 425 430

Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu
 435 440 445

Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu
 450 455 460

Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu
 465 470 475 480

Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn
 485 490 495

Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile
 500 505 510

Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn
515 520 525

Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys
530 535 540

Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val
545 550 555 560

Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu
565 570 575

Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys
580 585 590

Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn
595 600 605

Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys
610 615 620

[0021]

Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp
625 630 635 640

Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln
645 650 655

Gly Asp

<210> 6

<211> 60

<212> PRT

<213> 人工序列

<220>

<221> 来源

<223> /注释= “人工序列的描述：合成多肽”

<400> 6

Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Ile Thr Asp Asp Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu
50 55 60

<210> 7

<211> 650

<212> PRT

<213> 人工序列

<220>

<221> 来源

<223> /注释=“人工序列的描述：合成多肽”

<400> 7

Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys
1 5 10 15

Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val
20 25 30

Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu
35 40 45

[0022]

Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys
50 55 60

Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu
65 70 75 80

His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr
85 90 95

Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile
100 105 110

Asn Arg Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe
115 120 125

Leu Lys Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys
130 135 140

Asn Arg Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys
145 150 155 160

Met Lys Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln
165 170 175

Arg Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu
180 185 190

Leu Asp Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln
195 200 205

Ile Thr Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys
210 215 220

Tyr Asp Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu
225 230 235 240

Lys Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys
245 250 255

Val Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn
260 265 270

[0023]

Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser
275 280 285

Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile
290 295 300

Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe
305 310 315 320

Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn
325 330 335

Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly
340 345 350

Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg Lys Val
355 360 365

Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu Val Gln Thr
370 375 380

Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys
 385 390 395 400

Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe
 405 410 415

Asp Ser Pro Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu
 420 425 430

Lys Gly Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile
 435 440 445

Thr Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu
 450 455 460

Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu
 465 470 475 480

Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu
 485 490 495

[0024]

Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser
 500 505 510

Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys
 515 520 525

Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His
 530 535 540

Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys
 545 550 555 560

Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr
 565 570 575

Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile
 580 585 590

His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr
 595 600 605

Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val
 610 615 620

Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr
625 630 635 640

Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
645 650

<210> 8
<211> 34
<212> PRT
<213> 人工序列

<220>
<221> 来源
<223> /注释= “人工序列的描述：合成多肽”

<400> 8
Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys
1 5 10 15

Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys
20 25 30

[0025] Val Asp

<210> 9
<211> 86
<212> PRT
<213> 人工序列

<220>
<221> 来源
<223> /注释= “人工序列的描述：合成多肽”

<400> 9
Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp
1 5 10 15

Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val
20 25 30

Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu
35 40 45

Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu
50 55 60

Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu
65 70 75 80

Asn Pro Asp Asn Ser Asp
85

<210> 10

<211> 410

<212> PRT

<213> 人工序列

<220>

<221> 来源

<223> /注释=“人工序列的描述：合成多肽”

<400> 10

Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met Ile Lys Arg
1 5 10 15

Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala Leu Val Arg
20 25 30

Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys
35 40 45

[0026]

Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe
50 55 60

Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly Thr Glu Glu
65 70 75 80

Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr
85 90 95

Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly Glu Leu His
100 105 110

Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn
115 120 125

Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val
130 135 140

Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met Thr Arg Lys
145 150 155 160

Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val Val Asp Lys
 165 170 175

Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn Phe Asp Lys
 180 185 190

Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu Leu Tyr Glu
 195 200 205

Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr Val Thr Glu
 210 215 220

Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile
 225 230 235 240

Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val Lys Gln Leu
 245 250 255

Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser Val Glu Ile
 260 265 270

[0027] Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr Tyr His Asp
 275 280 285

Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn
 290 295 300

Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu Phe Glu Asp
 305 310 315 320

Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His Leu Phe Asp
 325 330 335

Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly
 340 345 350

Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly
 355 360 365

Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn
 370 375 380

Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile
 385 390 395 400

Gln Lys Ala Gln Val Ser Gly Gln Gly Asp
405 410

<210> 11
<211> 128
<212> PRT
<213> 人工序列

<220>
<221> 来源
<223> /注释= “人工序列的描述: 合成多肽”

<400> 11
Val Asp Lys Leu Asn Ile Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe
1 5 10 15

Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys Ala Ile Leu
20 25 30

Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln
35 40 45

[0028] Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu
50 55 60

Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu
65 70 75 80

Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp
85 90 95

Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala
100 105 110

Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile Leu Arg Val
115 120 125

<210> 12
<211> 59
<212> PRT
<213> 人工序列

<220>
<221> 来源
<223> /注释= “人工序列的描述: 合成多肽”

<400> 12

Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Ile Thr Asp Asp Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala
50 55

<210> 13

<211> 57

<212> PRT

<213> 人工序列

<220>

<221> 来源

<223> /注释= “人工序列的描述：合成多肽”

<400> 13

Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile
1 5 10 15

[0029]

Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys
20 25 30

Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg
35 40 45

Glu Asn Gln Thr Thr Gln Lys Gly Gln
50 55

<210> 14

<211> 190

<212> PRT

<213> 人工序列

<220>

<221> 来源

<223> /注释= “人工序列的描述：合成多肽”

<400> 14

Ser Glu Leu Asp Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr
1 5 10 15

Arg Gln Ile Thr Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn
20 25 30

Thr Lys Tyr Asp Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile
35 40 45

Thr Leu Lys Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe
50 55 60

Tyr Lys Val Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr
65 70 75 80

Leu Asn Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu
85 90 95

Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys
100 105 110

Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr
115 120 125

Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu
130 135 140

[0030]

Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu
145 150 155 160

Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg
165 170 175

Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr
180 185 190

<210> 15

<211> 134

<212> PRT

<213> 人工序列

<220>

<221> 来源

<223> /注释=“人工序列的描述：合成多肽”

<400> 15

Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu
1 5 10 15

Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu
20 25 30

Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met
35 40 45

Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val
50 55 60

Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn
65 70 75 80

Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn Val
85 90 95

Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg Gln Leu
100 105 110

Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu Thr Lys
115 120 125

Ala Glu Arg Gly Gly Leu
130

[0031]

<210> 16

<211> 270

<212> PRT

<213> 人工序列

<220>

<221> 来源

<223> /注释= “人工序列的描述: 合成多肽”

<400> 16

Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg
1 5 10 15

Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys
20 25 30

Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val Leu Val Val
35 40 45

Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser Val Lys Glu
50 55 60

Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro
65 70 75 80

Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu
85 90 95

Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg
100 105 110

Lys Arg Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu
115 120 125

Ala Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr
130 135 140

Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe
145 150 155 160

Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser
165 170 175

Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val
180 185 190

[0032]

Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala
195 200 205

Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala
210 215 220

Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser
225 230 235 240

Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly
245 250 255

Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
260 265 270

<210> 17

<211> 1409

<212> PRT

<213> 嗜热链球菌

<400> 17

Met Leu Phe Asn Lys Cys Ile Ile Ile Ser Ile Asn Leu Asp Phe Ser
1 5 10 15

Asn Lys Glu Lys Cys Met Thr Lys Pro Tyr Ser Ile Gly Leu Asp Ile
 20 25 30

Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Asn Tyr Lys Val
 35 40 45

Pro Ser Lys Lys Met Lys Val Leu Gly Asn Thr Ser Lys Lys Tyr Ile
 50 55 60

Lys Lys Asn Leu Leu Gly Val Leu Leu Phe Asp Ser Gly Ile Thr Ala
 65 70 75 80

Glu Gly Arg Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg
 85 90 95

Arg Asn Arg Ile Leu Tyr Leu Gln Glu Ile Phe Ser Thr Glu Met Ala
 100 105 110

Thr Leu Asp Asp Ala Phe Phe Gln Arg Leu Asp Asp Ser Phe Leu Val
 115 120 125

[0033]

Pro Asp Asp Lys Arg Asp Ser Lys Tyr Pro Ile Phe Gly Asn Leu Val
 130 135 140

Glu Glu Lys Val Tyr His Asp Glu Phe Pro Thr Ile Tyr His Leu Arg
 145 150 155 160

Lys Tyr Leu Ala Asp Ser Thr Lys Lys Ala Asp Leu Arg Leu Val Tyr
 165 170 175

Leu Ala Leu Ala His Met Ile Lys Tyr Arg Gly His Phe Leu Ile Glu
 180 185 190

Gly Glu Phe Asn Ser Lys Asn Asn Asp Ile Gln Lys Asn Phe Gln Asp
 195 200 205

Phe Leu Asp Thr Tyr Asn Ala Ile Phe Glu Ser Asp Leu Ser Leu Glu
 210 215 220

Asn Ser Lys Gln Leu Glu Glu Ile Val Lys Asp Lys Ile Ser Lys Leu
 225 230 235 240

Glu Lys Lys Asp Arg Ile Leu Lys Leu Phe Pro Gly Glu Lys Asn Ser
 245 250 255

Gly Ile Phe Ser Glu Phe Leu Lys Leu Ile Val Gly Asn Gln Ala Asp
 260 265 270

Phe Arg Lys Cys Phe Asn Leu Asp Glu Lys Ala Ser Leu His Phe Ser
 275 280 285

Lys Glu Ser Tyr Asp Glu Asp Leu Glu Thr Leu Leu Gly Tyr Ile Gly
 290 295 300

Asp Asp Tyr Ser Asp Val Phe Leu Lys Ala Lys Lys Leu Tyr Asp Ala
 305 310 315 320

Ile Leu Leu Ser Gly Phe Leu Thr Val Thr Asp Asn Glu Thr Glu Ala
 325 330 335

Pro Leu Ser Ser Ala Met Ile Lys Arg Tyr Asn Glu His Lys Glu Asp
 340 345 350

[0034] Leu Ala Leu Leu Lys Glu Tyr Ile Arg Asn Ile Ser Leu Lys Thr Tyr
 355 360 365

Asn Glu Val Phe Lys Asp Asp Thr Lys Asn Gly Tyr Ala Gly Tyr Ile
 370 375 380

Asp Gly Lys Thr Asn Gln Glu Asp Phe Tyr Val Tyr Leu Lys Asn Leu
 385 390 395 400

Leu Ala Glu Phe Glu Gly Ala Asp Tyr Phe Leu Glu Lys Ile Asp Arg
 405 410 415

Glu Asp Phe Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro
 420 425 430

Tyr Gln Ile His Leu Gln Glu Met Arg Ala Ile Leu Asp Lys Gln Ala
 435 440 445

Lys Phe Tyr Pro Phe Leu Ala Lys Asn Lys Glu Arg Ile Glu Lys Ile
 450 455 460

Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn
 465 470 475 480

Ser Asp Phe Ala Trp Ser Ile Arg Lys Arg Asn Glu Lys Ile Thr Pro
485 490 495

Trp Asn Phe Glu Asp Val Ile Asp Lys Glu Ser Ser Ala Glu Ala Phe
500 505 510

Ile Asn Arg Met Thr Ser Phe Asp Leu Tyr Leu Pro Glu Glu Lys Val
515 520 525

Leu Pro Lys His Ser Leu Leu Tyr Glu Thr Phe Asn Val Tyr Asn Glu
530 535 540

Leu Thr Lys Val Arg Phe Ile Ala Glu Ser Met Arg Asp Tyr Gln Phe
545 550 555 560

Leu Asp Ser Lys Gln Lys Lys Asp Ile Val Arg Leu Tyr Phe Lys Asp
565 570 575

Lys Arg Lys Val Thr Asp Lys Asp Ile Ile Glu Tyr Leu His Ala Ile
580 585 590

[0035]

Tyr Gly Tyr Asp Gly Ile Glu Leu Lys Gly Ile Glu Lys Gln Phe Asn
595 600 605

Ser Ser Leu Ser Thr Tyr His Asp Leu Leu Asn Ile Ile Asn Asp Lys
610 615 620

Glu Phe Leu Asp Asp Ser Ser Asn Glu Ala Ile Ile Glu Glu Ile Ile
625 630 635 640

His Thr Leu Thr Ile Phe Glu Asp Arg Glu Met Ile Lys Gln Arg Leu
645 650 655

Ser Lys Phe Glu Asn Ile Phe Asp Lys Ser Val Leu Lys Lys Leu Ser
660 665 670

Arg Arg His Tyr Thr Gly Trp Gly Lys Leu Ser Ala Lys Leu Ile Asn
675 680 685

Gly Ile Arg Asp Glu Lys Ser Gly Asn Thr Ile Leu Asp Tyr Leu Ile
690 695 700

Asp Asp Gly Ile Ser Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp
 705 710 715 720

Ala Leu Ser Phe Lys Lys Lys Ile Gln Lys Ala Gln Ile Ile Gly Asp
 725 730 735

Glu Asp Lys Gly Asn Ile Lys Glu Val Val Lys Ser Leu Pro Gly Ser
 740 745 750

Pro Ala Ile Lys Lys Gly Ile Leu Gln Ser Ile Lys Ile Val Asp Glu
 755 760 765

Leu Val Lys Val Met Gly Gly Arg Lys Pro Glu Ser Ile Val Val Glu
 770 775 780

Met Ala Arg Glu Asn Gln Tyr Thr Asn Gln Gly Lys Ser Asn Ser Gln
 785 790 795 800

Gln Arg Leu Lys Arg Leu Glu Lys Ser Leu Lys Glu Leu Gly Ser Lys
 805 810 815

[0036] Ile Leu Lys Glu Asn Ile Pro Ala Lys Leu Ser Lys Ile Asp Asn Asn
 820 825 830

Ala Leu Gln Asn Asp Arg Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Lys
 835 840 845

Asp Met Tyr Thr Gly Asp Asp Leu Asp Ile Asp Arg Leu Ser Asn Tyr
 850 855 860

Asp Ile Asp His Ile Ile Pro Gln Ala Phe Leu Lys Asp Asn Ser Ile
 865 870 875 880

Asp Asn Lys Val Leu Val Ser Ser Ala Ser Asn Arg Gly Lys Ser Asp
 885 890 895

Asp Phe Pro Ser Leu Glu Val Val Lys Lys Arg Lys Thr Phe Trp Tyr
 900 905 910

Gln Leu Leu Lys Ser Lys Leu Ile Ser Gln Arg Lys Phe Asp Asn Leu
 915 920 925

Thr Lys Ala Glu Arg Gly Gly Leu Leu Pro Glu Asp Lys Ala Gly Phe
 930 935 940

Ile Gln Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys His Val Ala
945 950 955 960

Arg Leu Leu Asp Glu Lys Phe Asn Asn Lys Lys Asp Glu Asn Asn Arg
965 970 975

Ala Val Arg Thr Val Lys Ile Ile Thr Leu Lys Ser Thr Leu Val Ser
980 985 990

Gln Phe Arg Lys Asp Phe Glu Leu Tyr Lys Val Arg Glu Ile Asn Asp
995 1000 1005

Phe His His Ala His Asp Ala Tyr Leu Asn Ala Val Ile Ala Ser
1010 1015 1020

Ala Leu Leu Lys Lys Tyr Pro Lys Leu Glu Pro Glu Phe Val Tyr
1025 1030 1035

Gly Asp Tyr Pro Lys Tyr Asn Ser Phe Arg Glu Arg Lys Ser Ala
1040 1045 1050

[0037]

Thr Glu Lys Val Tyr Phe Tyr Ser Asn Ile Met Asn Ile Phe Lys
1055 1060 1065

Lys Ser Ile Ser Leu Ala Asp Gly Arg Val Ile Glu Arg Pro Leu
1070 1075 1080

Ile Glu Val Asn Glu Glu Thr Gly Glu Ser Val Trp Asn Lys Glu
1085 1090 1095

Ser Asp Leu Ala Thr Val Arg Arg Val Leu Ser Tyr Pro Gln Val
1100 1105 1110

Asn Val Val Lys Lys Val Glu Glu Gln Asn His Gly Leu Asp Arg
1115 1120 1125

Gly Lys Pro Lys Gly Leu Phe Asn Ala Asn Leu Ser Ser Lys Pro
1130 1135 1140

Lys Pro Asn Ser Asn Glu Asn Leu Val Gly Ala Lys Glu Tyr Leu
1145 1150 1155

	Asp	Pro	Lys	Lys	Tyr	Gly	Gly	Tyr	Ala	Gly	Ile	Ser	Asn	Ser	Phe
	1160						1165					1170			
	Ala	Val	Leu	Val	Lys	Gly	Thr	Ile	Glu	Lys	Gly	Ala	Lys	Lys	Lys
	1175						1180					1185			
	Ile	Thr	Asn	Val	Leu	Glu	Phe	Gln	Gly	Ile	Ser	Ile	Leu	Asp	Arg
	1190						1195					1200			
	Ile	Asn	Tyr	Arg	Lys	Asp	Lys	Leu	Asn	Phe	Leu	Leu	Glu	Lys	Gly
	1205						1210					1215			
	Tyr	Lys	Asp	Ile	Glu	Leu	Ile	Ile	Glu	Leu	Pro	Lys	Tyr	Ser	Leu
	1220						1225					1230			
	Phe	Glu	Leu	Ser	Asp	Gly	Ser	Arg	Arg	Met	Leu	Ala	Ser	Ile	Leu
	1235						1240					1245			
	Ser	Thr	Asn	Asn	Lys	Arg	Gly	Glu	Ile	His	Lys	Gly	Asn	Gln	Ile
	1250						1255					1260			
[0038]	Phe	Leu	Ser	Gln	Lys	Phe	Val	Lys	Leu	Leu	Tyr	His	Ala	Lys	Arg
	1265						1270					1275			
	Ile	Ser	Asn	Thr	Ile	Asn	Glu	Asn	His	Arg	Lys	Tyr	Val	Glu	Asn
	1280						1285					1290			
	His	Lys	Lys	Glu	Phe	Glu	Glu	Leu	Phe	Tyr	Tyr	Ile	Leu	Glu	Phe
	1295						1300					1305			
	Asn	Glu	Asn	Tyr	Val	Gly	Ala	Lys	Lys	Asn	Gly	Lys	Leu	Leu	Asn
	1310						1315					1320			
	Ser	Ala	Phe	Gln	Ser	Trp	Gln	Asn	His	Ser	Ile	Asp	Glu	Leu	Cys
	1325						1330					1335			
	Ser	Ser	Phe	Ile	Gly	Pro	Thr	Gly	Ser	Glu	Arg	Lys	Gly	Leu	Phe
	1340						1345					1350			
	Glu	Leu	Thr	Ser	Arg	Gly	Ser	Ala	Ala	Asp	Phe	Glu	Phe	Leu	Gly
	1355						1360					1365			
	Val	Lys	Ile	Pro	Arg	Tyr	Arg	Asp	Tyr	Thr	Pro	Ser	Ser	Leu	Leu
	1370						1375					1380			

Lys Asp Ala Thr Leu Ile His Gln Ser Val Thr Gly Leu Tyr Glu
1385 1390 1395

Thr Arg Ile Asp Leu Ala Lys Leu Gly Glu Gly
1400 1405

<210> 18

<211> 1371

<212> PRT

<213> 停乳链球菌

<400> 18

Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Ile Thr Asp Asp Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

[0039]

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Arg
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Ser Glu Met Ser Lys Val Asp Asp Ser
85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Ala Asp
130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175

Asp Asn Ser Asp Met Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Arg Val Asp Ala
195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Arg Asn Gly Leu Phe Gly Asn
225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
275 280 285

[0040]

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
290 295 300

Ile Leu Arg Val Asn Ser Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
370 375 380

Gly Thr Glu Glu Leu Leu Ala Lys Leu Asn Arg Glu Asp Leu Leu Arg
385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
 405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
 420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
 435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
 450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
 465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
 485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
 500 505 510

[0041] Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
 515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Glu Phe Leu Ser Gly Lys Gln
 530 535 540

Lys Glu Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
 545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
 565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
 580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
 595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
 610 615 620

Leu Phe Glu Asp Lys Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
 625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg His Tyr
645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685

Ala Asn Arg Asn Phe Ile Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700

Lys Glu Ala Ile Gln Lys Ala Gln Val Ser Gly Gln Gly His Ser Leu
705 710 715 720

His Glu Gln Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735

Ile Leu Gln Ser Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750

[0042]

His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr
755 760 765

Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu
770 775 780

Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val
785 790 795 800

Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln
805 810 815

Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu
820 825 830

Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Ile Lys Asp
835 840 845

Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly
850 855 860

Lys Ser Asp Asp Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn
 865 870 875 880

Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe
 885 890 895

Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys
 900 905 910

Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys
 915 920 925

His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu
 930 935 940

Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys
 945 950 955 960

Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu
 965 970 975

[0043] Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val
 980 985 990

Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val
 995 1000 1005

Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys
 1010 1015 1020

Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Arg Phe Phe Tyr
 1025 1030 1035

Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn
 1040 1045 1050

Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Glu Glu Thr
 1055 1060 1065

Gly Glu Ile Val Trp Asn Lys Gly Arg Asp Phe Ala Thr Val Arg
 1070 1075 1080

Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu
 1085 1090 1095

Val	Gln	Thr	Gly	Ala	Leu	Thr	Asn	Glu	Ser	Ile	Tyr	Ala	Arg	Gly
1100						1105					1110			
Ser	Phe	Asp	Lys	Leu	Ile	Ser	Arg	Lys	His	Arg	Phe	Glu	Ser	Ser
1115						1120					1125			
Lys	Tyr	Gly	Gly	Phe	Gly	Ser	Pro	Thr	Val	Thr	Tyr	Ser	Val	Leu
1130						1135					1140			
Val	Val	Ala	Lys	Ser	Lys	Val	Gln	Asp	Gly	Lys	Val	Lys	Lys	Ile
1145						1150					1155			
Lys	Thr	Gly	Lys	Glu	Leu	Ile	Gly	Ile	Thr	Leu	Leu	Asp	Lys	Leu
1160						1165					1170			
Val	Phe	Glu	Lys	Asn	Pro	Leu	Lys	Phe	Ile	Glu	Asp	Lys	Gly	Tyr
1175						1180					1185			
Gly	Asn	Val	Gln	Ile	Asp	Lys	Cys	Ile	Lys	Leu	Pro	Lys	Tyr	Ser
1190						1195					1200			
Leu	Phe	Glu	Phe	Glu	Asn	Gly	Thr	Arg	Arg	Met	Leu	Ala	Ser	Val
1205						1210					1215			
Met	Ala	Asn	Asn	Asn	Ser	Arg	Gly	Asp	Leu	Gln	Lys	Ala	Asn	Glu
1220						1225					1230			
Met	Phe	Leu	Pro	Ala	Lys	Leu	Val	Thr	Leu	Leu	Tyr	His	Ala	His
1235						1240					1245			
Lys	Ile	Glu	Ser	Ser	Lys	Glu	Leu	Glu	His	Glu	Ala	Tyr	Ile	Leu
1250						1255					1260			
Asp	His	Tyr	Asn	Asp	Leu	Tyr	Gln	Leu	Leu	Ser	Tyr	Ile	Glu	Arg
1265						1270					1275			
Phe	Ala	Ser	Leu	Tyr	Val	Asp	Val	Glu	Lys	Asn	Ile	Ser	Lys	Val
1280						1285					1290			
Lys	Glu	Leu	Phe	Ser	Asn	Ile	Glu	Ser	Tyr	Ser	Ile	Ser	Glu	Ile
1295						1300					1305			

[0044]

Cys Ser Ser Val Ile Asn Leu Leu Thr Leu Thr Ala Ser Gly Ala
1310 1315 1320

Pro Ala Asp Phe Lys Phe Leu Gly Thr Thr Ile Pro Arg Lys Arg
1325 1330 1335

Tyr Gly Ser Pro Gln Ser Ile Leu Ser Ser Thr Leu Ile His Gln
1340 1345 1350

Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu
1355 1360 1365

Gly Gly Asp
1370

<210> 19

<211> 1345

<212> PRT

<213> 变形链球菌

<400> 19

[0045] Met Lys Lys Pro Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Val Thr Asp Asp Tyr Lys Val Pro Ala Lys Lys Met
20 25 30

Lys Val Leu Gly Asn Thr Asp Lys Ser His Ile Glu Lys Asn Leu Leu
35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Asn Thr Ala Glu Asp Arg Arg Leu
50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Arg Asn Arg Ile Leu
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Glu Glu Met Gly Lys Val Asp Asp Ser
85 90 95

Phe Phe His Arg Leu Glu Asp Ser Phe Leu Val Thr Glu Asp Lys Arg
100 105 110

Gly Glu Arg His Pro Ile Phe Gly Asn Leu Glu Glu Glu Val Lys Tyr
115 120 125

His Glu Asn Phe Pro Thr Ile Tyr His Leu Arg Gln Tyr Leu Ala Asp
 130 135 140

Asn Pro Glu Lys Val Asp Leu Arg Leu Val Tyr Leu Ala Leu Ala His
 145 150 155 160

Ile Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Lys Phe Asp Thr
 165 170 175

Arg Asn Asn Asp Val Gln Arg Leu Phe Gln Glu Phe Leu Ala Val Tyr
 180 185 190

Asp Asn Thr Phe Glu Asn Ser Ser Leu Gln Glu Gln Asn Val Gln Val
 195 200 205

Glu Glu Ile Leu Thr Asp Lys Ile Ser Lys Ser Ala Lys Lys Asp Arg
 210 215 220

Val Leu Lys Leu Phe Pro Asn Glu Lys Ser Asn Gly Arg Phe Ala Glu
 225 230 235 240

[0046] Phe Leu Lys Leu Ile Val Gly Asn Gln Ala Asp Phe Lys Lys His Phe
 245 250 255

Glu Leu Glu Glu Lys Ala Pro Leu Gln Phe Ser Lys Asp Thr Tyr Glu
 260 265 270

Glu Glu Leu Glu Val Leu Leu Ala Gln Ile Gly Asp Asn Tyr Ala Glu
 275 280 285

Leu Phe Leu Ser Ala Lys Lys Leu Tyr Asp Ser Ile Leu Leu Ser Gly
 290 295 300

Ile Leu Thr Val Thr Asp Val Gly Thr Lys Ala Pro Leu Ser Ala Ser
 305 310 315 320

Met Ile Gln Arg Tyr Asn Glu His Gln Met Asp Leu Ala Gln Leu Lys
 325 330 335

Gln Phe Ile Arg Gln Lys Leu Ser Asp Lys Tyr Asn Glu Val Phe Ser
 340 345 350

Asp Val Ser Lys Asp Gly Tyr Ala Gly Tyr Ile Asp Gly Lys Thr Asn
 355 360 365

Gln Glu Ala Phe Tyr Lys Tyr Leu Lys Gly Leu Leu Asn Lys Ile Glu
370 375 380

Gly Ser Gly Tyr Phe Leu Asp Lys Ile Glu Arg Glu Asp Phe Leu Arg
385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
405 410 415

Gln Glu Met Arg Ala Ile Ile Arg Arg Gln Ala Glu Phe Tyr Pro Phe
420 425 430

Leu Ala Asp Asn Gln Asp Arg Ile Glu Lys Leu Leu Thr Phe Arg Ile
435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Lys Ser Asp Phe Ala Trp
450 455 460

Leu Ser Arg Lys Ser Ala Asp Lys Ile Thr Pro Trp Asn Phe Asp Glu
465 470 475 480

[0047]

Ile Val Asp Lys Glu Ser Ser Ala Glu Ala Phe Ile Asn Arg Met Thr
485 490 495

Asn Tyr Asp Leu Tyr Leu Pro Asn Gln Lys Val Leu Pro Lys His Ser
500 505 510

Leu Leu Tyr Glu Lys Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525

Tyr Lys Thr Glu Gln Gly Lys Thr Ala Phe Phe Asp Ala Asn Met Lys
530 535 540

Gln Glu Ile Phe Asp Gly Val Phe Lys Val Tyr Arg Lys Val Thr Lys
545 550 555 560

Asp Lys Leu Met Asp Phe Leu Glu Lys Glu Phe Asp Glu Phe Arg Ile
565 570 575

Val Asp Leu Thr Gly Leu Asp Lys Glu Asn Lys Val Phe Asn Ala Ser
580 585 590

Tyr Gly Thr Tyr His Asp Leu Cys Lys Ile Leu Asp Lys Asp Phe Leu
 595 600 605

Asp Asn Ser Lys Asn Glu Lys Ile Leu Glu Asp Ile Val Leu Thr Leu
 610 615 620

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

Ser Asp Leu Leu Thr Lys Glu Gln Val Lys Lys Leu Glu Arg Arg His
 645 650 655

Tyr Thr Gly Trp Gly Arg Leu Ser Ala Glu Leu Ile His Gly Ile Arg
 660 665 670

Asn Lys Glu Ser Arg Lys Thr Ile Leu Asp Tyr Leu Ile Asp Asp Gly
 675 680 685

Asn Ser Asn Arg Asn Phe Met Gln Leu Ile Asn Asp Asp Ala Leu Ser
 690 695 700

[0048] Phe Lys Glu Glu Ile Ala Lys Ala Gln Val Ile Gly Glu Thr Asp Asn
 705 710 715 720

Leu Asn Gln Val Val Ser Asp Ile Ala Gly Ser Pro Ala Ile Lys Lys
 725 730 735

Gly Ile Leu Gln Ser Leu Lys Ile Val Asp Glu Leu Val Lys Ile Met
 740 745 750

Gly His Gln Pro Glu Asn Ile Val Val Glu Met Ala Arg Glu Asn Gln
 755 760 765

Phe Thr Asn Gln Gly Arg Arg Asn Ser Gln Gln Arg Leu Lys Gly Leu
 770 775 780

Thr Asp Ser Ile Lys Glu Phe Gly Ser Gln Ile Leu Lys Glu His Pro
 785 790 795 800

Val Glu Asn Ser Gln Leu Gln Asn Asp Arg Leu Phe Leu Tyr Tyr Leu
 805 810 815

Gln Asn Gly Arg Asp Met Tyr Thr Gly Glu Glu Leu Asp Ile Asp Tyr
 820 825 830

Leu Ser Gln Tyr Asp Ile Asp His Ile Ile Pro Gln Ala Phe Ile Lys
835 840 845

Asp Asn Ser Ile Asp Asn Arg Val Leu Thr Ser Ser Lys Glu Asn Arg
850 855 860

Gly Lys Ser Asp Asp Val Pro Ser Lys Asp Val Val Arg Lys Met Lys
865 870 875 880

Ser Tyr Trp Ser Lys Leu Leu Ser Ala Lys Leu Ile Thr Gln Arg Lys
885 890 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Thr Asp Asp Asp
900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
915 920 925

Lys His Val Ala Arg Ile Leu Asp Glu Arg Phe Asn Thr Glu Thr Asp
930 935 940

[0049]

Glu Asn Asn Lys Lys Ile Arg Gln Val Lys Ile Val Thr Leu Lys Ser
945 950 955 960

Asn Leu Val Ser Asn Phe Arg Lys Glu Phe Glu Leu Tyr Lys Val Arg
965 970 975

Glu Ile Asn Asp Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
980 985 990

Ile Gly Lys Ala Leu Leu Gly Val Tyr Pro Gln Leu Glu Pro Glu Phe
995 1000 1005

Val Tyr Gly Asp Tyr Pro His Phe His Gly His Lys Glu Asn Lys
1010 1015 1020

Ala Thr Ala Lys Lys Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe
1025 1030 1035

Lys Lys Asp Asp Val Arg Thr Asp Lys Asn Gly Glu Ile Ile Trp
1040 1045 1050

	Lys	Lys	Asp	Glu	His	Ile	Ser	Asn	Ile	Lys	Lys	Val	Leu	Ser	Tyr
	1055						1060					1065			
	Pro	Gln	Val	Asn	Ile	Val	Lys	Lys	Val	Glu	Glu	Gln	Thr	Gly	Gly
	1070						1075					1080			
	Phe	Ser	Lys	Glu	Ser	Ile	Leu	Pro	Lys	Gly	Asn	Ser	Asp	Lys	Leu
	1085						1090					1095			
	Ile	Pro	Arg	Lys	Thr	Lys	Lys	Phe	Tyr	Trp	Asp	Thr	Lys	Lys	Tyr
	1100						1105					1110			
	Gly	Gly	Phe	Asp	Ser	Pro	Ile	Val	Ala	Tyr	Ser	Ile	Leu	Val	Ile
	1115						1120					1125			
	Ala	Asp	Ile	Glu	Lys	Gly	Lys	Ser	Lys	Lys	Leu	Lys	Thr	Val	Lys
	1130						1135					1140			
	Ala	Leu	Val	Gly	Val	Thr	Ile	Met	Glu	Lys	Met	Thr	Phe	Glu	Arg
	1145						1150					1155			
[0050]	Asp	Pro	Val	Ala	Phe	Leu	Glu	Arg	Lys	Gly	Tyr	Arg	Asn	Val	Gln
	1160						1165					1170			
	Glu	Glu	Asn	Ile	Ile	Lys	Leu	Pro	Lys	Tyr	Ser	Leu	Phe	Lys	Leu
	1175						1180					1185			
	Glu	Asn	Gly	Arg	Lys	Arg	Leu	Leu	Ala	Ser	Ala	Arg	Glu	Leu	Gln
	1190						1195					1200			
	Lys	Gly	Asn	Glu	Ile	Val	Leu	Pro	Asn	His	Leu	Gly	Thr	Leu	Leu
	1205						1210					1215			
	Tyr	His	Ala	Lys	Asn	Ile	His	Lys	Val	Asp	Glu	Pro	Lys	His	Leu
	1220						1225					1230			
	Asp	Tyr	Val	Asp	Lys	His	Lys	Asp	Glu	Phe	Lys	Glu	Leu	Leu	Asp
	1235						1240					1245			
	Val	Val	Ser	Asn	Phe	Ser	Lys	Lys	Tyr	Thr	Leu	Ala	Glu	Gly	Asn
	1250						1255					1260			
	Leu	Glu	Lys	Ile	Lys	Glu	Leu	Tyr	Ala	Gln	Asn	Asn	Gly	Glu	Asp
	1265						1270					1275			

Leu Lys Glu Leu Ala Ser Ser Phe Ile Asn Leu Leu Thr Phe Thr
1280 1285 1290

Ala Ile Gly Ala Pro Ala Thr Phe Lys Phe Phe Asp Lys Asn Ile
1295 1300 1305

Asp Arg Lys Arg Tyr Thr Ser Thr Thr Glu Ile Leu Asn Ala Thr
1310 1315 1320

Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp
1325 1330 1335

Leu Asn Lys Leu Gly Gly Asp
1340 1345

<210> 20

<211> 1334

<212> PRT

<213> 英诺克李斯特菌

<400> 20

[0051] Met Lys Lys Pro Tyr Thr Ile Gly Leu Asp Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Leu Thr Asp Gln Tyr Asp Leu Val Lys Arg Lys Met
20 25 30

Lys Ile Ala Gly Asp Ser Glu Lys Lys Gln Ile Lys Lys Asn Phe Trp
35 40 45

Gly Val Arg Leu Phe Asp Glu Gly Gln Thr Ala Ala Asp Arg Arg Met
50 55 60

Ala Arg Thr Ala Arg Arg Arg Ile Glu Arg Arg Arg Asn Arg Ile Ser
65 70 75 80

Tyr Leu Gln Gly Ile Phe Ala Glu Glu Met Ser Lys Thr Asp Ala Asn
85 90 95

Phe Phe Cys Arg Leu Ser Asp Ser Phe Tyr Val Asp Asn Glu Lys Arg
100 105 110

Asn Ser Arg His Pro Phe Phe Ala Thr Ile Glu Glu Glu Val Glu Tyr
115 120 125

His Lys Asn Tyr Pro Thr Ile Tyr His Leu Arg Glu Glu Leu Val Asn
130 135 140

Ser Ser Glu Lys Ala Asp Leu Arg Leu Val Tyr Leu Ala Leu Ala His
145 150 155 160

Ile Ile Lys Tyr Arg Gly Asn Phe Leu Ile Glu Gly Ala Leu Asp Thr
165 170 175

Gln Asn Thr Ser Val Asp Gly Ile Tyr Lys Gln Phe Ile Gln Thr Tyr
180 185 190

Asn Gln Val Phe Ala Ser Gly Ile Glu Asp Gly Ser Leu Lys Lys Leu
195 200 205

Glu Asp Asn Lys Asp Val Ala Lys Ile Leu Val Glu Lys Val Thr Arg
210 215 220

Lys Glu Lys Leu Glu Arg Ile Leu Lys Leu Tyr Pro Gly Glu Lys Ser
225 230 235 240

[0052]

Ala Gly Met Phe Ala Gln Phe Ile Ser Leu Ile Val Gly Ser Lys Gly
245 250 255

Asn Phe Gln Lys Pro Phe Asp Leu Ile Glu Lys Ser Asp Ile Glu Cys
260 265 270

Ala Lys Asp Ser Tyr Glu Glu Asp Leu Glu Ser Leu Leu Ala Leu Ile
275 280 285

Gly Asp Glu Tyr Ala Glu Leu Phe Val Ala Ala Lys Asn Ala Tyr Ser
290 295 300

Ala Val Val Leu Ser Ser Ile Ile Thr Val Ala Glu Thr Glu Thr Asn
305 310 315 320

Ala Lys Leu Ser Ala Ser Met Ile Glu Arg Phe Asp Thr His Glu Glu
325 330 335

Asp Leu Gly Glu Leu Lys Ala Phe Ile Lys Leu His Leu Pro Lys His
340 345 350

Tyr Glu Glu Ile Phe Ser Asn Thr Glu Lys His Gly Tyr Ala Gly Tyr
 355 360 365

Ile Asp Gly Lys Thr Lys Gln Ala Asp Phe Tyr Lys Tyr Met Lys Met
 370 375 380

Thr Leu Glu Asn Ile Glu Gly Ala Asp Tyr Phe Ile Ala Lys Ile Glu
 385 390 395 400

Lys Glu Asn Phe Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ala Ile
 405 410 415

Pro His Gln Leu His Leu Glu Glu Leu Glu Ala Ile Leu His Gln Gln
 420 425 430

Ala Lys Tyr Tyr Pro Phe Leu Lys Glu Asn Tyr Asp Lys Ile Lys Ser
 435 440 445

Leu Val Thr Phe Arg Ile Pro Tyr Phe Val Gly Pro Leu Ala Asn Gly
 450 455 460

[0053] Gln Ser Glu Phe Ala Trp Leu Thr Arg Lys Ala Asp Gly Glu Ile Arg
 465 470 475 480

Pro Trp Asn Ile Glu Glu Lys Val Asp Phe Gly Lys Ser Ala Val Asp
 485 490 495

Phe Ile Glu Lys Met Thr Asn Lys Asp Thr Tyr Leu Pro Lys Glu Asn
 500 505 510

Val Leu Pro Lys His Ser Leu Cys Tyr Gln Lys Tyr Leu Val Tyr Asn
 515 520 525

Glu Leu Thr Lys Val Arg Tyr Ile Asn Asp Gln Gly Lys Thr Ser Tyr
 530 535 540

Phe Ser Gly Gln Glu Lys Glu Gln Ile Phe Asn Asp Leu Phe Lys Gln
 545 550 555 560

Lys Arg Lys Val Lys Lys Lys Asp Leu Glu Leu Phe Leu Arg Asn Met
 565 570 575

Ser His Val Glu Ser Pro Thr Ile Glu Gly Leu Glu Asp Ser Phe Asn
 580 585 590

Ser Ser Tyr Ser Thr Tyr His Asp Leu Leu Lys Val Gly Ile Lys Gln
595 600 605

Glu Ile Leu Asp Asn Pro Val Asn Thr Glu Met Leu Glu Asn Ile Val
610 615 620

Lys Ile Leu Thr Val Phe Glu Asp Lys Arg Met Ile Lys Glu Gln Leu
625 630 635 640

Gln Gln Phe Ser Asp Val Leu Asp Gly Val Val Leu Lys Lys Leu Glu
645 650 655

Arg Arg His Tyr Thr Gly Trp Gly Arg Leu Ser Ala Lys Leu Leu Met
660 665 670

Gly Ile Arg Asp Lys Gln Ser His Leu Thr Ile Leu Asp Tyr Leu Met
675 680 685

Asn Asp Asp Gly Leu Asn Arg Asn Leu Met Gln Leu Ile Asn Asp Ser
690 695 700

[0054]

Asn Leu Ser Phe Lys Ser Ile Ile Glu Lys Glu Gln Val Thr Thr Ala
705 710 715 720

Asp Lys Asp Ile Gln Ser Ile Val Ala Asp Leu Ala Gly Ser Pro Ala
725 730 735

Ile Lys Lys Gly Ile Leu Gln Ser Leu Lys Ile Val Asp Glu Leu Val
740 745 750

Ser Val Met Gly Tyr Pro Pro Gln Thr Ile Val Val Glu Met Ala Arg
755 760 765

Glu Asn Gln Thr Thr Gly Lys Gly Lys Asn Asn Ser Arg Pro Arg Tyr
770 775 780

Lys Ser Leu Glu Lys Ala Ile Lys Glu Phe Gly Ser Gln Ile Leu Lys
785 790 795 800

Glu His Pro Thr Asp Asn Gln Glu Leu Arg Asn Asn Arg Leu Tyr Leu
805 810 815

Tyr Tyr Leu Gln Asn Gly Lys Asp Met Tyr Thr Gly Gln Asp Leu Asp
 820 825 830

Ile His Asn Leu Ser Asn Tyr Asp Ile Asp His Ile Val Pro Gln Ser
 835 840 845

Phe Ile Thr Asp Asn Ser Ile Asp Asn Leu Val Leu Thr Ser Ser Ala
 850 855 860

Gly Asn Arg Glu Lys Gly Asp Asp Val Pro Pro Leu Glu Ile Val Arg
 865 870 875 880

Lys Arg Lys Val Phe Trp Glu Lys Leu Tyr Gln Gly Asn Leu Met Ser
 885 890 895

Lys Arg Lys Phe Asp Tyr Leu Thr Lys Ala Glu Arg Gly Gly Leu Thr
 900 905 910

Glu Ala Asp Lys Ala Arg Phe Ile His Arg Gln Leu Val Glu Thr Arg
 915 920 925

[0055] Gln Ile Thr Lys Asn Val Ala Asn Ile Leu His Gln Arg Phe Asn Tyr
 930 935 940

Glu Lys Asp Asp His Gly Asn Thr Met Lys Gln Val Arg Ile Val Thr
 945 950 955 960

Leu Lys Ser Ala Leu Val Ser Gln Phe Arg Lys Gln Phe Gln Leu Tyr
 965 970 975

Lys Val Arg Asp Val Asn Asp Tyr His His Ala His Asp Ala Tyr Leu
 980 985 990

Asn Gly Val Val Ala Asn Thr Leu Leu Lys Val Tyr Pro Gln Leu Glu
 995 1000 1005

Pro Glu Phe Val Tyr Gly Asp Tyr His Gln Phe Asp Trp Phe Lys
 1010 1015 1020

Ala Asn Lys Ala Thr Ala Lys Lys Gln Phe Tyr Thr Asn Ile Met
 1025 1030 1035

Leu Phe Phe Ala Gln Lys Asp Arg Ile Ile Asp Glu Asn Gly Glu
 1040 1045 1050

Ile Leu Trp Asp Lys Lys Tyr Leu Asp Thr Val Lys Lys Val Met
1055 1060 1065

Ser Tyr Arg Gln Met Asn Ile Val Lys Lys Thr Glu Ile Gln Lys
1070 1075 1080

Gly Glu Phe Ser Lys Ala Thr Ile Lys Pro Lys Gly Asn Ser Ser
1085 1090 1095

Lys Leu Ile Pro Arg Lys Thr Asn Trp Asp Pro Met Lys Tyr Gly
1100 1105 1110

Gly Leu Asp Ser Pro Asn Met Ala Tyr Ala Val Val Ile Glu Tyr
1115 1120 1125

Ala Lys Gly Lys Asn Lys Leu Val Phe Glu Lys Lys Ile Ile Arg
1130 1135 1140

Val Thr Ile Met Glu Arg Lys Ala Phe Glu Lys Asp Glu Lys Ala
1145 1150 1155

[0056]

Phe Leu Glu Glu Gln Gly Tyr Arg Gln Pro Lys Val Leu Ala Lys
1160 1165 1170

Leu Pro Lys Tyr Thr Leu Tyr Glu Cys Glu Glu Gly Arg Arg Arg
1175 1180 1185

Met Leu Ala Ser Ala Asn Glu Ala Gln Lys Gly Asn Gln Gln Val
1190 1195 1200

Leu Pro Asn His Leu Val Thr Leu Leu His His Ala Ala Asn Cys
1205 1210 1215

Glu Val Ser Asp Gly Lys Ser Leu Asp Tyr Ile Glu Ser Asn Arg
1220 1225 1230

Glu Met Phe Ala Glu Leu Leu Ala His Val Ser Glu Phe Ala Lys
1235 1240 1245

Arg Tyr Thr Leu Ala Glu Ala Asn Leu Asn Lys Ile Asn Gln Leu
1250 1255 1260

Phe Glu Gln Asn Lys Glu Gly Asp Ile Lys Ala Ile Ala Gln Ser
1265 1270 1275

Phe Val Asp Leu Met Ala Phe Asn Ala Met Gly Ala Pro Ala Ser
1280 1285 1290

Phe Lys Phe Phe Glu Thr Thr Ile Glu Arg Lys Arg Tyr Asn Asn
1295 1300 1305

Leu Lys Glu Leu Leu Asn Ser Thr Ile Ile Tyr Gln Ser Ile Thr
1310 1315 1320

Gly Leu Tyr Glu Ser Arg Lys Arg Leu Asp Asp
1325 1330

<210> 21

<211> 1053

<212> PRT

<213> 金黄色葡萄球菌

<400> 21

[0057]

Met Lys Arg Asn Tyr Ile Leu Gly Leu Asp Ile Gly Ile Thr Ser Val
1 5 10 15

Gly Tyr Gly Ile Ile Asp Tyr Glu Thr Arg Asp Val Ile Asp Ala Gly
20 25 30

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

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

Gln Arg Val Lys Lys Leu Leu Phe Asp Tyr Asn Leu Leu Thr Asp His
65 70 75 80

Ser Glu Leu Ser Gly Ile Asn Pro Tyr Glu Ala Arg Val Lys Gly Leu
85 90 95

Ser Gln Lys Leu Ser Glu Glu Glu Phe Ser Ala Ala Leu Leu His Leu
100 105 110

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

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

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

Asp Gly Glu Val Arg Gly Ser Ile Asn Arg Phe Lys Thr Ser Asp Tyr
 165 170 175

Val Lys Glu Ala Lys Gln Leu Leu Lys Val Gln Lys Ala Tyr His Gln
 180 185 190

Leu Asp Gln Ser Phe Ile Asp Thr Tyr Ile Asp Leu Leu Glu Thr Arg
 195 200 205

Arg Thr Tyr Tyr Glu Gly Pro Gly Glu Gly Ser Pro Phe Gly Trp Lys
 210 215 220

Asp Ile Lys Glu Trp Tyr Glu Met Leu Met Gly His Cys Thr Tyr Phe
 225 230 235 240

[0058] Pro Glu Glu Leu Arg Ser Val Lys Tyr Ala Tyr Asn Ala Asp Leu Tyr
 245 250 255

Asn Ala Leu Asn Asp Leu Asn Asn Leu Val Ile Thr Arg Asp Glu Asn
 260 265 270

Glu Lys Leu Glu Tyr Tyr Glu Lys Phe Gln Ile Ile Glu Asn Val Phe
 275 280 285

Lys Gln Lys Lys Lys Pro Thr Leu Lys Gln Ile Ala Lys Glu Ile Leu
 290 295 300

Val Asn Glu Glu Asp Ile Lys Gly Tyr Arg Val Thr Ser Thr Gly Lys
 305 310 315 320

Pro Glu Phe Thr Asn Leu Lys Val Tyr His Asp Ile Lys Asp Ile Thr
 325 330 335

Ala Arg Lys Glu Ile Ile Glu Asn Ala Glu Leu Leu Asp Gln Ile Ala
 340 345 350

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

Thr Asn Leu Asn Ser Glu Leu Thr Gln Glu Glu Ile Glu Gln Ile Ser
370 375 380

Asn Leu Lys Gly Tyr Thr Gly Thr His Asn Leu Ser Leu Lys Ala Ile
385 390 395 400

Asn Leu Ile Leu Asp Glu Leu Trp His Thr Asn Asp Asn Gln Ile Ala
405 410 415

Ile Phe Asn Arg Leu Lys Leu Val Pro Lys Lys Val Asp Leu Ser Gln
420 425 430

Gln Lys Glu Ile Pro Thr Thr Leu Val Asp Asp Phe Ile Leu Ser Pro
435 440 445

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

Ile Lys Lys Tyr Gly Leu Pro Asn Asp Ile Ile Ile Glu Leu Ala Arg
465 470 475 480

[0059]

Glu Lys Asn Ser Lys Asp Ala Gln Lys Met Ile Asn Glu Met Gln Lys
485 490 495

Arg Asn Arg Gln Thr Asn Glu Arg Ile Glu Glu Ile Ile Arg Thr Thr
500 505 510

Gly Lys Glu Asn Ala Lys Tyr Leu Ile Glu Lys Ile Lys Leu His Asp
515 520 525

Met Gln Glu Gly Lys Cys Leu Tyr Ser Leu Glu Ala Ile Pro Leu Glu
530 535 540

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

Arg Ser Val Ser Phe Asp Asn Ser Phe Asn Asn Lys Val Leu Val Lys
565 570 575

Gln Glu Glu Asn Ser Lys Lys Gly Asn Arg Thr Pro Phe Gln Tyr Leu
580 585 590

Ser Ser Ser Asp Ser Lys Ile Ser Tyr Glu Thr Phe Lys Lys His Ile
 595 600 605

Leu Asn Leu Ala Lys Gly Lys Gly Arg Ile Ser Lys Thr Lys Lys Glu
 610 615 620

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

Phe Ile Asn Arg Asn Leu Val Asp Thr Arg Tyr Ala Thr Arg Gly Leu
 645 650 655

Met Asn Leu Leu Arg Ser Tyr Phe Arg Val Asn Asn Leu Asp Val Lys
 660 665 670

Val Lys Ser Ile Asn Gly Gly Phe Thr Ser Phe Leu Arg Arg Lys Trp
 675 680 685

Lys Phe Lys Lys Glu Arg Asn Lys Gly Tyr Lys His His Ala Glu Asp
 690 695 700

[0060] Ala Leu Ile Ile Ala Asn Ala Asp Phe Ile Phe Lys Glu Trp Lys Lys
 705 710 715 720

Leu Asp Lys Ala Lys Lys Val Met Glu Asn Gln Met Phe Glu Glu Lys
 725 730 735

Gln Ala Glu Ser Met Pro Glu Ile Glu Thr Glu Gln Glu Tyr Lys Glu
 740 745 750

Ile Phe Ile Thr Pro His Gln Ile Lys His Ile Lys Asp Phe Lys Asp
 755 760 765

Tyr Lys Tyr Ser His Arg Val Asp Lys Lys Pro Asn Arg Glu Leu Ile
 770 775 780

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

Ile Val Asn Asn Leu Asn Gly Leu Tyr Asp Lys Asp Asn Asp Lys Leu
 805 810 815

Lys Lys Leu Ile Asn Lys Ser Pro Glu Lys Leu Leu Met Tyr His His
 820 825 830

Asp Pro Gln Thr Tyr Gln Lys Leu Lys Leu Ile Met Glu Gln Tyr Gly
835 840 845

Asp Glu Lys Asn Pro Leu Tyr Lys Tyr Tyr Glu Glu Thr Gly Asn Tyr
850 855 860

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

Lys Tyr Tyr Gly Asn Lys Leu Asn Ala His Leu Asp Ile Thr Asp Asp
885 890 895

Tyr Pro Asn Ser Arg Asn Lys Val Val Lys Leu Ser Leu Lys Pro Tyr
900 905 910

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

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

[0061]

Lys Cys Tyr Glu Glu Ala Lys Lys Leu Lys Lys Ile Ser Asn Gln Ala
945 950 955 960

Glu Phe Ile Ala Ser Phe Tyr Asn Asn Asp Leu Ile Lys Ile Asn Gly
965 970 975

Glu Leu Tyr Arg Val Ile Gly Val Asn Asn Asp Leu Leu Asn Arg Ile
980 985 990

Glu Val Asn Met Ile Asp Ile Thr Tyr Arg Glu Tyr Leu Glu Asn Met
995 1000 1005

Asn Asp Lys Arg Pro Pro Arg Ile Ile Lys Thr Ile Ala Ser Lys
1010 1015 1020

Thr Gln Ser Ile Lys Lys Tyr Ser Thr Asp Ile Leu Gly Asn Leu
1025 1030 1035

Tyr Glu Val Lys Ser Lys Lys His Pro Gln Ile Ile Lys Lys Gly
1040 1045 1050

<210> 22
 <211> 1400
 <212> PRT
 <213> 肺炎克雷伯氏菌

 <400> 22
 Met Ala Pro Lys Lys Lys Arg Lys Val Gly Ile His Gly Val Pro Ala
 1 5 10 15

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

 Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
 35 40 45

 Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
 50 55 60

 Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
 65 70 75 80

 Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
 85 90 95

 Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
 100 105 110

 Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
 115 120 125

 His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
 130 135 140

 His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
 145 150 155 160

 Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
 165 170 175

 Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
 180 185 190

 Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
 195 200 205

[0062]

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
 210 215 220

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
 225 230 235 240

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
 245 250 255

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
 260 265 270

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
 275 280 285

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
 290 295 300

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
 305 310 315 320

[0063] Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
 325 330 335

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
 340 345 350

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
 355 360 365

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
 370 375 380

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
 385 390 395 400

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
 405 410 415

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
 420 425 430

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
 435 440 445

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
450 455 460

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
465 470 475 480

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
485 490 495

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
500 505 510

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
515 520 525

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
530 535 540

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
545 550 555 560

[0064]

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
565 570 575

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
580 585 590

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
595 600 605

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
610 615 620

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
625 630 635 640

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
645 650 655

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
660 665 670

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
 675 680 685

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
 690 695 700

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
 705 710 715 720

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
 725 730 735

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
 740 745 750

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
 755 760 765

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
 770 775 780

[0065] Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
 785 790 795 800

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
 805 810 815

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
 820 825 830

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
 835 840 845

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
 850 855 860

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
 865 870 875 880

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
 885 890 895

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
 900 905 910

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
915 920 925

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
930 935 940

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
945 950 955 960

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
965 970 975

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
980 985 990

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
995 1000 1005

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu
1010 1015 1020

[0066]

Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile
1025 1030 1035

Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe
1040 1045 1050

Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu
1055 1060 1065

Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly
1070 1075 1080

Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr
1085 1090 1095

Val Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys
1100 1105 1110

Thr Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro
1115 1120 1125

	Lys	Arg	Asn	Ser	Asp	Lys	Leu	Ile	Ala	Arg	Lys	Lys	Asp	Trp	Asp
	1130						1135					1140			
	Pro	Lys	Lys	Tyr	Gly	Gly	Phe	Asp	Ser	Pro	Thr	Val	Ala	Tyr	Ser
	1145						1150					1155			
	Val	Leu	Val	Val	Ala	Lys	Val	Glu	Lys	Gly	Lys	Ser	Lys	Lys	Leu
	1160						1165					1170			
	Lys	Ser	Val	Lys	Glu	Leu	Leu	Gly	Ile	Thr	Ile	Met	Glu	Arg	Ser
	1175						1180					1185			
	Ser	Phe	Glu	Lys	Asn	Pro	Ile	Asp	Phe	Leu	Glu	Ala	Lys	Gly	Tyr
	1190						1195					1200			
	Lys	Glu	Val	Lys	Lys	Asp	Leu	Ile	Ile	Lys	Leu	Pro	Lys	Tyr	Ser
	1205						1210					1215			
	Leu	Phe	Glu	Leu	Glu	Asn	Gly	Arg	Lys	Arg	Met	Leu	Ala	Ser	Ala
	1220						1225					1230			
[0067]	Gly	Glu	Leu	Gln	Lys	Gly	Asn	Glu	Leu	Ala	Leu	Pro	Ser	Lys	Tyr
	1235						1240					1245			
	Val	Asn	Phe	Leu	Tyr	Leu	Ala	Ser	His	Tyr	Glu	Lys	Leu	Lys	Gly
	1250						1255					1260			
	Ser	Pro	Glu	Asp	Asn	Glu	Gln	Lys	Gln	Leu	Phe	Val	Glu	Gln	His
	1265						1270					1275			
	Lys	His	Tyr	Leu	Asp	Glu	Ile	Ile	Glu	Gln	Ile	Ser	Glu	Phe	Ser
	1280						1285					1290			
	Lys	Arg	Val	Ile	Leu	Ala	Asp	Ala	Asn	Leu	Asp	Lys	Val	Leu	Ser
	1295						1300					1305			
	Ala	Tyr	Asn	Lys	His	Arg	Asp	Lys	Pro	Ile	Arg	Glu	Gln	Ala	Glu
	1310						1315					1320			
	Asn	Ile	Ile	His	Leu	Phe	Thr	Leu	Thr	Asn	Leu	Gly	Ala	Pro	Ala
	1325						1330					1335			
	Ala	Phe	Lys	Tyr	Phe	Asp	Thr	Thr	Ile	Asp	Arg	Lys	Arg	Tyr	Thr
	1340						1345					1350			

Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile
1355 1360 1365

Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly
1370 1375 1380

Asp Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys Lys
1385 1390 1395

Lys Lys
1400

<210> 23

<211> 1629

<212> PRT

<213> 新凶手弗朗西斯菌

<400> 23

Met Asn Phe Lys Ile Leu Pro Ile Ala Ile Asp Leu Gly Val Lys Asn
1 5 10 15

[0068]

Thr Gly Val Phe Ser Ala Phe Tyr Gln Lys Gly Thr Ser Leu Glu Arg
20 25 30

Leu Asp Asn Lys Asn Gly Lys Val Tyr Glu Leu Ser Lys Asp Ser Tyr
35 40 45

Thr Leu Leu Met Asn Asn Arg Thr Ala Arg Arg His Gln Arg Arg Gly
50 55 60

Ile Asp Arg Lys Gln Leu Val Lys Arg Leu Phe Lys Leu Ile Trp Thr
65 70 75 80

Glu Gln Leu Asn Leu Glu Trp Asp Lys Asp Thr Gln Gln Ala Ile Ser
85 90 95

Phe Leu Phe Asn Arg Arg Gly Phe Ser Phe Ile Thr Asp Gly Tyr Ser
100 105 110

Pro Glu Tyr Leu Asn Ile Val Pro Glu Gln Val Lys Ala Ile Leu Met
115 120 125

Asp Ile Phe Asp Asp Tyr Asn Gly Glu Asp Asp Leu Asp Ser Tyr Leu
130 135 140

Lys Leu Ala Thr Glu Gln Glu Ser Lys Ile Ser Glu Ile Tyr Asn Lys
145 150 155 160

Leu Met Gln Lys Ile Leu Glu Phe Lys Leu Met Lys Leu Cys Thr Asp
165 170 175

Ile Lys Asp Asp Lys Val Ser Thr Lys Thr Leu Lys Glu Ile Thr Ser
180 185 190

Tyr Glu Phe Glu Leu Leu Ala Asp Tyr Leu Ala Asn Tyr Ser Glu Ser
195 200 205

Leu Lys Thr Gln Lys Phe Ser Tyr Thr Asp Lys Gln Gly Asn Leu Lys
210 215 220

Glu Leu Ser Tyr Tyr His His Asp Lys Tyr Asn Ile Gln Glu Phe Leu
225 230 235 240

Lys Arg His Ala Thr Ile Asn Asp Arg Ile Leu Asp Thr Leu Leu Thr
245 250 255

[0069]

Asp Asp Leu Asp Ile Trp Asn Phe Asn Phe Glu Lys Phe Asp Phe Asp
260 265 270

Lys Asn Glu Glu Lys Leu Gln Asn Gln Glu Asp Lys Asp His Ile Gln
275 280 285

Ala His Leu His His Phe Val Phe Ala Val Asn Lys Ile Lys Ser Glu
290 295 300

Met Ala Ser Gly Gly Arg His Arg Ser Gln Tyr Phe Gln Glu Ile Thr
305 310 315 320

Asn Val Leu Asp Glu Asn Asn His Gln Glu Gly Tyr Leu Lys Asn Phe
325 330 335

Cys Glu Asn Leu His Asn Lys Lys Tyr Ser Asn Leu Ser Val Lys Asn
340 345 350

Leu Val Asn Leu Ile Gly Asn Leu Ser Asn Leu Glu Leu Lys Pro Leu
355 360 365

Arg Lys Tyr Phe Asn Asp Lys Ile His Ala Lys Ala Asp His Trp Asp
 370 375 380

Glu Gln Lys Phe Thr Glu Thr Tyr Cys His Trp Ile Leu Gly Glu Trp
 385 390 395 400

Arg Val Gly Val Lys Asp Gln Asp Lys Lys Asp Gly Ala Lys Tyr Ser
 405 410 415

Tyr Lys Asp Leu Cys Asn Glu Leu Lys Gln Lys Val Thr Lys Ala Gly
 420 425 430

Leu Val Asp Phe Leu Leu Glu Leu Asp Pro Cys Arg Thr Ile Pro Pro
 435 440 445

Tyr Leu Asp Asn Asn Asn Arg Lys Pro Pro Lys Cys Gln Ser Leu Ile
 450 455 460

Leu Asn Pro Lys Phe Leu Asp Asn Gln Tyr Pro Asn Trp Gln Gln Tyr
 465 470 475 480

[0070] Leu Gln Glu Leu Lys Lys Leu Gln Ser Ile Gln Asn Tyr Leu Asp Ser
 485 490 495

Phe Glu Thr Asp Leu Lys Val Leu Lys Ser Ser Lys Asp Gln Pro Tyr
 500 505 510

Phe Val Glu Tyr Lys Ser Ser Asn Gln Gln Ile Ala Ser Gly Gln Arg
 515 520 525

Asp Tyr Lys Asp Leu Asp Ala Arg Ile Leu Gln Phe Ile Phe Asp Arg
 530 535 540

Val Lys Ala Ser Asp Glu Leu Leu Leu Asn Glu Ile Tyr Phe Gln Ala
 545 550 555 560

Lys Lys Leu Lys Gln Lys Ala Ser Ser Glu Leu Glu Lys Leu Glu Ser
 565 570 575

Ser Lys Lys Leu Asp Glu Val Ile Ala Asn Ser Gln Leu Ser Gln Ile
 580 585 590

Leu Lys Ser Gln His Thr Asn Gly Ile Phe Glu Gln Gly Thr Phe Leu
 595 600 605

His Leu Val Cys Lys Tyr Tyr Lys Gln Arg Gln Arg Ala Arg Asp Ser
610 615 620

Arg Leu Tyr Ile Met Pro Glu Tyr Arg Tyr Asp Lys Lys Leu His Lys
625 630 635 640

Tyr Asn Asn Thr Gly Arg Phe Asp Asp Asp Asn Gln Leu Leu Thr Tyr
645 650 655

Cys Asn His Lys Pro Arg Gln Lys Arg Tyr Gln Leu Leu Asn Asp Leu
660 665 670

Ala Gly Val Leu Gln Val Ser Pro Asn Phe Leu Lys Asp Lys Ile Gly
675 680 685

Ser Asp Asp Asp Leu Phe Ile Ser Lys Trp Leu Val Glu His Ile Arg
690 695 700

Gly Phe Lys Lys Ala Cys Glu Asp Ser Leu Lys Ile Gln Lys Asp Asn
705 710 715 720

[0071]

Arg Gly Leu Leu Asn His Lys Ile Asn Ile Ala Arg Asn Thr Lys Gly
725 730 735

Lys Cys Glu Lys Glu Ile Phe Asn Leu Ile Cys Lys Ile Glu Gly Ser
740 745 750

Glu Asp Lys Lys Gly Asn Tyr Lys His Gly Leu Ala Tyr Glu Leu Gly
755 760 765

Val Leu Leu Phe Gly Glu Pro Asn Glu Ala Ser Lys Pro Glu Phe Asp
770 775 780

Arg Lys Ile Lys Lys Phe Asn Ser Ile Tyr Ser Phe Ala Gln Ile Gln
785 790 795 800

Gln Ile Ala Phe Ala Glu Arg Lys Gly Asn Ala Asn Thr Cys Ala Val
805 810 815

Cys Ser Ala Asp Asn Ala His Arg Met Gln Gln Ile Lys Ile Thr Glu
820 825 830

Pro Val Glu Asp Asn Lys Asp Lys Ile Ile Leu Ser Ala Lys Ala Gln
 835 840 845

Arg Leu Pro Ala Ile Pro Thr Arg Ile Val Asp Gly Ala Val Lys Lys
 850 855 860

Met Ala Thr Ile Leu Ala Lys Asn Ile Val Asp Asp Asn Trp Gln Asn
 865 870 875 880

Ile Lys Gln Val Leu Ser Ala Lys His Gln Leu His Ile Pro Ile Ile
 885 890 895

Thr Glu Ser Asn Ala Phe Glu Phe Glu Pro Ala Leu Ala Asp Val Lys
 900 905 910

Gly Lys Ser Leu Lys Asp Arg Arg Lys Lys Ala Leu Glu Arg Ile Ser
 915 920 925

Pro Glu Asn Ile Phe Lys Asp Lys Asn Asn Arg Ile Lys Glu Phe Ala
 930 935 940

[0072] Lys Gly Ile Ser Ala Tyr Ser Gly Ala Asn Leu Thr Asp Gly Asp Phe
 945 950 955 960

Asp Gly Ala Lys Glu Glu Leu Asp His Ile Ile Pro Arg Ser His Lys
 965 970 975

Lys Tyr Gly Thr Leu Asn Asp Glu Ala Asn Leu Ile Cys Val Thr Arg
 980 985 990

Gly Asp Asn Lys Asn Lys Gly Asn Arg Ile Phe Cys Leu Arg Asp Leu
 995 1000 1005

Ala Asp Asn Tyr Lys Leu Lys Gln Phe Glu Thr Thr Asp Asp Leu
 1010 1015 1020

Glu Ile Glu Lys Lys Ile Ala Asp Thr Ile Trp Asp Ala Asn Lys
 1025 1030 1035

Lys Asp Phe Lys Phe Gly Asn Tyr Arg Ser Phe Ile Asn Leu Thr
 1040 1045 1050

Pro Gln Glu Gln Lys Ala Phe Arg His Ala Leu Phe Leu Ala Asp
 1055 1060 1065

Glu Asn Pro Ile Lys Gln Ala Val Ile Arg Ala Ile Asn Asn Arg
 1070 1075 1080

Asn Arg Thr Phe Val Asn Gly Thr Gln Arg Tyr Phe Ala Glu Val
 1085 1090 1095

Leu Ala Asn Asn Ile Tyr Leu Arg Ala Lys Lys Glu Asn Leu Asn
 1100 1105 1110

Thr Asp Lys Ile Ser Phe Asp Tyr Phe Gly Ile Pro Thr Ile Gly
 1115 1120 1125

Asn Gly Arg Gly Ile Ala Glu Ile Arg Gln Leu Tyr Glu Lys Val
 1130 1135 1140

Asp Ser Asp Ile Gln Ala Tyr Ala Lys Gly Asp Lys Pro Gln Ala
 1145 1150 1155

Ser Tyr Ser His Leu Ile Asp Ala Met Leu Ala Phe Cys Ile Ala
 1160 1165 1170

[0073]

Ala Asp Glu His Arg Asn Asp Gly Ser Ile Gly Leu Glu Ile Asp
 1175 1180 1185

Lys Asn Tyr Ser Leu Tyr Pro Leu Asp Lys Asn Thr Gly Glu Val
 1190 1195 1200

Phe Thr Lys Asp Ile Phe Ser Gln Ile Lys Ile Thr Asp Asn Glu
 1205 1210 1215

Phe Ser Asp Lys Lys Leu Val Arg Lys Lys Ala Ile Glu Gly Phe
 1220 1225 1230

Asn Thr His Arg Gln Met Thr Arg Asp Gly Ile Tyr Ala Glu Asn
 1235 1240 1245

Tyr Leu Pro Ile Leu Ile His Lys Glu Leu Asn Glu Val Arg Lys
 1250 1255 1260

Gly Tyr Thr Trp Lys Asn Ser Glu Glu Ile Lys Ile Phe Lys Gly
 1265 1270 1275

	Lys	Lys	Tyr	Asp	Ile	Gln	Gln	Leu	Asn	Asn	Leu	Val	Tyr	Cys	Leu
	1280						1285					1290			
	Lys	Phe	Val	Asp	Lys	Pro	Ile	Ser	Ile	Asp	Ile	Gln	Ile	Ser	Thr
	1295						1300					1305			
	Leu	Glu	Glu	Leu	Arg	Asn	Ile	Leu	Thr	Thr	Asn	Asn	Ile	Ala	Ala
	1310						1315					1320			
	Thr	Ala	Glu	Tyr	Tyr	Tyr	Ile	Asn	Leu	Lys	Thr	Gln	Lys	Leu	His
	1325						1330					1335			
	Glu	Tyr	Tyr	Ile	Glu	Asn	Tyr	Asn	Thr	Ala	Leu	Gly	Tyr	Lys	Lys
	1340						1345					1350			
	Tyr	Ser	Lys	Glu	Met	Glu	Phe	Leu	Arg	Ser	Leu	Ala	Tyr	Arg	Ser
	1355						1360					1365			
	Glu	Arg	Val	Lys	Ile	Lys	Ser	Ile	Asp	Asp	Val	Lys	Gln	Val	Leu
	1370						1375					1380			
[0074]	Asp	Lys	Asp	Ser	Asn	Phe	Ile	Ile	Gly	Lys	Ile	Thr	Leu	Pro	Phe
	1385						1390					1395			
	Lys	Lys	Glu	Trp	Gln	Arg	Leu	Tyr	Arg	Glu	Trp	Gln	Asn	Thr	Thr
	1400						1405					1410			
	Ile	Lys	Asp	Asp	Tyr	Glu	Phe	Leu	Lys	Ser	Phe	Phe	Asn	Val	Lys
	1415						1420					1425			
	Ser	Ile	Thr	Lys	Leu	His	Lys	Lys	Val	Arg	Lys	Asp	Phe	Ser	Leu
	1430						1435					1440			
	Pro	Ile	Ser	Thr	Asn	Glu	Gly	Lys	Phe	Leu	Val	Lys	Arg	Lys	Thr
	1445						1450					1455			
	Trp	Asp	Asn	Asn	Phe	Ile	Tyr	Gln	Ile	Leu	Asn	Asp	Ser	Asp	Ser
	1460						1465					1470			
	Arg	Ala	Asp	Gly	Thr	Lys	Pro	Phe	Ile	Pro	Ala	Phe	Asp	Ile	Ser
	1475						1480					1485			
	Lys	Asn	Glu	Ile	Val	Glu	Ala	Ile	Ile	Asp	Ser	Phe	Thr	Ser	Lys
	1490						1495					1500			

Asn Ile Phe Trp Leu Pro Lys Asn Ile Glu Leu Gln Lys Val Asp
1505 1510 1515

Asn Lys Asn Ile Phe Ala Ile Asp Thr Ser Lys Trp Phe Glu Val
1520 1525 1530

Glu Thr Pro Ser Asp Leu Arg Asp Ile Gly Ile Ala Thr Ile Gln
1535 1540 1545

Tyr Lys Ile Asp Asn Asn Ser Arg Pro Lys Val Arg Val Lys Leu
1550 1555 1560

Asp Tyr Val Ile Asp Asp Asp Ser Lys Ile Asn Tyr Phe Met Asn
1565 1570 1575

His Ser Leu Leu Lys Ser Arg Tyr Pro Asp Lys Val Leu Glu Ile
1580 1585 1590

Leu Lys Gln Ser Thr Ile Ile Glu Phe Glu Ser Ser Gly Phe Asn
1595 1600 1605

[0075]

Lys Thr Ile Lys Glu Met Leu Gly Met Lys Leu Ala Gly Ile Tyr
1610 1615 1620

Asn Glu Thr Ser Asn Asn
1625

<210> 24

<211> 5

<212> PRT

<213> 人工序列

<220>

<221> 来源

<223> /注释= “人工序列的描述: 合成肽”

<400> 24

Lys Phe Glu Arg Gln

1 5

<210> 25

<211> 5

<212> PRT

<213> 人工序列

<220>

<221> 来源
<223> /注释= “人工序列的描述：合成肽”

<400> 25
Arg Lys Val Glu Gln
1 5

<210> 26
<211> 5
<212> PRT
<213> 人工序列

<220>
<221> 来源
<223> /注释= “人工序列的描述：合成肽”

<400> 26
Gln Asp Leu Lys Phe
1 5

<210> 27
<211> 5
<212> PRT
<213> 人工序列

[0076] <220>
<221> 来源
<223> /注释= “人工序列的描述：合成肽”

<400> 27
Gln Arg Phe Phe Glu
1 5

<210> 28
<211> 5
<212> PRT
<213> 人工序列

<220>
<221> 来源
<223> /注释= “人工序列的描述：合成肽”

<400> 28
Asn Arg Val Val Asp
1 5

<210> 29
<211> 5
<212> PRT
<213> 人工序列

<220>
<221> 来源
<223> /注释= “人工序列的描述：合成肽”

<400> 29
Gln Arg Asp Lys Val
1 5

<210> 30
<211> 5
<212> PRT
<213> 人工序列

<220>
<221> 来源
<223> /注释= “人工序列的描述：合成肽”

<400> 30
Gln Lys Ile Leu Asp
1 5

<210> 31
<211> 5
<212> PRT
<213> 人工序列

<220>
<221> 来源
<223> /注释= “人工序列的描述：合成肽”

[0077] <400> 31
Gln Lys Lys Glu Leu
1 5

<210> 32
<211> 5
<212> PRT
<213> 人工序列

<220>
<221> 来源
<223> /注释= “人工序列的描述：合成肽”

<400> 32
Gln Phe Arg Glu Leu
1 5

<210> 33
<211> 5
<212> PRT
<213> 人工序列

<220>
<221> 来源
<223> /注释= “人工序列的描述：合成肽”

<400> 33

Ile Lys Leu Asp Gln
1 5

<210> 34
<211> 5
<212> PRT
<213> 人工序列

<220>
<221> 来源
<223> /注释= “人工序列的描述: 合成肽”

<400> 34
Asp Val Val Arg Gln
1 5

<210> 35
<211> 5
<212> PRT
<213> 人工序列

<220>
<221> 来源
<223> /注释= “人工序列的描述: 合成肽”

[0078] <400> 35
Gln Arg Ile Val Glu
1 5

<210> 36
<211> 5
<212> PRT
<213> 人工序列

<220>
<221> 来源
<223> /注释= “人工序列的描述: 合成肽”

<400> 36
Val Lys Glu Leu Gln
1 5

<210> 37
<211> 5
<212> PRT
<213> 人工序列

<220>
<221> 来源
<223> /注释= “人工序列的描述: 合成肽”

<400> 37
Gln Lys Val Phe Asp
1 5

<210> 38
<211> 5
<212> PRT
<213> 人工序列

<220>
<221> 来源
<223> /注释= “人工序列的描述: 合成肽”

<400> 38
Gln Glu Leu Leu Arg
1 5

<210> 39
<211> 5
<212> PRT
<213> 人工序列

<220>
<221> 来源
<223> /注释= “人工序列的描述: 合成肽”

<400> 39
Val Asp Lys Leu Asn
1 5

[0079]

<210> 40
<211> 5
<212> PRT
<213> 人工序列

<220>
<221> 来源
<223> /注释= “人工序列的描述: 合成肽”

<400> 40
Arg Ile Lys Glu Asn
1 5

<210> 41
<211> 5
<212> PRT
<213> 人工序列

<220>
<221> 来源
<223> /注释= “人工序列的描述: 合成肽”

<400> 41
Asn Lys Lys Phe Glu
1 5

<210> 42
 <211> 7
 <212> PRT
 <213> 猴病毒 40

<400> 42
 Pro Lys Lys Lys Arg Lys Val
 1 5

<210> 43
 <211> 20
 <212> PRT
 <213> 未知

<220>
 <221> 来源
 <223> /注释= “未知的描述：核质蛋白二分体 NLS 序列”

<400> 43
 Ala Val Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys Lys
 1 5 10 15

Lys Lys Leu Asp
 20

[0080]

<210> 44
 <211> 9
 <212> PRT
 <213> 未知

<220>
 <221> 来源
 <223> /注释= “未知的描述：C-myc NLS 序列”

<400> 44
 Pro Ala Ala Lys Arg Val Lys Leu Asp
 1 5

<210> 45
 <211> 25
 <212> PRT
 <213> 秀丽隐杆线虫

<400> 45
 Met Ser Arg Arg Arg Lys Ala Asn Pro Thr Lys Leu Ser Glu Asn Ala
 1 5 10 15

Lys Lys Leu Ala Lys Glu Val Glu Asn
 20 25

<210> 46
 <211> 9

	tttattatct gcacatgtat g	21
	<210> 51	
	<211> 21	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<221> 来源	
	<223> /注释= “人工序列的描述: 合成引物”	
	<400> 51	
	ctacctgtac atctgcacaa g	21
	<210> 52	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<221> 来源	
	<223> /注释= “人工序列的描述: 合成引物”	
	<400> 52	
	atgtgcttca acccatcacg	20
[0082]	<210> 53	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<221> 来源	
	<223> /注释= “人工序列的描述: 合成引物”	
	<400> 53	
	gttggctttc acaaggatgc	20
	<210> 54	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<221> 来源	
	<223> /注释= “人工序列的描述: 合成引物”	
	<400> 54	
	cacggataaa gacgetggga	20
	<210> 55	
	<211> 19	
	<212> DNA	

	<213> 人工序列	
	<220>	
	<221> 来源	
	<223> /注释= “人工序列的描述：合成引物”	
	<400> 55	
	tcccaggtgc tgacgtagg	19
	<210> 56	
	<211> 21	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<221> 来源	
	<223> /注释= “人工序列的描述：合成引物”	
	<400> 56	
	tagcactggg tgcttaatcc g	21
	<210> 57	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
[0083]	<220>	
	<221> 来源	
	<223> /注释= “人工序列的描述：合成引物”	
	<400> 57	
	gggtttgggtt ggctgctcat	20
	<210> 58	
	<211> 23	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<221> 来源	
	<223> /注释= “人工序列的描述：合成引物”	
	<400> 58	
	accggggcca ctagggacag gat	23
	<210> 59	
	<211> 23	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<221> 来源	
	<223> /注释= “人工序列的描述：合成引物”	

<400> 59 aaacatcctg tccctagtgg ccc	23
<210> 60 <211> 28 <212> DNA <213> 人工序列	
<220> <221> 来源 <223> /注释= “人工序列的描述: 合成引物”	
<400> 60 gataaagcag acctgaggct gatctatc	28
<210> 61 <211> 25 <212> DNA <213> 人工序列	
<220> <221> 来源 <223> /注释= “人工序列的描述: 合成引物”	
<400> 61 ctggcagctg agcgatcagg ttctc	25
[0084]	
<210> 62 <211> 20 <212> DNA <213> 人工序列	
<220> <221> 来源 <223> /注释= “人工序列的描述: 合成引物”	
<400> 62 gatgccacc ttcccctctc	20
<210> 63 <211> 20 <212> DNA <213> 人工序列	
<220> <221> 来源 <223> /注释= “人工序列的描述: 合成引物”	
<400> 63 agtgtggcc tcattctgga	20
<210> 64 <211> 22	

	<212> DNA	
	<213> 人工序列	
	<220>	
	<221> 来源	
	<223> /注释= “人工序列的描述: 合成引物”	
	<400> 64	
	ggcttcacga gaaaagttga tg	22
	<210> 65	
	<211> 21	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<221> 来源	
	<223> /注释= “人工序列的描述: 合成引物”	
	<400> 65	
	ggattcacag gcttcaccta c	21
	<210> 66	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
[0085]	<220>	
	<221> 来源	
	<223> /注释= “人工序列的描述: 合成寡核苷酸”	
	<400> 66	
	gagtccgagc agaagaagaa	20
	<210> 67	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<221> 来源	
	<223> /注释= “人工序列的描述: 合成寡核苷酸”	
	<400> 67	
	ggaatccctt ctgcagcacc	20
	<210> 68	
	<211> 19	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<221> 来源	
	<223> /注释= “人工序列的描述: 合成寡核苷酸”	

	<400> 68 cactttggga ggccgaggc	19
	<210> 69 <211> 40 <212> DNA <213> 人工序列	
	<220> <221> 来源 <223> /注释= “人工序列的描述：合成寡核苷酸”	
	<400> 69 gaacttggtg atgatatcgt gccctctggc cagtcctgat	40
	<210> 70 <211> 40 <212> DNA <213> 人工序列	
	<220> <221> 来源 <223> /注释= “人工序列的描述：合成寡核苷酸”	
[0086]	<400> 70 tttatagtag gatttacacg accaattcct tcattaatct	40
	<210> 71 <211> 20 <212> DNA <213> 人工序列	
	<220> <221> 来源 <223> /注释= “人工序列的描述：合成寡核苷酸”	
	<400> 71 caggttccat gggatgctct	20
	<210> 72 <211> 20 <212> DNA <213> 人工序列	
	<220> <221> 来源 <223> /注释= “人工序列的描述：合成寡核苷酸”	
	<400> 72 agcagcagcg gcggcaacag	20
	<210> 73	

	<211> 23	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<221> 来源	
	<223> /注释= “人工序列的描述: 合成寡核苷酸”	
	<400> 73	
	gagtcgagc agaagaagaa ggg	23
	<210> 74	
	<211> 23	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<221> 来源	
	<223> /注释= “人工序列的描述: 合成寡核苷酸”	
	<400> 74	
	gagtctaagc agaagaagaa gag	23
	<210> 75	
	<211> 23	
	<212> DNA	
[0087]	<213> 人工序列	
	<220>	
	<221> 来源	
	<223> /注释= “人工序列的描述: 合成寡核苷酸”	
	<400> 75	
	ggaatccctt ctgcagcacc tgg	23
	<210> 76	
	<211> 23	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<221> 来源	
	<223> /注释= “人工序列的描述: 合成寡核苷酸”	
	<400> 76	
	ggaaccccg tgcagcacc tgg	23
	<210> 77	
	<211> 23	
	<212> DNA	
	<213> 未知	
	<220>	
	<221> 来源	

	<223> /注释= “未知的描述：野生型序列”	
	<220>	
	<221> 修饰碱基	
	<222> (21)..(21)	
	<223> a, c, t, g, 未知或其他	
	<400> 77	
	agcagcagcg gcggcaacag ngg	23
	<210> 78	
	<211> 23	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<221> 来源	
	<223> /注释= “人工序列的描述：合成寡核苷酸”	
	<400> 78	
	agcagcagcg gcggcaacag cgg	23
	<210> 79	
	<211> 23	
	<212> DNA	
[0088]	<213> 人工序列	
	<220>	
	<221> 来源	
	<223> /注释= “人工序列的描述：合成寡核苷酸”	
	<400> 79	
	agcagcagca gcagcaacag cgg	23
	<210> 80	
	<211> 23	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<221> 来源	
	<223> /注释= “人工序列的描述：合成寡核苷酸”	
	<400> 80	
	agcagcagcg gcagcaacag cag	23
	<210> 81	
	<211> 23	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<221> 来源	

	<223> /注释= “人工序列的描述：合成寡核苷酸”	
	<400> 81 agcagcagca gcagcaacag ggg	23
	<210> 82 <211> 23 <212> DNA <213> 人工序列	
	<220> <221> 来源 <223> /注释= “人工序列的描述：合成寡核苷酸”	
	<400> 82 agcagcagca gcagcaacag gag	23
	<210> 83 <211> 23 <212> DNA <213> 人工序列	
	<220> <221> 来源 <223> /注释= “人工序列的描述：合成寡核苷酸”	
[0089]	<400> 83 agcagcagca gcagcaacag cgg	23
	<210> 84 <211> 23 <212> DNA <213> 人工序列	
	<220> <221> 来源 <223> /注释= “人工序列的描述：合成寡核苷酸”	
	<400> 84 agcagcagcg gcggaacag cag	23
	<210> 85 <211> 23 <212> DNA <213> 人工序列	
	<220> <221> 来源 <223> /注释= “人工序列的描述：合成寡核苷酸”	
	<400> 85 agcagcagca acagcaacag cgg	23

	<210> 86	
	<211> 23	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<221> 来源	
	<223> /注释= “人工序列的描述: 合成寡核苷酸”	
	<400> 86	
	agcagcagca gcagcaacag cag	23
[0090]		
	<210> 87	
	<211> 23	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<221> 来源	
	<223> /注释= “人工序列的描述: 合成寡核苷酸”	
	<400> 87	
	agcagcaggg gcggcaacag ggg	23

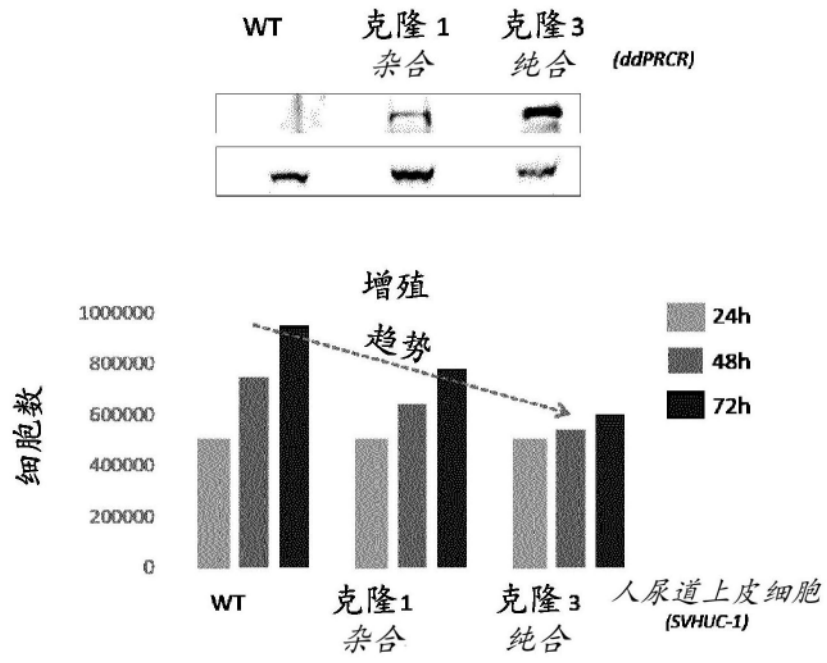


图1A

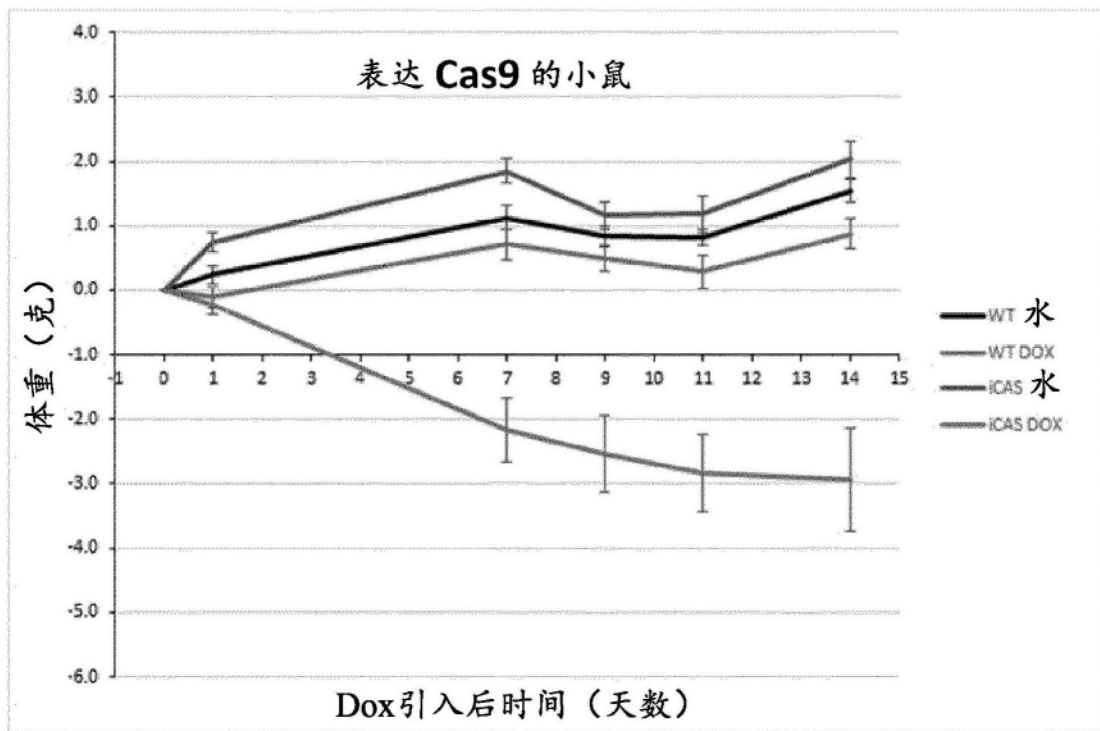


图1B

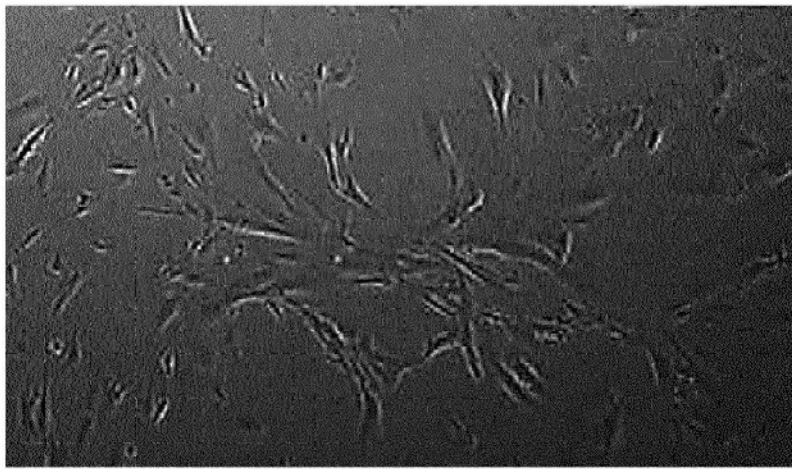
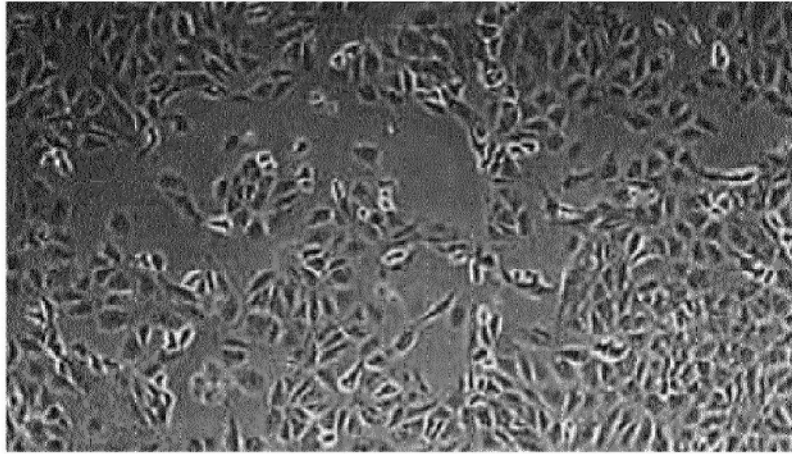


图2

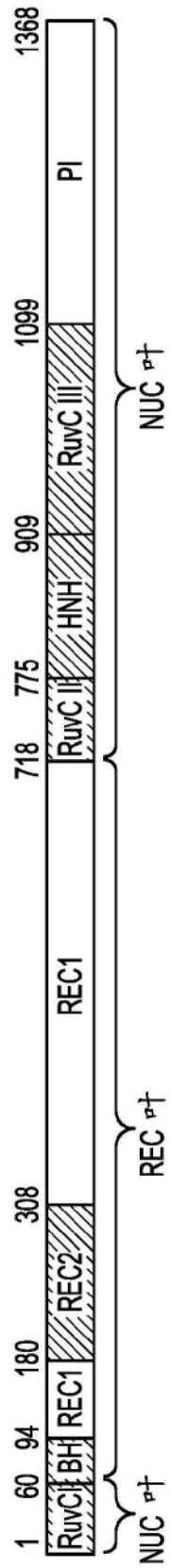


图3

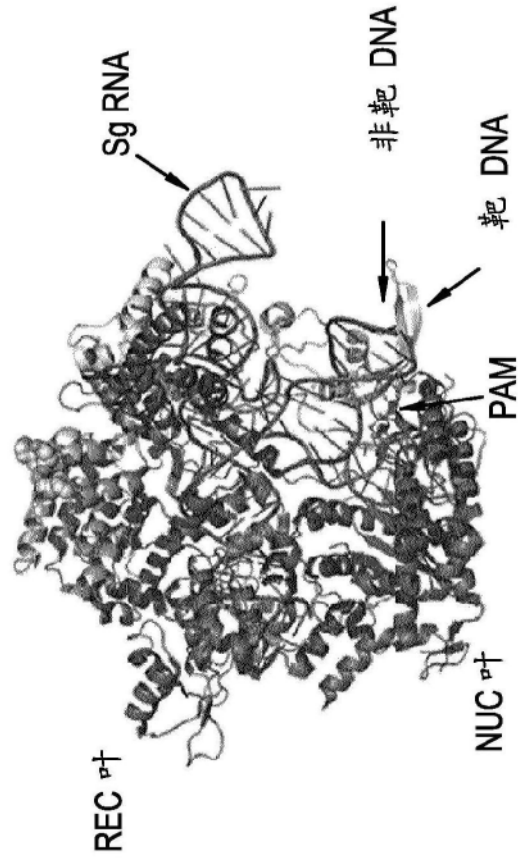


图4

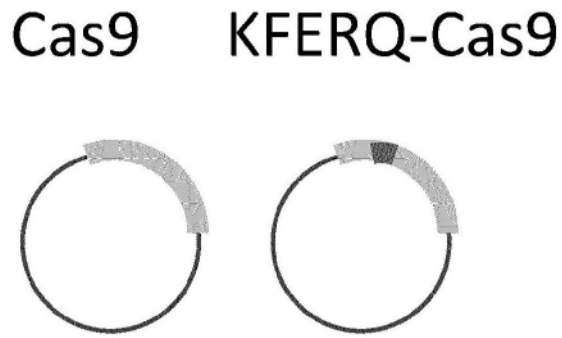


图5A

FLAG-Cas9 FLAG-FaDe

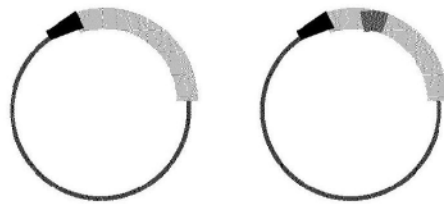


图5B

Cas9 KFERQ-Cas9

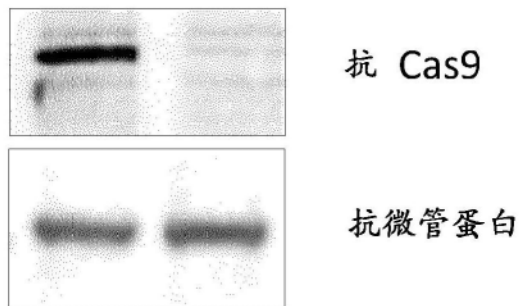


图6A

标记Cas9 标记FaDe

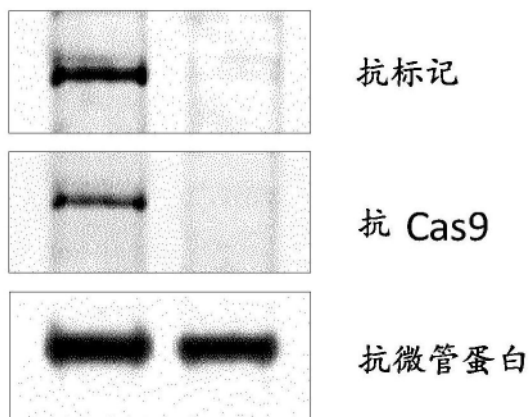


图6B

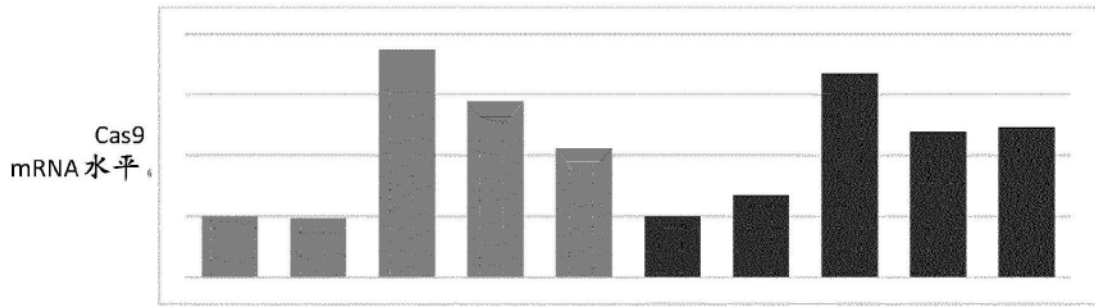


图7A

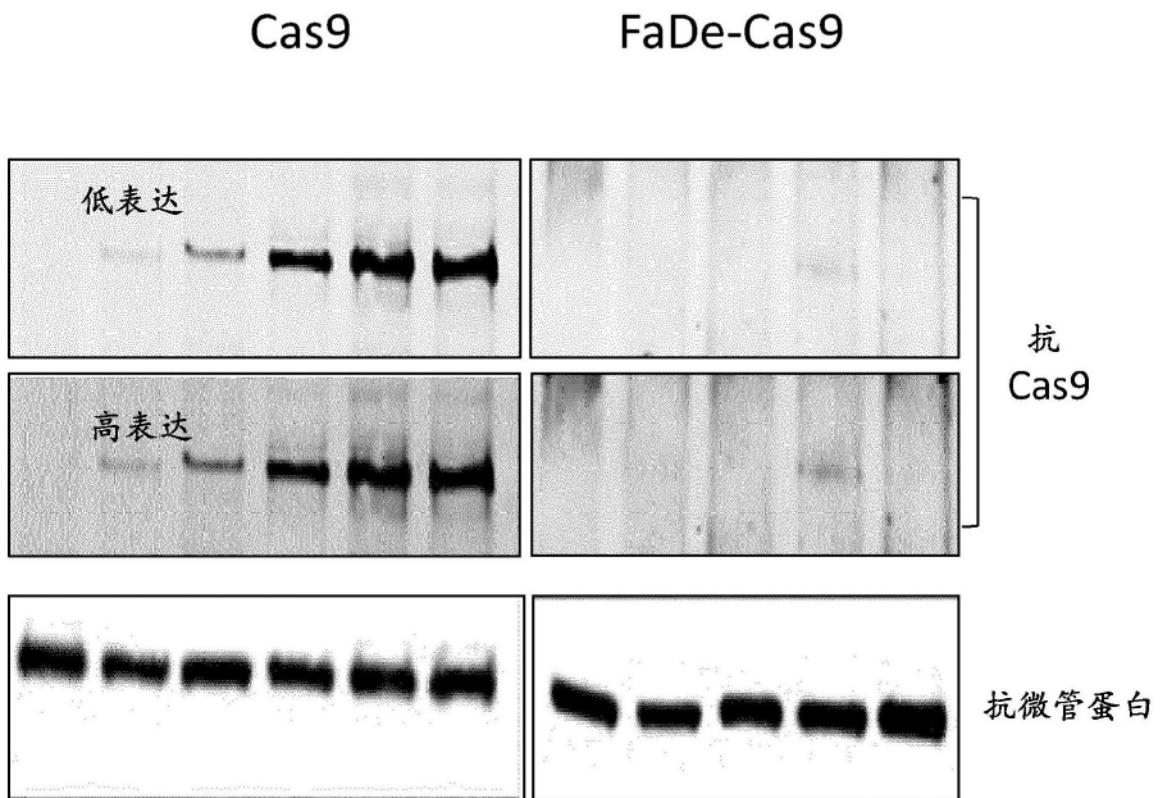


图7B

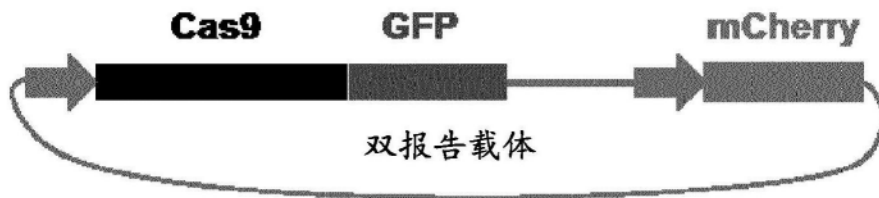
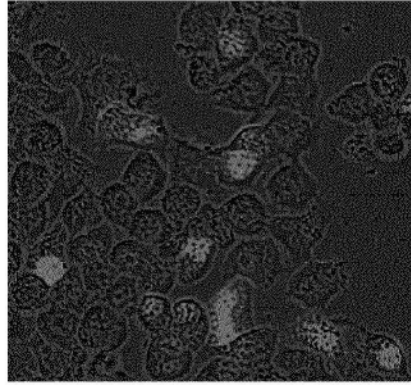
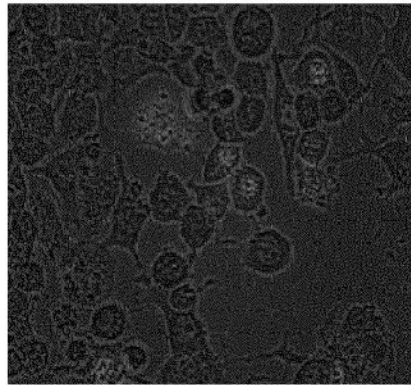


图8



Cas9 融合的 GFP

图9A



FaDe-Cas9 融合的 GFP

图9B

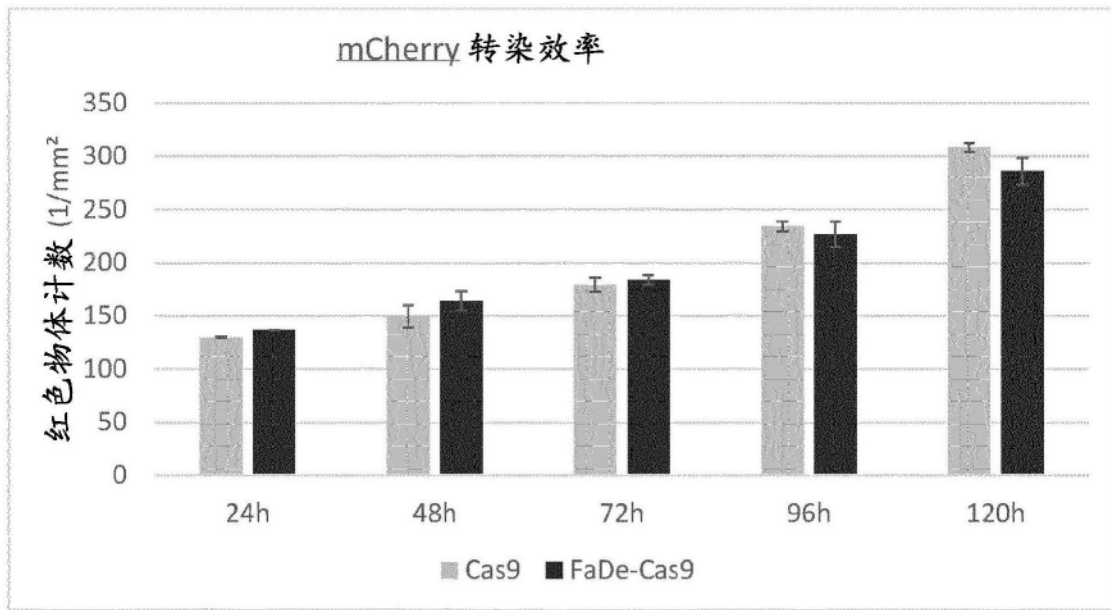


图10A

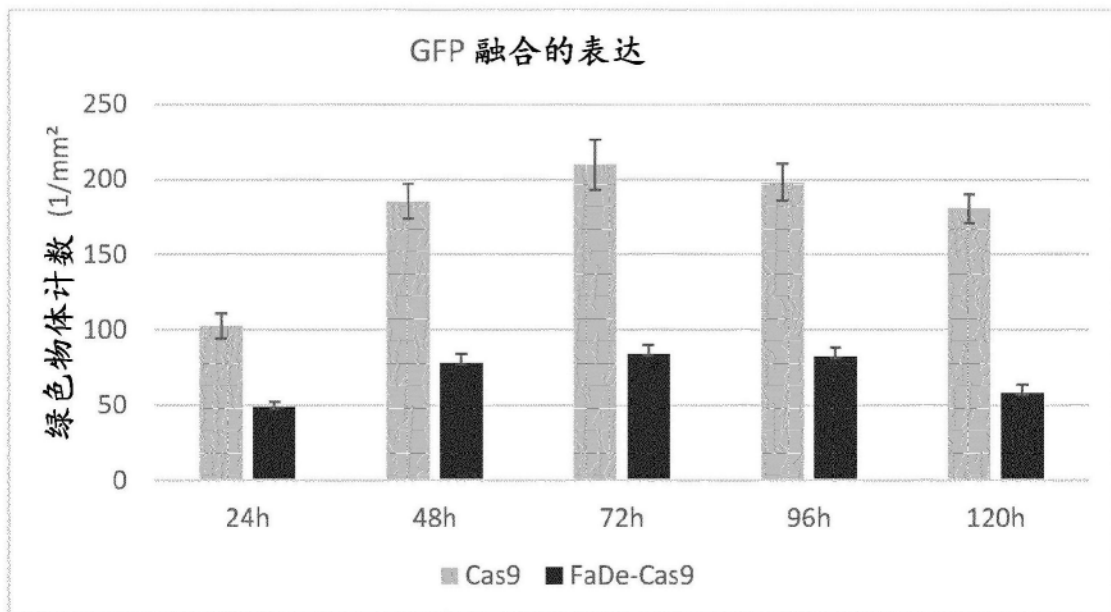
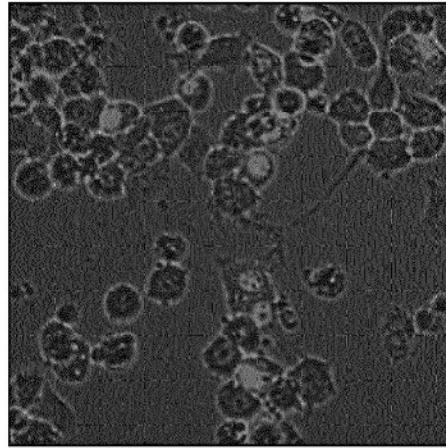
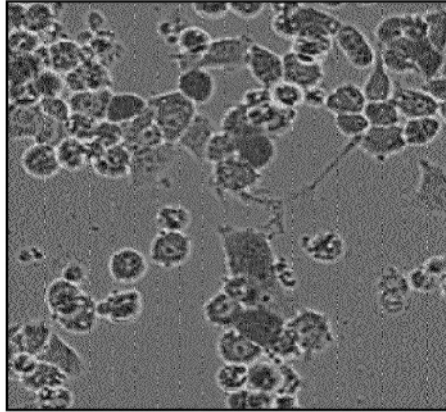


图10B



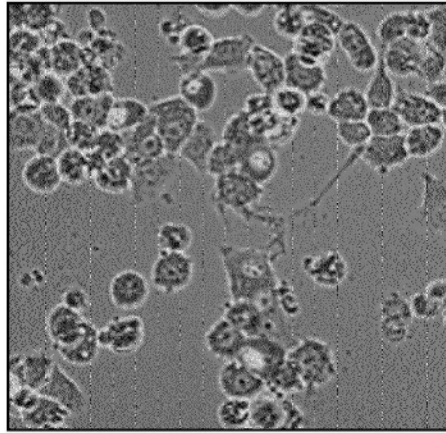
Cas9-GFP-融合

图11A



mCherry

图11B



合并

图11C

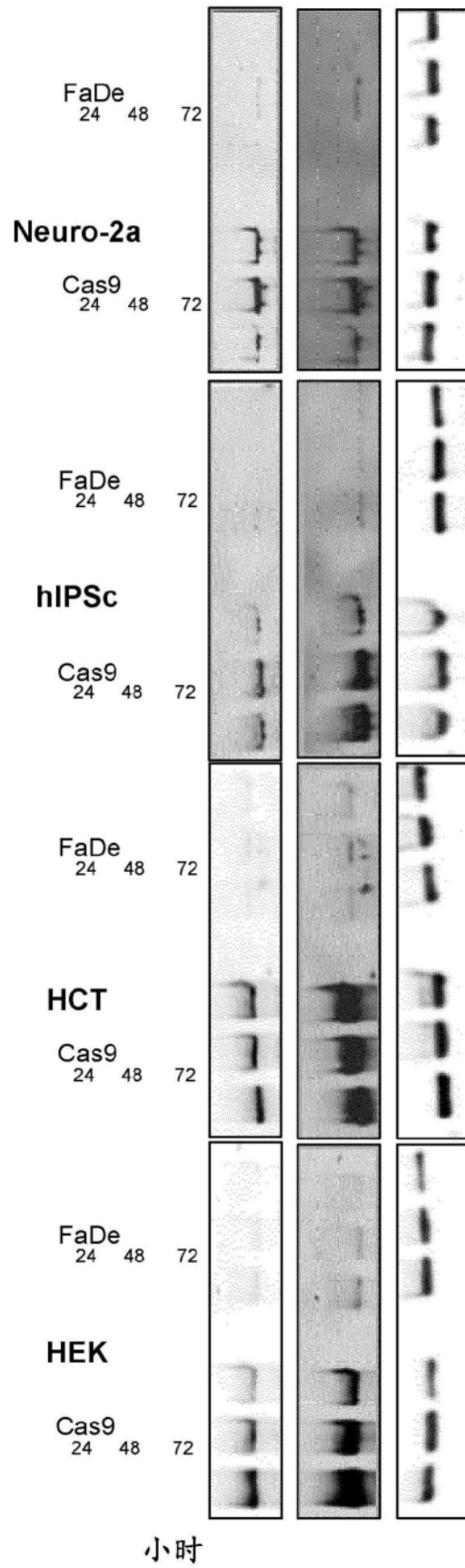


图12

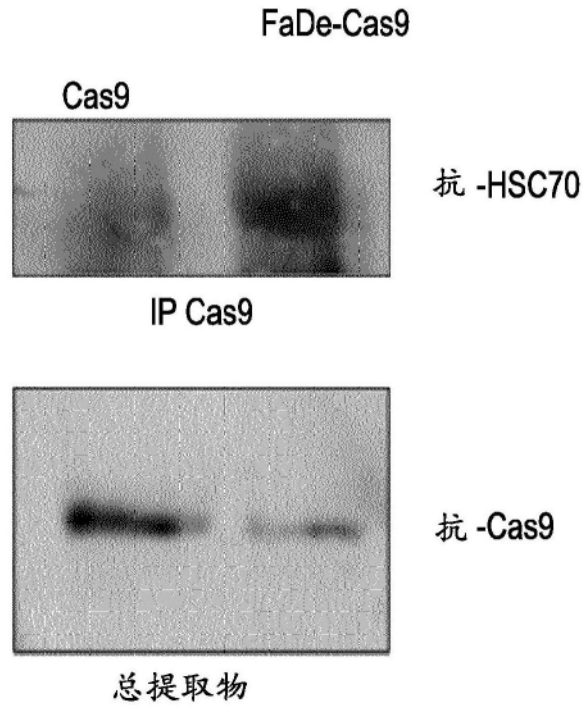


图13A

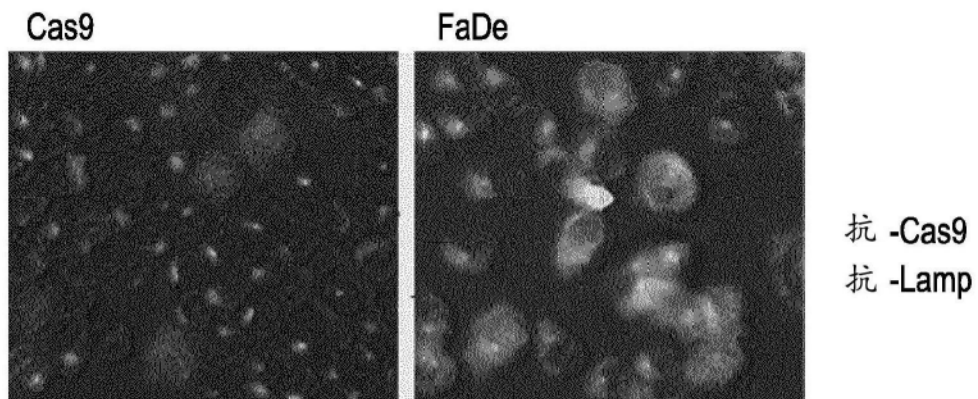


图13B

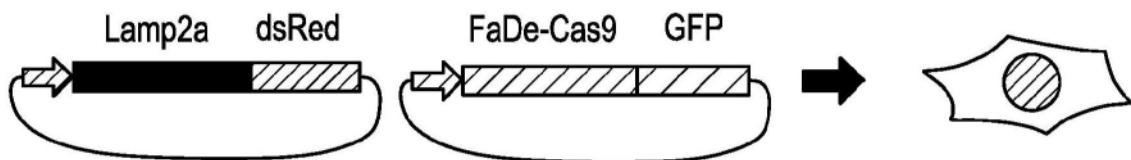


图14

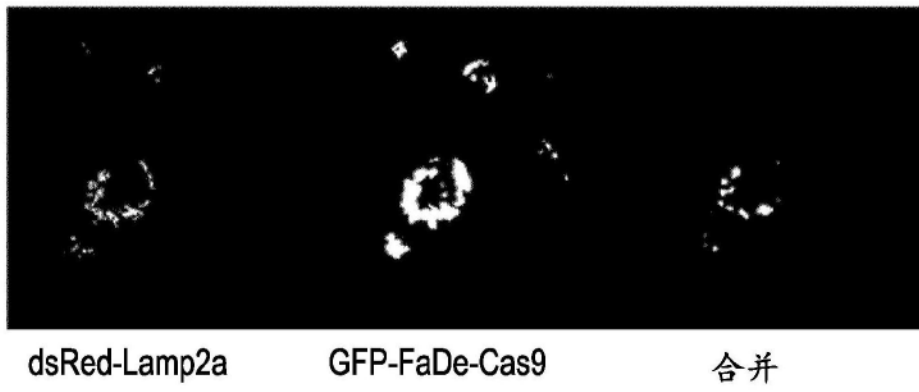


图15

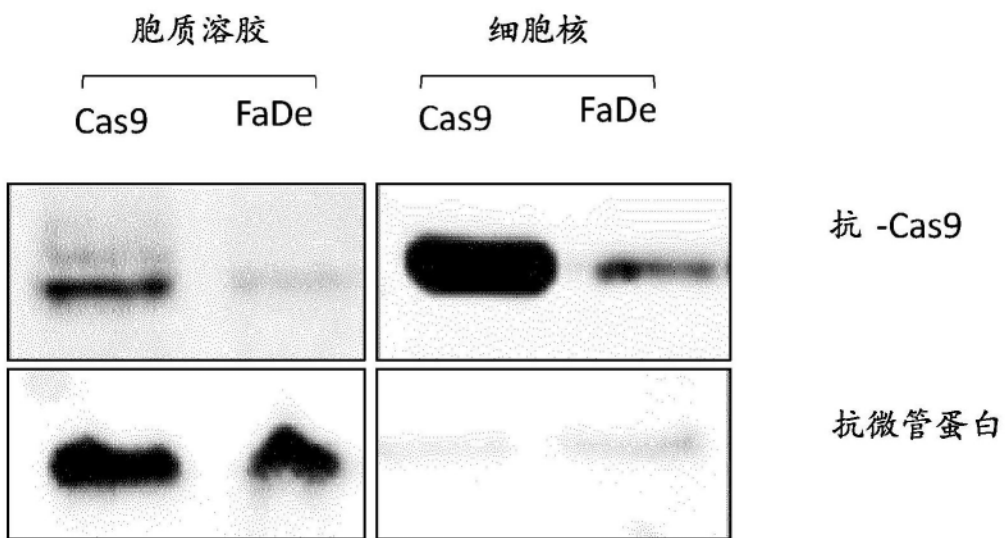


图16

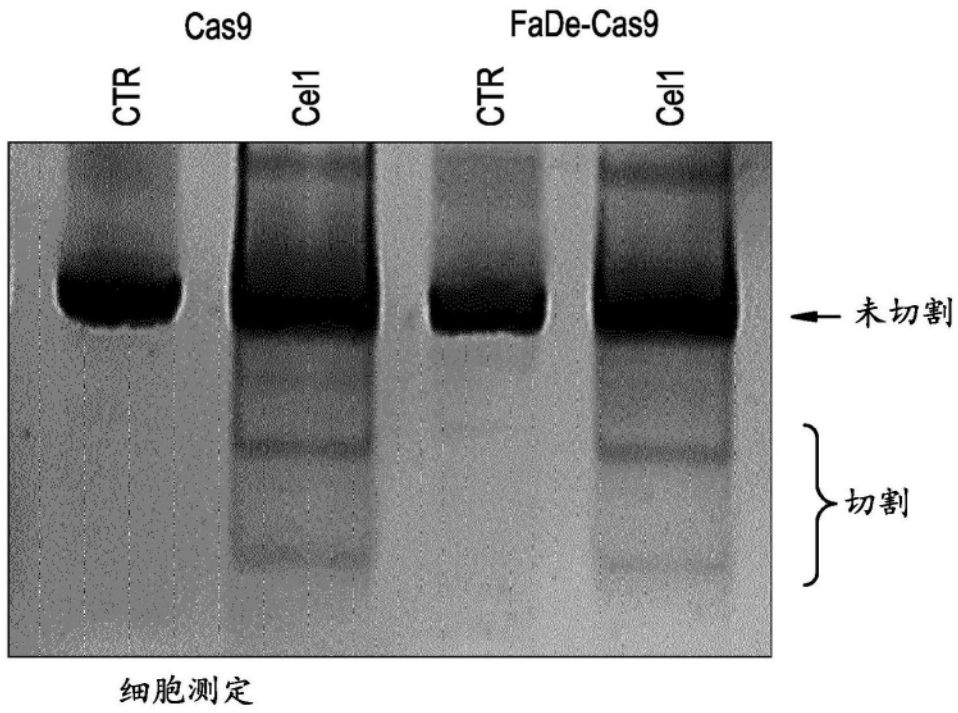


图17A

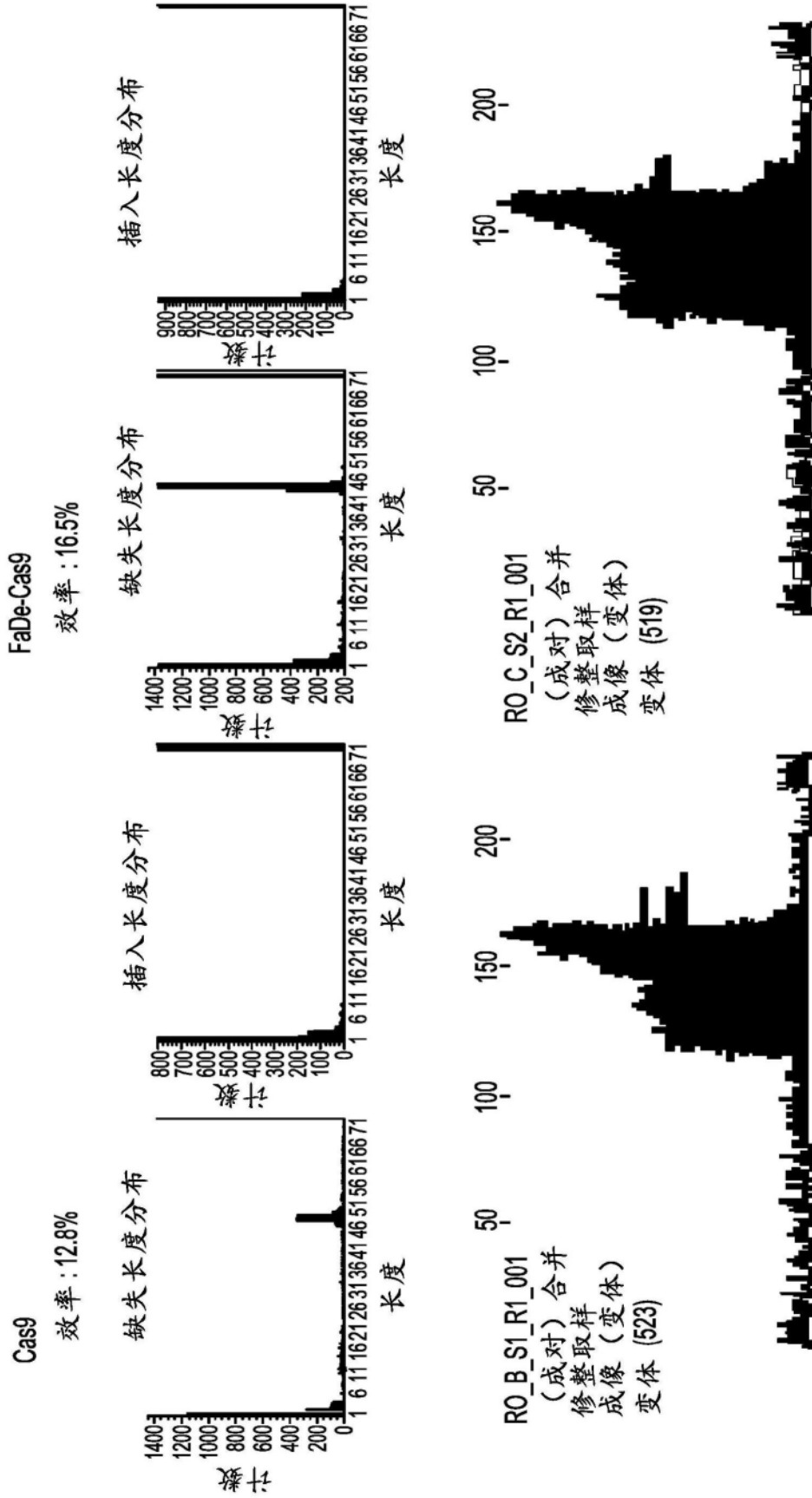


图17B

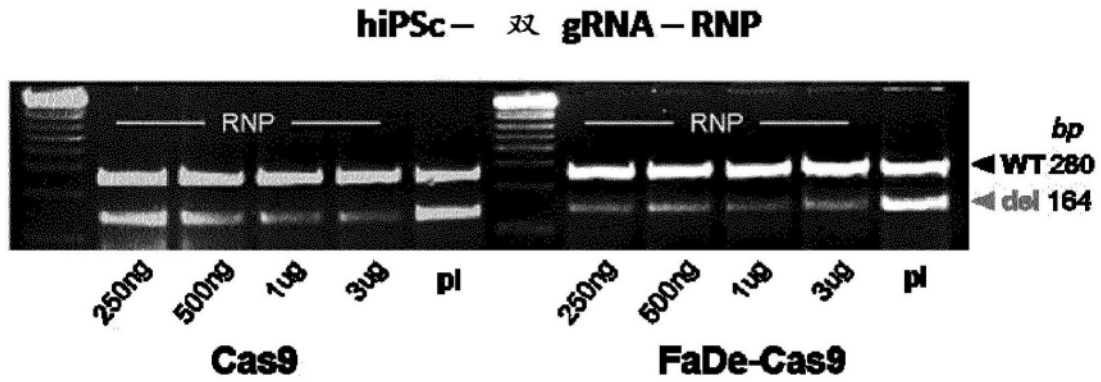


图18

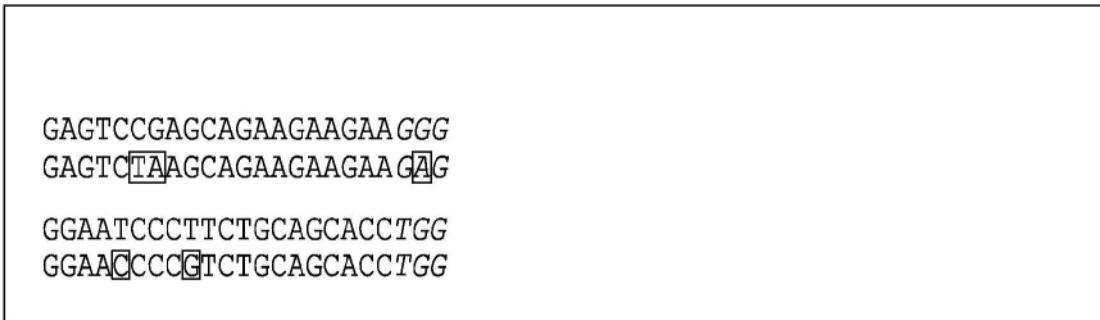
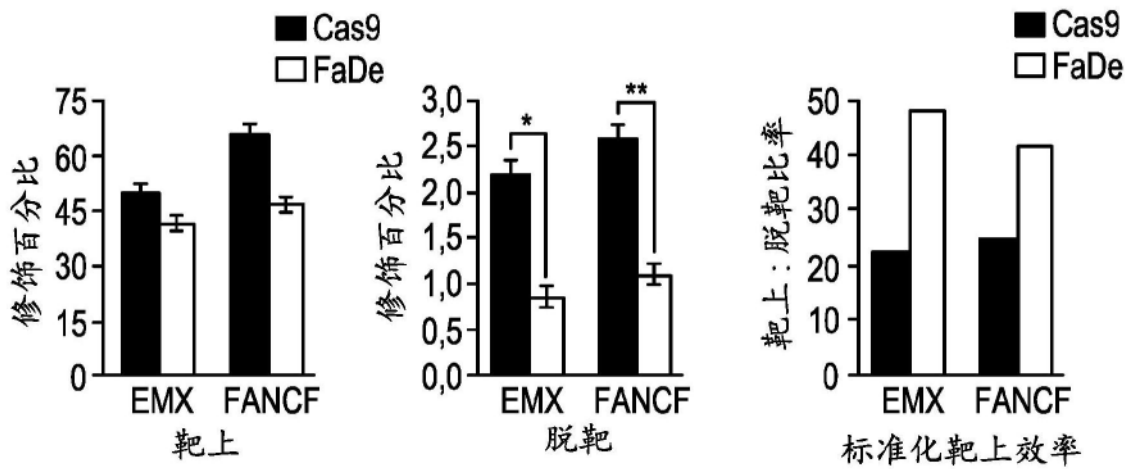


图19

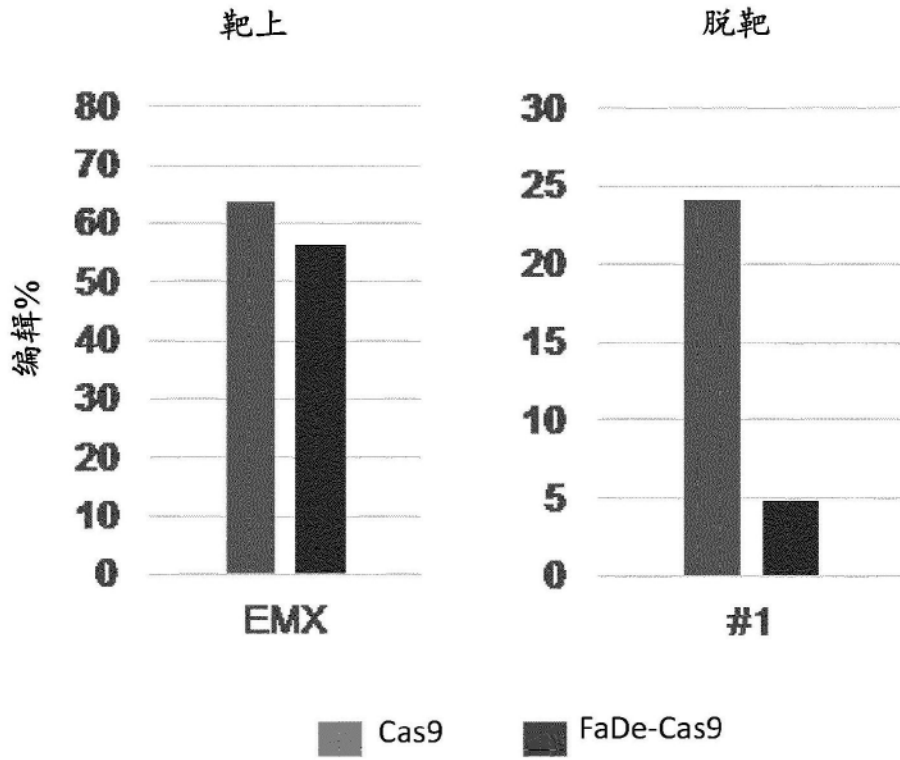


图20A

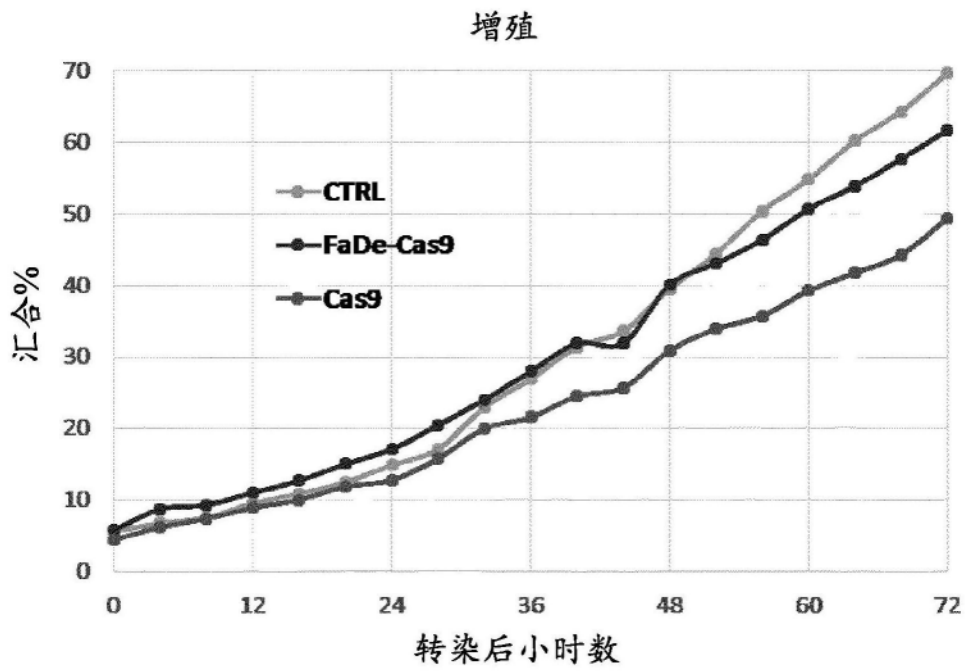


图20B

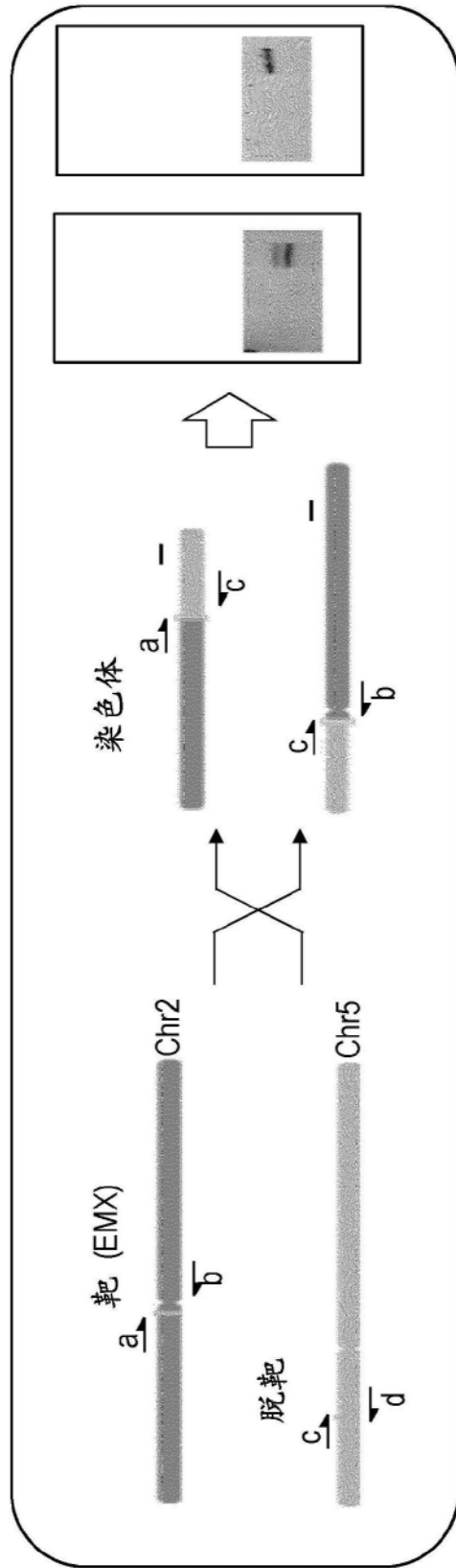


图21

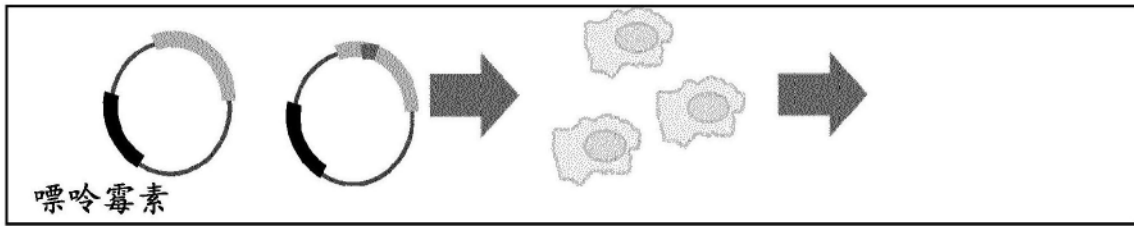


图22A

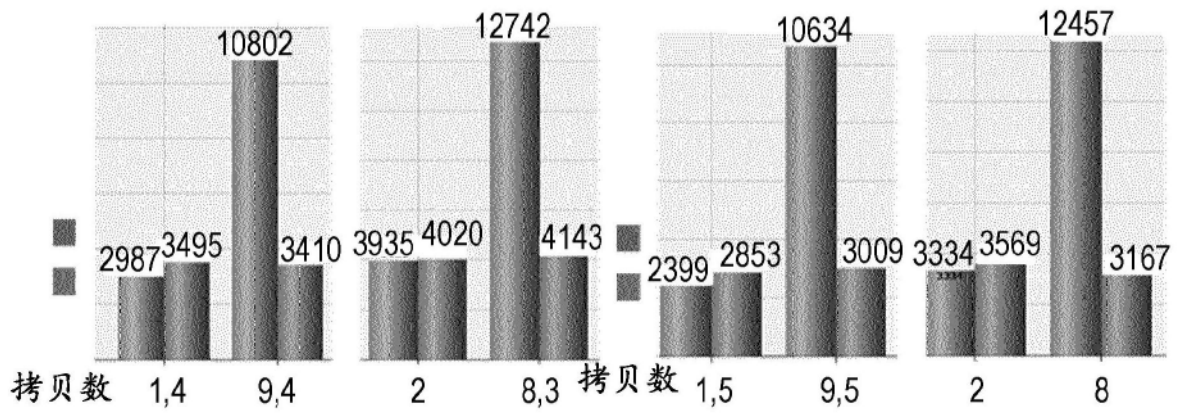


图22B

稳定池	拷贝数 Cas9	拷贝数嘌呤霉素
Cas9 池 1	1,5	1,4
Cas9 池 2	2	2
FaDe 池 1	9,5	9,4
FaDe 池 2	8	8,3

图22C

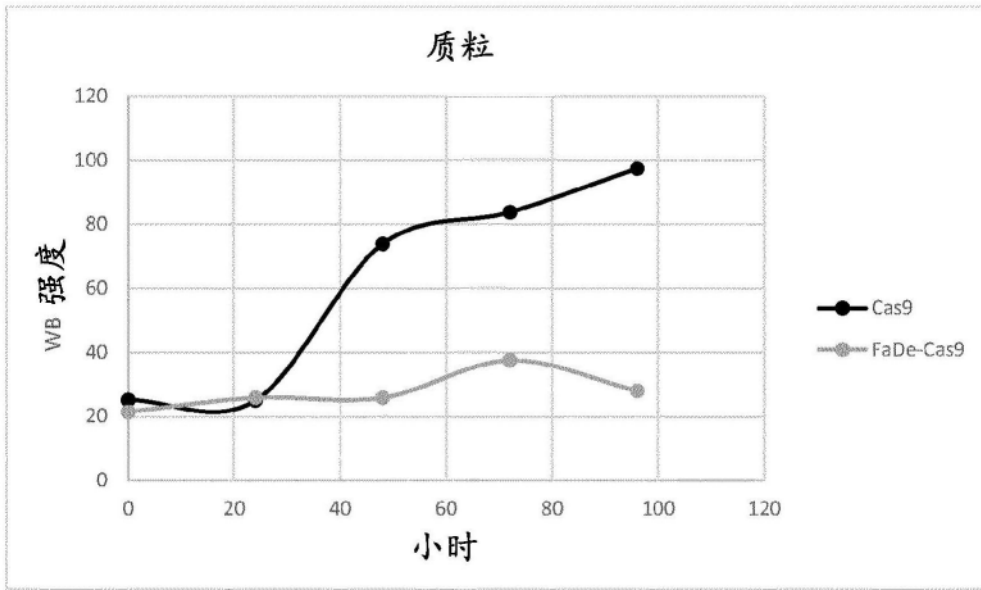


图23A

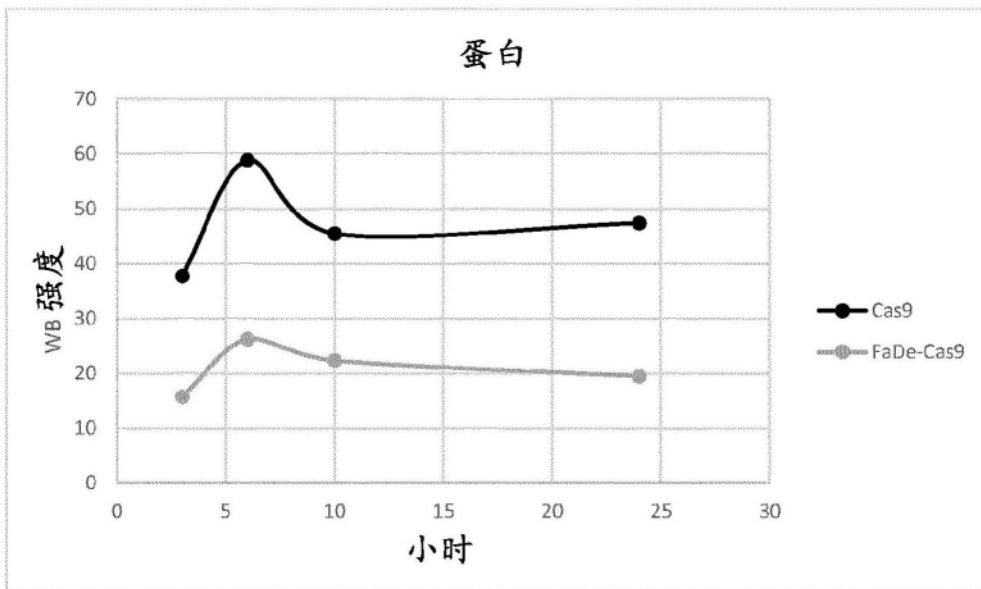


图23B

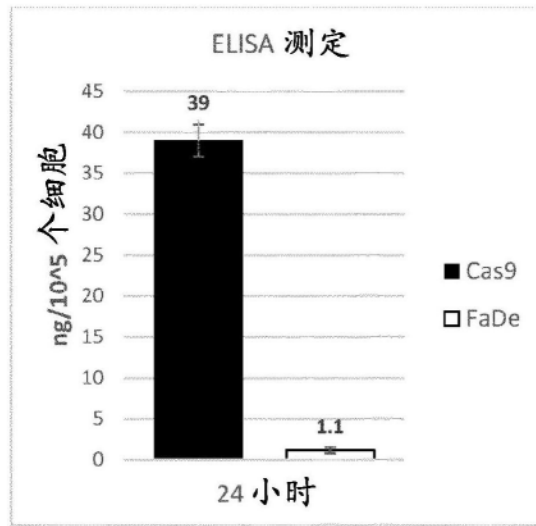


图24A

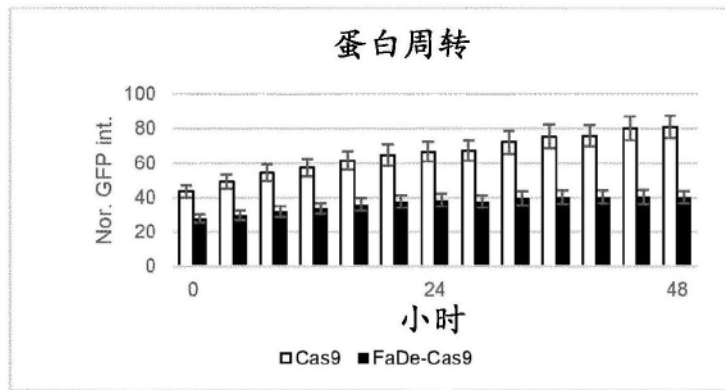


图24B

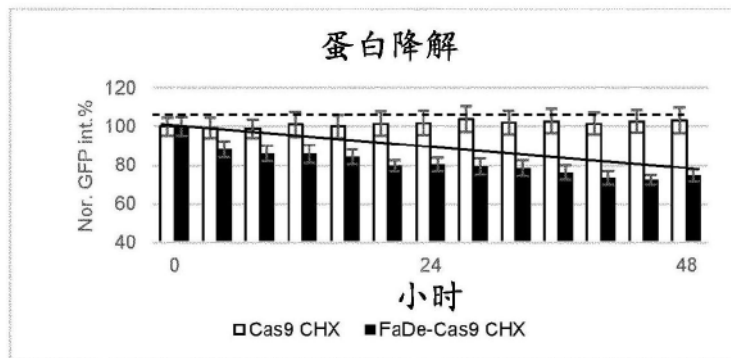


图24C

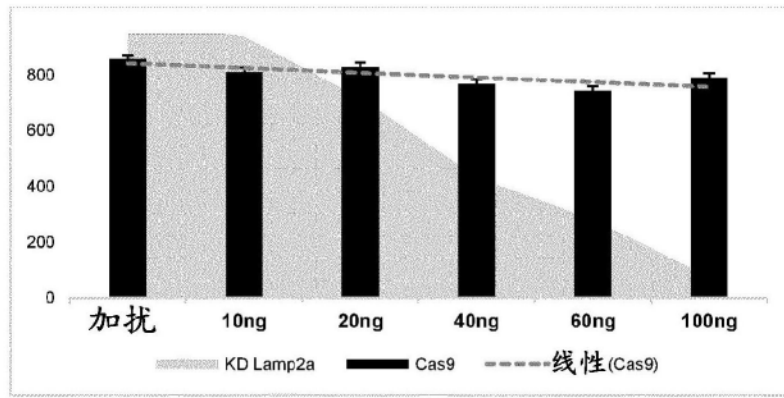


图25A

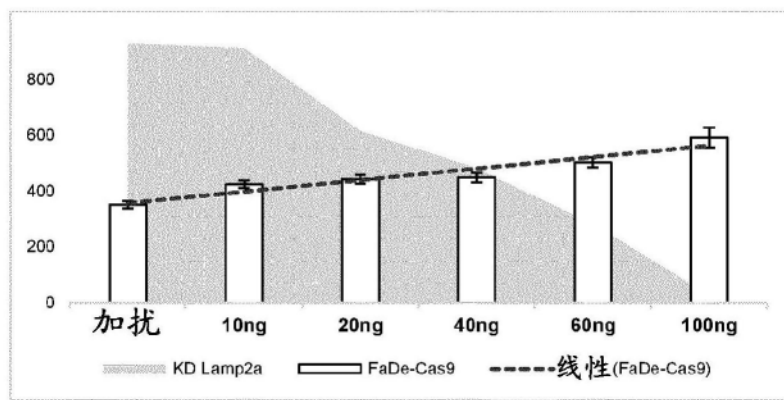


图25B

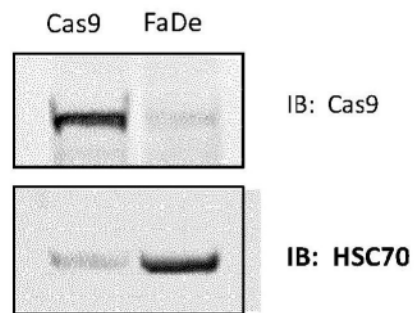


图26

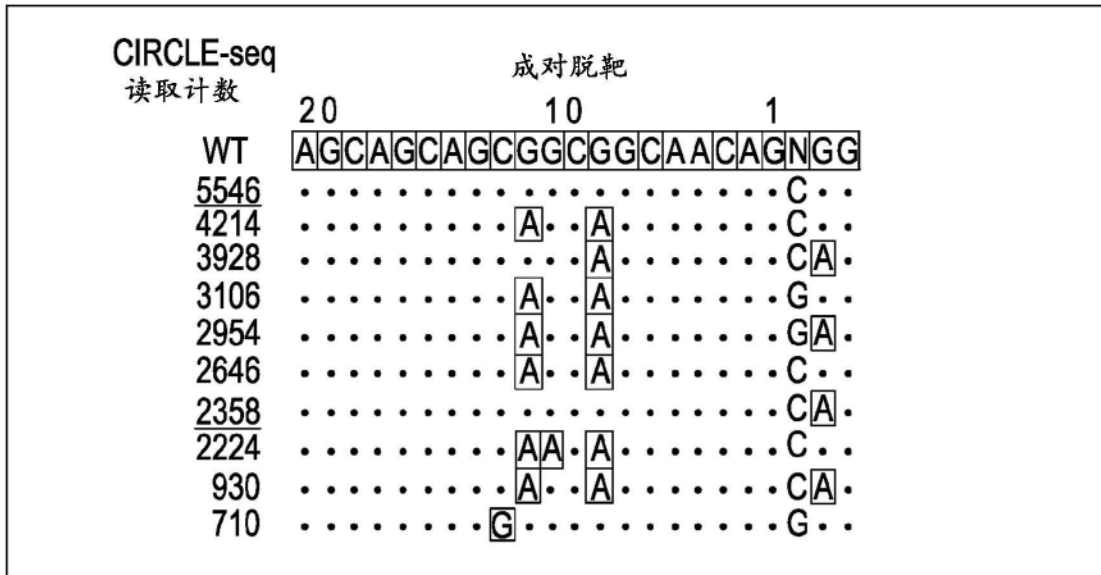


图27A

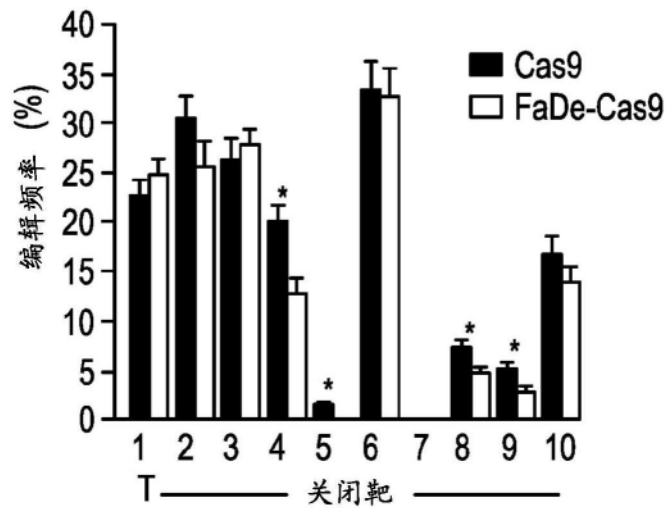


图27B

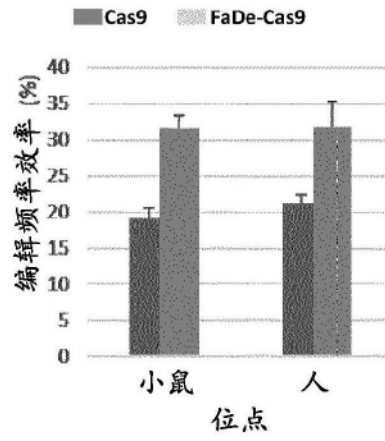


图28A

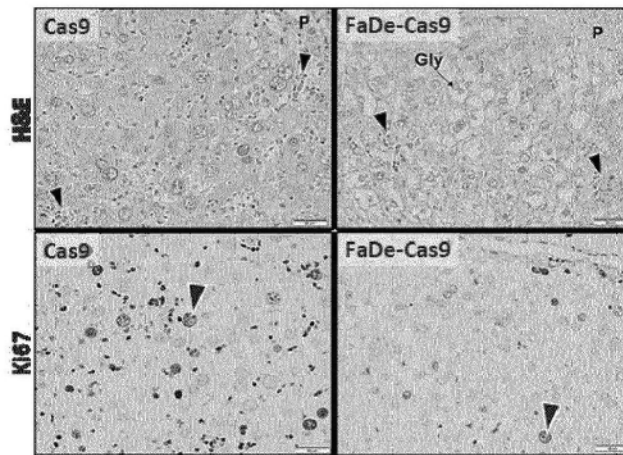


图28B

	Cas9-gMH	FaDe-gMH
小鼠数	3	3
糖原损失	3(2.3)	0
细胞浸润物	3(5)	3(2.3)
单细胞坏死	3(3.0)	3(1.7)
Ki67 表达	++	+

图28C

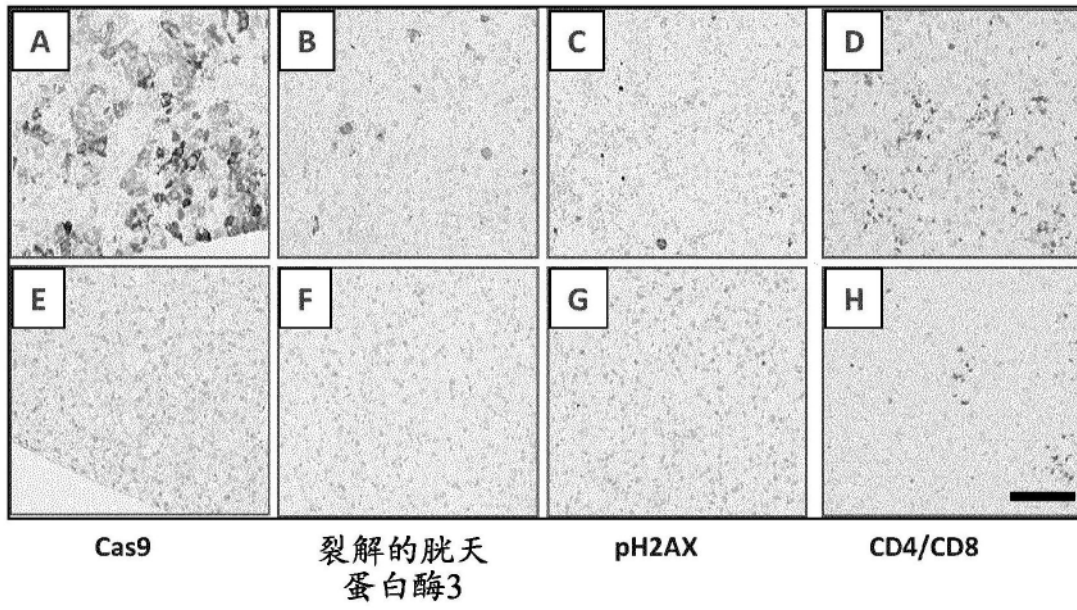


图28D