



(12) 发明专利

(10) 授权公告号 CN 114410625 B

(45) 授权公告日 2024. 06. 25

(21) 申请号 202111462778.0
(22) 申请日 2013.03.20
(65) 同一申请的已公布的文献号
申请公布号 CN 114410625 A
(43) 申请公布日 2022.04.29
(30) 优先权数据
61/613,373 2012.03.20 US
61/625,420 2012.04.17 US
(62) 分案原申请数据
201380023255.3 2013.03.20
(73) 专利权人 维尔纽斯大学
地址 立陶宛维尔纽斯
(72) 发明人 V·希克斯尼斯 G·加休那斯
T·卡维利斯 A·鲁比斯
L·扎里奥斯奇尼 M·格勒姆扎特
A·史密斯
(74) 专利代理机构 中国贸促会专利商标事务所
有限公司 11038
专利代理师 郑天松

C12N 9/22 (2006.01)
C12N 15/10 (2006.01)
C12N 15/90 (2006.01)
C12Q 1/6811 (2018.01)

(56) 对比文件
CN 110312792 A, 2019.10.08
CN 111448313 A, 2020.07.24
CN 112424824 A, 2021.02.26
CN 116555254 A, 2023.08.08
US 2023131972 A1, 2023.04.27
WO 2019217336 A2, 2019.11.14
WO 2022261115 A1, 2022.12.15
WO 2023025767 A1, 2023.03.02
孙明尧;白玮;梁宏玉;赵秋蓉;张悦榕;徐慰
倬.CRISPR-Cas9技术在医药领域中的应用.沈阳
药科大学学报.2019,(第12期),全文.
王玮玮;刘瑞琪;吴勇延;杨严格;王勇胜;卿
素珠.CRISPR/Cas9基因编辑系统研究进展及其
在动物基因编辑研究中的应用.畜牧兽医学报
.2016,(第07期),全文.

审查员 熊壮壮

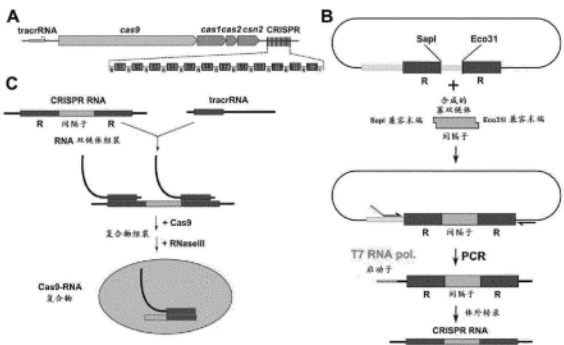
(51) Int.Cl.
C12N 15/113 (2010.01)

权利要求书2页 说明书37页
序列表34页 附图20页

(54) 发明名称
通过Cas9-crRNA复合物的RNA指导的DNA裂解

(57) 摘要
嗜热链球菌CRISPR3/Cas系统的Cas9-crRNA复合物的分离或体外组装及其用于DNA裂解的用途,所述DNA具有与crRNA互补的核苷酸序列和原型间隔子毗邻基序。使用RNA导向的DNA核酸内切酶在体外或体内对靶标DNA分子进行位点特异性修饰的方法,所述DNA核酸内切酶包含RNA序列及RuvC活性位点基序和HNH活性位点基序中至少一种;通过至少一个点突变使多肽中活性位点(RuvC或HNH)之一失活将Cas9多肽转化为双链DNA的切口酶裂解一条链的方法;在体内或体外

组装活性多肽-多核糖核苷酸复合物的方法;以及在体外对Cas9-crRNA复合物的特异性进行重新编程的方法,该方法中使用了含有单一的重复序列-间隔子-重复序列单元的序列盒子。



CN 114410625 B

[接上页]

(56) 对比文件

CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III.《Nature》.2011,第602-607页.

Rimantas Sapranas. The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli.《Nucleic Acids Research》.2011,第9275-9282页.

1. 制备针对靶DNA分子具有工程改造的特异性的Cas9-crRNA复合物的方法,所述方法包括:

将多核苷酸工程改造以编码能够与所述靶DNA分子的多核苷酸序列以位点特异性方式结合的成簇的规律间隔的短回文重复序列 (CRISPR) RNA (crRNA),其中所述工程改造包括将所述多核苷酸的序列重新编程以编码与所述靶DNA分子的多核苷酸序列互补的crRNA的多核苷酸序列,所述多核苷酸还编码反式激活RNA (tracrRNA);

将编码所述crRNA和所述tracrRNA的多核苷酸与编码CRISPR相关多肽9 (Cas9多肽) 的多核苷酸在体外组合;及

形成Cas9-crRNA复合物,

其中所述Cas9-crRNA复合物针对所述靶DNA分子的位点具有工程改造的特异性,用于修饰所述靶DNA分子。

2. 根据权利要求1所述的方法,其中与所述靶DNA分子的多核苷酸序列互补的crRNA的多核苷酸序列包含至少20个核苷酸。

3. 根据权利要求1所述的方法,其中由所述多核苷酸编码的Cas9多肽包含Cas9多肽的RuvC活性位点基序或HNH活性位点基序中的突变。

4. 根据权利要求1所述的方法,其中所述靶DNA分子是双链的。

5. 根据权利要求1所述的方法,其中所述靶DNA分子是基因组DNA。

6. 根据权利要求1所述的方法,其中编码crRNA的多核苷酸、编码Cas9多肽的多核苷酸或两者是质粒。

7. 根据权利要求1所述的方法,其中编码crRNA的多核苷酸、编码Cas9多肽的多核苷酸或两者在表达盒中。

8. 制备Cas9-crRNA复合物的方法,所述方法包括:

提供编码成簇的规律间隔的短回文重复序列序列 (CRISPR) 相关多肽9 (Cas9多肽) 的多核苷酸;

提供编码工程改造的CRISPR RNA (crRNA) 和反式激活RNA (tracrRNA) 的多核苷酸;

将所述多核苷酸体外组合;及

形成Cas9-crRNA复合物;

其中产生由所述多核苷酸编码的工程改造的crRNA,以将Cas9-crRNA复合物引导至包含靶DNA分子中的位点的区域,从而使所述Cas9-crRNA复合物与所述靶DNA分子结合;且

其中所述工程改造的crRNA不是通过细菌CRISPR重复序列-间隔子阵列的加工产生的。

9. 根据权利要求8所述的方法,其中所述工程改造的crRNA通过以下方式获得:

鉴定所述靶DNA分子的多核苷酸序列;和

通过将所述多核苷酸的多核苷酸序列重新编程来产生多核苷酸,以产生具有与所述靶DNA分子的多核苷酸序列互补的多核苷酸序列的工程改造的crRNA。

10. 根据权利要求9所述的方法,其中与所述靶DNA分子的多核苷酸序列互补的工程改造的crRNA的多核苷酸序列包含至少20个核苷酸。

11. 根据权利要求8所述的方法,其中由所述多核苷酸编码的Cas9多肽包含Cas9多肽的RuvC活性位点基序或HNH活性位点基序中的突变。

12. 根据权利要求8所述的方法,其中所述靶DNA分子是双链的。

13. 根据权利要求8所述的方法,其中所述靶DNA分子是基因组DNA。
14. 根据权利要求8所述的方法,其中编码工程改造的crRNA的多核苷酸、编码Cas9多肽的多核苷酸或两者是质粒。
15. 根据权利要求8所述的方法,其中编码crRNA的多核苷酸、编码Cas9多肽的多核苷酸或两者在表达盒中。

通过Cas9-crRNA复合物的RNA指导的DNA裂解

[0001] 本申请是2013年3月20日提交的同名发明专利申请201380023255.3的分案申请。

[0002] 本申请要求同时申请的美国申请序列号61/613,373(提交于2012年3月20日)及61/625,420(提交于2012年4月17日)的优先权,二者都明确地以其全文并入本文作为参考。

[0003] 摘要

[0004] CRISPR/Cas系统为细菌和古细菌提供针对病毒和质粒的适应性免疫。预先装载有小干扰crRNA的核糖核蛋白(RNP)复合物执行对入侵核酸的沉默,其中crRNA引导对外源核酸的靶向和降解。在本文中,我们描述了Cas9-crRNA复合物的分离,并证明其在体外产生靶标DNA分子中特异性位点上的双链断裂,其中靶标DNA分子与crRNA序列互补并在匹配序列邻近具有短的原型间隔子序列毗邻基序(proto-spacer adjacent motif,PAM)。我们表明,对DNA的裂解是由Cas9内部两个不同的活性位点(RuvC和HNH)进行的,二者在相反的DNA链上生成位点特异性切口。Cas9-crRNA复合物的序列特异性是由42nt的crRNA控制的,其包括与靶标DNA中原型间隔子序列互补的20nt片段。复合物可在体外或体内装配。总体而言,我们的数据证明Cas9-crRNA复合物起到RNA指导的核酸内切酶作用,识别序列特异性靶标位点并在两个不同的链切口处进行裂解。

[0005] 发明背景

[0006] 成簇的规律间隔的短回文重复序列(CRISPR)与cas(CRISPR相关)基因构成适应性免疫系统,其为细菌和古细菌提供针对入侵的外源核酸的获得性抗性(Barrangou等人,2007.Science 315:1709-12)。CRISPR由短保守重复序列的阵列构成,其中间隔有类似大小的独特可变DNA序列,称为间隔子,这些序列通常起源于噬菌体或质粒DNA(Barrangou等人,2007.Science 315:1709-12;Bolotin等人,2005.Microbiology 151:2551-61;Mojica等人,2005.J Mol Evol 60:174-82)。CRISPR-Cas系统发挥作用是通过获得外源DNA的短片段(间隔子),将其插入到CRISPR区域中,并在随后暴露于携带有匹配序列的噬菌体和质粒时为其提供免疫(Barrangou等人,2007.Science 315:1709-12;Brouns等人,2008.Science 321:960-4) CRISPR-Cas免疫通常是通过三个阶段进行的,称为i)适应/免疫化/间隔子获得,ii)CRISPR表达/crRNA生物生成,iii)干扰/免疫。(Horvath&Barrangou,2010.Science 327:167-70;Deveau等人,2010.Annu Rev Microbiol.64:475-93;Marraffini&Sontheimer,2010.Nat Rev Genet 11,181-90;Bhaya等人,Annu Rev Genet 45:273-97;Wiedenheft等人,2012.Nature 482:331-338)。本文中我们具体聚焦于使得crRNA介导的外源核酸沉默进行的干扰/免疫步骤。

[0007] CRISPR-Cas系统是高度多样化的,其分类为三种主要类型,这三种类型基于核心元件含量和序列又进一步划分为十种亚型(Makarova等人,2011.Nat Rev Microbiol 9:467-77)。crRNA介导的外源核酸沉默中所涉及的核蛋白复合物的结构组织和功能对于不同的CRISPR/Cas类型是有差别的(Wiedenheft等人,2012.Nature 482:331-338)。在以大肠埃希氏菌为例的I-E型系统中,crRNA被整合入称为Cascade(用于抗病毒防御的CRISPR相关复合物)的多亚基效应子复合物中(Brouns等人,2008.Science 321:960-4),其结合于靶标DNA并通过签名Cas3蛋白质引发降解(Sinkunas等人,2011.EMBO J 30:1335-42;

Beloglazova等人,2011.EMBO J 30:616-27)。在矿硫化叶菌(*Sulfolobus solfataricus*)和强烈炽热球菌(*Pyrococcus furiosus*)的III型CRISPR/Cas系统中,Cas RAMP模块(Cmr)和crRNA复合物在体外识别并裂解合成的RNA(Hale等人,2012.Mol Cell 45:292-302; Zhang等人,2012.Mol Cell,45:303-13),而表皮葡萄球菌(*Staphylococcus epidermidis*)的CRISPR/Cas系统在体内靶向DNA(Marraffini&Sontheimer,Science.322:1843-5)。

[0008] 在II型CRISPR/Cas系统,更具体地是嗜热链球菌(*Streptococcus thermophilus*) DGCC7710的CRISPR3/Cas系统中(Horvath&Barrangou,2010.Science 327:167-70)所进行的DNA沉默中涉及的RNP复合物由四种cas基因cas9、cas1、cas2和csn2组成,这些基因位于12个重复间隔子单元的上游(图1A)。Cas9(之前称为cas5或csn1)是II型系统的签名基因(Makarova等人,2011.Nat Rev Microbiol 9:467-77)。在与之密切相关的嗜热链球菌CRISPR1/Cas系统中,破坏cas9会消除crRNA介导的DNA干扰(Barrangou等人,2007.Science 315:1709-12)。我们在近期的研究中已经表明,嗜热链球菌的CRISPR3/Cas系统可被转移至大肠埃希氏菌(*Escherichia coli*)中,并且这种异源系统从头开始提供了针对质粒转化和噬菌体感染的保护(Sapranauskas等人,2011.Nucleic Acids Res 39:9275-82)。嗜热链球菌CRISPR3提供的针对噬菌体和质粒DNA的干扰要求靶标DNA中存在与间隔子衍生的crRNA互补的原型间隔子序列,以及保守的PAM(原型间隔子毗邻基序)序列NGGNG,其紧邻原型间隔子下游(Deveau等人,2008.J Bacteriol 190:1390-400;Horvath等人,2008.J Bacteriol 190:1401-12;Mojica等人,2009.Microbiology155:733-40)。PAM或限定的原型间隔子位置中单一点突变使得噬菌体或质粒能避免CRISPR介导的免疫(Deveau等人,2008.J Bacteriol 190:1390-400;Garneau等人,2010.Nature 468:67-71;Sapranauskas等人,2011.Nucleic Acids Res 39:9275-82)。我们已经确认在异源系统中,cas9是CRISPR编码的干扰中唯一需要的cas基因(Sapranauskas等人,2011.Nucleic Acids Res 39:9275-82),这表明此蛋白质参与了crRNA加工和/或crRNA介导的对入侵DNA的沉默。嗜热链球菌CRISPR3/Cas系统中的Cas9是一个大的多结构域蛋白质,由1,409aa残基构成(Sapranauskas等人,2011.Nucleic Acids Res 39:9275-82)。其含有两个核酸酶结构域,其一是靠近氨基末端的RuvC样核酸酶结构域,另一个是位于蛋白质中部的HNH样结构域。突变分析已经确认Cas9提供的体内干扰需要RuvC和HNH两种基序的存在(Sapranauskas等人,2011.Nucleic Acids Res 39:9275-82)。

[0009] 本文提供了嗜热链球菌CRISPR3/Cas系统中Cas9-crRNA复合物的分离以及分离组分在体外的复合物组装,且证明了其以PAM依赖的方式裂解具有与crRNA互补的核苷酸序列的合成寡脱氧核糖核苷酸或是质粒DNA。进一步地,我们提供了实验证据,证明PAM是在双链DNA中被识别的,并且对体外DNA结合和裂解起至关重要的作用。最后,我们表明,Cas9的RuvC-和HNH-活性位点负责相反DNA链的裂解。总的说来,我们的数据表明Cas9-crRNA复合物起RNA指导的核酸内切酶的作用,其使用RNA进行靶标位点识别,使用Cas9进行DNA裂解。Cas9-crRNA复合物的简单模块化组织为工程改造普适性的RNA指导的DNA核酸内切酶提供了灵活的平台,其中对DNA靶标的特异性是由小crRNA编码而裂解机制则由单个多结构域Cas蛋白质构成。事实上,我们提供证据证明,通过改变Cas9-crRNA复合物中的RNA序列可以为体外和体内应用设计可编程的内切核酸酶,我们还提供了所述新型应用的定制方案。这些发现为RNA指导的DNA手术开发新型分子工具奠定了基础。

[0010] 发明概述

[0011] 提供了用于对靶标DNA分子进行位点特异性修饰的方法,通过在合适的条件下将靶标多脱氧核糖核苷酸分子;与包含至少一种RNA序列及RuvC活性位点基序及HNH活性位点基序中至少之一的RNA指导的DNA核酸内切酶相接触;产生的靶标多脱氧核糖核苷酸分子在一定的区域得到修饰,该区域是由RNA序列向靶标DNA分子的互补结合确定的。该方法包括在合适的条件下对组合物进行孵育,所述组合物包括靶标双链多脱氧核糖核苷酸或单链多脱氧核糖核苷酸;其中双链多脱氧核糖核苷酸含有短的原型间隔子毗邻基序(PAM),这对于单链多脱氧核糖核苷酸不是需要的;其中的PAM包含5'NGGNG-3'序列;多核糖核苷酸(crRNA)包含3'和5'区域,其中3'区域包含含有CRISPR基因座的存在于微生物中的至少22nt的重复,而5'区域则包含紧邻CRISPR基因座中重复的下游的至少20nt的间隔子序列,其与靶标多核苷酸的部分大体上互补,可选地与其互补;多肽,其中多肽的氨基酸序列与SEQ ID NO:1的氨基酸序列具有至少80%的同一性,所述多肽分离自嗜热链球菌或遗传修饰的微生物包括遗传修饰的大肠杆菌,或者其中的多肽由选自重组DNA技术或化学合成的方法产生;核苷酸序列为SEQ ID NO:5(或与其有至少80%的同一性)的多核糖核苷酸tracrRNA,包含5'和3'区域,其中5'区域由与crRNA的22个核苷酸3'区域互补的至少22个核苷酸构成,以及3'区域。其中多核糖核苷酸是通过体外转录或化学合成产生的。其中,合适的条件指的是可能发生反应的体外或体内条件。

[0012] 提供了将Cas9多肽转化为切口酶(其仅仅裂解双链DNA的一条链)的方法,其通过产生至少一个点突变使多肽上活性位点(RuvC或HNH)之一失活,示例有D31A(SEQ ID NO:2),N891A(SEQ ID NO:3)和H868A(SEQ ID NO:4)点突变。RuvC基序突变体仅裂解对于5'NGGNG-3'基序而言的底部DNA链,而HNH基序突变体裂解顶部DNA链。

[0013] 多肽-多核糖核苷酸复合物可分离自经遗传修饰的微生物(例如大肠埃希氏菌或嗜热链球菌),或在体外由分开的组分组装而成。在遗传修饰的微生物组分中,复合物的组分可以是在含有来自遗传修饰微生物的宿主启动子或来自原生宿主基因组的启动子的一种、两种或三种分开的质粒上编码的。

[0014] 提供了用于在体外组装活性多肽-多核糖核苷酸复合物的方法,包含在适合复合物组装的条件下孵育组合物的组分。所述复合物可使用三或四种组分组装。用于三种组分组装的方法包括对Cas9多肽、78nt的tracrRNA多核糖核苷酸(SEQ ID NO:5)及42nt的crRNA多核糖核苷酸(5'-NNNNNNNNNNNNNNNNNNN GUUUUAGAGCUGUGUUGUUUCG-3')(SEQ ID NO:15)在适合复合物组装的条件下进行孵育。用于四种组分组装的方法包括对Cas9多肽;102nt的tracrRNA多核糖核苷酸(SEQ ID NO:6);含有序列5'-NNNNNNNNNNNNNNNNNNNGUUUUAGAGCUGUGUUGUUUCG-3'(SEQ ID NO:15)和邻侧区域以及裂解双链RNA多核苷酸的RNase III多肽进行孵育。含有序列5'-NNNNNNNNNNNNNNNNNNNGUUUUAGAGCUGUGUUGUUUCG-3'(SEQ ID NO:15)的多核糖核苷酸的实例为SEQ ID NO:8、SEQ ID NO:9、SEQ ID NO:10、SEQ ID NO:11和SEQ ID NO:12)。合适的RNaseIII的来源的实例包括大肠埃希氏菌或嗜热链球菌。

[0015] 提供了用于对Cas9-crRNA复合物特异性进行重新编程的方法,是通过将分开的组分相混合或使用含有单个重复序列-间隔子-重复序列单元的序列盒子(cassette)完成的。任何序列都可使用合适的限制性内切酶插入序列盒子中的两个重复序列之间。所述序列盒子可用于在体内靶向序列,或产生适合体外复合物组装的RNA核糖核苷酸。

[0016] 附图简述

[0017] 图1显示了与crRNA共纯化的Cas9蛋白质。(A)嗜热链球菌CRISPR3/Cas系统的示意图。四个cas基因(cas9、cas1、cas2、csn2)位于CRISPR重复序列-间隔子阵列的上游,由13个重复序列(R)和12个独特的间隔子(S1-S12)构成。II型CRISPR系统的crRNA成熟中需要的tracrRNA(Deltcheva等人,2011.Nature 471:602-7)位于cas9基因的上游,并在相对CRISPR3/Cas系统其他元件的相反DNA链(以箭头指出)上编码。(B)用于Cas9-crRNA共表达的两个质粒中异源基因座的示意图。大肠杆菌RR1菌株含有pCas9(-)1SP(编码Cas1、Cas2、Csn2、SP1和tracrRNA)及pASKIBA-Cas9(编码Cas9的Strep标签形式)质粒。(C)使用抗-crDNA寡核苷酸作为探针对Cas9-crRNA复合物进行的Northern分析。M1-对应于间隔子S1-重复序列单元的84nt的寡脱氧核糖核苷酸;M2-对应于推定的嗜热链球菌CRISPR3crRNA的42nt的合成寡核糖核苷酸(见图4);crRNA(wt)-分离自wt Cas9复合物的crRNA;K1-用DNase I对crRNA(wt)处理15min;K2-用RNase I处理15min的crRNA(wt),D31A-纯化自Cas9 D31A突变体复合物的crRNA;N891A-纯化自Cas9 N891 A突变体复合物的crRNA。

[0018] 图2显示了Cas9-crRNA复合物对DNA的裂解情况,该复合物是通过Cas9与全长CRISPR基因座共表达而获得。(A)携带内源SthCRISPR3/Cas系统的12间隔子-重复序列阵列的重组pCas9(-)质粒以及携带C末端具有Strep标签的cas9基因的pASKIBA-Cas9质粒的CRISPR/Cas基因座示意图。(B)寡双链体裂解测定。在大肠杆菌中共表达pCas9(-)和pASKIBA-Cas9两种质粒,纯化Cas9-crRNA复合物并对其进行裂解分析,使用的是在(+)链5'端以33P标记的SP1(第一原型间隔子)和SP2(第二原型间隔子)寡双链体。在PAA凝胶上对反应产物进行分析。

[0019] 图3显示了SthCRISPR3/Cas系统提供的大肠杆菌细胞中对质粒转化的免疫。(A)携带内源SthCRISPR3/Cas系统的12间隔子-重复序列阵列的重组pCRISPR3质粒以及携带1个间隔子-重复序列单元的工程改造pCRISPR3-SP1质粒的CRISPR/Cas基因座示意图。(B)大肠杆菌细胞中SthCRISPR3/Cas系统对质粒转化的干扰。用携带原型间隔子和PAM或pUC18(1)的质粒pSP1对携带质粒pACYC184,pCRISPR3或pCRISPR3-SP1的大肠埃希氏菌RR1受体菌株进行转化。转化效率以每纳克质粒DNA的cfu(平均值±SD)表示。

[0020] 图4显示了来自嗜热链球菌DGCC7710、LMD-9以及化脓链球菌(*S. pyogenes*) SF370菌株的IIA型CRISPR/Cas系统的比较。(A)CRISPR/Cas系统的组织示意图。对应化脓链球菌(2)的crRNA成熟所需要的tracrRNA的核苷酸序列存在于LMD-9和DGCC7710中。由虚线相连的对应的蛋白质序列之间相同和类似的残基(以括号表示)的百分比。(B)保守重复序列和tracrRNA的序列比对。来自DGCC7710和LMD-9的对应序列是相同的。在比对的序列下用星号标记在三种菌株中都相同的核苷酸位置。图4(B)以出现顺序分开地公开了SEQ ID NO 50、50-52和52-53。(C)crRNA序列的比较。化脓链球菌crRNA的序列和长度是通过深度测序分析确定的(2)。通过northern印迹分析确定来自嗜热链球菌LMD-9(2)和DGCC7710(本研究)菌株的crRNA的近似长度。图4(C)以出现顺序分别公开了SEQ ID NOS54-56。

[0021] 图5显示Cas9-crRNA复合物在体外裂解原型间隔子内的双链DNA。(A)裂解测定中使用的寡双链体底物。55nt的寡双链体SP1含有原型间隔子1(红色字体)、PAM(蓝色字体)并在两边都有一段与pSP1质粒中相同的10nt的邻侧序列。在SP1寡双链体DNA中,与crRNA(红色字体)的5'末端片段互补的DNA链命名为(+)链,相反的DNA链命名为(-)链。图5(A)以出现

顺序分别公开了SEQ ID NOS 31、7和34。(B) 寡双链体SP1的裂解情况。在37℃下将2.5nM Cas9-crRNA复合体和1nM SP1寡双链体(+)或(-)链的5'端以33P标记)加入反应缓冲液(10mM Tris-HCl, pH=7.5, 10mM NaCl, 10mM MgCl₂, 0.1mg/ml BSA)中孵育不同时间长度(30s-10min)并在20%的PAA凝胶中分析反应产物。M1和M2泳道含有化学合成的5'端以33P标记的37nt和18nt的寡脱氧核糖核苷酸,其分别对应于(-)和(+)DNA链的裂解产物。裂解位置以箭头指示。图5(B)公开了SEQ ID NO:31。(C) 用于质粒裂解测定的pSP1质粒(Sapranauskas等人,2011.Nucleic Acids Res 39:9275-82)示意图。图5(C)公开了SEQ ID NO:57。(D) pSP1质粒裂解情况。对pSP1裂解产物进行琼脂糖凝胶分析(左图)。SC-超螺旋质粒DNA, OC-在一条链上有切口的开环DNA, FLL-全长线性DNA, 在两条链上有剪口。37℃下最终的反应混合物在反应缓冲液(B部分)中含有2.5nM pSP1质粒和2.5nM Cas9-crRNA复合物。pSP1质粒裂解产物的(+)链(上部)和(-)链(下部)的直接测序电泳图谱(右图)。序列末端的腺嘌呤的非模板加入是由聚合酶引起的测序矫作物(此处显示的是反向互补序列中的T)。图5(D)以出现顺序分别公开了SEQ ID NO57-59、58和60。

[0022] 图6显示了缺少crRNA的Cas9-Chis蛋白质的DNA结合和裂解分析。

[0023] Cas9-Chis蛋白质结合到(A)双链SP1寡双链体和(B)单链的s(+)SP1寡核苷酸的电泳迁移率变动分析(EMSA)。电泳迁移率变动实验是在结合缓冲液中进行的(40mM Tris-乙酸, 25℃下pH 8.3, 0.1EDTA, 0.1mg/ml BSA, 10% v/v甘油)。反应物含有0.5nM的33P-标记的寡双链体,以及如每个泳道上标记的浓度的蛋白质。(C) .寡核苷酸裂解测定。将5nM Cas9-Chis蛋白质与1nM寡核苷酸在反应缓冲液中(10mM Tris-HCl, pH=7.5, 10mM NaCl, 10mM MgCl₂, 0.1mg/ml BSA)于37℃下孵育。用33P标记SP1寡双链体(+)或(-)链的5'端。用33P标记单链寡核苷酸的s(+)SP1的5'端。

[0024] 图7显示了对Cas9-crRNA复合物的重新编程。(A) 用于Cas9-crRNA复合物重新编程的两个质粒中异源基因座的示意图。通过插入新的间隔子序列(SN) (5'-CC ACC CAG CAA AAT TCG GTT TTC TGG CTG-3' (SEQ ID NO:16))并如(1)中描述的方法失活Cas9基因,由pCas9(-)质粒构建pCas(-)SPN(见图2A)。(B) 质粒DNA裂解产物的琼脂糖凝胶分析。将pSP1和pSP1+SPN(具有插入的新原型间隔子且AatII位点上有PAM的pSP1质粒)以2.5nM的浓度与2nM的Cas9-crRNA复合物在反应缓冲液中(10mM Tris-HCl pH=7.5, 10mM NaCl, 10mM MgCl₂, 0.1mg/ml BSA)于37℃下孵育不同的时间段,并在琼脂糖凝胶中分析反应产物。SC-超螺旋质粒DNA, OC-在一条DNA链上有切口的开环DNA, FLL-全长线性DNA, 在两条链上都有剪口。(C) 寡双链体SP1裂解情况。将2.5nM的Cas9-crRNA复合物和1nM以33P标记(+)或(-)链的5'端的SPN寡双链体(表S2)在反应缓冲液中(10mM Tris-HCl pH=7.5, 10mM NaCl, 10mM MgCl₂, 0.1mg/ml BSA)于37℃下孵育。M1-18nt长的标记物泳道。M1和M2含有化学合成的5'端以33P标记的18nt和37nt的寡脱氧核糖核苷酸,其分别对应于(-)和(+)DNA链的裂解产物。(D) SPN寡双链底物及裂解产物的示意图。SPN寡双链含有新的原型间隔子(红色字体)、PAM(蓝色字体)。以箭头标记出裂解位置。图7(D)公开了SEQ ID NO:39。

[0025] 图8显示了间隔子长度对CRISPR编码的免疫的影响。(A) 经转化的质粒中插入的精简形式的原型间隔子的示意图。图8(A)以出现顺序分别公开了SEQ ID NO 7和61-66。(B) 原型间隔子长度对质粒转化效率的效应。转化效率以每纳克质粒DNA的cfu(平均值±SD)表示。(C) .用于体外裂解和结合实验的寡双链体的示意图。图8(C)以出现顺序分别公开了SEQ

ID NO 31和38。(D) 27bp寡双链体(全长原型间隔子SP1,实心圈)和20bp寡双链体(截短的原型间隔子SP1-20,方块)被Cas9-crRNA复合物裂解的时间过程。(E) Cas9-crRNA复合物结合SP1和SP1-20寡双链体的电泳迁移率变动分析。

[0026] 图9显示PAM是Cas9-crRNA复合物进行体外DNA结合和裂解所需要的。(A) 对质粒DNA裂解产物进行琼脂糖凝胶分析。将三种不同的质粒:PAM+原型间隔子+(含有原型间隔子和PAM二者的pSP1质粒),PAM-原型间隔子-(含有多个PAM但没有原型间隔子的pUC18质粒)以及PAM-原型间隔子+(含有原型间隔子但不含PAM的pSP1-p Δ (Sapranauskas等人, 2011.Nucleic Acids Res 39:9275-82))以2.5nM的浓度与2nM的Cas9-crRNA复合物在反应缓冲液(10mM Tris-HCl pH=7.5,10mM NaCl,10mM MgCl₂,0.1mg/ml BSA)于37℃下中孵育不同的时间段并在琼脂糖凝胶中分析反应产物。SC-超螺旋质粒DNA,OC-在一条DNA链上有切口的开环DNA,FLL-全长线性DNA,在两条链上有剪口。(B) 单链和双链寡脱氧核苷酸的(+)链水解的时间过程。反应物含有2nM的Cas9-crRNA和1nM的寡脱氧核糖核苷酸,在反应缓冲液(A部分)于37℃下中进行反应。用SP1(实心圈)和SPI-p Δ (空心方块)寡双链体作为dsDNA.s(+)SP1(空心三角)和s(+)SPI-p Δ (实心方块)用作ssDNA。(C)和(D)通过Cas9-crRNA复合物的dsDNA和ssDNA(+)链的结合。反应物含有0.5nM的33P标记的ssDNA或dsDNA寡核苷酸以及如各泳道上标明的浓度的蛋白质。室温15min后,对样品进行2h的PAGE实验并如“材料和方法”中所述进行分析。

[0027] 图10显示了对Cas9-crRNA复合物进行的RNA结合及裂解分析。(A) 分别针对Cas9-crRNA复合物结合含有原型间隔子-1、PAM以及两侧均有一段24nt的邻侧序列的84nt RNA片段进行电泳迁移率变动分析(EMSA)。左图:RNA(-)链;中图:RNA(+)链;右图:双链RNA。用于分析的RNA片段是通过体外转录(TranscriptAid™ T7 High Yield Transcription Kit, Fermentas)从RNA编码序列前端插入有T7启动子的PCR片段生成的。编码(+)、(-)RNA链的PCR片段获自pSP1质粒(1),分别使用了下列引物:5'taatacgactcactataGggtaccgagctcgattg 3'(SEQ ID NO:17)/5'GGGAAACAGCTATGACCATGATTACGAATTC-3'(SEQ ID NO:18)及5'gggtaccgagctcgattgaaattcTAAACG3'(SEQ ID NO:19)/5'taatacgactcactataGggAAACAGCTATGACCATGATTACG 3'(SEQ ID NO:20)(下划线表示T7 RNA聚合酶启动子,黑体表示转录起点)。反应物含有1nM的33P标记的RNA片段以及如各泳道上标明的浓度的蛋白质。室温15min后,对样品进行2h的PAGE实验并如“材料和方法”中所述进行分析。(B) RNA裂解测定。在1nM的(+)和(-)RNA链(左侧)或在(+)和(-)链上有标记的双链RNA的存在下在反应缓冲液(10mM Tris-HCl pH=7.5,10mM NaCl,10mM MgCl₂,0.1mg/ml BSA)中于37℃下对2.5nM的Cas9-crRNA复合物进行孵育。在变性PAA凝胶中对反应产物进行分析。

[0028] 图11显示了Cas9的RuvC和HNH活性位点基序有助于对相反的DNA链的裂解。(A) Cas9蛋白质内保守活性位点基序的定位。标明了鉴定为对Cas9体内活性至关重要的氨基酸残基(Sapranauskas等人,2011.Nucleic Acids Res 39:9275-82)。(B) Cas9和突变蛋白质对pSP1质粒裂解后进行琼脂糖凝胶分析。如“材料和方法”所述进行反应。(C) D31A突变体的链偏好。如图2A和“方法和材料”中所述进行反应。D31突变体仅仅裂解SP1寡双链体的(+)链。图11(C)以出现顺序分别公开了SEQ ID NO 31和67。(D) N891A突变体的链偏好。N891突变体仅仅裂解SP1寡双链的(-)链。裂解位点以箭头指示。图11(D)以出现顺序分别公开了SEQ ID NO 31和68。

[0029] 图12显示了Cas9活性位点突变体-crRNA复合物的特性。(A) 用Cas9突变体D31A (RuvC样活性位点基序) 获得的反应产物的直接测序情况。图12 (A) 以出现顺序分别公开了SEQ ID NO 58、59、58和58。(B) 用Cas9突变体N891 A (HNH样活性位点基序) 获得的反应产物的直接测序情况。图12 (B) 以出现顺序分别公开了SEQ ID NO 58、58、58和60。(C) wt Cas9-crRNA及活性位点突变体复合物对SP1寡双链体结合的情况。(D) Cas9-crRNA突变体复合物对(+)SP1链的裂解情况。

[0030] 图13显示了wt Cas9-Chis蛋白质的分子质量。在室温下进行凝胶过滤实验,使用Superdex 200 10/300GL柱子(GE healthcare),其预先用含有500mM氯化钠的10mM磷酸钠缓冲液(pH 7.4)平衡。通过在一套已知Mw(黑色圆圈)的蛋白质(Bio-Rad凝胶过滤标准品)获得的标准曲线中获得内插值得到Cas9(黑色三角)的表观Mw值。

[0031] 图14显示了Cas9-crRNA复合物进行的crRNA导向的DNA裂解的图示流程及机制。图上部示意性显示了Cas9的结构域组织形式。Cas9-crRNA复合物结合至含有PAM的dsDNA。crRNA结合至互补的(+)链,导致DNA链的分开并形成R-环。在这个三元复合物中,Cas9的RuvC活性位点位于未配对的(-)链的易断的磷酸基团上,而HNH活性位点位于结合至crRNA的DNA(+)链上易断的磷酸基团上。两个活性位点的协同作用导致在距离PAM 4nt距离处发生双链断裂,生成平末端DNA。图14以出现顺序分别公开了SEQ ID NO:31和69。

[0032] 图15显示了Cas9-crRNA及裂解产物的非变性电泳情况。在0.5nM的SP1寡双链体存在时,将各泳道上标明的浓度的蛋白质在反应缓冲液(10mM Tris-HCl pH=7.5,10mM NaCl,10mM MgCl₂,0.1mg/ml BSA)中于37℃下孵育30min。将样品与上样染料溶液(0.01%溴酚蓝和75mM EDTA,溶于50%v/v的甘油)混合并通过非变性PAGE进行分析。凝胶泳道记作M——代表裂解反应产物的融解形式。凝胶各边的卡通图阐明蛋白质-DNA复合物及对应每条带的DNA,而凝胶下面的卡通图阐明了反应后主要的底物形式。

[0033] 图16显示了Cas9-crRNA复合物对质粒DNA的裂解情况。(A) pSP1和pUC18质粒DNA的裂解。将Cas9-crRNA复合物与pSP1和pUC18质粒在实施例1中提供的反应缓冲液中孵育。pSP1质粒含有原型间隔子1序列,邻侧是5'-GGNG-3' PAM序列。原型间隔子1序列不存在于pUC18中。在琼脂糖凝胶中对反应产物进行分析。在这些条件下,pSP1被转化为线性形式,而缺少原型间隔子1序列的pUC18质粒对裂解产生抗性。(B) 在缺少一种组分的情况下pSP1的裂解反应。在缺乏一种组分(分别是缺少Cas9、crRNA或tracrRNA)的反应混合物中,pSP1质粒没有被裂解。SC-超螺旋质粒DNA,OC-在一条链上有切口的开环DNA,FLL-全长线性DNA。

[0034] 图17显示了Cas9-crRNA复合物对DNA寡双链体的裂解。寡双链体中与crRNA互补的链被记做(+)链,另一条链记做(-)链。为监测裂解反应,将P33标记在寡双链的(+)或(-)链的5'末端。M1和M2为对应于(-)链的37nt以及(+)链的18nt的合成寡核苷酸标记物,用于确定裂解产物大小并为裂解位置定位。Cas9蛋白质在原型间隔子内第37个核苷酸之后,PAM上游4nt(5'-GGNG-3')处裂解寡双链体的全部两条链,留下平末端。将非特异性底物(K1和K2)与Cas9-crRNA复合物孵育30min后,两条链都没有被裂解。图17公开了SEQ ID NO:31。

[0035] 图18显示了无RNaseIII的情况下组装的Cas9-crRNA复合物对质粒DNA的裂解情况。将Cas9-crRNA复合物与pSP1质粒孵育并在琼脂糖凝胶中分析反应产物。pSP1质粒在无crRNA组装的复合物存在下对裂解有抗性(左图)。在使用了合成的42nt crRNA(无RNaseIII)组装的复合物存在下,pSP1质粒被转化为线性形式(中图)。在使用CRISPR RNA转

录物(无RNaseIII)组装的复合物存在下,pSP1质粒被转化为线性和环状DNA形式的混合物(右图)。

[0036] 图19显示了Cas9-crRNA复合物对DNA寡双链体的裂解情况。寡双链体中与crRNA互补的链被记做(+)链,另一条链记做(-)链。为监测裂解反应,将P33标记在寡双链的(+)或(-)链的5'末端。M1和M2为对应于(-)链的37nt以及(+)链的18nt的合成寡核苷酸标记物,用于确定裂解产物大小并为裂解位置定位。Cas9蛋白质在原型间隔子内裂解寡双链的全部两条链,位于从5'端起第37个核苷酸之后,PAM上游4nt(5'-GGNG-3')处,留下平末端。将非特异性底物(K1和K2)与Cas9-crRNA复合物孵育30min后,两条链都没有被裂解。图19公开了SEQ ID NO:31。

[0037] 图20显示了(A)嗜热链球菌DGCC7710的CRISPR3/Cas系统的示意图。四种cas基因(cas9、cas1、cas2、csn2)位于CRISPR重复序列-间隔子阵列上游,由13个重复序列(R)及12个独特的间隔子(S1-S12)构成。II型CRISPR/Cas系统中crRNA成熟所需的tracrRNA(Deltcheva等人,2011.Nature 471,602-7)位于cas9基因上游且编码在相对此系统其他元件相反的DNA链(以箭头表示)上。(B)用于新的间隔子插入CRISPR区域和CRISPR RNA合成的通路。编码所需间隔子序列并含有SapI和Eco31I限制性兼容末端的合成寡双链体被插入到两个重复片段之间。使用PCR技术对CRISPR区域进行扩增。通过体外转录获得编码CRISPR RNA的间隔子。(C)Cas9-RNA复合物的体外组装。CRISPR RNA和tracrRNA转录物被组装成为双链体。首先将Cas9蛋白质与RNA双链进行预孵育,随后与RNaseII孵育,生成具有催化能力的Cas9-RNA复合物。

[0038] 图21显示了A.pUC18质粒的示意图。SapI和AatII限制性位点之间距离为775bp,而两个间隔子之间的距离为612bp。B.经过重新编程的Cas9-crRNA复合物对pUC18质粒的裂解情况。“1”-pUC18质粒;“2”-以AatII裂解的pUC18;“3”-以含有crRNA匹配原型间隔子1的复合物裂解的pUC18;“4”-以SapI裂解的pUC18;“5”-以含有crRNA匹配原型间隔子2的复合物裂解的pUC18;“6”-以AatII和SapI裂解的pUC18;“7”-以第3、第5条中所用复合物的混合物裂解的pUC18。

[0039] 图22显示了以体外组装的Cas9-RNA复合物对基因组DNA的裂解情况。(A)对线性λDNA裂解产物进行琼脂糖凝胶分析。将噬菌体λDNA与Cas9-RNA复合物在反应缓冲液中孵育不同的时间段。Cas9-RNA复合物的靶向位点位于距离cos位点的8kb处。(B).为DNA杂交试验进行探针选择。用PstI酶处理将基因组DNA片段化。原型间隔子位于两个PstI位点之间。如果用Cas9-RNA复合物裂解基因组DNA,应当检测到466bp的片段。否则,探针则会与1499bp长的片段杂交。(C)对基因组DNA片段进行Southern印迹分析。C线-用PstI片段化的大肠杆菌基因组DNA。在片段化前,将Cas9-RNA-基因组DNA与Cas9-RNA复合物孵育。(D).Cas9-crRNA复合物对人基因组DNA的裂解情况。完整DNA和DNA片段的相对量通过qPCR进行估计。

[0040] 图23示意性地阐明了包含在报告子质粒(pMTC-DSR+eGFP)中的靶向序列。eGFP编码序列被一个来自GAPDH基因的内含子隔开。指出了5'和3'RFP编码序列。homol表示RFP基因中发生同源重组所需要的同源序列。A、B、C和D指出了Cas9-介导的裂解中四个不同的靶标位点。靶标A和B位于内含子中。靶标C和D位于eGFP的编码区域中。Cre表示Cre内切酶的靶标位点,位于内含子序列中。

[0041] 图24显示引入Cas9/RNA复合物后eGFP阳性细胞减少情况。用报告子质粒和含有靶

向eGFP序列A(内含子)、eGFP序列C(编码)或非特异性序列K的crRNA的Cas9/RNA复合物转染CHO-K1细胞。通过流式细胞术确定eGFP阳性细胞百分比。作为阴性对照,用未转染(NC)或仅以报告子质粒转染(DNA)或以报告子质粒与单独Cas9蛋白质共转染以及报告子质粒与Cas9非特异性crRNA复合物共转染(DNA+K)的细胞。

[0042] 图25显示了细胞的照片,其中RFP的出现表示存在同源重组(HR)引起的对Cas9/RNA-介导的双链断裂的修复。在与报告子质粒和靶向eGFP序列C的Cas9/RNA复合物共转染48小时后,通过荧光显微镜观察CHO-k1细胞中的eGFP and RFP表达情况。

[0043] 图26示意性地阐明了包含在报告子质粒(pMTC-DSR+eGFP)中的靶向序列。eGFP编码序列被来自基因组DNA的GAPDH内含子隔开。指出了RFP N-和C-编码序列。RFP基因中的同源序列为发生同源重组所必须。位于eGFP内含子内部的靶标E以黑体表示。

[0044] 图27为凝胶分析结果,显示使用了合成crRNA和tracrRNA的Cas9/RNA复合物的功能类似于使用合成crRNA及体外转录tracrRNA的Cas9/RNA复合物。在进行琼脂糖凝胶电泳后,可以观察到质粒的情况。C泳道:未剪切的质粒。1-3泳道:以Cas9+crRNA剪切的质粒,其中1:体外转录tracrRNA对照;2:未修饰合成tracrRNA(89nt);或3:未修饰合成tracrRNA(74nt)。

[0045] 图28A-E示意性地显示了报告子质粒(pMTC-DSR+eGFP)内含有的靶向序列及可能的加工/基因重排结果。

[0046] 图29在引入Cas9/RNA复合物后eGFP阳性细胞减少情况。用报告子质粒和含有靶向eGFP序列A(内含子)、eGFP序列C(编码)或非特异性序列K的crRNA的Cas9/RNA复合物转染CHO-K1细胞。通过流式细胞术确定eGFP阳性细胞百分比。作为阴性对照,未转染(NC)或仅以报告子质粒转染(DNA)或以报告子质粒与单独Cas9蛋白质共转染以及报告子质粒与Cas9非特异性crRNA复合物共转染(DNA+K)的细胞。

[0047] 图30示意性地显示了包含在报告子质粒(pMTC-DSR+eGFP)中的靶向序列。eGFP编码序列以黑色表示,由拷贝自基因组DNA的GAPDH内含子隔开。RFP N-和C-编码序列以灰色表示。RFP基因中的同源序列是发生同源重组所必须的。位于eGFP内含子内的靶标E以黑体表示。

[0048] 以下非限制性实例进一步描述了方法、组合物、用途和实施方式。

[0049] 阐述性实施方式详述

[0050] 实施例1

[0051] 在本实施例中,我们已经分离了嗜热链球菌CRISPR3/Cas系统的Cas9-crRNA复合物并表明其以PAM依赖的方式剪切具有与crRNA互补的核苷酸序列的合成寡脱氧核糖核苷酸和质粒DNA。进一步地,我们提供了实验证据,证明PAM是在双链DNA的情况下得到识别的,并且对于体外DNA结合和裂解至关重要。最终,我们表明Cas9的RuvC和HNH基序有助于对相反的DNA链进行裂解。综上所述,我们的数据表明Cas9-crRNA复合物起到RNA导向的核酸内切酶的作用,其使用RNA模块进行靶标位点的识别并采用蛋白质模块中的两个分开的活性位点进行DNA裂解。这些发现为把可编程的Cas9-crRNA复合物工程改造为通用的RNA导向的核酸内切酶奠定了基础。

[0052] 材料和方法

[0053] DNA操作。使用嗜热链球菌DGCC7710的基因组DNA作为PCR反应的模板以克隆cas9。为生成pASKIBA3-Cas9质粒用于表达C末端有Strep标签的Cas9蛋白质变体,以下列引物扩

增PCR片段:5'-ACGTCTCAAATGTTGTTTAATAAGTGTATAATAATTTTC-3' (SEQ ID NO:21) 和5'-ACGTCTCCGCGCTACCCTCTCCTAGTTTG-3' (SEQ ID NO:22), 并将片段经Esp31位点克隆至pASK-IBA3表达载体。为生成pBAD-Cas9质粒用于表达C末端有6xHis标签的Cas9蛋白质变体(如SEQ ID NO:23公开的"6xHis"), 以下列引物对扩增PCR片段:5'-ACGTCTCACATGACTAAGCCAT ACTCAATTGGAC-3' (SEQ ID NO:24) 和5'-ACTCGAGACCCTCTCCTAGTTTGCAA-3' (SEQ ID NO:25), 并将片段经NcoI和XhoI位点克隆入pBAD24-Chis表达载体。对pASKIBA3-Cas9和pBAD-Cas9质粒中cas9基因进行全长测序的结果表明其与起始的cas9序列之间无差异。为获得具有单个间隔子1的质粒pCas9(-) SP1 (图1B) 和pCRISPR3-SP1 (图2A), 以下列引物对自pCRISPR3质粒扩增PCR片段:5'GACCACTTATTGAGGTAAATGAG 3' (SEQ ID NO:26) /5'CAAACCAGGATCCAAGCTAATACAGCAG-3' (SEQ ID NO:27) ((BamHI (GGATCC) 位点以下划线表示) 并将其分别克隆至pCas9(-) 和pCRISPR3质粒中 (Sapranauskas等人, 2011. *Nucleic Acids Res* 39:9275-82)。

[0054] Cas9蛋白质及Cas9-crRNA复合物的表达和纯化。使用了用于嗜热链球菌CRISPR4/Cas系统中Cas3蛋白质所描述的流程 (Sinkunas等人, 2011. *EMBO J* 30:1335-42) 对具有(His)6标签的Cas9蛋白质形式(如SEQ ID NO:23中公开的"(His)6")进行表达和纯化。为纯化Cas9-crRNA复合物, 在具有pCas9(-) SP1质粒的大肠杆菌RR1菌株中表达Cas9蛋白质的Strep标签形式(图1B)。LB培养液中补充有Ap (100 μ g/ml) 和Cm (10 μ g/ml)。用于Cas9-crRNA复合物分离的大肠杆菌细胞以两个步骤进行培养。首先, 将4ml的细胞培养物在37 $^{\circ}$ C下培养至OD600值~0.5, 然后通过添加0.2 μ g/ml的无水四环素(AHT) (Sigma) 诱导表达。4h后, 将预先经诱导的培养物的1/400接种至补充有Ap (100 μ g/ml), Cm (12 μ g/ml) 和AHT (0.2 μ g/ml) 的新鲜LB培养基中并在37 $^{\circ}$ C下培养过夜。通过超声破坏所收获的细胞, 通过离心移除细胞碎片。将上清液加样至1ml的StrepTrap HP柱子 (GE Healthcare) 并以2.5mM脱硫生物素进行洗脱。以1L大肠杆菌培养物运行一次可获得大约1.5 μ g Cas9蛋白质。含有Cas9的部分在+4 $^{\circ}$ C下储存数日。通过SDS-PAGE估测蛋白质制备物的均质性。通过对含有Strep-Tactin纯化的Cas9蛋白质以及已知量的加His标签的Cas9蛋白质的样品的SDS-PAGE凝胶进行灰度分析确定Cas9-crRNA复合物中的蛋白质浓度。假定Cas9是单体形式并以1:1的摩尔比结合于在复合物中的crRNA, 以Cas9蛋白质的浓度来表示Cas9-crRNA复合物的浓度。

[0055] Northern印迹分析。将Cas9-结合的RNA从Strep-Tactin纯化的Cas9 (其与pCas9(-) SP1质粒共表达) 上分离, 使用的是miRNeasy Mini kit (Qiagen)。进行Northern印迹, 使RNA在10%聚丙烯酰胺凝胶(含有溶于20mM MOPS/NaOH pH 8缓冲液的7M尿素) 中进行。将RNA通过半干杂交使用Trans-blot SD (Bio-Rad) 转移至SensiBlotTM Plus Nylon Membrane (Fermentas) 上。在60 $^{\circ}$ C下, 使用0.16M的1-乙基-3-(3-二甲氨基丙基)-碳二亚胺(EDC) (Pierce)/0.13M 1-甲基咪唑 (Sigma) pH 8将RNA与膜交联1h。将膜与含有1% SDS和0.1mg/ml变性鱼精DNA (Ambion) 的2 $\times$ SSC缓冲液在40 $^{\circ}$ C下预杂交1h。用32P-5'-标记的42nt抗-crRNA DNA寡核苷酸(含有20nt的间隔子1和22nt的重复序列) (5'-TCGAAACAACACAGCTCTAA AACTGTCCTCTTCTCTTTAGC-3' (SEQ ID NO:28) 对印迹进行12h的探针检测。用含有0.2% SDS的0.2 $\times$ SSC缓冲液洗涤印迹3 $\times$ 15min, 使用感光成像进行观察。42nt的合成寡核糖核苷酸(5'-CGCUAAAGAGGAAGAGGACAGUUUAGAGCUGUGUUGUUUCG-3' (SEQ ID NO:7)) 和84nt的DNA寡核苷酸。

[0056] 寡核苷酸底物。本研究中所使用的所有寡核苷酸底物都在表1中给出。寡脱氧核苷酸购自Metabion(Martinsried, Germany)。使用PNK (Fermentas) 和[γ -33P]ATP (Hartmann Analytic) 对寡核苷酸的5'-端进行放射性标记。通过将两条寡核苷酸与互补序列退火产生双链体(SP1, SP1- Δ p, SP2)。在与未标记链退火前, 向单个DNA链的5'端引入放射性标签。

[0057] 与寡核苷酸底物的反应。反应通常这样进行, 在37°C下将2nM的Cas9-crRNA复合物添加至溶于10mM Tris-HCl的1nM标记的寡核苷酸(37°C下pH 7.5), 10mM的NaCl, 0.1mg/ml的BSA和10mM的MgCl₂中。在确定的时间间隔移出等分试样并用加样染料(95% v/v甲酰胺, 0.01%溴酚蓝, 25mM的EDTA, pH 9.0) 淬灭, 随后在20%聚丙烯酰胺凝胶中进行变性凝胶电泳, 然后进行FLA-5100感光成像仪(Fujilm) 检测。

[0058] 与质粒底物的反应。在37°C下在用于寡核苷酸底物反应的缓冲液中, 进行pUC18质粒及其衍生物的反应(Sapranauskas等人, 2011. Nucleic Acids Res 39:9275-82)。反应混合物通常含有2.5nM超螺旋质粒及2nM的Cas9-crRNA复合物。通过将蛋白质加入其他组分的混合物而起始反应。在确定的时间间隔移出等分试样并用苯酚/氯仿进行淬灭。将水相与加样染料溶液(0.01%溴酚蓝和75mM EDTA, 溶于50% v/v甘油) 混合, 通过琼脂糖凝胶电泳进行分析。

[0059] 质粒裂解位置确定。为对质粒底物进行完全裂解, 将8nM的Cas9-crRNA复合物与2.5nM超螺旋质粒在反应缓冲液中于37°C孵育10min。使用GeneJET PCR Purification Kit (Fermentas) 对反应产物进行纯化和浓缩。使用以下引物对由Cas9线性化及缺口化的质粒的间隔子1周围区域进行直接测序: 5'-ccgcatcaggcgccattcgcc-3' (SEQ ID NO:29) (对(+)链测序) 和5'-gcgaggaagcggaagagcgccc-3' (SEQ ID NO:30) (对(-)链测序)。

[0060] 结合测定。将量不断增加的蛋白质-crRNA复合物与0.5nM的33P-标记的双链和单链DNA底物(表1) 在结合缓冲液(40mM Tris-乙酸, 25°C下pH 8.3, 0.1的EDTA, 0.1mg/ml的BSA, 10% v/v甘油) 中于室温下孵育15min。游离DNA和蛋白质-DNA复合物在8%的非变性聚丙烯酰胺凝胶(丙烯酰胺//N,N'-甲叉双丙烯酰胺比例为29:1) 中分离, 使用的电泳缓冲液为补充有0.1mM EDTA的40mM Tris-乙酸(pH 8.3)。在室温下以6V/cm进行3h的电泳。

[0061] 诱变。通过先前描述的方法进行定点诱变获得突变体D31A和N891A(Tamulaitis等人, 2007. Nucleic Acids Res 35:4792-9)。对每个突变体的整个基因进行测序的结果确定仅仅引入了所设计的突变。

[0062] 表1. 寡核苷酸底物。原型间隔子序列以下划线表示, PAM以黑体示出。

寡核苷酸	序列	说明
SP1 (SEQ ID NO:31)	5' - GCTCGAATTGAAATTCCTAAACGCTAAAGAGGAAGAGGACATGGTGAATTCGTAAT -3' 3' - CGAGCTTAACCTTTAAGATTTGCGATTCTCCTTCTCCTGTACCACTTAAGCATT -5'	含有原型间隔子 1 和 PAM 的 55bp 的寡双链体
SP1-pΔ (SEQ ID NO:32)	5' - GCTCGAATTGAAATTCCTAAACGCTAAAGAGGAAGAGGACAAATTCGTAAT -3' 3' - CGAGCTTAACCTTTAAGATTTGCGATTCTCCTTCTCCTGTTTAAGCATT -5'	含有原型间隔子 2 的 50bp 的寡双链体
SP2 (SEQ ID NO:33)	5' - GCTCGAATTGTACTGCTGTATTAGCTTGGTTGTTGTTGTGGTGAATTCGTAAT -3' 3' - CGAGCTTAACATGACGACATAATCGAACCAACAACCAACACCACTTAAGCATT -5'	含有原型间隔子 2 和 PAM (没有原型间隔子 1 的寡双链体) 的 55bp 的寡双链体
[0063] S (+) SP1 (SEQ ID NO:34)	5' - ATTACGAATTCACCATGTCTCTTCTCTTTAGCGTTTAGAATTTCAATTCGAGC -3'	55bp 的 ssDNA 寡核苷酸底物, SP1 寡双链体的 (+) 链
S (+) SP1-pΔ (SEQ ID NO:35)	5' - ATTACGAATTGTCCTCTTCTCTTTAGCGTTTAGAATTTCAATTCGAGC -3'	50bp 的 ssDNA 寡核苷酸底物, SP1-pΔ 寡双链体的 (+) 链
S (+) SP2 (SEQ ID NO:36)	5' - ATTACGAATTCACCAACAACAACAACAAGCTAATACAGCAGTACAATTCGAGC -3'	55bp 的 ssDNA 寡核苷酸底物, SP2 寡双链体的 (+) 链
S (-) SP1 (SEQ ID NO:37)	5' - GCTCGAATTGAAATTCCTAAACGCTAAAGAGGAAGAGGACATGGTGAATTCGTAAT -3'	55bp 的 ssDNA 寡核苷酸底物, SP2 寡双链体的 (-) 链
[0064] SP1-20 (SEQ ID NO:38)	5' - GCTCGAATTGCGCTAAAGAGGAAGAGGACATGGTGAATTCGTAAT -3' 3' - CGAGCTTAACGCGATTCTCCTTCTCCTGTACCACTTAAGCATT -5'	含有 20nt 的原型间隔子 1 和 PAM 的 45nt 的寡双链体
SPN (SEQ ID NO:39)	5' - GCTCGAATTGCCACCCAGCAAAATTCGGTTTTCTGGCTGATGGTGAATTCGTAAT -3' 3' - CGAGCTTAACGGTGGGTCGTTTTAAGCCAAAAGACCGACTACCACTTAAGCATT -5'	含有原型间隔子 N 和 PAM 的 55nt 的寡双链体

[0065] 结果

[0066] Cas9-crRNA复合物的表达和纯化。将来自嗜热链球菌DGCC7710菌株CRISR3系统的cas9基因克隆入pASK-IBA3载体以产生编码含有C-末端的Strep(II)-标签(图1B)的Cas9蛋白质融合物的构建体。起始,我们尝试从大肠杆菌RR1菌株(其在pASK-IBA3载体上表达Cas9蛋白质并在pCas9(-)质粒上表达其他Cas蛋白质(除Cas9))中纯化Cas9-crRNA复合物(Sapranauskas等人,2011)。pCas9(-)也含有完整的CRISPR3阵列,由12个间隔子-重复序列的单元构成(图2A)。为使所有靶标基因同时得到转录,我们以两步法进行cas9基因的表达。

首先,我们把Cas9的表达引入少量体积的大肠杆菌培养物,4h后将预先经诱导的培养物中取出等分试样转移至已经含有诱导物的较大体积的新鲜LB培养基并孵育过夜。使用Strep-Tactin Sepharose从初始的细胞提取物中纯化Cas9蛋白质复合物。我们设法分离了少量对含有原型间隔子1和PAM的寡双链体SP1只显示极微核酸裂解活性的Cas9-crRNA复合物。我们假定低的裂解活性可能是由于12个间隔子-重复序列单元转录产生的Cas9-crRNA复合物内在的异质性。如果所有间隔子-重复序列单元都一致地转录为成熟crRNA,则含有针对间隔子-1的crRNA的Cas9复合物的浓度为总Cas9-crRNA浓度的1/12。Cas9-crRNA制备物对含有原型间隔子-2和PAM的SP2寡双链体的裂解活性符合Cas9-crRNA复合物的异质性(图2B)。为增加特异性Cas9-crRNA复合物的产出,我们对在CRISPR阵列含有单个R-间隔子1-R单元的pCas9(-) SP1质粒进行工程改造(图1B)。质粒转化干扰测定确定携带单一间隔子1的CRISPR3/Cas系统阻碍大肠杆菌中pSP1转化,效率等同于携带完整CRISPR区域的CRISPR3/Cas系统(图3B)。我们依照上文描述的步骤分离了Cas9-crRNA复合物并对结合于Cas9蛋白质的crRNA进行分析。

[0067] Cas9蛋白质与crRNA共纯化。嗜热链球菌的CRISPR3/Cas系统属于CRISPR/Cas系统的IIA亚型(之前的Nmeni或CASS4)(Makarova等人,2011.Nat Rev Microbiol 9:467-77)。有研究显示化脓链球菌的IIA型CRISPR/Cas系统的转编码小RNA(tracrRNA)和细菌的RNaseIII都参与crRNA的生成(Deltcheva等人,2011.Nature 471:602-7)。化脓链球菌crRNA长度只有42nt,并且没有I和III型CRISPR系统的crRNA中保守的“5'-把手”(Hale等人,2009.Cell139:945-56;Jore等人,2011.Nat Struct Mol Biol 18:529-36)。根据northern印迹分析,嗜热链球菌LMD-9CRISPR3/Cas系统中生成了类似长度的crRNA(Makarova等人,2011.Nat Rev Microbiol 9:467-77),其几乎与DGCC7710菌株的CRISPR3/Cas系统相同(图4A和B)。我们假定分离自在异源大肠杆菌菌株中表达的Cas9-crRNA复合物的crRNA(图1)可能具有相同的长度(图4)。因此,为对提取自Strep-Tactin纯化的Cas9复合物的核酸进行探针检测,我们使用了42nt抗-crRNA的DNA寡核苷酸,其包含了与重复序列3'-端对应的22nt区域以及SP1片段5'-端的20nt区域。Cas9复合物中存在的核酸与抗-crRNA寡核苷酸杂交,对RNase的处理敏感,但对DNase不敏感(图1C)。提取的crRNA大小与对应嗜热链球菌DGCC7710菌株CRISPR3系统中推定的crRNA的42nt合成寡核苷酸相同(图3A、图4C)。综合而言,这些数据确认Cas9Strep标签蛋白质与来源于CRISPR3区域的42nt crRNA共纯化。

[0068] Cas9蛋白质在原型间隔子内部裂解双链DNA。为测定纯化的Cas9-crRNA复合物的体外活性,我们首先使用的SP1寡双链(表1)含有与CRISPR3阵列中间隔子SP1相同的原型间隔子序列,含有原型间隔子下游PAM序列5'-TGGTG-3'以及来自pSP1质粒的10nt邻侧序列(Sapranauskas等人,2011.Nucleic Acids Res 39:9275-82)(图5A)。将与crRNA互补的寡双链体记为(+)链,而相反的双链体链被称为(-)链。为检测裂解反应,以P33标记SP1寡双链体的(+)或(-)链的5'-末端。显示于图5B的数据表明Cas9-crRNA复合物在固定位置裂解寡双链体的两条链。使用合成寡核苷酸作为大小的标记物对裂解位置进行作图分析,表明Cas9-crRNA复合物在PAM上游4nt的原型间隔子内部切断SP1寡双链体的两条链(图5B),产生平末端。值得注意的是,将SP1寡双链体与缺少crRNA的Cas9蛋白质孵育2h后未观察到裂解现象(图6C)。

[0069] 为测试Cas9-crRNA复合物是否能定位原型间隔子并在体外长的DNA底物(模拟体内侵入的外源DNA)中剪切DNA,我们分析了携带原型间隔子1和PAM的pSP1质粒的裂解情况(Sapranauskas等人,2011.Nucleic Acids Res 39:9275-82)(图5C)。在Cas9-crRNA复合物存在下,pSP1质粒的超螺旋形式被转化为线性形式(图5D),而缺少原型间隔1的pUC18质粒则不被裂解。这意味着pSP1质粒的两条链都是在原型间隔子区域内被特异性裂解的。我们进行直接测序以确定Cas9-crRNA裂解后形成的线性DNA形式的末端。测序结果确认质粒DNA的裂解发生在距离PAM序列4nt处,这与SP1寡双链体裂解类似(图5D)。在体外实验中鉴定出的嗜热链球菌CRISPR3/Cas系统的裂解位置(图4)与嗜热链球菌CRISPR1/Cas系统体内裂解实验中确定的位置相同(Garneau等人,2010.Nature 468:67-71)。为确定Cas9-crRNA诱导的裂解在其他间隔子序列中是否也发生于同一位置,我们分析了分离自携带12个间隔子-重复序列单元的宿主的异质Cas9-crRNA复合物对携带原型间隔子-2和PAM序列的寡双链体SP2的裂解情况。我们发现这种异质Cas9-crRNA复合物在SP2寡双链体(+)链上剪切的位置与SP1寡双链体上的位置刚好相同。

[0070] Cas9-crRNA裂解特异性是通过crRNA序列导向的。为直接证明Cas9-crRNA复合物的特异性可通过改变该核糖蛋白质复合物中crRNA得到重新编程,我们将新的间隔子(SN)代替间隔子S1插入CRISPR区域,生成pCas(-)SN质粒,其只含有最少的CRISPR区域和tracrRNA编码序列(图7);将该质粒与pASKIBA-Cas9共表达并纯化Cas9-crRNA复合物。Cas9-crRNA复合物的裂解特异性通过使用pSP1+SPN和pSP1进行分析。使用Cas9-crRNA复合物对pSP1+SPN质粒(含有在CRISPR区域中与SN间隔子匹配的原型间隔子序列)进行线性化,缺少互补序列的pSP1质粒则保持完好(图7B)。为确定SPN间隔子序列内部的裂解位置,我们用含与间隔子SN互补的原型间隔子及PAM的SPN寡双链体进行实验(图7D)。寡双链体裂解测定确认(图7C和D)具有重新工程改造的特异性的Cas9-crRNA复合物在PAM上游4nt处SN原型间隔子内部裂解两条DNA链,这一点与其他Cas9-crRNA复合物是一致的。

[0071] 嗜热链球菌CRISPR3区域的间隔子长度是30nt。根据图1C中提供的数据,与Cas9蛋白质共纯化的成熟crRNA由42nt构成。这意味着只有20nt的crRNA与原型间隔子的(+)链互补。为评估原型间隔子的5'-端对于嗜热链球菌CRISPR3系统的质粒干扰是否重要,我们工程改造了质粒pSP1-27、pSP1-23、pSP1-19、pSP1-15、pSP1-11,使其具有5'端截短的原型间隔子1(原型间隔子长度分别为27bp、23bp、19bp、15bp、11bp),并对含有pCRISPR3的受体菌株的转化效率进行分析(图8B)。在原型间隔子1的5'端含有4或7bp截短的质粒对受体菌株干扰质粒转化的能力无影响。更短形式的原型间隔子(11、15、19bp)消除了受体菌株阻止质粒转化的能力。这些数据显示与成熟crRNA无互补性的原型间隔子5'端对CRISPR3/Cas功能并不重要。只含有原型间隔子-1的20nt的SP1-20寡双链体能有效地被Cas9-crRNA的裂解,这一点全面支持了体内实验的结果(图8D和E)。

[0072] PAM对DNA结合及Cas9-crRNA的裂解是需要的。携带原型间隔子但不含PAM(pSPI-p Δ)或含有多个PAM却没有原型间隔子(pUC18)的质粒对Cas9-crRNA的裂解有抗性(图8A)。因此,与体内数据相一致的是,PAM和原型间隔子都是Cas9-crRNA复合物裂解双链DNA所需要的(Sapranauskas等人,2011.Nucleic Acids Res 39:9275-82)。为查明PAM是否是在双链或单链DNA的条件下被识别,我们分析了以下寡脱氧核苷酸的Cas9-crRNA结合和裂解情况i) SP1(含有原型间隔子和PAM),ii) SPI- Δ p(仅含有原型间隔子),及iii) SP2(仅含有

PAM)。这些寡脱氧核苷酸的(+)链被用作单链DNA底物(相应地为s(+)SP1,s(+)SP1- Δ p,s(+)SP2)(表1)。

[0073] 与质粒裂解实验一致,仅含有原型间隔子而不含PAM的寡双链体不会被Cas9-crRNA剪切(图9B)。另一方面,不管是否含有PAM,单链形式的(+)链受到剪切的速率是类似的(图9B)。这些数据清楚地显示PAM仅仅对于双链DNA裂解是需要的,而对于单链DNA裂解则不然。

[0074] 为测试PAM对Cas9-crRNA复合物的DNA结合重要与否,进行了电泳迁移率变动实验。为避免裂解的发生,结合实验在无 Mg^{2+} 离子(裂解所需要)的情况下进行。Cas9-crRNA显示了对双链和单链寡核苷酸有不同的结合形式。在SP1寡双链体的情况下,1nM浓度下已经观察到低迁移率的复合物(图9C)。另一方面,无PAM(SP1- Δ p)或无原型间隔子(SP2)的寡双链体在相同实验条件下未观察到结合。此外,在无crRNA的Cas9蛋白质的情况下未观察到低迁移率复合物(图6A),这确认了crRNA对复合物形成的重要性。因此,结合实验综合而言清楚地显示Cas9蛋白质复合物在无PAM的情况下不能结合双链DNA,即使其含有与原型间隔子互补的crRNA。换言之,缺少PAM的双链DNA底物不会被裂解,因为PAM对Cas9-crRNA结合是需要的。

[0075] 另一方面,Cas9-crRNA对单链寡核苷酸((+)链)的结合具有相同的亲和力,与PAM的存在与否无关(图9D)。再一次地,对无原型间隔子的单链DNA寡核苷酸(图9D)或缺少crRNA的Cas9蛋白质(图6C)都未观察到有结合发生。这些数据综合表明Cas9-crRNA复合物仅仅在双链DNA中PAM的存在与否则有区别。

[0076] 由于一些III型CRISPR系统提供RNA干扰而不是DNA干扰,所以我们研究了Cas9-crRNA复合物对RNA的结合和裂解。Cas9-crRNA未特异性裂解具有原型间隔子和PAM的裂解单链RNA或双链RNA(图10B)。该发现再一次确认了DNA是嗜热链球菌CRISPR3/Cas系统的主要靶标。Cas9-crRNA复合物结合含有原型间隔子的互补RNA,但这种相互作用很可能在功能上并不重要,因为Cas9并不在原型间隔子内特异性裂解单链RNA。

[0077] Cas9蛋白质RuvC和HNH基序的诱变。质粒转化实验表明RuvC和HNH基序(图11A)对Cas9功能很重要(Sapranauskas等人,2011.Nucleic Acids Res 39:9275-82)。为测试这些基序是否参与靶标DNA的裂解,我们表达并纯化了D31A和N891A突变体,依照对wt Cas9描述的步骤进行。两种突变体都与crRNA共纯化,此处的crRNA与wt Cas9复合物中crRNA是相同的(图11C)。为测试突变蛋白质是否保留了裂解活性,我们监测了突变体对pSP1质粒的裂解情况。

[0078] Cas9-crRNA复合物。令人惊讶的是,两种突变体都产生了切口的DNA形式而不是在使用wt Cas9蛋白质时观察到的线性反应产物(图11B),这表明两个活性位点突变体都仅仅在原型间隔子内裂解质粒底物的一条DNA链。

[0079] 为确定突变体蛋白质是否显示了链偏好,我们分析了D31A和N891A突变体对SP1寡双链体的裂解情况。RuvC活性位点突变体(D31A)与wt Cas9-crRNA蛋白质在寡双链体(+)链上相同位置处剪切,而(-)链保持完好(图11C)。反之亦然,HNH活性位点突变体(N891A)仅仅裂解SP1寡双链体的(-)链,但不裂解(+)链(图11D)。综合而言,这些数据表明RuvC和HNH活性位点在相反的DNA链上起作用,生成双链断裂。为测试质粒DNA裂解过程中的是否保留这相同的裂解形式,我们对切口质粒的原型间隔子区域进行测序。最终序列结果确认RuvC活

性位点突变体仅仅剪切(+) DNA链,而HNH/McrA突变体仅剪切(-)链(图12A和B)。进一步地,我们发现RuvC突变体裂解单链DNA的(+)链,但对于HNH突变体没有检测到裂解(图12D)。

[0080] 为测试突变是否改变了突变蛋白质-crRNA复合物的DNA结合亲和力,使用电泳迁移率变动测定对DNA结合进行研究。两种突变体蛋白质-crRNA复合物与寡双链体SP1结合的亲和力都与野生型蛋白质相同(图12C)。这样,在Cas9的推测活性位点中的突变对Cas9-crRNA复合物的双链DNA结合特性没有显著效应。由于42nt crRNA存在于突变体蛋白质复合物中(图12C),所以我们推断突变体Cas9-crRNA复合物由于活性位点突变而失去了剪切靶标DNA链之一的能力。由于Cas9-His标签蛋白质在溶液中为单体(图13),所以Cas9蛋白质很可能作为单体起作用,并使用两个活性位点裂解相反的DNA链。一些限制性内切酶也采用了类似的策略(Armalyte等人,2005.J Biol Chem 280:41584-94)。

[0081] 讨论

[0082] 嗜热链球菌CRISPR3/Cas系统的Cas9-crRNA复合物是crRNA引导的核酸内切酶。这项工作表明嗜热链球菌CRISPR3/Cas系统的Cas9-crRNA复合物是crRNA指导的核酸内切酶,其在 Mg^{2+} 离子存在下在PAM序列下游4nt处原型间隔子内部剪切两条DNA链,产生平末端裂解产物。Cas9-crRNA复合物的序列特异性是由所述42nt crRNA控制的,其包括与靶标DNA中原型间隔子序列互补的~20nt的片段。在这一点上,嗜热链球菌CRISPR3/Cas系统的Cas9复合物中的成熟crRNA与化脓链球菌的crRNA类似,后者具有重复序列的3'-把手,但缺少间隔子序列的部分以及对应重复片段的5'-把手(Deltcheva等人,2011)。因此,嗜热链球菌CRISPR3/Cas系统的Cas9-crRNA复合物中存在的crRNA仅仅与远离PAM的原型间隔子序列的部分互补。这样无疑,对原型间隔子序列3'端截短10个核苷酸对合成寡双链体或质粒DNA的Cas9-crRNA裂解无影响(图8)。

[0083] Cas9-crRNA复合物的裂解机制位于Cas9蛋白质中,其为磷双酯键裂解提供两个活性位点。Cas9蛋白质的RuvC-和HNH-样活性位点位于不同结构域,并对各个DNA链独立起作用。RuvC-和HNH-基序中活性位点的残基中进行丙氨酸取代使Cas9-crRNA复合物转化为链特异性缺口核酸内切酶,与缺口酶类相似(Chan等人,2011.Nucleic Acids Res 39:1-18)。与体内研究一致的是,Cas9-crRNA复合物体外的功能活性绝对依赖于原型间隔子邻近基序NGGNG的存在(位于原型间隔子序列上游)。图3中体现的数据显示PAM是Cas9-crRNA结合于双链DNA所需要的。如果PAM序列在双链DNA中缺失,那么Cas9-crRNA复合物即使含有互补的原型间隔子序列也不会结合于这种DNA。在另一方面,如果PAM(或多个PAM)存在但缺少原型间隔子序列时,Cas9-crRNA不会显示DNA结合。因此,与体内数据相一致,PAM和原型间隔子序列是双链DNA结合及后续裂解的必须前提。与Cas9-crRNA对双链DNA的结合相反,PAM序列基序对于单链DNA结合没有影响:含有原型间隔子且具有PAM或不具有PAM序列的单链寡脱氧核苷酸的结合同样好,只是比双链DNA的亲和力稍低。在 Mg^{2+} 离子的存在下,Cas9使用其HNH活性位点剪切结合于所述crRNA的单链DNA。

[0084] II型系统中的DNA干扰机制。我们的结果建立了嗜热链球菌CRISPR3/Cas系统的Cas9-crRNA复合物裂解双链DNA的机制的简单模型(图14)。Cas9-crRNA复合物位于并以PAM依赖的过程结合双链DNA中的原型间隔子序列,使用的机制至今仍未确定。有可能双链DNA中的PAM起到链分离的起始位点(信号)的作用,并促进后续crRNA与互补的DNA(+)链的配对。还需要确定的是Cas9蛋白质模块还是Cas9结合的crRNA(例如使用保守重复序列中"3'-

把手”中的核苷酸)识别PAM序列。尽管缺少这些机制上的细节,我们的数据清楚地表明在双链DNA的情况下,PAM是由Cas9-crRNA识别的。一般认为Cas9-crRNA结合于dsDNA中的靶标序列会导致R环结构的形成,其中(-)链被取代,互补的(+)DNA链与crRNA配对。在Mg²⁺离子的存在下,磷双酯键裂解发生于两条链的PAM序列5'-上游4nt处,生成DNA平末端。RuvC-或HNH-基序突变体的DNA裂解分析表明Cas9蛋白质的RuvC-和HNH-样活性位点分别作用于(-)和(+)链。因此,在有催化能力的Cas9-crRNA复合物中,含有RuvC基序的D31A催化残基的N末端位于受取代的(-)DNA链上,而含有HNH基序的Cas9中心部分位于与crRNA配对的DNA(+)链的易断裂磷双酯键附近。在DNA裂解后,Cas9-crRNA仍然结合于反应产物(图15)。综合而言,本文呈递的数据提出了嗜热链球菌CRISPR3/Cas系统进行DNA干扰步骤的第一个分子机制。由于cas9是IIA型和IIB型系统的签名基因(Makarova等人,2011.Nat Rev Microbiol 9:467-77),所以本文提出的裂解机制很可能在其他IIA型和IIB型系统中是保守的。生物信息学鉴别出单独存在形式的Cas9样蛋白质,其不是CRISPR系统的一部分(Makarova等人,2011.Biol Direct 6:38)。根据本文提供的数据,如果这些蛋白质装配有小的crRNA分子(其可通过不同于CRISPR的途径生成),便可提供类似于Cas9的针对外源DNA的干扰。

[0085] 与其他的RNA干扰复合物的比较。本文提出的Cas9-crRNA复合物裂解双链DNA的机制与I-E型(即之前的大肠杆菌或CASS2)系统的机制差异显著(Jore等人,2011.Nat Struct Mol Biol 18:529-36)。在大肠杆菌系统中,crRNA和Cas蛋白质组装成大的核糖蛋白质复合物,称作Cascade,其通过增强CRISPR RNA和互补靶标序列之间的序列特异性杂交而促进靶标识别(Jore等人,2011.Nat Struct Mol Biol 18:529-36)。靶标识别依赖于PAM,并受到位于间隔子区域5'端的“种子”crRNA序列的控制(Semenova等人,2011.Proc Natl Acad Sci USA 108:10098-103)。然而,虽然Cascade-crRNA复合物自身就能够结合含有PAM和原型间隔子的双链DNA,但其需要附属的Cas3蛋白质进行DNA断裂。Cas3是一种单链DNA核酸酶和螺旋酶,能够裂解单链DNA产生多个剪切口(Sinkunas等人,2011.EMBO J 30:1335-42)。Cas3作用于合适的生物学底物(例如Cascade-crRNA在R环样复合物中结合于双链DNA)的机制细节还有待确立。然而,最近有研究表明单独的甲烷球菌Cas3能够在模拟R环的合成底物上剪切两条DNA链(Beloglazova等人,2011.EMBO J 30:616-27)。有人提出Cas3可能在Cascade-crRNA复合物的存在下遵循相似的机制进行DNA裂解。因此,现在的数据清楚地显示I-E型系统的干扰步骤的机制细节与CRISPR3系统在催化机制及机制复杂性上都有不同。

[0086] 在很多古细菌和一些细菌中存在的CRISPR系统III-B亚型中,Cas模块RAMP(Cmr)蛋白质和crRNA组装成效应子复合物,其靶向侵入的RNA(Hale等人,2009.Cell 139:945-56; Hale等人,2012.Mol Cell 45:292-302)。在激烈火球菌(Pyrococcus furiosus)中,RNA沉默复合物由六种Cmr1-6蛋白质构成,crRNA结合于靶标RNA并在相对psiRNA的3'端固定距离处将其剪切。裂解活性依赖于Mg²⁺离子,然而负责靶标RNA裂解的(这些)Cmr蛋白质仍然需要鉴别。硫磺矿硫化叶菌(Sulfolobus solfataricus)的效应子复合物由七个Cmr1-7蛋白质和crRNA构成,其以核酸内切酶反应形式在UA双核苷酸处剪切侵入的RNA(Zhang等人,2012.Mol Cell 45:303-13)。重要的是,两种Cmr-crRNA复合物都以不依赖PAM的方式进行RNA裂解。

[0087] 本文提供的数据显示CRISPR3系统的Cas9-crRNA复合物是迄今最简单的DNA干扰系统,由单个Cas9蛋白质结合于crRNA分子构成。Cas9-crRNA复合物的简单的模块架构提供

了多能的平台用于改造普适性的RNA导向的DNA核酸内切酶,该复合物中对DNA靶标的特异性由crRNA编码,而裂解机制由Cas蛋白质施行。

[0088] 实施例2.

[0089] 由4种组分体外组装成Cas9-crRNA复合物

[0090] 在此实施例中,我们证明了通过在体外混合4种单独的组分能组装成具有催化活性的Cas9-crRNA复合物:C末端具有(His)6标签的Cas9蛋白质变体(" (His)6"公开于SEQ ID NO:23中),tracrRNA转录物(SEQ ID NO:5),CRISPR RNA转录物(SEQ ID NO:8)及大肠杆菌RNaseIII (Abgene)。首先将Cas9蛋白质与tracrRNA和CRISPR RNA转录物进行预孵育,接着将其与RNaseIII孵育生成有催化能力的Cas9-crRNA复合物,这可以用于进行位点特异性的DNA裂解。

[0091] 更具体地,通过PCR生成的在RNA编码序列的邻近末端含有T7启动子的片段进行体外转录来产生复合物组装所需的RNA片段(TranscriptAid™ T7 High Yield Transcription Kit, Fermentas)。使用pCas9(-)SP1质粒作为模板,以下列引物对产生编码CRISPR RNA和tracrRNA的PCR生产DNA片段:5'-taatacgactcactataGggtagaaaagatatcctacgagg-3' (SEQ ID NO:40)/5'-CAACAACCAAGCTAATACAGCAG-3' (SEQ ID NO:41)和5'-aaaaacaccgaatcggtgccac-3' (SEQ ID NO:42)/5'-taatacgactcactataGggTAATAATAATTGTGGTTTGAACCATTC-3' (SEQ ID NO:43) (T7 RNA聚合酶启动子以下划线示出,转录起始以黑体表示)。150nt的CRISPR RNA转录物由102nt的重复序列-间隔子1-重复序列邻侧的引物退火所需的23nt上游和25nt下游区域组成。tracrRNA的105nt的转录物由38nt的片段构成,其与嗜热链球菌DCGG7710 CRISPR3重复序列片段部分地互补(反重复序列),邻侧是16nt的上游和51nt的下游区域。使用RNeasy MinElute Cleanup Kit (Qiagen) 纯化通过体外转录产生的RNA片段。

[0092] 为了体外组装有催化功能的Cas9-crRNA复合物,将具有(His)6标签的Cas9蛋白质变体(" (His)6"公开于SEQ ID NO:23中)与CRISPR RNA及tracrRNA转录物以1:0.5:1的摩尔比混合,并在缓冲液(含有10mM Tris-HCl (37℃下pH 7.5), 100mM NaCl)中于37℃下预孵育30min,随后加入RNaseIII (Ambion), MgCl₂和DTT,然后额外孵育30min。组装混合物中这些组分的终浓度如下:100nM具有(His)6标签的Cas9蛋白质变体(" (His)6"公开于SEQ ID NO:23中)、50nM的CRISPR RNA、100nM的tracrRNA、50nM的RNaseIII、10mM的MgCl₂、1mM的DTT。

[0093] 下文我们提供了实验证据证明体外组装的Cas9-crRNA复合物在crRNA序列的引导下在特异性位点裂解DNA生成平末端。在此方面,Cas9-crRNA复合物可作为限制性内切酶或大范围核酸酶的替代物使用,进行体外的位点特异性DNA裂解。复合物的序列特异性是由crRNA序列指导的,可对其进行工程改造以针对所需的DNA靶标。

[0094] 首先,以质粒pSP1和pUC18为底物测定体外组装Cas9-crRNA复合物的DNA裂解活性。pSP1质粒含有原型间隔子1序列,邻侧为5'-GGNG-3' PAM序列。pUC18中不存在原型间隔子1序列。在37℃下于10mM Tris HCl (37℃下pH 7.5)、50mM NaCl、0.05mg/ml BSA、0.5mM DTT和10mM MgCl₂中对质粒pUC18和pSP1进行反应(Sapranauskas等人,2011.Nucleic Acids Res 39:9275-82)。反应混合物通常含有3.0nM的超螺旋质粒DNA。将50μl体积的Cas9-crRNA复合物和质粒DNA(1:1v/v比例)在反应缓冲液中混合以起始反应。在确定的时间间隔移出等分试样并用苯酚/氯仿淬灭反应。将水相与上样染料溶液(0.01%溴酚蓝和

75mM EDTA,溶于50%v/v的甘油)混合,通过琼脂糖电泳分析反应产物(图16)。为检查通过Cas9-crRNA复合物预先裂解的pSP1质粒是否能够被重新连接,我们使用GeneJET gel extraction Kit(Fermentas)从琼脂糖凝胶中纯化线性的pSP1裂解产物并使用T4 DNA连接酶(Fermentas)进行重新连接。在使用连接混合物转化大肠杆菌细胞后,从所得的转化子中选取五个单独的克隆,纯化质粒DNA并对其进行测序,使用的引物如下:5'-ccgcatcaggcgccattcgcc-3'(SEQ ID NO:29)(对(+)链测序)及5'-gcgaggaagcggaagagcgccc-3'(SEQ ID NO:30)(对(-)链测序)。序列分析显示在被Cas9-crRNA复合物裂解并重新连接的基因座的pSP1质粒的DNA序列与未处理质粒的序列相同。在无T4 DNA连接酶存在下由裂解混合物转化大肠杆菌不产生转化子,这表明没有痕量的超螺旋质粒与线性反应产物共纯化。

[0095] 然后,在合成的55bp的寡脱氧核苷酸双链体SP1上测定体外组装的Cas9-crRNA复合物的裂解活性,该双链体含有与crRNA间隔子序列匹配的原型间隔子序列(图17)。反应条件与上文对于质粒DNA裂解所描述的条件同一,只是其中使用的是1nM的寡双链体。反应产物分析显示体外组装的Cas9-crRNA复合物在固定位置裂解寡双链体的两条链,所述位置在原型间隔子中距离5'末端的第37个核苷酸后,位于PAM序列5'-GGNG-3'上游4nt处,得到平末端(图17)。

[0096] 实施例3.

[0097] 从3种组分体外组装成Cas9-crRNA复合物

[0098] 在此实施例中,我们证明了通过在体外混合3种单独的组分能组装成活性Cas9-crRNA复合物:C末端具有(His)6标签的Cas9蛋白质变体("His)6"公开于SEQ ID NO:23中),如实施例1中提供的tracrRNA转录物(SEQ ID NO:5和SEQ ID NO:6)以及实施例1中提供的CRISPR RNA转录物(SEQ ID NO:8)或对应于嗜热链球菌DGCC7710菌株CRISPR3/Cas系统的推定crRNA的合成crRNA(SEQ ID NO:8)。合成的42nt的寡核苷酸由5'末端与CRISPR3间隔子1区域的相同的20nt以及3'端的22nt的重复序列构成。更具体地,tracrRNA和CRISPR RNA转录物是按照实施例1中所述方法而获得。为生成Cas9-crRNA复合物,将具有(His)6标签的Cas9蛋白("His)6"公开于SEQ ID NO:23中)与CRISPR RNA及tracrRNA转录物或42nt的合成crRNA以1:0.5:1的摩尔比混合,并在缓冲液(含有10mM Tris-HCl(37℃下pH7.5)、100mM NaCl)中于37℃下孵育1h。组装混合物中这些组分的终浓度如下:100nM具有(His)6标签的Cas9蛋白质("His)6"公开于SEQ ID NO:23中),50nM的CRISPR RNA或42nt合成crRNA,100nM的tracrRNA。

[0099] 下文我们提供了实验证据证明体外组装的Cas9-crRNA复合物在crRNA序列的引导下在特异性位点裂解DNA生成平末端。在此方面,Cas9-crRNA复合物可作为限制性内切酶或大范围核酸酶的替代物使用,进行体外的位点特异性DNA裂解。复合物的序列特异性是由crRNA序列控制的,所述crRNA序列被工程改造以针对所需的DNA靶标。

[0100] 首先,以质粒pSP1和pUC18为底物测定体外组装的Cas9-crRNA复合物的DNA裂解活性。pSP1质粒含有原型间隔子1序列,邻侧为5'-GGNG-3'PAM序列。pUC18中不存在原型间隔子1序列。在37℃下于10mM Tris HCl(pH 7.5at 37℃)、50mM NaCl、0.05mg/ml BSA、0.5mM DTT和10mM MgCl₂中对质粒底物进行反应(Sapranauskas等人,2011.Nucleic Acids Res39:9275-82)。反应混合物通常含有3.0nM的超螺旋质粒DNA。将50μl体积的Cas9-crRNA

复合物和质粒DNA (1:1v/v比例) 在反应缓冲液中混合以起始反应。在确定的时间间隔移出等分试样并用苯酚/氯仿淬灭反应。将水相与上样染料溶液 (0.01% 溴酚蓝和75mM EDTA, 溶于50% v/v的甘油) 混合, 通过琼脂糖电泳分析反应产物 (图18)。

[0101] 然后, 在合成的55bp的寡脱氧核苷酸双链体SP1上测定体外组装的Cas9-crRNA复合物的裂解活性, 该双链体含有与crRNA间隔子序列匹配的原型间隔子序列 (图19)。反应条件与上文对于质粒DNA裂解所描述的条件同一, 只是其中使用的是1nM的寡双链体。反应产物分析显示体外组装的Cas9-crRNA复合物在固定位置裂解寡双链体的两条链, 所述位置在原型间隔子中距离5'末端的第37个核苷酸后, 位于PAM序列5'-GGNG-3'上游4nt处, 得到平末端 (图19)。

[0102] 实施例4.

[0103] 用于对Cas9-crRNA复合物特异性进行重新编程的可交换的间隔子序列盒子

[0104] 在此实施例中, 我们描述了一种可交换的间隔子序列盒子, 其可产生携带针对任何所需的DNA靶标的crRNA, 后者可用于组装实施例1和2中所描述的Cas9-crRNA复合物 (图20B)。所述序列盒子携带单个重复序列-间隔子-重复序列单元, 其允许插入寡双链体, 该双链体携带生成所需crRNA所需的间隔子序列。为工程改造序列盒子, 我们首先构建序列盒子, 其含有先导序列、重复序列并在重复序列附近具有唯一的SapI识别位点, 随后是BamHI位点 (图20C)。为生成含有所述独特的所需的间隔子的CRISPR区域, 我们将含有独特间隔子序列及重复序列单元的合成寡双链体插入由限制性酶SapI和BamHI预先裂解的质粒中。我们使用这个序列盒子产生了crRNA转录物, 其含有与pUC18质粒中存在的原型间隔子N1和N2互补的核苷酸序列 (见下文)。

[0105] 作为主要证据, 我们使用可交换的间隔子序列盒子生成crRNA1和crRNA2, 二者分别在原型间隔子1和原型间隔子2处被工程改造以靶向pUC18质粒, 如实施例1所描述的方法将crRNA1和crRNA2整合入Cas9复合物, 并使用所述复合物裂解pUC18质粒。原型间隔子N1位于SapI限制性内切酶位点处, 而原型间隔子N2在AatII位点附近。SapI和AatII限制性位点之间的距离为775bp, 而推定的位于间隔子N1和N2的Cas9-crRNA复合物裂解位点距离为612bp (图21A)。从对应的可交换间隔子序列盒子质粒获得在近端含有T7启动子的crRNA1和crRNA2 PCR片段, 并用于通过体外转录产生CRISPR RNA转录物, 后者携带匹配间隔子N1或间隔子N2的序列。如实施例1中所描述的, 将有催化活性的Cas9复合物与crRNA1和crRNA2进行组装用于DNA裂解。体外组装的复合物含有crRNA1或crRNA2线性化的pUC18质粒 (图21B)。当将两种复合物与pUC18质粒一同孵育时, 获得了两种DNA片段 (2074和612bp) (图21B), 这表明质粒裂解发生于复合物中存在的crRNA分子靶向的位点。

[0106] 实施例5.

[0107] 使用Cas9-crRNA复合物的克隆步骤。

[0108] 在此实施例中, 我们证明Cas9-crRNA复合物可用于制备载体进行克隆步骤。首先, 我们证明了由Cas9-crRNA复合物获得的裂解产物可由DNA连接酶重新连接。我们从琼脂糖凝胶中纯化得到线性pSP1裂解产物并使用DNA连接酶进行重新连接。由所述连接混合物转化大肠杆菌细胞后, 从所得转化子中选取五个单独的克隆, 纯化质粒DNA并对其测序。序列分析显示在受到Cas9-RNA复合物裂解并重新连接的基因座中pSP1质粒的DNA序列与未处理质粒的序列相同。在无T4 DNA连接酶存在下由裂解混合物转化大肠杆菌不产生转化子, 这

表明没有痕量的超螺旋质粒与线性反应产物共纯化。该结果阐明Cas9裂解生成的DNA末端是T4 DNA连接酶的底物,因此一定在5'末端含有磷酸,在3'末端具有游离的OH基团(Lehman,1974)。

[0109] 接着我们分析了如实施例5中所描述的装载crRNA1和crRNA2的Cas9复合物对pUC18质粒的裂解(图21A)。首先,用一个复合物裂解pUC18,然后对其进行纯化和重新连接。在每次实验中对10个克隆进行测序,确认受到裂解和重新连接的质粒的序列与未处理质粒的序列相同(图21C)。该实验表示Cas9-crRNA复合物裂解和连接后未引入其他突变,所述Cas9-crRNA复合物可用于克隆实验。当两种复合物都与pUC18质粒孵育时,获得两种DNA片段(2074和612bp)(图21B),表明质粒裂解发生于复合物中存在的crRNA分子靶向的位点。为证明由Cas9-RNA复合物裂解的pUC18质粒适合于遗传改造,我们将PCR片段克隆入预先由含有crRNA1或crRNA2的Cas9复合物混合物裂解的pUC18载体,该PCR片段含有来自pACYC184质粒的启动子和四环素抗性基因。从富含四环素和氨苄霉素的培养基上选择克隆。对4个所选克隆进行测序,确认完整的PCR片段已经插入到所需的位置(图21C)。

[0110] 更具体地,将2 μ g pUC18与分开组装的含有不同crRNA的Cas9-RNA复合物(分别为250nM)在100 μ l体积反应液中(10mM的Tris-HCl(37 $^{\circ}$ C下pH 7.5)、100mM的NaCl、1mM的DTT和10mM的MgCl₂)于37 $^{\circ}$ C下孵育1小时。使用GeneJET gel extraction Kit(Thermo Fisher scientific)从琼脂糖凝胶中纯化所获得的载体片段并将其分成两个相等的部分。用FastAP碱性磷酸酶对其中一部分预先裂解的载体进行去磷酸化,而另一部分未经处理。通过PCR从pACYC184质粒中获得1282bp的插入片段,其含有启动子和四环素抗性基因。在使用GeneJET PCR Purification Kit(Thermo Fisher scientific)进行纯化后,将含有所述PCR片段的溶液分成两个部分。用T4多核苷酸激酶(Thermo Fisher scientific)对其中一部分进行磷酸化,而另一部分不受处理。使用T4 DNA连接酶(Thermo Fisher scientific)对未处理载体与未处理PCR片段进行连接,同时将去磷酸化载体与磷酸化片段进行连接。在补充有100 μ g/ml Ap及25 μ g/ml Tc的培养基上筛选克隆。

[0111] 实施例6.

[0112] Cas9 crRNA复合物对长DNA底物的裂解。

[0113] 在此实施例中,我们证明Cas9-crRNA可用于裂解长DNA分子中的靶标,包括噬菌体 λ ,大肠杆菌和人基因组DNA。

[0114] 更具体地,我们使用Cas9-RNA复合物裂解 λ 细菌噬菌体(48kb)、大肠杆菌BL-21菌株(4.6Mb)和人(3.2Gb)基因组DNA中的特异性位点。如实施例2和3中描述的,组装Cas9-crRNA复合物。我们使用了42nt长的合成crRNAs,150nt前体-crRNA和tracrRNA,其是从根据实施例4中的描述生成的模板进行体外转录合成的。

[0115] λ DNA裂解反应通过将 λ DNA(Thermo Fisher Scientific)与组装的Cas9-RNA复合物(1:1v/v比例)混合并于37 $^{\circ}$ C下孵育而起始裂解反应。最终反应混合物含有2 μ g的 λ DNA、50nM的Cas9-RNA复合物、10mM的Tris-HCl(pH 7.5at 37 $^{\circ}$ C)、100mM的NaCl、1mM的DTT和10mM的MgCl₂,反应液体积为100 μ l。在确定的时间间隔移出等分试样并用苯酚/氯仿进行淬灭。将水相与3 $\times$ 加样染料溶液(0.01%溴酚蓝和75mM EDTA,溶于50%v/v的甘油)混合,通过琼脂糖凝胶电泳和溴化乙锭染色反应分析产物。在琼脂糖凝胶中对线性 λ 噬菌体基因组DNA裂解产物的分析确认了 $\sim$ 40bp长的DNA在单一位点受到有效的裂解(图22A)。

[0116] 使用基因组DNA纯化试剂盒(Thermo Fisher Scientific)分离来自大肠杆菌BL21(DE3)菌株的DNA。为进行裂解测定,将大肠杆菌基因组DNA与组装的Cas9-RNA复合物(1:1v/v比例)相组合,并于37℃下孵育3小时。最终的反应混合物含有30μg基因组DNA、1μM的Cas9-RNA复合物、10mM的Tris-HCl(37℃下pH 7.5)、100mM的NaCl、1mM的DTT和10mM的MgCl₂,反应体积为300μl。在孵育后,加入30μl的FastDigest PstI(Thermo Fisher Scientific),将反应混合物在37℃下再孵育16小时。通过对反应混合物在蛋白酶K(0.5mg/ml;Thermo Fisher Scientific)和SDS(0.5%,w/v)于55℃下进行30min的加热随后在室温下与RNase A(0.25mg/ml;Thermo Fisher Scientific)孵育30min,来终止反应。在苯酚/氯仿提取后,用异丙醇沉淀DNA并将其溶于TE缓冲液中(10mM的Tris-HCl,pH 8.0和1mM的EDTA)。将10μg的DNA与3×上样染料溶液中(0.01%溴酚蓝和75mM EDTA,溶于50%v/v甘油)混合,并在1%的琼脂糖凝胶中进行电泳。

[0117] 为分析大肠杆菌基因组DNA的Cas9-crRNA裂解产物,我们设计了针对DNA片段的探针,其含有Cas9-RNA复合物的靶标(原型间隔子)(图22B),并进行了Southern印迹分析。Southern印迹分析如(Sambrook等人,1989.Molecular Cloning:A Laboratory Manual)中所述,并具有如下改动。通过半干转移将分级的DNA从琼脂糖凝胶中转移至SensiBlot Plus Nylon膜(Thermo Fisher Scientific)上。将DNA在0.4M NaOH饱和的纸巾上放置10min使DNA变性并将其固定于膜上,用2×SSC冲洗并在空气中干燥。用含有0.5%SDS和100μg/ml变性鲑鱼精DNA(Amresco)的6×SSC缓冲液在65℃下对膜进行预先杂交1h。杂交探针是使用基因组大肠杆菌BL21(DE3)DNA作为模板进行PCR而生成的,产生397bp的产物。用FastAP磷酸酶(Thermo Fisher Scientific)对5'端去磷酸化,将其与[γ-³²P]ATP(Hartmann Analytic)及T4 PNK(Thermo Fisher Scientific)孵育进行放射性标记。使用GeneJET PCR Purification Kit(Thermo Fisher Scientific)来纯化经过标记的探针,将探针加热到95℃孵育5min使之变性,并在冰上迅速冷却,然后直接加入预杂交溶液中。于65℃下将探针在膜上杂交16小时,在室温下以2×SSC、0.5%SDS洗涤两次,以2×SSC、0.1%SDS洗涤两次,在空气中干燥并用感光成像进行观察(FLA-5100;Fujifilm)。

[0118] 探针被设计成靶向含有用于Cas9-RNA复合物的靶标(原型间隔子)的靶标的DNA片段(图22B)。两个PstI靶标之间的距离为~1500bp,而原型间隔子和左侧PstI靶标之间的距离是466bp。Cas9复合物裂解后,我们只检测到466bp DNA片段(图22C),这意味着所有DNA靶标都在期望的位置被Cas9蛋白质裂解。这些数据清楚地表明Cas9蛋白质在很长且复杂的分子如病毒和细菌DNA中有效地寻找到靶标。

[0119] 为分析人基因组DNA的Cas9-crRNA裂解产物,我们使用提取自人脑的DNA。将人基因组DNA与组装的Cas9-crRNA复合物(1:1v/v比例)合并,于37℃下孵育30min。最终反应混合物含有1μg基因组DNA、100nM的Cas9、10mM的Tris-HCl(37℃下pH 7.5)、100mM的NaCl、1mM的DTT和10mM的MgCl₂,反应体积为100μl。组装Cas9-crRNA-HS1(SeqID#13)和Cas9-crRNA-HS2(SeqID#14)复合物以分别靶向RASGEF1 C或ARL15基因座。使用qPCR(图22D)分析裂解产物。在用Cas9-crRNA复合物处理后,完整的DNA靶标的量降低了超过25倍。对获自qPCR数据的结果分析表明Cas9-RNA复合物在期望的基因座有效裂解人基因组DNA。这些数据清楚地表明Cas9蛋白质在很长且复杂的分子如病毒、细菌和哺乳动物DNA中有效地寻找到靶标。

[0120] 实施例7.

[0121] 证明在哺乳动物细胞中转染Cas9/RNA复合物后发生报告子质粒的基因编辑的证据。

[0122] 构建报告子质粒以监测双链断裂的修复,这或者通过非同源末端连接(NHEJ),或者通过同源重组(HR)进行的。该质粒含有带内含子的GFP,eGFP序列邻侧为RFP的5'和3'序列以及同源位点(图23)。使用该报告子质粒时,eGFP荧光的减少表示发生了NHEJ,在这个过程中靶标C或D处的Cas9/RNA介导的双链断裂通过NHEJ不甚完美地被修复,这样破坏了eGFP编码序列。若靶向内含子靶标A和B并通过NHEJ进行修复则不会导致eGFP荧光减少,因为NHEJ诱导的这些突变通常删除或插入的片段<20bp,因此不会影响eGFP编码区域或剪接位点连接处。另一方面,RFP荧光的出现则表示发生了HR,其中Cas9/RNA介导的双链断裂通过HR得到修复,使用了所示的RFP的同原序列。

[0123] crRNA使用42个核苷酸的RNA分子进行靶向,如上文所述,该RNA分子具有作为重复序列的22个核苷酸和20个核苷酸(间隔子序列)用于特异性靶向。如上文描述的,靶标DNA需要嗜热链球菌基序或PAM,其为靶标中原型间隔子下游的“NGGNG”序列。没有将GFP“工程改造”为含有所述PAM基序;eGFP中的几个靶标序列天然具有PAM序列,crRNA被设计为靶向邻近的间隔子序列。在通过Cas9/RNA产生eGFP中的双链断裂后,RFP是发生同源重组的标记物。

[0124] 图28A显示了用于在真核细胞中体内进行Cas9蛋白质活性分析的报告子基因构建体。内含子序列含有三个cas9靶标位点(A、E、B);GFP基因含有两个(C、D)cas9靶标位点。RFP基因在Y196位裂开,RFP荧光被消除。图28B显示,体内内含子加工后观察到GFP荧光。图28C显示在任何前述核酸酶靶向位点发生的Cas9/crRNA复合物促进的dsDNA断裂可诱导HR,导致RFP基因的重新组装及RFP荧光的出现。图28D和E显示任何前述核酸酶靶向位点发生的Cas9/crRNA复合物促进的dsDNA断裂可诱导NHEJ。GFP基因序列的突变会导致GFP荧光的丢失或消除;内含子的突变可能对GFP荧光无影响,然而在不同的情况下可能产生成熟的RNA,其内含子序列受到不恰当的剪接,导致GFP荧光的丢失或消除。

[0125] 将纯化自大肠杆菌的嗜热链球菌Cas9蛋白质与体外转录的tracrRNA和合成的未修饰crRNA复合,靶向eGFP的(内含子)序列A或(编码)序列C。为进行转染,将Cas9/RNA复合物(靶向A或C)与转染反应剂TurboFECT孵育,并将报告子质粒DNA和TurboFECT在分开的管中孵育,然后将二者都加至CHO-K1细胞。通过流式细胞术确定eGFP阳性细胞的百分比。如图24和29所示,当单独用报告子质粒转染细胞或用单独Cas9蛋白质与报告子质粒转染细胞时,GFP阳性细胞的百分比为大约40-50%,表明总体的转染效率。然而,当将靶向eGFP的序列C的Cas9/RNA复合物与报告子质粒一同加入细胞时,eGFP阳性细胞的百分比减少至大约15%。仅仅在靶向eGFP的序列C的Cas9/RNA复合物的情况下观察到eGFP阳性细胞的降低,在靶向eGFP的序列A的Cas9/RNA复合物或非特异性RNA的情况下没有观察到eGFP阳性细胞的显著减少。这个结果表明靶向eGFP的序列C的Cas9/RNA通过引入双链断裂和NHEJ的不完美修正而导致eGFP的基因编辑,产生在eGFP编码序列中的删除。

[0126] 除了分析eGFP阳性细胞的百分比外,还通过荧光显微镜观察转染的细胞以监测RFP阳性细胞的存在情况,这表明发生了通过HR而不是NHEJ方式发生的Cas9介导的双链断裂的修复。如图25中所见的,在用报告子质粒和靶向eGFP的序列C的Cas9/RNA复合物转染后,一些细胞中可见RFP,表示通过HR进行了双链断裂的修复。

[0127] 实施例8.

[0128] 用合成的未修饰tracrRNA和crRNA产生的Cas9/RNA复合物在体外具有功能性。

[0129] 上文实施例7中描述的实验使用由纯化的Cas9、合成的crRNA以及体外转录的tracrRNA组成的Cas9/RNA复合物。为确定用完全合成的RNA组分(crRNA和tracrRNA)产生的Cas9/RNA复合物是否有功能性,合成了未修饰的嗜热链球菌tracrRNA(预期维持功能性的内源的89-mer和较短的74-mer形式)。所生成的未修饰合成crRNA针对的是位于上文描述的报告子质粒中eGFP的内含子中的靶标E(见图26和30),还生成了Cas9/RNA(crRNA和tracrRNA)复合物。为测试这些复合物,将上文使用的报告子质粒与所述复合物在体外孵育并通过凝胶电泳监测限制性酶切情况。

[0130] 如图27中所见,由完全合成的RNA组成的Cas9/RNA复合物在体外测定中与由合成crRNA和体外转录tracrRNA组成的Cas9/RNA复合物有同等的功能性。

[0131] 序列

[0132] SEQ ID NO:1

[0133] WT_Cas9_嗜热链球菌DGCC7710 CRISPR3-Cas菌株

[0134] 单字母序列:

mlfnkciisinldfsnkeckmtkpsysigldigntsvgwavitdnykvpskkmkvlngtskkyikknlglvllfdsgitaegrllkrtrrrytrrr
nrilylqeifstematlddaffqrlddsflvpddkrskypifgnlveekvyhdefptiyhlrkyldstkkadrlrvylalahmikyrghfliegef
nsknnidqknfdldtynaifsdlsenskqleeivkdkisklekkdrilklfpgeknsigifseflklivgnqadfrckfnldekaslhfskesy
dedletlgyigddysdvflkakklydaillsgflvtvdnateaplssamikrynehkedlallkeyimislktynevfkdtdkngyagvidgktn
qedfyvylknllaefegadyflekidredflrkqrtfdngsipyqihlqemraildkqakfypflaknkeriekiltfripyvvgplargnsdfaws
irkrnekitpwnfedvidkessacaafinrmtsfldlylpeekvlpkhslyetfnvyneltkvrfaesmrddyqfldskqkdivrlyfkdkrkvt
kdiieylhaiygydgiekgiekqfnsslstyhdllniindkeflddssneaieeihtltifedremikqrlskfenifdksvlkklsrrhytgwgl
saklingirdeksgntildyliddgisnrmfmqlihddalsfkkkiqkaqiigdedkgnikevvkslpgspaikkqilqsikivdelvkvmmggrk
pesivvemarenqytnqgksnsqqlrkrlekslkelgskilkenipaklskidnnalqndrlylylqngkdmtygdldldrlsnydidhiip
qafldnsidnkvlvssasnrsgsddfpslevvkkrrktfwyqlkksklisqrkfdnltaerggllpedkagfiqrqlvetrqitkhvarlldekfn
nkkdennravrtvkiitlkstlvsqfrkdfelykvreindfhahdaylnaviasallkypklepefygydypkynsfrersatekvyfysni
mnifkksisladgrvierplievneetgesvwnkesdlatvrrvlsypqvnvkvveeqnhgldrgkpglfnanlsskpkpsnlenlvgak
eyldpkkygygagisnsfavlvgtickgakkkitnvlfeqgisildrinyrkdklnfllekgykdielielpkyslfelsdgsrrmlasilstnnkr
geihkgngqiflsqkfvklllyhakrisntinenhrkyvenhkkefeelfyilefnyevgakkngklinsafqswqnhsidelcssfigptgser
kglfeltsrgsaadfeflgvkiprydytpssllkdatlihqsvtglyetridlaklgeg

[0136] 三字母序列:

[0137]

MetLeuPheAsnLysCysIleIleIleSerIleAsnLeuAspPheSerAsnLysGluLys
CysMetThrLysProTyrSerIleGlyLeuAspIleGlyThrAsnSerValGlyTrpAla
ValIleThrAspAsnTyrLysValProSerLysLysMetLysValLeuGlyAsnThrSer
LysLysTyrIleLysLysAsnLeuLeuGlyValLeuLeuPheAspSerGlyIleThrAla
GluGlyArgArgLeuLysArgThrAlaArgArgArgTyrThrArgArgArgAsnArgIle
LeuTyrLeuGlnGluIlePheSerThrGluMetAlaThrLeuAspAspAlaPhePheGln
ArgLeuAspAspSerPheLeuValProAspAspLysArgAspSerLysTyrProIlePhe
GlyAsnLeuValGluGluLysValTyrHisAspGluPheProThrIleTyrHisLeuArg
LysTyrLeuAlaAspSerThrLysLysAlaAspLeuArgLeuValTyrLeuAlaLeuAla
HisMetIleLysTyrArgGlyHisPheLeuIleGluGlyGluPheAsnSerLysAsnAsn
AspIleGlnLysAsnPheGlnAspPheLeuAspThrTyrAsnAlaIlePheGluSerAsp
LeuSerLeuGluAsnSerLysGlnLeuGluGluIleValLysAspLysIleSerLysLeu
GluLysLysAspArgIleLeuLysLeuPheProGlyGluLysAsnSerGlyIlePheSer
GluPheLeuLysLeuIleValGlyAsnGlnAlaAspPheArgLysCysPheAsnLeuAsp
GluLysAlaSerLeuHisPheSerLysGluSerTyrAspGluAspLeuGluThrLeuLeu
GlyTyrIleGlyAspAspTyrSerAspValPheLeuLysAlaLysLysLeuTyrAspAla
IleLeuLeuSerGlyPheLeuThrValThrAspAsnGluThrGluAlaProLeuSerSer
AlaMetIleLysArgTyrAsnGluHisLysGluAspLeuAlaLeuLeuLysGluTyrIle
ArgAsnIleSerLeuLysThrTyrAsnGluValPheLysAspAspThrLysAsnGlyTyr
AlaGlyTyrIleAspGlyLysThrAsnGlnGluAspPheTyrValTyrLeuLysAsnLeu
LeuAlaGluPheGluGlyAlaAspTyrPheLeuGluLysIleAspArgGluAspPheLeu
ArgLysGlnArgThrPheAspAsnGlySerIleProTyrGlnIleHisLeuGlnGluMet
ArgAlaIleLeuAspLysGlnAlaLysPheTyrProPheLeuAlaLysAsnLysGluArg
IleGluLysIleLeuThrPheArgIleProTyrTyrValGlyProLeuAlaArgGlyAsn
SerAspPheAlaTrpSerIleArgLysArgAsnGluLysIleThrProTrpAsnPheGlu
AspValIleAspLysGluSerSerAlaGluAlaPheIleAsnArgMetThrSerPheAsp
LeuTyrLeuProGluGluLysValLeuProLysHisSerLeuLeuTyrGluThrPheAsn
ValTyrAsnGluLeuThrLysValArgPheIleAlaGluSerMetArgAspTyrGlnPhe
LeuAspSerLysGlnLysLysAspIleValArgLeuTyrPheLysAspLysArgLysVal
ThrAspLysAspIleIleGluTyrLeuHisAlaIleTyrGlyTyrAspGlyIleGluLeu
LysGlyIleGluLysGlnPheAsnSerSerLeuSerThrTyrHisAspLeuLeuAsnIle
IleAsnAspLysGluPheLeuAspAspSerSerAsnGluAlaIleIleGluGluIleIle
HisThrLeuThrIlePheGluAspArgGluMetIleLysGlnArgLeuSerLysPheGlu
AsnIlePheAspLysSerValLeuLysLysLeuSerArgArgHisTyrThrGlyTrpGly
LysLeuSerAlaLysLeuIleAsnGlyIleArgAspGluLysSerGlyAsnThrIleLeu
AspTyrLeuIleAspAspGlyIleSerAsnArgAsnPheMetGlnLeuIleHisAspAsp
AlaLeuSerPheLysLysLysIleGlnLysAlaGlnIleIleGlyAspGluAspLysGly
AsnIleLysGluValValLysSerLeuProGlySerProAlaIleLysLysGlyIleLeu
GlnSerIleLysIleValAspGluLeuValLysValMetGlyGlyArgLysProGluSer
IleValValGluMetAlaArgGluAsnGlnTyrThrAsnGlnGlyLysSerAsnSerGln
GlnArgLeuLysArgLeuGluLysSerLeuLysGluLeuGlySerLysIleLeuLysGlu
AsnIleProAlaLysLeuSerLysIleAspAsnAsnAlaLeuGlnAsnAspArgLeuTyr
LeuTyrTyrLeuGlnAsnGlyLysAspMetTyrThrGlyAspAspLeuAspIleAspArg
LeuSerAsnTyrAspIleAspHisIleIleProGlnAlaPheLeuLysAspAsnSerIle
AspAsnLysValLeuValSerSerAlaSerAsnArgGlyLysSerAspAspPheProSer
LeuGluValValLysLysArgLysThrPheTrpTyrGlnLeuLeuLysSerLysLeuIle
SerGlnArgLysPheAspAsnLeuThrLysAlaGluArgGlyGlyLeuLeuProGluAsp
LysAlaGlyPheIleGlnArgGlnLeuValGluThrArgGlnIleThrLysHisValAla

- ArgLeuLeuAspGluLysPheAsnAsnLysLysAspGluAsnAsnArgAlaValArgThr
ValLysIleIleThrLeuLysSerThrLeuValSerGlnPheArgLysAspPheGluLeu
TyrLysValArgGluIleAsnAspPheHisHisAlaHisAspAlaTyrLeuAsnAlaVal
IleAlaSerAlaLeuLeuLysLysTyrProLysLeuGluProGluPheValTyrGlyAsp
TyrProLysTyrAsnSerPheArgGluArgLysSerAlaThrGluLysValTyrPheTyr
SerAsnIleMetAsnIlePheLysLysSerIleSerLeuAlaAspGlyArgValIleGlu
ArgProLeuIleGluValAsnGluGluThrGlyGluSerValTrpAsnLysGluSerAsp
LeuAlaThrValArgArgValLeuSerTyrProGlnValAsnValValLysLysValGlu
GluGlnAsnHisGlyLeuAspArgGlyLysProLysGlyLeuPheAsnAlaAsnLeuSer
SerLysProLysProAsnSerAsnGluAsnLeuValGlyAlaLysGluTyrLeuAspPro
LysLysTyrGlyGlyTyrAlaGlyIleSerAsnSerPheAlaValLeuValLysGlyThr
[0138] IleGluLysGlyAlaLysLysLysIleThrAsnValLeuGluPheGlnGlyIleSerIle
LeuAspArgIleAsnTyrArgLysAspLysLeuAsnPheLeuLeuGluLysGlyTyrLys
AspIleGluLeuIleIleGluLeuProLysTyrSerLeuPheGluLeuSerAspGlySer
ArgArgMetLeuAlaSerIleLeuSerThrAsnAsnLysArgGlyGluIleHisLysGly
AsnGlnIlePheLeuSerGlnLysPheValLysLeuLeuTyrHisAlaLysArgIleSer
AsnThrIleAsnGluAsnHisArgLysTyrValGluAsnHisLysLysGluPheGluGlu
LeuPheTyrTyrIleLeuGluPheAsnGluAsnTyrValGlyAlaLysLysAsnGlyLys
LeuLeuAsnSerAlaPheGlnSerTrpGlnAsnHisSerIleAspGluLeuCysSerSer
PheIleGlyProThrGlySerGluArgLysGlyLeuPheGluLeuThrSerArgGlySer
AlaAlaAspPheGluPheLeuGlyValLysIleProArgTyrArgAspTyrThrProSer
SerLeuLeuLysAspAlaThrLeuIleHisGlnSerValThrGlyLeuTyrGluThrArg
IleAspLeuAlaLysLeuGlyGluGly
- [0139] SEQ ID NO:2
- [0140] D31A突变体
- [0141] 单字母序列:
- mlfnkciisinldfsnkekcmtpysiglaigtmsvgavitdnykvpskkmkvlngtskkyikknlvgllfdsgitaegrllkrtarrytrrr
nrilylqefstematlddaaffqrlddsflvpddkrdskeypifgnlveekvyhdefptiyhlrkyladstkkadrlrvylalahmikyrghfliegef
nsknndiqknqfdldtynaifedsllenskqleeivkdisklekkdrilklfpgeknsigifseflklivgnqadfrkcfnldekasilhfskesy
dedletllgyigddysdvflkakklydaillsgfltvtdnetcaplssamikrynehkedlallkeyimisltktynevfkdtdkngyagyidgktn
qedfyvylknllaefegadyflekidredflrkqrtfdngsipyqihlqemraildkqakfypflaknkeriekiltfripyyvgplargnsdfaws
irkrnekitpwnfedvidkessaea finrmtsfdlylpeekvlpkhslyetfnvyneltkvrfiaesmrdyqfldskqkkdivrlyfkdkrkvt
kdiieylhaiygydgelkgiekqfnsslstyhdlniindkefiddssneaieeihtltifedremikqriskfenifdksvlklssrrhytgwgl
[0142] saklingirdeksntildyliddgisnrmfnqlihddalsfkkiqkaqiigdedkgnikvkslpgspaikkgilqsikivdelvkvmggrk
pesivvemarenqytnqgksnsqqrklrlekslkelgskilkenipaklskidnnalqndrlylylqngkdmtygdldidrlsnydidhiip
qafldksnidnkvlvssasnrgksddfpslevvkkrtfwyqlksklisqrkfdnlkaerggllpedkagfiqrqlvetrqitkhvarlldekfn
nkkdennravrtvkiitlksflvsqfrkdfelykvreindfhahdaylnaviasallkypklepefvygdykpynsfrerksatekvfyfysni
mnifkksisladgrvierplievneetgesvwnkesdlatvrrvlsypqvnvkvkveeqnhglrdgkpgklfnanlsskpkpnsnenlvgak
eyldpkkygyagisnsfavlvgtickgakkkitnvlefqgisildrinyrkdklnflekgkykdielielpkyslfelsdgsrrmlasilstnkr
geihkgngqiflsqkfvllyhakrisntinenhrkyvenhkkfeelfyyilefnyvvgakkngklinsafqswqnhsidelcssfigptgser
kglfeltsrgsaadfeiflgvkipryrdytpssllkdatlihqsvtglyetridlaklgeg
- [0143] 三字母序列:

[0144]

MetLeuPheAsnLysCysIleIleIleSerIleAsnLeuAspPheSerAsnLysGluLys
CysMetThrLysProTyrSerIleGlyLeuAlaIleGlyThrAsnSerValGlyTrpAla
ValIleThrAspAsnTyrLysValProSerLysLysMetLysValLeuGlyAsnThrSer
LysLysTyrIleLysLysAsnLeuLeuGlyValLeuLeuPheAspSerGlyIleThrAla
GluGlyArgArgLeuLysArgThrAlaArgArgArgTyrThrArgArgArgAsnArgIle
LeuTyrLeuGlnGluIlePheSerThrGluMetAlaThrLeuAspAspAlaPhePheGln
ArgLeuAspAspSerPheLeuValProAspAspLysArgAspSerLysTyrProIlePhe
GlyAsnLeuValGluGluLysValTyrHisAspGluPheProThrIleTyrHisLeuArg
LysTyrLeuAlaAspSerThrLysLysAlaAspLeuArgLeuValTyrLeuAlaLeuAla
HisMetIleLysTyrArgGlyHisPheLeuIleGluGlyGluPheAsnSerLysAsnAsn
AspIleGlnLysAsnPheGlnAspPheLeuAspThrTyrAsnAlaIlePheGluSerAsp
LeuSerLeuGluAsnSerLysGlnLeuGluGluIleValLysAspLysIleSerLysLeu
GluLysLysAspArgIleLeuLysLeuPheProGlyGluLysAsnSerGlyIlePheSer
GluPheLeuLysLeuIleValGlyAsnGlnAlaAspPheArgLysCysPheAsnLeuAsp
GluLysAlaSerLeuHisPheSerLysGluSerTyrAspGluAspLeuGluThrLeuLeu
GlyTyrIleGlyAspAspTyrSerAspValPheLeuLysAlaLysLysLeuTyrAspAla
IleLeuLeuSerGlyPheLeuThrValThrAspAsnGluThrGluAlaProLeuSerSer
AlaMetIleLysArgTyrAsnGluHisLysGluAspLeuAlaLeuLeuLysGluTyrIle
ArgAsnIleSerLeuLysThrTyrAsnGluValPheLysAspAspThrLysAsnGlyTyr
AlaGlyTyrIleAspGlyLysThrAsnGlnGluAspPheTyrValTyrLeuLysAsnLeu
LeuAlaGluPheGluGlyAlaAspTyrPheLeuGluLysIleAspArgGluAspPheLeu
ArgLysGlnArgThrPheAspAsnGlySerIleProTyrGlnIleHisLeuGlnGluMet
ArgAlaIleLeuAspLysGlnAlaLysPheTyrProPheLeuAlaLysAsnLysGluArg
IleGluLysIleLeuThrPheArgIleProTyrTyrValGlyProLeuAlaArgGlyAsn
SerAspPheAlaTrpSerIleArgLysArgAsnGluLysIleThrProTrpAsnPheGlu

[0145]

AspValIleAspLysGluSerSerAlaGluAlaPheIleAsnArgMetThrSerPheAsp
LeuTyrLeuProGluGluLysValLeuProLysHisSerLeuLeuTyrGluThrPheAsn
ValTyrAsnGluLeuThrLysValArgPheIleAlaGluSerMetArgAspTyrGlnPhe
LeuAspSerLysGlnLysLysAspIleValArgLeuTyrPheLysAspLysArgLysVal
ThrAspLysAspIleIleGluTyrLeuHisAlaIleTyrGlyTyrAspGlyIleGluLeu
LysGlyIleGluLysGlnPheAsnSerSerLeuSerThrTyrHisAspLeuLeuAsnIle
IleAsnAspLysGluPheLeuAspAspSerSerAsnGluAlaIleIleGluGluIleIle
HisThrLeuThrIlePheGluAspArgGluMetIleLysGlnArgLeuSerLysPheGlu
AsnIlePheAspLysSerValLeuLysLysLeuSerArgArgHisTyrThrGlyTrpGly
LysLeuSerAlaLysLeuIleAsnGlyIleArgAspGluLysSerGlyAsnThrIleLeu
AspTyrLeuIleAspAspGlyIleSerAsnArgAsnPheMetGlnLeuIleHisAspAsp
AlaLeuSerPheLysLysLysIleGlnLysAlaGlnIleIleGlyAspGluAspLysGly
AsnIleLysGluValValLysSerLeuProGlySerProAlaIleLysLysGlyIleLeu
GlnSerIleLysIleValAspGluLeuValLysValMetGlyGlyArgLysProGluSer
IleValValGluMetAlaArgGluAsnGlnTyrThrAsnGlnGlyLysSerAsnSerGln
GlnArgLeuLysArgLeuGluLysSerLeuLysGluLeuGlySerLysIleLeuLysGlu
AsnIleProAlaLysLeuSerLysIleAspAsnAsnAlaLeuGlnAsnAspArgLeuTyr
LeuTyrTyrLeuGlnAsnGlyLysAspMetTyrThrGlyAspAspLeuAspIleAspArg
LeuSerAsnTyrAspIleAspHisIleIleProGlnAlaPheLeuLysAspAsnSerIle
AspAsnLysValLeuValSerSerAlaSerAsnArgGlyLysSerAspAspPheProSer
LeuGluValValLysLysArgLysThrPheTrpTyrGlnLeuLeuLysSerLysLeuIle
SerGlnArgLysPheAspAsnLeuThrLysAlaGluArgGlyGlyLeuLeuProGluAsp
LysAlaGlyPheIleGlnArgGlnLeuValGluThrArgGlnIleThrLysHisValAla
ArgLeuLeuAspGluLysPheAsnAsnLysLysAspGluAsnAsnArgAlaValArgThr
ValLysIleIleThrLeuLysSerThrLeuValSerGlnPheArgLysAspPheGluLeu
TyrLysValArgGluIleAsnAspPheHisHisAlaHisAspAlaTyrLeuAsnAlaVal
IleAlaSerAlaLeuLeuLysLysTyrProLysLeuGluProGluPheValTyrGlyAsp
TyrProLysTyrAsnSerPheArgGluArgLysSerAlaThrGluLysValTyrPheTyr
SerAsnIleMetAsnIlePheLysLysSerIleSerLeuAlaAspGlyArgValIleGlu
ArgProLeuIleGluValAsnGluGluThrGlyGluSerValTrpAsnLysGluSerAsp
LeuAlaThrValArgArgValLeuSerTyrProGlnValAsnValValLysLysValGlu
GluGlnAsnHisGlyLeuAspArgGlyLysProLysGlyLeuPheAsnAlaAsnLeuSer
SerLysProLysProAsnSerAsnGluAsnLeuValGlyAlaLysGluTyrLeuAspPro
LysLysTyrGlyGlyTyrAlaGlyIleSerAsnSerPheAlaValLeuValLysGlyThr
IleGluLysGlyAlaLysLysIleThrAsnValLeuGluPheGlnGlyIleSerIle
LeuAspArgIleAsnTyrArgLysAspLysLeuAsnPheLeuLeuGluLysGlyTyrLys
AspIleGluLeuIleIleGluLeuProLysTyrSerLeuPheGluLeuSerAspGlySer
ArgArgMetLeuAlaSerIleLeuSerThrAsnAsnLysArgGlyGluIleHisLysGly
AsnGlnIlePheLeuSerGlnLysPheValLysLeuLeuTyrHisAlaLysArgIleSer
AsnThrIleAsnGluAsnHisArgLysTyrValGluAsnHisLysLysGluPheGluGlu
LeuPheTyrTyrIleLeuGluPheAsnGluAsnTyrValGlyAlaLysLysAsnGlyLys
LeuLeuAsnSerAlaPheGlnSerTrpGlnAsnHisSerIleAspGluLeuCysSerSer
PheIleGlyProThrGlySerGluArgLysGlyLeuPheGluLeuThrSerArgGlySer
AlaAlaAspPheGluPheLeuGlyValLysIleProArgTyrArgAspTyrThrProSer
SerLeuLeuLysAspAlaThrLeuIleHisGlnSerValThrGlyLeuTyrGluThrArg
IleAspLeuAlaLysLeuGlyGluGly

[0146] SEQ ID NO:3

[0147] 891A突变体

[0148] 单字母序列:

mlfnkciisindfsnkekcmtpysigldigtnsvgwavitdnykvpsskmmkvlgnstskkyikknllgvllfdsgitaegrrlkrtrrrrr
 nrilylqefstematlddaaffqrlddsflvpddkrdsckypifgnlveekvyhdefptiyhlrkyladstkkadlrlvylalahmikyrghfliegef
 nsknndiqknfddfldtynafesdlsenskqlleeivkdisklekkdrilklfpgeknsqifseflklivgnqadfrkcfnldekaslhfksesy
 dedletllgyigddysdvflkakklydailsgflvtndeteapssamikrynehkedallkeyirnislktynevfkdtdkngyagidydgktn
 qedfyvylknllaefegadyflekidredflrkqrtfdngsipyqihlqemraildkqakfypflakneriekiltfripyvvgplargnsdfaws
 irkrnekitpwnfedvidkessaeafinrmstfdilylpeekvlpkhslllyetfnvyneltkvrfiaesmrdyqfldskqkddivlyfkdrkvtld
 kdiieylhaiyydgdielkgiekqfnsslstyhdllniindkefldssneaieeiihtlftidremikqrlskfenifdksvlkklsrrhytgwkl
 saklingirdeksgntildyldiddgisnrmfnqilhdalsfkkkiqkaqiigdedkgnikevvkslpgspaikkqilqsikivdelvkvmmggrk
 pesivvemarenqyntnqksnsqrrlkrlekslkelgskilenipaklskidnnalqndrilyllylqngkdmtygdldldirlnsydidhiip
 qaffkdsindknvlvssasargksddfpsevvkkrktfwyqlkksklsqrkfdnlkaerggllpedkagfiqrqlvetrqitkhvarildkfin
 nkdkdennravrtvkiitlksltvsqfrkdfelykvreindfhhahdaylnviasallkkyplefeydygdykynsfrerksatekvyfysni
 mnifkksisladgrvierplievneetgesvwnkesdlatvrrvlsypqvnvkvveeqnhgldrgkpglfnanlsskpkpnsnenlvgak
 eyldpkkygygagisnsfavlvktiekgakkkitnvlefqgisildrinyrkdklnfllekgykdielielpkyslfelsdgsrrmlasilstnnrk
 geihkngqiflsqkfvllyhakrisntinenhrkyvenhkkfeeflyyilefnenyvgakkngkllnsafsqwnhsidelcssfigptgser
 kgfelftsrgsaadfeffgvkipryrdytpssllkdatlihqsvgtlyetridlaklgeg

[0149] 三字母序列:

MetLeuPheAsnLysCysIleIleIleSerIleAsnLeuAspPheSerAsnLysGluLys
 CysMetThrLysProTyrSerIleGlyLeuAspIleGlyThrAsnSerValGlyTrpAla
 ValIleThrAspAsnTyrLysValProSerLysLysMetLysValLeuGlyAsnThrSer
 LysLysTyrIleLysLysAsnLeuLeuGlyValLeuLeuPheAspSerGlyIleThrAla
 GluGlyArgArgLeuLysArgThrAlaArgArgArgTyrThrArgArgArgAsnArgIle
 LeuTyrLeuGlnGluIlePheSerThrGluMetAlaThrLeuAspAspAlaPhePheGln
 ArgLeuAspAspSerPheLeuValProAspAspLysArgAspSerLysTyrProIlePhe
 GlyAsnLeuValGluGluLysValTyrHisAspGluPheProThrIleTyrHisLeuArg
 LysTyrLeuAlaAspSerThrLysLysAlaAspLeuArgLeuValTyrLeuAlaLeuAla
 HisMetIleLysTyrArgGlyHisPheLeulleGluGlyGluPheAsnSerLysAsnAsn
 AspIleGlnLysAsnPheGlnAspPheLeuAspThrTyrAsnAlaIlePheGluSerAsp
 LeuSerLeuGluAsnSerLysGlnLeuGluGluIleValLysAspLysIleSerLysLeu
 GluLysLysAspArgIleLeuLysLeuPheProGlyGluLysAsnSerGlyIlePheSer
 GluPheLeuLysLeuIleValGlyAsnGlnAlaAspPheArgLysCysPheAsnLeuAsp
 GluLysAlaSerLeuHisPheSerLysGluSerTyrAspGluAspLeuGluThrLeuLeu
 GlyTyrIleGlyAspAspTyrSerAspValPheLeuLysAlaLysLysLeuTyrAspAla
 IleLeuLeuSerGlyPheLeuThrValThrAspAsnGluThrGluAlaProLeuSerSer
 AlaMetIleLysArgTyrAsnGluHisLysGluAspLeuAlaLeuLeuLysGluTyrIle
 ArgAsnIleSerLeuLysThrTyrAsnGluValPheLysAspAspThrLysAsnGlyTyr

[0150]

AlaGlyTyrIleAspGlyLysThrAsnGlnGluAspPheTyrValTyrLeuLysAsnLeu
 LeuAlaGluPheGluGlyAlaAspTyrPheLeuGluLysIleAspArgGluAspPheLeu
 ArgLysGlnArgThrPheAspAsnGlySerIleProTyrGlnIleHisLeuGlnGluMet
 ArgAlaIleLeuAspLysGlnAlaLysPheTyrProPheLeuAlaLysAsnLysGluArg
 IleGluLysIleLeuThrPheArgIleProTyrTyrValGlyProLeuAlaArgGlyAsn
 SerAspPheAlaTrpSerIleArgLysArgAsnGluLysIleThrProTrpAsnPheGlu
 AspValIleAspLysGluSerSerAlaGluAlaPheIleAsnArgMetThrSerPheAsp
 LeuTyrLeuProGluGluLysValLeuProLysHisSerLeuLeuTyrGluThrPheAsn
 ValTyrAsnGluLeuThrLysValArgPheIleAlaGluSerMetArgAspTyrGlnPhe
 LeuAspSerLysGlnLysLysAspIleValArgLeuTyrPheLysAspLysArgLysVal
 ThrAspLysAspIleIleGluTyrLeuHisAlaIleTyrGlyTyrAspGlyIleGluLeu
 LysGlyIleGluLysGlnPheAsnSerSerLeuSerThrTyrHisAspLeuLeuAsnIle
 IleAsnAspLysGluPheLeuAspAspSerSerAsnGluAlaIleIleGluGluIleIle
 HisThrLeuThrIlePheGluAspArgGluMetIleLysGlnArgLeuSerLysPheGlu
 AsnIlePheAspLysSerValLeuLysLysLeuSerArgArgHisTyrThrGlyTrpGly
 LysLeuSerAlaLysLeuIleAsnGlyIleArgAspGluLysSerGlyAsnThrIleLeu
 AspTyrLeulleAspAspGlyIleSerAsnArgAsnPheMetGlnLeulleHisAspAsp
 AlaLeuSerPheLysLysLysIleGlnLysAlaGlnIleIleGlyAspGluAspLysGly
 AsnIleLysGluValValLysSerLeuProGlySerProAlaIleLysLysGlyIleLeu

[0151]

GlnSerIleLysIleValAspGluLeuValLysValMetGlyGlyArgLysProGluSer
IleValValGluMetAlaArgGluAsnGlnTyrThrAsnGlnGlyLysSerAsnSerGln
GlnArgLeuLysArgLeuGluLysSerLeuLysGluLeuGlySerLysIleLeuLysGlu
AsnIleProAlaLysLeuSerLysIleAspAsnAsnAlaLeuGlnAsnAspArgLeuTyr
LeuTyrTyrLeuGlnAsnGlyLysAspMetTyrThrGlyAspAspLeuAspIleAspArg
LeuSerAsnTyrAspIleAspHisIleIleProGlnAlaPheLeuLysAspAsnSerIle
AspAsnLysValLeuValSerSerAlaSerAlaArgGlyLysSerAspAspPheProSer
LeuGluValValLysLysArgLysThrPheTrpTyrGlnLeuLeuLysSerLysLeuIle
SerGlnArgLysPheAspAsnLeuThrLysAlaGluArgGlyGlyLeuLeuProGluAsp
LysAlaGlyPheIleGlnArgGlnLeuValGluThrArgGlnIleThrLysHisValAla
ArgLeuLeuAspGluLysPheAsnAsnLysLysAspGluAsnAsnArgAlaValArgThr
ValLysIleIleThrLeuLysSerThrLeuValSerGlnPheArgLysAspPheGluLeu
TyrLysValArgGluIleAsnAspPheHisHisAlaHisAspAlaTyrLeuAsnAlaVal
IleAlaSerAlaLeuLeuLysLysTyrProLysLeuGluProGluPheValTyrGlyAsp
TyrProLysTyrAsnSerPheArgGluArgLysSerAlaThrGluLysValTyrPheTyr
SerAsnIleMetAsnIlePheLysLysSerIleSerLeuAlaAspGlyArgValIleGlu
ArgProLeuIleGluValAsnGluGluThrGlyGluSerValTrpAsnLysGluSerAsp
LeuAlaThrValArgArgValLeuSerTyrProGlnValAsnValValLysLysValGlu
GluGlnAsnHisGlyLeuAspArgGlyLysProLysGlyLeuPheAsnAlaAsnLeuSer
SerLysProLysProAsnSerAsnGluAsnLeuValGlyAlaLysGluTyrLeuAspPro
LysLysTyrGlyGlyTyrAlaGlyIleSerAsnSerPheAlaValLeuValLysGlyThr
IleGluLysGlyAlaLysLysLysIleThrAsnValLeuGluPheGlnGlyIleSerIle
LeuAspArgIleAsnTyrArgLysAspLysLeuAsnPheLeuLeuGluLysGlyTyrLys
AspIleGluLeuIleIleGluLeuProLysTyrSerLeuPheGluLeuSerAspGlySer
ArgArgMetLeuAlaSerIleLeuSerThrAsnAsnLysArgGlyGluIleHisLysGly
AsnGlnIlePheLeuSerGlnLysPheValLysLeuLeuTyrHisAlaLysArgIleSer
AsnThrIleAsnGluAsnHisArgLysTyrValGluAsnHisLysLysGluPheGluGlu
LeuPheTyrTyrIleLeuGluPheAsnGluAsnTyrValGlyAlaLysLysAsnGlyLys
LeuLeuAsnSerAlaPheGlnSerTrpGlnAsnHisSerIleAspGluLeuCysSerSer
PheIleGlyProThrGlySerGluArgLysGlyLeuPheGluLeuThrSerArgGlySer
AlaAlaAspPheGluPheLeuGlyValLysIleProArgTyrArgAspTyrThrProSer
SerLeuLeuLysAspAlaThrLeuIleHisGlnSerValThrGlyLeuTyrGluThrArg
IleAspLeuAlaLysLeuGlyGluGly

[0152]

SEQ ID NO:4

[0153]

H868A突变体

[0154]

单字母序列:

mlfnkciislnldnsnkecmktpysigldigtntsvgwavitndnykvpsskmkvlgntskkyikknllgvlldsgitaegrrlkrtrrrrr
nrilylqefstematlddaffqrlddsflvpddkrdsypifgnlveekvyhdefptiyhlrkyladstkkadrlrvylalahmikyrghfliegef
nsknndiqknfqdfltdynaifesdlslenksqleeivkdiskilekkdrilklfpgeknsqifseflklivnqadfrkcfnldekaslhfskesy
dedletllyigiddysdvflkakklydaillsgflttvdnateaplssamikrynehkedlallkeyimisltkynevfkdtdkngyagvidgktn
qedfyvylknllaefegadyflekidredflrkqrtfdngsipyqihlqemraildkqakfypflaknkeriekiltfrippyvgplargnsdfaws
irkrnekitpwnfedvidkessaea finrmsfdlylpeekvlpkhslyetfnvyneltkvrfaesmrddyqfldskqkddivrllyfkdkrvtd
kdiieylhaiygydgiekgiekqfmslstyhdlniindkefldssneaieeihtlftifedremikqrlskfenifdksvlkklsrrhytgwgl
saklingirdeksgntildyliddgisnrmfinqlihddalsfkkiqkaqiigdedkgnikervvkslpgspaikkqilqsikivdelvkvimggrk
pesivvemarenqyntnqgksnsqrlkrlekslkelgskilkenipaklskidnnalqndrlylylqngkdmtygdldidrlsnydidaiipq
afldnsidnkvlvssasnrqksddfpsevvkkrktfwyqlksklisqrkfdnlkaerggllpedkagfiqrqlvetrqitkhvarldekfnn
kkdennravrtvkiitkltvsqfrkdfelykvreindfhahdaylnaviasallkypklepefygdydpkynsfrerksatekvyfysnim
nifkksisladgrvierplievneetgesvwnkesdlatvrrvlsypqvnvkvveeqnhdrgkpgklfnanlsskpkpnsnenlvgakey
ldpkkygygagisnsvfvlvktiekgakkkitnvlefgqisildrinyrkdklnflekgkydielielpkyslfelsdgsrrmlasilstnnkrge
ihkgngqiflsqfvlklyhakrisntinenhrkyvenhkkefeelfyilefmenyvgakngklinsafsqwnhsidelcssfignptgsergkl
feltsrgsaadfeifgvkipryrdytpssllkdatlihqsvgtglyetridlaklgeg

[0155]

三字母序列:

[0156]

MetLeuPheAsnLysCysIleIleIleSerIleAsnLeuAspPheSerAsnLysGluLys
CysMetThrLysProTyrSerIleGlyLeuAspIleGlyThrAsnSerValGlyTrpAla
ValIleThrAspAsnTyrLysValProSerLysLysMetLysValLeuGlyAsnThrSer
LysLysTyrIleLysLysAsnLeuLeuGlyValLeuLeuPheAspSerGlyIleThrAla
GluGlyArgArgLeuLysArgThrAlaArgArgArgTyrThrArgArgArgAsnArgIle
LeuTyrLeuGlnGluIlePheSerThrGluMetAlaThrLeuAspAspAlaPhePheGln
ArgLeuAspAspSerPheLeuValProAspAspLysArgAspSerLysTyrProIlePhe
GlyAsnLeuValGluGluLysValTyrHisAspGluPheProThrIleTyrHisLeuArg
LysTyrLeuAlaAspSerThrLysLysAlaAspLeuArgLeuValTyrLeuAlaLeuAla
HisMetIleLysTyrArgGlyHisPheLeuIleGluGlyGluPheAsnSerLysAsnAsn
AspIleGlnLysAsnPheGlnAspPheLeuAspThrTyrAsnAlaIlePheGluSerAsp
LeuSerLeuGluAsnSerLysGlnLeuGluGluIleValLysAspLysIleSerLysLeu
GluLysLysAspArgIleLeuLysLeuPheProGlyGluLysAsnSerGlyIlePheSer
GluPheLeuLysLeuIleValGlyAsnGlnAlaAspPheArgLysCysPheAsnLeuAsp
GluLysAlaSerLeuHisPheSerLysGluSerTyrAspGluAspLeuGluThrLeuLeu
GlyTyrIleGlyAspAspTyrSerAspValPheLeuLysAlaLysLysLeuTyrAspAla
IleLeuLeuSerGlyPheLeuThrValThrAspAsnGluThrGluAlaProLeuSerSer
AlaMetIleLysArgTyrAsnGluHisLysGluAspLeuAlaLeuLeuLysGluTyrIle
ArgAsnIleSerLeuLysThrTyrAsnGluValPheLysAspAspThrLysAsnGlyTyr
AlaGlyTyrIleAspGlyLysThrAsnGlnGluAspPheTyrValTyrLeuLysAsnLeu
LeuAlaGluPheGluGlyAlaAspTyrPheLeuGluLysIleAspArgGluAspPheLeu
ArgLysGlnArgThrPheAspAsnGlySerIleProTyrGlnIleHisLeuGlnGluMet
ArgAlaIleLeuAspLysGlnAlaLysPheTyrProPheLeuAlaLysAsnLysGluArg
IleGluLysIleLeuThrPheArgIleProTyrTyrValGlyProLeuAlaArgGlyAsn
SerAspPheAlaTrpSerIleArgLysArgAsnGluLysIleThrProTrpAsnPheGlu
AspValIleAspLysGluSerSerAlaGluAlaPheIleAsnArgMetThrSerPheAsp
LeuTyrLeuProGluGluLysValLeuProLysHisSerLeuLeuTyrGluThrPheAsn
ValTyrAsnGluLeuThrLysValArgPheIleAlaGluSerMetArgAspTyrGlnPhe
LeuAspSerLysGlnLysLysAspIleValArgLeuTyrPheLysAspLysArgLysVal
ThrAspLysAspIleIleGluTyrLeuHisAlaIleTyrGlyTyrAspGlyIleGluLeu
LysGlyIleGluLysGlnPheAsnSerSerLeuSerThrTyrHisAspLeuLeuAsnIle
IleAsnAspLysGluPheLeuAspAspSerSerAsnGluAlaIleIleGluGluIleIle
HisThrLeuThrIlePheGluAspArgGluMetIleLysGlnArgLeuSerLysPheGlu
AsnIlePheAspLysSerValLeuLysLysLeuSerArgArgHisTyrThrGlyTrpGly
LysLeuSerAlaLysLeuIleAsnGlyIleArgAspGluLysSerGlyAsnThrIleLeu
AspTyrLeuIleAspAspGlyIleSerAsnArgAsnPheMetGlnLeuIleHisAspAsp
AlaLeuSerPheLysLysLysIleGlnLysAlaGlnIleIleGlyAspGluAspLysGly
AsnIleLysGluValValLysSerLeuProGlySerProAlaIleLysLysGlyIleLeu
GlnSerIleLysIleValAspGluLeuValLysValMetGlyGlyArgLysProGluSer
IleValValGluMetAlaArgGluAsnGlnTyrThrAsnGlnGlyLysSerAsnSerGln
GlnArgLeuLysArgLeuGluLysSerLeuLysGluLeuGlySerLysIleLeuLysGlu
AsnIleProAlaLysLeuSerLysIleAspAsnAsnAlaLeuGlnAsnAspArgLeuTyr
LeuTyrTyrLeuGlnAsnGlyLysAspMetTyrThrGlyAspAspLeuAspIleAspArg
LeuSerAsnTyrAspIleAspAlaIleIleProGlnAlaPheLeuLysAspAsnSerIle
AspAsnLysValLeuValSerSerAlaSerAsnArgGlyLysSerAspAspPheProSer
LeuGluValValLysLysArgLysThrPheTrpTyrGlnLeuLeuLysSerLysLeuIle
SerGlnArgLysPheAspAsnLeuThrLysAlaGluArgGlyGlyLeuLeuProGluAsp
LysAlaGlyPheIleGlnArgGlnLeuValGluThrArgGlnIleThrLysHisValAla
ArgLeuLeuAspGluLysPheAsnAsnLysLysAspGluAsnAsnArgAlaValArgThr
ValLysIleIleThrLeuLysSerThrLeuValSerGlnPheArgLysAspPheGluLeu
TyrLysValArgGluIleAsnAspPheHisHisAlaHisAspAlaTyrLeuAsnAlaVal

- IleAlaSerAlaLeuLeuLysLysTyrProLysLeuGluProGluPheValTyrGlyAsp
TyrProLysTyrAsnSerPheArgGluArgLysSerAlaThrGluLysValTyrPheTyr
SerAsnIleMetAsnIlePheLysLysSerIleSerLeuAlaAspGlyArgValIleGlu
ArgProLeuIleGluValAsnGluGluThrGlyGluSerValTrpAsnLysGluSerAsp
LeuAlaThrValArgArgValLeuSerTyrProGlnValAsnValValLysLysValGlu
GluGlnAsnHisGlyLeuAspArgGlyLysProLysGlyLeuPheAsnAlaAsnLeuSer
SerLysProLysProAsnSerAsnGluAsnLeuValGlyAlaLysGluTyrLeuAspPro
LysLysTyrGlyGlyTyrAlaGlyIleSerAsnSerPheAlaValLeuValLysGlyThr
IleGluLysGlyAlaLysLysLysIleThrAsnValLeuGluPheGlnGlyIleSerIle
[0157] LeuAspArgIleAsnTyrArgLysAspLysLeuAsnPheLeuLeuGluLysGlyTyrLys
AspIleGluLeuIleIleGluLeuProLysTyrSerLeuPheGluLeuSerAspGlySer
ArgArgMetLeuAlaSerIleLeuSerThrAsnAsnLysArgGlyGluIleHisLysGly
AsnGlnIlePheLeuSerGlnLysPheValLysLeuLeuTyrHisAlaLysArgIleSer
AsnThrIleAsnGluAsnHisArgLysTyrValGluAsnHisLysLysGluPheGluGlu
LeuPheTyrTyrIleLeuGluPheAsnGluAsnTyrValGlyAlaLysLysAsnGlyLys
LeuLeuAsnSerAlaPheGlnSerTrpGlnAsnHisSerIleAspGluLeuCysSerSer
PheIleGlyProThrGlySerGluArgLysGlyLeuPheGluLeuThrSerArgGlySer
AlaAlaAspPheGluPheLeuGlyValLysIleProArgTyrArgAspTyrThrProSer
SerLeuLeuLysAspAlaThrLeuIleHisGlnSerValThrGlyLeuTyrGluThrArg
IleAspLeuAlaLysLeuGlyGluGly
- [0158] SEQ ID NO:5
- [0159] tra-crRNA, 未成熟 (102nt):
uaauaauuuugggguuugaaccauucgaaacaacacagcgaguuaaauaaggcuuaguccguacucaacuugaaaaggugggcac
cgauucgguguuuuu
- [0160] SEQ ID NO:6
- [0161] 78nt的成熟tracrRNA:
gggcgaacaacacagcgaguuaaaauaaggcuuaguccguacucaacuugaaaaggugggcaccgauucgguguuuuu
- [0162] 较短的变体:
gggcgaacaacacagcgaguuaaaauaaggcuuaguccguacucaacuugaaaaggugggcaccgauucggug (SEQ ID NO:
44)
gggcgaacaacacagcgaguuaaaauaaggcuuaguccguacucaacuugaaaaggugggcaccgauu (SEQ ID NO: 45)
- [0163] gggcgaacaacacagcgaguuaaaauaaggcuuaguccguacucaacuugaaaaggugggcac (SEQ ID NO: 46)
gggcgaacaacacagcgaguuaaaauaaggcuuaguccguacucaacuugaaaaggu (SEQ ID NO: 47)
gggcgaacaacacagcgaguuaaaauaaggcuuaguccguacucaacuugaa (SEQ ID NO: 48)
gggcgaacaacacagcgaguuaaaauaaggcuuaguccguacucaac (SEQ ID NO: 49)
- [0164] SEQ ID NO:7
- [0165] 来自间隔子1的42nt crRNA:
5'-CGCUAAAGAGGAAGAGGACAGUUUUAGAGCUGUGUUGUUUCG-3'
- [0166] SEQ ID NO:8
- [0167] 150nt前体-cr RNA
5'-
- [0168] gggugaaaaagauauccuacgagguuuuagagcuguguguuuucgaaugguuccaaaaacaaauucuaa
acgcuaaaagaggaagaggacaguuuuagagcuguguguuuucgaaugguuccaaaaacuacugcugauu
uagcuugguuguug-3'

[0169] SEQ ID NO:9

[0170] crRNA1

5' -

[0171] ggguaagaaaagauauccuacgagguuuuagagcuguguguuuucgaaugguuccaaaaacTGTCATGA
TAATAATGGTTTCTTAGACGTCguuuuagagcuguguguuuucgaaugguuccaaaaacuacugcug
uauuagcuugguuguug-3'

[0172] SEQ ID NO:10

[0173] crRNA2

5' -

[0174] ggguaagaaaagauauccuacgagguuuuagagcuguguguuuucgaaugguuccaaaaacacgagccg
gaagcataaaagtgtaaagcctgguuuuagagcuguguguuuucgaaugguuccaaaaacuacugcug
uauuagcuugguuguug-3'

[0175] SEQ ID NO:11

[0176] 抗-λ噬菌体CRISPR RNA

5' -

[0177] ggguaagaaaagauauccuacgagguuuuagagcuguguguuuucgaaugguuccaaaactcaagggga
gaatagaggctctcgttgcatgtguuuuagagcuguguguuuucgaaugguuccaaaacuacugcug
uauuagcuugguuguug-3'

[0178] SEQ ID NO:12

[0179] 抗-大肠杆菌CRISPR RNA

5' -

[0180] ggguaagaaaagauauccuacgagguuuuagagcuguguguuuucgaaugguuccaaaaccgggaggg
aagctgcatgatgcatgttatgtguuuuagagcuguguguuuucgaaugguuccaaaacuacugcug
uauuagcuugguuguug-3'

[0181] SEQ ID NO:13

[0182] crRNA-HS1

[0183] 5'-GCUCCCGGGCUCGAUGAAGGUUUUAGAGCUGUGUUGUUUCG-3'

[0184] SEQ ID NO:14

[0185] crRNA-HS2

[0186] UGAAUCGUGAAAUCUGCUCAGUUUAGAGCUGUGUUGUUUCG

[0187] 本申请含有序列列表,其以ASCII格式提交并以其全文并入本文作为参考。ASCII拷贝文件创建于2013年3月20日,命名为078981_6_SL.txt,大小为64.4kb。

[0188] 说明书中显示并描述的这些实施方式只是作为本领域技术人员的发明者们的具体实施方式,不以任何方式进行限制。因此,在不偏离以下权利要求范围内的本发明的精神下可对这些实施方式作出不同的改变、修饰或变动。所引用的参考文献清楚地以其全文并入本文作为参考。

[0189] 本说明书还包括下列内容:

[0190] 1.用于对靶标DNA分子进行位点特异性修饰的方法,所述方法包括在合适的条件下将靶标DNA分子和RNA导向的DNA核酸内切酶相接触,所述RNA导向的DNA核酸内切酶含有至少一个RNA序列以及RuvC活性位点基序和HNH活性位点基序中的至少一个;

[0191] 以使得靶标DNA分子在某个区域受到修饰,所述区域是由RNA序列与靶标DNA分子

的互补结合所确定的。

[0192] 2.实施方式1的方法,其中位点特异性修饰在体内或在体外发生。

[0193] 3.实施方式1的方法,其中靶标DNA是双链或单链的。

[0194] 4.蛋白质-RNA复合物,其包含

[0195] a) 与SEQ ID NO:1具有至少80%同一性的多肽,

[0196] b) 第一多核糖核苷酸crRNA,其包含3'区域和5'区域,其中3'区域包含CRISPR基因座中存在的至少22个核苷酸的重复序列,而5'区域包含CRISPR基因座中重复序列下游紧跟的至少20个核苷酸的间隔子序列,以及

[0197] c) 第二多核糖核苷酸tracrRNA,其包含5'和3'区域,其中5'区域的至少一部分与crRNA的3'区域互补。

[0198] 5.实施方式4的复合物,其中tracrRNA的5'区域为至少22个核苷酸,并与crRNA的22个核苷酸的3'区域互补。

[0199] 6.实施方式4的复合物,其分离自嗜热链球菌。

[0200] 7.实施方式4的复合物,其分离自遗传修饰的微生物。

[0201] 8.实施方式4的复合物,其中多肽分离自遗传修饰的大肠埃希氏杆菌。

[0202] 9.实施方式4的复合物,其中多肽是由选自重组DNA技术或化学合成的方法产生的。

[0203] 10.实施方式4的复合物,其中多肽或其片段是以融合多肽的形式表达的,其包含至少一种其他的氨基酸序列。

[0204] 11.实施方式10的复合物,其中所述其他的序列用于通过亲和色谱纯化融合多肽。

[0205] 12.实施方式4的复合物,其中多肽在RuvC活性位点基序含有一个点突变。

[0206] 13.实施方式12的复合物,其中点突变是D31A。

[0207] 14.实施方式4的复合物,其中多肽在HNH活性位点基序内含有有点突变。

[0208] 15.实施方式14的复合物,其中点突变是N891A。

[0209] 16.实施方式4的复合物,其中第一多核糖核苷酸(crRNA)序列包含5'-NNNNNNNNNNNNNNNNNNNNNGUUUAGAGCUGUGUUGUUUCG-3'(SEQ ID NO:15)。

[0210] 17.用于在遗传修饰的微生物中获得实施方式4的复合物的方法,所述复合物含有体内组装实施方式4的多肽和多核糖核苷酸复合物所需的第一多核糖核苷酸(crRNA)和第二多核糖核苷酸(tracrRNA),所述方法包括

[0211] 向微生物中引入至少一种编码实施方式4的多肽、实施方式4的第一多核糖核苷酸以及实施方式4的第二多核糖核苷酸的质粒,以产生遗传修饰的微生物,

[0212] 其中第二多核糖核苷酸包含与嗜热链球菌DGCC7710 CRISPR3系统的重复序列区域及下游区域显示至少80%互补的序列,且第一多核糖核苷酸(crRNA)由间隔子及邻侧的重复序列组成,

[0213] 其中实施方式4的多肽以及第一和第二多核糖核苷酸在遗传修饰微生物中至少一种质粒中编码,其中所述质粒含有所述遗传修饰微生物的宿主启动子或来自原生宿主基因组的启动子。

[0214] 18.实施方式17的方法,其进一步包含从所述遗传修饰微生物中分离复合物。

[0215] 19.实施方式17的方法,其中实施方式4的多肽和多核糖核苷酸是在遗传修饰微生物

物的两种分开的质粒中编码的。

[0216] 20.实施方式17的方法,其中实施方式4的多肽和多核糖核苷酸是在遗传修饰微生物的三种分开的质粒中编码的。

[0217] 21.实施方式17的方法,其进一步包括从遗传修饰微生物中分离所述多肽、第一多核糖核苷酸crRNA以及第二多核糖核苷酸tracrRNA,并将所述多肽、crRNA和tracrRNA在适合复合物组装的条件下进行孵育。

[0218] 22.实施方式21的方法,其中第一多核糖核苷酸crRNA序列包含5'-NNNNNNNNNNNNNNNNNNNNNGUUUAGAGCUGUGUUUCG-3'(SEQ ID NO:15)。

[0219] 23.用于裂解多核苷酸序列的方法,所述方法包括:

[0220] 将实施方式4的复合物与靶标DNA在适合于DNA裂解的条件下孵育,

[0221] 其中靶标DNA含有

[0222] 原型间隔子序列,其与复合物中第一多核糖核苷酸的间隔子序列至少80%互补,以及

[0223] 位于原型间隔子序列下游的原型间隔子邻近基序(PAM)序列NGGNG,

[0224] 其中多肽在位于PAM序列上游4个核苷酸处的裂解位点裂解两条靶标DNA链,形成平末端。

[0225] 24.实施方式23的方法,其中原型间隔子序列与间隔子序列100%互补。

[0226] 25.实施方式23的方法,其中裂解发生在体外或体内。

[0227] 26.用于裂解核苷酸序列的方法,所述方法包括

[0228] 将实施方式4的复合物与单链靶标DNA在适合DNA裂解的条件下孵育,其中靶标DNA含有与第一多核糖核苷酸中间隔子序列至少80%互补的原型间隔子序列,其中多肽在裂解位点裂解单一的DNA链。

[0229] 27.实施方式26的方法,其中原型间隔子序列与多核糖核苷酸中的间隔子序列100%互补。

[0230] 28.实施方式26的方法,其中裂解发生于体外或体内。

[0231] 29.用于对多核苷酸序列进行位点特异性切口的方法,所述方法包括

[0232] 将复合物与双链靶标DNA在适合进行DNA切口的条件下进行孵育,其中复合物包含

[0233] (a) 第一多核糖核苷酸crRNA,其包含3'区域和5'区域,其中3'区域包含CRISPR基因座中存在的至少22个核苷酸的重复序列,而5'区域包含CRISPR基因座中重复序列下游紧跟的至少20个核苷酸的间隔子序列,

[0234] (b) 第二多核糖核苷酸tracrRNA,其包含5'和3'区域,其中5'区域中至少一部分与第一多核糖核苷酸的3'区域互补,以及

[0235] (c) 多肽RuvC,其在活性位点基序中含有点突变,

[0236] 其中靶标DNA含有与第一多核糖核苷酸中的间隔子序列至少80%互补的原型间隔子序列和位于原型间隔子序列下游的原型间隔子邻近基序(PAM)序列NGGNG,其中多肽在位于PAM序列上游4个核苷酸处的裂解位点处裂解双链DNA的一条DNA链,产生位点特异性切口。

[0237] 30.实施方式29的方法,其中原型间隔子序列与多核糖核苷酸中间隔子序列100%互补。

[0238] 31.实施方式29的方法,其中位点特异性切口在体外或体内发生。

[0239] 32.用于对核苷酸序列进行位点特异性切口的方法,所述方法包括复合物与双链靶标DNA在适合进行DNA切口的条件下进行孵育,其中复合物包含

[0240] (a) 第一多核糖核苷酸crRNA,其包含3'区域和5'区域,其中3'区域包含CRISPR基因座中存在的至少22个核苷酸的重复序列,而5'区域包含CRISPR基因座中重复序列下游紧跟的至少20个核苷酸的间隔子序列,(b) 第二多核糖核苷酸tracrRNA,其包含5'和3'区域,其中5'区域的至少一部分与第一多核糖核苷酸的3'区域互补(c) 多肽HNH,其在活性位点基序中含有点突变,

[0241] 其中靶标DNA含有与第一多核糖核苷酸中的间隔子序列有至少80%互补的原型间隔子序列和位于原型间隔子序列下游的原型间隔子邻近基序(PAM)序列NGGNG,其中多肽在位于PAM序列上游4个核苷酸处的裂解位点处裂解双链DNA的一条DNA链。

[0242] 33.实施方式32的方法,其中原型间隔子序列与多核糖核苷酸中的间隔子序列100%互补。

[0243] 34.实施方式32的方法,其中位点特异性切口在体外或体内发生。

[0244] 35.生成实施方式4的复合物的体外方法,该方法包括

[0245] 将复合物的组分在适合复合物组装的条件下孵育,其中复合物包含第一多核糖核苷酸和任何所需的间隔子序列,所述第一多核糖核苷酸具有包含5'-NNNNNNNNNNNNNNNNNNNNGUUUUAGAGCUGUGUUGUUUCG-3'(SEQ ID NO:15)的序列。

[0246] 36.实施方式35的方法,其中第一多核糖核苷酸是通过体外转录或化学合成获得的,其中体外转录是从含有单一重复序列-间隔子-重复序列单元的DNA片段进行,其中间隔子具有任何所需的序列。

[0247] 37.实施方式35的方法,其中组分包含Cas9多肽(SEQ ID NO:1)、tracrRNA多核糖核苷酸(SEQ ID NO:5)以及crRNA多核糖核苷酸(5'-NNNNNNNNNNNNNNNNNNNNGUUUUAGAGCUGUGUUGUUUCG-3')(SEQ ID NO:15),并具有任何所需的间隔子序列。

[0248] 38.实施方式35的方法,其进一步包括向复合物中加入RNaseIII多肽。

[0249] 39.来自嗜热链球菌DGCC7710菌株CRISPR3/Cas系统的分离的Cas9-crRNA复合物。

[0250] 40.嗜热链球菌DGCC7710菌株CRISPR3系统的分离的Cas9蛋白质,以及显示有80%同一性并含有D31 A和N891 A突变的序列,这些突变体显示了DNA切口活性。

[0251] 41.嗜热链球菌DGCC7710菌株CRISPR3系统的分离的crRNA,其具有5'标签但缺少3'标签。

[0252] 42.使用含有单一重复序列-间隔子-重复序列单元的序列盒子对Cas9-crRNA特异性重新编程的方法,所述方法包括

[0253] (a) 通过体外转录从DNA序列中产生具有30个核苷酸长度的序列盒子,

[0254] (b) 将所述序列盒子插入crRNA的重复序列-间隔子-重复序列,以及

[0255] (c) 使用步骤(b)的crRNA组装Cas9-crRNA复合物。

[0256] 43.用于体内对靶标DNA分子进行选择位点特异性修饰的方法,该方法包括

[0257] 生成第一多核糖核苷酸crRNA,其含有对靶标DNA中所需修饰位点显示互补的间隔子序列,

[0258] 将第一多核糖核苷酸crRNA与具有与SEQ ID NO:1至少80%同一性的多肽及第二

多核糖核苷酸tracrRNA在形成复合物的适合条件下进行复合,以及

[0259] 将复合物引入至少一个细胞中以产生靶标DNA分子的选择性位点特异性修饰,发生修饰的区域是通过第一多核糖核苷酸crRNA与靶标DNA分子的互补间隔子序列确定的。

[0260] 44.实施方式43的方法,其中第一多核糖核苷酸crRNA是通过体外翻译生成的或化学合成的。

[0261] 45.实施方式43的方法,其中的修饰是靶标DNA分子的双链裂解,以至少产生靶标DNA分子的第一和第二部分。

[0262] 46.实施方式45的方法,其进一步包括通过同源重组或非同源末端连接之一的方法将靶标DNA分子的第一和第二部分的重新连接。

[0263] 47.实施方式43的方法,其中靶标DNA含有与复合物中第一多核糖核苷酸crRNA中的间隔子序列至少80%互补的原型间隔子序列和在原型间隔子序列下游的原型间隔子邻近基序(PAM)序列NGGNG,其中多肽在位于PAM序列上游4个核苷酸处的裂解位点裂解两条靶标DNA链,形成平末端。

[0264] 48.实施方式43的方法,其中多肽为Cas9 (SEQ ID NO:1)。

[0265] 49.实施方式43的方法,其进一步包含向复合物中加入RNaseIII多肽。

[0001] 序列表
 [0002] <110> VILNIUS UNIVERSITY
 [0003] <120> 通过Cas9-crRNA复合物的RNA指导的DNA裂解
 [0004] <130> 078981.6
 [0005] <140>
 [0006] <141>
 [0007] <150> 61/625,420
 [0008] <151> 2012-04-17
 [0009] <150> 61/613,373
 [0010] <151> 2012-03-20
 [0011] <160> 69
 [0012] <170> PatentIn version 3.5
 [0013] <210> 1
 [0014] <211> 1409
 [0015] <212> PRT
 [0016] <213> 嗜热链球菌(*Streptococcus thermophilus*)
 [0017] <400> 1
 [0018] Met Leu Phe Asn Lys Cys Ile Ile Ile Ser Ile Asn Leu Asp Phe Ser
 [0019] 1 5 10 15
 [0020] Asn Lys Glu Lys Cys Met Thr Lys Pro Tyr Ser Ile Gly Leu Asp Ile
 [0021] 20 25 30
 [0022] Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Asn Tyr Lys Val
 [0023] 35 40 45
 [0024] Pro Ser Lys Lys Met Lys Val Leu Gly Asn Thr Ser Lys Lys Tyr Ile
 [0025] 50 55 60
 [0026] Lys Lys Asn Leu Leu Gly Val Leu Leu Phe Asp Ser Gly Ile Thr Ala
 [0027] 65 70 75 80
 [0028] Glu Gly Arg Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg
 [0029] 85 90 95
 [0030] Arg Asn Arg Ile Leu Tyr Leu Gln Glu Ile Phe Ser Thr Glu Met Ala
 [0031] 100 105 110
 [0032] Thr Leu Asp Asp Ala Phe Phe Gln Arg Leu Asp Asp Ser Phe Leu Val
 [0033] 115 120 125
 [0034] Pro Asp Asp Lys Arg Asp Ser Lys Tyr Pro Ile Phe Gly Asn Leu Val
 [0035] 130 135 140
 [0036] Glu Glu Lys Val Tyr His Asp Glu Phe Pro Thr Ile Tyr His Leu Arg
 [0037] 145 150 155 160
 [0038] Lys Tyr Leu Ala Asp Ser Thr Lys Lys Ala Asp Leu Arg Leu Val Tyr

[0039]	165							170							175				
[0040]	Leu	Ala	Leu	Ala	His	Met	Ile	Lys	Tyr	Arg	Gly	His	Phe	Leu	Ile	Glu			
[0041]	180							185							190				
[0042]	Gly	Glu	Phe	Asn	Ser	Lys	Asn	Asn	Asp	Ile	Gln	Lys	Asn	Phe	Gln	Asp			
[0043]	195							200							205				
[0044]	Phe	Leu	Asp	Thr	Tyr	Asn	Ala	Ile	Phe	Glu	Ser	Asp	Leu	Ser	Leu	Glu			
[0045]	210							215							220				
[0046]	Asn	Ser	Lys	Gln	Leu	Glu	Glu	Ile	Val	Lys	Asp	Lys	Ile	Ser	Lys	Leu			
[0047]	225							230							235			240	
[0048]	Glu	Lys	Lys	Asp	Arg	Ile	Leu	Lys	Leu	Phe	Pro	Gly	Glu	Lys	Asn	Ser			
[0049]	245							250							255				
[0050]	Gly	Ile	Phe	Ser	Glu	Phe	Leu	Lys	Leu	Ile	Val	Gly	Asn	Gln	Ala	Asp			
[0051]	260							265							270				
[0052]	Phe	Arg	Lys	Cys	Phe	Asn	Leu	Asp	Glu	Lys	Ala	Ser	Leu	His	Phe	Ser			
[0053]	275							280							285				
[0054]	Lys	Glu	Ser	Tyr	Asp	Glu	Asp	Leu	Glu	Thr	Leu	Leu	Gly	Tyr	Ile	Gly			
[0055]	290							295							300				
[0056]	Asp	Asp	Tyr	Ser	Asp	Val	Phe	Leu	Lys	Ala	Lys	Lys	Leu	Tyr	Asp	Ala			
[0057]	305							310							315			320	
[0058]	Ile	Leu	Leu	Ser	Gly	Phe	Leu	Thr	Val	Thr	Asp	Asn	Glu	Thr	Glu	Ala			
[0059]	325							330							335				
[0060]	Pro	Leu	Ser	Ser	Ala	Met	Ile	Lys	Arg	Tyr	Asn	Glu	His	Lys	Glu	Asp			
[0061]	340							345							350				
[0062]	Leu	Ala	Leu	Leu	Lys	Glu	Tyr	Ile	Arg	Asn	Ile	Ser	Leu	Lys	Thr	Tyr			
[0063]	355							360							365				
[0064]	Asn	Glu	Val	Phe	Lys	Asp	Asp	Thr	Lys	Asn	Gly	Tyr	Ala	Gly	Tyr	Ile			
[0065]	370							375							380				
[0066]	Asp	Gly	Lys	Thr	Asn	Gln	Glu	Asp	Phe	Tyr	Val	Tyr	Leu	Lys	Asn	Leu			
[0067]	385							390							395			400	
[0068]	Leu	Ala	Glu	Phe	Glu	Gly	Ala	Asp	Tyr	Phe	Leu	Glu	Lys	Ile	Asp	Arg			
[0069]	405							410							415				
[0070]	Glu	Asp	Phe	Leu	Arg	Lys	Gln	Arg	Thr	Phe	Asp	Asn	Gly	Ser	Ile	Pro			
[0071]	420							425							430				
[0072]	Tyr	Gln	Ile	His	Leu	Gln	Glu	Met	Arg	Ala	Ile	Leu	Asp	Lys	Gln	Ala			
[0073]	435							440							445				
[0074]	Lys	Phe	Tyr	Pro	Phe	Leu	Ala	Lys	Asn	Lys	Glu	Arg	Ile	Glu	Lys	Ile			
[0075]	450							455							460				
[0076]	Leu	Thr	Phe	Arg	Ile	Pro	Tyr	Tyr	Val	Gly	Pro	Leu	Ala	Arg	Gly	Asn			
[0077]	465							470							475			480	

[0078]	Ser Asp Phe Ala Trp Ser Ile Arg Lys Arg Asn Glu Lys Ile Thr Pro		
[0079]		485	490 495
[0080]	Trp Asn Phe Glu Asp Val Ile Asp Lys Glu Ser Ser Ala Glu Ala Phe		
[0081]		500	505 510
[0082]	Ile Asn Arg Met Thr Ser Phe Asp Leu Tyr Leu Pro Glu Glu Lys Val		
[0083]		515	520 525
[0084]	Leu Pro Lys His Ser Leu Leu Tyr Glu Thr Phe Asn Val Tyr Asn Glu		
[0085]		530	535 540
[0086]	Leu Thr Lys Val Arg Phe Ile Ala Glu Ser Met Arg Asp Tyr Gln Phe		
[0087]	545	550	555 560
[0088]	Leu Asp Ser Lys Gln Lys Lys Asp Ile Val Arg Leu Tyr Phe Lys Asp		
[0089]		565	570 575
[0090]	Lys Arg Lys Val Thr Asp Lys Asp Ile Ile Glu Tyr Leu His Ala Ile		
[0091]		580	585 590
[0092]	Tyr Gly Tyr Asp Gly Ile Glu Leu Lys Gly Ile Glu Lys Gln Phe Asn		
[0093]		595	600 605
[0094]	Ser Ser Leu Ser Thr Tyr His Asp Leu Leu Asn Ile Ile Asn Asp Lys		
[0095]		610	615 620
[0096]	Glu Phe Leu Asp Asp Ser Ser Asn Glu Ala Ile Ile Glu Glu Ile Ile		
[0097]	625	630	635 640
[0098]	His Thr Leu Thr Ile Phe Glu Asp Arg Glu Met Ile Lys Gln Arg Leu		
[0099]		645	650 655
[0100]	Ser Lys Phe Glu Asn Ile Phe Asp Lys Ser Val Leu Lys Lys Leu Ser		
[0101]		660	665 670
[0102]	Arg Arg His Tyr Thr Gly Trp Gly Lys Leu Ser Ala Lys Leu Ile Asn		
[0103]		675	680 685
[0104]	Gly Ile Arg Asp Glu Lys Ser Gly Asn Thr Ile Leu Asp Tyr Leu Ile		
[0105]		690	695 700
[0106]	Asp Asp Gly Ile Ser Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp		
[0107]	705	710	715 720
[0108]	Ala Leu Ser Phe Lys Lys Lys Ile Gln Lys Ala Gln Ile Ile Gly Asp		
[0109]		725	730 735
[0110]	Glu Asp Lys Gly Asn Ile Lys Glu Val Val Lys Ser Leu Pro Gly Ser		
[0111]		740	745 750
[0112]	Pro Ala Ile Lys Lys Gly Ile Leu Gln Ser Ile Lys Ile Val Asp Glu		
[0113]		755	760 765
[0114]	Leu Val Lys Val Met Gly Gly Arg Lys Pro Glu Ser Ile Val Val Glu		
[0115]		770	775 780
[0116]	Met Ala Arg Glu Asn Gln Tyr Thr Asn Gln Gly Lys Ser Asn Ser Gln		

[0117]	785	790	795	800
[0118]	Gln Arg Leu Lys Arg Leu Glu Lys Ser Leu Lys Glu Leu Gly Ser Lys			
[0119]		805	810	815
[0120]	Ile Leu Lys Glu Asn Ile Pro Ala Lys Leu Ser Lys Ile Asp Asn Asn			
[0121]		820	825	830
[0122]	Ala Leu Gln Asn Asp Arg Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Lys			
[0123]		835	840	845
[0124]	Asp Met Tyr Thr Gly Asp Asp Leu Asp Ile Asp Arg Leu Ser Asn Tyr			
[0125]		850	855	860
[0126]	Asp Ile Asp His Ile Ile Pro Gln Ala Phe Leu Lys Asp Asn Ser Ile			
[0127]	865	870	875	880
[0128]	Asp Asn Lys Val Leu Val Ser Ser Ala Ser Asn Arg Gly Lys Ser Asp			
[0129]		885	890	895
[0130]	Asp Phe Pro Ser Leu Glu Val Val Lys Lys Arg Lys Thr Phe Trp Tyr			
[0131]		900	905	910
[0132]	Gln Leu Leu Lys Ser Lys Leu Ile Ser Gln Arg Lys Phe Asp Asn Leu			
[0133]		915	920	925
[0134]	Thr Lys Ala Glu Arg Gly Gly Leu Leu Pro Glu Asp Lys Ala Gly Phe			
[0135]		930	935	940
[0136]	Ile Gln Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys His Val Ala			
[0137]	945	950	955	960
[0138]	Arg Leu Leu Asp Glu Lys Phe Asn Asn Lys Lys Asp Glu Asn Asn Arg			
[0139]		965	970	975
[0140]	Ala Val Arg Thr Val Lys Ile Ile Thr Leu Lys Ser Thr Leu Val Ser			
[0141]		980	985	990
[0142]	Gln Phe Arg Lys Asp Phe Glu Leu Tyr Lys Val Arg Glu Ile Asn Asp			
[0143]		995	1000	1005
[0144]	Phe His His Ala His Asp Ala Tyr Leu Asn Ala Val Ile Ala Ser			
[0145]		1010	1015	1020
[0146]	Ala Leu Leu Lys Lys Tyr Pro Lys Leu Glu Pro Glu Phe Val Tyr			
[0147]		1025	1030	1035
[0148]	Gly Asp Tyr Pro Lys Tyr Asn Ser Phe Arg Glu Arg Lys Ser Ala			
[0149]		1040	1045	1050
[0150]	Thr Glu Lys Val Tyr Phe Tyr Ser Asn Ile Met Asn Ile Phe Lys			
[0151]		1055	1060	1065
[0152]	Lys Ser Ile Ser Leu Ala Asp Gly Arg Val Ile Glu Arg Pro Leu			
[0153]		1070	1075	1080
[0154]	Ile Glu Val Asn Glu Glu Thr Gly Glu Ser Val Trp Asn Lys Glu			
[0155]		1085	1090	1095

[0156]	Ser	Asp	Leu	Ala	Thr	Val	Arg	Arg	Val	Leu	Ser	Tyr	Pro	Gln	Val
[0157]		1100					1105					1110			
[0158]	Asn	Val	Val	Lys	Lys	Val	Glu	Glu	Gln	Asn	His	Gly	Leu	Asp	Arg
[0159]		1115					1120					1125			
[0160]	Gly	Lys	Pro	Lys	Gly	Leu	Phe	Asn	Ala	Asn	Leu	Ser	Ser	Lys	Pro
[0161]		1130					1135					1140			
[0162]	Lys	Pro	Asn	Ser	Asn	Glu	Asn	Leu	Val	Gly	Ala	Lys	Glu	Tyr	Leu
[0163]		1145					1150					1155			
[0164]	Asp	Pro	Lys	Lys	Tyr	Gly	Gly	Tyr	Ala	Gly	Ile	Ser	Asn	Ser	Phe
[0165]		1160					1165					1170			
[0166]	Ala	Val	Leu	Val	Lys	Gly	Thr	Ile	Glu	Lys	Gly	Ala	Lys	Lys	Lys
[0167]		1175					1180					1185			
[0168]	Ile	Thr	Asn	Val	Leu	Glu	Phe	Gln	Gly	Ile	Ser	Ile	Leu	Asp	Arg
[0169]		1190					1195					1200			
[0170]	Ile	Asn	Tyr	Arg	Lys	Asp	Lys	Leu	Asn	Phe	Leu	Leu	Glu	Lys	Gly
[0171]		1205					1210					1215			
[0172]	Tyr	Lys	Asp	Ile	Glu	Leu	Ile	Ile	Glu	Leu	Pro	Lys	Tyr	Ser	Leu
[0173]		1220					1225					1230			
[0174]	Phe	Glu	Leu	Ser	Asp	Gly	Ser	Arg	Arg	Met	Leu	Ala	Ser	Ile	Leu
[0175]		1235					1240					1245			
[0176]	Ser	Thr	Asn	Asn	Lys	Arg	Gly	Glu	Ile	His	Lys	Gly	Asn	Gln	Ile
[0177]		1250					1255					1260			
[0178]	Phe	Leu	Ser	Gln	Lys	Phe	Val	Lys	Leu	Leu	Tyr	His	Ala	Lys	Arg
[0179]		1265					1270					1275			
[0180]	Ile	Ser	Asn	Thr	Ile	Asn	Glu	Asn	His	Arg	Lys	Tyr	Val	Glu	Asn
[0181]		1280					1285					1290			
[0182]	His	Lys	Lys	Glu	Phe	Glu	Glu	Leu	Phe	Tyr	Tyr	Ile	Leu	Glu	Phe
[0183]		1295					1300					1305			
[0184]	Asn	Glu	Asn	Tyr	Val	Gly	Ala	Lys	Lys	Asn	Gly	Lys	Leu	Leu	Asn
[0185]		1310					1315					1320			
[0186]	Ser	Ala	Phe	Gln	Ser	Trp	Gln	Asn	His	Ser	Ile	Asp	Glu	Leu	Cys
[0187]		1325					1330					1335			
[0188]	Ser	Ser	Phe	Ile	Gly	Pro	Thr	Gly	Ser	Glu	Arg	Lys	Gly	Leu	Phe
[0189]		1340					1345					1350			
[0190]	Glu	Leu	Thr	Ser	Arg	Gly	Ser	Ala	Ala	Asp	Phe	Glu	Phe	Leu	Gly
[0191]		1355					1360					1365			
[0192]	Val	Lys	Ile	Pro	Arg	Tyr	Arg	Asp	Tyr	Thr	Pro	Ser	Ser	Leu	Leu
[0193]		1370					1375					1380			
[0194]	Lys	Asp	Ala	Thr	Leu	Ile	His	Gln	Ser	Val	Thr	Gly	Leu	Tyr	Glu

[0195]	1385	1390	1395
[0196]	Thr Arg Ile Asp Leu Ala Lys Leu Gly Glu Gly		
[0197]	1400	1405	
[0198]	<210> 2		
[0199]	<211> 1409		
[0200]	<212> PRT		
[0201]	<213> 嗜热链球菌 (Streptococcus thermophilus)		
[0202]	<400> 2		
[0203]	Met Leu Phe Asn Lys Cys Ile Ile Ile Ser Ile Asn Leu Asp Phe Ser		
[0204]	1 5 10 15		
[0205]	Asn Lys Glu Lys Cys Met Thr Lys Pro Tyr Ser Ile Gly Leu Ala Ile		
[0206]	20 25 30		
[0207]	Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Asn Tyr Lys Val		
[0208]	35 40 45		
[0209]	Pro Ser Lys Lys Met Lys Val Leu Gly Asn Thr Ser Lys Lys Tyr Ile		
[0210]	50 55 60		
[0211]	Lys Lys Asn Leu Leu Gly Val Leu Leu Phe Asp Ser Gly Ile Thr Ala		
[0212]	65 70 75 80		
[0213]	Glu Gly Arg Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg		
[0214]	85 90 95		
[0215]	Arg Asn Arg Ile Leu Tyr Leu Gln Glu Ile Phe Ser Thr Glu Met Ala		
[0216]	100 105 110		
[0217]	Thr Leu Asp Asp Ala Phe Phe Gln Arg Leu Asp Asp Ser Phe Leu Val		
[0218]	115 120 125		
[0219]	Pro Asp Asp Lys Arg Asp Ser Lys Tyr Pro Ile Phe Gly Asn Leu Val		
[0220]	130 135 140		
[0221]	Glu Glu Lys Val Tyr His Asp Glu Phe Pro Thr Ile Tyr His Leu Arg		
[0222]	145 150 155 160		
[0223]	Lys Tyr Leu Ala Asp Ser Thr Lys Lys Ala Asp Leu Arg Leu Val Tyr		
[0224]	165 170 175		
[0225]	Leu Ala Leu Ala His Met Ile Lys Tyr Arg Gly His Phe Leu Ile Glu		
[0226]	180 185 190		
[0227]	Gly Glu Phe Asn Ser Lys Asn Asn Asp Ile Gln Lys Asn Phe Gln Asp		
[0228]	195 200 205		
[0229]	Phe Leu Asp Thr Tyr Asn Ala Ile Phe Glu Ser Asp Leu Ser Leu Glu		
[0230]	210 215 220		
[0231]	Asn Ser Lys Gln Leu Glu Glu Ile Val Lys Asp Lys Ile Ser Lys Leu		
[0232]	225 230 235 240		
[0233]	Glu Lys Lys Asp Arg Ile Leu Lys Leu Phe Pro Gly Glu Lys Asn Ser		

[0234]		245		250		255
[0235]	Gly Ile Phe Ser Glu Phe Leu Lys Leu Ile Val Gly Asn Gln Ala Asp					
[0236]		260		265		270
[0237]	Phe Arg Lys Cys Phe Asn Leu Asp Glu Lys Ala Ser Leu His Phe Ser					
[0238]		275		280		285
[0239]	Lys Glu Ser Tyr Asp Glu Asp Leu Glu Thr Leu Leu Gly Tyr Ile Gly					
[0240]		290		295		300
[0241]	Asp Asp Tyr Ser Asp Val Phe Leu Lys Ala Lys Lys Leu Tyr Asp Ala					
[0242]	305		310		315	320
[0243]	Ile Leu Leu Ser Gly Phe Leu Thr Val Thr Asp Asn Glu Thr Glu Ala					
[0244]		325		330		335
[0245]	Pro Leu Ser Ser Ala Met Ile Lys Arg Tyr Asn Glu His Lys Glu Asp					
[0246]		340		345		350
[0247]	Leu Ala Leu Leu Lys Glu Tyr Ile Arg Asn Ile Ser Leu Lys Thr Tyr					
[0248]		355		360		365
[0249]	Asn Glu Val Phe Lys Asp Asp Thr Lys Asn Gly Tyr Ala Gly Tyr Ile					
[0250]		370		375		380
[0251]	Asp Gly Lys Thr Asn Gln Glu Asp Phe Tyr Val Tyr Leu Lys Asn Leu					
[0252]	385		390		395	400
[0253]	Leu Ala Glu Phe Glu Gly Ala Asp Tyr Phe Leu Glu Lys Ile Asp Arg					
[0254]		405		410		415
[0255]	Glu Asp Phe Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro					
[0256]		420		425		430
[0257]	Tyr Gln Ile His Leu Gln Glu Met Arg Ala Ile Leu Asp Lys Gln Ala					
[0258]		435		440		445
[0259]	Lys Phe Tyr Pro Phe Leu Ala Lys Asn Lys Glu Arg Ile Glu Lys Ile					
[0260]		450		455		460
[0261]	Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn					
[0262]	465		470		475	480
[0263]	Ser Asp Phe Ala Trp Ser Ile Arg Lys Arg Asn Glu Lys Ile Thr Pro					
[0264]		485		490		495
[0265]	Trp Asn Phe Glu Asp Val Ile Asp Lys Glu Ser Ser Ala Glu Ala Phe					
[0266]		500		505		510
[0267]	Ile Asn Arg Met Thr Ser Phe Asp Leu Tyr Leu Pro Glu Glu Lys Val					
[0268]		515		520		525
[0269]	Leu Pro Lys His Ser Leu Leu Tyr Glu Thr Phe Asn Val Tyr Asn Glu					
[0270]		530		535		540
[0271]	Leu Thr Lys Val Arg Phe Ile Ala Glu Ser Met Arg Asp Tyr Gln Phe					
[0272]	545		550		555	560

[0273]	Leu	Asp	Ser	Lys	Gln	Lys	Lys	Asp	Ile	Val	Arg	Leu	Tyr	Phe	Lys	Asp
[0274]					565					570					575	
[0275]	Lys	Arg	Lys	Val	Thr	Asp	Lys	Asp	Ile	Ile	Glu	Tyr	Leu	His	Ala	Ile
[0276]					580					585					590	
[0277]	Tyr	Gly	Tyr	Asp	Gly	Ile	Glu	Leu	Lys	Gly	Ile	Glu	Lys	Gln	Phe	Asn
[0278]					595					600					605	
[0279]	Ser	Ser	Leu	Ser	Thr	Tyr	His	Asp	Leu	Leu	Asn	Ile	Ile	Asn	Asp	Lys
[0280]					610					615					620	
[0281]	Glu	Phe	Leu	Asp	Asp	Ser	Ser	Asn	Glu	Ala	Ile	Ile	Glu	Glu	Ile	Ile
[0282]																
[0283]	His	Thr	Leu	Thr	Ile	Phe	Glu	Asp	Arg	Glu	Met	Ile	Lys	Gln	Arg	Leu
[0284]																
[0285]	Ser	Lys	Phe	Glu	Asn	Ile	Phe	Asp	Lys	Ser	Val	Leu	Lys	Lys	Leu	Ser
[0286]																
[0287]	Arg	Arg	His	Tyr	Thr	Gly	Trp	Gly	Lys	Leu	Ser	Ala	Lys	Leu	Ile	Asn
[0288]																
[0289]	Gly	Ile	Arg	Asp	Glu	Lys	Ser	Gly	Asn	Thr	Ile	Leu	Asp	Tyr	Leu	Ile
[0290]																
[0291]	Asp	Asp	Gly	Ile	Ser	Asn	Arg	Asn	Phe	Met	Gln	Leu	Ile	His	Asp	Asp
[0292]																
[0293]	Ala	Leu	Ser	Phe	Lys	Lys	Lys	Ile	Gln	Lys	Ala	Gln	Ile	Ile	Gly	Asp
[0294]																
[0295]	Glu	Asp	Lys	Gly	Asn	Ile	Lys	Glu	Val	Val	Lys	Ser	Leu	Pro	Gly	Ser
[0296]																
[0297]	Pro	Ala	Ile	Lys	Lys	Gly	Ile	Leu	Gln	Ser	Ile	Lys	Ile	Val	Asp	Glu
[0298]																
[0299]	Leu	Val	Lys	Val	Met	Gly	Gly	Arg	Lys	Pro	Glu	Ser	Ile	Val	Val	Glu
[0300]																
[0301]	Met	Ala	Arg	Glu	Asn	Gln	Tyr	Thr	Asn	Gln	Gly	Lys	Ser	Asn	Ser	Gln
[0302]																
[0303]	Gln	Arg	Leu	Lys	Arg	Leu	Glu	Lys	Ser	Leu	Lys	Glu	Leu	Gly	Ser	Lys
[0304]																
[0305]	Ile	Leu	Lys	Glu	Asn	Ile	Pro	Ala	Lys	Leu	Ser	Lys	Ile	Asp	Asn	Asn
[0306]																
[0307]	Ala	Leu	Gln	Asn	Asp	Arg	Leu	Tyr	Leu	Tyr	Tyr	Leu	Gln	Asn	Gly	Lys
[0308]																
[0309]	Asp	Met	Tyr	Thr	Gly	Asp	Asp	Leu	Asp	Ile	Asp	Arg	Leu	Ser	Asn	Tyr
[0310]																
[0311]	Asp	Ile	Asp	His	Ile	Ile	Pro	Gln	Ala	Phe	Leu	Lys	Asp	Asn	Ser	Ile

[0312]	865	870	875	880
[0313]	Asp Asn Lys Val Leu Val Ser Ser Ala Ser Asn Arg Gly Lys Ser Asp			
[0314]		885	890	895
[0315]	Asp Phe Pro Ser Leu Glu Val Val Lys Lys Arg Lys Thr Phe Trp Tyr			
[0316]		900	905	910
[0317]	Gln Leu Leu Lys Ser Lys Leu Ile Ser Gln Arg Lys Phe Asp Asn Leu			
[0318]		915	920	925
[0319]	Thr Lys Ala Glu Arg Gly Gly Leu Leu Pro Glu Asp Lys Ala Gly Phe			
[0320]		930	935	940
[0321]	Ile Gln Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys His Val Ala			
[0322]	945	950	955	960
[0323]	Arg Leu Leu Asp Glu Lys Phe Asn Asn Lys Lys Asp Glu Asn Asn Arg			
[0324]		965	970	975
[0325]	Ala Val Arg Thr Val Lys Ile Ile Thr Leu Lys Ser Thr Leu Val Ser			
[0326]		980	985	990
[0327]	Gln Phe Arg Lys Asp Phe Glu Leu Tyr Lys Val Arg Glu Ile Asn Asp			
[0328]		995	1000	1005
[0329]	Phe His His Ala His Asp Ala Tyr Leu Asn Ala Val Ile Ala Ser			
[0330]		1010	1015	1020
[0331]	Ala Leu Leu Lys Lys Tyr Pro Lys Leu Glu Pro Glu Phe Val Tyr			
[0332]		1025	1030	1035
[0333]	Gly Asp Tyr Pro Lys Tyr Asn Ser Phe Arg Glu Arg Lys Ser Ala			
[0334]		1040	1045	1050
[0335]	Thr Glu Lys Val Tyr Phe Tyr Ser Asn Ile Met Asn Ile Phe Lys			
[0336]		1055	1060	1065
[0337]	Lys Ser Ile Ser Leu Ala Asp Gly Arg Val Ile Glu Arg Pro Leu			
[0338]		1070	1075	1080
[0339]	Ile Glu Val Asn Glu Glu Thr Gly Glu Ser Val Trp Asn Lys Glu			
[0340]		1085	1090	1095
[0341]	Ser Asp Leu Ala Thr Val Arg Arg Val Leu Ser Tyr Pro Gln Val			
[0342]		1100	1105	1110
[0343]	Asn Val Val Lys Lys Val Glu Glu Gln Asn His Gly Leu Asp Arg			
[0344]		1115	1120	1125
[0345]	Gly Lys Pro Lys Gly Leu Phe Asn Ala Asn Leu Ser Ser Lys Pro			
[0346]		1130	1135	1140
[0347]	Lys Pro Asn Ser Asn Glu Asn Leu Val Gly Ala Lys Glu Tyr Leu			
[0348]		1145	1150	1155
[0349]	Asp Pro Lys Lys Tyr Gly Gly Tyr Ala Gly Ile Ser Asn Ser Phe			
[0350]		1160	1165	1170

[0351]	Ala Val Leu Val Lys Gly Thr Ile Glu Lys Gly Ala Lys Lys Lys
[0352]	1175 1180 1185
[0353]	Ile Thr Asn Val Leu Glu Phe Gln Gly Ile Ser Ile Leu Asp Arg
[0354]	1190 1195 1200
[0355]	Ile Asn Tyr Arg Lys Asp Lys Leu Asn Phe Leu Leu Glu Lys Gly
[0356]	1205 1210 1215
[0357]	Tyr Lys Asp Ile Glu Leu Ile Ile Glu Leu Pro Lys Tyr Ser Leu
[0358]	1220 1225 1230
[0359]	Phe Glu Leu Ser Asp Gly Ser Arg Arg Met Leu Ala Ser Ile Leu
[0360]	1235 1240 1245
[0361]	Ser Thr Asn Asn Lys Arg Gly Glu Ile His Lys Gly Asn Gln Ile
[0362]	1250 1255 1260
[0363]	Phe Leu Ser Gln Lys Phe Val Lys Leu Leu Tyr His Ala Lys Arg
[0364]	1265 1270 1275
[0365]	Ile Ser Asn Thr Ile Asn Glu Asn His Arg Lys Tyr Val Glu Asn
[0366]	1280 1285 1290
[0367]	His Lys Lys Glu Phe Glu Glu Leu Phe Tyr Tyr Ile Leu Glu Phe
[0368]	1295 1300 1305
[0369]	Asn Glu Asn Tyr Val Gly Ala Lys Lys Asn Gly Lys Leu Leu Asn
[0370]	1310 1315 1320
[0371]	Ser Ala Phe Gln Ser Trp Gln Asn His Ser Ile Asp Glu Leu Cys
[0372]	1325 1330 1335
[0373]	Ser Ser Phe Ile Gly Pro Thr Gly Ser Glu Arg Lys Gly Leu Phe
[0374]	1340 1345 1350
[0375]	Glu Leu Thr Ser Arg Gly Ser Ala Ala Asp Phe Glu Phe Leu Gly
[0376]	1355 1360 1365
[0377]	Val Lys Ile Pro Arg Tyr Arg Asp Tyr Thr Pro Ser Ser Leu Leu
[0378]	1370 1375 1380
[0379]	Lys Asp Ala Thr Leu Ile His Gln Ser Val Thr Gly Leu Tyr Glu
[0380]	1385 1390 1395
[0381]	Thr Arg Ile Asp Leu Ala Lys Leu Gly Glu Gly
[0382]	1400 1405
[0383]	<210> 3
[0384]	<211> 1409
[0385]	<212> PRT
[0386]	<213> 嗜热链球菌 (Streptococcus thermophilus)
[0387]	<400> 3
[0388]	Met Leu Phe Asn Lys Cys Ile Ile Ile Ser Ile Asn Leu Asp Phe Ser
[0389]	1 5 10 15

[0390]	Asn	Lys	Glu	Lys	Cys	Met	Thr	Lys	Pro	Tyr	Ser	Ile	Gly	Leu	Asp	Ile
[0391]				20					25					30		
[0392]	Gly	Thr	Asn	Ser	Val	Gly	Trp	Ala	Val	Ile	Thr	Asp	Asn	Tyr	Lys	Val
[0393]				35				40					45			
[0394]	Pro	Ser	Lys	Lys	Met	Lys	Val	Leu	Gly	Asn	Thr	Ser	Lys	Lys	Tyr	Ile
[0395]		50					55				60					
[0396]	Lys	Lys	Asn	Leu	Leu	Gly	Val	Leu	Leu	Phe	Asp	Ser	Gly	Ile	Thr	Ala
[0397]	65					70					75					80
[0398]	Glu	Gly	Arg	Arg	Leu	Lys	Arg	Thr	Ala	Arg	Arg	Arg	Tyr	Thr	Arg	Arg
[0399]					85					90					95	
[0400]	Arg	Asn	Arg	Ile	Leu	Tyr	Leu	Gln	Glu	Ile	Phe	Ser	Thr	Glu	Met	Ala
[0401]				100					105					110		
[0402]	Thr	Leu	Asp	Asp	Ala	Phe	Phe	Gln	Arg	Leu	Asp	Asp	Ser	Phe	Leu	Val
[0403]				115				120					125			
[0404]	Pro	Asp	Asp	Lys	Arg	Asp	Ser	Lys	Tyr	Pro	Ile	Phe	Gly	Asn	Leu	Val
[0405]		130					135				140					
[0406]	Glu	Glu	Lys	Val	Tyr	His	Asp	Glu	Phe	Pro	Thr	Ile	Tyr	His	Leu	Arg
[0407]	145					150				155					160	
[0408]	Lys	Tyr	Leu	Ala	Asp	Ser	Thr	Lys	Lys	Ala	Asp	Leu	Arg	Leu	Val	Tyr
[0409]					165					170					175	
[0410]	Leu	Ala	Leu	Ala	His	Met	Ile	Lys	Tyr	Arg	Gly	His	Phe	Leu	Ile	Glu
[0411]				180					185					190		
[0412]	Gly	Glu	Phe	Asn	Ser	Lys	Asn	Asn	Asp	Ile	Gln	Lys	Asn	Phe	Gln	Asp
[0413]			195					200					205			
[0414]	Phe	Leu	Asp	Thr	Tyr	Asn	Ala	Ile	Phe	Glu	Ser	Asp	Leu	Ser	Leu	Glu
[0415]		210					215					220				
[0416]	Asn	Ser	Lys	Gln	Leu	Glu	Glu	Ile	Val	Lys	Asp	Lys	Ile	Ser	Lys	Leu
[0417]	225					230					235				240	
[0418]	Glu	Lys	Lys	Asp	Arg	Ile	Leu	Lys	Leu	Phe	Pro	Gly	Glu	Lys	Asn	Ser
[0419]					245					250					255	
[0420]	Gly	Ile	Phe	Ser	Glu	Phe	Leu	Lys	Leu	Ile	Val	Gly	Asn	Gln	Ala	Asp
[0421]				260					265					270		
[0422]	Phe	Arg	Lys	Cys	Phe	Asn	Leu	Asp	Glu	Lys	Ala	Ser	Leu	His	Phe	Ser
[0423]			275					280					285			
[0424]	Lys	Glu	Ser	Tyr	Asp	Glu	Asp	Leu	Glu	Thr	Leu	Leu	Gly	Tyr	Ile	Gly
[0425]		290					295					300				
[0426]	Asp	Asp	Tyr	Ser	Asp	Val	Phe	Leu	Lys	Ala	Lys	Lys	Leu	Tyr	Asp	Ala
[0427]	305					310					315				320	
[0428]	Ile	Leu	Leu	Ser	Gly	Phe	Leu	Thr	Val	Thr	Asp	Asn	Glu	Thr	Glu	Ala

[0429]	325							330							335				
[0430]	Pro	Leu	Ser	Ser	Ala	Met	Ile	Lys	Arg	Tyr	Asn	Glu	His	Lys	Glu	Asp			
[0431]	340							345							350				
[0432]	Leu	Ala	Leu	Leu	Lys	Glu	Tyr	Ile	Arg	Asn	Ile	Ser	Leu	Lys	Thr	Tyr			
[0433]	355							360							365				
[0434]	Asn	Glu	Val	Phe	Lys	Asp	Asp	Thr	Lys	Asn	Gly	Tyr	Ala	Gly	Tyr	Ile			
[0435]	370							375							380				
[0436]	Asp	Gly	Lys	Thr	Asn	Gln	Glu	Asp	Phe	Tyr	Val	Tyr	Leu	Lys	Asn	Leu			
[0437]	385							390							395			400	
[0438]	Leu	Ala	Glu	Phe	Glu	Gly	Ala	Asp	Tyr	Phe	Leu	Glu	Lys	Ile	Asp	Arg			
[0439]	405							410							415				
[0440]	Glu	Asp	Phe	Leu	Arg	Lys	Gln	Arg	Thr	Phe	Asp	Asn	Gly	Ser	Ile	Pro			
[0441]	420							425							430				
[0442]	Tyr	Gln	Ile	His	Leu	Gln	Glu	Met	Arg	Ala	Ile	Leu	Asp	Lys	Gln	Ala			
[0443]	435							440							445				
[0444]	Lys	Phe	Tyr	Pro	Phe	Leu	Ala	Lys	Asn	Lys	Glu	Arg	Ile	Glu	Lys	Ile			
[0445]	450							455							460				
[0446]	Leu	Thr	Phe	Arg	Ile	Pro	Tyr	Tyr	Val	Gly	Pro	Leu	Ala	Arg	Gly	Asn			
[0447]	465							470							475			480	
[0448]	Ser	Asp	Phe	Ala	Trp	Ser	Ile	Arg	Lys	Arg	Asn	Glu	Lys	Ile	Thr	Pro			
[0449]	485							490							495				
[0450]	Trp	Asn	Phe	Glu	Asp	Val	Ile	Asp	Lys	Glu	Ser	Ser	Ala	Glu	Ala	Phe			
[0451]	500							505							510				
[0452]	Ile	Asn	Arg	Met	Thr	Ser	Phe	Asp	Leu	Tyr	Leu	Pro	Glu	Glu	Lys	Val			
[0453]	515							520							525				
[0454]	Leu	Pro	Lys	His	Ser	Leu	Leu	Tyr	Glu	Thr	Phe	Asn	Val	Tyr	Asn	Glu			
[0455]	530							535							540				
[0456]	Leu	Thr	Lys	Val	Arg	Phe	Ile	Ala	Glu	Ser	Met	Arg	Asp	Tyr	Gln	Phe			
[0457]	545							550							555			560	
[0458]	Leu	Asp	Ser	Lys	Gln	Lys	Lys	Asp	Ile	Val	Arg	Leu	Tyr	Phe	Lys	Asp			
[0459]	565							570							575				
[0460]	Lys	Arg	Lys	Val	Thr	Asp	Lys	Asp	Ile	Ile	Glu	Tyr	Leu	His	Ala	Ile			
[0461]	580							585							590				
[0462]	Tyr	Gly	Tyr	Asp	Gly	Ile	Glu	Leu	Lys	Gly	Ile	Glu	Lys	Gln	Phe	Asn			
[0463]	595							600							605				
[0464]	Ser	Ser	Leu	Ser	Thr	Tyr	His	Asp	Leu	Leu	Asn	Ile	Ile	Asn	Asp	Lys			
[0465]	610							615							620				
[0466]	Glu	Phe	Leu	Asp	Asp	Ser	Ser	Asn	Glu	Ala	Ile	Ile	Glu	Glu	Ile	Ile			
[0467]	625							630							635			640	

[0468]	His	Thr	Leu	Thr	Ile	Phe	Glu	Asp	Arg	Glu	Met	Ile	Lys	Gln	Arg	Leu
[0469]					645					650					655	
[0470]	Ser	Lys	Phe	Glu	Asn	Ile	Phe	Asp	Lys	Ser	Val	Leu	Lys	Lys	Leu	Ser
[0471]					660					665					670	
[0472]	Arg	Arg	His	Tyr	Thr	Gly	Trp	Gly	Lys	Leu	Ser	Ala	Lys	Leu	Ile	Asn
[0473]					675					680					685	
[0474]	Gly	Ile	Arg	Asp	Glu	Lys	Ser	Gly	Asn	Thr	Ile	Leu	Asp	Tyr	Leu	Ile
[0475]					690					695					700	
[0476]	Asp	Asp	Gly	Ile	Ser	Asn	Arg	Asn	Phe	Met	Gln	Leu	Ile	His	Asp	Asp
[0477]	705					710					715					720
[0478]	Ala	Leu	Ser	Phe	Lys	Lys	Lys	Ile	Gln	Lys	Ala	Gln	Ile	Ile	Gly	Asp
[0479]					725						730					735
[0480]	Glu	Asp	Lys	Gly	Asn	Ile	Lys	Glu	Val	Val	Lys	Ser	Leu	Pro	Gly	Ser
[0481]					740						745					750
[0482]	Pro	Ala	Ile	Lys	Lys	Gly	Ile	Leu	Gln	Ser	Ile	Lys	Ile	Val	Asp	Glu
[0483]					755											765
[0484]	Leu	Val	Lys	Val	Met	Gly	Gly	Arg	Lys	Pro	Glu	Ser	Ile	Val	Val	Glu
[0485]					770											780
[0486]	Met	Ala	Arg	Glu	Asn	Gln	Tyr	Thr	Asn	Gln	Gly	Lys	Ser	Asn	Ser	Gln
[0487]	785					790						795				800
[0488]	Gln	Arg	Leu	Lys	Arg	Leu	Glu	Lys	Ser	Leu	Lys	Glu	Leu	Gly	Ser	Lys
[0489]					805						810					815
[0490]	Ile	Leu	Lys	Glu	Asn	Ile	Pro	Ala	Lys	Leu	Ser	Lys	Ile	Asp	Asn	Asn
[0491]					820						825					830
[0492]	Ala	Leu	Gln	Asn	Asp	Arg	Leu	Tyr	Leu	Tyr	Tyr	Leu	Gln	Asn	Gly	Lys
[0493]					835											845
[0494]	Asp	Met	Tyr	Thr	Gly	Asp	Asp	Leu	Asp	Ile	Asp	Arg	Leu	Ser	Asn	Tyr
[0495]					850											860
[0496]	Asp	Ile	Asp	His	Ile	Ile	Pro	Gln	Ala	Phe	Leu	Lys	Asp	Asn	Ser	Ile
[0497]	865					870						875				880
[0498]	Asp	Asn	Lys	Val	Leu	Val	Ser	Ser	Ala	Ser	Ala	Arg	Gly	Lys	Ser	Asp
[0499]					885							890				895
[0500]	Asp	Phe	Pro	Ser	Leu	Glu	Val	Val	Lys	Lys	Arg	Lys	Thr	Phe	Trp	Tyr
[0501]					900											910
[0502]	Gln	Leu	Leu	Lys	Ser	Lys	Leu	Ile	Ser	Gln	Arg	Lys	Phe	Asp	Asn	Leu
[0503]					915											925
[0504]	Thr	Lys	Ala	Glu	Arg	Gly	Gly	Leu	Leu	Pro	Glu	Asp	Lys	Ala	Gly	Phe
[0505]					930											940
[0506]	Ile	Gln	Arg	Gln	Leu	Val	Glu	Thr	Arg	Gln	Ile	Thr	Lys	His	Val	Ala

[0507]	945	950	955	960
[0508]	Arg Leu Leu Asp Glu Lys Phe Asn Asn Lys Lys Asp Glu Asn Asn Arg			
[0509]	965	970	975	
[0510]	Ala Val Arg Thr Val Lys Ile Ile Thr Leu Lys Ser Thr Leu Val Ser			
[0511]	980	985	990	
[0512]	Gln Phe Arg Lys Asp Phe Glu Leu Tyr Lys Val Arg Glu Ile Asn Asp			
[0513]	995	1000	1005	
[0514]	Phe His His Ala His Asp Ala Tyr Leu Asn Ala Val Ile Ala Ser			
[0515]	1010	1015	1020	
[0516]	Ala Leu Leu Lys Lys Tyr Pro Lys Leu Glu Pro Glu Phe Val Tyr			
[0517]	1025	1030	1035	
[0518]	Gly Asp Tyr Pro Lys Tyr Asn Ser Phe Arg Glu Arg Lys Ser Ala			
[0519]	1040	1045	1050	
[0520]	Thr Glu Lys Val Tyr Phe Tyr Ser Asn Ile Met Asn Ile Phe Lys			
[0521]	1055	1060	1065	
[0522]	Lys Ser Ile Ser Leu Ala Asp Gly Arg Val Ile Glu Arg Pro Leu			
[0523]	1070	1075	1080	
[0524]	Ile Glu Val Asn Glu Glu Thr Gly Glu Ser Val Trp Asn Lys Glu			
[0525]	1085	1090	1095	
[0526]	Ser Asp Leu Ala Thr Val Arg Arg Val Leu Ser Tyr Pro Gln Val			
[0527]	1100	1105	1110	
[0528]	Asn Val Val Lys Lys Val Glu Glu Gln Asn His Gly Leu Asp Arg			
[0529]	1115	1120	1125	
[0530]	Gly Lys Pro Lys Gly Leu Phe Asn Ala Asn Leu Ser Ser Lys Pro			
[0531]	1130	1135	1140	
[0532]	Lys Pro Asn Ser Asn Glu Asn Leu Val Gly Ala Lys Glu Tyr Leu			
[0533]	1145	1150	1155	
[0534]	Asp Pro Lys Lys Tyr Gly Gly Tyr Ala Gly Ile Ser Asn Ser Phe			
[0535]	1160	1165	1170	
[0536]	Ala Val Leu Val Lys Gly Thr Ile Glu Lys Gly Ala Lys Lys Lys			
[0537]	1175	1180	1185	
[0538]	Ile Thr Asn Val Leu Glu Phe Gln Gly Ile Ser Ile Leu Asp Arg			
[0539]	1190	1195	1200	
[0540]	Ile Asn Tyr Arg Lys Asp Lys Leu Asn Phe Leu Leu Glu Lys Gly			
[0541]	1205	1210	1215	
[0542]	Tyr Lys Asp Ile Glu Leu Ile Ile Glu Leu Pro Lys Tyr Ser Leu			
[0543]	1220	1225	1230	
[0544]	Phe Glu Leu Ser Asp Gly Ser Arg Arg Met Leu Ala Ser Ile Leu			
[0545]	1235	1240	1245	

[0546]	Ser Thr Asn Asn Lys Arg Gly Glu Ile His Lys Gly Asn Gln Ile
[0547]	1250 1255 1260
[0548]	Phe Leu Ser Gln Lys Phe Val Lys Leu Leu Tyr His Ala Lys Arg
[0549]	1265 1270 1275
[0550]	Ile Ser Asn Thr Ile Asn Glu Asn His Arg Lys Tyr Val Glu Asn
[0551]	1280 1285 1290
[0552]	His Lys Lys Glu Phe Glu Glu Leu Phe Tyr Tyr Ile Leu Glu Phe
[0553]	1295 1300 1305
[0554]	Asn Glu Asn Tyr Val Gly Ala Lys Lys Asn Gly Lys Leu Leu Asn
[0555]	1310 1315 1320
[0556]	Ser Ala Phe Gln Ser Trp Gln Asn His Ser Ile Asp Glu Leu Cys
[0557]	1325 1330 1335
[0558]	Ser Ser Phe Ile Gly Pro Thr Gly Ser Glu Arg Lys Gly Leu Phe
[0559]	1340 1345 1350
[0560]	Glu Leu Thr Ser Arg Gly Ser Ala Ala Asp Phe Glu Phe Leu Gly
[0561]	1355 1360 1365
[0562]	Val Lys Ile Pro Arg Tyr Arg Asp Tyr Thr Pro Ser Ser Leu Leu
[0563]	1370 1375 1380
[0564]	Lys Asp Ala Thr Leu Ile His Gln Ser Val Thr Gly Leu Tyr Glu
[0565]	1385 1390 1395
[0566]	Thr Arg Ile Asp Leu Ala Lys Leu Gly Glu Gly
[0567]	1400 1405
[0568]	<210> 4
[0569]	<211> 1409
[0570]	<212> PRT
[0571]	<213> 嗜热链球菌 (Streptococcus thermophilus)
[0572]	<400> 4
[0573]	Met Leu Phe Asn Lys Cys Ile Ile Ile Ser Ile Asn Leu Asp Phe Ser
[0574]	1 5 10 15
[0575]	Asn Lys Glu Lys Cys Met Thr Lys Pro Tyr Ser Ile Gly Leu Asp Ile
[0576]	20 25 30
[0577]	Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Asn Tyr Lys Val
[0578]	35 40 45
[0579]	Pro Ser Lys Lys Met Lys Val Leu Gly Asn Thr Ser Lys Lys Tyr Ile
[0580]	50 55 60
[0581]	Lys Lys Asn Leu Leu Gly Val Leu Leu Phe Asp Ser Gly Ile Thr Ala
[0582]	65 70 75 80
[0583]	Glu Gly Arg Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg
[0584]	85 90 95

[0585]	Arg	Asn	Arg	Ile	Leu	Tyr	Leu	Gln	Glu	Ile	Phe	Ser	Thr	Glu	Met	Ala
[0586]				100					105					110		
[0587]	Thr	Leu	Asp	Asp	Ala	Phe	Phe	Gln	Arg	Leu	Asp	Asp	Ser	Phe	Leu	Val
[0588]			115					120					125			
[0589]	Pro	Asp	Asp	Lys	Arg	Asp	Ser	Lys	Tyr	Pro	Ile	Phe	Gly	Asn	Leu	Val
[0590]		130						135				140				
[0591]	Glu	Glu	Lys	Val	Tyr	His	Asp	Glu	Phe	Pro	Thr	Ile	Tyr	His	Leu	Arg
[0592]	145					150					155				160	
[0593]	Lys	Tyr	Leu	Ala	Asp	Ser	Thr	Lys	Lys	Ala	Asp	Leu	Arg	Leu	Val	Tyr
[0594]				165					170					175		
[0595]	Leu	Ala	Leu	Ala	His	Met	Ile	Lys	Tyr	Arg	Gly	His	Phe	Leu	Ile	Glu
[0596]			180					185					190			
[0597]	Gly	Glu	Phe	Asn	Ser	Lys	Asn	Asn	Asp	Ile	Gln	Lys	Asn	Phe	Gln	Asp
[0598]		195					200					205				
[0599]	Phe	Leu	Asp	Thr	Tyr	Asn	Ala	Ile	Phe	Glu	Ser	Asp	Leu	Ser	Leu	Glu
[0600]		210					215				220					
[0601]	Asn	Ser	Lys	Gln	Leu	Glu	Glu	Ile	Val	Lys	Asp	Lys	Ile	Ser	Lys	Leu
[0602]	225					230					235				240	
[0603]	Glu	Lys	Lys	Asp	Arg	Ile	Leu	Lys	Leu	Phe	Pro	Gly	Glu	Lys	Asn	Ser
[0604]				245						250				255		
[0605]	Gly	Ile	Phe	Ser	Glu	Phe	Leu	Lys	Leu	Ile	Val	Gly	Asn	Gln	Ala	Asp
[0606]			260						265				270			
[0607]	Phe	Arg	Lys	Cys	Phe	Asn	Leu	Asp	Glu	Lys	Ala	Ser	Leu	His	Phe	Ser
[0608]		275					280					285				
[0609]	Lys	Glu	Ser	Tyr	Asp	Glu	Asp	Leu	Glu	Thr	Leu	Leu	Gly	Tyr	Ile	Gly
[0610]		290					295				300					
[0611]	Asp	Asp	Tyr	Ser	Asp	Val	Phe	Leu	Lys	Ala	Lys	Lys	Leu	Tyr	Asp	Ala
[0612]	305					310				315				320		
[0613]	Ile	Leu	Leu	Ser	Gly	Phe	Leu	Thr	Val	Thr	Asp	Asn	Glu	Thr	Glu	Ala
[0614]				325					330				335			
[0615]	Pro	Leu	Ser	Ser	Ala	Met	Ile	Lys	Arg	Tyr	Asn	Glu	His	Lys	Glu	Asp
[0616]			340						345				350			
[0617]	Leu	Ala	Leu	Leu	Lys	Glu	Tyr	Ile	Arg	Asn	Ile	Ser	Leu	Lys	Thr	Tyr
[0618]		355						360				365				
[0619]	Asn	Glu	Val	Phe	Lys	Asp	Asp	Thr	Lys	Asn	Gly	Tyr	Ala	Gly	Tyr	Ile
[0620]		370					375				380					
[0621]	Asp	Gly	Lys	Thr	Asn	Gln	Glu	Asp	Phe	Tyr	Val	Tyr	Leu	Lys	Asn	Leu
[0622]	385					390				395			400			
[0623]	Leu	Ala	Glu	Phe	Glu	Gly	Ala	Asp	Tyr	Phe	Leu	Glu	Lys	Ile	Asp	Arg

[0624]		405		410		415
[0625]	Glu Asp Phe Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro					
[0626]		420		425		430
[0627]	Tyr Gln Ile His Leu Gln Glu Met Arg Ala Ile Leu Asp Lys Gln Ala					
[0628]		435		440		445
[0629]	Lys Phe Tyr Pro Phe Leu Ala Lys Asn Lys Glu Arg Ile Glu Lys Ile					
[0630]		450		455		460
[0631]	Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn					
[0632]	465		470		475	480
[0633]	Ser Asp Phe Ala Trp Ser Ile Arg Lys Arg Asn Glu Lys Ile Thr Pro					
[0634]		485		490		495
[0635]	Trp Asn Phe Glu Asp Val Ile Asp Lys Glu Ser Ser Ala Glu Ala Phe					
[0636]		500		505		510
[0637]	Ile Asn Arg Met Thr Ser Phe Asp Leu Tyr Leu Pro Glu Glu Lys Val					
[0638]		515		520		525
[0639]	Leu Pro Lys His Ser Leu Leu Tyr Glu Thr Phe Asn Val Tyr Asn Glu					
[0640]		530		535		540
[0641]	Leu Thr Lys Val Arg Phe Ile Ala Glu Ser Met Arg Asp Tyr Gln Phe					
[0642]	545		550		555	560
[0643]	Leu Asp Ser Lys Gln Lys Lys Asp Ile Val Arg Leu Tyr Phe Lys Asp					
[0644]		565		570		575
[0645]	Lys Arg Lys Val Thr Asp Lys Asp Ile Ile Glu Tyr Leu His Ala Ile					
[0646]		580		585		590
[0647]	Tyr Gly Tyr Asp Gly Ile Glu Leu Lys Gly Ile Glu Lys Gln Phe Asn					
[0648]		595		600		605
[0649]	Ser Ser Leu Ser Thr Tyr His Asp Leu Leu Asn Ile Ile Asn Asp Lys					
[0650]		610		615		620
[0651]	Glu Phe Leu Asp Asp Ser Ser Asn Glu Ala Ile Ile Glu Glu Ile Ile					
[0652]	625		630		635	640
[0653]	His Thr Leu Thr Ile Phe Glu Asp Arg Glu Met Ile Lys Gln Arg Leu					
[0654]		645		650		655
[0655]	Ser Lys Phe Glu Asn Ile Phe Asp Lys Ser Val Leu Lys Lys Leu Ser					
[0656]		660		665		670
[0657]	Arg Arg His Tyr Thr Gly Trp Gly Lys Leu Ser Ala Lys Leu Ile Asn					
[0658]		675		680		685
[0659]	Gly Ile Arg Asp Glu Lys Ser Gly Asn Thr Ile Leu Asp Tyr Leu Ile					
[0660]		690		695		700
[0661]	Asp Asp Gly Ile Ser Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp					
[0662]	705		710		715	720

[0663]	Ala	Leu	Ser	Phe	Lys	Lys	Lys	Ile	Gln	Lys	Ala	Gln	Ile	Ile	Gly	Asp
[0664]					725				730						735	
[0665]	Glu	Asp	Lys	Gly	Asn	Ile	Lys	Glu	Val	Val	Lys	Ser	Leu	Pro	Gly	Ser
[0666]					740				745						750	
[0667]	Pro	Ala	Ile	Lys	Lys	Gly	Ile	Leu	Gln	Ser	Ile	Lys	Ile	Val	Asp	Glu
[0668]					755				760						765	
[0669]	Leu	Val	Lys	Val	Met	Gly	Gly	Arg	Lys	Pro	Glu	Ser	Ile	Val	Val	Glu
[0670]					770				775						780	
[0671]	Met	Ala	Arg	Glu	Asn	Gln	Tyr	Thr	Asn	Gln	Gly	Lys	Ser	Asn	Ser	Gln
[0672]	785					790					795					800
[0673]	Gln	Arg	Leu	Lys	Arg	Leu	Glu	Lys	Ser	Leu	Lys	Glu	Leu	Gly	Ser	Lys
[0674]					805					810						815
[0675]	Ile	Leu	Lys	Glu	Asn	Ile	Pro	Ala	Lys	Leu	Ser	Lys	Ile	Asp	Asn	Asn
[0676]					820					825					830	
[0677]	Ala	Leu	Gln	Asn	Asp	Arg	Leu	Tyr	Leu	Tyr	Tyr	Leu	Gln	Asn	Gly	Lys
[0678]					835					840					845	
[0679]	Asp	Met	Tyr	Thr	Gly	Asp	Asp	Leu	Asp	Ile	Asp	Arg	Leu	Ser	Asn	Tyr
[0680]					850					855					860	
[0681]	Asp	Ile	Asp	Ala	Ile	Ile	Pro	Gln	Ala	Phe	Leu	Lys	Asp	Asn	Ser	Ile
[0682]	865					870					875					880
[0683]	Asp	Asn	Lys	Val	Leu	Val	Ser	Ser	Ala	Ser	Asn	Arg	Gly	Lys	Ser	Asp
[0684]					885					890						895
[0685]	Asp	Phe	Pro	Ser	Leu	Glu	Val	Val	Lys	Lys	Arg	Lys	Thr	Phe	Trp	Tyr
[0686]					900					905					910	
[0687]	Gln	Leu	Leu	Lys	Ser	Lys	Leu	Ile	Ser	Gln	Arg	Lys	Phe	Asp	Asn	Leu
[0688]					915					920					925	
[0689]	Thr	Lys	Ala	Glu	Arg	Gly	Gly	Leu	Leu	Pro	Glu	Asp	Lys	Ala	Gly	Phe
[0690]					930					935					940	
[0691]	Ile	Gln	Arg	Gln	Leu	Val	Glu	Thr	Arg	Gln	Ile	Thr	Lys	His	Val	Ala
[0692]	945					950					955					960
[0693]	Arg	Leu	Leu	Asp	Glu	Lys	Phe	Asn	Asn	Lys	Lys	Asp	Glu	Asn	Asn	Arg
[0694]					965					970						975
[0695]	Ala	Val	Arg	Thr	Val	Lys	Ile	Ile	Thr	Leu	Lys	Ser	Thr	Leu	Val	Ser
[0696]					980					985					990	
[0697]	Gln	Phe	Arg	Lys	Asp	Phe	Glu	Leu	Tyr	Lys	Val	Arg	Glu	Ile	Asn	Asp
[0698]					995					1000					1005	
[0699]	Phe	His	His	Ala	His	Asp	Ala	Tyr	Leu	Asn	Ala	Val	Ile	Ala	Ser	
[0700]					1010					1015					1020	
[0701]	Ala	Leu	Leu	Lys	Lys	Tyr	Pro	Lys	Leu	Glu	Pro	Glu	Phe	Val	Tyr	

[0702]	1025	1030	1035
[0703]	Gly Asp Tyr Pro Lys Tyr Asn Ser Phe Arg Glu Arg Lys Ser Ala		
[0704]	1040	1045	1050
[0705]	Thr Glu Lys Val Tyr Phe Tyr Ser Asn Ile Met Asn Ile Phe Lys		
[0706]	1055	1060	1065
[0707]	Lys Ser Ile Ser Leu Ala Asp Gly Arg Val Ile Glu Arg Pro Leu		
[0708]	1070	1075	1080
[0709]	Ile Glu Val Asn Glu Glu Thr Gly Glu Ser Val Trp Asn Lys Glu		
[0710]	1085	1090	1095
[0711]	Ser Asp Leu Ala Thr Val Arg Arg Val Leu Ser Tyr Pro Gln Val		
[0712]	1100	1105	1110
[0713]	Asn Val Val Lys Lys Val Glu Glu Gln Asn His Gly Leu Asp Arg		
[0714]	1115	1120	1125
[0715]	Gly Lys Pro Lys Gly Leu Phe Asn Ala Asn Leu Ser Ser Lys Pro		
[0716]	1130	1135	1140
[0717]	Lys Pro Asn Ser Asn Glu Asn Leu Val Gly Ala Lys Glu Tyr Leu		
[0718]	1145	1150	1155
[0719]	Asp Pro Lys Lys Tyr Gly Gly Tyr Ala Gly Ile Ser Asn Ser Phe		
[0720]	1160	1165	1170
[0721]	Ala Val Leu Val Lys Gly Thr Ile Glu Lys Gly Ala Lys Lys Lys		
[0722]	1175	1180	1185
[0723]	Ile Thr Asn Val Leu Glu Phe Gln Gly Ile Ser Ile Leu Asp Arg		
[0724]	1190	1195	1200
[0725]	Ile Asn Tyr Arg Lys Asp Lys Leu Asn Phe Leu Leu Glu Lys Gly		
[0726]	1205	1210	1215
[0727]	Tyr Lys Asp Ile Glu Leu Ile Ile Glu Leu Pro Lys Tyr Ser Leu		
[0728]	1220	1225	1230
[0729]	Phe Glu Leu Ser Asp Gly Ser Arg Arg Met Leu Ala Ser Ile Leu		
[0730]	1235	1240	1245
[0731]	Ser Thr Asn Asn Lys Arg Gly Glu Ile His Lys Gly Asn Gln Ile		
[0732]	1250	1255	1260
[0733]	Phe Leu Ser Gln Lys Phe Val Lys Leu Leu Tyr His Ala Lys Arg		
[0734]	1265	1270	1275
[0735]	Ile Ser Asn Thr Ile Asn Glu Asn His Arg Lys Tyr Val Glu Asn		
[0736]	1280	1285	1290
[0737]	His Lys Lys Glu Phe Glu Glu Leu Phe Tyr Tyr Ile Leu Glu Phe		
[0738]	1295	1300	1305
[0739]	Asn Glu Asn Tyr Val Gly Ala Lys Lys Asn Gly Lys Leu Leu Asn		
[0740]	1310	1315	1320

[0741]	Ser Ala Phe Gln Ser Trp Gln Asn His Ser Ile Asp Glu Leu Cys	
[0742]	1325 1330 1335	
[0743]	Ser Ser Phe Ile Gly Pro Thr Gly Ser Glu Arg Lys Gly Leu Phe	
[0744]	1340 1345 1350	
[0745]	Glu Leu Thr Ser Arg Gly Ser Ala Ala Asp Phe Glu Phe Leu Gly	
[0746]	1355 1360 1365	
[0747]	Val Lys Ile Pro Arg Tyr Arg Asp Tyr Thr Pro Ser Ser Leu Leu	
[0748]	1370 1375 1380	
[0749]	Lys Asp Ala Thr Leu Ile His Gln Ser Val Thr Gly Leu Tyr Glu	
[0750]	1385 1390 1395	
[0751]	Thr Arg Ile Asp Leu Ala Lys Leu Gly Glu Gly	
[0752]	1400 1405	
[0753]	<210> 5	
[0754]	<211> 102	
[0755]	<212> RNA	
[0756]	<213> 嗜热链球菌 (Streptococcus thermophilus)	
[0757]	<400> 5	
[0758]	uaauaauaau ugugguuuga aaccuucga aacaacacag cgaguuaaaa uaaggcuuag	60
[0759]	uccguacuca acuugaaaag guggcaccga uucgguguuu uu	102
[0760]	<210> 6	
[0761]	<211> 78	
[0762]	<212> RNA	
[0763]	<213> 嗜热链球菌 (Streptococcus thermophilus)	
[0764]	<400> 6	
[0765]	gggcgaaaca acacagcgag uaaaaauaag gcuuaguccg uacucaacuu gaaaaggugg	60
[0766]	caccgauucg guguuuuu	78
[0767]	<210> 7	
[0768]	<211> 42	
[0769]	<212> RNA	
[0770]	<213> 人工序列	
[0771]	<220>	
[0772]	<221> 来源	
[0773]	<223> /note="人工序列的描述：合成寡核苷酸"	
[0774]	<400> 7	
[0775]	cgcuaaagag gaagaggaca guuuuagagc uguguuguuu cg	42
[0776]	<210> 8	
[0777]	<211> 150	
[0778]	<212> RNA	
[0779]	<213> 人工序列	

[0780]	<220>	
[0781]	<221> 来源	
[0782]	<223> /note="人工序列的描述: 合成多肽"	
[0783]	<400> 8	
[0784]	ggguagaaaa gauauccuac gagguuuuag agcuguguug uuucgaaugg uuccaaaaca	60
[0785]	aaauucuaaac gcuaaagagg aagaggacag uuuuagagcu guguuguuuc gaaugguucc	120
[0786]	aaaacuacug cuguauuagc uugguuguug	150
[0787]	<210> 9	
[0788]	<211> 150	
[0789]	<212> DNA	
[0790]	<213> 人工序列	
[0791]	<220>	
[0792]	<221> 来源	
[0793]	<223> /note="人工序列的描述: 合成寡核苷酸"	
[0794]	<220>	
[0795]	<221> 来源	
[0796]	<223> /note="合并的DNA/RNA分子的描述: 合成寡核苷酸"	
[0797]	<400> 9	
[0798]	ggguagaaaa gauauccuac gagguuuuag agcuguguug uuucgaaugg uuccaaaact	60
[0799]	gtcatgataa taatggtttc ttagacgtcg uuuuagagcu guguuguuuc gaaugguucc	120
[0800]	aaaacuacug cuguauuagc uugguuguug	150
[0801]	<210> 10	
[0802]	<211> 150	
[0803]	<212> DNA	
[0804]	<213> 人工序列	
[0805]	<220>	
[0806]	<221> 来源	
[0807]	<223> /note="人工序列的描述: 合成寡核苷酸"	
[0808]	<220>	
[0809]	<221> 来源	
[0810]	<223> /note="合并的DNA/RNA分子的描述: 合成寡核苷酸"	
[0811]	<400> 10	
[0812]	ggguagaaaa gauauccuac gagguuuuag agcuguguug uuucgaaugg uuccaaaaca	60
[0813]	cgagccggaa gcataaagtg taaagcctgg uuuuagagcu guguuguuuc gaaugguucc	120
[0814]	aaaacuacug cuguauuagc uugguuguug	150
[0815]	<210> 11	
[0816]	<211> 150	
[0817]	<212> DNA	
[0818]	<213> 人工序列	

[0819]	<220>	
[0820]	<221> 来源	
[0821]	<223> /note="人工序列的描述：合成寡核苷酸"	
[0822]	<220>	
[0823]	<221> 来源	
[0824]	<223> /note="合并的DNA/RNA分子的描述：合成寡核苷酸"	
[0825]	<400> 11	
[0826]	ggguagaaaa gauauccuac gagguuuuag agcuguguug uuucgaaugg uuccaaaact	60
[0827]	caaggagaaa tagaggctct cgttgcatg uuuuagagcu guguuguuuc gaaugguucc	120
[0828]	aaaacuacug cuguauuagc uugguuguug	150
[0829]	<210> 12	
[0830]	<211> 150	
[0831]	<212> DNA	
[0832]	<213> 人工序列	
[0833]	<220>	
[0834]	<221> 来源	
[0835]	<223> /note="人工序列的描述：合成寡核苷酸"	
[0836]	<220>	
[0837]	<221> 来源	
[0838]	<223> /note="合并的DNA/RNA分子的描述：合成寡核苷酸"	
[0839]	<400> 12	
[0840]	ggguagaaaa gauauccuac gagguuuuag agcuguguug uuucgaaugg uuccaaaacc	60
[0841]	gggaggaag ctgcatgatg cgatgttatg uuuuagagcu guguuguuuc gaaugguucc	120
[0842]	aaaacuacug cuguauuagc uugguuguug	150
[0843]	<210> 13	
[0844]	<211> 42	
[0845]	<212> RNA	
[0846]	<213> 人工序列	
[0847]	<220>	
[0848]	<221> 来源	
[0849]	<223> /note="人工序列的描述：合成寡核苷酸"	
[0850]	<400> 13	
[0851]	gcucccgagg cucgaugaag guuuuagagc uguguuguu cg	42
[0852]	<210> 14	
[0853]	<211> 42	
[0854]	<212> RNA	
[0855]	<213> 人工序列	
[0856]	<220>	
[0857]	<221> 来源	

[0858]	<223> /note="人工序列的描述：合成寡核苷酸"	
[0859]	<400> 14	
[0860]	ugaaucguga aaucugcuca guuuuagagc uguguuguuu cg	42
[0861]	<210> 15	
[0862]	<211> 42	
[0863]	<212> RNA	
[0864]	<213> 人工序列	
[0865]	<220>	
[0866]	<221> 来源	
[0867]	<223> /note="人工序列的描述：合成寡核苷酸"	
[0868]	<220>	
[0869]	<221> modified_base	
[0870]	<222> (1) .. (20)	
[0871]	<223> a, c, u, g, 未知或其他	
[0872]	<400> 15	
[0873]	nnnnnnnnnnn nnnnnnnnnn guuuuagagc uguguuguuu cg	42
[0874]	<210> 16	
[0875]	<211> 29	
[0876]	<212> DNA	
[0877]	<213> 人工序列	
[0878]	<220>	
[0879]	<221> 来源	
[0880]	<223> /note="人工序列的描述：合成寡核苷酸"	
[0881]	<400> 16	
[0882]	ccaccagca aaattcggtt ttctggctg	29
[0883]	<210> 17	
[0884]	<211> 36	
[0885]	<212> DNA	
[0886]	<213> 人工序列	
[0887]	<220>	
[0888]	<221> 来源	
[0889]	<223> /note="人工序列的描述：合成引物"	
[0890]	<400> 17	
[0891]	taatacgact cactataggg taccgagctc gaattg	36
[0892]	<210> 18	
[0893]	<211> 31	
[0894]	<212> DNA	
[0895]	<213> 人工序列	
[0896]	<220>	

[0897]	<221> 来源	
[0898]	<223> /note="人工序列的描述: 合成引物"	
[0899]	<400> 18	
[0900]	gggaaacagc tatgaccatg attacgaatt c	31
[0901]	<210> 19	
[0902]	<211> 31	
[0903]	<212> DNA	
[0904]	<213> 人工序列	
[0905]	<220>	
[0906]	<221> 来源	
[0907]	<223> /note="人工序列的描述: 合成引物"	
[0908]	<400> 19	
[0909]	gggtaccgag ctcgattga aattctaaac g	31
[0910]	<210> 20	
[0911]	<211> 43	
[0912]	<212> DNA	
[0913]	<213> 人工序列	
[0914]	<220>	
[0915]	<221> 来源	
[0916]	<223> /note="人工序列的描述: 合成引物"	
[0917]	<400> 20	
[0918]	taatacgact cactataggg aaacagctat gaccatgatt acg	43
[0919]	<210> 21	
[0920]	<211> 38	
[0921]	<212> DNA	
[0922]	<213> 人工序列	
[0923]	<220>	
[0924]	<221> 来源	
[0925]	<223> /note="人工序列的描述: 合成引物"	
[0926]	<400> 21	
[0927]	acgtctcaaa tgttgtttaa taagtgtata ataatttc	38
[0928]	<210> 22	
[0929]	<211> 29	
[0930]	<212> DNA	
[0931]	<213> 人工序列	
[0932]	<220>	
[0933]	<221> 来源	
[0934]	<223> /note="人工序列的描述: 合成引物"	
[0935]	<400> 22	

[0936]	acgtctccgc gctaccctct cctagtttg	29
[0937]	<210> 23	
[0938]	<211> 6	
[0939]	<212> PRT	
[0940]	<213> 人工序列	
[0941]	<220>	
[0942]	<221> 来源	
[0943]	<223> /note="人工序列的描述: 合成6xHis标签"	
[0944]	<400> 23	
[0945]	His His His His His His	
[0946]	1 5	
[0947]	<210> 24	
[0948]	<211> 34	
[0949]	<212> DNA	
[0950]	<213> 人工序列	
[0951]	<220>	
[0952]	<221> 来源	
[0953]	<223> /note="人工序列的描述: 合成引物"	
[0954]	<400> 24	
[0955]	acgtctcaca tgactaagcc atactcaatt ggac	34
[0956]	<210> 25	
[0957]	<211> 27	
[0958]	<212> DNA	
[0959]	<213> 人工序列	
[0960]	<220>	
[0961]	<221> 来源	
[0962]	<223> /note="人工序列的描述: 合成引物"	
[0963]	<400> 25	
[0964]	actcgagacc ctctcctagt ttggcaa	27
[0965]	<210> 26	
[0966]	<211> 23	
[0967]	<212> DNA	
[0968]	<213> 人工序列	
[0969]	<220>	
[0970]	<221> 来源	
[0971]	<223> /note="人工序列的描述: 合成引物"	
[0972]	<400> 26	
[0973]	gaccacttat tgaggtaaat gag	23
[0974]	<210> 27	

[0975]	<211> 28	
[0976]	<212> DNA	
[0977]	<213> 人工序列	
[0978]	<220>	
[0979]	<221> 来源	
[0980]	<223> /note="人工序列的描述：合成引物"	
[0981]	<400> 27	
[0982]	caaaccagga tccaagctaa tacagcag	28
[0983]	<210> 28	
[0984]	<211> 42	
[0985]	<212> DNA	
[0986]	<213> 人工序列	
[0987]	<220>	
[0988]	<221> 来源	
[0989]	<223> /note="人工序列的描述：合成寡核苷酸"	
[0990]	<400> 28	
[0991]	tcgaaacaac acagctctaa aactgtcctc ttcctcttta gc	42
[0992]	<210> 29	
[0993]	<211> 21	
[0994]	<212> DNA	
[0995]	<213> 人工序列	
[0996]	<220>	
[0997]	<221> 来源	
[0998]	<223> /note="人工序列的描述：合成引物"	
[0999]	<400> 29	
[1000]	ccgcatcagg cgccattcgc c	21
[1001]	<210> 30	
[1002]	<211> 22	
[1003]	<212> DNA	
[1004]	<213> 人工序列	
[1005]	<220>	
[1006]	<221> 来源	
[1007]	<223> /note="人工序列的描述：合成引物"	
[1008]	<400> 30	
[1009]	gcgaggaagc ggaagagcgc cc	22
[1010]	<210> 31	
[1011]	<211> 55	
[1012]	<212> DNA	
[1013]	<213> 人工序列	

[1014]	<220>	
[1015]	<221> 来源	
[1016]	<223> /note="人工序列的描述：合成寡核苷酸"	
[1017]	<400> 31	
[1018]	gctcgaattg aaattctaaa cgctaaagag gaagaggaca tggatgaattc gtaat	55
[1019]	<210> 32	
[1020]	<211> 50	
[1021]	<212> DNA	
[1022]	<213> 人工序列	
[1023]	<220>	
[1024]	<221> 来源	
[1025]	<223> /note="人工序列的描述：合成寡核苷酸"	
[1026]	<400> 32	
[1027]	gctcgaattg aaattctaaa cgctaaagag gaagaggaca aattcgtaat	50
[1028]	<210> 33	
[1029]	<211> 55	
[1030]	<212> DNA	
[1031]	<213> 人工序列	
[1032]	<220>	
[1033]	<221> 来源	
[1034]	<223> /note="人工序列的描述：合成寡核苷酸"	
[1035]	<400> 33	
[1036]	gctcgaattg tactgctgta ttagcttggt tgttggtttg tggatgaattc gtaat	55
[1037]	<210> 34	
[1038]	<211> 55	
[1039]	<212> DNA	
[1040]	<213> 人工序列	
[1041]	<220>	
[1042]	<221> 来源	
[1043]	<223> /note="人工序列的描述：合成寡核苷酸"	
[1044]	<400> 34	
[1045]	attacgaatt caccatgtcc tcttcctctt tagcgtttag aatttcaatt cgagc	55
[1046]	<210> 35	
[1047]	<211> 50	
[1048]	<212> DNA	
[1049]	<213> 人工序列	
[1050]	<220>	
[1051]	<221> 来源	
[1052]	<223> /note="人工序列的描述：合成寡核苷酸"	

[1053]	<400> 35	
[1054]	attacgaatt tgtcctcttc ctcttttagcg tttagaattt caattcgagc	50
[1055]	<210> 36	
[1056]	<211> 55	
[1057]	<212> DNA	
[1058]	<213> 人工序列	
[1059]	<220>	
[1060]	<221> 来源	
[1061]	<223> /note="人工序列的描述: 合成寡核苷酸"	
[1062]	<400> 36	
[1063]	attacgaatt caccacaaac caacaaccaa gctaatacag cagtacaatt cgagc	55
[1064]	<210> 37	
[1065]	<211> 55	
[1066]	<212> DNA	
[1067]	<213> 人工序列	
[1068]	<220>	
[1069]	<221> 来源	
[1070]	<223> /note="人工序列的描述: 合成寡核苷酸"	
[1071]	<400> 37	
[1072]	gctcgaattg aaattctaaa cgctaaagag gaagaggaca tggatgaattc gtaat	55
[1073]	<210> 38	
[1074]	<211> 45	
[1075]	<212> DNA	
[1076]	<213> 人工序列	
[1077]	<220>	
[1078]	<221> 来源	
[1079]	<223> /note="人工序列的描述: 合成寡核苷酸"	
[1080]	<400> 38	
[1081]	gctcgaattg cgctaaagag gaagaggaca tggatgaattc gtaat	45
[1082]	<210> 39	
[1083]	<211> 55	
[1084]	<212> DNA	
[1085]	<213> 人工序列	
[1086]	<220>	
[1087]	<221> 来源	
[1088]	<223> /note="人工序列的描述: 合成寡核苷酸"	
[1089]	<400> 39	
[1090]	gctcgaattg ccaccagca aaattcggtt ttctggctga tggatgaattc gtaat	55
[1091]	<210> 40	

[1092]	<211> 41	
[1093]	<212> DNA	
[1094]	<213> 人工序列	
[1095]	<220>	
[1096]	<221> 来源	
[1097]	<223> /note="人工序列的描述: 合成引物"	
[1098]	<400> 40	
[1099]	taatacgact cactataggg tagaaaagat atcctacgag g	41
[1100]	<210> 41	
[1101]	<211> 23	
[1102]	<212> DNA	
[1103]	<213> 人工序列	
[1104]	<220>	
[1105]	<221> 来源	
[1106]	<223> /note="人工序列的描述: 合成引物"	
[1107]	<400> 41	
[1108]	caacaaccaa gctaatacag cag	23
[1109]	<210> 42	
[1110]	<211> 22	
[1111]	<212> DNA	
[1112]	<213> 人工序列	
[1113]	<220>	
[1114]	<221> 来源	
[1115]	<223> /note="人工序列的描述: 合成引物"	
[1116]	<400> 42	
[1117]	aaaaacaccg aatcggtgcc ac	22
[1118]	<210> 43	
[1119]	<211> 48	
[1120]	<212> DNA	
[1121]	<213> 人工序列	
[1122]	<220>	
[1123]	<221> 来源	
[1124]	<223> /note="人工序列的描述: 合成引物"	
[1125]	<400> 43	
[1126]	taatacgact cactataggg taataataat tgtggtttga aaccattc	48
[1127]	<210> 44	
[1128]	<211> 73	
[1129]	<212> RNA	
[1130]	<213> 嗜热链球菌 (Streptococcus thermophilus)	

[1131]	<400> 44	
[1132]	gggcgaaaca acacagcgag uaaaaauaag gcuuaguccg uacucaacuu gaaaaggugg	60
[1133]	caccgauucg gug	73
[1134]	<210> 45	
[1135]	<211> 68	
[1136]	<212> RNA	
[1137]	<213> 嗜热链球菌 (Streptococcus thermophilus)	
[1138]	<400> 45	
[1139]	gggcgaaaca acacagcgag uaaaaauaag gcuuaguccg uacucaacuu gaaaaggugg	60
[1140]	caccgauu	68
[1141]	<210> 46	
[1142]	<211> 63	
[1143]	<212> RNA	
[1144]	<213> 嗜热链球菌 (Streptococcus thermophilus)	
[1145]	<400> 46	
[1146]	gggcgaaaca acacagcgag uaaaaauaag gcuuaguccg uacucaacuu gaaaaggugg	60
[1147]	cac	63
[1148]	<210> 47	
[1149]	<211> 58	
[1150]	<212> RNA	
[1151]	<213> 嗜热链球菌 (Streptococcus thermophilus)	
[1152]	<400> 47	
[1153]	gggcgaaaca acacagcgag uaaaaauaag gcuuaguccg uacucaacuu gaaaaggu	58
[1154]	<210> 48	
[1155]	<211> 53	
[1156]	<212> RNA	
[1157]	<213> 嗜热链球菌 (Streptococcus thermophilus)	
[1158]	<400> 48	
[1159]	gggcgaaaca acacagcgag uaaaaauaag gcuuaguccg uacucaacuu gaa	53
[1160]	<210> 49	
[1161]	<211> 48	
[1162]	<212> RNA	
[1163]	<213> 嗜热链球菌 (Streptococcus thermophilus)	
[1164]	<400> 49	
[1165]	gggcgaaaca acacagcgag uaaaaauaag gcuuaguccg uacucaac	48
[1166]	<210> 50	
[1167]	<211> 36	
[1168]	<212> DNA	
[1169]	<213> 嗜热链球菌 (Streptococcus thermophilus)	

[1170]	<400> 50	
[1171]	gttttagagc tgtgttggtt cgaatggttc caaaac	36
[1172]	<210> 51	
[1173]	<211> 36	
[1174]	<212> DNA	
[1175]	<213> 化脓链球菌 (Streptococcus pyrogenes)	
[1176]	<400> 51	
[1177]	gttttagagc tatgctgttt tgaatggtcc caaaac	36
[1178]	<210> 52	
[1179]	<211> 36	
[1180]	<212> DNA	
[1181]	<213> 嗜热链球菌 (Streptococcus thermophilus)	
[1182]	<400> 52	
[1183]	tttaactcgc tgtgttggtt cgaatggttt caaac	36
[1184]	<210> 53	
[1185]	<211> 36	
[1186]	<212> DNA	
[1187]	<213> 化脓链球菌 (Streptococcus pyrogenes)	
[1188]	<400> 53	
[1189]	tttaacttgc tatgctgttt tgaatggttc caacaa	36
[1190]	<210> 54	
[1191]	<211> 42	
[1192]	<212> RNA	
[1193]	<213> 化脓链球菌 (Streptococcus pyrogenes)	
[1194]	<400> 54	
[1195]	gauuucuucu ugcgcuuuuu guuuuagagc uaugcuguuu ug	42
[1196]	<210> 55	
[1197]	<211> 42	
[1198]	<212> RNA	
[1199]	<213> 嗜热链球菌 (Streptococcus thermophilus)	
[1200]	<400> 55	
[1201]	guucacugua cgaguacuua guuuuagagc uguguuguuu cg	42
[1202]	<210> 56	
[1203]	<211> 42	
[1204]	<212> RNA	
[1205]	<213> 嗜热链球菌 (Streptococcus thermophilus)	
[1206]	<400> 56	
[1207]	cgcuaaagag gaagaggaca guuuuagagc uguguuguuu cg	42
[1208]	<210> 57	

[1209]	<211> 35	
[1210]	<212> DNA	
[1211]	<213> 人工序列	
[1212]	<220>	
[1213]	<221> 来源	
[1214]	<223> /note="人工序列的描述：合成寡核苷酸"	
[1215]	<400> 57	
[1216]	aaattctaaa cgctaaagag gaagaggaca tgggtg	35
[1217]	<210> 58	
[1218]	<211> 18	
[1219]	<212> DNA	
[1220]	<213> 人工序列	
[1221]	<220>	
[1222]	<221> 来源	
[1223]	<223> /note="人工序列的描述：合成寡核苷酸"	
[1224]	<400> 58	
[1225]	aggaagagga catggtga	18
[1226]	<210> 59	
[1227]	<211> 10	
[1228]	<212> DNA	
[1229]	<213> 人工序列	
[1230]	<220>	
[1231]	<221> 来源	
[1232]	<223> /note="人工序列的描述：合成寡核苷酸"	
[1233]	<400> 59	
[1234]	aggaagagga	10
[1235]	<210> 60	
[1236]	<211> 10	
[1237]	<212> DNA	
[1238]	<213> 人工序列	
[1239]	<220>	
[1240]	<221> 来源	
[1241]	<223> /note="人工序列的描述：合成寡核苷酸"	
[1242]	<400> 60	
[1243]	tacatggtga	10
[1244]	<210> 61	
[1245]	<211> 35	
[1246]	<212> DNA	
[1247]	<213> 人工序列	

[1248]	<220>	
[1249]	<221> 来源	
[1250]	<223> /note="人工序列的描述：合成寡核苷酸"	
[1251]	<400> 61	
[1252]	caccatgtcc tcttcctctt tagcgtttag aattt	35
[1253]	<210> 62	
[1254]	<211> 32	
[1255]	<212> DNA	
[1256]	<213> 人工序列	
[1257]	<220>	
[1258]	<221> 来源	
[1259]	<223> /note="人工序列的描述：合成寡核苷酸"	
[1260]	<400> 62	
[1261]	caccatgtcc tcttcctctt tagcgtttag aa	32
[1262]	<210> 63	
[1263]	<211> 28	
[1264]	<212> DNA	
[1265]	<213> 人工序列	
[1266]	<220>	
[1267]	<221> 来源	
[1268]	<223> /note="人工序列的描述：合成寡核苷酸"	
[1269]	<400> 63	
[1270]	caccatgtcc tcttcctctt tagcgttt	28
[1271]	<210> 64	
[1272]	<211> 24	
[1273]	<212> DNA	
[1274]	<213> 人工序列	
[1275]	<220>	
[1276]	<221> 来源	
[1277]	<223> /note="人工序列的描述：合成寡核苷酸"	
[1278]	<400> 64	
[1279]	caccatgtcc tcttcctctt tagc	24
[1280]	<210> 65	
[1281]	<211> 20	
[1282]	<212> DNA	
[1283]	<213> 人工序列	
[1284]	<220>	
[1285]	<221> 来源	
[1286]	<223> /note="人工序列的描述：合成寡核苷酸"	

[1287]	<400> 65	
[1288]	caccatgtcc tcttcctctt	20
[1289]	<210> 66	
[1290]	<211> 16	
[1291]	<212> DNA	
[1292]	<213> 人工序列	
[1293]	<220>	
[1294]	<221> 来源	
[1295]	<223> /note="人工序列的描述: 合成寡核苷酸"	
[1296]	<400> 66	
[1297]	caccatgtcc tcttcc	16
[1298]	<210> 67	
[1299]	<211> 55	
[1300]	<212> DNA	
[1301]	<213> 人工序列	
[1302]	<220>	
[1303]	<221> 来源	
[1304]	<223> /note="人工序列的描述: 合成寡核苷酸"	
[1305]	<400> 67	
[1306]	attacgaatt ctccttgtcc tcttcctctt tagcgtttag aatttcaatt cgagc	55
[1307]	<210> 68	
[1308]	<211> 55	
[1309]	<212> DNA	
[1310]	<213> 人工序列	
[1311]	<220>	
[1312]	<221> 来源	
[1313]	<223> /note="人工序列的描述: 合成寡核苷酸"	
[1314]	<400> 68	
[1315]	attacgaatt gtggttgtcc tcttcctctt tagcgtttag aatttcaatt cgagc	55
[1316]	<210> 69	
[1317]	<211> 42	
[1318]	<212> RNA	
[1319]	<213> 人工序列	
[1320]	<220>	
[1321]	<221> 来源	
[1322]	<223> /note="人工序列的描述: 合成寡核苷酸"	
[1323]	<400> 69	
[1324]	gcuaaagagg aagaggacag uuuuagagcu guguuguuuc ga	42

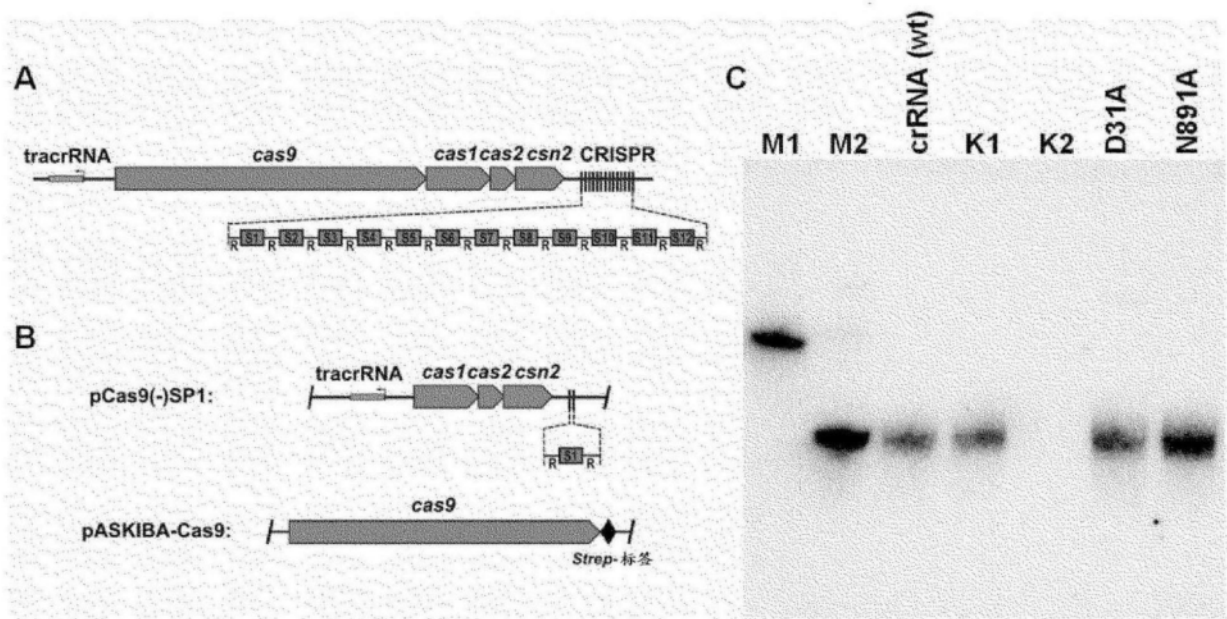


图1

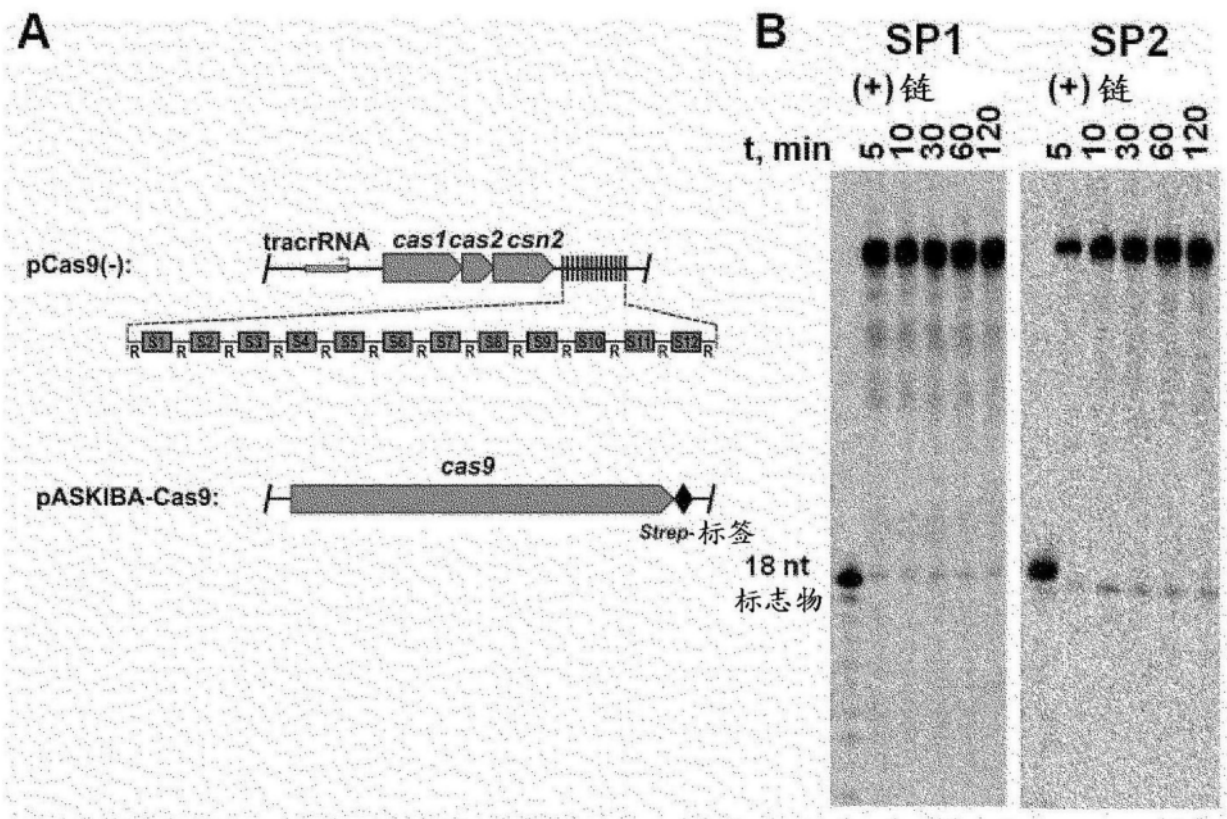


图2

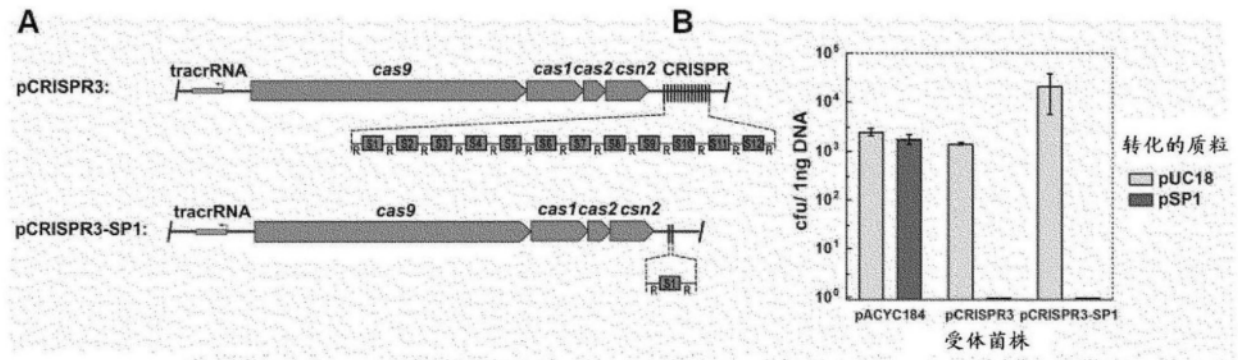


图3

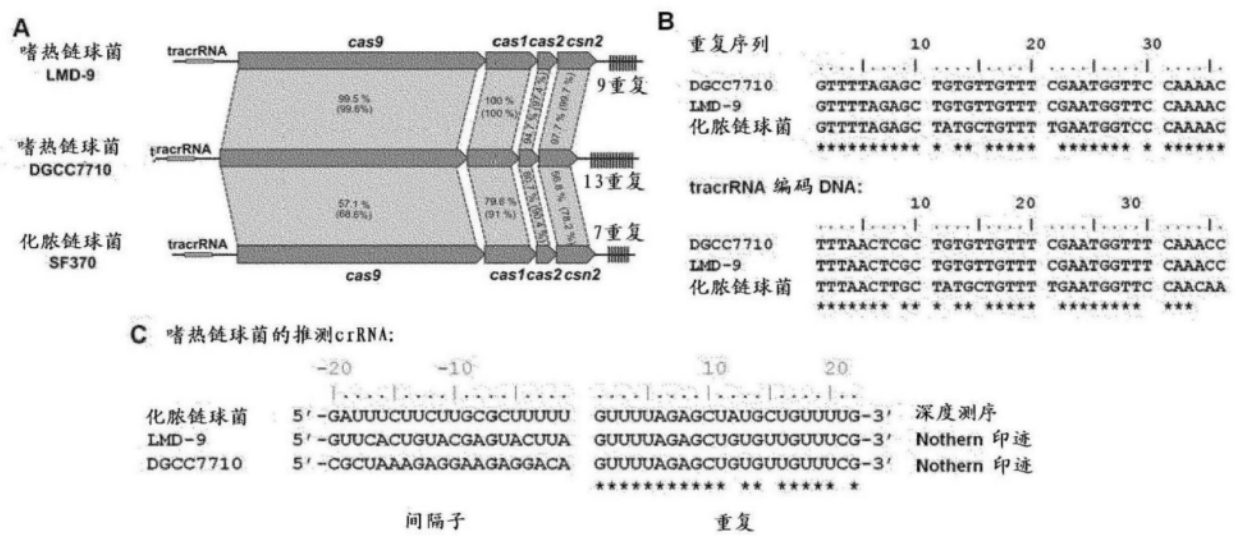


图4

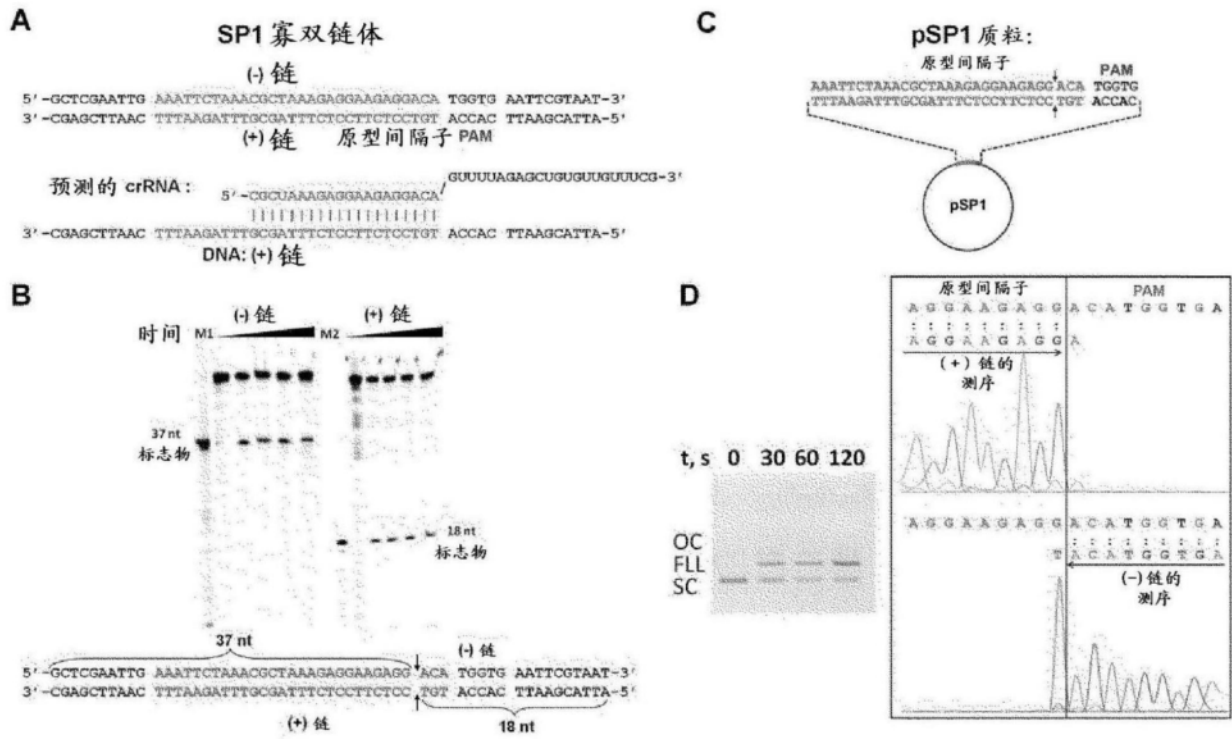


图5

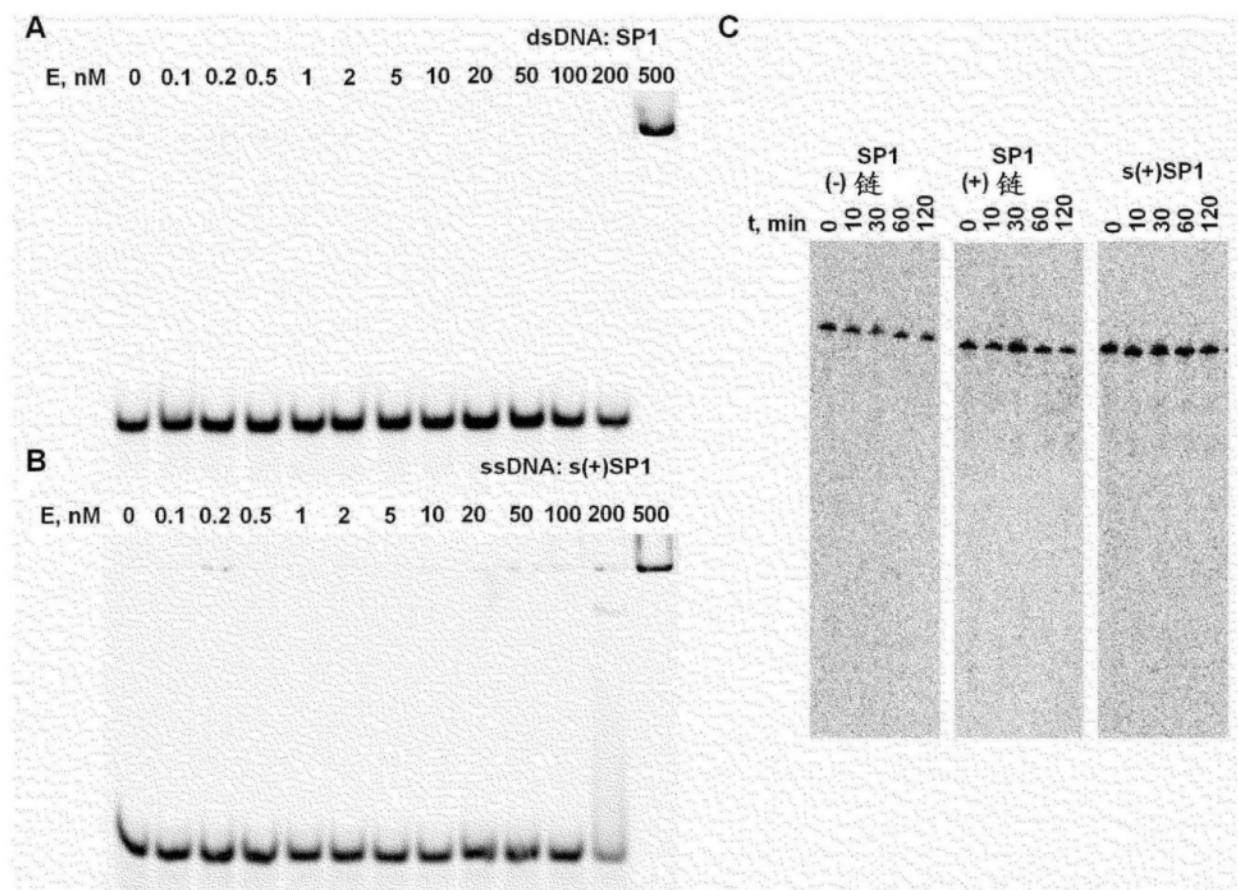


图6

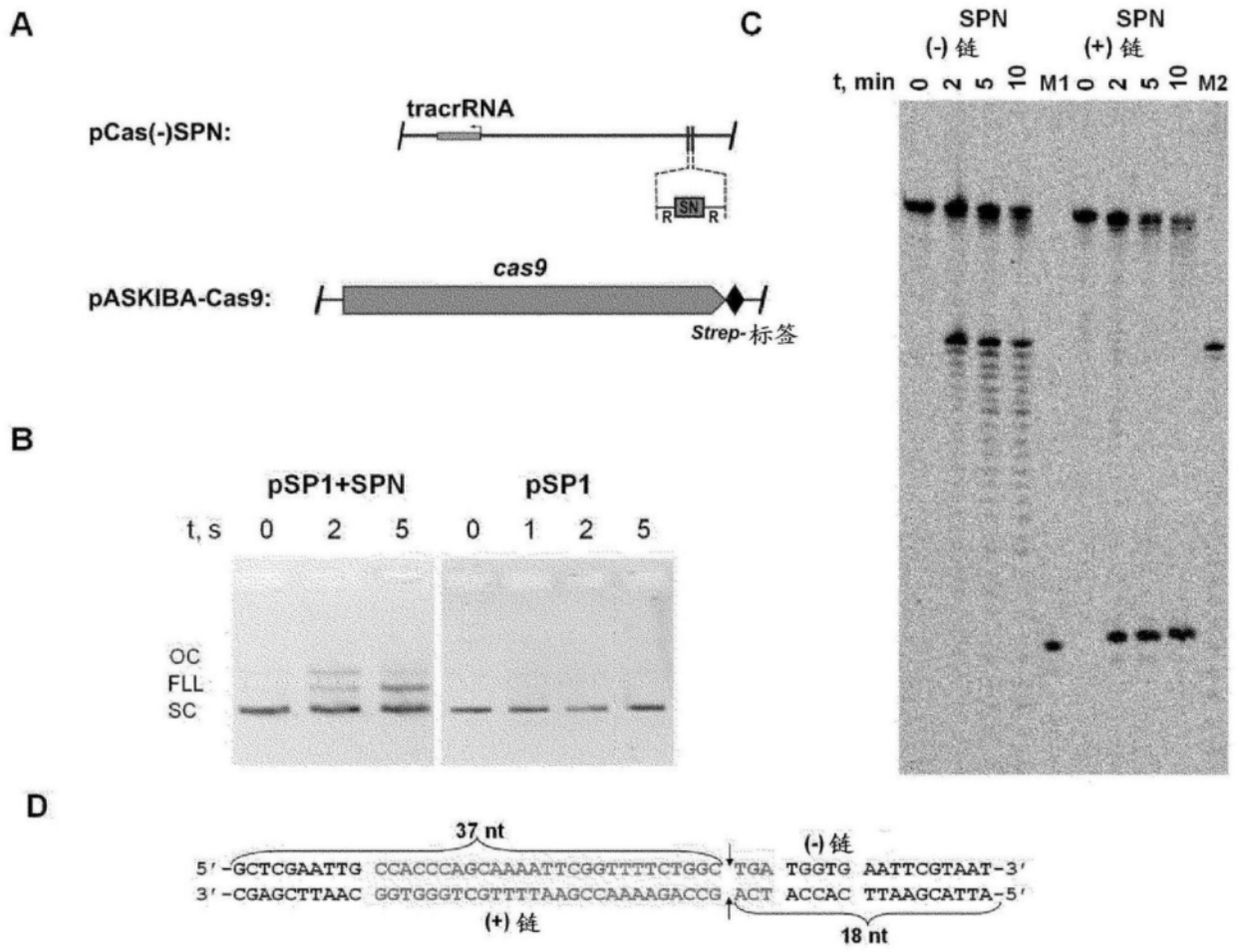


图7

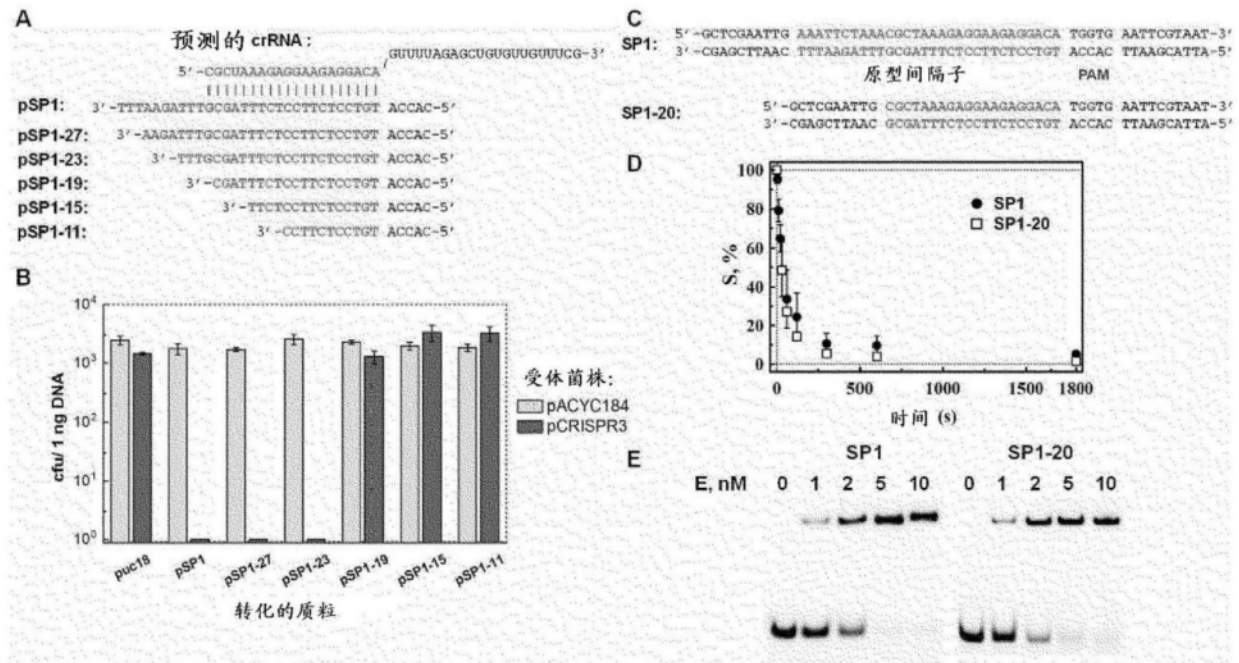


图8

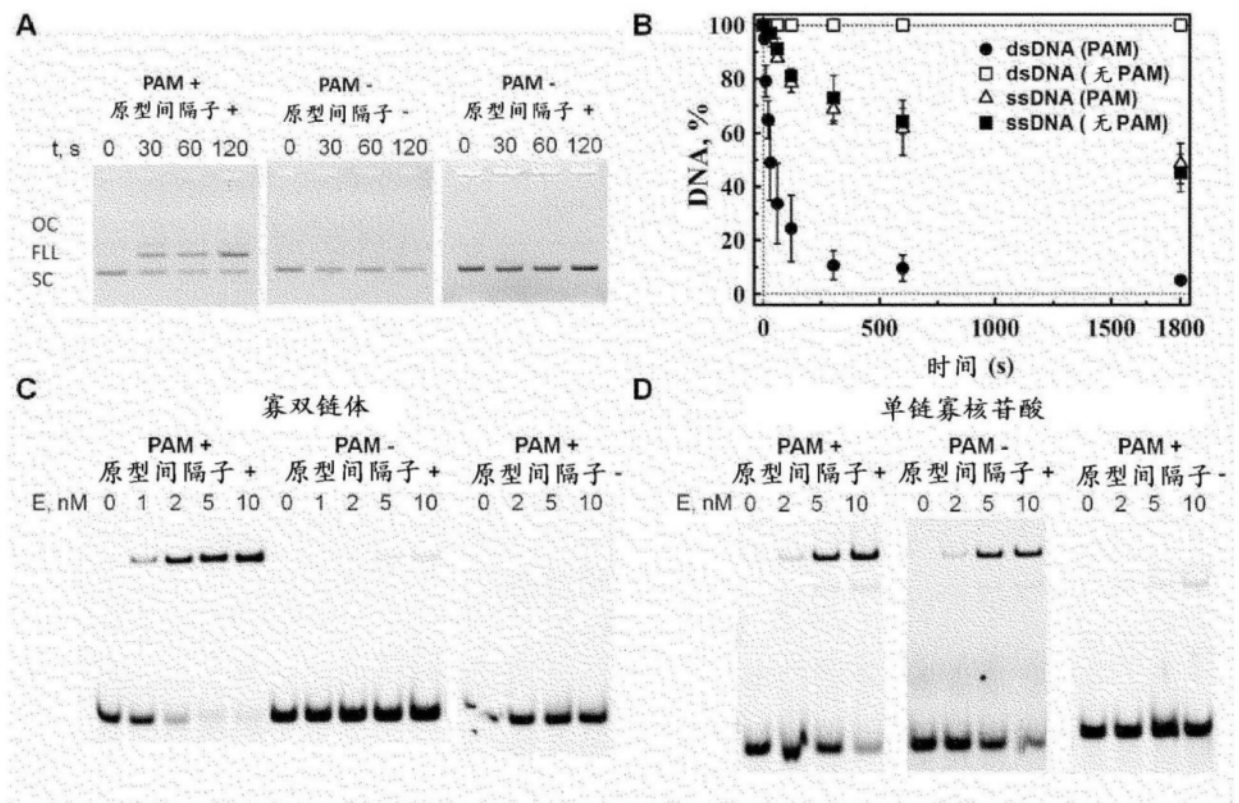


图9

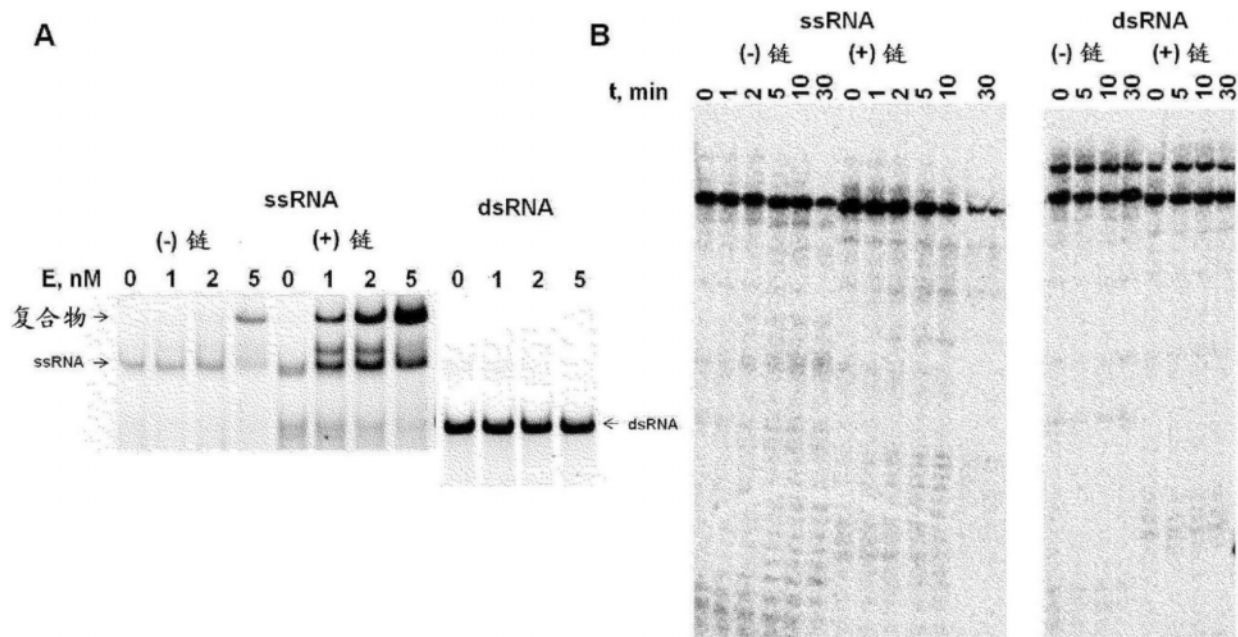


图10

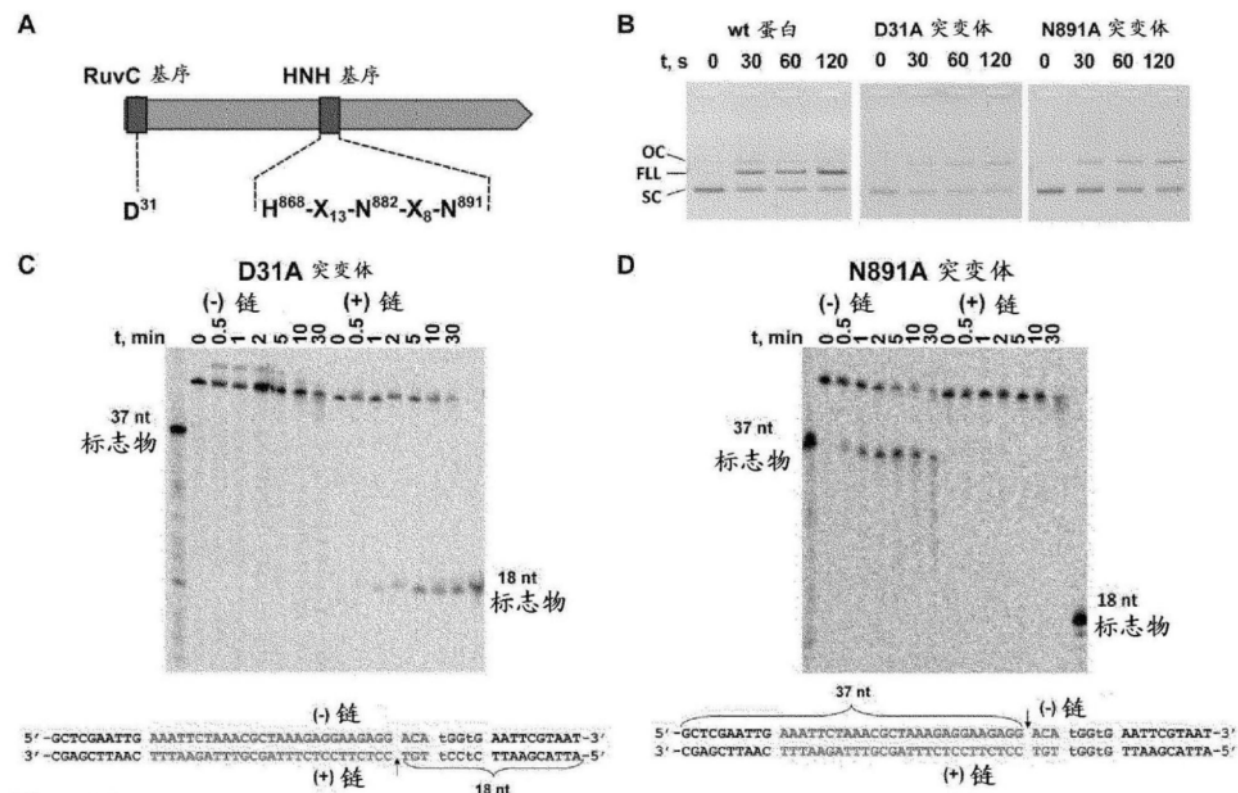


Figure 11.

图11

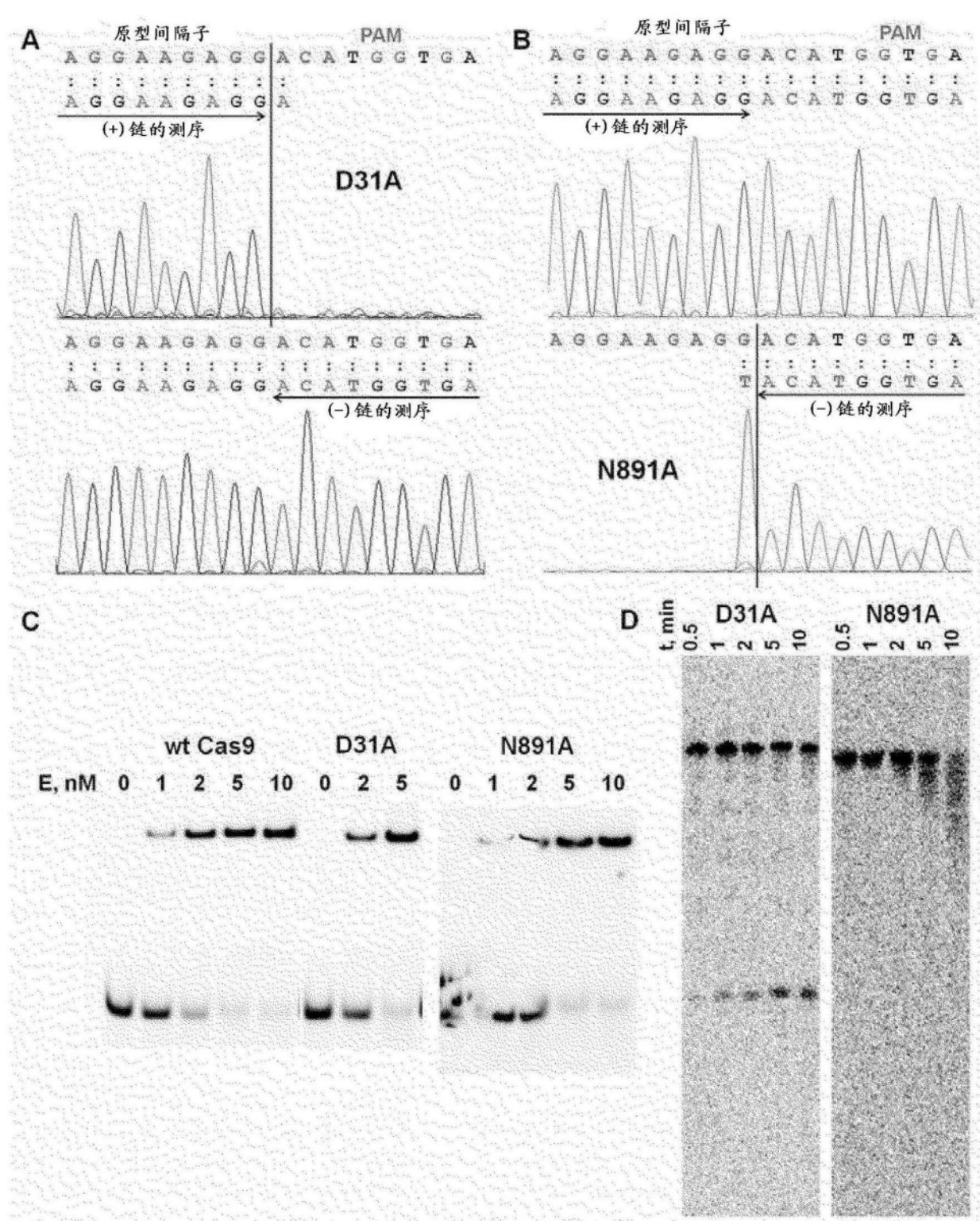


图12

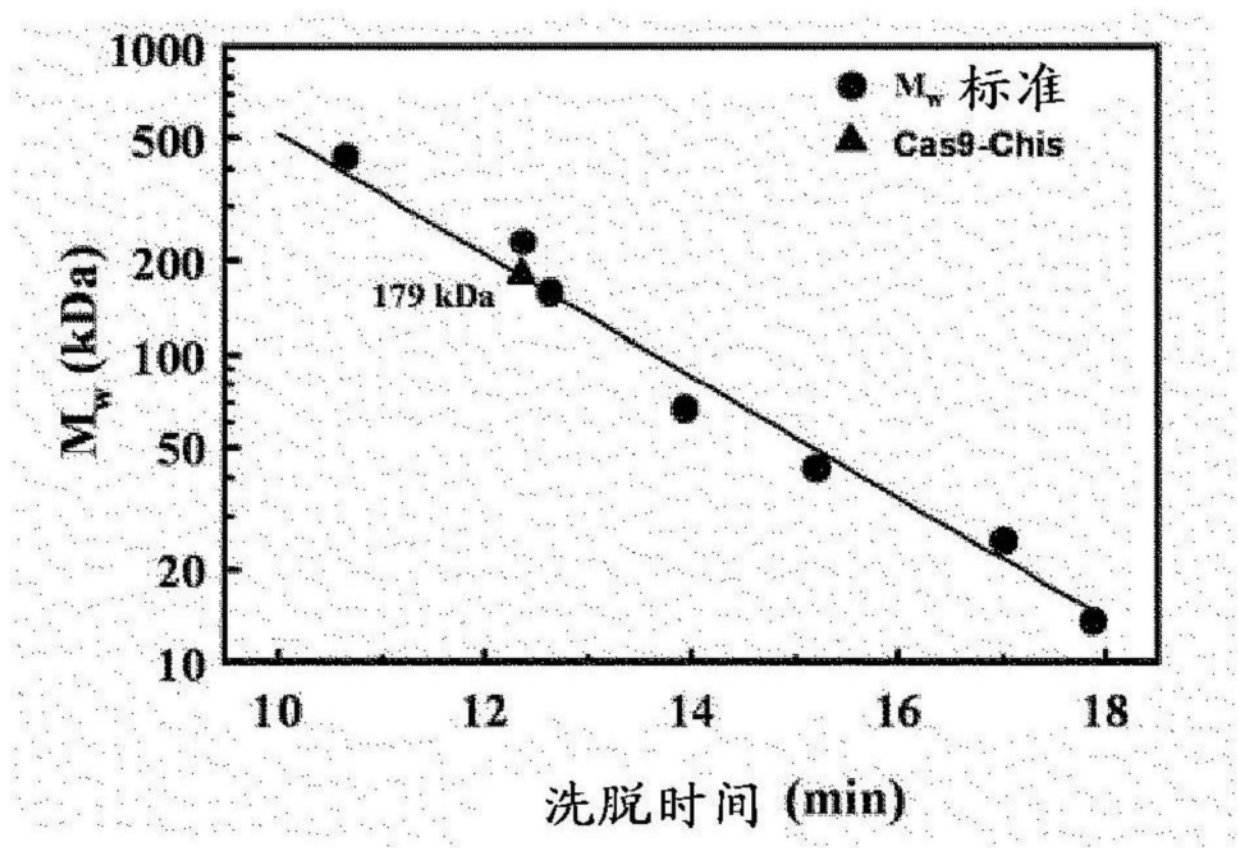


图13

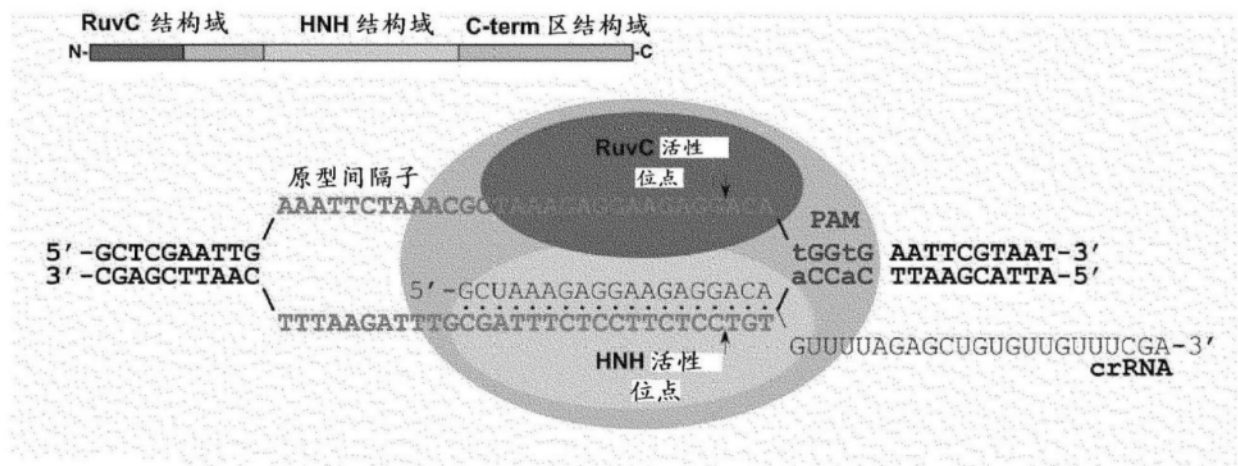


图14

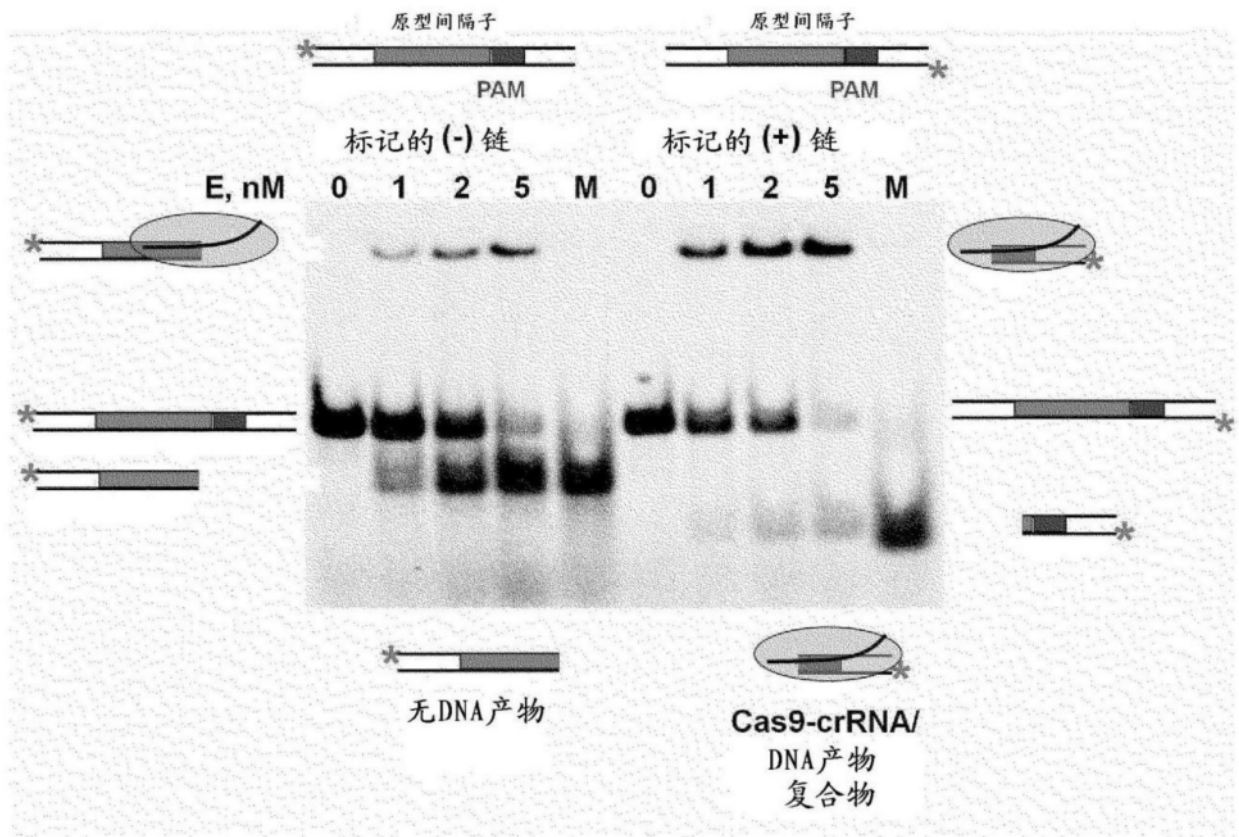


图15

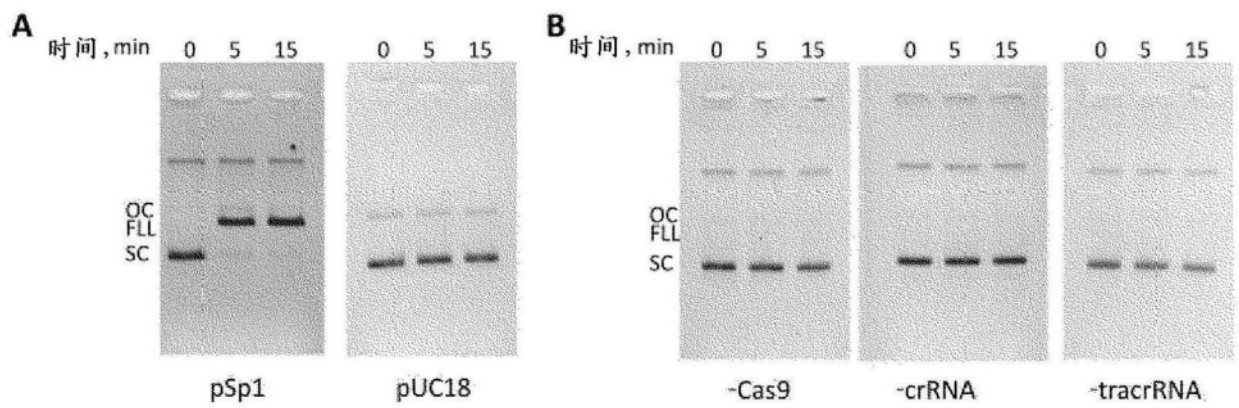


图16

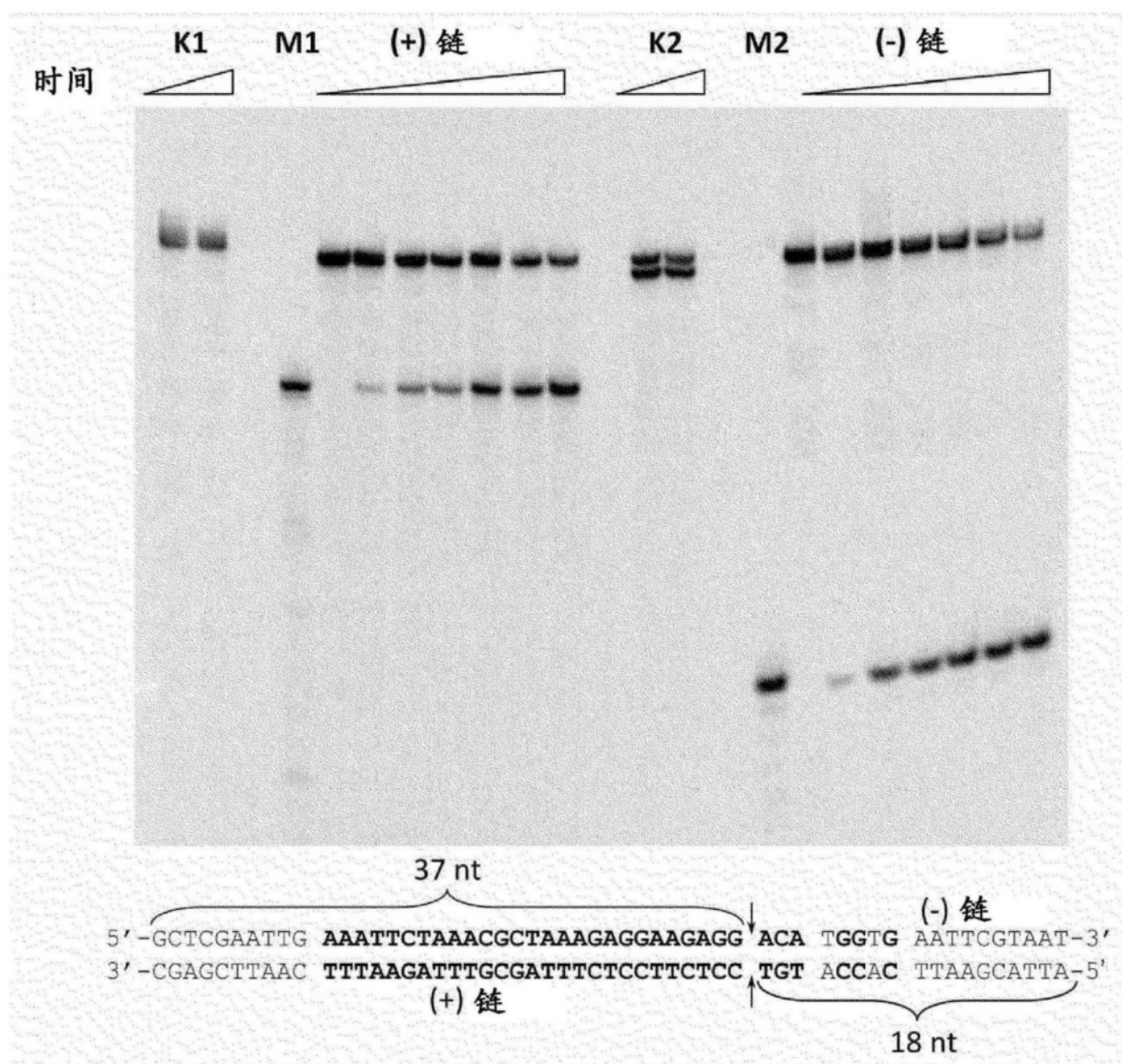


图17

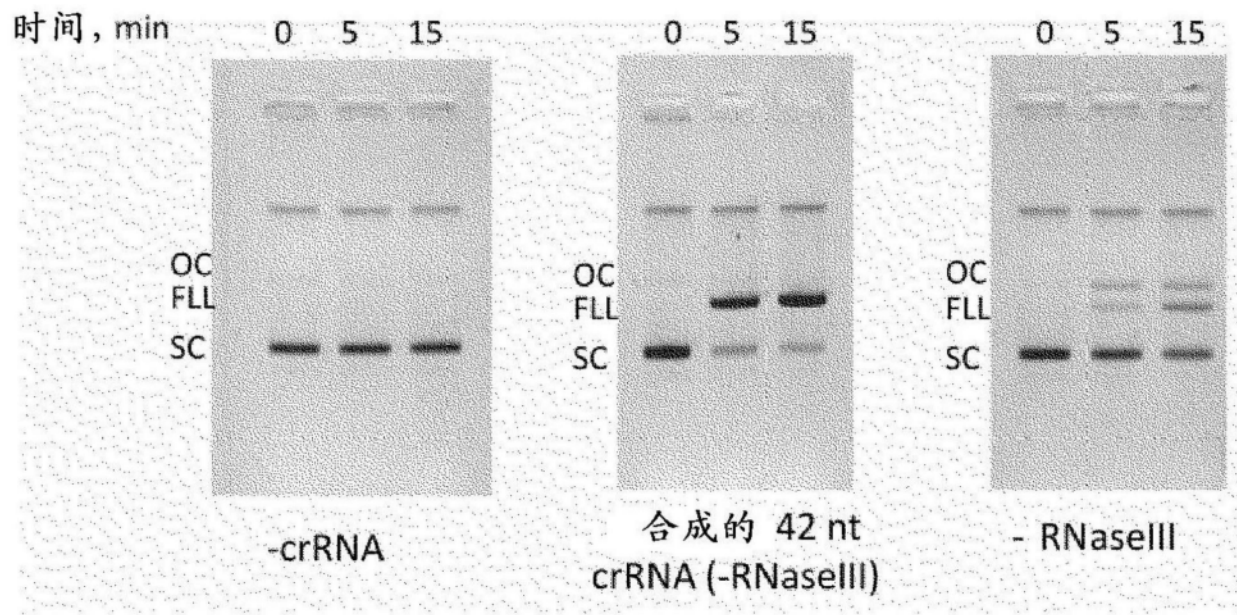


图18

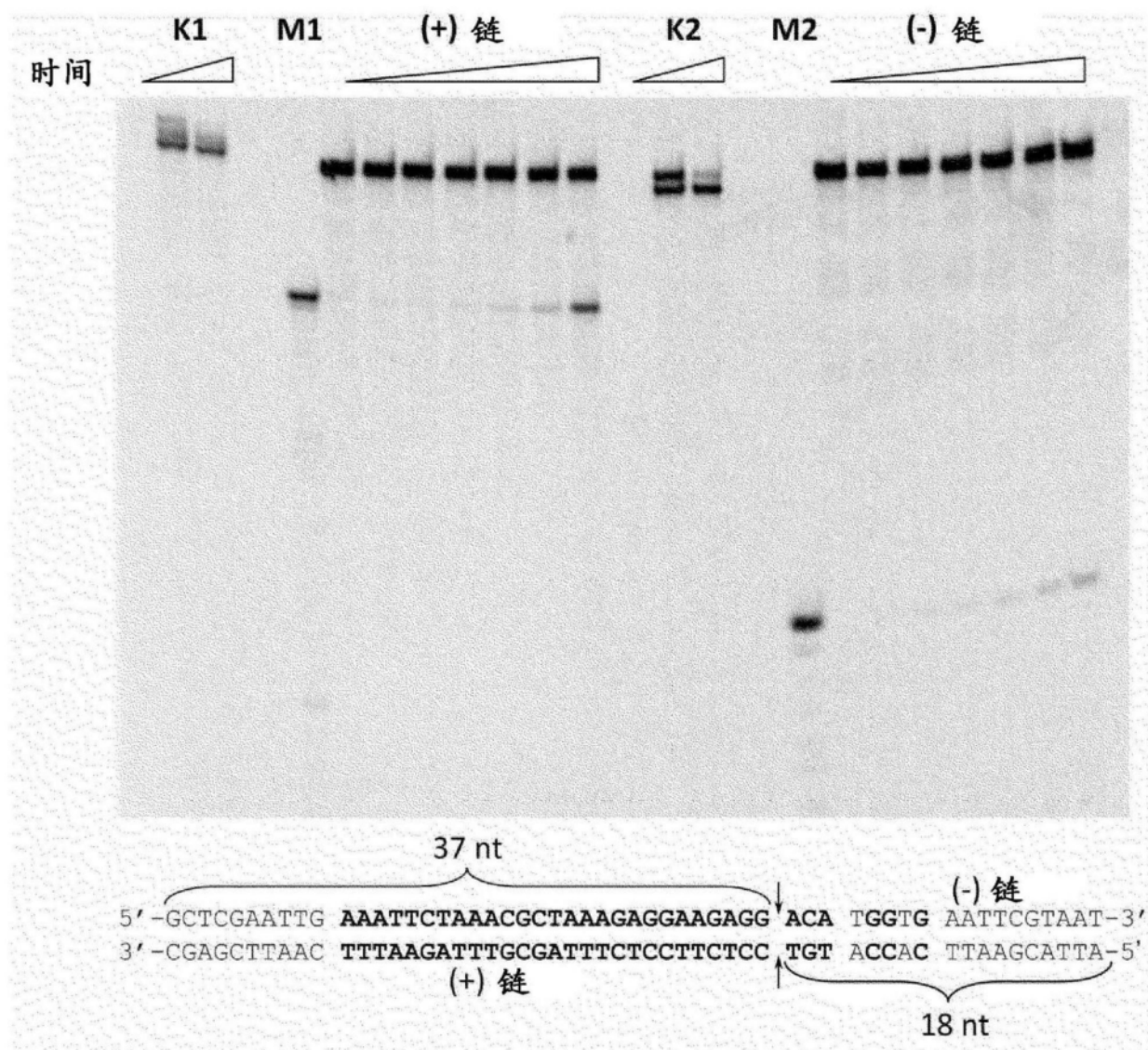


图19

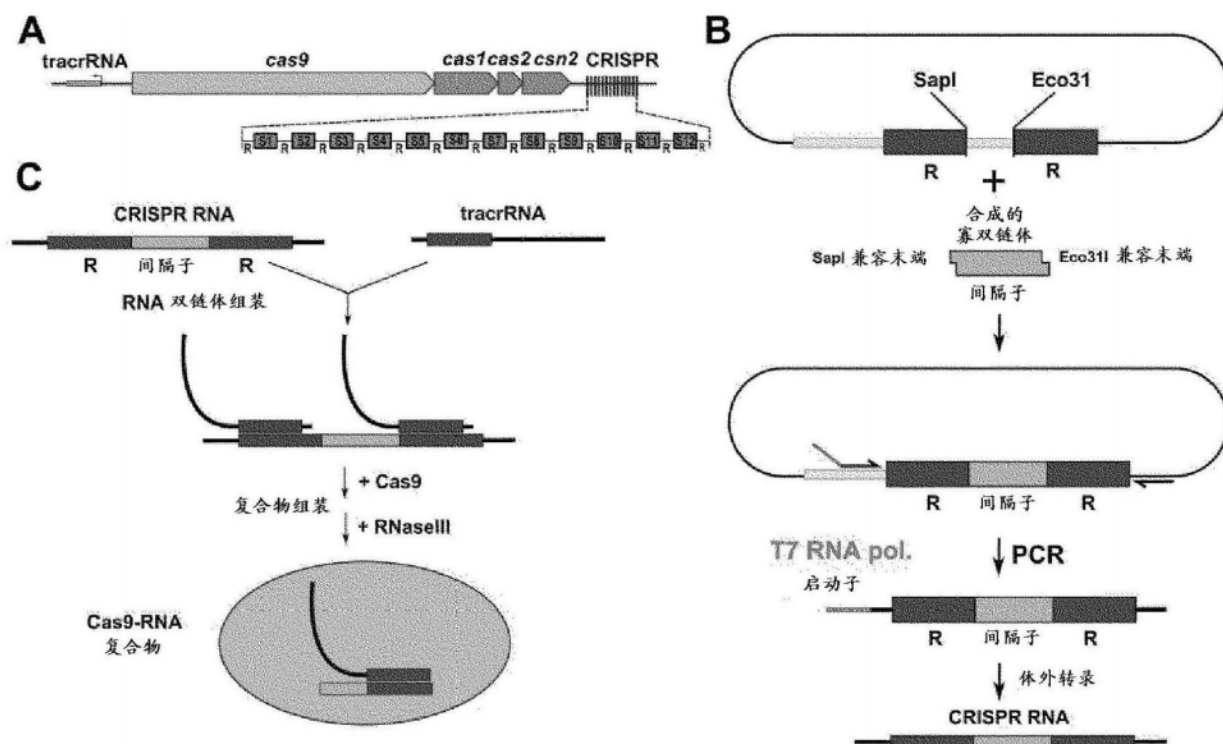


图20

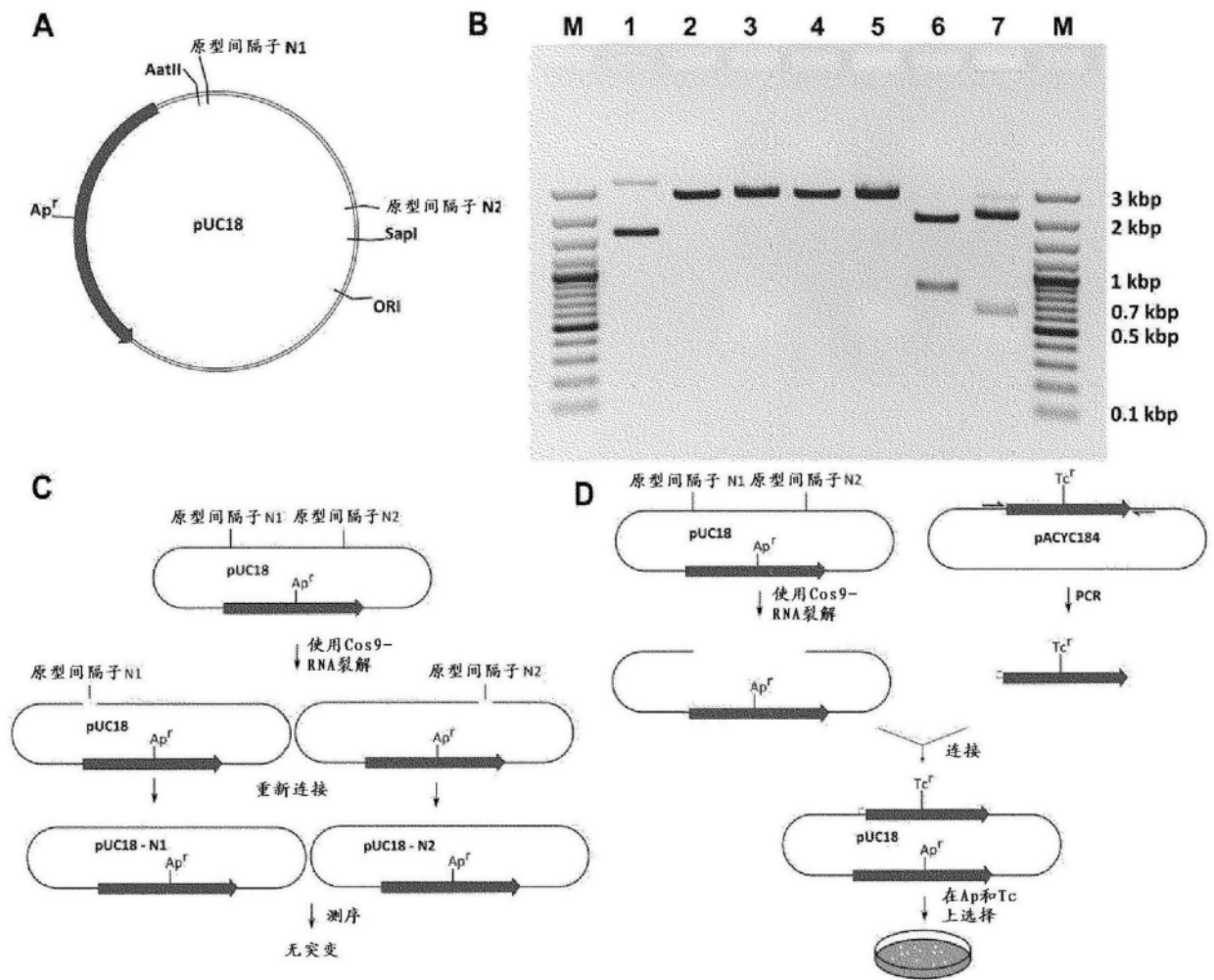


图21

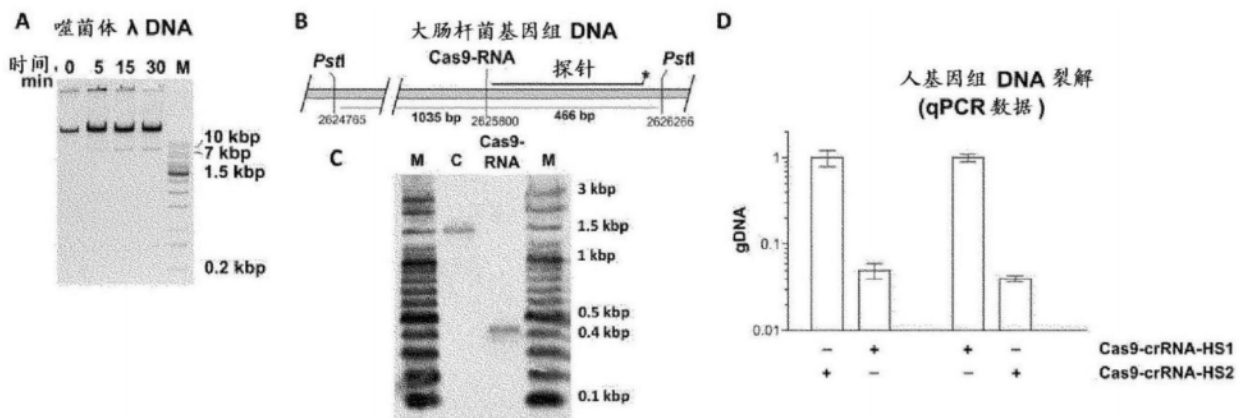


图22

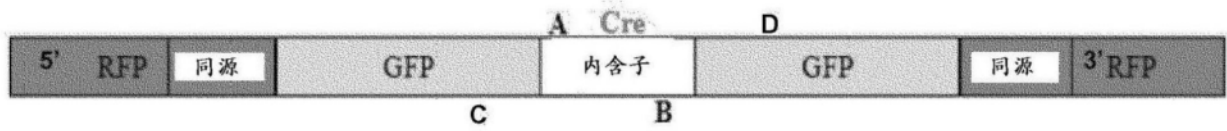


图23

pMTC-DSR+eGFP 和 Cas9/crRNA 共转染,
CHO-K1

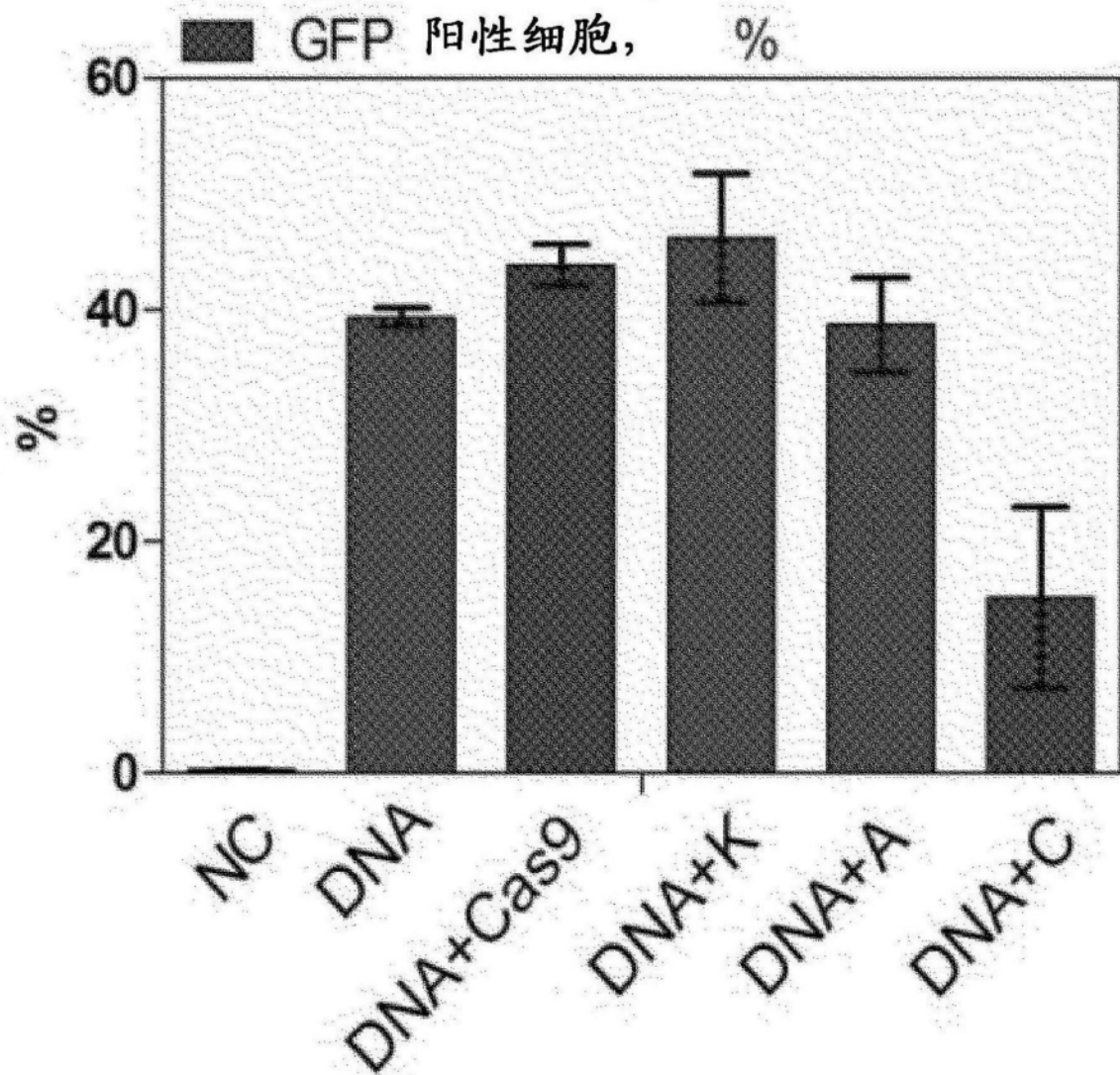
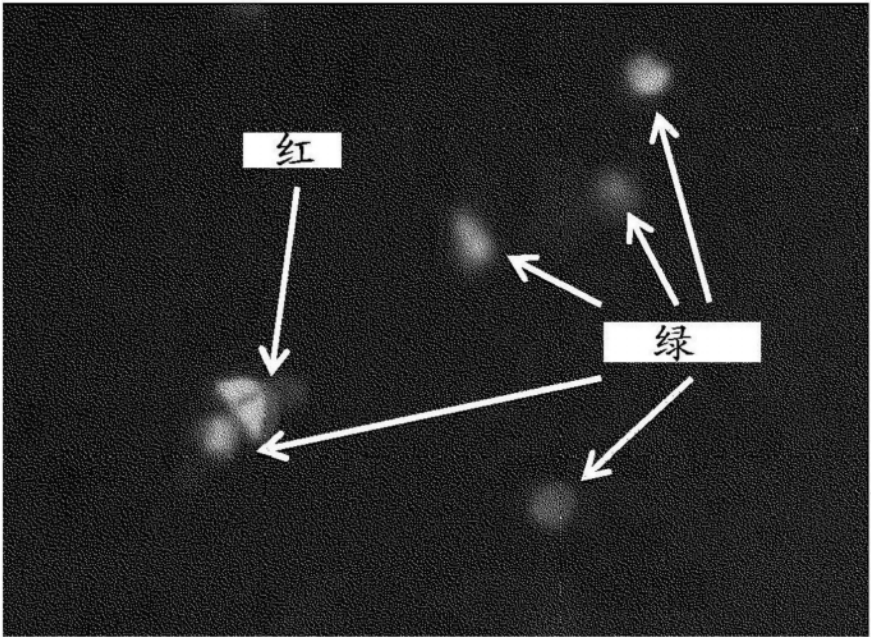


图24

A



B

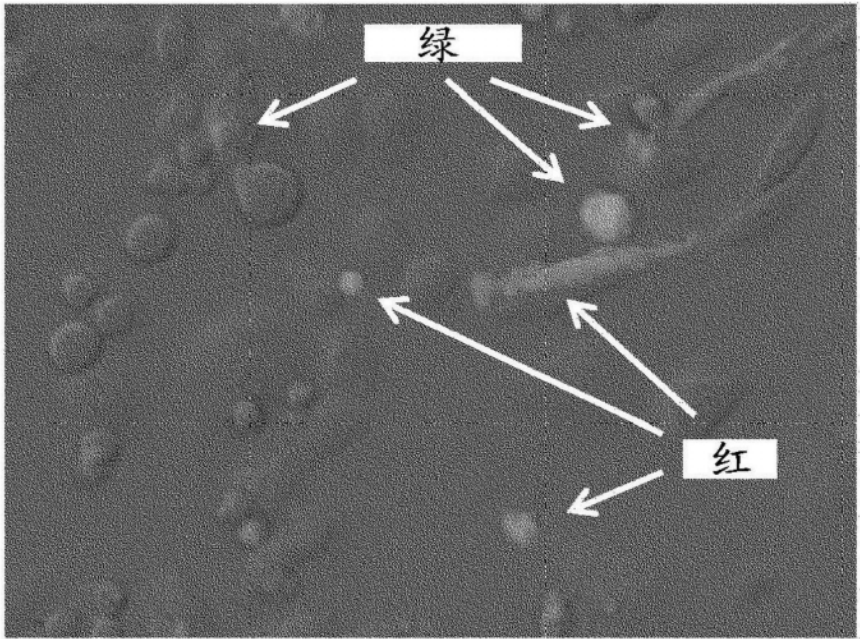


图25

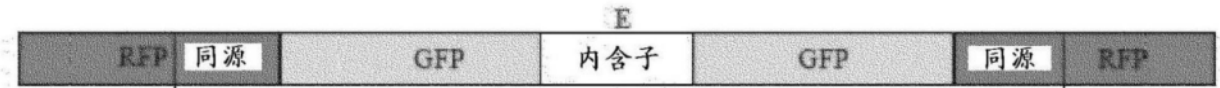


图26

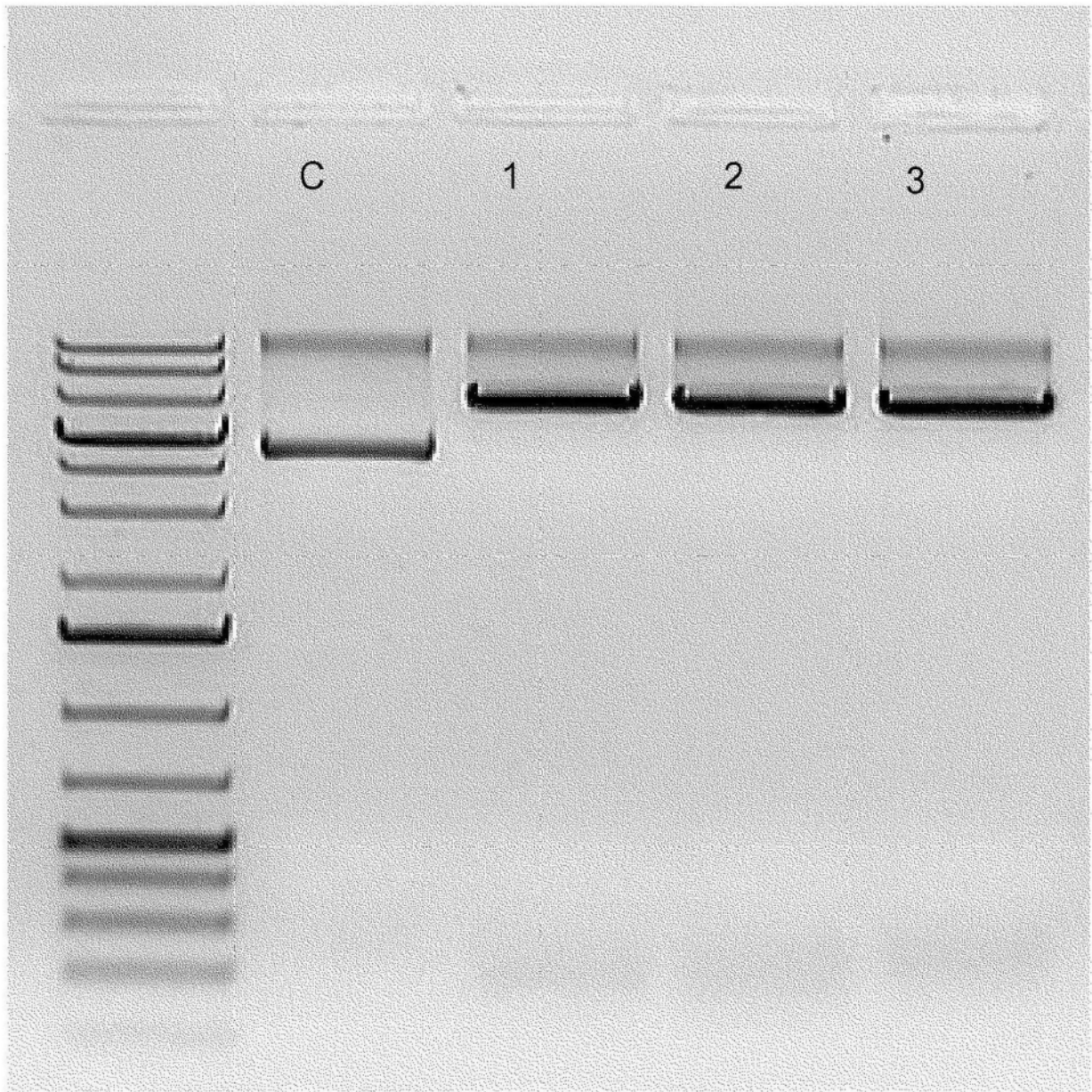


图27

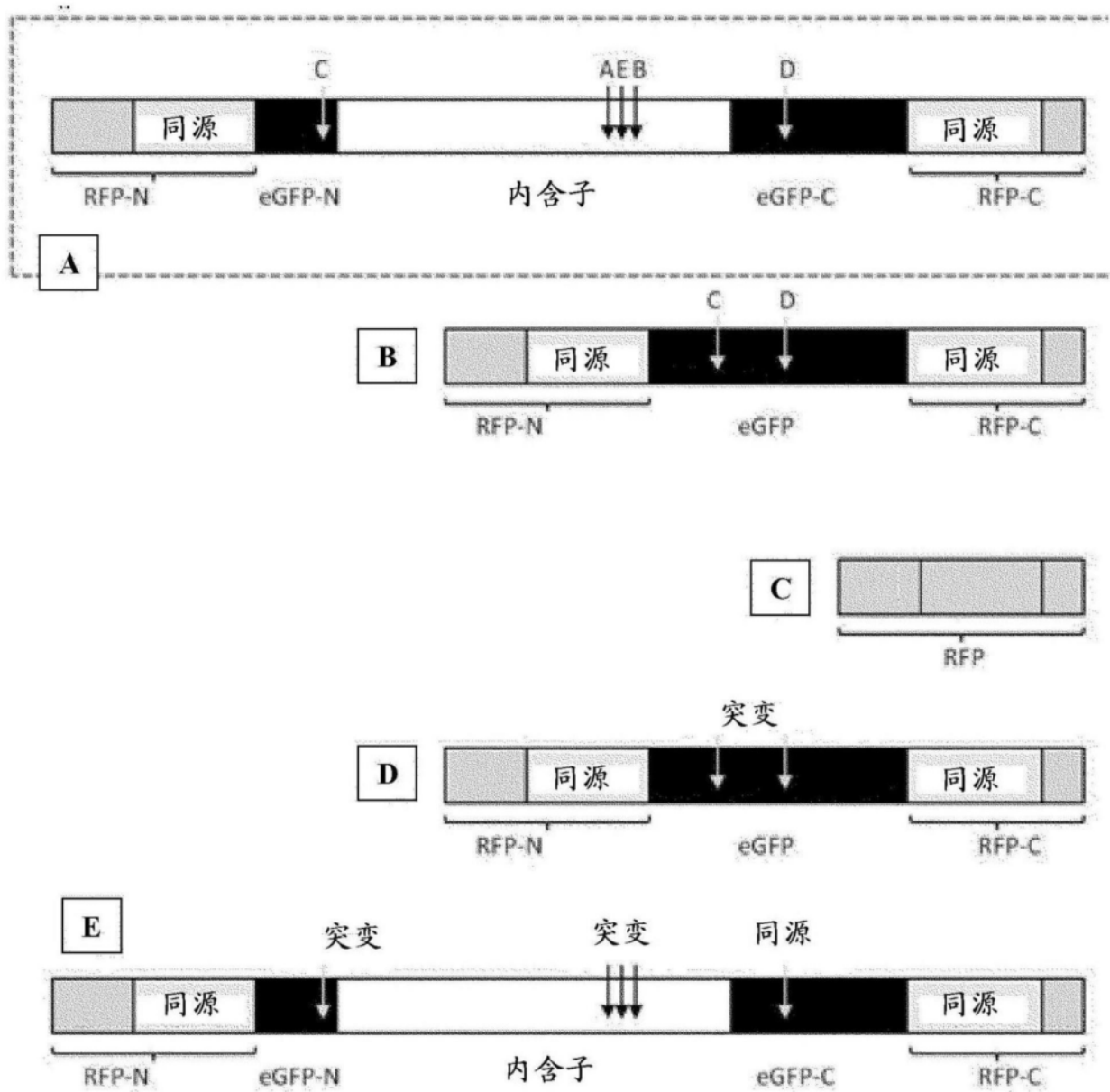


图28

pMTC-DSR+eGFP 和 Cas9/crRNA 共转染 CHO-k1

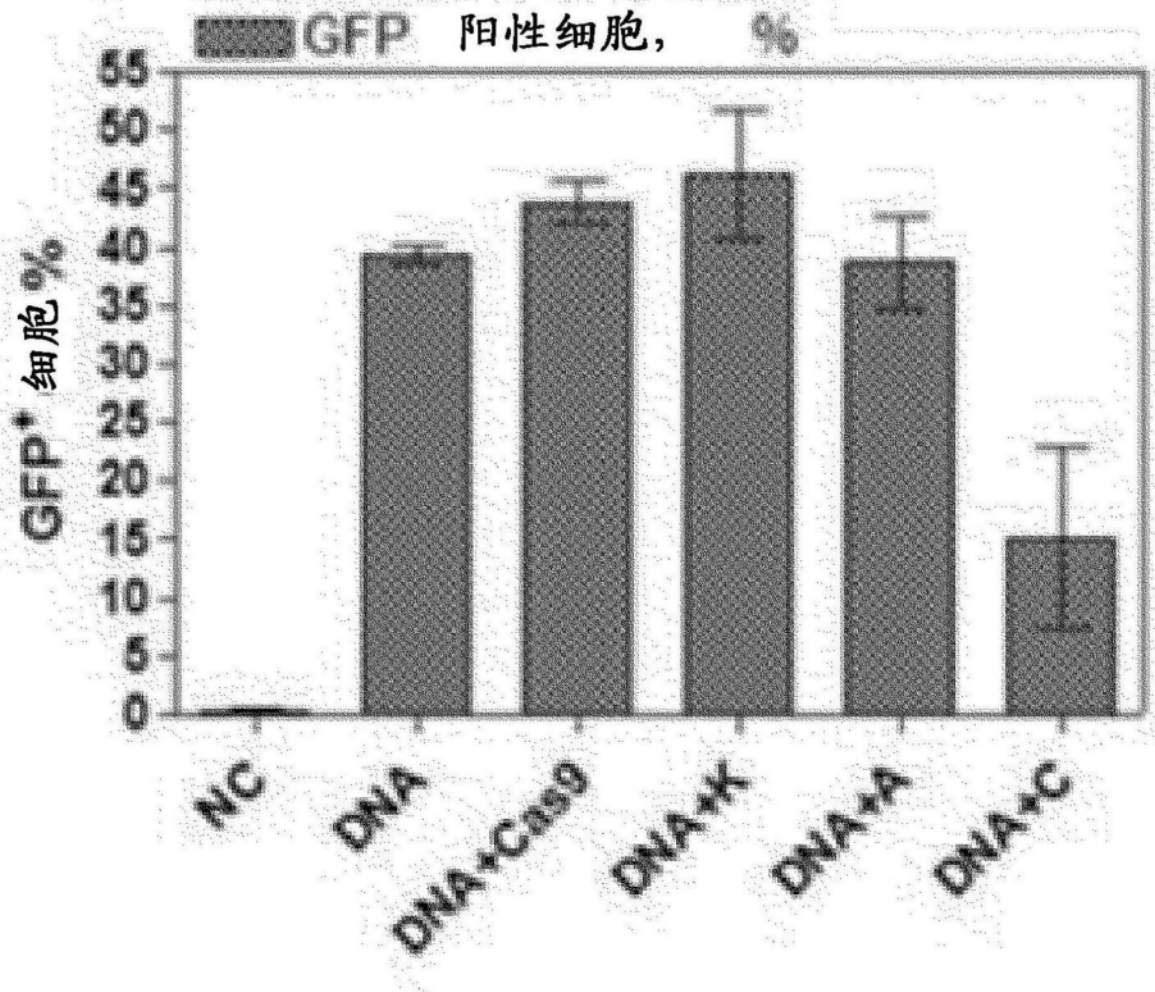


图29

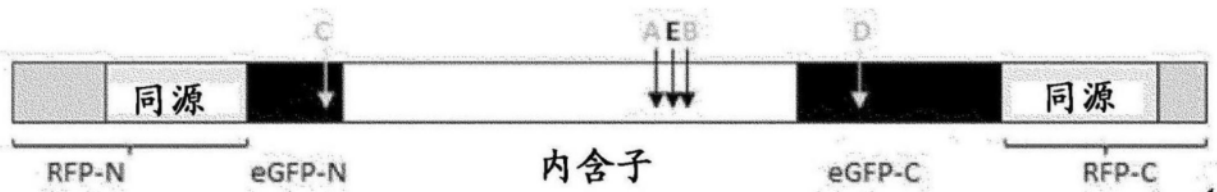


图30