Innovation, Sciences et
Développement économique Canada
Office de la Propriété Intellectuelle du Canada

Innovation, Science and
Economic Development Canada
Canadian Intellectual Property Office

CA 2867849 C 2021/09/07

(11)(21) **2 867 849**

(12) BREVET CANADIEN
CANADIAN PATENT

(13) **C**

(86) **Date de dépôt PCT/PCT Filing Date**: 2013/03/20
(87) **Date publication PCT/PCT Publication Date**: 2013/09/26
(45) **Date de délivrance/Issue Date**: 2021/09/07
(85) **Entrée phase nationale/National Entry**: 2014/09/18
(86) **N° demande PCT/PCT Application No.**: US 2013/033106
(87) **N° publication PCT/PCT Publication No.**: 2013/142578
(30) **Priorités/Priorities**: 2012/03/20 (US61/613,373);
2012/04/17 (US61/625,420)

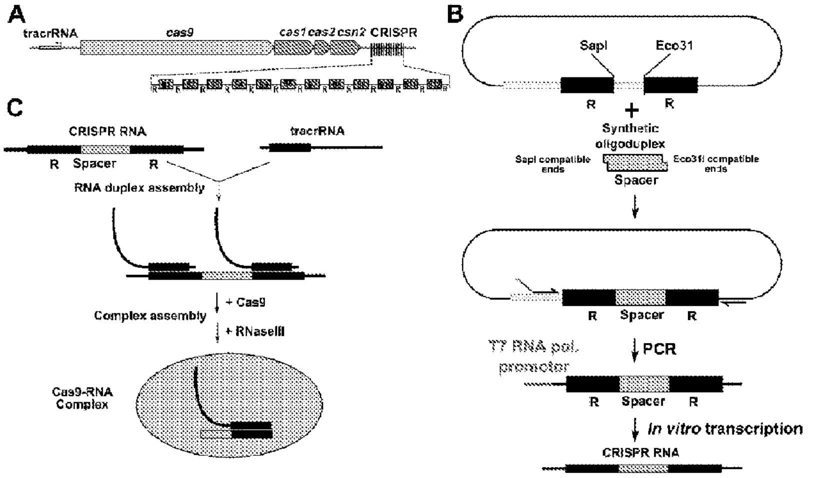
(51) **Cl.Int./Int.Cl.** *C12N 15/10* (2006.01),
C12N 15/113 (2010.01)

(72) **Inventeurs/Inventors**:
SIKSNYS, VIRGINIJUS, LT;
GASIUNAS, GIEDRIUS, LT;
KARVELIS, TAUTVYDAS, LT;
LUBYS, ARVYDAS, LT;
ZALIAUSKIENE, LOLITA, LT;
GLEMZAITE, MONIKA, LT;
SMITH, ANJA, US

(73) **Propriétaire/Owner**:
VILNIUS UNIVERSITY, LT

(74) **Agent**: MACRAE & CO.

(54) **Titre** : CLIVAGE D'ADN DIRIGE PAR ARN PAR LE COMPLEXE CAS9-ARNCR
(54) **Title**: RNA-DIRECTED DNA CLEAVAGE BY THE CAS9-CRRNA COMPLEX



(57) **Abrégé/Abstract**:

Isolation or in vitro assembly of the Cas9-crRNA complex of the Streptococcus thermophilus CRISPR3/Cas system and use for cleavage of DNA bearing a nucleotide sequence complementary to the crRNA and a proto-spacer adjacent motif. Methods for site-specific modification of a target DNA molecule in vitro or in vivo using an RNA-guided DNA endonuclease comprising RNA sequences and at least one of an RuvC active site motif and an HNH active site motif; for conversion of Cas9 polypeptide into a nickase cleaving one strand of double-stranded DNA by inactivating one of the active sites (RuvC or HNH) in the polypeptide by at least one point mutation; for assembly of active polypeptide-polyribonucleotides complex in vivo or in vitro; and for re-programming a Cas9-crRNA complex specificity in vitro and using a cassette containing a single repeat-spacer-repeat unit.

Canada *http://opic.gc.ca* · Ottawa-Hull K1A 0C9 · *http://cipo.gc.ca* OPIC CIPO
OPIC · CIPO 191

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
26 September 2013 (26.09.2013)

WIPO | PCT

(10) International Publication Number
WO 2013/142578 A1

(51) International Patent Classification:
C12N 15/10 (2006.01) *C12N 15/113* (2010.01)

(21) International Application Number:
PCT/US2013/033106

(22) International Filing Date:
20 March 2013 (20.03.2013)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

61/613,373	20 March 2012 (20.03.2012)	US
61/625,420	17 April 2012 (17.04.2012)	US

(71) Applicant *(for all designated States except US)*: **VILNIUS UNIVERSITY** [LT/LT]; Universiteto g. 3, LT-01513 Vilnius (LT).

(72) Inventors; and

(71) Applicants *(for US only)*: **SIKSNYS, Virginijus** [LT/LT]; Konarskio st. 32-26, Vilnius (LT). **GASIUNAS, Giedrius** [LT/LT]; Parko st. 4. Kvetkai, B, 41212 Birzai District (LT). **KARVELIS, Tautvydas** [LT/LT]; Vilniaus st. 28, Paberze, Vilnius District (LT). **LUBYS, Arvydas** [LT/LT]; Konarskio st. 32A-7, LT-03127 Vilnius (LT). **ZALIAUSKIENE, Lolita** [LT/LT]; Vyduno 18-39, LT-06205 Vilnius (LT). **GLEMZAITE, Monika** [LT/LT]; c/o Fermentas UAB, V. Graiciuno 8, LT-02241 Vilnius (LT). **SMITH, Anja** [US/US]; c/o Dharmacon, Inc., 2650 Crescent Drive, Suite 100, Lafayette, CO 80026 (US).

(74) Agents: LYMAN, Beverly, A. et al.; Thompson Hine LLP, 10050 Innovation Drive, Suite 400, Dayton, OH 45342-4934 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

- *with international search report (Art. 21(3))*
- *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

[Continued on next page]

(54) Title: RNA-DIRECTED DNA CLEAVAGE BY THE Cas9-crRNA COMPLEX

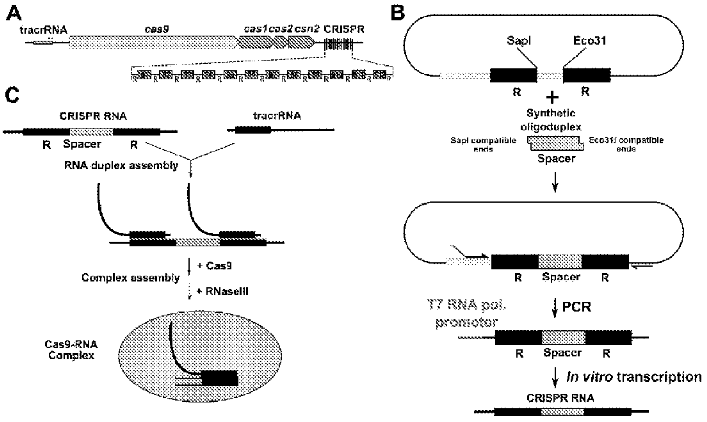


Figure 20

(57) Abstract: Isolation or in vitro assembly of the Cas9-crRNA complex of the Streptococcus *thermophilus* CRISPR3/Cas system and use for cleavage of DNA bearing a nucleotide sequence complementary to the crRNA and a proto-spacer adjacent motif. Methods for site-specific modification of a target DNA molecule in vitro or in vivo using an RNA-guided DNA endonuclease comprising RNA sequences and at least one of an RuvC active site motif and an HNH active site motif; for conversion of Cas9 polypeptide into a nickase cleaving one strand of double-stranded DNA by inactivating one of the active sites (RuvC or HNH) in the polypeptide by at least one point mutation; for assembly of active polypeptide-polyribonucleotides complex *in vivo* or *in vitro*; and for re-programming a Cas9-crRNA complex specificity *in vitro* and using a cassette containing a single repeat-spacer-repeat unit.

WO 2013/142578 A1

— *with sequence listing part of description (Rule 5.2(a))*

RNA-DIRECTED DNA CLEAVAGE BY THE Cas9-crRNA COMPLEX

Abstract

CRISPR/Cas systems provide adaptive immunity against viruses and plasmids in bacteria and archaea. The silencing of invading nucleic acids is executed by ribonucleoprotein (RNP) complexes pre-loaded with small interfering crRNAs that act as guides for foreign nucleic acid targeting and degradation. Here we describe an isolation of the Cas9-crRNA complex and demonstrate that it generates in vitro a double strand break at specific sites in target DNA molecules that are complementary to crRNA sequences and bear a short proto-spacer adjacent motif (PAM), in the direct vicinity of the matching sequence. We show that DNA cleavage is executed by two distinct active sites (RuvC and HNH) within Cas9, to generate site-specific nicks on opposite DNA strands. Sequence specificity of the Cas9-crRNA complex is dictated by the 42 nt crRNA which includes a 20 nt fragment complementary to the proto-spacer sequence in the target DNA. The complex can be assembled in vitro or in vivo. Altogether, our data demonstrate that the Cas9-crRNA complex functions as an RNA-guided endonuclease with sequence-specific target site recognition and cleavage through two distinct strand nicks.

Background

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) together with cas (CRISPR-associated) genes comprise an adaptive immune system that provides acquired resistance against invading foreign nucleic acids in bacteria and archaea (Barrangou et al., 2007. Science 315:1709-12). CRISPR consists of arrays of short conserved repeat sequences interspaced by unique variable DNA sequences of similar size called spacers, which often originate from phage or plasmid DNA (Barrangou et al., 2007. Science 315:1709-12; Bolotin et al., 2005. Microbiology 151 :2551-61 ; Mojica et al., 2005. J Mol Evol 60:174-82). The CRISPR-Cas system functions by acquiring short pieces of foreign DNA (spacers) which are inserted into the CRISPR region and provide immunity against subsequent exposures to phages and plasmids that carry matching sequences (Barrangou et al., 2007. Science 315:1709-12; Brouns et al., 2008. Science 321 : 960-4) The CRISPR-Cas immunity is generally carried out through three stages, referred to as i) adaptation/immunization/spacer acquisition, ii) CRISPR expression/crRNA biogenesis, iii) interference/immunity. (Horvath & Barrangou, 2010. Science 327:167-70; Deveau et al., 2010. Annu Rev Microbiol. 64:475-93;Marraffini & Sontheimer, 2010. Nat Rev Genet 1 1 , 181-90; Bhaya et al., Annu Rev Genet 45:273-97; Wiedenheft et al., 2012. Nature 482:331-338). Here, we specifically focus on the interference/immunity step which enables crRNA-mediated silencing of foreign nucleic acids.

The highly diverse CRISPR-Cas systems are categorized into three major types, which are further subdivided into ten subtypes, based on core element content and sequences (Makarova et al.,

2011. Nat Rev Microbiol 9:467–77). The structural organization and function of nucleoprotein complexes involved in crRNA-mediated silencing of foreign nucleic acids differ between distinct CRISPR/Cas types (Wiedenheft et al., 2012. Nature 482:331–338). In the Type I-E system, as exemplified by *Escherichia coli,* crRNAs are incorporated into a multisubunit effector complex called Cascade (CRISPR-associated complex for antiviral defence) (Brouns et al., 2008. Science 321: 960–4), which binds to the target DNA and triggers degradation by the signature Cas3 protein (Sinkunas et al.,2011. EMBO J 30:1335–42; Beloglazova et al., 2011. EMBO J 30:616–27). In Type III CRISPR/Cas systems of *Sulfolobus solfataricus* and *Pyrococcus furiosus,* Cas RAMP module (Cmr) and crRNA complex recognize and cleave synthetic RNA *in vitro* (Hale et al., 2012. Mol Cell 45:292–302; Zhang et al., 2012. Mol Cell, 45:303–13) while the CRISPR/Cas system of *Staphylococcus epidermidis* targets DNA *in vivo* (Marraffini & Sontheimer, Science. 322:1843–5).

RNP complexes involved in DNA silencing by Type II CRISPR/Cas systems, more specifically in the CRISPR3/Cas system of *Streptococcus thermophilus* DGCC7710 (Horvath & Barrangou, 2010. Science 327:167–70), consists of four *cas* genes *cas9, cas1, cas2,* and *csn2,* that are located upstream of 12 repeat-spacer units (Figure 1A). C*as9* (formerly named *cas5* or *csn1*) is the signature gene for Type II systems (Makarova et al., 2011. Nat Rev Microbiol 9:467–77). In the closely related *S. thermophilus* CRISPR1/Cas system, disruption of *cas9* abolishes crRNA-mediated DNA interference (Barrangou et al., 2007. Science 315:1709–12). We have shown recently that the *S. thermophilus* CRISPR3/Cas system can be transferred into *Escherichia coli*, and that this heterologous system provides protection against plasmid transformation and phage infection, *de novo* (Sapranauskas et al., 2011. Nucleic Acids Res 39:9275–82). The interference against phage and plasmid DNA provided by *S. thermophilus* CRISPR3 requires the presence, within the target DNA, of a proto-spacer sequence complementary to the spacer-derived crRNA, and a conserved PAM (Proto-spacer Adjacent Motif) sequence, NGGNG, located immediately downstream the proto-spacer (Deveau et al., 2008. J Bacteriol 190:1390–400; Horvath et al., 2008. J Bacteriol 190:1401–12; Mojica et al., 2009. Microbiology 155:733–40). Single point mutations in the PAM or defined proto-spacer positions allow the phages or plasmids to circumvent CRISPR-mediated immunity (Deveau et al., 2008. J Bacteriol 190:1390–400; Garneau et al., 2010. Nature 468:67–71; Sapranauskas et al., 2011. Nucleic Acids Res 39:9275–82). We have established that in the heterologous system, *cas9* is the sole *cas* gene necessary for CRISPR-encoded interference (Sapranauskas et al., 2011. Nucleic Acids Res 39:9275–82), suggesting that this protein is involved in crRNA processing and/or crRNA-mediated silencing of invasive DNA. Cas9 of *S. thermophilus* CRISPR3/Cas system is a large multi-domain protein comprised of 1,409 aa residues (Sapranauskas et al., 2011. Nucleic Acids Res 39:9275–82). It contains two nuclease domains, a RuvC-like nuclease domain near the amino terminus, and a HNH-like nuclease domain in the middle of the protein. Mutational analysis has established that interference provided *in vivo* by Cas9 requires both the RuvC- and HNH-motifs (Sapranauskas et al., 2011. Nucleic Acids Res 39:9275–82).

Isolation of the Cas9-crRNA complex of the *S. thermophilus* CRISPR3/Cas system as well as complex assembly *in vitro* from separate components and demonstration that it cleaves both synthetic oligodeoxynucleotide and plasmid DNA bearing a nucleotide sequence complementary to the crRNA, in a PAM-dependent manner, is provided. Furthermore, we provide experimental evidence that the PAM is recognized in the context of double-stranded DNA and is critical for *in vitro* DNA binding and cleavage. Finally, we show that the Cas9 RuvC- and HNH- active sites are responsible for the cleavage of opposite DNA strands. Taken together, our data demonstrate that the Cas9-crRNA complex functions as an RNA-guided endonuclease which uses RNA for the target site recognition and Cas9 for DNA cleavage. The simple modular organization of the Cas9-crRNA complex, where specificity for DNA targets is encoded by a small crRNA and the cleavage machinery consists of a single, multidomain Cas protein, provides a versatile platform for the engineering of universal RNA-guided DNA endonucleases. Indeed, we provide evidence that by altering the RNA sequence within the Cas9-crRNA complex, programmable endonucleases can be designed both for *in vitro* and *in vivo* applications, and we provide a proof of concept for this novel application. These findings pave the way for the development of novel molecular tools for RNA-directed DNA surgery.

Summary of the invention

A method for the site-specific modification of a target DNA molecule through contacting under suitable conditions, a target polydeoxynucleotide molecule; and an RNA-guided DNA endonuclease comprising at least one RNA sequences and at least one of an RuvC active site motif and an HNH active site motif; to result in the target polydeoxynucleotide molecule modified in a region that is determined by the complimentary binding of the RNA sequence to the target DNA molecule is provided. The method includes incubating under suitable conditions a composition that includes a target double stranded polydeoxynucleotide or single stranded polydeoxynucleotide; wherein a double stranded polydeoxynucleotide contains a short proto-spacer adjacent motif (PAM), which is non-obligatory for a single stranded polydeoxynucleotide; and where PAM comprises a 5'NGGNG-3' sequence; a polyribonucleotide (crRNA) comprising a 3' and 5' regions wherein the 3' region comprises at least 22 nt of the repeat present in a microbe containing CRISPR locus and 5'-region comprises of at least 20 nt of the spacer sequence immediately downstream of the repeat in the CRISPR locus, which is substantially complementary, optionally complementary, to a portion of the target polynucleotide, a polypeptide wherein the amino acid sequence of polypeptide and amino acid sequence of SEQ ID NO: 1 have at least 80% identity, isolated from *S. thermophilus*, or genetically modified microorganism, including a genetically modified *E. coli*, or wherein the polypeptide is produced by a method selected from recombinant DNA technology or chemical synthesis; a polyribonucleotide tracrRNA of nucleotide sequence SEQ ID NO: 5 (or have at least 80% identity) comprising a 5' and 3' regions wherein the 5' region is comprised of at least 22 nucleotides is complementary to the 22 nucleotides 3' region of crRNA, and 3' region. Wherein polyribonucleotides are produced by *in vitro* transcription or chemical synthesis. Wherein, suitable conditions means conditions *in vitro* or *in vivo* where reaction might occur.

A method for the conversion of Cas9 polypeptide into a nickase, cleaving only one strand of double-stranded DNA, by inactivating one of the active sites (RuvC or HNH) in the polypeptide by at least on point mutation, exemplified by D31A (SEQ ID NO: 2), N891A (SEQ ID NO: 3) and H868A (SEQ ID NO: 4) point mutations is provided. RuvC motif mutant cleaves only bottom DNA strand in respect to 5'NGGNG-3' motif, while HNH motif mutant cleaves top strand.

Polypeptide-polyribonucleotides complex might be isolated from a genetically modified microbe (for example *Escherichia coli* or *Streptoccocus thermophilus*), or assembled *in vitro* from separate components. In the genetically modified microbe components of the complex might be encoded on the one, two or three separate plasmids containing host promoters of the genetically modified microbe or promoters from a native host genome.

A method for assembly of active polypeptide-polyribonucleotides complex *in vitro*, comprising incubating the components of the complex under conditions suitable for complex assembly is provided. The complex might be assembled using three or four components. Method for three components assembly comprises incubating the Cas9 polypeptide, 78 nt tracrRNA polyribonucleotide (SEQ ID NO: 5), and 42 nt crRNA polyribonucleotide (5'-NNNNNNNNNNNNNNNNNNNN GUUUUAGAGCUGUGUUGUUUCG-3') (SEQ ID NO: 15) under conditions suitable for complex assembly. Method for four components assembly comprises incubating the Cas9 polypeptide; 102 nt tracrRNA polyribonucleotide (SEQ ID NO: 6); polyribonucleotide containing sequence 5'-NNNNNNNNNNNNNNNNNNNN GUUUUAGAGCUGUGUUGUUUCG-3' (SEQ ID NO: 15) and flanking regions and RNase III polypeptide, cleaving double stranded RNA polynucleotide. The examples for polyribonucleotide containing sequence 5'-NNNNNNNNNNNNNNNNNNNN GUUUUAGAGCUGUGUUGUUUCG-3' (SEQ ID NO: 15) are SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11 and SEQ ID NO: 12). Examples of source for suitable RNaseIII include *Escherichia coli* or *Streptococcus thermophilus*.

A method for re-programming of a Cas9-crRNA complex specificity by mixing separate components or using a cassette containing a single repeat-spacer-repeat unit is provided. Any sequence might be inserted between two repeats in the cassette using suitable restriction endonucleases. Cassette might be used to target sequences *in vivo*, or to produce RNA ribonucleotide suitable for complex assembly *in vitro*.

Brief description of the figures

Figure 1 shows Cas9 protein co-purifies with crRNA. (A) Schematic representation of CRISPR3/Cas system of *S. thermophilus*. Four cas genes (cas9, cas1, cas2, csn2) are located upstream of the CRISPR repeat-spacer array, consisting of 13 repeat (R) sequences and 12 unique spacers (S1-S12). The tracrRNA, required for crRNA maturation in Type II CRISPR systems (Deltcheva et al., 2011. Nature 471:602–7), is located upstream the cas9 gene and encoded on the opposite DNA strand (showed by an arrow) in respect to the other elements of CRISPR3/Cas system. (B) Schematic representation of heterologous loci in two plasmids used for the co-expression of the Cas9-crRNA complex. E.coli RR1 strain contained pCas9(-)1SP (encoding Cas1, Cas2, Csn2, SP1 and tracrRNA) and pASKIBA-Cas9 (encoding Strep-tagged version of Cas9) plasmids. (C) Northern analysis of Cas9-crRNA complexes using anti-crDNA oligonucleotide as a probe. M1 - 84 nt oligodeoxynucleotide corresponding to the spacer S1-repeat unit; M2 – 42 nt synthetic oligoribonucleotide corresponding to the predicted *S.thermophilus* CRISPR3 crRNA (See Figure 4); crRNA (wt) – crRNA isolated from the wt Cas9 complex; K1 - crRNA (wt) treated with Dnase I for 15 min; K2 - crRNA (wt) treated with RNaseI for 15 min, D31A - crRNA purified from the Cas9 D31A mutant complex; N891A - crRNA purified from the Cas9 N891A mutant complex.

Figure 2 shows DNA cleavage by Cas9-crRNA complexes obtained by Cas9 co-expression with full length CRISPR locus. (A) Schematic representation of CRISPR/Cas locus of recombinant pCas9(-) plasmid carrying indigenous 12 spacer-repeat array of SthCRISPR3/Cas system and pASKIBA-Cas9 plasmid carrying cas9 gene with a Strep-tag at the C-terminus. (B) Oligoduplex cleavage assay. Both pCas9(-) and pASKIBA-Cas9 plasmids were co-expressed in E.coli, Cas9-crRNA complexes were purified and subjected to cleavage analysis using SP1 (first proto-spacer) and SP2 (second proto-spacer) oligoduplexes labeled with 33P at the 5'-end of the (+) strand. Reaction products were analysed on PAA gel.

Figure 3 shows immunity against plasmid transformation in *E. coli* cells provided by the SthCRISPR3/Cas system. (A) Schematic representation of CRISPR/Cas locus of recombinant plasmid pCRISPR3 carrying indigenous 12 spacer-repeat array of SthCRISPR3/Cas system and engineered pCRISPR3-SP1 plasmid carrying 1 spacer-repeat unit. (B) Interference of plasmid transformation by SthCRISPR3/Cas system in *E. coli* cells. Escherichia coli RR1 recipient strains carrying plasmids pACYC184, pCRISPR3 or pCRISPR3-SP1, were transformed with plasmid pSP1 carrying proto-spacers and PAM or pUC18 (1). Transformation efficiency is expressed as cfu per nanogram of plasmid DNA (mean ± SD).

Figure 4 shows comparison of Type IIA CRISPR/Cas systems from *S. thermophilus* DGCC7710, LMD-9 and *S. pyogenes* SF370 strains. (A) Schematic organization of the CRISPR/Cas systems. Nucleotide sequences corresponding to the tracrRNA required for the crRNA maturation in of *S. pyogenes* (2) are present in LMD-9 and DGCC7710. Percentage of identical and similar (in parenthesis) residues between corresponding protein sequences that are connected by dashed lines. (B). Alignment of the conserved repeat sequences and tracrRNA. Corresponding sequences from DGCC7710 and LMD-9 are identical. Nucleotide positions which are identical in all three strains are labeled with an asterisk below aligned sequences. Figure 4(B) discloses SEQ ID NOS 50, 50-52, and 52-53, respectively, in order of appearance. (C) Comparison of crRNA sequences. The sequence and length of *S. pyogenes* crRNA was determined by deep sequencing analysis (2). The approximate length of crRNA from *S. thermophilus* LMD-9 (2) and DGCC7710 (this work) strains were determined by the northern blot analysis. Figure 4(C) discloses SEQ ID NOS 54-56, respectively, in order of appearance.

Figure 5 shows Cas9-crRNA complex cleaves *in vitro* double-stranded DNA within a proto-spacer. (A) Oligoduplex substrate used in the cleavage assay. 55 nt oligoduplex SP1 contains the proto-spacer1 (red letters), PAM (blue letters) and 10 nt flanking sequences on both sides identical to those in pSP1 plasmid. In the SP1 oligoduplex DNA strand complimentary to the 5'-terminal fragment of crRNA (red letters) is named (+)strand, an opposite DNA strand is named (-)strand. Figure 5(A) discloses SEQ ID NOS 31, 7, and 34, respectively, in order of appearance. (B) Oligoduplex SP1 cleavage. 2.5 nM of Cas9-crRNA complex and 1 nM SP1 oligoduplex labeled with 33P at the 5'-end of either (+) or (–)strand were incubated in the reaction buffer (10 mM Tris-HCl pH=7.5, 10 mM NaCl, 10 mM MgCl2, 0.1 mg/ml BSA) at 37°C for varied time intervals (30 s to 10 min) and reaction products analysed in the 20 % PAA gel. Lanes M1 and M2 contain chemically synthesized 5'-end 33P-labeled 37 nt and 18 nt oligodeoxynucleotides corresponding to the cleavage products of (-) and (+) DNA strands, respectively. Cleavage positions are designated by arrows. Figure 5(B) discloses SEQ ID NO: 31. (C) Schematic representation of pSP1 plasmid (Sapranauskas et al., 2011. Nucleic Acids Res 39:9275–82) used in the plasmid cleavage assay. Figure 5(C) discloses SEQ ID NO: 57. (D) pSP1 plasmid cleavage. Agarose gel analysis of pSP1 cleavage products (left panel). SC – super-coiled plasmid DNA, OC – open circular DNA nicked at one of the strands, FLL – full length linear DNA cut at both strands. Final reaction mixtures at 37°C contained 2.5 nM of pSP1 plasmid and 2.5 nM of Cas9-crRNA complex in the reaction buffer (section B). Direct sequencing electropherograms (right panel) of (+) (upper part) and (–) (lower part) strands of pSP1 plasmid cleavage product. The non-templated addition of adenine (T in the reverse complement sequence shown here) at the extremity of sequence is a sequencing artefact caused by the polymerase. Figure 5(D) discloses SEQ ID NOS 57-59, 58, and 60, respectively, in order of appearance.

Figure 6 shows DNA binding and cleavage analysis of Cas9-Chis protein lacking crRNA. Electrophoretic mobility shift analysis (EMSA) of Cas9-Chis protein binding to (A) the double stranded SP1 oligoduplex and (B) the single stranded s(+)SP1 oligonucleotide. Electrophoretic mobility shift experiments were performed in the binding buffer (40 mM Tris–acetate, pH 8.3 at 25 C, 0.1 EDTA, 0.1 mg/ml BSA, 10% v/v glycerol). The reactions contained 0.5 nM of the 33P-labelled oligoduplex, and the protein at concentrations as indicated above each lane. (C). Oligonucleotide cleavage assay. 5 nM of Cas9-Chis protein was incubated in the reaction buffer (10 mM Tris-HCl, pH=7.5, 10 mM NaCl, 10 mM MgCl2, 0.1 mg/ml BSA) at 37°C with 1 nM oligonucleotide. SP1 oligoduplex was labeled with 33P at the 5'-end of the (+) or (–) strand. Single stranded oligonucleotide s(+)SP1 was labeled with 33P at the 5'-end.

Figure 7 shows reprograming of Cas9-crRNA complex. (A) Schematic representation of heterologous loci in two plasmids used for reprogramming of Cas9-crRNA complex. pCas(-)SPN were constructed from pCas9(-) plasmid (See Figure 2A), by inserting new spacer sequence (SN) (5'-CC ACC CAG CAA AAT TCG GTT TTC TGG CTG-3' (SEQ ID NO: 16)) and inactivating Cas9 gene as described in (1). (B) Agarose gel analysis of plasmid DNA cleavage products. pSP1 and pSP1+SPN (pSP1 plasmid with inserted new proto-spacer and PAM over AatII site were incubated at 2.5 nM concentration with 2 nM of Cas9-crRNA complex in the reaction buffer (10 mM Tris-HCl pH=7.5, 10 mM NaCl, 10 mM MgCl2, 0.1 mg/ml BSA) at 37°C for varied time intervals and reaction products analysed in the agarose gel. SC – super-coiled plasmid DNA, OC – open circular DNA nicked at one of DNA strands, FLL – full length linear DNA cut at both strands. (C) Oligoduplex SP1 cleavage. 2.5 nM of Cas9-crRNA complex and 1 nM SPN oligoduplex (Table S2) labeled with 33P at the 5'-end of either (+) or (–)strand were incubated in the reaction buffer (10 mM Tris-HCl pH=7.5, 10 mM NaCl, 10 mM MgCl2, 0.1 mg/ml BSA) at 37°C. M1 – 18 nt length marker Lanes M1 and M2 contain chemically synthesized 5'-end 33P-labeled 18 nt and 37 nt oligodeoxynucleotides corresponding to the cleavage products of (+) and (-) DNA strands, respectively. (D) Schematic representation of SPN oligoduplex substrate and cleavage products. SPN oligoduplex contains the new proto-spacer (red letters), PAM (blue letters). Cleavage positions are designated by arrows. Figure 7(D) discloses SEQ ID NO: 39.

Figure 8 shows impact of spacer length on CRISPR-encoded immunity. (A) Schematic representation of shortened versions of proto-spacers inserted in the transformed plasmids. Figure 8(A) discloses SEQ ID NOS 7 and 61-66, respectively, in order of appearance. (B) Effect of proto-spacer length on the plasmid transformation efficiency. Transformation efficiency is expressed as cfu per nanogram of plasmid DNA (mean ± SD). (C). Schematic representation of oligoduplexes used in the *in vitro* cleavage and binding experiments. Figure 8(C) discloses SEQ ID NOS 31 and 38, respectively, in order of appearance. (D) Time courses of the 27 bp oligoduplex (full length protospacer SP1, filled circles) and the 20 bp oligoduplex (truncated protospacer SP1-20, square) cleavage by the Cas9-crRNA complex. (E) Electrophoretic mobility shift assay of SP1 and SP1-20 oligoduplex binding by the Cas9-crRNA complex.

Figure 9 shows PAM is required for *in vitro* DNA binding and cleavage by the Cas9-crRNA complex. (A) Agarose gel analysis of plasmid DNA cleavage products. Three different plasmids: PAM+Proto-spacer+ (pSP1 plasmid containing both the proto-spacer and PAM), PAM-Protospacer- (pUC18 plasmid containing multiple PAMs but no protospacer) and PAM-Protospacer+ (pSP1-pΔ (Sapranauskas et al., 2011. Nucleic Acids Res 39:9275–82) containing a proto-spacer without PAM) were incubated at 2.5 nM concentration with 2 nM of Cas9-crRNA complex in the reaction buffer (10 mM Tris-HCl pH=7.5, 10 mM NaCl, 10 mM MgCl2, 0.1 mg/ml BSA) at 37°C for varied time intervals and reaction products analysed in the agarose gel. SC – super-coiled plasmid DNA, OC – open circular DNA nicked at one of DNA strands, FLL – full length linear DNA cut at both strands. (B) Time courses of (+)strand hydrolysis in the single-stranded and double-stranded oligodeoxynucleotides. Reactions containing 2 nM Cas9-crRNA and 1 nM of oligodeoxynucleotide were conducted at 37°C in the reaction buffer (section A). SP1 (filled circles) and SP1-pΔ (open squares) oligoduplexes were used as dsDNA. s(+)SP1 (open triangles) and s(+) SP1-pΔ (filled squeres) were used as ssDNA. (C) and (D) dsDNA and ssDNA (+)strand) binding by Cas9-crRNA complex. The reactions contained 0.5 nM of the 33P-labelled ssDNA or dsDNA oligonucleotide, and the protein at concentrations as indicated above each lane. After 15 min at room temperature, the samples were subjected to PAGE for 2 h and analysed as described in 'Materials and Methods'

Figure 10 shows RNA binding and cleavage analysis of Cas9-crRNA complex. (A) Electrophoretic mobility shift analysis (EMSA) of Cas9-crRNA complex binding to 84 nt RNA fragment containing proto-spacer-1, PAM and 24 nt flanking sequences on both sides. Left panel: RNA (–) strand; center panel: RNA (+) strand; right panel: double stranded RNA. RNA fragments used for analysis were generated by *in vitro* transcription (TranscriptAid™ T7 High Yield Transcription Kit, Fermentas) from

PCR fragments with inserted T7 promoter at the front end of RNA coding sequence. PCR fragments coding (+) and (-) RNA strands were obtained from pSP1 plasmid (1) with following primer pairs accordingly: 5' taatacgactcactataGggtaccgagctcgaattg 3' (SEQ ID NO: 17)/5' GGGAAACAGCTATGACCATGATTACGAATTC -3' (SEQ ID NO: 18) and 5' gggtaccgagctcgaattgaaattcTAAACG 3' (SEQ ID NO: 19)/5' taatacgactcactataGggAAACAGCTATGACCATGATTACG 3' (SEQ ID NO: 20) (T7 RNA polymerase promoter underlined, transcription start on bold). The reactions contained 1 nM of the 33P-labelled RNA fragment, and the protein at concentrations as indicated above each lane. After 15 min at room temperature, the samples were subjected to PAGE for 2 h and analyzed as described in 'Materials and Methods'. (B) RNA cleavage assay. 2.5 nM of Cas9-crRNA complex was incubated in the reaction buffer (10 mM Tris-HCl pH=7.5, 10 mM NaCl, 10 mM MgCl2, 0.1 mg/ml BSA,) at 37°C in the presence of 1 nM (+) and (-) RNA strands(left panel) or double stranded RNA labeled on (+) or (–) strand (right panel). Reaction products were analysed on denaturing PAA gel.

Figure 11 shows RuvC and HNH active site motifs of Cas9 contribute to the cleavage of opposite DNA strands. (A) Localization of the conserved active site motifs within Cas9 protein. Amino acid residues identified as crucial for Cas9 *in vivo* activity (Sapranauskas et al., 2011. Nucleic Acids Res 39:9275–82) are indicated. (B). Agarose gel analysis of pSP1 plasmid cleavage by Cas9 and mutant proteins. Reactions were performed as described in and 'Materials and Methods' (C) Strand preference of D31A mutant. Reactions were performed as described in Figure 2A and 'Materials and Methods'. D31 mutant cleaves only (+)strand of SP1 oligoduplex . Figure 11(C) discloses SEQ ID NOS 31 and 67, respectively, in order of appearance. (D) Strand preference of N891A mutant. N891 mutant cleaves only (-)strand of SP1 oligoduplex. Cleavage positions are designated by arrows. Figure 11(D) discloses SEQ ID NOS 31 and 68, respectively, in order of appearance.

Figure 12 shows properties of Cas9 active site mutant-crRNA complexes. (A) Direct sequencing of reaction products obtained with Cas9 mutant D31A (RuvC-like active site motif). Figure 12(A) discloses SEQ ID NOS 58, 59, 58, and 58, respectively, in order of appearance. (B) Direct sequencing of reaction products obtained with Cas9 N891A mutant (HNH-like active site motif). Figure 12(B) discloses SEQ ID NOS 58, 58, 58, and 60, respectively, in order of appearance. (C) SP1 oligoduplex binding by the wt Cas9-crRNA and active site mutant complexes. (D) Cleavage of (+)SP1 strand by Cas9-crRNA mutant complexes.

Figure 13 shows molecular mass of the wt Cas9-Chis protein. Gel filtration experiments were carried out at room temperature using Superdex 200 10/300 GL column (GE healthcare) pre-equilibrated with 10 mM sodium phosphate (pH 7.4) buffer containing 500 mM sodium chloride.The apparent Mw of Cas9 (black triangle) were calculated by interpolation from the standard curve obtained using a set of proteins of known Mw (black circles) (Bio-Rad Gel Filtration Standards).

Figure 14 shows schematic arrangement and mechanism of crRNA-directed DNA cleavage by the Cas9-crRNA complex. Domain architecture of Cas9 is shown schematically on the top. Cas9-crRNA complex binds to the dsDNA containing PAM. crRNA binds to the complementary (+)strand resulting in DNA strand separation and the R-loop formation. In the ternary complex RuvC active site of Cas9 is positioned at the scissile phosphate on the unpaired (-)strand, while HNH active site is located at the scissile phosphate on the DNA (+)strand bound to crRNA. Coordinated action of both active sites results in the double strand break 4 nt away from the PAM generating blunt end DNA. Figure 14 discloses SEQ ID NOS. 31 and 69, respectively, in order of appearance.

Figure 15 shows native electrophoresis of Cas9-crRNA and cleavage products. The protein at concentrations as indicated above each lane, where incubated in the reaction buffer (10 mM Tris-HCl pH=7.5, 10 mM NaCl, 10 mM MgCl2, 0.1 mg/ml BSA) at 37°C for 30 min in the presence of 0.5 nM SP1 oligoduplex. Samples was mixed with loading dye solution (0.01 % bromphenol blue and 75 mM EDTA in 50 % v/v glycerol) and analysed by non-denaturing PAGE. The gel lanes marked M – melted form of cleavage reactions products. The cartoons in each side of the gel illustrate protein-DNA complexes and DNA that correspond to each band, while cartoons below the gel illustrate major substrate form after reaction.

Figure 16 shows plasmid DNA cleavage by Cas9-crRNA complex. (A) pSP1 and pUC18 plasmid DNA cleavage. Cas9-crRNA complex was incubated with pSP1 and pUC18 plasmids in a reaction buffer provided in the Example 1. pSP1 plasmid contained a proto-spacer1 sequence flanked by the the 5'-GGNG-3'PAM sequence. Proto-spacer1 sequence was not present in pUC18. Reaction products were analysed in the agarose gel. Under these conditions pSP1 plasmid is converted into a linear form while pUC18 plasmid lacking proto-spacer1 sequence is resistant to cleavage. (B) pSP1 cleavage reactions in the absence of one of the components. In the reaction mixes lacking one of the components (Cas9, crRNA or tracrRNA, respectively) pSP1 plasmid is not cleaved. SC – super-coiled plasmid DNA, OC – open circular DNA nicked at one of DNA strands, FLL – full length linear DNA cut at both strands.

Figure 17 shows DNA oligoduplex cleavage by Cas9-crRNA complex. The strand of oligoduplex which is complementary to crRNA is marked as (+) strand, while the other strand - (-) strand. To monitor cleavage reactions either (+) or (–) strand of the oligoduplex was P33-labeled at the 5'-terminus. M1 and M2 are synthetic oligonucleotide markers corresponding to the 37 nt of (-) strand and 18 nt of (+) strand which were used to determine the size of the cleavage products and map the cleavage position. Cas9 protein cleaves both strands of oligoduplex inside the proto-spacer, after the 37th nucleotide, 4 nt upstream of the PAM (5'-GGNG-3') leaving blunt ends. Both strands of non-specific substrate (K1 and K2) are not cleaved when incubated with Cas9-crRNA complex for 30 min. Figure 17 discloses SEQ ID NO: 31.

Figure 18 shows plasmid DNA cleavage by Cas9-crRNA complex assembled in the absence of RNaseIII. Cas9-crRNA complex was incubated with pSP1 plasmid and reaction products analysed in the agarose gels. The pSP1 plasmid is resistant for cleavage in the presence of complex assembled without crRNA (left panel). The pSP1 plasmid is converted into linear form in the presence of complex assembled using synthetic 42 nt crRNA (no RNAseIII) (middle panel). The pSP1 plasmid is converted into a mixture of linear and circular DNA forms in the presence of complex assembled using CRISPR RNA transcript (no RNAseIII) (right panel).

Figure 19 shows DNA oligoduplex cleavage by Cas9-crRNA complex. The strand of oligoduplex which is complementary to crRNA is marked as (+) strand, while the other strand - (-)strand. To monitor cleavage reaction either (+) or (–) strand of the oligoduplex was P33-labeled at the 5'-terminus. M1 and M2 are synthetic oligonucleotide markers corresponding to the 37 nt of (-) strand and 18 nt of (+) strand which were used to determine the size of the cleavage products and map the cleavage position. Cas9 protein cleaves both strands of oligoduplex inside the proto-spacer, after the 37th nucleotide form the 5'-end, 4 nt upstream of the PAM (5'-GGNG-3') leaving blunt ends. Both strands of non-specific substrate (K1 and K2) are not cleaved when incubated with Cas9-crRNA complex for 30 min. Figure 19 discloses SEQ ID NO: 31.

Figure 20 shows (A) Schematic representation of the CRISPR3/Cas system of *S. thermophilus* DGCC7710*.* Four *cas* genes (*cas9, cas1, cas2, csn2*) are located upstream of the CRISPR repeat-spacer array, consisting of 13 repeat (R) sequences and 12 unique spacers (S1-S12). The tracrRNA, required for crRNA maturation in Type II CRISPR/Cas systems (Deltcheva et al.,2011. Nature 471, 602–7), is located upstream the *cas9* gene and encoded on the opposite DNA strand (shown by an arrow) with respect to the other elements of this system. (B) The pathways for a new spacer insertion in to CRISPR region and CRISPR RNA synthesis. Synthetic oligoduplex encoding desired spacer sequence and containing SapI and Eco31I restriction compatible ends was inserted between two repeats. The CRISPR region was amplified using PCR. The new spacer encoding CRISPR RNA was obtained by *In vitro* transcription. (C) *In vitro* assembly of Cas9-RNA complex. The CRISPR RNA and tracrRNA transcripts were assembled in to duplex. The Cas9 protein was first pre-incubated with RNA duplex, followed by the subsequent incubation with RNAseIII to generate a catalytically competent Cas9-RNA complex.

Figure 21 shows A. Schematic representation of pUC18 plasmid. The distance between SapI and AatII restriction sites is 775 bp, while the distance between two spacers is 612 bp. B. pUC18 plasmid cleavage by re-programed Cas9-crRNA complexes. "1" – pUC18 plasmid; "2" – pUC18 cleaved with AatII; "3" – pUC18 cleaved with complex containing crRNA matching proto-spacer1; "4" – pUC18 cleaved with SapI; "5" – pUC18 cleaved with complex containing crRNA matching proto-spacer2; "6" – pUC18 cleaved with AatII and SapI; "7" – pUC18 cleaved with mix of the complexes used in the line 3 and 5.

Figure 22 shows genomic DNA cleavage with *in vitro* assembled Cas9-RNA complex. (A) Agarose gel analysis of linear λ DNA cleavage products. Phage λ DNA was incubated with Cas9-RNA complex in the reaction buffer for various time intervals. The target site for Cas9-RNA complex is located 8 kb away from the *cos* site. (B). Probe selection for Southern blot experiments. Genomic DNA was fragmented by treating with PstI enzyme. The proto-spacer is located between two PstI sites. If genomic DNA is cleaved with Cas9-RNA complex, 466 bp fragment should be detected. Otherwise the probe will hybridize with 1499 bp length fragment. (C) Southern blot analysis of genomic DNA fragments. C line - *E. coli* genomic DNA fragmented with PstI. Cas9-RNA – genomic DNA was incubated with Cas9-RNA complex before fragmentation. (D). Human genomic DNA cleavage by Cas9-crRNA complex. Relative amount of intact DNA DNA fragments were estimated by qPCR.

Figure 23 schematically illustrates targeting sequences contained in the reporter plasmid (pMTC-DSR+eGFP). eGFP coding sequence is separated by an intron from GAPDH gene. The 5' and 3' RFP coding sequences are indicated. *homol* indicates homologous sequences in the RFP gene necessary for homologous recombination to occur. A, B, C, and D indicate four distinct target sites for Cas9-mediated cleavage. Targets A and B are located in the intron. Targets C and D are located in the coding regions of eGFP. Cre indicates a target site for Cre endonuclease and is located in the intronic sequence.

Figure 24 shows reduction of eGFP-positive cells after introduction of Cas9/RNA complexes. CHO-K1 cells were transfected with the reporter plasmid and Cas9/RNA complexes containing crRNA targeting either eGFP sequence A (intronic), eGFP sequence C (coding), or a non-specific sequence K. The percentage of eGFP-positive cells was determined by flow cytometry. As negative controls, cells were untransfected (NC) or transfected with the reporter plasmid alone (DNA) or with reporter plasmid and Cas9 protein alone as well as with reporter plasmid and Cas9-nonspecific crRNA complex (DNA+K).

Figure 25 shows cell images where appearance of RFP suggested Cas9/RNA-mediated double-strand break repair by homologous recombination (HR). Forty-eight hours after co-transfection with the reporter plasmid and Cas9/RNA complexes targeting eGFP sequence C, CHO-k1 cells were visualized by fluorescence microscopy for eGFP and RFP.

Figure 26 schematically illustrates targeting sequences contained in the reporter plasmid (pMTC-DSR+eGFP). eGFP coding sequence is separated by GAPDH intron copied from genomic DNA. The RFP N- and C- coding sequences are as indicated. Homologous sequences in the RFP gene are necessary for homologous recombination to occur. Target E located within the intron of eGFP is indicated in bold.

Figure 27 is a gel showing Cas9/RNA complexes using synthetic crRNA and tracRNA function similarly to Cas9/RNA complexes using synthetic crRNA and *in vitro* transcribed tracrRNA. Plasmids were visualized after agarose gel electrophoresis. Lane C: uncut plasmid. Lanes 1-3: plasmids cut with Cas9+crRNA and either 1: control *in vitro*-transcribed tracrRNA; 2: unmodified synthetic tracrRNA (89 nt); or 3: unmodified synthetic tracrRNA (74 nt).

Figures 28A-E schematically show targeting sequences contained in the reporter plasmid (pMTC-DSR+eGFP) and potential processing / gene rearrangement outcomes.

Figure 29 shows reduction of eGFP-positive cells after introduction of Cas9/RNA complexes. CHO-K1 cells were transfected with the reporter plasmid and Cas9/RNA complexes containing crRNA targeting either eGFP sequence A (intronic), eGFP sequence C (coding), or a non-specific sequence K. The percentage of eGFP-positive cells was determined by flow cytometry. As negative controls, cells were untransfected (NC) or transfected with the reporter plasmid alone (DNA) or with reporter plasmid and Cas9 protein alone as well as with reporter plasmid and Cas9-nonspecific crRNA complex (DNA+K).

Figure 30 schematically shows targeting sequences contained in the reporter plasmid (pMTC-DSR+eGFP). eGFP coding sequence is indicated in black and is separated by GAPDH intron copied from genomic DNA. The RFP N- and C- coding sequences are indicated in gray. Homologous sequences in the RFP gene (light grey) are necessary for homologous recombination to occur. Target E located within the intron of eGFP is indicated in bold.

The following non-limiting examples further describe the methods, compositions, uses, and embodiments.

Detailed description of illustrative embodiments

Example 1.

In this example, we have isolated the Cas9-crRNA complex of *S. thermophilus* CRISPR3/Cas system and demonstrate that it cuts in a PAM dependent manner both synthetic oligodeoxynucleotide and plasmid DNA bearing a nucleotide sequence complementary to the crRNA. Furthermore, we provide experimental evidence that PAM is recognized in the context of double-stranded DNA and is critical for *in vitro* DNA binding and cleavage. Finally, we show that RuvC and HNH- motifs of Cas9 contribute to the cleavage of opposite DNA strands. Taken together, our data demonstrate that Cas9-crRNA complex functions as RNA-guided endonuclease which uses RNA module for the target site recognition and employs two separate active sites in the protein module for DNA cleavage. These findings pave the way for engineering of programable Cas9-crRNA complexes as universal RNA-guided endonucleases.

Materials and methods

DNA manipulations. Genomic DNA of *Streptococcus thermophilus* DGCC7710 strain was used as a template in PCR reactions to clone cas9. To generate a pASKIBA3-Cas9 plasmid which was used for the expression of the C-terminal Strep-tagged Cas9 protein variant, PCR fragment amplified with following primers: 5′- ACGTCTCAAATGTTGTTTAATAAGTGTATAATAATTTC-3′ (SEQ ID NO: 21) and 5′-ACGTCTCCGCGCTACCCTCTCCTAGTTTG-3′ (SEQ ID NO: 22) was cloned into the pASK-IBA3 expression vector via Esp3I sites. To generate a pBAD-Cas9 plasmid which was used for the expression of the C-terminal 6xHis-tagged Cas9 protein variant ("6xHis" disclosed as SEQ ID NO:

23), PCR fragment amplified with the following primer pair: 5′-ACGTCTCACATGACTAAGCCATACTCAATTGGAC -3′ (SEQ ID NO: 24) and 5′-ACTCGAGACCCTCTCCTAGTTTGGCAA -3′ (SEQ ID NO: 25) was cloned into the pBAD24-Chis expression vector via NcoI and XhoI sites. Full sequencing of cas9 gene in pASKIBA3-Cas9 and pBAD-Cas9 plasmids revealed no difference with the original cas9 sequence. To obtain plasmids pCas9(-)SP1 (Figure 1B) and pCRISPR3-SP1 (Figure 2A), bearing a single spacer1, PCR fragment amplified from pCRISPR3 plasmid with the following primer pair:5' GACCACTTATTGAGGTAAATGAG 3' (SEQ ID NO: 26)/5' CAAACCAGGATCCAAGCTAATACAGCAG-3' (SEQ ID NO: 27) ((BamHI (GGATCC) sites is underlined) was cloned into pCas9(-) and pCRISPR3 plasmids (Sapranauskas et al., 2011. Nucleic Acids Res 39:9275–82), respectively.

Expression and purification of Cas9 protein and Cas9-crRNA complex. (His)6-tagged ("(His)6" disclosed as SEQ ID NO: 23) version of Cas9 protein was expressed and purified using a scheme described for the Cas3 protein from *S. thermophilus* CRISPR4/Cas system (Sinkunas et al.,2011. EMBO J 30:1335–42). For purification of the Cas9-crRNA complex, Strep-tagged version of the Cas9 protein was expressed in *E. coli* RR1 strain, bearing pCas9(-)SP1 plasmid (Figure 1B). LB broth was supplemented with Ap (100 µg/ml) and Cm (10 µg/ml). E.coli cells for the Cas9-crRNA complex isolation were grown in two steps. First, 4 ml of cells culture were grown at 37°C to OD600 of ~0.5, and expression induced by adding 0.2 µg/ml of anhydrotetracycline (AHT) (Sigma). After for 4 h, 1/400 of the pre-induced culture was inoculated into fresh LB medium supplemented with Ap (100 µg/ml), Cm (12 µg/ml) and AHT (0.2 µg/ml) and was grown at 37°C overnight. Harvested cells were disrupted by sonication and cell debris removed by centrifugation. The supernatant was loaded onto the 1 ml StrepTrap HP column (GE Healthcare) and eluted with 2.5 mM of desthiobiotin. Approximately 1.5 µg of the Cas9 protein was obtained in a single run from 1 L of *E. coli* culture. The fractions containing Cas9 were stored at + 4°C for several days. The homogeneity of protein preparations was estimated by SDS-PAGE. Protein concentrations in the Cas9-crRNA complexes were determined by densitometric analysis of SDS-PAGE gels containing samples of Strep-Tactin purified Cas9 proteins along with known amounts of His-tagged Cas9 protein. The concentration of the Cas9-crRNA complexes is expressed as Cas9 protein concentration assuming that Cas9 is a monomer and binds crRNA in a complex with 1:1 stoichiometry.

Northern blot analysis. Cas9-bound RNA was isolated from Strep-Tactin purified Cas9, co-expressed with pCas9(-)SP1 plasmid using the miRNeasy Mini kit (Qiagen). Northern blots were performed by running RNA on a 10 % polyacrylamide gel with 7 M urea in 20 mM MOPS/NaOH pH 8 buffer. The RNA was transferred to a SensiBlot™ Plus Nylon Membrane (Fermentas) by semi-dry blotting using a Trans-blot SD (Bio-Rad). RNA was cross-linked to the membrane with 0.16 M l-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC) (Pierce)/0.13 M 1-methylimidazole (Sigma) pH 8 at 60°C for 1 h. The membrane was pre-hybridized with 2× SSC buffer containing 1% SDS and 0.1 mg/ml denatured DNA from fish testes (Ambion) for 1 h at 40°C. Blots were probed for 12 h with a ^{32}P-5′-labelled 42 nt anti-crRNA DNA oligonucleotide containing 20 nt of spacer1 and 22 nt of the repeat sequence (5′-TCGAAACAACACAGCTCTAAAACTGTCCTCTTCCTCTTTAGC-3′ (SEQ ID NO: 28)). The blots were washed 3× for 15 min with 0.2× SSC buffer containing 0.2% SDS, and were visualized using phosphorimaging. A 42 nt synthetic oligoribonucleotide (5′-CGCUAAAGAGGAAGAGGACAGUUUUAGAGCUGUGUUGUUUCG-3′ (SEQ ID NO: 7)) and 84 nt DNA oligonucleotide.

Oligonucleotide substrates. All oligonucleotide substrates used in this study are given in Table 1. Oligodeoxyribonucleotides were purchased from Metabion (Martinsried, Germany). The 5′-ends of oligonucleotides were radiolabelled using PNK (Fermentas) and [γ-33P]ATP (Hartmann Analytic). Duplexes were made by annealing two oligonucleotides with complementary sequences (SP1, SP1-Δp, SP2). Radioactive label was introduced at the 5' end of individual DNA strand prior to the annealing with unlabelled strand.

Reactions with oligonucleotide substrates. Reactions were typically carried out by adding 2 nM of Cas9-crRNA complex to 1 nM labeled oligonucleotide in 10 mM Tris-HCl (pH 7.5 at 37°C), 10 mM NaCl, 0.1 mg/ml BSA and 10 mM MgCl2 at 37 °C. Aliquots were removed at timed intervals and quenched with loading dye (95 % v/v formamide, 0.01 % bromphenol blue, 25 mM EDTA, pH 9.0) and subjected to denaturing gel electrophoresis through 20 % polyacrylamide followed by a FLA-5100 phosphorimager (Fujilm) detection.

Reactions with plasmid substrates. Reactions on pUC18 plasmid and its derivatives (Sapranauskas et al., 2011. Nucleic Acids Res 39:9275–82) were conducted at 37 °C in the buffer used for reactions on oligonucleotide substrates. Reaction mixtures typically contained 2.5 nM supercoiled plasmid and 2 nM of Cas9-crRNA complex. The reactions were initiated by adding protein to the mixture of the other components. Aliquots were removed at timed intervals and quenched with phenol/chloroform. The aqueous phase was mixed with loading dye solution (0.01 % bromphenol blue and 75 mM EDTA in 50 % v/v glycerol) and analyzed by electrophoresis through agarose.

Plasmid cleavage position determination. To achieve complete cleavage of plasmid substrate, 8 nM of Cas9-crRNA complex was incubated with 2.5 nM of supercoiled plasmid in the reaction buffer at 37 °C for 10 min. Reaction products were purified and concentrated using GeneJET PCR Purification Kit (Fermentas). Spacer1 surrounding region of Cas9 linearized and nicked plasmids were directly sequenced with the following primers: 5’-ccgcatcaggcgccattcgcc-3’ (SEQ ID NO: 29) (sequencing of (+)strand) and 5’-gcgaggaagcggaagagcgccc-3’ (SEQ ID NO: 30) (sequencing of (-)strand).

Binding assay. Increasing amounts of protein-crRNA complex were mixed with 0.5 nM of 33P-labeled double-stranded and single-stranded DNA substrates (Table 1) in the binding buffer (40 mM Tris–acetate, pH 8.3 at 25 C, 0.1 EDTA, 0.1 mg/ml BSA, 10% v/v glycerol) and incubated for 15 min at room temperature. Free DNA and protein–DNA complexes were separated on the non-denaturing 8% polyacrylamide gel (ratio of acrylamide/N,N'-methylenebisacrylamide 29:1) using 40 mM Tris–acetate (pH 8.3) supplemented with 0.1 mM EDTA as the running buffer. Electrophoresis was run at room temperature for 3 h at 6 V/cm.

Mutagenesis. The mutants D31A and N891A were obtained by the site-directed mutagenesis as previously described (Tamulaitis et al., 2007. Nucleic Acids Res 35:4792–9). Sequencing of the entire gene for each mutant confirmed that only the designed mutation had been introduced.

Table 1. Oligonucleotide substrates. Proto-spacer sequence is underlined, PAM is on bold.

Oligonucleotide	Sequence	Specification
SP1 (SEQ ID NO: 31)	5’ - GCTCGAATTGAAATTCTAAACGCTAAAGAGGAAGAGGACAT**GG**T**G**AATTCGTAAT -3' 3’ - CGAGCTTAACTTTAAGATTTGCGATTTCTCCTTCTCCTGTA**CC**A**C**TTAAGCATTA -5’	55 bp oligoduplex substrate containing proto-spacer1 and PAM
SP1-pΔ (SEQ ID NO: 32)	5’ - GCTCGAATTGAAATTCTAAACGCTAAAGAGGAAGAGGACAAATTCGTAAT -3’ 3’ - CGAGCTTAACTTTAAGATTTGCGATTTCTCCTTCTCCTGTTTAAGCATTA -5'	50 bp oligoduplex substrate containing proto-spacer2
SP2 (SEQ ID NO: 33)	5’- GCTCGAATTGTACTGCTGTATTAGCTTGGTTGTTGGTTTGT**GG**T**G**AATTCGTAAT -3' 3’- CGAGCTTAACATGACGACATAATCGAACCAACAACCAAACA**CC**A**C**TTAAGCATTA -5’	55 bp oligoduplex substrate containing proto-spacer2 and PAM (oligodublex without proto-spacer1)
s(+) SP1 (SEQ ID NO: 34)	5'- ATTACGAATTCACCATGTCCTCTTCCTCTTTAGCGTTTAGAATTTCAATTCGAGC-3'	55 nt ssDNA oligonucleotide substrate (+) strand of SP1 oligoduplex
s(+) SP1-pΔ (SEQ ID NO: 35)	5'- ATTACGAATTTGTCCTCTTCCTCTTTAGCGTTTAGAATTTCAATTCGAGC-3'	50 nt ssDNA oligonucleotide substrate (+) strand of SP1-pΔ oligoduplex
s(+) SP2 (SEQ ID NO: 36)	5'- ATTACGAATTCACCACAAACCAACAACCAAGCTAATACAGCAGTACAATTCGAGC-3'	55 nt ssDNA oligonucleotide substrate, (+) strand of SP2 oligoduplex
s(-) SP1 (SEQ ID NO:	5'- GCTCGAATTGAAATTCTAAACGCTAAAGAGGAAGAGGACAT**GG**T**G**AATTCGTAAT -3'	55 nt ssDNA oligonucleotide substrate, (-) strand of SP1

37)		oligoduplx
SP1-20 (SEQ ID NO: 38)	5′- GCTCGAATTGCGCTAAAGAGGAAGAGGACAT**GG**T**G**AATTCGTAAT -3' 3′- CGAGCTTAACGCGATTTCTCCTTCTCCTGTA**CC**A**C**TTAAGCATTA -5'	45 nt oligoduplex substrate containing 20 nt of proto-spacer1 and PAM
SPN (SEQ ID NO: 39)	5′- GCTCGAATTGCCACCCAGCAAAATTCGGTTTTCTGGCTGAT**GG**T**G**AATTCGTAAT -3′ 3′- CGAGCTTAACGGTGGGTCGTTTTAAGCCAAAAGACCGACTA**CC**A**C**TTAAGCATTA -5'	55 bp oligoduplex substrate containing proto-spacerN and PAM

Results

Expression and purification of the Cas9-crRNA complex. The cas9 gene from the CRISR3 system of *S. thermophilus* DGCC7710 strain was cloned into the pASK-IBA3 vector to produce a construct encoding a Cas9 protein fusion containing a C-terminal Strep(II)-tag (Figure 1B). Initially, we have tried to purify Cas9-crRNA complex from *E. coli* strain RR1 expressing Cas9 protein on the pASK-IBA3 vector and other Cas proteins (except Cas9) on pCas9(-) plasmid (Sapranauskas et al, 2011). pCas9(-) also contained a complete CRISPR3 array comprised of 12 spacer-repeat units (Figure 2A). To achieve simultaneous transcription of all target genes we performed cas9 gene expression in two steps. First, we induced Cas9 expression in a small volume of *E. coli* culture and after 4 h transferred an aliquot of pre-induced culture into a larger volume of fresh LB media already containing inductor and incubated overnight. Cas9 protein complex was purified from the crude cell extract using Strep-Tactin Sepharose. We managed to isolate a small amount of the Cas9-crRNA complex which showed only traces of nucleolytic activity on the oligoduplex SP1 containing a proto-spacer1 and PAM. We assumed that low cleavage activity could be due to the intrinsic heterogeneity of Cas9-crRNA complexes resulting from the transcription of 12 spacer-repeat units. If all spacer-repeat units are uniformly transcribed into a mature crRNA, the concentration of the Cas9 complex containing crRNA against spacer-1 will make 1/12th fraction of the total Cas9-crRNA concentration. The cleavage activity of the Cas9-crRNA preparation against the SP2 oligoduplex containing a proto-spacer-2 and PAM is consistent with the heterogeneity of Cas9-crRNA complexes (Figure 2B). To increase the yield of the specific Cas9-crRNA complex we engineered a pCas9(-)SP1 plasmid which contains a single R-spacer1-R unit in the CRISPR array (Figure 1B). Plasmid transformation interference assay confirmed that the CRISPR3/Cas system carrying a single spacer1 prevents plasmid pSP1 transformation in *E. coli* with the same efficiency as the CRISPR3/Cas system carrying a complete CRISPR region (Figure 3B). We have isolated Cas9-crRNA complex following the procedure described above and analysed crRNA bound to Cas9 protein.

Cas9 protein co-purifies with crRNA. CRISPR3/Cas system of *S. thermophilus* belongs to the Type IIA subtype (former Nmeni or CASS4) of CRISPR/Cas systems (Makarova et al., 2011. Nat Rev Microbiol 9:467–77). It has been shown that in the Type IIA CRISPR/Cas system of *Streptococcus pyogenes* trans-encoded small RNA (tracrRNA) and bacterial RNaseIII are involved in the generation of crRNA (Deltcheva et al., 2011. Nature 471:602-7). *Streptococcus pyogenes* crRNA is only 42 nt in length and has no “5’-handle” which is conserved in crRNA’s from Type I and III CRISPR systems (Hale et al., 2009. Cell 139:945–56; Jore et al., 2011. Nat Struct Mol Biol 18:529–36). According to the northern blot analysis crRNA of similar length is generated in the *S. thermophilus* LMD-9 CRISPR3/Cas system (Makarova et al., 2011. Nat Rev Microbiol 9:467–77), which is almost identical to the CRISPR3/Cas system of DGCC7710 strain (Figures 4A and B). We assumed that crRNA isolated from the Cas9-crRNA complex expressed in the heterologous *E. coli* strain (Figure 1) may have the same length (Figure 4). Therefore, to probe nucleic acids extracted from the Strep-Tactin purified Cas9 complex we used 42 nt anti-crRNA DNA oligonucleotide comprised of 22 nt region corresponding to the 3'-end of the repeat sequence and 20 nt at the 5'-end of SP1 fragment. Nucleic acid present in the Cas9 complex hybridized with anti-crRNA oligonucleotide, and was sensitive to RNAse but not DNAse treatment (Figure 1C). The size of extracted crRNA was identical to the 42 nt synthetic oligoribonucleotide corresponding to the putative crRNA of the CRISPR3 system of *S. thermophilus* DGCC7710 strain (Figure 3A, Figure 4C). Taken together, these data confirm that Cas9 Strep-tag protein co-purifies with 42 nt crRNA, which is derived from CRISPR3 region.

Cas9 protein cleaves double-stranded DNA within a proto-spacer. To test *in vitro* activity of purified Cas9-crRNA complex we first used the SP1 oligoduplex (Table 1) containing the proto-spacer sequence identical to spacer SP1 in the CRISPR3 array, the PAM sequence 5' –TGGTG-3' downstream of the proto-spacer, and 10 nt flanking sequences from pSP1 plasmid (Sapranauskas et al., 2011. Nucleic Acids Res 39:9275–82) (Figure 5A). The oligoduplex strand complementary to crRNA is named (+) strand, while the opposite duplex strand is called the (-) strand. To monitor cleavage reaction either (+) or (-) strand of the SP1 oligoduplex was P33-labeled at the 5'-terminus. Data shown in Figure 5B demonstrate that the Cas9-crRNA complex cleaves both strands of oligoduplex at fixed position. Mapping of the cleavage position using synthetic oligonucleotides as size markers revealed that the Cas9-crRNA complex cuts both strands of the SP1 oligoduplex within the proto-spacer 4 nt upstream of the PAM (Figure 5B) leaving blunt ends. It is worth to note, that no cleavage is observed after the 2 h incubation of the SP1 oligoduplex with the Cas9 protein lacking crRNA (Figure 6C).

To test whether the Cas9-crRNA complex can locate the proto-spacer and cut DNA *in vitro* in long DNA substrates mimicking *in vivo* invading foreign DNA we analyzed cleavage of pSP1 plasmid (Sapranauskas et al., 2011. Nucleic Acids Res 39:9275–82) (Figure 5C) carrying proto-spacer1 and PAM. In the presence of Cas9-crRNA complex supercoiled form of pSP1 plasmid was converted into a linear form (Figure 5D), while pUC18 plasmid lacking proto-spacer1 was not cleaved. This means that both strands of the pSC1 plasmid were cleaved specifically within the proto-spacer region. We used direct sequencing to determine the ends of linear DNA form formed after the Cas9-crRNA cleavage. Sequencing results confirmed that cleavage of plasmid DNA occurred 4 nt away from PAM sequence similarly to the SP1 oligoduplex cleavage (Figure 5D). The cleavage positions identified in the *in vitro* experiments (Figure 4) for the CRISPR3/Cas system of *S. thermophilus* are identical to those determined in the *in vivo* cleavage experiments for the CRISPR1/Cas system in *S. thermophilus* (Garneau et al., 2010. Nature 468:67–71). To check if Cas9-crRNA induced cleavage occurs at the same position in other proto-spacer sequences, we analysed cleavage of the SP2 oligoduplex carrying a protospacer-2 and PAM sequences by the heterogeneous Cas9-crRNA complex isolated from the host carrying 12 spacer-repeat units. We have found that this heterogeneous Cas9-crRNA complex cuts (+)strand of SP2 oligoduplex exactly at the same position as in the SP1 oligoduplex.

Cas9-crRNA cleavage specificity is directed by the crRNA sequence. To demonstrate directly that Cas9-crRNA complex specificity can be re-programmed by changing crRNA in the ribonucleoprotein complex we inserted a new spacer (SN) instead of spacer S1 in the CRISPR region generating pCas(-)SN plasmid containing only a minimal CRISPR region and tracrRNA encoding sequence (Figure 7), co-expressed this plasmid together with pASKIBA-Cas9 and purified Cas9-crRNA complex. The cleavage specificity of Cas9-crRNA complex was analysed using plasmids pSP1+SPN and pSP1. pSP1+SPN plasmid containing the proto-spacer sequence matching the SN spacer in the CRISPR region, was linearized by the Cas9-crRNA complex, while pSP1 plasmid which lacks complimentary sequence remained intact (Figure 7B). To determine the cleavage position within the SPN spacer sequence, we performed experiments with SPN oligoduplex, containing proto-spacer complementary to spacer SN and PAM (Figure 7D). Oligoduplex cleavage assay confirmed (Figure 7C and D) that Cas9-crRNA complex with re-engineered specificity cleaves both DNA strands within the SN proto-spacer 4 nt upstream of the PAM identically to other Cas9-crRNA complexes.

The length of the spacer in the CRISPR3 region of *S. thermophilus* is 30 nt. According to the data provided in the Figure 1C, the mature crRNA copurified with the Cas9 protein is comprised of 42 nt. It means that only 20 nt of crRNA is complementary to the (+)strand of proto-spacer. To assess whether 5'-end of proto-spacer is important for the plasmid interference by the CRISPR3 system of *S. thermophilus* we engineered plasmids pSP1-27, pSP1-23, pSP1-19, pSP1-15, pSP1-11 with the 5'-truncated proto-spacer1 (the length of proto-spacer 27 bp, 23 bp, 19 bp, 15 bp, 11 bp, respectively), and analyzed transformation efficiency of the recipient strain containing pCRISPR3 (Figure 8B). Plasmids containing 4 or 7 bp truncations at the 5' end of proto-spacer1, had no effect on the recipient strain ability to interfere with plasmid transformation. Shorter versions of proto-spacer (11, 15, 19 bp) abolished recipient strain ability to prevent plasmid transformation. These data shows that 5' end of the proto-spacer, which has no complementarity to mature crRNA is not important for CRISPR3/Cas function. In full support to the *in vivo* experiments, the SP1-20 oligoduplex containing only 20 nt of the protospacer-1 is efficiently cleaved by Cas9-crRNA (Figure 8 D and E).

PAM is required for DNA binding and cleavage by Cas9-crRNA. Plasmids carrying a proto-spacer but not PAM (pSP1-pΔ) or multiple PAM's but no proto-spacer (pUC18) are resistant for Cas9-crRNA cleavage (Figure 8A). Hence, in accordance with *in vivo* data both PAM and proto-spacer are required for double-stranded DNA cleavage by Cas9-crRNA complex (Sapranauskas et al., 2011. Nucleic Acids Res 39:9275–82). To find out, whether PAM is recognized in a context of a double-stranded or a single-stranded DNA, we analyzed Cas9-crRNA binding and cleavage of oligodeoxynucleotides i) SP1 (containing both proto-spacer and PAM), ii) SP1-Δp (contains only proto-spacer), and iii) SP2 (contains only PAM). The (+)strands of these oligodeoxynucleotides were used as single-stranded DNA substrates (s(+)SP1, s(+)SP1-Δp, s(+)SP2, accordingly) (Table 1).

Consistent with the plasmid cleavage experiments, oligoduplexes which have only proto-spacer, but not PAM are not cut by Cas9-crRNA (Figure 9B). On the other hand, (+)strand in the single-stranded form is cut at the similar rate independently whether it has or has not PAM (Figure 9B). These data clearly show that PAM is required only for a double-stranded but not for a single-stranded DNA cleavage.

To test if PAM is important for DNA binding by the Cas9-crRNA complex, electrophoretic mobility shift experiments were performed. To avoid cleavage, binding experiments were performed in the absence of Mg2+ ions which are necessary for cleavage. Cas9-crRNA showed different binding patterns for double-stranded and single-stranded oligonucleotides. In the case of the SP1 oligoduplex a low mobility complex is observed already at 1 nM concentration (Figure 9C). On the other hand, no binding is observed under the same experimental conditions for oligoduplexes without PAM (SP1-Δp) or without proto-spacer (SP2). Moreover, no low mobility complex is observed in the case of Cas9 protein without crRNA (Figure 6A), confirming that crRNA is important for complex formation. Thus, taken together binding experiments clearly show that the Cas9 protein complex is unable to bind double-stranded DNA in the absence of PAM, even if it contains crRNA complementary to proto-spacer. To put it into other words, double-stranded DNA substrates lacking PAM are not cleaved because PAM is required for Cas9-crRNA binding.

On the other hand, single-stranded oligonucleotides ((+)strand) are bound by Cas9-crRNA with the same affinity independently of the PAM presence (Figure 9D). Again, no binding was observed for single-stranded DNA oligonucleotide without proto-spacer (Figure 9D), or for Cas9 protein lacking crRNA (Figure 6C). Taken together these data indicate that Cas9-crRNA complex discriminates PAM only in the double-stranded but not a single-stranded DNA.

Since some Type III CRISPR systems provide RNA rather than DNA interference, we have studied RNA binding and cleavage by the Cas9-crRNA complex. The Cas9-crRNA did not cleave specifically either single-stranded RNA, or double-stranded RNA bearing a proto-spacer and PAM (Figure 10B).This finding confirms confirms once more that DNA is a primary target for the CRISPR3/Cas system of *S.thermophilus*. Cas9-crRNA complex binds a complementary RNA containing a proto-spacer, but this interaction is probably functionally not important, because single stranded RNA is not cleaved specifically by Cas9 within a proto-spacer.

Mutagenesis of Cas9 protein RuvC and HNH motifs. Plasmid transformation experiments indicate that RuvC and HNH motifs (Figure 11A) are important for Cas9 function (Sapranauskas et al., 2011. Nucleic Acids Res 39:9275–82). To test if these motifs are involved in the target DNA cleavage, we expressed and purified D31A and N891A mutants following procedure described for wt Cas9. Both mutants co-purified with crRNA identical to crRNA in the wt Cas9 complex (Figure 11C). To test whether mutant proteins retained cleavage activity, we monitored pSP1 plasmid cleavage by mutant

Cas9-crRNA complexes. Surprisingly, instead of linear reaction product observed for the wt Cas9 protein, both mutants produced nicked DNA form (Figure 11B) indicating that both active sites mutants cleave only one DNA strand of plasmid substrate within a proto-spacer.

To determine whether mutant proteins exhibit a strand preference, we analysed D31A and N891A mutant cleavage of the SP1 oligoduplex. RuvC active site mutant (D31A) cut (+) strand of oligoduplex at the same position as wt Cas9-crRNA protein, while the (-)strand stayed intact (Figure 11C). And vice versa, HNH active site mutant (N891A) cleaved only (-)strand, but not (+) strand of the SP1 oligoduplex (Figure 11D). Taken together these data indicate that RuvC and HNH active sites act on opposite DNA strands to generate a double strand break. To test, whether the same cleavage pattern is conserved during the plasmid DNA cleavage, we sequenced proto-spacer regions of nicked plasmids. Run-off sequence data confirmed that RuvC active site mutant cut only (+) DNA strand while HNH/McrA mutant - only (-)strand (Figure 12A and B). Furthermore, we found that RuvC mutant cleaved (+) strand of a single-stranded DNA but no such cleavage was detected for the HNH mutant (Figure 12D).

To test whether mutations altered DNA-binding affinity of mutant protein-crRNA complexes, DNA binding was studied using the electrophoretic mobility shift assay. Both mutant protein-crRNA complexes bound oligoduplex SP1 with the same affinity as wild type protein (Figure 12C.). Thus, mutations in the putative active sites of Cas9 have no significant effect on double-stranded DNA-binding properties of the Cas9-crRNA complex. Since 42 nt crRNA was present in the mutant protein complexes (Figure 12C), we conclude that mutant Cas9-crRNA complexes lost ability to cut one of the target DNA strand due to active site mutation. Since Cas9-HisTag protein is a monomer in solution (Figure 13), it is likely that Cas9 protein is functional as a monomer and uses two active sites for the cleavage of opposite DNA strands. Similar strategy is exploited by some restriction endonucleases (Armalyte et al., 2005. J Biol Chem 280: 41584–94).

Discussion

Cas9-crRNA complex of CRISPR3/Cas system of *S. thermophilus* is crRNA-guided endonuclease. This work demonstrates that Cas9-crRNA complex of CRISPR3/Cas system of *S. thermophilus* is crRNA-directed endonuclease which cuts both DNA strands in the presence of Mg2+-ions within a protospacer 4 nt downstream of the PAM sequence to produce blunt end cleavage products. Sequence specificity of the Cas9-crRNA complex is dictated by the 42 nt crRNA which include ~ 20 nt fragment complementary to the proto-spacer sequence in the target DNA. In this respect the mature crRNA in the Cas9 complex of CRISPR3/Cas system of *S. thermophilus* is similar to crRNA of Streptoccocus pyogenes which has a 3'-handle of repeat sequence but lacks part of the spacer sequence and 5'-handle corresponding to the repeat fragment (Deltcheva et al, 2011). Therefore, crRNA present in the Cas9-crRNA complex of CRISPR3/Cas system of *S. thermophilus* is complementary only to the part of the proto-spacer sequence distal to PAM. Not surprisingly, truncation of the 3'-end of the proto-spacer sequence by 10 nucleotides has no effect on Cas9-crRNA cleavage of synthetic oligoduplexes or plasmid DNA (Figure 8).

The cleavage machinery of Cas9-crRNA complex resides in the Cas9 protein which provides two active sites for the phosphodiester bond cleavage. The RuvC- and HNH-like active sites of Cas9 protein are located on different domains and act independently on individual DNA strands. Alanine replacement of the active site residues in the RuvC- and HNH-motifs transforms Cas9-crRNA complex into a strand-specific nicking endonucleases similar to the nicking enzymes (Chan et al., 2011. Nucleic Acids Res 39:1–18). Consistent with *in vivo* studies, a functional activity of the Cas9-crRNA complex *in vitro* is absolutely dependent on the presence of the proto-spacer adjacent motif NGGNG upstream of the proto-spacer sequence. Data presented in the Figure 3 show that PAM is required for Cas9-crRNA binding to the double-stranded DNA. If PAM sequence is missing in double-stranded DNA, the Cas9-crRNA complex does not bind such DNA even if it contains a complementary proto-spacer sequence. On the other hand, Cas9-crRNA does not display DNA binding if PAM (or multiple PAM's) is present but proto-spacer sequence is absent. Thus, in consistence with the *in vivo* data, both PAM and proto-spacer sequences are necessary prerequisite for double-stranded DNA binding and subsequent cleavage. Contrary to the Cas9-crRNA binding to the double-stranded DNA, PAM sequence motif has no effect on the single-stranded DNA binding by: a single-stranded oligodeoxynucleotide containing proto-spacer with or without PAM sequence is bound equally well but with lower affinity than double-stranded DNA. In the presence of Mg2+ ions Cas9 cuts single-stranded DNA bound to the crRNA using its HNH-active site.

Mechanism of DNA interference in the Type II systems. Our results establish a simple model for the mechanism of double-stranded DNA cleavage by Cas9-crRNA complex in the *S. thermophilus* CRISPR3/Cas system (Figure 14). Cas9-crRNA complexes using a mechanism that yet has to be defined locates and binds to a proto-spacer sequence within the double-stranded DNA in a PAM-dependent process. It is possible that PAM in the double-stranded DNA serves as an initiation site (signal) for the strand separation and promotes subsequent pairing of crRNA to the complementary (+)strand of DNA. It remains to be established whether a Cas9 protein module or Cas9-bound crRNA (for example, using nucleotides in the conserved the "3'-handle" of the conserved repeat sequence) recognizes the PAM sequence. Despite of the lack of these mechanistic details, our data clearly demonstrate that PAM is recognized by Cas9-crRNA in the context of double-stranded DNA. The Cas9-crRNA binding to the target sequence in the ds DNA presumably results in the R-loop structure where (-)strand is displaced and the complementary (+) DNA strand is paired to the crRNA. In the presence of Mg2+ ions phosphodiester bond cleavage occurs on both strands 4 nt 5'-upstream of the PAM sequence to generate blunt DNA ends. DNA cleavage analysis by the RuvC- or HNH-motif mutants demonstrate that RuvC- and HNH-like active sites of Cas9 protein act on the (-) and (+)strands, respectively. Therefore, in the catalytically competent the Cas9-crRNA complex, the Nterminal domain containing the catalytic D31A residue of the RuvC motif is positioned at the displaced (–) DNA strand, while the central part of Cas9 containing the HNH motif is located in the vicinity of the scissile phosphodiester bond of (+) DNA strand paired to crRNA. After DNA cleavage Cas9-crRNA remains bound to the reaction products (Figure 15). Taken together data presented here suggest a first molecular mechanism for the DNA interference step by the CRISPR3/Cas system of *S.thermophilus*. Since cas9 is a signature gene (Makarova et al., 2011. Nat Rev Microbiol 9:467–77) for Type IIA and Type IIB systems the cleavage mechanism proposed here is likely to be conserved in other Type IIA and Type IIB systems. Stand-alone versions of Cas9-like proteins which are not a part of the CRISPR system were identified by bioinformatics (Makarova et al., 2011. Biol Direct 6: 38). In the light of the data provided here we suggest that these proteins can provide interference against foreign DNA similarly to Cas9 if loaded with small crRNA molecules which may be generated through the pathway different from CRISPR.

Comparison to other RNA interference complexes. The mechanism proposed here for the double-stranded DNA cleavage by the Cas9-crRNA complex differs significantly from that for the Type I-E (former Ecoli or CASS2) system (Jore et al., 2011. Nat Struct Mol Biol 18:529–36). In the E.coli system crRNA and Cas proteins assemble into a large ribonucleoprotein complex named Cascade that facilitates target recognition by enhancing sequence-specific hybridization between the CRISPR RNA and complementary target sequences (Jore et al., 2011. Nat Struct Mol Biol 18:529–36). Target recognition is dependent on PAM and governed by the “seed” crRNA sequence located at the 5’-end of the spacer region (Semenova et al., 2011. Proc Natl Acad Sci USA 108:10098–103). However, while Cascade-crRNA complex alone is able to bind double-stranded DNA containing PAM and proto-spacer, it requires an accessory Cas3 protein for DNA cleavage. Cas3 is a single-stranded DNA nuclease and helicase which is able to cleave single-stranded DNA producing multiple cuts (Sinkunas et al.,2011. EMBO J 30:1335–42). The mechanistic details of the Cas3 action on a proper biological substrate (e.g., Cascade-crRNA bound to the double-stranded DNA in the R–loop like complex) have yet to be established. However, it has been demonstrated recently that Cas3 of M. jannaschii alone is able to cut both DNA strands in the synthetic substrate mimicking R-loop (Beloglazova et al., 2011. EMBO J 30:616–27). It is proposed that Cas3 may follow similar mechanism for DNA cleavage in the presence of Cascade-crRNA complex. Thus, current data clearly show that mechanistic details of the interference step for the Type I-E system differs from that of CRISPR3 system both by the catalytic machinery and mechanism and complexity.

In the III-B subtype CRISPR systems present in many archea and some bacteria, Cas module RAMP (Cmr) proteins and cRNA assemble into the effector complex that targets invading RNA (Hale et al., 2009. Cell 139:945–56; Hale et al., 2012. Mol Cell 45:292–302). In Pyroccus furiosus RNA silencing complex comprised of six Cmr1-6 proteins and crRNA binds to the target RNA and cuts it at fixed distance in respect to 3'-end the psiRNA. The cleavage activity depends on Mg2+ -ions however individual Cmr protein(-s) responsible for target RNA cleavage has yet to be identified. The effector complex of Sulfolobus solfataricus comprised of seven Cmr1-7 proteins and crRNA cuts invading RNA in an endonucleolytic reaction at UA dinucleotides (Zhang et al., 2012. Mol Cell 45: 303–13). Importantly, both Cmr-crRNA complexes perform RNA cleavage in a PAM independent manner.

The data provided here show that Cas9-crRNA complex of CRISPR3 system is so far the most simple DNA interference system comprised of a single Cas9 protein bound to the crRNA molecule. The simple modular organization of the Cas9-crRNA complex where specificity for DNA target is encoded by the crRNA and cleavage machinery is brought by the Cas protein provides a versatile platform for engineering of universal RNA-guided DNA endonucleases.

Example 2.

In vitro assembly of Cas9-crRNA complex from 4 components

In this example we demonstrate that the catalytically active Cas9-crRNA complex can be assembled *in vitro* by mixing 4 individual components: the C-terminal (His)6-tagged variant of Cas9 protein ("(His)6" disclosed as SEQ ID NO: 23), tracrRNA transcript (SEQ ID NO: 5), CRISPR RNA transcript (SEQ ID NO: 8) and *E. coli* RNAseIII (Abgene). Cas9 protein is first pre-incubated with tracrRNA and CRISPR RNA transcripts, followed by the subsequent incubation with RNAseIII to generate a catalytically competent Cas9-crRNA complex which is used for the site-specific DNA cleavage.

More specifically, RNA fragments required for complex assembly were produced by *in vitro* transcription (TranscriptAid™ T7 High Yield Transcription Kit, Fermentas) of PCR-generated fragment containing a T7 promoter at the proximal end of RNA coding sequence. PCR-generated DNA fragments encoding CRISPR RNA and tracrRNA were produced using pCas9(-)SP1 plasmid as a template with a following primer pair: 5'-taatacgactcactataGggtagaaaagatatcctacgagg-3' (SEQ ID NO: 40)/5'-CAACAACCAAGCTAATACAGCAG-3' (SEQ ID NO: 41) and 5'-aaaaacaccgaatcggtgccac-3' (SEQ ID NO: 42)/ 5'-taatacgactcactataGggTAATAATAATTGTGGTTTGAAACCATTC-3' (SEQ ID NO: 43) (T7 RNA polymerase promoter underlined, transcription start shown in bold). The 150 nt CRISPR RNA transcript is comprised of 102 nt Repeat-Spacer1-Repeat sequences flanked by the 23 nt upstream and 25 nt downstream regions required for primer annealing. The 105 nt transcript of tracrRNA is comprised of a 38 nt stretch partially complimentary to the *S. thermophilus* DCGG7710 CRISPR3 repeat sequence fragment (anti-repeat sequence), flanked by the 16 nt upstream and 51 nt downstream region. RNA fragments produced by *in vitro* transcription were purified using RNeasy MinElute Cleanup Kit (Qiagen).

For *in vitro* assembly of catalytically competent Cas9-crRNA complex, the (His)6-tagged Cas9 protein ("(His)6" disclosed as SEQ ID NO: 23) was mixed with CRISPR RNA and tracrRNA transcripts at 1:0.5:1 molar ratio and pre-incubated in a buffer containing 10 mM Tris-HCl (pH 7.5 at 37°C), 100 mM

NaCl at 37°C for 30 min followed by addition of RNAseIII (Ambion), MgCl2 and DTT and subsequent incubation for additional 30 min. The final concentrations of the components in the assembly mix were the following: 100 nM of (His)6-tagged Cas9 protein ("(His)6" disclosed as SEQ ID NO: 23), 50 nM of CRISPR RNA, 100 nM of tracrRNA, 50 nM RNAseIII, 10 mM MgCl2 and 1 mM DTT.

Below we provide experimental evidences that *in vitro* assembled Cas9-crRNA complex guided by the crRNA sequence cleaves DNA at the specific site to generate blunt ends. In this respect Cas9-crRNA complex can be used an alternative for a restriction endonuclease or meganuclease for the site-specific DNA cleavage *in vitro*. The sequence specificity of the complex is dictated by the crRNA sequence which can be engineered to address a desirable DNA target.

First, the DNA cleavage activity of the *in vitro* assembled Cas9-crRNA complex was assayed on the plasmid substrates pSP1 and pUC18. The pSP1 plasmid contained a proto-spacer1 sequence flanked by the 5'-GGNG-3'PAM sequence. Proto-spacer1 sequence was not present in pUC18. Reactions on pUC18 and pSP1 plasmids (Sapranauskas et al., 2011. Nucleic Acids Res 39:9275–82) were conducted at 37 °C in the 10 mM Tris HCl (pH 7.5 at 37°C), 50 mM NaCl, 0.05 mg/ml BSA, 0.5 mM DTT and 10 mM MgCl2. Reaction mixtures typically contained 3.0 nM of supercoiled plasmid DNA. The reactions were initiated by mixing 50 µl volumes of Cas9-crRNA complex and plasmid DNA (1:1 v/v ratio) in a reaction buffer. Aliquots were removed at timed intervals and quenched with phenol/chloroform. The aqueous phase was mixed with loading dye solution (0.01 % bromphenol blue and 75 mM EDTA in 50 % v/v glycerol) and reaction products analyzed by electrophoresis through agarose (Figure 16). To check whether the pSP1 plasmid pre-cleaved by Cas9-crRNA complex can be re-ligated, we purified linear pSP1 cleavage product from agarose gel using GeneJET gel extraction Kit (Fermentas) and re-ligated using T4 DNA ligase (Fermentas). After transformation of E.coli cells by the ligation mix, five individual clones were selected from resulting transformants, plasmid DNA was purified and subjected to sequencing using the following primers: 5'-ccgcatcaggcgccattcgcc-3' (SEQ ID NO: 29) (sequencing of (+)strand) and 5'-gcgaggaagcggaagagcgccc-3' (SEQ ID NO: 30) (sequencing of (-)strand). Sequence analysis revealed that the DNA sequence of the pSP1 plasmid in the locus that was cleaved by Cas9-crRNA complex and re-ligated was identical to the sequence of the non-treated plasmid. *E. coli* transformation by the ligation mix in the absence of T4 DNA ligase did not produce transformants indicating that no traces of supercoiled plasmid are co-purified with the linear reaction product.

Next, the cleavage activity of the *in vitro* assembled Cas9-crRNA complex was assayed on a synthetic 55 bp oligodeoxynucleotide duplex SP1 containing a proto-spacer sequence matching to the spacer sequence of crRNA (Figure 17). Reactions conditions were identical to those described above for the plasmid DNA cleavage, except that 1 nM of oligoduplex was used. Reaction product analysis revealed that *in vitro* assembled Cas9-crRNA complex cleaved both strands of the oligoduplex at fixed position, inside the proto-spacer, after the 37th nucleotide from the 5'-terminus, 4 nt upstream of the PAM sequence 5'-GGNG-3' leaving blunt ends (Figure 17).

Example 3.

In vitro assembly of Cas9-crRNA complex from 3 components

In this example we demonstrate that active Cas9-crRNA complex can be assembled *in vitro* by mixing 3 individual components: the C-terminal (His)6-tagged variant of Cas9 protein ("(His)6" disclosed as SEQ ID NO: 23), tracrRNA transcript provided in Example 1 (SEQ ID NO: 5 and SEQ ID NO: 6), and CRISPR RNA transcript (SEQ ID NO: 8) provided in Example 1 or synthetic crRNA (SEQ ID NO: 8) which corresponds to the putative crRNA of CRISPR3/Cas system of *S.thermophilus* DGCC7710 strain. Synthetic 42 nt oligoribonucleotide is comprised of 20 nt of identical to the spacer1 of CRISPR3 region at the 5' terminus and 22 nt of repeat sequence at the 3' end. More specifically, tracrRNA and CRISPR RNA transcripts were obtained as described in Example 1. To generate the Cas9-crRNA complex the (His)6-tagged Cas9 protein ("(His)6" disclosed as SEQ ID NO: 23) was mixed with tracrRNA and CRISPR RNA transcript, or 42 nt synthetic crRNA, at 1:0.5:1 molar ratio and incubated in a buffer containing 10 mM Tris-HCl (pH 7.5 at 37°C), 100 mM NaCl at 37 °C for 1 h. The final concentrations of the components in the assembly mix were the following: 100 nM of (His)6-tagged Cas9 protein ("(His)6" disclosed as SEQ ID NO: 23), 50 nM of CRISPR RNA or 42 nt synthetic crRNA, 100 nM of tracrRNA.

Below we provide experimental evidences that *in vitro* assembled Cas9-crRNA complex guided by the crRNA sequence cleaves DNA at the specific site to generate blunt ends. In this respect Cas9-crRNA complex can be used an alternative for a restriction endonuclease or meganuclease for the site-specific DNA cleavage *in vitro*. The sequence specificity of the complex is dictated by the crRNA sequence which can be engineered to address a desirable DNA target.

First, the DNA cleavage activity of the *in vitro* assembled Cas9-crRNA complex was assayed on the plasmid substrates pSP1 and pUC18. The pSP1 plasmid contained a proto-spacer1 sequence flanked by the 5'-GGNG-3'PAM sequence. Proto-spacer1 sequence was not present in pUC18. Reactions on plasmid substrates (Sapranauskas et al., 2011. Nucleic Acids Res 39:9275–82) were conducted at 37 ºC in the 10 mM Tris-HCl (pH 7.5 at 37°C), 50 mM NaCl, 0.05 mg/ml BSA, 0.5 mM of DTT and 10 mM MgCl2. Reaction mixtures typically contained 3.0 nM of supercoiled plasmid DNA. The reactions were initiated by mixing 50 µl volumes of Cas9-crRNA complex and plasmid DNA (1:1 v/v ratio) in a reaction buffer. Aliquots were removed at timed intervals and quenched with phenol/chloroform. The aqueous phase was mixed with loading dye solution (0.01 % bromphenol blue and 75 mM EDTA in 50 % v/v glycerol) and reaction products analyzed by electrophoresis through agarose (Figure 18).

Next, the cleavage activity of the *in vitro* assembled Cas9-crRNA complex was assayed on a synthetic 55 bp oligodeoxynucleotide duplex SP1 containing a a proto-spacer sequence matching to the spacer sequence of crRNA (Figure 19). Reactions conditions were identical to those described above for the plasmid DNA cleavage, except that 1 nM of oligoduplex was used. Reaction product analysis revealed that *in vitro* assembled Cas9-crRNA complex cleaved both strands of the oligoduplex at fixed position, inside the proto-spacer, after the 37th nucleotide form the 5'-end, 4 nt upstream of the PAM sequence 5'-GGNG-3' leaving blunt ends (Figure 19).

Example 4.

Interchangeable spacer cassette for the re-programing of the Cas9-crRNA complex specificity.

In this example we describe an interchangeable spacer cassette which allows to produce crRNA carrying a nucleotide sequence against any desirable DNA target to be used for assembly of the Cas9-crRNA complex described in Examples 1 and 2 (Figure 20B). The cassette caries a single repeat-spacer-repeat unit which allows insertion of the oligoduplex carrying the new spacer sequence required to generate a desired crRNA. To engineer a cassette, first we constructed a cassette containing a leader sequence, a repeat sequence and a unique SapI recognition site in the vicinity of the repeat sequence followed by BamHI site (Figure 20C). To generate CRISPR region containing the unique desired spacer, we inserted a synthetic oligoduplex containing a unique spacer sequence and a repeat unit into the plasmid precleaved with SapI and BamHI restriction enzymes. Using this cassette we produced crRNA transcripts which contained nucleotide sequences complementary to the proto-spacers N1 and N2 present in pUC18 plasmid (see below).

As proof of the principle demonstration, we used an interchangeable spacer cassette to generate crRNA1 and crRNA2 which were engineered to target pUC18 plasmid at proto-spacer1 and proto-spacer2, respectively, incorporated crRNA1 and crRNA2 into Cas9 complex as described in the Example 1 and used these complexes for the cleavage of pUC18 plasmid. The proto-spacer N1 is located near the SapI restriction endonuclease site, while the proto-spacer N2 is in the vicinity of AatII site. The distance between SapI and AatII restriction sites is 775 bp, while the distance between the putative Cas9-crRNA complex cleavage sites located in the spacers N1 and N2 is 612 bp (Figure 21A). The crRNA1 and crRNA2 PCR fragments containing T7 promoter at the proximal end were obtained from the corresponding interchangeable spacer cassette plasmids and used to produce by *in vitro* transcription CRISPR RNA transcripts carrying sequences matching spacer N1 or spacer N2 sequences. The catalytically active complexes of Cas9 with crRNA1 and crRNA2 were assembled for DNA cleavage as described in Example 1. *In vitro* assembled complexes containing either crRNA1 or crRNA2 linearized pUC18 plasmid (Figure 21B). When both complexes were incubated with the pUC18plasmid, two DNA fragments (2074 and 612 bp) were obtained (Figure 21B), indicating that plasmid cleavage occurred at sites targeted by the crRNA molecules present in the complexes.

Example 5.

Cloning procedure using Cas9-crRNA complex.

In this example we demonstrate that Cas9-crRNA complex may be used to prepare a vector for cloning procedure. First we demonstrated that cleavage products obtained by the Cas9-crRNA complex can be re-ligated by DNA ligase. We purified linear pSP1 cleavage product from agarose gel and re-ligated it using DNA ligase. After transformation of *E.coli* cells by the ligation mix, five individual clones were selected from resulting transformants, plasmid DNA was purified and subjected to sequencing. Sequence analysis revealed that the DNA sequence of the pSP1 plasmid in the locus that was cleaved by Cas9-RNA complex and re-ligated was identical to the sequence of the non-treated plasmid. *E. coli* transformation by the ligation mix in the absence of T4 DNA ligase did not produce transformants indicating that no traces of supercoiled plasmid are co-purified with the linear reaction product. This result illustrates, that the DNA ends generated by the Cas9 cleavage are substrates for T4 DNA ligase, and therefore must contain a phosphate at the 5' terminus and a free OH group at the 3' terminus (Lehman, 1974).

Next we analyzed cleavage of pUC18 plasmid with Cas9 complex loaded with crRNA1 and crRNA2 described in Example 5 (Figure 21A). First, pUC18 was cleaved with one complex, purified and re-ligated. Sequencing of 10 clones in each case confirmed, that sequence of cleaved and re-ligated plasmid was identical to the sequence of the non-treated plasmid (Figure 21C). This experiment suggests that additional mutations are not introduced after cleavage by Cas9-crRNA complex and ligation, and the Cas9-crRNA complex can be used for cloning experiments. When both complexes were incubated with the pUC18 plasmid, two DNA fragments (2074 and 612 bp) were obtained (Figure 21B), indicating that plasmid cleavage occurred at sites targeted by the crRNA molecules present in the complexes. To demonstrate that the pUC18 plasmid cleaved with Cas9-RNA complexes is suitable for a genetic engineering we cloned PCR fragment containing a promoter and a tetracycline resistance gene from the pACYC184 plasmid to the pUC18 vector pre-cleaved with the Cas9 complex mix containing both crRNA1 or crRNA2. The clones were selected on the media enriched by tetracycline and ampicillin. Sequencing of 4 selected clones confirmed that the intact PCR fragment was inserted into a desired position ((Figure 21C).

More specifically, the 2 µg pUC18 was incubated with the mix of separately assembled Cas9-RNA complexes (250 nM each) containing different crRNAs for 1 hour at 37°C in 100 µl reaction volume (10 mM Tris-HCl (pH 7.5 at 37°C), 100 mM NaCl, 1 mM DTT and 10 mM $MgCl_2$). Obtained vector fragment was purified from agarose gel using GeneJET gel extraction Kit (Thermo Fisher scientific) and divided in to two equal parts. One part of pre-cleaved vector was dephosphorylated with the FastAP alkaline phosphatase while another part was untreated. 1282 bp insert containing a promoter and a tetracycline resistance gene was obtained from the pACYC184 plasmid by PCR. After purification using the GeneJET PCR Purification Kit (Thermo Fisher scientific), a solution containing the PCR fragment was divided in to two parts. One part was phosphorylated with T4 polynucleotide kinase (Thermo Fisher scientific) while another part remained untreated. Untreated vector was ligated with the untreated PCR fragment, while a dephosphorylated vector was ligated with a phosphorylated fragment using the T4 DNA ligase (Thermo Fisher scientific). Clones were selected on a media supplemented with 100 μg/ml of Ap and 25 μg/ml Tc.

Example 6.

Cleavage of long DNA substrates by Cas9 crRNA complex.

In this example we demonstrate that Cas9-crRNA may be addressed to cleave targets in long DNA molecules, including phage λ, *E. coli* and human genomic DNAs.

More specifically, we addressed Cas9-RNA complex to cleave specific sites in λ bacteriophage (48 kb), *E. coli* BL-21 strain (4.6 Mb) and human (3.2 Gb) genomic DNAs. Cas9-crRNA complex was assembled as described in Examples 2 and 3. We used 42 nt long synthetic crRNAs, 150 nt pre-crRNAs and tracrRNAs synthesized using *in vitro* transcription from templates generated as described in Example 4.

λ DNA cleavage reactions were initiated by mixing λ DNA (Thermo Fisher Scientific) with assembled Cas9-RNA complex (1:1 v/v ratio) and incubating at 37°C. Final reaction mixture contained 2 μg λ DNA, 50 nM Cas9-RNA complex, 10 mM Tris-HCl (pH 7.5 at 37°C), 100 mM NaCl, 1 mM DTT and 10 mM $MgCl_2$ in 100 μl reaction volume. Aliquots were removed at timed intervals and quenched with phenol/chloroform. The aqueous phase was mixed with 3X loading dye solution (0.01 % bromphenol blue and 75 mM EDTA in 50 % v/v glycerol) and reaction products analyzed by electrophoresis through agarose gels and ethidium bromide staining. The analysis of linear λ phage genomic DNA cleavage products in agarose gel confirmed that ~40 bp length DNA is efficiently cleaved at a single site (Figure 22A).

DNA from *E. coli* BL21 (DE3) strain was isolated using the Genomic DNA purification kit (Thermo Fisher Scientific). For cleavage assay, *E. coli* genomic DNA was combined with assembled Cas9-RNA complex (1:1 v/v ratio) and incubated for 3 hours at 37°C. Final reaction mixture contained 30 μg genomic DNA, 1 μM Cas9-RNA complex, 10 mM Tris-HCl (pH 7.5 at 37°C), 100 mM NaCl, 1 mM DTT and 10 mM $MgCl_2$ in 300 μl reaction volume. Following incubation, 30 μl of FastDigest *Pst*I (Thermo Fisher Scientific) was added and the reaction mix was incubated for additional 16 hours at 37°C. The reaction was terminated by heating the reaction mixture for 30 min at 55°C with Proteinase K (0.5 mg/ml; Thermo Fisher Scientific) and SDS (0.5%, w/v) followed by 30 min incubation at room temperature with RNase A (0.25 mg/ml; Thermo Fisher Scientific). After phenol/chloroform extraction, DNA was precipitated by isopropanol and dissolved in TE buffer (10 mM Tris–HCl, pH 8.0 and 1 mM EDTA). 10 µg of DNA was mixed with 3X loading dye solution (0.01 % bromphenol blue and 75 mM EDTA in 50 % v/v glycerol) and electrophoresed on 1% agarose gel.

To analyse Cas9-crRNA cleavage products of *E.coli* genomic DNA, we designed a probe against DNA fragment containing a Cas9-RNA complex target (a proto-spacer) (Figure 22B) and performed Southern blot analysis. Southern blot analysis was performed as described in (Sambrook et al, 1989. Molecular Cloning: A Laboratory Manual) with the following modifications. Fractionated DNA was transferred from agarose gel onto SensiBlot Plus Nylon membrane (Thermo Fisher Scientific) via semi-dry transfer. DNA was denatured and fixed on the membrane by placing it on paper towel saturated with 0.4 M NaOH for 10 min, rinsed with 2X SSC and air dried. The membrane was prehybridized with 6X SSC buffer containing 0.5% SDS and 100 µg/ml denatured salmon sperm DNA (Amresco) for 1 h at 65°C. The hybridization probe was generated by PCR using the genomic *E. coli* BL21(DE3) DNA as a template yielding 397 bp product. 5'-ends were dephosphorylated with FastAP phosphatase (Thermo Fisher Scientific) and radiolabelled by incubating with [γ-^{32}P]ATP (Hartmann Analytic) and T4 PNK (Thermo Fisher Scientific). The labeled probe was purified using GeneJET PCR Purification Kit (Thermo Fisher Scientific), denatured by heating to 95°C for 5 min, rapidly cooled on ice and added directly to the prehybridization solution. The membrane was probed for 16 hours at 65°C and washed twice with 2X SSC, 0.5% SDS and twice with 2X SSC, 0.1% SDS at room temperature, air dried and visualized by phosphorimaging (FLA-5100; Fujifilm).

The probe was designed to target DNA fragment containing a target (a proto-spacer) for the Cas9-RNA complex (Figure 22B). The distance between two PstI targets is ~ 1500 bp, while the distance between proto-spacer and left PstI target is 466 bp. After cleavage with Cas9 complex we detected only 466 bp DNA fragment (Figure 22C), which means that all DNA targets were cleaved by Cas9 protein in the desired position. These data clearly demonstrates that Cas9 protein effectively finds targets in very long and complex molecules such as viral and bacterial DNA.

To analyze Cas9-crRNA cleavage products of human genomic DNA we used DNA extracted from human brain. Human genomic DNA was combined with assembled Cas9-crRNA complex (1:1 v/v ratio) and incubated for 30 min at 37°C. Final reaction mixture contained 1 µg genomic DNA, 100 nM Cas9, 10 mM Tris-HCl (pH 7.5 at 37°C), 100 mM NaCl, 1 mM DTT and 10 mM $MgCl_2$ in 100 µl reaction volume. Cas9-crRNA-HS1 (SeqID#13) and Cas9-crRNA-HS2 (SeqID#14) complexes were assembled to target RASGEF1C or ARL15 loci, respectively. Cleavage products were analyzed using qPCR (Figure 22D). After treatment with Cas9-crRNA complex, the amount of intact DNA targets decreased more than 25 times. The analysis of the results obtained from qPCR data revealed that

Cas9-RNA complexes cleave human genomic DNA efficiently in the desired loci. These data clearly demonstrates that Cas9 protein effectively finds targets in very long and complex molecules such as viral, bacterial and mammal DNA.

Example 7.

Evidence for gene editing of a reporter plasmid in mammalian cells after transfection of Cas9/RNA complexes.

A reporter plasmid was constructed to monitor double-strand break repair either through non-homologous end-joining (NHEJ) or homologous recombination (HR). The plasmid contained GFP with an intron and flanking the eGFP sequences are 5' and 3' sequences of RFP as well as sites of homology (Figure 23). The reduction of eGFP fluorescence using this reporter plasmid was an indication of NHEJ in which a Cas9/RNA-mediated double-strand break at targets C or D was repaired imperfectly by NHEJ, thereby disrupting the eGFP coding sequence. Targeting of intronic targets A and B and repair by NHEJ would likely not result in a reduction in eGFP fluorescence because the mutations induced by NHEJ usually delete or insert <20 bps and would therefore not affect the eGFP coding regions or splice site junctions. The appearance of RFP fluorescence, on the other hand, was an indication of HR where the Cas9/RNA-mediated double strand break is repaired by HR using the homologous sequences of RFP indicated.

The crRNA targeting used 42 nucleotide RNA molecules, as described above, having 22 nucleotides that are the repeat sequence, and 20 nucleotides (spacer sequence) are for the specific target. As described above, the target DNA needs the *S. thermophilus* motif or PAM which is “NGGNG” downstream of the protospacer in the target. GFP was not “engineered” to contain this PAM motif; several target sequences within eGFP naturally occur with the PAM sequence and crRNAs were designed to target the adjacent spacer sequences. RFP was a marker for homologous recombination after a double strand break in eGFP was created by Cas9/RNA.

Figure 28A shows reporter gene construct for Cas9 protein activity analysis in eukaryotic cells *in vivo.* Intron sequence contains three cas9 target sites (A, E, B); GFP gene contains two (C, D) cas9 target sites. The RFP gene is split at Y196 position, where RFP fluorescence is abolished. Figure 28B shows that GFP fluorescence is observed following intron processing *in vivo*. Figure 28C shows that the Cas9/crRNA complex facilitated dsDNA breaks in any of aforementioned nuclease target sites may induce HR, result in reassembly of RFP gene and appearance of RFP fluorescence. Figures 28D and E show that the Cas9/crRNA complex facilitated dsDNA breaks in any of aforementioned nuclease target sites may induce NHEJ. Mutations in GFP gene sequence would result in lost or diminished GFP fluorescence; mutations in intron may have no affect on GFP fluorescence, however, in distinct cases may yield mature messenger RNA with improperly spliced intron sequences and result in lost or diminished GFP fluorescence.

S. thermophilus Cas9 protein, purified from *E. coli*, was complexed with *in vitro*-transcribed tracrRNA and synthetic unmodified crRNA targeting either sequence A (intronic) or sequence C (coding) of eGFP. For transfection, the Cas9/RNA complexes (either targeting A or C) were incubated with the transfection reagent TurboFECT and the reporter plasmid DNA was also incubated with TurboFECT in separate tubes and they were both added to CHO-K1 cells. The percentage of eGFP-positive cells was determined by flow cytometry. As shown in Figures 24 and 29, when cells were transfected with the reporter plasmid alone or with the reporter plasmid with Cas9 protein alone, the percentage of GFP-positive cells was about 40-50%, indicative of the overall transfection efficiency. However, when Cas9/RNA complexes targeting sequence C of eGFP were added to cells along with the reporter plasmid, the percentage of eGFP-positive cells was reduced to about 15%. This decrease in eGFP-positive cells was seen only with Cas9/RNA complexes targeting sequence C and there was no significant decrease in eGFP-positive cells seen with the Cas9/RNA complexes targeting sequence A or with a non-specific RNA. This result indicated that the Cas9/RNA targeting sequence C of eGFP resulted in gene editing of eGFP by introduction of a double-strand break and imperfect correction by NHEJ, creating a deletion in the coding sequence of eGFP.

In addition to analyzing the percentage of eGFP-positive cells, transfected cells were also visualized by fluorescent microscopy to monitor the appearance of RFP-positive cells, an indication of repair of Cas9-mediated double strand break by HR rather than NHEJ. As seen in Figure 25, RFP is seen in some cells after transfection with the reporter plasmid and Cas9/RNA complexes targeting eGFP sequence C, suggesting double-strand break repair by HR.

Example 8.

Cas9/RNA complexes made using synthetic unmodified tracrRNAs and crRNAs are functional *in vitro*.

The experiments described in Example 7 above used Cas9/RNA complexes comprised of purified Cas9, synthetic crRNAs, and *in vitro*-transcribed tracrRNA. To determine whether Cas9/RNA complexes were functional when made using fully synthetic RNA components (crRNA and tracrRNA), unmodified *S. thermophilus* tracrRNAs (both endogenous 89-mer and a shorter 74-mer version that is expected to maintain functionality) were syntheized. The unmodified synthetic crRNAs were generated against target E (see Figures 26 and 30) located within the intron of eGFP in the reporter plasmid described above and Cas9/RNA (crRNA and tracrRNA) complexes were generated. To test these complexes, the reporter plasmid used above was incubated with the complexes *in vitro* and monitored for restriction by gel electrophoresis.

As seen in Figure 27, Cas9/RNA complexes comprised of fully synthetic RNAs were equally functional in the *in vitro* assay as Cas9/RNA complexes comprised of synthetic crRNA and *in vitro*-transcribed tracrRNA.

Sequences

SEQ ID NO: 1

WT_Cas9_*S. thermophilus* DGCC7710 CRISPR3-Cas strain

One letter:

mlfnkciiisinldfsnkekcmtkpysigldigtnsvgwavitdnykvpskkmkvlgntskkyikknllgvllfdsgitaegrrlkrtarrrytrrr
nrilylqeifstematlddaffqrlddsflvpddkrdskypifgnlveekvyhdefptiyhlrkyladstkkadlrlvylalahmikyrghfliegef
nsknndiqknfqdfldtynaifesdlslenskqleeivkdkisklekkdrilklfpgeknsgifseflklivgnqadfrkcfnldekaslhfskesy
dedletllgyigddysdvflkakklydaillsgfltvtdneteaplssamikrynehkedlallkeyirnislktynevfkddtkngyagyidgktn
qedfyvylknllaefegadyflekidredflrkqrtfdngsipyqihlqemraildkqakfypflaknkeriekiltfripyyvgplargnsdfaws
irkrnekitpwnfedvidkessaeafinrmtsfdlylpeekvlpkhsllyetfnvyneltkvrfiaesmrdyqfldskqkkdivrlyfkdkrkvtd
kdiieylhaiygydgielkgiekqfnsslstyhdllniindkeflddssneaiieeiihtltifedremikqrlskfenifdksvlkklsrrhytgwgkl
saklingirdeksgntildyliddgisnrnfmqlihddalsfkkkiqkaqiigdedkgnikevvkslpgspaikkgilqsikivdelvkvmggrk
pesivvemarenqytnqgksnsqqrlkrlekslkelgskilkenipaklskidnnalqndrlylyylqngkdmytgddldidrlsnydidhiip
qaflkdnsidnkvlvssasnrgksddfpslevvkkrktfwyqllksklisqrkfdnltkaergglllpedkagfiqrqlvetrqitkhvarlldekfn
nkkdennravrtvkiitlkstlvsqfrkdfelykvreindfhhahdaylnaviasallkkypklepefvygdypkynsfrerksatekvyfysni
mnifkksisladgrvierplievneetgesvwnkesdlatvrrvlsypqvnvvkkveeqnhgldrgkpkglfnanlsskpkpnsnenlvgak
eyldpkkyggyagisnsfavlvkgtiekgakkkitnvlefqgisildrinyrkdklnfllekgykdieliielpkyslfelsdgsrrmlasilstnnkr
geihkgnqiflsqkfvkllyhakrisntinenhrkyvenhkkefeelfyyilefnenyvgakkngkllnsafqswqnhsidelcssfigptgser
kglfeltsrgsaadfeflgvkipryrdytpssllkdatlihqsvtglyetridlaklgeg
Three letters:

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
IleGluLysGlyAlaLysLysLysIleThrAsnValLeuGluPheGlnGlyIleSerIle
LeuAspArgIleAsnTyrArgLysAspLysLeuAsnPheLeuLeuGluLysGlyTyrLys
AspIleGluLeuIleIleGluLeuProLysTyrSerLeuPheGluLeuSerAspGlySer
ArgArgMetLeuAlaSerIleLeuSerThrAsnAsnLysArgGlyGluIleHisLysGly
AsnGlnIlePheLeuSerGlnLysPheValLysLeuLeuTyrHisAlaLysArgIleSer
AsnThrIleAsnGluAsnHisArgLysTyrValGluAsnHisLysLysGluPheGluGlu
LeuPheTyrTyrIleLeuGluPheAsnGluAsnTyrValGlyAlaLysLysAsnGlyLys
LeuLeuAsnSerAlaPheGlnSerTrpGlnAsnHisSerIleAspGluLeuCysSerSer
PheIleGlyProThrGlySerGluArgLysGlyLeuPheGluLeuThrSerArgGlySer
AlaAlaAspPheGluPheLeuGlyValLysIleProArgTyrArgAspTyrThrProSer
SerLeuLeuLysAspAlaThrLeuIleHisGlnSerValThrGlyLeuTyrGluThrArg
IleAspLeuAlaLysLeuGlyGluGly

SEQ ID NO: 2

D31A mutant

One letter:

mlfnkciiisinldfsnkekcmtkpysiglaigtnsvgwavitdnykvpskkmkvlgntskkyikknllgvllfdsgitaegrrlkrtarrrytrrr
nrilylqeifstematlddaffqrlddsflvpddkrdskypifgnlveekvyhdefptiyhlrkyladstkkadlrlvylalahmikyrghfliegef
nsknndiqknfqdfldtynaifesdlslenskqleeivkdkisklekkdrilklfpgeknsgifseflklivgnqadfrkcfnldekaslhfskesy
dedletllgyigddysdvflkakklydaillsgfltvtdneteaplssamikrynehkedlallkeyirnislktynevfkddtkngyagyidgktn
qedfyvylknllaefegadyflekidredflrkqrtfdngsipyqihlqemraildkqakfypflaknkeriekiltfripyyvgplargnsdfaws
irkrnekitpwnfedvidkessaeafinrmtsfdlylpeekvlpkhsllyetfnvyneltkvrfiaesmrdyqfldskqkkdivrlyfkdkrkvtd

kdiieylhaiygydgielkgiekqfnsslstyhdllniindkeflddssneaiieeiihtltifedremikqrlskfenifdksvlkklsrrhytgwgkl
saklingirdeksgntildyliddgisnrnfmqlihddalsfkkkiqkaqiigdedkgnikevvkslpgspaikkgilqsikivdelvkvmggrk
pesivvemarenqytnqgksnsqqrlkrlekslkelgskilkenipaklskidnnalqndrlylyylqngkdmytgddldidrlsnydidhiip
qaflkdnsidnkvlvssasnrgksddfpslevvkkrktfwyqllksklisqrkfdnltkaergpllpedkagfiqrqlvetrqitkhvarlldekfn
nkkdennravrtvkiitlkstlvsqfrkdfelykvreindfhhahdaylnaviasallkkypklepefvygdypkynsfrerksatekvyfysni
mnifkksisladgrvierplievneetgesvwnkesdlatvrrvlsypqvnvvkkveeqnhgldrgkpkglfnanlsskpkpnsnenlvgak
eyldpkkyggyagisnsfavlvkgtiekgakkkitnvlefqgisildrinyrkdklnfllekgykdieliielpkyslfelsdgsrrmlasilstnnkr
geihkgnqiflsqkfvkllyhakrisntinenhrkyvenhkkefeelfyyilefnenyvgakkngkllnsafqswqnhsidelcssfigptgser
kglfeltsrgsaadfeflgvkipryrdytpssllkdatlihqsvtglyetridlaklgeg
Three letters:

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

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

IleGluLysGlyAlaLysLysLysIleThrAsnValLeuGluPheGlnGlyIleSerIle

LeuAspArgIleAsnTyrArgLysAspLysLeuAsnPheLeuLeuGluLysGlyTyrLys

AspIleGluLeuIleIleGluLeuProLysTyrSerLeuPheGluLeuSerAspGlySer
ArgArgMetLeuAlaSerIleLeuSerThrAsnAsnLysArgGlyGluIleHisLysGly
AsnGlnIlePheLeuSerGlnLysPheValLysLeuLeuTyrHisAlaLysArgIleSer
AsnThrIleAsnGluAsnHisArgLysTyrValGluAsnHisLysLysGluPheGluGlu
LeuPheTyrTyrIleLeuGluPheAsnGluAsnTyrValGlyAlaLysLysAsnGlyLys
LeuLeuAsnSerAlaPheGlnSerTrpGlnAsnHisSerIleAspGluLeuCysSerSer
PheIleGlyProThrGlySerGluArgLysGlyLeuPheGluLeuThrSerArgGlySer
AlaAlaAspPheGluPheLeuGlyValLysIleProArgTyrArgAspTyrThrProSer
SerLeuLeuLysAspAlaThrLeuIleHisGlnSerValThrGlyLeuTyrGluThrArg
IleAspLeuAlaLysLeuGlyGluGly

SEQ ID NO: 3

N891A mutant

One letter:

mlfnkciiisinldfsnkekcmtkpysigldigtnsvgwavitdnykvpskkmkvlgntskkyikknllgvllfdsgitaegrrlkrtarrytrrr
nrilylqeifstematlddaffqrlddsflvpddkrdskypifgnlveekvyhdefptiyhlrkyladstkkadlrlvylalahmikyrghfliegef
nsknndiqknfqdfldtynaifesdlslenskqleeivkdkisklekkdrilklfpgeknsgifseflklivgnqadfrkcfnldekaslhfskesy
dedletllgyigddysdvflkakklydaillsgfltvtdneteaplssamikrynehkedlallkeyirnislktynevfkddtkngyagyidgktn
qedfyvvylknllaefegadyflekidredflrkqrtfdngsipyqihlqemraildkqakfypflaknkeriekiltfripyyvgplargnsdfaws
irkrnekitpwnfedvidkessaeafinrmtsfdlylpeekvlpkhsllyetfnvyneltkvrfiaesmrdyqfldskqkkdivrlyfkdkrkvtd
kdiieylhaiygydgielkgiekqfnsslstyhdllniindkeflddssneaiieeiihtltifedremikqrlskfenifdksvlkklsrrhytgwgkl
saklingirdeksgntildyliddgisnrnfmqlihddalsfkkkiqkaqiigdedkgnikevvkslpgspaikkgilqsikivdelvkvmggrk
pesivvemarenqytnqgksnsqqrlkrlekslkelgskilkenipaklskidnnalqndrlylyylqngkdmytgddldidrlsnydidhiip
qaflkdnsidnkvlvssasargksddfpslevvkkrktfwyqllksklisqrkfdnltkaergglleedkagfiqrqlvetrqitkhvarlldekfn
nkkdennravrtvkiitlkstlvsqfrkdfelykvreindfhhahdaylnaviasallkkypklepefvygdypkynsfrerksatekvyfysni
mnifkksisladgrvierplievneetgesvwnkesdlatvrrvlsypqvnvvkkveeqnhgldrgkpkglfnanlsskpkpnsnenlvgak
eyldpkkyggyagisnsfavlvkgtiekgakkkitnvlefqgisildrinyrkdklnfllekgykdieliielpkyslfelsdgsrrmlasilstnnkr
geihkgnqiflsqkfvkllyhakrisntinenhrkyvenhkkefeelfyyilefnenyvgakkngkllnsafqswqnhsidelcssfigptgser
kglfeltsrgsaadfeflgvkipryrdytpssllkdatlihqsvtglyetridlaklgeg

Three letters:

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
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

SEQ ID NO: 4

H868A mutant

One letter

mlfnkciiisinldfsnkekcmtkpysigldigtnsvgwavitdnykvpskkmkvlgntskkyikknllgvllfdsgitaegrrlkrtarrrytrrr
nrilylqeifstematlddaffqrlddsflvpddkrdskypifgnlveekvyhdefptiyhlrkyladstkkadlrlvylalahmikyrghfliegef
nsknndiqknfqdfldtynaifesdlslenskqleeivkdkisklekkdrilklfpgeknsgifseflklivgnqadfrkcfnldekaslhfskesy
dedletllgyigddysdvflkakklydaillsgfltvtdneteaplssamikrynehkedlallkeyirnislktynevfkddtkngyagyidgktn
qedfyvylknllaefegadyflekidredflrkqrtfdngsipyqihlqemraildkqakfypflaknkeriekiltfripyyvgplargnsdfaws
irkrnekitpwnfedvidkessaeafinrmtsfdlylpeekvlpkhsllyetfnvyneltkvrfiaesmrdyqfldskqkkdivrlyfkdkrkvtd
kdiieylhaiygydgielkgiekqfnsslstyhdllniindkeflddssneaiieeiihtltifedremikqrlskfenifdksvlkklsrrhytgwgkl
saklingirdeksgntildyliddgisnrnfmqlihddalsfkkkiqkaqiigdedkgnikevvkslpgspaikkgilqsikivdelvkvmggrk
pesivvemarenqytnqgksnsqqrlkrlekslkelgskilkenipaklskidnnalqndrlylyylqngkdmytgddldidrlsnydidaiipq
aflkdnsidnkvlvssasnrgksddfpslevvkkrktfwyqllksklisqrkfdnltkaergglledkagfiqrqlvetrqitkhvarlldekfnn
kkdennravrtvkiitlkstlvsqfrkdfelykvreindfhhahdaylnaviasallkkypklepefvygdypkynsfrerksatekvyfysnim
nifkksisladgrvierplievneetgesvwnkesdlatvrrvlsypqvnvvkkveeqnhgldrgkpkglfnanlsskpkpnsnenlvgakey
ldpkkyggyagisnsfavlvkgtiekgakkkitnvlefqgisildrinyrkdklnfllekgykdieliielpkyslfelsdgsrrmlasilstnnkrge
ihkgnqiflsqkfvkllyhakrisntinenhrkyvenhkkefeelfyyilefnenyvgakkngkllnsafqswqnhsidelcssfigptgserkgl
feltsrgsaadfeflgvkipryrdytpssllkdatlihqsvtglyetridlaklgeg

Three letters:

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
LeuAspArgIleAsnTyrArgLysAspLysLeuAsnPheLeuLeuGluLysGlyTyrLys
AspIleGluLeuIleIleGluLeuProLysTyrSerLeuPheGluLeuSerAspGlySer
ArgArgMetLeuAlaSerIleLeuSerThrAsnAsnLysArgGlyGluIleHisLysGly
AsnGlnIlePheLeuSerGlnLysPheValLysLeuLeuTyrHisAlaLysArgIleSer
AsnThrIleAsnGluAsnHisArgLysTyrValGluAsnHisLysLysGluPheGluGlu
LeuPheTyrTyrIleLeuGluPheAsnGluAsnTyrValGlyAlaLysLysAsnGlyLys
LeuLeuAsnSerAlaPheGlnSerTrpGlnAsnHisSerIleAspGluLeuCysSerSer
PheIleGlyProThrGlySerGluArgLysGlyLeuPheGluLeuThrSerArgGlySer
AlaAlaAspPheGluPheLeuGlyValLysIleProArgTyrArgAspTyrThrProSer
SerLeuLeuLysAspAlaThrLeuIleHisGlnSerValThrGlyLeuTyrGluThrArg
IleAspLeuAlaLysLeuGlyGluGly

SEQ ID NO: 5

Tra-crRNA, Unmature (102 nt):

uaauaauaauuguggguuuugaaaccauucgaaacaacacagcgaguuaaaauaaggcuuaguccguacucaacuugaaaagguggcac
cgauucgguguuuuu

SEQ ID NO: 6

Mature 78 nt tracrRNA:

gggcgaaacaacacagcgaguuaaaauaaggcuuaguccguacucaacuugaaaagguggcaccgauucgguguuuuu

Shorter variants:

gggcgaaacaacacagcgaguuaaaauaaggcuuaguccguacucaacuugaaaagguggcaccgauucggug (SEQ ID NO: 44)
gggcgaaacaacacagcgaguuaaaauaaggcuuaguccguacucaacuugaaaagguggcaccgauu (SEQ ID NO: 45)
gggcgaaacaacacagcgaguuaaaauaaggcuuaguccguacucaacuugaaaagguggcac (SEQ ID NO: 46)
gggcgaaacaacacagcgaguuaaaauaaggcuuaguccguacucaacuugaaaaggu (SEQ ID NO: 47)
gggcgaaacaacacagcgaguuaaaauaaggcuuaguccguacucaacuugaa (SEQ ID NO: 48)
gggcgaaacaacacagcgaguuaaaauaaggcuuaguccguacucaac (SEQ ID NO: 49)

SEQ ID NO: 7

42 nt crRNA from spacer 1:

5'-CGCUAAAGAGGAAGAGGACAGUUUUAGAGCUGUGUUGUUUCG-3'

SEQ ID NO: 8

150 nt pre-crRNA

5′-

ggguagaaaagauauccuacgag*guuuuagagcuguguuguuucgaaugguuccaaaac***aaauucuaaacgcuaaagaggaagaggaca***guuuuagagcuguguuguuucgaaugguuccaaaac*uacugcuguauuagcuugguuguug-3′

SEQ ID NO: 9

crRNA1

5′-

ggguagaaaagauauccuacgag*guuuuagagcuguguuguuucgaaugguuccaaaac*TGTCATGATAATAATGGTTTCTTAGACGTC*guuuuagagcuguguuguuucgaaugguuccaaaac*uacugcuguauuagcuugguuguug-3′

SEQ ID NO: 10

crRNA2

5′-ggguagaaaagauauccuacgag*guuuuagagcuguguuguuucgaaugguuccaaaac*acgagccggaagcataaagtgtaaagcctg*guuuuagagcuguguuguuucgaaugguuccaaaac*uacugcuguauuagcuugguuguug-3′

SEQ ID NO: 11

Anti-λ phage CRISPR RNA

5′-ggguagaaaagauauccuacgag*guuuuagagcuguguuguuucgaaugguuccaaaac*tcaagggagaatagaggctctcgttgcatt*guuuuagagcuguguuguuucgaaugguuccaaaac*uacugcuguauuagcuugguuguug-3′

SEQ ID NO: 12

Anti *E. coli* CRISPR RNA

5′-ggguagaaaagauauccuacgag*guuuuagagcuguguuguuucgaaugguuccaaaac*cgggagggaagctgcatgatgcgatgttat*guuuuagagcuguguuguuucgaaugguuccaaaac*uacugcuguauuagcuugguuguug-3′

SEQ ID NO: 13

crRNA-HS1

5’-GCUCCCGGGGCUCGAUGAAGGUUUUAGAGCUGUGUUGUUUCG-3’

SEQ ID NO: 14

crRNA-HS2

UGAAUCGUGAAAUCUGCUCAGUUUUAGAGCUGUGUUGUUUCG

The application contains a Sequence Listing which has been submitted in ASCII format. The ASCII copy, created on March 20, 2013, is named 078981_6_SL.txt and is 64.4 kilobytes in size.

The embodiments shown and described in the specification are only specific embodiments of inventors who are skilled in the art and are not limiting in any way. Therefore, various changes, modifications, or alterations to those embodiments may be made without departing from the spirit of the invention in the scope of the following claims.

What is claimed is:

CLAIMS:

1. A method for site-specific modification of a target DNA molecule, the method comprising:

combining a Cas9 protein, a crRNA, and a tracrRNA *in vitro* to form a Cas9-crRNA complex, a sequence of the crRNA being complementary to a region of the target DNA molecule comprising the site, wherein the sequence of the crRNA comprises at least 20 nucleotides; and

contacting the target DNA molecule with the Cas9-crRNA complex,

wherein the crRNA is engineered to guide the Cas9-crRNA complex to the site in the target DNA molecule, and

wherein the modification occurs *in vitro* or in an isolated cell.

2. The method of claim 1, wherein the cell is an isolated mammalian cell.

3. The method of claim 1, wherein the Cas9 protein comprises at least one of an RuvC active site motif or an HNH active site motif.

4. The method of claim 1, wherein the Cas9 protein has at least 80% identity with SEQ ID NO: 1.

5. The method of claim 1, wherein the crRNA comprises a 3' region and a 5' region, wherein the 3' region comprises at least 22 nucleotides of a CRISPR repeat and the 5' region comprises the sequence that is complementary to the region of the target DNA molecule.

6. The method of claim 1, wherein the tracrRNA comprises a 5' region and a 3' region, wherein a portion of the 5' region comprising at least 22 nucleotides is complementary to a 3' region of the crRNA.

7. The method of claim 1, wherein the target DNA molecule comprises a proto-spacer adjacent motif (PAM) sequence downstream of a proto-spacer sequence.

8. The method of claim 7, wherein the PAM sequence comprises a nucleic acid sequence including 5'-NGGNG-3'.

9. The method of any one of claims 1-8, wherein the site-specific modification of the target DNA molecule is cleavage of the target DNA molecule.

10. The method of claim 3, wherein the Cas9 protein contains a point mutation in the RuvC motif or the HNH motif, and wherein the modification of the target DNA molecule is site-specific nicking of the target DNA molecule.

11. The method of claim 10, wherein the point mutation in the RuvC motif is D31A and the point mutation in the HNH motif is N891A.

12. The method of claim 1, wherein the target DNA molecule is double stranded or single stranded.

13. The method of claim 1, wherein the Cas9 and/or crRNA is generated by recombinant DNA technology, *in vitro* translation or is chemically synthesized.

14. The method of claim 1, wherein the crRNA has a sequence comprising 5'-NNNNNNNNNNNNNNNNNNNNGUUUUAGAGCUGUGUUGUUUCG- 3' corresponding to SEQ ID NO: 15 with any desirable spacer sequence.

15. A method for directing a Cas9-RNA-mediated double stranded cleavage of a target DNA molecule in an isolated cell, the method comprising:

contacting a target DNA molecule in the cell with a recombinant Cas9-crRNA complex comprising a Cas9 protein, a crRNA heterologous to the Cas9 protein, and a tracrRNA, a sequence of the crRNA being complementary to a region of the target DNA molecule, wherein a region of the target DNA molecule complementary to the crRNA comprises at least 20 nucleotides, wherein the crRNA is engineered to guide the Cas9-crRNA complex to the target DNA molecule, thereby cleaving the target DNA module.

16. The method of claim 15, wherein the cell is a mammalian cell.

17. A method for directing a Cas9-RNA-mediated homologous recombination (HR) at a target DNA site in an isolated cell, the method comprising:

contacting a target DNA molecule in the cell with a Cas9-crRNA complex comprising a Cas9 protein, a crRNA, and a tracrRNA, a sequence of the crRNA being at least 80% complementary to a region of the target DNA molecule comprising the target DNA site, wherein the sequence of the crRNA comprises at least 20 nucleotides,

wherein the crRNA is engineered to guide the Cas9-crRNA complex to the target DNA site of the target DNA molecule; and

wherein a first site and a second site of homology flank the target DNA site of the target DNA molecule.

18. The method of claim 17, wherein the cell is a mammalian cell.

19. The method of claim 17, wherein the region of the target DNA molecule comprises a proto-spacer sequence, and wherein the target DNA molecule further comprises a proto-spacer adjacent motif (PAM) sequence NGGNG downstream from the proto-spacer sequence, and wherein the Cas9 protein cleaves both target DNA molecule strands at a cleavage site located 3 nucleotides upstream of the PAM sequence to create blunt ends.

20. A Cas9-crRNA complex comprising:

a Cas9 protein,

a crRNA polynucleotide heterologous to the Cas9 protein, the crRNA polynucleotide comprising a 3' region and a 5' region, wherein the 3' region comprises a repeat sequence present in a CRISPR locus and the 5' region comprises at least 20 nucleotides of a spacer sequence immediately downstream of the repeat in the CRISPR locus, and a tracrRNA polynucleotide comprising a 5' region and a 3' region wherein a sequence of the 5' region of the tracrRNA is complementary to the 3' region of the crRNA polynucleotide,

wherein the spacer sequence of the crRNA polynucleotide is engineered to be complementary to a region in a target DNA molecule, the target DNA molecule having a proto-spacer adjacent motif sequence.

21. The Cas9-crRNA complex of claim 20, wherein the Cas9 protein comprises at least one of an RuvC active site motif or an HNH active site motif.

22. The Cas9-crRNA complex of claim 21, wherein the Cas9 protein contains a point mutation in the RuvC motif and/or a point mutation in the HNH motif.

23. The Cas9-crRNA complex of claim 22, wherein the point mutation in the RuvC motif is D31A and the point mutation in the HNH motif is N891A.

24. The Cas9-crRNA complex of claim 20, wherein the complex is formed *in vivo* by introducing at least one plasmid encoding the Cas9 protein, the crRNA polynucleotide, and the tracrRNA polynucleotide into a microorganism, to result in a genetically modified microorganism, and isolating the complex from the genetically modified microorganism.

25. The Cas9-crRNA complex of claim 24, further comprising incubating the isolated Cas9 protein, crRNA polynucleotide, and tracrRNA polynucleotide under conditions suitable for complex assembly.

26. The Cas9-crRNA complex of claim 24, wherein the Cas9 protein, the crRNA polynucleotide, and the tracrRNA polynucleotide are encoded in two or three separate plasmids.

27. The Cas9-crRNA complex of claim 20, wherein the complex is formed *in vitro* by incubating the components of the complex under conditions suitable for complex assembly.

28. The Cas9-crRNA complex of claim 27, wherein the crRNA polynucleotide is obtained by *in vitro* transcription from a DNA fragment containing a single repeat-spacer- repeat unit, where the spacer has any desirable sequence, or is chemically synthesized.

29. The Cas9-crRNA complex of claim 20, wherein the Cas9 protein comprises a sequence corresponding to SEQ ID NO: l, the tracrRNA polynucleotide comprises a sequence corresponding to SEQ ID NO: 5, and the crRNA polynucleotide comprises a sequence corresponding to 5'-NNNNNNNNNNNNNNNNNNNNGUUUUAGAGCUGUGUUGUUUCG-3' with any desirable spacer sequence.

30. A method for site-specific modification of a target DNA molecule, the method comprising:

assembling a recombinant Cas9-crRNA complex *in vitro* by combining a Cas9 protein, an engineered crRNA, and a tracrRNA under conditions suitable for formation of the complex, and

contacting a target DNA molecule with the recombinant Cas9-crRNA complex *in vitro* or in an isolated cell, wherein the engineered crRNA is capable of universal targeting and programmed to guide the recombinant Cas9-crRNA complex to a region comprising a site in the target DNA molecule, wherein the engineered crRNA is reprogrammed to be heterologous to the Cas9 protein, and

wherein the site-specific modification of the target DNA molecule is cleavage of the target DNA molecule.

31. The method of claim 30, wherein a sequence of the crRNA comprising at least 20 nucleotides is substantially complementary to the target DNA molecule.

32. The method of claim 30, wherein the Cas9 protein comprises at least one of an RuvC active site motif or an HNH active site motif.

33. The method of claim 30, wherein the crRNA comprises a 3' and a 5' region, wherein the 3' region comprises at least 22 nucleotides of a CRISPR repeat and the 5' region

comprises at least 20 nucleotides of a spacer sequence engineered to be substantially complementary to the region of the target DNA.

34. The method of claim 30, wherein the target DNA molecule comprises a proto-spacer adjacent motif (PAM) sequence.

35. The method of claim 34, wherein the PAM sequence comprises a nucleic acid molecule having the nucleic acid sequence 5'-NGGNG.

36. The method of claim 32, wherein one of the RuvC active site motif or the HNH active site motif of the Cas9 protein is inactivated, and wherein the target DNA is double stranded and the modification of the target DNA molecule is site-specific nicking of a single strand of the target DNA molecule.

37. The method of claim 30, wherein the target DNA is double stranded or single stranded.

38. The method of claim 34, wherein the PAM sequence comprises a proto-spacer upstream of the PAM.

39. The method of claim 30, wherein the recombinant Cas9-crRNA complex remains associated with the target DNA molecule after cleavage.

40. The method of claim 37, wherein the target DNA is double stranded and the proto-spacer adjacent motif (PAM) sequence is recognized in the context of double stranded DNA.

41. The method of claim 36, wherein the RuvC active site motif or the HNH active site motif of the Cas9 protein is inactivated by a point mutation.

42. The method of claim 41, wherein the point mutation in the RuvC motif is D31A and the point mutation in the HNH motif is N891A.

43. The method of claim 30 wherein, prior to assembly of the recombinant Cas9-crRNA complex, the engineered crRNA and the tracrRNA are produced by *in vitro* transcription or chemical synthesis; and Cas9 protein is produced by a method selected from recombinant DNA technology and chemical synthesis.

44. The method of claim 43 wherein Cas9 protein is isolated from a genetically modified organism.

45. The method of claim 30 wherein, prior to assembly of the recombinant Cas9-crRNA complex, Cas9 protein, the engineered crRNA, and the tracrRNA are isolated from a genetically modified microorganism.

46. The method of claim 45 wherein the genetically modified microorganism is created by introducing at least one plasmid encoding Cas9 protein, the engineered crRNA, and the tracrRNA into the microorganism, to result in the genetically modified microorganism.

47. The method of claim 46 wherein Cas9 protein, the engineered crRNA, and the tracrRNA are encoded in two separate plasmids.

48. The method of claim 46 wherein Cas9 protein, the engineered crRNA, and the tracrRNA are encoded in three separate plasmids.

49. The method of claim 30 wherein Cas9 is at least 80% identical to SEQ ID NO: 1.

50. The method of claim 30 wherein the engineered crRNA and the tracrRNA are assembled into a duplex, and then Cas9 protein is incubated with the duplex, to result in the recombinant Cas9-crRNA complex.

51. A composition comprising:

a recombinant Cas9-crRNA complex comprising:

a Cas9 protein containing a mutation in a RuvC nuclease domain or an HNH nuclease domain of the Cas9 protein;

a tracrRNA; and

a crRNA containing a single repeat-spacer unit, wherein the crRNA is heterologous to the Cas9 protein and is engineered to have a sequence complementary to a proto-spacer sequence in the target DNA molecule, to guide the recombinant Cas9-crRNA complex to a region comprising a site in a target DNA molecule, wherein the sequence of the crRNA comprises at least 20 nucleotides;

wherein the Cas9-crRNA complex is capable of cleaving the target DNA molecule in a cell-free condition, and

wherein the composition is cell-free.

52. The composition of claim 51, wherein the target DNA molecule is a double-stranded DNA molecule.

53. The composition of claim 51, wherein the target DNA molecule is a plasmid DNA.

54. The composition of claim 51, wherein the Cas9 protein comprises a fusion polypeptide comprising at least one additional amino acid sequence.

55. The composition of claim 54, wherein the additional amino acid sequence is used for purification of the fusion polypeptide by affinity chromatography.

56. The composition of claim 51, wherein the Cas9 protein contains a mutation in the RuvC nuclease domain and in the HNH nuclease domain of the Cas9 protein.

57. The composition of claim 56, wherein the mutation of the Cas9 protein is a point mutation in the RuvC nuclease domain corresponding to D31A of SEQ ID NO: 1.

58. The composition of claim 51, wherein the Cas9 protein contains a point mutation in the HNH nuclease domain of the Cas9 protein.

59. The composition of claim 51, wherein the mutation of the Cas9 protein is a point mutation in the HNH nuclease domain corresponding to N891A of SEQ ID NO: 1.

60. The composition of claim 51, wherein the engineered crRNA and the tracrRNA are produced by *in vitro* transcription or chemical synthesis, and the Cas9 protein is produced by recombinant DNA technology or chemical synthesis.

61. The composition of claim 51, wherein each of the engineered crRNA and the tracrRNA is isolated from a genetically modified microorganism.

62. The composition of claim 51, wherein the recombinant Cas9-crRNA complex consists essentially of the Cas9 protein, the tracrRNA, and the engineered crRNA.

63. A composition comprising

a recombinant Cas9-crRNA complex comprising:

a Cas9 protein containing a mutation in a RuvC nuclease domain or an HNH nuclease domain of the Cas9 protein;

a tracrRNA; and

a crRNA containing no more than one repeat-spacer unit, wherein the engineered crRNA is heterologous to the Cas9 protein, comprises a sequence of at least 20 consecutive nucleotides complementary to a heterologous target DNA molecule, and is reprogrammed to guide the recombinant Cas9-crRNA complex to the target DNA molecule;

wherein the composition is cell-free.

64. The composition of claim 63, wherein the Cas9-crRNA complex is formed *in vitro*.

65. The composition of claim 63, wherein the DNA molecule is a double-stranded DNA molecule.

66. A universal RNA-guided DNA endonuclease system comprising:

a. a Cas9 protein; and

b. a plasmid encoding a tracrRNA and a separate crRNA, wherein the plasmid contains a single crRNA repeat-spacer unit, the spacer engineered to be complementary to a target DNA sequence for site specific cleavage of the target DNA sequence in the absence of RNaseIII;

wherein the plasmid does not encode the Cas9 protein; and

wherein the system does not comprise RNaseIII.

67. The universal RNA-guided DNA endonuclease system of Claim 66, wherein the crRNA comprises an approximately 20 nucleotide fragment complementary to a sequence in the target DNA sequence.

68. The universal RNA-guided DNA endonuclease system of Claim 66, wherein the target DNA sequence is a plasmid DNA.

69. The universal RNA-guided DNA endonuclease system of Claim 66, wherein the target DNA sequence is double stranded DNA.

70. The universal RNA-guided DNA endonuclease system of Claim 66, wherein the Cas9 protein comprises a fusion polypeptide comprising at least one additional amino acid sequence.

71. The universal RNA-guided DNA endonuclease system of Claim 70, wherein the additional amino acid sequence is used for purification of the fusion polypeptide by affinity chromatography.

72. The universal RNA-guided DNA endonuclease system of Claim 66, wherein the Cas9 protein contains a point mutation in an RuvC active site motif of the Cas9 protein.

73. The universal RNA-guided DNA endonuclease system of Claim 72, wherein the Cas9 protein is at least 80% identical to SEQ ID NO: 1 and the point mutation is D31A.

74. The universal RNA-guided DNA endonuclease system of Claim 66, wherein the Cas9 protein contains a point mutation in an HNH active site motif of the Cas9 protein.

75. The universal RNA-guided DNA endonuclease system of Claim 74, wherein the Cas9 protein is at least 80% identical to SEQ ID NO: 1 and the point mutation is N891A.

76. The universal RNA-guided DNA endonuclease system of Claim 66, wherein the crRNA comprises 42 nucleotides.

77. The universal RNA-guided DNA endonuclease system of Claim 66, wherein the spacer is proximate a 5'-end of the crRNA.

78. The universal RNA-guided DNA endonuclease system of Claim 66, wherein the crRNA comprises a fragment of at least 20 consecutive nucleotides complementary to the target DNA sequence.

79. A method of preparing a Cas9-crRNA complex for modification of a target DNA molecule, the method comprising:

contacting a Cas9 polypeptide with a crRNA and a tracrRNA *in vitro* to form the Cas9-crRNA complex, wherein the crRNA is engineered to site-specifically bind to a polynucleotide sequence of the target DNA molecule, the crRNA comprising a polynucleotide sequence complementary to the polynucleotide sequence of the target DNA molecule, wherein the polynucleotide sequence of the crRNA comprises at least 20 nucleotides.

80. The method of claim 79, further comprising:

contacting the Cas9-crRNA complex with the target DNA molecule *in vitro* or in an isolated cell, whereby the Cas9-crRNA complex site-specifically binds to the target DNA molecule.

81. The method of claim 80, wherein the target DNA molecule is double stranded.

82. The method of claim 79, wherein the Cas9 polypeptide is produced by recombinant DNA technology or chemical synthesis.

83. The method of claim 79, wherein the Cas9 polypeptide is isolated from a genetically modified microorganism.

84. The method of claim 79, further comprising:

contacting the Cas9-crRNA complex with the target DNA molecule *in vitro* or in an isolated cell, whereby the Cas9-crRNA complex site-specifically cleaves the target DNA molecule.

85. The method of claim 79, further comprising producing the crRNA by chemical synthesis.

86. The method of claim 79, wherein the Cas9-crRNA complex is formed in the absence of RNaseIII.

87. The method of claim 79, wherein the Cas9 polypeptide comprises a mutation in a RuvC active site motif or a HNH active site motif of the Cas9 polypeptide.

88. A method of preparing a Cas9-crRNA complex for modification of a heterologous target DNA molecule, the method comprising:

contacting a Cas9 polypeptide with a crRNA and a tracrRNA *in vitro* to form the Cas9-crRNA complex, wherein the crRNA is heterologous to the Cas9 polypeptide and engineered to bind the Cas9-crRNA complex to the target DNA molecule, and wherein the crRNA is not generated through processing of a bacterial CRISPR repeat-spacer array; and

contacting the target DNA molecule with the Cas9-crRNA complex *in vitro* or in an isolated cell, whereby the Cas9-crRNA complex modifies the target DNA molecule.

89. The method of claim 88, wherein the crRNA is obtained by:

identifying a polynucleotide sequence of the target DNA molecule; and

preparing the crRNA, such that the crRNA includes a polynucleotide sequence complementary to the polynucleotide sequence of the DNA molecule, wherein the polynucleotide sequence comprises at least 20 nucleotides.

90. The method of claim 88, wherein the crRNA is prepared by chemical synthesis.

91. The method of claim 88, wherein the Cas9 polypeptide comprises a mutation in a RuvC active site motif or a HNH active site motif of the Cas9 polypeptide.

92 The method of claim 88, wherein the Cas9 polypeptide is contacted with the crRNA and the tracrRNA in the absence of RNaseIII.

93. The method of claim 88, wherein the target DNA molecule includes a protospacer-adjacent motif.

94. The method of claim 88, wherein the target DNA molecule is double stranded, and wherein the Cas9-crRNA complex modifies the target DNA molecule by site-specific double stranded cleavage of the target DNA molecule.

95. A method of preparing a Cas9-crRNA complex for modification of a target DNA molecule, the method comprising:

identifying a polynucleotide sequence of the target DNA molecule;

preparing a crRNA having a polynucleotide sequence complementary to the polynucleotide sequence of the DNA molecule, wherein the polynucleotide sequence of the crRNA comprises at least 20 nucleotides; and

contacting a Cas9 polypeptide with the crRNA and a tracrRNA to form the Cas9-crRNA complex, wherein the crRNA is engineered to site-specifically bind to the target DNA molecule and is heterologous to the Cas9 polypeptide.

96. The method of claim 95, wherein the Cas9 polypeptide comprises a mutation in a RuvC active site motif or a HNH active site motif of the Cas9 polypeptide.

97. The method of claim 95, wherein the Cas9 polypeptide is isolated from a genetically modified microorganism, or wherein the Cas9 polypeptide is produced by recombinant DNA technology or chemical synthesis.

98. The method of claim 95, further comprising:

contacting the target DNA molecule with the Cas9-crRNA complex *in vitro* or in an isolated cell, whereby the Cas9-crRNA complex cleaves the target DNA molecule.

99. The method of claim 98, wherein the crRNA is heterologous to the target DNA molecule.

100. A method of preparing a Cas9-crRNA complex for modification of a target DNA molecule, the method comprising:

contacting a recombinant Cas9 polypeptide with a crRNA and a tracrRNA to form the Cas9-crRNA complex, wherein the crRNA is programmed to site-specifically bind to the target DNA molecule, and wherein the crRNA is heterologous to the Cas9 polypeptide and is not generated through processing of a bacterial CRISPR repeat-spacer array; and

contacting the target DNA molecule with the Cas9-crRNA complex, whereby the Cas9-crRNA complex site-specifically modifies the target DNA molecule;

wherein the modification occurs *in vitro* or in an isolated cell.

101. The method of claim 100, wherein the Cas9 polypeptide comprises a mutation in a RuvC active site motif or a HNH active site motif, whereby the Cas9 polypeptide is a nickase; and wherein the Cas9-crRNA complex site-specifically modifies the target DNA molecule by nicking the target DNA molecule.

102. The method of claim 100, wherein the Cas9-crRNA complex is formed in the absence of RNaseIII.

103. A method for site-specific modification of a DNA molecule, the method comprising:

introducing a Cas9 protein, a tracrRNA, and a crRNA into an isolated cell, the crRNA comprising at least 20 nucleotides and engineered to have a nucleotide sequence complementary to a nucleotide sequence of the DNA molecule;

wherein a recombinant Cas9-crRNA complex formed from the Cas9 protein, the tracrRNA, and the crRNA cleaves a target site of the DNA molecule *in vitro* or in the isolated cell; and

wherein the cleaved target site of the DNA molecule is modified by non-homologous end joining or homologous recombination.

104. The method of claim 103, wherein the nucleotide sequence of the crRNA comprises 42 nucleotides.

105. The method of claim 103, wherein the crRNA is synthetic.

106. The method of claim 103, wherein the tracrRNA is synthetic.

107. The method of claim 103, wherein the tracrRNA is an *in-vitro* transcribed RNA.

108. The method of claim 103, wherein the DNA molecule is a plasmid.

109. The method of claim 103, wherein introducing the Cas9 protein, the tracrRNA, and the crRNA into the cell comprises transfecting the cell with the Cas9 protein, the tracrRNA, and the crRNA.

110. The method of claim 103, wherein the target site of the DNA molecule is located in a coding portion of the nucleotide sequence of the DNA molecule.

111. The method of claim 103, wherein the target site of the DNA molecule is located in an intron.

112. The method of claim 103, wherein the cleaved target site of the DNA molecule is modified by non-homologous end joining.

113. The method of claim 112, wherein the non-homologous end joining results in deletion of a coding portion of the nucleotide sequence of the DNA molecule.

114. The method of claim 113, wherein the coding portion comprises fewer than 20 nucleotides.

115. A method for site-specific modification of a DNA molecule, the method comprising:

combining a Cas9 protein, a tracrRNA, and a crRNA *in vitro*, the crRNA having at least 20 nucleotides including a nucleotide sequence complementary to a nucleotide sequence of the DNA molecule; and

contacting the combined Cas9 protein, tracrRNA, and crRNA with the DNA molecule *in vitro* or in an isolated cell,

wherein a recombinant Cas9-crRNA complex formed from the Cas9 protein, the tracrRNA, and the crRNA cleaves the DNA molecule; and

wherein the DNA molecule is modified by non-homologous end joining or homologous recombination.

116. The method of claim 115, wherein the recombinant Cas9-crRNA complex cleaves the DNA molecule *in vitro*.

117. The method of claim 115, wherein contacting the combined Cas9 protein, tracrRNA, and crRNA with the DNA molecule comprises transfecting the cell with the recombinant Cas9- crRNA complex.

118. The method of claim 115, wherein a cleaved target site of the DNA molecule is modified by non-homologous end joining.

119. The method of claim 118, wherein the non-homologous end joining results in deletion of a coding portion of the nucleotide sequence of the DNA molecule, and wherein the coding portion comprises fewer than 20 nucleotides.

120. A method for site-specific modification of a DNA molecule, the method comprising:

combining a Cas9 protein, a tracrRNA, and a crRNA *in vitro*, the crRNA having at least 20 nucleotides including a nucleotide sequence complementary to a nucleotide sequence of the DNA molecule;

introducing the combined Cas9 protein, tracrRNA, and crRNA into an isolated cell; and

introducing the DNA molecule into the cell,

wherein a recombinant Cas9-crRNA complex formed from the Cas9 protein, the tracrRNA, and the crRNA cleaves a target site of the DNA molecule; and

wherein the cleaved target site of the DNA molecule is modified by non-homologous end joining or homologous recombination.

121. The method of claim 120, wherein the DNA molecule is a plasmid.

122. The method of claim 120, wherein introducing the DNA molecule into the cell comprises transfecting the cell with the DNA molecule.

123. The method of claim 122, wherein the cleaved target site of the DNA molecule is modified by non-homologous end joining.

124. A method for directing a Cas9-crRNA-mediated cleavage of a target DNA molecule in an isolated prokaryotic cell, the method comprising,

assembling a recombinant Cas9-crRNA complex *in vitro* by combining a Cas9 protein, an engineered crRNA, and a tracrRNA under conditions suitable for formation of the complex, and

contacting a target DNA molecule in the cell with the recombinant Cas9-crRNA complex, wherein the engineered crRNA is programmed to guide the recombinant Cas9-crRNA complex to a region comprising a desired specific site in the target DNA molecule,

wherein the recombinant Cas9-crRNA complex cleaves the target DNA molecule at the desired specific site.

125. The method of claim 124, wherein the engineered crRNA comprises at least 20 nucleotides having substantial complementarity to the target DNA.

126. The method of claim 124, wherein the Cas9 protein comprises an RuvC active site motif and an HNH active site motif.

127. The method of claim 124, wherein the engineered crRNA comprises at least 22 nucleotides of a repeat sequence present in a CRISPR locus.

128. The method of claim 124, wherein the tracrRNA comprises a 5' region and a 3' region, wherein at least a portion of the 5' region is complementary to the 3' region of the engineered crRNA.

129. The method of claim 124, wherein the target DNA molecule comprises a proto-spacer adjacent motif (PAM) sequence downstream of a proto-spacer sequence.

130. The method of claim 129, wherein the PAM sequence comprises a nucleic acid molecule having the nucleic acid sequence 5'-NGGNG.

131. The method of claim 126, wherein the Cas9 protein contains a point mutation in the RuvC motif or the HNH motif, and wherein the cleavage of the target DNA molecule is site-specific nicking of the target DNA molecule.

132. The method of claim 131, wherein the point mutation in the RuvC motif corresponds to a D31A mutation in SEQ ID NO: 1 and the point mutation in the HNH motif corresponds to a N891A mutation in SEQ ID NO:1.

133. The method of claim 124, wherein the target DNA is double stranded or single stranded.

134. The method of claim 124, wherein the Cas9 and/or engineered crRNA is generated by recombinant DNA technology, *in vitro* transcription/translation or chemical synthesis.

135. The method of claim 124, wherein the engineered crRNA has a sequence comprising 5'-NNNN GUUUUAGAGCUGUGUUGUUUCG-3' (SEQ ID NO: 15) with any desirable spacer sequence.

136. Use of the Cas9-crRNA complex of claim 20 to site-specifically modify the target DNA molecule in an isolated cell.

137. Use of the Cas9-crRNA complex of claim 20 to site-specifically modify the target DNA molecule in an isolated mammalian cell.

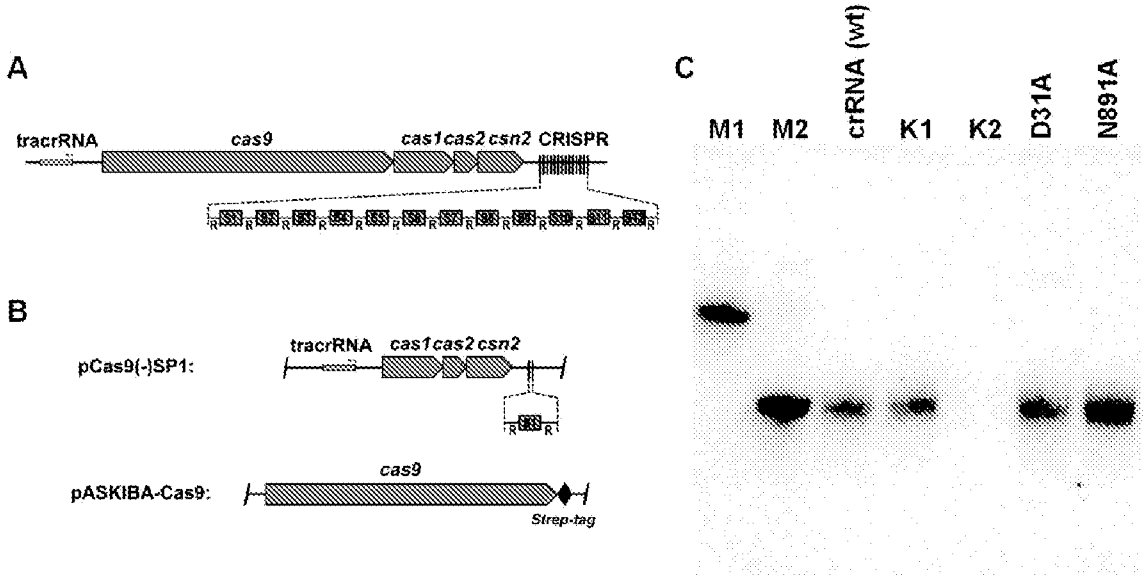

A
tracrRNA
cas9
cas1 cas2 csn2 CRISPR
R
B
pCas9(-)SP1:
tracrRNA
cas1 cas2 csn2
R
pASKIBA-Cas9:
cas9
Strep-tag
C
M1
M2
crRNA (wt)
K1
K2
D31A
N891A

Figure 1.

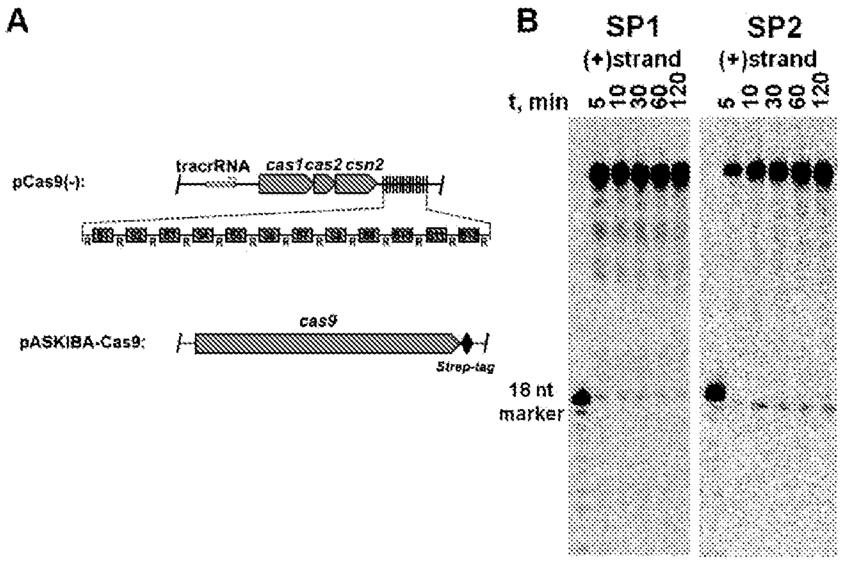

A
pCas9(-):
tracrRNA
cas1
cas2
csn2
R
pASKIBA-Cas9:
cas9
Strep-tag
B
SP1
(+)strand
SP2
(+)strand
t, min
5
10
30
60
120
5
10
30
60
120
18 nt
marker

Figure 2.

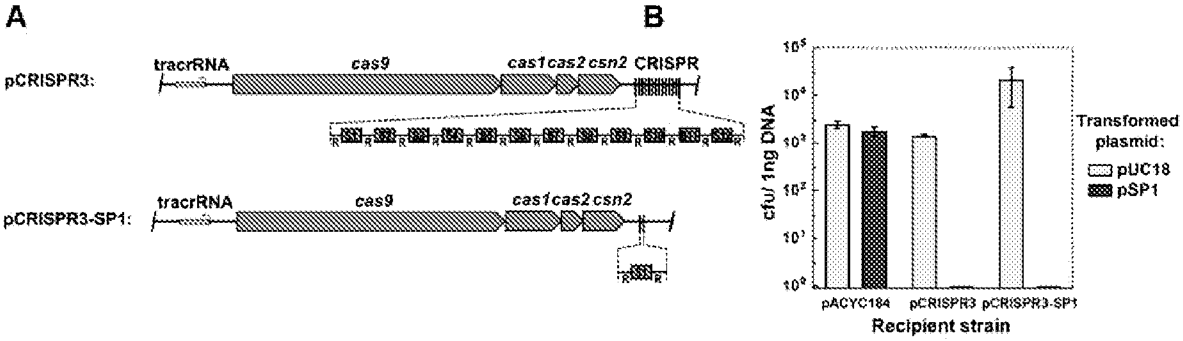
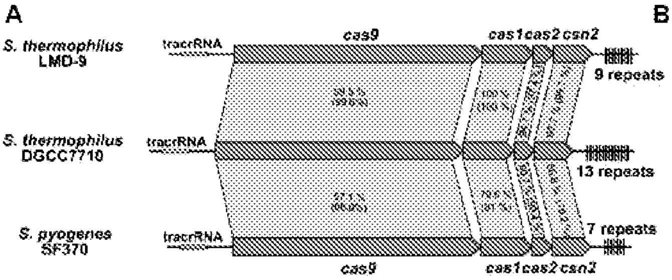

A
B
pCRISPR3:
tracrRNA
cas9
cas1
cas2
csn2
CRISPR
R
pCRISPR3-SP1:
tracrRNA
cas9
cas1
cas2
csn2
R
R
cfu/ 1ng DNA
10^0
10^1
10^2
10^3
10^4
10^5
pACYC184
pCRISPR3
pCRISPR3-SP1
Recipient strain
Transformed plasmid:
pUC18
pSP1

Figure 3.

A



B Repeat sequences:

```
                 10         20         30
          ....|....| ....|....| ....|....| ....|.
DGCC7710    GTTTTAGAGC TGTGTTGTTT CGAATGGTTC CAAAAC
LMD-9       GTTTTAGAGC TGTGTTGTTT CGAATGGTTC CAAAAC
S.pyrogenes GTTTTAGAGC TATGCTGTTT TGAATGGTCC CAAAAC
            ********** * ** *****  ******* * ******
```

tracrRNA encoding DNA:

```
                 10         20         30
          ....|....| ....|....| ....|....| ....|.
DGCC7710    TTTAACTCGC TGTGTTGTTT CGAATGGTTT CAAACC
LMD-9       TTTAACTCGC TGTGTTGTTT CGAATGGTTT CAAACC
S.pyrogenes TTTAACTTGC TATGCTGTTT TGAATGGTTC CAACAA
            ******* ** * ** *****  ********  ***
```

C Putative crRNA of *S. thermophilus*:

```
               -20       -10                    10        20
               :....:....:....:....   ....:....:....:....:..
S.pyogenes  5'-GAUUUCUUCUUGCGCUUUUU   GUUUUAGAGCUAUGCUGUUUUG-3'  Deep sequencing
LMD-9       5'-GUUCACUGUACGAGUACUUA   GUUUUAGAGCUGUGUUGUUUCG-3'  Nothern blot
DGCC7710    5'-CGCUAAAGAGGAAGAGGACA   GUUUUAGAGCUGUGUUGUUUCG-3'  Nothern blot
                                      *********** ** ***** *
                     Spacer                    Repeat
```

Figure 4.

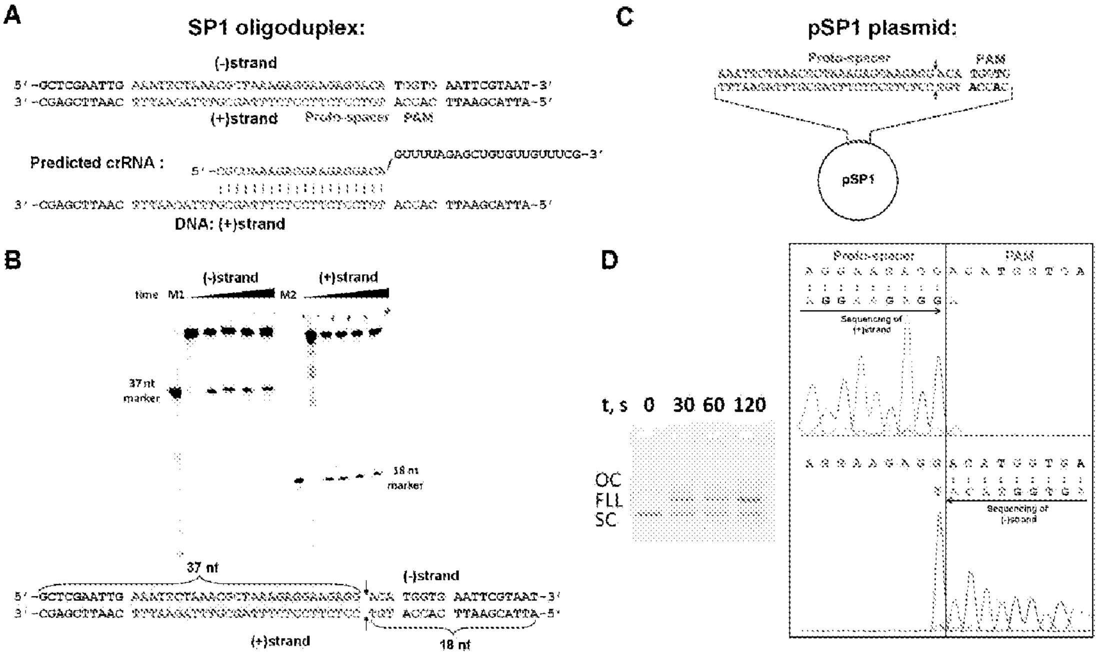

A
SP1 oligoduplex:
(-)strand
5'-GCTCGAATTG AAATTCTAAACGCTAAAGAGGAAGAGGACA TGGTG AATTCGTAAT-3'
3'-CGAGCTTAAC TTTAAGATTTGCGATTTCTCCTTCTCCTGT ACCAC TTAAGCATTA-5'
(+)strand
Proto-spacer
PAM
Predicted crRNA :
5'-CGCUAAAGAGGAAGAGGACA
GUUUUAGAGCUGUGUUGUUUCG-3'
3'-CGAGCTTAAC TTTAAGATTTGCGATTTCTCCTTCTCCTGT ACCAC TTAAGCATTA-5'
DNA: (+)strand
B
(-)strand
(+)strand
time
M1
M2
37 nt marker
18 nt marker
37 nt
(-)strand
5'-GCTCGAATTG AAATTCTAAACGCTAAAGAGGAAGAGG ACA TGGTG AATTCGTAAT-3'
3'-CGAGCTTAAC TTTAAGATTTGCGATTTCTCCTTCTCC TGT ACCAC TTAAGCATTA-5'
(+)strand
18 nt
C
pSP1 plasmid:
Proto-spacer
PAM
AAATTCTAAACGCTAAAGAGGAAGAGG ACA TGGTG
TTTAAGATTTGCGATTTCTCCTTCTCC TGT ACCAC
pSP1
D
t, s 0 30 60 120
OC
FLL
SC
Proto-spacer
PAM
A G G A A G A G G A C A T G G T G A
A G G A A G A G G A
Sequencing of (+)strand
A G G A A G A G G A C A T G G T G A
T A C A T G G T G A
Sequencing of (-)strand

Figure 5.

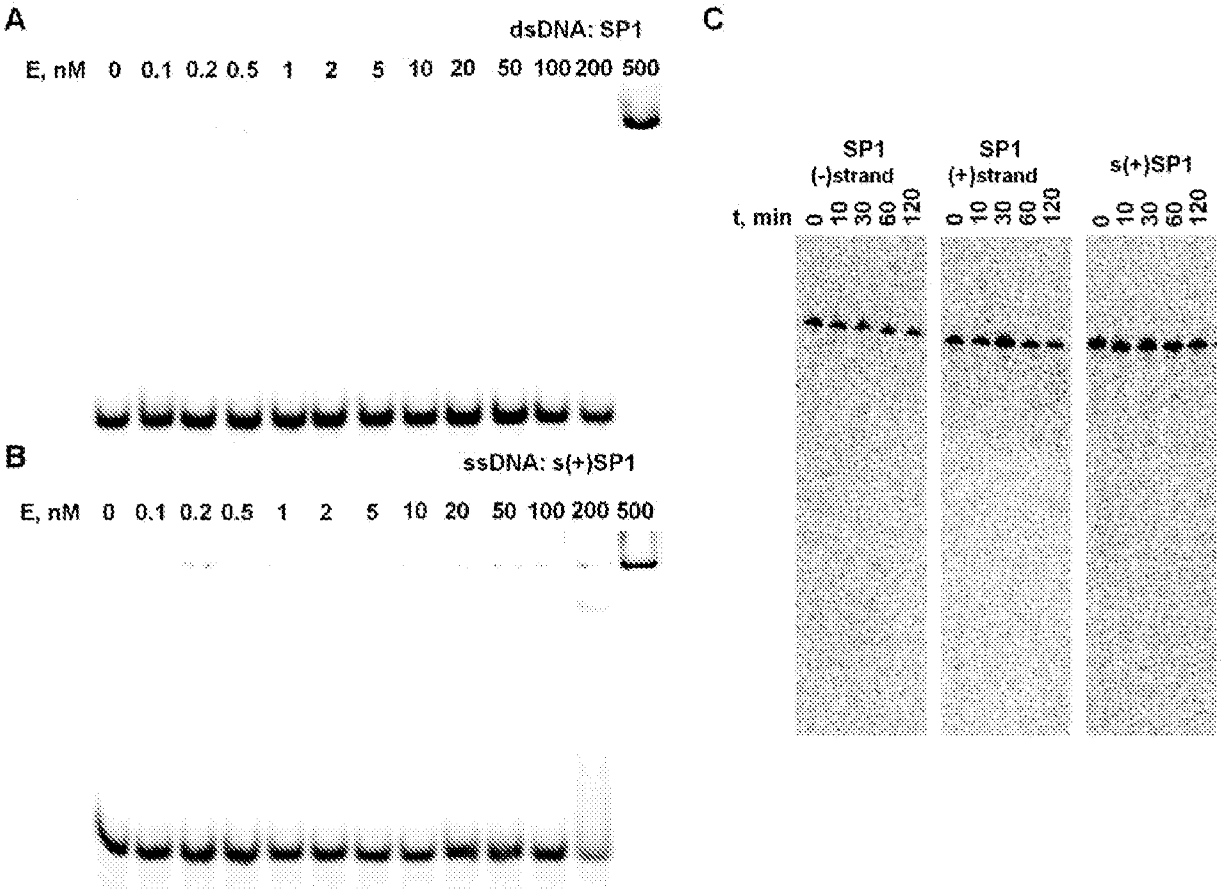

A
dsDNA: SP1
E, nM 0 0.1 0.2 0.5 1 2 5 10 20 50 100 200 500
B
ssDNA: s(+)SP1
E, nM 0 0.1 0.2 0.5 1 2 5 10 20 50 100 200 500
C
SP1 (-)strand
SP1 (+)strand
s(+)SP1
t, min
0 10 30 60 120
0 10 30 60 120
0 10 30 60 120

Figure 6.

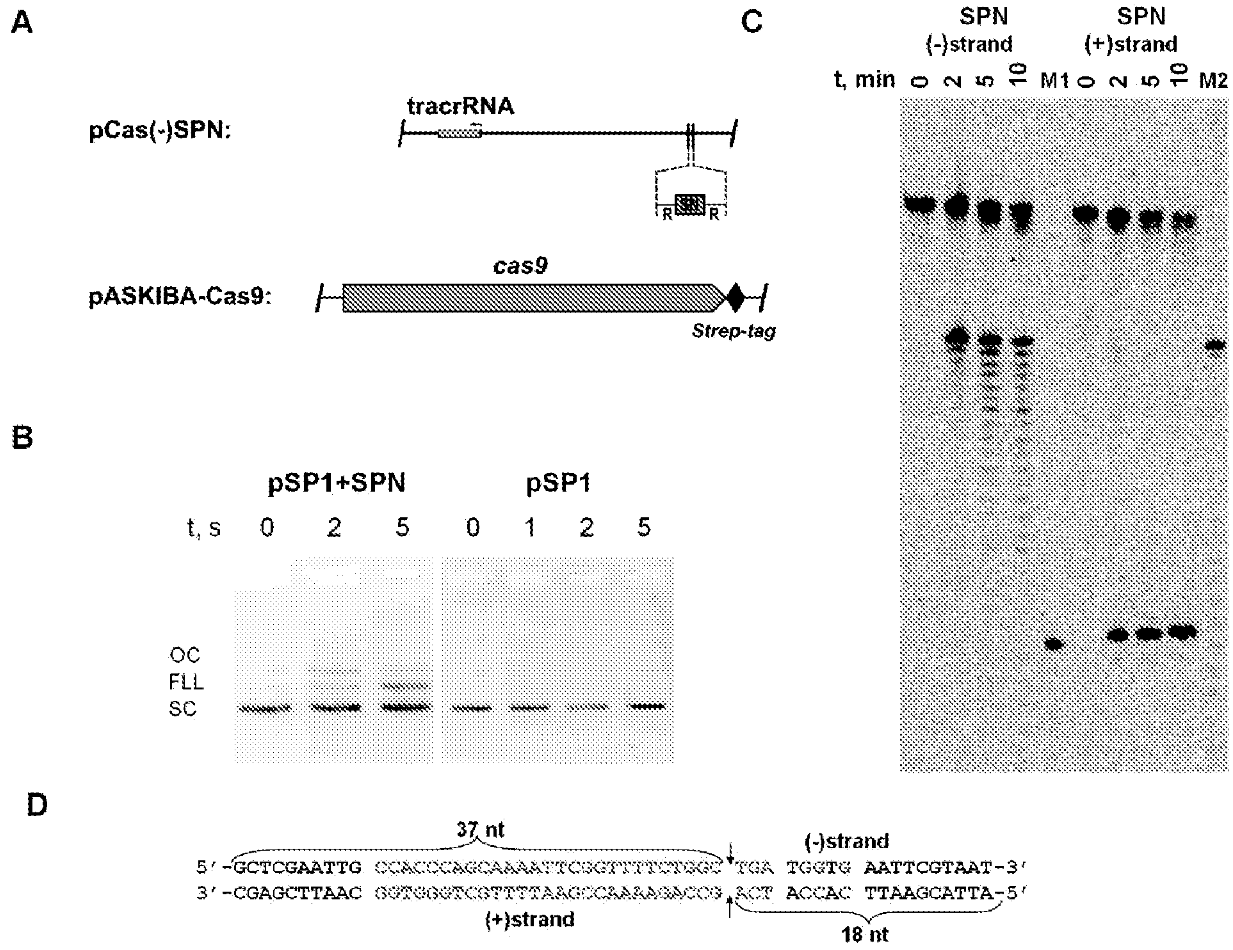

A
pCas(-)SPN:
tracrRNA
R
SPN
R
pASKIBA-Cas9:
cas9
Strep-tag
B
pSP1+SPN
pSP1
t, s
0
2
5
0
1
2
5
OC
FLL
SC
C
SPN
(-)strand
SPN
(+)strand
t, min
0
2
5
10
M1
0
2
5
10
M2
D
37 nt
(-)strand
5′-GCTCGAATTG CCACCCAGCAAAATTCGGTTTTCTGGC TGA TGGTG AATTCGTAAT-3′
3′-CGAGCTTAAC GGTGGGTCGTTTTAAGCCAAAAGACCG ACT ACCAC TTAAGCATTA-5′
(+)strand
18 nt

Figure 7.

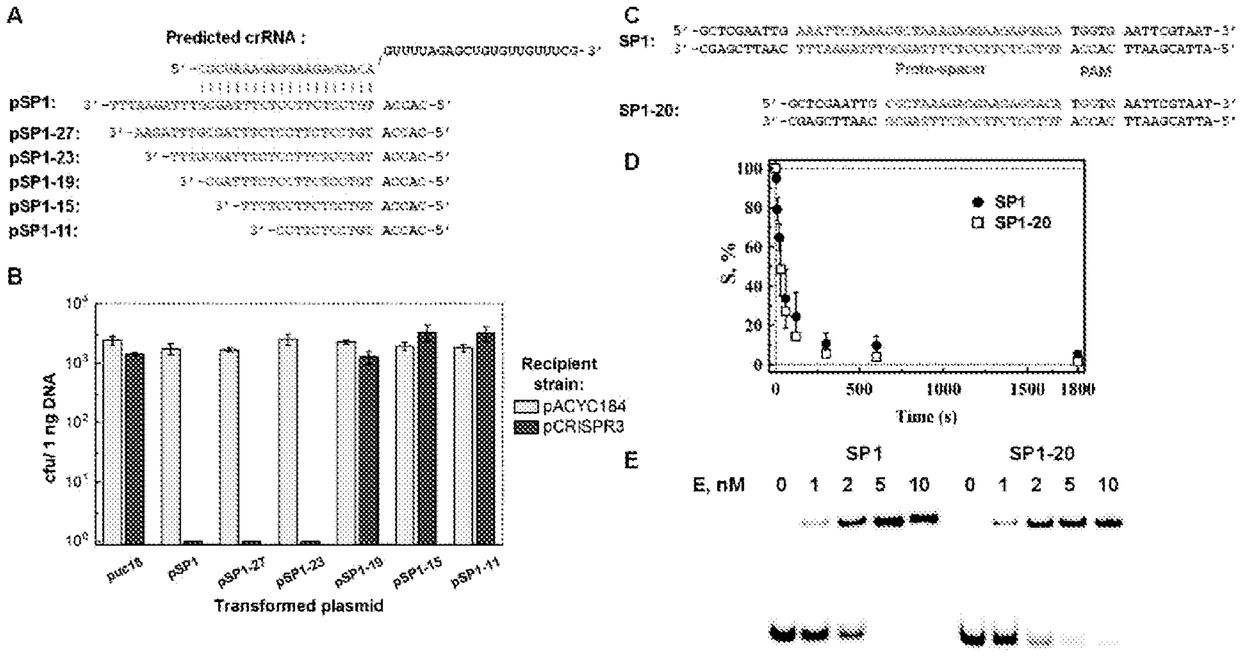

A
Predicted crRNA :
GUUUUAGAGCUGUGUUGUUUCG-3'
pSP1:
pSP1-27:
pSP1-23:
pSP1-19:
pSP1-15:
pSP1-11:
B
cfu/ 1 ng DNA
Recipient strain:
pACYC184
pCRISPR3
puc18
pSP1
pSP1-27
pSP1-23
pSP1-19
pSP1-15
pSP1-11
Transformed plasmid
C
SP1:
5'-GCTCGAATTG
TGGTG AATTCGTAAT-3'
3'-CGAGCTTAAC
ACCAC TTAAGCATTA-5'
Proto-spacer
PAM
SP1-20:
5'-GCTCGAATTG
TGGTG AATTCGTAAT-3'
3'-CGAGCTTAAC
ACCAC TTAAGCATTA-5'
D
S, %
SP1
SP1-20
Time (s)
0
500
1000
1500
1800
E
E, nM
SP1
SP1-20
0 1 2 5 10
0 1 2 5 10

Figure 8.

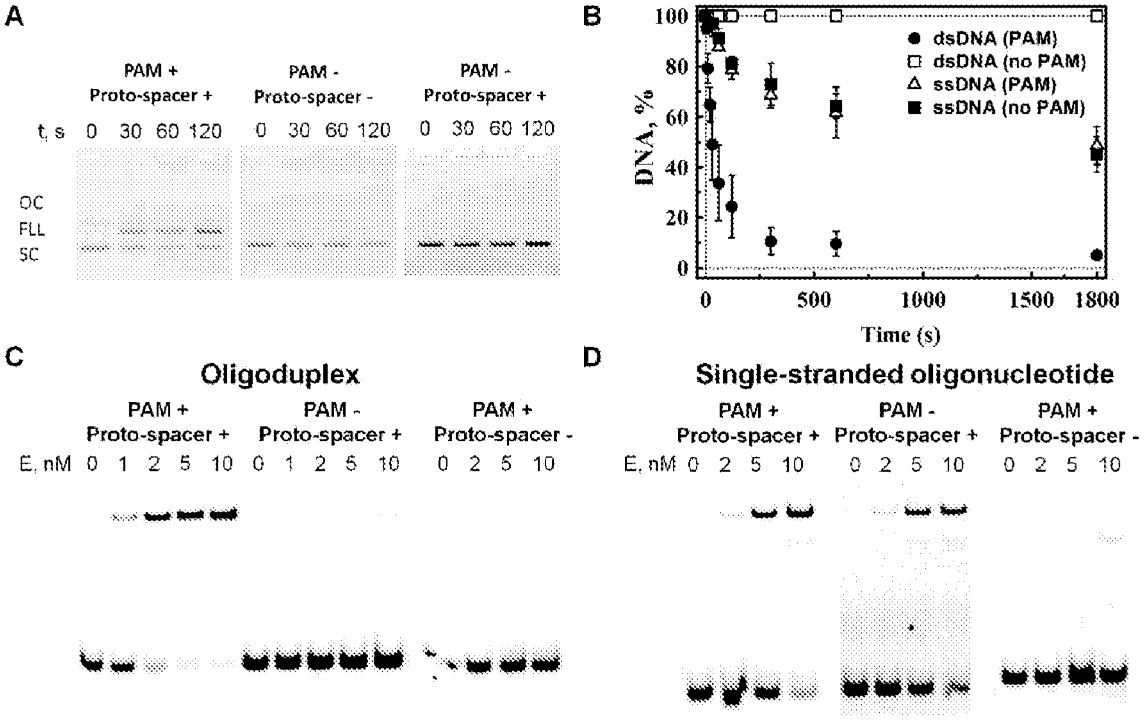

A
PAM +
Proto-spacer +
PAM -
Proto-spacer -
PAM -
Proto-spacer +
t, s 0 30 60 120 0 30 60 120 0 30 60 120
OC
FLL
SC
B
100
80
60
40
20
0
DNA, %
0
500
1000
1500
1800
Time (s)
● dsDNA (PAM)
□ dsDNA (no PAM)
△ ssDNA (PAM)
■ ssDNA (no PAM)
C
Oligoduplex
PAM +
Proto-spacer +
PAM -
Proto-spacer +
PAM +
Proto-spacer -
E, nM 0 1 2 5 10 0 1 2 5 10 0 2 5 10
D
Single-stranded oligonucleotide
PAM +
Proto-spacer +
PAM -
Proto-spacer +
PAM +
Proto-spacer -
E, nM 0 2 5 10 0 2 5 10 0 2 5 10

Figure 9.

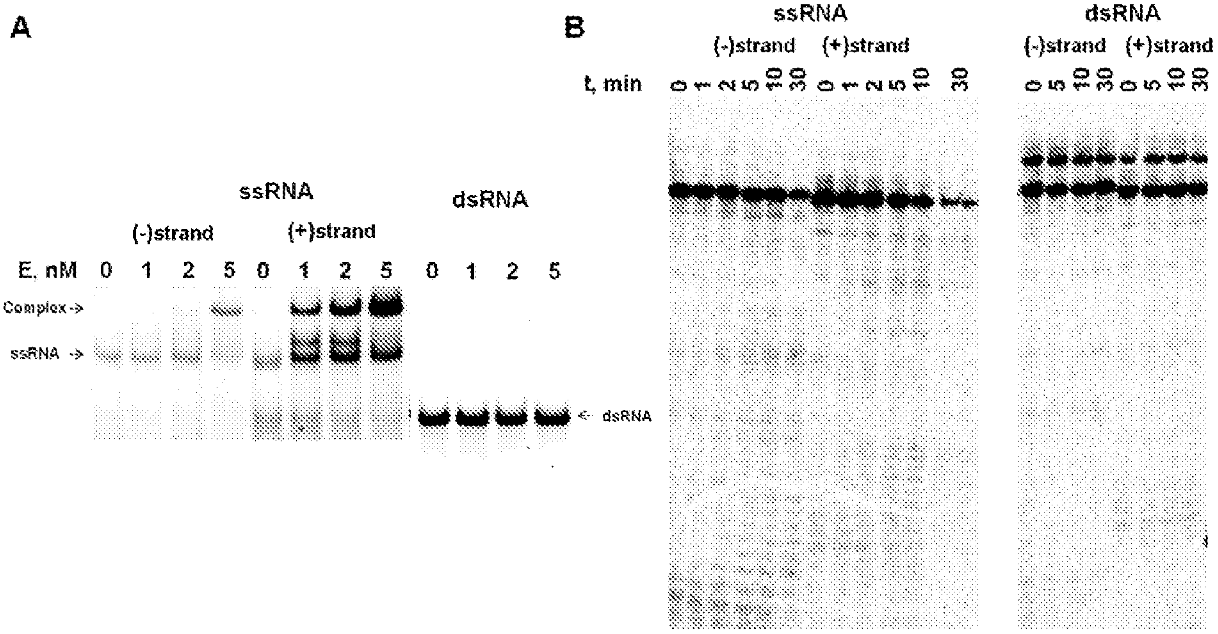

A
ssRNA
(-)strand
(+)strand
dsRNA
E, nM 0 1 2 5 0 1 2 5 0 1 2 5
Complex ->
ssRNA ->
<- dsRNA
B
ssRNA
(-)strand
(+)strand
t, min 0 1 2 5 10 30 0 1 2 5 10 30
dsRNA
(-)strand
(+)strand
0 5 10 30 0 5 10 30

Figure 10

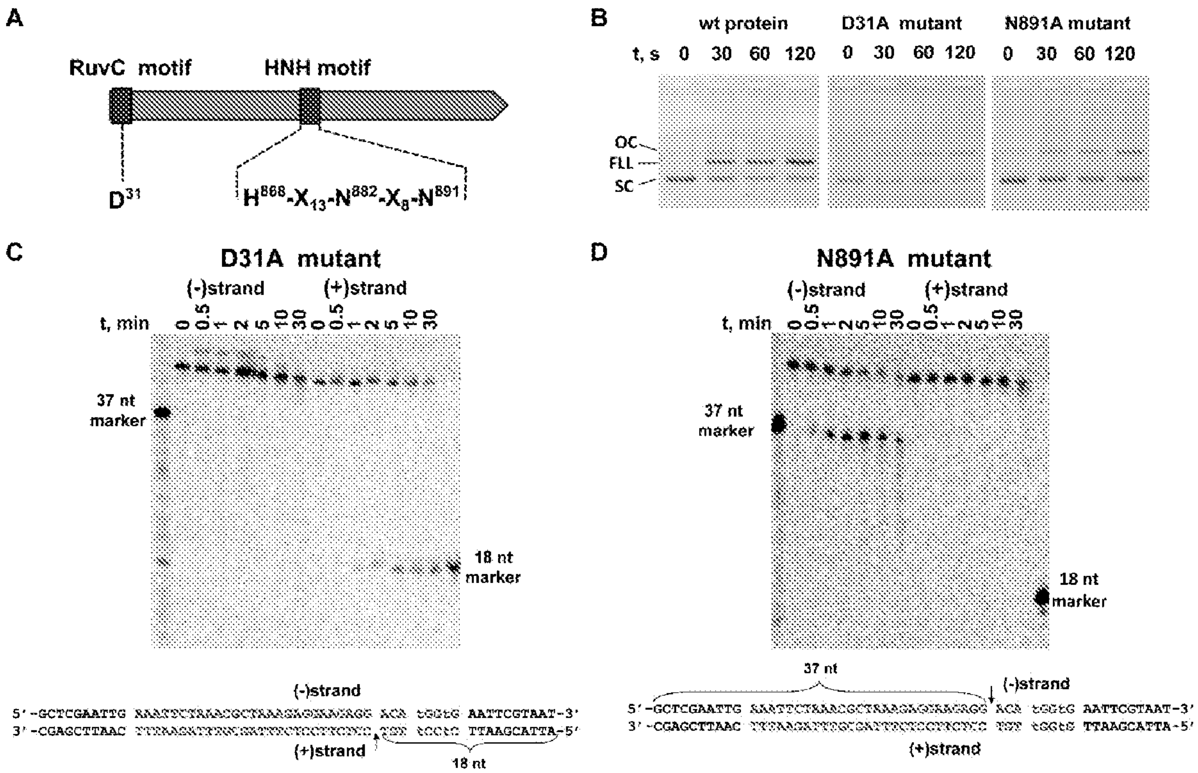

A
RuvC motif
HNH motif
D^{31}
H^{868}-X_{13}-N^{882}-X_8-N^{891}
B
wt protein
D31A mutant
N891A mutant
t, s 0 30 60 120 0 30 60 120 0 30 60 120
OC
FLL
SC
C
D31A mutant
(-)strand
(+)strand
t, min 0 0.5 1 2 5 10 30 0 0.5 1 2 5 10 30
37 nt marker
18 nt marker
D
N891A mutant
(-)strand
(+)strand
t, min 0 0.5 1 2 5 10 30 0 0.5 1 2 5 10 30
37 nt marker
18 nt marker
(-)strand
5'-GCTCGAATTG AAATTCTAAACGCTAAAGAGGAAGAGG ACA tGGtG AATTCGTAAT-3'
3'-CGAGCTTAAC TTTAAGATTTGCGATTTCTCCTTCTCC TGT tCCtC TTAAGCATTA-5'
(+)strand
18 nt
37 nt
(-)strand
5'-GCTCGAATTG AAATTCTAAACGCTAAAGAGGAAGAGG ACA tGGtG AATTCGTAAT-3'
3'-CGAGCTTAAC TTTAAGATTTGCGATTTCTCCTTCTCC TGT tGGtG TTAAGCATTA-5'
(+)strand

Figure 11.

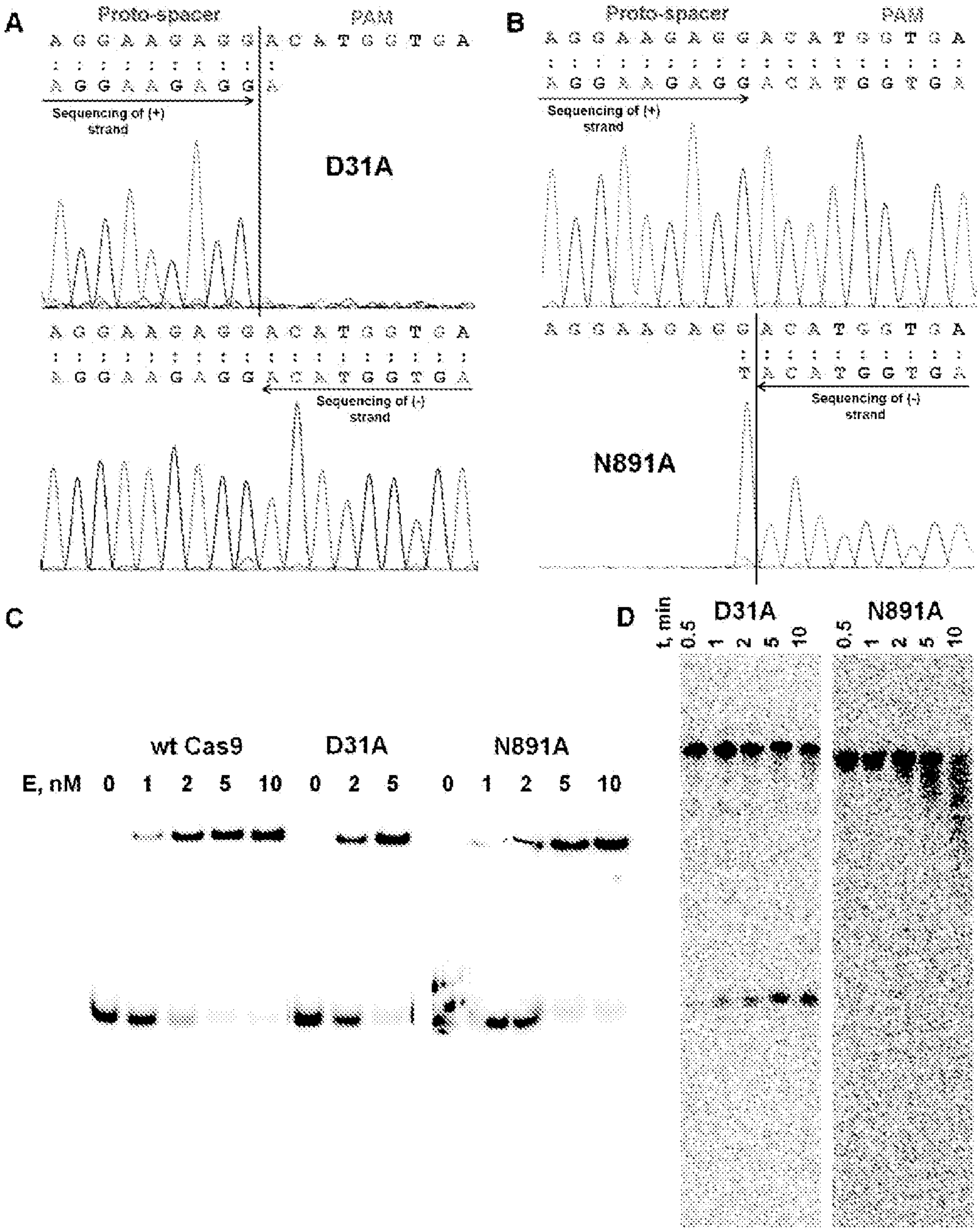

A
Proto-spacer
PAM
A G G A A G A G G A C A T G G T G A
A G G A A G A G G A
Sequencing of (+) strand
D31A
A G G A A G A G G A C A T G G T G A
A G G A A G A G G A C A T G G T G A
Sequencing of (-) strand
B
Proto-spacer
PAM
A G G A A G A G G A C A T G G T G A
A G G A A G A G G A C A T G G T G A
Sequencing of (+) strand
A G G A A G A G G A C A T G G T G A
T A C A T G G T G A
Sequencing of (-) strand
N891A
C
wt Cas9
D31A
N891A
E, nM 0 1 2 5 10 0 2 5 0 1 2 5 10
D
t, min
D31A
0.5 1 2 5 10
N891A
0.5 1 2 5 10

Figure 12.

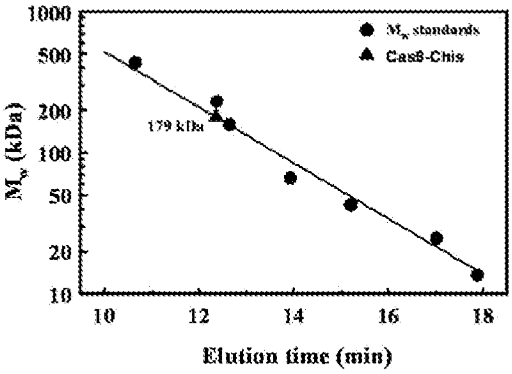

1000
500
200
100
50
20
10
M_w (kDa)
10
12
14
16
18
Elution time (min)
M_w standards
Cas8-Chis
179 kDa

Figure 13.

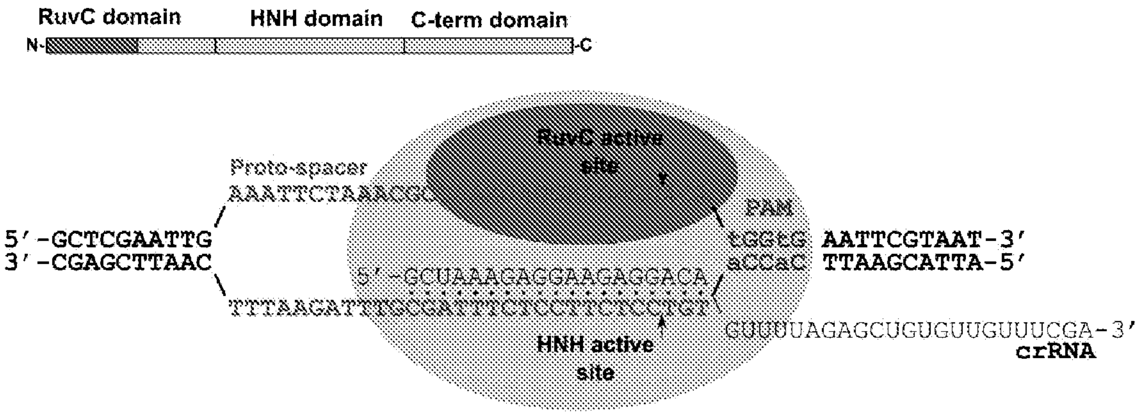

RuvC domain
HNH domain
C-term domain
N-
-C
RuvC active site
Proto-spacer
AAATTCTAAACG
PAM
5′-GCTCGAATTG
3′-CGAGCTTAAC
tGGtG
aCCaC
AATTCGTAAT-3′
TTAAGCATTA-5′
5′-GCUAAAGAGGAAGAGGACA
TTTAAGATTTGCGATTTCTCCTTCTCCTGT
GUUUUAGAGCUGUGUUGUUUCGA-3′
crRNA
HNH active site

Figure 14.

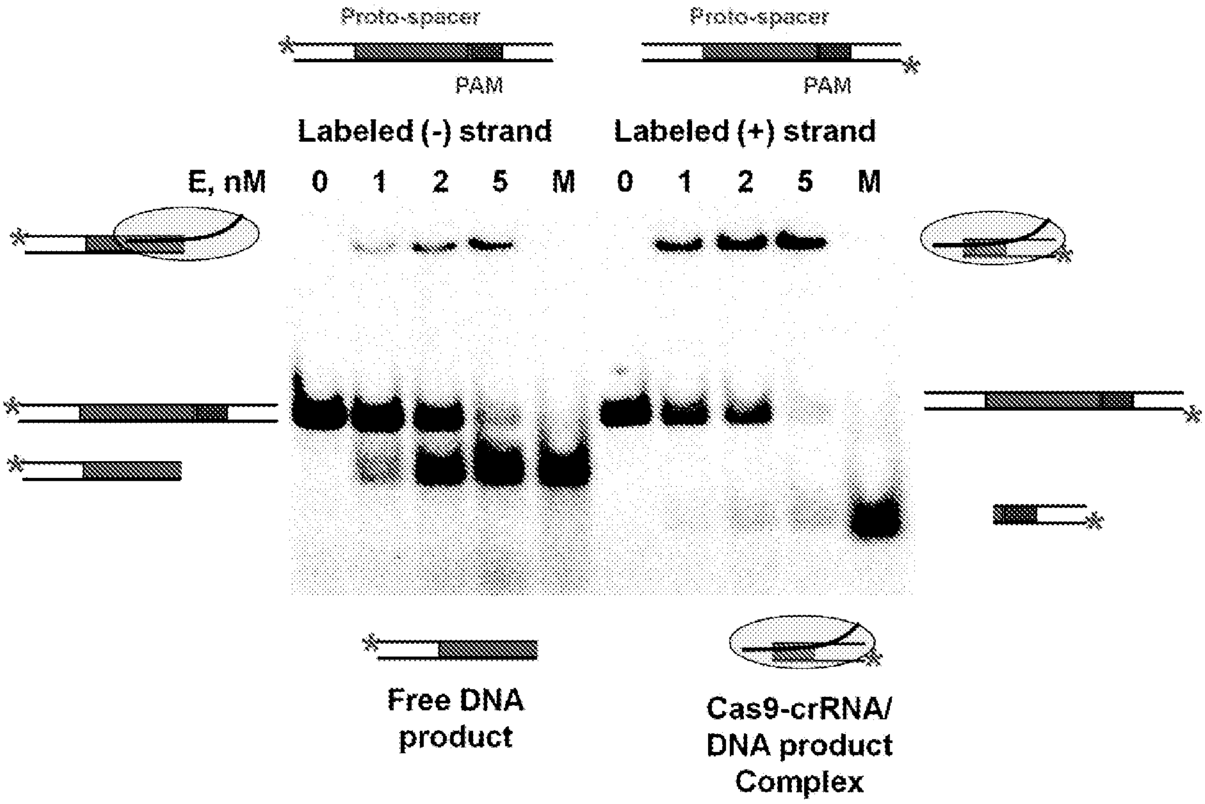

Proto-spacer
PAM
Proto-spacer
PAM
Labeled (-) strand
Labeled (+) strand
E, nM
0
1
2
5
M
0
1
2
5
M
Free DNA product
Cas9-crRNA/ DNA product Complex

Figure 15.

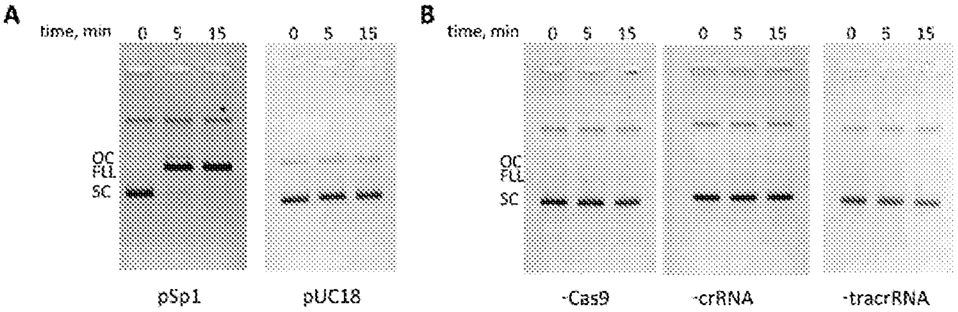

A
time, min
0
5
15
0
5
15
OC
FLL
SC
pSp1
pUC18
B
time, min
0
5
15
0
5
15
0
5
15
OC
FLL
SC
-Cas9
-crRNA
-tracrRNA

Figure 16.

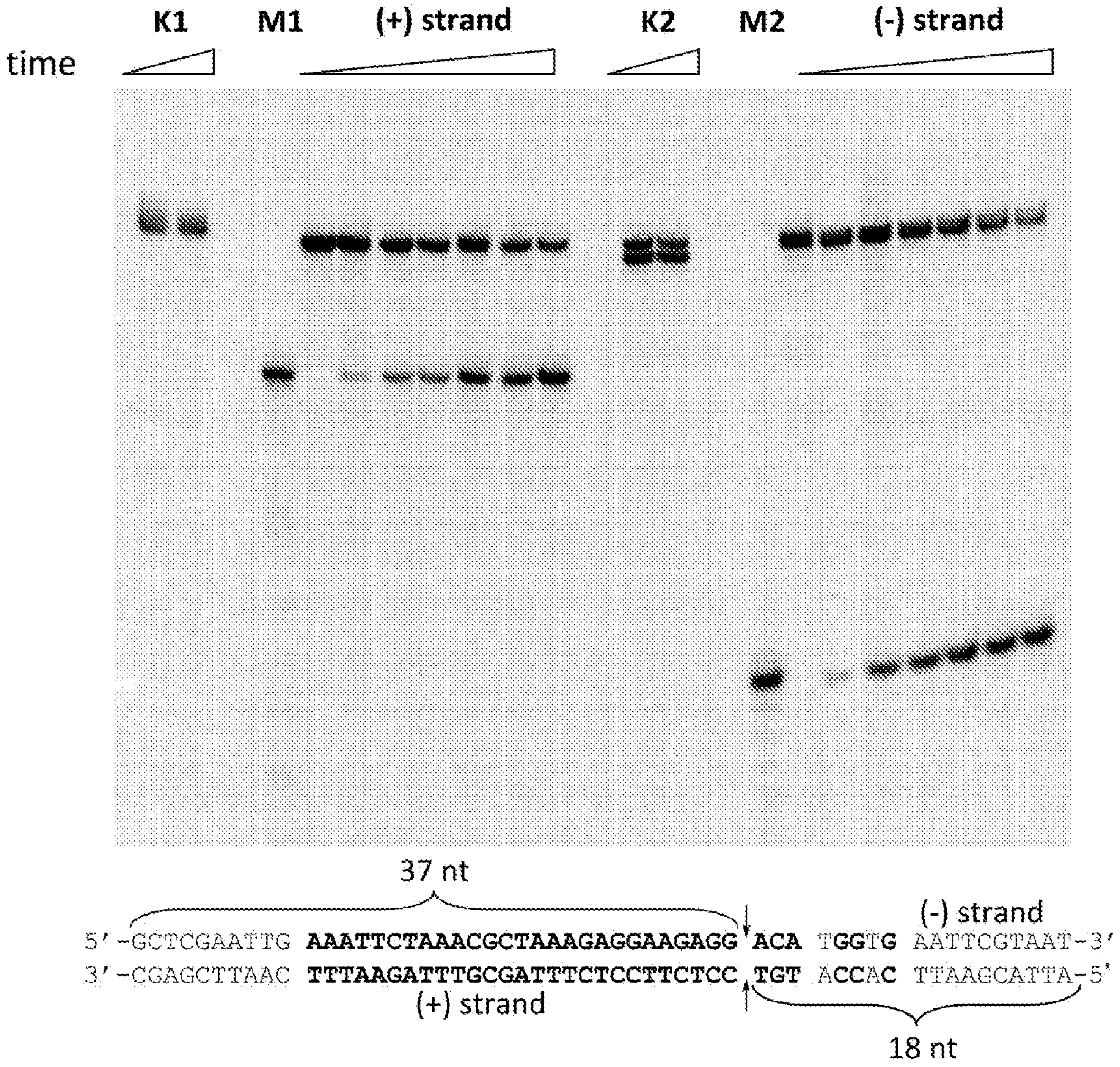

K1
M1
(+) strand
K2
M2
(-) strand
time
37 nt
(-) strand
5'-GCTCGAATTG AAATTCTAAACGCTAAAGAGGAAGAGG ACA TGGTG AATTCGTAAT-3'
3'-CGAGCTTAAC TTTAAGATTTGCGATTTCTCCTTCTCC TGT ACCAC TTAAGCATTA-5'
(+) strand
18 nt

Figure 17.

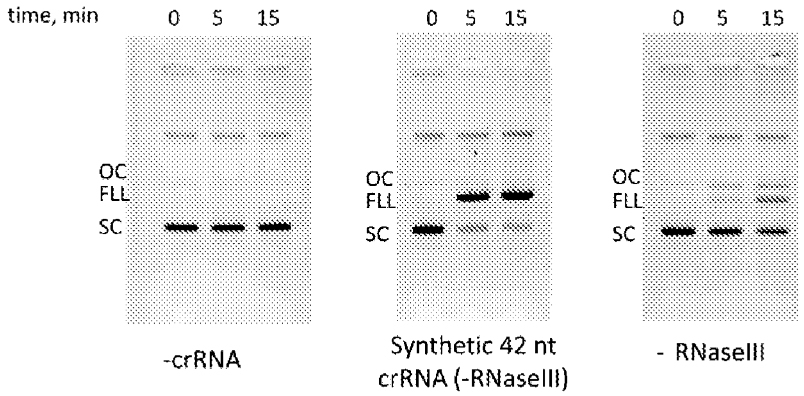

time, min
0
5
15
0
5
15
0
5
15
OC
FLL
SC
OC
FLL
SC
OC
FLL
SC
-crRNA
Synthetic 42 nt crRNA (-RNaseIII)
- RNaseIII

Figure 18.

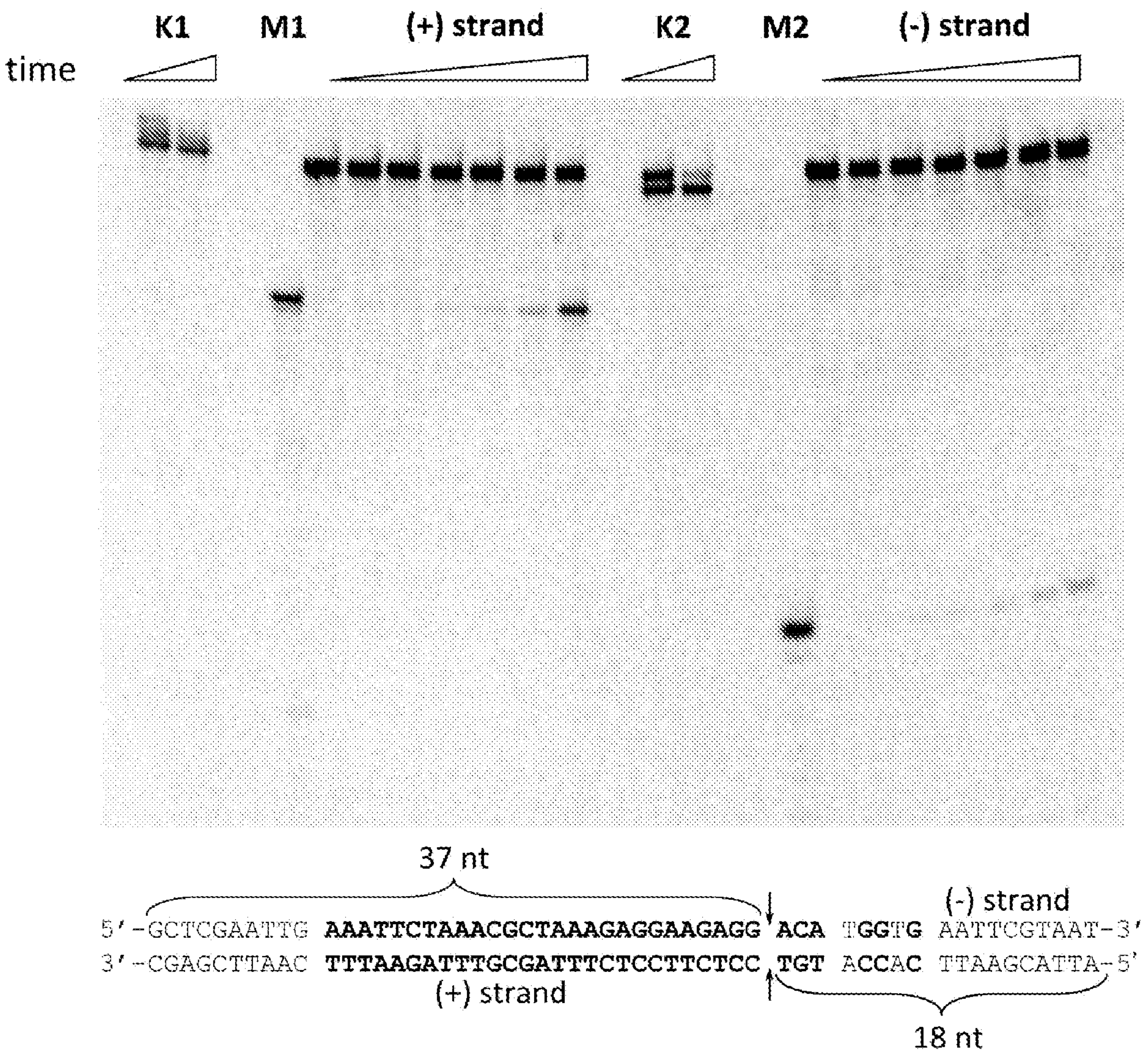

K1
M1
(+) strand
K2
M2
(-) strand
time
37 nt
(-) strand
5'-GCTCGAATTG AAATTCTAAACGCTAAAGAGGAAGAGG ACA TGGTG AATTCGTAAT-3'
3'-CGAGCTTAAC TTTAAGATTTGCGATTTCTCCTTCTCC TGT ACCAC TTAAGCATTA-5'
(+) strand
18 nt

Figure 19.

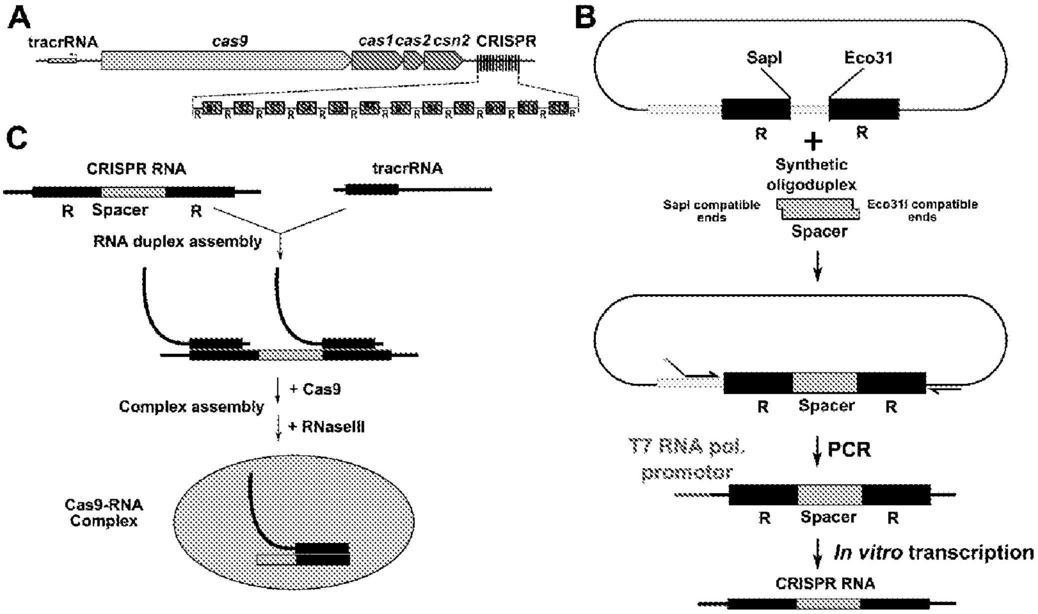

A
tracrRNA
cas9
cas1 cas2 csn2 CRISPR
R
B
SapI
Eco31
R
+
R
Synthetic
oligoduplex
SapI compatible ends
Eco31I compatible ends
Spacer
R
Spacer
R
T7 RNA pol. promotor
PCR
R
Spacer
R
In vitro transcription
CRISPR RNA
C
CRISPR RNA
R
Spacer
R
tracrRNA
RNA duplex assembly
+ Cas9
Complex assembly
+ RNaseIII
Cas9-RNA Complex

Figure 20

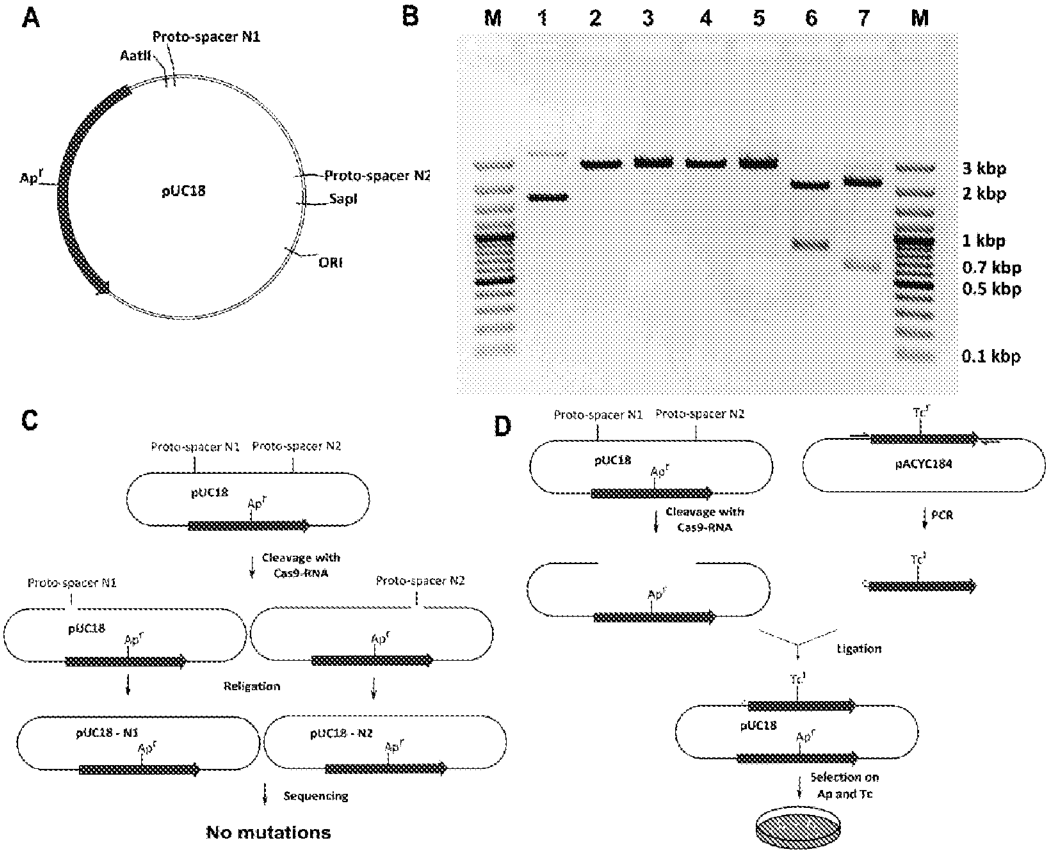

A
Proto-spacer N1
AatII
Ap^r
pUC18
Proto-spacer N2
SapI
ORI
B
M 1 2 3 4 5 6 7 M
3 kbp
2 kbp
1 kbp
0.7 kbp
0.5 kbp
0.1 kbp
C
Proto-spacer N1
Proto-spacer N2
pUC18
Ap^r
Cleavage with Cas9-RNA
Religation
pUC18 - N1
pUC18 - N2
Sequencing
No mutations
D
pACYC184
Tc^r
PCR
ligation
Selection on Ap and Tc

Figure 21

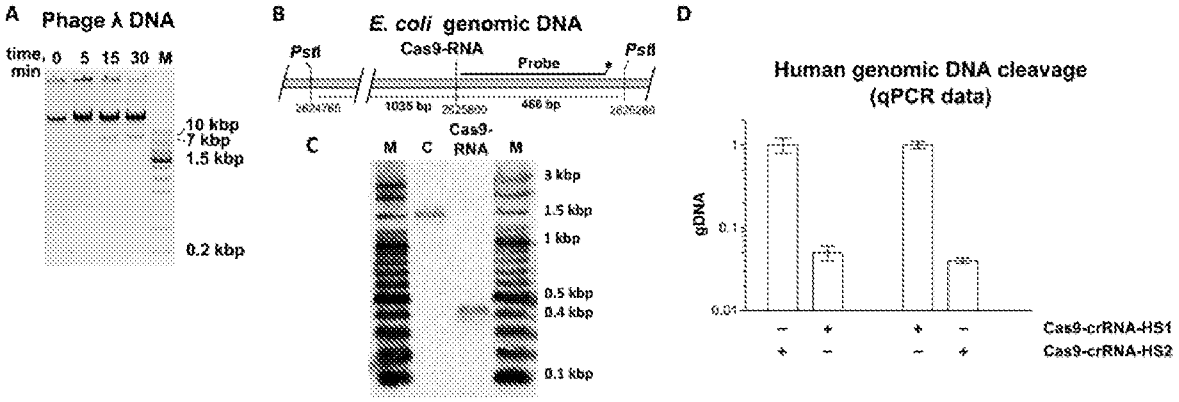


Figure 22

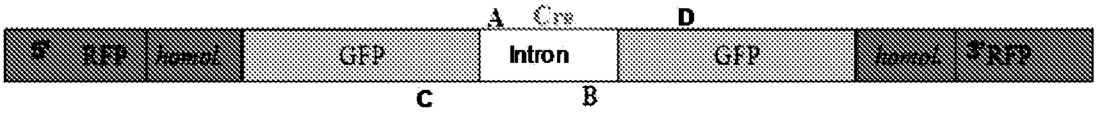


Figure 23

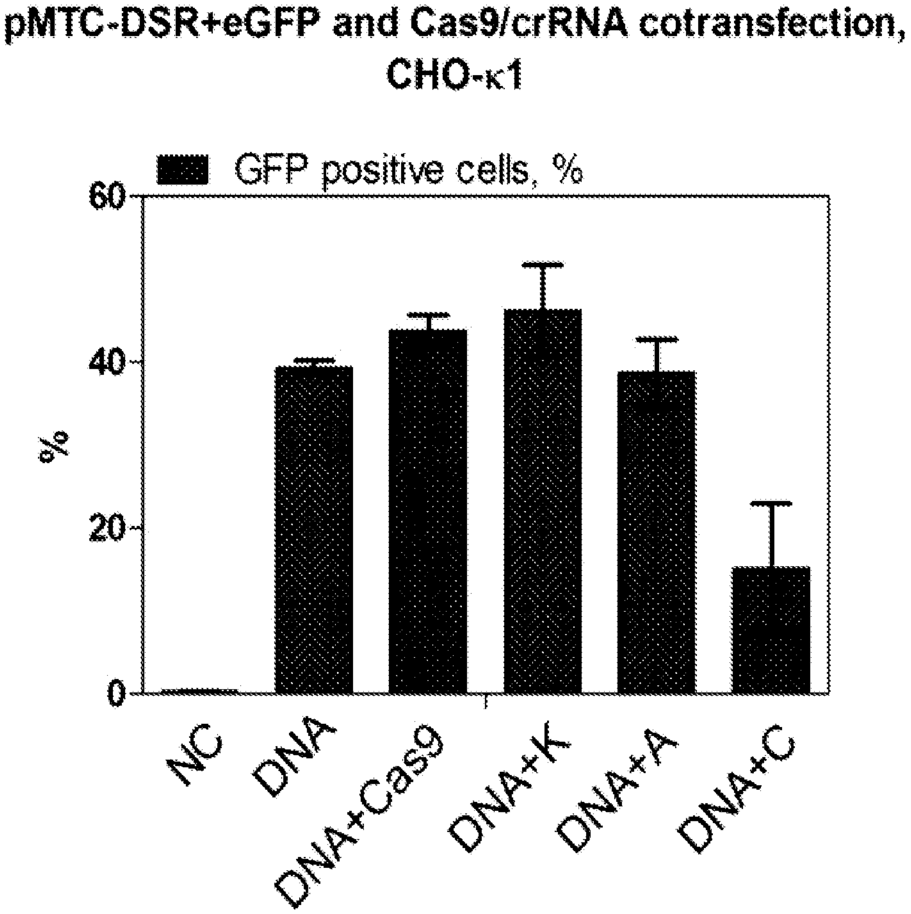
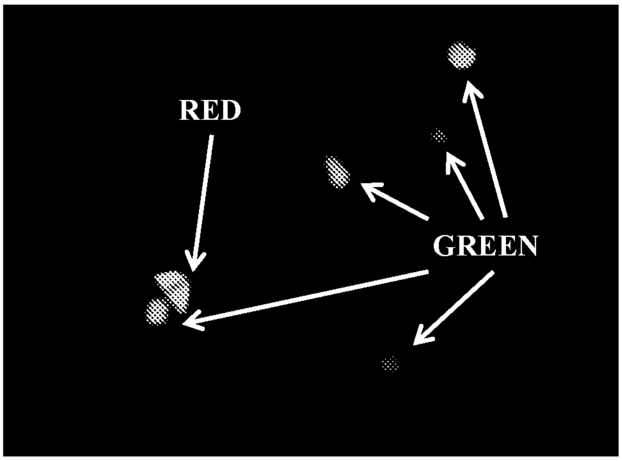

pMTC-DSR+eGFP and Cas9/crRNA cotransfection,
CHO-κ1
GFP positive cells, %
60
40
20
0
%
NC
DNA
DNA+Cas9
DNA+K
DNA+A
DNA+C

Figure 24

A



B

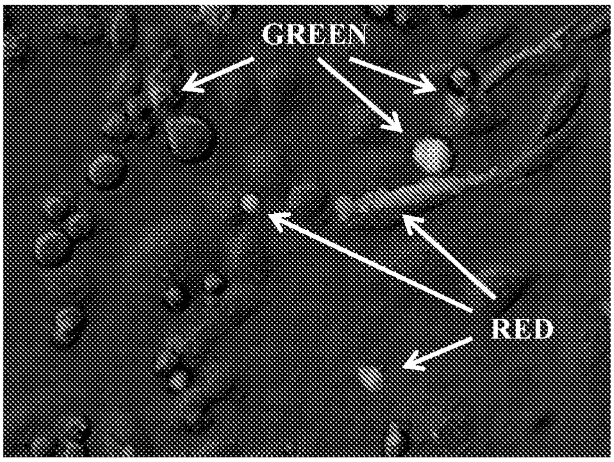


Figure 25

Figure 26

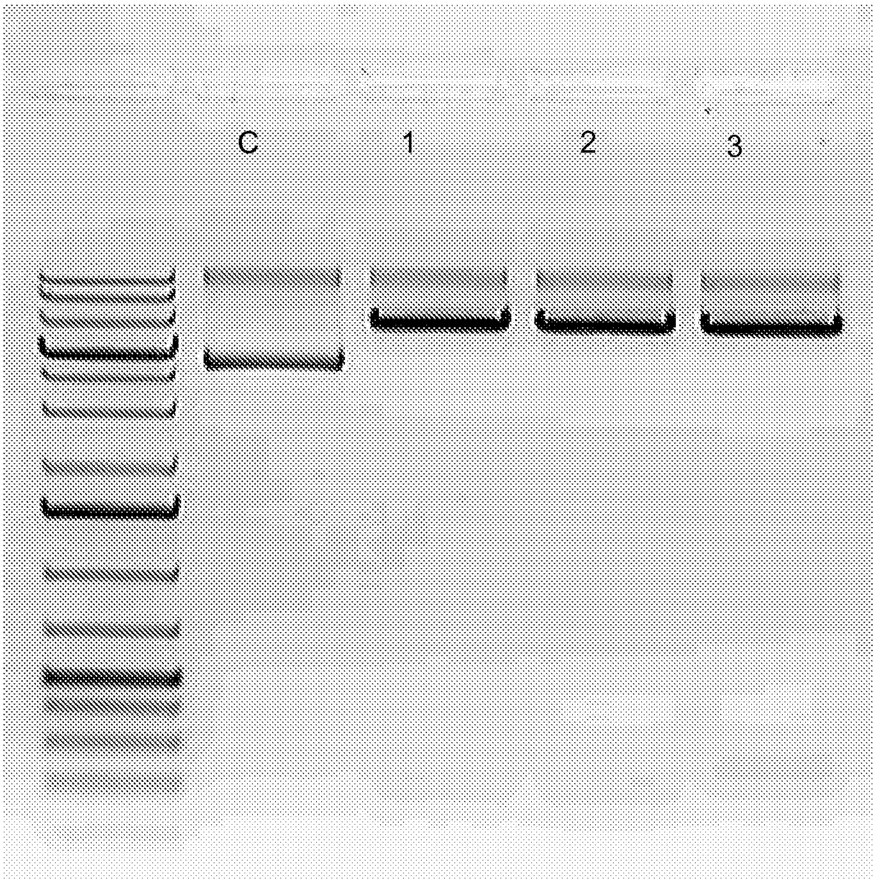


Figure 27

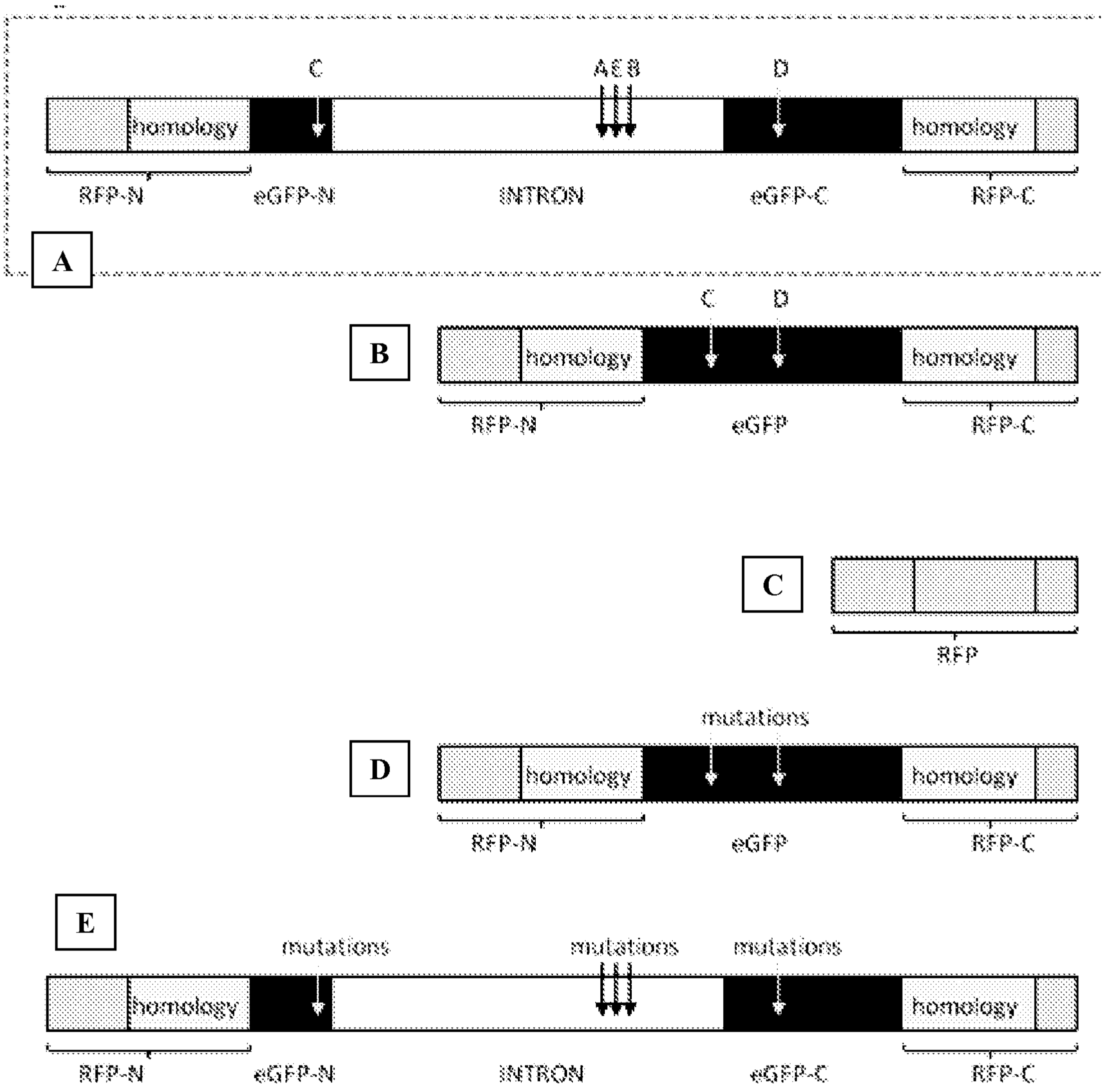

C
A E B
D
homology
homology
RFP-N
eGFP-N
INTRON
eGFP-C
RFP-C
A
C
D
B
homology
homology
RFP-N
eGFP
RFP-C
C
RFP
mutations
D
homology
homology
RFP-N
eGFP
RFP-C
E
mutations
mutations
mutations
homology
homology
RFP-N
eGFP-N
INTRON
eGFP-C
RFP-C

Figure 28

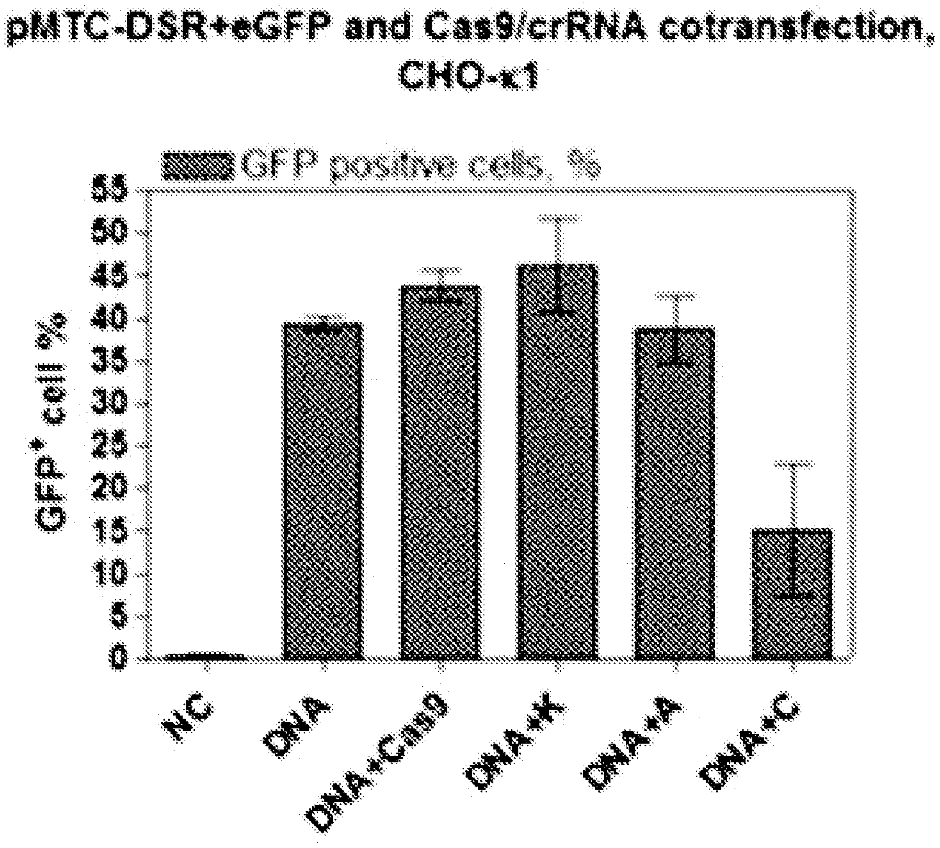


Figure 29

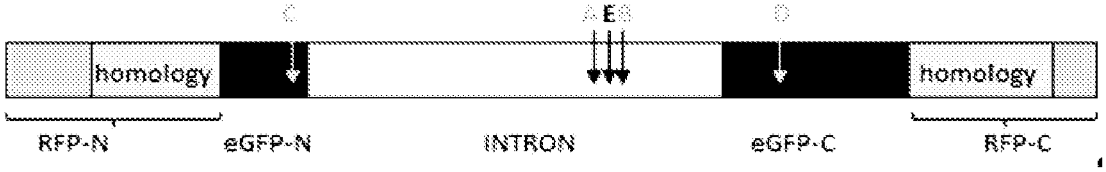
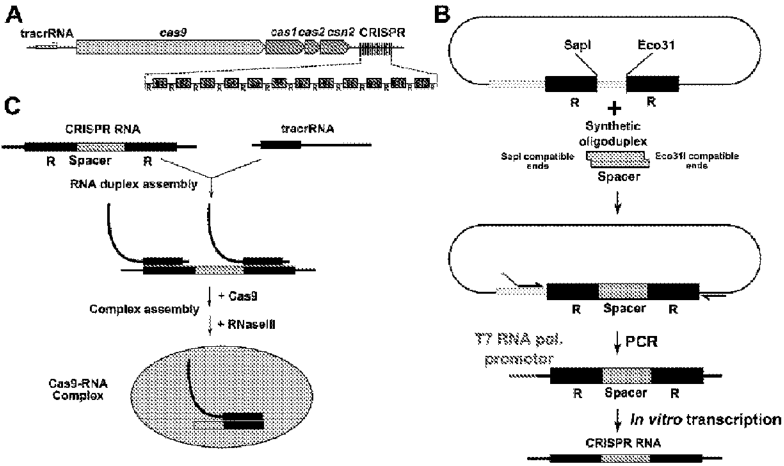


Figure 30


A
tracrRNA
cas9
cas1
cas2
csn2
CRISPR
R
B
SapI
Eco31
R
R
+
Synthetic oligoduplex
SapI compatible ends
Eco31I compatible ends
Spacer
R
Spacer
R
T7 RNA pol. promoter
PCR
R
Spacer
R
In vitro transcription
CRISPR RNA
C
CRISPR RNA
tracrRNA
R
Spacer
R
RNA duplex assembly
+ Cas9
Complex assembly
+ RNaseIII
Cas9-RNA Complex