



US 20150240261A1

(19) **United States**(12) **Patent Application Publication**
SIKSNYS et al.(10) **Pub. No.: US 2015/0240261 A1**(43) **Pub. Date: Aug. 27, 2015**(54) **RNA-DIRECTED DNA CLEAVAGE BY THE
CAS9-CRRNA COMPLEX***C12N 15/113* (2006.01)*C12N 15/10* (2006.01)(71) Applicant: **VILNIUS UNIVERSITY**, Vilnius (LT)(52) **U.S. Cl.**(72) Inventors: **Virginijus SIKSNYS**, Vilnius (LT);
Giedrius GASIŪNAS, Birzu r. (LT);
Tautvydas KARVELIS, Vilniaus r. (LT)CPC *C12N 15/902* (2013.01); *C12N 15/102*
(2013.01); *C12N 9/22* (2013.01); *C12N 15/113*
(2013.01); *C12N 2310/12* (2013.01); *C12N*
2800/80 (2013.01)(21) Appl. No.: **14/683,443**

(57)

ABSTRACT(22) Filed: **Apr. 10, 2015****Related U.S. Application Data**(63) Continuation of application No. 14/385,241, filed on
Sep. 15, 2014, filed as application No. PCT/LT2013/
000006 on Mar. 15, 2013.(60) Provisional application No. 61/625,420, filed on Apr.
17, 2012, provisional application No. 61/613,373,
filed on Mar. 20, 2012.**Publication Classification**(51) **Int. Cl.***C12N 15/90* (2006.01)*C12N 9/22* (2006.01)

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 using an RNA-guided DNA endonuclease comprising at least one RNA sequence 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 or using a cassette containing a single repeat-spacer-repeat unit.

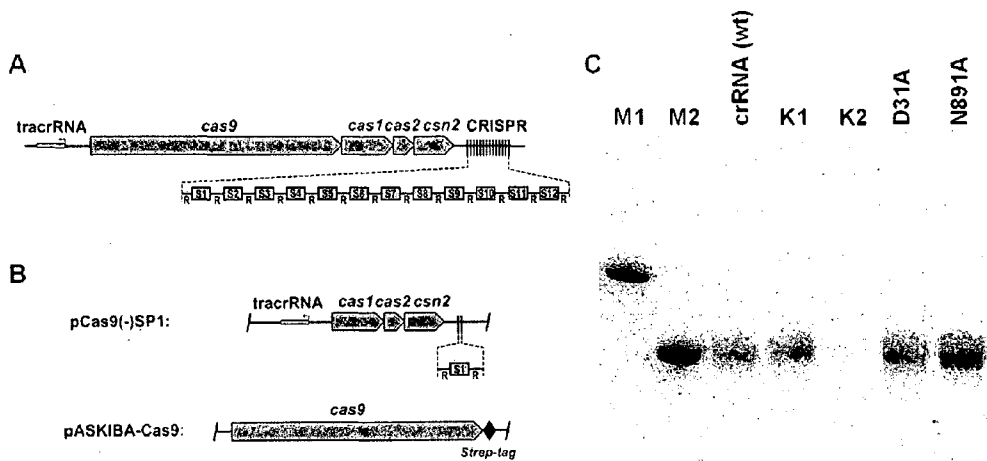


Figure 1

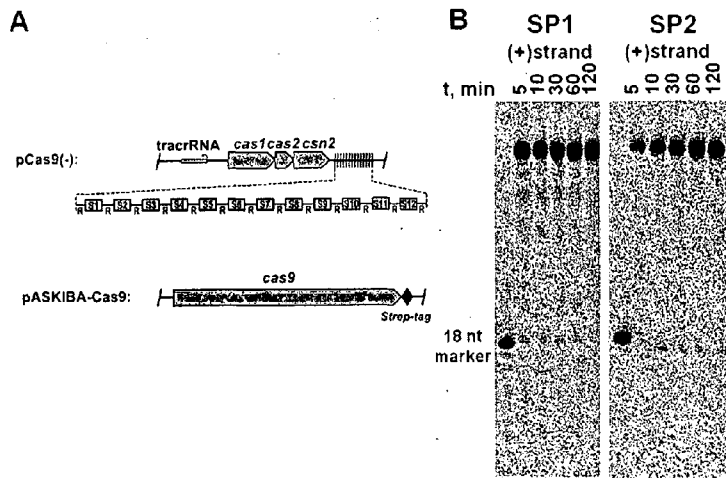


Figure 2

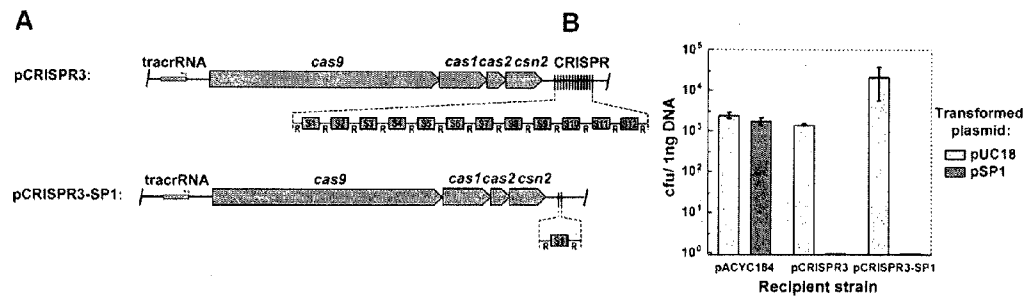


Figure 3

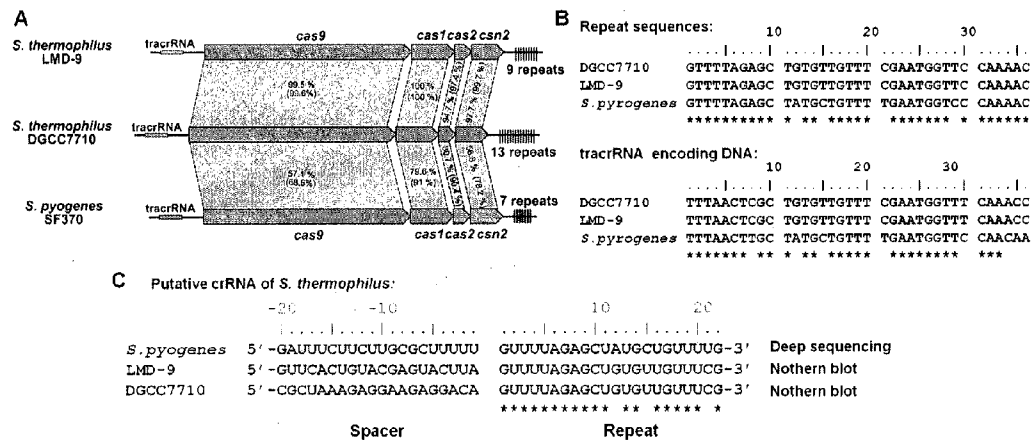


Figure 4

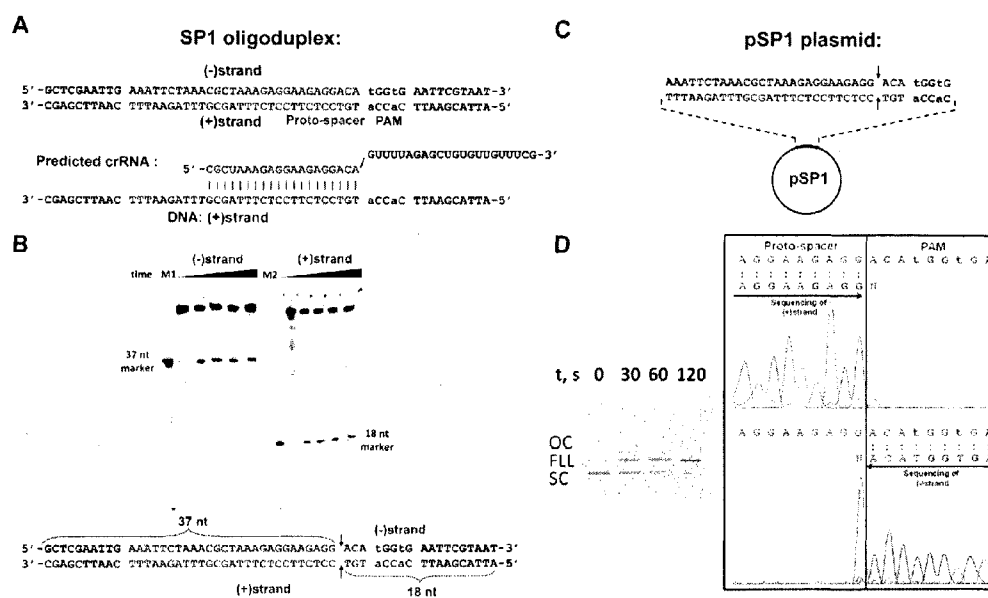


Figure 5

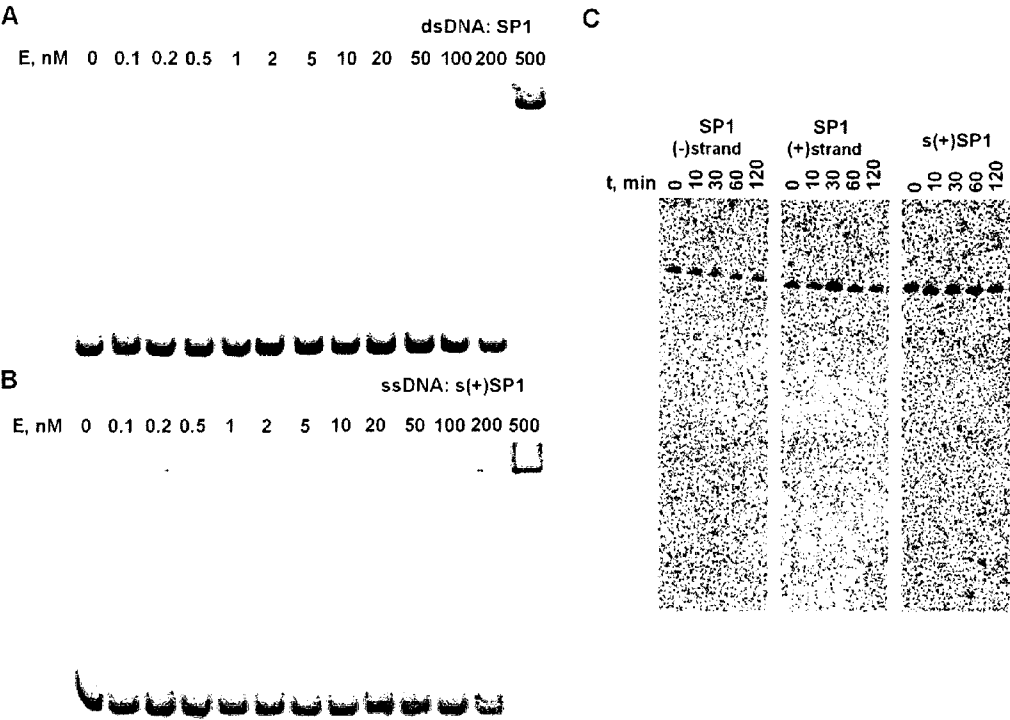


Figure 6

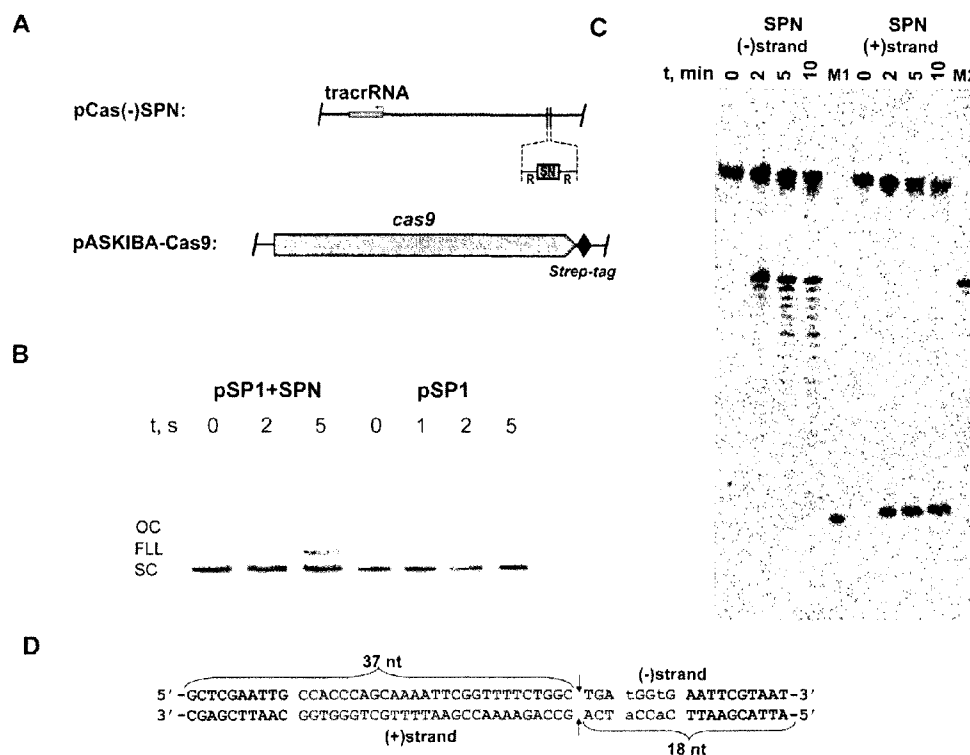


Figure 7

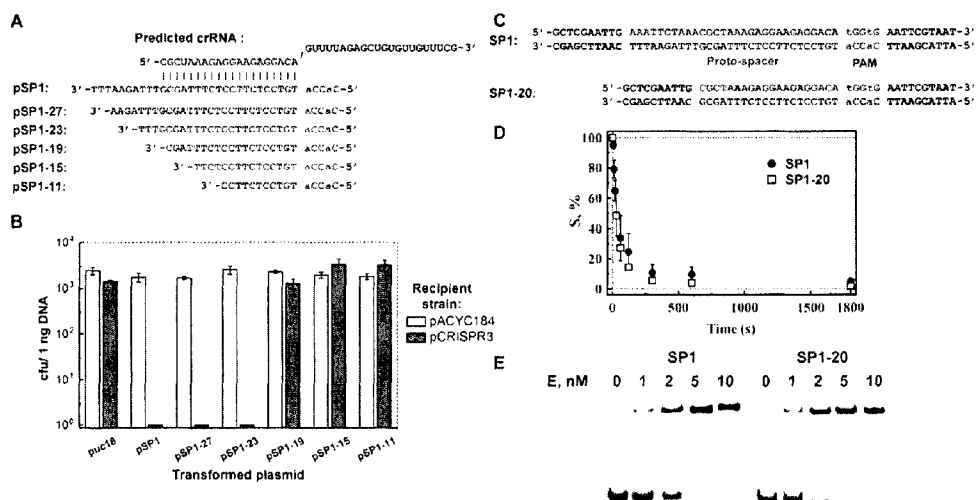


Figure 8

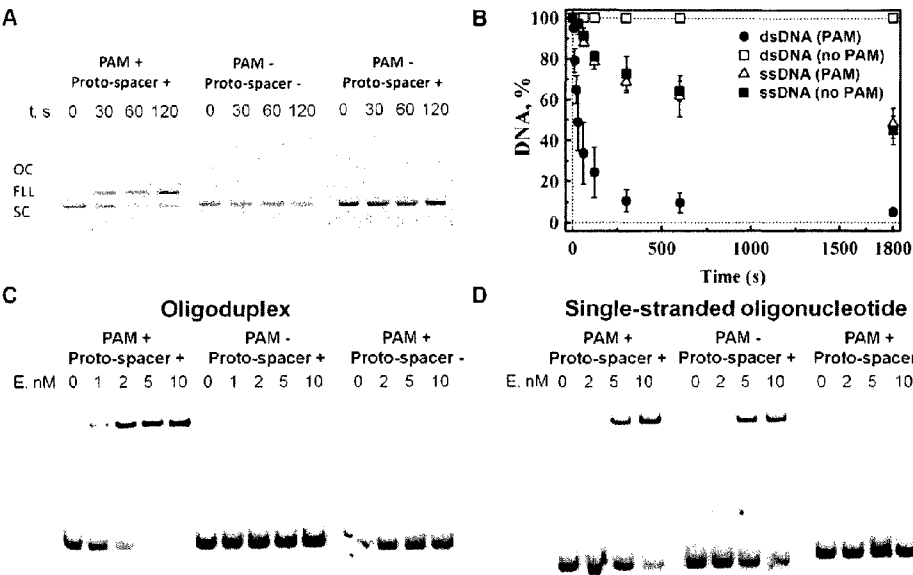


Figure 9

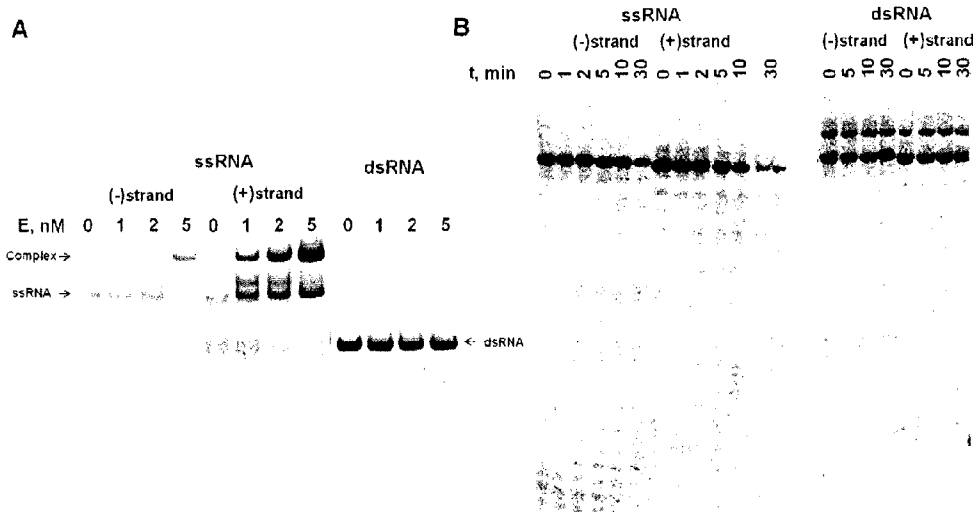


Figure 10

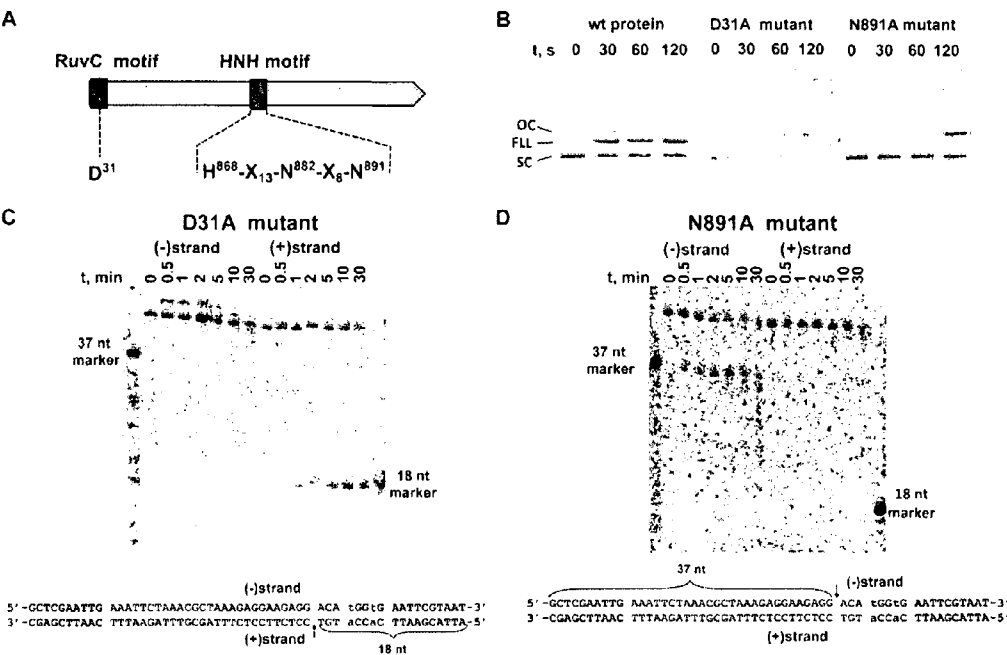
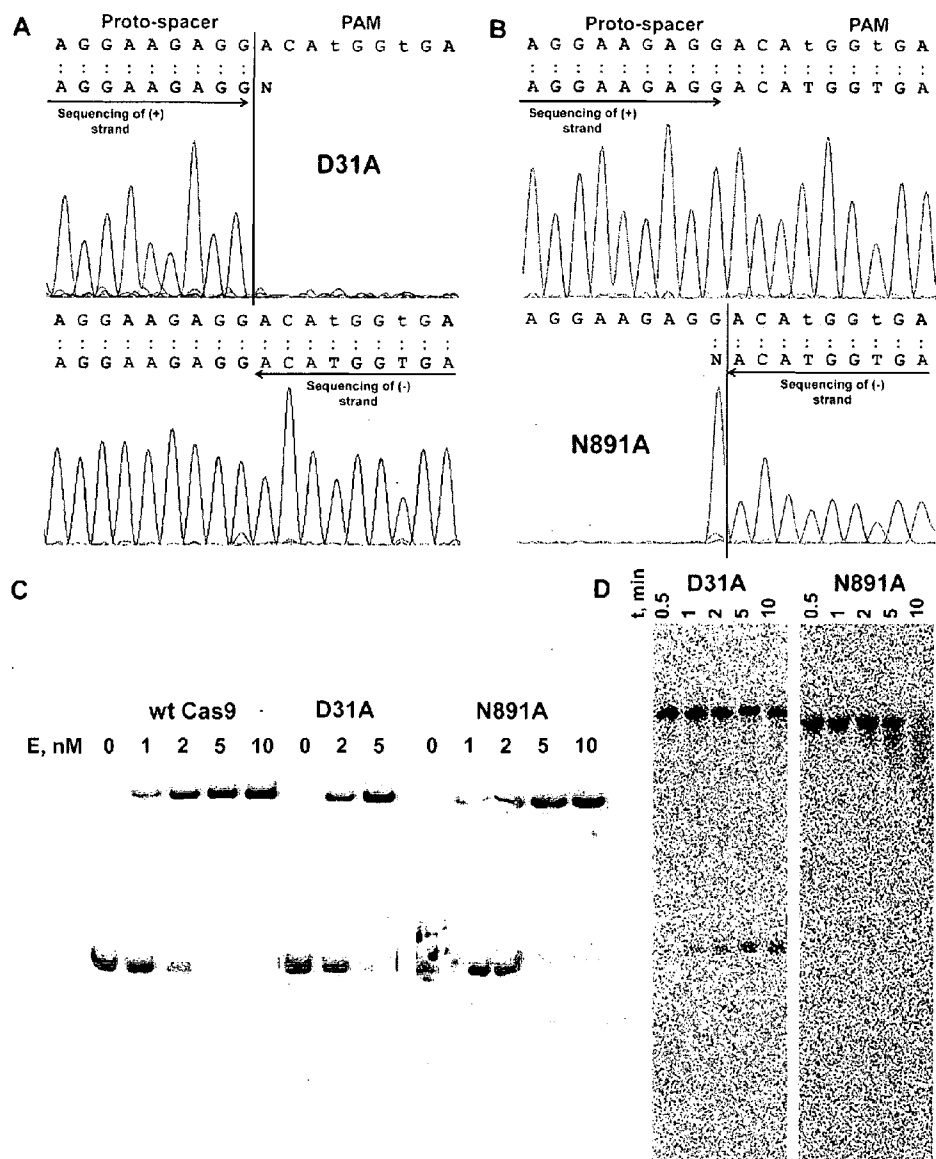


Figure 11



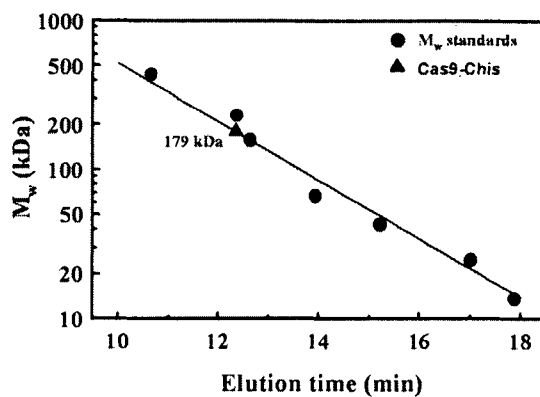


Figure 13

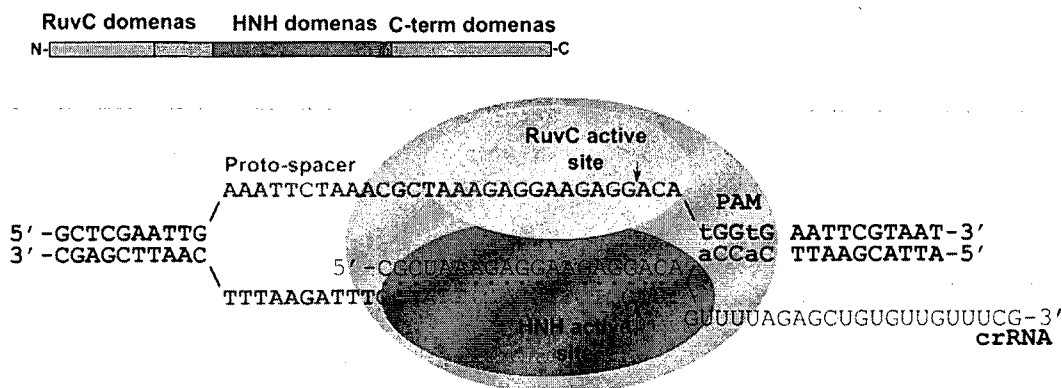


Figure 14

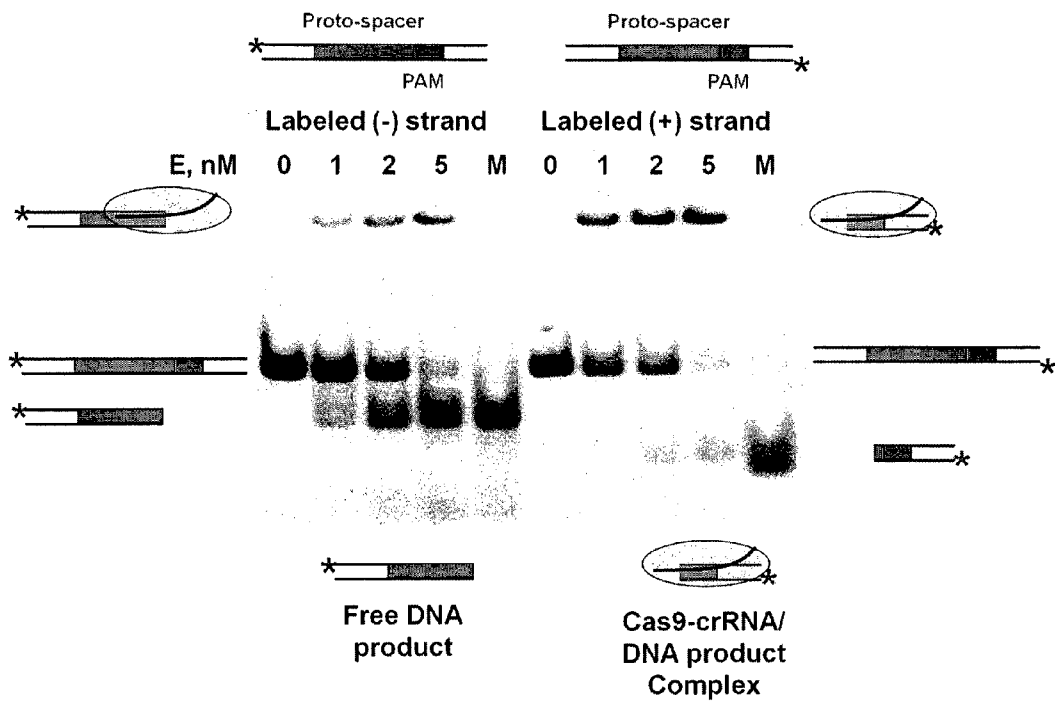


Figure 15

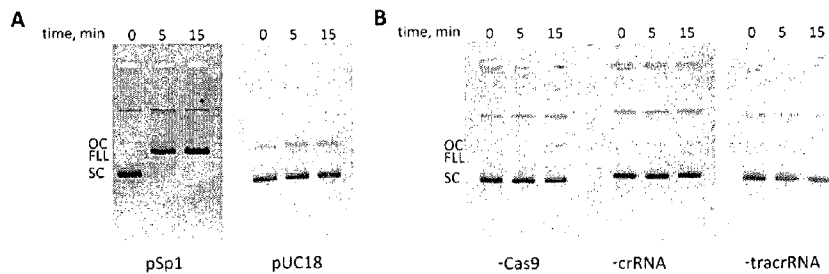


Figure 16

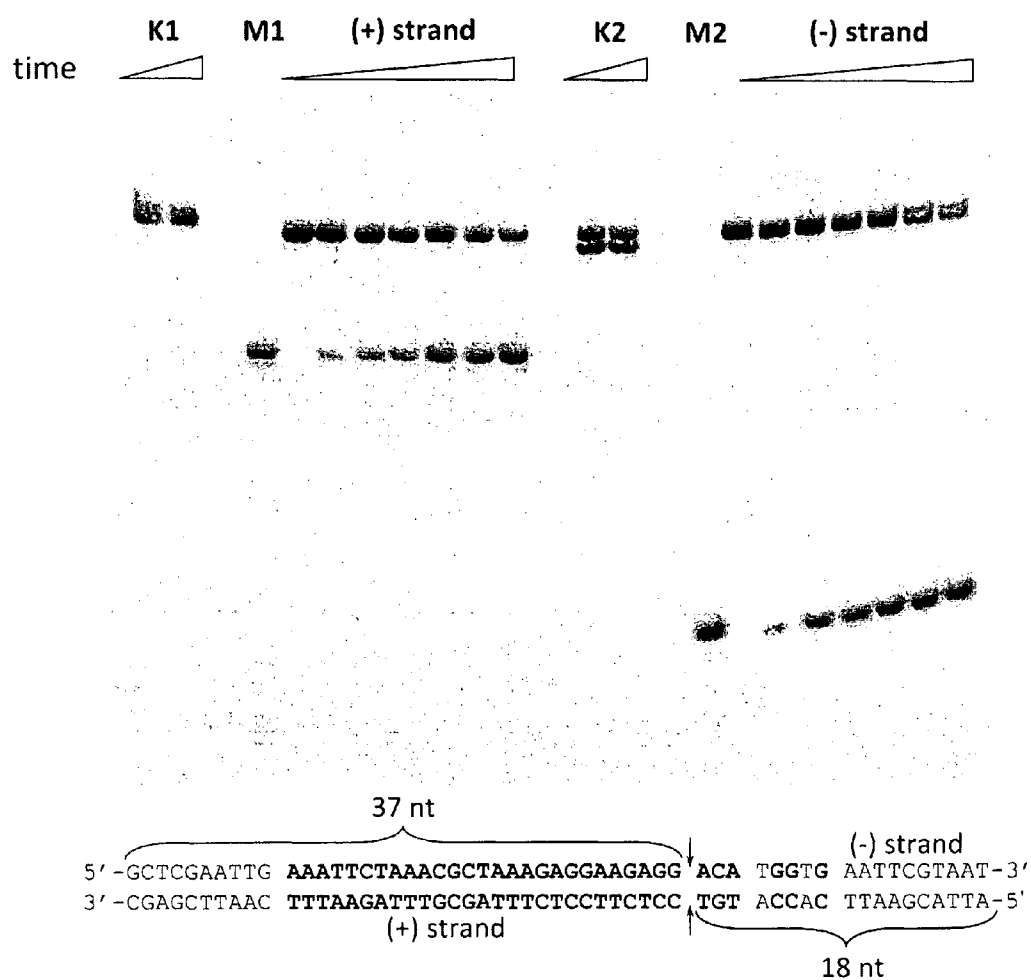


Figure 17

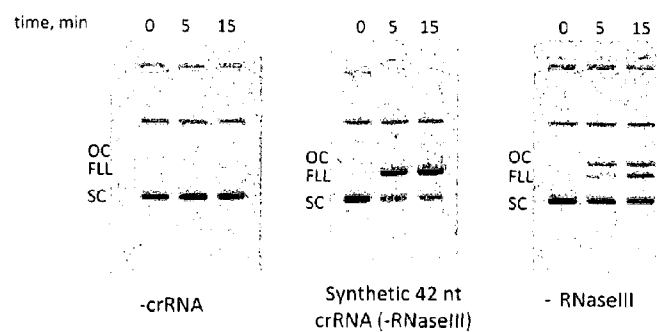


Figure 18

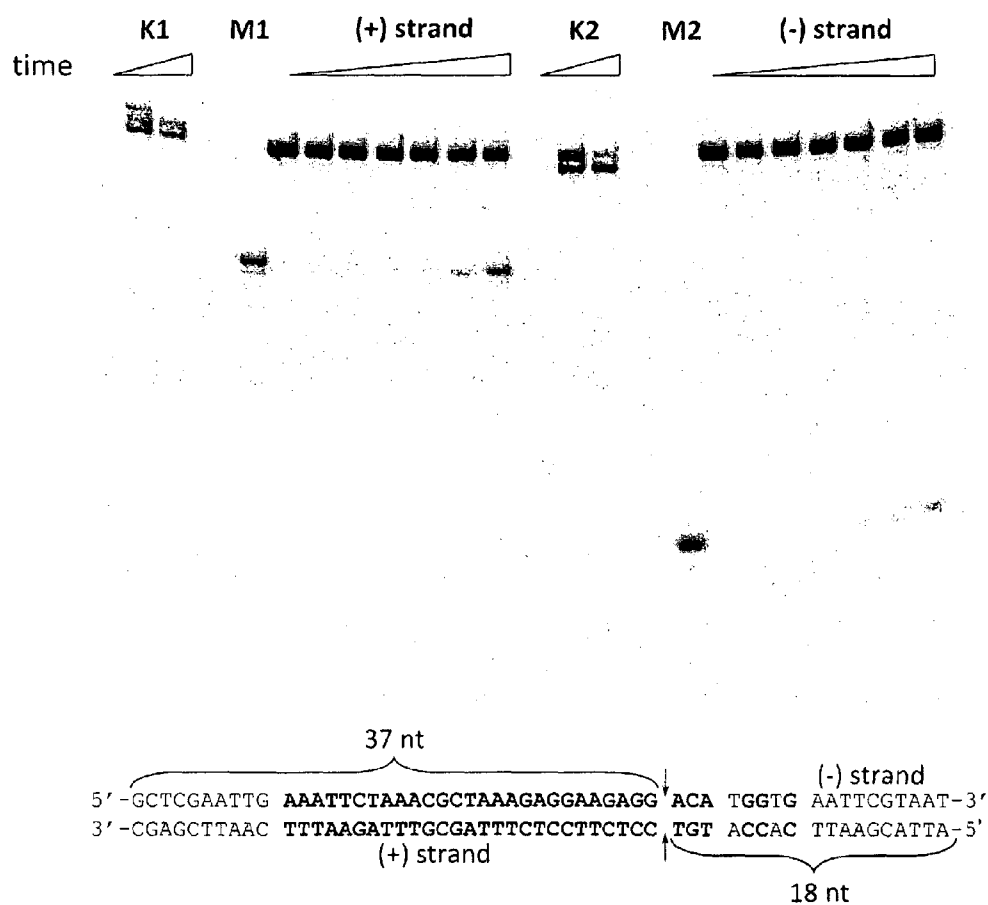


Figure 19

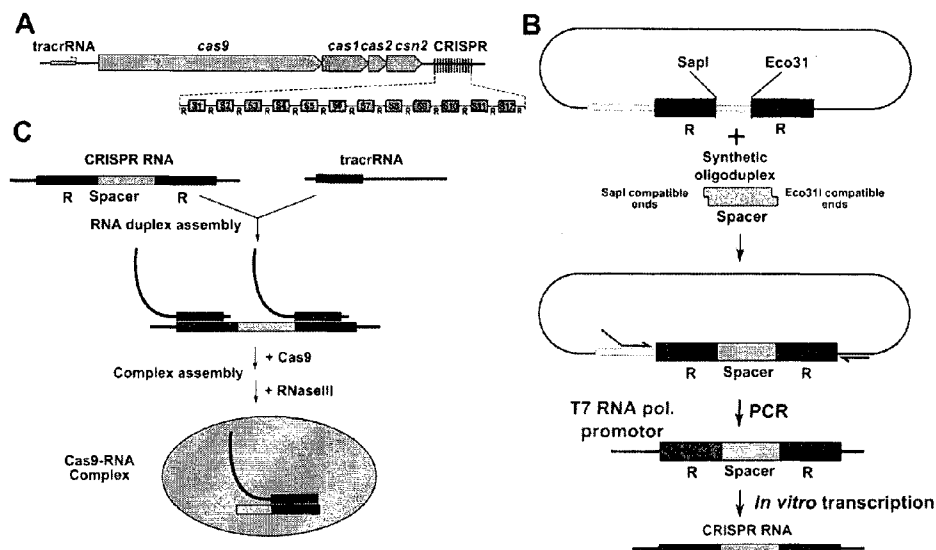


Figure 20

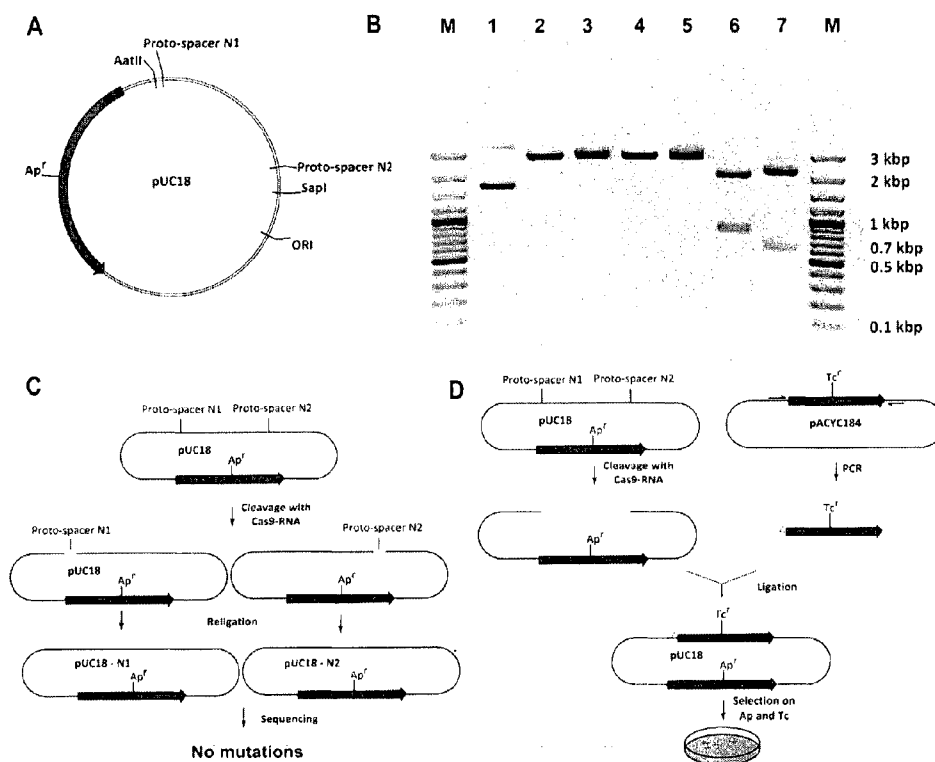


Figure 21

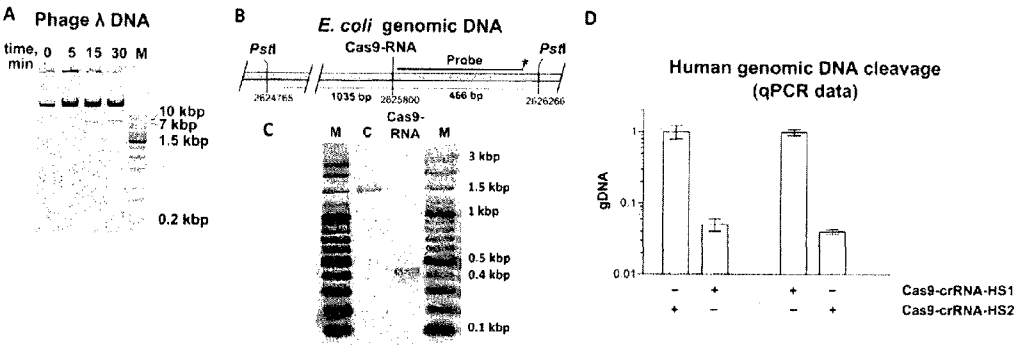


Figure 22

RNA-DIRECTED DNA CLEAVAGE BY THE CAS9-CRRNA COMPLEX

[0001] This application claims priority to co-pending U.S. application Ser. Nos. 61/613,373 filed Mar. 20, 2012, and 61/625,420 filed Apr. 17, 2012, each of which is expressly incorporated by reference herein in its entirety.

BACKGROUND

[0002] 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 11:181-90; 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.

[0003] 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).

[0004] 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 (FIG. 1A). Cas9 (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 (Sapranaukas 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; Sapranaukas 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 (Sapranaukas 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 (Sapranaukas 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 (Sapranaukas et al., 2011. Nucleic Acids Res 39:9275-82).

[0005] 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

[0006] 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 sequence 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'-NG-GNG-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 poly-nucleotide, 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.

[0007] 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.

[0008] Polypeptide-polyribonucleotides complex might be isolated from a genetically modified microbe (for example *Escherichia coli* or *Streptococcus 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.

[0009] 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'-NNNNNNNNNNNNNNNNNNNNNGUUUA-GAGCUGUGUUGUUUCG-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'-NNNNNNNNNNNNNNNNNNNNNGUUUA-GAGCUGUGUUGUUUCG-3' (SEQ ID NO: 15) and flanking regions and RNase III polypeptide, cleaving double stranded RNA polynucleotide. The examples for polyribonucleotide containing sequence 5'-NNNNNNNNNNNNNNNNNNNNNGUUUA-GAGCUGUGUUGUUUCG-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*.

[0010] 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

[0011] FIG. 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 Strept-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 FIG. 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.

[0012] FIG. 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 *Sth*CRISPR3/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.

[0013] FIG. 3 shows immunity against plasmid transformation in *E. coli* cells provided by the *Sth*CRISPR3/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 $\pm$ SD).

[0014] FIG. 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. (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. (C) discloses SEQ ID NOS 54-56, respectively, in order of appearance.

[0015] FIG. 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 complementary to the 5'-terminal fragment of crRNA (red letters) is named (+) strand, an opposite DNA strand is named (−) strand. (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. (B) discloses SEQ ID NO: 31. (C) Schematic representation of pSP1 plasmid (Sapranaukas et al., 2011. Nucleic Acids Res 39:9275-82) used in the plasmid cleavage assay. (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. (D) discloses SEQ ID NOS 58, 59, 58, and 60, respectively, in order of appearance.

[0016] FIG. 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.

[0017] FIG. 7 shows reprogramming 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 FIG. 2), 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. (D) discloses SEQ ID NO: 39.

[0018] FIG. 8 shows impact of spacer length on CRISPR-encoded immunity. (A) Schematic representation of shortened versions of proto-spacers inserted in the transformed plasmids. (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 $\pm$ SD). (C). Schematic representation of oligoduplexes used in the in vitro cleavage and binding experiments. (C) discloses SEQ ID NOS 31 and 38, respectively, in order of appearance. (D) Time courses of the 27 by oligoduplex (full length protospacer SP1, filled circles) and the 20 by 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.

[0019] FIG. 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 (Sapranaukas 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 squares) 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'.

[0020] FIG. 10 shows RNA binding and cleavage analysis of Cas9-crRNA complex. (A)

[0021] 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' taatagactactataGggtaccgagctcgaattg 3' (SEQ ID NO: 17)/5' GGGAAA-CAGCTATGACCATGATTACGAATTC-3' (SEQ ID NO: 18) and 5' ggggtaccgagctcgaattgaaattcTAAACG 3' (SEQ ID NO: 19)/5' taatagactactataGggAAACAGC-TATGACCATGATTACG 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.

[0022] FIG. 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 (Sapranaukas 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 FIG. 2 and 'Materials and Methods'. D31 mutant cleaves only (+)strand of SP1 oligoduplex. (C) discloses SEQ ID NOS 31. (D) Strand preference of N891A mutant. N891 mutant cleaves only (−)strand of SP1 oligoduplex. Cleavage positions are designated by arrows. (D) discloses SEQ ID NOS 31.

[0023] FIG. 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). (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). (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.

[0024] FIG. 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).

[0025] FIG. 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 3 nt away from the 5'-NGGNG-3' PAM generating blunt end DNA. FIG. 14 discloses SEQ ID NOS. 31 and 7, respectively, in order of appearance.

[0026] FIG. 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.

[0027] FIG. 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 5'-NGGNG-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.

[0028] FIG. 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, 3 nt upstream of the PAM (5'-NGGNG-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. FIG. 17 discloses SEQ ID NO: 31.

[0029] FIG. 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).

[0030] FIG. 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 from the 5'-end, 3 nt upstream of the PAM (5'-NGGNG-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. FIG. 19 discloses SEQ ID NO: 31.

[0031] FIG. 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.

[0032] FIG. 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.

[0033] FIG. 22 shows genomic DNA cleavage with in vitro assembled Cas9-RNA complex. (A) Agarose gel analysis of linear A DNA cleavage products. Phage A 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 by fragment should be detected. Otherwise the probe will hybridize with 1499 by 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.

[0034] The following non-limiting examples further describe the methods, compositions, uses, and embodiments.

DETAILED DESCRIPTION OF ILLUSTRATIVE EMBODIMENTS

EXAMPLE 1

[0035] 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 programmable Cas9-crRNA complexes as universal RNA-guided endonucleases.

Materials and Methods

[0036] 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'-ACGTCTCAAATGTTGTT-TAATAAGTGTATAATAATTC-3' (SEQ ID NO: 21) and 5'-ACGTCTCCGCGCTACCTCTCCTAGTTT-3' (SEQ ID NO: 22) was cloned into the pASK-IBA3 expression vector via Esp31 sites. To generate a pBAD-Cas9 plasmid which was used for the expression of the C-terminal 6× His-tagged Cas9 protein variant ("6× His" 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 (FIG. 1B) and pCRISPR3-SP1 (FIG. 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 (Sapranaukas et al., 2011. Nucleic Acids Res 39:9275-82), respectively.

[0037] Expression and purification of Cas9 protein and Cas9-crRNA complex. 6xHis-tagged ("6× His" 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 (FIG. 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.

[0038] 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 1-ethyl-3-(3-dimethylaminopropyl) carbo-

diimide (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 ³²P-5'-labelled 42 nt anti-crRNA DNA oligonucleotide containing 20 nt of spacer1 and 22 nt of the repeat sequence (5'-TCGAAACAACACAGCTCTAAACTGTCCTCTTCCTCTTTAGC-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'-CGCUAAAGAGGAAGAGGACAGUUUUA-GAGCUGUGUUGUUUCG-3' (SEQ ID NO: 7)) and 84 nt DNA oligonucleotide.

[0039] 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 radio-labelled 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.

[0040] 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.

[0041] Reactions with plasmid substrates. Reactions on pUC18 plasmid and its derivatives (Sapranaukas 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.

[0042] 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'-ccgcattcagcgcattcggc-3' (SEQ ID NO: 29) (sequencing of (+)strand) and 5'-gcgaggaagcgaagagcgccc-3' (SEQ ID NO: 30) (sequencing of (-)strand).

[0043] 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'-methylenebisacry-

lamide 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.

[0044] 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.

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 inducer 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

TABLE 1

Oligonucleotide substrates. Proto-spacer sequence is underlined, PAM is on bold.		
Oligonucleotide	Sequence	Specification
SP1 (SEQ ID NO: 31)	5'-GCTCGAATTGAAATTCATAACGCTAAAGAGGAAGAGGACATGGTGAATTCGTAAT-3' 3'-CGAGCTTAAC <u>TTAAGATTGCGATTCTCCTTCTCCTGT</u> ACCACTTAAGCATT-5'	55 bp oligoduplex substrate containing proto-spacer1 and PAM
SP1-pA (SEQ ID NO: 32)	5'-GCTCGAATTGAAATTCATAACGCTAAAGAGGAAGAGGACAAATTCGTAAT-3' 3'-CGAGCTTAAC <u>TTAAGATTGCGATTCTCCTTCTCCTGT</u> TTAAGCATT-5'	50 bp oligoduplex substrate containing proto-spacer2
SP2 (SEQ ID NO: 33)	5'-GCTCGAATTGACTGCTGTATTAGCTTGGTTGTTGGTTTGGTGAATTCGTAAT-3' 3'-CGAGCTTAACATGACGACATAATCGAACCAACAACCAACACCACTTAAGCATT-5'	55 bp oligoduplex substrate containing proto-spacer2 and PAM (oligoduplex without proto-spacer1)
s(+) SP1 (SEQ ID NO: 34)	5'-ATTACGAATTCACCATGTCCTCTTCTCTTTAGCGTTTAGAATTCAATTCGAGC-3'	55 nt ssDNA oligonucleotide substrate (+) strand of SP1 oligoduplex
s(+) SP1-p (SEQ ID NO: 35)	5'-ATTACGAATTGTCCTCTTCTCTTTAGCGTTTAGAATTCAATTCGAGC-3'	50 nt ssDNA oligonucleotide substrate (+) strand of SP1-A oligoduplex
s(+) SP2 (SEQ ID NO: 36)	5'-ATTACGAATTCACCAACAACAACAACAAGCTAATACAGCAGTACAATTCGAGC-3'	55 nt ssDNA oligonucleotide substrate (+) strand of SP2 oligoduplex
s(-) SP1 (SEQ ID NO: 37)	5'-GCTCGAATTGAAATTCATAACGCTAAAGAGGAAGAGGACATGGTGAATTCGTAAT-3'	55 nt ssDNA oligonucleotide substrate, ② strand of SP1 oligoduplex
SP1-20 (SEQ ID NO: 38)	5'-GCTCGAATTGCGCTAAAGAGGAAGAGGACATGGTGAATTCGTAAT-3' 3'-CGAGCTTAACGCGATTCTCCTTCTCCTGTACCACTTAAGCATT-5'	45 nt oligoduplex substrate containing 2 nt of proto-spacer1 and PAM
SPN (SEQ ID NO: 39)	5'-GCTCGAATTGCCACCCAGCAAAATTCGGTTTCTGGCTGATGGTGAATTCGTAAT-3' 3'-CGAGCTTAACGGTGGGTCGTTTAAAGCCAAAGACCGACTACCACTTAAGCATT-5'	55 by oligoduplex substrate containing proto-spacerN and PAM

② indicates text missing or illegible when filed

Results

[0045] Expression and purification of the Cas9-crRNA complex. The cas9 gene from the CRISPR3 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 (FIG. 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 (Sapranaukas et al, 2011). pCas9(-) also contained a complete CRISPR3 array comprised of 12 spacer-repeat units (FIG. 2A). To achieve simultaneous transcription of all target

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 (FIG. 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 (FIG. 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 (FIG. 3B). We have isolated Cas9-crRNA complex following the procedure described above and analysed crRNA bound to Cas9 protein.

[0046] 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 (FIGS. 4A and B). We assumed that crRNA isolated from the Cas9-crRNA complex expressed in the heterologous *E. coli* strain (FIG. 1) may have the same length (FIG. 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 (FIG. 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 (FIG. 3A, FIG. 4C). Taken together, these data confirm that Cas9 Strep-tag protein co-purifies with 42 nt crRNA, which is derived from CRISPR3 region.

[0047] 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 (Sapranaukas et al., 2011. Nucleic Acids Res 39:9275-82) (FIG. 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 FIG. 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 3 nt upstream of the PAM (FIG. 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 (FIG. 6C).

[0048] 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 (Sapranaukas et al., 2011. Nucleic Acids Res 39:9275-82) (FIG. 5C) carrying proto-spacer1 and PAM. In the presence of Cas9-crRNA complex supercoiled

form of pSP1 plasmid was converted into a linear form (FIG. 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 3 nt away from 5'-NGGNG-3' PAM sequence similarly to the SP1 oligoduplex cleavage (FIG. 5D). The cleavage positions identified in the in vitro experiments (FIG. 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.

[0049] 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 (FIG. 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 complementary sequence remained intact (FIG. 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 (FIG. 7D). Oligoduplex cleavage assay confirmed (FIG. 7C and D) that Cas9-crRNA complex with re-engineered specificity cleaves both DNA strands within the SN proto-spacer 3 nt upstream of the PAM identically to other Cas9-crRNA complexes.

[0050] The length of the spacer in the CRISPR3 region of *S. thermophilus* is 30 nt. According to the data provided in the FIG. 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 (FIG. 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 experi-

ments, the SP1-20 oligoduplex containing only 20 nt of the protospacer-1 is efficiently cleaved by Cas9-crRNA (FIGS. 8D and E).

[0051] 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 (FIG. 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).

[0052] Consistent with the plasmid cleavage experiments, oligoduplexes which have only proto-spacer, but not PAM are not cut by Cas9-crRNA (FIG. 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 (FIG. 9B). These data clearly show that PAM is required only for a double-stranded but not for a single-stranded DNA cleavage.

[0053] 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 (FIG. 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 (FIG. 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.

[0054] On the other hand, single-stranded oligonucleotides ((+)strand) are bound by Cas9-crRNA with the same affinity independently of the PAM presence (FIG. 9D). Again, no binding was observed for single-stranded DNA oligonucleotide without proto-spacer (FIG. 9D), or for Cas9 protein lacking crRNA (FIG. 6C). Taken together these data indicate that Cas9-crRNA complex discriminates PAM only in the double-stranded but not a single-stranded DNA.

[0055] 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 (FIG. 10B). This finding 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.

[0056] Mutagenesis of Cas9 protein RuvC and HNH motifs. Plasmid transformation experiments indicate that RuvC and HNH motifs (FIG. 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 (FIG. 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 (FIG. 11B) indicating that both active sites mutants cleave only one DNA strand of plasmid substrate within a proto-spacer.

[0057] 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 (FIG. 11C). And vice versa, HNH active site mutant (N891A) cleaved only (−)strand, but not (+)strand of the SP1 oligoduplex (FIG. 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 (FIGS. 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 (FIG. 12D).

[0058] 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 (FIG. 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 (FIG. 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 (FIG. 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

[0059] 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 3 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 protospacer 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 *Streptococcus 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 (FIG. 8).

[0060] 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 FIG. 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.

[0061] 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 (FIG. 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 3 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 N-terminal 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 (FIG. 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.

[0062] 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 *E. coli* 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.

[0063] In the III-B subtype CRISPR systems present in many archaea and some bacteria, Cas module RAMP (Cmr) proteins and crRNA 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 *Pyrococcus 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 of 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 invad-

ing 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.

[0064] 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

[0065] In Vitro assembly of Cas9-crRNA Complex from 4 Components

[0066] 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 6× His-tagged variant of Cas9 protein ("6× His" 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.

[0067] 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'-taatcagactactataGggtagaaaagatatcctacgagg-3' (SEQ ID NO: 40)/5'-CAACAAC-CAAGCTAATACAGCAG-3' (SEQ ID NO: 41) and 5'-aaaaaacaccgaatcggtgccac-3' (SEQ ID NO: 42)/5'-taatcagactactataGggTAATAATAATTGTGGTTTGAAACCATTC-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 complementary 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).

[0068] For in vitro assembly of catalytically competent Cas9-crRNA complex, the 6× His-tagged Cas9 protein ("6× His" 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 6× His-tagged Cas9 protein ("6× His" 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.

[0069] 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 as 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.

[0070] 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'-NGGNG-3' PAM sequence. Proto-spacer1 sequence was not present in pUC18. Reactions on pUC18 and pSP1 plasmids (Sapranaukas 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% bromophenol blue and 75 mM EDTA in 50% v/v glycerol) and reaction products analyzed by electrophoresis through agarose (FIG. 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'-ccgcatcaggcgcattcgcc-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.

[0071] Next, the cleavage activity of the in vitro assembled Cas9-crRNA complex was assayed on a synthetic 55 by oligodeoxynucleotide duplex SP1 containing a proto-spacer sequence matching to the spacer sequence of crRNA (FIG. 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, 3 nt upstream of the PAM sequence 5'-NGGNG-3' leaving blunt ends (FIG. 17).

EXAMPLE 3

[0072] In Vitro Assembly of Cas9-crRNA Complex from 3 Components

[0073] In this example we demonstrate that active Cas9-crRNA complex can be assembled in vitro by mixing 3 individual components: the C-terminal 6× His-tagged variant of Cas9 protein ("6× His" 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: 7) 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 6× His-tagged Cas9 protein ("6× His" 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 6× His-tagged Cas9 protein ("6× His" disclosed as SEQ ID NO: 23), 50 nM of CRISPR RNA or 42 nt synthetic crRNA, 100 nM of tracrRNA.

[0074] 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 as 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.

[0075] 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'-NGGNG-3' PAM sequence. Proto-spacer1 sequence was not present in pUC18. Reactions on plasmid substrates (Sapranaukas 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 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% bromophenol blue and 75 mM EDTA in 50% v/v glycerol) and reaction products analyzed by electrophoresis through agarose (FIG. 18).

[0076] Next, the cleavage activity of the in vitro assembled Cas9-crRNA complex was assayed on a synthetic 55 by oligodeoxynucleotide duplex SP1 containing a proto-spacer sequence matching to the spacer sequence of crRNA (FIG. 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 from the 5'-end, 3 nt upstream of the PAM sequence 5'-NGGNG-3' leaving blunt ends (FIG. 19).

EXAMPLE 4

[0077] Interchangeable Spacer Cassette for the Re-Programming of the Cas9-crRNA Complex Specificity.

[0078] 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 (FIG. 20B). The cassette carries 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 (FIG. 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).

[0079] 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 (FIG. 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 (FIG. 21B). When both complexes were incubated with the pUC18 plasmid, two DNA fragments (2074 and 612 bp) were obtained (FIG. 21B), indicating that plasmid cleavage occurred at sites targeted by the crRNA molecules present in the complexes.

EXAMPLE 5

[0080] Cloning Procedure Using Cas9-crRNA Complex.

[0081] 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. Science 186:790-7).

[0082] Next we analyzed cleavage of pUC18 plasmid with Cas9 complex loaded with crRNA1 and crRNA2 described in

Example 5 (FIG. 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 (FIG. 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 (FIG. 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 (FIG. 21C).

[0083] 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 MgCl₂). 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

[0084] Cleavage of Long DNA Substrates by Cas9 crRNA Complex.

[0085] In this example we demonstrate that Cas9-crRNA may be addressed to cleave targets in long DNA molecules, including phage A, *E. coli* and human genomic DNAs.

[0086] 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.

[0087] 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 MgCl₂ in 100 µl reaction volume.

Aliquots were removed at timed intervals and quenched with phenol/chloroform. The aqueous phase was mixed with 3× 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 by length DNA is efficiently cleaved at a single site (FIG. 22A).

[0088] 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 MgCl₂ in 300 µl reaction volume. Following incubation, 30 µl of FastDigest PstI (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 3× loading dye solution (0.01% bromphenol blue and 75 mM EDTA in 50% v/v glycerol) and electrophoresed on 1% agarose gel.

[0089] 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) (FIG. 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 2×SSC and air dried. The membrane was prehybridized with 6×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-³²P]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 2×SSC, 0.5% SDS and twice with 2×SSC, 0.1% SDS at room temperature, air dried and visualized by phosphorimaging (FLA-5100; Fujifilm).

[0090] The probe was designed to target DNA fragment containing a target (a proto-spacer) for the Cas9-RNA complex (FIG. 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 (FIG. 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.

[0091] 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 MgCl₂ in 100 µl reaction volume. Cas9-crRNA-HS1 (SEQ ID NO: 13) and Cas9-crRNA-HS2 (SEQ ID NO: 14) complexes were assembled to target RASGEF1C or ARL15 loci, respectively. Cleavage products were analyzed using qPCR (FIG. 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.

[0092] The application contains a Sequence Listing which has been submitted in ASCII format and is hereby incorporated by reference in its entirety. The ASCII copy, created on Mar. 13, 2013, is named 078981.4_SL.txt and is 63.6 kilobytes in size.

[0093] 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. The references cited are expressly incorporated by reference herein in their entirety.

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 69

<210> SEQ ID NO 1

<211> LENGTH: 1409

<212> TYPE: PRT

<213> ORGANISM: Streptococcus thermophilus

<400> SEQUENCE: 1

```

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

```

Glu 245	Lys	Lys	Asp	Arg 245	Ile	Leu	Lys	Leu	Phe 250	Pro	Gly	Glu	Lys	Asn 255	Ser
Gly	Ile	Phe	Ser 260	Glu	Phe	Leu	Lys	Leu 265	Ile	Val	Gly	Asn 270	Gln	Ala	Asp
Phe	Arg	Lys 275	Cys	Phe	Asn	Leu	Asp 280	Glu	Lys	Ala	Ser	Leu 285	His	Phe	Ser
Lys	Glu 290	Ser	Tyr	Asp	Glu	Asp 295	Leu	Glu	Thr	Leu	Leu 300	Gly	Tyr	Ile	Gly
Asp 305	Asp	Tyr	Ser	Asp	Val 310	Phe	Leu	Lys	Ala	Lys 315	Lys	Leu	Tyr	Asp	Ala 320
Ile	Leu	Leu	Ser	Gly 325	Phe	Leu	Thr	Val 330	Thr	Asp	Asn	Glu	Thr	Glu 335	Ala
Pro	Leu	Ser	Ser 340	Ala	Met	Ile	Lys	Arg 345	Tyr	Asn	Glu	His	Lys 350	Glu	Asp
Leu	Ala 355	Leu	Leu	Lys	Glu	Tyr	Ile 360	Arg	Asn	Ile	Ser	Leu 365	Lys	Thr	Tyr
Asn 370	Glu	Val	Phe	Lys	Asp	Asp 375	Thr	Lys	Asn	Gly	Tyr 380	Ala	Gly	Tyr	Ile
Asp 385	Gly	Lys	Thr	Asn	Gln 390	Glu	Asp	Phe	Tyr	Val 395	Tyr	Leu	Lys	Asn	Leu 400
Leu	Ala	Glu	Phe	Glu 405	Gly	Ala	Asp	Tyr	Phe	Leu	Glu	Lys	Ile	Asp 415	Arg
Glu	Asp	Phe	Leu	Arg	Lys	Gln	Arg	Thr 425	Phe	Asp	Asn	Gly	Ser 430	Ile	Pro
Tyr	Gln 435	Ile	His	Leu	Gln	Glu	Met 440	Arg	Ala	Ile	Leu	Asp 445	Lys	Gln	Ala
Lys	Phe 450	Tyr	Pro	Phe	Leu	Ala 455	Lys	Asn	Lys	Glu	Arg 460	Ile	Glu	Lys	Ile
Leu 465	Thr	Phe	Arg	Ile	Pro 470	Tyr	Tyr	Val	Gly	Pro 475	Leu	Ala	Arg	Gly	Asn 480
Ser	Asp	Phe	Ala 485	Trp	Ser	Ile	Arg	Lys	Arg 490	Asn	Glu	Lys	Ile	Thr 495	Pro
Trp	Asn	Phe	Glu 500	Asp	Val	Ile	Asp	Lys	Glu	Ser	Ser	Ala	Glu 510	Ala	Phe
Ile	Asn	Arg	Met 515	Thr	Ser	Phe	Asp 520	Leu	Tyr	Leu	Pro	Glu 525	Glu	Lys	Val
Leu 530	Pro	Lys	His	Ser	Leu	Leu 535	Tyr	Glu	Thr	Phe	Asn	Val 540	Tyr	Asn	Glu
Leu 545	Thr	Lys	Val	Arg	Phe 550	Ile	Ala	Glu	Ser	Met 555	Arg	Asp	Tyr	Gln	Phe 560
Leu	Asp	Ser	Lys	Gln 565	Lys	Lys	Asp	Ile	Val	Arg	Leu	Tyr	Phe	Lys 575	Asp
Lys	Arg	Lys	Val 580	Thr	Asp	Lys	Asp	Ile 585	Ile	Glu	Tyr	Leu	His	Ala	Ile
Tyr	Gly	Tyr 595	Asp	Gly	Ile	Glu	Leu 600	Lys	Gly	Ile	Glu	Lys 605	Gln	Phe	Asn
Ser	Ser 610	Leu	Ser	Thr	Tyr	His	Asp 615	Leu	Leu	Asn	Ile	Ile	Asn	Asp	Lys
Glu 625	Phe	Leu	Asp	Asp	Ser 630	Ser	Asn	Glu	Ala	Ile	Ile	Glu	Glu	Ile	Ile 640

His	Thr	Leu	Thr	Ile	Phe	Glu	Asp	Arg	Glu	Met	Ile	Lys	Gln	Arg	Leu		
				645					650							655	
Ser	Lys	Phe	Glu	Asn	Ile	Phe	Asp	Lys	Ser	Val	Leu	Lys	Lys	Leu	Ser		
				660					665							670	
Arg	Arg	His	Tyr	Thr	Gly	Trp	Gly	Lys	Leu	Ser	Ala	Lys	Leu	Ile	Asn		
				675					680							685	
Gly	Ile	Arg	Asp	Glu	Lys	Ser	Gly	Asn	Thr	Ile	Leu	Asp	Tyr	Leu	Ile		
				690					695							700	
Asp	Asp	Gly	Ile	Ser	Asn	Arg	Asn	Phe	Met	Gln	Leu	Ile	His	Asp	Asp		
				705					710							715	720
Ala	Leu	Ser	Phe	Lys	Lys	Lys	Ile	Gln	Lys	Ala	Gln	Ile	Ile	Gly	Asp		
				725					730							735	
Glu	Asp	Lys	Gly	Asn	Ile	Lys	Glu	Val	Val	Lys	Ser	Leu	Pro	Gly	Ser		
				740					745							750	
Pro	Ala	Ile	Lys	Lys	Gly	Ile	Leu	Gln	Ser	Ile	Lys	Ile	Val	Asp	Glu		
				755					760							765	
Leu	Val	Lys	Val	Met	Gly	Gly	Arg	Lys	Pro	Glu	Ser	Ile	Val	Val	Glu		
				770					775							780	
Met	Ala	Arg	Glu	Asn	Gln	Tyr	Thr	Asn	Gln	Gly	Lys	Ser	Asn	Ser	Gln		
				785					790							795	800
Gln	Arg	Leu	Lys	Arg	Leu	Glu	Lys	Ser	Leu	Lys	Glu	Leu	Gly	Ser	Lys		
				805					810							815	
Ile	Leu	Lys	Glu	Asn	Ile	Pro	Ala	Lys	Leu	Ser	Lys	Ile	Asp	Asn	Asn		
				820					825							830	
Ala	Leu	Gln	Asn	Asp	Arg	Leu	Tyr	Leu	Tyr	Tyr	Leu	Gln	Asn	Gly	Lys		
				835					840							845	
Asp	Met	Tyr	Thr	Gly	Asp	Asp	Leu	Asp	Ile	Asp	Arg	Leu	Ser	Asn	Tyr		
				850					855							860	
Asp	Ile	Asp	His	Ile	Ile	Pro	Gln	Ala	Phe	Leu	Lys	Asp	Asn	Ser	Ile		
				865					870							875	880
Asp	Asn	Lys	Val	Leu	Val	Ser	Ser	Ala	Ser	Asn	Arg	Gly	Lys	Ser	Asp		
				885					890							895	
Asp	Phe	Pro	Ser	Leu	Glu	Val	Val	Lys	Lys	Arg	Lys	Thr	Phe	Trp	Tyr		
				900					905							910	
Gln	Leu	Leu	Lys	Ser	Lys	Leu	Ile	Ser	Gln	Arg	Lys	Phe	Asp	Asn	Leu		
				915					920							925	
Thr	Lys	Ala	Glu	Arg	Gly	Gly	Leu	Leu	Pro	Glu	Asp	Lys	Ala	Gly	Phe		
				930					935							940	
Ile	Gln	Arg	Gln	Leu	Val	Glu	Thr	Arg	Gln	Ile	Thr	Lys	His	Val	Ala		
				945					950							955	960
Arg	Leu	Leu	Asp	Glu	Lys	Phe	Asn	Asn	Lys	Lys	Asp	Glu	Asn	Asn	Arg		
				965					970							975	
Ala	Val	Arg	Thr	Val	Lys	Ile	Ile	Thr	Leu	Lys	Ser	Thr	Leu	Val	Ser		
				980					985							990	
Gln	Phe	Arg	Lys	Asp	Phe	Glu	Leu	Tyr	Lys	Val	Arg	Glu	Ile	Asn	Asp		
				995					1000							1005	
Phe	His	His	Ala	His	Asp	Ala	Tyr	Leu	Asn	Ala	Val	Ile	Ala	Ser			
				1010					1015							1020	
Ala	Leu	Leu	Lys	Lys	Tyr	Pro	Lys	Leu	Glu	Pro	Glu	Phe	Val	Tyr			
				1025					1030							1035	
Gly	Asp	Tyr	Pro	Lys	Tyr	Asn	Ser	Phe	Arg	Glu	Arg	Lys	Ser	Ala			

-continued

1040	1045	1050
Thr Glu Lys Val Tyr Phe	Tyr Ser Asn Ile Met	Asn Ile Phe Lys
1055	1060	1065
Lys Ser Ile Ser Leu Ala	Asp Gly Arg Val Ile	Glu Arg Pro Leu
1070	1075	1080
Ile Glu Val Asn Glu Glu	Thr Gly Glu Ser Val	Trp Asn Lys Glu
1085	1090	1095
Ser Asp Leu Ala Thr Val	Arg Arg Val Leu Ser	Tyr Pro Gln Val
1100	1105	1110
Asn Val Val Lys Lys Val	Glu Glu Gln Asn His	Gly Leu Asp Arg
1115	1120	1125
Gly Lys Pro Lys Gly Leu	Phe Asn Ala Asn Leu	Ser Ser Lys Pro
1130	1135	1140
Lys Pro Asn Ser Asn Glu	Asn Leu Val Gly Ala	Lys Glu Tyr Leu
1145	1150	1155
Asp Pro Lys Lys Tyr Gly	Gly Tyr Ala Gly Ile	Ser Asn Ser Phe
1160	1165	1170
Ala Val Leu Val Lys Gly	Thr Ile Glu Lys Gly	Ala Lys Lys Lys
1175	1180	1185
Ile Thr Asn Val Leu Glu	Phe Gln Gly Ile Ser	Ile Leu Asp Arg
1190	1195	1200
Ile Asn Tyr Arg Lys Asp	Lys Leu Asn Phe Leu	Leu Glu Lys Gly
1205	1210	1215
Tyr Lys Asp Ile Glu Leu	Ile Ile Glu Leu Pro	Lys Tyr Ser Leu
1220	1225	1230
Phe Glu Leu Ser Asp Gly	Ser Arg Arg Met Leu	Ala Ser Ile Leu
1235	1240	1245
Ser Thr Asn Asn Lys Arg	Gly Glu Ile His Lys	Gly Asn Gln Ile
1250	1255	1260
Phe Leu Ser Gln Lys Phe	Val Lys Leu Leu Tyr	His Ala Lys Arg
1265	1270	1275
Ile Ser Asn Thr Ile Asn	Glu Asn His Arg Lys	Tyr Val Glu Asn
1280	1285	1290
His Lys Lys Glu Phe Glu	Glu Leu Phe Tyr Tyr	Ile Leu Glu Phe
1295	1300	1305
Asn Glu Asn Tyr Val Gly	Ala Lys Lys Asn Gly	Lys Leu Leu Asn
1310	1315	1320
Ser Ala Phe Gln Ser Trp	Gln Asn His Ser Ile	Asp Glu Leu Cys
1325	1330	1335
Ser Ser Phe Ile Gly Pro	Thr Gly Ser Glu Arg	Lys Gly Leu Phe
1340	1345	1350
Glu Leu Thr Ser Arg Gly	Ser Ala Ala Asp Phe	Glu Phe Leu Gly
1355	1360	1365
Val Lys Ile Pro Arg Tyr	Arg Asp Tyr Thr Pro	Ser Ser Leu Leu
1370	1375	1380
Lys Asp Ala Thr Leu Ile	His Gln Ser Val Thr	Gly Leu Tyr Glu
1385	1390	1395
Thr Arg Ile Asp Leu Ala	Lys Leu Gly Glu Gly	
1400	1405	

-continued

```

<211> LENGTH: 1409
<212> TYPE: PRT
<213> ORGANISM: Streptococcus thermophilus

<400> SEQUENCE: 2
Met Leu Phe Asn Lys Cys Ile Ile Ile Ser Ile Asn Leu Asp Phe Ser
 1             5             10             15
Asn Lys Glu Lys Cys Met Thr Lys Pro Tyr Ser Ile Gly Leu Ala Ile
 20             25             30
Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Asn Tyr Lys Val
 35             40             45
Pro Ser Lys Lys Met Lys Val Leu Gly Asn Thr Ser Lys Lys Tyr Ile
 50             55             60
Lys Lys Asn Leu Leu Gly Val Leu Leu Phe Asp Ser Gly Ile Thr Ala
 65             70             75             80
Glu Gly Arg Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg
 85             90             95
Arg Asn Arg Ile Leu Tyr Leu Gln Glu Ile Phe Ser Thr Glu Met Ala
100            105            110
Thr Leu Asp Asp Ala Phe Phe Gln Arg Leu Asp Asp Ser Phe Leu Val
115            120            125
Pro Asp Asp Lys Arg Asp Ser Lys Tyr Pro Ile Phe Gly Asn Leu Val
130            135            140
Glu Glu Lys Val Tyr His Asp Glu Phe Pro Thr Ile Tyr His Leu Arg
145            150            155            160
Lys Tyr Leu Ala Asp Ser Thr Lys Lys Ala Asp Leu Arg Leu Val Tyr
165            170            175
Leu Ala Leu Ala His Met Ile Lys Tyr Arg Gly His Phe Leu Ile Glu
180            185            190
Gly Glu Phe Asn Ser Lys Asn Asn Asp Ile Gln Lys Asn Phe Gln Asp
195            200            205
Phe Leu Asp Thr Tyr Asn Ala Ile Phe Glu Ser Asp Leu Ser Leu Glu
210            215            220
Asn Ser Lys Gln Leu Glu Glu Ile Val Lys Asp Lys Ile Ser Lys Leu
225            230            235            240
Glu Lys Lys Asp Arg Ile Leu Lys Leu Phe Pro Gly Glu Lys Asn Ser
245            250            255
Gly Ile Phe Ser Glu Phe Leu Lys Leu Ile Val Gly Asn Gln Ala Asp
260            265            270
Phe Arg Lys Cys Phe Asn Leu Asp Glu Lys Ala Ser Leu His Phe Ser
275            280            285
Lys Glu Ser Tyr Asp Glu Asp Leu Glu Thr Leu Leu Gly Tyr Ile Gly
290            295            300
Asp Asp Tyr Ser Asp Val Phe Leu Lys Ala Lys Lys Leu Tyr Asp Ala
305            310            315            320
Ile Leu Leu Ser Gly Phe Leu Thr Val Thr Asp Asn Glu Thr Glu Ala
325            330            335
Pro Leu Ser Ser Ala Met Ile Lys Arg Tyr Asn Glu His Lys Glu Asp
340            345            350
Leu Ala Leu Leu Lys Glu Tyr Ile Arg Asn Ile Ser Leu Lys Thr Tyr
355            360            365
Asn Glu Val Phe Lys Asp Asp Thr Lys Asn Gly Tyr Ala Gly Tyr Ile

```

-continued

370	375	380
Asp Gly Lys Thr Asn Gln Glu Asp Phe Tyr Val Tyr Leu Lys Asn Leu		
385	390	395 400
Leu Ala Glu Phe Glu Gly Ala Asp Tyr Phe Leu Glu Lys Ile Asp Arg		
	405	410 415
Glu Asp Phe Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro		
	420	425 430
Tyr Gln Ile His Leu Gln Glu Met Arg Ala Ile Leu Asp Lys Gln Ala		
	435	440 445
Lys Phe Tyr Pro Phe Leu Ala Lys Asn Lys Glu Arg Ile Glu Lys Ile		
	450	455 460
Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn		
465	470	475 480
Ser Asp Phe Ala Trp Ser Ile Arg Lys Arg Asn Glu Lys Ile Thr Pro		
	485	490 495
Trp Asn Phe Glu Asp Val Ile Asp Lys Glu Ser Ser Ala Glu Ala Phe		
	500	505 510
Ile Asn Arg Met Thr Ser Phe Asp Leu Tyr Leu Pro Glu Glu Lys Val		
	515	520 525
Leu Pro Lys His Ser Leu Leu Tyr Glu Thr Phe Asn Val Tyr Asn Glu		
	530	535 540
Leu Thr Lys Val Arg Phe Ile Ala Glu Ser Met Arg Asp Tyr Gln Phe		
545	550	555 560
Leu Asp Ser Lys Gln Lys Lys Asp Ile Val Arg Leu Tyr Phe Lys Asp		
	565	570 575
Lys Arg Lys Val Thr Asp Lys Asp Ile Ile Glu Tyr Leu His Ala Ile		
	580	585 590
Tyr Gly Tyr Asp Gly Ile Glu Leu Lys Gly Ile Glu Lys Gln Phe Asn		
	595	600 605
Ser Ser Leu Ser Thr Tyr His Asp Leu Leu Asn Ile Ile Asn Asp Lys		
	610	615 620
Glu Phe Leu Asp Asp Ser Ser Asn Glu Ala Ile Ile Glu Glu Ile Ile		
625	630	635 640
His Thr Leu Thr Ile Phe Glu Asp Arg Glu Met Ile Lys Gln Arg Leu		
	645	650 655
Ser Lys Phe Glu Asn Ile Phe Asp Lys Ser Val Leu Lys Lys Leu Ser		
	660	665 670
Arg Arg His Tyr Thr Gly Trp Gly Lys Leu Ser Ala Lys Leu Ile Asn		
	675	680 685
Gly Ile Arg Asp Glu Lys Ser Gly Asn Thr Ile Leu Asp Tyr Leu Ile		
	690	695 700
Asp Asp Gly Ile Ser Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp		
705	710	715 720
Ala Leu Ser Phe Lys Lys Lys Ile Gln Lys Ala Gln Ile Ile Gly Asp		
	725	730 735
Glu Asp Lys Gly Asn Ile Lys Glu Val Val Lys Ser Leu Pro Gly Ser		
	740	745 750
Pro Ala Ile Lys Lys Gly Ile Leu Gln Ser Ile Lys Ile Val Asp Glu		
	755	760 765
Leu Val Lys Val Met Gly Gly Arg Lys Pro Glu Ser Ile Val Val Glu		
	770	775 780

-continued

Met	Ala	Arg	Glu	Asn	Gln	Tyr	Thr	Asn	Gln	Gly	Lys	Ser	Asn	Ser	Gln	785	790	795	800
Gln	Arg	Leu	Lys	Arg	Leu	Glu	Lys	Ser	Leu	Lys	Glu	Leu	Gly	Ser	Lys	805	810	815	
Ile	Leu	Lys	Glu	Asn	Ile	Pro	Ala	Lys	Leu	Ser	Lys	Ile	Asp	Asn	Asn	820	825	830	
Ala	Leu	Gln	Asn	Asp	Arg	Leu	Tyr	Leu	Tyr	Tyr	Leu	Gln	Asn	Gly	Lys	835	840	845	
Asp	Met	Tyr	Thr	Gly	Asp	Asp	Leu	Asp	Ile	Asp	Arg	Leu	Ser	Asn	Tyr	850	855	860	
Asp	Ile	Asp	His	Ile	Ile	Pro	Gln	Ala	Phe	Leu	Lys	Asp	Asn	Ser	Ile	865	870	875	880
Asp	Asn	Lys	Val	Leu	Val	Ser	Ser	Ala	Ser	Asn	Arg	Gly	Lys	Ser	Asp	885	890	895	
Asp	Phe	Pro	Ser	Leu	Glu	Val	Val	Lys	Lys	Arg	Lys	Thr	Phe	Trp	Tyr	900	905	910	
Gln	Leu	Leu	Lys	Ser	Lys	Leu	Ile	Ser	Gln	Arg	Lys	Phe	Asp	Asn	Leu	915	920	925	
Thr	Lys	Ala	Glu	Arg	Gly	Gly	Leu	Leu	Pro	Glu	Asp	Lys	Ala	Gly	Phe	930	935	940	
Ile	Gln	Arg	Gln	Leu	Val	Glu	Thr	Arg	Gln	Ile	Thr	Lys	His	Val	Ala	945	950	955	960
Arg	Leu	Leu	Asp	Glu	Lys	Phe	Asn	Asn	Lys	Lys	Asp	Glu	Asn	Asn	Arg	965	970	975	
Ala	Val	Arg	Thr	Val	Lys	Ile	Ile	Thr	Leu	Lys	Ser	Thr	Leu	Val	Ser	980	985	990	
Gln	Phe	Arg	Lys	Asp	Phe	Glu	Leu	Tyr	Lys	Val	Arg	Glu	Ile	Asn	Asp	995	1000	1005	
Phe	His	His	Ala	His	Asp	Ala	Tyr	Leu	Asn	Ala	Val	Ile	Ala	Ser		1010	1015	1020	
Ala	Leu	Leu	Lys	Lys	Tyr	Pro	Lys	Leu	Glu	Pro	Glu	Phe	Val	Tyr		1025	1030	1035	
Gly	Asp	Tyr	Pro	Lys	Tyr	Asn	Ser	Phe	Arg	Glu	Arg	Lys	Ser	Ala		1040	1045	1050	
Thr	Glu	Lys	Val	Tyr	Phe	Tyr	Ser	Asn	Ile	Met	Asn	Ile	Phe	Lys		1055	1060	1065	
Lys	Ser	Ile	Ser	Leu	Ala	Asp	Gly	Arg	Val	Ile	Glu	Arg	Pro	Leu		1070	1075	1080	
Ile	Glu	Val	Asn	Glu	Glu	Thr	Gly	Glu	Ser	Val	Trp	Asn	Lys	Glu		1085	1090	1095	
Ser	Asp	Leu	Ala	Thr	Val	Arg	Arg	Val	Leu	Ser	Tyr	Pro	Gln	Val		1100	1105	1110	
Asn	Val	Val	Lys	Lys	Val	Glu	Glu	Gln	Asn	His	Gly	Leu	Asp	Arg		1115	1120	1125	
Gly	Lys	Pro	Lys	Gly	Leu	Phe	Asn	Ala	Asn	Leu	Ser	Ser	Lys	Pro		1130	1135	1140	
Lys	Pro	Asn	Ser	Asn	Glu	Asn	Leu	Val	Gly	Ala	Lys	Glu	Tyr	Leu		1145	1150	1155	
Asp	Pro	Lys	Lys	Tyr	Gly	Gly	Tyr	Ala	Gly	Ile	Ser	Asn	Ser	Phe		1160	1165	1170	

```

<210> SEQ ID NO 3
<211> LENGTH: 1409
<212> TYPE: PRT
<213> ORGANISM: Streptococcus thermophilus

<400> SEQUENCE: 3

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

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

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

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

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

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

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

```

Thr	Leu	Asp	Asp	Ala	Phe	Phe	Gln	Arg	Leu	Asp	Asp	Ser	Phe	Leu	Val
	115						120					125			
Pro	Asp	Asp	Lys	Arg	Asp	Ser	Lys	Tyr	Pro	Ile	Phe	Gly	Asn	Leu	Val
	130					135					140				
Glu	Glu	Lys	Val	Tyr	His	Asp	Glu	Phe	Pro	Thr	Ile	Tyr	His	Leu	Arg
145					150					155					160
Lys	Tyr	Leu	Ala	Asp	Ser	Thr	Lys	Lys	Ala	Asp	Leu	Arg	Leu	Val	Tyr
				165					170					175	
Leu	Ala	Leu	Ala	His	Met	Ile	Lys	Tyr	Arg	Gly	His	Phe	Leu	Ile	Glu
			180					185					190		
Gly	Glu	Phe	Asn	Ser	Lys	Asn	Asn	Asp	Ile	Gln	Lys	Asn	Phe	Gln	Asp
		195					200					205			
Phe	Leu	Asp	Thr	Tyr	Asn	Ala	Ile	Phe	Glu	Ser	Asp	Leu	Ser	Leu	Glu
210						215					220				
Asn	Ser	Lys	Gln	Leu	Glu	Glu	Ile	Val	Lys	Asp	Lys	Ile	Ser	Lys	Leu
225					230					235					240
Glu	Lys	Lys	Asp	Arg	Ile	Leu	Lys	Leu	Phe	Pro	Gly	Glu	Lys	Asn	Ser
				245					250					255	
Gly	Ile	Phe	Ser	Glu	Phe	Leu	Lys	Leu	Ile	Val	Gly	Asn	Gln	Ala	Asp
			260					265					270		
Phe	Arg	Lys	Cys	Phe	Asn	Leu	Asp	Glu	Lys	Ala	Ser	Leu	His	Phe	Ser
		275					280					285			
Lys	Glu	Ser	Tyr	Asp	Glu	Asp	Leu	Glu	Thr	Leu	Leu	Gly	Tyr	Ile	Gly
290						295					300				
Asp	Asp	Tyr	Ser	Asp	Val	Phe	Leu	Lys	Ala	Lys	Lys	Leu	Tyr	Asp	Ala
305					310					315					320
Ile	Leu	Leu	Ser	Gly	Phe	Leu	Thr	Val	Thr	Asp	Asn	Glu	Thr	Glu	Ala
				325				330						335	
Pro	Leu	Ser	Ser	Ala	Met	Ile	Lys	Arg	Tyr	Asn	Glu	His	Lys	Glu	Asp
			340					345					350		
Leu	Ala	Leu	Leu	Lys	Glu	Tyr	Ile	Arg	Asn	Ile	Ser	Leu	Lys	Thr	Tyr
			355				360					365			
Asn	Glu	Val	Phe	Lys	Asp	Asp	Thr	Lys	Asn	Gly	Tyr	Ala	Gly	Tyr	Ile
370						375					380				
Asp	Gly	Lys	Thr	Asn	Gln	Glu	Asp	Phe	Tyr	Val	Tyr	Leu	Lys	Asn	Leu
385					390					395					400
Leu	Ala	Glu	Phe	Glu	Gly	Ala	Asp	Tyr	Phe	Leu	Glu	Lys	Ile	Asp	Arg
				405					410					415	
Glu	Asp	Phe	Leu	Arg	Lys	Gln	Arg	Thr	Phe	Asp	Asn	Gly	Ser	Ile	Pro
			420					425					430		
Tyr	Gln	Ile	His	Leu	Gln	Glu	Met	Arg	Ala	Ile	Leu	Asp	Lys	Gln	Ala
							440					445			
Lys	Phe	Tyr	Pro	Phe	Leu	Ala	Lys	Asn	Lys	Glu	Arg	Ile	Glu	Lys	Ile
450						455					460				
Leu															

Ile	Asn	Arg	Met	Thr	Ser	Phe	Asp	Leu	Tyr	Leu	Pro	Glu	Glu	Lys	Val
		515					520					525			
Leu	Pro	Lys	His	Ser	Leu	Leu	Tyr	Glu	Thr	Phe	Asn	Val	Tyr	Asn	Glu
	530					535					540				
Leu	Thr	Lys	Val	Arg	Phe	Ile	Ala	Glu	Ser	Met	Arg	Asp	Tyr	Gln	Phe
545					550					555					560
Leu	Asp	Ser	Lys	Gln	Lys	Lys	Asp	Ile	Val	Arg	Leu	Tyr	Phe	Lys	Asp
				565					570					575	
Lys	Arg	Lys	Val	Thr	Asp	Lys	Asp	Ile	Ile	Glu	Tyr	Leu	His	Ala	Ile
			580					585					590		
Tyr	Gly	Tyr	Asp	Gly	Ile	Glu	Leu	Lys	Gly	Ile	Glu	Lys	Gln	Phe	Asn
		595					600					605			
Ser	Ser	Leu	Ser	Thr	Tyr	His	Asp	Leu	Leu	Asn	Ile	Ile	Asn	Asp	Lys
	610					615					620				
Glu	Phe	Leu	Asp	Asp	Ser	Ser	Asn	Glu	Ala	Ile	Ile	Glu	Glu	Ile	Ile
625					630					635					640
His	Thr	Leu	Thr	Ile	Phe	Glu	Asp	Arg	Glu	Met	Ile	Lys	Gln	Arg	Leu
				645					650					655	
Ser	Lys	Phe	Glu	Asn	Ile	Phe	Asp	Lys	Ser	Val	Leu	Lys	Lys	Leu	Ser
			660					665					670		
Arg	Arg	His	Tyr	Thr	Gly	Trp	Gly	Lys	Leu	Ser	Ala	Lys	Leu	Ile	Asn
		675					680					685			
Gly	Ile	Arg	Asp	Glu	Lys	Ser	Gly	Asn	Thr	Ile	Leu	Asp	Tyr	Leu	Ile
	690					695					700				
Asp	Asp	Gly	Ile	Ser	Asn	Arg	Asn	Phe	Met	Gln	Leu	Ile	His	Asp	Asp
705					710					715					720
Ala	Leu	Ser	Phe	Lys	Lys	Lys	Ile	Gln	Lys	Ala	Gln	Ile	Ile	Gly	Asp
				725					730					735	
Glu	Asp	Lys	Gly	Asn	Ile	Lys	Glu	Val	Val	Lys	Ser	Leu	Pro	Gly	Ser
			740					745					750		
Pro	Ala	Ile	Lys	Lys	Gly	Ile	Leu	Gln	Ser	Ile	Lys	Ile	Val	Asp	Glu
		755				760						765			
Leu	Val	Lys	Val	Met	Gly	Gly	Arg	Lys	Pro	Glu	Ser	Ile	Val	Val	Glu
	770				775						780				
Met	Ala	Arg	Glu	Asn	Gln	Tyr	Thr	Asn	Gln	Gly	Lys	Ser	Asn	Ser	Gln
785					790					795					800
Gln	Arg	Leu	Lys	Arg	Leu	Glu	Lys	Ser	Leu	Lys	Glu	Leu	Gly	Ser	Lys
				805					810					815	
Ile	Leu	Lys	Glu	Asn	Ile	Pro	Ala	Lys	Leu	Ser	Lys	Ile	Asp	Asn	Asn
			820					825					830		
Ala	Leu	Gln	Asn	Asp	Arg	Leu	Tyr	Leu	Tyr	Tyr	Leu	Gln	Asn	Gly	Lys
		835					840					845			
Asp	Met	Tyr	Thr	Gly	Asp	Asp	Leu	Asp	Ile	Asp	Arg	Leu	Ser	Asn	Tyr
	850					855					860				
Asp	Ile														

-continued

915					920					925					
Thr	Lys	Ala	Glu	Arg	Gly	Gly	Leu	Leu	Pro	Glu	Asp	Lys	Ala	Gly	Phe
930						935					940				
Ile	Gln	Arg	Gln	Leu	Val	Glu	Thr	Arg	Gln	Ile	Thr	Lys	His	Val	Ala
945					950					955					960
Arg	Leu	Leu	Asp	Glu	Lys	Phe	Asn	Asn	Lys	Lys	Asp	Glu	Asn	Asn	Arg
				965					970					975	
Ala	Val	Arg	Thr	Val	Lys	Ile	Ile	Thr	Leu	Lys	Ser	Thr	Leu	Val	Ser
			980					985					990		
Gln	Phe	Arg	Lys	Asp	Phe	Glu	Leu	Tyr	Lys	Val	Arg	Glu	Ile	Asn	Asp
		995					1000					1005			
Phe	His	His	Ala	His	Asp	Ala	Tyr	Leu	Asn	Ala	Val	Ile	Ala	Ser	
1010						1015						1020			
Ala	Leu	Leu	Lys	Lys	Tyr	Pro	Lys	Leu	Glu	Pro	Glu	Phe	Val	Tyr	
1025						1030						1035			
Gly	Asp	Tyr	Pro	Lys	Tyr	Asn	Ser	Phe	Arg	Glu	Arg	Lys	Ser	Ala	
1040						1045						1050			
Thr	Glu	Lys	Val	Tyr	Phe	Tyr	Ser	Asn	Ile	Met	Asn	Ile	Phe	Lys	
1055						1060						1065			
Lys	Ser	Ile	Ser	Leu	Ala	Asp	Gly	Arg	Val	Ile	Glu	Arg	Pro	Leu	
1070						1075						1080			
Ile	Glu	Val	Asn	Glu	Glu	Thr	Gly	Glu	Ser	Val	Trp	Asn	Lys	Glu	
1085						1090						1095			
Ser	Asp	Leu	Ala	Thr	Val	Arg	Arg	Val	Leu	Ser	Tyr	Pro	Gln	Val	
1100						1105						1110			
Asn	Val	Val	Lys	Lys	Val	Glu	Glu	Gln	Asn	His	Gly	Leu	Asp	Arg	
1115						1120						1125			
Gly	Lys	Pro	Lys	Gly	Leu	Phe	Asn	Ala	Asn	Leu	Ser	Ser	Lys	Pro	
1130						1135						1140			
Lys	Pro	Asn	Ser	Asn	Glu	Asn	Leu	Val	Gly	Ala	Lys	Glu	Tyr	Leu	
1145						1150						1155			
Asp	Pro	Lys	Lys	Tyr	Gly	Gly	Tyr	Ala	Gly	Ile	Ser	Asn	Ser	Phe	
1160						1165						1170			
Ala	Val	Leu	Val	Lys	Gly	Thr	Ile	Glu	Lys	Gly	Ala	Lys	Lys	Lys	
1175						1180						1185			
Ile	Thr	Asn	Val	Leu	Glu	Phe	Gln	Gly	Ile	Ser	Ile	Leu	Asp	Arg	
1190						1195						1200			
Ile	Asn	Tyr	Arg	Lys	Asp	Lys	Leu	Asn	Phe	Leu	Leu	Glu	Lys	Gly	
1205						1210						1215			
Tyr	Lys	Asp	Ile	Glu	Leu	Ile	Ile	Glu	Leu	Pro	Lys	Tyr	Ser	Leu	
1220						1225						1230			
Phe	Glu	Leu	Ser	Asp	Gly	Ser	Arg	Arg	Met	Leu	Ala	Ser	Ile	Leu	
1235						1240						1245			
Ser	Thr	Asn	Asn	Lys	Arg	Gly	Glu	Ile	His	Lys	Gly	Asn	Gln	Ile	
1250						1255						1260			
Phe	Leu	Ser	Gln	Lys	Phe	Val	Lys	Leu	Leu	Tyr	His	Ala	Lys	Arg	
1265						1270						1275			
Ile	Ser	Asn	Thr	Ile	Asn	Glu	Asn	His	Arg	Lys	Tyr	Val	Glu	Asn	
1280						1285						1290			
His	Lys	Lys	Glu	Phe	Glu	Glu	Leu	Phe	Tyr	Tyr	Ile	Leu	Glu	Phe	
1295						1300						1305			

-continued

```

Asn Glu  Asn Tyr Val Gly Ala  Lys Lys Asn Gly Lys  Leu Leu Asn
1310                      1315                      1320

Ser Ala  Phe Gln Ser Trp Gln  Asn His Ser Ile Asp  Glu Leu Cys
1325                      1330                      1335

Ser Ser  Phe Ile Gly Pro Thr  Gly Ser Glu Arg Lys  Gly Leu Phe
1340                      1345                      1350

Glu Leu  Thr Ser Arg Gly Ser  Ala Ala Asp Phe Glu  Phe Leu Gly
1355                      1360                      1365

Val Lys  Ile Pro Arg Tyr Arg  Asp Tyr Thr Pro Ser  Ser Leu Leu
1370                      1375                      1380

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

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

```

```

<210> SEQ ID NO 4
<211> LENGTH: 1409
<212> TYPE: PRT
<213> ORGANISM: Streptococcus thermophilus

```

```

<400> SEQUENCE: 4

```

```

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Glu Lys Lys Asp Arg Ile Leu Lys Leu Phe Pro Gly Glu Lys Asn Ser

```

-continued

245								250					255				
Gly	Ile	Phe	Ser	Glu	Phe	Leu	Lys	Leu	Ile	Val	Gly	Asn	Gln	Ala	Asp		
			260					265				270					
Phe	Arg	Lys	Cys	Phe	Asn	Leu	Asp	Glu	Lys	Ala	Ser	Leu	His	Phe	Ser		
			275				280				285						
Lys	Glu	Ser	Tyr	Asp	Glu	Asp	Leu	Glu	Thr	Leu	Leu	Gly	Tyr	Ile	Gly		
			290	295						300							
Asp	Asp	Tyr	Ser	Asp	Val	Phe	Leu	Lys	Ala	Lys	Lys	Leu	Tyr	Asp	Ala		
			305	310						315			320				
Ile	Leu	Leu	Ser	Gly	Phe	Leu	Thr	Val	Thr	Asp	Asn	Glu	Thr	Glu	Ala		
			325				330						335				
Pro	Leu	Ser	Ser	Ala	Met	Ile	Lys	Arg	Tyr	Asn	Glu	His	Lys	Glu	Asp		
			340				345				350						
Leu	Ala	Leu	Leu	Lys	Glu	Tyr	Ile	Arg	Asn	Ile	Ser	Leu	Lys	Thr	Tyr		
			355	360						365							
Asn	Glu	Val	Phe	Lys	Asp	Asp	Thr	Lys	Asn	Gly	Tyr	Ala	Gly	Tyr	Ile		
			370	375						380							
Asp	Gly	Lys	Thr	Asn	Gln	Glu	Asp	Phe	Tyr	Val	Tyr	Leu	Lys	Asn	Leu		
			385	390						395			400				
Leu	Ala	Glu	Phe	Glu	Gly	Ala	Asp	Tyr	Phe	Leu	Glu	Lys	Ile	Asp	Arg		
			405				410						415				
Glu	Asp	Phe	Leu	Arg	Lys	Gln	Arg	Thr	Phe	Asp	Asn	Gly	Ser	Ile	Pro		
			420				425						430				
Tyr	Gln	Ile	His	Leu	Gln	Glu	Met	Arg	Ala	Ile	Leu	Asp	Lys	Gln	Ala		
			435	440						445							
Lys	Phe	Tyr	Pro	Phe	Leu	Ala	Lys	Asn	Lys	Glu	Arg	Ile	Glu	Lys	Ile		
			450	455						460							
Leu	Thr	Phe	Arg	Ile	Pro	Tyr	Tyr	Val	Gly	Pro	Leu	Ala	Arg	Gly	Asn		
			465	470						475			480				
Ser	Asp	Phe	Ala	Trp	Ser	Ile	Arg	Lys	Arg	Asn	Glu	Lys	Ile	Thr	Pro		
			485				490						495				
Trp	Asn	Phe	Glu	Asp	Val	Ile	Asp	Lys	Glu	Ser	Ser	Ala	Glu	Ala	Phe		
			500				505						510				
Ile	Asn	Arg	Met	Thr	Ser	Phe	Asp	Leu	Tyr	Leu	Pro	Glu	Glu	Lys	Val		
			515				520						525				
Leu	Pro	Lys	His	Ser	Leu	Leu	Tyr	Glu	Thr	Phe	Asn	Val	Tyr	Asn	Glu		
			530	535						540							
Leu	Thr	Lys	Val	Arg	Phe	Ile	Ala	Glu	Ser	Met	Arg	Asp	Tyr	Gln	Phe		
			545	550						555			560				
Leu	Asp	Ser	Lys	Gln	Lys	Lys	Asp	Ile	Val	Arg	Leu	Tyr	Phe	Lys	Asp		
			565				570						575				
Lys	Arg	Lys	Val	Thr	Asp	Lys	Asp	Ile	Ile	Glu	Tyr	Leu	His	Ala	Ile		
			580				585						590				
Tyr	Gly	Tyr	Asp	Gly	Ile	Glu	Leu	Lys	Gly	Ile	Glu	Lys	Gln	Phe	Asn		
			595	600						605							
Ser	Ser	Leu	Ser	Thr	Tyr	His	Asp	Leu	Leu	Asn	Ile	Ile	Asn	Asp	Lys		
			610	615						620							
Glu	Phe	Leu	Asp	Asp	Ser	Ser	Asn	Glu	Ala	Ile	Ile	Glu	Glu	Ile	Ile		
			625	630						635			640				
His	Thr	Leu	Thr	Ile	Phe	Glu	Asp	Arg	Glu	Met	Ile	Lys	Gln	Arg	Leu		
			645				650						655				

-continued

Ser	Lys	Phe	Glu	Asn	Ile	Phe	Asp	Lys	Ser	Val	Leu	Lys	Lys	Leu	Ser	660	665	670	
Arg	Arg	His	Tyr	Thr	Gly	Trp	Gly	Lys	Leu	Ser	Ala	Lys	Leu	Ile	Asn	675	680	685	
Gly	Ile	Arg	Asp	Glu	Lys	Ser	Gly	Asn	Thr	Ile	Leu	Asp	Tyr	Leu	Ile	690	695	700	
Asp	Asp	Gly	Ile	Ser	Asn	Arg	Asn	Phe	Met	Gln	Leu	Ile	His	Asp	Asp	705	710	715	720
Ala	Leu	Ser	Phe	Lys	Lys	Lys	Ile	Gln	Lys	Ala	Gln	Ile	Ile	Gly	Asp	725	730	735	
Glu	Asp	Lys	Gly	Asn	Ile	Lys	Glu	Val	Val	Lys	Ser	Leu	Pro	Gly	Ser	740	745	750	
Pro	Ala	Ile	Lys	Lys	Gly	Ile	Leu	Gln	Ser	Ile	Lys	Ile	Val	Asp	Glu	755	760	765	
Leu	Val	Lys	Val	Met	Gly	Gly	Arg	Lys	Pro	Glu	Ser	Ile	Val	Val	Glu	770	775	780	
Met	Ala	Arg	Glu	Asn	Gln	Tyr	Thr	Asn	Gln	Gly	Lys	Ser	Asn	Ser	Gln	785	790	795	800
Gln	Arg	Leu	Lys	Arg	Leu	Glu	Lys	Ser	Leu	Lys	Glu	Leu	Gly	Ser	Lys	805	810	815	
Ile	Leu	Lys	Glu	Asn	Ile	Pro	Ala	Lys	Leu	Ser	Lys	Ile	Asp	Asn	Asn	820	825	830	
Ala	Leu	Gln	Asn	Asp	Arg	Leu	Tyr	Leu	Tyr	Tyr	Leu	Gln	Asn	Gly	Lys	835	840	845	
Asp	Met	Tyr	Thr	Gly	Asp	Asp	Leu	Asp	Ile	Asp	Arg	Leu	Ser	Asn	Tyr	850	855	860	
Asp	Ile	Asp	Ala	Ile	Ile	Pro	Gln	Ala	Phe	Leu	Lys	Asp	Asn	Ser	Ile	865	870	875	880
Asp	Asn	Lys	Val	Leu	Val	Ser	Ser	Ala	Ser	Asn	Arg	Gly	Lys	Ser	Asp	885	890	895	
Asp	Phe	Pro	Ser	Leu	Glu	Val	Val	Lys	Lys	Arg	Lys	Thr	Phe	Trp	Tyr	900	905	910	
Gln	Leu	Leu	Lys	Ser	Lys	Leu	Ile	Ser	Gln	Arg	Lys	Phe	Asp	Asn	Leu	915	920	925	
Thr	Lys	Ala	Glu	Arg	Gly	Gly	Leu	Leu	Pro	Glu	Asp	Lys	Ala	Gly	Phe	930	935	940	
Ile	Gln	Arg	Gln	Leu	Val	Glu	Thr	Arg	Gln	Ile	Thr	Lys	His	Val	Ala	945	950	955	960
Arg	Leu	Leu	Asp	Glu	Lys	Phe	Asn	Asn	Lys	Lys	Asp	Glu	Asn	Asn	Arg	965	970	975	
Ala	Val	Arg	Thr	Val	Lys	Ile	Ile	Thr	Leu	Lys	Ser	Thr	Leu	Val	Ser	980	985	990	
Gln	Phe	Arg	Lys	Asp	Phe	Glu	Leu	Tyr	Lys	Val	Arg	Glu	Ile	Asn	Asp	995	1000	1005	
Phe	His	His	Ala	His	Asp	Ala	Tyr	Leu	Asn	Ala	Val	Ile	Ala	Ser		1010	1015	1020	
Ala	Leu	Leu	Lys	Lys	Tyr	Pro	Lys	Leu	Glu	Pro	Glu	Phe	Val	Tyr		1025	1030	1035	
Gly	Asp	Tyr	Pro	Lys	Tyr	Asn	Ser	Phe	Arg	Glu	Arg	Lys	Ser	Ala		1040	1045	1050	

-continued

Thr 1055	Glu	Lys	Val	Tyr	Phe	Tyr 1060	Ser	Asn	Ile	Met	Asn 1065	Ile	Phe	Lys
Lys 1070	Ser	Ile	Ser	Leu	Ala	Asp 1075	Gly	Arg	Val	Ile	Glu 1080	Arg	Pro	Leu
Ile 1085	Glu	Val	Asn	Glu	Glu	Thr 1090	Gly	Glu	Ser	Val	Trp 1095	Asn	Lys	Glu
Ser 1100	Asp	Leu	Ala	Thr	Val	Arg 1105	Arg	Val	Leu	Ser	Tyr 1110	Pro	Gln	Val
Asn 1115	Val	Val	Lys	Lys	Val	Glu 1120	Glu	Gln	Asn	His	Gly 1125	Leu	Asp	Arg
Gly 1130	Lys	Pro	Lys	Gly	Leu	Phe 1135	Asn	Ala	Asn	Leu	Ser 1140	Ser	Lys	Pro
Lys 1145	Pro	Asn	Ser	Asn	Glu	Asn 1150	Leu	Val	Gly	Ala	Lys 1155	Glu	Tyr	Leu
Asp 1160	Pro	Lys	Lys	Tyr	Gly	Gly 1165	Tyr	Ala	Gly	Ile	Ser 1170	Asn	Ser	Phe
Ala 1175	Val	Leu	Val	Lys	Gly	Thr 1180	Ile	Glu	Lys	Gly	Ala 1185	Lys	Lys	Lys
Ile 1190	Thr	Asn	Val	Leu	Glu	Phe 1195	Gln	Gly	Ile	Ser	Ile 1200	Leu	Asp	Arg
Ile 1205	Asn	Tyr	Arg	Lys	Asp	Lys 1210	Leu	Asn	Phe	Leu	Leu 1215	Glu	Lys	Gly
Tyr 1220	Lys	Asp	Ile	Glu	Leu	Ile 1225	Ile	Glu	Leu	Pro	Lys 1230	Tyr	Ser	Leu
Phe 1235	Glu	Leu	Ser	Asp	Gly	Ser 1240	Arg	Arg	Met	Leu	Ala 1245	Ser	Ile	Leu
Ser 1250	Thr	Asn	Asn	Lys	Arg	Gly 1255	Glu	Ile	His	Lys	Gly 1260	Asn	Gln	Ile
Phe 1265	Leu	Ser	Gln	Lys	Phe	Val 1270	Lys	Leu	Leu	Tyr	His 1275	Ala	Lys	Arg
Ile 1280	Ser	Asn	Thr	Ile	Asn	Glu 1285	Asn	His	Arg	Lys	Tyr 1290	Val	Glu	Asn
His 1295	Lys	Lys	Glu	Phe	Glu	Glu 1300	Leu	Phe	Tyr	Tyr	Ile 1305	Leu	Glu	Phe
Asn 1310	Glu	Asn	Tyr	Val	Gly	Ala 1315	Lys	Lys	Asn	Gly	Lys 1320	Leu	Leu	Asn
Ser 1325	Ala	Phe	Gln	Ser	Trp	Gln 1330	Asn	His	Ser	Ile	Asp 1335	Glu	Leu	Cys
Ser 1340	Ser	Phe	Ile	Gly	Pro	Thr 1345	Gly	Ser	Glu	Arg	Lys 1350	Gly	Leu	Phe
Glu 1355	Leu	Thr	Ser	Arg	Gly	Ser 1360	Ala	Ala	Asp	Phe	Glu 1365	Phe	Leu	Gly
Val 1370	Lys	Ile	Pro	Arg	Tyr	Arg 1375	Asp	Tyr	Thr	Pro	Ser 1380	Ser	Leu	Leu
Lys 1385	Asp	Ala	Thr	Leu	Ile	His 1390	Gln	Ser	Val	Thr	Gly 1395	Leu	Tyr	Glu
Thr 1400	Arg	Ile	Asp	Leu	Ala	Lys 1405	Leu	Gly	Glu	Gly				

<210> SEQ ID NO 5

<211> LENGTH: 102

<212> TYPE: RNA

-continued

<213> ORGANISM: Streptococcus thermophilus

<400> SEQUENCE: 5

uaauaauuuu uguguuuga aaccuucga aacaacacag cgaguuaaaa uaaggcuuag 60

uccguacuca acuugaaaag guggcaccga uucgguguuu uu 102

<210> SEQ ID NO 6

<211> LENGTH: 78

<212> TYPE: RNA

<213> ORGANISM: Streptococcus thermophilus

<400> SEQUENCE: 6

gggcgaaca acacagcgag uaaaaaag gcuuaguccg uacucaacuu gaaaaggugg 60

caccgauucg guguuuuu 78

<210> SEQ ID NO 7

<211> LENGTH: 42

<212> TYPE: RNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<221> NAME/KEY: source

<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic oligonucleotide"

<400> SEQUENCE: 7

cgcuaaaagag gaagaggaca guuuuagagc uguguuguuu cg 42

<210> SEQ ID NO 8

<211> LENGTH: 150

<212> TYPE: RNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<221> NAME/KEY: source

<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic polynucleotide"

<400> SEQUENCE: 8

ggguagaaaa gauauccuac gagguuuuag agcuguguug uuucgaaugg uuccaaaaaca 60

aaucuaaac gcuaaagagg aagaggacag uuuuagagcu guguuuguuc gaaugguucc 120

aaaacuacug cuguauuagc uugguuguug 150

<210> SEQ ID NO 9

<211> LENGTH: 150

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<221> NAME/KEY: source

<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic oligonucleotide"

<220> FEATURE:

<221> NAME/KEY: source

<223> OTHER INFORMATION: /note="Description of Combined DNA/RNA
Molecule: Synthetic oligonucleotide"

<400> SEQUENCE: 9

ggguagaaaa gauauccuac gagguuuuag agcuguguug uuucgaaugg uuccaaaact 60

gtcatgataa taatggtttc ttagacgtcg uuuuagagcu guguuuguuc gaaugguucc 120

aaaacuacug cuguauuagc uugguuguug 150

<210> SEQ ID NO 10

-continued

```

<211> LENGTH: 150
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
    Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Combined DNA/RNA
    Molecule: Synthetic oligonucleotide"

<400> SEQUENCE: 10

ggguagaaaa gauauccuac gagguuuuag agcuguguug uuucgaaugg uuccaaaaca      60
cgagccggaa gcataaagt taaagcctgg uuuuagagcu guguuguuuc gaaugguucc      120
aaaacuacug cuguauuagc uugguuguug                                     150

<210> SEQ ID NO 11
<211> LENGTH: 150
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
    Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Combined DNA/RNA
    Molecule: Synthetic oligonucleotide"

<400> SEQUENCE: 11

ggguagaaaa gauauccuac gagguuuuag agcuguguug uuucgaaugg uuccaaaact      60
caaggaggaa tagagctct cgttgcattg uuuuagagcu guguuguuuc gaaugguucc      120
aaaacuacug cuguauuagc uugguuguug                                     150

<210> SEQ ID NO 12
<211> LENGTH: 150
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
    Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Combined DNA/RNA
    Molecule: Synthetic oligonucleotide"

<400> SEQUENCE: 12

ggguagaaaa gauauccuac gagguuuuag agcuguguug uuucgaaugg uuccaaaacc      60
gggagggag ctgcatgatg cgatgttatg uuuuagagcu guguuguuuc gaaugguucc      120
aaaacuacug cuguauuagc uugguuguug                                     150

<210> SEQ ID NO 13
<211> LENGTH: 42
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
    Synthetic oligonucleotide"

<400> SEQUENCE: 13

```

-continued

gcucccgagg cucgaugaag guuuuagagc uguguuguuu cg 42

<210> SEQ ID NO 14
<211> LENGTH: 42
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic oligonucleotide"

<400> SEQUENCE: 14

ugaaucguga aaucgcuca guuuuagagc uguguuguuu cg 42

<210> SEQ ID NO 15
<211> LENGTH: 42
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic oligonucleotide"

<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(20)
<223> OTHER INFORMATION: a, c, u, g, unknown or other

<400> SEQUENCE: 15

nnnnnnnnnn nnnnnnnnnn guuuuagagc uguguuguuu cg 42

<210> SEQ ID NO 16
<211> LENGTH: 29
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic oligonucleotide"

<400> SEQUENCE: 16

ccaccagca aaattcggtt ttctggctg 29

<210> SEQ ID NO 17
<211> LENGTH: 36
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic primer"

<400> SEQUENCE: 17

taatacgact cactataggg taccgagctc gaattg 36

<210> SEQ ID NO 18
<211> LENGTH: 31
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic primer"

<400> SEQUENCE: 18

gggaaacagc tatgacctg attacgaatt c 31

-continued

<210> SEQ ID NO 19
<211> LENGTH: 31
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic primer"

<400> SEQUENCE: 19

gggtaccgag ctcgaattga aattctaaac g 31

<210> SEQ ID NO 20
<211> LENGTH: 43
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic primer"

<400> SEQUENCE: 20

taatacgact cactataggg aaacagctat gaccatgatt acg 43

<210> SEQ ID NO 21
<211> LENGTH: 38
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic primer"

<400> SEQUENCE: 21

acgtctcaaa tggtgtttaa taagtgtata ataatttc 38

<210> SEQ ID NO 22
<211> LENGTH: 29
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic primer"

<400> SEQUENCE: 22

acgtctccgc gctaccctct cctagtttg 29

<210> SEQ ID NO 23
<211> LENGTH: 6
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic 6xHis tag"

<400> SEQUENCE: 23

His His His His His His
1 5

<210> SEQ ID NO 24
<211> LENGTH: 34
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence

-continued

<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic primer"

<400> SEQUENCE: 24

acgtctcaca tgactaagcc atactcaatt ggac 34

<210> SEQ ID NO 25
<211> LENGTH: 27
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic primer"

<400> SEQUENCE: 25

actcgagacc ctctcctagt ttggcaa 27

<210> SEQ ID NO 26
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic primer"

<400> SEQUENCE: 26

gaccacttat tgaggtaaataat gag 23

<210> SEQ ID NO 27
<211> LENGTH: 28
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic primer"

<400> SEQUENCE: 27

caaaccagga tccaagctaa tacagcag 28

<210> SEQ ID NO 28
<211> LENGTH: 42
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic oligonucleotide"

<400> SEQUENCE: 28

tcgaaacaac acagctctaa aactgtcctc ttcctcttta gc 42

<210> SEQ ID NO 29
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic primer"

<400> SEQUENCE: 29

-continued

ccgcatcagg cgccattcgc c 21

<210> SEQ ID NO 30
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: "Description of Artificial Sequence:
Synthetic primer"

<400> SEQUENCE: 30

gcgaggaagc ggaagagcgc cc 22

<210> SEQ ID NO 31
<211> LENGTH: 55
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic oligonucleotide"

<400> SEQUENCE: 31

gctcgaattg aaattctaaa cgctaaagag gaagaggaca tggatgaattc gtaat 55

<210> SEQ ID NO 32
<211> LENGTH: 50
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic oligonucleotide"

<400> SEQUENCE: 32

gctcgaattg aaattctaaa cgctaaagag gaagaggaca aattcgtaat 50

<210> SEQ ID NO 33
<211> LENGTH: 55
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic oligonucleotide"

<400> SEQUENCE: 33

gctcgaattg tactgctgta ttagcttggt tggtaggttg tggatgaattc gtaat 55

<210> SEQ ID NO 34
<211> LENGTH: 55
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic oligonucleotide"

<400> SEQUENCE: 34

attacgaatt caccatgtcc tttctctctt tagcgttttag aatttcaatt cgagc 55

<210> SEQ ID NO 35
<211> LENGTH: 50

-continued

<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic oligonucleotide"

<400> SEQUENCE: 35

attacgaatt tgtcctcttc ctcttttagcg tttagaattt caattcgagc 50

<210> SEQ ID NO 36
<211> LENGTH: 55
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic oligonucleotide"

<400> SEQUENCE: 36

attacgaatt caccacaaac caacaaccaa gctaatacag cagtacaatt cgagc 55

<210> SEQ ID NO 37
<211> LENGTH: 55
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic oligonucleotide"

<400> SEQUENCE: 37

gctcgaattg aaattctaaa cgctaaagag gaagaggaca tggatgaattc gtaat 55

<210> SEQ ID NO 38
<211> LENGTH: 45
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic oligonucleotide"

<400> SEQUENCE: 38

gctcgaattg cgctaaagag gaagaggaca tggatgaattc gtaat 45

<210> SEQ ID NO 39
<211> LENGTH: 55
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic oligonucleotide"

<400> SEQUENCE: 39

gctcgaattg ccaccagca aaattcggtt ttctggctga tggatgaattc gtaat 55

<210> SEQ ID NO 40
<211> LENGTH: 41
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic primer"

-continued

<400> SEQUENCE: 40

taatacgact cactataggg tagaaaagat atcctacgag g 41

<210> SEQ ID NO 41

<211> LENGTH: 23

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<221> NAME/KEY: source

<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic primer"

<400> SEQUENCE: 41

caacaaccaa gctaatacag cag 23

<210> SEQ ID NO 42

<211> LENGTH: 22

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<221> NAME/KEY: source

<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic primer"

<400> SEQUENCE: 42

aaaaacaccg aatcggtgcc ac 22

<210> SEQ ID NO 43

<211> LENGTH: 48

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<221> NAME/KEY: source

<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic primer"

<400> SEQUENCE: 43

taatacgact cactataggg taataataat tgtgggttga aaccattc 48

<210> SEQ ID NO 44

<211> LENGTH: 73

<212> TYPE: RNA

<213> ORGANISM: Streptococcus thermophilus

<400> SEQUENCE: 44

gggcgaaaca acacagcgag uaaaaaag gcuuaguccg uacucaacuu gaaaaggugg 60

caccgauucg gug 73

<210> SEQ ID NO 45

<211> LENGTH: 68

<212> TYPE: RNA

<213> ORGANISM: Streptococcus thermophilus

<400> SEQUENCE: 45

gggcgaaaca acacagcgag uaaaaaag gcuuaguccg uacucaacuu gaaaaggugg 60

caccgauu 68

<210> SEQ ID NO 46

<211> LENGTH: 63

<212> TYPE: RNA

<213> ORGANISM: Streptococcus thermophilus

-continued

<400> SEQUENCE: 46

gggcgaaaca acacagcgag uaaaaaag gcuuaguccg uacucaacuu gaaaaggugg 60

cac 63

<210> SEQ ID NO 47

<211> LENGTH: 58

<212> TYPE: RNA

<213> ORGANISM: Streptococcus thermophilus

<400> SEQUENCE: 47

gggcgaaaca acacagcgag uaaaaaag gcuuaguccg uacucaacuu gaaaaggu 58

<210> SEQ ID NO 48

<211> LENGTH: 53

<212> TYPE: RNA

<213> ORGANISM: Streptococcus thermophilus

<400> SEQUENCE: 48

gggcgaaaca acacagcgag uaaaaaag gcuuaguccg uacucaacuu gaa 53

<210> SEQ ID NO 49

<211> LENGTH: 48

<212> TYPE: RNA

<213> ORGANISM: Streptococcus thermophilus

<400> SEQUENCE: 49

gggcgaaaca acacagcgag uaaaaaag gcuuaguccg uacucaac 48

<210> SEQ ID NO 50

<211> LENGTH: 36

<212> TYPE: DNA

<213> ORGANISM: Streptococcus thermophilus

<400> SEQUENCE: 50

gttttagagc tgtgtgtgtt cgaatgggtc caaaac 36

<210> SEQ ID NO 51

<211> LENGTH: 36

<212> TYPE: DNA

<213> ORGANISM: Streptococcus pyrogenes

<400> SEQUENCE: 51

gttttagagc tatgtgtgtt tgaatgggtc caaaac 36

<210> SEQ ID NO 52

<211> LENGTH: 36

<212> TYPE: DNA

<213> ORGANISM: Streptococcus thermophilus

<400> SEQUENCE: 52

tttaactcgc tgtgtgtgtt cgaatgggtt caaacc 36

<210> SEQ ID NO 53

<211> LENGTH: 36

<212> TYPE: DNA

<213> ORGANISM: Streptococcus pyrogenes

<400> SEQUENCE: 53

tttaacttgc tatgtgtgtt tgaatgggtc caacaa 36

-continued

<210> SEQ ID NO 54
<211> LENGTH: 42
<212> TYPE: RNA
<213> ORGANISM: Streptococcus pyrogenes

<400> SEQUENCE: 54

gauuucuucu ugcgcuuuuu guuuuagagc uaugcuguuu ug 42

<210> SEQ ID NO 55
<211> LENGTH: 42
<212> TYPE: RNA
<213> ORGANISM: Streptococcus thermophilus

<400> SEQUENCE: 55

guucacugua cgaguacuua guuuuagagc uguguuguuu cg 42

<210> SEQ ID NO 56
<211> LENGTH: 42
<212> TYPE: RNA
<213> ORGANISM: Streptococcus thermophilus

<400> SEQUENCE: 56

cgcuaaagag gaagaggaca guuuuagagc uguguuguuu cg 42

<210> SEQ ID NO 57
<211> LENGTH: 35
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic oligonucleotide"

<400> SEQUENCE: 57

aaattctaaa cgctaaagag gaagaggaca tgggtg 35

<210> SEQ ID NO 58
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic oligonucleotide"

<400> SEQUENCE: 58

aggaagagga catggtga 18

<210> SEQ ID NO 59
<211> LENGTH: 10
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic oligonucleotide"

<400> SEQUENCE: 59

aggaagagga 10

<210> SEQ ID NO 60
<211> LENGTH: 10

-continued

<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic oligonucleotide"

<400> SEQUENCE: 60

tacatggtga 10

<210> SEQ ID NO 61
<211> LENGTH: 35
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic oligonucleotide"

<400> SEQUENCE: 61

caccatgtcc tcttcctctt tagcgtttag aattt 35

<210> SEQ ID NO 62
<211> LENGTH: 32
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic oligonucleotide"

<400> SEQUENCE: 62

caccatgtcc tcttcctctt tagcgtttag aa 32

<210> SEQ ID NO 63
<211> LENGTH: 28
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic oligonucleotide"

<400> SEQUENCE: 63

caccatgtcc tcttcctctt tagcgttt 28

<210> SEQ ID NO 64
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic oligonucleotide"

<400> SEQUENCE: 64

caccatgtcc tcttcctctt tagc 24

<210> SEQ ID NO 65
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic oligonucleotide"

-continued

```

<400> SEQUENCE: 65

caccatgtcc tcttcctctt                                20

<210> SEQ ID NO 66
<211> LENGTH: 16
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
        Synthetic oligonucleotide"

<400> SEQUENCE: 66

caccatgtcc tcttcc                                      16

<210> SEQ ID NO 67
<211> LENGTH: 55
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
        Synthetic oligonucleotide"

<400> SEQUENCE: 67

attacgaatt ctctctgtcc tcttcctctt tagcgtttag aatttcaatt cgagc        55

<210> SEQ ID NO 68
<211> LENGTH: 55
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
        Synthetic oligonucleotide"

<400> SEQUENCE: 68

attacgaatt gtggtgtgcc tcttcctctt tagcgtttag aatttcaatt cgagc        55

<210> SEQ ID NO 69
<211> LENGTH: 42
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
        Synthetic oligonucleotide"

<400> SEQUENCE: 69

gcuaaagagg aagaggacag uuuuagagcu guguuguuuc ga                                42

```

1-78. (canceled)

79. A Cas9-crRNA complex comprising:

- a. a Cas9 protein;
 - b. a tracrRNA transcript; and
 - c. a CRISPR RNA transcript that is engineered to guide the Cas9-crRNA complex to a target DNA sequence for site specific modification of the target DNA sequence,
- wherein (a), (b), and (c) are assembled in vitro.

80-113. (canceled)

114. The Cas9-crRNA complex of claim **79**, wherein the crRNA comprises an approximately 20 nucleotide fragment complementary to a proto-spacer sequence in the target DNA sequence.

115. The Cas9-crRNA complex of claim **79**, wherein the target DNA sequence is a plasmid DNA.

116. The Cas9-crRNA complex of claim **79**, wherein the target DNA sequence is a human genomic DNA.

117. The Cas9-crRNA complex of claim **79**, wherein the Cas9 protein comprises a fusion polypeptide comprising at least one additional amino acid sequence.

118. The Cas9-crRNA complex of claim **117**, wherein the additional amino acid sequence is used for purification of the fusion polypeptide by affinity chromatography.

119. The Cas9-crRNA complex of claim **79**, wherein the Cas9 protein contains a point mutation in the RuvC active site motif.

120. The Cas9-crRNA complex of claim **119**, wherein the point mutation is D31A.

121. The Cas9-crRNA complex of claim **79**, wherein the Cas9 protein contains a point mutation in the HNH active site motif.

122. The Cas9-crRNA complex of claim **121**, wherein the point mutation is N891A.

* * * * *