



(51) International Patent Classification:

C12N 9/22 (2006.01) C12N 15/09 (2006.01)

TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

(21) International Application Number:

PCT/SG2017/050523

Published:

- with international search report (Art. 21(3))
- with sequence listing part of description (Rule 5.2(a))

(22) International Filing Date:

17 October 2017 (17.10.2017)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

10201608681P 17 October 2016 (17.10.2016) SG

(71) Applicants: **NANYANG TECHNOLOGICAL UNIVERSITY** [SG/SG]; 50 Nanyang Avenue, Singapore 639798 (SG). **AGENCY FOR SCIENCE, TECHNOLOGY AND RESEARCH** [SG/SG]; 1 Fusionopolis Way, #20-10, Connexis North Tower, Singapore 138632 (SG).

(72) Inventors: **TAN, Meng How**; c/o Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798 (SG). **OOL, Kean Hean**; c/o Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798 (SG).

(74) Agent: **VIERING, JENTSCHURA & PARTNER LLP**; P.O Box 1088, Rochor Post Office, Rochor Road, Singapore 911833 (SG).

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

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM,

(54) Title: TRUNCATED CRISPR-CAS PROTEINS FOR DNA TARGETING

(57) Abstract: The present invention relates to a polypeptide comprising at least one at least one deletion selected from the group consisting of ΔHNH (Δ775-909), ARuvCIII-b (Δ1002-1074), AREC1-a (Δ510-655), AREC1-b (Δ525-587), AREC1-c (Δ662-710), AREC2 (Δ180-308), AREC2-a (Δ212-244), AREC2-b (Δ244-276), AREC2-c (Δ276-308), AREC2-d (Δ199-283), AREC2-e (Δ198-257), AREC2-f (Δ235-286), AREC2-g (Δ217-266), AREC3 (Δ498-712) and combinations thereof, wherein the position numbering is in accordance with SEQ ID NO: 1 encoding for *S. pyogenes* Cas9, and wherein the polypeptide has CRISPR-Cas DNA-binding activity. The polypeptide may further comprises a missense mutations selected from G12R, T13K, T13R, N14K, N497K, T657K, T657R, N767K, T770K, T770R, Q920K, Q920R, S1109R, D1135K, D1135R, S1338R and combinations thereof. Also claimed are nucleic acid molecules encoding for said polypeptides, compositions and method of site-directed engineering of a target DNA thereof.



5

TRUNCATED CRISPR-CAS PROTEINS FOR DNA TARGETING**CROSS-REFERENCE TO RELATED APPLICATION**

10

This application makes reference to and claims the benefit of priority of the Singapore Patent Application No. 10201608681P filed on 17 October 2016, the content of which is incorporated herein by reference for all purposes, including an incorporation of any element or part of the description, claims or drawings not contained herein and referred to in Rule 20.5(a) of the PCT, pursuant to Rule 4.18 of the PCT.

15

FIELD OF THE INVENTION

The present invention generally relates to engineered and optimized CRISPR-Cas proteins and their applications in the site-specific modification of a target DNA.

20

BACKGROUND OF THE INVENTION

The ability to engineer the genome of any living organism has many biomedical and biotechnological applications, such as the correction of disease-causing mutations, construction of accurate cellular models for disease studies, or generation of agricultural crops with desirable traits. Various technologies have been developed over the past 10-20 years, including meganucleases, zinc finger nucleases, or transcription activator-like effector nucleases (TALENs). However, these technologies are either difficult to program or difficult to assemble, thereby limiting their widespread adoption.

30

Recently, the CRISPR (clustered regularly interspaced short palindromic repeats)-Cas system has emerged as a powerful new genome engineering technology. Originally an adaptive immune system in bacteria, it has been successfully deployed for genome modification in animals and plants, including human. A key reason why CRISPR-Cas has gained so much popularity in such a short time is that the element that targets the Cas endonuclease to the correct location in the genome is simply a short piece of guide RNA, which is straightforward to design and cheap to synthesize.

35

The Cas enzyme performs two important and distinct functions. First, it is able to bind to specific loci in the genome with the aid of the guide RNA and a short sequence in the targeted DNA, called the protospacer adjacent motif (PAM). Cas proteins from different bacteria will recognize different PAMs in general. Various protein domains are involved in interacting with the DNA, including the REC1, REC2, and PI domains. Second, the Cas enzyme can cleave both strands

40

of the DNA upon binding, thereby generating a double-stranded break. The Cas protein contains two nuclease domains, the HNH and RuvC domains, each of which cut opposite strands of the DNA. The overall domain architecture of the Cas9 enzyme from *Streptococcus pyogenes* is shown in Figure 1.

Since the two functions of the Cas enzyme may be separated, it is possible to utilize the CRISPR-Cas system only for the purpose of DNA targeting and not for DNA cleavage. To disable the cleavage function, the common practice is to mutate a catalytic residue in each of the two nuclease domains (D10A and H840A in SpCas9). Various laboratories have successfully fused different effector domains to the catalytically dead Cas enzyme (dCas) for a range of applications, including gene regulation, imaging, and base conversion.

Despite the tremendous potential of CRISPR-Cas for genome engineering, the system suffers from several shortcomings that restricts its use for basic biomedical research and therapeutic applications. In particular, the large size of the Cas protein affects the delivery of the CRISPR-Cas system both in vitro and even more so in vivo. For the purpose of gene therapy in vivo, adeno-associated virus (AAV) vectors are extensively used because the virus can transduce many different cell types, appears to elicit little immune response, and reliably and permanently inserts its gene into a safe genomic locus on human chromosome 19. However, the packaging limit of AAV vectors is around 4.5 kb, excluding the inverted terminal repeats, and this is close to the size of the commonly used SpCas9, which is around 4.2 kb. As a result, there is little room for customization and addition of other genetic elements. Notably, it is currently not possible to package dSpCas9 with a fused effector domain into an AAV. Hence, there is an urgent need to explore alternative strategies to utilize the CRISPR-Cas system for in vivo genome engineering.

SUMMARY OF THE INVENTION

In a first aspect, the invention relates to a polypeptide having at least 50% sequence identity to *Streptococcus pyogenes* Cas9 having the amino acid sequence set forth in SEQ ID NO:1 over its entire length, wherein said polypeptide comprises at least one deletion selected from the group consisting of Δ HNH (Δ 775-909), Δ RuvCIII-b (Δ 1002-1074), Δ REC1-a (Δ 510-655), Δ REC1-b (Δ 525-587), Δ REC1-c (Δ 662-710), Δ REC2 (Δ 180-308), Δ REC2-a (Δ 212-244), Δ REC2-b (Δ 244-276), Δ REC2-c (Δ 276-308), Δ REC2-d (Δ 199-283), Δ REC2-e (Δ 198-257), Δ REC2-f (Δ 235-286), Δ REC2-g (Δ 217-266), Δ REC3 (Δ 498-712) and combinations thereof, wherein the position numbering is in accordance with SEQ ID NO:1, and wherein the polypeptide has CRISPR-Cas DNA-binding activity.

5

In various embodiments, the polypeptide comprises at least one combination of deletions selected from the group consisting of Δ REC1-a Δ HNH (Δ 510-655 Δ 775-909), Δ REC1-c Δ HNH (Δ 662-710 Δ 775-909), Δ REC2 Δ HNH (Δ 180-308 Δ 775-909), Δ REC2-d Δ HNH (Δ 199-283 Δ 775-909), Δ HNH Δ RuvCIII-b (Δ 775-909 Δ 1002-1074), Δ REC3 Δ HNH (Δ 498-712 Δ 775-909), Δ REC3 Δ HNH Δ RuvCIII-b (Δ 498-712 Δ 775-909 Δ 1002-1074), Δ REC2 Δ REC3 Δ HNH Δ RuvCIII-b (Δ 180-308 Δ 498-712 Δ 775-909 Δ 1002-1074), and Δ REC2 Δ HNH Δ RuvCIII-b (Δ 180-308 Δ 775-909 Δ 1002-1074).

10

15

In various embodiments, the polypeptide further comprises at least one missense mutation selected from the group consisting of G12R, T13K, T13R, N14K, N497K, T657K, T657R, N767K, T770K, T770R, Q920K, Q920R, S1109R, D1135K, D1135R, S1338R and combinations thereof.

20

25

In various embodiments, the polypeptide comprises at least one combination of modifications selected from the group consisting of Δ REC1-c Δ HNH G12R, Δ REC1-c Δ HNH T13K, Δ REC1-c Δ HNH T13R, Δ REC1-c Δ HNH N14K, Δ REC1-c Δ HNH N497K, Δ REC1-c Δ HNH T657K, Δ REC1-c Δ HNH T657R, Δ REC1-c Δ HNH N767K, Δ REC1-c Δ HNH T770K, Δ REC1-c Δ HNH T770R, Δ REC1-c Δ HNH Q920K, Δ REC1-c Δ HNH Q920R, Δ REC1-c Δ HNH S1109R, Δ REC1-c Δ HNH D1135K, Δ REC1-c Δ HNH D1135R, Δ REC1-c Δ HNH S1338R, Δ REC1-c Δ HNH T657R T13K, Δ REC1-c Δ HNH T657R N497K, Δ REC1-c Δ HNH T657R T770K, Δ REC1-c Δ HNH T657R Q920K, Δ REC1-c Δ HNH T657R S1109R, Δ REC1-c Δ HNH T657R D1135K, Δ REC2 Δ HNH Δ RuvCIII-b T13K, Δ REC2 Δ HNH Δ RuvCIII-b T657K, Δ REC2 Δ HNH Δ RuvCIII-b T657R, Δ REC2 Δ HNH Δ RuvCIII-b 770K, Δ REC2 Δ HNH Δ RuvCIII-b Q920K, Δ REC2 Δ HNH Δ RuvCIII-b S1109R, and Δ REC2 Δ HNH Δ RuvCIII-b D1135K.

30

In preferred embodiments, the polypeptide has the amino acid sequence set forth in any one of SEQ ID NOs:14-30.

35

In various embodiments, the polypeptide is further codon optimized for expression in a eukaryotic cell, preferably in a mammalian or human cell.

40

In a second aspect, the invention relates to a polypeptide having at least 50% sequence identity to *Staphylococcus aureus* Cas9 having the amino acid sequence set forth in SEQ ID NO:2, *Neisseria meningitidis* Cas9 having the amino acid sequence set forth in SEQ ID NO:3, *Acidaminococcus* sp. Cpf1 having the amino acid sequence set forth in SEQ ID NO:4, *Streptococcus thermophilus* Cas9 having the amino acid sequence set forth in SEQ ID NO:5,

- 5 *Sutterella wadsworthensis* Cas9 having the amino acid sequence set forth in SEQ ID NO:6, *Filifactor alocis* Cas9 having the amino acid sequence set forth in SEQ ID NO:7, *Lactobacillus johnsonii* Cas9 having the amino acid sequence set forth in SEQ ID NO:8, *Campylobacter lari* Cas9 having the amino acid sequence set forth in SEQ ID NO:9, *Parvibaculum lavamentivorans* Cas9 having the amino acid sequence set forth in SEQ ID NO:10, *Mycoplasma gallisepticum*
- 10 Cas9 having the amino acid sequence set forth in SEQ ID NO:11, *Treponema denticola* Cas9 having the amino acid sequence set forth in SEQ ID NO:12, or *Lachnospiraceae* bacterium ND2006 Cpf1 having the amino acid sequence set forth in SEQ ID NO:13 over its entire length, wherein said polypeptide comprises at least one deletion of said amino acid sequence, wherein said at least one deletion corresponds to at least one deletion of SEQ ID NO:1 selected from the
- 15 group consisting of Δ HNH (Δ 775-909), Δ RuvCIII-b (Δ 1002-1074), Δ REC1-a (Δ 510-655), Δ REC1-b (Δ 525-587), Δ REC1-c (Δ 662-710), Δ REC2 (Δ 180-308), Δ REC2-a (Δ 212-244), Δ REC2-b (Δ 244-276), Δ REC2-c (Δ 276-308), Δ REC2-d (Δ 199-283), Δ REC2-e (Δ 198-257), Δ REC2-f (Δ 235-286), Δ REC2-g (Δ 217-266), Δ REC3 (Δ 498-712) and combinations thereof, wherein the position numbering is in accordance with SEQ ID NO:1, wherein the polypeptide has CRISPR-Cas DNA-
- 20 binding activity.

In various embodiments, the polypeptide comprises at least one combination of deletions selected from the group consisting of Δ REC1-a Δ HNH (Δ 510-655 Δ 775-909), Δ REC1-c Δ HNH (Δ 662-710 Δ 775-909), Δ REC2 Δ HNH (Δ 180-308 Δ 775-909), Δ REC2-d Δ HNH (Δ 199-283 Δ 775-909), Δ HNH Δ RuvCIII-b (Δ 775-909 Δ 1002-1074), Δ REC3 Δ HNH (Δ 498-712 Δ 775-909), Δ REC3 Δ HNH Δ RuvCIII-b (Δ 498-712 Δ 775-909 Δ 1002-1074), Δ REC2 Δ REC3 Δ HNH Δ RuvCIII-b (Δ 180-308 Δ 498-712 Δ 775-909 Δ 1002-1074), and Δ REC2 Δ HNH Δ RuvCIII-b (Δ 180-308 Δ 775-909 Δ 1002-1074).

- 30 In various embodiments, the polypeptide further comprises at least one missense mutation selected from the group consisting of G12R, T13K, T13R, N14K, N497K, T657K, T657R, N767K, T770K, T770R, Q920K, Q920R, S1109R, D1135K, D1135R, S1338R and combinations thereof, wherein the position numbering is in accordance with SEQ ID NO:1.

- 35 In various embodiments, the polypeptide comprises at least one combination of modifications selected from the group consisting of Δ REC1-c Δ HNH G12R, Δ REC1-c Δ HNH T13K, Δ REC1-c Δ HNH T13R, Δ REC1-c Δ HNH N14K, Δ REC1-c Δ HNH N497K, Δ REC1-c Δ HNH T657K, Δ REC1-c Δ HNH T657R, Δ REC1-c Δ HNH N767K, Δ REC1-c Δ HNH T770K, Δ REC1-c Δ HNH T770R, Δ REC1-c Δ HNH Q920K, Δ REC1-c Δ HNH Q920R, Δ REC1-c Δ HNH S1109R, Δ REC1-c Δ HNH D1135K, Δ REC1-c Δ HNH D1135R, Δ REC1-c Δ HNH S1338R, Δ REC1-c Δ HNH T657R T13K, Δ REC1-c Δ HNH T657R N497K, Δ REC1-c Δ HNH T657R T770K, Δ REC1-c Δ HNH T657R
- 40

5 Q920K, Δ REC1-c Δ HNH T657R S1109R, Δ REC1-c Δ HNH T657R D1135K, Δ REC2 Δ HNH Δ RuvCIII-b T13K, Δ REC2 Δ HNH Δ RuvCIII-b T657K, Δ REC2 Δ HNH Δ RuvCIII-b T657R, Δ REC2 Δ HNH Δ RuvCIII-b 770K, Δ REC2 Δ HNH Δ RuvCIII-b Q920K, Δ REC2 Δ HNH Δ RuvCIII-b S1109R, and Δ REC2 Δ HNH Δ RuvCIII-b D1135K.

10 In various embodiments, the polypeptide is further codon optimized for expression in a eukaryotic cell, preferably in a mammalian or human cell, by any techniques known in the art.

In a third aspect, the invention relates to a nucleic acid molecule comprising or consisting of a nucleic acid encoding a polypeptide described herein.

15

In various embodiments, the nucleic acid molecule is a recombinant expression vector.

In various embodiments, the vector further comprises at least one regulatory element for controlling expression of the polypeptide.

20

In a fourth aspect, the invention relates to a composition comprising the polypeptide or nucleic acid molecule described herein.

In a fifth aspect, the invention relates to a method of site-specific engineering of a target DNA, the method comprising contacting the target DNA with a CRISPR-Cas system comprising the polypeptide, nucleic acid molecule or composition described herein, wherein the CRISPR-Cas DNA-binding activity of the polypeptide recruits an effector domain to the target DNA. The effector domain includes but is not limited to transcriptional regulatory domains, histone-modifying domains, chromatin remodelers, fluorescent proteins, and deaminases.

30

In various embodiments, the target DNA is extrachromosomal or is part of a chromosome in vitro, in vivo, or in a cell.

In various embodiments, the CRISPR-Cas nuclease activity of the polypeptide introduces a single-stranded nick or double stranded break in the target DNA.

35

In a sixth aspect, the invention relates to use of the polypeptide, nucleic acid molecule or composition described herein in site-directed nucleic acid modification, preferably as a component in a CRISPR-Cas system.

40

5

BRIEF DESCRIPTION OF THE DRAWINGS

The invention will be better understood with reference to the detailed description when considered in conjunction with the non-limiting examples and the accompanying drawings.

10 Figure 1 shows the schematic illustration of the different domains of *Streptococcus pyogenes* Cas9 (SpCas9).

Figure 2 shows the schematic illustration of Δ RuvCIII, Δ RuvCIII-a, and Δ RuvCIII-b.

15 Figure 3 shows the schematic illustration of Δ REC1-a, Δ REC1-b, Δ REC1-c, and Δ REC3.

Figure 4 shows the schematic illustration of Δ REC2, Δ REC2-a, Δ REC2-b, Δ REC2-c, Δ REC2-d, Δ REC2-e, Δ REC-f, and Δ REC2-g.

20 Figure 5 shows the schematic illustration of Δ PI-a, Δ PI-b, Δ PI-c, Δ PI-d, Δ PI-e, and Δ PI-f.

Figure 6 shows the results of reporter assays of 41 different truncated constructs, with deletions in various parts of the REC or NUC lobe. (a) A minimal promoter containing two binding sites for a dSpCas9-VPR activator was used and the cells were analysed 24 hours after transfection. (b)
25 A minimal promoter containing two binding sites was used and the cells were analysed 48 hours after transfection. (c) A minimal promoter containing three binding sites was used and the cells were analysed 24 hours after transfection.

Figure 7 shows a boxplot that summarizes the results of reporter assays of different truncated
30 constructs. 2x indicates that a fluorescence reporter with two binding sites for the dCas9-VPR enzyme was used, while 3x indicates that a fluorescence reporter with three binding sites was used. 24hr and 48hr indicates the time after transfection when the assays were performed.

Figure 8 shows the results of quantitative real-time PCR (qRT-PCR) experiments to measure
35 the expression levels of (a) ACTC, (b) ASCL1, (c) MIAT, and (d) TTN in HEK293 cells, when the cells were transfected with different truncated constructs and a relevant guide RNA targeting the endogenous gene-of-interest. In these experiments, the cells were harvested 48 hours after transfection.

40 Figure 9 shows the results of experiments evaluating the effect of deleting the REC3 domain, either by itself or in combination with other domains. (a) Reporter assays of various REC3 truncations in comparison with the original full-length protein (WT) and three other truncations

- 5 (ΔHNH, ΔRuvCIII-b ΔHNH, and ΔREC1-c) that we had found to perform similarly to wildtype. (b) Expression levels of either MIAT or TTN in HEK293 cells measured by qRT-PCR, when the cells were transfected with different REC3 truncated constructs and a corresponding gRNA targeting the gene-of-interest.
- 10 Figure 10 shows a ribbon diagram of dSpCas9 bound to a gRNA and a target DNA. The residues that are within 5Å of the DNA are represented as sticks.
- Figure 11 shows the reporter assays of constructs, each of which has 1 mutation introduced into deletions of HNH and REC1-c ("2ple").
- 15 Figure 12 shows the reporter assays of constructs, each of which has 2 different mutations introduced into deletions of HNH and REC1-c ("2ple").
- Figure 13 shows the reporter assays of constructs, each of which has 1 mutation introduced into HNH, REC2, and RuvCIII-b ("3ple").
- 20

DETAILED DESCRIPTION OF THE INVENTION

25 The following detailed description refers to, by way of illustration, specific details and embodiments in which the invention may be practiced. These embodiments are described in sufficient detail to enable those skilled in the art to practice the invention. Other embodiments may be utilized and structural, and logical changes may be made without departing from the scope of the invention. The various embodiments are not necessarily mutually exclusive, as some embodiments can be combined with one or more other embodiments to form new

30 embodiments.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art. The singular terms "a," "an," and "the" include plural referents unless context clearly indicates otherwise. Similarly, the word "or"

35 is intended to include "and" unless the context clearly indicates otherwise. The term "comprises" means "includes." In case of conflict, the present specification, including explanations of terms, will control.

In a first aspect, the invention relates to a polypeptide having at least 50%, 51%, 52%, 53%,

40 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 99.5%

5 sequence identity to *Streptococcus pyogenes* Cas9 having the amino acid sequence set forth in SEQ ID NO:1 over its entire length, wherein said polypeptide comprises at least one deletion selected from the group consisting of Δ HNH (Δ 775-909), Δ RuvCIII-b (Δ 1002-1074), Δ REC1-a (Δ 510-655), Δ REC1-b (Δ 525-587), Δ REC1-c (Δ 662-710), Δ REC2 (Δ 180-308), Δ REC2-a (Δ 212-244), Δ REC2-b (Δ 244-276), Δ REC2-c (Δ 276-308), Δ REC2-d (Δ 199-283), Δ REC2-e (Δ 198-257),
 10 Δ REC2-f (Δ 235-286), Δ REC2-g (Δ 217-266), Δ REC3 (Δ 498-712) and combinations thereof, wherein the position numbering is in accordance with SEQ ID NO:1, and wherein the polypeptide has CRISPR-Cas DNA-binding activity.

15 The term "polypeptide" is used interchangeably herein with "peptide" and "protein", and refers to polymers of at least two amino acids connected by peptide bonds. The polymer may comprise amino acid analogues or modified amino acids, it may be linear or branched, and it may be interrupted by non-amino acids. The term also encompasses an amino acid polymer that has been modified naturally or artificially; for example, by disulfide bond formation, glycosylation, lipidation, acetylation, phosphorylation, or any other manipulation, such as conjugation to a
 20 labeling moiety. However, in preferred embodiments, this term relates to polymers of naturally occurring amino acids, as defined below, which may optionally be modified as defined above, but does not comprise non-amino acid moieties in the polymer backbone.

25 The term "amino acid" as used herein refers to natural and/or unnatural or synthetic amino acids, including both the D and L optical isomers, amino acid analogs (for example norleucine is an analog of leucine) and derivatives known in the art. The term "naturally occurring amino acid", as used herein, relates to the 20 naturally occurring L-amino acids, namely Gly (G), Ala (A), Val (V), Leu (L), Ile (I), Phe (F), Cys (C), Met (M), Pro (P), Thr (T), Ser (S), Glu (E), Gln (Q), Asp (D), Asn (N), His (H), Lys (K), Arg (R), Tyr (Y), and Trp (W). Generally, in the context of the
 30 present application, the peptides and polypeptides are shown in the N- to C-terminal orientation.

The term "sequence identity" as used herein refers to the relatedness between two amino acid sequences or between two nucleotide sequences. For purposes of the present invention, the sequence identity between two amino acid sequences is determined using the Needleman-Wunsch algorithm (Needleman and Wunsch, 1970, *J. Mol. Biol.* 48: 443-453) as implemented in
 35 the Needle program of the EMBOSS package (EMBOSS: The European Molecular Biology Open Software Suite, Rice *et al.*, 2000, *Trends Genet.* 16: 276-277), preferably version 5.0.0 or later. The parameters used are gap open penalty of 10, gap extension penalty of 0.5, and the EBLOSUM62 (EMBOSS version of BLOSUM62) substitution matrix. The output of Needle
 40 labeled "longest identity" (obtained using the *-nobrief* option) is used as the percent identity and is calculated as follows: (Identical Residues x 100)/(Length of Alignment – Total Number of Gaps

5 in Alignment).

The term "deletion" as used herein refers to a removal of a portion of one or more amino acid residues from the amino acid sequence of the reference polypeptide.

10 In the context of the present application, the polypeptide is truncated compared to the reference amino acid sequence in that it comprises one or more deletions described herein. A polypeptide having a deletion denoted as $\Delta a-b$ is one whose amino acid residues of positions $a-b$ inclusive are deleted, while a first fragment of amino acid residues 1 to $(a-1)$ and a second fragment of amino acid residues $(b+1)$ to the C-terminus are joined together by a peptide bond
15 via an optional linker. The polypeptide has a length of at least 500 aa, preferably up to 600, 700, 800, 900, 1000, 1100, 1200, 1300, or 1360 aa.

In the context of the present application, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, or 14 of the deletions ΔHNH ($\Delta 775-909$), $\Delta RuvCIII-b$ ($\Delta 1002-1074$), $\Delta REC1-a$ ($\Delta 510-655$), $\Delta REC1-b$ ($\Delta 525-587$), $\Delta REC1-c$ ($\Delta 662-710$), $\Delta REC2$ ($\Delta 180-308$), $\Delta REC2-a$ ($\Delta 212-244$), $\Delta REC2-b$ ($\Delta 244-276$), $\Delta REC2-c$ ($\Delta 276-308$), $\Delta REC2-d$ ($\Delta 199-283$), $\Delta REC2-e$ ($\Delta 198-257$), $\Delta REC2-f$ ($\Delta 235-286$), $\Delta REC2-g$ ($\Delta 217-266$), and $\Delta REC3$ ($\Delta 498-712$) are present in the polypeptide described herein, meaning that any one or more of the deletions described herein may be present in the polypeptide individually or in different combinations.

25 In various embodiments, the polypeptide comprises at least one combination of deletions selected from the group consisting of $\Delta REC1-a$ ΔHNH ($\Delta 510-655$ $\Delta 775-909$), $\Delta REC1-c$ ΔHNH ($\Delta 662-710$ $\Delta 775-909$), $\Delta REC2$ ΔHNH ($\Delta 180-308$ $\Delta 775-909$), $\Delta REC2-d$ ΔHNH ($\Delta 199-283$ $\Delta 775-909$), ΔHNH $\Delta RuvCIII-b$ ($\Delta 775-909$ $\Delta 1002-1074$), $\Delta REC3$ ΔHNH ($\Delta 498-712$ $\Delta 775-909$), $\Delta REC3$ ΔHNH $\Delta RuvCIII-b$ ($\Delta 498-712$ $\Delta 775-909$ $\Delta 1002-1074$), $\Delta REC2$ $\Delta REC3$ ΔHNH $\Delta RuvCIII-b$ ($\Delta 180-308$ $\Delta 498-712$ $\Delta 775-909$ $\Delta 1002-1074$), and $\Delta REC2$ ΔHNH $\Delta RuvCIII-b$ ($\Delta 180-308$ $\Delta 775-909$ $\Delta 1002-1074$).

35 In various embodiments, the polypeptide further comprises at least one missense mutation selected from the group consisting of G12R, T13K, T13R, N14K, N497K, T657K, T657R, N767K, T770K, T770R, Q920K, Q920R, S1109R, D1135K, D1135R, S1338R and combinations thereof. These mutations are intended to introduce positively charged amino acids (lysine or arginine) to enhance the binding between the polypeptide and the target DNA. They were selected based on their location in poorly conserved regions of *Streptococcus pyogenes* Cas9,
40 which are unlikely to be important residues for the nuclease, as well as their proximity to the target DNA as revealed by computational simulations.

5

In various embodiments, the polypeptide comprises at least one combination of modifications selected from the group consisting of Δ REC1-c Δ HNH G12R, Δ REC1-c Δ HNH T13K, Δ REC1-c Δ HNH T13R, Δ REC1-c Δ HNH N14K, Δ REC1-c Δ HNH N497K, Δ REC1-c Δ HNH T657K, Δ REC1-c Δ HNH T657R, Δ REC1-c Δ HNH N767K, Δ REC1-c Δ HNH T770K, Δ REC1-c Δ HNH T770R, Δ REC1-c Δ HNH Q920K, Δ REC1-c Δ HNH Q920R, Δ REC1-c Δ HNH S1109R, Δ REC1-c Δ HNH D1135K, Δ REC1-c Δ HNH D1135R, Δ REC1-c Δ HNH S1338R, Δ REC1-c Δ HNH T657R T13K, Δ REC1-c Δ HNH T657R N497K, Δ REC1-c Δ HNH T657R T770K, Δ REC1-c Δ HNH T657R Q920K, Δ REC1-c Δ HNH T657R S1109R, Δ REC1-c Δ HNH T657R D1135K, Δ REC2 Δ HNH Δ RuvCIII-b T13K, Δ REC2 Δ HNH Δ RuvCIII-b T657K, Δ REC2 Δ HNH Δ RuvCIII-b T657R, Δ REC2 Δ HNH Δ RuvCIII-b 770K, Δ REC2 Δ HNH Δ RuvCIII-b Q920K, Δ REC2 Δ HNH Δ RuvCIII-b S1109R, and Δ REC2 Δ HNH Δ RuvCIII-b D1135K.

10

15

In preferred embodiments, the polypeptide has the amino acid sequence set forth in any one of SEQ ID NOs:14-30.

20 Table 1. Preferred polypeptides (catalytically dead) in accordance with the first aspect of the invention.

Name	SEQ ID NO	Amino acid sequence
Δ HNH	14	MDKKYSIGLA IGTVSVGWAV ITDEYKVPSK KFKVLGNTDR HSIKKNLIGA LLFDSGETAE ATRLKRTARR RYTRRKNRIC YLQEIFSNEK AKVDDSFHHR LEESFLVEED KKHERHPIFG NIVDEVAYHE KYPTIYHLRK KLVDDSTKAD LRLIYLALAH MIKFRGHFLI EGDLPDNDSD VDKLFIQLVQ TYNQLFEENP INASGVDAKA ILSARLSKSR RLENLIAQLP GEKKNGLFGN LIALSLGLTP NFKSNFDLAE DAKLQLSKDT YDDDLNLLA QIGDQYADLF LAAKNLSDAI LLSDILRVNT EITKAPLSAS MIKRYDEHHQ DLTLLKALVR QQLPEKYKEI FFDQSKNGYA GYIDGGASQE EFKYFKIPIL EKMDGTEELL VKLNREDLLR KQRTFDNGSI PHQIHLGELH AILRRQEDFY PFLKDNREKI EKILTFRIPY YVGPLARGNS RFAWMTRKSE ETITPWNFEE VVDKGASAQS FIERMTNFDK NLPNEKVLPK HSLLEYEFTV YNELTKVKYV TEGMRKPAFL SGEQKKAIVD LLFKTNRKVT VKQLKEDYFK KIECFDSVEI SGVEDRFNAS LGTYHDLKI IKDKDFLDNE ENEDILEDIV LTLTLFEDRE MIEERLKYA HLFDDKVMKQ LKRRRYTGWG RLSRKLINGI RDKQSGKTIL DFLKSDGFAN RNFMQLIHDD SLTFKEDIQK AQVSGQGDSL HEHIANLAGS PAIKKGILQT VKVVDLVKVV MGRHKPENIV IEMARENQTT QKGQKGGGSE LDKAGFIKQ LVETRQITKH VAQILDSRMN TKYDENDKLI REVKVITLKS KLVSDFRKDF QFYKVREINN YHHAHDAYLN AVVGTALIKK YPKLESEFVY GDYKVYDVRK MIAKSEQEIG KATAKYFFYS NIMNFFKTEI TLANGEIRKR PLIETNGETG EIVWDKGRDF ATVRKVLSMP QVNIVKKTEV QTGGFSKESI LPKRNSDKLI ARKKDWDPPK YGGFDSPTVA YSVLVVAKVE KGKSKKLKSV KELLGITIME RSSFEKNPID FLEAKGYKEV KKDIIKLPK YSLFELENGR KRMLASAGEL QKGNELALPS KYVNFYLYAS HYEKLKGSPE DNEQKQLFVE QHKHYLDEII EQISEFSKRV ILADANLDKV LSAYNKHHRDK PIREQAENII HLFTLTNLGA PAAFKYFDTT

		IDRKRYTSTK EVLDATLIHQ SITGLYETRI DLSQLGGD
Δ RuvCIII-b	15	MDKKYSIGLA IGTNSVGWAV ITDEYKVPSK KFKVLGNTDR HSIKKNLIGA LLFDSGETAE ATRLKRTARR RYTRRNKRIC YLQEIFSNEK AKVDDSFHHR LEESFLVEED KKHERHPIFG NIVDEVAYHE KYPTIYHLRK KLVDSSTKAD LRLIYLALAH MIKFRGHFLI EGDLPDNDSD VDKLFIQLVQ TYNQLFEENP INASGVDAKA ILSARLSKSR RLENLIAQLP GEKKNGLFGN LIALSLGLTP NFKSNFDLAE DAKLQLSKDT YDDDLNLLA QIGDQYADLF LAAKNLSDAI LLSDILRVNT EITKAPLSAS MIKRYDEHHQ DLTLLKALVR QQLPEKYKEI FFDQSKNGYA GYIDGGASQE EFKYFIKPIL EKMDGTEELL VKLNREDLLR KQRTFDNGSI PHQIHLGELH AILRRQEDFY PFLKDNREKI EKILTFRIPY YVGPLARGNS RFAWMTRKSE ETITPWNFEE VVDKGASAQS FIERMTNFDK NLPNEKVLPAK HSLLYEYFTV YNELTKVKYV TEGMRKPAFL SGEQKKAIVD LLFKTNRKVT VKQLKEDYFK KIECFDSVEI SGVEDRFNAS LGTYHDLLKI IKDKDFLDNE ENEDILEDIV LTLTLFEDRE MIEERLKTYA HLFDDKVMKQ LKRRRYTGWG RLSRKLINGI RDKQSGKTIL DFLKSDGFAN RNFMQLIHDD SLTFKEDIQK AQVSGQGDSL HEHIANLAGS PAIKKGILQT VKVVDLVKVV MGRHKPENIV IEMARENQTT QKGQKNSRER MKRIEEGIKE LGSQILKEHP VENTQLQNEK LYLYYLQNGR DMYVDQELDI NRLSDYDVAA IVPQSFLKDD SIDNKVLTSS DKARGKSDNV PSEEVVKMMK NYWRQLLNAK LITQRKFDNL TKAERGGLSE LDKAGFIKRO LVETRQITKH VAQILDSRMN TKYDENDKLI REVKVITLKS KLVSDFRKDF QFYKVRINN YHHAHDAYLN AVVGTALIKK YPGGSGGGGS DKGRDFATVR KVLSPQVNI VKKTEVQTGG FSKESILPKR NSDKLIARKK DWDPKKYGGF DSPTVAYSVL VVAKVEKGS KKLKSVKELL GITIMERSSE EKNPIDFLEA KGYKEVKKDL IIKLPKYSLF ELNGRKRML ASAGELQKGN ELALPSKYVN FLYLASHYK LKGSPEDEQ KQLFVEQHKH YLDEIEEQIS EFSKRVLAD ANLDKVL SAY NKHRDKPIRE QAENIIHLFT LTNLGAPAAF KYFDTTIDRK RYTSTKEVLD ATLIHQ SITG LYETRIDLSQ LGGD
Δ HNH Δ RuvCIII-b	16	MDKKYSIGLA IGTNSVGWAV ITDEYKVPSK KFKVLGNTDR HSIKKNLIGA LLFDSGETAE ATRLKRTARR RYTRRNKRIC YLQEIFSNEK AKVDDSFHHR LEESFLVEED KKHERHPIFG NIVDEVAYHE KYPTIYHLRK KLVDSSTKAD LRLIYLALAH MIKFRGHFLI EGDLPDNDSD VDKLFIQLVQ TYNQLFEENP INASGVDAKA ILSARLSKSR RLENLIAQLP GEKKNGLFGN LIALSLGLTP NFKSNFDLAE DAKLQLSKDT YDDDLNLLA QIGDQYADLF LAAKNLSDAI LLSDILRVNT EITKAPLSAS MIKRYDEHHQ DLTLLKALVR QQLPEKYKEI FFDQSKNGYA GYIDGGASQE EFKYFIKPIL EKMDGTEELL VKLNREDLLR KQRTFDNGSI PHQIHLGELH AILRRQEDFY PFLKDNREKI EKILTFRIPY YVGPLARGNS RFAWMTRKSE ETITPWNFEE VVDKGASAQS FIERMTNFDK NLPNEKVLPAK HSLLYEYFTV YNELTKVKYV TEGMRKPAFL SGEQKKAIVD LLFKTNRKVT VKQLKEDYFK KIECFDSVEI SGVEDRFNAS LGTYHDLLKI IKDKDFLDNE ENEDILEDIV LTLTLFEDRE MIEERLKTYA HLFDDKVMKQ LKRRRYTGWG RLSRKLINGI RDKQSGKTIL DFLKSDGFAN RNFMQLIHDD SLTFKEDIQK AQVSGQGDSL HEHIANLAGS PAIKKGILQT VKVVDLVKVV MGRHKPENIV IEMARENQTT QKGQKGGGSE LDKAGFIKRO LVETRQITKH VAQILDSRMN TKYDENDKLI REVKVITLKS KLVSDFRKDF QFYKVRINN YHHAHDAYLN AVVGTALIKK YPGGSGGGGS DKGRDFATVR KVLSPQVNI VKKTEVQTGG FSKESILPKR NSDKLIARKK DWDPKKYGGF DSPTVAYSVL

		VVAKVEKGKS KKLKSVKELL GITIMERSSE EKNPIDFLEA KGYKEVKKDL IIKLPKYSLF ELENGRKRML ASAGELQKGN ELALPSKYVN FLYLASHYK LKGSPEDEQ KQLFVEQHKH YLDEIEQIS EFSKRVLAD ANLDKVL SAY NKH RD KPIRE QAENIIHLFT LTNLGAPAAF KYFDTTIDRK RYTSTKEVLD ATLIHQ SITG LYETRIDLSQ LGGD
Δ REC1-a	17	MDKKYSIGLA IGNSVGWAV ITDEYKVPSK KFKVLGNTDR HSIKKNLIGA LLFDSGETAE ATRLKRTARR RYTRRKNRIC YLQEIFSNEK AKVDDSFHHR LEESFLVEED KKHERHPIFG NIVDEVAYHE KYPTIYHLRK KLVDSTDKAD LRILIYALAH MIKFRGHFLI EGDLPNDNSD VDKLFIQLVQ TYNQLFEENP INASGVDAKA ILSARLSKSR RLENLIAQLP GEKKNGLFGN LIALSLGLTP NFKSNFDLAE DAKLQLSKDT YDDDLNLLA QIGDQYADLF LAANKLSDAI LLSDILRVNT EITKAPLSAS MIKRYDEHHQ DLTLLKALVR QQLPEKYKEI FFDQSKNGYA GYIDGGASQE EFYKFIKPIL EKMDGTEELL VKLNREDLLR KQRTFDNGSI PHQIHLGELH AILRRQEDFY PFLKDNREKI EKILTFRIPY YVGPLARGNS RFAWMTRKSE ETITPWNFEE VVDKGASAQS FIERMTNFEDK NLPNEKVL PK GGSYTGWGR LSRKLINGIR DKQSGKTILD FLKSDGFANR NFMQLIHDDS LTFKEDIQKA QVSGQGDSLH EHIANLAGSP AIKKGILQTV KVVDLVKVM GRHKPENIVI EMARENQTTQ KGQKNSRERM KRIEEGKEL GSQILKEHPV ENTQLQNEKL YLYYLQNGRD MYVDQELDIN RLSYDVAAI VPQSFLKDDS IDNKVLTRSD KARGKSDNVP SEEVVKMKMN YWRQLLNAKL ITQRKFDNLT KAERGGLSEL DKAGFIKRL VETRQITKHV AQILDSRMNT KYDENDKLIR EVKVITLKS LVSDFRKDFQ FYKREINNY HHAHDAYLNA VVG TALIKKY PKLESEFVYG DYKVYDVRKM IAKSEQEIGK ATAKYFFYSN IMNFFKTEIT LANGEIRKRP LIETNGETGE IVWDKGRDFA TVRKVLSMPQ VNIVKKTEVQ TGGFSKESIL PKRNSDKLIA RKKDWDPKKY GGFDSPTVAY SVLVVAKVEK GSKSKLKS VK ELLGITIMER SSFEKNPIDF LEAKGYKEVK KDIIKL PKY SLFELENGRK RMLASAGELQ KGNELALPSK YVNFY LASH YEKLKGSPEDEQ NEQQLFVEQ HKHYLDEIE QISEFSKRVI LADANLDKVL SAYNKHRDKP IREQAENIIH LFTLTNLGAP AAFKYFDTTI DRKRYTSTKE VLDATLIHQ SITGLYETRID LSQ LGGD
Δ REC1-c	18	MDKKYSIGLA IGNSVGWAV ITDEYKVPSK KFKVLGNTDR HSIKKNLIGA LLFDSGETAE ATRLKRTARR RYTRRKNRIC YLQEIFSNEK AKVDDSFHHR LEESFLVEED KKHERHPIFG NIVDEVAYHE KYPTIYHLRK KLVDSTDKAD LRILIYALAH MIKFRGHFLI EGDLPNDNSD VDKLFIQLVQ TYNQLFEENP INASGVDAKA ILSARLSKSR RLENLIAQLP GEKKNGLFGN LIALSLGLTP NFKSNFDLAE DAKLQLSKDT YDDDLNLLA QIGDQYADLF LAANKLSDAI LLSDILRVNT EITKAPLSAS MIKRYDEHHQ DLTLLKALVR QQLPEKYKEI FFDQSKNGYA GYIDGGASQE EFYKFIKPIL EKMDGTEELL VKLNREDLLR KQRTFDNGSI PHQIHLGELH AILRRQEDFY PFLKDNREKI EKILTFRIPY YVGPLARGNS RFAWMTRKSE ETITPWNFEE VVDKGASAQS FIERMTNFEDK NLPNEKVL PK HSLLYEYFTV YNELTKVKYV TEGMRKPAFL SGEQKKAIVD LLFKTNRKVT VKQLKEDYFK KIECFDSVEI SGVEDRFNAS LGTYHDLKI IKDKDFLDNE ENEDILEDIV LTLTLFEDRE MIEERLKYA HLFDDKVMKQ LKRRTYTGWG RLGGGSAQVS GQGDSLHEHI ANLAGSPAIA KGILQTVKVV DELVKVMGRH KPNIVIEMA RENQTTQKGQ KNSRERMKRI EEGIKELGSQ ILKEHPVENT QLQNEKLYLY YLQNGRDMYV DQELDINRLS DYDVAAIVPQ SFLKDDSIDN KVLTRSDKAR GKSDNVPSEE VVKMKMYWR QLLNAKLITQ

		RKFDNLTKE RGLSELDA GFIKRQLVET RQITKHVAQI LDSRMNTKYD ENDKLIREVK VITLKSCLVS DFRKDFQFYK VREINNYHHA HDAYLNAVVG TALIKKYPKL ESEFVYGDYK VYDVRKMIK SEQEIGKATA KYFFYSNIMN FFKTEITLAN GEIRKRPLIE TNGETGEIVW DKGRDFATVR KVLSPQVNI VKKTEVQTGG FSKESILPKR NSDKLIARKK DWDPKKYGGF DSPTVAYSVL VVAKVEKGKS KKLKSVKELL GITIMERSSE EKNPIDFLEA KGYKEVKKDL IIKLPKYSLF ELENGRKRML ASAGELQKGN ELALPSKYVN FLYLASHYEK LKGSPEDNEQ KQLFVEQHKH YLDEIEEQIS EFSKRVILAD ANLDKVL SAY NKHRDKPIRE QAENIIHLFT LTNLGAPAAF KYFDTTIDRK RYTSTKEVLD ATLIHQSI TG LYETRIDLSQ LGGD
ΔREC2	19	MDKKYSIGLA IGTSVGVAV ITDEYKVPSK KFKVLGNTDR HSIKKNLIGA LLFDSGETAE ATRLKRTARR RYTRRNKRIC YLQEIFSNEK AKVDDSFHHR LEESFLVEED KKHERHPIFG NIVDEVAYHE KYPTIYHLRK KLVDSTDKAD LRLIYLALAH MIKFRGHFLI EGDLPDNDSD NTEITKAPLS ASMIKRYDEH HQDLTLLKAL VRQQLPEKYK EIFFDQSKNG YAGYIDGGAS QEEFYKFIKP ILEKMDGTEE LLVKLNREDL LRKQRTFDNG SIPHQIHLGE LHAILRRQED FYFPLKDNRE KIEKILTFRI PYYVGPLARG NSRFAWMTRK SEETITPWNF EEVVDKGASA QSFIERMTNF DKNLPNEKVL PKHSLLYEYF TVYNELTKVK YVTEGMRKPA FLSGEQKKAI VDLLFKTNRK VTKVQLKEDY PKKIECFDSV EISGVEDRFN ASLGTYHDLI KIIKDKDFLD NEENEDILED IVTLTLTFED REMIEERLKT YAHLFDDKVM KQLKRRRYTG WGRLSRKLIN GIRDKQSGKT ILDFLKSDGF ANRNFMLIH DDSLTFKEDI QKAQVSGQGD SLHEHIANLA GSPAIIKKGIL QTVKVDELV KVMGRHKPEN IVIEMARENQ TTQKGQKNSR ERMKRIEEGI KELGSQILKE HPVENTQLQN EKLYLYYLQN GRDMYVDQEL DINRLSDYDV AAIVPQSFLK DDSIDNKVLT RSDKARGKSD NVPSEEVVKK MKNYWRQLLN AKLITQRKFD NLTKAERGGL SELDKAGFIK RQLVETRQIT KHVAQILDSR MNTKYDENDK LIREVKVITL KSKLVSDFRK DFQFYKVREI NNYHHAHDAY LNAVVGTAI KKYPKLESEF VYGDYKVYDV RKMIKSEQE IGKATAKYFF YSNIMNFFKT EITLANGEIR KRPLIETNGE TGEIVWDKGR DFATVRKVL MPQVNIVKKT EVQTGGFSKE SILPKRNSDK LIARKKDWD KKYGGFDSPT VAYSVLVAK VEKGSKKLK SVKELLGITI MERSSEFKNP IDFLEAKGYK EVKDLIIKL PKYSLFELEN GRKRMLASAG ELQKGNELAL PSKYVNFLYL ASHYEKLKGS PEDNEQKQLF VEQHKHYLDE IIEQISEFSK RVILADANLD KVL SAYNKH RDKPIREQAEN IHLFTLTNL GAPAAFKYFD TTIDRKRYTS TKEVLDTLI HQSITGLYET RIDLSQLGGD
ΔHNH ΔREC1-a	20	MDKKYSIGLA IGTSVGVAV ITDEYKVPSK KFKVLGNTDR HSIKKNLIGA LLFDSGETAE ATRLKRTARR RYTRRNKRIC YLQEIFSNEK AKVDDSFHHR LEESFLVEED KKHERHPIFG NIVDEVAYHE KYPTIYHLRK KLVDSTDKAD LRLIYLALAH MIKFRGHFLI EGDLPDNDSD VDKLFIQLVQ TYNQLFEENP INASGVDAKA ILSARLSKSR RLENLIAQLP GEKKNGLFGN LIALSLGLTP NFKSNFDAE DAKLQLSKDT YDDDLNLLA QIGDQYADLF LAAKNLSDAI LLSDILRVNT EITKAPLSAS MIKRYDEHHQ DLTLLKALVR QQLPEKYKEI FFDQSKNGYA GYIDGGASQE EFKYFIKPIL EKMDGTEELL VKLNREDLLR KQRTFDNGSI PHQIHLGELH AILRRQEDFY PFLKDNREKI EKILTFRI YVGPLARGNS RFAWMTRKSE ETITPWNFEE VVDKGASQAS FIERMTNFDK NLPNEKVL PK GGSYTGWGR LSRKLINGIR DKQSGKTILD FLKSDGFANR NFMQLIHDDS LTFKEDIQKA QVSGQGDSLH EHIANLAGSP AIKKGILQTV

		KVVDELVKVM GRHKPENIVI EMARENQTTQ KGQKGGGSEL DKAGFIKRQL VETRQITKHV AQILDSRMNT KYDENDKLIR EVKVITLKS LVSDFRKDFQ FYKVBREINNY HHAHDAYLNA VVGITALIKKY PKLESEFVYG DYKVYDVRKM IAKSEQEIGK ATAKYFFYSN IMNFFKTEIT LANGEIRKRP LIETNGETGE IVWDKGRDFA TVRKVLSMPQ VNIVKKTEVQ TGGFSKESIL PKRNSDKLIA RKKDWDPKKY GGFDSPTVAY SVLVVAKVEK GKSKKLKSVK ELLGITIMER SSFEKNPIDF LEAKGYKEVK KDLIIKLPKY SLFELENGRK RMLASAGELQ KGNELALPSK YVNFLYLASH YEKLKGSPEQ NEQKQLFVEQ HKHYLDEIIE QISEFSKRVI LADANLDKVL SAYNKHDKP IREQAENIIH LFTLTNLGAP AAFKYFDTTI DRKRYTSTKE VLDATLIHQ S ITGLYETRID LSQLGGD
Δ HNH Δ REC1-c	21	MDKKYSIGLA IGTNSVGWAV ITDEYKVPSK KFKVLGNTDR HSIKKNLIGA LLFDSGETAE ATRLKRTARR RYTRRNKRIC YLQEIFSNEK AKVDDSFHHR LEESFLVEED KKHERHPIFG NIVDEVAYHE KYPTIYHLRK KLVDSTDKAD LRLIYLALAH MIKFRGHFLI EGDLPDNDSD VDKLFIQLVQ TYNQLFEENP INASGVDAKA ILSARLSKSR RLENLIAQLP GEKKNGLFGN LIALSLGLTP NFKSNDLAE DAKLQLSKDT YDDDLNLLA QIGDQYADLF LAAKNLSDAI LLSDILRVNT EITKAPLSAS MIKRYDEHHQ DLTLLKALVR QQLPEKYKEI FFDQSKNGYA GYIDGGASQE EFYKFIKPIL EKMDGTEELL VKLNREDLLR KQRTFDNGSI PHQIHLGELH AILRRQEDFY PFLKDNREKI EKILTFRIPY YVGPLARGNS RFAWMTRKSE ETITPWNFEE VVDKGASAQS FIERMTNFDK NLPNEKVLPK HSLLYEYFTV YNELTKVKYV TEGMRKPAFL SGEQKKAIVD LLFKTNRKVT VKQLKEDYFK KIECFDSVEI SGVEDRFNAS LGTYHDLKI IKDKDFLDNE ENEDILEDIV LTLTLFEDRE MIEERLKYA HLFDDKVMKQ LKRRRYTGWG RLGGGSAQVS GQGDLSLHEHI ANLAGSPAIA KGILQTVKVV DELVKVMGRH KPNIVIEMA RENQTTQKGQ KGGGSELDKA GFIKRQLVET RQITKHVAQI LDSRMNTKYD ENDKLIREVK VITLKSCLVS DFRKDFQFYK VREINNYHHA HDAYLNAVVG TALIKKYPKL ESEFVYG DYK VYDVRKMIK SEQEIGKATA KYFFYSNIMN FFKTEITLAN GEIRKRPLIE TNGETGEIVW DKGRDFATVR KVLSPQVNI VKKTEVQTGG FSKESILPKR NSDKLIARKK DWDPKKYGGF DSPTVAYSVL VVAKVEKGKS KKLKSVKELL GITIMERSSE EKNPIDFLEA KGYKEVKKDL IIKLPKYSLF ELNGRKRML ASAGELQKGN ELALPSKYVN FLYLASHYEK LKGSPEDNEQ KQLFVEQHKH YLDEIIEQIS EFSKRVILAD ANLDKVL SAY NKHRDKPIRE QAENIIHLFT LTNLGAPAAF KYFDTTIDRK RYTSTKEVLD ATLIHQ SITG LYETRIDL S LGGD
Δ HNH Δ REC2	22	MDKKYSIGLA IGTNSVGWAV ITDEYKVPSK KFKVLGNTDR HSIKKNLIGA LLFDSGETAE ATRLKRTARR RYTRRNKRIC YLQEIFSNEK AKVDDSFHHR LEESFLVEED KKHERHPIFG NIVDEVAYHE KYPTIYHLRK KLVDSTDKAD LRLIYLALAH MIKFRGHFLI EGDLPDNDSD NTEITKAPLS ASMIKRYDEH HQDLTLLKAL VRQQLPEKYK EIFFDQSKNG YAGYIDGGAS QEEFYKFIKP ILEKMDGTEE LLVKLNREDL LRKQRTFDNG SIPHQIHLGE LHAILRRQED FYPFLKDNRE KIEKILTFRI PYYVGPLARG NSRFAWMTRK SEETITPWNF EEVVDKGASA QSFIERMTNF DKNLPNEKVL PKHSLLYEYF TVYNELTKVK YVTEGMRKPA FLSGEQKKA VDLFKTNRK VTKQLKEDY FKKIECFDSV EISGVEDRFN ASLGTYHDLI KIIKDKDFLD NEENEDILED IVLTLTLFED REMIEERLKT YAHLFDDKVM KQLKRRRYTG WGRLSRKLIN GIRDKQSGKT ILDFLKSDGF ANRNFMLIH DDSLTFKEDI QKAQVSGQGD SLHEHIANLA GSPAIAKGIL QTQKVVDELV KVMGRHKPEN IVIEMARENQ TTQKGQKGGG

		SELDKAGFIK RQLVETRQIT KHVAQILDSR MNTKYDENDK LIREVKVITL KSKLVSDFRK DFQFYKVREI NNYHHAHDAY LNAVVGTAI KKYPKLESEF VYGDYKVYDV RKMIKSEQE IGKATAKYFF YSNIMNFFKT EITLANGEIR KRPLIETNGE TGEIVWDKGR DFATVRKVLS MPQVNIVKKT EVQTGGFSKE SILPKRNSDK LIARKKDWD P KKYGGFDSPT VAYSVLVVA VEKGSKKLK SVKELLGITI MERSSFENP IDFLEAKGYK EVKCDLIKL PKYSLFELEN GRKRMLASAG ELQKGNELAL PSKYVNFYLY ASHYEKLKGS PEDNEQKQLF VEQHKHYLDE IIEQISEFSK RVILADANLD KVL SAYNKH R DKPIREQAEN IHLFTLTNL GAPAAFKYFD TTIDRKRYTS TKEVLDTALI HOSITGLYET RIDLSQLGGD
Δ HNH Δ RuvCIII-b Δ REC2	23	MDKYSIGLA IGTSVGVAV ITDEYKVPSK KFKVLGNTDR HSIKKNLIGA LLFDSGETAE ATRLKRTARR RYTRRNKRIC YLQEIFSNEK AKVDDSFHHR LEESFLVEED KKHHPHPIFG NIVDEVAYHE KYPTIYHLRK KLVDSTDKAD LRLIYLALAH MIKFRGHFLI EGDLPNDNSD NTEITKAPLS ASMIKRYDEH HQDLTLLKAL VRQQLPEKYK EIFFDQSKNG YAGYIDGGAS QEEFYKFIKP ILEKMDGTEE LLVKLNREDL LRKQRTFDNG SIPHQIHLGE LHAILRRQED FYPFLKDNRE KIEKILTFRI PYYVGPLARG NSRFAWMTRK SEETITPWNF EEVVDKGASA QSFIERMTNF DKNLPNEKVL PKHSLLYEYF TVYNELTKVK YVTEGMRKPA FLSGEQKAI VDLLFKTNRK VTKVQLKEDY PKKIECFDSV EISGVEDRFN ASLGYTHDLL KIIKDKDFLD NEENEDILED IVLTTLTFED REMIEERLKT YAHLFDDKVM KQLKRRRYTG WGRLSRKLIN GIRDKQSGKT ILDFLKSDGF ANRNFMLIH DDSLTFKEDI QKAQVSGQGD SLHEHIANLA GSPAIIKKGIL QTVKVDELV KVMGRHKPEN IVIEMARENQ TTQKGQKGGG SELDKAGFIK RQLVETRQIT KHVAQILDSR MNTKYDENDK LIREVKVITL KSKLVSDFRK DFQFYKVREI NNYHHAHDAY LNAVVGTAI KKYPGGSGG GSDKGRDFAT VRKVLSPQV NIVKKTEVQT GGFSKESILP KRNSDKLIAR KKDWDPKKYG GFDSPVAYS VLVAKEVKG KSKKLKSVKE LLGITIMERS SFEKNPIDFL EAKGYKEVKK DLIKLPKYS LFELENGRKR MLASAGELQK GNELALPSKY VNFYLYASHY EKLKGS PEDN EQKQLFVEQH KHYLDEIEEQ ISEFSKRVL ADANLDKVL AYNKHRDKPI REQAENIHL FTLTNLGAPA AFKYFDTTID RKRYTSTKEV LDATLIHQSI TGLYETRIDL SQLGGD
Δ HNH Δ RuvCIII-b Δ REC2 Δ REC1-a	24	MDKYSIGLA IGTSVGVAV ITDEYKVPSK KFKVLGNTDR HSIKKNLIGA LLFDSGETAE ATRLKRTARR RYTRRNKRIC YLQEIFSNEK AKVDDSFHHR LEESFLVEED KKHHPHPIFG NIVDEVAYHE KYPTIYHLRK KLVDSTDKAD LRLIYLALAH MIKFRGHFLI EGDLPNDNSD NTEITKAPLS ASMIKRYDEH HQDLTLLKAL VRQQLPEKYK EIFFDQSKNG YAGYIDGGAS QEEFYKFIKP ILEKMDGTEE LLVKLNREDL LRKQRTFDNG SIPHQIHLGE LHAILRRQED FYPFLKDNRE KIEKILTFRI PYYVGPLARG NSRFAWMTRK SEETITPWNF EEVVDKGASA QSFIERMTNF DKNLPNEKVL PKGGGSYTGW GRLSRKLING IRDKQSGKTI LDFLKSDGFA NRNFMLIHD DSLTFKEDIQ KAQVSGQGDS LHEHIANLAG SPAIKKGILQ TVKVDELVK VMGRHKPENI VIEMARENQT TQKGQKGGGS ELDKAGFIK RQLVETRQIT KHVAQILDSR MNTKYDENDK LIREVKVITL KSKLVSDFRK DFQFYKVREI NNYHHAHDAY LNAVVGTAI KYPGGSGGG SDKGRDFATV RKVLSPQVN IVKKTEVQTG GFSKESILPK RNSDKLIAR KDWDPKKYG GFDSPVAYS LVVAKEVKG KSKKLKSVKEL LGITIMERS FKNPIDFL EAKGYKEVKD LIIKLPKYS LFELENGRKR LASAGELQK GNELALPSKY VNFYLYASHY EKLKGS PEDN EQKQLFVEQH

		HYLDEIIEQI SEFSKRVILA DANLDKVLSA YNKHRRDKPIR EQAENIIHLF TLTNLGAPAA FKYFDTTIDR KRYTSTKEVL DATLIHQSIIT GLYETRIDLS QLGGD
Δ HNH Δ RuvCIII-b Δ REC2 Δ REC1-c	25	MDKKYSIGLA IGTNSVGWAV ITDEYKVPSK KFKVLGNTDR HSIKKNLIGA LLFDSGETAE ATRLKRTARR RYTRRNKRIC YLQEIFSNEK AKVDDSFHHR LEESFLVEED KKHERHPIFG NIVDEVAYHE KYPTIYHLRK KLVDSTDKAD LRLIYLALAH MIKFRGHFLI EGDLPDNDSD NTEITKAPLS ASMIKRYDEH HQDLTLLKAL VRQQLPEKYK EIFFDQSKNG YAGYIDGGAS QEEFYKFIKP ILEKMDGTEE LLVKNREDL LRKQRTFDNG SIPHQIHLGE LHAILRRQED FYFPLKDNRE KIEKILTFRI PYYVGPLARG NSRFAWMTRK SEETITPWNF EEVVDKGASA QSFIERMTNF DKNLPNEKVL PKHSLLYEYF TVYNELTKVK YVTEGMRKPA FLSGEQKKAI VDLLFKTNRK VTVKQLKEDY FKKIECFDSV EISGVEDRFN ASLGYTHDLL KIIKDKDFLD NEENEDILED IVLTLTLFED REMIEERLKT YAHLFDDKVM KQLKRRRYTG WGRLGGSQAQ VSGQGDSLHE HIANLAGSPA IKKGILQTVK VVDELVKVMG RHKPNIVIE MARENQTTQK GQKGGGSELD KAGFIKRQLV ETRQITKHVA QILDSRMNTK YDENDKLIRE VKVITLKS KL VSDFRKDFQF YKVREINNYH HAHDAYLNAV VGTALIKKYP GGGSGGSDK GRDFATVRKV LSMPQVNIVK KTEVQTGGFS KESILPKRNS DKLIARKKDW DPKKYGGFDS PTVAYSVLVV AKVEKGSKK LKSVKELLGI TIMERSSEFEK NPIDFLEAKG YKEVKKDLII KLPKYSLFEL ENGRKRLAS AGELQKGNEL ALPSKYVNFL YLASHYEKLG GSPEDNEQKQ LFVEQHKHYL DEIEQISEF SKRVILADAN LDKVLSAYNK HRDKPIREQA ENIIHLFTLT NLGAPAAFKY FDTTIDRKRY TSTKEVLDTL LIHQSIITGLY ETRIDLSQLG GD
Δ HNH Δ RuvCIII-b Δ REC2 Δ REC1-a Δ REC1-c	26	MDKKYSIGLA IGTNSVGWAV ITDEYKVPSK KFKVLGNTDR HSIKKNLIGA LLFDSGETAE ATRLKRTARR RYTRRNKRIC YLQEIFSNEK AKVDDSFHHR LEESFLVEED KKHERHPIFG NIVDEVAYHE KYPTIYHLRK KLVDSTDKAD LRLIYLALAH MIKFRGHFLI EGDLPDNDSD NTEITKAPLS ASMIKRYDEH HQDLTLLKAL VRQQLPEKYK EIFFDQSKNG YAGYIDGGAS QEEFYKFIKP ILEKMDGTEE LLVKNREDL LRKQRTFDNG SIPHQIHLGE LHAILRRQED FYFPLKDNRE KIEKILTFRI PYYVGPLARG NSRFAWMTRK SEETITPWNF EEVVDKGASA QSFIERMTNF DKNLPNEKVL PKGGGSYTGW GRLGGSQAQV SGQGDSLHEH IANLAGSPA KKGILQTVKV VDELVKVMGR HKPENIVIE ARENQTTQKG QKGGGSELDK AGFIKRQLVE TRQITKHVAQ ILDSRMNTKY DENDKLIREV KVITLKS KL VSDFRKDFQFY KVRINNYHH AHDAYLNAV GTALIKKYPG GSGGGSDKG RDFATVRKVL SMPQVNIVKK TEVQTGGFSK ESILPKRNSD KLIARKKDW DPKKYGGFDS PTVAYSVLVVA KVEKGSKSKL KSVKELLGIT IMERSSEFEK NPIDFLEAKG YKEVKKDLIIK LPKYSLFEL NGRKRMLASA GELQKGNELA LPSKYVNFL YLASHYEKLG GSPEDNEQKQ FVEQHKHYLD EIEQISEFS KRVILADAN LDKVLSAYNK HRDKPIREQA NIIHLFTLTN LGAPAAFKY FDTTIDRKRY TSTKEVLDTL LIHQSIITGLY TRIDLSQLGG D
Δ REC3	27	MDKKYSIGLA IGTNSVGWAV ITDEYKVPSK KFKVLGNTDR HSIKKNLIGA LLFDSGETAE ATRLKRTARR RYTRRNKRIC YLQEIFSNEK AKVDDSFHHR LEESFLVEED KKHERHPIFG NIVDEVAYHE KYPTIYHLRK KLVDSTDKAD LRLIYLALAH MIKFRGHFLI EGDLPDNDSD VDKLFIQLVQ TYNQLFEENP INASGVDAKA ILSARLSKSR RLENLIAQLP GEKKNGLFGN LIALSLGLTP

		NFKSNFDAE DAKLQLSKDT YDDDLNLLA QIGDQYADLF LAAKNLSDAI LLSDILRVNT EITKAPLSAS MIKRYDEHHQ DLTLLKALVR QQLPEKYKEI FFDQSKNGYA GYIDGGASQE EFYKFIKPIL EKMDGTEELL VKLNREDLLR KQRTFDNGSI PHQIHLGELH AILRRQEDFY PFLKDNREKI EKILTFRIPY YVGPLARGNS RFAWMTRKSE ETITPWNFEE VVDKGASAQS FIERMTNGGS VSGQGDSLHE HIANLAGSPA IKKGILQTVK VVDELVKVMG RHKPENIVIE MARENQTTQK GQKNSRERMK RIEEGIKELG SQILKEHPVE NTQLQNEKLY LYYLQNGRDM YVDQELDINR LSDYDVAAIV PQSFLKDDSI DNKVLTRSDK ARGKSDNVPS EEVVKMKMNY WRQLLNAKLI TQRKFDNLTK AERGGELSELD KAGFIKRQLV ETRQITKHVA QILDSRMNTK YDENDKLIRE VKVITLKS KL VSDFRKDFQF YKVREINNYH HAHDAYLNAV VGTALIKKYP KLESEFVYGD YKVYDVRKMI AKSEQEIGKA TAKYFFYSNI MNFFKTEITL ANGEIRKRPL IETNGETGEI VWDKGRDFAT VRKVLSPQV NIVKKTEVQT GGFSKESILP KRNSDKLIAR KKDWDPKKYG GFDSPTVAYS VLVVAKVEKG KSKKLKSVKE LLGITIMERS SFEKNPIDFL EAKGYKEVKK DLIKLPKYS LFELENGRKR MLASAGELQK GNEALPSKY VNFLYLASHY EKLKGSPEDN EQKQLFVEQH KHYLDEIEQ ISEFSKRVL ADANLDKVL AYNKHRDKPI REQAENIIHL FTLTNLGAPA AFKYFDTTID RKRYTSTKEV LDATLIHQSI TGLYETRIDL SQLGGD
Δ REC3 Δ HNH	28	MDKKYSIGLA IGTNSVGWAV ITDEYKVPSK KFKVLGNTDR HSIKKNLIGA LLFDSGETAE ATRLKRTARR RYTRRNKRIC YLQEIFSNEK AKVDDSFHHR LEESFLVEED KKHERHPIFG NIVDEVAYHE KYPTIYHLRK KLVDSTDKAD LRILIYALAH MIKFRGHFLI EGDLPNDNSD VDKLFIQLVQ TYNQLFEENP INASGVDAKA ILSARLSKSR RLENLIAQLP GEKKNGLFGN LIALSLGLTP NFKSNFDAE DAKLQLSKDT YDDDLNLLA QIGDQYADLF LAAKNLSDAI LLSDILRVNT EITKAPLSAS MIKRYDEHHQ DLTLLKALVR QQLPEKYKEI FFDQSKNGYA GYIDGGASQE EFYKFIKPIL EKMDGTEELL VKLNREDLLR KQRTFDNGSI PHQIHLGELH AILRRQEDFY PFLKDNREKI EKILTFRIPY YVGPLARGNS RFAWMTRKSE ETITPWNFEE VVDKGASAQS FIERMTNGGS VSGQGDSLHE HIANLAGSPA IKKGILQTVK VVDELVKVMG RHKPENIVIE MARENQTTQK GQKGGSELD KAGFIKRQLV ETRQITKHVA QILDSRMNTK YDENDKLIRE VKVITLKS KL VSDFRKDFQF YKVREINNYH HAHDAYLNAV VGTALIKKYP KLESEFVYGD YKVYDVRKMI AKSEQEIGKA TAKYFFYSNI MNFFKTEITL ANGEIRKRPL IETNGETGEI VWDKGRDFAT VRKVLSPQV NIVKKTEVQT GGFSKESILP KRNSDKLIAR KKDWDPKKYG GFDSPTVAYS VLVVAKVEKG KSKKLKSVKE LLGITIMERS SFEKNPIDFL EAKGYKEVKK DLIKLPKYS LFELENGRKR MLASAGELQK GNEALPSKY VNFLYLASHY EKLKGSPEDN EQKQLFVEQH KHYLDEIEQ ISEFSKRVL ADANLDKVL AYNKHRDKPI REQAENIIHL FTLTNLGAPA AFKYFDTTID RKRYTSTKEV LDATLIHQSI TGLYETRIDL SQLGGD
Δ REC3 Δ HNH Δ RuvCIII-b	29	MDKKYSIGLA IGTNSVGWAV ITDEYKVPSK KFKVLGNTDR HSIKKNLIGA LLFDSGETAE ATRLKRTARR RYTRRNKRIC YLQEIFSNEK AKVDDSFHHR LEESFLVEED KKHERHPIFG NIVDEVAYHE KYPTIYHLRK KLVDSTDKAD LRILIYALAH MIKFRGHFLI EGDLPNDNSD VDKLFIQLVQ TYNQLFEENP INASGVDAKA ILSARLSKSR RLENLIAQLP GEKKNGLFGN LIALSLGLTP NFKSNFDAE DAKLQLSKDT YDDDLNLLA QIGDQYADLF LAAKNLSDAI LLSDILRVNT EITKAPLSAS MIKRYDEHHQ DLTLLKALVR QQLPEKYKEI

		FFDQSKNGYA GYIDGGASQE EFYKFIKPIL EKMDGTEELL VKLNREDLLR KQRTFDNGSI PHQIHLGELH AILRRQEDFY PFLKDNREKI EKILTFRIPIY YVGPLARGNS RFAWMTRKSE ETITPWNFEE VVDKGASAQS FIERMTNGGS VSGQGDSLHE HIANLAGSPA IKKGILQTVK VVDELVKVMG RHKPENIVIE MARENQTTQK GQKGGGSELD KAGFIKRQLV ETRQITKHVA QILDSRMNTK YDENDKLIRE VKVITLKS KL VSDFRKDFQF YKVREINNYH HAHDAYLNAV VGTALIKKYP GGGSGGSDK GRDFATVRKV LSMPQVNIVK KTEVQTGGFS KESILPKRNS DKLIARKKDW DPKKYGGFDS PTVAYSVLVV AKVEKGKSKK LKS VKELGI TIMERSSEFEK NPIDFLEAKG YKEVKKDLII KLPKYSLFEL ENGRKRMLAS AGELQKGNEL ALPSKYVNFL YLASHYEK LK GSPEDNEQKQ LFVEQHKHYL DEIEQISEF SKRVILADAN LDKVLSAYNK HRDKPIREQA ENIIHLFTLT NLGAPAAFKY FDTTIDRKRY TSTKEVL DAT LIHQ SITGLY ETRIDLSQLG GD
ΔREC3 ΔHNN ΔRuvCIII-b ΔREC2	30	MDKKYSIGLA IGNSVGWAV ITDEYKVPSK KFKVLGNTDR HSIKKNLIGA LLFDSGETAE ATRLKRTARR RYTRRNKRIC YLQEIFSNEM AKVDDSFHHR LEESFLVEED KKHERHPIFG NIVDEVAYHE KYPTIYHLRK KLVSDTKAD LRLIYLALAH MIKFRGHFLI EGDLPDNDSD NTEITKAPLS ASMIKRYDEH HQDLTLLKAL VRQQLPEKYK EIFFDQSKNG YAGYIDGGAS QEEFYKFIKP ILEKMDGTEE LLVKLNREDL LKQRTFDNG SIPHQIHLGE LHAILRRQED FYFPLKDNRE KIEKILTFRI PYYVGPLARG NSRFAWMTRK SEETITPWNF EEVVDKGASA QSFIERMTNG GSVSGQGDSL HEHIANLAGS PAIKKGILQT VKVDELVKV MGRHKPENIV IEMARENQTT QKGQKGGGSE LDKAGFIKRQ LVETRQITKH VAQILDSRMN TKYDENDKLI REVKVITLKS KLVSDFRKDF QFYKVREINN YHHAHDAYLN AVVGTALIKK YPGGGSGGGS DKGRDFATVR KVL SMPQVNI VKKTEVQTGG FSKESILPKR NSDKLIARKK DWDPKKG GGF DSPTVAYSVL VVAKVEKGKS KKLKSVKELL GITIMERSSE EKNPIDFLEA KGYKEVKKDL IIKLPKYSLF ELNGRKRML ASAGELQKGN ELALPSKYVN FLYLASHYEK LK GSPEDNEQ KQLFVEQHKH YLDEIEQIS EFSKRVLAD ANLDKVLSAY NKHRDKPIRE QAENIIHLFT LTNLGAPAAF KYFDTTIDRK RYTSTKEVL DAT LIHQ SITG LYETRIDLSQ LGGD

5

It is to be understood that the various polypeptides having at least one of the aforementioned deletions and/or mutations, even if their amino acid sequences are not explicitly described herein for the sake of conciseness, are contemplated to be within the scope of the present invention.

10

In a second aspect, the invention relates to a polypeptide having at least 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 99.5%

15

sequence identity to *Staphylococcus aureus* Cas9 having the amino acid sequence set forth in SEQ ID NO:2, *Neisseria meningitidis* Cas9 having the amino acid sequence set forth in SEQ ID NO:3, *Acidaminococcus* sp. Cpf1 having the amino acid sequence set forth in SEQ ID NO:4, *Streptococcus thermophilus* Cas9 having the amino acid sequence set forth in SEQ ID NO:5,

- 5 *Sutterella wadsworthensis* Cas9 having the amino acid sequence set forth in SEQ ID NO:6, *Filifactor alocis* Cas9 having the amino acid sequence set forth in SEQ ID NO:7, *Lactobacillus johnsonii* Cas9 having the amino acid sequence set forth in SEQ ID NO:8, *Campylobacter lari* Cas9 having the amino acid sequence set forth in SEQ ID NO:9, *Parvibaculum lavamentivorans* Cas9 having the amino acid sequence set forth in SEQ ID NO:10, *Mycoplasma gallisepticum*
- 10 Cas9 having the amino acid sequence set forth in SEQ ID NO:11, *Treponema denticola* Cas9 having the amino acid sequence set forth in SEQ ID NO:12, or *Lachnospiraceae* bacterium ND2006 Cpf1 having the amino acid sequence set forth in SEQ ID NO:13 over its entire length, wherein said polypeptide comprises at least one truncation of said amino acid sequence, wherein said at least one truncation corresponds to at least one truncation of SEQ ID NO:1
- 15 selected from the group consisting of Δ HNH (Δ 775-909), Δ RuvCIII-b (Δ 1002-1074), Δ REC1-a (Δ 510-655), Δ REC1-b (Δ 525-587), Δ REC1-c (Δ 662-710), Δ REC2 (Δ 180-308), Δ REC2-a (Δ 212-244), Δ REC2-b (Δ 244-276), Δ REC2-c (Δ 276-308), Δ REC2-d (Δ 199-283), Δ REC2-e (Δ 198-257), Δ REC2-f (Δ 235-286), Δ REC2-g (Δ 217-266), Δ REC3 (Δ 498-712) and combinations thereof, wherein the position numbering is in accordance with SEQ ID NO:1, wherein the polypeptide
- 20 has CRISPR-Cas DNA-binding activity.

Table 2. Cas proteins of different organisms

Name	UniProt Entry number	SEQ ID NO	Amino acid sequence
<i>Streptococcus pyogenes</i> Cas9	Q99ZW2	1	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAE ATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSSFFHRLEESFLVEEDKKHERHPIFG NIVDEVAYHEKYPTIYHLRKKLV DSTDKADLRLLYLALAHMIKFRGHFLIEGDLNPDNSD VDKLFQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFNG LIALSLGLTPNFKSNFDLAEDAKLQLSKD TYDDDLNLLAQIGDQYADFLAAKNLSDAI LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELH AILRRQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEE VVDKGASQAQSFIERMTNFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFL SGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKI IKDKDFLDNEENEDILEDIVLTTLTFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG RLSRKLINGIRDKQSGKTILDFLKSDFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSL HEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIAMARENQTTQKGQKNSRER MKRIEIGIKELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVDH IVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKMKMKNYWRQLLNAKLITQRKFDNL TKAERGGLSELDKAGFIKRLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS KLVSDFRKDFQFYKVVREINNYHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVYDVRK MIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDF ATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKGFGDSPTVA YSVLVAKVEKGKSKKLKSVKELLGITIMERSSEFKNPIDFLEAKGYKEVKKDLIILPK YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGS PEDNEQKQLFV

			E QHKHYLDEIEQISEFSKRVLADANLDKVL SAYNKH RD KPIREQAENIIHLFTLTNLGA PAAFKYFDTTIDRKRYTSTKEVL DATLIHQ SITGLYETRIDLSQLGGD
<i>Staphylococcus aureus</i> Cas9	J7RUA5	2	MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRR RHRIQRVKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEFSAALLHLAKRRGVHN VNEVEEDTGNELSTKEQISRN SKALEEKYVAELQLERLKKDGEVRGSINRFKTS DYVKEA KQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTY F PEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYEYEFQIEN VFKQKKKPTLKQIA KEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIENAELLDQIAKILTIYQS SEDIQEELTNLNSLTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNR LKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIKKYGLPNDIIELAR EKNSKDAQKMINEMQKRNRQTNERIEEII RTTGKENAKYLIEKIKLHDMQEGKCLYSLEA IPLEDLLNNPFNYEVDHII PRSVSFDNSFNNKVLVKQEENSKGNRTPFQYLSSSDSKIS YETFKKHILNLA KGGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNL RSYFRVNNLDVKVKSINGGFTSFLRRKWKFKERNKGYKHAEDALIANADFIKWK LDKAKKVMENQMFEKQAESMPEIETE QEYKEIFITPHQIKHIKDFKDYKYS HRVDKKPN RELINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLKLINKSPEKLLMYHHD PQTYQKL KLIMEQYGDENPLYKYEETGNYLTKYSKKDNGPVIKKIKYGNKLNALHDITDDYPNS RNKVVKLSLKP YRFDVYLDNGVYKFVTVKNLDVIKKENYEVNSKCYEEAKKLKISNQA EFIASFYNNDLIKINGEL YRIGVNNDDLNRIEVNMIDITYREYLENMNDKRPPRIIKTI ASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG
<i>Neisseria meningitidis</i> Cas9	C9X1G5	3	MAAFKPNSINYILGLDIGIASVGWAMVEIDEEENPIRLIDLGVRFERAEVPTGDSLAM ARRLARSVRRLTRRRRAHRLLRTRRLKREGVLQAANFDENGLIKSLPNTPWQLRAAALD R KLTPLEWSAVLLHLIKHRGYLSQRKNEGETADKELGALLKG VAGNAHALQTGDFRTPAEL ALNKFESGHIRNQRSDYSHTFSRKDLQAE LILLFEKQKEFGNPHVSGGLKEGIETLLM TQRPALSGDAVQKMLGHCTFEPAEPKAAKN TYAERFIWLTKLNNLRILEQGSERPLTDT ERATLMDEPYRKS KLTYAQARKLLGLEDTAFFKGLRYGKDNAEASTLMEMKAYHAISRAL EKEGLKDKKSPLNLSPELQDEIGTAFSLFKTDEDITGRLKDRIQPEILEALLKHISFDKF VQISLKALRRIVPLMEQ GKRYDEACAEIYGDHYGKKNTEEKIYLPPIPADEIRNPVVLRA LSQARKVINGVVRRYGSPARIHIETAREVGKSFKDRKEIEKRQEENRKDREKAAAKFREY FPNFVGEPKSKDILKRLYEQQHGKCLYSGKEINLGR LNEKGYVEIDHALPFSRTWDDSF NNKVLVLGSENQNKGNQTPYEYFNGKDNSREWQE FKARVETSRFP RSKQRILLQKFD ED GFKERNLNDTRYVNRFLCQFVADRMRLTGKGKRVFASNGQITNLLRGFWGLRKVRAE ND RHHALDAVVACSTVAMQQKITRFVRYKEMNAFDGKTIDKETGEVLHQKTHFPQPWEFF A QEV MIRVFGKPDGKPEFEEADTLEKLRTL LAEKLSSRPEAVHEYVTPLFVS RAPNRKMSG QGHMETVKSARLDEGVSLRVPLTQLKLKDLEKMVNREREPKLYEALKARLEAHKDDP A KAFAEPFYKYDKAGNRTQQVKAVRVEQVQKTGVWVRNHNGIADNATMVRVDVFEKGD KYY LVPIYSWQVAKGILPDRAVVQ GKDEEDWQLIDDSFNFKFSLHPNDLVEVITKKAR MFGYF ASCHRG TGNNIRIHDLDHKIGKNGILEGIGVKTALS FQKYQIDELGKEIRPCRLKKRPP VR

<i>Acida minoc occus sp. Cpf1</i>	U2UMQ 6	4	MTQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDDHYKELKPIIDRIYKT YADQCLQLVQLDWENLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDFIGRTDNLTD INKRHAEIYKGLFKAELFNGKVLKQLGTVTTTEHENALLRSFDKFTTYFSGFYENRKNVF SAEDISTAIPHRIVQDNFPKFKENCHIFTRLITAVPSLREHFENVKKAIGIFVSTSIEEV FSFPFYNQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQKNDETAHIIASLPH RFIPLFKQLSDRNTLSFILEEFKSDEEVIQSFCKYKTLRNENVLETAELFNLNSID LTHIFISHKKLETISSALCDHWDTLRNALYERRISELTGKITKSAKEKVQRSLKHEDINL QEIIAAGKELSEAFKQKTSEILSHAHAAALDQPLPTTLKKQEEKEILKSQDLSLLGLYHL LDWFAVDESNEVDPEFSARLTGIKLEMEPSLSFYNKARNYATKPKYSVEKFLNFQMPTL ASGWDVNKEKNNGAILFVKNGLYYLGIMPKQKGRYKALSFEPTTEKTSSEGFDMYYDYFP D AAKMIPKCSTQLKAVTAHFQTHHTPILLSNNFIEPLEITKEIYDLNNPEKEPKKFQTAYA KKTGDQKGYREALCKWIDFTRDFLSKYTKTTSIDLSSLRPSSQYKDLGEYYAELNPLLYH ISFQRIAEKEIMDAVETGKLYLFQIYNKDFAKGHGKPNLHTLYWTGLFSPENLAKTSIK LNGQAELFYRPKSRMKRMAHRLGEKMLNKKLKDQKTPIDTLYQELYDYVNHRLSHDLS D EARALLPNVITKEVSHEIIKDRRFTSDKFFFHVPITLNYQAANSPPSKFNQRVNAYLKEHP ETPIIGIDRGERNLIYITVIDSTGKILEQRSNTIQQFDYQKKLDNREKERVAAARQAWSV VGTIKDLKQGYLSQVIHEIVDLMIHYQAVVLENLNGFKSKRTGIAEKAVYQQFEKMLI DKLNCLVLKDYPAEKVGGLNPNYQLTDQFTSFAKMGTSQSGFLFYVPAPYTSKIDPLTGFV DPFVWTKIKNHESRKHFLEGDFDLHYDVKTGDFILHFKMNRNLSFQRLPGFMPAWDIVF EKNETQFDAQGTPFIAGKRIVPIENHRFTGRYRDLYPANELIALLEEKGIVFRDGSNIL PKLENDSDSHADTMVALIRSVLQMRNSNAATGEDYINSPVRDLNGVCFDSRFQNPPEWP M DADANGAYHIALKGQLLLNLHLESKDLKLQNGISNQDWLAYIQELRN
<i>Strept ococc us thermo philus Cas9</i>	G3ECR 1	5	MLFNKCIISINLDFSNEKEKCMTPYSIGLDIGTNSVGWAVITDNYKVPSKKMKVLGNTS KKYIKKNLLGVLLFDSGITAEGRRLKRTARRRYTRRRNRILYLQEIFSTEMATLDDAFFQ RLDDSFLVPDDKRDSPYIFGNLVEEKVYHDEFPTIYHLRKYLDSTKKADRLRLVYALA HMIKYRGHFLIEGEFNSKNNDIQNFQDFLDTYNAIFESDLSLENSKQLEEVKDKISKL EKKDRILKLPGEKNSGIFSEFLKLVGNQADFRKCFNLDEKASLHFSKESYDEDETLTLL GYIGDDYSDVFLKAKKLYDAILLGLTVTDNETEAPLSSAMIKRYNEHKEDLALLKEYI RNISLKYNEVFKDDTKNGYAGYIDGKTNQEDFYVYLKNLLAEFEGADYFLEKIDREDFL RKQRTFDNGSIPYQIHLQEMRAILDKQAKFYFPLAKNKERIEKILTFRIPIYVYVGLARGN SDFAWSIRKRNEKITPWNFEDVIDKESSAEAFINRMTSFDLYLPEEKVLPKHSLLYETFN VYNELTKVRFIAESMRDYQFLDSKQKKDIVRLYFKDKRKVTDKDIIEYLHAIYGYDGIEL KGIEKQFNSSLSTYHLLNIINDKEFLDDSSNEAIIIEIHTLTIFEDREMIKQRLSKFE NIFDKSVLKKLSRRHYTGWGKLSAKLINGIRDEKSGNTILDYLIDDGISNRNFMQLIHDD ALSFKKKIKQAQIIGDEDKGNIKEVVKSLPGSPAIIKKGILQSIKIVDELVKVMGGRKPES IVVEMARENQYTNQGSNSQQLRLEKSLKELGSKILKENIPAKLSKIDNNALQNDRLY LYYLQNGKDMYTGDDLDIDRLSNYDIDHIIPAFLKDNSIDNKVLVSSASNRGKSDDFPS LEVVKRRTFWYQLLKSLLISQKFDNLTKAERGGLLPEDKAGFIQRQLVETRQITKHVA RLLDEKFNNKKDENNRVTRVKIITLSTLVSQFRKDFELYKVREINDFHHAHDAYLNAV IASALLKKYPKLEPEFVYGDYPKYNSFRERKSATEKVYFYSNIMNIFKKSISLADGRVIE RPLIEVNEETGESVWNKESDLATVRRVLSYPQVNVVKKVEEQNHGLDRGKPKGLFNANL S SKPKPNSNENLVGAKEYLDPKKYGGYAGISNSFAVLVKGTEKGAKKKITNVLEFQGISI LDRINYRKDKLNFLEKGYKDIELIIELPKYSLEFSDGSRRMLASILSTNNKRGEIHKG

			NQIFLSQKFVKLLYHAKRISNTINENHRKYVENHKKFEELFYYLEFNENYVGAKKNGK LLNSAFQSWQNHSIDELCSSFIGPTGSEKGLFELTSRGSAADFEFLGVKIPRYRDYTPS SLLKDATLIHQSVTGLYETRIDLAKLGEG
<i>Sutter ella wadsw orthen sis Cas9</i>	E7H313	6	MTQSERRFSCSIGIDMGAKYTGIFYALFDREELPTNLNSKAMTLVMPETGPRYVQAQRT A VRHRLRGQKRYTLARKLAFLVDDMIKKQEKRLTDEEWKRGREALSGLLKRRGYSRPNA D GEDLTPLENVRADVFAAHPAFSTYFSEVRS�AEQWEEFTANISNVEKFLGDPNIPADKEF IEFAVAEGLIDKTEKKAYQSALSTLRANANVLTGLRQMGHKPRSEYFKAIEADLKKSRL AKINEAFGGAERLARLLGNLSNLQLRAERWYFNAPDIMKDRGWEPDRFKKTLVRAFKFF H PAKDQNKQHLELIKQIENSEDIIETLCTLDPNRTIPPYEDQNNRRPPLDQTLLSPEKLT RQYGEIWKTWSARLTSAEPTLAPAAEILERSTDRKSRVAVNGHEPLPTLAYQLSYALQRA FDRSKALDPYALRALAAGSKSNKLTSALENCIGGQNVKTFLDCAARRYREADDAKV G LWFDNADGLLERSDLHPPMKKKILPLLVANILQTDETTQKQFLDEIWRKQIKGRETVASR CARIETVRKSFGGGFNIAYNATAQYREVNKLPRNAQDKELLTIRDVAETADFIAANLGLS DEQKRKFANPFSLAQFYTLIETEVSGFSATTAVHLENAWRMTIKDAVINGETVRAAQCS RLPAETARPFDDLVRRLVDRQAWEIARVSTDIQSKVDFSNGIVDSIFVEENKFEFSAS VADLKKNKRVKDKMLSEAEKLETRWLIKNERIKKASRGTCPYTGDRLAEGGEIDHILPRS LIKDARGIVNAEPNLIYASSRGNQLKKNQRYSLDLKANYRNEIFKTSNIAAITAEIED VVTKLQQTHRLKFFDLLNEHEQDCVRHALFLDDGSEARDAVLELLATQRRTRVNGTQIW M IKNLANKIREELQNWCKTTNNRLHFQAAATNVSDAKNLRLKLAQNQPDFEKPDIQPIASH SIDALCSFAVGSADAERDQNGFDYLDGKTVLGLYPQSCEVIHLQAKPQEEKSHFDSVAIF KEGIYAEQFLPIFTLNEKIWIGYETLNAKGERCGAIEVSGKQPKELLEMLAPFFNKPVGD LSAHATYRILKKPAYEFLAKAALQPLSAEEKRLAALLDALRYCTSRKSLMSLFMAANGKS LKKREDVLKPKLFQLKVELKGEKSFKLNGSLTPVKQDWLRICDSPELADAFGKPCSADE LTSKLARIWKRPVMRDLAHAPVRREFSLPAIDNPSSGGFRIRRTNLFGNELYQVHAINAKK YRGFASAGSNVDWSKGILFNLQHENLTECGGRFITSADVTMPSEWRKVVAEDNLSIWI A PGTEGRRYVRVETTFIQASHWFEQSVENWAITSPLSLPASFKVDKPAEFQKAVGTELSEL LGQPRSEIFIENVGNAKHIRFWYIVVSSNKKMNESYNNVSKS
<i>Filifact or alocis Cas9</i>	D6GRK 4	7	MTKEYYLGLDVGTNSVGWAVTDSQYNLCKFKKKDMWGIRLFESANTAKDRRLQRGNRR RL ERKKQRIDLLQEIFSPEICKIDPTFFIRLNESSLHLEDKSNDFKYPLFIEKDYSIDIEYYK EFPTIFHLRKHLIESEEKQDIRLIYALHNIKTRGHFLIDGDLQSAKQLRPILDTFLLS LQEEQNLSVLSSENQKDEYEEILKNRSIAKSEKVKKLNLFESDELEKEEKKAQSAVIE NFCKFIVGNKGDVCKFLRVSKEELEIDSFSFSEGKYEDDIVKNLEEKVPEKVYLFEQMKA MYDWNILVDILETEEYISFAKVKQYEHKTNLRLRLDILKYCTKDEYNRMFNDEKEAGS YTAYVGKLKKNKKYWIEKKRNPEEFYKSLGKLKDIEPLKEDLEVTMMIEECKNHTLL PIQKNKDNQVIPHQVHEVELKKILENAKKYYSFLTETDKDGYSVVQKIESIFRFRIPYYV GPLSTRHQEKGSNVWMVRKPGREDRIYPWNMEEIDFEKSNNFITRMTNKCTYLIGEDV LPKHSLLYSKYMVLNELLNNVKVRGKKLPTSLKQKVFEDLFENKSKVTGKNLLEYLQIQDK DIQIDDLSGFDKDFKTSLSYLDKFKQIFGEEIEKESIQNMIEDIIKWITYGNDKEMLK RVIRANYSNQLTEEQMKKITGFQYSGWGNFSKMFLKGISGSDVSTGETFDITAMWETDN NLMQILSKKFTFMDNVEDFNSGKVGKIDKITYDSTVKEMFLSPENKRAVWQTQIVAAEIK

			KVMGCEPKKIFIEMARGGEKVKKRTKSRAQELLE YAAACEEDCRELIKEIEDRDERDFNS MKLFLYYTQFGKCMYSGDDIDINELIRGNSKWDRDHIYPQSKIKDDSIDNLVLVNKTYNA KKSNELLEDIQKKMHSFWLSLLNKKLITKSKYDRLTRKGDFTDEELSGFIARQLVETRO STKAIADIFKQIYSSEVVYKSSLVSDFRKKPLNLYKSRRVNDYHHAKEYLNIVGVNVY NKKFTSNPIQWMKKNRDNTNYSLNKVFEHDVVINGEVIWEKCTYHEDTNTYDGGTLDRIRK IVERDNILYTEYAYCEKGELFNATIONKNGNSTVSLKKGLDVKKYGGYFSANTSIFYSLIE FEDKKGDRARHIIGVPIYIANMLEHSPSAFLEYCEQKGYQNVIRLVEKIKKNSLLIINGY PLRIRGENEVDTSFKRAIQLKLDQKNYELVRNIEKFLEKYVEKKGNYPIDENRDHITHEK MNQLYEVLLSKMKKFNKKGMADPSDRIEKS PKFIKLEDLIDKINVINKMLNLLRCDNDT KADLSLIELPKNAGSFVVKNTIGKSKILVNQSVTGLYENRREL
<i>Lactobacillus johnsonii</i> Cas9	F4AF10	8	MTKIKDDYIVGLDIGTDCSGWVAMNSNNDILKLGKTAIGSRLFEGGKSAERRLFRTH RRIKRRRWRLKLLLEFFDPYMAEVDPYFFARLKESGLSPLDKRKT VSSIVFPTS AEDKKF YDDYPTIYHLRYKLMTEDEKFDLREVYLAIHIIKYRGNFLYNTSVKDFKASKIDVKSSI EKLNELYENLGLDLNVEFNISNTAEIEKVLKDKQIFKRDKVKKIAELFAIKTDNKEQSKR IKDISKQVANAVLGYKTRFDTIALKEISKDELSDWNFKLS DIDADSKFEALMGNL DENEQ AILLTIKELFNEVTNLNGIVEDGNTLSESMINKYNDHRDDLKLLKEVIENHIDRKKAKELA LAYDLYVNNRHGQLLQAKKKLGKIKPRSKEDFYKVVNKNLDDSRASKEIKKKIELDSFMP KQRTNANGVIPYQLQLELDKIIENQSKYYPFLKEINPVSSHLKEAPYKLDLIRFRVPY YVGPLISPNESTKDIQTKKNQNFAMWIRKEEGRITPWNFDQKVDRIESANKFIKRMTTKD TYLFGEDVLPANSLLYQKFTVLNELLNIRINGKRISVDLKQEIYENLFKKHTTVTVKKLE NYLKENHNLVKVEIKGLADEKKFNSGLTTYNRFKNLNIFDNQIDDLKYRNDFEKIIEWST IFEDKSIYKEKLRSDWLNEKQINALSNIRLQGWGRLSKKLLAQLHDHNGQTIIQLWDS QNNFMQIVTQADFKDAIAKANQNLLVATSVEDILNNAYTSPANKKAIRQVIKVVDDIVKA ASGKVPKQIAIEFTRDADENPKRSQTRGSKLQKVYKDLSTELASKTIAEELNEAIKDKKL VQDKYYLYFMYQLGRDAYTGEPINIDEIQKYDIDHILPQSFIKDDALDNRLVSVRAVNNGK SDNVPVKLFGNEMAANLGMTIRKMWEWKNIGLISKTKYNNLLTDPHINKYKSAGFIRR QLVETSQIIKL VSTILQSRYPNTEIITVKAKYNHYLREKFDLYKSREVNDYHHAYLS AICGNLLYQNYPNLRPFVYGYKFFSSDPDEKAIFNKTRKFSFISQLLNKSENSKEI AKKLKRAYQFKYMLVSRETETRDQEMFKMTVYPRFSDHTVKAPRNLPKKMGMSPIYGG G YTNNSDAYMVIVRIDKKKGTEYKILGIPTREL VNLKKA EKEDHYKSYLKEILTPRILYNK NGKRDKKITSFEIVKSKIPYKQVIQDGDKKFMLGSSTYVYNAKQLTLSTESMAITNNFD KDS DENDALIKAYDEILDVKYLP LFDINKFREKLHSGREKFIKLSLEDKKDTILKVLE GLHDNAVMTKIPTIGLSTPLGFMQFPNGVILSENAKLIYQSPTGLFKKSVKISDL
<i>Campylobacter lari</i> Cas9	G1UFN 3	9	MRILGFDIGINSIGWAFVENDELKDCGVRIPTKAENPKNKESLALPRRNARSSRRRLKRR KARLIAIKRILAKELKNYKDYAADGELPKAYEGSLASVYELRYKALTQNLTKDLARV ILHIAKHRGYMNKNEKKSNDACKGKILSALKNNALKLENYQSVGEYFYKEFFQKYKNTK NFIKIRNTKDYN NCVLSSDLEKELKLILEKQKEFGYNYSED FINEILKVAFFQRPLKDF SHLVGACTFFEEEEKRACKNSYSAWEFVALTKIINEIKSLEKISGEIVPTQTINEVLNLIL DKGSITYKKFRSCINLHESISFKSLKYDKENAENAKLIDFRKLVEFKKALGVHLSRQEL DQISTHITLIKDNVCLKTVLEKYNLSNEQINNLLIEFN DYINLSFKALGMILPLMREGK RYDEACEIANLKPKTVDEKKDFLPAFCDSIFAHELSPVNVRAISEYRKVLNALLKKYKG VHKHLELARDVGLSKKAREKIEKEQKENQAVNAWALKECENIGLKASAKNILKLWKE QKEICIYSGNKISIEHLKDEKALEVDHIYPYSRFSFDDSFINKVLVFTKENQEKL NKT PFE AFGKNIEKWSKIQT LAQNLPYKKKNKILDENFKDKQQEDFISRNLDTRYIATLIAKYTK EYLNFLLLSENEANANKSGEKGSKIHVQTISGMLTSVLRHTWGFDDKDRNNHLHHALDAI IVAYSTNSIIKAFSDFRKNQELLKARFYAKELTSDNYKHQVKFFEPFKSFREKILSKIDE

			IFVSKPPRKRARRALHKDTFHSENKIIDKCSYNSKEGLQIALSCGRVRKIGTKYVENDTI VRVDIFKKQNKFYAIIPIYAMDFALGILPNKIVITGDKNNNPKQWQTIDESYEFCSLYK NDLILLQKKNMQEPEFAYYNDFSISTSSICVEKHDNKFENLTSNQKLLFSNAKEGSVKVE SLGIQNLKVFEKYIITPLGDKIKADFQPRENISLKTSSKYGLR
<i>Parvib aculu m lavam entivor ans Cas9</i>	A7HP89	10	MERIFGFDIGTTSIGFSVIDYSSTQSAGNIQRLGVRIFPEARDPDGTPLNQRRQKRM RQLRRRRIRRKALNETLHEAGFLPAYGSADWPVVMADPEYELRRRGLEEGLSAYEFGRA I YHLAQHRHFKGRELEESDTPDPDVDEKEAANERAATLKALKNEQTTLGAWLARRPPS DR KRGIAHRNVVAEEFERLWEVQSKFHPALKSEEMRARISDTIFAQRPVFWRKNTLGECH F MPGEPLCPKGSWLSQQRRMLEKLNLAAGGNARPLDAEERDAILSKLQQQASMSWPG VR SALKALYKQRGEPGAEKSLKFNLELGGESKLLGNALEAKLADMFGPDWPAHPRKQEIRH A VHERLWAADYGETPDKKRVILSEKDRKAHREAAANSFVADFGITGEQAAQLQALKLPTG WEPYSIPALNLFALAELEKGERFGALVNGPDWEGWRRNTNPHRNQPTGEILDKLPSPASK E ERERISQLRNPTVVRTQNELRKVVNNLIGLYGKPDRIEVRDVGKSKREREIISGIR RNEKQRRKATEDLIKNGIANPSRDDVEKWILWKEGQERCPYTDQIGFNALFREGRYEV E HIWPRSRSFDNSPRNKTLCKRDVNIEKGNRMPFEAFGHDEDRWSAIQIRLQGMVSAKG GT GMSPGKVKRFLAKTMPEDFAARQLNDTRYAAKQILAQLKRLWPDMPGPEAPVKVEAVTG QV TAQLRKLWTLNNILADDGEKTRADHRHHAIDALTVACTHPGMTNKLSTRYWQLRDDPRAE K PALTPPWDTIRADAEKAVSEIVVSHRVRKKVSGPLHKETTYGDTGTDIKTKSGTYRQFVT RKKIESLSKGELDEIRDPRKEIVAHHVAGRGDPKKAFFPPYPCVSPGPGPEIRKVRILT SK QQLNLMAQTGNGYADLGSNHIIAYRLPDGKADFEIVSLFDASRRLAQRNPVQRTTRADG ASFVMSLAAGEAIMPEGSKKGIWIVQGVWASGQVVLERDADHSTTRPMPNPILKDD AKKVSIDPIGRVRSND
<i>Mycop lasma gallise pticum Cas9</i>	A0A0F6 CLF2	11	MNNSIKSKPEVTIGLDLGVGSVGWAIVDNETNIIHHLGSRLFSQAKTAEDRRSFRGVRRL IRRRKYKLKRFVNLWIKYNSYFGFNKEDILNNYQEQQKLHNTVLNLKSEALNAKIDPKA LSWILHDYLNKRGHFDYEDNRDFNVYPTKELAKYFDKYGYKGIIDSKEDNDNLEEELTK YKFSNKHWELEVKKVLSNQTGLPEKFKEEYESLSYVRNYSEGPGSINSVSPYGIYHLDE KEGKVVQKYNNIWDKTIGKCNIFPDEYRAPKNSPIAMIFNEINELSTIRSYIYLTGWFI NQEFKKAYLNKLLDLIKTNGEKPIDARQFKKLREETIAESIGKETLKDVENEEKLEKED HKWKLGKGLNTNGKIQYNDLSSLAKFVHKLKQHLKLDLLEDQYATLDKINFLQSLFVY LGKHLRYNNRVDSANLKEFSDSNRFLFERVLQEQQDGLFKLFEQTDKDDKILAQTHSLST KAMLLAITRMTNLDNEDNQNNDKGWNFEAIKNFDQKFIDITKTNNNLSLKQDKRYLDD RFINDAILSPGVKRILREATKVFNAILKQFSQEYDVTKVVIELARELSEEKELENNKNYK KLIKNSDKISEGLKALDIAEDKIEDILKSPTKSYKVLLWLQQDHIDPYSQKEIAFEDIL TKTEKTEIDHIIPYSISFDDSSSNKLLVLAESNQAKSNQTPYEFITSGNAGIKWEDYEAY CRKFKDGDTSLLDSTQRSKKFAKMMKTDTSKYDIGFLARNLNDTRYATIVFRDALKDYA NNHLVEDKPMFKVVCINGGVTSFLRKNFDSWYAKKDRDKNIHHAVDASIISFSNKTKT LFDQLTQFADYKLFKNTDGSWKKIDPKTGTVTEVDENWKQIRVRNQVSKIAEEDKCIQ

			DSNIERKARYSRKIENKTNISLFNDTVYSAKKVGYDDQIKRKNLKTLDIDESVEENKNSK VKKQFVYRKLNVNLSLLNNDKLADLFAEKEDILMYRANPWVINLAEQIFNEYTENRRIKSQ NVFGKYMLDLTKEFPEKFSEAFVKSMLRNKTAIYNVEKKVVRHRIKRLKLSSELKENKL SNVIIRSKNESGTKLSYQDTINSVALMIMRSIDPTAKKQYIRVPLNTLNLHLGDHDFDLH NIDAYLKPKPKFVKYKANEIGDEYKPWRVLISGSLLIHKRDKKLMYISSFQNLNDLIEIK NLIEYKENVSDPKKKKKKASQILRSLSTILNDYILLDAKDNFDILGLSKNRIDEILNS KLDLDKIAK
<i>Trep nema dentic ola Cas9</i>	Q73QW 6	12	MKKEIKDYFLGLDVGTGSVGWAVTDDYKLLKANRKDLWGMRCFETAETAEVRRRLHRG AR RRIERRKKRIKLLQELFSQEIAKTDEGFFQRMKESPFYAEDKTILQENTLFNDKDFADKT YHKAYPTINHLIKAWIENKVKPDPRLLYLACHNIIKKRGHFLFEGDFDSENQFDTSIQAL FEYLREDMEVIDADSQVKEILKDSSLKNSEKQSRLNKLGLKPSDKQKKAITNLISGN KINFADLYDNPDLKDAEKNSISFSKDDFDALSDDLASILGDSFELLLKAKAVYNCSVLSK VIGDEQYLSFAKVKIYEHKTDLTCLKNVIKKHFPKDYKKVFGYNKNEKNNNNYSYGVGV CKTKSKKLIINNSVNQEDFYKFLKTILSAKSEIKEVNDILTEIETGTFLPKQISKSNAEI PYQLRKMELEKILSNAEKHFSFLKQKDEKGLSHSEKIIMLLTFKIPYYIGPINDNHKKFF PDRCWVVKKEKSPSGKTPWNFFDHIDKEKTAFAFITSRTNFCTYLVGESVLPKSSLLYS EYTVLNEINNLIIDGKNICDIKQKIYEDLFKKYKKITQKQISTFIKHEGICNKTDE VIILGIDKECTSSLKSYIELKNIFGKQVDEISTKNMLEEIRWATYDEGEGKTILTKI KAIEYGYCSDEQIKILNLKFSGWGRLSRKFLFTVTSEMPGFSEPVNIITAMRETQNNLM ELLSSEFTFTENIKKINSGFEDAQKQFSYDGLVKPLFLSPSVKMLWQTLKLVEKISHIT QAPPKKIFIEMAKGAELPARTKRLKILQDLYNNCKNDADAFSSEIKDLSGKIENEDNL RLRSDKLYLYYTQLGKCMYCGKPIEIGHVFDTSNYDIDHIYPQSKIKDSSISNRVLVCSS CNKNKEDKYPLKSEIQSKQRGFWNFLQRNNFISLEKLNRLTRATPISDDETAQFIARQLV ETRQATKVAKVLEKMFPEKIVYSAETVSMFRNKFDIVKCREINDFHHAHDAYLNIVV GNVYNTKFTNNPWNFIKEKRDNPKIADTYNYYKVFYDVKRNNITAWKEGKTIITVKDML KRNTPIYTRQAACKKGELFNQTIMKKGLGQHPLKKEGPFNSISKYGGYNKVSAAYTLIE YEEKGNKIRSLETIPLYLVKDIQKDQDVLKSYLTDLLGKKEFKILVPKIKINSLLKINGF PCHITGKTNDSSFLLRPVAVQFCCSNNEVLYFKKIIRFSEIRSQREKIGKITSIPYEDLSFRS YIKENLWKKTKNDEIGEKEFYDLLQKKNLEIYDMLLTKHKDTIYKKRPN SATIDILVKGK EKFKSLIENQFEVILEILKLSATRNVSDLQHIGGSKYSGVAKIGNKISSLDNCILYQ SITGIFEKRIDLKLV
<i>Lach nosp race ae bacte rium ND2 006 Cpf1</i>	A0A182 DWE3	13	AASKLEKFTNCYSLSKTLRFKAIPVGKTQENIDNKRLLVEDEKRAEDYKGVKKLLDRYYL SFINDVLHSIKLKNLNNYISLFRKKTREKENKELENLEINLRKEIAKAFKGAAGYKSLF KKDIETILPEAADDKDEIALVNSFNGFTTAFTGFFDNRENMFSEEAKSTSIAFRNINEN LTRYISNMDIFEKVDAIFDKHEVQEIKEKILNSDYDVEDFFEGEFFNFVLTQEGIDVYNA IIGGFVTESGEKIKGLNEYINLYNAKTKQALPKFKPLYKQVLSDRESLSFYGEGYTSDEE VLEVFRNTLNKNSEIFSSIKKLEKLFKNFDEYSSAGIFVKNGPAISTISKDIFGEWNLIR DKWNAEYDDIHLKKKAVVTEKYEDDRRSFKKIGSFSLEQLQEYADADLSVVEKLKEIII QKVDEIYKVYGSSEKLFDAFVLEKSLKKNDVAVAIMKDLLDSVKSFENYKAFEGEGKE TNRDESFYGDFVLAIDILLKVDHIYDAIRNYVTQKPYSKDKFKLYFQNPQFMGGWDKDKE TDYRATILRYGSKYYLAIMDKYAKCLQKIDKDDVNGNYEKINYKLLPGPNKMLPKVFFS KKWMAYYNPSEDIQKIYKNGTFKKGDMFNLNDCHKLIDFFKDSISRYPKWSNAYDFNFSE TEKYKDIAGFYREVVEEQYKVSFESASKKEVDKLVVEEGKLYMFQIYNKDFSDKSHGTPNL HTMYFKLLFDENNHGQIRLSGGAELFMRRASLKKEELVHPANSPIANKNPDNPKKTTTL SYDVYKDKRFSEDQYELHIPIAINKCPKNIFKINTEVRVLLKHDDNPYVIGIDRGERNLL YIVVVDGKGNIVEQYSLNEIINNFGIRIKTDYHSLLDKKEKERFEARQNWTSIENIKEL

			KAGYISQVVHKICELVEKYDAVIALEDLNSGFKNSRVKVEKQVYQKFEKMLIDKLNMYMD KKSNPCATGGALKGYQITNKFESFKSMSTQNGFIFYIPAWLTSKIDPSTGFVNLLKTKYT SIADSKKFISDFRIMYVPEEDLFEFALDYKNFSRTDADYIKWKLYSYGNRIRIFAAAK KNNVFAWEEVCLTSAYKELFNKYGINYQQGDIRALLCEQSDKAFYSSFMALMSLMLQMR N SITGRTDVDFLISPVKNSDGIFYDSRNYEAQENAILPKNADANGAYNIARKVLWAIGQFK KAEDKLDKVKIAISNKEWLEYAQTSVK
--	--	--	---

5

It should be noted that all the amino acid modifications described herein were derived in the context of the amino acid sequence of *Streptococcus pyogenes* Cas9 (SEQ ID NO:1) through extensive experimentation. However, one skilled in the art would readily appreciate that these modifications are transferrable to other CRISPR-Cas nucleases, which share significant

10 homology with the *Streptococcus pyogenes* Cas9.

To determine whether these modifications could be transferred to another nuclease, the equivalent or corresponding counterpart residues between the two amino acid sequences are determined, typically based on the sequence or structural homology between the sequences.

15

To establish homology, the amino acid sequence of the *Streptococcus pyogenes* Cas9 is directly compared to the sequence of a second nuclease. After aligning the sequences, using one or more of the homology alignment programs well known in the art, such as CLUSTALW (for example using conserved residues between species), allowing for necessary insertions and deletions in order to maintain alignment (i. e. avoiding the elimination of conserved residues

20

through arbitrary deletion and insertion), the residues equivalent or corresponding to particular amino acid residues in the primary sequence of the *Streptococcus pyogenes* Cas9 are defined. Alignment of conserved residues preferably should conserve at least 20%, preferably at least 30%, preferably at least 40%, preferably at least 50%, preferably at least 60%, preferably at least 70%, preferably at least 80%, preferably at least 90%, preferably at least 95%, more

25

preferably at least 96%, even more preferably at least 98%, even more preferably at least 99% and most preferably 100% of residues. Equivalent or corresponding counterpart residues may also be defined by determining structural homology between the nucleases, that is at the level of tertiary structure for the nucleases or certain domains thereof whose structures have been determined. In this case, equivalent or corresponding residues are defined as those, for which

30

the atomic coordinates of two or more of the mainchain atoms of a particular amino acid residue of the *Streptococcus pyogenes* Cas9 (N on N, CA on CA, C on C and O on O) are within 0.13 nm and preferably 0.1 nm after alignment. Alignment is achieved after the best model has been oriented and positioned to give the maximum overlap of atomic coordinates of non-hydrogen protein atoms of the proteins. Regardless of how equivalent or corresponding residues are determined, and regardless of the identity of the *Streptococcus pyogenes* Cas9 in which the

35 modifications are introduced, what is meant to be conveyed is that the polypeptides according to

5 the present invention may be constructed into any other CRISPR-Cas nuclease which has a significant sequence or structural homology with the *Streptococcus pyogenes* Cas9.

In various embodiments, the polypeptide comprises at least one combination of deletions selected from the group consisting of Δ REC1-a Δ HNH (Δ 510-655 Δ 775-909), Δ REC1-c Δ HNH
 10 (Δ 662-710 Δ 775-909), Δ REC2 Δ HNH (Δ 180-308 Δ 775-909), Δ REC2-d Δ HNH (Δ 199-283 Δ 775-909), Δ HNH Δ RuvCIII-b (Δ 775-909 Δ 1002-1074), Δ REC3 Δ HNH (Δ 498-712 Δ 775-909), Δ REC3 Δ HNH Δ RuvCIII-b (Δ 498-712 Δ 775-909 Δ 1002-1074), Δ REC2 Δ REC3 Δ HNH Δ RuvCIII-b (Δ 180-308 Δ 498-712 Δ 775-909 Δ 1002-1074), and Δ REC2 Δ HNH Δ RuvCIII-b (Δ 180-308 Δ 775-909 Δ 1002-1074).

15 In various embodiments, the polypeptide further comprises at least one missense mutation selected from the group consisting of G12R, T13K, T13R, N14K, N497K, T657K, T657R, N767K, T770K, T770R, Q920K, Q920R, S1109R, D1135K, D1135R, S1338R and combinations thereof, wherein the position numbering is in accordance with SEQ ID NO:1.

20 In various embodiments, the polypeptide comprises at least one combination of modifications selected from the group consisting of Δ REC1-c Δ HNH G12R, Δ REC1-c Δ HNH T13K, Δ REC1-c Δ HNH T13R, Δ REC1-c Δ HNH N14K, Δ REC1-c Δ HNH N497K, Δ REC1-c Δ HNH T657K, Δ REC1-c Δ HNH T657R, Δ REC1-c Δ HNH N767K, Δ REC1-c Δ HNH T770K, Δ REC1-c Δ HNH T770R,
 25 Δ REC1-c Δ HNH Q920K, Δ REC1-c Δ HNH Q920R, Δ REC1-c Δ HNH S1109R, Δ REC1-c Δ HNH D1135K, Δ REC1-c Δ HNH D1135R, Δ REC1-c Δ HNH S1338R, Δ REC1-c Δ HNH T657R T13K, Δ REC1-c Δ HNH T657R N497K, Δ REC1-c Δ HNH T657R T770K, Δ REC1-c Δ HNH T657R Q920K, Δ REC1-c Δ HNH T657R S1109R, Δ REC1-c Δ HNH T657R D1135K, Δ REC2 Δ HNH Δ RuvCIII-b T13K, Δ REC2 Δ HNH Δ RuvCIII-b T657K, Δ REC2 Δ HNH Δ RuvCIII-b T657R, Δ REC2
 30 Δ HNH Δ RuvCIII-b 770K, Δ REC2 Δ HNH Δ RuvCIII-b Q920K, Δ REC2 Δ HNH Δ RuvCIII-b S1109R, and Δ REC2 Δ HNH Δ RuvCIII-b D1135K.

Without wishing to be bound to any theory, it is believed that the polypeptide according to the
 35 afore-mentioned aspects can be used as a site-directed modifying polypeptide, i.e. a CRISPR-Cas protein, and functions as an essential component of a CRISPR-Cas system in site-directed modification of a target DNA. Since smaller Cas proteins are easier to package in delivery vehicles, higher delivery efficiencies can be expected.

In general, "CRISPR-Cas system" refers collectively to transcripts and other elements involved
 40 in the expression of or directing the activity of CRISPR-associated ("Cas") genes, including sequences encoding a Cas gene, a tracr (trans-activating CRISPR) sequence (e.g. tracrRNA or

an active partial tracrRNA), a tracr-mate sequence (encompassing a “direct repeat” and a tracrRNA-processed partial direct repeat in the context of an endogenous CRISPR-Cas system), a guide sequence (also referred to as a “spacer” in the context of an endogenous CRISPR-Cas system), or other sequences and transcripts from a CRISPR locus. In some embodiments, one or more elements of a CRISPR-Cas system is derived from a type I, type II, or type III CRISPR-Cas system. In some embodiments, one or more elements of a CRISPR-Cas system are derived from an organism comprising an endogenous CRISPR-Cas system, such as *Streptococcus pyogenes*. In general, a CRISPR-Cas system is characterized by elements that promote the formation of a CRISPR complex at the site of a target sequence (also referred to as a protospacer in the context of an endogenous CRISPR-Cas system). In the context of formation of a CRISPR complex, “target sequence” refers to a sequence to which a guide sequence is designed to have complementarity, where hybridization between a target sequence and a guide sequence promotes the formation of a CRISPR complex. A target sequence may comprise any polynucleotide, such as DNA or RNA polynucleotides. In some embodiments, a target sequence is located in the nucleus or cytoplasm of a cell.

In preferred embodiments of the invention, the CRISPR-Cas system is a type II CRISPR-Cas system and the Cas enzyme is Cas9, which binds to a target genomic locus and catalyzes DNA cleavage. Enzymatic action by Cas9 derived from *Streptococcus pyogenes* or any closely related Cas9 generates double stranded breaks at target site sequences, which hybridize to guide sequences of length 17 to 24 nucleotides inclusive. The target locus contains a protospacer-adjacent motif (PAM), whose sequence varies according to the bacteria species where the Cas9 originates from. For the Cas9 from *Streptococcus pyogenes*, the PAM sequence is NGG/NRG, while for the Cas9 from *Staphylococcus aureus*, the PAM sequence is NNGRRT. The PAM of Cas9 follows the 17-24 nucleotides of the target sequence. CRISPR activity through Cas9 for site-specific DNA recognition or cleavage is defined in part by the guide sequence, the tracr sequence, and the PAM sequence. More aspects of the CRISPR-Cas system are described in Karginov and Hannon, The CRISPR-Cas system: small RNA-guided defense in bacteria and archae, *Mol Cell* 2010, Jan. 15; 37(1): 7.

The term “site-directed modifying polypeptide” as used herein refers to a polypeptide that binds RNA and is targeted to a specific DNA sequence. A site-directed modifying polypeptide as described herein is targeted to a specific DNA sequence by the RNA molecule to which it is bound. The RNA molecule comprises a sequence that is complementary to a target sequence within the target DNA, thus targeting the bound polypeptide to a specific location within the target DNA (the target sequence).

5 The term “cleavage” as used herein refers to the breakage of the covalent backbone of a DNA molecule. Cleavage can be initiated by a variety of methods including, but not limited to, enzymatic or chemical hydrolysis of a phosphodiester bond. Both single-stranded cleavage and double-stranded cleavage are possible, and double-stranded cleavage can occur as a result of two distinct single-stranded cleavage events. DNA cleavage can result in the production of
10 either blunt ends or staggered ends. In certain embodiments, a complex comprising a DNA-targeting RNA and a site-directed modifying polypeptide is used for targeted double-stranded DNA cleavage.

15 The term “nuclease” as used herein refers to an enzyme possessing catalytic activity for DNA cleavage.

The RNA molecule that binds to the site-directed modifying polypeptide and targets the polypeptide to a specific location within the target DNA is referred to herein as the “DNA-targeting RNA” or “DNA-targeting RNA polynucleotide” (also referred to herein as a “guide RNA”
20 or “gRNA”). A subject DNA-targeting RNA comprises two segments, a “DNA-targeting segment” and a “protein-binding segment.” By “segment” it is meant a segment/section/region of a molecule, e.g., a contiguous stretch of nucleotides in an RNA. A segment can also mean a region/section of a complex such that a segment may comprise regions of more than one molecule. For example, in some cases the protein-binding segment of a DNA-targeting RNA is
25 one RNA molecule and the protein-binding segment therefore comprises a region of that RNA molecule. In other cases, the protein-binding segment of a DNA-targeting RNA comprises two separate molecules that are hybridized along a region of complementarity. As an illustrative, non-limiting example, a protein-binding segment of a DNA-targeting RNA that comprises two separate molecules can comprise (i) base pairs 40-75 of a first RNA molecule that is 100 base
30 pairs in length; and (ii) base pairs 10-25 of a second RNA molecule that is 50 base pairs in length. The definition of “segment,” unless otherwise specifically defined in a particular context, is not limited to a specific number of total base pairs, is not limited to any particular number of base pairs from a given RNA molecule, is not limited to a particular number of separate molecules within a complex, and may include regions of RNA molecules that are of any total
35 length and may or may not include regions with complementarity to other molecules.

The DNA-targeting segment (or “DNA-targeting sequence”) comprises a nucleotide sequence that is complementary to a specific sequence within a target DNA (the complementary strand of the target DNA). The protein-binding segment (or “protein-binding sequence”) interacts with a
40 site-directed modifying polypeptide. When the site-directed modifying polypeptide is a Cas9 or Cas9 related polypeptide, e.g. the polypeptide described herein, site-specific cleavage of or recruitment of an effector domain to the target DNA occurs at locations determined by both (i)

- 5 base-pairing complementarity between the DNA-targeting RNA and the target DNA; and (ii) a short motif (referred to as the protospacer adjacent motif (PAM)) in the target DNA.

The protein-binding segment of a subject DNA-targeting RNA comprises two complementary stretches of nucleotides that hybridize to one another to form a double stranded RNA duplex
10 (dsRNA duplex).

In some embodiments, a subject nucleic acid (e.g., a DNA-targeting RNA, a nucleic acid comprising a nucleotide sequence encoding a DNA-targeting RNA; a nucleic acid encoding a site-directed polypeptide; etc.) comprises a modification or sequence that provides for an
15 additional desirable feature (e.g., modified or regulated stability; subcellular targeting; tracking, e.g., a fluorescent label; a binding site for a protein or protein complex; etc.). Non-limiting examples include: a 5' cap (e.g., a 7-methylguanylate cap (m7G)); a 3' polyadenylated tail (i.e., a 3' poly(A) tail); a riboswitch sequence (e.g., to allow for regulated stability and/or regulated accessibility by proteins and/or protein complexes); a stability control sequence; a sequence
20 that forms a dsRNA duplex (i.e., a hairpin)); a modification or sequence that targets the RNA to a subcellular location (e.g., nucleus, mitochondria, chloroplasts, and the like); a modification or sequence that provides for tracking (e.g., direct conjugation to a fluorescent molecule, conjugation to a moiety that facilitates fluorescent detection, a sequence that allows for fluorescent detection, etc.); a modification or sequence that provides a binding site for proteins
25 (e.g., proteins that act on DNA, including transcriptional activators, transcriptional repressors, DNA methyltransferases, DNA demethylases, histone acetyltransferases, histone deacetylases, and the like); and combinations thereof.

In some embodiments, a DNA-targeting RNA comprises an additional segment at either the 5' or 3' end that provides for any of the features described above. For example, a suitable third segment can comprise a 5' cap (e.g., a 7-methylguanylate cap (m7G)); a 3' polyadenylated tail (i.e., a 3' poly(A) tail); a riboswitch sequence (e.g., to allow for regulated stability and/or regulated accessibility by proteins and protein complexes); a stability control sequence; a sequence that forms a dsRNA duplex (i.e., a hairpin)); a sequence that targets the RNA to a
35 subcellular location (e.g., nucleus, mitochondria, chloroplasts, and the like); a modification or sequence that provides for tracking (e.g., direct conjugation to a fluorescent molecule, conjugation to a moiety that facilitates fluorescent detection, a sequence that allows for fluorescent detection, etc.); a modification or sequence that provides a binding site for proteins (e.g., proteins that act on DNA, including transcriptional activators, transcriptional repressors,
40 DNA methyltransferases, DNA demethylases, histone acetyltransferases, histone deacetylases, and the like); and combinations thereof.

5 A subject DNA-targeting RNA and a subject site-directed modifying polypeptide (i.e., site-directed polypeptide) form a complex (i.e., bind via non-covalent interactions). The DNA-targeting RNA provides target specificity to the complex by comprising a nucleotide sequence that is complementary to a sequence of a target DNA. The site-directed modifying polypeptide of the complex provides the site-specific activity. In other words, the site-directed modifying
10 polypeptide is guided to a target DNA sequence (e.g. a target sequence in a chromosomal nucleic acid; a target sequence in an extrachromosomal nucleic acid, e.g. an episomal nucleic acid, a minicircle, etc.; a target sequence in a mitochondrial nucleic acid; a target sequence in a chloroplast nucleic acid; a target sequence in a plasmid; etc.) by virtue of its association with the protein-binding segment of the DNA-targeting RNA.

15 In some embodiments, a subject DNA-targeting RNA comprises two separate RNA molecules (RNA polynucleotides: an “activator-RNA” and a “targeter-RNA”, see below) and is referred to herein as a “double-molecule DNA-targeting RNA” or a “two-molecule DNA-targeting RNA.” In other embodiments, the subject DNA-targeting RNA is a single RNA molecule (single RNA
20 polynucleotide) and is referred to herein as a “single-molecule DNA-targeting RNA,” a “single-guide RNA,” or an “sgRNA.” The term “DNA-targeting RNA” or “gRNA” is inclusive, referring both to double-molecule DNA-targeting RNAs and to single-molecule DNA-targeting RNAs (i.e., sgRNAs).

25 An exemplary two-molecule DNA-targeting RNA comprises a crRNA-like (“CRISPR RNA” or “targeter-RNA” or “crRNA” or “crRNA repeat”) molecule and a corresponding tracrRNA-like (“trans-acting CRISPR RNA” or “activator-RNA” or “tracrRNA”) molecule. A crRNA-like molecule (targeter-RNA) comprises both the DNA-targeting segment (single stranded) of the DNA-targeting RNA and a stretch (“duplex-forming segment”) of nucleotides that forms one-half of the
30 dsRNA duplex of the protein-binding segment of the DNA-targeting RNA. A corresponding tracrRNA-like molecule (activator-RNA) comprises a stretch of nucleotides (duplex-forming segment) that forms the other half of the dsRNA duplex of the protein-binding segment of the DNA-targeting RNA. In other words, a stretch of nucleotides of a crRNA-like molecule are complementary to and hybridize with a stretch of nucleotides of a tracrRNA-like molecule to
35 form the dsRNA duplex of the protein-binding domain of the DNA-targeting RNA. As such, each crRNA-like molecule can be said to have a corresponding tracrRNA-like molecule. The crRNA-like molecule additionally provides the single stranded DNA-targeting segment. Thus, a crRNA-like and a tracrRNA-like molecule (as a corresponding pair) hybridize to form a DNA-targeting RNA. The exact sequence of a given crRNA or tracrRNA molecule is characteristic of the
40 species in which the RNA molecules are found. A subject double-molecule DNA-targeting RNA can comprise any corresponding crRNA and tracrRNA pair. A subject double-molecule DNA-targeting RNA can comprise any corresponding crRNA and tracrRNA pair.

5

The term “activator-RNA” is used herein to mean a tracrRNA-like molecule of a double-molecule DNA-targeting RNA. The term “targeter-RNA” is used herein to mean a crRNA-like molecule of a double-molecule DNA-targeting RNA. The term “duplex-forming segment” is used herein to mean the stretch of nucleotides of an activator-RNA or a targeter-RNA that contributes to the formation of the dsRNA duplex by hybridizing to a stretch of nucleotides of a corresponding activator-RNA or targeter-RNA molecule. In other words, an activator-RNA comprises a duplex-forming segment that is complementary to the duplex-forming segment of the corresponding targeter-RNA. As such, an activator-RNA comprises a duplex-forming segment while a targeter-RNA comprises both a duplex-forming segment and the DNA-targeting segment of the DNA-targeting RNA. Therefore, a subject double-molecule DNA-targeting RNA can be comprised of any corresponding activator-RNA and targeter-RNA pair.

10

15

20

The polypeptide of the invention may be prepared recombinantly or produced by in vitro transcription/translation. The polypeptide may also be prepared synthetically, preferably using a commercially available peptide synthesizer. Methods of synthetic peptide synthesis include, but are not limited to liquid-phase peptide synthesis and solid-phase peptide synthesis. Methods to produce peptides synthetically and according protocols are well-known in the art (Nilsson, B L et al. (2005) Annu Rev BiophysBiomolStruct, 34, 91).

25

The CRISPR-Cas nuclease activity of the polypeptide can be determined by a variety of assays known in the art, for example by Surveyor cleavage assay or by deep sequencing the targeted genomic loci.

30

The CRISPR-Cas DNA-binding activity of the polypeptide can be determined by a variety of assays known in the art, depending on what effector domain is fused to the polypeptide. If a transcriptional regulatory domain is fused to the polypeptide, a fluorescence reporter assay or a qRT-PCR experiment to measure the expression of endogenous genes can be performed, as described in the examples of the present application.

35

It should be noted that proteins comprising the polypeptide described herein are also encompassed in the present application.

40

In a third aspect, the invention relates to a nucleic acid molecule comprising or consisting of a nucleic acid encoding a polypeptide described herein.

In various embodiments, the nucleic acid molecule is a recombinant expression vector.

5 The term “nucleic acid molecule” or “nucleic acid” as used herein refers to a polymeric form of nucleotides of any length, either deoxyribonucleotides or ribonucleotides, or analogs thereof. Nucleic acid molecules may have any three-dimensional structure, and may perform any function, known or unknown. The term also encompasses nucleic-acid-like structures with synthetic backbones, see, e.g., Eckstein, 1991; Baserga et al., 1992; Milligan, 1993; WO
10 97/03211; WO 96/39154; Mata, 1997; Strauss-Soukup, 1997; and Samstag, 1996. A nucleic acid molecule may comprise one or more modified nucleotides, such as methylated nucleotides and nucleotide analogs. If present, modifications to the nucleotide structure may be imparted before or after assembly of the polymer. The sequence of nucleotides may be interrupted by non-nucleotide components. A polynucleotide may be further modified after polymerization,
15 such as by conjugation with a labeling component.

As used herein, a “vector” is a tool that allows or facilitates the transfer of an entity from one environment to another. It is a replicon, such as a plasmid, phage, or cosmid, into which another DNA segment may be inserted so as to bring about the replication of the inserted segment.
20 Generally, a vector is capable of replication when associated with the proper control elements. In general, the term “vector” refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. Vectors include, but are not limited to, nucleic acid molecules that are single-stranded, double-stranded, or partially double-stranded; nucleic acid molecules that comprise one or more free ends, no free ends (e.g. circular); nucleic acid
25 molecules that comprise DNA, RNA, or both; and other varieties of polynucleotides known in the art. One type of vector is a “plasmid,” which refers to a circular double stranded DNA loop into which additional DNA segments can be inserted, such as by standard molecular cloning techniques. Another type of vector is a viral vector, wherein virally-derived DNA or RNA sequences are present in the vector for packaging into a virus (e.g. retroviruses, replication
30 defective retroviruses, adenoviruses, replication defective adenoviruses, and adeno-associated viruses). Viral vectors also include polynucleotides carried by a virus for transfection into a host cell. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g. bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (e.g., non-episomal mammalian vectors) are integrated into
35 the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors can direct the expression of genes to which they are operatively-linked. Such vectors are referred to herein as “expression vectors.” Common expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. Further discussion of vectors is provided herein.

40 Recombinant expression vectors can comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression

5 vectors include one or more regulatory elements, which may be selected on the basis of the host cells to be used for expression, that is operatively-linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory element(s) in a manner that allows for expression of the nucleotide sequence (e.g. in an in vitro transcription/translation system or
10 in a host cell when the vector is introduced into the host cell). With regards to recombination and cloning methods, mention is made of U.S. patent application Ser. No. 10/815,730, published Sep. 2, 2004 as US 2004-0171156 A1, the contents of which are herein incorporated by reference in their entirety.

15 In various embodiments, the vector further comprises regulatory elements for controlling expression of the polypeptide.

The term "regulatory element" is intended to include promoters, enhancers, internal ribosomal entry sites (IRES), and other expression control elements (e.g. transcription termination signals, such as polyadenylation signals and poly-U sequences). Such regulatory elements are described, for example, in Goeddel, Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, Calif. (1990). Regulatory elements include those that direct constitutive expression of a nucleotide sequence in many types of host cell and those that direct expression of the nucleotide sequence only in certain host cells (e.g., tissue-specific regulatory sequences). A tissue-specific promoter may direct expression primarily in a desired tissue of interest, such as muscle, neuron, bone, skin, blood, specific organs (e.g. liver, pancreas), or particular cell types (e.g. lymphocytes). Regulatory elements may also direct expression in a temporal-dependent manner, such as in a cell-cycle dependent or developmental stage-dependent manner, which may or may not also be tissue or cell-type specific. In some
20 embodiments, a vector comprises one or more pol III promoter (e.g. 1, 2, 3, 4, 5, or more pol III promoters), one or more pol II promoters (e.g. 1, 2, 3, 4, 5, or more pol II promoters), one or more pol I promoters (e.g. 1, 2, 3, 4, 5, or more pol I promoters), or combinations thereof. Examples of pol III promoters include, but are not limited to, U6 and H1 promoters. Examples of pol II promoters include, but are not limited to, the retroviral Rous sarcoma virus (RSV) LTR promoter (optionally with the RSV enhancer), the cytomegalovirus (CMV) promoter (optionally with the CMV enhancer) [see, e.g., Boshart et al, Cell, 41:521-530 (1985)], the SV40 promoter, the dihydrofolate reductase promoter, the β -actin promoter, the phosphoglycerol kinase (PGK) promoter, and the EF1 α promoter. Also encompassed by the term "regulatory element" are enhancer elements, such as WPRE; CMV enhancers; the R-U5' segment in LTR of HTLV-I
35 (Mol. Cell. Biol., Vol. 8(1), p. 466-472, 1988); SV40 enhancer; and the intron sequence between exons 2 and 3 of rabbit β -globin (Proc. Natl. Acad. Sci. USA., Vol. 78(3), p. 1527-31, 1981). It will be appreciated by those skilled in the art that the design of the expression vector can
40

5 depend on such factors as the choice of the host cell to be transformed, the level of expression desired, etc. A vector can be introduced into host cells to thereby produce transcripts, proteins, or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein. With regards to regulatory sequences, mention is made of U.S. patent application Ser. No. 10/491,026, the contents of which are incorporated by reference herein in their entirety. With
10 regards to promoters, mention is made of PCT publication WO 2011/028929 and U.S. application Ser. No. 12/511,940, the contents of which are incorporated by reference herein in their entirety.

Vectors can be designed for expression of CRISPR transcripts (e.g. nucleic acid transcripts,
15 proteins, or enzymes) in prokaryotic or eukaryotic cells. For example, CRISPR transcripts can be expressed in bacterial cells such as *Escherichia coli*, insect cells (using baculovirus expression vectors), yeast cells, or mammalian cells. Suitable host cells are discussed further in Goeddel, Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, Calif. (1990). Alternatively, the recombinant expression vector can be transcribed and
20 translated in vitro, for example using T7 promoter regulatory sequences and T7 polymerase.

Vectors may be introduced and propagated in a prokaryote or prokaryotic cell. In some embodiments, a prokaryote is used to amplify copies of a vector to be introduced into a eukaryotic cell or as an intermediate vector in the production of a vector to be introduced into a
25 eukaryotic cell (e.g. amplifying a plasmid as part of a viral vector packaging system). In some embodiments, a prokaryote is used to amplify copies of a vector and express one or more nucleic acids, such as to provide a source of one or more proteins for delivery to a host cell or host organism. Expression of proteins in prokaryotes is most often carried out in *Escherichia coli* with vectors containing constitutive or inducible promoters directing the expression of either
30 fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, such as to the amino terminus of the recombinant protein. Such fusion vectors may serve one or more purposes, such as: (i) to increase expression of recombinant protein; (ii) to increase the solubility of the recombinant protein; and (iii) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression
35 vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Example fusion expression vectors include pGEX (Pharmacia Biotech Inc, Smith and Johnson, 1988. Gene 67: 31-40), pMAL (New
40 England Biolabs, Beverly, Mass.) and pRIT5 (Pharmacia, Piscataway, N.J.) that fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

5

Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amrann et al., (1988) Gene 69:301-315) and pET 11d (Studier et al., Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, Calif. (1990) 60-89).

10

In some embodiments, a vector is a yeast expression vector. Examples of vectors for expression in yeast *Saccharomyces cerevisiae* include pYepSec1 (Baldari, et al., 1987. EMBO J. 6: 229-234), pMFa (Kuijan and Herskowitz, 1982. Cell 30: 933-943), pJRY88 (Schultz et al., 1987. Gene 54: 113-123), pYES2 (Invitrogen Corporation, San Diego, Calif.), and picZ (Invitrogen Corp, San Diego, Calif.).

15

In some embodiments, a vector drives protein expression in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., SF9 cells) include the pAc series (Smith, et al., 1983. Mol. Cell. Biol. 3: 2156-2165) and the pVL series (Lucklow and Summers, 1989. Virology 170: 31-39).

20

In some embodiments, a vector is capable of driving expression of one or more sequences in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, 1987. Nature 329: 840) and pMT2PC (Kaufman, et al., 1987. EMBO J. 6: 187-195). When used in mammalian cells, the expression vector's control functions are typically provided by one or more regulatory elements. For example, commonly used promoters are derived from polyoma, adenovirus 2, cytomegalovirus, simian virus 40, and others disclosed herein and known in the art. For other suitable expression systems for both prokaryotic and eukaryotic cells see, e.g., Chapters 16 and 17 of Sambrook, et al., Molecular Cloning: A Laboratory Manual. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989.

30

In some embodiments, the recombinant mammalian expression vector can direct expression of the nucleic acid preferentially in a particular cell type (e.g., tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art.

35

Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert, et al., 1987. Genes Dev. 1: 268-277), lymphoid-specific promoters (Calame and Eaton, 1988. Adv. Immunol. 43: 235-275), in particular promoters of T cell receptors (Winoto and Baltimore, 1989. EMBO J. 8: 729-733) and immunoglobulins (Baneiji, et al., 1983. Cell 33: 729-740; Queen and Baltimore, 1983. Cell 33: 741-748), neuron-specific promoters (e.g., the neurofilament promoter; Byrne and Ruddle, 1989. Proc. Natl. Acad. Sci. USA 86: 5473-5477), pancreas-specific promoters (Edlund, et al., 1985. Science 230: 912-916), and mammary gland-specific promoters (e.g., milk whey promoter; U.S. Pat. No. 4,873,316 and

40

European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, e.g., the murine hox promoters (Kessel and Gruss, 1990. Science 249: 374-379) and the α -fetoprotein promoter (Campes and Tilghman, 1989. Genes Dev. 3: 537-546). With regards to these prokaryotic and eukaryotic vectors, mention is made of U.S. Pat. No. 6,750,059, the contents of which are incorporated by reference herein in their entirety. Other embodiments of the invention may relate to the use of viral vectors, with regards to which mention is made of U.S. patent application Ser. No. 13/092,085, the contents of which are incorporated by reference herein in their entirety. Tissue-specific regulatory elements are known in the art and in this regard, mention is made of U.S. Pat. No. 7,776,321, the contents of which are incorporated by reference herein in their entirety.

In some embodiments, a regulatory element is operably linked to one or more elements of a CRISPR-Cas system so as to drive expression of the one or more elements of the CRISPR-Cas system. In general, CRISPRs constitute a family of DNA loci that are usually specific to a particular bacterial species. The CRISPR locus comprises a distinct class of interspersed short sequence repeats (SSRs) that were recognized in *E. coli* (Ishino et al., J. Bacteriol., 169:5429-5433 [1987]; and Nakata et al., J. Bacteriol., 171:3553-3556 [1989]), and associated genes. Similar interspersed SSRs have been identified in *Haloferax mediterranei*, *Streptococcus pyogenes*, *Anabaena*, and *Mycobacterium tuberculosis* (See, Groenen et al., Mol. Microbiol., 10:1057-1065 [1993]; Hoe et al., Emerg. Infect. Dis., 5:254-263 [1999]; Masepohl et al., Biochim. Biophys. Acta 1307:26-30 [1996]; and Mojica et al., Mol. Microbiol., 17:85-93 [1995]). The CRISPR loci typically differ from other SSRs by the structure of the repeats, which have been termed short regularly spaced repeats (SRSRs) (Janssen et al., OMICS J. Integ. Biol., 6:23-33 [2002]; and Mojica et al., Mol. Microbiol., 36:244-246 [2000]). In general, the repeats are short elements that occur in clusters that are regularly spaced by unique intervening sequences with a substantially constant length (Mojica et al., [2000], supra). Although the repeat sequences are highly conserved between strains, the number of interspersed repeats and the sequences of the spacer regions typically differ from strain to strain (van Embden et al., J. Bacteriol., 182:2393-2401 [2000]). CRISPR loci have been identified in more than 40 prokaryotes (See e.g., Jansen et al., Mol. Microbiol., 43:1565-1575 [2002]; and Mojica et al., [2005]) including, but not limited to *Aeropyrum*, *Pyrobaculum*, *Sulfolobus*, *Archaeoglobus*, *Halocarcula*, *Methanobacterium*, *Methanococcus*, *Methanosarcina*, *Methanopyrus*, *Pyrococcus*, *Picrophilus*, *Thermoplasma*, *Corynebacterium*, *Mycobacterium*, *Streptomyces*, *Aquifex*, *Porphyromonas*, *Chlorobium*, *Thermus*, *Bacillus*, *Listeria*, *Staphylococcus*, *Clostridium*, *Thermoanaerobacter*, *Mycoplasma*, *Fusobacterium*, *Azarcus*, *Chromobacterium*, *Neisseria*, *Nitrosomonas*, *Desulfovibrio*, *Geobacter*, *Myxococcus*, *Campylobacter*, *Wolinella*, *Acinetobacter*, *Erwinia*, *Escherichia*, *Legionella*, *Methylococcus*, *Pasteurella*, *Photobacterium*, *Salmonella*, *Xanthomonas*, *Yersinia*, *Treponema*, and *Thermotoga*.

5

The polypeptides and nucleic acid molecules described above are preferably non-naturally occurring. The present invention provides the polypeptides and nucleic acid molecules preferably in recombinant, synthetic, isolated, and/or purified form.

10

In a fourth aspect, to a composition comprising the polypeptide and/or the nucleic acid molecule described above.

The composition preferably also comprises other essential components of a CRISPR-Cas system, e.g. a guide RNA.

15

In a fifth aspect, the invention relates to a method of site-specific engineering of a target DNA, the method comprising contacting the target DNA with a CRISPR-Cas system comprising the polypeptide, nucleic acid molecule or composition described herein, and recruiting an effector domain to the target site.

20

Without wishing to be bound to any theory, the polypeptide described herein comprises: (a) an RNA-binding portion that interacts with the DNA-targeting RNA of a CRISPR-Cas system; and (b) DNA-binding portion that interacts with the target DNA. It may also comprise an activity portion that exhibits nuclease activity that creates a single-stranded nick or double stranded break in the target DNA of the CRISPR-Cas system, wherein the site of the double stranded break is determined by the DNA-targeting RNA, wherein preferably the contacting occurs under conditions that are permissive for nonhomologous end joining (NHEJ) or homology-directed repair (HDR), preferably HDR, and the target DNA is cleaved and rejoined to produce a modified DNA sequence.

30

It is within the knowledge of the skilled person to utilize the polypeptide, the recombinant expression vector, or the composition of the invention effectively in the site-directed engineering of a target DNA, including genome editing or base editing, in the light of the afore-described aspects and the common general knowledge regarding CRISPR-Cas systems.

35

In various embodiments, the target DNA is extrachromosomal (e.g. an episomal sequence, a minicircle sequence, a mitochondrial sequence, a chloroplast sequence, etc.) or is part of a chromosome in vitro, in an isolated or cultured cell, or in a cell in an organism.

40

In various embodiments, the cell is selected from the group consisting of: an archaeal cell, a bacterial cell, a eukaryotic cell, a eukaryotic single-cell organism, a somatic cell, a germ cell, a stem cell, a plant cell, an algal cell, an animal cell, an invertebrate cell, a vertebrate cell, a fish

- 5 cell, a frog cell, a bird cell, a mammalian cell, a pig cell, a cow cell, a goat cell, a sheep cell, a rodent cell, a rat cell, a mouse cell, a non-human primate cell, and a human cell. In various embodiments, the organism is selected from the group consisting of: an archaea, a bacterium, a eukaryotic single-cell organism, an algae, a plant, an animal, an invertebrate, a fly, a worm, a cnidarian, a vertebrate, a fish, a frog, a bird, a mammal, an ungulate, a rodent, a rat, a mouse,
10 and a non-human primate.

Also encompassed in the present application are cells and non-human organisms genetically engineered by the method described herein.

- 15 In a sixth aspect, the invention relates to use of the polypeptide, nucleic acid molecule or composition described herein in site-directed nucleic acid modification, preferably as a component in a CRISPR-Cas system.

- The present invention is further illustrated by the following examples. However, it should be
20 understood, that the invention is not limited to the exemplified embodiments.

5

EXAMPLES**Materials and methods****Example1: CRISPR-Cas Nucleases with Missense Mutations and Deletions**

10

Overall, the inventors aimed to truncate different Cas enzymes, including Cas9 from *Streptococcus pyogenes*, rationally guided by the solved crystal structures of these proteins (Nishimasu, H. *et al. Cell* 156, 935-949 (2014); Nishimasu, H. *et al. Cell* 162, 1113-1126 (2015); Yamano, T. *et al. Cell* 165, 949-962 (2016)). For example, disrupting protein secondary structures (α -helices and β -sheets) was avoided.

15

As a proof-of-concept, the inventors deleted different parts of a catalytically dead Cas9 gene from *Streptococcus pyogenes* (dSpCas9) fused to a tripartite activator, VP64-p65-Rta (VPR) (Chavez, A. *et al. Nat Methods* 12, 326-328, doi:10.1038/nmeth.3312 (2015)), and then tested the ability of each truncated dSpCas9-VPR fusion to activate the expression of a zsGreen reporter that was under the control of a minimal promoter with multiple binding sites for the enzyme. Using published crystal structures of SpCas9 as a guide, the inventors focused primarily on eliminating regions of the protein that appear to be globular and are not in contact with either the DNA or the gRNA. Furthermore, the start and end residues of each target region must be spatially close together so that the deleted portion of the protein could be replaced with a short linker. The regions of the dSpCas9 protein that were deleted are summarized in Figures 1-5.

20

25

The results of the reporter assays for all the truncated constructs (except those involving REC3) are shown in Figure 6. Within the nuclease lobe (Figure 6), it was found that removal of the entire HNH nuclease domain did not affect the ability of dSpCas9-VPR to activate gene expression significantly. However, removal of the RuvCI, RuvCII, or RuvCIII domain resulted in complete abolishment of reporter gene activation. Subsequently, smaller fragments of the RuvCIII domain were deleted and it was found that removal of one of the fragments, RuvCIII-b, did not hamper the regulatory activity of dSpCas9-VPR significantly.

30

35

For the REC lobe, it was examined whether any segments of the REC1, REC2, or PI domains, which are involved in DNA recognition, may be removed without causing the dSpCas9 scaffold to lose its function completely (Figure 6). Previous mutational analysis of the REC1 domain found that $\Delta 97-150$ and $\Delta 312-409$, which are regions of Cas9 that interact with a repeat:anti-repeat duplex of the gRNA, abolished DNA cleavage of the enzyme. Hence, deletions that were located towards the back portion of the REC1 domain were tested instead and it was found that

40

5 $\Delta 510-655$ ($\Delta \text{REC1-a}$) and $\Delta 662-710$ ($\Delta \text{REC1-c}$) did not affect the transcriptional activation function of dSpCas9-VPR significantly. Additionally, while a previous study showed that a Cas9 mutant lacking the REC2 domain retained only 50% of the DNA cleavage activity of the wildtype enzyme, it was found that removal of the REC2 domain from dSpCas9-VPR reduced reporter gene activation by only around 20%. Surprisingly, deletions of smaller fragments of the REC2
10 domain were more deleterious than removal of the entire domain. It was also found that removal of any segment of the PI domain abolished the activity of dSpCas9-VPR completely.

The inventors further asked whether different deletions could be combined to yield smaller but functional dSpCas9-VPR enzymes (Figure 6). While pairing of ΔHNNH with ΔREC2 gave the
15 same extent of reporter gene activation as ΔREC2 alone, combining ΔHNNH with either $\Delta \text{REC1-a}$ or $\Delta \text{REC1-c}$ significantly reduced the activity of dSpCas9-VPR, although each of the individual deletions did not have an obvious effect. Nevertheless, it was found that five of the combinatorial deletions tested still resulted in smaller dSpCas9-VPR constructs that gave at least 30% of the activity of the full-length enzyme. Of these, the smallest scaffold contains three
20 separate deletions, namely ΔHNNH , ΔREC2 , and $\Delta \text{RuvCIII-b}$.

In the reporter experiments, the inventors tested two or three binding sites for the enzyme as well as different time durations post-transfection before performing the assays. Overall, it was observed that a longer time duration led to a more significant improvement in the extent of gene
25 activation than an additional binding site (Figure 7). Some of the truncated enzymes appeared to be non-functional when assayed at 24 hours post-transfection, but were still able to activate zsGreen expression at 48 hours after transfection (Figure 6). The slower kinetics may be a result of poorer binding between the enzyme and the DNA when different parts of the dSpCas9 scaffold have been removed. Taken together, the data indicated the possibility to shrink the size
30 of dSpCas9 and possibly other Cas9 proteins for DNA recognition.

Next, the inventors evaluated the performance of all the truncated constructs (except those involving REC3) in a native cellular context. It was tested whether the various constructs can activate four different endogenous genes (ACTC, ASCL1, MIAT, and TTN) in HEK293T cells.
35 From quantitative real-time PCR (qRT-PCR) experiments (Figure 8), it was found that ΔHNNH , $\Delta \text{RuvCIII-b}$, $\Delta \text{HNNH} \Delta \text{RuvCIII-b}$, and $\Delta \text{REC1-c}$ performed as well as the original full-length enzyme at all four endogenous genomic loci. Unexpectedly, it was also observed that several truncated constructs, despite being able to switch on the reporter gene robustly, did not perform well in the native cellular context, either failing to activate some of the genes completely or up-regulating gene expression at levels much lower than the full-length dCas9-VPR protein. This
40 may be because there are multiple binding sites for the enzyme in the promoter driving the

5 reporter gene and that the sequence for this artificial binding site is known to be able to recruit the CRISPR-Cas9 system very efficiently.

Most recently, an important non-catalytic domain within the Cas9 REC lobe (termed REC3) was reported to recognize target complementarity and govern the cleavage specificity of the
10 enzyme. Importantly, in the absence of REC3, it was found that the binding affinity of Cas9 for the target DNA was still close to wildtype, although cleavage rate decreased by 1000-fold. Hence, the inventors tested whether deletion of REC3, either by itself or in combination with other truncations, will affect the ability of the Cas9 protein to function as a robust scaffold. The inventors fused the VPR tripartite activator to different REC3 deletion constructs and performed
15 reporter assays and qRT-PCR experiments (Figure 9). Overall, it was found that removal of REC3 alone marginally affected the ability of dSpCas9-VPR to activate either the fluorescence reporter or two endogenous genes, MIAT and TTN. Removal of both REC3 and HNH further degraded the performance of the scaffold, but the dSpCas9-VPR enzyme was still able to activate the reporter and the two endogenous genes reasonably well. This indicates that the
20 Δ REC3 Δ HNH construct is a suitable candidate for further engineering and optimization.

It was hypothesized that the performance of the dCas9 scaffolds may be improved by enhancing the binding between each scaffold and the underlying DNA. Since the DNA backbone is negatively charged, the inventors sought to introduce positively charged amino
25 acids (lysine or arginine) into the dCas9 protein. From existing crystal structures of SpCas9, the inventors selected a handful of residues that are poorly conserved and thus may not be important in the CRISPR-Cas9 system, that can accommodate a mutation to either lysine or arginine based on computational simulations, and that are sufficiently close to the targeted DNA for further testing (Figure 10). The inventors introduced the mutations individually into the Δ HNH
30 Δ REC1-c construct and performed the reporter assay. Notably, it was observed that most of the selected mutations could enhance the performance of Δ HNH Δ REC1-c (Figure 11). The inventors also introduced a subset of the mutations into the Δ HNH Δ REC2 Δ RuvCIII-b construct and performed the reporter assay. It was observed that most of the tested mutations could enhance the performance of Δ HNH Δ REC2 Δ RuvCIII-b (Figure 13). The next steps are to
35 comprehensively evaluate the mutations in combinations (Figure 12) and to further test whether they could help to improve the ability of a truncated dCas9-VPR protein to activate gene expression in an endogenous cellular context.

The invention has been described broadly and generically herein. Each of the narrower species
40 and subgeneric groupings falling within the generic disclosure also form part of the invention. This includes the generic description of the invention with a proviso or negative limitation

5 removing any subject matter from the genus, regardless of whether or not the excised material is specifically recited herein. Other embodiments are within the following claims.

One skilled in the art would readily appreciate that the present invention is well adapted to carry out the objects and obtain the ends and advantages mentioned, as well as those inherent
10 therein. Further, it will be readily apparent to one skilled in the art that varying substitutions and modifications may be made to the invention disclosed herein without departing from the scope and spirit of the invention. The compositions, methods, procedures, treatments, molecules and specific compounds described herein are presently representative of preferred
15 embodiments are exemplary and are not intended as limitations on the scope of the invention. Changes therein and other uses will occur to those skilled in the art which are encompassed within the spirit of the invention are defined by the scope of the claims. The listing or discussion of a previously published document in this specification should not necessarily be taken as an acknowledgement that the document is part of the state of the art or is common general
20 knowledge.

The invention illustratively described herein may suitably be practiced in the absence of any element or elements, limitation or limitations, not specifically disclosed herein. Thus, for example, the terms "comprising", "including," "containing", etc. shall be read expansively and without limitation. The word "comprise" or variations such as "comprises" or "comprising" will
25 accordingly be understood to imply the inclusion of a stated integer or groups of integers but not the exclusion of any other integer or group of integers. Additionally, the terms and expressions employed herein have been used as terms of description and not of limitation, and there is no intention in the use of such terms and expressions of excluding any equivalents of the features shown and described or portions thereof, but it is recognized that various
30 modifications are possible within the scope of the invention claimed. Thus, it should be understood that although the present invention has been specifically disclosed by exemplary embodiments and optional features, modification and variation of the inventions embodied therein herein disclosed may be resorted to by those skilled in the art, and that such modifications and variations are considered to be within the scope of this invention.

35 The content of all documents and patent documents cited herein is incorporated by reference in their entirety.

5 CLAIMS

What is claimed is:

1. Polypeptide having at least 50% sequence identity to *Streptococcus pyogenes* Cas9 having the amino acid sequence set forth in SEQ ID NO:1 over its entire length, wherein said polypeptide comprises at least one deletion selected from the group consisting of Δ HNH (Δ 775-909), Δ RuvCIII-b (Δ 1002-1074), Δ REC1-a (Δ 510-655), Δ REC1-b (Δ 525-587), Δ REC1-c (Δ 662-710), Δ REC2 (Δ 180-308), Δ REC2-a (Δ 212-244), Δ REC2-b (Δ 244-276), Δ REC2-c (Δ 276-308), Δ REC2-d (Δ 199-283), Δ REC2-e (Δ 198-257), Δ REC2-f (Δ 235-286), Δ REC2-g (Δ 217-266), Δ REC3 (Δ 498-712) and combinations thereof, wherein the position numbering is in accordance with SEQ ID NO:1, and wherein the polypeptide has CRISPR-Cas DNA-binding activity.
2. The polypeptide of claim 1, wherein the polypeptide comprises at least one combination of deletions selected from the group consisting of Δ REC1-a Δ HNH (Δ 510-655 Δ 775-909), Δ REC1-c Δ HNH (Δ 662-710 Δ 775-909), Δ REC2 Δ HNH (Δ 180-308 Δ 775-909), Δ REC2-d Δ HNH (Δ 199-283 Δ 775-909), Δ HNH Δ RuvCIII-b (Δ 775-909 Δ 1002-1074), Δ REC3 Δ HNH (Δ 498-712 Δ 775-909), Δ REC3 Δ HNH Δ RuvCIII-b (Δ 498-712 Δ 775-909 Δ 1002-1074), Δ REC2 Δ REC3 Δ HNH Δ RuvCIII-b (Δ 180-308 Δ 498-712 Δ 775-909 Δ 1002-1074), and Δ REC2 Δ HNH Δ RuvCIII-b (Δ 180-308 Δ 775-909 Δ 1002-1074).
3. The polypeptide of claim 1 or 2, wherein the polypeptide further comprises at least one missense mutation selected from the group consisting of G12R, T13K, T13R, N14K, N497K, T657K, T657R, N767K, T770K, T770R, Q920K, Q920R, S1109R, D1135K, D1135R, S1338R and combinations thereof.
4. The polypeptide of any one of claims 1-3, wherein the polypeptide comprises at least one combination of modifications selected from the group consisting of Δ REC1-c Δ HNH G12R, Δ REC1-c Δ HNH T13K, Δ REC1-c Δ HNH T13R, Δ REC1-c Δ HNH N14K, Δ REC1-c Δ HNH N497K, Δ REC1-c Δ HNH T657K, Δ REC1-c Δ HNH T657R, Δ REC1-c Δ HNH N767K, Δ REC1-c Δ HNH T770K, Δ REC1-c Δ HNH T770R, Δ REC1-c Δ HNH Q920K, Δ REC1-c Δ HNH Q920R, Δ REC1-c Δ HNH S1109R, Δ REC1-c Δ HNH D1135K, Δ REC1-c Δ HNH D1135R, Δ REC1-c Δ HNH S1338R, Δ REC1-c Δ HNH T657R T13K, Δ REC1-c Δ HNH T657R N497K, Δ REC1-c Δ HNH T657R T770K, Δ REC1-c Δ HNH T657R Q920K, Δ REC1-c Δ HNH T657R S1109R, Δ REC1-c Δ HNH T657R D1135K, Δ REC2 Δ HNH Δ RuvCIII-b T13K, Δ REC2 Δ HNH Δ RuvCIII-b T657K, Δ REC2 Δ HNH Δ RuvCIII-b T657R, Δ REC2 Δ HNH Δ RuvCIII-b 770K, Δ REC2 Δ HNH Δ RuvCIII-b Q920K, Δ REC2 Δ HNH Δ RuvCIII-b S1109R, and Δ REC2 Δ HNH Δ RuvCIII-b D1135K.

5

5. The polypeptide of any one of claims 1-4, wherein the polypeptide has the amino acid sequence set forth in any one of SEQ ID NOs:14-30.

10

6. The polypeptide of any one of claims 1-5, wherein the polypeptide is further codon optimized for expression in a eukaryotic cell, preferably in a mammalian or human cell.

15

20

25

30

7. Polypeptide having at least 50% sequence identity to *Staphylococcus aureus* Cas9 having the amino acid sequence set forth in SEQ ID NO:2, *Neisseria meningitides* Cas9 having the amino acid sequence set forth in SEQ ID NO:3, *Acidaminococcus* sp. Cpf1 having the amino acid sequence set forth in SEQ ID NO:4, *Streptococcus thermophilus* Cas9 having the amino acid sequence set forth in SEQ ID NO:5, *Sutterella wadsworthensis* Cas9 having the amino acid sequence set forth in SEQ ID NO:6, *Filifactor alocis* Cas9 having the amino acid sequence set forth in SEQ ID NO:7, *Lactobacillus johnsonii* Cas9 having the amino acid sequence set forth in SEQ ID NO:8, *Campylobacter lari* Cas9 having the amino acid sequence set forth in SEQ ID NO:9, *Parvibaculum lavamentivorans* Cas9 having the amino acid sequence set forth in SEQ ID NO:10, *Mycoplasma gallisepticum* Cas9 having the amino acid sequence set forth in SEQ ID NO:11, *Treponema denticola* Cas9 having the amino acid sequence set forth in SEQ ID NO:12, or *Lachnospiraceae* bacterium ND2006 Cpf1 having the amino acid sequence set forth in SEQ ID NO:13 over its entire length, wherein said polypeptide comprises at least one deletion of said amino acid sequence, wherein said at least one deletion corresponds to at least one deletion of SEQ ID NO:1 selected from the group consisting of Δ HNH (Δ 775-909), Δ RuvCIII-b (Δ 1002-1074), Δ REC1-a (Δ 510-655), Δ REC1-b (Δ 525-587), Δ REC1-c (Δ 662-710), Δ REC2 (Δ 180-308), Δ REC2-a (Δ 212-244), Δ REC2-b (Δ 244-276), Δ REC2-c (Δ 276-308), Δ REC2-d (Δ 199-283), Δ REC2-e (Δ 198-257), Δ REC2-f (Δ 235-286), Δ REC2-g (Δ 217-266), Δ REC3 (Δ 498-712) and combinations thereof, wherein the position numbering is in accordance with SEQ ID NO:1, wherein the polypeptide has CRISPR-Cas DNA-binding activity.

35

8. The polypeptide of claim 7, wherein the polypeptide comprises at least one combination of deletions selected from the group consisting of Δ REC1-a Δ HNH (Δ 510-655 Δ 775-909), Δ REC1-c Δ HNH (Δ 662-710 Δ 775-909), Δ REC2 Δ HNH (Δ 180-308 Δ 775-909), Δ REC2-d Δ HNH (Δ 199-283 Δ 775-909), Δ HNH Δ RuvCIII-b (Δ 775-909 Δ 1002-1074), Δ REC3 Δ HNH (Δ 498-712 Δ 775-909), Δ REC3 Δ HNH Δ RuvCIII-b (Δ 498-712 Δ 775-909 Δ 1002-1074), Δ REC2 Δ REC3 Δ HNH Δ RuvCIII-b (Δ 180-308 Δ 498-712 Δ 775-909 Δ 1002-1074), and Δ REC2 Δ HNH Δ RuvCIII-b (Δ 180-308 Δ 775-909 Δ 1002-1074).

40

- 5 9. The polypeptide of claim 7 or 8, wherein the polypeptide further comprises at least one missense mutation selected from the group consisting of G12R, T13K, T13R, N14K, N497K, T657K, T657R, N767K, T770K, T770R, Q920K, Q920R, S1109R, D1135K, D1135R, S1338R and combinations thereof, wherein the position numbering is in accordance with SEQ ID NO:1.
- 10 10. The polypeptide of any one of claims 7-9, wherein the polypeptide comprises at least one combination of modifications selected from the group consisting of Δ REC1-c Δ HNH G12R, Δ REC1-c Δ HNH T13K, Δ REC1-c Δ HNH T13R, Δ REC1-c Δ HNH N14K, Δ REC1-c Δ HNH N497K, Δ REC1-c Δ HNH T657K, Δ REC1-c Δ HNH T657R, Δ REC1-c Δ HNH N767K, Δ REC1-c Δ HNH T770K, Δ REC1-c Δ HNH T770R, Δ REC1-c Δ HNH Q920K, Δ REC1-c Δ HNH Q920R, Δ REC1-c Δ HNH S1109R, Δ REC1-c Δ HNH D1135K, Δ REC1-c Δ HNH D1135R, Δ REC1-c Δ HNH S1338R, Δ REC1-c Δ HNH T657R T13K, Δ REC1-c Δ HNH T657R N497K, Δ REC1-c Δ HNH T657R T770K, Δ REC1-c Δ HNH T657R Q920K, Δ REC1-c Δ HNH T657R S1109R, Δ REC1-c Δ HNH T657R D1135K, Δ REC2 Δ HNH Δ RuvCIII-b T13K, Δ REC2 Δ HNH Δ RuvCIII-b T657K, Δ REC2 Δ HNH Δ RuvCIII-b T657R, Δ REC2 Δ HNH Δ RuvCIII-b 770K, Δ REC2 Δ HNH Δ RuvCIII-b Q920K, Δ REC2 Δ HNH Δ RuvCIII-b S1109R, and Δ REC2 Δ HNH Δ RuvCIII-b D1135K.
11. The polypeptide of any one of claims 7-10, wherein the polypeptide is further codon optimized for expression in a eukaryotic cell, preferably in a mammalian or human cell.
- 25 12. Nucleic acid molecule comprising or consisting of a nucleic acid encoding a polypeptide of any one of claims 1-11.
13. The nucleic acid molecule of claim 12, wherein the nucleic acid molecule is a recombinant expression vector.
- 30 14. The nucleic acid molecule of claim 13, wherein the vector further comprises at least one regulatory element for controlling expression of the polypeptide.
15. Composition comprising the polypeptide of any one of claims 1-11 or the nucleic acid molecule of any one of claim 12-14.
- 35 16. Method of site-specific engineering of a target DNA, the method comprising contacting the target DNA with a CRISPR-Cas system comprising the polypeptide of any one of claims 1-11, the nucleic acid molecule of any one of claims 12-14, or the composition of claim 15, wherein the CRISPR-Cas DNA-binding activity of the polypeptide recruits an effector domain to the target DNA.
- 40

5 17. The method of claim 16, wherein the target DNA is extrachromosomal or is part of a chromosome in vitro, in vivo, or in a cell.

18. The method of claim 16 or 17, wherein any CRISPR-Cas nuclease activity of the polypeptide introduces a single-stranded nick or a double stranded break in the target DNA.

10

19. Use of the polypeptide of any one of claims 1-11, the nucleic acid molecule of any one of claims 12-14, or the composition of claim 15 in site-directed nucleic acid modification, preferably as a component in a CRISPR-Cas system.

15

Figure 1

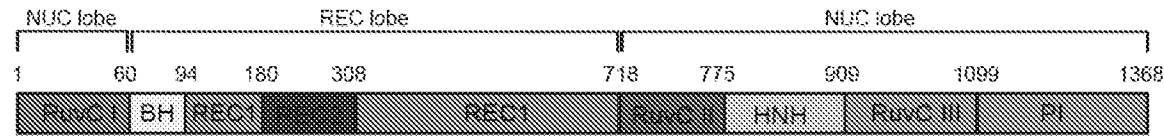
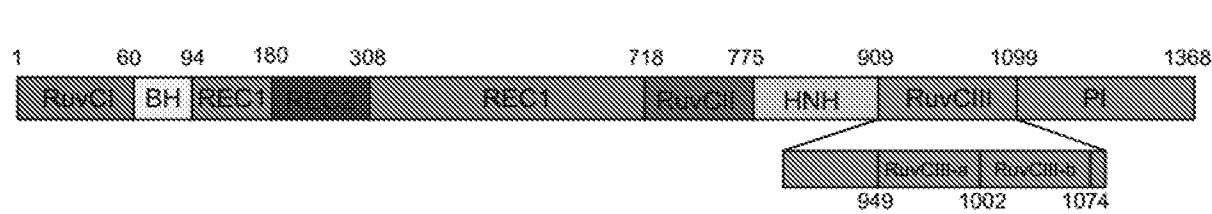


Figure 2

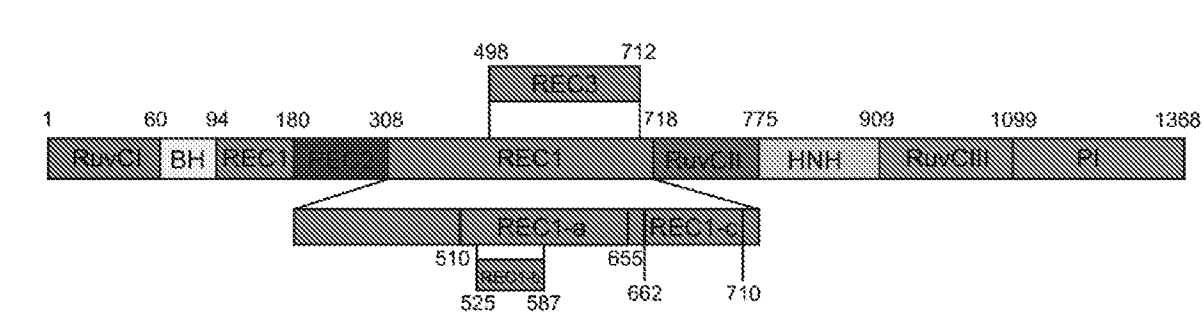


Δ RuvCIII

Δ RuvCIII-a

Δ RuvCIII-b

Figure 3



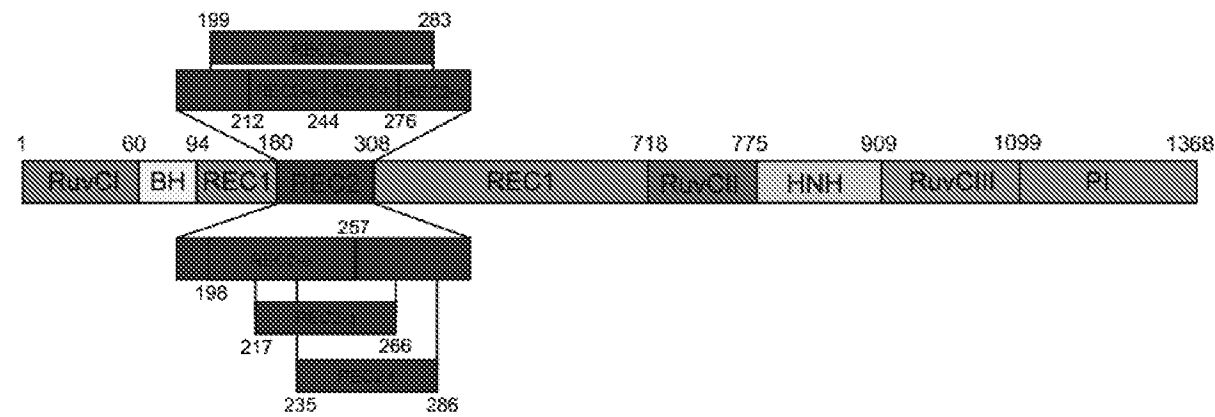
Δ REC1-a

Δ REC1-b

Δ REC1-c

Δ REC3

Figure 4



- Δ REC2
- Δ REC2-a
- Δ REC2-b
- Δ REC2-c
- Δ REC2-d
- Δ REC2-e
- Δ REC2-f
- Δ REC2-g

Figure 5

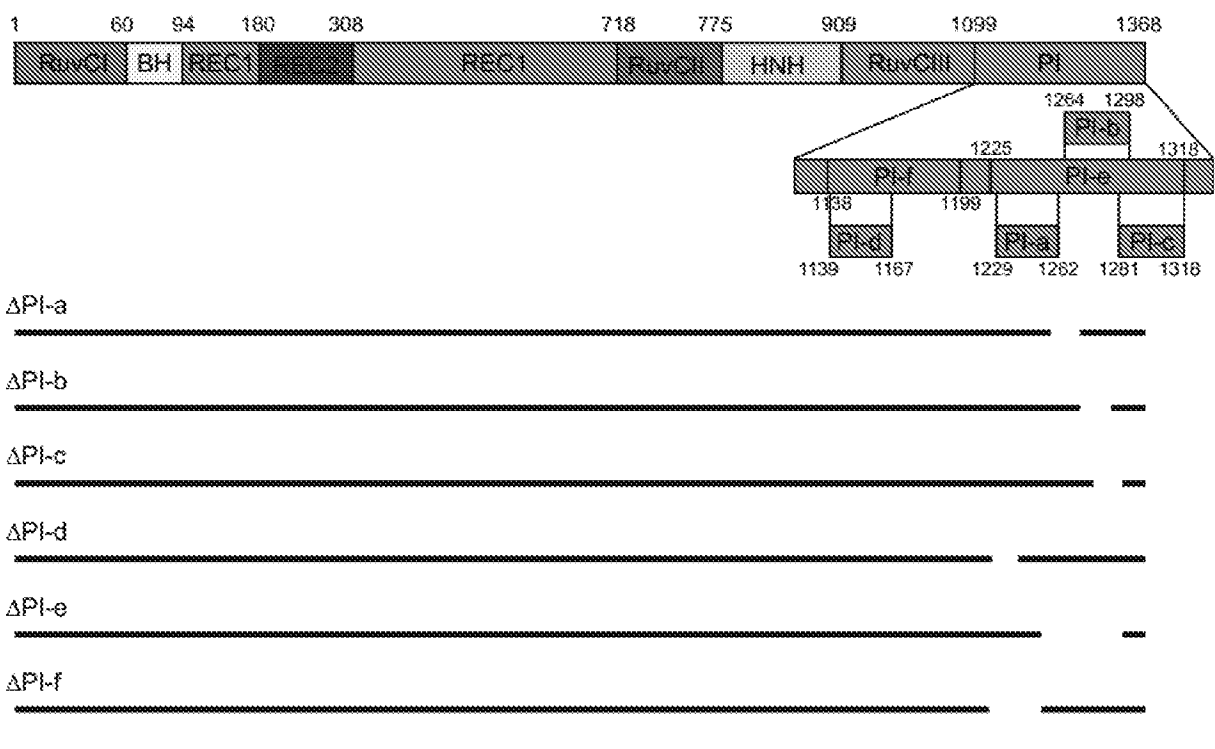
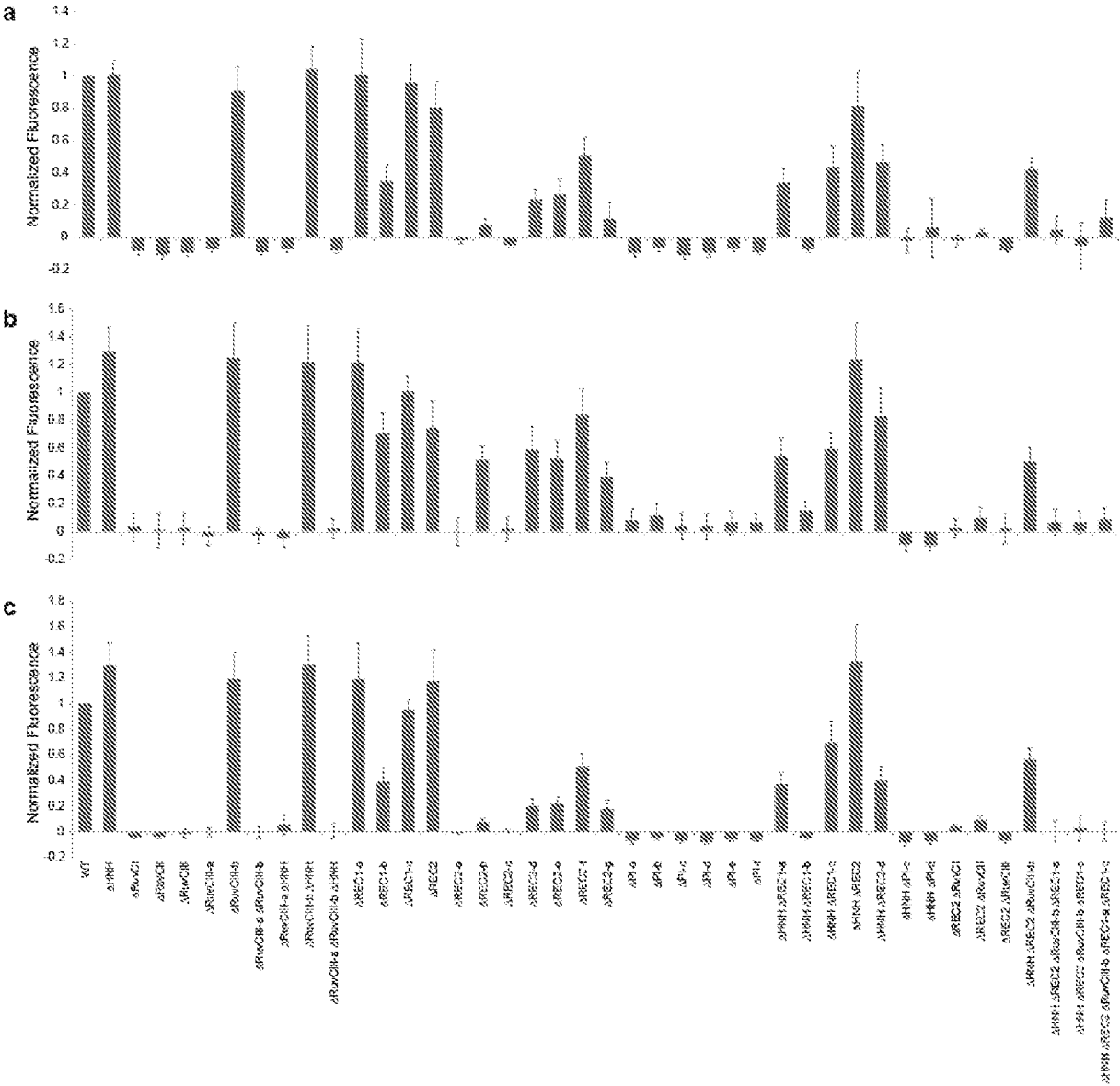


Figure 6



5/10

Figure 7

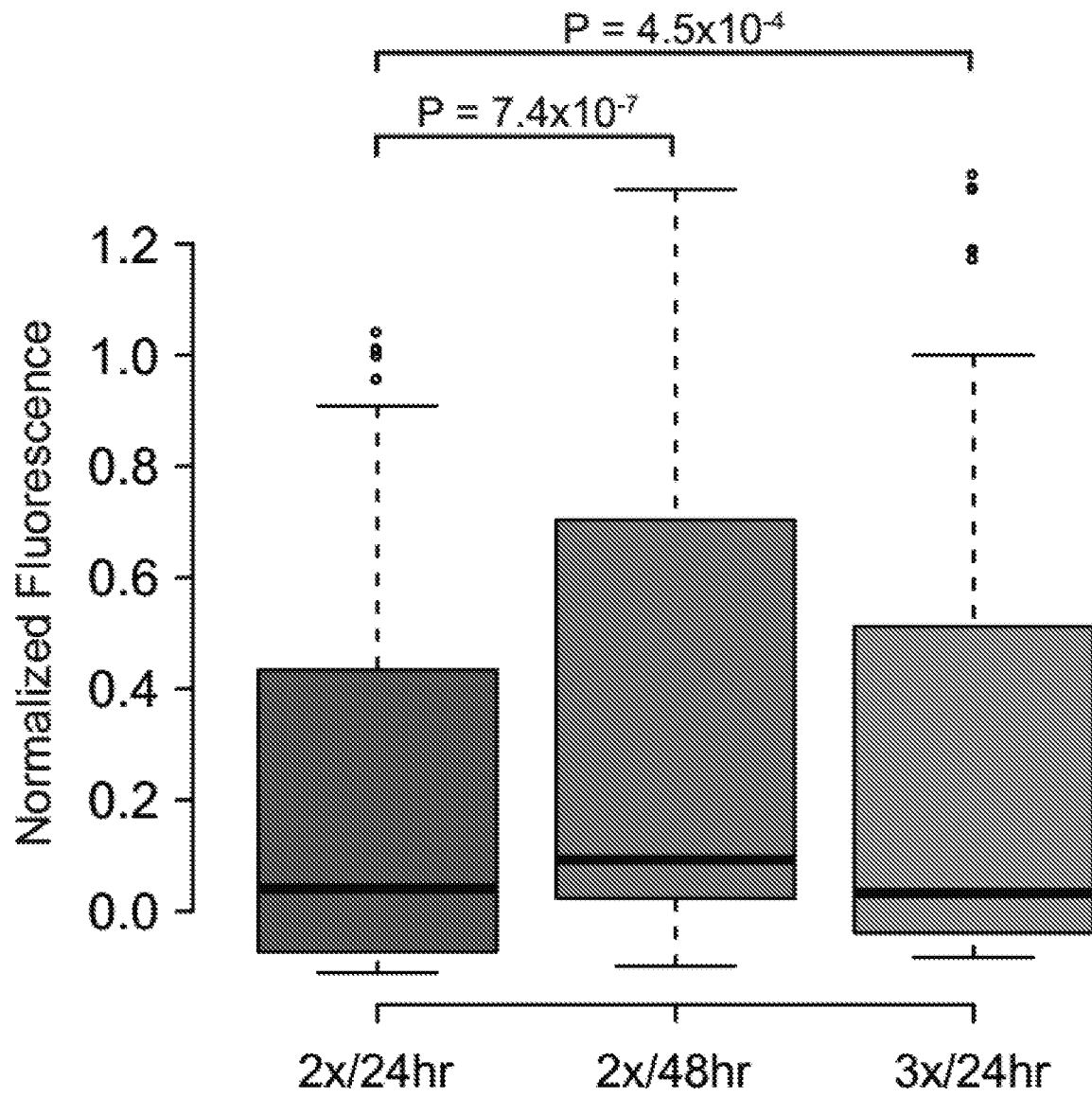


Figure 8

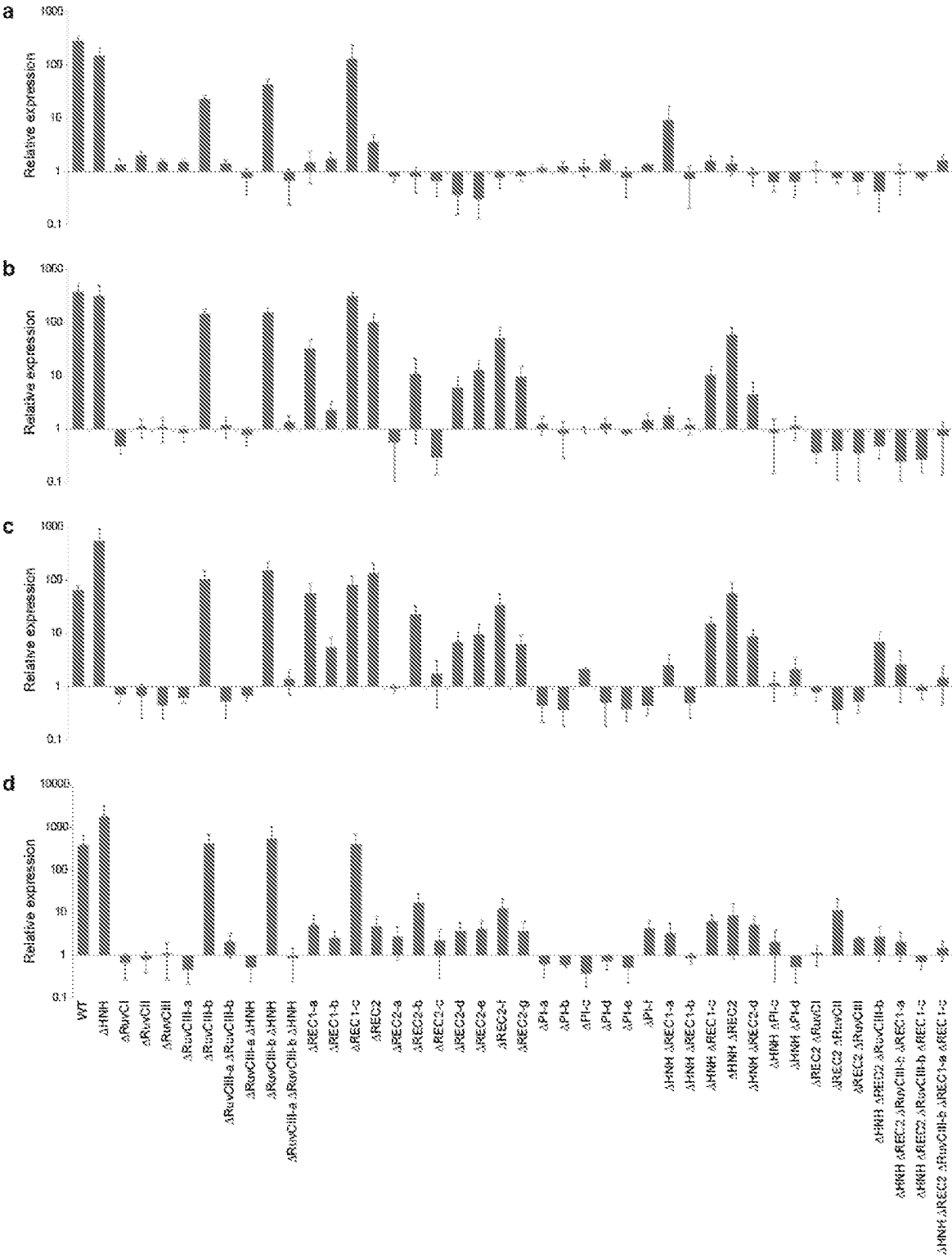
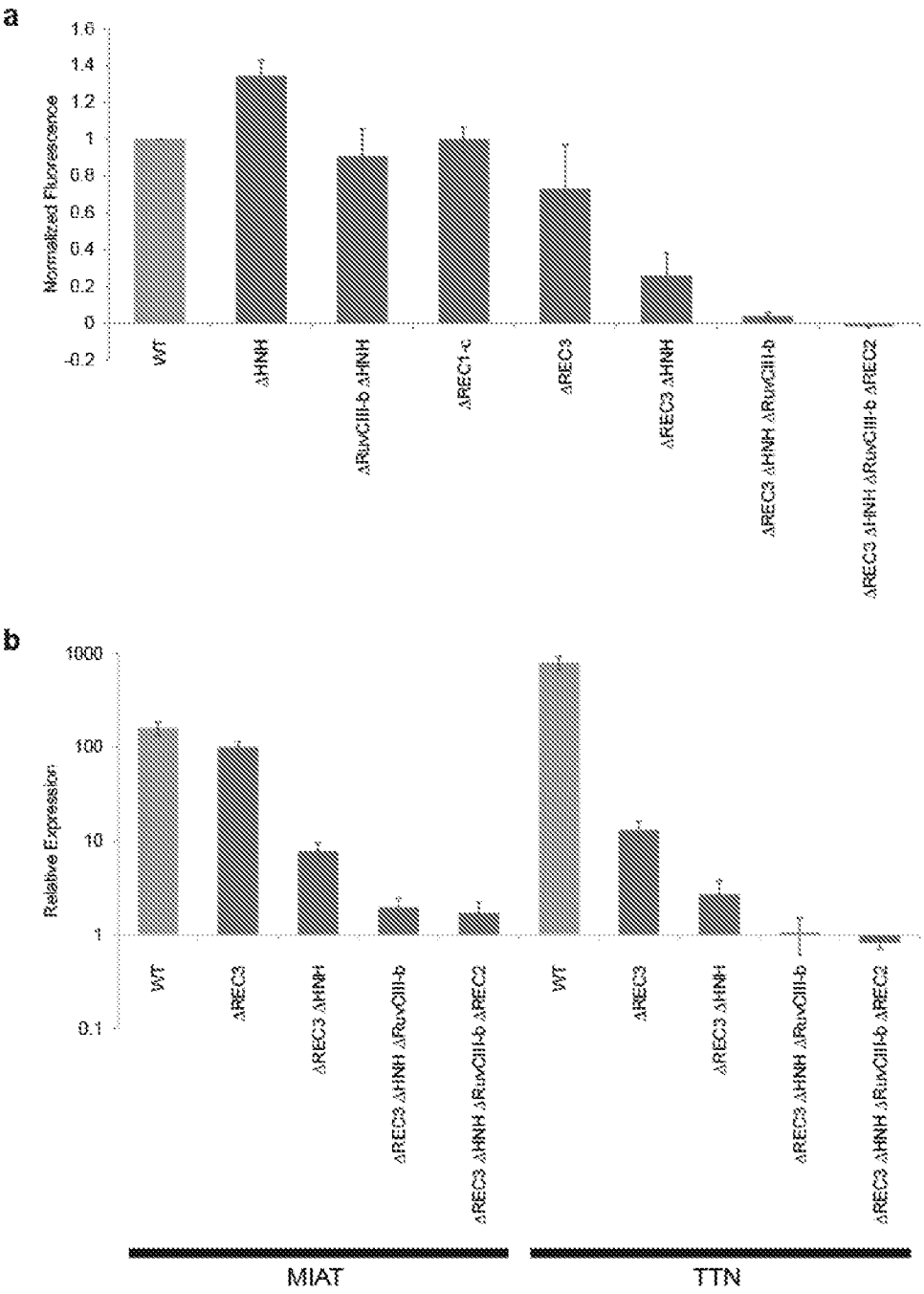


Figure 9



8/10

Figure 10



Figure 11

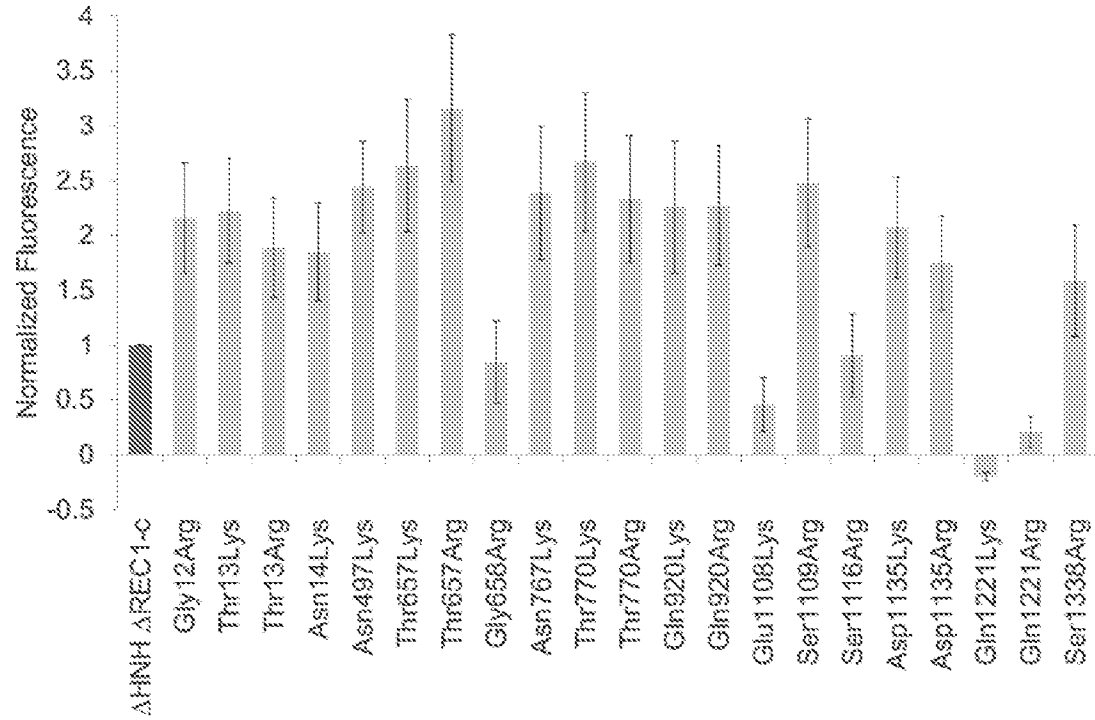


Figure 12

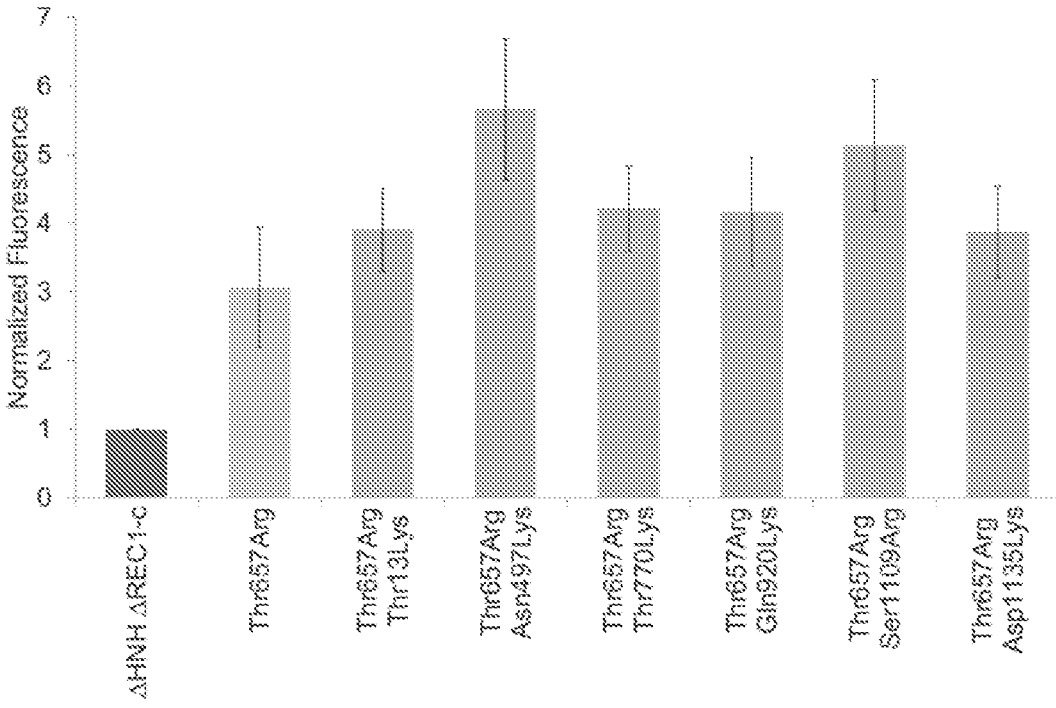
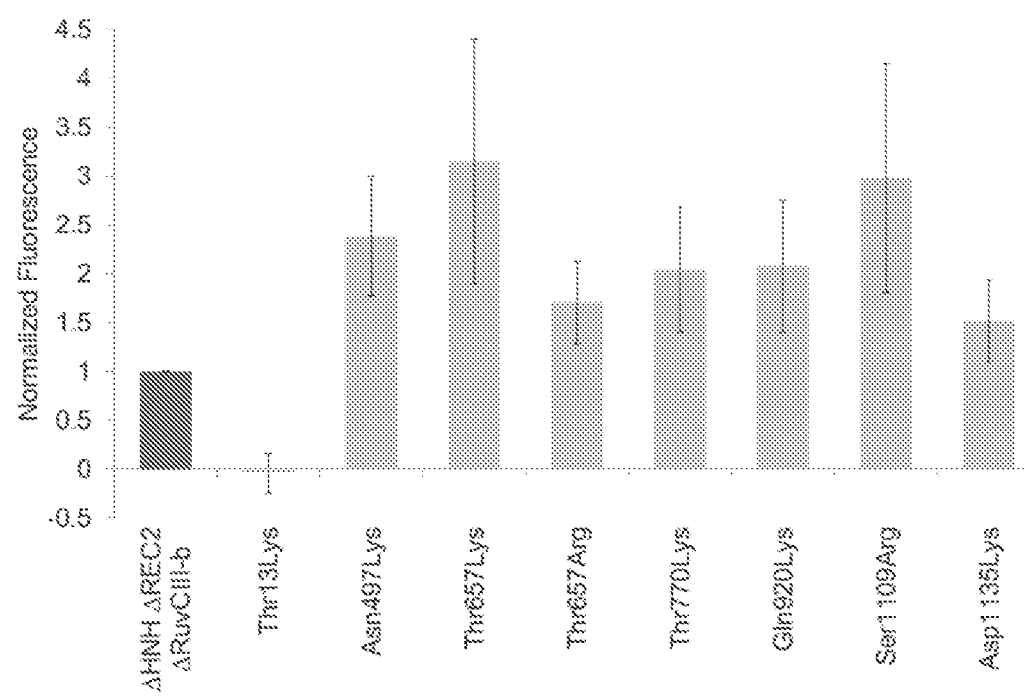


Figure 13



INTERNATIONAL SEARCH REPORT

International application No.

PCT/SG2017/050523

A. CLASSIFICATION OF SUBJECT MATTER

C12N 9/22 (2006.01) C12N 15/09 (2006.01)

According to International Patent Classification (IPC)

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPODOC/WPI/CAPLUS/BIOSIS/EMBASE/MEDLINE: CRISPR, Cas9, Cpf1, truncation, deletion, removal, substitution, minimal, compact and similar terms.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	NISHIMASU, H. ET AL., Crystal Structure of Cas9 in Complex with Guide RNA and Target DNA. <i>Cell</i> , 13 February 2014, Vol. 156, No. 5, pages 935-949 [Retrieved on 2017-12-18] <DOI: 10.1016/J.CELL.2014.02.001 > Whole document.	1, 5-7 & 11-19
X	WO 2015/077318 A1 (PRESIDENT AND FELLOWS OF HARVARD COLLEGE) 28 May 2015 Whole document.	1, 5-7 & 11-19
A	WO 2014/191521 A2 (CELLECTIS) 4 December 2014 Whole document.	-

☒ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

*Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

18/12/2017

(day/month/year)

Date of mailing of the international search report

28/12/2017

(day/month/year)

Name and mailing address of the ISA/SG

IPOS Intellectual Property Office of Singapore
51 Bras Basah Road
#01-01 Manulife Centre
Singapore 189554
Email: pct@ipos.gov.sg

Authorized officer

Teclise Nq (Dr)

IPOS Customer Service Tel. No.: (+65) 6339 8616

INTERNATIONAL SEARCH REPORT

International application No.

PCT/SG2017/050523

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	Truncation of the dCas9 protein. Team Freiburg. 10 November 2013 [Retrieved on 2017-12-18 from https://web.archive.org/web/20130101000000/http://2013.igem.org/Team:Freiburg/Project/truncation] Whole document.	-
A	HIRANO, S., Structural Basis for the Altered PAM Specificities of Engineered CRISPR-Cas9. <i>Mol Cell</i> , 17 March 2016, Vol. 61, No. 6, pages 886-894 [Retrieved on 2017-12-18] <DOI: 10.1016/J.MOLCEL.2016.02.018> Whole document.	-
A	KLEINSTIVER, B.P. ET AL., High-fidelity CRISPR-Cas9 variants with undetectable genome-wide off-targets. <i>Nature</i> , 28 January 2016, Vol. 529, No. 7587, pages 490-495 [Retrieved on 2017-12-18] <DOI: 10.1038/NATURE16526.> Whole document, particularly Figure 5.	-
A	QI, L.S. ET AL., Repurposing CRISPR as an RNA-Guided Platform for Sequence Specific Control of Gene Expression. <i>Cell</i> , 28 February 2013, Vol. 152, No. 5, pages 1173-1183 NIH Author Access Manuscript [Retrieved on 2017-12-18 from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3664290/] <DOI: 10.1016/J.CELL.2013.02.022> Whole document.	-
A	GUILINGER, J.P. ET AL., Fusion of catalytically inactive Cas9 to <i>FokI</i> nuclease improves the specificity of genome modification. <i>Nat Biotechnol</i> , 1 June 2014, Vol. 32, No. 6, pages 577-582 NIH Author Access Manuscript [Retrieved on 2017-12-18 from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4263420/] <DOI: 10.1038/NBT.2909> Whole document.	-
A	RAN, F.A. ET AL., <i>In vivo</i> genome editing using <i>Staphylococcus aureus</i> Cas9. <i>Nature</i> , 9 April 2015, Vol. 520, No. 7546, pages 186-191 NIH Author Access Manuscript [Retrieved on 2017-12-18 from https://www.ncbi.nlm.nih.gov/pubmed/25830891] <DOI: 10.1038/NATURE14299> Whole document.	-
P,X	WO 2016/196655 A1 (THE REGENTS OF THE UNIVERSITY OF CALIFORNIA) 8 December 2016 Whole document, particularly paragraphs [00592]-[0594].	1, 5-7 & 11-19

INTERNATIONAL SEARCH REPORT

International application No.

PCT/SG2017/050523

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:
 - a. ☐ forming part of the international application as filed:
 - ☐ in the form of an Annex C/ST.25 text file.
 - ☐ on paper or in the form of an image file.
 - b. ☐ furnished together with the international application under PCT Rule 13~~ter~~.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.
 - c. ☐ furnished subsequent to the international filing date for the purposes of international search only:
 - ☐ in the form of an Annex C/ST.25 text file (Rule 13~~ter~~.1(a)).
 - ☐ on paper or in the form of an image file (Rule 13~~ter~~.1(b) and Administrative Instructions, Section 713).
2. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that in the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments:

Although a sequence listing has been filed or furnished, it was not used for the purposes of this search.

Since only one version or copy of a sequence listing has been filed or furnished, the statements under item 2 are not required.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/SG2017/050523

Note: This Annex lists known patent family members relating to the patent documents cited in this International Search Report. This Authority is in no way liable for these particulars which are merely given for the purpose of information.

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2015/077318 A1	28/05/2015	AU 2014353128 A1 CA 2930829 A1 EP 3071694 A1 JP 2017501688 A KR 20160079119 A US 9074199 B1 US 2016/298097 A1	09/06/2016 28/05/2015 28/09/2016 19/01/2017 05/07/2016 07/07/2015 13/10/2016
WO 2014/191521 A2	04/12/2014	AU 2014273085 A1 CA 2913869 A1 EP 3004339 A2 JP 2016/520318 A US 2016/102324 A1	10/12/2015 28/05/2015 13/04/2016 14/07/2016 14/04/2016
WO 2016/196655 A1	08/12/2016	NONE	