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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
18 March 2021 (18.03.2021)

WIPO | PCT

(10) International Publication Number
WO 2021/050755 A1

(51) International Patent Classification:
C12N 9/22 (2006.01) *C12Q 1/682* (2018.01)
C12N 15/113 (2010.01) *C12Q 1/70* (2006.01)

(21) International Application Number:
PCT/US2020/050237

(22) International Filing Date:
10 September 2020 (10.09.2020)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/898,340 10 September 2019 (10.09.2019) US
63/058,448 29 July 2020 (29.07.2020) US

(71) Applicants: **CASPR BIOTECH CORPORATION** [US/US]; 479 Jessie Street, San Francisco, California 94103 (US). **CONSEJO NACIONAL DE INVESTIGACIONES CIENTIFICAS Y TÉCNICAS (CONICET)** [AR/AR]; Godoy Cruz 2290, Buenos Aires, C1425FQB (AR).

(72) Inventors: **GIMENEZ, Carla Alejandra**; c/o IndieBio, 479 Jessie Street, San Francisco, California 94103 (US). **REPIZO, Guillermo Daniel**; c/o IndieBio, 479 Jessie Street, San Francisco, California 94103 (US). **PEREYRA BONNET, Federico Alberto**; c/o IndieBio, 479 Jessie Street, San Francisco, California 94103 (US). **CURTI, Lucía Ana**; c/o IndieBio, 479 Jessie Street, San Francisco, California 94103 (US). **GOYTIA, Franco**; c/o IndieBio, 479 Jessie Street, San Francisco, California 94103 (US). **FARÍAS, Maria Eugenia**; Federico Rossi 871, Yerba Buena, Tucumán (AR).

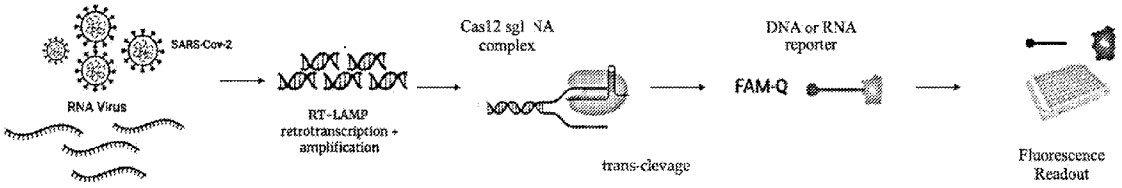
(74) Agent: **ROY, Madhuri** et al.; Cooley LLP, 1299 Pennsylvania Ave., Suite 700, Washington, District of Columbia 20004 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, 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, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, 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, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

(54) Title: NOVEL CLASS 2 TYPE II AND TYPE V CRISPR-CAS RNA-GUIDED ENDONUCLEASES

FIG. 18



(57) Abstract: Provided herein are novel Class 2 Type II and Type V CRISPR-Cas RNA-guided endonucleases, e.g. Cas9 and Casl2 endonucleases, and systems comprising the same. Provided also are methods of making, and methods of use thereof. Exemplary methods of use include modifying target DNAs and detecting targeting DNAs, useful for therapeutic and diagnostic applications. Some of the diagnostic applications may utilise the collateral nuclease activity of an enzyme bound to a target sequence.

[Continued on next page]

WO 2021/050755 A1

Declarations under Rule 4.17:

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

Published:

— *with international search report (Art. 21(3))*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

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

NOVEL CLASS 2 TYPE II AND TYPE V CRISPR-CAS RNA-GUIDED ENDONUCLEASES

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority to United States Provisional Patent Application Serial No. 63/058,448 filed July 29, 2020 and United States Provisional Patent Application Serial No. 62/898,340 filed September 10, 2019, each of which are herein incorporated by reference in their entirety.

DESCRIPTION OF THE TEXT FILE SUBMITTED ELECTRONICALLY

[0002] The sequence listing associated with this application is provided in text format in lieu of a paper copy, and is hereby incorporated by reference into the specification. The name of the text file containing the sequence listing is "CABI_002_02WO_SeqList_ST25.txt". The text file is 456 kb, was created on September 10, 2020, and is being submitted electronically via EFS-Web.

BACKGROUND

[0003] Bacterial adaptive immune systems have in place CRISPRs (clustered regularly interspaced short palindromic repeats) and CRISPR-associated (Cas) proteins for RNA-guided nucleic acid cleavage. The CRISPR-Cas systems act to confer adaptive immunity in bacteria and archaea via RNA-guided nucleic acid interference. To provide immunity against invaders, processed CRISPR array transcripts (crRNAs) assemble with Cas protein-containing surveillance complexes that recognize nucleic acids bearing sequence complementarity to the invader's derived segment of the crRNAs, known as the spacer.

[0004] Class 2 CRISPR-Cas systems are streamlined versions in which a single Cas protein (an effector endonuclease protein) bound to RNA is responsible for binding to and cleavage of a targeted sequence. The programmable nature of these minimal systems has facilitated their use as a versatile technology that continues to revolutionize the field of genome manipulation.

[0005] There however is a need for improved Class 2 Type II and Type V CRISPR-Cas RNA-guided endonuclease variants. Provided herein are such variants, methods of making, methods of testing, and methods of using the same.

SUMMARY

[0006] Provided herein are novel Class 2 Type II and novel Type V CRISPR-Cas RNA-guided systems, methods of making, and methods of use. More specifically, provided are novel Cas9 variants, novel Cas12a variants, and novel Cas12 subtypes.

[0007] In one aspect provided herein is an engineered system comprising: (a) a Cas9.1, Cas9.2, Cas9.3 or Cas9.4 protein, or a nucleic acid encoding the a Cas9.1, Cas9.2, Cas9.3 or Cas9.4 protein; and (b) a Cas9.1, Cas9.2, Cas9.3 or Cas9.4 guide RNA (gRNA), or a nucleic acid encoding the Cas9.1, Cas9.2, Cas9.3 or Cas9.4 gRNA, wherein the gRNA and the Cas9.1, Cas9.2, Cas9.3 or Cas9.4 protein do not naturally occur together, wherein the gRNA is capable of hybridizing to a target sequence in a target DNA, and the gRNA is capable of forming a complex with the Cas9.1, Cas9.2, Cas9.3 or Cas9.4 protein.

[0008] In another aspect, provided herein is an engineered single-molecule gRNA, comprising: (a) a targeter-RNA comprising a spacer sequence that is capable of hybridizing with a target sequence in a target DNA; and (b) an activator-RNA that is capable of hybridizing with the targeter-RNA to form a double-stranded RNA duplex, the activator-RNA comprising a activator-RNA, wherein the targeter-RNA and the activator-RNA are covalently linked to one another, wherein the single-molecule gRNA is capable of forming a complex with a Cas9.1, Cas9.2, Cas9.3 or Cas9.4 protein, and wherein hybridization of the spacer sequence to the target sequence is capable of targeting the Cas9.1, Cas9.2, Cas9.3 or Cas9.4 protein to the target DNA.

[0009] In another aspect, provided herein is an engineered system comprising: a Class 2 Type V CRISPR-Cas RNA-guided endonuclease protein and a single guide RNA, wherein the gRNA and the Class 2 Type V CRISPR-Cas RNA-guided endonuclease protein do not naturally occur together, wherein the gRNA is capable of hybridizing to a target sequence in a target DNA, wherein the gRNA is capable of forming a complex with the Class 2 Type V CRISPR-Cas RNA-guided endonuclease protein, and wherein the Class 2 Type V CRISPR-Cas RNA-guided endonuclease protein possesses collateral activity and is capable of collaterally cleaving a single stranded polynucleotide comprising RNA, without the use of a tracrRNA. In some embodiments, the Class 2 Type V CRISPR-Cas RNA-guided endonuclease protein comprises the amino acid sequence of SEQ ID NO: 4, or at least 70% sequence identity thereto. In some embodiments, the Class 2 Type V CRISPR-Cas RNA-guided endonuclease protein is capable of collaterally cleaving a single stranded RNA. In some embodiments, the Class 2 Type V

CRISPR-Cas RNA-guided endonuclease protein is capable of collaterally cleaving a single stranded DNA/RNA hybrid.

[0010] In another aspect, provided herein is an engineered system comprising: (a) a Cas12a.1, Cas12p, or Cas12q protein, or a nucleic acid encoding the Cas12a.1, Cas12p, or Cas12q protein; and (b) a Cas12a.1, Cas12p, or Cas12q gRNA, or a nucleic acid encoding a Cas12a.1, Cas12p, or Cas12q gRNA, wherein the gRNA and the Cas12a.1, Cas12p, or Cas12q protein do not naturally occur together, wherein the gRNA is capable of hybridizing to a target sequence in a target DNA, and the gRNA is capable of forming a complex with the Cas12a.1, Cas12p, or Cas12q protein.

[0011] In another aspect, provided herein is an engineered single-molecule gRNA, comprising the scaffold sequence of SEQ ID NO: 116 or SEQ ID NO: 117 and a spacer sequence that is capable of hybridizing with a target sequence in a target DNA. In some embodiments, the target DNA is viral DNA, plant DNA, fungal DNA, or bacterial DNA. In some embodiments, the target sequence is a sequence of a target provided in any of Tables 6a-6f. In some embodiments, the target is a coronvavirus. In some embodiments, the target is a SARS-CoV-2 virus. In some embodiments, the target DNA is cDNA, and has been obtained by reverse transcription.

[0012] In another aspect, provided herein is a method of detecting a target DNA in a sample, the method comprising: (a) contacting the sample with: (i) a Cas12a.1, Cas12p, or Cas12q protein; (ii) a Cas12a.1, Cas12p, or Cas12q gRNA comprising a spacer sequence that is capable of hybridizing with a target sequence in a target DNA; and (iii) a labeled detector oligonucleotide that does not hybridize with the spacer sequence of the gRNA; and (b) measuring a detectable signal produced by cleavage of the labeled detector by the Cas12a.1, Cas12p, or Cas12q protein, thereby detecting the target DNA. This method is useful for diagnostics, e.g. detection of a viral or bacterial pathogen in a sample.

[0013] In another aspect, provided herein is a method of modifying a target DNA, the method comprising (a) contacting the target DNA with (i) a Cas9.1, Cas9.2, Cas9.3, Cas9.4, Cas12a.1, Cas12p, or Cas12q protein or a nucleotide encoding the same; and (ii) a Cas9.1, Cas9.2, Cas9.3, Cas9.4, Cas12a.1, Cas12p, or Cas12q gRNA comprising a spacer sequence that is capable of hybridizing with a target sequence in a target DNA. This method is useful for gene therapeutic applications, and generation of cells for therapeutic delivery purposes and for the preparation of cell lines.

[0014] In various embodiments, provided herein are compositions, pharmaceutical compositions, vectors, host cells, and kits comprising any of the proteins or polynucleotides of the engineered systems described herein.

BRIEF DESCRIPTION OF THE DRAWINGS

[0015] **FIGS. 1A-1B** show expression vector maps for Cas9.1 and Cas9.2.

[0016] **FIGS. 2A-2C** show expression vector maps for Cas12a.1, Cas12p, and Cas12q.

[0017] **FIG. 3A** is a schematic representation of the CRISPR Cas cluster around the novel Cas9.1 gene. **FIG. 3B** shows the secondary structure of the direct repeat for the Cas9.1 pre-crRNA. **FIG. 3C** is a schematic representation of the CRISPR Cas cluster around the novel Cas9.2 gene. **FIG. 3D** is a schematic representation of the CRISPR Cas cluster around the novel Cas9.3 gene. **FIG. 3E** shows the secondary structure of the direct repeat for the Cas9.3 pre-crRNA. **FIG. 3F** is a schematic representation of the CRISPR Cas cluster around the novel Cas9.4 gene. **FIG. 3G** shows the secondary structure of the direct repeat for the Cas9.4 pre-crRNA.

[0018] **FIG. 4A** shows the key catalytic amino acids for Cas9 proteins (SEQ ID NOs: 137-168), and alignments of conserved motifs in selected representatives of the Cas9 protein family. **FIG. 4B** shows the alignment of RuvC1, Bridge Helix, RuvCII, and RuvCIII domains for Cas9.1 (SEQ ID NO: 1) and other selected representatives of the Cas9 protein family (SEQ ID NOs: 169-176). **FIG. 4C** shows the alignment of RuvC1, Bridge Helix, RuvCII, and RuvCIII domains for Cas9.2 (SEQ ID NO: 2) and other selected representatives of the Cas9 protein family (SEQ ID NOs: 170-174 and 169). **FIG. 4D** shows the alignment of RuvC1, Bridge Helix, RuvCII, and RuvCIII domains for Cas9.3 (SEQ ID NO: 10) and other selected representatives of the Cas9 protein family (SEQ ID NOs: 169-176). **FIG. 4E** shows the alignment of RuvC1, Bridge Helix, RuvCII, and RuvCIII domains for Cas9.4 (SEQ ID NO: 11) and other selected representatives of the Cas9 protein family (SEQ ID NOs: 169-176).

[0019] **FIG. 5A** is a schematic representation of the CRISPR Cas cluster around the novel Cas12a.1 gene. **FIG. 5B** shows the secondary structure of the direct repeat for the Cas12a.1 pre-crRNA (SEQ ID NO: 177). **FIG. 5C** is a schematic representation of the CRISPR Cas cluster around the novel Cas12p gene. **FIG. 5D** shows the secondary structure of the direct repeat for a first Cas12p pre-crRNA (SEQ ID NO: 178) and a second Cas12p pre-crRNA (SEQ ID NO: 179). **FIG. 5E** is a schematic representation of the CRISPR Cas cluster around the novel

Cas12q gene. **FIG. 5F** shows the secondary structure of the direct repeat for the Cas12q pre-crRNA (SEQ ID NOs: 180 and 181).

[0020] **FIG. 6A** shows the key catalytic amino acids for Cas12 proteins (SEQ ID NOs: 182-217, and alignments of conserved motifs in selected representatives of the Cas12a protein family.

[0021] **FIG. 6B** shows the alignment of Cas12a.1 (SEQ ID NO: 3) vs. SEQ ID NO: 81 of US20160208243 (SEQ ID NO: 218), and has a 46.8% sequence identity; and **FIG. 6C** shows the alignment of Cas12a.1 (SEQ ID NO: 3) vs. SEQ ID NO: 3 of US 10,253,365 (SEQ ID NO: 219), and has a 46.5% sequence identity.

[0022] **FIG. 6D** shows the amino acid sequence of Cas12p (SEQ ID NO: 4) with the RuvC motifs underlined. The FnCas12a sequence referenced in Shmakov et al., 2015 was used as a reference for identification of the Ruv motifs.

[0023] **FIG 6E** shows the alignment of Cas12p (SEQ ID NO: 4) with Cas12g1 (SEQ ID NO: 220). This figure shows an alignment of Cas12p with Cas12g1.

[0024] In the following figures, the structure of Cas12p protein was modeled based on Fn Cas12a structure with Swiss Model server. **FIG 6F** shows a structural analysis of Cas12p using the Swiss Model server. **FIG 6G** shows a spatial prediction of non-conserved amino acid residues in Cas12p. **FIG 6H** shows the approximation of charge distribution over the surface of Cas12p. **FIG 6I** shows predicted structural differences between Cas12p (SEQ ID NO: 4) and FnCas12a (SEQ ID NO: 221) based on protein sequences. **FIG 6J** shows RuvCIII domain structural analysis of Cas12p (SEQ ID NO: 4) and Cas12a proteins (AsCas12a (SEQ ID NO: 223), LbCas12a (SEQ ID NO: 224) and FnCas12a (SEQ ID NO: 221)) based on structural analysis with Swiss Model server.

[0025] **FIG. 6K** shows the amino acid sequence of Cas12q (SEQ ID NO: 5) with the RuvC motifs underlined.

[0026] **FIGS. 7A, 7B, 7C** show predicted RNA secondary structures of non-naturally occurring direct repeats (artificial variants; SEQ ID NOs: 225-239), generated to improve stem-loop stability of guides of the disclosure.

[0027] **FIG. 8** shows bar graphs for the PAM sequence preferences of Cas12a.1 and Cas12p for the ten PAM motifs, measuring the performance of the Cas12a.1 and the Cas12p using fluorescence assays.

[0028] **FIG. 9A** shows specific cleavage activity of the Ca12a.1 (designated as Cas12.1 in the figure) and Cas12p proteins of the disclosures with an exemplary Hanta virus target. **FIG. 9B** shows that both Cas12a.1 and Cas12p exhibit collateral activity and can cut non-target containing ssDNA. **FIG 9C** shows that Cas12p exhibits both ssDNA and RNA reporter collateral clevage using as a SARS-CoV-2 inactivated virus as sample as the target.

[0029] **FIG.10** shows activity of the novel cas12 proteins at 25°C.

[0030] **FIG. 11** shows the activity of the novel Cas12 proteins at various salt concentrations.

[0031] **FIG. 12** shows the performance of the Cas12a.1 and the Cas12p of the disclosure in three different commercial buffers.

[0032] **FIG. 13** shows sensitivity curves without RPA of the Cas12a.1 and the Cas12p of the disclosure, for various target concentrations measured for 30 minutes.

[0033] **FIG. 14** shows that the amount of fluorescence detection by Cas12a.1 and Cas12p for a target DNA reverse transcribed from SARS-CoV-2 RNA was equal at both 37°C and 25°C, indicative of thermostability and function and room temperature.

[0034] **FIG. 15** shows the differential performance of Cas12p vs. LbCas12a at 25°C.

[0035] **FIG. 16** shows the differential performance of Cas12p vs. LbCas12a at 25°C, using SARS-CoV-2 as a target, described in Example 10.

[0036] **FIG. 17** shows the ability of Cas12p to cleave both a ssDNA and RNA reporter.

[0037] **FIG. 18** shows a schematic workflow for the detection of SARS-CoV-2 described herein.

[0038] **FIG. 19** shows a schematic workflow for the detection of SARS-CoV-2 described herein, from a sample.

[0039] **FIG. 20** shows that Cas12p has a minimal background signal after 30-60 minutes of cleavage activity. This provides advantages at low viral concentrations, and indicates stability of the lyophilized format.

[0040] **FIG. 21** shows that a diagnostics assay using Cas12p at room temperature, can be read out on a paper format.

[0041] **FIG. 22** shows that a diagnostics assay using Cas12p at room temperature can be read in well plate with a fluorescent detector.

[0042] **FIG. 23** shows exemplary lyophilized beads of the disclosure.

[0043] **FIG. 24** shows the results of SARS-CoV-2 detection using a Cas12p/guide, using a RNA reporter from patient samples and negative control samples in lyophilized format.

[0044] **FIG. 25** shows specific dsDNA cleavage time courses of the Ca12a.1 and Cas12p proteins of the disclosures, complexed with a sgRNA for an exemplary Hanta virus target. Time points: 0, 30, 60 and 90 minutes.

[0045] **FIG. 26** shows specific ssDNA cleavage time courses of the Ca12a.1 and Cas12p proteins of the disclosures, complexed with a sgRNA for an exemplary Hanta virus target. **(S)**: 3′FAM-ssDNA target substrate. **(P)**: 3′FAM-ssDNA target product. **(NTC)**: ASssDNA non-target control. Time points: 0, 0.5,1 and 5 minutes.

[0046] **FIG. 27** shows specific ssRNA cleavage time courses of the Ca12a.1 and Cas12p proteins of the disclosures, complexed with a sgRNA for an exemplary Hanta virus target. **(S)**: ssRNA target substrate. **(TC)**: ssDNA target control. **(NTC)**: ssRNA non-target control. Time points: 0, 1 and 3 h.

[0047] **FIG. 28 s**hows the mass spectra data of Cas12p reactions using a DNA oligo as the reporter.

[0048] **FIG. 29 s**hows the mass spectra data of Cas12p reactions using a DNA oligo as the reporter.

[0049] **FIG. 30 s**hows the mass spectra data of Cas12p reactions using a RNA oligo as the reporter.

[0050] **FIG. 31 s**hows the mass spectra data of Cas12p reactions using a RNA oligo as the reporter.

[0051] **FIG. 32** shows that DNA–RNA chimeric guides enable efficient collateral activity, when used with Cas 12p.

[0052] **FIG. 33** shows agarose gels demonstrating the collateral activity for Cas12a.1 and Cas12p, for ssDNA, but not dsDNA.

[0053] **FIG. 34** shows differential efficiency of cleavage of homopolymeric reporters, at 25°C and 37°C. The results show that Cas12p clearved poly T, poly A and poly C, whereas Cas12a.1 showed a preference for polyC cleavage.

[0054] **FIG. 35** shows the the collateral cleavage (also referred to herein as *trans*-cleavage) ability of Cas12p but not of Cas12a.1, to cleave a RNA reporter.

[0055] **FIG. 36** shows the kinetics of collateral cleavage activity of Cas12p and Cas12a.1, using DNA and RNA as reporters.

[0056] **FIG. 37** shows the collateral cleavage of Cas12p and Cas12a.1 using a FAMQ DNA–RNA chimeric reporter.

[0057] **FIG. 38** shows the sequences and secondary structures of mature guide scaffolds for Cas12a.1 (SEQ ID NO: 116) and Cas12p (SEQ ID NO: 117).

[0058] **FIG. 39** shows the validation of the use of the mature guide scaffolds to detect SARS-CoV-2 using Cas12a.1 and Cas12p, when used in conjuction with a spacer targeting the N gene of SARS-CoV-2.

DETAILED DESCRIPTION

[0059] Provided herein are novel Class 2 Type II and novel Type V CRISPR-Cas RNA-guided systems, methods of making, and methods of use.

Definitions

[0060] The terms "polynucleotide" and "nucleic acid," used interchangeably herein, refer to a polymeric form of nucleotides of any length, either ribonucleotides or deoxyribonucleotides. Thus, terms "polynucleotide" and "nucleic acid" encompass single-stranded DNA; double-stranded DNA; multi-stranded DNA; single-stranded RNA; double-stranded RNA; multi-stranded RNA; genomic DNA; cDNA; DNA-RNA hybrids; and a polymer comprising purine and pyrimidine bases or other natural, chemically or biochemically modified, non-natural, or derivatized nucleotide bases.

[0061] By "hybridizable" or "complementary" or "substantially complementary" it is meant that a nucleic acid (e.g. RNA, DNA) comprises a sequence of nucleotides that enables it to non-covalently bind, i.e. form Watson-Crick base pairs and/or G/U base pairs, "anneal", or "hybridize," to another nucleic acid in a sequence-specific, antiparallel, manner (i.e., a nucleic acid specifically binds to a complementary nucleic acid) under the appropriate in vitro and/or in vivo conditions of temperature and solution ionic strength.

[0062] It is understood that the sequence of a polynucleotide need not be 100% complementary to that of its target nucleic acid to be specifically hybridizable. Moreover, a polynucleotide may hybridize over one or more segments such that intervening or adjacent segments are not involved in the hybridization event (e.g., a loop structure or hairpin structure, a 'bulge', and the like).

[0063] Percent complementarity between particular stretches of nucleic acid sequences within nucleic acids can be determined using any convenient method. Example methods include BLAST programs (basic local alignment search tools) and PowerBLAST programs (Altschul et al., J. Mol. Biol., 1990, 215, 403-410; Zhang and Madden, Genome Res., 1997, 7, 649-656) or

by using the Gap program (Wisconsin Sequence Analysis Package, Version 8 for Unix, Genetics Computer Group, University Research Park, Madison Wis.), e.g., using default settings, which uses the algorithm of Smith and Waterman (Adv. Appl. Math., 1981, 2, 482-489).

[0064] The terms "peptide," "polypeptide," and "protein" are used interchangeably herein, and refer to a polymeric form of amino acids of any length, which can include coded and non-coded amino acids, chemically or biochemically modified or derivatized amino acids, and polypeptides having modified peptide backbones.

[0065] A "vector" or "expression vector" is a replicon, such as plasmid, phage, virus, or cosmid, to which another DNA segment, i.e. an "insert", may be attached so as to bring about the replication of the attached segment in a cell.

[0066] General methods in molecular and cellular biochemistry can be found in such standard textbooks as Molecular Cloning: A Laboratory Manual, 3rd Ed. (Sambrook et al., HaRBor Laboratory Press 2001); Short Protocols in Molecular Biology, 4th Ed. (Ausubel et al. eds., John Wiley & Sons 1999); Protein Methods (Bollag et al., John Wiley & Sons 1996); Nonviral Vectors for Gene Therapy (Wagner et al. eds., Academic Press 1999); Viral Vectors (Kaplift & Loewy eds., Academic Press 1995); Immunology Methods Manual (I. Lefkovits ed., Academic Press 1997); and Cell and Tissue Culture: Laboratory Procedures in Biotechnology (Doyle & Griffiths, John Wiley & Sons 1998), the disclosures of which are incorporated herein by reference.

[0067] Where a range of values is provided, it is understood that each intervening value, to the tenth of the unit of the lower limit unless the context clearly dictates otherwise, between the upper and lower limit of that range and any other stated or intervening value in that stated range, is encompassed within the invention. The upper and lower limits of these smaller ranges may independently be included in the smaller ranges, and are also encompassed within the invention, subject to any specifically excluded limit in the stated range. Where the stated range includes one or both of the limits, ranges excluding either or both of those included limits are also included in the invention.

[0068] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. All publications mentioned herein are incorporated herein by reference to disclose and describe the methods and/or materials in connection with which the publications are cited.

[0069] It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to "a Cas12a.1 protein" includes a plurality of such Cas12a.1 proteins and reference to "the gRNA" or "the guide RNA" includes reference to one or more gRNAs and equivalents thereof known to those skilled in the art, and so forth. It is further noted that the claims may be drafted to exclude any optional element. As such, this statement is intended to serve as antecedent basis for use of such exclusive terminology as "solely," "only" and the like in connection with the recitation of claim elements, or use of a "negative" limitation.

I. *Class 2 Type II CRISPR-Cas RNA-Guided Systems*

[0070] Provided herein are novel Class 2 Type II CRISPR-Cas RNA-guided proteins and their guide RNAs (a "guide RNA" is interchangeably referred to herein as "gRNA"), constituting the Class 2 Type II CRISPR-Cas RNA-guided systems of the disclosure. As used herein a gRNA may comprise only RNA nucleotides, may comprise RNA and DNA nucleotides, or may comprise only DNA nucleotides, and thus while referred to as a gRNA, may comprise non RNA-nucleotides.

[0071] Accordingly, provided herein are systems comprising (a) a Cas9.1, a Cas9.2, a Cas9.3 or a Cas9.4 protein, or a nucleic acid encoding the Cas9.1, the Cas9.2, the Cas9.3 or the Cas9.4 protein; and (b) a Cas9.1, a Cas9.2, a Cas9.3 or a Cas9.4 gRNA, or a nucleic acid encoding the Cas9.1,the Cas9.2, the Cas9.3 or the Cas9.4 molecule RNA, wherein the gRNA and the Cas9.1the Cas9.2, the Cas9.3 or the Cas9.4 protein do not naturally occur together, wherein the gRNA is capable of hybridizing to a target sequence in a target DNA, and the gRNA is capable of forming a complex with the Cas9.1,the Cas9.2, the Cas9.3 or the Cas9.4 protein. It should be understood that "Cas9.1 – Cas9.4" as used herein refers to the following: Cas9.1, Cas9.2, Cas9.3, Cas9.4.

[0072] These components are described in turn below.

a. *Class 2 Type II CRISPR-Cas RNA-Guided Proteins*

[0073] Provided herein are novel Class 2 Type II and Type V CRISPR-Cas RNA-guided endonucleases, e.g. novel Cas9 proteins (Cas9 variants) and novel Cas12a proteins (Cas12a variants), and novel Cas12 subtypes.

[0074] Table 1 shows the protein sequences for the novel Cas9 proteins of the disclosure. In some embodiments the novel Cas9 proteins of the disclosure have been deduced using bioinformatics methods from metagenomics samples.

[0075] SEQ ID NO: 1 represents a novel Cas9 variant of the disclosure, Cas9.1, (1038 amino acids in length). FIG. 3A is a schematic representation of the CRISPR Cas cluster around the novel Cas9.1 gene. FIG. 4A shows the key catalytic amino acids for Cas9 proteins, and alignments of conserved motifs in selected representatives of the Cas9 protein family. FIG. 4B shows the alignment of RuvC1, Bridge Helix, RuvCII, and RuvCIII domains for Cas9.1 and other selected representatives of the Cas9 protein family.

[0076] SEQ ID NO: 2 represents a novel Cas9 variant of the disclosure, Cas9.2, (1375 amino acids in length). FIG. 3C is a schematic representation of the CRISPR Cas cluster around the novel Cas9.2 gene. FIG. 4C shows the alignment of RuvC1, Bridge Helix, RuvCII, and RuvCIII domains for Cas9.2 and other selected representatives of the Cas9 protein family.

[0077] SEQ ID NO: 10 represents a novel Cas9 variant of the disclosure, Cas9.3, (1031 amino acids in length). FIG. 3D is a schematic representation of the CRISPR Cas cluster around the novel Cas9.3 gene. FIG. 4D shows the alignment of RuvC1, Bridge Helix, RuvCII, and RuvCIII domains for Cas9.3 and other selected representatives of the Cas9 protein family.

[0078] SEQ ID NO: 11 represents a novel Cas9 variant of the disclosure, Cas9.4, (1329 amino acids in length). FIG. 3F is a schematic representation of the CRISPR Cas cluster around the novel Cas9.4 gene. FIG. 4E shows the alignment of RuvC1, Bridge Helix, RuvCII, and RuvCIII domains for Cas9.4 and other selected representatives of the Cas9 protein family.

Table 1

Cas9.1
MQRIFGLDIGTTSIGFAVIDHDRDQGVGRIHRLGARIFPEARDEKGTPLNQHRRQKR
LARRQLRRRRLRRKALNELLSARGMLPRFGTSAWHDAMALDPYALRARGTEEAL
QPVEVGRALYHLAQRRHFKPRDEAAEADEQEVGDQEAETKREKLLQALRRSGRT
LGQELAARGPHERKRHEHALRSTVETEFERLLTAQARHHEILRDPEFVEELRETIFA
QRPVFWRTSTLGTCPFVPGAPLCPKGAWLSRQRRMLEQVNNLAITGGNARPLDHE
ERRAILAVLQTQASMSWGAVRTALKPLFKARGEAGAERRLRFNLEEGGGKTLLG
NPLEAKLARIFGEAWATHPHRDAIRETIHDRLFAATYNAKGAQRIVILPASQRAER
MRGVIAGLQADFGLSHEQAMALAELPLTPGWEPYSSEALRALMPKLEEGVRFGAL
VVAPEWEDWREATFPQRERPTGEVLDLLPSPKCHDESRRQTRLRNPTVLRTQNEL
RKVVNNLIRAHGKPDIIRVEVAREVGLSKREREDRYNGMRRQERQRQAAIKDLQA
KGFAEPSRADVEKWLLWKESKETCPYTGDKICFDALFRRGEFQVEHIWPRSRSFD
DSFRNKTLCRRDVNLAKGNQTPFEFFESRPEEWEAVKRRLDGLQAKRAGGEGMA
RGKVKRFVASTLPDDFAQRQLNDTGWAAREAVAFLKRLWPDEGQAAPVRVQAV
TGRVTAQLRHLGGLDGVLSDGARKTRDDHRHHAVDALVVACTHPGMTERLSRY

WQQKEDERAERPQLDPPWPTIRADAEAAKDLIVVSHRVRKKISGPFHKETVYGAT
DEREVTRGLEYEKFVTRKRVEDLTKSMLADIRDDRVRQIVTAWVAERGGDPKKA
FPPYPTLGSSGPEIRKVRVLIRRQPTLMARAATGFADLGANHHVAIYKTADERFAF
EVVSLLEVARRVDRGEPPVKRQRGDEKLVMSLAQGDLIRFAKTPDAEAAIWRVQ
KIATKGQISLLHHDDASPKEPSLFEPMVGGLMARNPEKLAVDPIGRVRKAGD **(SEQ ID NO: 1)**

Cas9.2
MKKEKVYMGLDLGTNSVGWAVTDNDYKVLKFKRRAMWGVRLFNEANPAVERR
VARSNRRRLARKKQRVAWLKEIFKNSISEIDPEFFDRLEQSALWAEDKNVAGKYS
LFNEKKLTDKTFYRKFPTVFHLKKALMDGKIKKPDIRFVYLALSHYLQNRGHFLLE
NELNSVEDIDIRDIFNSLNERIHVLIDSGDDMVPAFDLTNLDDLKQIATDTNISGKTQ
EKEAFIKTLLNGAKQPALEAIIKLCTGGSANLSKIFGDMFEFESEIKSISFEKANFEDE
IAPKLQDCLGDYYQIIELAQQIYSWYTLYKVCSGRPSVSHAKVEDYEKHKEQLSHL
KVLVRKHFSKNVYREIFRKEDDKIHNYVSYISGKKDRDEFYKYLKKTLEKKSTFK
KTSEFENISRAIEQQNYLPKQRVKDNSVVPQQLYKQEIVKILNNLSSHYPFLSQKTD
GISNREKIIKIFEYRIPYYVGPLCDIHRAGDDGFSWLVRDCSKKITPWNFEQVVDIPQ
SAENFIKNMTRKCTYLKQYNVLPKNSLLYSEYSVLNELNNVRIKTKKLTPKLKEK
MLNTLFRQKKNISITSLIHWLVSEGVYEKGEIEKSDVSGVDSNFTSSLSAAISFDRIIG
EKMKNKKTQKMVEEIINWLALFSDKKILQQKIVEKYQDKVSQEQIGKILRLNLSG
WGRLSSEFLQLKNSQPGEHDGKTLINIMRQTQMNLMEIIHSPQFSFNTVIETEAKKQ
LTGHITHSHVEALYCSPVVKKQIWQALQIALELKKTLKKDPNKIFVETTRHEGEKK
RTTSRHKQLLELYQAAKSHLPDLTKSIKELNDALKDTEPEKMKRKKLFHYYKQLG
RCMYTGRPISLEDLFTNKYDIDHIYPQSLTKDDSFTNTVLVERLSNAEKSDAFPLDS
KTRKDRQGLWRCLRRNGLITKEKYYRLTRETPLSEEEKAAFIRRQLVETSQTTKEV
IRFLATLFPKSKVVYVKSGNVSDFRRDFSPSLPENKTNGKDPKGITDYSMIKVREIN
DLHHAKDAYLNIVVGNVYDTKFRYRGKDLTAIVREKARQYHLSRLFLYSTDGAW
IGAADENRGKQRPSIETVIAEMRRNSCQVTWEAVFKKGQLWDMNAKSKRPGLLPI
KKELSDTAKYGGYQGKTASYFVVVEYENKKGEREKKLESVPIYVKALSKQKPDA
VNSFLRDTLGLEKPSVMVDNIKIGSIVEINGARMVLTGNNEVLVFGRIASQLILDIT
MAAYLKRMFKLLADTAKIKENNVYFKNCGYLDKETNLAVYDTFIAKLKLPRYAQ
IITHSLYEKMESNRDVFINLSLADQCNLLAGVLPALQCNSQNADLSLLGEGKAVGN
IAFSKNAILKKNQVRLVDCSITGLFENSRNMA **(SEQ ID NO: 2)**

Cas9.3
MIFGLDVGTTSIGFALISLDEDKETGCIVHSGCRVFPEGVTEDKKESRNKARREARL
RRRQLRRKKENRKRLAQFLHETSLLPVFGSTEWKNLMDNTHSNPYELRSAALKKQ
LQPFELGKVIYHLAKHRGFKATKLDELMAESDEKKELGVVKDGIKELDHKLGDQT
LGVYLASIPPSEKKRGRYLGRYMIQEELEQILEYQKHYNPELITSTFKKHLNSLIFSQ
RPTFWRLNTLGTCSLEQNESVCPKHSWIGQQFIMMQKVNDLRIVEPHPRHLTMEE
RTQLIQGLCKQKIMSFGGIRKLLHLPKGTVFNFETYQDKEDKRGLPGNAIEAALSTI
FGSEWKHLPHKDAIRSSLSNRIWSISYNRVGNKRIEIRADESYQNQRQTVKQEMMK
DWNIAEDQAEQLVQLPIPPQWLRFSEKAIQKLLPDLESGVPLQTAIKEHYPETLKSS
EVEHELLPSSPHLVPELRNPTVNRALNELRKVVNNIIRSYGKPDIIRIELARDLKLGK
KKKLEITKKNRQREQERKEAKNQLEKEGVKPTGMNIEKFLLWQESDGLDLYTGQ

KISFAALFKQTEYDIEHIIPRSRSFNNTFFNKTLAHNEINRQKGNMIPKEFFGDGETW
HAFVTRVNQSKLPLEKKEKLLIPHYDAIASEEMTERQLRDTAYIATEAKTYLQTLGI
PVQPTNGRATASLRRVWGINSIWATEFGLEEESKKAAGEKIRDDHRHHAVDAAVV
ALTSPGRIKRLSTFYQYRKEMKPDDFPLPWETFRADLITSLHKIIISHRVQRKISGPL
HEETAYGFTKKKSETDPTAYYFVTRKTLDKDFKPNKVKDIVDPAVRHLIGEHLQK
FDNNPAVAFAPENRPHMPLRKGGWGPPIKKVRIQIARNPQFMVSRQKNPISYYDSG
DNHHMAIYGTHLDDGTVDPETVSFEVVSRFEVNQRASKNEPLVKPQNENGVPLLF
TLVKNNVLIWNEPGEEEQMHLVRWTTANKGRIFHKPLWMSGTPPIEISISVKNLISY
GGRKVSVDPIGNIFPCND **(SEQ ID NO: 10)**

Cas9.4
MKKILGLDLGTDSIGWTIVQQNEEKKFKLIDKGVRIFQKGVGEEKNNEFSLAKERT
THRNTRKKYRRTKQRKVRLLRELIKHGMCPLSFDELELWSKYRKGKPYIYPLSNK
GFTQWLKLNPYDLRERAIKPDEKLTPLELGRIFYHITQRRGFKSNRKDNSEDSEGV
VKTSISQLREEMEGKTLGQFFNNELKKGNKVRKKYTAREDYHHEFNEICNIQKIDN
KTKAALEREIFFQRALKSQRHLVGKCTLEPKKPRCPLSAIPYEEFRALQFINSIRIKD
AEENLMPLTQKEREVIQSLFFRKSKPSFPFNDIKKILEKHNGQRLTFNYPEKLQIIGS
PTIALLKSVFGEEWASLSVAYTKKDGTTGTINSEDVWHALFEFEHNDKLEDFLKQR
LKLSDDNIQKLIKGNLKQGYASLSRKAINNILPFLKDGHIYTHAVFLAKIPEIIGRKQ
WLHSKDQIVNWFLKSAEELPLKNRLCKIVNNLITEFNETYANADPKYILDDSDKKS
INRSLQHDFGPKTWNKFSSEKKDELQKETERLFLSQINKGNASAPYIKPYRQDEEL
KQYLIDNFNIKQEEAERIYHPSAIDIFDEAPYNDDGIKLLQSPRTPSARNPMAMRAL
HELRYLLNQLLSQRGIDEHTVIHLEMSRELNNQNKRLAIQRYQQARNEEHQEYAK
EIKKIFKEQTQKEIEPTEADILKYRLWKEQEHNCLYTGRKIGIADFIGDNSNVDIEHT
WPRSKSFDNSTANKTLCDSHYNRNIKKNKIPYDLPNFKESAIIEGKQYDPIKARLKD
WEEKCNHLKELAAKYRYNAKRASTKEQKDKALQNAHFYQMHHEYWKDKIFRFT
GKEIRNSFKNSQLVDTGIINKYARAYLQTVFNKVFTIKGTLTADFRKAWGIQNPDT
SKSRQRHTHHAIDAAVVACLTRDRYDFLTQWYRAEEKGNERKKHIIQERMKPWT
TFVQDIKAFENSILVSHHTRKTSAKQTRKRLRENGKIVKDPNGNPIYSKGDTFRNRL
HKDTFYGAILRPQIDKEGKTVTDENGNPKLTTQYVVKKPVTDLKETDIKNIVDSKI
KSLFESKKLNEIQKEGISIPPSKPEGKETPIKSVRLKQPFNPIPLREHTHLSQKPHKQY
YHVQNEGNFLMAIYEETSASKKPEKTFELISNLQAADYYKASNKENREQYPIVPER
KFITKRNKEIELPLKQIIYIGQMVMLYENSPEELKSKNEEELFKCLYKIVGITSMTIQ
AKYEYGVFILKHHAISTPYSELKPKDGDFSWEGNIEAMRKQLHSRIKVVIENLDFKI
TPTGKIEWLF **(SEQ ID NO: 11)**

[0079] As used herein, Cas9.1 includes SEQ ID NO: 1 and proteins with at least 70% - 99.5% sequence identity thereto. Accordingly, provided herein are proteins comprising the amino acid sequence of SEQ ID NO: 1 and proteins with at least 70%, at least 71%, at least 72%, at least 73%, at least 74%, at least 75%, at least 76%, at least 77%, at least 78%, at least 79%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% sequence identity thereto. Also provided herein are nucleic acids encoding the proteins

comprising the amino acid sequence of SEQ ID NO: 1 and proteins with at least 70% - 99.5% sequence identity thereto.

[0080] As used herein, Cas9.2 includes SEQ ID NO: 2 and proteins with at least 70% - 99.5% sequence identity thereto. Accordingly, provided herein are proteins comprising the amino acid sequence of SEQ ID NO: 2 and proteins with at least 70%, at least 71%, at least 72%, at least 73%, at least 74%, at least 75%, at least 76%, at least 77%, at least 78%, at least 79%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% sequence identity thereto. Also provided herein are nucleic acids encoding the proteins comprising the amino acid sequence of SEQ ID NO: 2 and proteins with at least 70% - 99.5% sequence identity thereto.

[0081] As used herein, Cas9.3 includes SEQ ID NO: 10 and proteins with at least 70% - 99.5% sequence identity thereto. Accordingly, provided herein are proteins comprising the amino acid sequence of SEQ ID NO: 10 and proteins with at least 70%, at least 71%, at least 72%, at least 73%, at least 74%, at least 75%, at least 76%, at least 77%, at least 78%, at least 79%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% sequence identity thereto. Also provided herein are nucleic acids encoding the proteins comprising the amino acid sequence of SEQ ID NO: 10 and proteins with at least 70% - 99.5% sequence identity thereto.

[0082] As used herein, Cas9.4 includes SEQ ID NO: 11 and proteins with at least 70% - 99.5% sequence identity thereto. Accordingly, provided herein are proteins comprising the amino acid sequence of SEQ ID NO: 11 and proteins with at least 70%, at least 71%, at least 72%, at least 73%, at least 74%, at least 75%, at least 76%, at least 77%, at least 78%, at least 79%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% sequence identity thereto. Also provided herein are nucleic acids encoding the proteins comprising the amino acid sequence of SEQ ID NO: 11 and proteins with at least 70% - 99.5% sequence identity thereto.

[0083] In some embodiments, the Cas9 protein of the disclosure is a catalytically active Cas9 protein, e.g. a catalytically active Cas9.1, Cas9.2, Cas9.3 or Cas9.4 protein.

[0084] In some embodiments, the Cas9 protein of the disclosure cleaves at a site distal to the target sequence, e.g. the Cas9.1, Cas9.2, Cas9.3 or Cas9.4.4 protein cleaves at a site distal to the target sequence.

[0085] In some embodiments, the Cas9 protein of the disclosure is a catalytically dead Cas9 protein, e.g. the Cas9.1, Cas9.2, Cas9.3 or Cas9.4 protein is catalytically dead (dCas9.1, dCas9.2, dCas9.3 or dCas9.4 protein)..

[0086] In some embodiments, the Cas9 protein of the disclosure is a nickase Cas9 protein, e.g. a Cas9.1 nickase, Cas9.2 nickase, Cas9.3 nickase or Cas9.4 nickase protein.

[0087] The Cas9 proteins of the disclosure can be modified to include an aptamer.

[0088] The Cas9 proteins of the disclosure can be further fused to domains, e.g. catalytic domains to produce dual action Cas proteins. In some embodiments, a Cas9 protein is further fused to a base editor.

b. gRNAs for Class 2 Type II CRISPR-Cas RNA-Guided Proteins

[0089] The present disclosure provides DNA-targeting RNAs that direct the activities of the novel Cas9 proteins of the disclosure to a specific target sequence within a target DNA. These DNA-targeting RNAs are referred to herein as “gRNAs” or “gRNAs” Generally, as provided herein, a Cas9 variant gRNA comprises a first segment (also referred to herein as a “targeter-RNA”, a "DNA-targeting segment" or a "DNA-targeting sequence") and a second segment (also referred to herein as a “activator-RNA”, a "activator-RNA" or a "protein-binding sequence"). Also provided herein are nucleotide sequences encoding the Cas9 gRNAs of the disclosure.

i. Targeter-RNA

[0090] The targeter-RNA of a Cas9 variant gRNA of the disclosure comprises a nucleotide sequence that is complementary to a sequence in a target DNA (targeting sequence of the gRNA; DNA-targeting sequence; spacer sequence). The targeter-RNA can interchangeably be referred to as a crRNA. The targeter-RNA of a gRNA interacts with a target DNA in a sequence-specific manner via hybridization (i.e., base pairing). As such, the nucleotide sequence of the targeter-RNA may vary and determines the location within the target DNA that the gRNA and the target DNA will interact. The targeter-RNA of a subject gRNA can be modified (e.g., by genetic engineering) to hybridize to any desired sequence within a target DNA.

[0091] The targeter-RNA can have a length of from about 12 nucleotides to about 100 nucleotides. For example, the targeter-RNA can have a length of from about 12 nucleotides (nt) to about 80 nt, from about 12 nt to about 50nt, from about 12 nt to about 40 nt, from about 12 nt to about 30 nt, from about 12 nt to about 25 nt, from about 12 nt to about 20 nt, or from about 12 nt to about 19 nt. For example, the targeter-RNA can have a length of from about 19 nt to about 20 nt, from about 19 nt to about 25 nt, from about 19 nt to about 30 nt, from about 19 nt to about 35 nt, from about 19 nt to about 40 nt, from about 19 nt to about 45 nt, from about 19 nt to about 50 nt, from about 19 nt to about 60 nt, from about 19 nt to about 70 nt, from about 19 nt to about 80 nt, from about 19 nt to about 90 nt, from about 19 nt to about 100 nt, from about 20 nt to about 25 nt, from about 20 nt to about 30 nt, from about 20 nt to about 35 nt, from about 20 nt to about 40 nt, from about 20 nt to about 45 nt, from about 20 nt to about 50 nt, from about 20 nt to about 60 nt, from about 20 nt to about 70 nt, from about 20 nt to about 80 nt, from about 20 nt to about 90 nt, or from about 20 nt to about 100 nt.

[0092] Generally, a naturally unprocessed pre-crRNA for Cas9 comprises a direct repeat and an adjacent spacer (the portion of the crRNA that allows for targeting to a DNA molecule). In some embodiments, inclusion of direct repeats, and direct repeat mutations from unprocessed pre-crRNA into the mature gRNA may improve gRNA stability.

[0093] Table 2 shows the naturally occurring direct repeat sequences for the naturally occurring crRNAs of the Cas9 variants of the disclosure.

Table 2: Direct repeat sequences

Description	Name	Sequence
Direct Repeat	Cas9.1	ACTGTAGCAAGACGAAGGGCCGGCGCAATCCGCAGC (SEQ ID NO: 9)

[0094] In some embodiments, the gRNAs of the disclosure include non-naturally occurring, engineered direct repeat sequences which can be incorporated into the engineered gRNAs of the disclosure.

ii. Spacer Sequences

[0095] gRNAs of the disclosure comprise spacer sequences, complementary to the target DNA. More specifically, the nucleotide sequence of the targeter-RNA that is complementary to a target nucleotide sequence (the DNA-targeting sequence or spacer sequence) of the target DNA can have a length at least about 12 nt. For example, the DNA-targeting sequence of the

targeter-RNA that is complementary to a target sequence of the target DNA can have a length at least about 12 nt, at least about 15 nt, at least about 18 nt, at least about 19 nt, at least about 20 nt, at least about 25 nt, at least about 30 nt, at least about 35 nt or at least about 40 nt. For example, the DNA-targeting sequence of the targeter-RNA that is complementary to a target sequence of the target DNA can have a length of from about 12 nucleotides (nt) to about 80 nt, from about 12 nt to about 50nt, from about 12 nt to about 45 nt, from about 12 nt to about 40 nt, from about 12 nt to about 35 nt, from about 12 nt to about 30 nt, from about 12 nt to about 25 nt, from about 12 nt to about 20 nt, from about 12 nt to about 19 nt, from about 19 nt to about 20 nt, from about 19 nt to about 25 nt, from about 19 nt to about 30 nt, from about 19 nt to about 35 nt, from about 19 nt to about 40 nt, from about 19 nt to about 45 nt, from about 19 nt to about 50 nt, from about 19 nt to about 60 nt, from about 20 nt to about 25 nt, from about 20 nt to about 30 nt, from about 20 nt to about 35 nt, from about 20 nt to about 40 nt, from about 20 nt to about 45 nt, from about 20 nt to about 50 nt, or from about 20 nt to about 60 nt. The nucleotide sequence (the DNA-targeting sequence) of the targeter-RNA that is complementary to a nucleotide sequence (target sequence) of the target DNA can have a length at least about 12 nt. In some embodiments, the DNA-targeting sequence of the targeter-RNA that is complementary to a target sequence of the target DNA is 20 nucleotides in length. In some embodiments, the DNA-targeting sequence of the targeter-RNA that is complementary to a target sequence of the target DNA is 19 nucleotides in length.

[0096] The percent complementarity between the spacer sequence of the targeter-RNA and the target sequence of the target DNA can be at least 60% (e.g., at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, at least 99%, or 100%). In some embodiments, the percent complementarity between the DNA-targeting sequence of the targeter-RNA and the target sequence of the target DNA is 100% over the 1-25 contiguous 5 '-most nucleotides of the target sequence of the complementary strand of the target DNA. In some embodiments, the percent complementarity between the DNA-targeting sequence of the targeter-RNA and the target sequence of the target DNA is at least 60% over about 1 – 25 contiguous nucleotides. In some embodiments, the percent complementarity between the DNA-targeting sequence of the targeter-RNA and the target sequence of the target DNA is 100% over the 1 – 25 contiguous 5 '-most nucleotides of the target sequence of the complementary strand of the target DNA and as low as 0% over the remainder. In such a case, the DNA-targeting sequence can be considered to be 1 – 25 nucleotides in length.

[0097] In some embodiments the spacer sequence of a Cas9 gRNA of the disclosure is directed to a target sequence in a mammalian organism. In some embodiments the spacer sequence is directed to a target sequence in a non-mammalian organism.

[0098] In some embodiments, the spacer sequence of a Cas9 gRNA of the disclosure is directed to a target sequence which is a sequence of a human. In some embodiments, the target sequence is a sequence of a non-human primate.

[0099] In some embodiments the spacer sequence of a Cas9 gRNA of the disclosure is directed to a target sequence selected of a therapeutic target.

[00100] In some embodiments the spacer sequence of a Cas9 gRNA of the disclosure is directed to a target sequence selected of a diagnostic target – for example in such embodiments a labeled dCas9 of the disclosure and a gRNA directed to a diagnostic target DNA is contacted with the target DNA, or a cell comprising the target DNA, or a sample comprising the target DNA.

iii. Activator-RNA

[00101] The activator-RNA of a Cas9 variant gRNA of the disclosure binds with its cognate Cas9 variant of the disclosure. The activator-RNA can interchangeably be referred to as a tracrRNA. The gRNA guides the bound Cas9 protein to a specific nucleotide sequence within target DNA via the above described targeter-RNA. The activator-RNA of a Cas9 variant gRNA comprises two stretches of nucleotides that are complementary to one another.

iv. Dual-molecule Cas9 gRNAs

[00102] In some embodiments, provided herein are dual molecule (two-molecule) Cas9 gRNAs for the novel Cas9 proteins of the disclosure. Such gRNAs comprise two separate RNA molecules (activator RNA-tracRNA; and the targeting RNA- crRNA). Each of the two RNA molecules of a subject double-molecule gRNA comprises a stretch of nucleotides that are complementary to one another such that the complementary nucleotides of the two RNA molecules hybridize to form the double stranded RNA duplex of the gRNA.

[00103] A dual-molecule gRNA can be designed to allow for controlled (i.e., conditional) binding of a targeter-RNA with an activator-RNA. Because a dual-molecule gRNA is not functional unless both the activator-RNA and the targeter-RNA are bound in a functional complex with Cas9 variant of the disclosure, a dual-molecule gRNA can be inducible (e.g., drug inducible) by rendering the binding between the activator-RNA and the targeter-RNA to be inducible. As one non-limiting example, RNA aptamers can be used to regulate (i.e., control) the

binding of the activator-RNA with the targeter-RNA. Accordingly, the activator-RNA and/or the targeter-RNA can comprise an RNA aptamer sequence.

[00104] The dual-molecule guide can be modified to include an aptamer

v. Single-molecule Cas9 Variant gRNAs

[00105] In some embodiments, provided herein are Cas9 gRNAs that comprises a single-molecule gRNA (interchangeably referred to herein as a sgRNA), for the novel Cas9 proteins of the disclosure.

[00106] Accordingly provided herein is an engineered single-molecule gRNA, comprising:

a. a targeter-RNA that is capable of hybridizing with a target sequence in a target DNA; and

b. an activator-RNA that is capable of hybridizing with the targeter-RNA to form a double-stranded RNA duplex, the activator-RNA comprising a activator-RNA,

wherein the targeter-RNA and the activator-RNA are covalently linked to one another, wherein the single-molecule gRNA is capable of forming a complex with a novel Cas9 protein of the disclosure, and wherein hybridization of the targeter-RNA to the target sequence is capable of targeting the Cas9 protein of the disclosure to the target DNA.

[00107] A subject single-molecule gRNA comprises two segments of nucleotides (a targeter-RNA and an activator-RNA) that are complementary to one another, can be covalently linked by intervening nucleotides ("linkers" or "linker nucleotides"), and hybridize to form the double stranded RNA duplex (dsRNA duplex) of the activator-RNA, whereby resulting in a stem-loop structure. In some embodiments, the targeter-RNA and the activator-RNA are covalently linked via the 3' end of the targeter-RNA and the 5' end of the activator-RNA. In other embodiments, the activator-RNA is covalently linked via the 5' end of the targeter-RNA and the 3' end of the activator-RNA.

[00108] In some embodiments, the targeter-RNA and the activator-RNA are arranged in a 5' to 3' orientation.

[00109] In some embodiments, the activator-RNA and the targeter-RNA are arranged in a 5' to 3' orientation.

[00110] In some embodiments, the single molecule gRNA comprises one or more sequence modifications compared to a sequence of a corresponding wild type tracrRNA and/or crRNA.

[00111] In some embodiments, the targeter-RNA and the activator-RNA are covalently linked to one another via a linker.

[00112] When present, the linker of a single-molecule gRNA can have a length of from about 3 nucleotides to about 30 nucleotides. In exemplary embodiments, the linker of a single- molecule gRNA is 4, 5, 6, or 7 nt.

[00113] An exemplary single-molecule gRNA comprises two complementary stretches of nucleotides that hybridize to form a dsRNA duplex. In some embodiments, one of the two complementary stretches of nucleotides of the single -molecule gRNA (or the DNA encoding the stretch) is at least about 60% identical to one of the activator-RNA. For example, one of the two complementary stretches of nucleotides of the single-molecule gRNA (or the DNA encoding the stretch) is at least about 65% identical, at least about 70% identical, at least about 75% identical, at least about 80% identical, at least about 85% identical, at least about 90% identical, at least about 95% identical, at least about 98% identical, at least about 99% identical or 100 % identical to an activator-RNA.

[00114] The activator-RNA and targeter-RNA segments can be engineered, while ensuring that the structure of the protein-binding domain of the gRNA is conserved. Thus, RNA folding structure of a naturally occurring protein-binding domain of a DNA-targeting RNA can be taken into account in order to design artificial protein- binding domains (either dual-molecule or single-molecule versions).

[00115] The activator-RNA in a single-molecule gRNA can have a length of from about 10 nucleotides to about 100 nucleotides. For example, the activator-RNA can have a length of from about 15 nucleotides (nt) to about 80 nt, from about 15 nt to about 50 nt, from about 15 nt to about 40 nt, from about 15 nt to about 30 nt or from about 15 nt to about 25 nt.

[00116] Also with regard to both the single-molecule and double-molecule gRNAs of the disclosure, the dsRNA duplex of the activator-RNA can have a length from about 6 nucleotides (nt) to about 50bp. For example, the dsRNA duplex of the activator-RNA can have a length from about 6 nt to about 40 nt, from about 6 nt to about 30bp, from about 6 nt to about 25 nt, from about 6 nt to about 20 nt, from about 6 nt to about 15 nt, from about 8 nt to about 40 nt, from about 8 nt to about 30bp, from about 8 nt to about 25 nt, from about 8 nt to about 20 nt or from about 8 nt to about 15 nt. For example, the dsRNA duplex of the activator-RNA can have a length from about from about 8 nt to about 10 nt, from about 10 nt to about 15 nt, from about 15 nt to about 18 nt, from about 18 nt to about 20 nt, from about 20 nt to about 25 nt, from about 25 nt to about 30 nt, from about 30 nt to about 35 nt, from about 35 nt to about 40 nt, or from about 40 nt to about 50 nt. In some embodiments, the dsRNA duplex of the activator-RNA has a length

of 8 – 15 base pairs. The percent complementarity between the nucleotide sequences that hybridize to form the dsRNA duplex of the activator-RNA can be at least about 60%. For example, the percent complementarity between the nucleotide sequences that hybridize to form the dsRNA duplex of the activator-RNA can be at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 98%, or at least about 99% . In some embodiments, the percent complementarity between the nucleotide sequences that hybridize to form the dsRNA duplex of the activator-RNA is 100%.

[00117] In some embodiments, the spacer sequence of a Cas9 gRNA (whether it is a single molecule gRNA or a dual molecule gRNA) of the disclosure is directed to a target sequence in a mammalian organism, e.g. a human or non-human primate. In some embodiments, the spacer sequence of a Cas9 gRNA of the disclosure is directed to a target sequence in a bacteria.

[00118] In some embodiments, the spacer sequence of a Cas9 gRNA of the disclosure is directed to a target sequence in a virus. In some embodiments, the spacer sequence of a Cas9 gRNA of the disclosure is directed to a target sequence in a plant.

[00119] In some embodiments, the single-molecule Cas9 gRNAs of the disclosure can be modified to include an aptamer.

vi. gRNA Arrays

[00120] The Cas9 gRNAs of the disclosure can be provided as gRNA arrays.

[00121] gRNA arrays include more than one gRNA arrayed in tandem, and can be processed into into two or more individual gRNAs. Thus, in some embodiments a precursor Cas9 gRNA array comprises two or more (e.g., 3 or more, 4 or more, 5 or more, 2, 3, 4, or 5) gRNAs (e.g., arrayed in tandem as precursor molecules). In some embodiments, two or more gRNAs can be present on an array (a precursor gRNA array). A Cas9 protein of the disclosure can cleave the precursor gRNA array into individual gRNAs.

[00122] In some embodiments a Cas9 gRNA array includes 2 or more gRNAs (e.g., 3 or more, 4 or more, 5 or more, 6 or more, or 7 or more, gRNAs). The gRNAs of a given array can target (i.e., can include guide sequences that hybridize to) different target sites of the same target DNA. In some embodiments, two or more gRNAs of a precursor gRNA array have the same guide sequence. In some embodiments, the precursor gRNA array comprises two or more gRNAs that target different target sites within the same target DNA. In some embodiments, the precursor gRNA array comprises two or more gRNAs that target different target DNAs.

II. Class 2 Type V CRISPR-Cas RNA-Guided Systems

[00123] Provided herein are novel Class 2 Type V CRISPR-Cas RNA-guided proteins and their gRNAs, constituting the novel Class 2 Type V CRISPR-Cas RNA-guided systems of the disclosure.

[00124] Provided herein are engineered systems comprising: a Class 2 Type V CRISPR-Cas RNA-guided endonuclease protein and a single guide RNA, wherein the gRNA and the Class 2 Type V CRISPR-Cas RNA-guided endonuclease protein do not naturally occur together, wherein the gRNA is capable of hybridizing to a target sequence in a target DNA, wherein the gRNA is capable of forming a complex with the Class 2 Type V CRISPR-Cas RNA-guided endonuclease protein, and wherein the Class 2 Type V CRISPR-Cas RNA-guided endonuclease protein possesses collateral activity and is capable of collaterally cleaving a single stranded polynucleotide comprising RNA, without the use of a tracrRNA. In some embodiments, the Class 2 Type V CRISPR-Cas RNA-guided endonuclease protein comprises the amino acid sequence of SEQ ID NO: 4, or at least 70% sequence identity thereto. In some embodiments, the Class 2 Type V CRISPR-Cas RNA-guided endonuclease protein is capable of collaterally cleaving a single stranded RNA. In some embodiments, the Class 2 Type V CRISPR-Cas RNA-guided endonuclease protein is capable of collaterally cleaving a single stranded DNA/RNA hybrid.

[00125] Also provided herein are engineered systems comprising: (a) a Cas12a.1, Cas12p, or Cas12q protein, or a nucleic acid encoding the Cas12a.1, Cas12p, or Cas12q protein; and (b) a Cas12a.1, Cas12p, or Cas12q gRNA, or a nucleic acid encoding a Cas12a.1, Cas12p, or Cas12q gRNA, wherein the gRNA and the Cas12a.1, Cas12p, or Cas12q protein do not naturally occur together, wherein the gRNA is capable of hybridizing to a target sequence in a target DNA, and the gRNA is capable of forming a complex with the Cas12a.1, Cas12p, or Cas12q protein.

[00126] The components are described in turn below.

a. Class 2 Type V CRISPR-Cas RNA-Guided Proteins

[00127] Provided herein are novel Class 2 Type V CRISPR-Cas RNA-guided endonucleases, e.g. novel Cas12 proteins of the disclosure, including novel Cas12a variants, and novel Cas12 subtypes. In some embodiments the novel Cas12 proteins of the disclosure have been deduced using bioinformatics methods.

[00128] Table 3a shows the protein sequences for the novel Cas12 proteins of the disclosure. Table 2b shows the nucleotide sequences encoding the novel Cas12a proteins of the disclosure.

[00129] SEQ ID NO: 3 represents a novel Cas12a variant of the disclosure, Cas12a.1 (1254 amino acids in length). Cas12a.1 was isolated from a metagenomics sample and deduced to be from *Candidatus Micrarchaeota* archaeon. Based on sequence, function, and structural features it is believed that Cas12a.1 is a Cas12a subtype. FIG. 5A is a schematic representation of the CRISPR Cas cluster around the novel Cas12a.1 gene. FIG. 6A shows the key catalytic amino acids for Cas12a proteins, and alignments of conserved motifs in selected representatives of the Cas12a protein family. FIG. 6B shows the alignment of RuvC1, Bridge Helix, RuvCII, and RuvCIII domains for Cas12a.1 and other selected representatives of the Cas12a protein family. SEQ ID NO: 13 shows the nucleotide sequence encoding the Cas12a.1 of the disclosure.

[00130] SEQ ID NO: 4 represents a novel Cas12 subytpe of the disclosure, Cas12p (1281 amino acids in length). Cas12a.1 was isolated from a metagenomics sample and deduced to be from *Candidatus Peregrinibacteria* bacterium. Based on sequence, function, and structural features described herein, Cas12p differs from the other members of the Cas12 family identified to date and thus is a novel Cas12 enzyme. This novel Cas12 subtype possesses unique properties, not seen in other Cas12 proteins, for example, the ability to collaterally cleave a RNA or DNA containing sequence, e.g. single stranded DNA, singled stranded RNA, and single stranded chimeric RNA/DNA, without the use of a tracrRNA. It is noted that SEQ ID NO: 222 also in Table 3a is N-terminal truncation of the Cas12p of SEQ ID NO: 4.

[00131] SEQ ID NO: 14 provides a nucleotide sequence encoding the Cas12p of the disclosure. FIG. 5C is a schematic representation of the CRISPR Cas cluster around the novel Cas12p gene. FIG. 6B.1 shows the alignment of Cas12a.1 vs. SEQ ID NO: 81 of US20160208243, and has a 46.8% sequence identity; and FIG. 6C shows the alignment of Cas12a.1 vs. SEQ ID NO: 3 of US 10,253365, and has a 46.5% sequence identity.

[00132] FIG. 6D shows the amino acid sequence of Cas12p with the RuvC motifs underlined (SEQ ID NO: 4). The FnCas12a sequence referenced in Shmakov et al., 2015 was used as a reference for identification of the Ruv motifs. FIG 6E shows the alignment of Cas12p with Cas12g1, another Cas12 enzyme. This figure shows an alignment of Cas12p with Cas12g1. Although Cas12g1 has been reported to possess the ability to collaterally cleave RNA (*trans*-cleavage), the sequence homology is less than 8.9% as retrieved by the program Clustal Omega. The very low homology between the enzymes and the lack of conserved domains indicate that

they are members of different enzyme families. Moreover, Cas12g1 requires the presence of a tracr sequence, Cas12p does now, providing an additional functional distinction.

[00133] In the following figures, the structure of Cas12p protein was modeled based on Fn Cas12a structure with Swiss Model server. The sequence identity between the proteins is 38.34%. The model covered the entire sequence of the Cas12p protein. FIG 6F shows a structural analysis of Cas12p using the Swiss Model server. FIG 6G shows a spatial prediction of non-conserved amino acid residues in Cas12p. It is seen that the non-conserved residues are located on protein exposed surface. These differences could reflect changes on first contact with substrates and solvent interactions. FIG 6H shows the approximation of charge distribution over the surface of Cas12p. Using the model showed in FIG 6F, vacuum electrostatics generated by Pymol software allowed for the modeling of the approximation of charge distribution over the surface of the proteins. The positive to negative charge is represented from white to black, the white zones representing the most positive ones. The white oval highlights the active site groove on both positions. The figure shows a slight increase of positive charges on the active site groove of Cas12p protein in comparison to FnCas12a. An increase of positive charge could be related to a stronger interaction with a negative charge substrate and could explain the increased affinity of Cas12p to RNA and DNA substrates. FIG 6I shows predicted structural differences between Cas12p and FnCas12a based on protein sequences. On FnCas12a, the region 696-706 on PAM-interacting domain is related to the binding and cleavage of target DNA and the region 842-852 on Wedge III region is related to pre-cRNA processing (Swarts et al, 2017). When compared to Cas12p, the enzyme presents low homology on those regions, given the deletion of the sequences KNGNPQKGY (SEQ ID NO: 113) on position 699 and PAKE (SEQ ID NO: 114) on position 844. Due to the catalytic relevance of those regions, it is possible to relate the sequence changes to changes seen with the catalysis. The deletions are predicted to impact on the secondary structure of Cas12p. The figures show a superimposition of the model of Cas12p (light grey) and the structure of FnCas12a (dark grey), the deleted sequences are shown in black. The lack of the sequence KNGNPQY (SEQ ID NO: 115) is reflected on a loop shortening. The lack of PAKE sequence (SEQ ID NO: 114; added to additional changes on the loop), decrease a loop length and reduce the negative charges on that position of Cas12p. FIG 6J shows RuvCIII domain structural analysis of Cas12p based on structural analysis with Swiss Model server. The FnCas12a sequence referenced in Shmakov et al., 2015 was used as a reference for identification of the Ruv motifs. Although the RuvCIII region is conserved on Cas12p and the prototypic

Cas12a proteins, the Cas12p has several differences on the sequences surrounding the domain. The presence of these changes impact on the secondary structure of Cas12p (as is shown in black) and could account for the differential RNA cleavage activity of this enzyme. In the structural model depicted in the figure, the superimposition of the structure of the RuvCIII region of the studied Cas12a enzymes and the model of Cas12p. Changes on the secondary structure of Cas12p are circled and shown in black. FIGS. 9B, 9C and 17 show the unique collateral activity of this novel Cas12p enzyme.

[00134] SEQ ID NO: 5 represents a novel Cas12 of the disclosure, Cas12q (1137 amino acids in length). FIG. 5E is a schematic representation of the CRISPR Cas cluster around the novel Cas12q gene. FIG. 6K shows the Cas12q sequence with RuvC motifs underlined for the novel Cas12 protein of the disclosure, Cas12q. The FnCas12a sequence referenced in Shmakov et al., 2015 was used as a reference for identification of the Ruv motifs. SEQ ID NO: 15 shows the nucleotide sequence encoding the Cas12q of the disclosure.

Table 3a

Cas12a.1 MKVSTWDSFTNQYPLTKTLRFELKPVGKTLQKIQDRNLITEDEQRQKDFNKVKKI MDGYYKQFIEECLEGAKIPLKKLEENNNAYTKLKKDPYNKKLREEYAKLQKQLR KLIHDEINKKEEFKYLFKKEFIKKILPEWLEKKGKKEELKEIEKFDKWVTYFSGFFN NRKNVFSSDEISTSMIYRIVNDNLPKFLDDVSRFGEITRYKEFDANQIEENFESELNG EKLKDFFNLKNFNNCLNQEGIEKFNLIIGGKSEEGNNKIKGLNELVNELAQKQADK NEQKKVRKLKLAPLFKQILSDRKSSSFAFEKFEENTEVFDAIDEFYDKISLETLKKIE ATLEKLEEKDLELVYLKNDRCLTGISQEVFGDRERVLQALREYAKTELGLKTDKKI EKWMKKGRYSIHEIESGLKKIGSTGHPICNYFSKLEEKKTNLIQEIKKARTEYEKIS DKKKKLTAESQEPNVARIKALLDSIMRLYHFIKPLNINFKNKKEKDSEALETDNDF YNDFDESFAELGNIIPLYNQVRNYVTQKPFSTEKFKLNFENPKLLSGWDKNKEKD YYSVILRKEESYYLAIMTPKQKNVFDELERLPAGKNYFEKIDYKLLPTPEKNLPRIL FAKKNISFYKPSKEIEAIRNHSAHTKHGNPQNGFKKRDFRLSDCHKMIDFYKKSIQ KHPEWKEYDFQFKKTEDYVDISEFYKEVSDQGYKIEFKKISEKYLLDLVEEGKLYL FQIWNKDFSKYSEGRKNLHTIYWKELFSKENLSDITYKLNGEAEIFYRPKSMERKV THPKNQKIENKDPIKGKKFSKFKYDFIKNKRYTEDRFFFHCPITLNFQARDGSKTIN KRVNDHIRETKDDIFVLSIDRGERHLAYYTLLNSKGEIQEQGSFNVISDDKERKRD YHEKLDEREKERDKARKSWQKIETIKKLKDGYLSQIVHKIAKLAIEKNAIIVLEDLN LDFKRGRLKIEKQVYQKFEKKLIDKLNYLVFKERTEKEAGGSLNAYQLTGKFEGF KKLGKETGIIYYVPAAYTSKICPKTGFVNLLRPKFKNIEKAKEFFKKFNYIKYDSSE GLFEFNFDYSKFIKNGKKETKIIQDNWSVYSNGTKLVGFRNKNKNNSWDTKEVKP NEKLKILFKEYGVSFQKDENIISQIASQNKKAFFENLIKIFKTILMLRNSRKDPEEDY VLSCVKDENGEFFDSRKAKDNEPKDADANGAYHIGLKGLMLLERIKANKGKKKL DLLISRNDFINFAVERSK **(SEQ ID NO: 3)**
Cas12p

MKKSIFDQFVNQYALSKTLRFELKPVGETGRMLEEAKVFAKDETIKKKYEATKPF
FNKLHREFVEEALNEVELAGLPEYFEIFKYWKRYKKKFEKDLQKKEKELRKSVVG
FFNAQAKEWAKKYETLGVKKKDVGLLFEENVFAILKERYGNEEGSQIVDESTGKD
VSIFDSWKGFTGYFIKFQETRKNFYKDDGTATALATRIIDQNLKRFCDNLLIFESIRD
KIDFSEVEQTMGNSIDKVFSVIFYSSCLLQEGIDFYNCVLGGETLPNGEKRQGINELI
NLYRQKTSEKVPFLKLLDKQILSEKEKFMDEIENDEALLDTLKIFRKSAEEKTTLLK
NIFGDFVMNQGKYDLAQIYISRESLNTISRKWTSETDIFEDSLYEVLKKSKIVSASV
KKKDGGYAFPEFIALIYVKSALEQIPTEKFWKERYYKNIGDVLNKGFLNGKEGVW
LQFLLIFDFEFNSLFEREIIDENGDKKVAGYNLFAKGFDDLLNNFKYDQKAKVVIK
DFADEVLHIYQMGKYFAIEKKRSWLADYDIDSFYTDPEKGYLKFYENAYEEIIQVY
NKLRNYLTKKPYSEDKWKLNFENPTLADGWDKNKEADNSTVILKKDGRYYLGL
MARGRNKLFDDRNLPKILEGVENGKYEKVVYKYFPDQAKMFPKVCFSTKGLEFF
QPSEEVITIYKNSEFKKGYTFNVRSMQRLIDFYKDCLVRYEGWQCYDFRNLRKTE
DYRKNIEEFFSDVAMDGYKISFQDVSESYIKEKNQNGDLYLFEIKNKDWNEGANG
KKNLHTIYFESLFSADNIAMNFPVKLNGQAEIFYRPRTEGLEKERIITKKGNVLEKG
DKAFHKRRYTENKVFFHVPITLNRTKKNPFQFNAKINDFLAKNSDINVIGVDRGEK
QLAYFSVISQRGKILDRGSLNVINGVNYAEKLEEKARGREQARKDWQQIEGIKDL
KKGYISQVVRKLADLAIQYNAIIVFEDLNMRFKQIRGGIEKSVYQQLEKALIDKLTF
LVEKEEKDVEKAGHLLKAYQLAAPFETFQKMGKQTGIVFYTQAAYTSRIDPVTG
WRPHLYLKYSSAEKAKADLLKFKKIKFVDGRFEFTYDIKSFREQKEHPKATVWTV
CSCVERFRWNRYLNSNKGGYDHYSDVTKFLVELFQEYGIDFERGDIVGQIEVLET
KGNEKFFKNFVFFFNLICQIRNTNASELAKKDGKDDFILSPVEPFFDSRNSEKFGED
LPKNGDDNGAFNIARKGLVIMDKITKFADENGGCEKMKWGDLYVSNVEWDNFV
ANK **(SEQ ID NO: 4)**

Cas12p truncated
NSIDKVFSVIFYSSCLLQEGIDFYNCVLGGETLPNGEKRQGINELINLYRQKTSEKVP
FLKLLDKQILSEKEKFMDEIENDEALLDTLKIFRKSAEEKTTLLKNIFGDFVMNQGK
YDLAQIYISRESLNTISRKWTSETDIFEDSLYEVLKKSKIVSASVKKKDGGYAFPEFI
ALIYVKSALEQIPTEKFWKERYYKNIGDVLNKGFLNGKEGVWLQFLLIFDFEFNSL
FEREIIDENGDKKVAGYNLFAKGFDDLLNNFKYDQKAKVVIKDFADEVLHIYQMG
KYFAIEKKRSWLADYDIDSFYTDPEKGYLKFYENAYEEIIQVYNKLRNYLTKKPYS
EDKWKLNFENPTLADGWDKNKEADNSTVILKKDGRYYLGLMARGRNKLFDDRN
LPKILEGVENGKYEKVVYKYFPDQAKMFPKVCFSTKGLEFFQPSEEVITIYKNSEFK
KGYTFNVRSMQRLIDFYKDCLVRYEGWQCYDFRNLRKTEDYRKNIEEFFSDVAM
DGYKISFQDVSESYIKEKNQNGDLYLFEIKNKDWNEGANGKKNLHTIYFESLFSAD
NIAMNFPVKLNGQAEIFYRPRTEGLEKERIITKKGNVLEKGDKAFHKRRYTENKVF
FHVPITLNRTKKNPFQFNAKINDFLAKNSDINVIGVDRGEKQLAYFSVISQRGKILD
RGSLNVINGVNYAEKLEEKARGREQARKDWQQIEGIKDLKKGYISQVVRKLADL
AIQYNAIIVFEDLNMRFKQIRGGIEKSVYQQLEKALIDKLTFLVEKEEKDVEKAGH
LLKAYQLAAPFETFQKMGKQTGIVFYTQAAYTSRIDPVTGWRPHLYLKYSSAEKA
KADLLKFKKIKFVDGRFEFTYDIKSFREQKEHPKATVWTVCSCVERFRWNRYLNS
NKGGYDHYSDVTKFLVELFQEYGIDFERGDIVGQIEVLETKGNEKFFKNFVFFFNLI
CQIRNTNASELAKKDGKDDFILSPVEPFFDSRNSEKFGEDLPKNGDDNGAFNIARK
GLVIMDKITKFADENGGCEKMKWGDLYVSNVEWDNFVANK **(SEQ ID NO: 222)**

Cas12q
MINIDELKNLYKVQKTITFELKNKWENKNDENDRVEFLKTQEWVESLFKVDEENF
DEKESIPNLLDFGQKIASLFYKLSEDIANNQIDTRVLKVSKFLLEEIDRNQYHEKKN
KPTKVKEMNPNTNKSYIKEYKLSDQNTLYVLLKIMEDEGRGLQKFLYDKADRLN
LYNQKVRRDFALKESNEQQKFSGNANYYGNIKLLIDSLEDAVRIIGYFTFDDQAEN
AQINEFKSVKQEMNNNEASYQALKDFAIDNAKKEIELTTLNHRAVNKDPKKIQEQI
EEVENFEEDINQLKHQISALNDKKFDVVSRLKHALIKMLPELNLLDAESEQGREVQ
QIYQDKKNGLELDDFKFNLLKHHQWQKTIFKYIKLEGLVLPDLYAENKQDKIKVY
IENYRQSGERISKKAREELGKIDKREEFNGNDELKKAWYEYKDFCRDKRNKSVEL
GNKKSLYNAIKREVLRQKMCNHFAVLVSDGEDTSPYYYLILIPNENSDEMNRTFK
ELKASEGNWKMLDYNRLTFKALEKLALLRSSTFEIADQELQEEAKKIWEEYKEKA
YKDFKNKKLLQGLSGRQREEKKQELQKESLNRVINYLIRCIQSLPDSGKYNFNFKE
PHQYQSLEEFAEEIDRQGYHCAWKNVSKDKLMELEAMEKIKVFKLHNKDFRKVK
LNDSKHNPNLFTLYWLDAMNLDKVNVRLLPEVDLYKRAKETQLKLFERDVKCNI
NNQKIKSIKEKNRLFQDKLYASFKLEFYPENEGLGFEQVNDKVNNFCGSDTAYYL
GLDRGEKELVTFCLVDSDGRLVKNGDWTKFKEVNYADKLKQFYYSKGEIESTQQ
QLLEARDNIKQATNTEDKESMKLNYKKLELKLKQQNLLAQEFIKKAYCGYLIDSI
NEILREYPNTYLVLEDLDIAGKADPESGMTNKEQNLNKTMGASVYQAIENAIVNK
FKYRTVKLSDIKGLQTVPNVVKVEDLREVKEVEDGEHKFGLIRSVKSKDQIGNILF
VDEGETSNTCPNCGFNSDWFKRDVDFDLEIVATVNGQKNAVIEQNDKKYCFPGEI
YKLEIINKEYETNKRNLAMIFKPRAKACRKFINNNLDKNDYFYCPYCAFSSKNCNN
PKLQNGDFVVYSGDDVAAYNVAIRGINLLNNIK **(SEQ ID NO: 5)**

[00135] As used herein, Cas12a.1 includes SEQ ID NO: 3 and proteins with at least 70% - 99.5% sequence identity thereto. Accordingly, provided herein are proteins comprising the amino acid sequence of SEQ ID NO: 3 and proteins with at least 70%, at least 71%, at least 72%, at least 73%, at least 74%, at least 75%, at least 76%, at least 77%, at least 78%, at least 79%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% sequence identity thereto. Also provided herein are nucleic acids encoding the proteins comprising the amino acid sequence of SEQ ID NO: 3 and proteins with at least 70% - 99.5% sequence identity thereto.

[00136] As used herein, Cas12p includes SEQ ID NO: 4 and proteins with at least 70% - 99.5% sequence identity thereto. Accordingly, provided herein are proteins comprising the amino acid sequence of SEQ ID NO: 4 and proteins with at least 70%, at least 71%, at least 72%, at least 73%, at least 74%, at least 75%, at least 76%, at least 77%, at least 78%, at least 79%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least

94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% sequence identity thereto. Also provided herein are nucleic acids encoding the proteins comprising the amino acid sequence of SEQ ID NO: 4 and proteins with at least 70% - 99.5% sequence identity thereto.

[00137] Also provided herein are proteins comprising the amino acid sequence of SEQ ID NO: 222 and proteins with at least 70% - 99.5% sequence identity thereto. Also provided herein are nucleic acids encoding the proteins comprising the amino acid sequence of SEQ ID NO: 222 and proteins with at least 70% - 99.5% sequence identity thereto.

[00138] As used herein, Cas12q includes SEQ ID NO: 5 and proteins with at least 70% - 99.5% sequence identity thereto. Accordingly, provided herein are proteins comprising the amino acid sequence of SEQ ID NO: 5 and proteins with at least 70%, at least 71%, at least 72%, at least 73%, at least 74%, at least 75%, at least 76%, at least 77%, at least 78%, at least 79%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% sequence identity thereto. Also provided herein are nucleic acids encoding the proteins comprising the amino acid sequence of SEQ ID NO: 5 and proteins with at least 70% - 99.5% sequence identity thereto.

[00139] Table 3b shows exemplary nucleotide sequences, and exemplary codon optimized nucleic acid sequences for the novel Cas12 proteins of the disclosure.

Table 3b

Cas12a.1
ATGAAGGTCTCGACTTGGGATTCGTTTACAAACCAATACCCCCTAACGAAAACTCTA
CGCTTCGAATTAAAGCCAGTCGGCAAAACACTGCAGAAAATTCAAGATCGCAACCTG
ATTACAGAAGACGAACAACGCCAAAAAGATTTCAACAAAGTCAAAAAAATAATGGA
CGGATACTACAAGCAATTCATAGAAGAATGCTTGGAAGGTGCCAAGATACCGTTAAA
AAAATTGGAAGAAAACAACAACGCTTACACGAAACTGAAAAAAGACCCTTACAACA
AAAAATTAAGGGAAGAATACGCAAAACTCCAAAAACAATTAAGGAAACTAATTCAC
GACGAAATAAATAAAAAAGAAGAATTCAAATACTTGTTCAAGAAAGAATTCATCAA
AAAAATATTGCCGGAATGGCTCGAAAAAAAAGGGAAAAAAGAGGAACTCAAAGAA
ATCGAAAAATTCGATAAATGGGTTACCTACTTTAGCGGTTTTTTTAACAACCGCAAA
AACGTTTTTTCAAGCGATGAAATTTCGACGTCAATGATTTACAGGATAGTCAACGAC
AACCTACCGAAATTCCTAGATGACGTTTCACGCTTCGGAGAAATAACCAGATACAAG
GAATTTGACGCCAACCAAATAGAAGAAAACTTTGAAAGCGAGTTGAACGGAGAGAA
ATTAAAAGATTTTTTCAACTTGAAAAACTTCAACAACTGCCTTAACCAAGAAGGAAT
AGAAAAATTCAACTTAATCATAGGAGGCAAAAGCGAAGAAGGCAACAATAAAATAA
AGGGCTTAAACGAATTAGTCAACGAACTCGCCCAAAAACAAGCGGACAAAAACGAG
CAAAAAAAGGTTAGAAAATTAAAACTCGCGCCGTTATTCAAGCAAATCTTAAGTGAC

CGCAAATCCTCCTCGTTCGCATTCGAAAAATTCGAGGAAAATACGGAGGTATTCGAT
GCAATAGACGAATTTTACGATAAAATAAGCTTGGAAACACTCAAAAAAATAGAAGC
GACCCTCGAAAAGCTAGAAGAAAAAGATTTGGAATTAGTTTACTTGAAAAACGATA
GATGCCTAACAGGAATTTCACAAGAAGTATTCGGGGATCGGGAAAGAGTACTTCAA
GCCCTAAGGGAATACGCGAAAACCGAACTCGGCCTCAAAACCGACAAAAAAATAGA
AAAATGGATGAAAAAAGGCAGGTATTCAATCCACGAAATAGAGAGCGGCCTCAAAA
AAATCGGTTCAACCGGACACCCGATATGTAATTATTTCTCAAAACTAGAAGAAAAAA
AGACAAACTTGATTCAAGAAATAAAAAAAGCGCGCACTGAATATGAAAAAATAAGT
GACAAAAAAAAGAAATTAACTGCTGAAAGCCAAGAGCCCAACGTCGCAAGAATAAA
GGCGTTACTGGACTCAATAATGCGGCTATACCACTTCATAAAACCCCTCAACATCAA
CTTCAAAAACAAGAAAGAAAAGGATTCAGAGGCACTTGAAACCGATAACGATTTCT
ATAACGATTTCGACGAATCGTTTGCGGAACTAGGGAATATAATCCCACTATACAATC
AAGTCAGAAACTATGTTACGCAAAAACCGTTCAGCACCGAAAAATTCAAGTTAAACT
TTGAAAATCCCAAACTCCTAAGCGGCTGGGACAAAAACAAGGAAAAAGACTATTAT
TCTGTTATATTGAGAAAAGAGGAGTCATACTACTTAGCCATTATGACCCCAAAACAA
AAAAACGTTTTTGACGAACTGGAACGGCTTCCGGCTGGAAAAAATTATTTTGAAAAA
ATAGACTACAAATTATTGCCTACCCCAGAAAAAAATCTACCTAGAATATTATTTGCA
AAAAAAAACATTTCATTTTACAAGCCATCAAAAGAAATCGAAGCGATTCGTAATCAC
TCTGCCCACACCAAGCATGGAAACCCACAAAACGGGTTCAAAAAAAGGGATTTCCG
ATTAAGCGATTGCCATAAAATGATTGACTTTTACAAAAAGAGCATTCAAAAACACCC
CGAATGGAAAGAATACGATTTCCAATTCAAAAAAACGGAAGATTACGTCGACATATC
AGAATTTTATAAAGAAGTATCCGACCAAGGCTATAAAATAGAATTCAAAAAAATAA
GCGAAAAATATTTGCTTGACTTGGTCGAAGAAGGAAAACTTTACTTATTCCAAATTT
GGAACAAGGACTTTTCGAAGTATTCGGAAGGCCGTAAAAACCTGCACACAATTTACT
GGAAAGAACTATTCTCCAAAGAAAACCTTTCAGACATAACTTACAAATTAAACGGCG
AAGCCGAAATATTCTACCGCCCAAAGTCAATGGAAAGGAAAGTAACTCACCCAAAA
AACCAAAAAATAGAAAACAAAGACCCGATTAAAGGGAAAAAATTCAGTAAATTCAA
ATACGACTTTATAAAAAACAAAAGGTACACCGAAGACCGTTTCTTCTTCCACTGCCC
GATAACCTTGAACTTCCAGGCGCGCGATGGCAGCAAAACGATTAACAAGCGGGTCA
ACGACCACATACGCGAAACAAAAGATGACATTTTCGTGTTAAGCATTGACCGCGGGG
AAAGGCACTTGGCGTACTACACGCTATTGAATTCAAAAGGAGAAATCCAAGAACAA
GGCTCTTTCAACGTAATCTCGGACGACAAAGAAAGAAAACGTGATTACCACGAAAA
ACTGGATGAACGCGAAAAAGAACGCGACAAAGCAAGGAAAAGCTGGCAGAAAATC
GAGACCATAAAGAAATTGAAGGATGGCTACCTATCCCAAATCGTACACAAAATCGCT
AAACTCGCAATAGAAAAAAACGCGATAATCGTCTTGGAAGACCTGAACTTAGACTTC
AAGCGCGGGAGATTAAAAATCGAGAAGCAAGTATACCAAAAGTTCGAGAAAAAACT
AATAGACAAACTCAATTACTTGGTTTTCAAGGAAAGAACCGAAAAAGAAGCCGGCG
GATCCCTAAACGCATACCAACTAACCGGAAAATTTGAAGGATTTAAGAAACTCGGAA
AAGAAACAGGTATAATATACTACGTTCCCGCGGCGTACACCTCGAAGATTTGCCCGA
AAACAGGCTTCGTAAATCTGTTAAGACCTAAATTCAAGAACATAGAAAAAGCTAAG
GAATTCTTCAAAAAATTCAACTACATCAAATACGATTCGAGCGAAGGCTTATTCGAA
TTCAACTTCGACTACTCCAAATTCATTAAAAACGGAAAAAAAGAAACAAAAATAATT
CAAGACAATTGGTCGGTTTACTCGAACGGAACGAAACTAGTCGGCTTCAGAAACAAG
AATAAAAACAATTCATGGGATACAAAGGAAGTCAAACCGAACGAAAAACTAAAAAT
ATTGTTCAAAGAATACGGGGTTTCCTTCCAAAAAGACGAAAATATTATAAGCCAAAT
AGCCAGCCAAAACAAAAAAGCTTTCTTTGAAAACCTCATTAAAATCTTTAAAACGAT
TTTAATGTTACGCAACTCAAGAAAAGACCCCGAAGAAGATTACGTACTTTCCTGCGT
AAAAGACGAAAACGGCGAATTCTTCGACTCAAGAAAAGCTAAAGACAACGAGCCCA

AGGACGCCGACGCGAACGGCGCTTACCACATAGGGTTGAAAGGATTAATGCTCTTGG
AAAGAATAAAGGCCAACAAAGGAAAGAAAAAACTCGATTTACTAATCAGCAGGAAC
GACTTCATCAACTTCGCAGTTGAACGGAGCAAGTAA **(SEQ ID NO: 13)**

Cas12p

ATGAAAAAATCTATTTTTGATCAGTTTGTAAATCAGTATGCTCTTTCTAAAACGTTGC
GGTTTGAATTGAAGCCGGTGGGGGAGACGGGGAGGATGCTTGAGGAGGCGAAGGTT
TTTGCTAAAGATGAAACAATCAAGAAAAAATATGAGGCAACCAAGCCTTTTTTTAAT
AAATTGCATCGTGAATTTGTAGAGGAGGCTTTAAATGAGGTGGAATTAGCTGGTTTG
CCTGAATATTTTGAAATATTTAAATATTGGAAAAGGTATAAAAAGAAGTTTGAAAAG
GATTTGCAGAAGAAAGAAAAAGAATTGCGGAAATCAGTTGTAGGTTTTTTTAATGCA
CAGGCAAAGGAATGGGCGAAAAAATATGAAACTTTGGGTGTGAAGAAAAAAGATGT
GGGACTTTTATTTGAAGAAAATGTTTTTGCTATATTGAAGGAAAGGTACGGAAATGA
GGAGGGATCACAAATTGTTGATGAAAGTACAGGAAAAGATGTTTCGATATTTGATAG
TTGGAAGGGCTTTACAGGGTATTTTATTAAATTCCAGGAAACTCGTAAGAATTTTTAT
AAGGATGATGGCACGGCTACTGCTTTGGCTACAAGGATTATTGATCAAAATTTGAAG
CGTTTTTGTGATAATTTACTAATATTTGAAAGTATTAGAGATAAAATTGATTTTTCAG
AGGTAGAACAAACTATGGGAAACTCTATTGATAAGGTTTTTTCAGTAATTTTTTATAG
TTCCTGTTTACTTCAGGAAGGAATTGATTTTTATAATTGTGTTTTAGGTGGGGAGACT
CTGCCAAATGGTGAAAAGAGACAGGGAATAAATGAGCTTATTAATCTCTATAGGCAA
AAAACTAGTGAGAAAGTACCTTTTTTAAAGTTGCTTGATAAGCAGATTTTGAGTGAA
AAAGAGAAGTTTATGGATGAAATTGAAAATGATGAGGCTCTCTTGGATACTCTTAAA
ATATTTAGAAAATCGGCTGAAGAAAAAACCACTTTGTTAAAAAATATTTTTGGTGAT
TTTGTTATGAATCAGGGTAAGTATGATTTAGCGCAGATTTATATTTCCAGAGAATCTT
TAAATACTATTTCACGGAAATGGACCAGTGAAACAGATATATTTGAGGATTCATTAT
ATGAAGTGTTAAAGAAATCAAAAATAGTTTCTGCCTCTGTAAAAAAGAAAGATGGA
GGGTACGCTTTCCCTGAGTTTATTGCGCTTATTTATGTGAAAAGTGCTCTTGAACAAA
TTCCTACTGAAAAATTTTGGAAGGAGCGATATTATAAAAATATTGGAGATGTTTTGA
ATAAAGGGTTTTTGAATGGTAAGGAAGGTGTCTGGTTACAATTTTTATTGATTTTTGA
TTTTGAATTTAATTCTCTTTTTGAAAGAGAAATAATTGATGAAAATGGAGACAAGAA
AGTGGCCGGATATAATTTGTTTGCCAAGGGTTTTGATGATCTTTTGAATAACTTTAAA
TATGATCAAAAAGCTAAGGTTGTTATTAAGGATTTTGCAGATGAGGTTTTACATATTT
ATCAGATGGGAAAATATTTTGCTATTGAAAAGAAACGTTCTTGGTTGGCTGATTATG
ATATTGATTCATTTTATACTGATCCTGAAAAAGGTTATTTGAAGTTTTATGAAAATGC
GTATGAAGAGATTATTCAAGTTTATAATAAATTGCGAAATTACCTAACGAAGAAACC
TTATAGTGAGGATAAATGGAAACTTAATTTTGAGAATCCAACTTTAGCTGATGGGTG
GGACAAAAATAAAGAAGCTGATAATTCTACAGTTATTTTGAAAAAGGATGGTCGCTA
TTATTTAGGGTTGATGGCTCGCGGGCGAAATAAACTTTTTGATGATAGAAATTTACCA
AAAATTTTGGAGGGCGTTGAGAATGGGAAATATGAGAAAGTTGTATATAAGTATTTT
CCGGATCAGGCAAAAATGTTTCCAAAAGTTTGTTTTTCAACTAAAGGTTTGGAGTTTT
TCCAACCTTCGGAGGAAGTCATTACTATTTACAAAAATTCTGAATTCAAAAAAGGGT
ATACTTTTAATGTAAGGAGTATGCAGAGGCTTATTGATTTTTATAAAGATTGTCTTGT
TAGATATGAGGGGTGGCAATGTTATGATTTTAGAAATTTGAGAAAGACAGAAGATTA
TCGGAAGAATATTGAAGAGTTTTTCAGCGATGTTGCTATGGATGGGTATAAAATATC
CTTTCAGGATGTCTCGGAAAGTTATATTAAAGAGAAAAATCAGAATGGGGATTTATA
TTTATTTGAGATAAAAAATAAAGATTGGAATGAAGGCGCAAATGGAAAGAAAAATT
TGCACACTATATATTTTGAATCTCTTTTTTCGGCTGATAATATTGCCATGAATTTTCCC
GTTAAGTTGAATGGACAAGCGGAAATTTTTTATCGGCCAAGAACAGAGGGGCTGGA

GAAAGAAAGGATAATCACTAAAAAGGGTAATGTTTTGGAAAAAGGAGATAAAGCTT
TTCATAAAAGAAGGTATACGGAAAACAAAGTTTTTTTTCATGTTCCGATTACACTTAA
TCGAACAAAAAAAAAATCCATTTCAATTTAATGCAAAAATTAATGATTTTTTGGCTAA
AAATTCTGATATAAATGTTATTGGGGTCGATCGTGGGGAGAAGCAATTAGCATATTT
TTCTGTTATTTCACAGAGAGGCAAAATTTTGGATAGGGGTAGTTTAAATGTGATAAA
TGGAGTTAATTATGCAGAGAAATTAGAAGAAAAAGCTAGAGGGCGTGAGCAGGCGC
GTAAGGATTGGCAGCAGATTGAAGGTATTAAAGATTTAAAGAAGGGATATATTTCTC
AGGTAGTTAGAAAGCTAGCCGATTTAGCAATTCAGTATAATGCGATTATTGTTTTTGA
AGATTTGAACATGCGGTTTAAGCAGATTCGTGGAGGTATTGAAAAAAGTGTTTATCA
GCAGTTGGAGAAGGCTTTGATTGATAAATTAACTTTTTTGGTTGAAAAGGAAGAAAA
AGATGTAGAAAAGGCAGGTCATTTGTTAAAAGCTTACCAGCTTGCTGCTCCGTTTGA
GACTTTTCAGAAAATGGGTAAACAAACGGGGATTGTTTTTTATACACAGGCTGCATA
TACTTCACGAATTGATCCTGTTACAGGTTGGCGGCCTCACTTGTATTTGAAATATTCC
AGTGCGGAGAAGGCAAAGGCGGATTTATTAAAATTTAAAAAGATAAAGTTTGTGGA
TGGCCGGTTTGAGTTTACTTATGATATTAAGAGTTTTCGTGAACAAAAGGAACATCC
AAAGGCGACTGTCTGGACGGTGTGTTCTTGCGTGGAGAGATTTCGTTGGAATAGATA
TTTAAATAGCAATAAAGGTGGTTATGACCATTACAGTGATGTGACGAAGTTCTTGGT
AGAGCTTTTTCAAGAGTATGGGATTGATTTTGAAAGAGGGGATATTGTCGGGCAAAT
TGAGGTTTTGGAAACGAAGGGAAATGAAAAATTTTTTAAGAATTTCGTTTTTTTCTTT
AATTTGATTTGTCAGATAAGAAATACTAATGCGTCGGAGTTGGCAAAAAAAGATGGA
AAAGATGATTTTATTCTTTCACCGGTGGAACCGTTTTTTGATAGCAGAAATTCGGAGA
AGTTTGGGGAGGATTTGCCAAAAAATGGGGATGATAATGGGGCATTTAATATTGCGA
GGAAAGGGCTTGTTATTATGGATAAAATTACAAAATTTGCAGATGAGAATGGTGGGT
GCGAGAAGATGAAGTGGGGAGATTTGTATGTTTCTAATGTGGAGTGGGATAATTTTG
TAGCTAATAAATGA **(SEQ ID NO: 14)**

Cas12q

ATGATAAATATTGACGAATTAAAAAATTTATATAAAGTTCAAAAAACAATTACTTTT
GAATTAAAAAATAAATGGGAAAATAAGAATGATGAAAATGATAGAGTTGAGTTTTT
AAAGACTCAAGAATGGGTGGAATCTTTATTCAAAGTTGATGAGGAGAATTTTGATGA
AAAGGAGTCAATTCCGAACTTGTTAGATTTCGGCCAAAAGATTGCGAGTCTTTTTTAT
AAGTTGAGTGAAGATATCGCTAATAATCAAATTGATACACGGGTTTTAAAAGTGAGC
AAGTTTTTGTTGGAGGAGATCGATAGAAATCAATATCATGAGAAAAAAAATAAACC
AACAAAGGTTAAGGAGATGAATCCAAATACAAATAAGAGTTATATTAAGGAGTATA
AGTTATCAGATCAAAATACATTGTATGTTCTGTTGAAGATAATGGAAGATGAAGGGC
GGGGTTTACAAAAATTTTTATATGATAAGGCAGACAGATTAAATTTATATAATCAGA
AGGTAAGAAGAGATTTCGCTTTAAAAGAAAGTAACGAACAGCAGAAGTTTTCGGGT
AACGCTAATTATTACGGAAACATAAAATTGTTGATTGATTCATTGGAAGACGCTGTT
CGTATTATTGGTTATTTCACGTTTGATGATCAAGCAGAAAATGCTCAAATAAATGAAT
TCAAGAGCGTTAAGCAGGAAATGAATAACAATGAAGCTTCGTATCAGGCTTTGAAAG
ATTTTGCTATTGATAACGCAAAAAAAGAAATTGAACTTACAACTCTAAATCATAGGG
CTGTTAACAAGGATCCAAAAAAGATACAAGAACAGATTGAAGAAGTGGAAAATTTT
GAAGAAGATATAAATCAATTGAAGCACCAAATTTCTGCGCTTAATGATAAAAAATTT
GATGTAGTGTCAAGATTAAAGCATGCATTAATTAAAATGTTACCGGAGTTGAATTTG
TTAGATGCTGAAAGCGAGCAAGGTAGAGAGGTTCAGCAAATATATCAAGATAAAAA
GAATGGTTTGGAATTAGACGATTTTAAGTTCAATTTGCTTAAACATCATCAATGGCAG
AAAACCATTTTTAAATACATTAAATTAGAGGGTTTGGTTTTACCTGATTTATATGCCG
AAAACAAACAAGATAAGATTAAAGTGTATATTGAAAATTATCGACAAAGCGGAGAA

AGGATAAGTAAAAAGGCACGCGAGGAGTTGGGCAAGATCGATAAAAGAGAGGAATT
TAATGGTAATGATGAACTAAAGAAAGCGTGGTACGAATACAAAGATTTTTGCAGAG
ACAAGCGTAATAAATCCGTGGAATTGGGCAATAAGAAATCACTGTACAATGCCATCA
AGCGTGAGGTTTTAAGGCAGAAAATGTGTAATCATTTTGCCGTATTGGTGAGTGATG
GGGAAGATACATCGCCTTATTATTATTTGATATTAATTCCCAATGAAAACAGTGATG
AAATGAACAGGACATTCAAAGAGCTTAAAGCATCCGAAGGAAATTGGAAGATGCTC
GATTATAACAGATTAACTTTTAAAGCTTTGGAAAAATTGGCATTATTGCGCAGCTCTA
CATTTGAAATTGCAGACCAAGAACTACAAGAAGAAGCTAAAAAAATTTGGGAAGAA
TATAAAGAAAAGGCGTATAAAGATTTTAAGAATAAAAAATTATTACAAGGGCTATCC
GGTCGCCAAAGAGAAGAAAAAAAACAAGAATTGCAAAAAGAAAGTTTAAATCGAGT
TATAAATTATTTAATTCGTTGCATTCAGTCGTTGCCGGATAGCGGTAAATACAATTTT
AATTTTAAAGAACCGCATCAATATCAGAGCTTGGAAGAGTTTGCGGAAGAAATTGAT
AGACAGGGTTATCATTGCGCTTGGAAGAATGTAAGCAAAGACAAGCTTATGGAGCTG
GAGGCGATGGAAAAAATTAAAGTATTTAAATTGCATAATAAGGATTTTAGAAAAGTT
AAACTTAACGATTCGAAACACAATCCGAATCTTTTTACTTTATATTGGCTTGACGCGA
TGAATTTGGATAAAGTCAATGTTCGTTTATTGCCCGAGGTGGATTTATATAAAAGAG
CCAAAGAAACGCAACTAAAATTATTCGAAAGAGATGTAAAGTGCAATATTAATAATC
AAAAAATAAAATCAATTAAAGAAAAAAATAGATTATTTCAAGATAAACTTTACGCTT
CATTCAAGCTGGAATTTTATCCAGAAAACGAAGGTTTGGGTTTTGAACAAGTCAATG
ATAAAGTGAATAATTTTTGCGGAAGTGATACAGCGTATTATTTGGGTTTGGATAGGG
GTGAGAAAGAATTGGTTACGTTTTGCTTGGTTGATTCTGATGGGCGGTTGGTTAAGA
ACGGAGATTGGACGAAGTTTAAAGAGGTTAACTATGCGGATAAATTAAAGCAATTTT
ATTATTCAAAAGGTGAAATAGAATCTACTCAACAACAACTTTTGGAAGCTCGAGACA
ATATTAAACAAGCTACTAACACGGAGGATAAAGAATCGATGAAATTAAACTATAAA
AAATTAGAGTTGAAACTAAAACAACAGAATTTGTTAGCGCAGGAGTTTATTAAAAAA
GCTTATTGCGGTTATTTGATAGATTCAATAAATGAAATATTACGGGAATATCCAAAT
ACGTATCTTGTATTAGAGGATTTGGATATAGCAGGTAAAGCTGACCCCGAAAGCGGC
ATGACCAATAAAGAACAAAATTTAAATAAAACAATGGGTGCCAGCGTTTATCAAGCT
ATTGAAAATGCCATAGTAAATAAGTTTAAATACCGTACTGTTAAATTATCCGATATC
AAAGGTTTGCAAACTGTACCGAATGTAGTGAAGGTGGAAGATTTGCGCGAAGTTAAG
GAAGTGGAAGATGGTGAGCATAAATTTGGTTTGATAAGATCCGTGAAATCAAAGGAT
CAAATTGGCAATATTCTGTTTGTGGATGAAGGAGAAACATCTAATACTTGCCCGAAT
TGCGGATTTAACAGCGATTGGTTTAAGCGGGATGTTGATTTTGATTTGGAGATTGTGG
CTACTGTAAACGGTCAGAAAAATGCGGTTATAGAACAAAACGACAAAAAGTACTGT
TTTCCCGGTGAAATTTATAAGTTAGAAATAATTAATAAAGAATACGAAACAAATAAA
CGGAATTTAGCCATGATTTTTAAACCGCGCGCAAAAGCTTGTAGAAAATTTATAAAT
AATAATTTGGATAAGAATGACTATTTTTATTGCCCGTATTGCGCTTTTTCTAGCAAGA
ACTGCAATAATCCAAAATTGCAAAACGGTGATTTTGTGGTATATTCGGGTGATGATG
TGGCGGCATACAATGTAGCGATCAGAGGTATTAACCTTTTAAACAATATAAAATAG
(SEQ ID NO: 15)

Cas12a.1 codon optimized version:
ATGGGGTCAAGTCATCACCACCACCACCACTCAAGTGGACTAGTACCCCGTGGCAGC
ATGAAAGTTAGCACCTGGGATAGCTTCACCAACCAGTACCCGCTGACCAAGACCCTG
CGTTTTGAGCTGAAGCCGGTGGGTAAAACCCTGCAGAAGATCCAAGACCGTAACCTG
ATTACCGAGGACGAACAGCGTCAAAAGGATTTCAACAAGGTTAAGAAAATCATGGA
TGGTTACTACAAGCAGTTCATCGAGGAATGCCTGGAAGGCGCGAAGATCCCGCTGAA
GAAACTGGAGGAAAACAACAACGCGTACACCAAACTGAAGAAAGACCCGTATAACA

AGAAACTGCGTGAGGAATACGCGAAGCTGCAGAAACAACTGCGTAAACTGATCCAC
GATGAGATTAACAAGAAAGAGGAATTCAAGTACCTGTTTAAGAAAGAATTCATCAA
GAAAATTCTGCCGGAATGGCTGGAGAAGAAAGGTAAGAAAGAGGAACTGAAAGAG
ATCGAAAAGTTCGACAAATGGGTGACCTACTTTAGCGGCTTCTTTAACAACCGTAAG
AACGTTTTCAGCAGCGACGAGATTAGCACCAGCATGATCTATCGTATTGTGAACGAT
AACCTGCCGAAATTCCTGGACGATGTTAGCCGTTTTGGTGAAATTACCCGTTACAAG
GAGTTCGACGCGAACCAGATCGAGGAAAACTTTGAGAGCGAACTGAACGGTGAAAA
ACTGAAGGATTTCTTTAACCTGAAAAACTTCAACAACTGCCTGAACCAAGAAGGCAT
TGAGAAATTTAACCTGATCATTGGTGGCAAGAGCGAGGAAGGTAATAACAAAATCA
AGGGCCTGAACGAACTGGTGAACGAGCTGGCGCAAAAACAAGCGGACAAGAACGA
GCAGAAGAAAGTTCGTAAACTGAAGCTGGCGCCGCTGTTCAAACAAATCCTGAGCG
ATCGTAAGAGCAGCAGCTTTGCGTTCGAAAAATTTGAGGAAAACACCGAGGTGTTCG
ACGCGATCGATGAATTTTATGACAAGATTAGCCTGGAGACCCTGAAGAAAATCGAAG
CGACCCTGGAGAAACTGGAGGAAAAGGACCTGGAACTGGTTTACCTGAAAAACGAT
CGTTGCCTGACCGGTATCAGCCAGGAAGTGTTCGGCGACCGTGAGCGTGTTCTGCAA
GCGCTGCGTGAATACGCGAAAACCGAGCTGGGTCTGAAGACCGATAAGAAAATCGA
GAAGTGGATGAAGAAAGGTCGTTATAGCATCCACGAGATTGAAAGCGGCCTGAAGA
AAATCGGTAGCACCGGCCACCCGATTTGCAACTACTTCAGCAAACTGGAGGAAAAG
AAAACCAACCTGATCCAGGAAATTAAGAAAGCGCGTACCGAGTATGAAAAGATCAG
CGACAAGAAAAAGAAACTGACCGCGGAAAGCCAAGAGCCGAACGTGGCGCGTATCA
AAGCGCTGCTGGATAGCATTATGCGTCTGTATCACTTCATCAAGCCGCTGAACATCA
ACTTCAAGAACAAGAAAGAGAAGGACAGCGAAGCGCTGGAGACCGACAACGATTTT
TACAACGACTTCGATGAAAGCTTTGCGGAGCTGGGCAACATCATTCCGCTGTACAAC
CAAGTGCGTAACTATGTTACCCAAAAACCGTTCAGCACCGAGAAATTCAAGCTGAAC
TTTGAAAACCCGAAGCTGCTGAGCGGTTGGGACAAAAACAAGGAAAAAGATTACTA
TAGCGTGATTCTGCGTAAAGAGGAAAGCTACTATCTGGCGATCATGACCCCGAAGCA
GAAAAACGTTTTCGACGAGCTGGAACGTCTGCCGGCGGGCAAAAATTACTTCGAGAA
GATCGATTACAAGCTGCTGCCGACCCCGGAAAAGAACCTGCCGCGTATCCTGTTCGC
GAAGAAAAACATTAGCTTTTACAAGCCGAGCAAAGAGATCGAAGCGATTCGTAACC
ACAGCGCGCACACCAAACACGGTAACCCGCAGAACGGCTTCAAGAAACGTGACTTT
CGTCTGAGCGATTGCCACAAGATGATCGACTTCTACAAGAAAAGCATTCAGAAACAC
CCGGAATGGAAGGAGTATGATTTTCAATTCAAGAAAACCGAGGACTACGTGGATATC
AGCGAATTCTATAAAGAGGTTTCTGACCAGGGTTACAAGATCGAATTCAAGAAAATT
AGCGAGAAATACCTGCTGGACCTGGTGGAGGAAGGTAAACTGTACCTGTTCCAAATC
TGGAACAAGGATTTCAGCAAGTACAGCGAAGGCCGTAAAAACCTGCACACCATCTAT
TGGAAAGAACTGTTCAGCAAGGAGAACCTGAGCGATATTACCTATAAGCTGAACGG
CGAGGCGGAAATCTTTTACCGTCCGAAAAGCATGGAGCGTAAGGTTACCCACCCGAA
GAACCAGAAAATCGAAAACAAAGACCCGATCAAGGGTAAGAAATTCAGCAAGTTCA
AGTATGACTTCATCAAGAACAAGCGTTACACCGAGGATCGTTTCTTTTTCCACTGCCC
GATCACCCTGAACTTTCAAGCGCGTGACGGCAGCAAAACCATCAACAAGCGTGTGAA
CGATCACATTCGTGAGACCAAAGACGATATCTTCGTTCTGAGCATTGATCGTGGTGA
ACGTCACCTGGCGTACTATACCCTGCTGAACAGCAAGGGTGAAATTCAGGAGCAAGG
CAGCTTTAACGTGATCAGCGACGATAAGGAGCGTAAACGTGACTATCACGAAAAACT
GGATGAGCGTGAAAAGGAGCGTGACAAGGCGCGTAAAAGCTGGCAGAAAATCGAG
ACCATTAAGAAACTGAAGGATGGCTACCTGAGCCAAATCGTGCACAAGATTGCGAA
ACTGGCGATCGAGAAAAACGCGATCATTGTTCTGGAAGACCTGAACCTGGATTTCAA
GCGTGGTCGTCTGAAGATTGAGAAACAGGTGTACCAAAAATTCGAAAAGAAACTGA
TCGACAAGCTGAACTATCTGGTTTTTAAAGAACGTACCGAAAAAGAGGCGGGTGGTA

GCCTGAACGCGTATCAGCTGACCGGTAAATTCGAGGGCTTTAAGAAACTGGGCAAGG
AAACCGGCATCATTTACTATGTGCCGGCGGCGTACACCAGCAAAATCTGCCCGAAGA
CCGGCTTCGTTAACCTGCTGCGTCCGAAGTTCAAGAACATCGAAAAGGCGAAGGAGT
TTTTCAAGAAGTTCAACTACATCAAGTACGACAGCAGCGAAGGTCTGTTTGAGTTCA
ACTTCGATTACAGCAAGTTCATCAAGAACGGCAAGAAAGAGACCAAAATCATTCAG
GACAACTGGAGCGTGTATAGCAACGGTACCAAGCTGGTTGGCTTCCGTAACAAGAAC
AAAAACAACAGCTGGGATACCAAGGAAGTGAAACCGAACGAGAAGCTGAAAATTCT
GTTCAAAGAGTACGGTGTTAGCTTTCAAAAGGACGAAAACATCATTAGCCAGATCGC
GAGCCAAAACAAGAAAGCGTTTTTCGAGAACCTGATCAAGATTTTCAAAACCATTCT
GATGCTGCGTAACAGCCGTAAAGACCCGGAGGAAGATTACGTGCTGAGCTGCGTTAA
GGACGAAAACGGCGAGTTTTTCGACAGCCGTAAGGCGAAAGATAACGAGCCGAAAG
ACGCGGATGCGAACGGCGCGTACCACATTGGTCTGAAGGGCCTGATGCTGCTGGAAC
GTATCAAGGCGAACAAAGGTAAGAAAAAGCTGGACCTGCTGATCAGCCGTAACGAT
TTCATTAACTTTGCGGTTGAGCGTAGCAAGTAA **(SEQ ID NO: 16)**

Cas12p codon optimized:

ATGGGATCAAGTCATCACCACCACCACCACTCAAGTGGACTAGTACCCAGGGGAAGC
ATGAAGAAGAGCATTTTCGATCAGTTCGTTAACCAGTACGCGCTGAGCAAGACCCTG
CGTTTCGAGCTGAAACCGGTGGGTGAAACCGGCCGTATGCTGGAGGAAGCGAAGGT
TTTCGCGAAGGATGAAACCATTAAGAAAAAGTACGAAGCGACCAAGCCGTTCTTTAA
CAAACTGCACCGTGAATTCGTGGAGGAAGCGCTGAACGAGGTTGAACTGGCGGGCC
TGCCGGAGTACTTCGAAATCTTCAAGTACTGGAAGCGTTACAAAAAGAAATTCGAGA
AGGACCTGCAGAAGAAAGAGAAGGAACTGCGTAAAAGCGTGGTTGGTTTCTTTAAC
GCGCAAGCGAAGGAGTGGGCGAAGAAATATGAAACCCTGGGCGTGAAGAAAAAGG
ATGTTGGTCTGCTGTTCGAGGAAAACGTGTTTGCGATTCTGAAAGAACGTTACGGTA
ACGAGGAAGGCAGCCAGATTGTGGACGAGAGCACCGGCAAGGATGTTAGCATCTTC
GACAGCTGGAAGGGTTTTACCGGCTATTTCATCAAATTTCAGGAAACCCGTAAGAAC
TTCTACAAAGATGATGGTACCGCGACCGCGCTGGCGACCCGTATCATTGATCAAAAC
CTGAAACGTTTCTGCGACAACCTGCTGATCTTTGAGAGCATTCGTGATAAGATCGACT
TCAGCGAGGTTGAACAGACCATGGGCAACAGCATCGATAAGGTGTTCAGCGTTATCT
TTTATAGCAGCTGCCTGCTGCAAGAAGGTATCGACTTTTACAACTGCGTGCTGGGTG
GTGAAACCCTGCCGAACGGTGAAAAGCGTCAGGGCATTAACGAACTGATCAACCTGT
ACCGTCAAAAGACCAGCGAGAAAGTTCCGTTCCTGAAGCTGCTGGACAAACAGATTC
TGAGCGAGAAGGAAAAATTTATGGATGAGATCGAAAACGACGAGGCGCTGCTGGAT
ACCCTGAAGATTTTCCGTAAAAGCGCGGAGGAAAAGACCACCCTGCTGAAAAACAT
CTTCGGCGATTTTGTGATGAACCAGGGTAAATATGACCTGGCGCAAATCTACATTAG
CCGTGAAAGCCTGAACACCATTAGCCGTAAGTGGACCAGCGAAACCGATATCTTCGA
AGACAGCCTGTACGAGGTGCTGAAAAAGAGCAAAATCGTGAGCGCGAGCGTTAAAA
AGAAAGACGGTGGCTACGCGTTCCCGGAGTTTATCGCGCTGATTTATGTTAAAAGCG
CGCTGGAACAGATTCCGACCGAGAAGTTCTGGAAAGAACGTTACTATAAGAACATCG
GCGATGTGCTGAACAAGGGTTTCCTGAACGGTAAAGAAGGCGTTTGGCTGCAATTTC
TGCTGATCTTTGACTTCGAATTTAACAGCCTGTTCGAGCGTGAAATCATTGATGAGAA
CGGCGACAAGAAAGTGGCGGGTTATAACCTGTTCGCGAAGGGTTTTGACGATCTGCT
GAACAACTTCAAATACGACCAGAAGGCGAAAGTGGTTATTAAGGATTTTGCGGACG
AAGTTCTGCACATTTATCAAATGGGCAAATACTTCGCGATCGAGAAGAAACGTAGCT
GGCTGGCGGACTATGATATTGACAGCTTCTACACCGATCCGGAGAAGGGTTACCTGA
AATTTTATGAAAACGCGTACGAGGAAATCATTCAGGTTTATAACAAGCTGCGTAACT
ACCTGACCAAGAAACCGTATAGCGAGGACAAGTGGAAACTGAACTTCGAAAACCCG

ACCCTGGCGGATGGTTGGGACAAGAACAAAGAGGCGGATAACAGCACCGTGATTCT
GAAGAAAGACGGTCGTTACTATCTGGGCCTGATGGCGCGTGGTCGTAACAAGCTGTT
CGACGATCGTAACCTGCCGAAAATCCTGGAGGGTGTTGAAAACGGCAAGTACGAAA
AGGTGGTTTACAAGTACTTCCCGGATCAGGCGAAGATGTTCCCGAAAGTGTGCTTTA
GCACCAAAGGCCTGGAATTCTTTCAACCGAGCGAGGAAGTTATCACCATTTACAAGA
ACAGCGAGTTCAAGAAAGGTTATACCTTTAACGTGCGTAGCATGCAGCGTCTGATTG
ATTTCTATAAAGACTGCCTGGTTCGTTACGAAGGTTGGCAATGCTATGATTTTCGTAA
CCTGCGTAAGACCGAGGACTACCGTAAAAACATCGAGGAATTCTTTAGCGATGTGGC
GATGGACGGCTACAAGATTAGCTTCCAGGACGTTAGCGAGAGCTATATCAAGGAGA
AGAACCAAAACGGTGATCTGTACCTGTTTGAGATCAAGAACAAAGACTGGAACGAA
GGTGCGAACGGCAAGAAAAACCTGCACACCATTTATTTCGAGAGCCTGTTTAGCGCG
GATAACATCGCGATGAACTTCCCGGTGAAACTGAACGGCCAGGCGGAGATCTTTTAC
CGTCCGCGTACCGAAGGTCTGGAGAAGGAACGTATCATTACCAAGAAAGGCAACGT
TCTGGAAAAGGGTGACAAAGCGTTCCACAAGCGTCGTTACACCGAGAACAAAGTGTT
CTTTCACGTTCCGATTACCCTGAACCGTACCAAGAAAAACCCGTTCCAATTTAACGCG
AAGATCAACGACTTCCTGGCGAAAAACAGCGATATCAACGTGATTGGTGTTGACCGT
GGCGAGAAACAGCTGGCGTATTTTAGCGTGATTAGCCAACGTGGCAAGATCCTGGAC
CGTGGTAGCCTGAACGTGATCAACGGCGTTAACTACGCGGAGAAGCTGGAGGAAAA
AGCGCGTGGTCGTGAACAGGCGCGTAAGGATTGGCAGCAAATCGAGGGCATTAAAG
ACCTGAAGAAAGGTTATATTAGCCAGGTGGTTCGTAAACTGGCGGATCTGGCGATCC
AATACAACGCGATCATTGTGTTCGAGGACCTGAACATGCGTTTTAAGCAAATTCGTG
GTGGCATCGAGAAAAGCGTTTATCAGCAACTGGAAAAGGCGCTGATCGATAAACTG
ACCTTCCTGGTGGAGAAGGAAGAAAAGGACGTTGAAAAGGCGGGTCACCTGCTGAA
AGCGTACCAGCTGGCGGCGCCGTTCGAAACCTTTCAGAAGATGGGTAAACAAACCG
GCATTGTGTTTTATACCCAAGCGGCGTACACCAGCCGTATCGATCCGGTTACCGGCTG
GCGTCCGCACCTGTACCTGAAATATAGCAGCGCGGAAAAGGCGAAAGCGGACCTGC
TGAAGTTCAAGAAAATTAAGTTCGTGGATGGTCGTTTCGAGTTTACCTACGACATCA
AGAGCTTCCGTGAGCAGAAGGAACACCCGAAAGCGACCGTGTGGACCGTTTGCAGC
TGCGTTGAGCGTTTTCGTTGGAACCGTTATCTGAACAGCAACAAAGGTGGCTACGAT
CACTATAGCGACGTGACCAAGTTCCTGGTTGAGCTGTTTCAGGAATACGGCATCGAC
TTCGAACGTGGTGATATTGTGGGCCAAATCGAGGTTCTGGAAACCAAGGGTAACGAG
AAGTTCTTTAAGAACTTCGTGTTCTTTTTCAACCTGATCTGCCAGATTCGTAACACCA
ACGCGAGCGAACTGGCGAAGAAAGACGGCAAGGACGATTTCATTCTGAGCCCGGTT
GAGCCGTTTTTCGATAGCCGTAACAGCGAGAAGTTCGGCGAAGACCTGCCGAAAAAC
GGTGACGATAACGGCGCGTTTAACATCGCGCGTAAAGGTCTGGTTATTATGGATAAG
ATCACCAAATTCGCGGACGAGAACGGTGGCTGCGAAAAGATGAAATGGGGTGACCT
GTATGTGAGCAATGTGGAGTGGGATAACTTTGTGGCGAATAAATAA **(SEQ ID NO: 17)**

Cas12q codon optimized

ATGGGGTCCTCCCATCATCACCACCACCACTCTTCAGGCTTGGTACCGCGTGGTTCCA
TGATCAACATAGACGAATTGAAAAATTTATATAAGGTGCAAAAGACCATCACTTTCG
AACTTAAGAACAAGTGGGAGAACAAAAATGATGAGAACGACAGAGTAGAGTTCTTG
AAGACTCAGGAGTGGGTCGAAAGCCTTTTCAAGGTCGATGAAGAGAACTTTGATGAG
AAAGAGTCTATCCCTAACTTGTTAGACTTCGGACAGAAGATTGCGTCCTTGTTTTACA
AGCTGAGCGAGGACATAGCGAACAACCAAATTGATACGCGGGTATTGAAAGTCTCG
AAATTCCTTTTAGAGGAAATTGATAGAAATCAATACCACGAGAAAAAAAACAAGCC
CACAAAGGTAAAAGAAATGAATCCCAACACAAACAAAAGTTATATAAAAGAATATA
AGCTGTCCGACCAAAACACACTGTACGTGTTATTAAAGATAATGGAAGATGAAGGTC

GGGGATTACAAAAATTTTTGTACGATAAAGCGGACCGGTTAAACCTGTACAATCAAA
AAGTTCGGAGAGACTTCGCCTTAAAGGAATCAAATGAGCAACAAAAATTCTCTGGAA
ATGCCAACTACTATGGGAATATAAAGCTGCTTATAGATAGCTTAGAAGATGCAGTCC
GGATCATTGGGTATTTCACTTTCGACGATCAAGCAGAAAACGCACAAATCAATGAAT
TTAAGTCCGTTAAACAGGAAATGAATAATAATGAAGCGTCTTACCAAGCACTGAAAG
ACTTCGCTATTGATAACGCAAAAAAAGAGATAGAATTGACGACGTTGAACCACCGG
GCGGTCAACAAGGATCCAAAAAAGATTCAAGAACAGATTGAGGAAGTCGAAAATTT
CGAAGAAGATATTAACCAGTTAAAGCATCAGATATCAGCCTTGAATGATAAGAAGTT
TGACGTGGTTAGCAGATTAAAGCACGCTCTTATAAAAATGTTACCAGAACTGAATCT
TTTGGATGCTGAGTCGGAACAGGGCCGTGAAGTCCAGCAGATATATCAAGACAAAA
AAAACGGGTTGGAGCTTGATGACTTTAAATTTAACCTTTTAAAACATCATCAATGGC
AAAAAACGATCTTCAAGTATATTAAGCTTGAGGGCTTAGTTCTGCCAGACCTTTACG
CGGAAAACAAACAAGATAAAATCAAGGTTTATATTGAGAATTATAGACAGAGTGGT
GAGCGTATTTCTAAGAAGGCGAGAGAGGAATTAGGAAAAATCGATAAACGCGAAGA
GTTCAATGGAAATGACGAACTTAAGAAGGCATGGTATGAGTATAAGGACTTCTGTAG
AGACAAACGTAATAAGAGCGTGGAACTTGGCAATAAGAAGTCGCTGTACAATGCCA
TAAAGCGCGAAGTTTTGCGGCAAAAAATGTGCAACCATTTCGCTGTGCTGGTGTCCG
ACGGTGAAGATACTTCCCCTTATTATTATCTGATATTAATCCCGAACGAGAACTCCGA
TGAAATGAATAGAACGTTCAAGGAATTGAAGGCCTCCGAGGGGAATTGGAAGATGT
TGGATTACAATCGTCTGACCTTCAAAGCCTTGGAGAAATTGGCCCTGTTACGGTCGTC
TACCTTCGAGATAGCGGATCAGGAACTGCAAGAAGAGGCAAAAAAGATCTGGGAGG
AGTACAAGGAAAAGGCGTACAAAGACTTCAAAAACAAAAAGTTATTACAGGGTTTA
TCGGGAAGACAGCGGGAGGAGAAAAAGCAAGAATTGCAAAAGGAGAGCCTGAATA
GAGTAATCAATTACTTGATCAGATGCATTCAGTCATTGCCCGACAGCGGAAAATACA
ACTTTAACTTTAAAGAGCCTCATCAATACCAATCGCTTGAAGAGTTTGCCGAGGAGA
TTGATCGGCAAGGTTATCACTGTGCTTGGAAAAACGTTTCTAAAGATAAACTGATGG
AATTGGAAGCGATGGAAAAGATTAAGGTTTTCAAACTTCATAACAAAGACTTTCGCA
AGGTAAAACTGAACGACTCCAAGCACAACCCTAATCTTTTTACTTTGTACTGGTTAGA
CGCCATGAATTTGGATAAGGTTAACGTCCGCCTGTTACCGGAAGTTGACCTTTACAA
GAGAGCTAAGGAAACACAGCTGAAATTGTTCGAACGTGATGTGAAATGCAATATCA
ATAACCAAAAGATTAAATCTATCAAGGAGAAGAATAGACTGTTTCAGGACAAGTTGT
ATGCTAGTTTTAAGTTAGAGTTTTATCCAGAAAACGAAGGATTAGGTTTCGAGCAGG
TAAATGACAAGGTCAATAACTTCTGCGGTAGCGATACGGCCTATTATCTTGGGCTTG
ATCGTGGAGAGAAAGAGCTTGTTACATTCTGCCTGGTGGACTCTGATGGCCGCCTGG
TAAAAAACGGAGACTGGACCAAGTTTAAAGAGGTGAACTATGCCGACAAACTGAAG
CAATTCTACTACTCAAAAGGCGAAATAGAGAGTACCCAACAACAGCTGTTAGAAGCC
CGGGACAATATTAAACAAGCGACCAACACGGAAGATAAGGAGTCCATGAAACTGAA
TTATAAGAAACTGGAACTGAAGTTAAAACAACAGAATTTGCTGGCGCAAGAATTCAT
AAAAAAAGCGTACTGCGGCTACCTTATCGATAGCATTAATGAGATTCTGAGAGAATA
TCCAAATACTTATCTTGTCTTAGAGGATTTGGATATCGCGGGTAAAGCGGATCCAGA
GTCGGGGATGACTAATAAAGAGCAGAACTTAAACAAAACGATGGGGGCTTCAGTAT
ACCAGGCCATTGAGAATGCGATCGTAAATAAATTCAAATATCGCACCGTGAAATTGT
CCGATATCAAGGGCCTTCAGACTGTACCTAATGTAGTGAAGGTCGAAGACTTACGGG
AAGTGAAAGAGGTTGAAGATGGGGAACACAAGTTCGGGTTAATAAGATCAGTTAAG
AGCAAGGATCAAATCGGTAACATACTTTTTGTCGACGAGGGGGAGACCAGTAACACT
TGTCCGAATTGCGGTTTTAATAGTGATTGGTTTAAACGCGATGTTGATTTTGACTTAG
AAATAGTCGCTACTGTAAACGGGCAAAAGAATGCCGTGATTGAGCAAAATGACAAA
AAATACTGTTTCCCGGGCGAAATATATAAATTGGAAATCATTAATAAAGAGTACGAA

ACAAACAAGCGTAATCTTGCCATGATTTTTAAACCTCGGGCCAAAGCGTGCCGTAAA TTTATCAATAATAATTTAGATAAGAACGATTATTTCTATTGTCCCTACTGCGCCTTCTC GTCGAAGAATTGTAACAACCCGAAACTGCAGAACGGCGATTTCGTGGTATATTCAGG AGACGATGTTGCTGCTTACAATGTTGCTATCAGAGGAATTAACCTGCTGAACAATAT TAAATAG **(SEQ ID NO: 18)**

[00140] Table 4a shows the structural and functional characteristics of the novel Cas12 proteins of the disclosure as exemplified herein. Table 4b shows the number and sequence of the natural spacers of the corresponding CRISPR arrays. Blank cells in the tables do not indicate that no value/property exists, but rather that it has not been exemplified herein.

Table 4a

	Cas12a.1	**Cas12p**	**Cas12q**
Structural			
Size aa	1254	1281	1137
Kda			
Identity protein seq to NCBI Protein DB Best hit	46.7	56.3	23.7
Nuclease domains	yes	yes	yes
Bridge Helix	yes	yes	yes
Cas gene cluster	yes	yes	yes
Cas1 (length in aa of Cas1 proteins encoded in each corresponding cluster)	325 aa	354 aa	334 aa
CRISPR array	806 bp	551 bp	1650 bp
Repeats (number and sequence of the natural repeats included in the corresponding CRISPR array)	13 (GTTTAAGGCC TTGACAAAAT TTCTACTGTAG TAGAT) **(SEQ ID NO: 19)**	9 (CTCGAATATC CCTATTAGATT TCTACTTTTGT AGAT) **(SEQ ID NO: 20)**	27 (ATCTACAAAA GTAGAAATTA AATAGGTCTA TTTGAG) **(SEQ ID NO: 21)**
Target	GTGGCAGCTC AAAAATTGGC TACAAAAC **(SEQ ID NO: 25)**	GTGGCAGCTCA AAAATTGGCTA CAAAAC **(SEQ ID NO: 25)**	
Collateral Cleavage shown in examples	yes	yes	

Table 4b

Cas12a.1	**Cas12p**	**Cas12q**
Repeats: 12	Repeats: 8	Repeats: 26
CCCGATTGACGCTATAG TAAGCATCGAG **(SEQ ID NO: 71)**	TTCAGATGTTTGCTCTTTG ACATATCG **(SEQ ID NO: 82)**	AATCGTAGCGATAACCGA AGAACAAAT **(SEQ ID NO: 89)**

Cas12a.1	Cas12p	Cas12q
GCGTCCCATAAAGGTATGACTTGTATT **(SEQ ID NO: 72)**	ATGGTGATTTAAAAACAAAACTCGGCGCGA **(SEQ ID NO: 83)**	AACCCATATGTTTTATTATCCTGCTGA **(SEQ ID NO: 90)**
ACGCACGCAGTATTGAATACGCGAATAG **(SEQ ID NO: 73)**	CCTTGTGCAAAATAGACAGGTTAGACCGT **(SEQ ID NO: 84)**	TACAAAATTAAGGCGGTCTAGGAGA **(SEQ ID NO: 91)**
GTTCACGTAAAACTTAATCGTTGAACT **(SEQ ID NO: 74)**	CGAAAATCCAGCTAAACTCATTCTCTGATT **(SEQ ID NO: 85)**	TCCATTTGATGATAACCATAAGAAT **(SEQ ID NO: 92)**
TGGTAGTGCCAACACGTGCGCCCACCA **(SEQ ID NO: 75)**	TTGATTGGAGGAACAAGCTACATAAA **(SEQ ID NO: 86)**	AAAATAATGTAATATAATACAATAT **(SEQ ID NO: 93)**
TCGGTGGTGGGCGAACATTGACTGTTGGT **(SEQ ID NO: 76)**	CGTATGGGTGTAATTTAATCGGTTTG **(SEQ ID NO: 87)**	CTCGACACTGGGACAACTTCCGTAT **(SEQ ID NO: 94)**
CCGATTCCTTCTCGTTCGCCCGTGACCA **(SEQ ID NO: 77)**	ATGAATCGTATAAGATATGATCTGAAT **(SEQ ID NO: 88)**	CAATAAATACTGATTAGAAGAAGATAT **(SEQ ID NO: 95)**
GTTGCGGGAGATACTACTTCAATATACA **(SEQ ID NO: 78)**	ATTAAATTACATAATGAGCCAACACGGCGACC **(SEQ ID NO: 23**	CTGTCAAAGCCATAGTCTTGATCCAGC **(SEQ ID NO: 96)**
GGTAGCCGAAATGAATTCGGTATAACCCG **(SEQ ID NO: 79)**		GCGCAAAGCATCAGCGCAATGGCTCG **(SEQ ID NO: 97)**
AACCAGTATCCTACCGTGAAGTTGTCGC **(SEQ ID NO: 80)**		TGGCAGAGTTCGGGCCAAGTATCAT **(SEQ ID NO: 98)**
GGGGGTTTGAGTGGGCAACGCAAGGAA **(SEQ ID NO: 81)**		GTAGCGTTCTGTTACGTGCCAGCGA **(SEQ ID NO: 99)**
CGTCGTATTGAGTGCTAGTACTGGTTTGAG **(SEQ ID NO: 22)**		CGGAATAATGTATGTCTTACCGAGGC **(SEQ ID NO: 100)**
		CTACGATTACCTTAACGACCCTAAC **(SEQ ID NO: 101)**
		ATGATTGACACAATAATTAACTGGTT **(SEQ ID NO: 102)**
		GCTTGGAAATATGTCTTATTTATCA **(SEQ ID NO: 12)**
		ATACCGTCTGTACCTATTGGGGGCA **(SEQ ID NO: 103)**
		AAAAGTGCTAAAATTCTTAACGGAA **(SEQ ID NO: 104)**

Cas12a.1	Cas12p	Cas12q
		CTTGCTGAATTCGGCTCA AGCATCAT **(SEQ ID NO: 105)**
		CAGCATGGGATAGAACGC TTCCGAGC **(SEQ ID NO: 106)**
		CCACTAGCATCTCCTAGG ATAGTTGGA **(SEQ ID NO: 107)**
		ATATAAGACAGCTCCAAG CTCCCGTT **(SEQ ID NO: 108)**
		TACCTCTGGAGTTTAATCT TTGATAGA **(SEQ ID NO: 109)**
		AATGAAAAACCAAAATCC GCACCTTA **(SEQ ID NO: 110)**
		GACGCATATTGCATAGCG GTTTATGC **(SEQ ID NO: 111)**
		ATAAATTCACAAACTAAC TTGTAAC **(SEQ ID NO: 112)**
		CTAGCTCCTCTACGTCTTT ATTTTCACCCTCAT **(SEQ ID NO: 24)**

[00141] In some embodiments, the Cas12 protein of the disclosure is a catalytically active Cas12 protein, e.g. a catalytically active Cas12a.1, Cas12p, or Cas12q protein.

[00142] In some embodiments, the Cas12 protein of the disclosure cleaves at a site distal to the target sequence, e.g. the Cas12a.1, Cas12p, or Cas12q protein cleaves at a site distal to the target sequence.

[00143] In some embodiments, the Cas12 protein of the disclosure is a catalytically dead Cas12 protein, e.g. the Cas12a.1, Cas12p, or Cas12q protein is a catalytically dead (dCas12a.1, dCas12p, or a dCas12q protein).

[00144] In some embodiments, the Cas12 protein of the disclosure is a nickase Cas12 protein, e.g. a Cas12a.1 nickase, a Cas12p nickase, or a Cas12q nickase protein.

[00145] In some embodiments, the Cas12 proteins of the disclosure can be modified to include an aptamer.

[00146] In some embodiments, theCas12 proteins of the disclosure can be further fused to domains, e.g. catalytic domains to produce dual action Cas proteins. In some embodiments, a Cas12a protein is further fused to a base editor.

b. Collateral activity of Class 2 Type V CRISPR-Cas RNA-Guided Proteins

[00147] In addition to the ability to cleave a target sequence in a targeted DNA, the Cas12 proteins of the disclosure also possess collateral (*trans*-cleavage activity), i.e. the ability to promiscuously cleave non-targeted single stranded DNA (ssDNA) or RNA once activated by detection of a target DNA. Without being bound to any theory or mechanism, generally once a Cas12 protein of the disclosure is activated by a gRNA, which occurs when a sample includes a target sequence to which the gRNA hybridizes (i.e., the sample includes the targeted DNA), the Cas12 can become a nuclease that promiscuously cleaves oligonucleotides (e.g. ssDNAs, RNAs, chimeric RNA/DNAs) not comprising the target sequence of the gRNA (non-target oligonucleotides, to which the guide sequence of the gRNA does not hybridize). Thus, when the targeted DNA (double or single stranded) is present in the sample (e.g., in some embodiments above a threshold amount), the result can be cleavage of single stranded oligonucleotides (e.g. ssDNAs, ssRNAs, single stranded chimeric RNA/DNAs) in the sample, which can be detected using any convenient detection method (e.g., using a labeled detector DNA, RNA, or DNA/RNA chimera).

[00148] Accordingly, provided herein are methods and compositions for detecting a target DNA (dsDNA or ssDNA) in a sample. Also provided are methods and compositions for cleaving non-target oligonucleotides, which can be utilized detectors. These embodiments are described in further detail below.

c. gRNAs for Class 2 Type V CRISPR-Cas RNA-Guided Proteins

[00149] The present disclosure provides DNA-targeting RNAs that direct the activities of the novel Cas12 proteins of the disclosure to a specific target sequence within a target DNA. As above for the novel Cas9 proteins of the disclosure, these DNA-targeting RNAs are referred to herein as “gRNAs” or “gRNAs” Generally, as provided herein, a Cas12’s gRNA comprises a single segment comprising both a spacer (DNA-targeting sequence) and a Cas12a "protein-binding sequence" together referred to as a crRNA. Also provided herein are nucleotide sequences encoding the Cas12a gRNAs of the disclosure.

i. Spacer Sequences

[00150] The Cas12 proteins of the disclosure are single crRNA-guided endonucleases (single guide RNA, sgRNA, while the Cas9 proteins of the disclosure are guided by a dual-RNA system consisting of a crRNA and a trans-activating crRNA (tracrRNA). The crRNA of the Cas12 guides of the disclosure comprises a nucleotide sequence that is complementary to a sequence in a target DNA (DNA-targeting sequence or spacer).

[00151] The crRNA portion of the Cas12 gRNAs of the disclosure can have a length of from about 25 – 50 nt. In some embodiments, the length can be about 40 – 43 nt.

[00152] The mature guide scaffolds for Cas12a.1 and Cas12p were deduced *in silico* from the corresponding CRISPR loci. FIG. 38 shows the secondary structure of the scaffolds for Cas12a.1 (5' aaauuucuacuguaguagau 3') (SEQ ID NO: 116; Panel A) and Cas12p (5' agauuucuacuuuuguagau3')(SEQ ID NO: 117; Panel B). These mature scaffolds can then be joined with variable targeting spacer sequences, giving rise to a sgRNA. Accordingly, in some embodiments, provided herein is an engineered single-molecule gRNA, comprising the scaffold sequence of SEQ ID NO: 116 or SEQ ID NO: 117 and a spacer sequence that is capable of hybridizing with a target sequence in a target DNA. In some embodiments, the target DNA is viral DNA, plant DNA, fungal DNA, or bacterial DNA. In some embodiments, the target sequence is a sequence of a target provided in any of Tables 6a-6f. In some embodiments, the target is a coronvavirus. In some embodiments, the target is a SARS-CoV-2 virus. In some embodiments, the target DNA is cDNA, and has been obtained by reverse transcription.

[00153] The DNA-targeting spacer sequence of a Cas12 gRNA generally interacts with a target DNA in a sequence-specific manner via hybridization (i.e., base pairing). As such, the nucleotide sequence of the DNA-targeting sequence may vary and determines the location within the target DNA that the gRNA and the target DNA will interact. The DNA-targeting sequence of a subject Cas12 gRNA can be modified (e.g., by genetic engineering) to hybridize to a desired sequence within a target DNA.

[00154] The DNA-targeting sequence of a subject Cas12 gRNA can have a length of from about 8 nucleotides to about 30 nucleotides. For example, the length can be 23 nucleotides.

[00155] The percent complementarity between the DNA-targeting spacer sequence of the crRNA and the target sequence of the target DNA can be at least 60% (e.g., at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, at least 99%, or 100%). In some embodiments, the percent complementarity between the

DNA-targeting sequence of the crRNA -RNA and the target sequence of the target DNA is 100% over the 1 – 23 contiguous 5 '-most nucleotides of the target sequence of the complementary strand of the target DNA. In some embodiments, the percent complementarity between the DNA-targeting sequence of the crRNA and the target sequence of the target DNA is at least 60% over about 1 - 23 contiguous nucleotides. In some embodiments, the percent complementarity between the DNA-targeting sequence of the crRNA and the target sequence of the target DNA is 100% over the 1 – 23 contiguous 5 '-most nucleotides of the target sequence of the complementary strand of the target DNA and as low as 0% over the remainder. In such a case, the DNA-targeting sequence can be considered to be 1 – 23 nucleotides in length.

[00156] Generally, a naturally unprocessed pre-crRNA of Cas12 comprises a direct repeat and an adjacent spacer (the portion of the crRNA that allows for targeting to a DNA molecule). In some embodiments, direct repeats, and direct repeat mutations from unprocessed pre-crRNA are included into the Cas12 gRNAs of the disclosure, and improve gRNA stability.

[00157] Table 5a shows the predicted (putative) naturally occurring direct repeat sequences in the CRISPR locus, as found in bacterial DNA, of the Cas12 proteins of the disclosure. These are the predicted natural sequences in the CRISPR locus contig, as found in bacterial DNA. The gRNAs of the disclosure have a part of the direct repeat joined to the spacer.

Table 5a: Direct repeat sequences

Description	**Name**	**Sequence**
Direct Repeat	Cas12a.1	GTTTAAGGCCTTGACAAAATTTCTACTGTAGTAGAT (SEQ ID NO: 6)
Direct Repeat	Cas12p	ATCTACAAAAGTAGAAATCTAATAGGGATATTCGAG (SEQ ID NO: 7)
Direct Repeat	Cas12p	CTCGAATATCCCTATTAGATTTCTACTTTTGTAGAT (SEQ ID NO: 26)
Direct Repeat	Cas12q	ATCTACAAAAGTAGAAATTAAATAGGTCTATTTGAG (SEQ ID NO: 8) CTCAAATAGACCTATTTAATTTCTACTTTTGTAGAT (SEQ ID NO: 27)

[00158] In some embodiments, the crRNAs include non-naturally occurring, engineered direct repeat sequences. Table 5b shows non-naturally occurring, engineered direct repeat sequences which can be incorporated into the engineered gRNAs of the disclosure.

[00159] The predicted RNA secondary structures of non-naturally occurring, engineered direct repeat sequences are shown in FIGS. 7A-7C.

Table 5b

Description	Name	Sequence of non-naturally ocurring direct repeat sequences
Direct Repeat	Cas12a.1	**A)** GTTTAAGGCCTTGACAAAATTTCCACTGTAGTGGAT (SEQ ID NO: 28) **B)**GGTTTAAGGCCTTGACAAAATTTCTCCTGTAGGAGAT (SEQ ID NO: 29) **C)**GTTTAAGGCCTTGACAAAATTTCCCCTGTAGGGGAT (SEQ ID NO: 30)
Direct Repeat	Cas12p	**A)** ATCTACAAAAGTAGAAGTCTAATAGGGACATTCGAG (SEQ ID NO: 31) **B)** ATCTACAAAAGTAGAAAGCTAATAGGGCTATTCGAG (SEQ ID NO: 32) **C)**ATCTACAAAAGTAGAAGGCTAATAGGGCCATTCGAG (SEQ ID NO: 33) **D)** CTCGAATATCCCTATTAGATTTCGACTTTTGTCGAT (SEQ ID NO: 34) **E)** CTCGAATATCCCTATTAGATTTCTCCTTTTGGAGAT (SEQ ID NO: 35) **F)** CTCGAATATCCCTATTAGATTTCGGCTTTTGCCGAT (SEQ ID NO: 36)
Direct Repeat	Cas12q	**A)** ATCTACAAAAGTAGAAATTGAATAGGTCTATTCGAG (SEQ ID NO: 37) **B)** ATCTACAAAAGTAGAAATTAAAGAGGTCTCTTTGAG (SEQ ID NO: 38) **C)**ATCTACAAAAGTAGAAATTGGGTAGGTCTACCCGAG (SEQ ID NO: 39) **D)** CTCAAATAGACCTATTTAATTTCCACTTTTGTGGAT (SEQ ID NO: 40) **E)** CTCAAATAGACCTATTTAATTTCTCCTTTTGGAGAT (SEQ ID NO: 41) **F)**CTCAAATAGACCTATTTAATTTCCCCTTTTGGGGAT (SEQ ID NO: 42)

[00160] In some embodiments the spacer sequence of a Cas12 gRNA of the disclosure is directed to a target sequence in a mammalian organism. In some embodiments the spacer sequence is directed to a target sequence in a non-mammalian organism.

[00161] In some embodiments, the spacer sequence of a Cas12 gRNA of the disclosure is directed to a target sequence which is a sequence of a human. In some embodiments, the target sequence is a sequence of a non-human primate.

[00162] In some embodiments, the spacer sequence of a Cas12 gRNA of the disclosure is directed to a target sequence in a mammalian organism, e.g. a human or non-human primate.

[00163] In some embodiments, the spacer sequence of a Cas12 gRNA of the disclosure is directed to a target sequence in a bacteria.

[00164] In some embodiments, the spacer sequence of a Cas12 gRNA of the disclosure is directed to a target sequence in a virus.

[00165] In some embodiments, the spacer sequence of a Cas12 gRNA of the disclosure is directed to a target sequence in a plant.

[00166] The Cas12 gRNAs of the disclosure can be modified to include an aptamer.

ii. PAM Specificities

[00167] TCTN and TGTN are identified to be efficient PAM sequences for Cas12a.1 and Cas12p, respectively.

iii. gRNA Arrays

[00168] In some embodiments, the Cas12 gRNAs of the disclosure can be provided as gRNA arrays.

[00169] Such gRNA arrays of the disclosure include more than one gRNA arrayed in tandem, and can be processed into into two or more individual gRNAs. Thus, in some embodiments a precursor Cas12 gRNA array comprises two or more (e.g., 3 or more, 4 or more, 5 or more, 2, 3, 4, or 5) gRNAs (e.g., arrayed in tandem as precursor molecules). In some embodiments, two or more gRNAs can be present on an array (a precursor gRNA array). A Cas12 protein of the disclosure can cleave the precursor gRNA array into individual gRNAs.

[00170] In some embodiments a Cas12 gRNA array includes 2 or more gRNAs (e.g., 3 or more, 4 or more, 5 or more, 6 or more, or 7 or more, gRNAs). The gRNAs of a given array can target (i.e., can include guide sequences that hybridize to) different target sites of the same target DNA. In some embodiments, two or more gRNAs of a precursor gRNA array have the same guide sequence. In some embodiments, the precursor gRNA array comprises two or more gRNAs that target different target sites within the same target DNA. In some embodiments, the precursor gRNA array comprises two or more gRNAs that target different target DNAs.

III. Methods of Use – Modification and Therapeutics

a. Modification of Target DNA

[00171] Provided herein are uses of the novel Cas9 and Cas12 proteins of the disclosure. Accordingly, provided herein is a method of modifying a target DNA, the method comprising contacting the target DNA with any one Cas9 systems or Cas12 systems described herein. Such methods are useful for therapeutic application

[00172] In some embodiments, the target DNA is part of a chromosome *in vitro*. In some embodiments, the target DNA is part of a chromosome *in vivo*.

[00173] In some embodiments, the target DNA is part of a chromosome in a cell.

[00174] In some embodiments, the target DNA is extrachromosomal DNA.

[00175] In some embodiments, the target DNA is in a cell, wherein 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, in invertebrate cell, a vertebrate cell, a fish 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.

[00176] In some embodiments, the target DNA is the DNA of a parasite.

[00177] In some embodiments, the target DNA is a viral DNA.

[00178] In some embodiments, the target DNA is a bacterial DNA.

[00179] In some embodiments, the modifying comprises introducing a double strand break in the target DNA.

[00180] In some embodiments, the contacting occurs under conditions that are permissive for non-homologous end joining or homology-directed repair.

[00181] In some embodiments, the method comprises contacting the target DNA with a donor polynucleotide, wherein the donor polynucleotide, a portion of the donor polynucleotide, a copy of the donor polynucleotide, or a portion of a copy of the donor polynucleotide integrates into the target DNA.

[00182] In some embodiments, the method does not comprise contacting the cell with a donor polynucleotide, wherein the target DNA is modified such that nucleotides within the target DNA are deleted.

b. Therapeutic Applications

[00183] The disclosure provides novel Cas9 proteins, novel Cas12a proteins, and novel Cas12 proitein subtypes, engineered systems, one or more polynucleotides encoding components of said system, and vector or delivery systems comprising one or more polynucleotides encoding components of said system for use in therapeutic methods. The therapeutic methods may comprise gene or genome editing, or gene therapy. The therapeutic methods comprise use and delivery of the novel Cas9 and Cas12 proteins of the disclosure. Accordingly, in some embodiments, provided herein is a method of modifying a target DNA, the method comprising contacting a target DNA, a cell comprising the target DNA, or a subject with cells with the target DNA, with any one Cas9 systems or Cas12 systems described herein.

[00184] In some embodiments, the target DNA is part of a chromosome *in vitro*. In some embodiments, the target DNA is part of a chromosome *in vivo*.

[00185] In some embodiments, the target DNA is part of a chromosome in a cell.

[00186] In some embodiments, the target DNA is extrachromosomal DNA.

[00187] In some embodiments, the target DNA is in a cell, wherein 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, in invertebrate cell, a vertebrate cell, a fish 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.

[00188] In some embodiments, the target DNA is outside of a cell.

[00189] In some embodiments, the target DNA is *in vitro* inside of a cell.

[00190] In some embodiments, the target DNA is *in vivo,* inside of a cell.

[00191] In some embodiments, the modifying comprises introducing a double strand break in the target DNA.

[00192] In some embodiments, the contacting occurs under conditions that are permissive for non-homologous end joining or homology-directed repair.

[00193] In some embodiments, the method comprises contacting the target DNA with a donor polynucleotide, wherein the donor polynucleotide, a portion of the donor polynucleotide, a copy of the donor polynucleotide, or a portion of a copy of the donor polynucleotide integrates into the target DNA.

[00194] In some embodiments, the method does not comprise contacting the cell with a donor polynucleotide, wherein the target DNA is modified such that nucleotides within the target DNA are deleted.

[00195] In some embodiments, the therapeutic methods involve modifying a target DNA comprising a target sequence of a gene of interest and/or the regulatory region of the gene of interest, the method comprising delivering to a cell comprising the target DNA, a Cas9 protein of the disclosure and one or more Cas9 gRNAs, a Cas12 protein of the disclosure and one or more Cas12 gRNAs, one or more nucleotides encoding the Cas9 protein of the disclosure and one or more Cas9 gRNAs, or one or more nucleotides encoding a Cas12 protein of the disclosure and one or more Cas12 gRNAs.

[00196] In some embodiments, the gene of interest is within a eukaryotic cell, e.g. a human or non-human primate cell.

[00197] In some embodiments, the gene of interest is within a plant cell.

[00198] In some embodiments, the delivering comprises delivering to the cell a Cas9 protein of the disclosure (or one or more nucleotides encoding the same) and one or more Cas9 gRNAs.

[00199] In some embodiments, the delivering comprises delivering to the cell a Cas12 protein of the disclosure (or one or more nucleotides encoding the same) and one or more Cas12 gRNAs.

[00200] In some embodiments, the delivering comprises delivering to the cell one or more nucleotides encoding the Cas9 protein of the disclosure and one or more Cas9 gRNAs.

[00201] In some embodiments, the delivering comprises delivering to the cell one or more nucleotides encoding a Cas12 protein of the disclosure and one or more Cas12 gRNAs.

[00202] Delivery of the Cas9 or Cas12 components to a cell can be achieved by any variety of delivery methods known to those of skill in the art. As a non-limiting example, the components can be combined with a lipid. As another non-limiting example, the components combined with a particle, or formulated into a particle, e.g. a nanoparticle.

[00203] Methods of introducing a nucleic acid and/or protein into a host cell are known in the art, and any convenient method can be used to introduce a subject nucleic acid (e.g., an expression construct/vector) into a target cell (e.g., prokaryotic cell, eukaryotic cell, plant cell, animal cell, mammalian cell, human cell, and the like). Suitable methods include, e.g., viral infection, transfection, conjugation, protoplast fusion, lipofection, electroporation, calcium phosphate precipitation, polyethyleneimine (PEI)-mediated transfection, DEAE-dextran

mediated transfection, liposome-mediated transfection, particle gun technology, calcium phosphate precipitation, direct micro injection, nanoparticle-mediated nucleic acid delivery and the like.

[00204] A gRNA can be introduced, e.g., as a DNA molecule encoding the gRNA, or can be provided directly as an RNA molecule (or a chimeric/hybrid molecule when applicable).

[00205] In some embodiments, a Cas9 or Cas12 protein is provided as a nucleic acid (e.g., an mRNA, a DNA, a plasmid, an expression vector, a viral vector, etc.) that encodes the protein.

[00206] In some embodiments, the Cas9 or Cas12 protein is provided directly as a protein (e.g., without an associated gRNA or with an associate gRNA, i.e., as a ribonucleoprotein complex—RNP). Like a gRNA, a Cas9 or Cas12 protein of the disclosure can be introduced into a cell (provided to the cell) by any convenient method; such methods are known to those of ordinary skill in the art. As an illustrative example, a Cas9 or Cas12 protein of the disclosure can be injected directly into a cell (e.g., with or without a gRNA or nucleic acid encoding a gRNA). As another example, a pre-formed complex of a Cas9 or Cas12 protein and a gRNA can be introduced into a cell (e.g., eukaryotic cell) (e.g., via injection, via nucleofection; via a protein transduction domain (PTD) conjugated to one or more components, e.g., conjugated to the Cas9 or Cas12 protein of the disclosure, conjugated to a gRNA; etc.).

[00207] In some embodiments, a nucleic acid (e.g., a gRNA; a nucleic acid comprising a nucleotide sequence encoding a Cas9 or Cas12 protein of the disclosure; etc.) and/or a polypeptide (e.g., a Cas9 or Cas12 protein of the disclosure) is delivered to a cell (e.g., a target host cell) in a particle, or associated with a particle. In some embodiments, the particle is a nanoparticle.

[00208] A Cas9 or Cas12 protein of the disclosure (or an mRNA comprising a nucleotide sequence encoding the protein) and/or gRNA (or a nucleic acid such as one or more expression vectors encoding the gRNA) may be delivered simultaneously using particles or lipid envelopes.

i. Target Cells of Interest

[00209] Suitable target cells (which can comprise target DNA such as genomic DNA) include, but are not limited to: a bacterial cell; an archaeal cell; a cell of a single-cell eukaryotic organism; a plant cell; an algal cell, e.g., Botryococcus braunii, Chlamydomonas reinhardtii, Nannochloropsis gaditana, Chlorella pyrenoidosa, Sargassum patens, C. agardh, and the like; a fungal cell (e.g., a yeast cell); an animal cell; a cell from an invertebrate animal (e.g. fruit fly, a cnidarian, an echinoderm, a nematode, etc.); a cell of an insect (e.g., a mosquito; a bee; an

agricultural pest; etc.); a cell of an arachnid (e.g., a spider; a tick; etc.); a cell from a vertebrate animal (e.g., a fish, an amphibian, a reptile, a bird, a mammal); a cell from a mammal (e.g., a cell from a rodent; a cell from a human; a cell of a non-human mammal; a cell of a rodent (e.g., a mouse, a rat); a cell of a lagomorph (e.g., a rabbit); a cell of an ungulate (e.g., a cow, a horse, a camel, a llama, a vicuna, a sheep, a goat, etc.); a cell of a marine mammal (e.g., a whale, a seal, an elephant seal, a dolphin, a sea lion; etc.) and the like.

[00210] Any type of cell may be of interest (e.g. a stem cell, e.g. an embryonic stem (ES) cell, an induced pluripotent stem cell (iPSC), a germ cell (e.g., an oocyte, a sperm, an oogonia, a spermatogonia, etc.), an adult stem cell, a somatic cell, e.g. a fibroblast, a hematopoietic cell, a neuron, a muscle cell, a bone cell, a hepatocyte, a pancreatic cell; an in vitro or in vivo embryonic cell of an embryo at any stage, e.g., a 1-cell, 2-cell, 4-cell, 8-cell, etc. stage zebrafish embryo; etc.).

[00211] Cells may be from cell lines or primary cells. Target cells can be unicellular organisms and/or can be grown in culture. If the cells are primary cells, they may be harvest from an individual by any convenient method. For example, leukocytes may be conveniently harvested by apheresis, leukocytapheresis, density gradient separation, etc., while cells from tissues such as skin, muscle, bone marrow, spleen, liver, pancreas, lung, intestine, stomach, etc. can be conveniently harvested by biopsy.

[00212] Because the gRNA provides specificity by hybridizing to target nucleic acid, a mitotic and/or post-mitotic cell of interest in the disclosed methods may include a cell of any organism (e.g. a bacterial cell, an archaeal cell, a cell of a single-cell eukaryotic organism, a plant cell, an algal cell, e.g., Botryococcus braunii, Chlamydomonas reinhardtii, Nannochloropsis gaditana, Chlorella pyrenoidosa, Sargassum patens, C. agardh, and the like, a fungal cell (e.g., a yeast cell), an animal cell, a cell of an invertebrate animal (e.g. fruit fly, cnidarian, echinoderm, nematode, etc.), a cell of a vertebrate animal (e.g., fish, amphibian, reptile, bird, mammal), a cell of a mammal, a cell of a rodent, a cell of a human, etc.).

[00213] Plant cells include cells of a monocotyledon, and cells of a dicotyledon. The cells can be root cells, leaf cells, cells of the xylem, cells of the phloem, cells of the cambium, apical meristem cells, parenchyma cells, collenchyma cells, sclerenchyma cells, and the like. Plant cells include cells of agricultural crops such as wheat, corn, rice, sorghum, millet, soybean, etc. Plant cells include cells of agricultural fruit and nut plants, e.g., plant that produce apricots, oranges, lemons, apples, plums, pears, almonds, etc.

[00214] Non-limiting examples of cells (target cells) include: a prokaryotic cell, eukaryotic cell, a bacterial cell, an archaeal cell, a cell of a single-cell eukaryotic organism, a protozoa cell, a cell from a plant (e.g., cells from plant crops, fruits, vegetables, grains, soy bean, corn, maize, wheat, seeds, tomatos, rice, cassava, sugarcane, pumpkin, hay, potatos, cotton, cannabis, tobacco, flowering plants, conifers, gymnosperms, angiosperms, ferns, clubmosses, hornworts, liverworts, mosses, dicotyledons, monocotyledons, etc.), an algal cell, (e.g., Botryococcus braunii, Chlamydomonas reinhardtii, Nannochloropsis gaditana, Chlorella pyrenoidosa, Sargassum patens, C. agardh, and the like), seaweeds (e.g. kelp) a fungal cell (e.g., a yeast cell, a cell from a mushroom), an animal cell, a cell from an invertebrate animal (e.g., fruit fly, cnidarian, echinoderm, nematode, etc.), a cell from a vertebrate animal (e.g., fish, amphibian, reptile, bird, mammal), a cell from a mammal (e.g., an ungulate (e.g., a pig, a cow, a goat, a sheep); a rodent (e.g., a rat, a mouse); a non-human primate; a human; a feline (e.g., a cat); a canine (e.g., a dog); etc.), and the like. In some embodiments, the cell is a cell that does not originate from a natural organism (e.g., the cell can be a synthetically made cell; also referred to as an artificial cell).

[00215] A cell can be an in vitro cell (e.g., established cultured cell line). A cell can be an ex vivo cell (cultured cell from an individual). A cell can be and in vivo cell (e.g., a cell in an individual). A cell can be an isolated cell. A cell can be a cell inside of an organism. A cell can be an organism.

[00216] Suitable cells include human embryonic stem cells, fetal cardiomyocytes, myofibroblasts, mesenchymal stem cells, autotransplated expanded cardiomyocytes, adipocytes, totipotent cells, pluripotent cells, blood stem cells, myoblasts, adult stem cells, bone marrow cells, mesenchymal cells, embryonic stem cells, parenchymal cells, epithelial cells, endothelial cells, mesothelial cells, fibroblasts, osteoblasts, chondrocytes, exogenous cells, endogenous cells, stem cells, hematopoietic stem cells, bone-marrow derived progenitor cells, myocardial cells, skeletal cells, fetal cells, undifferentiated cells, multi-potent progenitor cells, unipotent progenitor cells, monocytes, cardiac myoblasts, skeletal myoblasts, macrophages, capillary endothelial cells, xenogenic cells, allogenic cells, and post-natal stem cells.

[00217] In some embodiments, the cell is an immune cell, a neuron, an epithelial cell, and endothelial cell, or a stem cell. In some embodiments, the immune cell is a T cell, a B cell, a monocyte, a natural killer cell, a dendritic cell, or a macrophage. In some embodiments, the

immune cell is a cytotoxic T cell. In some embodiments, the immune cell is a helper T cell. In some embodiments, the immune cell is a regulatory T cell (Treg).

[00218] In some embodiments, the cell is a stem cell. Stem cells include adult stem cells. Adult stem cells are also referred to as somatic stem cells.

[00219] Adult stem cells are resident in differentiated tissue, but retain the properties of self-renewal and ability to give rise to multiple cell types, usually cell types typical of the tissue in which the stem cells are found. Numerous examples of somatic stem cells are known to those of skill in the art, including muscle stem cells; hematopoietic stem cells; epithelial stem cells; neural stem cells; mesenchymal stem cells; mammary stem cells; intestinal stem cells; mesodermal stem cells; endothelial stem cells; olfactory stem cells; neural crest stem cells; and the like.

[00220] Stem cells of interest include mammalian stem cells, where the term "mammalian" refers to any animal classified as a mammal, including humans; non-human primates; domestic and farm animals; and zoo, laboratory, sports, or pet animals, such as dogs, horses, cats, cows, mice, rats, rabbits, etc. In some embodiments, the stem cell is a human stem cell. In some embodiments, the stem cell is a rodent (e.g., a mouse; a rat) stem cell. In some embodiments, the stem cell is a non-human primate stem cell.

ii. Targets

[00221] Any gene of interest can serve as a target for modification.

[00222] In particular embodiments, the target is a gene implicated in cancer. In particular embodiments, the target is a gene implicated in an immune disease, e.g. an autoimmune disease. In particular embodiments, the target is a gene implicated in a neurodegenerative disease. In particular embodiments, the target is a gene implicated in a neuropsychiatric diseaese. In particular embodiments, the target is a gene implicated in a muscular disease. In particular embodiments, the target is a gene implicated in a cardiac disease. In particular embodiments, the target is a gene implicated in diabetes. In particular embodiments, the target is a gene implicated in kidney disease.

iii. Precursor gRNA Arrays

[00223] The therapeutic methods provided herein can include delivery of precursor gRNA arrays. A Cas9 or Cas12 protein of the disclosure can cleave a precursor gRNA into a mature gRNA, e.g., by endoribonucleolytic cleavage of the precursor. A Cas9 or Cas12 protein of the

disclosure can cleave a precursor gRNA array (that includes more than one gRNA arrayed in tandem) into two or more individual gRNAs.

IV. Methods of Use – Detection and Diagnostic Applications

[00224] In addition to the ability to cleave a target sequence in a targeted DNA, the Cas12 proteins of the disclosure also possess collateral (trans-cleavage activity), i.e. the ability to promiscuously cleave non-targeted oligonucleotides (ssDNA, RNA, DNA/RNA hybrids) once activated by detection of a target DNA. Without being bound to any theory or mechanism, generally once a Cas12 protein of the disclosure is activated by a gRNA, which occurs when a sample includes a target sequence to which the gRNA hybridizes (i.e., the sample includes the targeted DNA), the Cas12 becomes a nuclease that promiscuously cleaves single stranded oligonucleotides (i.e., non-target single stranded oligonucleotides, i.e., single stranded oligonucleotides to which the guide sequence of the gRNA does not hybridize). Thus, when the targeted DNA (double or single stranded) is present in the sample (e.g., in some embodiments above a threshold amount), the result can be cleavage (collateral) of oligonucleotidesin the sample, which can be detected using any convenient detection method (e.g., using a labeled single stranded detector DNA, labeled detector RNA, or labeled detector DNA/RNA chimeric oligonucleotides).

[00225] Accordingly, provided herein are methods and compositions for detecting a target DNA (dsDNA or ssDNA) in a sample. Also provided are methods and compositions for cleaving non-target oligonucleotides (e.g. used as detectors).

[00226] As used herein a "detector" comprises a oligonucleotide of any nature, single or double stranded and does not hybridize with the guide sequence of the gRNA (i.e., the detector oligonucleotide that is a non-target). Exemplary detectors include, but are not limited to ssDNA, dsDNA, ssRNA, ss DNA/RNA chimeras, dsRNA, RNA comprising ss and ds regions, and ss or ds oligonucleotides containing RNA and DNA nucleotides (as used herein ss=single stranded; and ds= double stranded).

[00227] The detection methods based on the collateral activity of the Cas12 proteins of the disclosure can include:

(a) contacting the sample with: (i) a Cas12 protein of the disclosure; (ii) a gRNA comprising: a region that binds to the Cas12 protein, and a guide sequence that hybridizes with

the target DNA; and (iii) a detector that does not hybridize with the guide sequence of the gRNA; and

(b) measuring a detectable signal produced by cleavage of the detector by the Cas12 protein, thereby detecting the target DNA.

[00228] Once a subject Cas12 protein is activated by a gRNA, which can occur when the sample includes a target DNA to which the gRNA hybridizes (i.e., the sample includes the targeted sequence in the target DNA), the Cas12 can be activated to function as an endoribonuclease that non-specifically cleaves detector oligonucleotides (including non-target ss oligonucleotides) present in the sample. Thus, when the target DNA is present in the sample, the result is cleavage of a detector oligonucleotide in the sample, which can be detected using any convenient detection method (e.g., using a labeled detector oligonucleotides).

[00229] Also provided are methods and compositions for cleaving detector oligonucleotides(e.g., ssDNAs, ssRNAs, ssDNA/RNA chimeras or detectors comprising ss and ds regions). Such methods can include contacting a population of nucleic acids, wherein said population comprises a target DNA and a plurality of non-target ss oligonucleotides, with: (i) a Cas12protein of the disclosure; and (ii) a gRNA comprising: a region that binds to the Cas12 effector protein, and a guide sequence that hybridizes with the target DNA, wherein the Cas12 protein cleaves non-target ss oligonucleotides

[00230] Accordingly, provided herein is a method of detecting a target DNA in a sample, the method comprising:

(a) contacting the sample with:

(i) a Cas12 protein of the disclosure (e.g. Cas12a.1, Cas12p, or Cas12q protein);

(ii) a gRNA comprising a spacer sequence that is capable of hybridizing with a target sequence in a target DNA; and

(iii) a labeled detector oligonucleotide that does not hybridize with the spacer sequence of the gRNA; and

(b) measuring a detectable signal produced by cleavage of the labeled detector oligonucleotide by the Cas12 protein, thereby detecting the target oligonucleotide.

[00231] In some embodiments, the method further comprises the above along with detecting a positive control target DNA in a positive control sample, the detecting comprising the additional steps of:

(c) contacting the positive control sample with:

(i) a Cas12 protein of the disclosure(e.g. Cas12a.1, Cas12p, or Cas12q protein);

(ii) a positive control gRNA comprising: a region that binds to the Cas12a.1, Cas12p, or Cas12q protein, and a positive control spacer sequence that hybridizes with the positive control target DNA; and

(iii) a labeled detector oligonucleotide that does not hybridize with the positive control spacer sequence of the positive control gRNA; and

(d) measuring a detectable signal produced by cleavage of the labeled detector by the Cas12 protein, thereby detecting the positive control target DNA.

[00232] In some embodiments, the contacting step can be carried out in an acellular environment, e.g., outside of a cell. In other embodiments, contacting step can be carried out inside a cell. The contacting step can be carried out in a cell *in vitro*. The contacting step can be carried out in a cell *in vivo*. The contacting step of a detection method can be carried out in a composition comprising divalent metal ions.

[00233] The gRNA can be provided as RNA or as a nucleic acid encoding the gRNA (e.g., a DNA such as a recombinant expression vector), described herein.

[00234] The contacting, prior to the measuring step, can last for any period of time, e.g from 5 seconds to 2 hours or more, prior to the measuring step. In some embodiments the sample is contacted for 45 minutes or less prior to the measuring step. In some embodiments the sample is contacted for 30 minutes or less prior to the measuring step. In some embodiments the sample is contacted for 10 minutes or less prior to the measuring step. In some embodiments the sample is contacted for 5 minutes or less prior to the measuring step. In some embodiments the sample is contacted for 1 minute or less prior to the measuring step. In some embodiments the sample is contacted for from 50 seconds to 60 seconds prior to the measuring step. In some embodiments the sample is contacted for from 40 seconds to 50 seconds prior to the measuring step. In some embodiments the sample is contacted for from 30 seconds to 40 seconds prior to the measuring step. In some embodiments the sample is contacted for from 20 seconds to 30 seconds prior to the measuring step. In some embodiments the sample is contacted for from 10 seconds to 20 seconds prior to the measuring step.

[00235] The detection methods provided herein can detect a target DNA with a high degree of sensitivity. Accordingly, in some embodiments, the detection methods of the disclosure can be used to detect a target DNA present in a sample comprising a plurality of DNAs (including the

target DNA and a plurality of non-target DNAs), where the target DNA is present at one or more copies per 5 to 10^9 copies of the non-target DNAs

[00236] In some embodiments, the threshold of detection, for a detection method of detecting a target DNA in a sample, is 10 nM or less. The term "threshold of detection" is used herein to describe the minimal amount of target DNA that must be present in a sample in order for detection to occur. Thus, as an illustrative example, when a threshold of detection is 10 nM, then a signal can be detected when a target DNA is present in the sample at a concentration of 10 nM or more. In some embodiments, a subject composition or method exhibits an attomolar (aM) sensitivity of detection. In some embodiments, a subject composition or method exhibits a femtomolar (fM) sensitivity of detection. In some embodiments, a subject composition or method exhibits a picomolar (pM) sensitivity of detection. In some embodiments, a subject composition or method exhibits a nanomolar (nM) sensitivity of detection.

a. Target DNA

[00237] A target DNA can be single stranded (ssDNA) or double stranded (dsDNA). There need not be any preference or requirement for a PAM sequence in a single stranded target DNA.

[00238] The source of the target DNA can be any source. In some embodiments the target DNA is a viral or bacterial DNA (e.g., a genomic DNA of a DNA virus or bacteria). As such, detection method can be for detecting the presence of a viral or bacterial DNA amongst a population of nucleic acids (e.g., in a sample). In the case of a RNA-carrying organism, for example, a RNA virus (e.g. a coronavirus) – it is understood that a step such as reverse transcription may be carried out on a sample comprising the RNA-carrying organism to generated cDNA, and the cDNA is then the target DNA, for the purposes of this disclosure.

[00239] Exemplary non-limiting sources for target DNA are provided in Tables 6a – 6f.

Table 6a

Bacterial Resistance Gene Targets
KPC: carbapenem-hydrolyzing class A beta-lactamase
NDM: metallo-beta-lactamase
OXA: oxacillin-hydrolyzing class D beta-lactamase
MecA: PBP2a family beta-lactam-resistant peptidoglycan transpeptidase
vanA/B: Vancomycin resistance

Table 6b

Virus Genome Targets
Dengue (DENV) fever virus (subtypes 1,2, 3 and 4)
Zika Virus
Chikungunya virus
Coronoavirus

Respiratory Targets

[00240] DNA obtained from viruses and bacteria related to respiratory infections may also be targeted. A list of targets of interest may include the examples shown in Table 6c.

Table 6c

Respiratory Targets
Adenovirus
Coronoavirus
SARS-CoV
SARS-CoV-2
MERS-CoV
Coronavirus HKU1
Coronavirus NL63
Coronavirus 229E
Coronavirus OC43
Coronovirus HKU1
Human Metapneumovirus
Human Rhinovirus/Enterovirus
Influenza A
Influenza A/H1
Influenza A/H3
Influenza A/H1-2009
Influenza B
Parainfluenza Virus 1
Parainfluenza Virus 2

Respiratory Targets
Parainfluenza Virus 3
Parainfluenza Virus 4
Respiratory Syncytial Virus
BACTERIA:
Bordetella parapertussis
Bordetella pertussis
Chlamydia pneumoniae
Mycoplasma pneumoniae

Sexually Transmitted Disease Targets

[00241] DNA obtained from viruses and bacteria related to sexually transmitted diseases may also be targeted. A list of targets of interest may include the examples shown in Table 6d.

Table 6d

Sexually Transmitted Disease Targets
HIV (Type 1 and type 2)
Herpes Simplex Virus 1 (HSV-1)
Herpes Simplex Virus 2 (HSV-2)
Hepatitis A
Hepatitis B
Hepatitis C
BACTERIA
Treponema pallidum
Chlamydia
Neisseria gonorrhoeae

Other targets

[00242] Other DNAs may also be targeted. As another example, male genes to determine the sex of the embryo of a pregnant woman/animal, and the male genes to determine the sex of plants and seeds may also be targeted. Examples of further targets of interest may include the following shown in Table 6e.

Table 6e

Viral
Papovavirus (e.g., human papillomavirus (HPV), polyomavirus)
Hepadnavirus (e.g., Hepatitis B Virus (HBV))
Herpesvirus (e.g., herpes simplex virus (HSV)
Varicella zoster virus (VZV)
Epstein-barr virus (EBV)
Cytomegalovirus (CMV)
Herpes lymphotropic virus, Pityriasis Rosea, kaposi's sarcoma-associated herpesvirus);
Adenovirus (e.g., atadenovirus, aviadenovirus, ichtadenovirus, mastadenovirus, siadenovirus)
Poxvirus (e.g., smallpox, vaccinia virus, cowpox virus, monkeypox virus, orf virus, pseudocowpox, bovine papular stomatitis virus; tanapox virus, yaba monkey tumor virus; molluscum contagiosum virus (MCV))
Parvovirus (e.g., adeno-associated virus (AAV), Parvovirus B19, human bocavirus, bufavirus, human parv4 G1); Geminiviridae; Nanoviridae; Phycodnaviridae; and the like.
Dengue fever virus (subtypes 1, 2, 3, and 4)
Zika virus
Hantavirus
Chikungunya virus

[00243] Other miscellaneous targets of interest that provide sources for DNA targets are shown in Table 6f.

Table 6f

Sex determination targets
SRY genes of mammals and non-mammal animals
Other miscellaneous targets of interest
hHPRT1 (hypoxanthine phosphoribosyltransferase 1)
16S E. coli

[00244] A list of non-limiting exemplary target sequences is provided in Tables 6g.

Table 6g

Description	Name	Sequence	Targets
Target sequence	KPC 1	TTGCTGAAGGAGTTGGGCGGCCC (SEQ ID NO: 51)	KPC
Target sequence	NDM 1	GCGATCTGGTTTTCCGCCAGCTC (SEQ ID NO: 52)	NDM
Target sequence	Ctrol + hHPRT1 1	GGTTAAAGATGGTTAAATGAT (SEQ ID NO: 53)	hHPRT1
Target sequence	S16 cntl Ecoli 1	CAGTAGTTATCCCCCTCCATCAG (SEQ ID NO: 54)	16S E. coli
Target sequence	DENV1	CTTCTGTCCAGTGAGCATGGTCT (SEQ ID NO: 55)	Dengue virus
Target sequence	DENV2	TGGTTCAAAGAGAGCTGGTATAA (SEQ ID NO: 56)	Dengue Virus
Target sequence	ZIK1	GGCATGTGCGTCCTTGAACTCTA (SEQ ID NO: 57)	Zika
Target sequence	ZIK2	CCTTTTGGCATGTGCGTCCTTGA (SEQ ID NO: 58)	Zika
Target sequence	OXA1	AGCCCGAATAATATAGTCACCAT (SEQ ID NO: 59)	Oxa-1
Target sequence	OXA 1b	AGCCCGAATAATATAGTCGCCAT (SEQ ID NO: 60)	Oxa-1
Target sequence	HANTAndes1	GTGGCAGCTCAAAAATTGGCTAC (SEQ ID NO: 61)	Hantavirus
Target sequence	HANTAndes2	GATGATCATCAGGCTCAAGCCCT (SEQ ID NO: 62)	Hantavirus
Target sequence	MecA1	TCTTTTTGCCAACCTTTACCATC (SEQ ID NO: 63)	MecA1
Target sequence	SARS-CoV-2	GATCGCGCCCCACTGCGTTCTCC (SEQ ID NO: 120)	SARS-CoV-2

b. Samples

[00245] The term "sample" is used herein to mean any sample that includes DNA (e.g., in order to determine whether a target DNA is present among a population of DNAs). As noted above, the DNA can be single stranded DNA, double stranded DNA, complementary DNA, and the like.

[00246] A sample intended for detection comprises a plurality of nucleic acids. Thus, in some embodiments a sample includes two or more (e.g., 3 or more, 5 or more, 10 or more, 20 or more, 50 or more, 100 or more, 500 or more, 1,000 or more, or 5,000 or more) nucleic acids (e.g.,

DNAs). A detection method can be used as a very sensitive way to detect a target DNA present in a sample (e.g., in a complex mixture of nucleic acids such as DNAs).

[00247] In some embodiments the sample includes 5 or more DNAs (e.g., 10 or more, 20 or more, 50 or more, 100 or more, 500 or more, 1,000 or more, or 5,000 or more DNAs) that differ from one another in sequence. In some embodiments, the sample includes 10 or more, 20 or more, 50 or more, 100 or more, 500 or more, 10^3 or more, 5×10^3 or more, 10^4 or more, 5×10^4 or more, 10^5 or more, 5×10^5 or more, 10^6 or more 5×10^6 or more, or 10^7 or more, DNAs. In some embodiments, the sample comprises from 10 to 20, from 20 to 50, from 50 to 100, from 100 to 500, from 500 to 10^3, from 10^3 to 5×10^3, from 5×10^3 to 10^4, from 10^4 to 5×10^4, from 5×10^4 to 10^5, from 10^5 to 5×10^5, from 5×10^5 to 10^6, from 10^6 to 5×10^6, or from 5×10^6 to 10^7, or more than 10^7, DNAs. In some embodiments, the sample comprises from 5 to 10^7 DNAs (e.g., that differ from one another in sequence)(e.g., from 5 to 10^6, from 5 to 10^5, from 5 to 50,000, from 5 to 30,000, from 10 to 10^6, from 10 to 10^5, from 10 to 50,000, from 10 to 30,000, from 20 to 10^6, from 20 to 10^5, from 20 to 50,000, or from 20 to 30,000 DNAs).

[00248] In some embodiments the sample includes 20 or more DNAs that differ from one another in sequence. In some embodiments, the sample includes DNAs from a cell lysate (e.g., a eukaryotic cell lysate, a mammalian cell lysate, a human cell lysate, a prokaryotic cell lysate, a plant cell lysate, and the like). For example, in some embodiments the sample includes DNA from a cell such as a eukaryotic cell, e.g., a mammalian cell such as a human cell.

[00249] The sample can be derived from any source, e.g., the sample can be a synthetic combination of purified DNAs; the sample can be a cell lysate, a DNA-enriched cell lysate, or DNAs isolated and/or purified from a cell lysate. The sample can be from a patient (e.g., for the purpose of diagnosis). The sample can be from permeabilized cells. The sample can be from crosslinked cells. The sample can be in tissue sections.

[00250] A sample can include a target DNA and a plurality of non-target DNAs. In some embodiments, the target DNA is present in the sample at one or more copies per 5 to 10^9 copies of the non-target DNAs.

[00251] Suitable samples include but are not limited to urine, blood, serum, plasma, lymphatic fluid, cerebrospinal fluid, saliva, nasopharyngeal, oropharyngeal, nasopharyngeal/ oropharyngeal, aspirate, or biopsy sample. Thus, the term “sample” with respect to a patient encompasses blood and other liquid samples of biological origin, solid tissue samples such as a

biopsy specimen or tissue cultures or cells derived therefrom and the progeny thereof. Samples also can be samples that have been manipulated in any way after their procurement, such as by treatment with reagents; washed; or enrichment for certain cell populations, such as cancer cells. The samples can be obtained by use of a swab, for example, a nasopharyngeal swab, an oropharyngeal swab, or a nasopharyngeal/ oropharyngeal swab. Samples also can be samples that have been enriched for particular types of molecules, e.g., DNAs. Samples encompasses biological samples such as a clinical sample such as blood, plasma, serum, aspirate, cerebral spinal fluid (CSF), and also includes tissue obtained by surgical resection, tissue obtained by biopsy, cells in culture, cell supernatants, cell lysates, tissue samples, organs, bone marrow, and the like. A "biological sample" includes biological fluids derived therefrom (e.g., cancerous cell, infected cell, etc.), e.g., a sample comprising DNAs that is obtained from such cells (e.g., a cell lysate or other cell extract comprising DNAs).

[00252] A sample can comprise, or can be obtained from, any of a variety of cells, tissues, organs, or acellular fluids. Suitable sample sources include eukaryotic cells, bacterial cells, and archaeal cells. Suitable sample sources include single-celled organisms and multi-cellular organisms. Suitable sample sources include single-cell eukaryotic organisms; a plant or a plant cell; an algal cell; a fungal cell; an animal cell, tissue, or organ; a cell, tissue, or organ from an invertebrate animal; a cell, tissue, fluid, or organ from a vertebrate animal; a cell, tissue, fluid, or organ from a mammal (e.g., a human; a non-human primate; an ungulate; a feline; a bovine; an ovine; a caprine; etc.). Suitable sample sources include nematodes, protozoans, and the like. Suitable sample sources include parasites such as helminths, malarial parasites, etc.

[00253] Suitable sample sources include a cell, tissue, or organism of any of the six kingdoms.

[00254] Suitable sources of a sample include cells, fluid, tissue, or organ taken from an organism; from a particular cell or group of cells isolated from an organism; etc. For example, where the organism is a plant, suitable sources include xylem, the phloem, the cambium layer, leaves, roots, etc. Where the organism is an animal, suitable sources include particular tissues (e.g., lung, liver, heart, kidney, brain, spleen, skin, fetal tissue, etc.), or a particular cell type (e.g., neuronal cells, epithelial cells, endothelial cells, astrocytes, macrophages, glial cells, islet cells, T lymphocytes, B lymphocytes, etc.).

[00255] In some embodiments, the source of the sample is a (or is suspected of being a diseased cell, fluid, tissue, or organ.

[00256] In some embodiments, the source of the sample is a normal (non-diseased) cell, fluid, tissue, or organ.

[00257] In some embodiments, the source of the sample is a (or is suspected of being a pathogen-infected cell, tissue, or organ. For example, the source of a sample can be an individual who may or may not be infected—and the sample could be any biological sample (e.g., blood, saliva, biopsy, plasma, serum, bronchoalveolar lavage, sputum, a fecal sample, cerebrospinal fluid, a fine needle aspirate, a swab sample (e.g., a buccal swab, a cervical swab, a nasal swab), interstitial fluid, synovial fluid, nasal discharge, tears, buffy coat, a mucous membrane sample, an epithelial cell sample (e.g., epithelial cell scraping), etc.) collected from the individual. In some embodiments, the sample is a cell-free liquid sample.

[00258] In some embodiments, the sample is a liquid sample that can comprise cells (urine, blood, serum, plasma, lymphatic fluid, cerebrospinal fluid, saliva, nasopharyngeal, oropharyngeal, nasopharyngeal/ oropharyngeal, aspirate, and biopsy). Pathogens include viruses, fungi, helminths, protozoa, malarial parasites, Plasmodium parasites, Toxoplasma parasites, Schistosoma parasites, and the like. “Helminths” include roundworms, heartworms, and phytophagous nematodes (Nematoda), flukes (Tematoda), Acanthocephala, and tapeworms (Cestoda). Protozoan infections include infections from Giardia spp., Trichomonas spp., African trypanosomiasis, amoebic dysentery, babesiosis, balantidial dysentery, Chaga's disease, coccidiosis, malaria and toxoplasmosis. Examples of pathogens such as parasitic/protozoan pathogens include, but are not limited to: Plasmodium falciparum, Plasmodium vivax, Trypanosoma cruzi and Toxoplasma gondii. Fungal pathogens include, but are not limited to: Cryptococcus neoformans, Histoplasma capsulatum, Coccidioides immitis, Blastomyces dermatitidis, Chlamydia trachomatis, and Candida albicans. Pathogenic viruses include RNA or DNA viruses, e.g., coronoavirus (e.g. SARS-CoV, SARS-CoV-2, MERS-CoV); immunodeficiency virus (e.g., HIV); influenza virus; dengue; West Nile virus; herpes virus; yellow fever virus; Hepatitis Virus C; Hepatitis Virus A; Hepatitis Virus B; papillomavirus; and the like. Pathogenic viruses can include DNA viruses such as: a papovavirus (e.g., human papillomavirus (HPV), polyomavirus); a hepadnavirus (e.g., Hepatitis B Virus (HBV)); a herpesvirus (e.g., herpes simplex virus (HSV), varicella zoster virus (VZV), epstein-barr virus (EBV), cytomegalovirus (CMV), herpes lymphotropic virus, Pityriasis Rosea, kaposi's sarcoma-associated herpesvirus); an adenovirus (e.g., atadenovirus, aviadenovirus, ichtadenovirus, mastadenovirus, siadenovirus); a poxvirus (e.g., smallpox, vaccinia virus, cowpox virus,

monkeypox virus, orf virus, pseudocowpox, bovine papular stomatitis virus; tanapox virus, yaba monkey tumor virus; molluscum contagiosum virus (MCV)); a parvovirus (e.g., adeno-associated virus (AAV), Parvovirus B19, human bocavirus, bufavirus, human parv4 G1); Geminiviridae; Nanoviridae; Phycodnaviridae; and the like. Pathogens can include, e.g., DNAviruses [e.g.: a papovavirus (e.g., human papillomavirus (HPV), polyomavirus); a hepadnavirus (e.g., Hepatitis B Virus (HBV)); a herpesvirus (e.g., herpes simplex virus (HSV), varicella zoster virus (VZV), epstein-barr virus (EBV), cytomegalovirus (CMV), herpes lymphotropic virus, Pityriasis Rosea, kaposi's sarcoma-associated herpesvirus); an adenovirus (e.g., atadenovirus, aviadenovirus, ichtadenovirus, mastadenovirus, siadenovirus); a poxvirus (e.g., smallpox, vaccinia virus, cowpox virus, monkeypox virus, orf virus, pseudocowpox, bovine papular stomatitis virus; tanapox virus, yaba monkey tumor virus; molluscum contagiosum virus (MCV)); a parvovirus (e.g., adeno-associated virus (AAV), Parvovirus B19, human bocavirus, bufavirus, human parv4 G1); Geminiviridae; Nanoviridae; Phycodnaviridae; and the like], Mycobacterium tuberculosis, Streptococcus agalactiae, methicillin-resistant Staphylococcus aureus, Legionella pneumophila, Streptococcus pyogenes, Escherichia coli, Neisseria gonorrhoeae, Neisseria meningitidis, Pneumococcus, Cryptococcus neoformans, Histoplasma capsulatum, Hemophilus influenzae B, Treponema pallidum, Lyme disease spirochetes, Pseudomonas aeruginosa, Mycobacterium leprae, Brucella abortus, rabies virus, influenza virus, cytomegalovirus, herpes simplex virus I, herpes simplex virus II, human serum parvo-like virus, respiratory syncytial virus, varicella-zoster virus, hepatitis B virus, hepatitis C virus, measles virus, adenovirus, human T-cell leukemia viruses, Epstein-Barr virus, murine leukemia virus, mumps virus, vesicular stomatitis virus, Sindbis virus, lymphocytic choriomeningitis virus, wart virus, blue tongue virus, Sendai virus, feline leukemia virus, Reovirus, polio virus, simian virus 40, mouse mammary tumor virus, dengue virus, rubella virus, West Nile virus, Plasmodium falciparum, Plasmodium vivax, Toxoplasma gondii, Trypanosoma rangeli, Trypanosoma cruzi, Trypanosoma rhodesiense, Trypanosoma brucei, Schistosoma mansoni, Schistosoma japonicum, Babesia bovis, Eimeria tenella, Onchocerca volvulus, Leishmania tropica, Mycobacterium tuberculosis, Trichinella spiralis, Theileria parva, Taenia hydatigena, Taenia ovis, Taenia saginata, Echinococcus granulosus, Mesocestoides corti, Mycoplasma arthritidis, M. hyorhinis, M. orale, M. arginini, Acholeplasma laidlawii, M. salivarium and M. pneumoniae.

c. Measuring a Detectable Signal

[00259] The detection method generally includes a step of measuring (e.g., measuring a detectable signal produced by the Cas12 of the disclosure. A detectable signal can be any signal that is produced when ss oliogonucleotide is cleaved. The step of detection can involve a fluorescence-based detection. The readout of such detection methods can be any convenient readout. Examples of possible readouts include but are not limited to: a measured amount of detectable fluorescent signal; a visual analysis of bands on a gel (e.g., bands that represent cleaved product versus uncleaved substrate), a visual or sensor based detection of the presence or absence of a color (i.e., color detection method), the presence or absence of (or a particular amount of) a magnetic signal and the presence or absence of (or a particular amount of) an electrical signal.

[00260] The measuring can in some embodiments be quantitative, e.g., in the sense that the amount of signal detected can be used to determine the amount of target DNA present in the sample. The measuring can in some embodiments be qualitative, e.g., in the sense that the presence or absence of detectable signal can indicate the presence or absence of targeted DNA (e.g., virus, SNP, etc.). In some embodiments, a detectable signal will not be present (e.g., above a given threshold level) unless the targeted DNA(s) (e.g., virus, SNP, etc.) is present above a particular threshold concentration. In some embodiments, the threshold of detection can be titrated by modifying the amount of the Cas12 protein provided.

[00261] The compositions and methods of this disclosure can be used to detect any DNA target.

[00262] In some embodiments, the detection methods of the disclosure can be used to determine the amount of a target DNA in a sample (e.g., a sample comprising the target DNA and a plurality of non-target DNAs). Determining the amount of a target DNA in a sample can comprise comparing the amount of detectable signal generated from a test sample to the amount of detectable signal generated from a reference sample. Determining the amount of a target DNA in a sample can comprise: measuring the detectable signal to generate a test measurement; measuring a detectable signal produced by a reference sample to generate a reference measurement; and comparing the test measurement to the reference measurement to determine an amount of target DNA present in the sample.

[00263] In some embodiments, the detectable signal is detectable in less than 1, 2, 3, 4, 5, 10, 15, 20, 30, 60, 90, 120, 150, 180, 210, or 240 minutes.

[00264] In some embodiments, sensitivity of a subject composition and/or method (e.g., for detecting the presence of a target DNA, such as viral DNA or a SNP, in cellular genomic DNA) can be increased by coupling detection with nucleic acid amplification.

[00265] In some embodiments, the nucleic acids in a sample are amplified prior to contact with a Cas12; in particular embodiments, the Cas12 remains in an inactive state until amplification has concluded. In some embodiments, the nucleic acids in a sample are amplified simultaneous with contact with Cas12. Amplification can be carried out using primers. As it relates to the overall processing time for the detection method, amplification can occur for 5 seconds or more, up to 240 minutes or more.

[00266] Various amplification methods and components will be known to one of ordinary skill in the art and any convenient method can be used.

[00267] Nucleic acid amplification can comprise polymerase chain reaction (PCR), reverse transcription PCR (RT-PCR), quantitative PCR (qPCR), reverse transcription qPCR (RT-qPCR), isothermal PCR, nested PCR, multiplex PCR, asymmetric PCR, touchdown PCR, random primer PCR, hemi-nested PCR, polymerase cycling assembly (PCA), colony PCR, ligase chain reaction (LCR), digital PCR, methylation specific-PCR (MSP),co-amplification at lower denaturation temperature-PCR (COLD-PCR), allele-specific PCR, intersequence-specific PCR (ISS-PCR), whole genome amplification (WGA), inverse PCR, and thermal asymmetric interlaced PCR (TAIL-PCR).

[00268] In some embodiments the amplification is isothermal amplification. Isothermal nucleic acid amplification methods can therefore be carried out inside or outside of a laboratory environment. Examples of isothermal amplification methods include but are not limited to: loop-mediated isothermal Amplification (LAMP), helicase-dependent Amplification (HDA), recombinase polymerase amplification (RPA), strand displacement amplification (SDA), nucleic acid sequence-based amplification (NASBA), transcription mediated amplification (TMA), nicking enzyme amplification reaction (NEAR), rolling circle amplification (RCA), multiple displacement amplification (MDA), Ramification (RAM), circular helicase-dependent amplification (cHDA), single primer isothermal amplification (SPIA), signal mediated amplification of RNA technology (SMART), self-sustained sequence replication (3SR), genome exponential amplification reaction (GEAR) and isothermal multiple displacement amplification (IMDA).

d. Detector Oligonucleotides

[00269] The novel Cas12 proteins of the disclosure possess collateral cleavage (*trans*-cleavage) activity. As in the case of Cas12a.1, the protein possesses the ability to collaterally cleave ssDNAs upon the binding of the DNA targeted by the guide. In the case of Cas12p, the protein possesses the dual ability to collaterally cleave all types of oligonucleotides inclusive of ssDNAs, ssRNAs, chimeric ss DNA/RNAs, and other oligonucleotides comprising RNAs. These characteristics are taken into account when designing the detector oligonucleotides when using the assay.

[00270] In some embodiments, a detection method includes contacting a sample (e.g., a sample comprising a target DNA and a plurality of non-target ssDNAs) with: i) a Cas12 protein of the disclosure; ii) a gRNA (or precursor gRNA array); and iii) a detector that does not hybridize with the guide sequence of the gRNA. For example, in some embodiments, a detection method includes contacting a sample with a labeled detector (detector ssDNA in the case of Cas12a.1 or a detector comprising RNA, DNA, and combinations of the same in the case of Cas12p) that includes a fluorescence-emitting dye pair; the Cas12 protein of the disclosure has the ability to cleave the labeled detector after it is activated (by gRNA hybridizing to a target DNA); and the detectable signal that is measured is produced by the fluorescence-emitting dye pair. For example, in some embodiments, a detection method includes contacting a sample with a labeled detector comprising a fluorescence resonance energy transfer (FRET) pair or a quencher/fluor pair, or both. In some embodiments, a detection method includes contacting a sample with a labeled detector comprising a FRET pair. In some embodiments, a detection method includes contacting a sample with a labeled detector comprising a fluor/quencher pair.

[00271] Fluorescence-emitting dye pairs comprise a FRET pair or a quencher/fluor pair. In both embodiments of a FRET pair and a quencher/fluor pair, the emission spectrum of one of the dyes overlaps a region of the absorption spectrum of the other dye in the pair. As used herein, the term "fluorescence-emitting dye pair" is a generic term used to encompass both a "fluorescence resonance energy transfer (FRET) pair" and a "quencher/fluor pair". The term "fluorescence-emitting dye pair" is used interchangeably with the phrase "a FRET pair and/or a quencher/fluor pair."

[00272] In some embodiments (e.g., when the detector includes a FRET pair) the labeled detector produces an amount of detectable signal prior to being cleaved, and the amount of detectable signal that is measured is reduced when the labeled detector is cleaved. In some

embodiments, the labeled detector produces a first detectable signal prior to being cleaved (e.g., from a FRET pair) and a second detectable signal when the labeled detector is cleaved (e.g., from a quencher/fluor pair). As such, in some embodiments, the labeled detector comprises a FRET pair and a quencher/fluor pair.

[00273] In some embodiments, the labeled detector comprises a FRET pair.

[00274] FRET donor and acceptor moieties (FRET pairs) will be known to one of ordinary skill in the art and any convenient FRET pair (e.g., any convenient donor and acceptor moiety pair) can be used. Examples of suitable FRET pairs include but are not limited to those presented in Table 7. FRET pairs provided in US10,253,365 are incorporate by reference herein in their entirety. In some embodiments, the FRET pair is 5' 6-FAM and 3IABkFQ (Iowa Black (Registred)-FQ).

Table 7

Examples of FRET pairs (donor and and acceptor pairs)

	Donor	**Acceptor**	
	Tryptophan	Dansyl	
	IAEDANS (1)	DDPM (2)	
	BFP	DsRFP	
	Dansyl	Fluorescein	
		isothiocyanate (FITC)	
	Dansyl	Octadecylrhodamine	
	Cyan fluorescent	Green fluorescent protein	
	protein (CFP)	(GFP)	
	CF (3)	Texas Red	
	Fluorescein	Tetramethylrhodamine	
	Cy3	Cy5	
	GFP	Yellow fluorescent	
		protein (YFP)	
	BODIPY FL (4)	BODIPY FL (4)	
	Rhodamine 110	Cy3	
	Rhodamine 6G	Malachite Green	
	FITC	Eosin Thiosemicarbazide	
	B-Phycoerythrin	Cy5	
	Cy5	Cy5.5	
	(1) 5-(2-iodoacetylaminoethyl)aminonaphthalene-1-sulfonic acid		
	(2) N-(4-dimethylamino-3,5-dinitrophenyl)maleimide		
	(3) carboxyfluorescein succinimidyl ester		
	(4) 4,4-difluoro-4-bora-3a,4a-diaza-s-indacene		

[00275] In some embodiments, a detectable signal is produced when the labeled detector is cleaved (e.g., in some embodiments, the labeled detector comprises a quencher/fluor pair).

[00276] Any fluorescent label can be utilized. Examples of fluorescent labels include, but are not limited to: an Alexa Fluor® dye, an ATTO dye (e.g., ATTO 390, ATTO 425, ATTO 465, ATTO 488, ATTO 495, ATTO 514, ATTO 520, ATTO 532, ATTO Rho6G, ATTO 542, ATTO 550, ATTO 565, ATTO Rho3B, ATTO Rho11, ATTO Rho12, ATTO Thio12, ATTO Rho101, ATTO 590, ATTO 594, ATTO Rho13, ATTO 610, ATTO 620, ATTO Rho14, ATTO 633, ATTO 647, ATTO 647N, ATTO 655, ATTO Oxa12, ATTO 665, ATTO 680, ATTO 700, ATTO 725, ATTO 740), a DyLight dye, a cyanine dye (e.g., Cy2, Cy3, Cy3.5, Cy3b, Cy5, Cy5.5, Cy7, Cy7.5), a FluoProbes dye, a Sulfo Cy dye, a Seta dye, an IRIS Dye, a SeTau dye, an SRfluor dye, a Square dye, fluorescein isothiocyanate (FITC), fluorescein amidite (FAM), tetramethylrhodamine (TRITC), Texas Red, Oregon Green, Pacific Blue, Pacific Green, Pacific Orange, quantum dots, and a tethered fluorescent protein.

[00277] Examples of quencher moieties include, but are not limited to: a dark quencher, a Black Hole Quencher® (BHQ®) (e.g., BHQ-0, BHQ-1, BHQ-2, BHQ-3), a Qxl quencher, an ATTO quencher (e.g., ATTO 540Q, ATTO 580Q, and ATTO 612Q), dimethylaminoazobenzenesulfonic acid (Dabsyl), Iowa Black RQ, Iowa Black FQ, IRDye QC-1, a QSY dye (e.g., QSY 7, QSY 9, QSY 21), AbsoluteQuencher, Eclipse, and metal clusters such as gold nanoparticles, and the like.

[00278] In some embodiments, a quencher moiety is selected from: a dark quencher, a Black Hole Quencher® (BHQ®) (e.g., BHQ-0, BHQ-1, BHQ-2, BHQ-3), a Qxl quencher, an ATTO quencher (e.g., ATTO 540Q, ATTO 580Q, and ATTO 612Q), dimethylaminoazobenzenesulfonic acid (Dabsyl), Iowa Black RQ, Iowa Black FQ, IRDye QC-1, a QSY dye (e.g., QSY 7, QSY 9, QSY 21), AbsoluteQuencher, Eclipse, and a metal cluster.

[00279] In some embodiments, cleavage of a labeled detector can be detected by measuring a colorimetric read-out. For example, the liberation of a fluorophore (e.g., liberation from a FRET pair, liberation from a quencher/fluor pair, and the like) can result in a wavelength shift (and thus color shift) of a detectable signal. Thus, in some embodiments, cleavage of a subject labeled detector can be detected by a color-shift. Such a shift can be expressed as a loss of an amount of signal of one color (wavelength), a gain in the amount of another color, a change in the ration of one color to another, and the like.

[00280] As provided herein, a labeled detector can be a nucleic acid mimetic. Polynucleotide mimics include PNAs, LNAs, CeNAs, and morpholino nucleic acids.

[00281] A labeled detector can also include one or more substituted sugar moieties.

[00282] A labeled detector may also include modified nucleotides.

e. Positive Controls

[00283] The detection methods provided herein can also include a positive control target DNA. In some embodiments, the methods include using a positive control gRNA that comprises a nucleotide sequence that hybridizes to a control target DNA. In some embodiments, the positive control target DNA is provided in various amounts. In some embodiments, the positive control target DNA is provided in various known concentrations, along with control non-target DNAs.

f. gRNA Arrays

[00284] In some embodiments, the method comprises contacting the sample with a precursor gRNA array, wherein the novel Cas12 protein of the disclosure cleaves the precursor gRNA array to produce said gRNA.

[00285] In some embodiments a such a gRNA array includes 2 or more gRNAs (e.g., 3 or more, 4 or more, 5 or more, 6 or more, or 7 or more, gRNAs). The gRNAs of a given array can target (i.e., can include guide sequences that hybridize to) different target sites of the same target DNA (e.g., which can increase sensitivity of detection) and/or can target different target DNAs (e.g., single nucleotide polymorphisms (SNPs), different strains of a particular virus, etc.), and such could be used for example to detect multiple strains of a virus. In some embodiments, each gRNA of a precursor gRNA array has a different guide sequence.

[00286] In some embodiments, the precursor gRNA array comprises two or more gRNAs that target different target sites within the same target DNA. For example, such a scenario can in some embodiments increase sensitivity of detection by activating Cas9 or Cas12 protein of the disclosure when either one hybridizes to the target DNA. As such, in some embodiments as subject composition (e.g., kit) or method includes two or more gRNAs (in the context of a precursor gRNA array, or not in the context of a precursor gRNA array, e.g., the gRNAs can be mature gRNAs).

[00287] In some embodiments, the precursor gRNA array comprises two or more gRNAs that target different target DNAs. For example, such a scenario can result in a positive signal when any one of a family of potential target DNAs is present. Such an array could be used for targeting a family of transcripts, e.g., based on variation such as single nucleotide

polymorphisms (SNPs) (e.g., for diagnostic purposes). Such could also be useful for detecting whether any one of a number of different strains of virus is present. Such could also be useful for detecting whether any one of a number of different species, strains, isolates, or variants of a bacterium or virus is present As such, in some embodiments as subject composition (e.g., kit) or method includes two or more gRNAs (in the context of a precursor gRNA array, or not in the context of a precursor gRNA array, e.g., the gRNAs can be mature gRNAs).

V. Compositions of Matter

[00288] Provided herein are compositions and pharmaceutical compositions comprising the Cas9 proteins and/or the Cas9 gRNAs of the disclosure, which can optionally include a pharmaceutically acceptable carrier and/or a protein stabilizing buffer, and/or a nucleic acid stabilizing buffer. In some embodiments, the Cas9 proteins and/or the Cas9 gRNAs are provided in a lyophilized form.

[00289] Provided herein are compositions and pharmaceutical compositions comprising the Cas12 proteins and/or the Cas12 gRNAs of the disclosure, which can optionally include a pharmaceutically acceptable carrier and/or a protein stabilizing buffer, and/or a nucleic acid stabilizing buffer. In some embodiments, the Cas12 proteins and/or the Cas12 gRNAs are provided in a lyophilized form.

[00290] Provided herein are compositions comprising gRNAs and/or gRNA arrays of the disclosure (compatible for use with Cas9 proteins of the disclosure, and/or Cas12 proteins of the disclosure), and optionally a protein stabilizing buffer.

[00291] Provided herein are proteins comprising an amino acid sequence with 70%-99.5% homology to SEQ ID NO: 1, 2, 3, 4, 222, 5, 10, 11, or 12. Provided herein are compositions comprising these proteins, and optionally a pharmaceutically acceptable carrier. Provided herein are these proteins and optionally a protein stabilizing buffer.

[00292] Provided herein are DNA polynucleotides encoding a sequence that encodes any of the Cas9 or Cas12 proteins of the disclosure. Also provided are recombinant expression vectors comprising such DNA polynucleotides. In some embodiments, a nucleotide sequence encoding a Cas9 or Cas12 of the disclosure is operably linked to a promoter. In some embodiments, the nucleic acid encoding the Cas9 or Cas12 further comprises a nuclear localization signal (NLS), useful for expression in eukaryotic systems.

[00293] Provided herein are DNA polynucleotides or RNAs comprising a sequence that encodes any of the gRNAs of the disclosure. Also provided are recombinant expression vectors comprising such DNA polynucleotides. In some embodiments, a nucleotide sequence encoding a gRNA of the disclosure is operably linked to a promoter.

[00294] Also provided herein are host cells comprising any of the recombinant vectors provided herein.

VI. Kits

[00295] Provided herein are kits comprising one or more components of the Cas9 and Cas12 engineered systems described herein, useful for a variety of applications including, but not limited to, therapeutic and diagnostic applications.

[00296] In some embodiments provided herein is a kit comprising: (a) a Cas9.1, Cas9.2, Cas9.3 or Cas9.4 protein, or a nucleic acid encoding the Cas9.1, Cas9.2, Cas9.3 or Cas9.4 protein; and (b) a Cas9.1, Cas9.2, Cas9.3 or Cas9.4 gRNA, or a nucleic acid encoding the Cas9.1, Cas9.2, Cas9.3 or Cas9.4 gRNA, wherein the gRNA and the Cas9.1, Cas9.2, Cas9.3 or Cas9.4 protein do not naturally occur together, wherein the gRNA is capable of hybridizing to a target sequence in a target DNA, and the gRNA is capable of forming a complex with the Cas9.1, Cas9.2, Cas9.3 or Cas9.4 protein.

[00297] In some embodiments provided herein is a kit comprising: (a) a Cas12a.1, Cas12p, or Cas12q protein, or a nucleic acid encoding the Cas12a.1, Cas12p, or Cas12q protein; and (b) a Cas12a.1, Cas12p, or Cas12q gRNA, or a nucleic acid encoding a Cas12a.1, Cas12p, or Cas12q gRNA, wherein the gRNA and the Cas12a.1, Cas12p, or Cas12q protein do not naturally occur together, wherein the gRNA is capable of hybridizing to a target sequence in a target DNA, and the gRNA is capable of forming a complex with the Cas12a.1, Cas12p, or Cas12q protein.

[00298] In exemplary embodiments, provided herein are diagnostic kits. In exemplary embodiments, the reagent components are provided in lyophilized form. In some embodiments, the reagent components are provided individually (either lyophilized or not lyophilized), in other embodiments, the reagent componets are provided in a pre-mixed format (either lyophilized or not lyophilized).

[00299] The following are exemplary kit reagent components useful for the detection of SARS-CoV-2, a RNA virus, using one of the novel Cas12 proteins of the disclosure (Cas12a.1, Cas12p, and Cas12q), exemplified in Example 10.

[00300] (1) Lyophilized reaction mix containing reagents, SARS-CoV-2 primer sets and enzymes for reverse transcription and loop-mediated isothermal amplification (RT-LAMP) of a gene of diseasSARS-CoV-2 genome.

[00301] (2) Lyophilized reaction mix containing reagents, control RNAse P primer sets and enzymes for reverse transcription and RT-LAMP amplification of human housekeeping gene RNAse P.

[00302] (3) Lyophilized reaction mix containing reagents and Cas12p-gRNA RNP complexes for detection of a SARS-CoV-2 amplification product. Such mix may also include a labeled reporter, e.g. a 5'FAM-3'Quencher ssRNA-based oligonucleotide reporter, or a 5'FAM-3'Quencher single stranded DNA/RNA chimera-based oligonucleotide reporter.

[00303] (4) Lyophilized reaction mix containing reagents and Cas12p-gRNA RNP complexes for detection of RNAse P amplification product. Such mix may also include a labeled reporter, e.g. a 5'FAM-3'Quencher RNA-based oligonucleotide reporter.

[00304] FIG. 23 shows an exemplary strip of lyophilized beads of the disclosure included in exemplary kits. Each bead can be resuspended with water, and used for a detection assay. Exemplary beads each comprise a CRISPR protein (e.g. Cas12p), a gRNA for a desired target (e.g. gRNA for SARS-CoV-2), a labeled reporter, a buffer, and nuclease free water.

VII. Enumerated Embodiments

[00305] Provided herein are illustrative, non-limiting, enumerated embodiments of the disclosure.

[00306] Embodiment 1. An engineered system comprising:

a. a Cas9.1, Cas9.2, Cas9.3, or Cas9.4 protein, or a nucleic acid encoding the a Cas9.1, Cas9.2, Cas9.3, or Cas9.4 protein; and

b. a Cas9.1, Cas9.2, Cas9.3, or Cas9.4 guide RNA (gRNA), or a nucleic acid encoding the Cas9.1, Cas9.2, Cas9.3, or Cas9.4 gRNA, wherein the gRNA and the Cas9.1, Cas9.2, Cas9.3, or Cas9.4 protein do not naturally occur together, wherein the gRNA is capable of hybridizing to a target sequence in a target DNA, and the gRNA is capable of forming a complex with the Cas9.1, Cas9.2, Cas9.3, or Cas9.45 protein.

[00307] Embodiment 2. The system of embodiment 1, comprising:

a. a Cas9.1, Cas9.2, Cas9.3, Cas9.4 protein; and

b. a Cas9.1, Cas9.2, Cas9.3, or Cas9.4 gRNA.

[00308] Embodiment 3. The system of embodiment 1, comprising:

a. a nucleic acid encoding the Cas9.1, Cas9.2, Cas9.3, or Cas9.4 protein; and

b. a nucleic acid encoding the Cas9.1, Cas9.2, Cas9.3, or Cas9.4 gRNA.

[00309] Embodiment 4. The system of any one of embodiments 1 to 3, wherein the gRNA is a single-molecule gRNA.

[00310] Embodiment 5. The system of any one of embodiments 1 to 3, wherein the gRNA is a dual-molecule gRNA.

[00311] Embodiment 6. The system of any one of embodiments 1 to 5, wherein the Cas9.1 protein comprises the amino acid sequence of SEQ ID NO: 1, or at least 70% sequence identity thereto.

[00312] Embodiment 7. The system of any one of embodiments 1 to 5, wherein the Cas9.2 protein comprises the amino acid sequence of SEQ ID NO: 2 or at least 70% sequence identity thereto.

[00313] Embodiment 8. The system of any one of embodiments 1 to 5, wherein the Cas9.3 protein comprises the amino acid sequence of SEQ ID NO: 10, or at least 70% sequence identity thereto.

[00314] Embodiment 9. The system of any one of embodiments 1 to 5, wherein the Cas9.4 protein comprises the amino acid sequence of SEQ ID NO: 11, or at least 70% sequence identity thereto.

[00315] Embodiment 10. The system of any one of embodiments 1 to 7, wherein the target sequence is a sequence of a target provided in any of Tables 6a-6f.

[00316] Embodiment 11. The system of any one of embodiments 1 to 7, wherein the target sequence is a sequence of a human.

[00317] Embodiment 12. The system of any one of embodiments 1 to 7, wherein the target sequence is a sequence of a non-human primate.

[00318] Embodiment 13. The system of any one of embodiments 1 to 12, wherein the Cas9.1, Cas9.2, Cas9.3, or Cas9.4 protein is a catalytically active protein.

[00319] Embodiment 14. The system of embodiment 13, wherein the Cas9.1, Cas9.2, Cas9.3, or Cas9.4 protein cleaves at a site distal to the target sequence.

[00320] Embodiment 15. The system of any one of embodiments 1 to 12, wherein the Cas9.1, Cas9.2, Cas9.3, or Cas9.4 protein is a catalytically dead protein.

[00321] Embodiment 16. The system of any one of embodiments 1 to 12, wherein the Cas9.1, Cas9.2, Cas9.3, or Cas9.4 protein comprises nickase activity.

[00322] Embodiment 17. An engineered system comprising:

a. a Class 2 Type V CRISPR-Cas RNA-guided endonuclease protein; and

b. a single guide RNA (gRNA),

wherein the gRNA and the Class 2 Type V CRISPR-Cas RNA-guided endonuclease protein do not naturally occur together, wherein the gRNA is capable of hybridizing to a target sequence in a target DNA, wherein the gRNA is capable of forming a complex with the Class 2 Type V CRISPR-Cas RNA-guided endonuclease protein, and wherein the Class 2 Type V CRISPR-Cas RNA-guided endonuclease protein possesses collateral activity and is capable of collaterally cleaving a single stranded polynucleotide comprising RNA without a tracrRNA.

[00323] Embodiment 18. The system of embodiment 17, wherein the Class 2 Type V CRISPR-Cas RNA-guided endonuclease protein comprises the amino acid sequence of SEQ ID NO: 4, or at least 70% sequence identity thereto.

[00324] Embodiment 19. The system of any one of embodiments 17 to 18, wherein the target sequence is a sequence of a target provided in any of Tables 6a-6f.

[00325] Embodiment 20. The system of any one of embodiments 17 to 18, wherein the target sequence is a sequence of a human.

[00326] Embodiment 21. The system of any one of embodiments 17 to 18, wherein the target sequence is a sequence of a non-human primate.

[00327] Embodiment 22. The system of any one of embodiments 17 to 18, wherein the target sequence is a bacterial or viral sequence.

[00328] Embodiment 23. The system of any one of embodiments 17 to 22, wherein the Class 2 Type V CRISPR-Cas RNA-guided endonuclease protein is capable of collaterally cleaving a single stranded RNA.

[00329] Embodiment 24. The system of any one of embodiments 17 to 22, wherein the Class 2 Type V CRISPR-Cas RNA-guided endonuclease protein is capable of collaterally cleaving a single stranded DNA/RNA hybrid.

[00330] Embodiment 25. An engineered system comprising:

a. a Cas12a.1, Cas12p, or Cas12q protein, or a nucleic acid encoding the Cas12a.1, Cas12p, or Cas12q protein; and

b. a Cas12a.1, Cas12p, or Cas12q gRNA, or a nucleic acid encoding a Cas12a.1, Cas12p, or Cas12q gRNA,

wherein the gRNA and the Cas12a.1, Cas12p, or Cas12q protein do not naturally occur together, wherein the gRNA is capable of hybridizing to a target sequence in a target DNA, and the gRNA is capable of forming a complex with the Cas12a.1, Cas12p, or Cas12q protein.

[00331] Embodiment 26. The system of embodiment 25, comprising:

a. a Cas12a.1, Cas12p, or Cas12q protein; and

b. a Cas12a.1, Cas12p, or Cas12q gRNA.

[00332] Embodiment 27. The system of embodiment 25, comprising:

a. a nucleic acid encoding the Cas12a.1, Cas12p, or Cas12q protein; and

b. a nucleic acid encoding a Cas12a.1, Cas12p, or Cas12q gRNA.

[00333] Embodiment 28. The system of any one of embodiments 25 to 27, wherein the Cas12a.1 protein comprises the amino acid sequence of SEQ ID NO: 3, or at least 70% sequence identity thereto.

[00334] Embodiment 29. The system of any one of embodiments 25 to 27, wherein the Cas12p protein comprises the amino acid sequence of SEQ ID NO: 4, or at least 70% sequence identity thereto.

[00335] Embodiment 30. The system of any one of embodiments 25 to 27, wherein the Cas12q protein comprises the amino acid sequence of SEQ ID NO: 222, or at least 70% sequence identity thereto.

[00336] Embodiment 31. The system of any one of embodiments 25 to 27, wherein the Cas12q protein comprises the amino acid sequence of SEQ ID NO: 5, or at least 70% sequence identity thereto.

[00337] Embodiment 32. The system of any one of embodiments 25 to 31, wherein the target sequence is a sequence of a target provided in any of Tables 6a-6f.

[00338] Embodiment 33. The system of any one of embodiments 25 to 31, wherein the target sequence is a sequence of a human.

[00339] Embodiment 34. The system of any one of embodiments 25 to 31, wherein the target sequence is a sequence of a non-human primate.

[00340] Embodiment 35. The system of any one of embodiments 25 to 31, wherein the target sequence is a bacterial or viral sequence.

[00341] Embodiment 36. The system of any one of embodiments 25 to 34, wherein the Cas12a.1, Cas12p, or Cas12q protein is a catalytically active Cas12a.1, Cas12p, or Cas12q protein.

[00342] Embodiment 37. The system of embodiment 36, wherein the Cas12a.1, Cas12p, or Cas12q protein cleaves at a site distal to the target sequence.

[00343] Embodiment 38. The system of any one of embodiments 25 to 34, wherein the Cas12a.1, Cas12p, or Cas12q protein is a catalytically dead Cas12a.1, Cas12p, or Cas12q protein.

[00344] Embodiment 39. The system of any one of embodiments 25 to 34, wherein the Cas12a.1, Cas12p, or Cas12q protein comprises nickase activity.

[00345] Embodiment 40. An engineered single-molecule gRNA, comprising:

a. a targeter-RNA comprising a spacer sequence that is capable of hybridizing with a target sequence in a target DNA; and

b. an activator-RNA that is capable of hybridizing with the targeter-RNA to form a double-stranded RNA duplex, the activator-RNA comprising a activator-RNA,

wherein the targeter-RNA and the activator-RNA are covalently linked to one another, wherein the single-molecule gRNA is capable of forming a complex with a Cas9.1, Cas9.2, Cas9.3, or Cas9.4 protein, and wherein hybridization of the spacer sequence to the target sequence is capable of targeting the Cas9.1, Cas9.2, Cas9.3, or Cas9.4 protein to the target DNA.

[00346] Embodiment 41. The gRNA of embodiment 40, wherein the targeter-RNA and the activator-RNA are arranged in a 5' to 3' orientation.

[00347] Embodiment 42. The gRNA of embodiment 40, wherein the activator-RNA and the targeter-RNA are arranged in a 5' to 3' orientation.

[00348] Embodiment 43. The gRNA of any one of embodiments 40 to 42, wherein the targeter-RNA and the activator-RNA are covalently linked to one another via a linker.

[00349] Embodiment 44. The gRNA of ay one of embodiments 40 to 43, wherein the single-molecule gRNA comprises one or more sequence modifications compared to a sequence of a corresponding wild type tracrRNA and/or crRNA.

[00350] Embodiment 45. The gRNA of ay one of embodiments 40 to 44, wherein the targeter-RNA comprises a spacer sequence of about 10-50 nucleotides that have 100% complementarity to a sequence in the target DNA.

[00351] Embodiment 46. The gRNA of any one of embodiments 40 to 44, wherein the targeter-RNA comprises a spacer sequence of about 10-50 nucleotides that have less than 100% complementarity to a sequence in the target DNA.

[00352] Embodiment 47. The gRNA of any one of embodiments 40 to 46, wherein the target sequence is a sequence of a target provided in any of Tables 6a-6f.

[00353] Embodiment 48. The gRNA of any one of embodiments 40 to 47, wherein the Cas9.1 protein comprises the sequence of SEQ ID NO: 1 or a sequence with at least 70% sequence identity thereto.

[00354] Embodiment 49. The gRNA of any one of embodiments 40 to 47, wherein the Cas9.2 protein comprises the sequence of SEQ ID NO: 2 or a sequence with at least 70% sequence identity thereto.

[00355] Embodiment 50. The gRNA of any one of embodiments 40 to 47, wherein the Cas9.3 protein comprises the sequence of SEQ ID NO: 10 or a sequence with at least 70% sequence identity thereto.

[00356] Embodiment 51. The gRNA of any one of embodiments 40 to 47, wherein the Cas9.4 protein comprises the sequence of SEQ ID NO: 11 or a sequence with at least 70% sequence identity thereto.

[00357] Embodiment 52. An engineered single-molecule gRNA, comprising the scaffold sequence of SEQ ID NO: 116 or SEQ ID NO: 117 and a spacer sequence that is capable of hybridizing with a target sequence in a target DNA.

[00358] Embodiment 53. The gRNA of embodiment 52, wherein the target DNA comprises viral DNA, plant DNA, fungal DNA, or bacterial DNA.

[00359] Embodiment 54. The gRNA of embodiment 52, wherein the target sequence is a sequence of a target provided in any of Tables 6a-6f.

[00360] Embodiment 55. The gRNA of embodiment 52, wherein the target is a coronvavirus.

[00361] Embodiment 56. The gRNA of embodiment 52, wherein the target is a SARS-CoV-2 virus.

[00362] Embodiment 57. The gRNA of embodiment 52, wherein the target DNA is cDNA, and has been obtained by reverse transcription.

[00363] Embodiment 58. A method of modifying a target DNA, the method comprising contacting the target DNA with any one of the systems of embodiments 1 to 39, wherein the gRNA hybridizes with the target sequence whereby modification of the target DNA occurs.

[00364] Embodiment 59. The method of embodiment 58, wherein the target DNA is extrachromosomal DNA.

[00365] Embodiment 60. The method of embodiment 58, wherein the target DNA is part of a chromosome.

[00366] Embodiment 61. The method of embodiment 58, wherein the target DNA is part of a chromosome in vitro.

[00367] Embodiment 62. The method of embodiment 58, wherein the target DNA is part of a chromosome in vivo.

[00368] Embodiment 63. The method of embodiment 58, wherein the target DNA is outside a cell.

[00369] Embodiment 64. The method of embodiment 58, wherein the target DNA is inside a cell.

[00370] Embodiment 65. The method of embodiment 64, wherein the target DNA comprises a gene and/or its regulatory region.

[00371] Embodiment 66. The method of embodiment 64 or 65, wherein 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, in invertebrate cell, a vertebrate cell, a fish 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.

[00372] Embodiment 67. The method of any of the embodiments of 58 to 66, wherein the modifying comprises introducing a double strand break in the target DNA.

[00373] Embodiment 68. The method of any of the embodiments of 58 to 67, wherein the contacting occurs under conditions that are permissive for non-homologous end joining or homology-directed repair.

[00374] Embodiment 69. The method of any of the embodiments of 58 to 67, wherein the contacting the target DNA with a donor polynucleotide, wherein the donor polynucleotide, a portion of the donor polynucleotide, a copy of the donor polynucleotide, or a portion of a copy of the donor polynucleotide integrates into the target DNA.

[00375] Embodiment 70. The method of any of the embodiments of 58 to 67, wherein the method does not comprise contacting the cell with a donor polynucleotide, or wherein the target DNA is modified such that nucleotides within the target DNA are deleted.

[00376] Embodiment 71. A method of detecting a target DNA in a sample, the method comprising:

a. contacting the sample with:

i. a Cas12a.1, Cas12p, or Cas12q protein;

ii. a Cas12a.1, Cas12p, or Cas12q gRNA comprising a spacer sequence that is capable of hybridizing with a target sequence in a target DNA; and

iii. a labeled detector that does not hybridize with the spacer sequence of the gRNA; and

b. measuring a detectable signal produced by cleavage of the labeled detector by the Cas12a.1, Cas12p, or Cas12q protein, thereby detecting the target DNA.

[00377] Embodiment 72. The method of embodiment 71, wherein the labeled detector comprises a labeled single stranded DNA.

[00378] Embodiment 73. The method of embodiment 71, wherein the labeled detector comprises a labeled RNA.

[00379] Embodiment 74. The method of embodiment 72, wherein the labeled RNA is a single stranded RNA.

[00380] Embodiment 75. The method of embodiment 71, wherein the labeled detector comprises a labeled single stranded DNA/RNA chimera.

[00381] Embodiment 76. The method of any one of embodiments 71 to 75, wherein the labeled detector comprises one or more modified nucleotides.

[00382] Embodiment 77. The method of any one of embodiments 71 to 76, comprising contacting the sample with a precursor gRNA array, wherein the Cas12a.1, Cas12p, or Cas12q protein cleaves the precursor gRNA array to produce said gRNA.

[00383] Embodiment 78. The method of any one of embodiments 71 to 77, wherein the target DNA is single stranded.

[00384] Embodiment 79. The method of any one of embodiments 71 to 78, wherein the target DNA is double stranded.

[00385] Embodiment 80. The method of any one of embodiments 71 to 79, wherein the target DNA is viral DNA, plant DNA, fungal DNA, or bacterial DNA.

[00386] Embodiment 81. The method of embodiment 80, wherein the target sequence is a sequence of a target provided in any of Tables 6a-6f.

[00387] Embodiment 82. The method of embodiment 81, wherein the target is a coronvavirus.

[00388] Embodiment 83. The method of embodiment 82, wherein the target is a SARS-CoV-2 virus.

[00389] Embodiment 84. The method of any one of embodiments 71 to 83, wherein the target DNA is cDNA, and has been obtained by reverse transcription.

[00390] Embodiment 85. The method of any one of embodiments 71 to 79, wherein the target DNA is from a human cell.

[00391] Embodiment 86. The method of embodiment 85, wherein the target DNA is human fetal or cancer cell DNA.

[00392] Embodiment 87. The method of any one of embodiments 71 to 86, wherein the protein is Cas12a.1 comprising the amino acid sequence of SEQ ID NO: 3, or at least 70% sequence identity thereto.

[00393] Embodiment 88. The method of any one of embodiments 71 to 86, wherein the protein is Cas12p comprising the amino acid sequence of SEQ ID NO: 4, or at least 70% sequence identity thereto.

[00394] Embodiment 89. The method of any one of embodiments 71 to 86, wherein the protein is Cas12p comprising the amino acid sequence of SEQ ID NO: 222, or at least 70% sequence identity thereto.

[00395] Embodiment 90. The method of any one of embodiments 71 to 86, wherein the protein is Cas12q comprising the amino acid sequence of SEQ ID NO: 5, or at least 70% sequence identity thereto.

[00396] Embodiment 91. The method of any one of embodiments 71 to 87, wherein the sample comprises DNA from a cell lysate.

[00397] Embodiment 92. The method of any one of embodiments 71 to 87, wherein the sample comprises cells.

[00398] Embodiment 93. The method of any one of embodiments 71 to 87, wherein the sample is a urine, blood, serum, plasma, lymphatic fluid, cerebrospinal fluid, saliva, nasopharyngeal, oropharyngeal, nasopharyngeal/ oropharyngeal, aspirate, or biopsy sample.

[00399] Embodiment 94. The method of any one of embodiments 71 to 93, comprising determining an amount of the target DNA present in the sample.

[00400] Embodiment 95. The method of embodiment 94, wherein said measuring a detectable signal comprises one or more of: visual based detection, sensor based detection, color detection,

gold nanoparticle based detection, fluorescence polarization, colloid phase transition/dispersion, electrochemical detection, and semiconductor-based sensing.

[00401] Embodiment 96. The method of any one of embodiments 71 to 95, wherein the labeled detector comprises a modified nucleobase, a modified sugar moiety, and/or a modified nucleic acid linkage.

[00402] Embodiment 97. The method of any one of embodiments 71 to 96, further comprising detecting a positive control target DNA in a positive control sample, the detecting comprising:

a. contacting the positive control sample with:

i. the Cas12a.1, Cas12p, or Cas12q protein;

ii. a positive control gRNA comprising: a region that binds to the Cas12a.1, Cas12p, or Cas12q protein, and a positive control spacer sequence that hybridizes with the positive control target DNA; and

iii. a labeled detector that does not hybridize with the positive control spacer sequence of the positive control gRNA; and

b. measuring a detectable signal produced by cleavage of the labeled detector by the Cas12a.1, Cas12p, or Cas12q protein, thereby detecting the positive control target DNA.

[00403] Embodiment 98. The method of any one of embodiments 71 to 97, wherein the detectable signal is detectable in less than 15, 30, 45, 60, 90, 120, 150, 180, 210, or 240 minutes.

[00404] Embodiment 99. The method of any one of embodiments 71 to 98, further comprising amplifying the target DNA in the sample by loop-mediated isothermal amplification (LAMP), helicase-dependent amplification (HDA), recombinase polymerase amplification (RPA), strand displacement amplification (SDA), nucleic acid sequence-based amplification (NASBA), transcription mediated amplification (TMA), nicking enzyme amplification reaction (NEAR), rolling circle amplification (RCA), multiple displacement amplification (MDA), Ramification (RAM), circular helicase-dependent amplification (cHDA), single primer isothermal amplification (SPIA), signal mediated amplification of RNA technology (SMART), self-sustained sequence replication (3SR), genome exponential amplification reaction (GEAR), or isothermal multiple displacement amplification (IMDA).

[00405] Embodiment 100. The method of any one of embodiments 71 to 99, wherein target DNA in the sample is present at a concentration of less than 100 uM.

[00406] Embodiment 101. A protein comprising an amino acid sequence with 70%-99.5% homology to SEQ ID NO: 1, 2, 3, 4, 5, 10, 11, or 222.

[00407] Embodiment 102. A protein of embodiment 101, wherein the sequence of the protein has been deduced bioinformatically.

[00408] Embodiment 103. A composition comprising any of the proteins of embodiment 101, and optionally a pharmaceutically acceptable carrier.

[00409] Embodiment 104. A composition comprising any of the proteins of embodiment 101, optionally comprising a pharamaceutically aceptable carrier, a nucleic acid stabilizing buffer and/or or a protein stabilizing buffer.

[00410] Embodiment 105. A composition comprising any of the proteins of embodiment 101, wherein the protein is lyopholized, and optionally further comprises any one or more of a labeled detector, a reverse transcriptase enzyme, and reagents for loop-mediated isothermal amplification.

[00411] Embodiment 106. A DNA polynucleotide comprising a nucleotide sequence that encodes any of the proteins of embodiment 101.

[00412] Embodiment 107. A recombinant expression vector comprising the DNA polynucleotide of embodiment 106.

[00413] Embodiment 108. The recombinant expression vector of embodiment 107, wherein the nucleotide sequence encoding the single protein is operably linked to a promoter.

[00414] Embodiment 109. A host cell comprising the DNA polynucleotide of any one of embodiments 106 to 108.

[00415] Embodiment 110. A pharmaceutical composition comprising any of the engineered systems of embodiments 1 to 39, and optionally a pharmaceutically acceptable carrier.

[00416] Embodiment 111. A composition comprising any of the engineered systems of embodiments 1 to 39, and optionally comprising a nucleic acid stabilizing buffer and/or or a protein stabilizing buffer.

[00417] Embodiment 112. A pharmaceutical composition comprising any of the single molecule gRNAs of embodiments 40 to 57, and optionally pharmaceutically acceptable carrier.

[00418] Embodiment 113. A composition comprising any of the singe molecule gRNAs of embodiments 40 to 51, and optionally a nucleic acid stabilizing buffer and/or or a protein stabilizing buffer.

[00419] Embodiment 114. A DNA polynucleotide comprising a nucleotide sequence that encodes any of the nucleic acids of embodiments 3, 27, or the gRNAs of embodiments 40 to 51.

[00420] Embodiment 115. A recombinant expression vector comprising the DNA polynucleotide of embodiment 114.

[00421] Embodiment 116. The recombinant expression vector of embodiment 115, wherein the nucleotide sequence encoding the single gRNA is operably linked to a promoter.

[00422] Embodiment 117. A host cell comprising the DNA polynucleotide of any one of embodiments 114 to 116.

[00423] Embodiment 118. A kit comprising one or more components of any of the engineered systems of embodiments 1 to 39.

[00424] Embodiment 119. The kit of embodiment 118, wherein one or more components are lyopholized.

[00425] Embodiment 120. The kit of any one of embodiments 118 to 119, wherein the one or more components comprise Cas12p, a labeled RNA reporter, and a gRNA directed to SARS-CoV-2.

[00426] Embodiment 121. A method of isolating a Class 2 Type II or Class 2 Type V CRISPR-Cas protein from a metagenomics sample comprising the use of a bioinformatics-based method.

[00427] Embodiment 122. The method of embodiment 121, wherein the Class 2 Type II or Class 2 Type V CRISPR-Cas protein is selected from the group consisting of SEQ ID NO: 1, 2, 3, 4, 5, 10, 11, and 222.

EXAMPLES

[00428] The following examples are included for illustrative purposes and are not intend to limit the scope of the invention.

Example 1: Identification and Validation of Novel Class II type II and V Endonucleases

Class 2 type II and type V CRISPR-Cas loci identification

[00429] Metagenome sequences were obtained from NCBI, and compiled to construct a database of putative CRISPR-Cas loci. CRISPR arrays were identified using CrisprCasFinder software. The criteria of filtering were putative Class II type II and V effectors > 500 aa, which were adjacent to cas genes and CRISPR arrays. Sequences were aligned with Clustal Omega using HMM profiles. The novel Cas9.1, Cas9.2, Cas9.3, Cas9.4, Cas12a.1, Cas12p and Cas12q proteins described herein were identified.

Generation of expression plasmids and non-coding elements

[00430] Minimal conditions to validate the Cas proteins were established into a cloning strategy. Minimal CRISPR loci were designed by removing acquisition proteins and generating minimal arrays with a single spacer (Sp1). The natural Sp1 sequence was replaced by a known specific target sequence with the length of the naturally occurring sequence (GTGGCAGCTCAAAAATTGGCTACAAAACCAGTT; SEQ ID NO: 118) for target detection and PAM screening assays. The E. coli codon-optimized protein sequences of CRISPR effectors and/or accessory proteins were placed under the transcriptional control of lac and IPTG-inducible T7 promoters into a pET-based expression vector (EMD-Millipore).

Artificial Synthesis

[00431] For Cas12a.1, Cas12p, Cas9.1 and Cas9.2, expression vectors were artificially synthesized. The effector plasmid codon optimization, synthesis, and cloning were generated by a provider (GeneScript). To consider both putative transcription directions, flanking restriction sites were added in the CRISPR array to clone a DNA fragment (IDT). This was done with the same element in the opposite direction to create a second construct variant. FIGS. 1A-1B show expression vector maps for Cas9.1 and Cas9.2. FIGS. 2A-2C show expression vector maps for Cas12a.1, Cas12p, and Cas12q. Vector sequences are provided in Table 8.

Table 8. Expression vector sequences

Protein	Vector Sequence
Cas12a.1	TGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGT GTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGC GCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGC TTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTT AGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGG TTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGAC GTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAA CAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTT TGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAA TTTAACGCGAATTTTAACAAAATATTAACGTTTACAATTTCAGGTGG CACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTA AATACATTCAAATATGTATCCGCTCATGAATTAATTCTTAGAAAAAC TCATCGAGCATCAAATGAAACTGCAATTTATTCATATCAGGATTATC AATACCATATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAAACT CACCGAGGCAGTTCCATAGGATGGCAAGATCCTGGTATCGGTCTGCG ATTCCGACTCGTCCAACATCAATACAACCTATTAATTTCCCCTCGTCA AAAATAAGGTTATCAAGTGAGAAATCACCATGAGTGACGACTGAAT

Protein	Vector Sequence
	CCGGTGAGAATGGCAAAAGTTTATGCATTTCTTTCCAGACTTGTTCA ACAGGCCAGCCATTACGCTCGTCATCAAAATCACTCGCATCAACCAA ACCGTTATTCATTCGTGATTGCGCCTGAGCGAGACGAAATACGCGAT CGCTGTTAAAAGGACAATTACAAACAGGAATCGAATGCAACCGGCG CAGGAACACTGCCAGCGCATCAACAATATTTTCACCTGAATCAGGAT ATTCTTCTAATACCTGGAATGCTGTTTTCCCGGGGATCGCAGTGGTG AGTAACCATGCATCATCAGGAGTACGGATAAAATGCTTGATGGTCG GAAGAGGCATAAATTCCGTCAGCCAGTTTAGTCTGACCATCTCATCT GTAACATCATTGGCAACGCTACCTTTGCCATGTTTCAGAAACAACTC TGGCGCATCGGGCTTCCCATACAATCGATAGATTGTCGCACCTGATT GCCCGACATTATCGCGAGCCCATTTATACCCATATAAATCAGCATCC ATGTTGGAATTTAATCGCGGCCTAGAGCAAGACGTTTCCCGTTGAAT ATGGCTCATAACACCCCTTGTATTACTGTTTATGTAAGCAGACAGTTT TATTGTTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAG CGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTT TTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACC AGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGA AGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTA GTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCC TACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTG GCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCG GATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGC CCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCG TGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGAC AGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGG GAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTT TCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGG GGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTC CTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCC CTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACC GCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGG AAGCGGAAGAGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGC GGTATTTCACACCGCATATATGGTGCACTCTCAGTACAATCTGCTCT GATGCCGCATAGTTAAGCCAGTATACACTCCGCTATCGCTACGTGAC TGGGTCATGGCTGCGCCCCGACACCCGCCAACACCCGCTGACGCGCC CTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGA CCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCG AAACGCGCGAGGCAGCTGCGGTAAAGCTCATCAGCGTGGTCGTGAA GCGATTCACAGATGTCTGCCTGTTCATCCGCGTCCAGCTCGTTGAGTT TCTCCAGAAGCGTTAATGTCTGGCTTCTGATAAAGCGGGCCATGTTA AGGGCGGTTTTTTCCTGTTTGGTCACTGATGCCTCCGTGTAAGGGGG ATTTCTGTTCATGGGGGTAATGATACCGATGAAACGAGAGAGGATG CTCACGATACGGGTTACTGATGATGAACATGCCCGGTTACTGGAACG TTGTGAGGGTAAACAACTGGCGGTATGGATGCGGCGGGACCAGAGA AAAATCACTCAGGGTCAATGCCAGCGCTTCGTTAATACAGATGTAGG

Protein	Vector Sequence
	TGTTCCACAGGGTAGCCAGCAGCATCCTGCGATGCAGATCCGGAAC ATAATGGTGCAGGGCGCTGACTTCCGCGTTTCCAGACTTTACGAAAC ACGGAAACCGAAGACCATTCATGTTGTTGCTCAGGTCGCAGACGTTT TGCAGCAGCAGTCGCTTCACGTTCGCTCGCGTATCGGTGATTCATTCT GCTAACCAGTAAGGCAACCCCGCCAGCCTAGCCGGGTCCTCAACGA CAGGAGCACGATCATGCGCACCCGTGGGGCCGCCATGCCGGCGATA ATGGCCTGCTTCTCGCCGAAACGTTTGGTGGCGGGACCAGTGACGAA GGCTTGAGCGAGGGCGTGCAAGATTCCGAATACCGCAAGCGACAGG CCGATCATCGTCGCGCTCCAGCGAAAGCGGTCCTCGCCGAAAATGAC CCAGAGCGCTGCCGGCACCTGTCCTACGAGTTGCATGATAAAGAAG ACAGTCATAAGTGCGGCGACGATAGTCATGCCCCGCGCCCACCGGA AGGAGCTGACTGGGTTGAAGGCTCTCAAGGGCATCGGTCGAGATCC CGGTGCCTAATGAGTGAGCTAACTTACATTAATTGCGTTGCGCTCAC TGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGA ATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCCAGG GTGGTTTTTCTTTTCACCAGTGAGACGGGCAACAGCTGATTGCCCTTC ACCGCCTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCACGCTGGTTTG CCCCAGCAGGCGAAAATCCTGTTTGATGGTGGTTAACGGCGGGATAT AACATGAGCTGTCTTCGGTATCGTCGTATCCCACTACCGAGATATCC GCACCAACGCGCAGCCCGGACTCGGTAATGGCGCGCATTGCGCCCA GCGCCATCTGATCGTTGGCAACCAGCATCGCAGTGGGAACGATGCCC TCATTCAGCATTTGCATGGTTTGTTGAAAACCGGACATGGCACTCCA GTCGCCTTCCCGTTCCGCTATCGGCTGAATTTGATTGCGAGTGAGAT ATTTATGCCAGCCAGCCAGACGCAGACGCGCCGAGACAGAACTTAA TGGGCCCGCTAACAGCGCGATTTGCTGGTGACCCAATGCGACCAGAT GCTCCACGCCCAGTCGCGTACCGTCTTCATGGGAGAAAATAATACTG TTGATGGGTGTCTGGTCAGAGACATCAAGAAATAACGCCGGAACAT TAGTGCAGGCAGCTTCCACAGCAATGGCATCCTGGTCATCCAGCGGA TAGTTAATGATCAGCCCACTGACGCGTTGCGCGAGAAGATTGTGCAC CGCCGCTTTACAGGCTTCGACGCCGCTTCGTTCTACCATCGACACCA CCACGCTGGCACCCAGTTGATCGGCGCGAGATTTAATCGCCGCGACA ATTTGCGACGGCGCGTGCAGGGCCAGACTGGAGGTGGCAACGCCAA TCAGCAACGACTGTTTGCCCGCCAGTTGTTGTGCCACGCGGTTGGGA ATGTAATTCAGCTCCGCCATCGCCGCTTCCACTTTTTCCCGCGTTTTC GCAGAAACGTGGCTGGCCTGGTTCACCACGCGGGAAACGGTCTGAT AAGAGACACCGGCATACTCTGCGACATCGTATAACGTTACTGGTTTC ACATTCACCACCCTGAATTGACTCTCTTCCGGGCGCTATCATGCCAT ACCGCGAAAGGTTTTGCGCCATTCGATGGTGTCCGGGATCTCGACGC TCTCCCTTATGCGACTCCTGCATTAGGAAGCAGCCCAGTAGTAGGTT GAGGCCGTTGAGCACCGCCGCCGCAAGGAATGGTGCATGCAAGGAG ATGGCGCCCAACAGTCCCCCGGCCACGGGGCCTGCCACCATACCCAC GCCGAAACAAGCGCTCATGAGCCCGAAGTGGCGAGCCCGATCTTCC CCATCGGTGATGTCGGCGATATAGGCGCCAGCAACCGCACCTGTGGC GCCGGTGATGCCGGCCACGATGCGTCCGGCGTAGAGGATCGAGATC TCGATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCG

Protein	Vector Sequence
	GATAACAATTCCCCTCTAGATTTACACTTTATGCTTCCGGCTCGTATG TTTCTAGAGAATTCAAATAATTTTGTTTAACTTTAAGAAGGAGATTT AAATATGGGGTCAAGTCATCACCACCACCACCACTCAAGTGGACTA GTACCCCGTGGCAGCATGAAAGTTAGCACCTGGGATAGCTTCACCAA CCAGTACCCGCTGACCAAGACCCTGCGTTTTGAGCTGAAGCCGGTGG GTAAAACCCTGCAGAAGATCCAAGACCGTAACCTGATTACCGAGGA CGAACAGCGTCAAAAGGATTTCAACAAGGTTAAGAAAATCATGGAT GGTTACTACAAGCAGTTCATCGAGGAATGCCTGGAAGGCGCGAAGA TCCCGCTGAAGAAACTGGAGGAAAACAACAACGCGTACACCAAACT GAAGAAAGACCCGTATAACAAGAAACTGCGTGAGGAATACGCGAAG CTGCAGAAACAACTGCGTAAACTGATCCACGATGAGATTAACAAGA AAGAGGAATTCAAGTACCTGTTTAAGAAAGAATTCATCAAGAAAAT TCTGCCGGAATGGCTGGAGAAGAAAGGTAAGAAAGAGGAACTGAA AGAGATCGAAAAGTTCGACAAATGGGTGACCTACTTTAGCGGCTTCT TTAACAACCGTAAGAACGTTTTCAGCAGCGACGAGATTAGCACCAG CATGATCTATCGTATTGTGAACGATAACCTGCCGAAATTCCTGGACG ATGTTAGCCGTTTTGGTGAAATTACCCGTTACAAGGAGTTCGACGCG AACCAGATCGAGGAAAACTTTGAGAGCGAACTGAACGGTGAAAAAC TGAAGGATTTCTTTAACCTGAAAAACTTCAACAACTGCCTGAACCAA GAAGGCATTGAGAAATTTAACCTGATCATTGGTGGCAAGAGCGAGG AAGGTAATAACAAAATCAAGGGCCTGAACGAACTGGTGAACGAGCT GGCGCAAAAACAAGCGGACAAGAACGAGCAGAAGAAAGTTCGTAA ACTGAAGCTGGCGCCGCTGTTCAAACAAATCCTGAGCGATCGTAAG AGCAGCAGCTTTGCGTTCGAAAAATTTGAGGAAAACACCGAGGTGT TCGACGCGATCGATGAATTTTATGACAAGATTAGCCTGGAGACCCTG AAGAAAATCGAAGCGACCCTGGAGAAACTGGAGGAAAAGGACCTG GAACTGGTTTACCTGAAAAACGATCGTTGCCTGACCGGTATCAGCCA GGAAGTGTTCGGCGACCGTGAGCGTGTTCTGCAAGCGCTGCGTGAAT ACGCGAAAACCGAGCTGGGTCTGAAGACCGATAAGAAAATCGAGAA GTGGATGAAGAAAGGTCGTTATAGCATCCACGAGATTGAAAGCGGC CTGAAGAAAATCGGTAGCACCGGCCACCCGATTTGCAACTACTTCAG CAAACTGGAGGAAAAGAAAACCAACCTGATCCAGGAAATTAAGAAA GCGCGTACCGAGTATGAAAAGATCAGCGACAAGAAAAAGAAACTGA CCGCGGAAAGCCAAGAGCCGAACGTGGCGCGTATCAAAGCGCTGCT GGATAGCATTATGCGTCTGTATCACTTCATCAAGCCGCTGAACATCA ACTTCAAGAACAAGAAAGAGAAGGACAGCGAAGCGCTGGAGACCG ACAACGATTTTTACAACGACTTCGATGAAAGCTTTGCGGAGCTGGGC AACATCATTCCGCTGTACAACCAAGTGCGTAACTATGTTACCCAAAA ACCGTTCAGCACCGAGAAATTCAAGCTGAACTTTGAAAACCCGAAG CTGCTGAGCGGTTGGGACAAAAACAAGGAAAAAGATTACTATAGCG TGATTCTGCGTAAAGAGGAAAGCTACTATCTGGCGATCATGACCCCG AAGCAGAAAAACGTTTTCGACGAGCTGGAACGTCTGCCGGCGGGCA AAAATTACTTCGAGAAGATCGATTACAAGCTGCTGCCGACCCCGGA AAAGAACCTGCCGCGTATCCTGTTCGCGAAGAAAAACATTAGCTTTT ACAAGCCGAGCAAAGAGATCGAAGCGATTCGTAACCACAGCGCGCA

Protein	Vector Sequence
	CACCAAACACGGTAACCCGCAGAACGGCTTCAAGAAACGTGACTTT CGTCTGAGCGATTGCCACAAGATGATCGACTTCTACAAGAAAAGCAT TCAGAAACACCCGGAATGGAAGGAGTATGATTTTCAATTCAAGAAA ACCGAGGACTACGTGGATATCAGCGAATTCTATAAAGAGGTTTCTGA CCAGGGTTACAAGATCGAATTCAAGAAAATTAGCGAGAAATACCTG CTGGACCTGGTGGAGGAAGGTAAACTGTACCTGTTCCAAATCTGGAA CAAGGATTTCAGCAAGTACAGCGAAGGCCGTAAAAACCTGCACACC ATCTATTGGAAAGAACTGTTCAGCAAGGAGAACCTGAGCGATATTA CCTATAAGCTGAACGGCGAGGCGGAAATCTTTTACCGTCCGAAAAG CATGGAGCGTAAGGTTACCCACCCGAAGAACCAGAAAATCGAAAAC AAAGACCCGATCAAGGGTAAGAAATTCAGCAAGTTCAAGTATGACT TCATCAAGAACAAGCGTTACACCGAGGATCGTTTCTTTTTCCACTGC CCGATCACCCTGAACTTTCAAGCGCGTGACGGCAGCAAAACCATCA ACAAGCGTGTGAACGATCACATTCGTGAGACCAAAGACGATATCTTC GTTCTGAGCATTGATCGTGGTGAACGTCACCTGGCGTACTATACCCT GCTGAACAGCAAGGGTGAAATTCAGGAGCAAGGCAGCTTTAACGTG ATCAGCGACGATAAGGAGCGTAAACGTGACTATCACGAAAAACTGG ATGAGCGTGAAAAGGAGCGTGACAAGGCGCGTAAAAGCTGGCAGA AAATCGAGACCATTAAGAAACTGAAGGATGGCTACCTGAGCCAAAT CGTGCACAAGATTGCGAAACTGGCGATCGAGAAAAACGCGATCATT GTTCTGGAAGACCTGAACCTGGATTTCAAGCGTGGTCGTCTGAAGAT TGAGAAACAGGTGTACCAAAAATTCGAAAAGAAACTGATCGACAAG CTGAACTATCTGGTTTTTAAAGAACGTACCGAAAAAGAGGCGGGTG GTAGCCTGAACGCGTATCAGCTGACCGGTAAATTCGAGGGCTTTAAG AAACTGGGCAAGGAAACCGGCATCATTTACTATGTGCCGGCGGCGT ACACCAGCAAAATCTGCCCGAAGACCGGCTTCGTTAACCTGCTGCGT CCGAAGTTCAAGAACATCGAAAAGGCGAAGGAGTTTTTCAAGAAGT TCAACTACATCAAGTACGACAGCAGCGAAGGTCTGTTTGAGTTCAAC TTCGATTACAGCAAGTTCATCAAGAACGGCAAGAAAGAGACCAAAA TCATTCAGGACAACTGGAGCGTGTATAGCAACGGTACCAAGCTGGTT GGCTTCCGTAACAAGAACAAAAACAACAGCTGGGATACCAAGGAAG TGAAACCGAACGAGAAGCTGAAAATTCTGTTCAAAGAGTACGGTGT TAGCTTTCAAAAGGACGAAAACATCATTAGCCAGATCGCGAGCCAA AACAAGAAAGCGTTTTTCGAGAACCTGATCAAGATTTTCAAAACCAT TCTGATGCTGCGTAACAGCCGTAAAGACCCGGAGGAAGATTACGTG CTGAGCTGCGTTAAGGACGAAAACGGCGAGTTTTTCGACAGCCGTA AGGCGAAAGATAACGAGCCGAAAGACGCGGATGCGAACGGCGCGT ACCACATTGGTCTGAAGGGCCTGATGCTGCTGGAACGTATCAAGGCG AACAAAGGTAAGAAAAAGCTGGACCTGCTGATCAGCCGTAACGATT TCATTAACTTTGCGGTTGAGCGTAGCAAGTAATAAGGATCCCTCGAG TTGACAGCTAGCTCAGTCCTAGGTATAATGCTAGCGTTTAAGGCCTT GACAAAATTTCTACTGTAGTAGATGTGGCAGCTCAAAAATTGGCTAC AAAACGTTTAAGGCCTTGACAAAATTTCTACTGTAGTAGATCTAGCA TAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCATAT GCTGAAAGGAGGAACTATATCCGGAT **(SEQ ID NO: 64)**

Protein	Vector Sequence
Cas12p	TGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGT GTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGC GCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGC TTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTT AGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGG TTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGAC GTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAA CAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTT TGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAA TTTAACGCGAATTTTAACAAAATATTAACGTTTACAATTTCAGGTGG CACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTA AATACATTCAAATATGTATCCGCTCATGAATTAATTCTTAGAAAAAC TCATCGAGCATCAAATGAAACTGCAATTTATTCATATCAGGATTATC AATACCATATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAAACT CACCGAGGCAGTTCCATAGGATGGCAAGATCCTGGTATCGGTCTGCG ATTCCGACTCGTCCAACATCAATACAACCTATTAATTTCCCCTCGTCA AAAATAAGGTTATCAAGTGAGAAATCACCATGAGTGACGACTGAAT CCGGTGAGAATGGCAAAAGTTTATGCATTTCTTTCCAGACTTGTTCA ACAGGCCAGCCATTACGCTCGTCATCAAAATCACTCGCATCAACCAA ACCGTTATTCATTCGTGATTGCGCCTGAGCGAGACGAAATACGCGAT CGCTGTTAAAAGGACAATTACAAACAGGAATCGAATGCAACCGGCG CAGGAACACTGCCAGCGCATCAACAATATTTTCACCTGAATCAGGAT ATTCTTCTAATACCTGGAATGCTGTTTTCCCGGGGATCGCAGTGGTG AGTAACCATGCATCATCAGGAGTACGGATAAAATGCTTGATGGTCG GAAGAGGCATAAATTCCGTCAGCCAGTTTAGTCTGACCATCTCATCT GTAACATCATTGGCAACGCTACCTTTGCCATGTTTCAGAAACAACTC TGGCGCATCGGGCTTCCCATACAATCGATAGATTGTCGCACCTGATT GCCCGACATTATCGCGAGCCCATTTATACCCATATAAATCAGCATCC ATGTTGGAATTTAATCGCGGCCTAGAGCAAGACGTTTCCCGTTGAAT ATGGCTCATAACACCCCTTGTATTACTGTTTATGTAAGCAGACAGTTT TATTGTTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAG CGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTT TTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACC AGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGA AGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTA GTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCC TACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTG GCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCG GATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGC CCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCG TGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGAC AGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGG GAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTT TCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGG GGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTC

Protein	Vector Sequence
	CTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCC CTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACC GCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGG AAGCGGAAGAGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGC GGTATTTCACACCGCATATATGGTGCACTCTCAGTACAATCTGCTCT GATGCCGCATAGTTAAGCCAGTATACACTCCGCTATCGCTACGTGAC TGGGTCATGGCTGCGCCCCGACACCCGCCAACACCCGCTGACGCGCC CTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGA CCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCG AAACGCGCGAGGCAGCTGCGGTAAAGCTCATCAGCGTGGTCGTGAA GCGATTCACAGATGTCTGCCTGTTCATCCGCGTCCAGCTCGTTGAGTT TCTCCAGAAGCGTTAATGTCTGGCTTCTGATAAAGCGGGCCATGTTA AGGGCGGTTTTTTCCTGTTTGGTCACTGATGCCTCCGTGTAAGGGGG ATTTCTGTTCATGGGGGTAATGATACCGATGAAACGAGAGAGGATG CTCACGATACGGGTTACTGATGATGAACATGCCCGGTTACTGGAACG TTGTGAGGGTAAACAACTGGCGGTATGGATGCGGCGGGACCAGAGA AAAATCACTCAGGGTCAATGCCAGCGCTTCGTTAATACAGATGTAGG TGTTCCACAGGGTAGCCAGCAGCATCCTGCGATGCAGATCCGGAAC ATAATGGTGCAGGGCGCTGACTTCCGCGTTTCCAGACTTTACGAAAC ACGGAAACCGAAGACCATTCATGTTGTTGCTCAGGTCGCAGACGTTT TGCAGCAGCAGTCGCTTCACGTTCGCTCGCGTATCGGTGATTCATTCT GCTAACCAGTAAGGCAACCCCGCCAGCCTAGCCGGGTCCTCAACGA CAGGAGCACGATCATGCGCACCCGTGGGGCCGCCATGCCGGCGATA ATGGCCTGCTTCTCGCCGAAACGTTTGGTGGCGGGACCAGTGACGAA GGCTTGAGCGAGGGCGTGCAAGATTCCGAATACCGCAAGCGACAGG CCGATCATCGTCGCGCTCCAGCGAAAGCGGTCCTCGCCGAAAATGAC CCAGAGCGCTGCCGGCACCTGTCCTACGAGTTGCATGATAAAGAAG ACAGTCATAAGTGCGGCGACGATAGTCATGCCCCGCGCCCACCGGA AGGAGCTGACTGGGTTGAAGGCTCTCAAGGGCATCGGTCGAGATCC CGGTGCCTAATGAGTGAGCTAACTTACATTAATTGCGTTGCGCTCAC TGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGA ATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCCAGG GTGGTTTTTCTTTTCACCAGTGAGACGGGCAACAGCTGATTGCCCTTC ACCGCCTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCACGCTGGTTTG CCCCAGCAGGCGAAAATCCTGTTTGATGGTGGTTAACGGCGGGATAT AACATGAGCTGTCTTCGGTATCGTCGTATCCCACTACCGAGATATCC GCACCAACGCGCAGCCCGGACTCGGTAATGGCGCGCATTGCGCCCA GCGCCATCTGATCGTTGGCAACCAGCATCGCAGTGGGAACGATGCCC TCATTCAGCATTTGCATGGTTTGTTGAAAACCGGACATGGCACTCCA GTCGCCTTCCCGTTCCGCTATCGGCTGAATTTGATTGCGAGTGAGAT ATTTATGCCAGCCAGCCAGACGCAGACGCGCCGAGACAGAACTTAA TGGGCCCGCTAACAGCGCGATTTGCTGGTGACCCAATGCGACCAGAT GCTCCACGCCCAGTCGCGTACCGTCTTCATGGGAGAAAATAATACTG TTGATGGGTGTCTGGTCAGAGACATCAAGAAATAACGCCGGAACAT TAGTGCAGGCAGCTTCCACAGCAATGGCATCCTGGTCATCCAGCGGA

Protein	Vector Sequence
	TAGTTAATGATCAGCCCACTGACGCGTTGCGCGAGAAGATTGTGCAC CGCCGCTTTACAGGCTTCGACGCCGCTTCGTTCTACCATCGACACCA CCACGCTGGCACCCAGTTGATCGGCGCGAGATTTAATCGCCGCGACA ATTTGCGACGGCGCGTGCAGGGCCAGACTGGAGGTGGCAACGCCAA TCAGCAACGACTGTTTGCCCGCCAGTTGTTGTGCCACGCGGTTGGGA ATGTAATTCAGCTCCGCCATCGCCGCTTCCACTTTTTCCCGCGTTTTC GCAGAAACGTGGCTGGCCTGGTTCACCACGCGGGAAACGGTCTGAT AAGAGACACCGGCATACTCTGCGACATCGTATAACGTTACTGGTTTC ACATTCACCACCCTGAATTGACTCTCTTCCGGGCGCTATCATGCCAT ACCGCGAAAGGTTTTGCGCCATTCGATGGTGTCCGGGATCTCGACGC TCTCCCTTATGCGACTCCTGCATTAGGAAGCAGCCCAGTAGTAGGTT GAGGCCGTTGAGCACCGCCGCCGCAAGGAATGGTGCATGCAAGGAG ATGGCGCCCAACAGTCCCCCGGCCACGGGGCCTGCCACCATACCCAC GCCGAAACAAGCGCTCATGAGCCCGAAGTGGCGAGCCCGATCTTCC CCATCGGTGATGTCGGCGATATAGGCGCCAGCAACCGCACCTGTGGC GCCGGTGATGCCGGCCACGATGCGTCCGGCGTAGAGGATCGAGATC TCGATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCG GATAACAATTCCCCTCTAGATTTACACTTTATGCTTCCGGCTCGTATG TTTCTAGAGAATTCAAATAATTTTGTTTAACTTTAAGAAGGAGATTT AAATATGGGATCAAGTCATCACCACCACCACCACTCAAGTGGACTA GTACCCAGGGGAAGCATGAAGAAGAGCATTTTCGATCAGTTCGTTA ACCAGTACGCGCTGAGCAAGACCCTGCGTTTCGAGCTGAAACCGGT GGGTGAAACCGGCCGTATGCTGGAGGAAGCGAAGGTTTTCGCGAAG GATGAAACCATTAAGAAAAAGTACGAAGCGACCAAGCCGTTCTTTA ACAAACTGCACCGTGAATTCGTGGAGGAAGCGCTGAACGAGGTTGA ACTGGCGGGCCTGCCGGAGTACTTCGAAATCTTCAAGTACTGGAAGC GTTACAAAAAGAAATTCGAGAAGGACCTGCAGAAGAAAGAGAAGG AACTGCGTAAAAGCGTGGTTGGTTTCTTTAACGCGCAAGCGAAGGA GTGGGCGAAGAAATATGAAACCCTGGGCGTGAAGAAAAAGGATGTT GGTCTGCTGTTCGAGGAAAACGTGTTTGCGATTCTGAAAGAACGTTA CGGTAACGAGGAAGGCAGCCAGATTGTGGACGAGAGCACCGGCAAG GATGTTAGCATCTTCGACAGCTGGAAGGGTTTTACCGGCTATTTCAT CAAATTTCAGGAAACCCGTAAGAACTTCTACAAAGATGATGGTACC GCGACCGCGCTGGCGACCCGTATCATTGATCAAAACCTGAAACGTTT CTGCGACAACCTGCTGATCTTTGAGAGCATTCGTGATAAGATCGACT TCAGCGAGGTTGAACAGACCATGGGCAACAGCATCGATAAGGTGTT CAGCGTTATCTTTTATAGCAGCTGCCTGCTGCAAGAAGGTATCGACT TTTACAACTGCGTGCTGGGTGGTGAAACCCTGCCGAACGGTGAAAA GCGTCAGGGCATTAACGAACTGATCAACCTGTACCGTCAAAAGACC AGCGAGAAAGTTCCGTTCCTGAAGCTGCTGGACAAACAGATTCTGA GCGAGAAGGAAAAATTTATGGATGAGATCGAAAACGACGAGGCGCT GCTGGATACCCTGAAGATTTTCCGTAAAAGCGCGGAGGAAAAGACC ACCCTGCTGAAAAACATCTTCGGCGATTTTGTGATGAACCAGGGTAA ATATGACCTGGCGCAAATCTACATTAGCCGTGAAAGCCTGAACACCA TTAGCCGTAAGTGGACCAGCGAAACCGATATCTTCGAAGACAGCCT

Protein	Vector Sequence
	GTACGAGGTGCTGAAAAAGAGCAAAATCGTGAGCGCGAGCGTTAAA AAGAAAGACGGTGGCTACGCGTTCCCGGAGTTTATCGCGCTGATTTA TGTTAAAAGCGCGCTGGAACAGATTCCGACCGAGAAGTTCTGGAAA GAACGTTACTATAAGAACATCGGCGATGTGCTGAACAAGGGTTTCCT GAACGGTAAAGAAGGCGTTTGGCTGCAATTTCTGCTGATCTTTGACT TCGAATTTAACAGCCTGTTCGAGCGTGAAATCATTGATGAGAACGGC GACAAGAAAGTGGCGGGTTATAACCTGTTCGCGAAGGGTTTTGACG ATCTGCTGAACAACTTCAAATACGACCAGAAGGCGAAAGTGGTTATT AAGGATTTTGCGGACGAAGTTCTGCACATTTATCAAATGGGCAAATA CTTCGCGATCGAGAAGAAACGTAGCTGGCTGGCGGACTATGATATTG ACAGCTTCTACACCGATCCGGAGAAGGGTTACCTGAAATTTTATGAA AACGCGTACGAGGAAATCATTCAGGTTTATAACAAGCTGCGTAACTA CCTGACCAAGAAACCGTATAGCGAGGACAAGTGGAAACTGAACTTC GAAAACCCGACCCTGGCGGATGGTTGGGACAAGAACAAAGAGGCGG ATAACAGCACCGTGATTCTGAAGAAAGACGGTCGTTACTATCTGGGC CTGATGGCGCGTGGTCGTAACAAGCTGTTCGACGATCGTAACCTGCC GAAAATCCTGGAGGGTGTTGAAAACGGCAAGTACGAAAAGGTGGTT TACAAGTACTTCCCGGATCAGGCGAAGATGTTCCCGAAAGTGTGCTT TAGCACCAAAGGCCTGGAATTCTTTCAACCGAGCGAGGAAGTTATCA CCATTTACAAGAACAGCGAGTTCAAGAAAGGTTATACCTTTAACGTG CGTAGCATGCAGCGTCTGATTGATTTCTATAAAGACTGCCTGGTTCG TTACGAAGGTTGGCAATGCTATGATTTTCGTAACCTGCGTAAGACCG AGGACTACCGTAAAAACATCGAGGAATTCTTTAGCGATGTGGCGAT GGACGGCTACAAGATTAGCTTCCAGGACGTTAGCGAGAGCTATATC AAGGAGAAGAACCAAAACGGTGATCTGTACCTGTTTGAGATCAAGA ACAAAGACTGGAACGAAGGTGCGAACGGCAAGAAAAACCTGCACA CCATTTATTTCGAGAGCCTGTTTAGCGCGGATAACATCGCGATGAAC TTCCCGGTGAAACTGAACGGCCAGGCGGAGATCTTTTACCGTCCGCG TACCGAAGGTCTGGAGAAGGAACGTATCATTACCAAGAAAGGCAAC GTTCTGGAAAAGGGTGACAAAGCGTTCCACAAGCGTCGTTACACCG AGAACAAAGTGTTCTTTCACGTTCCGATTACCCTGAACCGTACCAAG AAAAACCCGTTCCAATTTAACGCGAAGATCAACGACTTCCTGGCGAA AAACAGCGATATCAACGTGATTGGTGTTGACCGTGGCGAGAAACAG CTGGCGTATTTTAGCGTGATTAGCCAACGTGGCAAGATCCTGGACCG TGGTAGCCTGAACGTGATCAACGGCGTTAACTACGCGGAGAAGCTG GAGGAAAAAGCGCGTGGTCGTGAACAGGCGCGTAAGGATTGGCAGC AAATCGAGGGCATTAAAGACCTGAAGAAAGGTTATATTAGCCAGGT GGTTCGTAAACTGGCGGATCTGGCGATCCAATACAACGCGATCATTG TGTTCGAGGACCTGAACATGCGTTTTAAGCAAATTCGTGGTGGCATC GAGAAAAGCGTTTATCAGCAACTGGAAAAGGCGCTGATCGATAAAC TGACCTTCCTGGTGGAGAAGGAAGAAAAGGACGTTGAAAAGGCGGG TCACCTGCTGAAAGCGTACCAGCTGGCGGCGCCGTTCGAAACCTTTC AGAAGATGGGTAAACAAACCGGCATTGTGTTTTATACCCAAGCGGC GTACACCAGCCGTATCGATCCGGTTACCGGCTGGCGTCCGCACCTGT ACCTGAAATATAGCAGCGCGGAAAAGGCGAAAGCGGACCTGCTGAA

Protein	Vector Sequence
	GTTCAAGAAAATTAAGTTCGTGGATGGTCGTTTCGAGTTTACCTACG ACATCAAGAGCTTCCGTGAGCAGAAGGAACACCCGAAAGCGACCGT GTGGACCGTTTGCAGCTGCGTTGAGCGTTTTCGTTGGAACCGTTATCT GAACAGCAACAAAGGTGGCTACGATCACTATAGCGACGTGACCAAG TTCCTGGTTGAGCTGTTTCAGGAATACGGCATCGACTTCGAACGTGG TGATATTGTGGGCCAAATCGAGGTTCTGGAAACCAAGGGTAACGAG AAGTTCTTTAAGAACTTCGTGTTCTTTTTCAACCTGATCTGCCAGATT CGTAACACCAACGCGAGCGAACTGGCGAAGAAAGACGGCAAGGAC GATTTCATTCTGAGCCCGGTTGAGCCGTTTTTCGATAGCCGTAACAG CGAGAAGTTCGGCGAAGACCTGCCGAAAAACGGTGACGATAACGGC GCGTTTAACATCGCGCGTAAAGGTCTGGTTATTATGGATAAGATCAC CAAATTCGCGGACGAGAACGGTGGCTGCGAAAAGATGAAATGGGGT GACCTGTATGTGAGCAATGTGGAGTGGGATAACTTTGTGGCGAATAA ATAATAAGGATCCCTCGAGTTGACAGCTAGCTCAGTCCTAGGTATAA TGCTAGCATCTACAAAAGTAGAAATCTAATAGGGATATTCGAGGTG GCAGCTCAAAAATTGGCTACAAAACATCTACAAAAGTAGAAATCTA ATAGGGATATTCGAGCTAGCATAACCCCTTGGGGCCTCTAAACGGGT CTTGAGGGGTTTTTTGCATATGCTGAAAGGAGGAACTATATCCGGAT **(SEQ ID NO: 65)**
Cas12q	TGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGT GTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGC GCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGC TTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTT AGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGG TTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGAC GTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAA CAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTT TGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAA TTTAACGCGAATTTTAACAAAATATTAACGTTTACAATTTCAGGTGG CACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTA AATACATTCAAATATGTATCCGCTCATGAATTAATTCTTAGAAAAAC TCATCGAGCATCAAATGAAACTGCAATTTATTCATATCAGGATTATC AATACCATATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAAACT CACCGAGGCAGTTCCATAGGATGGCAAGATCCTGGTATCGGTCTGCG ATTCCGACTCGTCCAACATCAATACAACCTATTAATTTCCCCTCGTCA AAAATAAGGTTATCAAGTGAGAAATCACCATGAGTGACGACTGAAT CCGGTGAGAATGGCAAAAGTTTATGCATTTCTTTCCAGACTTGTTCA ACAGGCCAGCCATTACGCTCGTCATCAAAATCACTCGCATCAACCAA ACCGTTATTCATTCGTGATTGCGCCTGAGCGAGACGAAATACGCGAT CGCTGTTAAAAGGACAATTACAAACAGGAATCGAATGCAACCGGCG CAGGAACACTGCCAGCGCATCAACAATATTTTCACCTGAATCAGGAT ATTCTTCTAATACCTGGAATGCTGTTTTCCCGGGGATCGCAGTGGTG AGTAACCATGCATCATCAGGAGTACGGATAAAATGCTTGATGGTCG GAAGAGGCATAAATTCCGTCAGCCAGTTTAGTCTGACCATCTCATCT GTAACATCATTGGCAACGCTACCTTTGCCATGTTTCAGAAACAACTC

Protein	Vector Sequence
	TGGCGCATCGGGCTTCCCATACAATCGATAGATTGTCGCACCTGATT GCCCGACATTATCGCGAGCCCATTTATACCCATATAAATCAGCATCC ATGTTGGAATTTAATCGCGGCCTAGAGCAAGACGTTTCCCGTTGAAT ATGGCTCATAACACCCCTTGTATTACTGTTTATGTAAGCAGACAGTTT TATTGTTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAG CGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTT TTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACC AGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGA AGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTA GTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCC TACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTG GCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCG GATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGC CCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCG TGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGAC AGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGG GAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTT TCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGG GGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTC CTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCC CTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACC GCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGG AAGCGGAAGAGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGC GGTATTTCACACCGCATATATGGTGCACTCTCAGTACAATCTGCTCT GATGCCGCATAGTTAAGCCAGTATACACTCCGCTATCGCTACGTGAC TGGGTCATGGCTGCGCCCCGACACCCGCCAACACCCGCTGACGCGCC CTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGA CCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCG AAACGCGCGAGGCAGCTGCGGTAAAGCTCATCAGCGTGGTCGTGAA GCGATTCACAGATGTCTGCCTGTTCATCCGCGTCCAGCTCGTTGAGTT TCTCCAGAAGCGTTAATGTCTGGCTTCTGATAAAGCGGGCCATGTTA AGGGCGGTTTTTTCCTGTTTGGTCACTGATGCCTCCGTGTAAGGGGG ATTTCTGTTCATGGGGGTAATGATACCGATGAAACGAGAGAGGATG CTCACGATACGGGTTACTGATGATGAACATGCCCGGTTACTGGAACG TTGTGAGGGTAAACAACTGGCGGTATGGATGCGGCGGGACCAGAGA AAAATCACTCAGGGTCAATGCCAGCGCTTCGTTAATACAGATGTAGG TGTTCCACAGGGTAGCCAGCAGCATCCTGCGATGCAGATCCGGAAC ATAATGGTGCAGGGCGCTGACTTCCGCGTTTCCAGACTTTACGAAAC ACGGAAACCGAAGACCATTCATGTTGTTGCTCAGGTCGCAGACGTTT TGCAGCAGCAGTCGCTTCACGTTCGCTCGCGTATCGGTGATTCATTCT GCTAACCAGTAAGGCAACCCCGCCAGCCTAGCCGGGTCCTCAACGA CAGGAGCACGATCATGCGCACCCGTGGGGCCGCCATGCCGGCGATA ATGGCCTGCTTCTCGCCGAAACGTTTGGTGGCGGGACCAGTGACGAA GGCTTGAGCGAGGGCGTGCAAGATTCCGAATACCGCAAGCGACAGG CCGATCATCGTCGCGCTCCAGCGAAAGCGGTCCTCGCCGAAAATGAC

Protein	Vector Sequence
	CCAGAGCGCTGCCGGCACCTGTCCTACGAGTTGCATGATAAAGAAG ACAGTCATAAGTGCGGCGACGATAGTCATGCCCCGCGCCCACCGGA AGGAGCTGACTGGGTTGAAGGCTCTCAAGGGCATCGGTCGAGATCC CGGTGCCTAATGAGTGAGCTAACTTACATTAATTGCGTTGCGCTCAC TGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGA ATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCCAGG GTGGTTTTTCTTTTCACCAGTGAGACGGGCAACAGCTGATTGCCCTTC ACCGCCTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCACGCTGGTTTG CCCCAGCAGGCGAAAATCCTGTTTGATGGTGGTTAACGGCGGGATAT AACATGAGCTGTCTTCGGTATCGTCGTATCCCACTACCGAGATATCC GCACCAACGCGCAGCCCGGACTCGGTAATGGCGCGCATTGCGCCCA GCGCCATCTGATCGTTGGCAACCAGCATCGCAGTGGGAACGATGCCC TCATTCAGCATTTGCATGGTTTGTTGAAAACCGGACATGGCACTCCA GTCGCCTTCCCGTTCCGCTATCGGCTGAATTTGATTGCGAGTGAGAT ATTTATGCCAGCCAGCCAGACGCAGACGCGCCGAGACAGAACTTAA TGGGCCCGCTAACAGCGCGATTTGCTGGTGACCCAATGCGACCAGAT GCTCCACGCCCAGTCGCGTACCGTCTTCATGGGAGAAAATAATACTG TTGATGGGTGTCTGGTCAGAGACATCAAGAAATAACGCCGGAACAT TAGTGCAGGCAGCTTCCACAGCAATGGCATCCTGGTCATCCAGCGGA TAGTTAATGATCAGCCCACTGACGCGTTGCGCGAGAAGATTGTGCAC CGCCGCTTTACAGGCTTCGACGCCGCTTCGTTCTACCATCGACACCA CCACGCTGGCACCCAGTTGATCGGCGCGAGATTTAATCGCCGCGACA ATTTGCGACGGCGCGTGCAGGGCCAGACTGGAGGTGGCAACGCCAA TCAGCAACGACTGTTTGCCCGCCAGTTGTTGTGCCACGCGGTTGGGA ATGTAATTCAGCTCCGCCATCGCCGCTTCCACTTTTTCCCGCGTTTTC GCAGAAACGTGGCTGGCCTGGTTCACCACGCGGGAAACGGTCTGAT AAGAGACACCGGCATACTCTGCGACATCGTATAACGTTACTGGTTTC ACATTCACCACCCTGAATTGACTCTCTTCCGGGCGCTATCATGCCAT ACCGCGAAAGGTTTTGCGCCATTCGATGGTGTCCGGGATCTCGACGC TCTCCCTTATGCGACTCCTGCATTAGGAAGCAGCCCAGTAGTAGGTT GAGGCCGTTGAGCACCGCCGCCGCAAGGAATGGTGCATGCAAGGAG ATGGCGCCCAACAGTCCCCCGGCCACGGGGCCTGCCACCATACCCAC GCCGAAACAAGCGCTCATGAGCCCGAAGTGGCGAGCCCGATCTTCC CCATCGGTGATGTCGGCGATATAGGCGCCAGCAACCGCACCTGTGGC GCCGGTGATGCCGGCCACGATGCGTCCGGCGTAGAGGATCGAGATC TCGATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCG GATAACAATTCCCCTCTAGATTTACACTTTATGCTTCCGGCTCGTATG TTTCTAGAGAATTCAAATAATTTTGTTTAACTTTAAGAAGGAGATTT AAATATGGGGTCCTCCCATCATCACCACCACCACTCTTCAGGCTTGG TACCGCGTGGTTCCATGATCAACATAGACGAATTGAAAAATTTATAT AAGGTGCAAAAGACCATCACTTTCGAACTTAAGAACAAGTGGGAGA ACAAAAATGATGAGAACGACAGAGTAGAGTTCTTGAAGACTCAGGA GTGGGTCGAAAGCCTTTTCAAGGTCGATGAAGAGAACTTTGATGAG AAAGAGTCTATCCCTAACTTGTTAGACTTCGGACAGAAGATTGCGTC CTTGTTTTACAAGCTGAGCGAGGACATAGCGAACAACCAAATTGATA

Protein	Vector Sequence
	CGCGGGTATTGAAAGTCTCGAAATTCCTTTTAGAGGAAATTGATAGA AATCAATACCACGAGAAAAAAAACAAGCCCACAAAGGTAAAAGAA ATGAATCCCAACACAAACAAAAGTTATATAAAAGAATATAAGCTGT CCGACCAAAACACACTGTACGTGTTATTAAAGATAATGGAAGATGA AGGTCGGGGATTACAAAAATTTTTGTACGATAAAGCGGACCGGTTA AACCTGTACAATCAAAAAGTTCGGAGAGACTTCGCCTTAAAGGAAT CAAATGAGCAACAAAAATTCTCTGGAAATGCCAACTACTATGGGAA TATAAAGCTGCTTATAGATAGCTTAGAAGATGCAGTCCGGATCATTG GGTATTTCACTTTCGACGATCAAGCAGAAAACGCACAAATCAATGA ATTTAAGTCCGTTAAACAGGAAATGAATAATAATGAAGCGTCTTACC AAGCACTGAAAGACTTCGCTATTGATAACGCAAAAAAAGAGATAGA ATTGACGACGTTGAACCACCGGGCGGTCAACAAGGATCCAAAAAAG ATTCAAGAACAGATTGAGGAAGTCGAAAATTTCGAAGAAGATATTA ACCAGTTAAAGCATCAGATATCAGCCTTGAATGATAAGAAGTTTGAC GTGGTTAGCAGATTAAAGCACGCTCTTATAAAAATGTTACCAGAACT GAATCTTTTGGATGCTGAGTCGGAACAGGGCCGTGAAGTCCAGCAG ATATATCAAGACAAAAAAAACGGGTTGGAGCTTGATGACTTTAAATT TAACCTTTTAAAACATCATCAATGGCAAAAAACGATCTTCAAGTATA TTAAGCTTGAGGGCTTAGTTCTGCCAGACCTTTACGCGGAAAACAAA CAAGATAAAATCAAGGTTTATATTGAGAATTATAGACAGAGTGGTG AGCGTATTTCTAAGAAGGCGAGAGAGGAATTAGGAAAAATCGATAA ACGCGAAGAGTTCAATGGAAATGACGAACTTAAGAAGGCATGGTAT GAGTATAAGGACTTCTGTAGAGACAAACGTAATAAGAGCGTGGAAC TTGGCAATAAGAAGTCGCTGTACAATGCCATAAAGCGCGAAGTTTTG CGGCAAAAAATGTGCAACCATTTCGCTGTGCTGGTGTCCGACGGTGA AGATACTTCCCCTTATTATTATCTGATATTAATCCCGAACGAGAACTC CGATGAAATGAATAGAACGTTCAAGGAATTGAAGGCCTCCGAGGGG AATTGGAAGATGTTGGATTACAATCGTCTGACCTTCAAAGCCTTGGA GAAATTGGCCCTGTTACGGTCGTCTACCTTCGAGATAGCGGATCAGG AACTGCAAGAAGAGGCAAAAAAGATCTGGGAGGAGTACAAGGAAA AGGCGTACAAAGACTTCAAAAACAAAAAGTTATTACAGGGTTTATC GGGAAGACAGCGGGAGGAGAAAAAGCAAGAATTGCAAAAGGAGAG CCTGAATAGAGTAATCAATTACTTGATCAGATGCATTCAGTCATTGC CCGACAGCGGAAAATACAACTTTAACTTTAAAGAGCCTCATCAATAC CAATCGCTTGAAGAGTTTGCCGAGGAGATTGATCGGCAAGGTTATCA CTGTGCTTGGAAAAACGTTTCTAAAGATAAACTGATGGAATTGGAAG CGATGGAAAAGATTAAGGTTTTCAAACTTCATAACAAAGACTTTCGC AAGGTAAAACTGAACGACTCCAAGCACAACCCTAATCTTTTTACTTT GTACTGGTTAGACGCCATGAATTTGGATAAGGTTAACGTCCGCCTGT TACCGGAAGTTGACCTTTACAAGAGAGCTAAGGAAACACAGCTGAA ATTGTTCGAACGTGATGTGAAATGCAATATCAATAACCAAAAGATTA AATCTATCAAGGAGAAGAATAGACTGTTTCAGGACAAGTTGTATGCT AGTTTTAAGTTAGAGTTTTATCCAGAAAACGAAGGATTAGGTTTCGA GCAGGTAAATGACAAGGTCAATAACTTCTGCGGTAGCGATACGGCC TATTATCTTGGGCTTGATCGTGGAGAGAAAGAGCTTGTTACATTCTG

Protein	Vector Sequence
	CCTGGTGGACTCTGATGGCCGCCTGGTAAAAAACGGAGACTGGACC AAGTTTAAAGAGGTGAACTATGCCGACAAACTGAAGCAATTCTACT ACTCAAAAGGCGAAATAGAGAGTACCCAACAACAGCTGTTAGAAGC CCGGGACAATATTAAACAAGCGACCAACACGGAAGATAAGGAGTCC ATGAAACTGAATTATAAGAAACTGGAACTGAAGTTAAAACAACAGA ATTTGCTGGCGCAAGAATTCATAAAAAAAGCGTACTGCGGCTACCTT ATCGATAGCATTAATGAGATTCTGAGAGAATATCCAAATACTTATCT TGTCTTAGAGGATTTGGATATCGCGGGTAAAGCGGATCCAGAGTCGG GGATGACTAATAAAGAGCAGAACTTAAACAAAACGATGGGGGCTTC AGTATACCAGGCCATTGAGAATGCGATCGTAAATAAATTCAAATATC GCACCGTGAAATTGTCCGATATCAAGGGCCTTCAGACTGTACCTAAT GTAGTGAAGGTCGAAGACTTACGGGAAGTGAAAGAGGTTGAAGATG GGGAACACAAGTTCGGGTTAATAAGATCAGTTAAGAGCAAGGATCA AATCGGTAACATACTTTTTGTCGACGAGGGGGAGACCAGTAACACTT GTCCGAATTGCGGTTTTAATAGTGATTGGTTTAAACGCGATGTTGAT TTTGACTTAGAAATAGTCGCTACTGTAAACGGGCAAAAGAATGCCGT GATTGAGCAAAATGACAAAAAATACTGTTTCCCGGGCGAAATATAT AAATTGGAAATCATTAATAAAGAGTACGAAACAAACAAGCGTAATC TTGCCATGATTTTTAAACCTCGGGCCAAAGCGTGCCGTAAATTTATC AATAATAATTTAGATAAGAACGATTATTTCTATTGTCCCTACTGCGC CTTCTCGTCGAAGAATTGTAACAACCCGAAACTGCAGAACGGCGATT TCGTGGTATATTCAGGAGACGATGTTGCTGCTTACAATGTTGCTATC AGAGGAATTAACCTGCTGAACAATATTAAATAGCTAGCATAACCCCT TGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAGGAA CTATATCCGGAT **(SEQ ID NO: 66)**
Cas9.1	TGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGT GTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGC GCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGC TTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTT AGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGG TTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGAC GTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAA CAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTT TGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAA TTTAACGCGAATTTTAACAAAATATTAACGTTTACAATTTCAGGTGG CACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTA AATACATTCAAATATGTATCCGCTCATGAATTAATTCTTAGAAAAAC TCATCGAGCATCAAATGAAACTGCAATTTATTCATATCAGGATTATC AATACCATATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAAACT CACCGAGGCAGTTCCATAGGATGGCAAGATCCTGGTATCGGTCTGCG ATTCCGACTCGTCCAACATCAATACAACCTATTAATTTCCCCTCGTCA AAAATAAGGTTATCAAGTGAGAAATCACCATGAGTGACGACTGAAT CCGGTGAGAATGGCAAAAGTTTATGCATTTCTTTCCAGACTTGTTCA ACAGGCCAGCCATTACGCTCGTCATCAAAATCACTCGCATCAACCAA ACCGTTATTCATTCGTGATTGCGCCTGAGCGAGACGAAATACGCGAT

Protein	Vector Sequence
	CGCTGTTAAAAGGACAATTACAAACAGGAATCGAATGCAACCGGCG CAGGAACACTGCCAGCGCATCAACAATATTTTCACCTGAATCAGGAT ATTCTTCTAATACCTGGAATGCTGTTTTCCCGGGGATCGCAGTGGTG AGTAACCATGCATCATCAGGAGTACGGATAAAATGCTTGATGGTCG GAAGAGGCATAAATTCCGTCAGCCAGTTTAGTCTGACCATCTCATCT GTAACATCATTGGCAACGCTACCTTTGCCATGTTTCAGAAACAACTC TGGCGCATCGGGCTTCCCATACAATCGATAGATTGTCGCACCTGATT GCCCGACATTATCGCGAGCCCATTTATACCCATATAAATCAGCATCC ATGTTGGAATTTAATCGCGGCCTAGAGCAAGACGTTTCCCGTTGAAT ATGGCTCATAACACCCCTTGTATTACTGTTTATGTAAGCAGACAGTTT TATTGTTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAG CGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTT TTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACC AGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGA AGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTA GTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCC TACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTG GCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCG GATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGC CCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCG TGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGAC AGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGG GAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTT TCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGG GGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTC CTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCC CTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACC GCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGG AAGCGGAAGAGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGC GGTATTTCACACCGCATATATGGTGCACTCTCAGTACAATCTGCTCT GATGCCGCATAGTTAAGCCAGTATACACTCCGCTATCGCTACGTGAC TGGGTCATGGCTGCGCCCCGACACCCGCCAACACCCGCTGACGCGCC CTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGA CCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCG AAACGCGCGAGGCAGCTGCGGTAAAGCTCATCAGCGTGGTCGTGAA GCGATTCACAGATGTCTGCCTGTTCATCCGCGTCCAGCTCGTTGAGTT TCTCCAGAAGCGTTAATGTCTGGCTTCTGATAAAGCGGGCCATGTTA AGGGCGGTTTTTTCCTGTTTGGTCACTGATGCCTCCGTGTAAGGGGG ATTTCTGTTCATGGGGGTAATGATACCGATGAAACGAGAGAGGATG CTCACGATACGGGTTACTGATGATGAACATGCCCGGTTACTGGAACG TTGTGAGGGTAAACAACTGGCGGTATGGATGCGGCGGGACCAGAGA AAAATCACTCAGGGTCAATGCCAGCGCTTCGTTAATACAGATGTAGG TGTTCCACAGGGTAGCCAGCAGCATCCTGCGATGCAGATCCGGAAC ATAATGGTGCAGGGCGCTGACTTCCGCGTTTCCAGACTTTACGAAAC ACGGAAACCGAAGACCATTCATGTTGTTGCTCAGGTCGCAGACGTTT

Protein	Vector Sequence
	TGCAGCAGCAGTCGCTTCACGTTCGCTCGCGTATCGGTGATTCATTCT GCTAACCAGTAAGGCAACCCCGCCAGCCTAGCCGGGTCCTCAACGA CAGGAGCACGATCATGCGCACCCGTGGGGCCGCCATGCCGGCGATA ATGGCCTGCTTCTCGCCGAAACGTTTGGTGGCGGGACCAGTGACGAA GGCTTGAGCGAGGGCGTGCAAGATTCCGAATACCGCAAGCGACAGG CCGATCATCGTCGCGCTCCAGCGAAAGCGGTCCTCGCCGAAAATGAC CCAGAGCGCTGCCGGCACCTGTCCTACGAGTTGCATGATAAAGAAG ACAGTCATAAGTGCGGCGACGATAGTCATGCCCCGCGCCCACCGGA AGGAGCTGACTGGGTTGAAGGCTCTCAAGGGCATCGGTCGAGATCC CGGTGCCTAATGAGTGAGCTAACTTACATTAATTGCGTTGCGCTCAC TGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGA ATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCCAGG GTGGTTTTTCTTTTCACCAGTGAGACGGGCAACAGCTGATTGCCCTTC ACCGCCTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCACGCTGGTTTG CCCCAGCAGGCGAAAATCCTGTTTGATGGTGGTTAACGGCGGGATAT AACATGAGCTGTCTTCGGTATCGTCGTATCCCACTACCGAGATATCC GCACCAACGCGCAGCCCGGACTCGGTAATGGCGCGCATTGCGCCCA GCGCCATCTGATCGTTGGCAACCAGCATCGCAGTGGGAACGATGCCC TCATTCAGCATTTGCATGGTTTGTTGAAAACCGGACATGGCACTCCA GTCGCCTTCCCGTTCCGCTATCGGCTGAATTTGATTGCGAGTGAGAT ATTTATGCCAGCCAGCCAGACGCAGACGCGCCGAGACAGAACTTAA TGGGCCCGCTAACAGCGCGATTTGCTGGTGACCCAATGCGACCAGAT GCTCCACGCCCAGTCGCGTACCGTCTTCATGGGAGAAAATAATACTG TTGATGGGTGTCTGGTCAGAGACATCAAGAAATAACGCCGGAACAT TAGTGCAGGCAGCTTCCACAGCAATGGCATCCTGGTCATCCAGCGGA TAGTTAATGATCAGCCCACTGACGCGTTGCGCGAGAAGATTGTGCAC CGCCGCTTTACAGGCTTCGACGCCGCTTCGTTCTACCATCGACACCA CCACGCTGGCACCCAGTTGATCGGCGCGAGATTTAATCGCCGCGACA ATTTGCGACGGCGCGTGCAGGGCCAGACTGGAGGTGGCAACGCCAA TCAGCAACGACTGTTTGCCCGCCAGTTGTTGTGCCACGCGGTTGGGA ATGTAATTCAGCTCCGCCATCGCCGCTTCCACTTTTTCCCGCGTTTTC GCAGAAACGTGGCTGGCCTGGTTCACCACGCGGGAAACGGTCTGAT AAGAGACACCGGCATACTCTGCGACATCGTATAACGTTACTGGTTTC ACATTCACCACCCTGAATTGACTCTCTTCCGGGCGCTATCATGCCAT ACCGCGAAAGGTTTTGCGCCATTCGATGGTGTCCGGGATCTCGACGC TCTCCCTTATGCGACTCCTGCATTAGGAAGCAGCCCAGTAGTAGGTT GAGGCCGTTGAGCACCGCCGCCGCAAGGAATGGTGCATGCAAGGAG ATGGCGCCCAACAGTCCCCCGGCCACGGGGCCTGCCACCATACCCAC GCCGAAACAAGCGCTCATGAGCCCGAAGTGGCGAGCCCGATCTTCC CCATCGGTGATGTCGGCGATATAGGCGCCAGCAACCGCACCTGTGGC GCCGGTGATGCCGGCCACGATGCGTCCGGCGTAGAGGATCGAGATC TCGATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCG GATAACAATTCCCCTCTAGATTTACACTTTATGCTTCCGGCTCGTATG TTTCTAGAGAATTCAAATAATTTTGTTTAACTTTAAGAAGGAGATTT AAATATGGGCAGCAGCCATCATCATCATCACAGCAGCGGCCTG

Protein	Vector Sequence
	GTGCCGCGCGGCAGCATGCAGAGGATTTTCGGCCTCGATATCGGCAC CACGTCCATCGGCTTTGCGGTCATCGACCACGACCGCGACCAAGGCG TCGGCCGCATCCACCGGCTGGGCGCGCGCATCTTCCCGGAAGCGCGC GACGAGAAGGGAACACCGCTCAACCAGCATCGGCGGCAAAAGCGTC TCGCGCGCCGCCAATTGCGCCGGCGCCGGCTTCGGCGCAAGGCGCTC AACGAACTGCTTTCGGCCCGCGGGATGCTGCCGCGCTTCGGCACGTC CGCTTGGCACGACGCGATGGCGCTCGACCCTTACGCGCTCCGTGCAC GGGGTACGGAGGAGGCGTTGCAGCCGGTAGAGGTCGGTCGGGCTCT CTATCACCTCGCCCAGCGTCGCCACTTCAAGCCACGGGACGAGGCTG CGGAAGCCGACGAGCAGGAGGTGGGCGATCAGGAGGCCGAGACCA AGCGTGAGAAGCTGCTGCAGGCGTTGCGCCGCAGCGGTCGAACGCT GGGCCAGGAACTGGCGGCGCGCGGTCCGCACGAGCGCAAGCGGCAC GAGCACGCTTTGCGCTCGACCGTCGAGACCGAGTTCGAGCGGCTCCT CACCGCGCAAGCGCGGCATCACGAGATCCTTCGCGATCCCGAGTTCG TCGAGGAACTGAGAGAGACCATCTTCGCGCAACGGCCCGTCTTTTGG CGGACGAGCACGCTCGGCACGTGCCCGTTCGTTCCAGGCGCACCGTT GTGCCCGAAGGGTGCTTGGCTCTCCCGCCAGCGGCGCATGCTGGAGC AGGTCAACAACCTCGCCATCACCGGCGGCAACGCGCGTCCGCTCGA CCACGAGGAGCGACGAGCGATCCTCGCCGTCTTACAGACGCAGGCC AGCATGAGCTGGGGCGCGGTCCGAACCGCGCTTAAGCCGCTCTTCAA GGCACGCGGCGAGGCGGGCGCCGAGCGTCGGCTCCGGTTCAATCTC GAAGAGGGCGGCGGTAAGACGCTGCTCGGGAACCCGCTGGAAGCGA AGCTCGCCCGGATCTTCGGCGAAGCCTGGGCCACGCACCCTCACCGC GACGCGATCCGTGAGACGATCCATGACCGCCTTTTCGCCGCGACCTA TAACGCGAAGGGCGCGCAGCGCATCGTCATCCTTCCGGCATCCCAAC GCGCTGAACGGATGCGGGGGGTCATCGCCGGCCTCCAAGCGGATTT CGGCCTTTCCCACGAGCAGGCGATGGCGCTTGCGGAGCTGCCGCTGA CGCCCGGCTGGGAACCCTATTCGAGCGAAGCCCTTCGCGCGTTAATG CCGAAGCTGGAGGAAGGCGTGCGCTTCGGCGCCCTCGTCGTGGCCCC TGAATGGGAAGATTGGCGCGAGGCCACCTTCCCCCAGCGCGAGCGG CCGACCGGCGAGGTGCTCGACCTCTTGCCTTCACCGAAATGCCACGA TGAGAGCCGCCGGCAGACGCGGCTGCGGAACCCGACGGTGCTGCGC ACGCAGAACGAGCTGCGCAAGGTCGTCAACAACCTGATCCGGGCGC ACGGCAAGCCCGACATCATCCGCGTCGAGGTCGCCCGCGAGGTGGG GCTTTCCAAGCGCGAGCGTGAAGATCGCTACAACGGGATGCGGCGC CAGGAGCGCCAGCGGCAAGCGGCGATCAAAGACCTCCAAGCCAAGG GCTTCGCCGAGCCGTCGCGCGCCGACGTCGAGAAGTGGCTTTTGTGG AAGGAGAGCAAGGAGACCTGCCCTTACACGGGGGACAAGATCTGCT TCGACGCTCTGTTTCGCCGCGGTGAGTTTCAAGTGGAGCACATCTGG CCGCGCTCGCGCTCGTTCGACGACAGCTTCCGCAACAAGACCCTGTG TCGGCGCGACGTGAACCTCGCCAAGGGTAACCAAACGCCCTTCGAG TTCTTCGAGAGCCGACCCGAGGAGTGGGAGGCCGTGAAGCGCCGCC TCGATGGCTTGCAGGCCAAGCGGGCAGGCGGTGAGGGGATGGCGCG CGGCAAGGTGAAGCGCTTCGTCGCGAGCACGTTGCCGGACGATTTCG CGCAGCGTCAGCTCAACGACACGGGCTGGGCGGCGCGCGAGGCGGT

Protein	Vector Sequence
	GGCCTTCCTCAAGCGGCTGTGGCCGGACGAGGGGCAAGCCGCGCCG GTCCGCGTCCAGGCGGTCACGGGGCGGGTGACGGCGCAGCTTCGCC ACCTGGGGGGCCTCGATGGCGTGCTGTCGGACGGTGCTCGAAAGAC GCGTGACGACCACCGCCATCACGCCGTCGATGCGCTGGTCGTCGCCT GCACGCATCCGGGCATGACCGAGCGGCTCAGCCGCTACTGGCAGCA GAAGGAGGACGAGCGCGCCGAACGACCGCAGCTGGACCCACCGTGG CCCACGATCCGAGCGGACGCCGAGGCGGCCAAGGACTTAATCGTCG TCTCGCACCGGGTGCGCAAGAAGATCTCGGGACCGTTCCACAAGGA AACCGTCTATGGCGCGACCGACGAGCGCGAGGTCACGCGCGGGCTT GAGTACGAGAAATTCGTCACGCGGAAGCGCGTCGAGGACCTGACGA AATCCATGCTCGCCGACATCCGCGACGACAGGGTGCGGCAAATTGT GACGGCGTGGGTGGCCGAGCGCGGCGGCGACCCGAAGAAGGCGTTT CCGCCCTATCCGACGCTGGGGTCGAGCGGACCCGAGATCCGCAAGG TGCGCGTTCTGATCCGCCGGCAGCCCACCTTGATGGCACGGGCAGCG ACGGGCTTCGCTGATCTCGGAGCGAACCACCATGTCGCCATCTACAA GACCGCCGACGAGCGATTCGCCTTCGAGGTCGTCAGCTTGCTGGAGG TCGCCAGGCGCGTCGACCGCGGTGAACCGCCCGTGAAGAGACAGCG AGGCGACGAGAAGCTCGTGATGTCTTTGGCGCAGGGCGATCTGATA CGGTTCGCCAAAACGCCCGATGCGGAAGCAGCAATTTGGCGTGTTCA GAAAATCGCAACTAAAGGTCAGATATCGCTCCTTCACCACGATGACG CTTCGCCGAAGGAGCCGAGTCTCTTTGAACCGATGGTTGGTGGGTTG ATGGCTCGGAACCCGGAGAAGCTGGCAGTCGATCCCATCGGCCGAG TGCGCAAGGCAGGCGACTGACTAGCATAACCCCTTGGGGCCTCTAA ACGGGTCTTGAGGGGTTTTTTGCATATGCTGAAAGGAGGAACTATAT CCGGAT **(SEQ ID NO: 67)**
Cas 9.2	TGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGT GTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGC GCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGC TTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTT AGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGG TTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGAC GTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAA CAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTT TGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAA TTTAACGCGAATTTTAACAAAATATTAACGTTTACAATTTCAGGTGG CACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTA AATACATTCAAATATGTATCCGCTCATGAATTAATTCTTAGAAAAAC TCATCGAGCATCAAATGAAACTGCAATTTATTCATATCAGGATTATC AATACCATATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAAACT CACCGAGGCAGTTCCATAGGATGGCAAGATCCTGGTATCGGTCTGCG ATTCCGACTCGTCCAACATCAATACAACCTATTAATTTCCCCTCGTCA AAAATAAGGTTATCAAGTGAGAAATCACCATGAGTGACGACTGAAT CCGGTGAGAATGGCAAAAGTTTATGCATTTCTTTCCAGACTTGTTCA ACAGGCCAGCCATTACGCTCGTCATCAAAATCACTCGCATCAACCAA ACCGTTATTCATTCGTGATTGCGCCTGAGCGAGACGAAATACGCGAT

Protein	Vector Sequence
	CGCTGTTAAAAGGACAATTACAAACAGGAATCGAATGCAACCGGCG CAGGAACACTGCCAGCGCATCAACAATATTTTCACCTGAATCAGGAT ATTCTTCTAATACCTGGAATGCTGTTTTCCCGGGGATCGCAGTGGTG AGTAACCATGCATCATCAGGAGTACGGATAAAATGCTTGATGGTCG GAAGAGGCATAAATTCCGTCAGCCAGTTTAGTCTGACCATCTCATCT GTAACATCATTGGCAACGCTACCTTTGCCATGTTTCAGAAACAACTC TGGCGCATCGGGCTTCCCATACAATCGATAGATTGTCGCACCTGATT GCCCGACATTATCGCGAGCCCATTTATACCCATATAAATCAGCATCC ATGTTGGAATTTAATCGCGGCCTAGAGCAAGACGTTTCCCGTTGAAT ATGGCTCATAACACCCCTTGTATTACTGTTTATGTAAGCAGACAGTTT TATTGTTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAG CGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTT TTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACC AGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGA AGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTA GTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCC TACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTG GCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCG GATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGC CCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCG TGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGAC AGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGG GAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTT TCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGG GGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTC CTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCC CTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACC GCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGG AAGCGGAAGAGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGC GGTATTTCACACCGCATATATGGTGCACTCTCAGTACAATCTGCTCT GATGCCGCATAGTTAAGCCAGTATACACTCCGCTATCGCTACGTGAC TGGGTCATGGCTGCGCCCCGACACCCGCCAACACCCGCTGACGCGCC CTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGA CCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCG AAACGCGCGAGGCAGCTGCGGTAAAGCTCATCAGCGTGGTCGTGAA GCGATTCACAGATGTCTGCCTGTTCATCCGCGTCCAGCTCGTTGAGTT TCTCCAGAAGCGTTAATGTCTGGCTTCTGATAAAGCGGGCCATGTTA AGGGCGGTTTTTTCCTGTTTGGTCACTGATGCCTCCGTGTAAGGGGG ATTTCTGTTCATGGGGGTAATGATACCGATGAAACGAGAGAGGATG CTCACGATACGGGTTACTGATGATGAACATGCCCGGTTACTGGAACG TTGTGAGGGTAAACAACTGGCGGTATGGATGCGGCGGGACCAGAGA AAAATCACTCAGGGTCAATGCCAGCGCTTCGTTAATACAGATGTAGG TGTTCCACAGGGTAGCCAGCAGCATCCTGCGATGCAGATCCGGAAC ATAATGGTGCAGGGCGCTGACTTCCGCGTTTCCAGACTTTACGAAAC ACGGAAACCGAAGACCATTCATGTTGTTGCTCAGGTCGCAGACGTTT

Protein	Vector Sequence
	TGCAGCAGCAGTCGCTTCACGTTCGCTCGCGTATCGGTGATTCATTCT GCTAACCAGTAAGGCAACCCCGCCAGCCTAGCCGGGTCCTCAACGA CAGGAGCACGATCATGCGCACCCGTGGGGCCGCCATGCCGGCGATA ATGGCCTGCTTCTCGCCGAAACGTTTGGTGGCGGGACCAGTGACGAA GGCTTGAGCGAGGGCGTGCAAGATTCCGAATACCGCAAGCGACAGG CCGATCATCGTCGCGCTCCAGCGAAAGCGGTCCTCGCCGAAAATGAC CCAGAGCGCTGCCGGCACCTGTCCTACGAGTTGCATGATAAAGAAG ACAGTCATAAGTGCGGCGACGATAGTCATGCCCCGCGCCCACCGGA AGGAGCTGACTGGGTTGAAGGCTCTCAAGGGCATCGGTCGAGATCC CGGTGCCTAATGAGTGAGCTAACTTACATTAATTGCGTTGCGCTCAC TGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGA ATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCCAGG GTGGTTTTTCTTTTCACCAGTGAGACGGGCAACAGCTGATTGCCCTTC ACCGCCTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCACGCTGGTTTG CCCCAGCAGGCGAAAATCCTGTTTGATGGTGGTTAACGGCGGGATAT AACATGAGCTGTCTTCGGTATCGTCGTATCCCACTACCGAGATATCC GCACCAACGCGCAGCCCGGACTCGGTAATGGCGCGCATTGCGCCCA GCGCCATCTGATCGTTGGCAACCAGCATCGCAGTGGGAACGATGCCC TCATTCAGCATTTGCATGGTTTGTTGAAAACCGGACATGGCACTCCA GTCGCCTTCCCGTTCCGCTATCGGCTGAATTTGATTGCGAGTGAGAT ATTTATGCCAGCCAGCCAGACGCAGACGCGCCGAGACAGAACTTAA TGGGCCCGCTAACAGCGCGATTTGCTGGTGACCCAATGCGACCAGAT GCTCCACGCCCAGTCGCGTACCGTCTTCATGGGAGAAAATAATACTG TTGATGGGTGTCTGGTCAGAGACATCAAGAAATAACGCCGGAACAT TAGTGCAGGCAGCTTCCACAGCAATGGCATCCTGGTCATCCAGCGGA TAGTTAATGATCAGCCCACTGACGCGTTGCGCGAGAAGATTGTGCAC CGCCGCTTTACAGGCTTCGACGCCGCTTCGTTCTACCATCGACACCA CCACGCTGGCACCCAGTTGATCGGCGCGAGATTTAATCGCCGCGACA ATTTGCGACGGCGCGTGCAGGGCCAGACTGGAGGTGGCAACGCCAA TCAGCAACGACTGTTTGCCCGCCAGTTGTTGTGCCACGCGGTTGGGA ATGTAATTCAGCTCCGCCATCGCCGCTTCCACTTTTTCCCGCGTTTTC GCAGAAACGTGGCTGGCCTGGTTCACCACGCGGGAAACGGTCTGAT AAGAGACACCGGCATACTCTGCGACATCGTATAACGTTACTGGTTTC ACATTCACCACCCTGAATTGACTCTCTTCCGGGCGCTATCATGCCAT ACCGCGAAAGGTTTTGCGCCATTCGATGGTGTCCGGGATCTCGACGC TCTCCCTTATGCGACTCCTGCATTAGGAAGCAGCCCAGTAGTAGGTT GAGGCCGTTGAGCACCGCCGCCGCAAGGAATGGTGCATGCAAGGAG ATGGCGCCCAACAGTCCCCCGGCCACGGGGCCTGCCACCATACCCAC GCCGAAACAAGCGCTCATGAGCCCGAAGTGGCGAGCCCGATCTTCC CCATCGGTGATGTCGGCGATATAGGCGCCAGCAACCGCACCTGTGGC GCCGGTGATGCCGGCCACGATGCGTCCGGCGTAGAGGATCGAGATC TCGATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCG GATAACAATTCCCCTCTAGATTTACACTTTATGCTTCCGGCTCGTATG TTTCTAGAGAATTCAAATAATTTTGTTTAACTTTAAGAAGGAGATTT AAATATGGGCAGCAGCCATCATCATCATCACAGCAGCGGCCTG

Protein	Vector Sequence
	GTGCCGCGCGGCAGCATGAAGAAAGAAAAAGTGTACATGGGGCTAG ATCTTGGCACGAACTCTGTCGGCTGGGCGGTAACGGATAACGACTAT AAGGTGCTCAAGTTTAAACGGCGCGCTATGTGGGGGGTTCGGCTCTT TAATGAAGCCAATCCGGCTGTCGAGAGGCGTGTTGCCCGTTCAAATC GCCGTCGCCTGGCCCGAAAAAAACAACGCGTGGCTTGGCTGAAAGA AATATTTAAGAATTCCATTAGCGAAATTGATCCGGAATTTTTCGACC GTCTTGAACAAAGCGCGCTTTGGGCAGAAGACAAAAATGTCGCCGG AAAATACTCCCTTTTTAATGAGAAAAAATTAACCGATAAGACATTCT ATCGGAAATTTCCCACCGTTTTTCACCTAAAAAAAGCGCTTATGGAC GGCAAAATAAAAAAACCTGATATTCGCTTTGTATATCTTGCCTTGTC CCACTATCTGCAAAACAGAGGCCATTTTCTCTTGGAAAATGAGCTGA ACAGTGTTGAAGATATAGACATTCGGGATATTTTTAACAGTCTTAAT GAAAGAATTCATGTTCTTATTGACAGCGGTGATGATATGGTTCCTGC TTTTGATTTGACAAACCTTGATGATTTGAAACAAATTGCCACAGACA CAAATATATCCGGGAAAACGCAGGAAAAAGAAGCCTTTATAAAAAC CCTGTTAAATGGGGCCAAACAGCCTGCCTTAGAGGCAATTATTAAAT TATGTACAGGCGGCTCGGCTAATTTATCAAAAATCTTTGGTGATATG TTTGAATTTGAAAGTGAAATCAAATCAATATCATTCGAAAAGGCCAA CTTCGAAGATGAAATCGCTCCCAAGCTGCAAGATTGTCTGGGAGATT ACTATCAGATTATTGAGCTGGCTCAGCAGATTTACAGCTGGTACACG CTTTATAAGGTATGCAGCGGTCGACCGTCGGTCTCTCACGCCAAAGT GGAGGATTACGAAAAACACAAAGAACAGCTGTCCCACCTAAAAGTG CTGGTAAGAAAACACTTTTCGAAAAATGTCTACCGGGAAATATTCCG AAAAGAAGACGACAAAATCCATAACTATGTATCCTACATATCCGGC AAAAAGGACCGCGACGAATTTTATAAATATCTCAAAAAAACGTTAG AAAAAAAATCTACATTCAAGAAAACGTCTGAATTTGAGAATATTTCT CGCGCCATTGAACAGCAAAACTACCTGCCGAAACAACGGGTCAAAG ACAACTCTGTGGTGCCTCAGCAGCTATACAAACAAGAAATCGTAAA AATCCTCAACAACCTTTCATCACACTACCCCTTTTTATCACAAAAAA CAGACGGGATCAGCAATCGAGAAAAGATTATCAAAATCTTTGAATA CCGCATCCCATACTATGTCGGTCCTCTTTGCGATATCCATCGTGCGGG GGATGACGGGTTCTCCTGGCTGGTTCGTGACTGCAGTAAAAAGATTA CTCCTTGGAACTTCGAGCAAGTCGTCGATATCCCCCAGTCTGCTGAA AATTTCATTAAGAACATGACCCGTAAATGCACCTATTTAAAACAGTA TAATGTGCTGCCGAAAAATTCTCTCCTCTATAGCGAGTATAGCGTAC TAAATGAACTTAACAATGTGCGCATCAAAACTAAAAAGCTGACCCCT AAGCTAAAAGAAAAAATGCTCAACACATTATTTCGCCAAAAGAAGA ATATTTCGATAACGAGCTTGATTCATTGGCTTGTCAGTGAAGGAGTG TATGAGAAAGGGGAGATTGAAAAATCAGACGTCAGCGGTGTTGATT CCAATTTTACCAGCTCTCTTTCTGCAGCCATTTCTTTTGATCGTATCA TTGGTGAAAAGATGAAAAACAAAAAAACCCAAAAAATGGTCGAGG AGATCATAAACTGGCTCGCCCTTTTTTCGGACAAAAAAATACTACAA CAAAAGATTGTAGAGAAATATCAAGATAAAGTCTCGCAAGAACAAA TCGGAAAAATTCTGCGCCTCAACCTAAGCGGATGGGGACGACTTTCT TCGGAGTTTCTGCAACTGAAAAACTCCCAACCGGGAGAACACGACG

Protein	Vector Sequence
	GAAAAACGCTCATCAATATCATGCGGCAGACCCAGATGAATCTGAT GGAGATTATTCACTCTCCCCAGTTCAGTTTCAATACCGTTATTGAAAC GGAGGCCAAAAAACAGCTAACGGGACACATTACCCACAGTCATGTT GAGGCGCTGTACTGCTCTCCTGTGGTCAAGAAACAGATATGGCAGGC CCTGCAAATCGCCCTGGAGCTAAAGAAAACCTTAAAGAAAGACCCG AACAAAATTTTTGTGGAGACAACCCGGCATGAAGGGGAGAAAAAAC GGACCACAAGCCGTCACAAACAACTACTCGAGTTATACCAAGCCGC CAAGTCCCATCTGCCCGACCTGACGAAAAGTATAAAGGAACTAAAC GATGCGCTAAAAGATACAGAGCCGGAGAAGATGAAACGGAAAAAA CTGTTTCACTACTACAAACAACTGGGACGTTGTATGTATACAGGCAG GCCCATCAGTCTAGAGGATCTGTTTACCAATAAATATGACATTGATC ATATTTATCCCCAGAGTTTAACCAAAGATGACAGTTTTACTAATACT GTACTGGTGGAACGGCTATCAAACGCGGAGAAATCAGACGCATTCC CTCTTGACAGTAAAACAAGAAAAGACCGTCAAGGACTGTGGCGCTG TTTACGACGGAACGGACTAATTACCAAAGAAAAGTACTACCGCTTA ACACGGGAAACACCTCTAAGCGAAGAAGAAAAAGCGGCCTTTATTC GTCGTCAGCTGGTGGAAACCAGCCAGACAACCAAGGAAGTAATCCG ATTTCTGGCGACCCTTTTCCCAAAGTCAAAAGTTGTGTATGTAAAGA GCGGCAACGTCAGCGACTTTCGCCGTGACTTTTCCCCGTCCCTGCCC GAAAACAAAACTAACGGCAAAGACCCCAAGGGGATAACCGACTACA GCATGATTAAAGTGAGGGAAATCAATGATTTGCACCACGCGAAAGA CGCGTATTTAAACATCGTGGTCGGCAATGTCTACGACACCAAATTTC GCTACCGAGGCAAAGACCTCACGGCCATAGTGCGCGAAAAAGCGAG GCAGTACCATTTATCCCGTTTGTTTCTTTACTCTACCGACGGCGCCTG GATCGGAGCGGCTGATGAAAACAGAGGGAAGCAACGACCGAGTATT GAAACCGTGATCGCGGAAATGCGGCGAAATAGCTGTCAGGTAACGT GGGAAGCCGTCTTTAAAAAAGGGCAGCTGTGGGACATGAACGCCAA AAGTAAGCGGCCGGGACTGTTGCCGATCAAGAAAGAACTATCCGAT ACGGCAAAATATGGAGGGTACCAGGGGAAGACCGCGTCTTATTTTG TGGTCGTTGAGTATGAGAATAAAAAAGGCGAACGTGAAAAAAAACT GGAATCGGTCCCGATTTATGTGAAAGCGCTCAGTAAACAAAAGCCG GACGCTGTCAATTCTTTCCTACGGGATACACTGGGTCTGGAGAAACC AAGCGTCATGGTCGACAACATCAAAATCGGCTCCATCGTCGAGATCA ACGGGGCCCGAATGGTCCTTACGGGGAATAATGAAGTTCTAGTATTT GGGCGTATCGCGTCCCAACTGATCCTGGATATAACGATGGCCGCCTA TCTAAAACGAATGTTTAAGCTGCTTGCTGACACAGCCAAGATCAAAG AGAACAATGTCTACTTTAAAAACTGCGGCTATCTGGATAAGGAGAC GAACCTGGCAGTATACGATACGTTTATTGCCAAGCTGAAACTGCCCC GGTATGCTCAGATTATCACCCATAGCCTATATGAGAAGATGGAAAGC AATCGTGATGTGTTTATCAACCTTTCACTGGCCGACCAGTGTAATCT GCTGGCCGGCGTACTGCCTGCGCTACAGTGTAACAGCCAAAATGCCG ATCTGTCTCTTCTTGGTGAAGGTAAAGCGGTCGGAAATATCGCATTT TCAAAAAACGCGATCCTGAAAAAGAATCAGGTCCGTCTTGTTGATTG CTCCATTACCGGGCTCTTCGAAAACAGCAGAAATATGGCATAACTAG CATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCAT

Protein	Vector Sequence
	ATGCTGAAAGGAGGAACTATATCCGGAT **(SEQ ID NO: 68)**

Protein expression and purification

[00432] Cas12 coding sequences were codon-optimized and synthesized by GeneScript and then cloned into pET28a (Novagen) with N-terminal 6×His tagging. Cas12 expression plasmids were transformed into E. coli NiCo21 (DE3) (NEB). For protein expression, a single clone was first cultured overnight in 5-mL liquid LB tubes and then inoculated into 400 ml of fresh liquid LB (OD 600 0.1). Cells were grown with shaking at 200 rpm and 37 °C until the OD 600 reached 0.8, and IPTG was then added to a final concentration of 0.1 mM followed by further culture of the cells at 37 °C for about 2 h before the cell harvesting. Cells were resuspended in 20 mL of buffer A (50 mM Tris-HCl pH 8.0, 0.5 M NaCl, 1 mM DTT and 5% glycerol) with protease inhibitor cocktail (Promega) and 5 mg/ml lysozyme. After a 15 min incubation at 37 °C, cells were lysed by sonication for 10 minutes with 10 s on and 10 s off cycle. Cell debris and insoluble particles were removed by centrifugation (15,000 rpm for 30 min). After centrifuging, the supernatant was loaded onto a 5 mL Crude HisTrap column (GE Healthcare) equilibrated in buffer A with 20 mM imidazol on an AKTA Pure 25L device (GE Healthcare Life Sciences). The elution was performed by a step gradient of buffer B (buffer A plus 0.5 M imidazole). The elution was dialysed with dialysis buffer (50 mM Tris-HCl pH 8.0, 200 mM NaCl, 1 mM DTT and 5% glycerol).

Guide RNAs (gRNAs) and Variants

[00433] Inclusion of direct repeat mutations may improve gRNA stability. The direct repeats from the three CRISPR Cas12 systems provided herein have two A:U base pairs within the stem-loop region. Increasing the thermal stability of the stem-loop is expected to increase the fraction of properly folded crRNA for loading into its cognate Cas12 and thereby nuclease activity (Pengpeng et al., 2019). Those A:U base pairs were replaced with C:G in the direct repeats of the CRISPR systems of the disclosure to create new, more stable non-naturally occurring variants based on the minimum free energy prediction for the RNA folding.

[00434] The predicted (putative) naturally occurring direct repeat sequences in the CRISPR locus, as found in bacterial DNA, of the Cas proteins of the disclosure are shown in Table 2 and 5a, above (shown as DNA sequences). Novel variants are shown in Table 5b above (represented

as DNA sequences). The predicted secondary structure are shown in FIGS. 7A-7C. The entire direct repeat sequence, or part of the direct repeat sequence is expected to form a functional non-naturally occurring gRNA, and bind to a Cas protein of the disclosure. RNAs forming the direct repeat variants and spacers used in this example were synthesized by Synthego.

[00435] FIGS. 3B, 3E, 3G, 5B, 5D, and 5F shows the predicted secondary structures (folding) of the repeat sequence for the Cas9.1, Cas9.3, Cas9.4, Cas12a.1, Cas12p, and Cas12q pre-crRNA. To assemble these predictions, the openly available RNAfold webserver tool was used.

In Vitro Transcription Reaction (IVT)

[00436] In Vitro Transcription was carried out using MEGAscript™ T7 Transcription Kit (Ambion, Invitrogen) according to the manufacturer's instructions and were cleaned with Monarch® RNA Cleanup Kit (New England Biolab) according the manufacturer's instructions. RNAs were visualized in a 2% agarose gel using Gel Loading Buffer II (Ambion, Invitrogen).

In vitro target cleavage assays

[00437] The following template sequences used for *in vitro* target cleavage assays are shown in Table 9.

Table 9

Name	**Sequence**	**Nucleic acid**
KPC gB template 1	TTCAAGGGCTTTCTTGCTGCCGCTGTGCTGGCTCGCAG CCAGCAGCAGGCCGGCTTGCTGGACACACCCATCCGT TACGGCAAAAATGCGCTGGTTCCGTGGTCACCCATCTC GGAAAAATATCTGACAACAGGCATGACGGTGGCGGAG CTGTCCGCGGCCGCCGTGCAATACAGTGATAACGCCG CCGCCAATTTGTTGCTGAAGGAGTTGGGCGGCCCGGC CGGGCTGACGGCCTTCATGCGCTCTATCGGCGATACCA CGTTCCGTCTGGACCGCTGGGAGCTGGAGCTGAACTC CGCCATCCCAGGCGATGCGCGCGATACCTCATCGCCG CGCGCCGTGACGGAAAGCTTACAAAAACTGACACTGG GCTCTGCACTGGCTGCGCCGCAGCGGCAGCAGTTTGTT GATTGGCTAAAGGGAAACACGACCGGCAACCACCGCA TCCGCGCGGCGGTGCCGGCAGACTGGGCAGTCGGAGA CA (SEQ ID NO: 43)	DNA
NDM gB template 1	CCAAATTAAGATCATCTATTTACTAGGCCTCGCATTTG CGGGGTTTTTAATGCTGAATAAAAGGAAAACTTGATG GAATTGCCCAATATTATGCACCCGGTCGCGAAGCTGA GCACCGCATTAGCCGCTGCATTGATGCTGAGCGGGTG CATGCCCGGTGAAATCCGCCCGACGATTGGCCAGCAA	DNA

Name	Sequence	Nucleic acid
	ATGGAAACTGGCGACCAACGGTTTGGCGATCTGGTTTT CCGCCAGCTCGCACCGAATGTCTGGCAGCACACTTCCT ATCTCGACATGCCGGGTTTCGGGGCAGTCGCTTCCAAC GGTTTGATCGTCAGGGATGGCGGCCGCGTGCTGGTGG TCGATACCGCCTGGACCGATGACCAGACCGCCCAGAT CCTCAACTGGATCAAGCAGGAGATCAACCTGCCGGTC GCGCTGGCGGTGGTGACTCACGCGCATCAGGACAAGA TGGGCGGTATGGACGCGCTGCATGCGGCGGGG (SEQ ID NO: 44)	
OXA gBlock 1	CGAAGCCAATGGTGACTATATTATTCGGGCTAAAACT GGATACTCGACTAGAATCGAACCTAAGATTGGCTGGT GGGTCGGTTGGGTTGAACTTGATGATAATGTGTGGTTT TTTGCGATGAATA (SEQ ID NO: 45)	DNA
MecA gBlock1	TACAACTTCACCAGGTTCAACTCAAAAAATATTAACA GCAATGATTGGGTTAAATAACAAAACATTAGACGATA AAACAAGTTATAAAATCGATGGTAAAGGTTGGCAAAA AGATAAATCTTGGGGTGGTTACAACGTTACAAGATAT GAAGTGGTAAATGGTAATATCGACTTAAAACAAGCAA TAGAATCATCAGATAACATTTTCTTTGCTAGAGTAGCA CTCGAATTAGGCAGTAAGAAATTTGAAAAAGGCATGA AAAAACTAGGTGTTGGTGAAGATATACCAAGTGATTA TCCATTTTATAATGCTCAAATTTCAAACAAAAATTTAG ATAATGAAATATTATTAGCTGATTCAGGTTACGGACAA GGTGAAATACTGATTAACCCAGTACAGATCCTTTCAAT CTATAGCGCATTAGAAAATAATGGCAATATTAACGCA CCTCACTTATTAAAAGACACGAAAAACAAAGTTTGGA AGAAA (SEQ ID NO: 46)	
hHPRT1	CTCTGTATGTTATATGTCACATTTTGTAATTAACAGCTT GCTGGTGAAAAGGACCCCACGAAGTGTTGGATATAAG CCAGACTGTAAGTGAATTACTTTTTTTGTCAATCATTT AACCATCTTTAACCTAAAAGAGTTTTATGTGAAATGGC TTATAATTGCTTAGAGAATATTTGTAGAGAGGCACATT TGCCAGTATTAGATTTAAAAGTGATGTTTTCTTTATCT AAAT (SEQ ID NO: 47)	DNA
DENV ssRNA target	UGACGAAGACCAUGCUCACUGGACAGAAGCAAAAAU GCUGCUGGACAACAUCAACACACCAGAAGGGAUUAU ACCAGCUCUCUUUGAACCAGAAAGGGAG (SEQ ID NO: 48)	RNA
ZIK ssRNA target	CCACACUGGAACAACAAAGAAGCACUGGUAGAGUUC AAGGACGCACAUGCCAAAAGGCAAACUGUCGUGGUU CUAGGGAGUCAAGAAGGAGCAGUUCACA (SEQ ID NO: 49)	RNA
HANT ssRNA target	AGAGGCAACUUGCAGAUUUGGUGGCAGCUCAAAAAU UGGCUACAAAACCAGUUGAUCCAACAGGGCUUGAGC	RNA

Name	Sequence	Nucleic acid
	CUGAUGAUCAUCUAAAGGAAAAAUCAUC (SEQ ID NO: 50)	

[00438] gBlocks (in Table 9) are double stranded DNA templates synthetize by IDT of about 100-500 nt, whose sequences include the target of interest. The specific cleavage assay containing 1 ug of gBlock target sequences is conducted in buffer NEB 3 with 30nM Cas (Cas9.1, Cas9.2, Cas9.3, Cas9.4 Cas12a.1, Cas12p, Cas12q), 30 nM crRNA against the specific sequences, during 2 h at 37 °C. Reactions are stopped by 10 min at 70°C. The products are cleaned up using PCR purification columns (QIAGEN) and visualized in 1% agarose gel pre-stained with SYBER Gold (Invitrogen). To identify the type of cut (staggered/blunt) aliquots of digestion products are run in 1% of agarose gel and bands corresponding to cleaved target were gel extracted using DNA Clean & Concentrator kit (Zymo Research). The purified products were sequenced using specific primers and analyzed by DNASTART. For collateral activity assays, we used buffer NEB 3 with 30nM Cas (Cas9.1, Cas9.2, Cas9.3, Cas9.4, Cas12a.1, Cas12p, Cas12q), 30 nM crRNA and 1nM ssDNA activator containing the target sequence during 10, 20, 40 and 60 min at 37C. The reactions were initiated by addition of 250 nM M13 ssDNA or M13 dsDNA plasmid (NEB). Reactions were stopped by 10 min at 70°C. Products were separated by 2% agarose gel pre-stained with SYBER Gold (Invitrogen)

Fluorescence detection of collateral activity

[00439] Fluorescence detection can be conducted to determine collateral activity. 30 nM Cas12 was complexed with 30 nM crRNA and 50 nM DNaseAlertTM substrate (IDT) in Buffer NEB 2.1 at 37°C in a 40 µl reaction final volume. The reaction can be monitored in a fluorescence plate reader for up to 30 min at 37°C with fluorescence measurements taken every 2 min in HEX channel (λex: 536 nm; λem: 556 nm). The resulting data can be background-corrected using the readings obtained in the absence of target. For the FQ detection of collateral cleavage of dsDNA/ssDNA and dsRNA/ssRNA DNaseAlertTM (IDT) and RNaseAlert®-1 was used respectively.

Velocity of cleavage cis and trans

[00440] The initial velocity (V0) can be calculated by fitting a linear regression and plotted against the substrate concentration to determine the Michaelis-Menten constants (GraphPad

Software), according to the following equation: $Y = (Vmax \times X)/(Km + X)$, where X is the substrate concentration and Y is the enzyme velocity. The turnover number (kcat) is determined by the following equation: $kcat = Vmax/Et$, where $Et = 0.1$ nM.

Example 2: Determination of endonuclease activity

[00441] It was investigated whether the novel Cas12a.1 and the Cas12p of the disclosure supplied only with crRNA could cleave target DNA *in vitro*. The Cas12a.1 and the Cas12p were designed, overexpressed, purified *in* vitro and used to form a complex with a crRNA against a specific target. It was found that the presence of the Cas12 protein and the cRNA are sufficient for forming an active complex for mediating DNA cleavage.

Example 3: Determination of PAM sequence specificities

[00442] To demonstrate a PAM sequence cleavage-dependent action of the Cas12a.1 and the Cas12p of the disclosure, ten different PAM motifs were designed, following a specific target sequence. Using these, of the ten motifs tested, TCTN and TGTN were identified as efficient PAM sequences for Cas12a.1 and Cas12p, respectively. FIG. 8 shows bar graphs for the PAM sequence preferences of Cas12a.1 and Cas12p for the ten PAM motifs, measuring the performance of the Cas12a.1 and the Cas12p using fluorescence assays. The resulting fluorescence data were background-subtracted.

Example 4: Demonstration of collateral activity of Cas12a.1 and Cas12p, and their ability to cut ssDNA and RNA reporters

[00443] It was investigated whether the Cas12a.1 and the Cas12p proteins of the disclosure were able to cut dsDNA or RNA. Cas12a.1-gRNA or Casp-gRNA complexes were mixed with sample (positive and negative) and a reporter to react in presence of a target. In these examples, a custom ssDNA fluorescently labeled reporter (5' FAM-TTATTATT- 3IABkFQ 3'- IDT) (SEQ ID NO: 121) and a commercial fluorescently labeled reporter RNA reporter (Cat N 11-04-03-03-IDT) were used.

[00444] FIG. 9B shows collateral activity of the Cas12a.1 and Cas12p proteins of the disclosure, using the Hanta virus as an exemplary target. Cas12a.1 and Cas12p were incubated with their respective gRNAs to target Hanta to form a 1uM complex and were exposed to the DNA target at concentration of 10 nM; added to the mix were fluorescently labeled ssDNA or

RNA reporters, at a concentration between 1 and 0.5 uM. Controls did not contain the specific DNA target. Collateral activity was observed only in the presence of target. Cas12a.1 shows ssDNA collateral cleavage for ssDNA but not for RNA, under these conditions. On the other hand, Cas12p exhibited collateral cleavage activity of both ssDNA and RNA reporters. The RNA substrate used for this and other examples provided herein was RNaseAlert®-1 Substrate (25 single use tubes. Catalog No. 11-04-03-03- IDT). The exemplary ssDNA reporter used for this and other examples provided herein was (5' FAM-TTATTATT- 3IABkFQ 3'- IDT) (SEQ ID NO: 121).

[00445] FIG 9C shows that Cas12p exhibits both ssDNA and RNA reporter collateral clevage using as a SARS-CoV-2 inactivated virus as sample as the target.

Example 5: Testing of Thermostability

[00446] The activities of Cas12a.1and Cas12p were tested at different temperatures.

[00447] FIG. 10 shows activity of the Cas12a.1 and Cas12p proteins at 25°C, using1 uM complex, 300nM Reporter SARS-CoV-2 (Spn2 target) at1 minute and 5 minutes as endpoint for the readout.

[00448] FIGS. 10 and 14 shows that Cas12p perform equally well at 25°C as it does at 37°C.

[00449] FIG. 15 shows the differential performance of Cas12p vs. LbCas12a in producing a fluorescence signal by reporter cleavage at 25°C. LbCas12a and Cas12p were incubated with their respective gRNAs to target N gene of SARS-CoV-2 to form a 1 uM complex. The target was the same for both and was provided at a concentration of 10nM. 600nM ssDNA reporter was added into the reaction mix (50 mM NaCl, 10 mM Tris-HCl, 10 mM MgCl2 and 100 µg/ml BSA). Collateral clevage was mesuared by fluorescence and the readout was performed in real time. FIG. 16 shows the differential performance of Cas12p vs. LbCas12a at 25°C, using SARS-CoV-2 as a target, described in Example 10.

Example 6: Testing of various salt concentrations

[00450] The activities of Cas12a.1and Cas12p were tested at various NaCl concentrations; Cas12a.1 and Cas12p and were shown to maintain functionality, FIG. 11 shows the activity of the two proteins at various NaCl concentrations. The resulting fluorescence data was background-subtracted.

Example 7: Testing of various commercial buffers

[00451] In various commercial buffers, the Cas12a.1 and the Cas12p of the disclosure showed different performances. FIG. 12 shows the performance of the Cas12a.1 and the Cas12p of the disclosure in three different commercial buffers. The resulting fluorescence data was background-subtracted.

Example 8: Use of Cas12a.1 and Cas12p for the detection of Hantavirus

[00452] Hantaviruses are a family of viruses spread mainly by rodents and can cause various disease symptoms in people worldwide. Infection with any hantavirus can produce hantavirus disease in people. Described below is the use of the novel Cas12a.1 and Cas12p proteins of the disclosure for the detection of Hantavirus.

Primer Design and CRISPR RNA Guide Selection

[00453] Provided below is the Hantavirus genome, Andes virus segment S complete sequence

(NCBI Reference Sequence: NC_003466.1)

TAGTAGTAGACTCCTTGAGAAGCTACTGCTGCGAAAGCTGGAATGAGCACCCTCCAAGAATTGCAGGAAA
ACATCACAGCACACGAACAACAGCTCGTGACTGCTCGGCAAAAGCTTAAGGATGCCGAGAAGGCAGTGGAGGT
GGACCCGGATGACGTTAACAAGAGCACACTACAAAGTAGACGGGCAGCTGTGTCTACATTGGAGACCAAACTC
GGAGAACTTAAGAGGCAACTTGCAGATTTGGTGGCAGCTCAAAAATTGGCTACAAAACCAGTTGATCCAACAG
GGCTTGAGCCTGATGATCATCTAAAGGAAAAATCATCTCTGAGATATGGGAATGTCCTGGATGTTAATTCAAT
TGATTTGGAAGAACCGAGTGGACAGACTGCTGATTGGAAGGCTATAGGAGCATACATCTTAGGGTTTGCAATT
CCGATCATCCTAAAGGCCTTATACATGCTGTCAACCCGTGGGAGACAAACTGTGA
AAGACAACAAAGGGACCAGGATAAGGTTTAAGGATGATTCTTCCTTTGAAGAAGTCAATGGGATACGTAA
ACCAAAACACCTTTACGTCTCAATGCCAACTGCACAGTCCACTATGAAGGCTGAAGAAATCACGCCAGGA
CGATTTAGGACAATTGCTTGTGGCCTTTTTCCAGCACAGGTCAAAGCCCGAAATATAATAAGTCCTGTAA
TGGGAGTAATTGGATTTGGCTTCTTTGTAAAGGATTGGATGGATCGGATAGAAGAGTTTCTGGCTGCAGA
GTGTCCATTCTTACCTAAGCCAAAGGTCGCCTCAGAAGCCTTCATGTCTACCAATAAGATGTATTTTCTG
AACAGACAGAGACAAGTCAATGAATCTAAGGTTCAAGATATTATCGATTTGATAGACCATGCTGAGACCG
AGTCTGCTACCTTGTTTACAGAGATTGCAACACCCCATTCAGTCTGGGTGTTTGCATGTGCACCTGACCG
GTGCCCTCCAACTGCATTGTATGTTGCAGGGGTACCGGAACTTGGTGCATTTTTTTCTATCCTTCAGGAC
ATGCGTAATACCATCATGGCATCTAAATCTGTAGGGACTGCAGAAGAGAAGCTAAAGAAAAAATCTGCCT
TCTACCAATCATACCTAAGAAGGACACAATCTATGGGAATCCAACTGGACCAGAAGATCATAATCCTTTA
CATGCTATCATGGGGTAAAGAAGCTGTGAATCACTTCCATCTTGGTGATGATATGGACCCTGAACTCAGG
CAGCTAGCACAATCTCTGATCGATACTAAGGTGAAGGAGATCTCCAACCAAGAGCCACTTAAGTTGTAGG
TGCTTAATGAAATCATGATTGAAGAAAGACTTTCCGGGCTTGTGCCACATATTAATCATCTCAGGACCTA
TCCTTAATGTGATTAATAGGGTTTTATTATAAGGGCAGTTAATGGGGTTGGTTACTAACTATGGGTAAGG
GTTCATTACCATTTTTGCACTAGGGTTAAAGGGCCACTACATTGTATTTGCACTAAGGGAAATGGGAGGT
GGGTTAGTTTGTATTTAGTTGTTAAGTTTTTTATAATCATATGTTAATGAGGAATTAGCTATATGATATC
ACTGATTGATTGGCTATTTTTAGGTTAAGTAATTGTAGTTAAATAGTTGTGTTAAGTTAGTATGTTAAGG
TTTATAGGTTAAGATTTACTAACAATCATATTATGTCATTAGATGTAAATTTCATTCCTGGCTTGCTTCT
GCTTTCGCATTGCTAACCTACAACAAGACTACCTCACCCACTACCCCTCCCCTATTCTACCTCAACACAT
ACTACCTCACATTTGATTTTTCTTGATTGCTTTTCAAGGAGCATACTACTA (SEQ ID NO: 69)

[00454] The following exemplary sequence was selected as the target of the spacer gRNA, for Hantavirus detection: GTGGCAGCTCAAAAATTGGCTAC (SEQ ID NO: 70) (underlined above). Other sequences can be selected for targeting.

CRISPR guide design and synthesis

[00455] A gRNA was designed, with a spacer specific to the Hantavirus target sequence. Shown below is the guide (includes direct repeat (single underline) + target complementary sequence (double underline)): AAATTTCTACTGTAGTAGAT GTGGCAGCTCAAAAATTGGCTAC (SEQ ID NO: 249)

[00456] For natural expression and processing of the gRNA, a minimal array with direct repeat from Cas12a.1 and Cas12p and the target complementary sequence was cloned in the Cas expression vector. The CRISPR complex was formed *in vivo* in the expressing bacteria NiCo21(DE3) Competent E. coli and purified from bacteria extracts. In other variations, the guide can be synthesized and complexed with a Cas protein *in vitro*.

[00457] The complex was added to a mix which contained a molecular reporter with a fluorochrome. The sample to be tested was added to the mix. The sample to be tested may be: a sample directly obtained from a subject; a sample obtained from a subject and then diluted and/or treated; DNA (may be amplified) or RNA from a sample taken from a subject; or the sample to be tested may be cDNA made from RNA from the sample. The sample may be further amplified, for example using RPA (Recombinase Polymerase Amplification, e.g. using RPA TwistAmp Basic (TABAS03)).

[00458] The components for formation of the CRISPR complex is shown in Table 10, mixed in that order. The complex was made, and allowed to incubate for 10 minutes at room temperature.

Table 10

Component	[Stock]	[Final]	Volume (1X)
Nuclease free water			15.05
Buffer NEB 2.1	10X	1X	2.0
RNA guide Work solution	300nM	30 nM	2.2
Cas12a.1 Work solution	1uM	30nM	0.8

Component	[Stock]	[Final]	Volume (1X)
TOTAL			20.0

[00459] The components for formation of the CRISPR mix is shown in Table 11, mixed in that order.

Table 11

Component	[Stock]	[Final]	Volume (1X)
Nuclease free water			16.00
Buffer NEB 3.1	10X	1X	2.0
CRISPR complex	300nM	30 nM	20.0
Molecular reporter FAM	50uM	2.5uM	2.0
DNA background	400ng/ul	200 ng	0.5
Amplified Sample	*		4.0
TOTAL			40.0

Results readout

[00460] The reaction was monitored in a fluorescence plate reader for up to 30 min at 37°C with fluorescence measurements taken every 2 min or in the final endpoint in HEX channel (λex: 536 nm; λem: 556 nm). The resulting data are background-corrected using the readings obtained in the absence of target.

[00461] FIG. 9A shows specific cleavage activity of the Ca12a.1 and Cas12p proteins of the disclosures with the Hanta target. A pGEM plasmid was cloned with the Hanta target (pGEM-Hanta) and used to demonstrate specific cleavage activity of Cas12a.1 and Cas12p. Cas12a.1 and Cas12p were incubated with their respective gRNAs to target the Hanta target and exposed to gGEM-Hanta plasmid or gGEM plasmid without target for 2 hours at 37°C. Arrows shows that pGEM-Hanta plasmid is cut but pGEM is not, demonstrating that the cleavage is specific to the Hanta target.

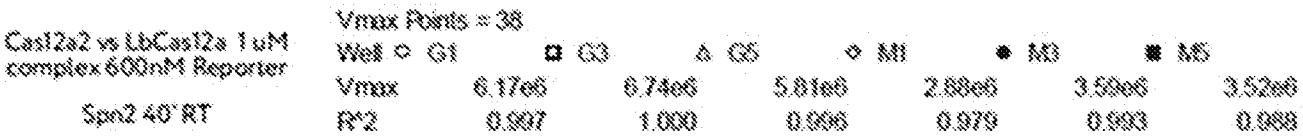
[00462] Using collateral activity, Hantavirus RNA was able to be detected in a picomolar concentration in less than one hour, as shown in FIG. 13. FIG. 13 shows sensitivity curves without RPA of the Cas12a.1 and the Cas12p of the disclosure, for various target concentrations measured for 30 minutes.

Example 9: Cas12p characterization

[00463] Cas12p was further characterized and compared to LbCas12a (SEQ ID NO: 122 (SEQ ID NO: 242 from US9790490)) to support the characteristics of this novel Cas12 subtype.

[00464] FIG. 14 shows that the amount of fluorescence detection by Cas12p for a target DNA reverse transcribed from SARS-CoV-2 RNA was equal at both 37°C and 25°C, indicative of thermostability and function and room temperature.

[00465] FIG. 15 and the below show the kinetic performance of Cas12p vs. LbCas12a at room termperature.



[00466] FIG. 16 further shows the differential performance of Cas12p vs. LbCas12a at room termperature.

[00467] As noted above, FIG. 9A shows specific cleavage activity of the Ca12a.1 and Cas12p proteins of the disclosures with an exemplary Hanta virus target, as described in the above example. FIG. 9B shows collateral activity of the Cas12a.1 and Cas12p proteins of the disclosure, using the Hanta virus as an exemplary target, as described in the above example. FIG. 9C shows collateral activity of the novel Cas12p protein for SARS-CoV-2 target described in Exampe 10.

[00468] FIG. 17 shows the ability of Cas12p to cleave both a ssDNA and RNA reporter, as tested across various targets as exemplary (Hanta virus, SARS-CoV-2). Cas12p was incubated with a gRNAs directed to the Hanta virus or SARS-CoV-2 virus to form a 1uM complex and was exposed to the DNA target at 10 nM concentration adding into the mix a ssDNA or RNA fluorescence marked reporter at a concentration between 1 and 0,5 uM. Controls did not have the specific DNA target. Collateral activity is seen only in the presence of target for both ssDNA and RNA.

Example 10: Use of Cas12a.1for the detection of SARS-CoV-2

[00469] Provided here is an example of the use of Cas12p for the detection of SARS-CoV-2 in upper respiratory specimens during the acute phase of infections. Positive results are indicative of the presentence of SARS-CoV-2 RNA. Further clnical correlation with patient history and other diagnostic information could be utilized to determine patient infection status.

Assay

[00470] RNA was purified from 140 µl of nasopharyngeal/oropharyngeal sample using QIAmp Viral RNA Mini Kit (QIAGEN) as instructed in the user guide and eluting in 60 µl. If RNA was not tested immediately, the RNA was stored at < -70°C.

[00471] After RNA purification, detection of SARS-CoV-2 genomic RNA using CASPR Lyo-CRISPR SARS-CoV-2 kit was carried out using a two-step procedure as summarized in FIG. 18 and outlined below.

[00472] Step 1: The purified RNA was subject to reverse transcription and amplification. Reverse transcription and amplificaiton of 5 µl of purified RNA using reverse transcription loop-mediated isothermal amplification (RT-LAMP) with primer sets specifically designed to target a highly conserved N gene of the SARS-CoV-2 viral genome were carried out.

[00473] The RT-LAMP reaction was based on a total of three (3) pair of primers that amplify a specific sequence in the N gene of SARS-CoV-2 RNA.

[00474] The RT-LAMP reaction was performed by incubating the reaction mix at 62°C for 30 minutes.

[00475] Step 2: Following the RT-LAMP reaction, the detection of amplified viral target was carried out using a Cas12a.1 ribonucleoprotein complex (RNP complex) comprising Cas12a.1 + a gRNA (single molecule guide) targeting the amplified viral N gene sequences from Step 1. The sequence targeted by the gRNA in the cDNA made from the viral RNA was as follows: GATCGCGCCCCACTGCGTTCTCC (SEQ ID NO: 119)

[00476] If the SARS-CoV-2 genomic RNA was present in the sample and was amplified during the RT-LAMP reaction, the gRNA from the RNP complex can bind to the DNA target and trigger the collateral cleavage activity of Cas12a.1, which degrades a 5'FAM-3'Quencher single stranded DNA (ss-DNA) reporter molecule causing the emission of fluorescence. Fluorescence measurements can be performed in standard plate readers with fluorescence capabilities.

[00477] The assay was carried out in less than 60 minutes from start to finish - from obtaining the sample to a readout of the results. FIG. 18 shows a schematic workflow for the detection of SARS-CoV-2 described in this example.

[00478] The additional negative, positive, and extraction controls were included.

[00479] Negative Control: Nuclease-free water was used to identify any potential contamination of the assay run.

[00480] Positive Control: A synthetic sequence identical to the target sequence was provided at a concentration of 2000 cp/ml, in a separate vial. The positive control verified that the assay was performing as expected.

[00481] Extraction controls: Primer sets that target human housekeeping gene RNAse P (for example) were included in the RT-LAMP reaction mix to ensure the proper performance of extraction procedure.

[00482] The reagents used were providedin lyophilized form, reducing manual sources of operator error.

Results

[00483] For negative controls (NTC) a ratio value was calculated between fluorescence (IF) measured at end-point (t=20 min) over fluorescence at the beginning of the run (t=0 min)

$$\text{Ratio Value } (X) = \frac{IF_{NTC}\, t^{20min}}{IF_{NTC} t^{0min}}$$

[00484] For positive control and clinical samples, a ratio value was calculated between sample reaction fluorescence (IF) measured at end-point (t=20 min) over the corresponding valid negative template control reaction fluorescence measurement at 20 minutes.

- *For positive control (PC)*

$$\text{Ratio Value } (A) = \frac{IF_{PC}\, t^{20min}}{IF_{NTC} t^{20min}}$$

- *For clinical samples*

$$\text{Ratio Value } (A) = \frac{IF_{Sample}\, t^{20min}}{IF_{NTC} t^{20min}}$$

[00485] Once ratio values for controls and samples were calculated results were calculated according to the following criteria for control assays:

Control Assay	Target	Cutoff Value	Result	
			Valid	Invalid
NTC	SARS-CoV-2	2	If ratio if < 2	If ratio > 2
Positive Control	SARS-CoV-2	3	If ratio if > 3	If ratio < 3

[00486] In this example, for unknown clinical samples: a Positive sample would have a ratio value > 3 (a minimum 3-fold increase in fluorescence emission between sample reaction and negative control reaction at t=20 min).

[00487] In this example, a Negative sample would have a ratio value < 3 (less than 3-fold increase in fluorescence emission between a sample reaction and negative control reaction at t=20 min). To confirm a negative result, RNAse P would be expected to have a value >3 (an increase in fluorescence emission between sample reaction and negative control reaction at t=20 min).

Performace Evaluation – Analytical Sensitivity, Limit of Detection

[00488] The Limit of Detection (LoD) study established the lowest concentration of SARS-CoV-2 (genome copies(cp)/µL of input) that could be detected at least 95% of the time.

[00489] To determine LoD, serial dilutions of whole inactivated SARS-CoV-2 were spiked into negative nasopharyngeal samples and processed according to the procedure described above.

[00490] A LoD was determined by testing three (3) replicates of three (3) different dilutions (10 copies/ µl, 5 copies/µl, 2.5 copies/µl) and corresponded to the lowest concentration (5 copies/ µl) at which 3/3 replicates were tested positive. This preliminary LoD (5 copies/µl) was confirmed by testing at 0.5X - 1X – 1.5X - 2X of the preliminary LoD in twenty (20) replicates for each concentration. The LoD was the lowest concentration at which at least 19/20 replicates were tested positive for the target.

[00491] LoD was confirmed at 7.5 copies/ µLwith a detection rate of 95% (19/20). Results are summarized in the following table:

Table 12

Replicates	Ratio Value	Result
1	>3	Positive
2	>3	Positive
3	>3	Positive
4	>3	Positive
5	>3	Positive
6	>3	Positive
7	>3	Positive
8	>3	Positive
9	>3	Positive
10	>3	Positive
11	>3	Positive
12	<3	Negative
13	>3	Positive
14	>3	Positive
15	>3	Positive
16	>3	Positive
17	>3	Positive
18	>3	Positive
19	>3	Positive
20	>3	Positive

Performace Evaluation – Analytical Sensitivity, Inclusivity

[00492] Inclusivity was demonstrated by comparing the SARS-CoV-2 assay primers and gRNA to an alignment of 4703 SARS-CoV-2 sequences available in GISAID as of May 16, 2020. The

dataset was further refined by considering only whole genome sequences (>29000 bp) and by removing low-quality sequences with ambiguous sequencing data (N's) and animal origin. This in-silico analysis indicatd that the that primers and gRNA sequences utilized have a 99.9% homology to all available circulating SARS-CoV-2 sequences.

Performace Evaluation – Analytical Specificity

[00493] The assay 2 was based on a set of primers and a unique gRNA designed for specific detection of SARS-CoV-2.

[00494] To evaluate the analytical specificity, an in-silico analysis using NCBI Blast tool was first performed to confirm the absence of any potential cross-reactivity between any of the primer/gRNA sequences with normal and pathogenic organisms of the respiratory tract.

[00495] Results are summarized in Table 13:

Table 13

Pathogen	% homology for SARS-CoV- 2 gRNA	% homology for SARS-CoV-2 primers
Human coronavirus 229E	<80%	<80%
Human coronavirus HKU1	<80%	<80%
Human coronavirus NL63	<80%	<80%
Human coronavirus OC43	<80%	<80%
MERS-coronavirus	<80%	<80%
SARS-coronavirus	>80%	>80%
Adenovirus	<80%	<80%
Human Metapneumovirus (hMPV)	<80%	<80%
Parainfluenza virus 1-4	<80%	<80%
Influenza A & B	<80%	<80%
Respiratory syncytial virus	<80%	<80%
Enterovirus	<80%	<80%
Rhinovirus	<80%	<80%
Chlamydia pneumoniae	<80%	>80%
Haemophilus influenzae	<80%	<80%
Legionella pneumophila	<80%	>80%
Mycobacterium tuberculosis	<80%	>80%
Streptococcus pneumoniae	<80%	<80%
Streptococcus pyogenes	<80%	<80%
Bordetella pertussis	<80%	>80%
Mycoplasma pneumoniae	<80%	<80%
Pneumocystis jirovecii	<80%	<80%
Candida albicans	<80%	<80%
Pseudomonas aeruginosa	<80%	>80%
Staphylococcus epidermidis	<80%	<80%

Pathogen	% homology for SARS-CoV- 2 gRNA	% homology for SARS-CoV-2 primers
Streptococcus salivarius	>80%	<80%

[00496] These results showed that only a few microorganisms have >80% homology between their genome sequences and at least one of the SARS-CoV-2 primers or gRNA included in the assay.

[00497] To confirm the in-silico evaluation, the same pathogens were *in vitro* to check for both potential cross-reactivity and interference.

[00498] The analysis was performed on a total of 22 pathogens by spiking either genomic DNA/RNA or inactivated strains into the SARS-CoV-2 negative nasopharyngeal sample at the concentration indicated in Table 15 during the lysis step of the extraction procedure and tested using the assay described herein. Each pathogen was tested in triplicate. To discard any false negative results an RNAseP assay was run in parallel to each sample,

[00499] An interference analysis was also evaluated on the microorganisms that showed >80% homology with either the SARS-CoV-2 primers or the gRNA included in the kit. To detect any potential interference the analysis was performed by following the same protocol used for cross-reactivity testing in presence of 3X LoD SARS-CoV-2 (22.5 cp/µl).

[00500] All negative results for pathogens tested were confirmed by a positive result in the RNAseP assay.

Table 14

Pathogens	Source	Tested Concentration	Cross-Reactivity	Interference	RNAse P
Human coronavirus 229E	0810229CFHI	10^5 TCID50/mL	0/3	N/A	3/3
Human coronavirus OC43	0810024CFHI	10^5 TCID50/mL	0/3	N/A	3/3
Human coronavirus HKU1	ATCC VR-1580DQ	10^5 copies/mL	0/3	N/A	3/3
Human coronavirus NL63	ATCC® 3263SD	10^5 copies/mL	0/3	N/A	3/3
SARS-coronavirus	NATSARS-ST	10^5 TCID50/mL	0/3	0/3	3/3
Respiratory syncytial virus	ATCC® VR-1580DQ	10^5 copies/mL	0/3	N/A	3/3
Influenza A	VR-95DQ	10^5 copies/mL	0/3	N/A	3/3

Pathogens	Source	Tested Concentration	Cross-Reactivity	Interference	RNAse P
Influenza B	VR-1885DQ	10^5 copies/mL	0/3	N/A	3/3
Mycobacterium tuberculosis	NR-14867	10^6 copies/mL	0/3	0/3	3/3
Candida albicans	ATCC® 10231D-5	10^6 copies/mL	0/3	N/A	3/3
Pseudomonas aeruginosa	ATCC® 27853D-5	10^6 copies/mL	0/3	0/3	3/3
Staphylococcus epidermis	ATCC® 12228D-5	10^6 copies/mL	0/3	N/A	3/3

[00501] In conclusion, based on in-silico and in vitro analysis it was anticipated no cross-reactivity nor interference between primers/gRNA included in the assay and most common pathogens in the respiratory tract.

Clinical Evaluation

[00502] Clinical evaluation of the assay was performed using nasopharyngeal swabs as clinical samples from male and female adult patients with signs and symptoms of an upper respiratory infection.

[00503] A total of 30 positive samples and 30 negative samples were collected to assess the performance and tested using QIAmp Viral RNA Mini kit for RNA extraction followed by the procedure as described (noted as "Cas12a.1-Based Assay" in Table 15). All samples were also tested using an RT-PCR Test as a comparison method to obtain positive and negative percent agreement values. Results are presented in Table 15 and show 100% positive percent agreement (PPA) and 100% negative percent agreement (NPA) with comparator method.

Table 15

Cas12a.1-Based Assay		FDA EUA RT-PCR Test (Comparator)		PPA (95% CI)	NPA (95% CI)
Positive	Negative	Positive	Negative		
30/30	30/30	30/30	30/30	100% (90.55-100)	100% (90.55-100)

Example 11: Use of Cas12p for the detection of SARS-CoV-2

[00504] Provided here is an example of the use of Cas12p for the detection of SARS-CoV-2 in upper respiratory specimens during the acute phase of infections. Positive results are indicative of the presence of SARS-CoV-2 RNA. Further clnical correlation with patient history and other diagnostic information could be utilized to determine patient infection status.

Assay

[00505] Nasopharyngeal/nasal swab is inserted in 500 uL of Lysis Buffer, vortex is applied for 2 minutes and 100 uL lysed sample is transported into 1,5 mL capacity tube and heated at 95 C for 5 minutes.

[00506] After sample treatment, detection of SARS-CoV-2 genomic RNA using CASPR Direct Lyo-CRISPR SARS-CoV-2 kit was carried out using a two-step procedure as summarized in FIG. 19 and outlined below.

[00507] Step 1: The lysed sample was subject to reverse transcription and amplification. Reverse transcription and amplificaiton of 10 µl of lysed sample using reverse transcription loop-mediated isothermal amplification (RT-LAMP) with primer sets specifically designed to target two highly conserved N gene and one highly conserved ORF1ab gene of the SARS-CoV-2 viral genome were carried out.

[00508] The RT-LAMP reaction was based on a total of three (9) pair of primers that amplify two specific sequences in the N gene and one specific sequence in the ORF1ab gene of SARS-CoV-2 RNA.

[00509] The RT-LAMP reaction was performed by incubating the reaction mix at 62□C for 60 minutes.

[00510] Step 2: Following the RT-LAMP reaction, the detection of amplified viral target was carried out using a Cas12p ribonucleoprotein complex (RNP complex) comprising Cas12p + three gRNAs (single molecule guide) targeting the amplified viral N and ORF1ab gene sequences from Step 1. The sequences targeted by the gRNAs in the cDNA made from the viral RNA were as follows: GATCGCGCCCCACTGCGTTCTCC (SEQ ID NO: 119), AUGGCACCUGUGUAGGUCAACCA (SEQ ID NO:120) and UGUGCUGACUCUAUCAUUAUUGG (SEQ ID NO:123).

[00511] If the SARS-CoV-2 genomic RNA was present in the sample and was amplified during the RT-LAMP reaction, the gRNA from the RNP complex can bind to the DNA target and

trigger the collateral cleavage activity of Cas12p, which degrades a 5'FAM-3'Quencher single stranded reporter molecule causing the emission of fluorescence. Fluorescence measurements can be performed in standard plate readers with fluorescence capabilities.

[00512] The assay was carried out in less than 75 minutes from start to finish - from obtaining the sample to a readout of the results. FIG. 18 and FIG. 19 show a schematic workflow for the detection of SARS-CoV-2.

[00513] The additional negative, positive, and extraction controls were included.

[00514] Negative Control: Nuclease-free water was used to identify any potential contamination of the assay run.

[00515] Positive Control: A synthetic sequence identical to the target sequences was provided at a concentration of 2000 cp/ml, in a separate vial. The positive control verified that the assay was performing as expected.

[00516] Extraction controls: Primer sets that target human housekeeping gene RNAse P (for example) were included in the RT-LAMP reaction mix to ensure the proper performance of extraction procedure.

[00517] The reagents used were providedin lyophilized form, reducing manual sources of operator error.

Results

[00518] For negative controls (NTC) a ratio value was calculated between fluorescence (IF) measured at end-point (t=5 min) over fluorescence at the beginning of the run (t=0 min).

$$\text{Ratio Value (X)} = \frac{IF_{NTC}\, t^{5min}}{IF_{NTC}\, t^{0min}}$$

[00519] For positive control and clinical samples, a ratio value was calculated between sample reaction fluorescence (IF) measured at end-point (t=5 min) over the corresponding valid negative template control reaction fluorescence measurement at 5 minutes.

$$\text{Ratio Value (A)} = \frac{IF_{PC}\, t^{5min}}{IF_{NTC}\, t^{5min}} \qquad \frac{IF_{Sample}\, t^{5min}}{IF_{NTC}\, t^{5min}}$$

[00520] Once ratio values for controls and samples were calculated results were calculated according to the following criteria for control assays:

Table 16

Control Assay	Target	Cutoff Value	Result	
			Valid	Invalid
NTC	SARS-CoV-2	2	If ratio if ≤ 2.5	If ratio > 2.5
Positive Control	SARS-CoV-2	3	If ratio if ≥ 2.5	If ratio < 2.5

[00521] In this example, for unknown clinical samples: a Positive sample would have a ratio value ≥ 2.5 (a minimum 2.5-fold increase in fluorescence emission between sample reaction and negative control reaction at t=5 min).

[00522] In this example, a Negative sample would have a ratio value ≤ 2.5 (less than 2.5-fold increase in fluorescence emission between a sample reaction and negative control reaction at t=5 min). To confirm a negative result, RNAse P would be expected to have a value ≥ 2.5 (an increase in fluorescence emission between sample reaction and negative control reaction at t=5 min).

Performace Evaluation – Analytical Sensitivity, Limit of Detection

[00523] The Limit of Detection (LoD) study established the lowest concentration of SARS-CoV-2 (genome copies(cp)/µL of input) that could be detected at least 95% of the time.

[00524] To determine LoD, serial dilutions of whole inactivated SARS-CoV-2 were spiked into lysis buffer with negative nasal matrix and processed according to the procedure described above.

[00525] A LoD was determined by testing three (5) replicates of three (3) different dilutions (25 copies/ µl, 12.5 copies/µl, 6.125 copies/µl) and corresponded to the lowest concentration (25 copies/ µl) at which 3/3 replicates were tested positive. This preliminary LoD (25 copies/µl) was confirmed in twenty (20) replicates. The LoD was the lowest concentration at which at least 20/20 replicates were tested positive for the target.

[00526] LoD was confirmed at 25 copies/ µLwith a detection rate of 100% (20/20). Results are summarized in the following table:

Table 17

Replicates	Ratio Value	Result
1	>2.5	Positive
2	>2.5	Positive
3	>2.5	Positive
4	>2.5	Positive
5	>2.5	Positive
6	>2.5	Positive
7	>2.5	Positive
8	>2.5	Positive
9	>2.5	Positive
10	>2.5	Positive
11	>2.5	Positive
12	>2.5	Positive
13	>2.5	Positive
14	>2.5	Positive
15	>2.5	Positive
16	>2.5	Positive
17	>2.5	Positive
18	>2.5	Positive
19	>2.5	Positive
20	>2.5	Positive

Performace Evaluation – Analytical Sensitivity, Inclusivity

[00527] Inclusivity was demonstrated by comparing the SARS-CoV-2 assay primers and gRNAs to an alignment of 4703 SARS-CoV-2 sequences available in GISAID as of May 16,

2020. The dataset was further refined by considering only whole genome sequences (>29000 bp) and by removing low-quality sequences with ambiguous sequencing data (N's) and animal origin. This in-silico analysis indicatd that the that primers and gRNA sequences overall utilized have a 100% homology to all available circulating SARS-CoV-2 sequences.

Performace Evaluation – Analytical Specificity

[00528] The assay 2 was based on a set of primers and gRNAs designed for specific detection of SARS-CoV-2.

[00529] To evaluate the analytical specificity, an in-silico analysis using NCBI Blast tool was first performed to confirm the absence of any potential cross-reactivity between any of the primer/gRNA sequences with normal and pathogenic organisms of the respiratory tract.

[00530] Results are summarized in Table 18:

Table 18

Pathogen	Target 1 (N)		Target 2 (N)		Target 3 (Orf1ab)	
	% Homology with sgRNA	% Homology primers	% Homology with sgRNA	% Homology primers	% Homology with sgRNAs	% Homology primers
Coronavirus 229E	<80	<80	<80	<80	<80	<80
Coronavirus HKU1	<80	<80	<80	<80	<80	<80
Coronavirus NL63	<80	<80	<80	<80	<80	<80
Coronavirus OC43	<80	<80	<80	<80	<80	<80
MERS-coronavirus	<80	<80	<80	<80	<80	<80
SARS-coronavirus	**>80**	**>80**	<80	<80	**>80**	<80
Adenovirus	<80	<80	<80	<80	<80	<80
Human Metapneumovirus (hMPV)	<80	<80	<80	<80	<80	<80
Parainfluenza virus 1-4	<80	<80	<80	<80	<80	<80
Influenza A & B	<80	<80	<80	<80	<80	<80
Respiratory syncytial virus	<80	<80	<80	<80	<80	<80
Enterovirus	<80	<80	<80	<80	<80	<80
Rhinovirus	<80	<80	<80	<80	<80	<80
Chlamydia pneumoniae	<80	**>80**	<80	<80	<80	<80

Pathogen	Target 1 (N)		Target 2 (N)		Target 3 (Orf1ab)	
	% Homology with sgRNA	% Homology primers	% Homology with sgRNA	% Homology primers	% Homology with sgRNAs	% Homology primers
Haemophilus influenzae	<80	<80	<80	<80	<80	<80
Legionella pneumophila	<80	**>80**	<80	<80	**>80**	<80
Mycobacterium tuberculosis	<80	**>80**	<80	<80	<80	<80
Streptococcus pneumoniae	<80	<80	<80	<80	<80	<80
Streptococcus pyogenes	<80	<80	<80	<80	<80	**>80**
Bordetella pertussis	<80	**>80**	<80	<80	<80	<80
Mycoplasma pneumoniae	<80	<80	<80	<80	<80	<80
Pneumocystis jirovecii	<80	<80	<80	<80	<80	<80
Candida albicans	<80	<80	<80	<80	<80	**>80**
Pseudomonas aeruginosa	<80	**>80**	<80	<80	<80	<80
Staphylococcus epidermidis	<80	<80	<80	<80	<80	<80
Streptococcus salivarius	**>80**	<80	<80	**>80**	<80	<80

[00531] These results showed that only a few microorganisms have >80% homology between their genome sequences and at least one of the SARS-CoV-2 primers or gRNA included in the assay.

[00532] To confirm the in-silico evaluation, the same pathogens were *in vitro* to check for both potential cross-reactivity and interference.

[00533] The analysis was performed on a total of 22 pathogens by spiking either genomic DNA/RNA or inactivated strains into the SARS-CoV-2 negative lysed ssample at the concentration indicated in Table 19 and tested using the assay described herein. Each pathogen was tested in triplicate. To discard any false negative results an RNAseP assay was run in parallel to each sample,

[00534] An interference analysis was also evaluated on the microorganisms that showed >80% homology with either the SARS-CoV-2 primers or the gRNA included in the kit. To detect any potential interference the analysis was performed by following the same protocol used for cross-reactivity testing in presence of 3X LoD SARS-CoV-2 (75 cp/µl).

[00535] All negative results for pathogens tested were confirmed by a positive result in the RNAseP assay.

Table 19

Pathogens	Source	Tested Concentration	Cross-Reactivity	Interference	RNAse P
Human coronavirus 229E	0810229CFHI	10^5 TCID50/mL	0/3	N/A	3/3
Human coronavirus OC43	0810024CFHI	10^5 TCID50/mL	0/3	N/A	3/3
Human coronavirus HKU1	ATCC VR-1580DQ	10^5 copies/mL	0/3	N/A	3/3
Human coronavirus NL63	ATCC® 3263SD	10^5 copies/mL	0/3	N/A	3/3
SARS-coronavirus	NATSARS-ST	10^5 TCID50/mL	0/3	3/3	3/3
Respiratory syncytial virus	ATCC® VR-1580DQ	10^5 copies/mL	0/3	N/A	3/3
Influenza A	VR-95DQ	10^5 copies/mL	0/3	N/A	3/3
Influenza B	VR-1885DQ	10^5 copies/mL	0/3	N/A	3/3
Mycobacterium tuberculosis	NR-14867	10^6 copies/mL	0/3	3/3	3/3
Candida albicans	ATCC® 10231D-5	10^6 copies/mL	0/3	3/3	3/3
Pseudomonas aeruginosa	ATCC® 27853D-5	10^6 copies/mL	0/3	3/3	3/3
Staphylococcus epidermis	ATCC® 12228D-5	10^6 copies/mL	0/3	N/A	3/3
Streptococcus salivarius	HM-121D	10^6 copies/mL	0/3	3/3	3/3
Pooled human nasal fluid	991-13-P-1	N/A	0/3	N/A	3/3
Legionella pneumophila	ATCC® 33152D-5	10^6 copies/mL	0/3	3/3	3/3
Haemophilus influenzae	ATCC® 51907D-5™	10^5 copies/mL	0/3	N/A	3/3
Streptococcus pyogenes	ATCC® 12344D-5™	10^6 copies/mL	0/3	3/3	3/3
Streptococcus pneumoniae	ATCC® 700669D-5™	10^6 copies/mL	0/3	N/A	3/3
MERS-CoV	NR-45843	10^5 copies/mL	0/3	N/A	3/3
Rhinovirus	NR-51453	10^5 copies/mL	0/3	N/A	3/3
Chlamydia pneumoniae	10^6 copies/mL	10^6 copies/mL	0/3	3/3	3/3
Bordetella pertussis	9797D-5	10^6 copies/mL	0/3	3/3	3/3

[00536] In conclusion, based on in-silico and in vitro analysis it was anticipated no cross-reactivity nor interference between primers/gRNAs included in the assay and most common pathogens in the respiratory tract.

Clinical Evaluation

[00537] Clinical evaluation of the assay was performed using nasopharyngeal swabs as clinical samples from male and female adult patients with signs and symptoms of an upper respiratory infection.

[00538] A total of 47 positive samples and 43 negative samples were collected to assess the performance. All samples were also tested using an RT-PCR Test as a comparison method to obtain positive and negative percent agreement values. Results are presented in Table 20 and show 97.9% positive percent agreement (PPA) and 100% negative percent agreement (NPA) with comparator method.

Table 20

		RT-PCR Reference Method		
		Positive	**Negative**	**Total**
Direct CASPR Lyo-CRISPR SARS-CoV-2	Positive	46	0	46
	Negative	1	43	44
	Total	47	43	90

[00539] FIG. 20 shows that Cas12p has a minimal background signal after 30-60 minutes of cleavage activity. This provides advantages at low viral concentrations, and indicates stability of the lyophilized format. FIG. 21 shows that a diagnostics assay using Cas12p at room temperature, can be read out on a paper format. FIG. 22 shows that a diagnostics assay using Cas12p at room temperature can be read in well plate with a fluorescent detector.

Example 12: SARS-CoV-2 detection using a Cas12p and a RNA guide

[00540] Lyophilized beads with a RNA based reporter were used to detect SARS-CoV-2 RNA in patient and control samples. A subset of the samples described in Example 11 were used for this example. Cas12p was pre-incubated with their respective sgRNA and labeled RNA reporter was added before the lyophilization process. Pre amplified RT-LAMP product was used as input. Input for the RT-LAMP reaction were lysed sample from patient and negative control nasopharyngeal swabs. FIG. 19 shows the workflow for SARS-CoV-2 detection using a

Cas12p/guide complex, using a RNA reporter, from a sample. FIG. 24 shows the results of SARS-CoV-2 detection using a Cas12p and a RNA reporter from patient samples and negative control samples in lyophilized format, at 30 minutes at 37°C (n=16).

Example 13: Testing specific cleavage activity

[00541] FIG. 25: It was investigated whether the Cas12a.1 and the Cas12p of the disclosure are able to cut dsDNA when complexed with its guide. In these examples, the target was a Hanta virus dsDNA sequence (100 pb) cloned into the commercial pGEM®-T Easy vector from Promega (Cat.# A1360). Negative controls included the empty pGEM®-T Easy vector. The positive control included the pGEM®-T Easy vector/Hanta dsDNA target linearized by cut with NdeI restriction endonuclease from NEB (Cat.# R0111L). The procedure was as follows: 100 nM of Cas12a.1 or Cas12p were complexed with 100 nM of sgRNA to target the Hanta sequence, in a commercial NEBuffer™ 2.1 (Cat.# B7202S) for 15 min at RT. Controls with Cas enzyme not complexed with its guide were included. Then, 5 ng/uL of target was added, in a final reaction volume of 20 uL. Reactions were incubated at 37 or 25 °C for 0, 30, 60 or 90 min, and ended by addition of 50 mM EDTA. Then, the samples were centrifuged at 12000 g for 10 min and mixed with 6X Gel Loading Dye from NEB (Cat# B7024S). Samples were analyzed in a 0.8% TBE-agarose gel. Fast DNA Ladder from NEB (Cat.# N3238S) was used to assess the molecular weight of the species. After electrophoresis the gel was stained for 30 min with a fresh solution of SYBR™ Gold Nucleic Acid Gel Stain from Invitrogen (Cat # S11494) and imaged on VersaDoc™ Imaging System (Bio-Rad). Figure 1 shows the results of the assay. Cas12a.1 could linearize the totality of the plasmid after 90 min at 37°C, while Cas12p lasted only 60 min to achieve comparable results.

[00542] FIG. 26: It was investigated whether the Cas12a.1 and Cas12p of the disclosure are able to cut ssDNA when complexed with its guide. In these examples, the target consisted of a custom ssDNA fluorescence marked sequence (3´FAM-ssDNA) of 70 nucleotide length from IDT (5´-TCA TTT AGA AAG TAG ATA TTG ATT GAT TTT AGC GAA AGC CAA TTT TTG AGC TGC CAC TGA TGT AAA AGT T-3´-6-FAM; SEQ ID NO: 124) targeted to Hanta virus. Negative control included a custom anti-sense ssDNA sequence (ASssDNA) of 120 nucleotide length from IDT (5´- GCT ATC TTA ATC CTT AAT CTA TCC TCA AAC GTT CTA TTA ATG GCC GTG TCA ATC AAT ATC TAC TTT CTA AAT GAA ACT TTT ACA TCA GTG GCA GCT CAA AAA TTG GCT TTC GCT AAA ATC-3´; SEQ ID NO: 125) also

targeted to Hanta virus. The procedure was as follows: 10 pmol of Cas12a.1 or Cas12p, were complexed with 10 pmol of sgRNA to target Hanta sequence, in commercial NEBuffer™ 2.1 (Cat.# B7202S) for 15 min at RT. Controls with Cas enzyme not complexed with its guide were included. Then, 10 pmol of 3´FAM-ssDNA or alternatively ASssDNA was added, in a final reaction volume of 10 uL. Reactions were incubated at 37 °C for 0, 0.5, 1 or 5 min, and ended by addition of 2X Novex™ TBE-Urea Sample Buffer from Invitrogen (Cat # LC6876) followed by heating at 95 °C for 3 min. Samples were centrifuged at 12000 g for 10 min and analyzed on 15% Mini-PROTEAN® TBE-Urea Gel from Bio-Rad (Cat.# 4566056). Oligo length standards from IDT (Cat.# 51-05-15-02) was used to assess the molecular weight of the species. Gels were first imaged on VersaDoc™ Imaging System (Bio-Rad) and then were stained for 30 min with a fresh solution of SYBR™ Gold Nucleic Acid Gel Stain from Invitrogen (Cat # S11494) to visualize the non-fluorescence marked sequence of ASssDNA and the non-fluorescence marked ladder. Figure 2 shows the results of the assay. Cas12a.1 and Cas12p demonstrated specific ssDNA cleavage of the 3´FAM- ssDNA substrate (S), with the production of a ~40 nucleotide length product (P). The two Cas enzymes were unable to cut the ASssDNA sequence (NTC). The reactions took place in the timeframe of seconds to few minutes.

[00543] FIG. 27**:** It was investigated whether the Cas12a.1 and Cas12p of the disclosure are able to cut ssRNA, when complexed with its guide. In these examples, the target consisted of a ssRNA sequence obtained by *in vitro* transcription (IVT) and targeted to Hanta virus. Negative control included a custom non-target ssRNA sequence of 65 nucleotide length from IDT (5´-TAA GCG CCC TTG CGC TTT CCC CAG CCT TCG GGT TGG TTG CCT TTT AGT GCA AGG GCG CGA TTA TT -3´; SEQ ID NO: 126). Positive control included a custom ssDNA sequence of 120 nucleotide length from IDT (5´- GAT TTT AGC GAA AGC CAA TTT TTG AGC TGC CAC TGA TGT AAA AGT TTC ATT TAG AAA GTA GAT ATT GAT TGA CAC GGC CAT TAA TAG AAC GTT TGA GGA TAG ATT AAG GAT TAA GAT AGC- 3´; SEQ ID NO: 127), targeted to Hanta Virus. The procedure was as follows: 150 nM of Cas12a.1 or Cas12p were complexed with 150 nM of sgRNA to target Hanta sequence, in commercial NEBuffer™ 2.1 (Cat.# B7202S) for 15 min at RT. Controls with Cas enzyme not complexed with its guide were included. Then, 5 ng/uL of ssRNA or alternatively non-target ssRNA or ssDNA was added, in a final reaction volume of 10 uL. Reactions were incubated at 37 °C for 0, 1 or 3 h, and ended by addition of 2X Novex™ TBE-Urea Sample Buffer from Invitrogen (Cat # LC6876) followed by heating at 65 °C for 3 min. Samples were centrifuged at 12000 g for 10

min and analyzed on 15% Mini-PROTEAN® TBE-Urea Gel from Bio-Rad (Cat.# 4566056). Low Range ssRNA Ladder from NEB (Cat.# N0364S) was used to assess the molecular weight of the species. Gels were stained for 30 min with a fresh solution of SYBR™ Gold Nucleic Acid Gel Stain from Invitrogen (Cat # S11494) and imaged on VersaDoc™ Imaging System (Bio-Rad). Figure 3 shows the results of the assay. Neither Cas12a.1 nor Cas12p demonstrated specific ssRNA cleavage activity.

Example 14: Characterization of non-specific nuclease activity of Cas12p

[00544] MALDI-TOF MS experiment description**:** Matrix-assisted laser desorption/ionization-time of flight mass spectrometry (MALDI-TOF MS) was employed to monitor the products generated by the unspecific nuclease activity of Cas12p enzyme. Protected DNA (C* C* C* C* C* C* C* C* C* C* C* C* C* C* C* C* C* C* C* C* C*TTATT; SEQ ID NO: 128) and RNA (rC* rC* rC* rC* rC* rC* rC* rC* rC* rC* rC* rC* rC* rC* rC* rC* rC* rC* rC* rC* rC* rC* rC*rUrUrArUrU; SEQ ID NO: 129) reporters were used to ensure a minimal length and minimize the number of possible hydrolysis products. The symbol (*) on C and rC bases indicates the presence of phosphorothioate bonds that are resistant to nuclease degradation. CRISPR reactions with the corresponding reporter were performed with complexes to a final concentration of 75 nM Cas12p : 75 nM sgRNA : 20 nM activator : 2.5 uM DNA reporter or 75 nM Cas12p : 75 nM sgRNA : 10 nM activator : 1.25 uM RNA reporter in a solution containing 1 X Binding Buffer (50 mM NaCl, 10 mM Tris-HCl, 10 mM MgCl2, 1 mM DTT, 100 g/ml BSA, pH 7.9). The reactions were incubated during 1 h at 25 °C for DNA reporter or 6 h at 37 °C for RNA reporter (T1 of reaction, FIG. 28 and FIG. 30). The time zero (T0, FIG. 29 and FIG. 31) of reaction was made as a negative control by heating Crispr reaction before reporter addition. The reactions were purified and analyzed on a PerSeptive Biosystems (ABI)-Voyager-DE RP-MALDI-TOF mass spectrometer, Stanford University. For each reaction, a list was generated with the predicted m/z (mass to charge ratio) of all the possible DNA/RNA cleavage products and all the expected overhangs, as was proposed by Joyner et al. 2012. The observed m/z were correlated to the aforementioned list by the use of a Perl script. Relative intensity of peaks was calculated in relation to the predominant signal of the reaction. DNA reporter hydrolysis gave rise to a unique cleavage product (FIG. 28) meanwhile the hydrolysis of RNA reporter generated multiple fragments, including both 3' hydroxide and phosphate ends (FIG. 30). In all cases, the

predominant hydrolysis species was the one that contained two nucleotides after the protected sequence and 3' hydroxide ends.

[00545] FIG. 28-29 show the mass spectra data of Cas12p reactions using a DNA oligo as the reporter. FIG. 30-31 shows the mass spectra data of Cas12p reactions using a RNA oligo as the reporter.

Example 15: Characterization of the use of chimeric (hybrid) guides

[00546] Guide sequences (hybrid guides, chimeric guides) partially composed of DNA and RNA nucleotides were tested and determined that they can support efficient collateral Cas12p activity. Partial replacement with DNA nucleotides at 3` of sgRNA (Hybrid 4 DNA; 5'AGAUUUCUACUUUUGUAGAUGUGGCAGCUCAAAAAU(TGGC)3'; SEQ ID NO: 130) or a replacement with DNA nucleotides at both 5` and 3` (Hybrid 3/4 DNA; 5'(AGA)UUUCUACUUUUGUAGAU GUGGCAGCUCAAAAAU(TGGC)3'; SEQ ID NO: 131) maintained its activity compared to the unmodified guide sequence (sgRNA; 5'AGAUUUCUACUUUUGUAGAU GUGGCAGCUCAAAAAUUGGC3'; SEQ ID NO: 132). A partial replacement of 8 DNA nucleotides at 3` led to a complete loss of Cas12p collateral cleavage activity (Hybrid 8 DNA; 5'AGAUUUCUACUUUUGUAGAU GUGGCAGCUCAA(AAATTGGC)3'; SEQ ID NO: 133).

[00547] Cas12p was pre-incubated with their respective sgRNA or hybrid guides (1 uM complex). The reaction was initiated by diluting Cas12p complexes to a final concentration of 37.5 nM Cas12p : 37.5 nM sgRNA : 10 nM activator in a solution containing 1× Binding Buffer (50 mM NaCl, 10 mM Tris-HCl, 10 mM MgCl2, 1 mM DTT, 100 g/ml BSA, pH 7.9) and 600 nM TTATTATT ssDNA FQ reporter (SEQ ID NO: 121) substrates in a 40 µl reaction. Reactions (40 µl, 384-well microplate format) were incubated in a fluorescence plate reader (SpectraMax ® M2) for 40 minutes at 25°C with fluorescence measurements taken every 1 minute (ssDNA FQ substrates = λex: 485 nm; λem: 538 nm). The result showed the quantification of maximum fluorescence signal generated after 30 minutes. Non-template negative control (NTC) fluorescence values were calculated from reactions carried out in the absence of target plasmid. Error bars represent the mean ± s.d., where n = 3 replicates.

[00548] FIG. 32 shows that the DNA–RNA chimeric guides used enable efficient collateral Cas12p activity.

Example 16: Characterization of collateral cleavage activity

[00549] FIG. 33 shows agarose gels showing the collateral activity for Cas12a.1 and Cas12p protein/guide complexes using the following substrates: (A) M13mp18 single-stranded DNA (Cat# N4040S, NEB); and (B) M13mp18 RF I double-stranded DNA (Cat# N4018S, NEB). Cas12a.1 and Cas12p exhibit collateral activity and cleavage ssDNA circular DNA (FIG. 33, Panel A), but not dsDNA circular DNA (FIG.33, Panel B). The reaction was initiated by diluting Cas12p/guide or Cas12a.1/guide complexes to a final concentration of 37.5 nM Cas12p : 37.5 nM sgRNA : 10 nM activator or 75 nM Cas12a.1 : 75 nM sgRNA : 10 nM activator in a solution containing 1× Binding Buffer (50 mM NaCl, 10 mM Tris-HCl, 10 mM MgCl2, 1 mM DTT, 100 g/ml BSA, pH 7.9) and 1 uL of M13mp18 single-stranded DNA (Cat# N4040S, NEB) and M13mp18 RF I double-stranded DNA (Cat# N4018S, NEB) at 25°C for 1 h. Control groups without the Cas enzyme, guide or activator were included and non-collateral cleavage was observed.

[00550] Cleavage efficiency Cas12p showed a similar cleavage efficiency for at least the T, A, or C homopolymeric reporter (7 nt in length), whereas Cas12a.1 demonstrated a higher efficiency in poly C cleavage but also cleaved polyA and poly T sequences. Cas12p displayed cleavage at 25°C for T, A, or C homopolymeric reporter evidenced by increased fluorescence, whereas Cas12a.1 only demonstrated cleavage response at 37°C with the 5'6-FAM-TTATTATT-3IABkFQ3' reporter sequence (SEQ ID NO: 121).

[00551] The reaction was initiated by diluting Cas12p or Cas12a.1 complexes to a final concentration of 37.5 nM Cas12p : 37.5 nM sgRNA : 10 nM activator or 75 nM Cas12a.1 : 75 nM sgRNA : 10 nM activator in a solution containing 1× Binding Buffer (50 mM NaCl, 10 mM Tris-HCl, 10 mM MgCl2, 1 mM DTT, 100 g/ml BSA, pH 7.9) and 600 nM ssDNA FQ reporter substrates (5'6-FAM-TTATTATT-3IABkFQ3' (SEQ ID NO: 121), 5'6-FAM-AAAAAAA-3IABkFQ3', 5'6-FAM-TTTTTTT-3IABkFQ3', 5'6-FAM-CCCCCCC-3IABkFQ3' or 5'6-FAM-C*GGGC*GGG-3IABkFQ3' from IDT (Integrated DNA Technologies, Inc)) in a 40 µl reaction. Reactions (40 µl, 384-well microplate format) were incubated in a fluorescence plate reader (SpectraMax ® M2) at 25°C or 37°C with fluorescence measurements taken every 1 minute (ssDNA FQ substrates = λex: 485 nm; λem: 538 nm). Background-corrected fluorescence values were calculated by subtracting fluorescence values obtained from reactions carried out in the absence of target plasmid. Error bars represent the mean ± s.d., where n = 3 replicates. FIG. 34 shows the differential efficiency in cleavage of homopolymeric reporters, at

25°C and 37°C. The results show that Cas12p cleaved poly T, poly A and poly C, whereas Cas12a.1 showed a preference for polyC cleavage.

[00552] The specificity of *trans*-cleavage activity (collateral activity) was tested using a customized ssRNA 5'6-FAM rArUrArUrArUrA-3IABkFQ3' and RNaseAlert™ (a commercially available RNA reporter) from IDT (Integrated DNA Technologies, Inc) as RNA reporters. The results showed that Cas12p is able to cleave RNA reporters used but Cas12a.1 is not. Detection assays were performed at 37°C using Cas12p or Cas12a.1 complexes to a final concentration of 37.5 nM Cas12p : 37.5 nM sgRNA : 10 nM activator or 75 nM Cas12a.1 : 75 nM sgRNA : 10 nM activator in a solution containing 1 × Binding Buffer (50 mM NaCl, 10 mM Tris-HCl, 10 mM MgCl2, 1 mM DTT, 100 g/ml BSA, pH 7.9) and 600 nM of RNA FAMQ reporter substrates (ssRNA 5'6-FAM rArUrArUrArUrA-3IABkFQ3 and RNaseAlert (Cat N 11-04-03-03- IDT)) in a 40 µl reaction. Reactions were incubated in a fluorescence plate reader (SpectraMax ® M2) and background-corrected fluorescence values were calculated by subtracting fluorescence values obtained from reactions carried out in the absence of target plasmid. Error bars represent the mean ± s.d., where n = 3 replicates. FIG. 35 shows the result of these data, and shows the the collateral cleavage ability of Cas12p but not of Cas12a.1, to cleave a RNA reporter.

[00553] The kinetics of collateral cleavage (*trans*-cleavage) activity using DNA and RNA reporters was assessed for Cas12p. Experiments with an RNA substrate showed a cleavage rate of ssRNA only 3-fold slower than a ssDNA reporter. The cleavage rate of Cas12a.1 for the ssRNA substrate was at least 1.10^4-fold slower than for ssDNA, confirming that ssDNA is the choice substrate for Cas12a.1 collateral cleavage. Detection assays were performed at 37°C using Cas12p or Cas12a.1 complexes to a final concentration of 37.5 nM Cas12p : 37.5 nM sgRNA : 10 nM activator or 75 nM Cas12a.1 : 75 nM sgRNA : 10 nM activator in a solution containing 1 × Binding Buffer (50 mM NaCl, 10 mM Tris-HCl, 10 mM MgCl2, 1 mM DTT, 100 g/ml BSA, pH 7.9) and 600 nM of ssDNA FAMQ reporter substrates (ssDNA 5'6-FAM TTATTATT-3IABkFQ3 (SEQ ID NO: 121)) or RNaseAlert (Cat N 11-04-03-03- IDT)) in a 40 µl reaction. Reactions were incubated in a fluorescence plate reader (SpectraMax ® M2) for up to 40 minutes with fluorescence measurements taken every 1 minute (λex: 535 nm; λem: 595 nm). Background-corrected fluorescence values were calculated by subtracting fluorescence values obtained from reactions carried out in the absence of target plasmid. The resulting data were fit to a single exponential decay curve (GraphPad Software), according to the following

equation: Fraction cleaved = A × (1 – exp(–k × t)), where A is the amplitude of the curve, k is the first-order rate constant, and t is time. Error bars represent the mean ± s.d., where n = 3 replicates. FIG. 36 shows the results of these data, and shows the kinetics of collateral cleavage activity of Cas12p and Cas12a.1, using DNA and RNA as reporters.

[00554] Reporters composed of DNA and RNA nucleotides led to efficient collateral Cas12p and Cas12a.1 activity. FQ Hybrid /56-FAM/TT rArUrU ATT /3IABkFQ/ or /56-FAM/TT ATrU rArUrU /3IABkFQ/ led to a maintained Cas12p collateral activity compared to the ssDNA or RNA reporters (ssDNA FAMQ reporter substrates (ssDNA 5'6-FAM TTATTATT-3IABkFQ3 (SEQ ID NO: 121)) or RNaseAlert (Cat N 11-04-03-03- IDT))). Whereas Cas12a.1 showed a slight decrease efficiency in *trans*-cleavage of chimeric reporters in comparison with the ssDNA. The reaction was initiated by diluting Cas12p or Cas12a.1 complexes to a final concentration of 37.5 nM Cas12p : 37.5 nM sgRNA : 10 nM activator or 75 nM Cas12a.1 : 75 nM sgRNA : 10 nM activator in a solution containing 1 × Binding Buffer (50 mM NaCl, 10 mM Tris-HCl, 10 mM MgCl2, 1 mM DTT, 100 g/ml BSA, pH 7.9) and 600 nM of ssDNA FAMQ reporter substrates (ssDNA 5'6-FAM TTATTATT-3IABkFQ3 (SEQ ID NO: 121), DNA-RNA chimeric reporters (/56-FAM/TT rArUrU ATT /3IABkFQ/, /56-FAM/TT ATrU rArUrU /3IABkFQ/ or RNaseAlert (Cat N 11-04-03-03- IDT)) in a 40 µl reaction. Reactions were incubated in a fluorescence plate reader (SpectraMax ® M2) for up to 40 minutes with fluorescence measurements taken every 1 minute (λex: 535 nm; λem: 595 nm). Background-corrected fluorescence values were calculated by subtracting fluorescence values obtained from reactions carried out in the absence of target plasmid. Error bars represent the mean ± s.d., where n = 3 replicates. FIG. 37 shows the results of these data.

[00555] While the inventions have been described in conjunction with the detailed description thereof, the foregoing description is intended to illustrate and not limit the scope of the invention, which is defined by the scope of the appended claims. Other aspects, advantages, and modifications are within the scope of the following.

Example 17: Cas12a.1 and Cas12p mature guide characterization and validation

[00556] The scaffold sequences of the mature guides were deduced *in silico* from the corresponding CRISPR loci. FIG. 38 shows the secondary structure of the mature guide scaffold

for Cas12a.1 (5' aaauuucuacuguaguagau 3') (SEQ ID NO: 116; Panel A) and Cas12p (5' agauuucuacuuuuguagau3') (SEQ ID NO: 117; Panel B). These were validated below.

[00557] The mature guide scaffolds for Cas12a.1 and Cas12p were evaluated *in vitro*. These mature scaffold sequences, along with a spacer targeting the N gene from SARS-CoV-2 virus were used in this example. The reactions were initiated by diluting Cas12p or Cas12a.1 complexes to a final concentration of 37.5 nM Cas12p : 37.5 nM sgRNA : 10 nM activator or 75 nM Cas12a.1 : 75 nM sgRNA : 10 nM activator in a solution containing 1 × Binding Buffer (50 mM NaCl, 10 mM Tris-HCl, 10 mM MgCl2, 1 mM DTT, 100 g/ml BSA, pH 7.9) and 600 nM of ssDNA FAMQ reporter substrates (ssDNA 5'6-FAM TTATTATT-3IABkFQ3 (SEQ ID NO: 121)) (in a 40 µl reaction. Reactions were incubated in a fluorescence plate reader (SpectraMax ® M2) for up to 20 minutes with fluorescence measurements taken every 1 minute (λex: 535 nm; λem: 595 nm). Background-corrected fluorescence values were calculated by subtracting fluorescence values obtained from reactions carried out in the absence of target plasmid. Error bars represent the mean ± s.d., where n = 3 replicates (FIG. 39). The data in this figure show that these mature scaffold sequences provide for CRISPR-mediated detection of SARS-CoV-2.

CLAIMS

1. An engineered system comprising:

a. a Cas9.1, Cas9.2, Cas9.3, or Cas9.4 protein, or a nucleic acid encoding the a Cas9.1, Cas9.2, Cas9.3, or Cas9.4 protein; and

b. a Cas9.1, Cas9.2, Cas9.3, or Cas9.4 guide RNA (gRNA), or a nucleic acid encoding the Cas9.1, Cas9.2, Cas9.3, or Cas9.4 gRNA, wherein the gRNA and the Cas9.1, Cas9.2, Cas9.3, or Cas9.4 protein do not naturally occur together, wherein the gRNA is capable of hybridizing to a target sequence in a target DNA, and the gRNA is capable of forming a complex with the Cas9.1, Cas9.2, Cas9.3, or Cas9.45 protein.

2. The system of claim 1, comprising:

a. a Cas9.1, Cas9.2, Cas9.3, Cas9.4 protein; and

b. a Cas9.1, Cas9.2, Cas9.3, or Cas9.4 gRNA.

3. The system of claim 1, comprising:

a. a nucleic acid encoding the Cas9.1, Cas9.2, Cas9.3, or Cas9.4 protein; and

b. a nucleic acid encoding the Cas9.1, Cas9.2, Cas9.3, or Cas9.4 gRNA.

4. The system of any one of claims 1 to 3, wherein the gRNA is a single-molecule gRNA.

5. The system of any one of claims 1 to 3, wherein the gRNA is a dual-molecule gRNA.

6. The system of any one of claims 1 to 5, wherein the Cas9.1 protein comprises the amino acid sequence of SEQ ID NO: 1, or at least 70% sequence identity thereto.

7. The system of any one of claims 1 to 5, wherein the Cas9.2 protein comprises the amino acid sequence of SEQ ID NO: 2 or at least 70% sequence identity thereto.

8. The system of any one of claims 1 to 5, wherein the Cas9.3 protein comprises the amino acid sequence of SEQ ID NO: 10, or at least 70% sequence identity thereto.

9. The system of any one of claims 1 to 5, wherein the Cas9.4 protein comprises the amino acid sequence of SEQ ID NO: 11, or at least 70% sequence identity thereto.

10. The system of any one of claims 1 to 7, wherein the target sequence is a sequence of a target provided in any of Tables 6a-6f.

11. The system of any one of claims 1 to 7, wherein the target sequence is a sequence of a human.

12. The system of any one of claims 1 to 7, wherein the target sequence is a sequence of a non-human primate.

13. The system of any one of claims 1 to 12, wherein the Cas9.1, Cas9.2, Cas9.3, or Cas9.4 protein is a catalytically active protein.

14. The system of claim 13, wherein the Cas9.1, Cas9.2, Cas9.3, or Cas9.4 protein cleaves at a site distal to the target sequence.

15. The system of any one of claims 1 to 12, wherein the Cas9.1, Cas9.2, Cas9.3, or Cas9.4 protein is a catalytically dead protein.

16. The system of any one of claims 1 to 12, wherein the Cas9.1, Cas9.2, Cas9.3, or Cas9.4 protein comprises nickase activity.

17. An engineered system comprising:

a. a Class 2 Type V CRISPR-Cas RNA-guided endonuclease protein; and

b. a single guide RNA (gRNA),

wherein the gRNA and the Class 2 Type V CRISPR-Cas RNA-guided endonuclease protein do not naturally occur together, wherein the gRNA is capable of hybridizing to a target sequence in a target DNA, wherein the gRNA is capable of forming a complex with the Class 2 Type V CRISPR-Cas RNA-guided endonuclease protein, and wherein the Class 2 Type V CRISPR-Cas RNA-guided endonuclease protein possesses collateral activity and is capable of collaterally cleaving a single stranded polynucleotide comprising RNA without a tracrRNA.

18. The system of claim 17, wherein the Class 2 Type V CRISPR-Cas RNA-guided endonuclease protein comprises the amino acid sequence of SEQ ID NO: 4, or at least 70% sequence identity thereto.

19. The system of any one of claims 17 to 18, wherein the target sequence is a sequence of a target provided in any of Tables 6a-6f.

20. The system of any one of claims 17 to 18, wherein the target sequence is a sequence of a human.

21. The system of any one of claims 17 to 18, wherein the target sequence is a sequence of a non-human primate.

22. The system of any one of claims 17 to 18, wherein the target sequence is a bacterial or viral sequence.

23. The system of any one of claims 17 to 22, wherein the Class 2 Type V CRISPR-Cas RNA-guided endonuclease protein is capable of collaterally cleaving a single stranded RNA.

24. The system of any one of claims 17 to 22, wherein the Class 2 Type V CRISPR-Cas RNA-guided endonuclease protein is capable of collaterally cleaving a single stranded DNA/RNA hybrid.

25. An engineered system comprising:

a. a Cas12a.1, Cas12p, or Cas12q protein, or a nucleic acid encoding the Cas12a.1, Cas12p, or Cas12q protein; and

b. a Cas12a.1, Cas12p, or Cas12q gRNA, or a nucleic acid encoding a Cas12a.1, Cas12p, or Cas12q gRNA,

wherein the gRNA and the Cas12a.1, Cas12p, or Cas12q protein do not naturally occur together, wherein the gRNA is capable of hybridizing to a target sequence in a target DNA, and the gRNA is capable of forming a complex with the Cas12a.1, Cas12p, or Cas12q protein.

26. The system of claim 25, comprising:

a. a Cas12a.1, Cas12p, or Cas12q protein; and

b. a Cas12a.1, Cas12p, or Cas12q gRNA.

27. The system of claim 25, comprising:

a. a nucleic acid encoding the Cas12a.1, Cas12p, or Cas12q protein; and

b. a nucleic acid encoding a Cas12a.1, Cas12p, or Cas12q gRNA.

28. The system of any one of claims 25 to 27, wherein the Cas12a.1 protein comprises the amino acid sequence of SEQ ID NO: 3, or at least 70% sequence identity thereto.

29. The system of any one of claims 25 to 27, wherein the Cas12p protein comprises the amino acid sequence of SEQ ID NO: 4, or at least 70% sequence identity thereto.

30. The system of any one of claims 25 to 27, wherein the Cas12q protein comprises the amino acid sequence of SEQ ID NO: 222, or at least 70% sequence identity thereto.

31. The system of any one of claims 25 to 27, wherein the Cas12q protein comprises the amino acid sequence of SEQ ID NO: 5, or at least 70% sequence identity thereto.

32. The system of any one of claims 25 to 31, wherein the target sequence is a sequence of a target provided in any of Tables 6a-6f.

33. The system of any one of claims 25 to 31, wherein the target sequence is a sequence of a human.

34. The system of any one of claims 25 to 31, wherein the target sequence is a sequence of a non-human primate.

35. The system of any one of claims 25 to 31, wherein the target sequence is a bacterial or viral sequence.

36. The system of any one of claims 25 to 34, wherein the Cas12a.1, Cas12p, or Cas12q protein is a catalytically active Cas12a.1, Cas12p, or Cas12q protein.

37. The system of claim 36, wherein the Cas12a.1, Cas12p, or Cas12q protein cleaves at a site distal to the target sequence.

38. The system of any one of claims 25 to 34, wherein the Cas12a.1, Cas12p, or Cas12q protein is a catalytically dead Cas12a.1, Cas12p, or Cas12q protein.

39. The system of any one of claims 25 to 34, wherein the Cas12a.1, Cas12p, or Cas12q protein comprises nickase activity.

40. An engineered single-molecule gRNA, comprising:

a. a targeter-RNA comprising a spacer sequence that is capable of hybridizing with a target sequence in a target DNA; and

b. an activator-RNA that is capable of hybridizing with the targeter-RNA to form a double-stranded RNA duplex, the activator-RNA comprising a activator-RNA,

wherein the targeter-RNA and the activator-RNA are covalently linked to one another, wherein the single-molecule gRNA is capable of forming a complex with a Cas9.1, Cas9.2, Cas9.3, or Cas9.4 protein, and wherein hybridization of the spacer sequence to the target sequence is capable of targeting the Cas9.1, Cas9.2, Cas9.3, or Cas9.4 protein to the target DNA.

41. The gRNA of claim 40, wherein the targeter-RNA and the activator-RNA are arranged in a 5' to 3' orientation.

42. The gRNA of claim 40, wherein the activator-RNA and the targeter-RNA are arranged in a 5' to 3' orientation.

43. The gRNA of any one of claims 40 to 42, wherein the targeter-RNA and the activator-RNA are covalently linked to one another via a linker.

44. The gRNA of ay one of claims 40 to 43, wherein the single-molecule gRNA comprises one or more sequence modifications compared to a sequence of a corresponding wild type tracrRNA and/or crRNA.

45. The gRNA of ay one of claims 40 to 44, wherein the targeter-RNA comprises a spacer sequence of about 10-50 nucleotides that have 100% complementarity to a sequence in the target DNA.

46. The gRNA of any one of claims 40 to 44, wherein the targeter-RNA comprises a spacer sequence of about 10-50 nucleotides that have less than 100% complementarity to a sequence in the target DNA.

47. The gRNA of any one of claims 40 to 46, wherein the target sequence is a sequence of a target provided in any of Tables 6a-6f.

48. The gRNA of any one of claims 40 to 47, wherein the Cas9.1 protein comprises the sequence of SEQ ID NO: 1 or a sequence with at least 70% sequence identity thereto.

49. The gRNA of any one of claims 40 to 47, wherein the Cas9.2 protein comprises the sequence of SEQ ID NO: 2 or a sequence with at least 70% sequence identity thereto.

50. The gRNA of any one of claims 40 to 47, wherein the Cas9.3 protein comprises the sequence of SEQ ID NO: 10 or a sequence with at least 70% sequence identity thereto.

51. The gRNA of any one of claims 40 to 47, wherein the Cas9.4 protein comprises the sequence of SEQ ID NO: 11 or a sequence with at least 70% sequence identity thereto.

52. An engineered single-molecule gRNA, comprising the scaffold sequence of SEQ ID NO: 116 or SEQ ID NO: 117 and a spacer sequence that is capable of hybridizing with a target sequence in a target DNA.

53. The gRNA of claim 52, wherein the target DNA comprises viral DNA, plant DNA, fungal DNA, or bacterial DNA.

54. The gRNA of claim 52, wherein the target sequence is a sequence of a target provided in any of Tables 6a-6f.

55. The gRNA of claim 52, wherein the target is a coronvavirus.

56. The gRNA of claim 52, wherein the target is a SARS-CoV-2 virus.

57. The gRNA of claim 52, wherein the target DNA is cDNA, and has been obtained by reverse transcription.

58. A method of modifying a target DNA, the method comprising contacting the target DNA with any one of the systems of claims 1 to 39, wherein the gRNA hybridizes with the target sequence whereby modification of the target DNA occurs.

59. The method of claim 58, wherein the target DNA is extrachromosomal DNA.

60. The method of claim 58, wherein the target DNA is part of a chromosome.

61. The method of claim 58, wherein the target DNA is part of a chromosome *in vitro*.

62. The method of claim 58, wherein the target DNA is part of a chromosome *in vivo*.

63. The method of claim 58, wherein the target DNA is outside a cell.

64. The method of claim 58, wherein the target DNA is inside a cell.

65. The method of claim 64, wherein the target DNA comprises a gene and/or its regulatory region.

66. The method of claim 64 or 65, wherein 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, in invertebrate cell, a vertebrate cell, a fish 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.

67. The method of any of the claims of 58 to 66, wherein the modifying comprises introducing a double strand break in the target DNA.

68. The method of any of the claims of 58 to 67, wherein the contacting occurs under conditions that are permissive for non-homologous end joining or homology-directed repair.

69. The method of any of the claims of 58 to 67, wherein the contacting the target DNA with a donor polynucleotide, wherein the donor polynucleotide, a portion of the donor polynucleotide,

a copy of the donor polynucleotide, or a portion of a copy of the donor polynucleotide integrates into the target DNA.

70. The method of any of the claims of 58 to 67, wherein the method does not comprise contacting the cell with a donor polynucleotide, or wherein the target DNA is modified such that nucleotides within the target DNA are deleted.

71. A method of detecting a target DNA in a sample, the method comprising:

a. contacting the sample with:

i. a Cas12a.1, Cas12p, or Cas12q protein;

ii. a Cas12a.1, Cas12p, or Cas12q gRNA comprising a spacer sequence that is capable of hybridizing with a target sequence in a target DNA; and

iii. a labeled detector that does not hybridize with the spacer sequence of the gRNA; and

b. measuring a detectable signal produced by cleavage of the labeled detector by the Cas12a.1, Cas12p, or Cas12q protein, thereby detecting the target DNA.

72. The method of claim 71, wherein the labeled detector comprises a labeled single stranded DNA.

73. The method of claim 71, wherein the labeled detector comprises a labeled RNA.

74. The method of claim 72, wherein the labeled RNA is a single stranded RNA.

75. The method of claim 71, wherein the labeled detector comprises a labeled single stranded DNA/RNA chimera.

76. The method of any one of claims 71 to 75, wherein the labeled detector comprises one or more modified nucleotides.

77. The method of any one of claims 71 to 76, comprising contacting the sample with a precursor gRNA array, wherein the Cas12a.1, Cas12p, or Cas12q protein cleaves the precursor gRNA array to produce said gRNA.

78. The method of any one of claims 71 to 77, wherein the target DNA is single stranded.

79. The method of any one of claims 71 to 78, wherein the target DNA is double stranded.

80. The method of any one of claims 71 to 79, wherein the target DNA is viral DNA, plant DNA, fungal DNA, or bacterial DNA.

81. The method of claim 80, wherein the target sequence is a sequence of a target provided in any of Tables 6a-6f.

82. The method of claim 81, wherein the target is a coronvavirus.

83. The method of claim 82, wherein the target is a SARS-CoV-2 virus.

84. The method of any one of claims 71 to 83, wherein the target DNA is cDNA, and has been obtained by reverse transcription.

85. The method of any one of claims 71 to 79, wherein the target DNA is from a human cell.

86. The method of claim 85, wherein the target DNA is human fetal or cancer cell DNA.

87. The method of any one of claims 71 to 86, wherein the protein is Cas12a.1 comprising the amino acid sequence of SEQ ID NO: 3, or at least 70% sequence identity thereto.

88. The method of any one of claims 71 to 86, wherein the protein is Cas12p comprising the amino acid sequence of SEQ ID NO: 4, or at least 70% sequence identity thereto.

89. The method of any one of claims 71 to 86, wherein the protein is Cas12p comprising the amino acid sequence of SEQ ID NO: 222, or at least 70% sequence identity thereto.

90. The method of any one of claims 71 to 86, wherein the protein is Cas12q comprising the amino acid sequence of SEQ ID NO: 5, or at least 70% sequence identity thereto.

91. The method of any one of claims 71 to 87, wherein the sample comprises DNA from a cell lysate.

92. The method of any one of claims 71 to 87, wherein the sample comprises cells.

93. The method of any one of claims 71 to 87, wherein the sample is a urine, blood, serum, plasma, lymphatic fluid, cerebrospinal fluid, saliva, nasopharyngeal, oropharyngeal, nasopharyngeal/ oropharyngeal, aspirate, or biopsy sample.

94. The method of any one of claims 71 to 93, comprising determining an amount of the target DNA present in the sample.

95. The method of claim 94, wherein said measuring a detectable signal comprises one or more of: visual based detection, sensor based detection, color detection, gold nanoparticle based detection, fluorescence polarization, colloid phase transition/dispersion, electrochemical detection, and semiconductor-based sensing.

96. The method of any one of claims 71 to 95, wherein the labeled detector comprises a modified nucleobase, a modified sugar moiety, and/or a modified nucleic acid linkage.

97. The method of any one of claims 71 to 96, further comprising detecting a positive control target DNA in a positive control sample, the detecting comprising:

a. contacting the positive control sample with:

i. the Cas12a.1, Cas12p, or Cas12q protein;

ii. a positive control gRNA comprising: a region that binds to the Cas12a.1, Cas12p, or Cas12q protein, and a positive control spacer sequence that hybridizes with the positive control target DNA; and

iii. a labeled detector that does not hybridize with the positive control spacer sequence of the positive control gRNA; and

b. measuring a detectable signal produced by cleavage of the labeled detector by the Cas12a.1, Cas12p, or Cas12q protein, thereby detecting the positive control target DNA.

98. The method of any one of claims 71 to 97, wherein the detectable signal is detectable in less than 15, 30, 45, 60, 90, 120, 150, 180, 210, or 240 minutes.

99. The method of any one of claims 71 to 98, further comprising amplifying the target DNA in the sample by loop-mediated isothermal amplification (LAMP), helicase-dependent amplification (HDA), recombinase polymerase amplification (RPA), strand displacement amplification (SDA), nucleic acid sequence-based amplification (NASBA), transcription mediated amplification (TMA), nicking enzyme amplification reaction (NEAR), rolling circle amplification (RCA), multiple displacement amplification (MDA), Ramification (RAM), circular helicase-dependent amplification (cHDA), single primer isothermal amplification (SPIA), signal mediated amplification of RNA technology (SMART), self-sustained sequence replication (3SR), genome exponential amplification reaction (GEAR), or isothermal multiple displacement amplification (IMDA).

100. The method of any one of claims 71 to 99, wherein target DNA in the sample is present at a concentration of less than 100 uM.

101. A protein comprising an amino acid sequence with 70%-99.5% homology to SEQ ID NO: 1, 2, 3, 4, 5, 10, 11, or 222.

102. A protein of claim 101, wherein the sequence of the protein has been deduced bioinformatically.

103. A composition comprising any of the proteins of claim 101, and optionally a pharmaceutically acceptable carrier.

104. A composition comprising any of the proteins of claim 101, optionally comprising a pharamaceutically aceptable carrier, a nucleic acid stabilizing buffer and/or or a protein stabilizing buffer.

105. A composition comprising any of the proteins of claim 101, wherein the protein is lyopholized, and optionally further comprises any one or more of a labeled detector, a reverse transcriptase enzyme, and reagents for loop-mediated isothermal amplification.

106. A DNA polynucleotide comprising a nucleotide sequence that encodes any of the proteins of claim 101.

107. A recombinant expression vector comprising the DNA polynucleotide of claim 106.

108. The recombinant expression vector of claim 107, wherein the nucleotide sequence encoding the single protein is operably linked to a promoter.

109. A host cell comprising the DNA polynucleotide of any one of claims 106 to 108.

110. A pharmaceutical composition comprising any of the engineered systems of claims 1 to 39, and optionally a pharmaceutically acceptable carrier.

111. A composition comprising any of the engineered systems of claims 1 to 39, and optionally comprising a nucleic acid stabilizing buffer and/or or a protein stabilizing buffer.

112. A pharmaceutical composition comprising any of the single molecule gRNAs of claims 40 to 57, and optionally pharmaceutically acceptable carrier.

113. A composition comprising any of the singe molecule gRNAs of claims 40 to 51, and optionally a nucleic acid stabilizing buffer and/or or a protein stabilizing buffer.

114. A DNA polynucleotide comprising a nucleotide sequence that encodes any of the nucleic acids of claims 3, 27, or the gRNAs of claims 40 to 51.

115. A recombinant expression vector comprising the DNA polynucleotide of claim 114.

116. The recombinant expression vector of claim 115, wherein the nucleotide sequence encoding the single gRNA is operably linked to a promoter.

117. A host cell comprising the DNA polynucleotide of any one of claims 114 to 116.

118. A kit comprising one or more components of any of the engineered systems of claims 1 to 39.

119. The kit of claim 118, wherein one or more components are lyopholized.

120. The kit of any one of claims 118 to 119, wherein the one or more components comprise Cas12p, a labeled RNA reporter, and a gRNA directed to SARS-CoV-2.

121. A method of isolating a Class 2 Type II or Class 2 Type V CRISPR-Cas protein from a metagenomics sample comprising the use of a bioinformatics-based method.

122. The method of claim 121, wherein the Class 2 Type II or Class 2 Type V CRISPR-Cas protein is selected from the group consisting of SEQ ID NO: 1, 2, 3, 4, 5, 10, 11, and 222.

Cas9.1

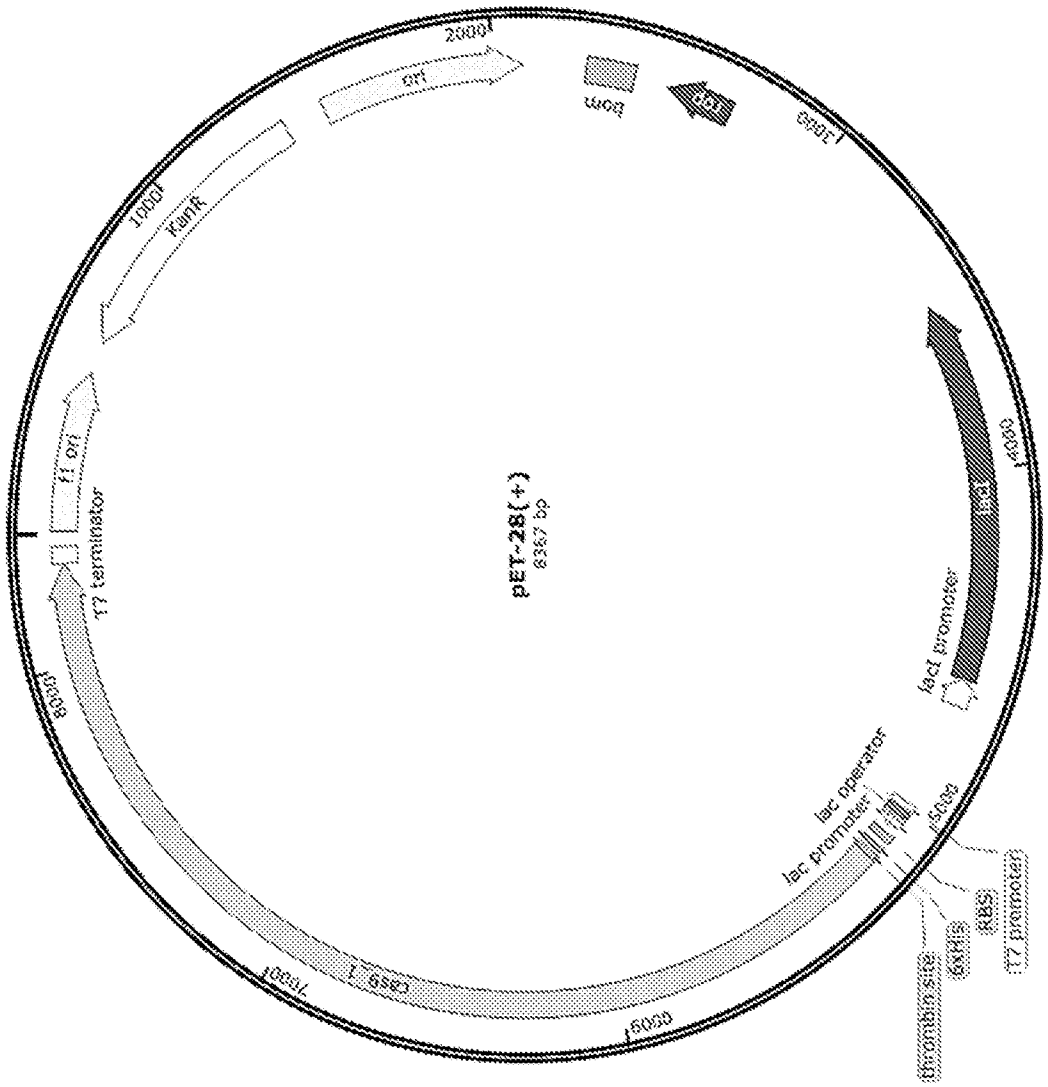
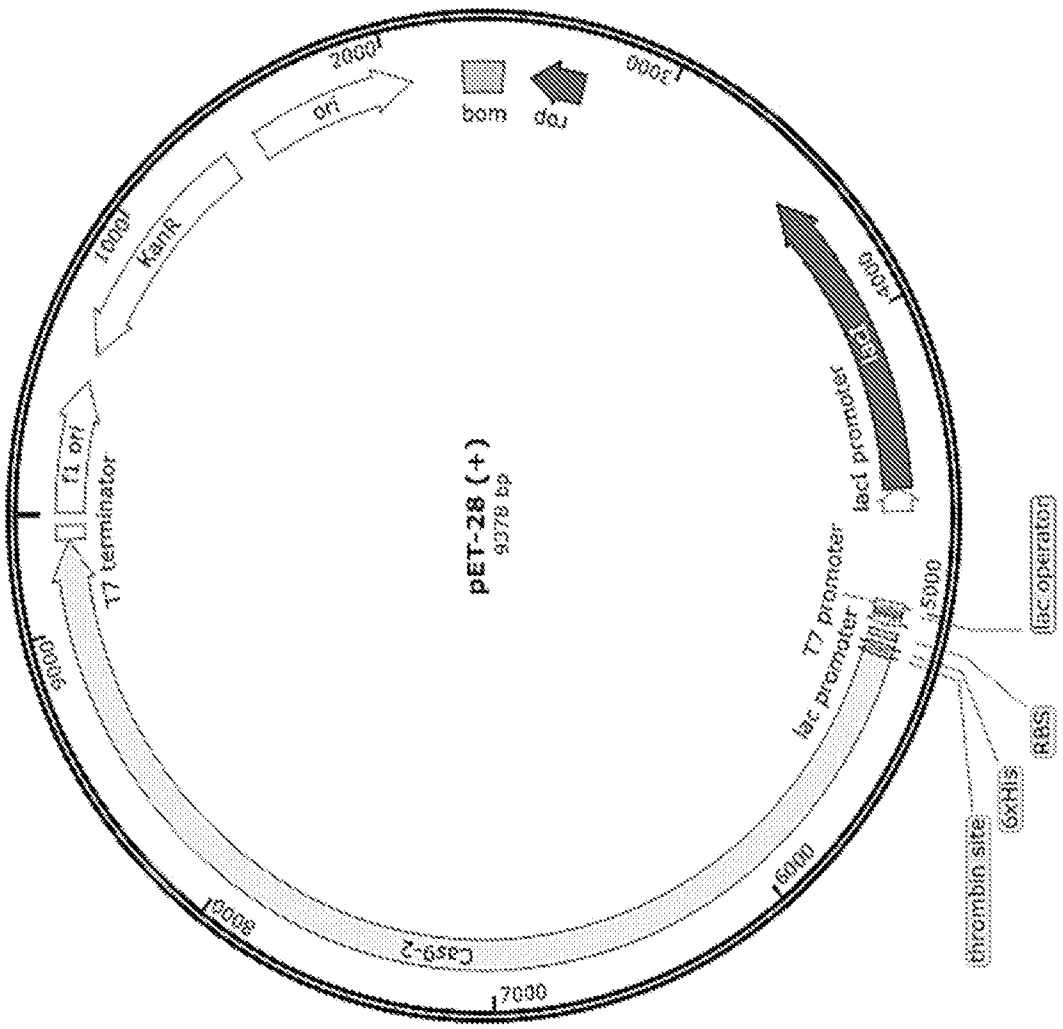

pET-28(+)
5367 bp
2000
ori
bom
lacI
1000
KanR
f1 ori
T7 terminator
lacI promoter
lac operator
lac promoter
T7 promoter
RBS
6xHis
thrombin site
T7 tag

FIG. 1A


pET-28 (+)
5378 bp
1000
2000
3000
4000
5000
6000
7000
8000
9000
rop
bom
ori
KanR
f1 ori
T7 terminator
lacI
lacI promoter
T7 promoter
lac promoter
lac operator
RBS
6xHis
enzyme site
Cas9-2

Cas9.2

FIG. 1B

Cas12a.1

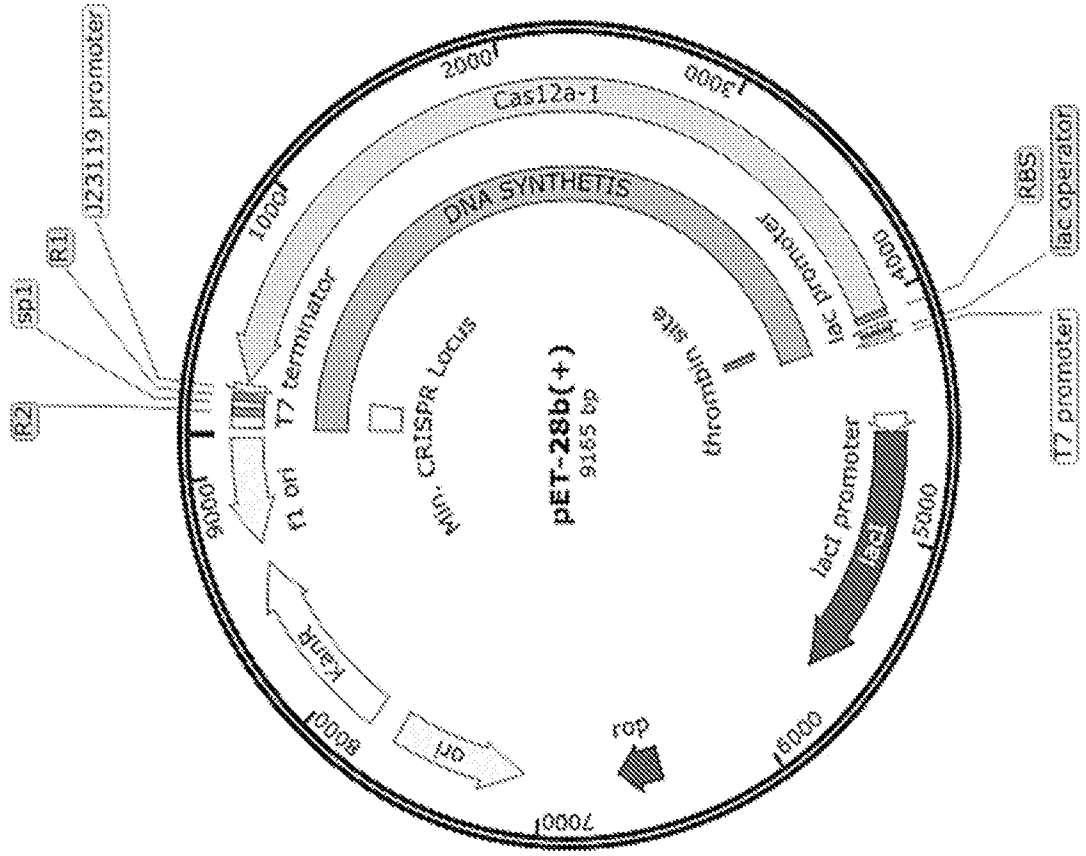


FIG. 2A

Cas12p

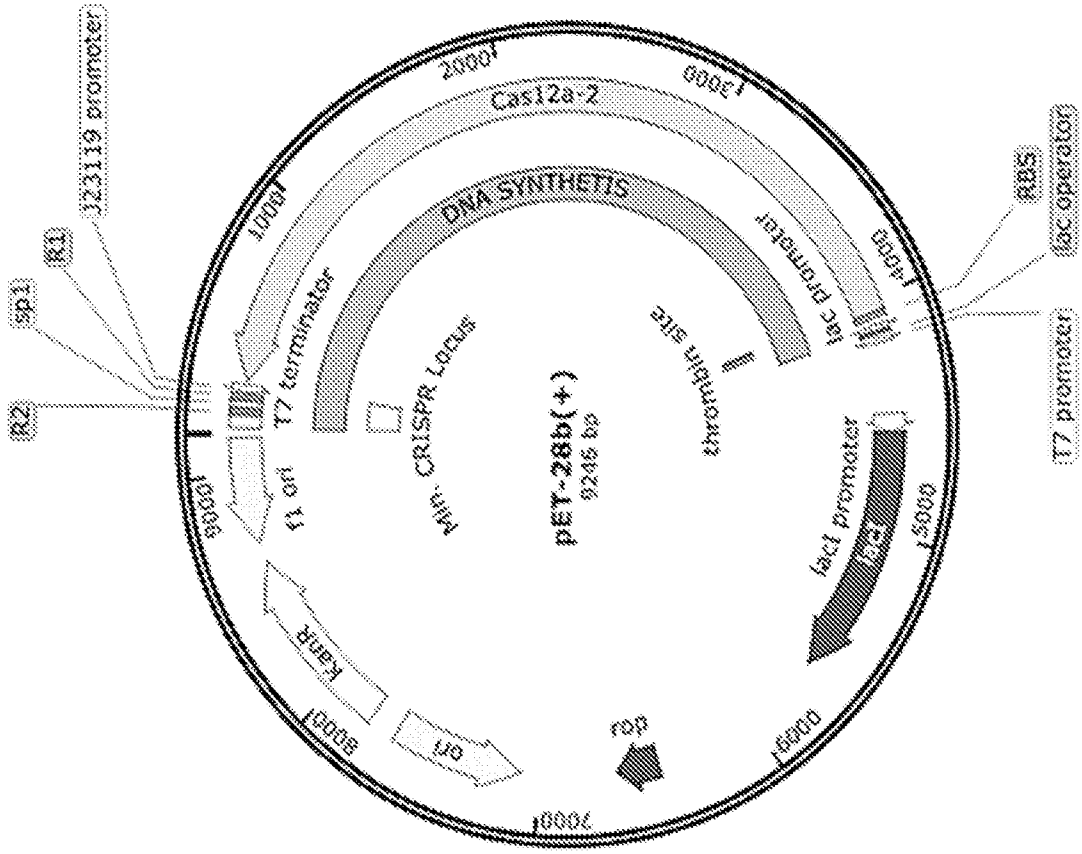


FIG. 2B

Cas12q

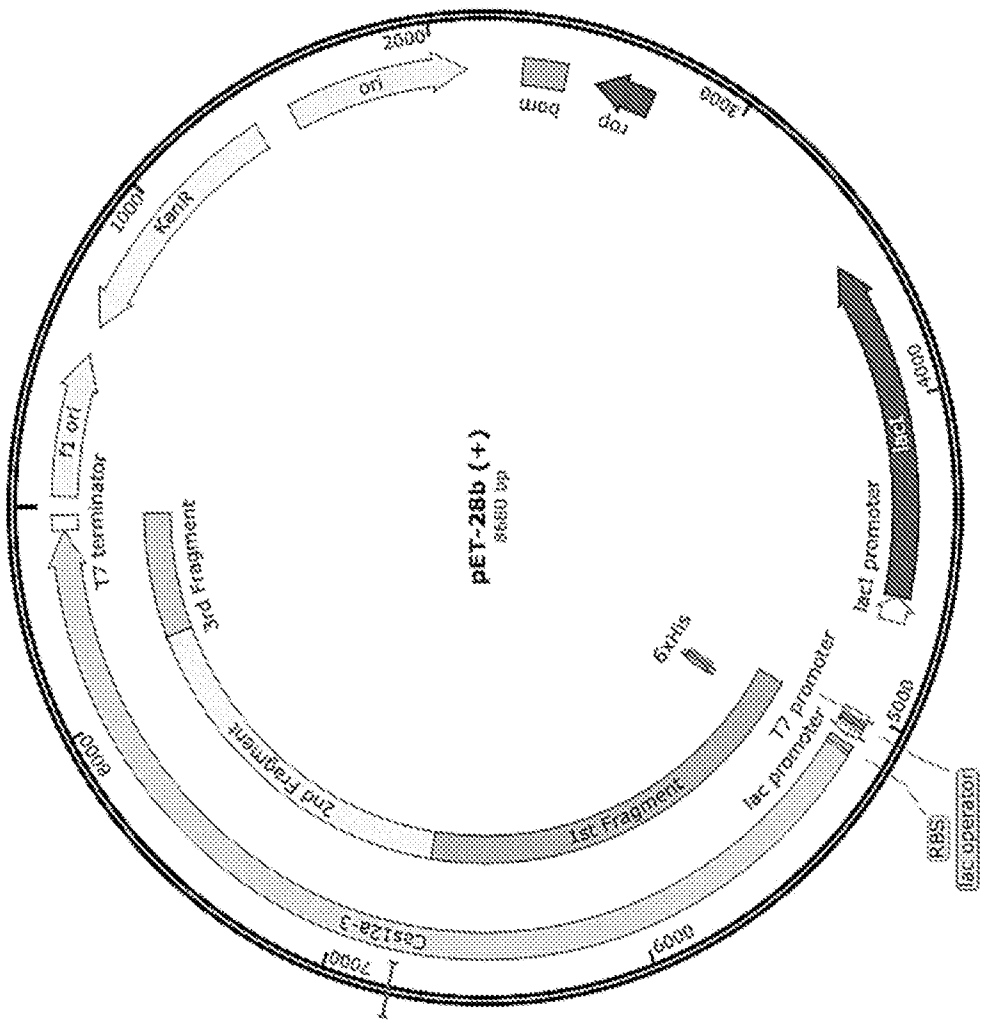
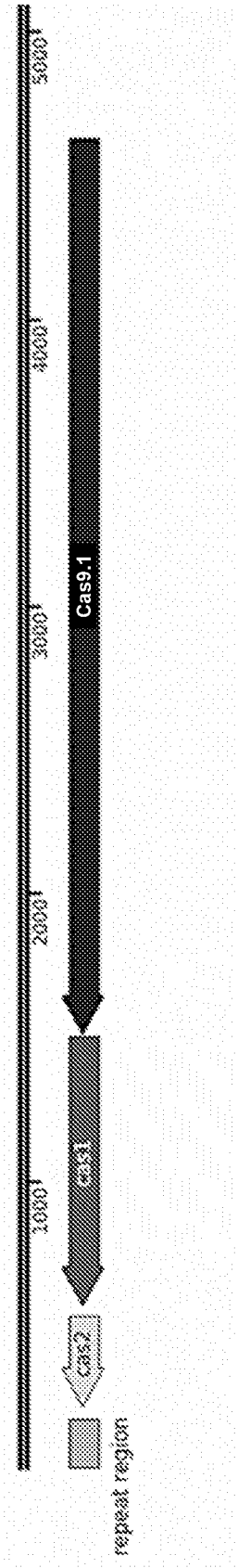


FIG. 2C

FIG. 3A

Cas9.1


50001
40001
30001
20001
10001
Cas9.1
cas1
cas2
repeat region

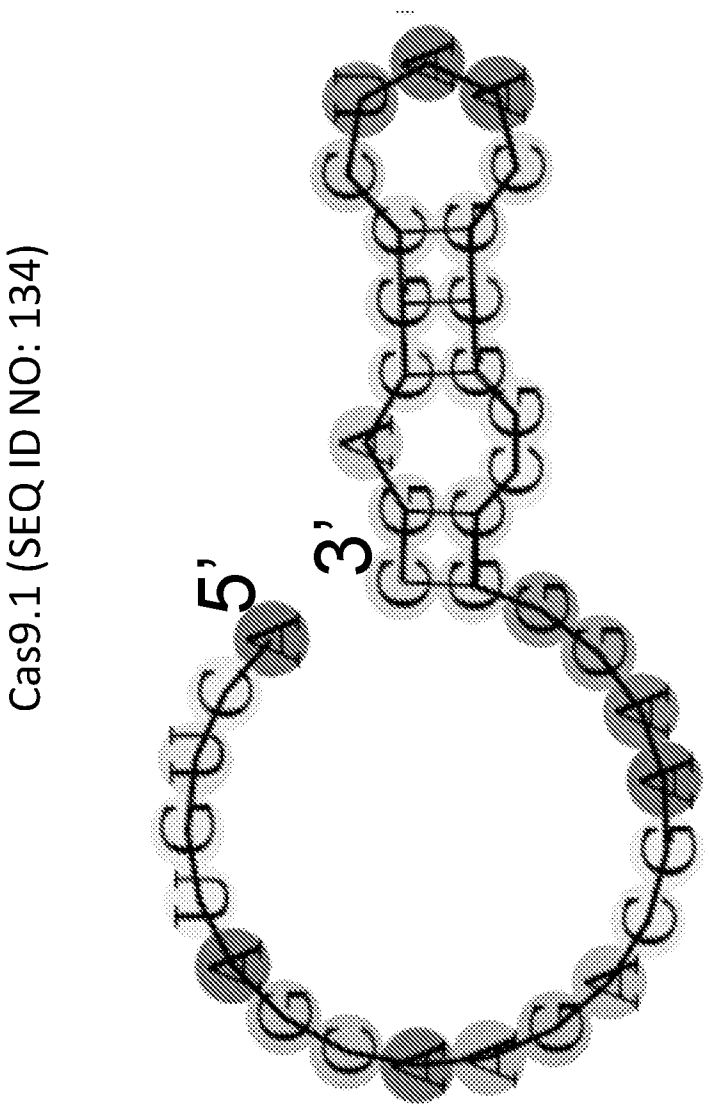
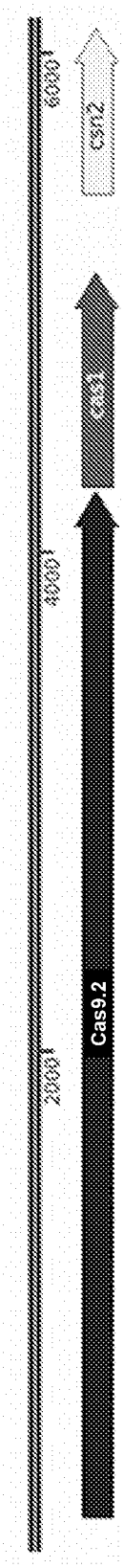


FIG. 3B



Cas9.2

FIG. 3C

Cas9.3

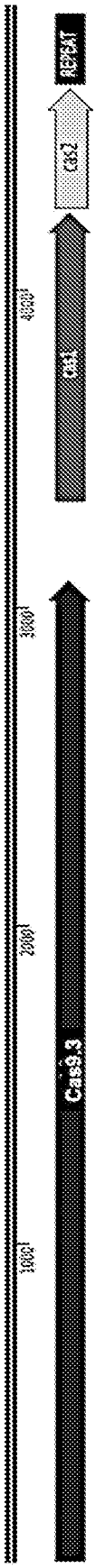


FIG. 3D



FIG. 3E

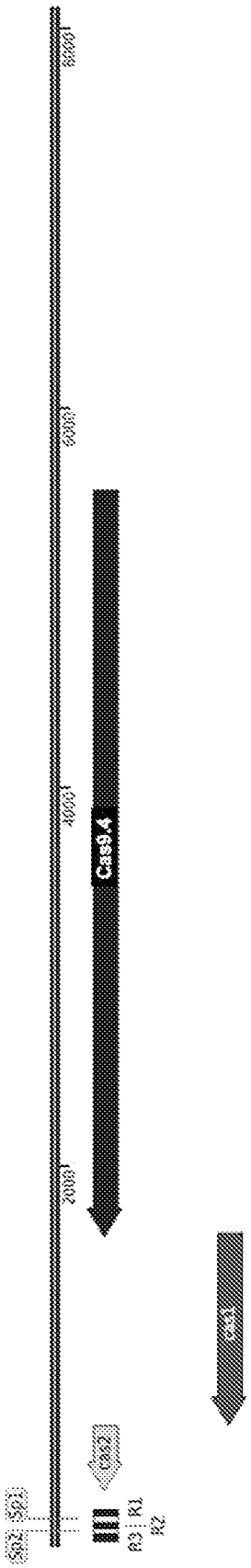


FIG. 3F

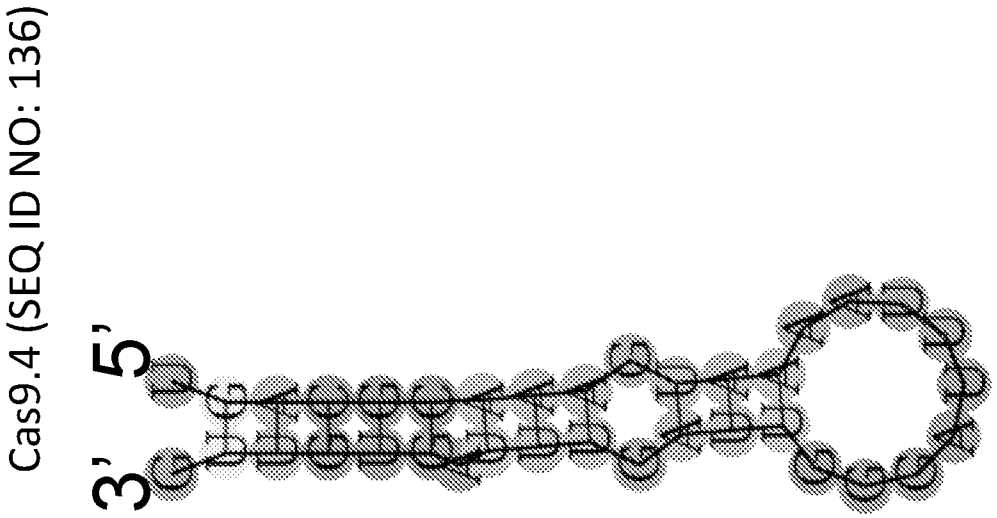
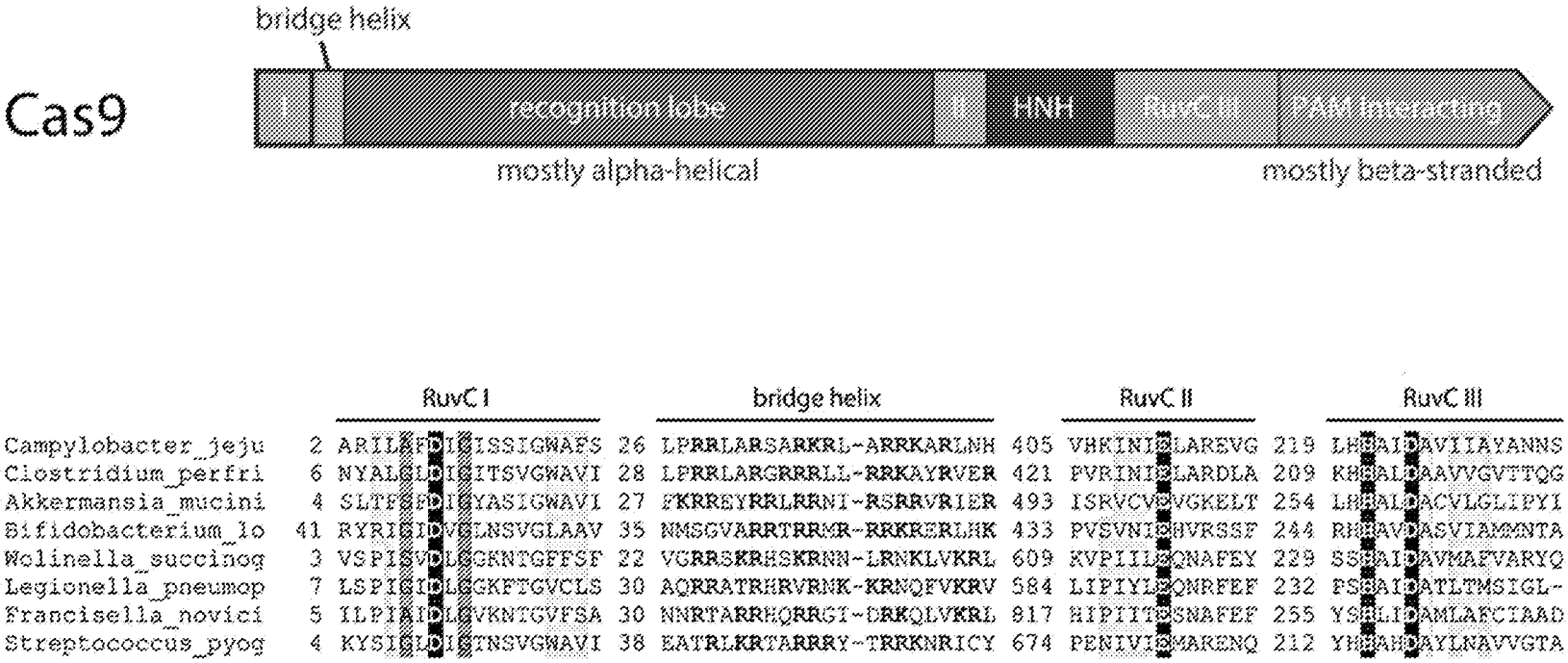


FIG. 3G

FIG. 4A



SEQ ID NOS: 137-168

FIG. 4B

Cas9.1

Domain Structure (RuvC I &+ Bridge Helix)

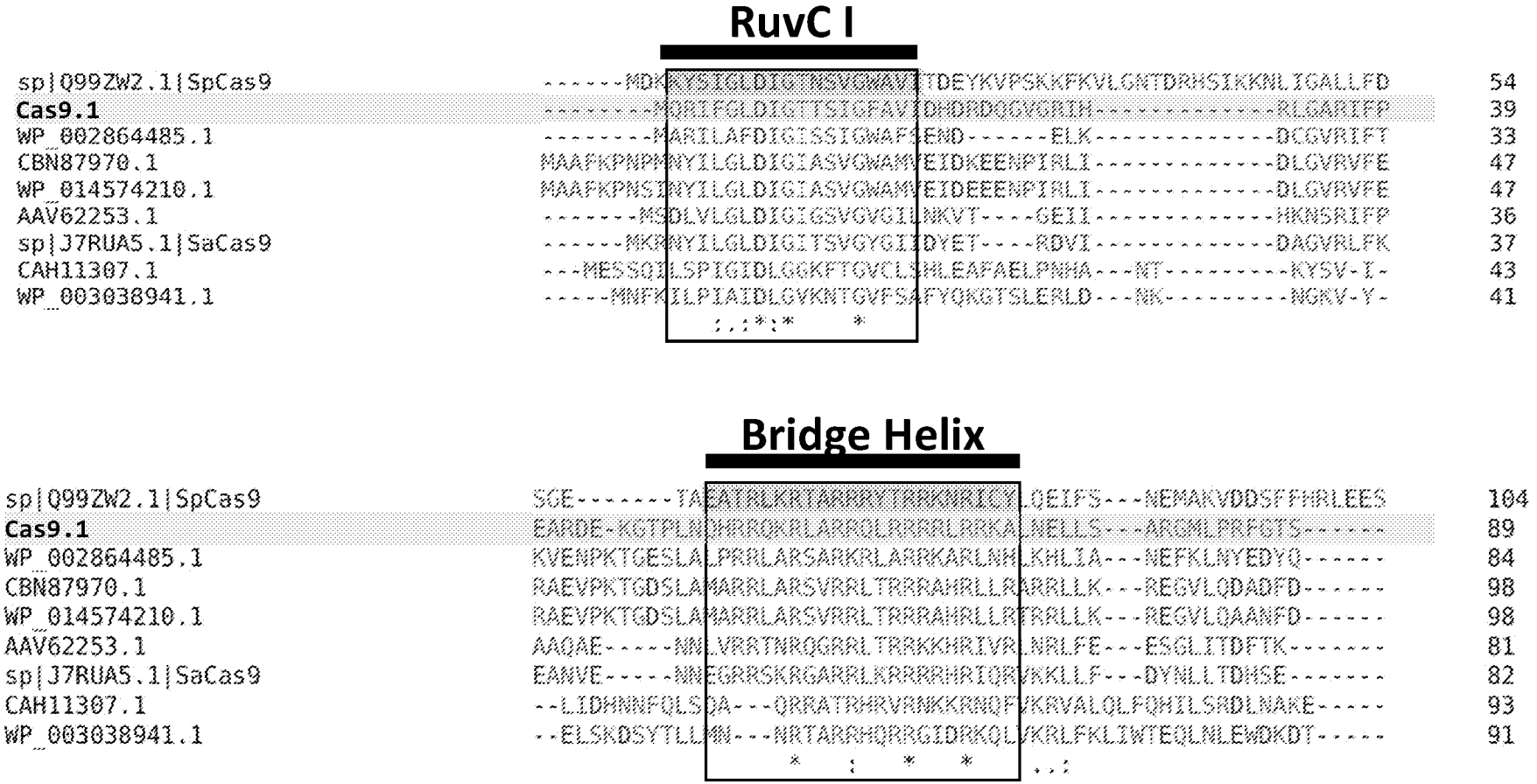


FIG. 4B continued

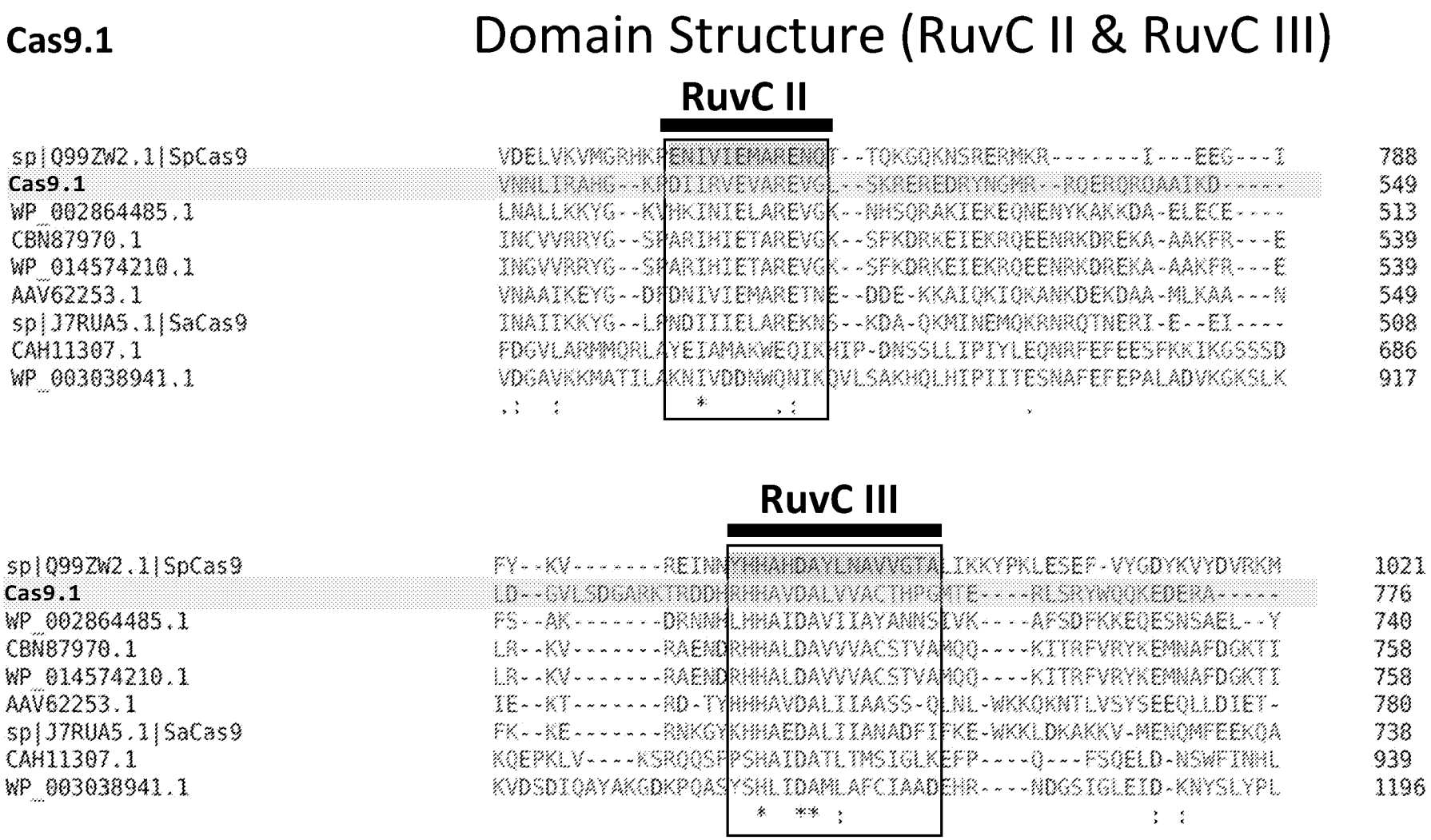


FIG. 4C

Cas9.2

Domain Structure (RuvC I &+ Bridge Helix)

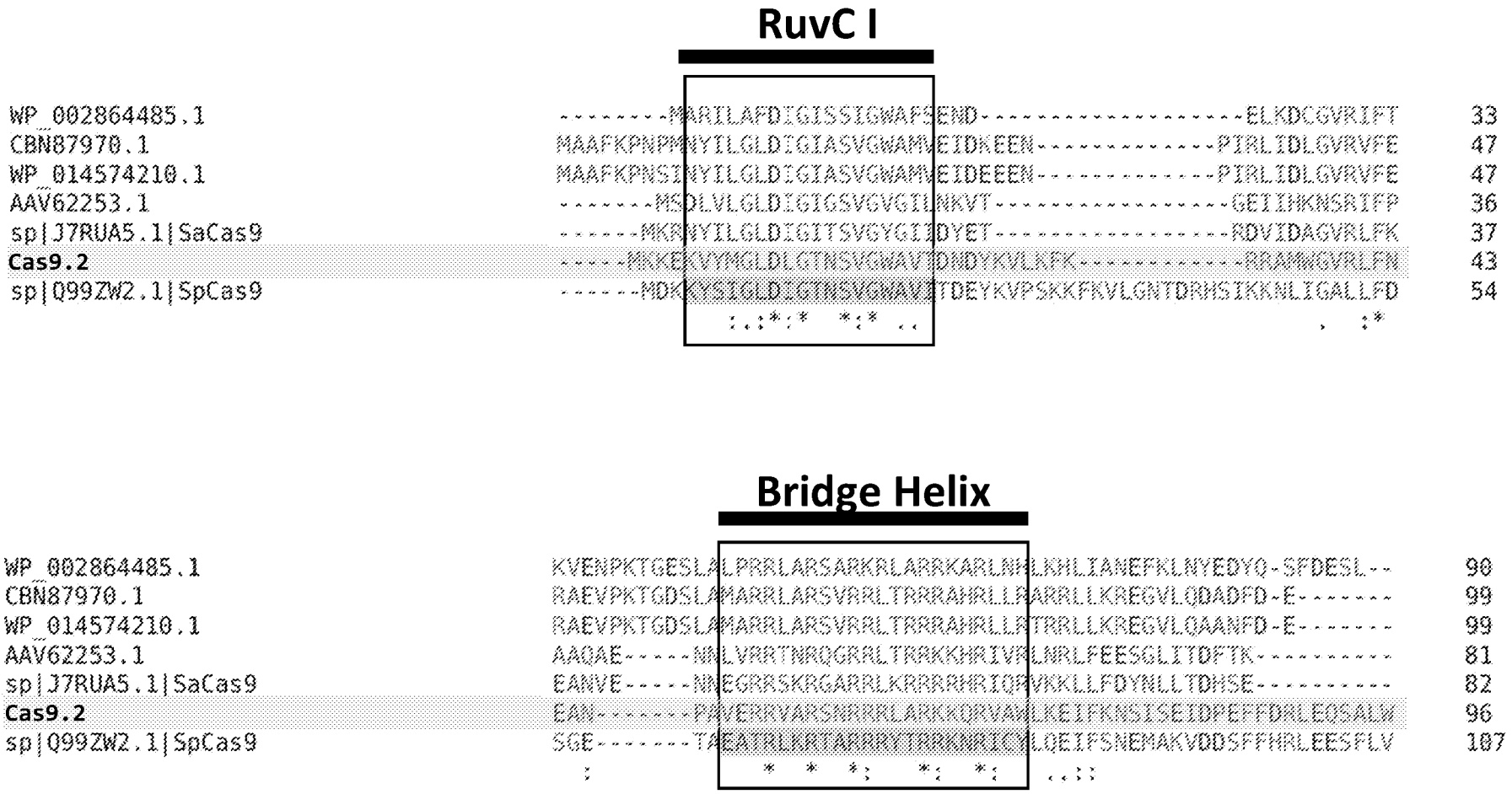


FIG. 4C continued

Cas9.2

Domain Structure (RuvC II & RuvC III)

RuvC II

```
WP_002864485.1          AI--NEDKKDFLPAFNETYYKDEVTNPVVLRAIKEYRKVLNALLKKYG--KVHKINIELA 481 REVG
CBN87970.1              YCKKNAEEKIYLPP----IPADEIRNPVVLRALSQARKVINCVVRRYG--SPARIHIETA 506 REVG
WP_014574210.1          YGKKNTEEKIYLPP----IPADEIRNPVVLRALSQARKVINGVVRRYG--SPARIHIETA 506 REVG
AAV62253.1              QIKQNISN--KTKYIDEKLLTEEIYNPVVAKSVRQAIKIVNAAIKEYG--DFDNIVIEMA 517 RETN
sp|J7RUA5.1|SaCas9      --KKVDLS--QQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYG--LPNDIIIELA 479 REKN
Cas9.2                  VIE-TEAKKQLTGHITHSHVEALYCSPVVKKQIWQALQIALELKKTLKK-DPNKIFVETT 781 RHEG
sp|Q99ZW2.1|SpCas9      DIQ-KA-QVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMA 764 RENQ
                              .                .*.: : . :  ::     :        * :* :     *.
```

RuvC III

```
WP_002864485.1          ----------------TWGFSAKDRNNHLHHAIDAVIIAYANNSIV-K---AFSDFKKE   731
CBN87970.1              ----------------FWGLRKVRAENDRHHALDAVVVACSTVAMQ-Q---KITRFVRY   747
WP_014574210.1          ----------------FWGLRKVRAENDRHHALDAVVVACSTVAMQ-Q---KITRFVRY   747
AAV62253.1              ----------------HWGIEKTRD-TYHHHAVDALIIAASS-QLN-LWKKQKNTLVSY   770
sp|J7RUA5.1|SaCas9      ----------------KWKFKKERNKGYKHHAEDALIIANADFIFK-EWKKLDKAKKV-   727
Cas9.2                  PSLPENKTNGKDPKGITDYSMIKVREINDLHHAKDAYLNIVVGNVYDTKFRYRGKDL---   1042
sp|Q99ZW2.1|SpCas9      ----------------DFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFV-Y   1010
                                      : :        *** ** :                .
```

FIG. 4D

Cas9.3

Domain Structure (RuvC I &+ Bridge Helix)

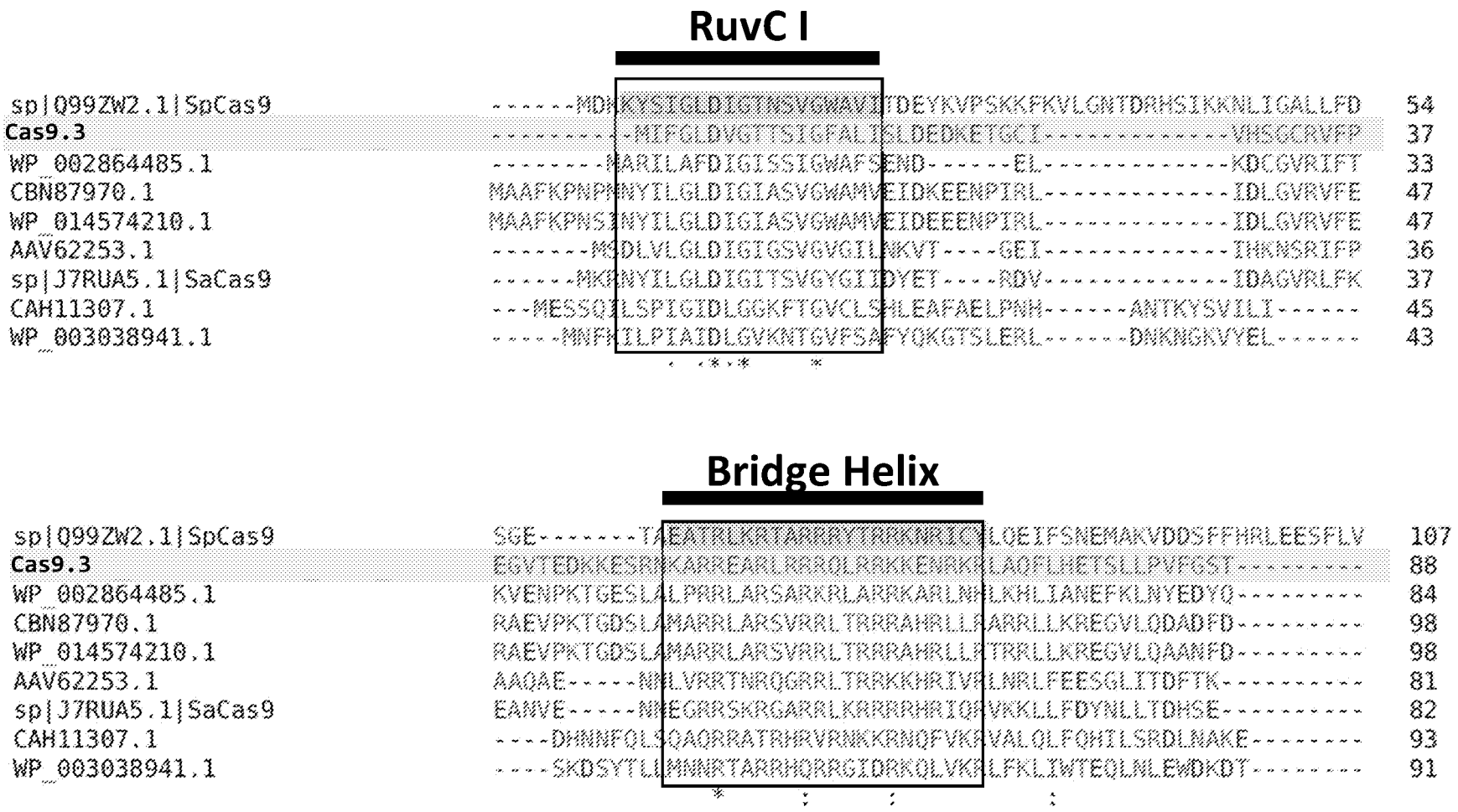


FIG. 4D continued

Cas9.3

Domain Structure (RuvC II & RuvC III)

RuvC II

```
sp|Q99ZW2.1|SpCas9      ENIVI EMARENQT--TQKGQKNSRERMKRIEEGI---------KELGS---QILKEHPVENTQLQ   807
Cas9.3                  DIIRI ELARDLKL--GKKKKLEITKKNRQREQERKEAK-------------NQLEKEGVKPTGMN   543
WP_002864485.1          HKINI ELAREVGK--NHSQRAKIEKEQNENYKAKKDAELECE-------------KLGLKINSKN   523
CBN87970.1              ARIHI ETAREVGK--SFKDRKEIEKRQEENRKDREKAAAKFRE---------YFPNFVGEPKSKD   552
WP_014574210.1          ARIHI ETAREVGK--SFKDRKEIEKRQEENRKDREKAAAKFRE---------YFPNFVGEPKSKD   552
AAV62253.1              DNIVI EMARETNE--DDE-KKAIQKIQKANKDEKDAAMLKAANQYNGK---AELPHSVFHGHKQL   568
sp|J7RUA5.1|SaCas9      NDIII ELAREKNS--KDA-QKMINEMQKRNRQTNERIE--EI----------IRTTGKENAKYL   520
CAH11307.1              YEIAM AKWEQIKHIP-DNSSLLIPIYLEQNRFEFEESFKKIKGSSSDKTLEQAIEKQN---IQWE   701
WP_003038941.1          KNIVD DNWQNIKQVLSAKHQLHIPIITESNAFEFEPALADVKGKSLKDRRKKALERIS-PENIFK   934
                          *      .:
```

RuvC III

```
sp|Q99ZW2.1|SpCas9      --EINNKHHAHDAYLNAVVGTALIKKYPKLESEFVYGD--YKVYDVRKMIAKSEQEIGK-   1031
Cas9.3                  KIRDDHRHHAVDAAVVALTSPGRIK---RLSTFYQYRK---------------------   751
WP_002864485.1          --RNNHLHHAIDAVIIAYANNSIVK---AFSDFKKEQE--SNSAEL--Y---AK-KISEL   747
CBN87970.1              --AENDKHHALDAVVVACSTVAMQQ---KITRFVRYKE--MNAFDGKTI---DK-ETG-E   764
WP_014574210.1          --AENDKHHALDAVVVACSTVAMQQ---KITRFVRYKE--MNAFDGKTI---DK-ETG-E   764
AAV62253.1              --D-TYKHHAVDALIIAASS-QLNLWKKQKNTLVSYSE--EQLLDIET----GE-LISDD   787
sp|J7RUA5.1|SaCas9      --NKGYKHHAEDALIIANADFIFKEWKKLDKAKKV-ME--NQMFEEKQAESMPE-IETEQ   748
CAH11307.1              SRQQSFPSHAIDATLTMSIGLKEFP---Q---FSQELDNSWFINHLMP----DE------   943
WP_003038941.1          DKPQASYSHLIDAMLAFCIAADEHR---NDGSIGLEIDKNYSLYPLDK--NTGE------   1202
                                 *  ** ;
```

FIG. 4E

Cas9.4

Domain Structure (RuvC I &+ Bridge Helix)

RuvC I

```
Cas9.4                  --------MKKILGLDLGTDSIGWTIVQQNEEK-KFKLID-------------KGVRIFQ   38
sp|Q99ZW2.1|SpCas9      ------MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFD   54
WP_002864485.1          --------MARILAFDIGISSIGWAFSEND------ELKD-------------CGVRIFT   33
CBN87970.1              MAAFKPNPMNYILGLDIGIASVGWAMVEIDKEENPIRLID-------------LGVRVFE   47
WP_014574210.1          MAAFKPNSINYILGLDIGIASVGWAMVEIDEEENPIRLID-------------LGVRVFE   47
AAV62253.1              -------MSDLVLGLDIGIGSVGVGILNKVT----GEIIH-------------KNSRIFP   36
sp|J7RUA5.1|SaCas9      ------MKRNYILGLDIGITSVGYGIIDYET----RDVID-------------AGVRLFK   37
CAH11307.1              ---MESSQILSPIGIDLGGKFTGVCLSHLEAFA---ELPNHANT---------KY-----   40
WP_003038941.1          -----MNFKILPIAIDLGVKNTGVFSAFYQKGT---SLERLDNK---------NG-----   38
                                  :.:*:*    *
```

Bridge Helix

```
Cas9.4                  KGVGEEKNNEFSLAKERTTHRNTRKKYRRTKQRKVRLLRELIK--HGMCPLSFDEL----   92
sp|Q99ZW2.1|SpCas9      SG-------E--TAEATRLKRTARRRYTRRKNRICYLQEIFS---NEMAKVDDSFFHRLE  102
WP_002864485.1          KVENPKTGES--LALPRRLARSARKRLARRKARLNHLKHLIA---NEFKLNYEDYQ-SFD   87
CBN87970.1              RAEVPKTGDS--LAMARRLARSVRRLTRRRAHRLLRARRLLK---REGVLQDADFD-E--   99
WP_014574210.1          RAEVPKTGDS--LAMARRLARSVRRLTRRRAHRLLRTRRLLK---REGVLQAANFD-E--   99
AAV62253.1              AAQAE-------NNLVRRTNRQGRRLTRRKKHRIVRLNRLFE---ESGLITDFTK-----   81
sp|J7RUA5.1|SaCas9      EANVE-------NNEGRRSKRGARRLKRRRRHRIQRVKKLLF---DYNLLTDHSE-----   82
CAH11307.1              -SVILIDHNNFQLSQAQRRAT---RHRVRNKKRNQRVKRVALQLFQHILSRDLNA-----   91
WP_003038941.1          -KVYELSKDSYTLLMNNRTAR---RHQRRGIDRKQLVKRLFKLIWTEQLNLEWDK-----   89
                                          :   *   *
```

FIG. 4E continued

Domain Structure (RuvC II & RuvC III)

Cas9.4

RuvC II

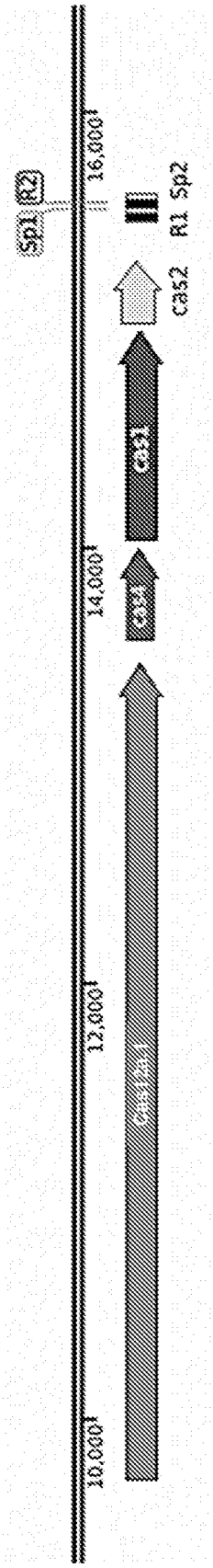
```
Cas9.4                  RGIDETVIHLEMSRELKN--ONKRLAIQRYQQARNEENQEYAK-EIK-KIF--KEQTQK   689
sp|Q99ZW2.1|SpCas9      MGRHKPENIVIEMAREKNT--TQKGQKNSRERMK-------RIEEGI--KELGSQILKEH   799
WP_002864485.1          YG--KVHKINIELAREVGK--NHSQRAKIEKEQNENYKAKKDAELEC----------KL   515
CBN87970.1              YG--SPARIHIETAREVGK--SFKDRKEIEKRQEENRKDREKAAAKFRE------YFPNF   544
WP_014574210.1          YG--SPARIHIETAREVGK--SFKDRKEIEKRQEENRKDREKAAAKFRE------YFPNF   544
AAV62253.1              YG--DPDNIVIEMARETNE--DDE-KKAIQKIQKANKDEKDAAMLKAANQYNGKAELPHS   560
sp|J7RUA5.1|SaCas9      YG--LPNDIIIELAREKNS--KDA-QKMINEMQKRNRQTNERIE--EI--------IRTT   512
CAH11307.1              MMQRLAYEIAMAKWEQIHHIP-DNSSLLIPIYLEQNRFEFEESFKKIKGSSS--DKTLEQ   691
WP_003038941.1          MATILAKNIVDDNWQNIKQVLSAKHQLHIPIITESNAFEFEPALADVKGKSL--KDRRKK   922
                                 *     .:
```

RuvC III

```
Cas9.4                  TADFRKAWGIQNPD-----TSKSRQRHTHHAIDAAVVACLTRDRYDFLTQWYRAEE----   937
sp|Q99ZW2.1|SpCas9      VSDFRKDFQFYKVR---------EINDYHHAHDAYLNAVVGTALIK---KYPKLESEF-V   1009
WP_002864485.1          TSALRHTWGFSAKD---------RNNPLHHAIDAVIIAYANNSIVK-------AFSDFKK   730
CBN87970.1              TNLLRGFWGLRKVR---------AENDRHHALDAVVVACSTVAMQQ-------KITRFVR   746
WP_014574210.1          TNLLRGFWGLRKVR---------AENDRHHALDAVVVACSTVAMQQ-------KITRFVR   746
AAV62253.1              TSQLRHMGIEKTR---------D-TYHHMVDALIIAASS-QLNL----WKKQNMTLVS   769
sp|J7RUA5.1|SaCas9      TSFLRRKWKFKKER---------NKQYKHHAEDALIIANADFIFKE----WKKLDKAKKV   727
CAH11307.1              VHDHRELLSKQEPKLV----KSRQQSTPSHAIDATLTMSIGLKEFP-------Q---FSQ   928
WP_003038941.1          IAEIRQLYEKVDSDIQAYAKGDKPQASYSHLIDAMLAFCIAADEHR-------NDGSIGL   1185
                            *                       *  ** :
```

FIG. 5A

Cas12a.1



Cas12a.1 (SEQ ID NO: 177)

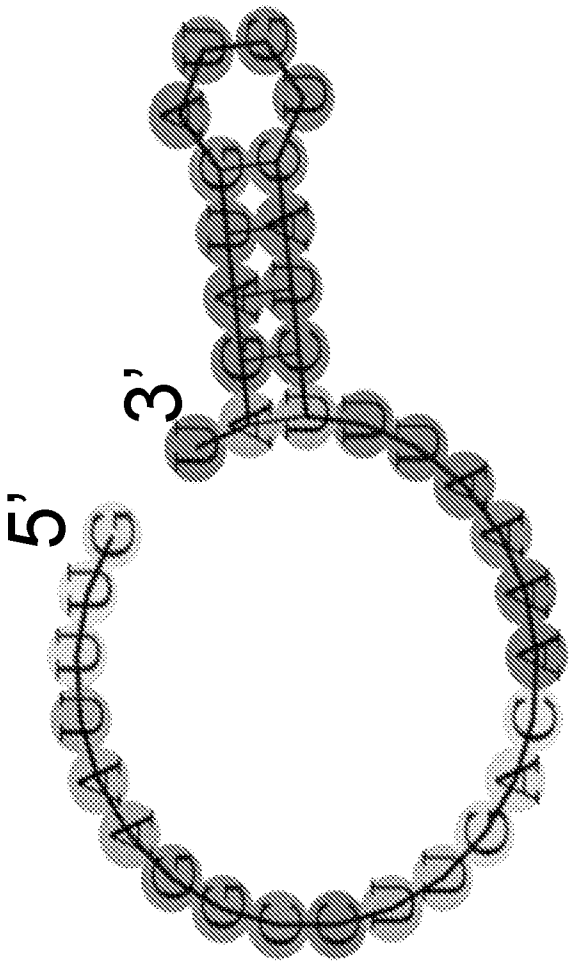


FIG. 5B

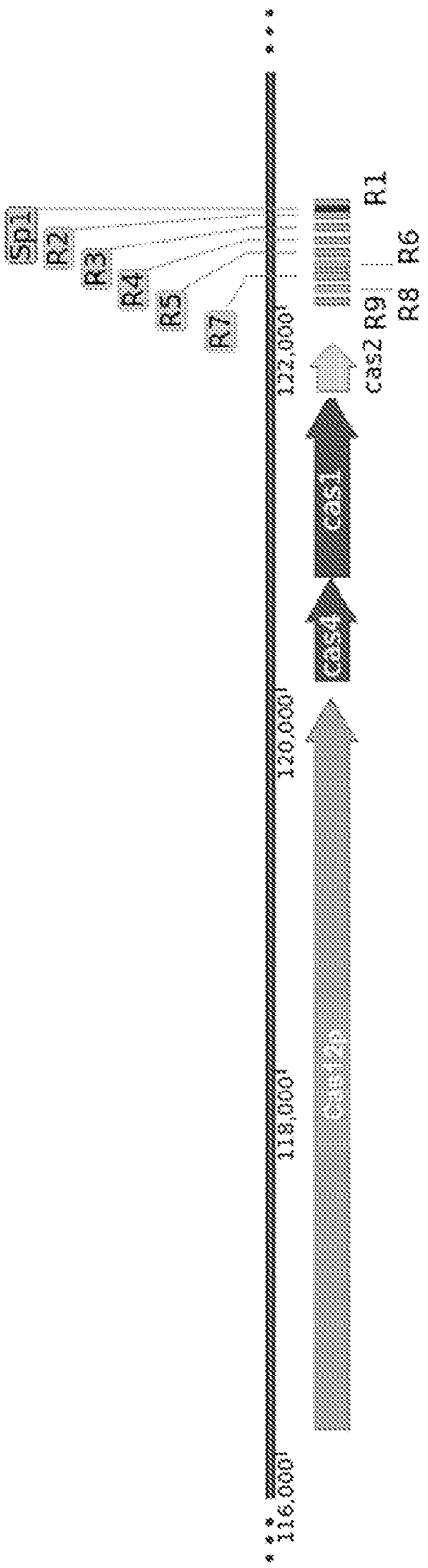


FIG. 5C

Cas12p

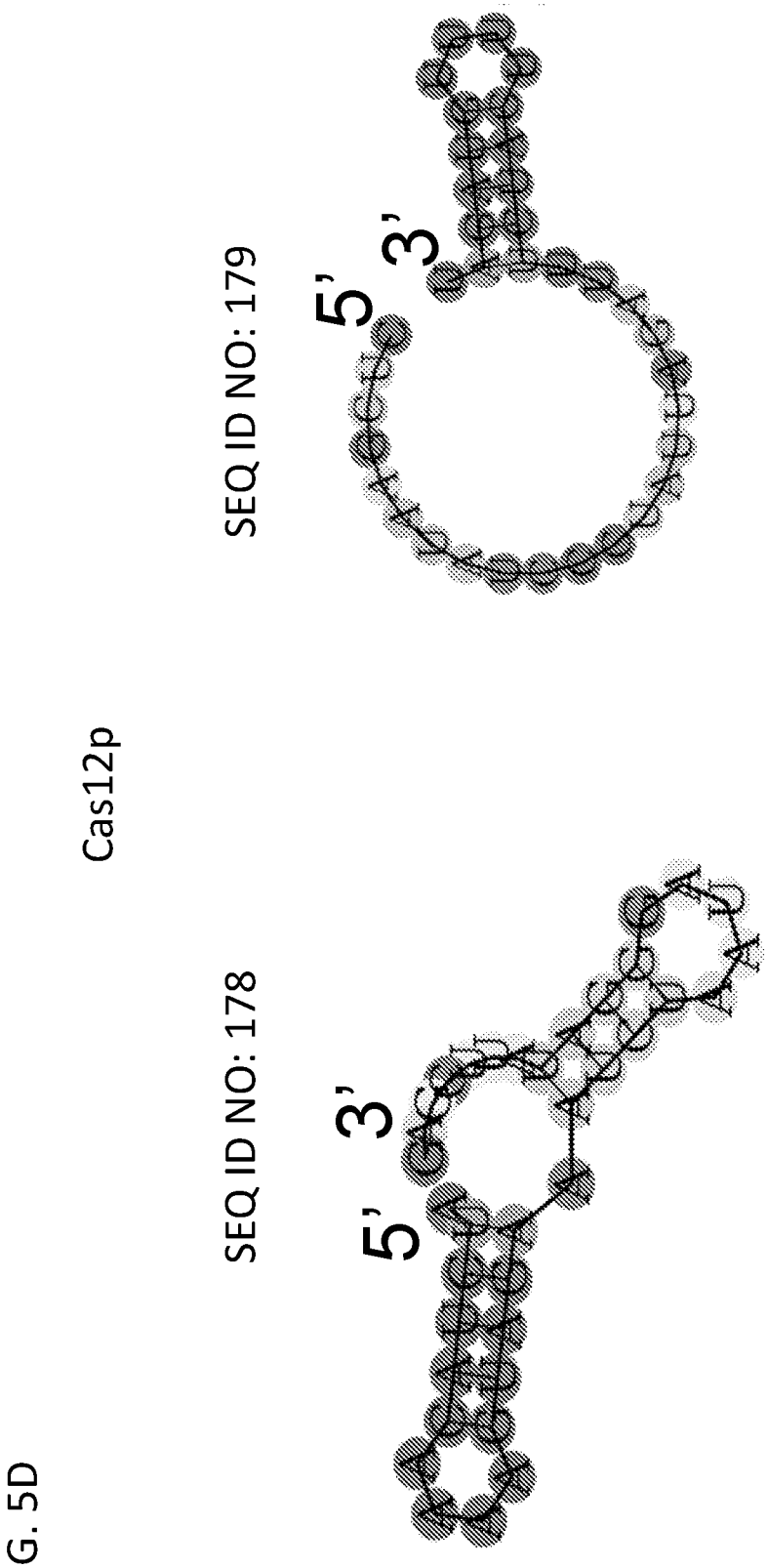


FIG. 5D

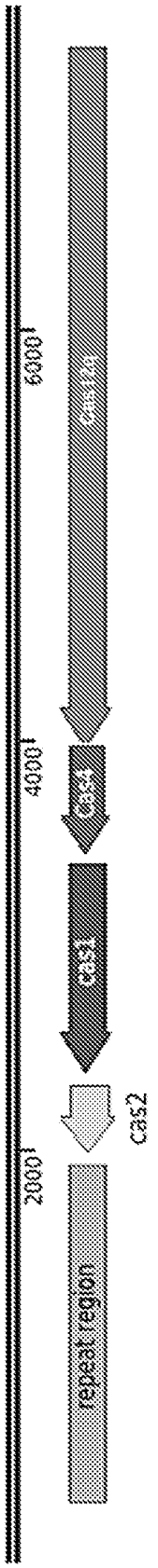


FIG. 5E

Cas12q

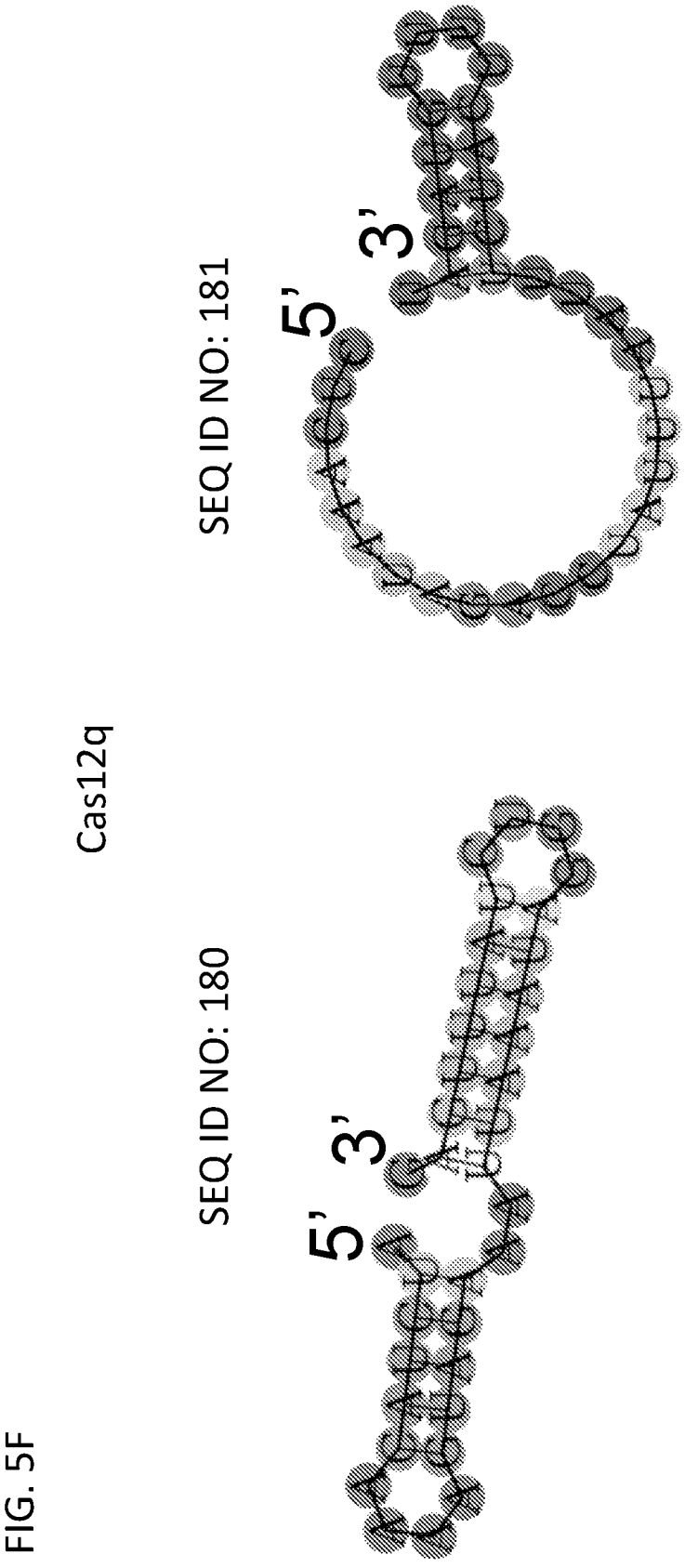
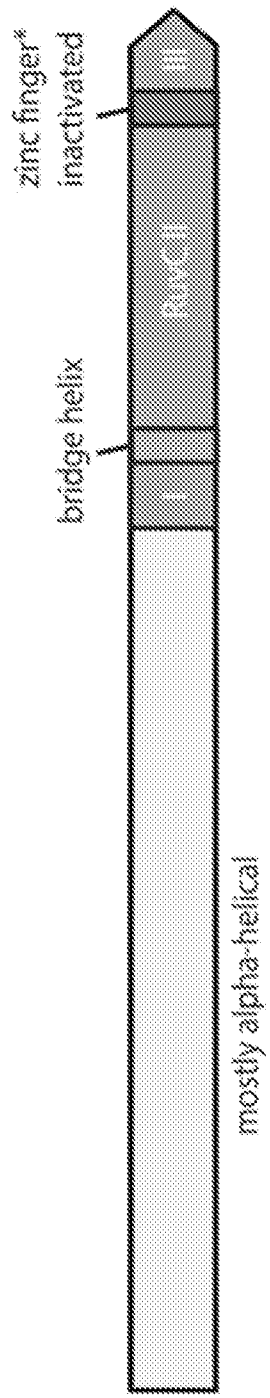


FIG. 5F

FIG. 6A


zinc finger*
inactivated
bridge helix
RuvC III
mostly alpha-helical

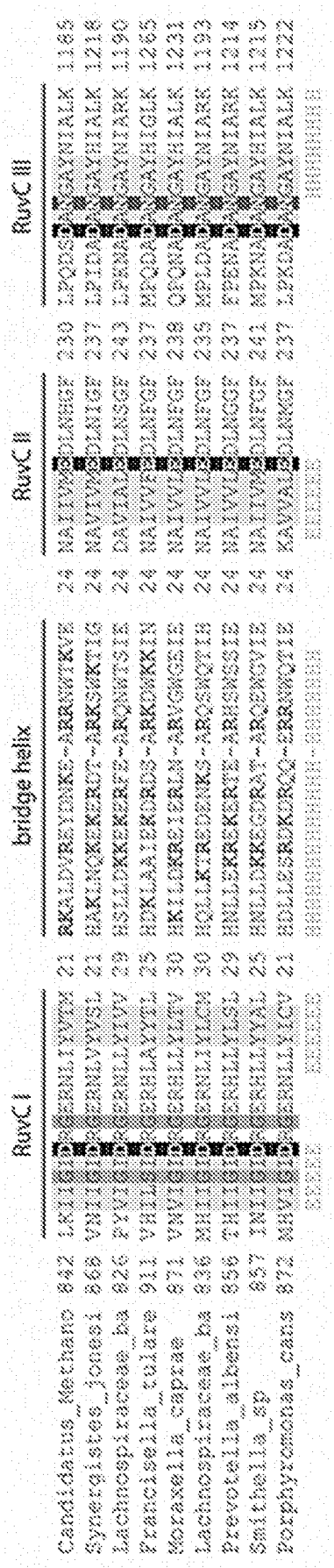

RuvC I
bridge helix
RuvC II
RuvC III
Candidatus_Methano 842 LKLIGERMLYYTDK 21 KKALDVREVIENKE-ARRNWTKYE 24 NAIIVMDQLNEGF 230 LPQDSKAK*GAYMIALK 1185
Synergistes_jonesi 868 VWTIGERMLYYTSL 21 RAKLNQKEKERQT-ARKSWTKTIG 24 NAVIVMDQLNIGF 237 LPIDAKSK*GAYHIALK 1218
Lachnospiraceae_ba 826 FYVIGERMLYYTDV 29 RSLLDKKEKERFE-ARQNWTSIE 24 DAVIAIDQLNSGF 243 LPEMAKAK*GAYNIARK 1190
Francisella_tularе 911 VRLLGERELAYYTL 25 RDKLAAIEKQRDS-ARKDWKKIN 24 NAIVVFDQLNFGF 237 MPQDAKAK*GAYHIGLK 1265
Moraxella_caprae 871 VNVIGERLLAYYTV 30 HKILDKREIERLN-ARVGWGEIE 24 NAIVVLDQLNFGF 239 QPQMAKAK*GAYHIALK 1231
Lachnospiraceae_ba 836 RHLIGERMLYYLCM 30 HQLLKTRGDEWKS-ARQSWQTIE 24 NAIVVLDQLNFGF 235 MPLDAKAK*GAYNIARK 1193
Prevotella_albensi 856 THLIGERLLYLSL 29 RNLLSKRCKERTE-ARHSWSSIE 24 NAIVVLDQLNGGF 237 FPEMAKAK*GAYNIARK 1014
Smithella_sp 857 INLIGERLLYYAL 25 RNLLDKREGDRAT-ARQEWGVIE 24 NAIVVFDQLNFGF 241 MPKNAKAK*GAYHIALK 1215
Porphyromonas_cans 872 MHVIGERMLYICV 21 RDLLESEKQKQRQ-ERRKWQTIE 24 KAVVALDQLNKGF 237 LPKQAKAK*GAYMIALK 1222

FIG. 6B

FIG. 6B continued

```
Cas12a1                TLNFQARDGSKTIGKNVKDHIRETKDDIFVLSIDRGEPHLAYYTLLNSKGEIQEQGSFNV  890
US_2016_0208243_A1_01  TINFKSS-GAMKFKDIFNLLEKAKDVKITLSIDRGERHLAYYTLVDGKGNIIKQDTFNI  943
                       *:::::  *:::::*:**  :::*.:**:*********************:::*:*:::*:

Cas12a1                ISDDKERKKDYHEKLDERLKGVREKFDKARSWKKIETIKKLKDGYLSQIVHKIAKLAIEYNAI  950
US_2016_0208243_A1_01  IGNDR-MKTNYHDKLAAIEKDRDSARKDWKKIENIKEMKEGYLSQVVHEIAKLVIEHNAI  1002
                       *.:*:  *.:*****  ********.***.**:*:*:*:*:*:*:*:*:*********:****

Cas12a1                IVLEDLNLGFMRGRQKVEKQVYQKFEKMLIDKLNYLVFKEKETEKEAGGVLNAYQLTAPFE  1010
US_2016_0208243_A1_01  VVFEDLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDKEFDKTGGVLRAYQLTAPFE  1062
                       *:*******:*********** **:*********:  :***.*******.  **

Cas12a1                GPKKLGKETQIIYYPAAYTSKICPKTGFVNLLRPKYMNIEKAKEFFKKFDKIRYNSEKG  1070
US_2016_0208243_A1_01  TPKMMKQTGIIYYPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKG  1122
                       **********.*******.***** * **:::::.*****.***: * *: :.*

Cas12a1                LFEFNFDYSNFIKNKKETKIIQDKEVYSKDTKLVGFRKMMVKSKQTKEKKPNEKLKI  1130
US_2016_0208243_A1_01  YFEFSFDYKNFGDK------AAKGKITIASFGSRLIWFDSKKNMGQYREVYPTKELEK  1176
                       **.***:*.*.:        .:.***: * *:::*:.***:.***:.******.*:::*:

Cas12a1                LFKEYGVSFQKDENIISQIASQMKKAFFELIKPKTTLMLRNSKKDFEEDYVLSCVKDE  1180
US_2016_0208243_A1_01  LLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLMTILQMRNSKTGTELDYLISPVADV  1236
                       *:::*:::.:.*.*.*.*.:.*.*.*.** :* .:::**** ****:. *.***:*.*.*

Cas12a1                NGEFFDSRKAADNKPKDADANGAYHIGLKGLMLLERINKNKGKKKLDLLISNKDFINFAV  1250
US_2016_0208243_A1_01  NGNFFDSRQAPKNMPQDADANGAYHIGLKGLMLLDRIKKNQEGKKLNLVIKNEEYFEFVQ  1296
                       **:*******:*.*.*:*********************:**** *: *:*****.*******.

Cas12a1                EKSK 1254
US_2016_0208243_A1_01  NPNN 1300
                       :*.:
```

FIG. 6C

FIG. 6C continued

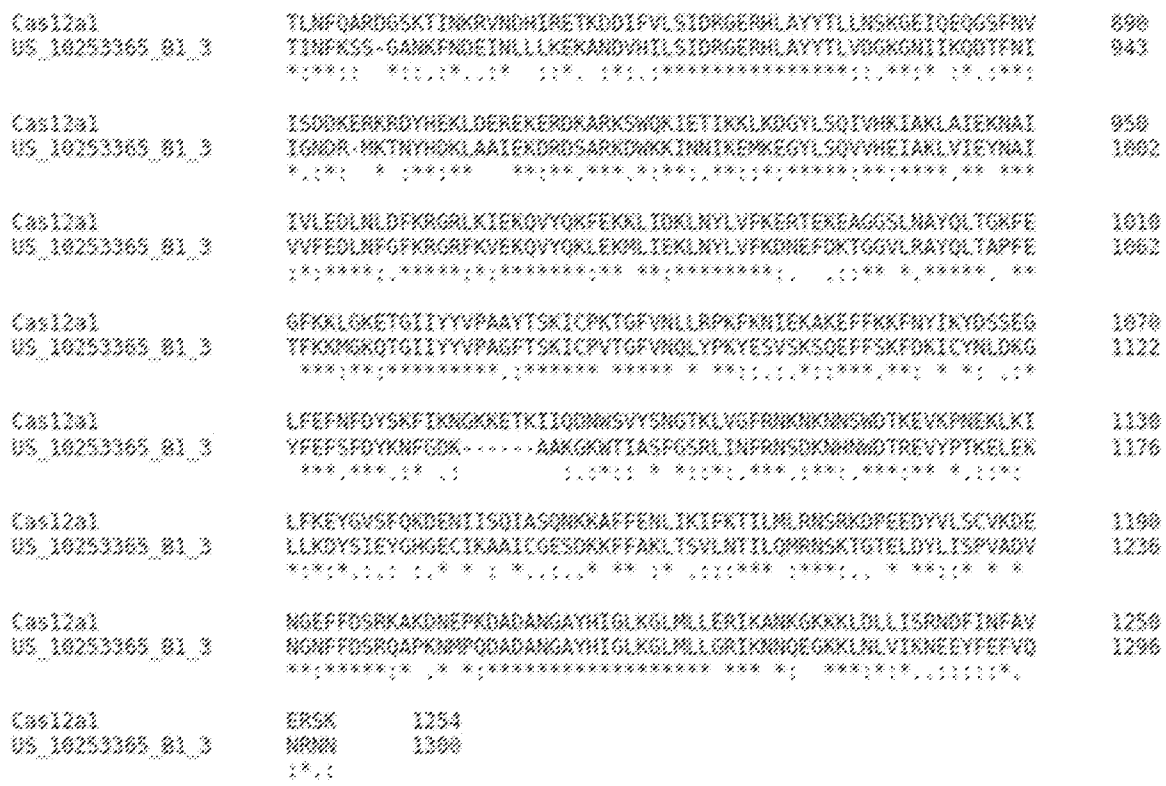
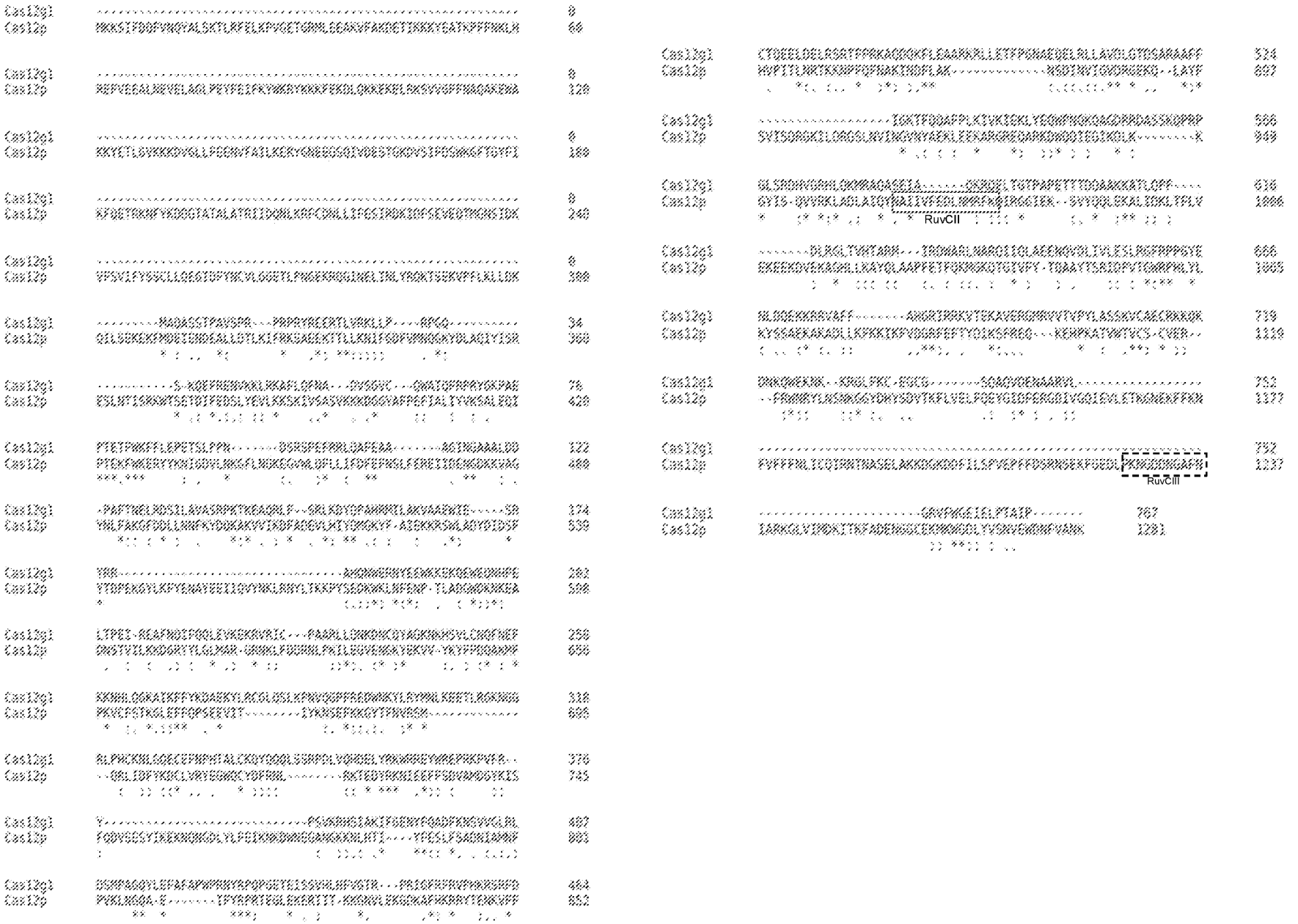


FIG. 6D

>Cas12p (SEQ ID NO: 4)

MKKSIFDQFVNQYALSKTLRFELKPVGETGRMLEEAKVFAKDETIKKKYEATKPFFNKLHREFVEEALNEVELAGLPEYFEIFKYWKRYKKKFEKDLQK
KEKELRKSVVGFFNAQAKEWAKKYETLGVKKKDVGLLFEENVFAILKERYGNEEGSQIVDESTGKDVSIFDSWKGFTGYFIKFQETRKNFYKDDGTAT
ALATRIIDQNLKRFCDNLLIFESIRDKIDFSEVEQTMGNSIDKVFSVIFYSSCLLQEGIDFYNCVLGGETLPNGEKRQGINELINLYRQKTSEKVPFLKLLD
KQILSEKEKFMDEIENDEALLDTLKIFRKSAEEKTTLLKNIFGDFVMNQGKYDLAQIYISRESLNTISRKWTSETDIFEDSLYEVLKKSKIVSASVKKKDGG
YAFPEFIALIYVKSALEQIPTEKFWKERYYKNIGDVLNKGFLNGKEGVWLQFLLIFDFEFNSLFEREIIDENGDKKVAGYNLFAKGFDDLLNNFKYDQK
AKVVIKDFADEVLHIYQMGKYFAIEKKRSWLADYDIDSFYTDPEKGYLKFYENAYEEIIQVYNKLRNYLTKKPYSEDKWKLNFENPTLADGWDKNKE
ADNSTVILKKDGRYYLGLMARGRNKLFDDRNLPKILEGVENGKYEKVVYKYFPDQAKMFPKVCFSTKGLEFFQPSEEVITIYKNSEFKKGYTFNVRSM
QRLIDFYKDCLVRYEGWQCYDFRNLRKTEDYRKNIEEFFSDVAMDGYKISFQDVSESYIKEKNQNGDLYLFEIKNKDWNEGANGKKNLHTIYFESLFS
ADNIAMNFPVKLNGQAEIFYRPRTEGLEKERIITKKGNVLEKGDKAFHKRRYTENKVFFHVPITLNRTKKNPFQFNAKINDFLAKNSDINVIGVDRGE
KQLAYFSVISQRGKILDRGSLNVINGVNYAEKLEEKARGREQARKDWQQIEGIKDLKKGYISQVVRKLADLAIQYNAIIVFEDLNMRFKQIRGGIEKSV
YQQLEKALIDKLTFLVEKEEKDVEKAGHLLKAYQLAAPFETFQKMGKQTGIVFYTQAAYTSRIDPVTGWRPHLYLKYSSAEKAKADLLKFKKIKFVDG
RFEFTYDIKSFREQKEHPKATVWTVCSCVERFRWNRYLNSNKGGYDHYSDVTKFLVELFQEYGIDFERGDIVGQIEVLETKGNEKFFKNFVFFFNLIC
QIRNTNASELAKKDGKDDFILSPVEPFFDSRNSEKFGEDLPKNGDDNGAFNIARKGLVIMDKITKFADENGGCEKMKWGDLYVSNVEWDNFVAN
K

FIG. 6E



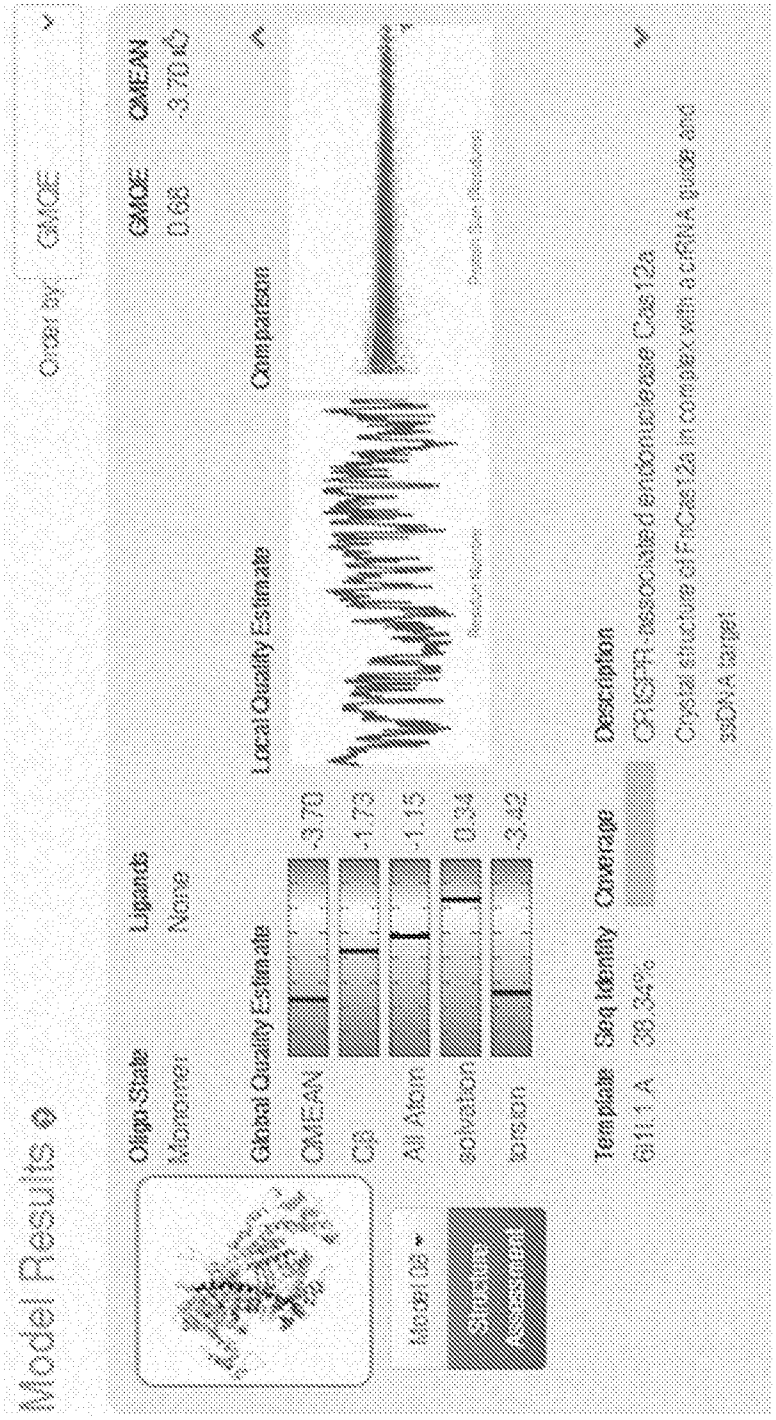

Model Results
Order by: GMQE
Oligo State
Monomer
Ligands
None
GMQE
0.68
GMEAN
-3.70 ±0
Global Quality Estimate
GMEAN
C8
All Atom
solvation
torsion
-3.70
-1.73
-1.15
0.04
-3.42
Local Quality Estimate
Comparison
Model 02
Structure
Alignment
Template
Seq Identity
Coverage
Description
6l11.1.A
38.24%
CRISPR-associated endonuclease Cas12a
Crystal structure of FnCas12a in complex with a crRNA guide and
ssDNA target

FIG. 6F

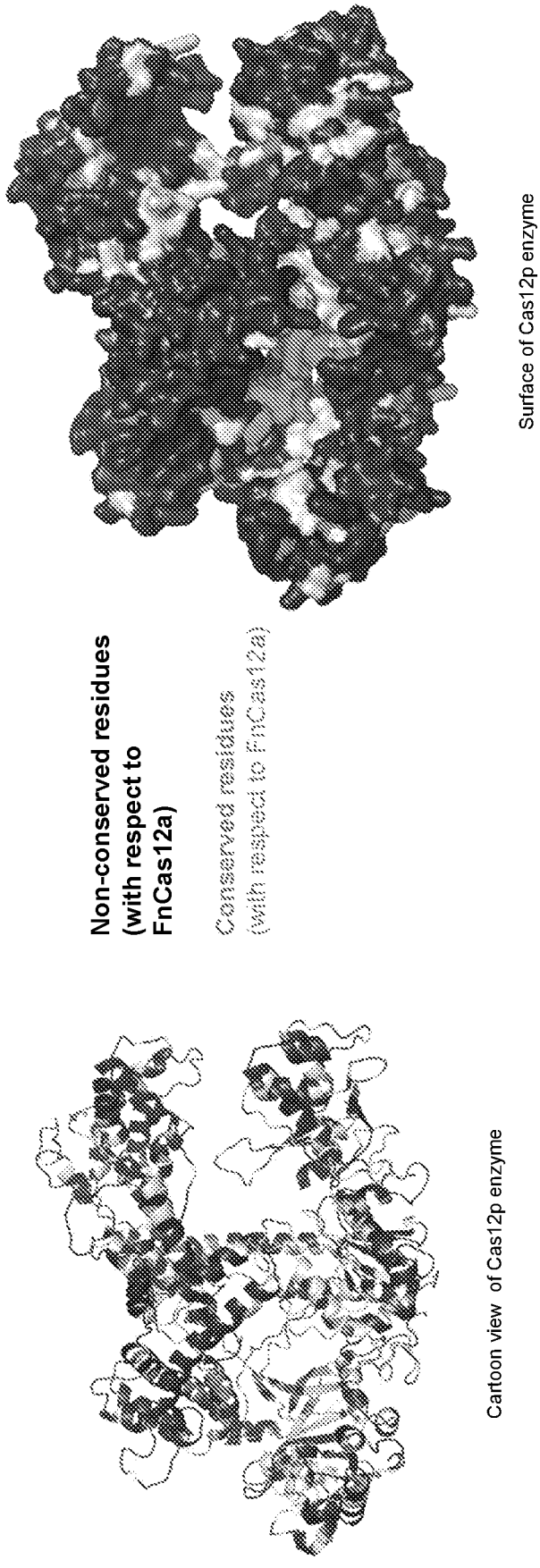

Surface of Cas12p enzyme
Non-conserved residues (with respect to FnCas12a)
Conserved residues (with respect to FnCas12a)
Cartoon view of Cas12p enzyme

FIG. 6G

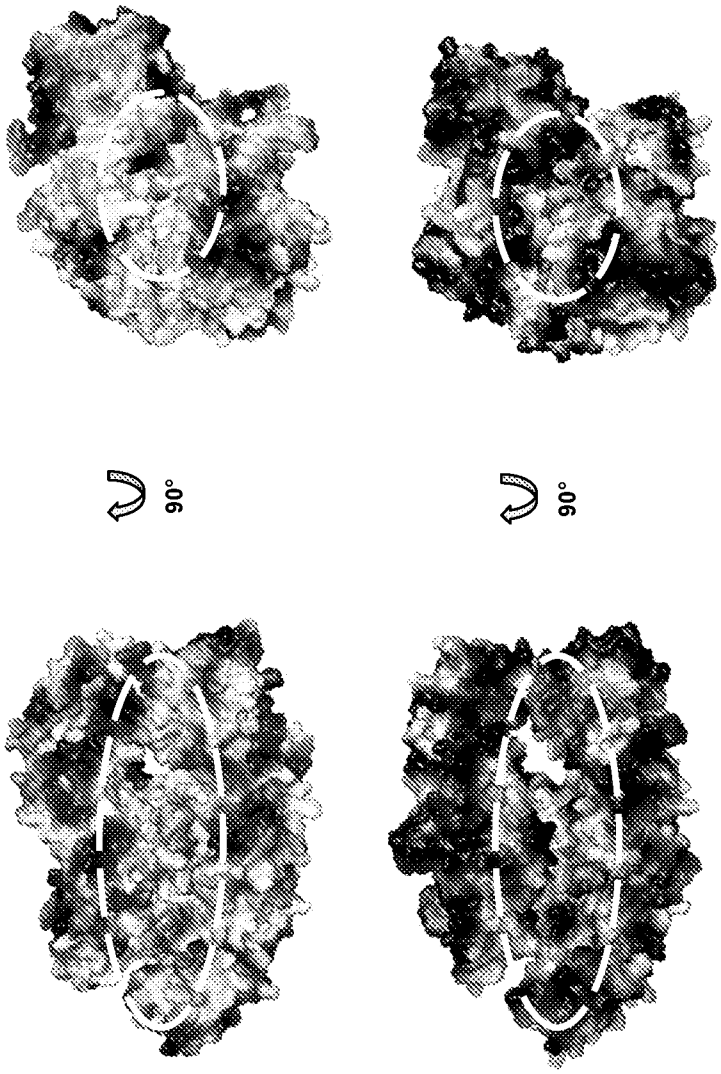


FIG. 6H

FIG. 6I

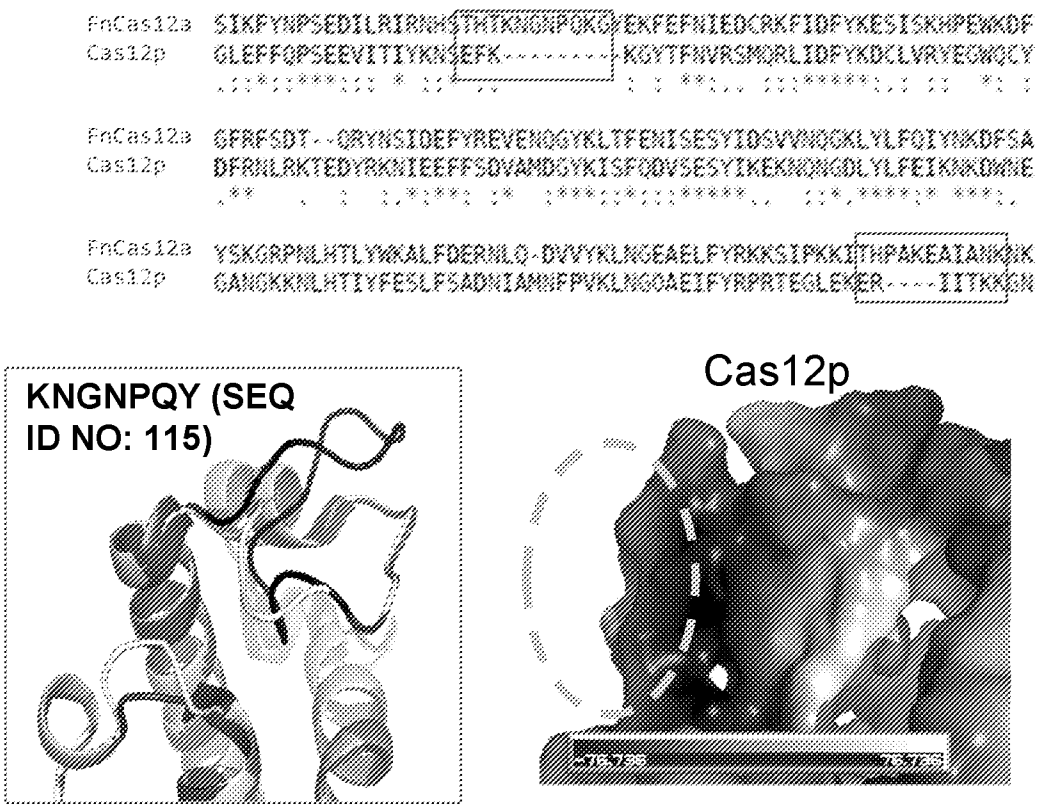

FnCas12a SIKPYNPSEDILRINNKSTHTKNGNPQKGYEKFEFNIEDCRKFIDFYKESISKHPEWKDF 737
Cas12p GLEPFQPSEEVITIYKNSEFK-------KGYTFNVRSNQRLIDPYKDCLVRYEGWQCY 715
FnCas12a GFRFSDT--GRYNSIDEFYREVENQGYKLTFENISESYIDSVVNQGKLYLFQIYNKDFSA 795
Cas12p DFRNLRKTEDYRKNIEEFFSQVANDGYKISFQDVSESYIKEKNQNGDLYLFEIKNKDWNE 775
FnCas12a YSKGRPNLHTLYWKALFDERNLQ-DVVYKLNGEAELFYRKKSIPKKITHPAKEAIANKNK 854
Cas12p GANGKKNLHTIYFESLFSADNIANNFPVKLNGQAEIFYRPRTEGLEKER----IITKKGN 831
PAKE
KNGNPQY (SEQ ID NO: 115)
Cas12p
FnCas12
PAKE SEQ ID NO: 114

FIG. 6J

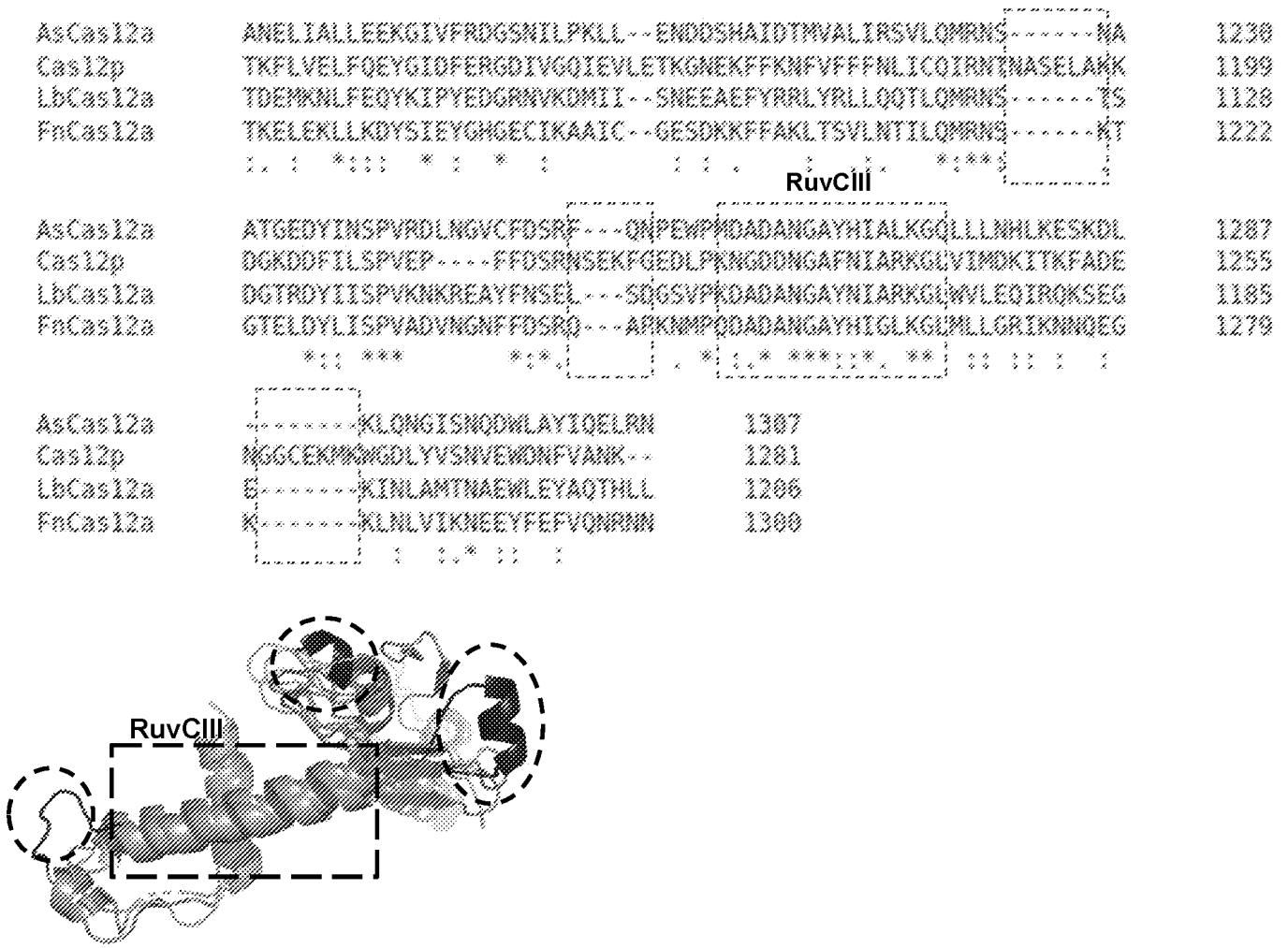


FIG. 6K

>Cas12q (SEQ ID NO: 5)

MINIDELKNLYKVQKTITFELKNKWENKNDENDRVEFLKTQEWVESLFKVDEENFDEKESIPNLLDFGQKIASLFYKLSEDIANNQIDTRVLKVSKFLLEEIDRNQY
HEKKNKPTKVKEMNPNTNKSYIKEYKLSDQNTLYVLLKIMEDEGRGLQKFLYDKADRLNLYNQKVRRDFALKESNEQQKFSGNANYYGNIKLLIDSLEDAVRIIG
YFTFDDQAENAQINEFKSVKQEMNNNEASYQALKDFAIDNAKKEIELTTLNHRAVNKDPKKIQEQIEEVENFEEDINQLKHQISALNDKKFDVVSRLKHALIKML
PELNLLDAESEQGREVQQIYQDKKNGLELDDFKFNLLKHHQWQKTIFKYIKLEGLVLPDLYAENKQDKIKVYIENYRQSGERISKKAREELGKIDKREEFNGNDEL
KKAWYEYKDFCRDKRNKSVELGNKKSLYNAIKREVLRQKMCNHFAVLVSDGEDTSPYYYLILIPNENSDEMNRTFKELKASEGNWKMLDYNRLTFKALEKLALL
RSSTFEIADQELQEEAKKIWEEYKEKAYKDFKNKKLLQGLSGRQREEKKQELQKESLNRVINYLIRCIQSLPDSGKYNFNFKEPHQYQSLEEFAEEIDRQGYHCAW
KNVSKDKLMELEAMEKIKVFKLHNKDFRKVKLNDSKHNPNLFTLYWLDAMNLDKVNVRLLPEVDLYKRAKETQLKLFERDVKCNINNQKIKSIKEKNRLFQDKL
YASFKLEFYPENEGLGFEQVNDKVNNFCGSDTAYYLGL**D**RGEKELVTFCLVDSDGRLVKNGDWTKFKEVNYADKLKQFYYSKGEIESTQQQLLEARDNIKQAT
NTEDKESM**K**LNYKKLELKLKQQNLLAQEFIKKAYCGYLIDSINEILREYPNTYLVL**E**DLDIAGKADPESGMTNKEQNLNKTMGASVYQAIENAIVNKFKYRTVKLS
DIKGLQTVPNVVKVEDLREVKEVEDGEHKFGLIRSVKSKDQIGNILFVDEGETSNTCPNCGFNSDWFKRDVDFDLEIVATVNGQKNAVIEQNDKKYCFPGEIYK
LEIINKEYETNKRNLAMIFKPRAKACRKFINNNLDKNDYFYCPYCAFSSKNCNNPKLQNGDFVVYSG**D**DVAAYNVAIRGINLLNNIK

FIG. 7A

Cas12a.1

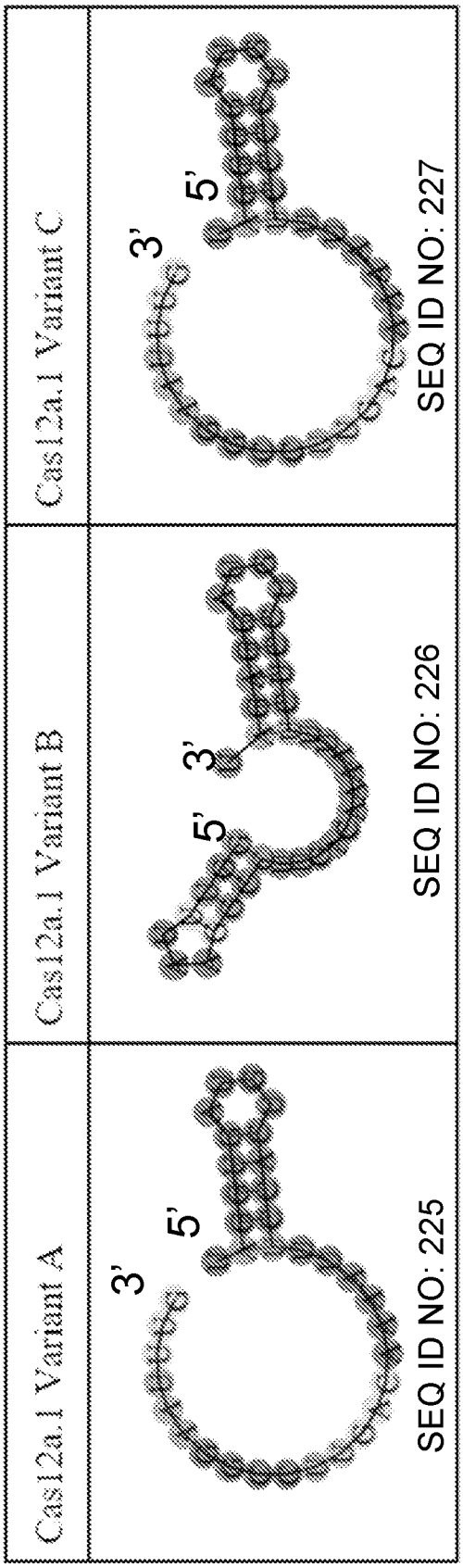
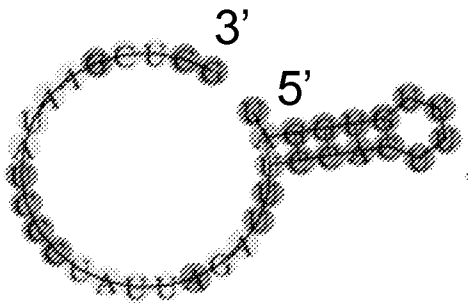
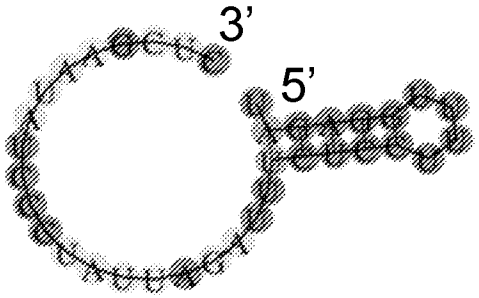
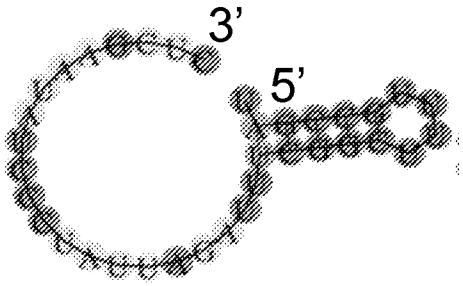
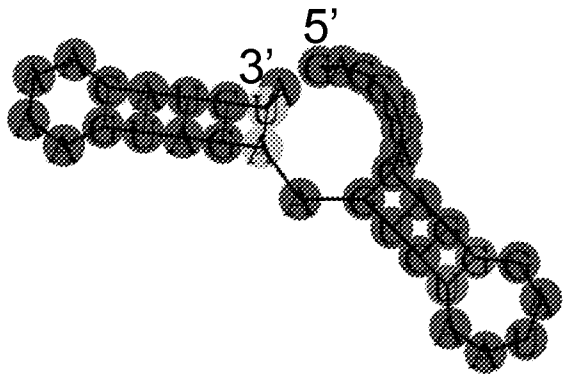


FIG. 7B

Cas12p

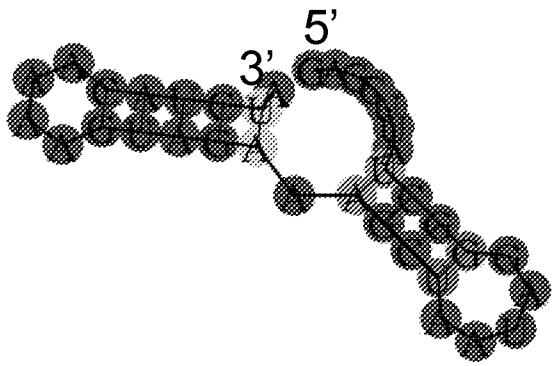
Cas12p Variant A	Cas12p Variant B	Cas12p Variant C
		
SEQ ID NO: 228	SEQ ID NO: 229	SEQ ID NO: 230

Cas12p Variant D



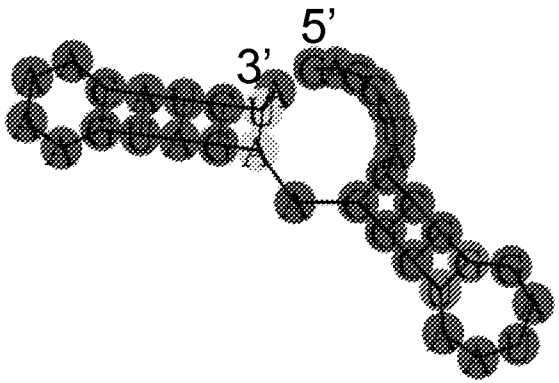
SEQ ID NO: 231

Cas12p Variant E



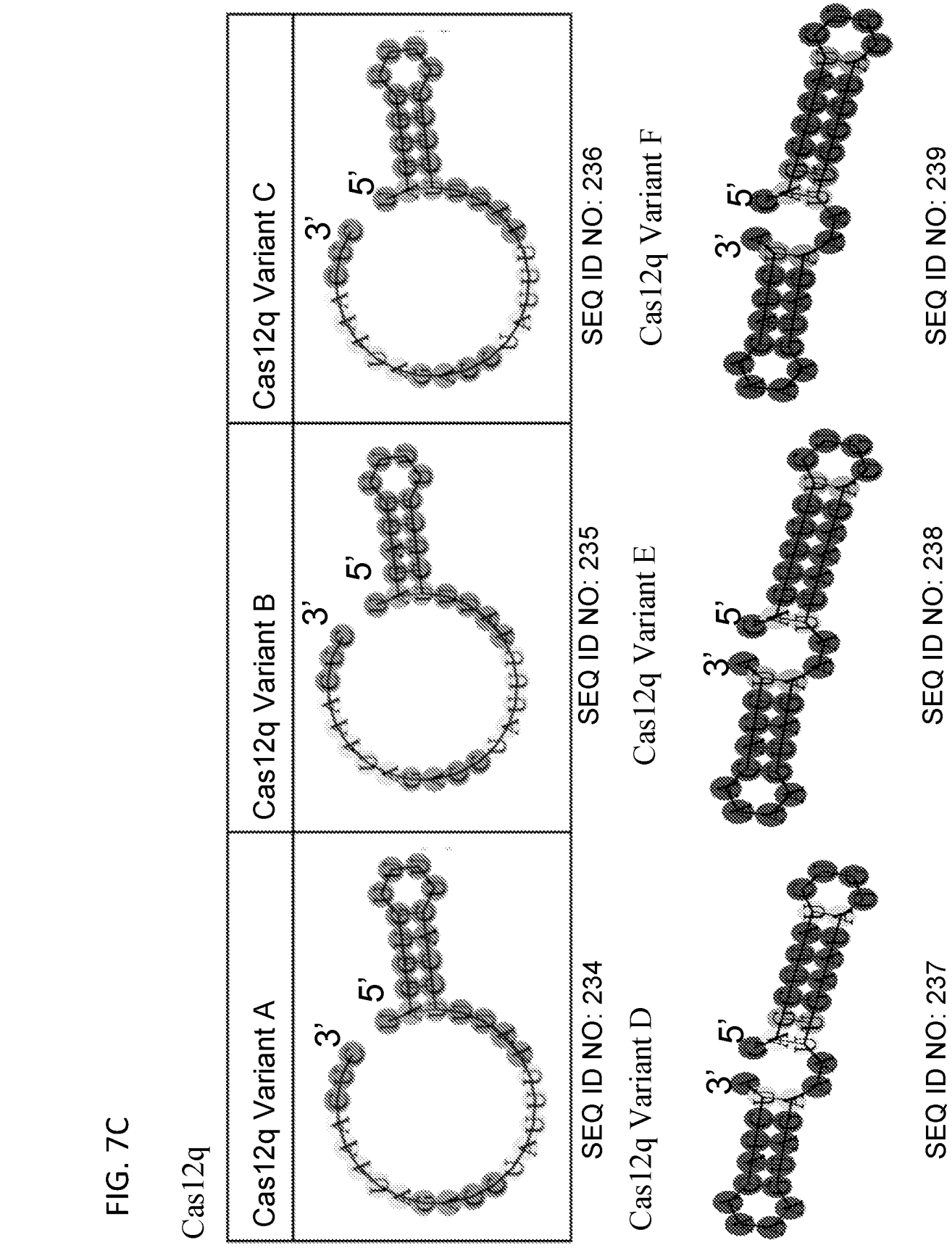
SEQ ID NO: 232

Cas12p Variant F



SEQ ID NO: 233

FIG. 7C


Cas12q
Cas12q Variant A
Cas12q Variant B
Cas12q Variant C
3'
5'
SEQ ID NO: 234
SEQ ID NO: 235
SEQ ID NO: 236
Cas12q Variant D
Cas12q Variant E
Cas12q Variant F
SEQ ID NO: 237
SEQ ID NO: 238
SEQ ID NO: 239

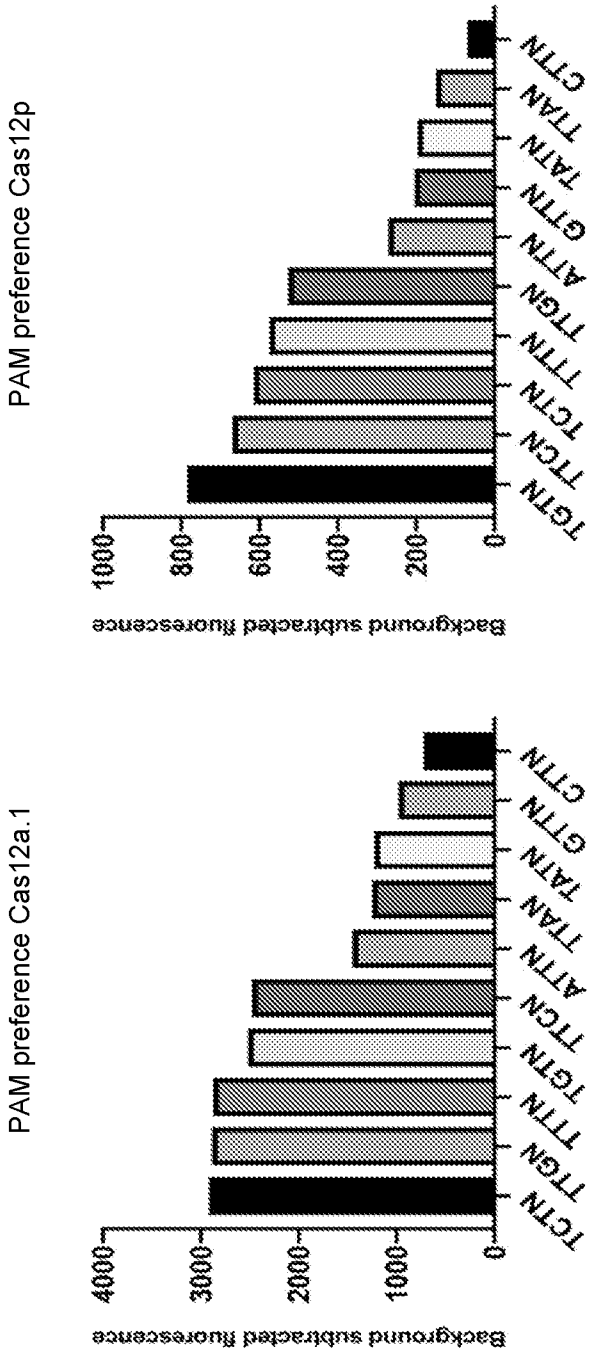

PAM preference Cas12p
Background subtracted fluorescence
1000
800
600
400
200
0
TGTN
TTCN
TCTN
TTTN
TTGN
ATTN
GTTN
TATN
TTAN
CTTN
PAM preference Cas12a.1
Background subtracted fluorescence
4000
3000
2000
1000
0
TCTN
TTGN
TTTN
TGTN
TTCN
ATTN
TTAN
TATN
GTTN
CTTN

FIG. 8

FIG. 9A

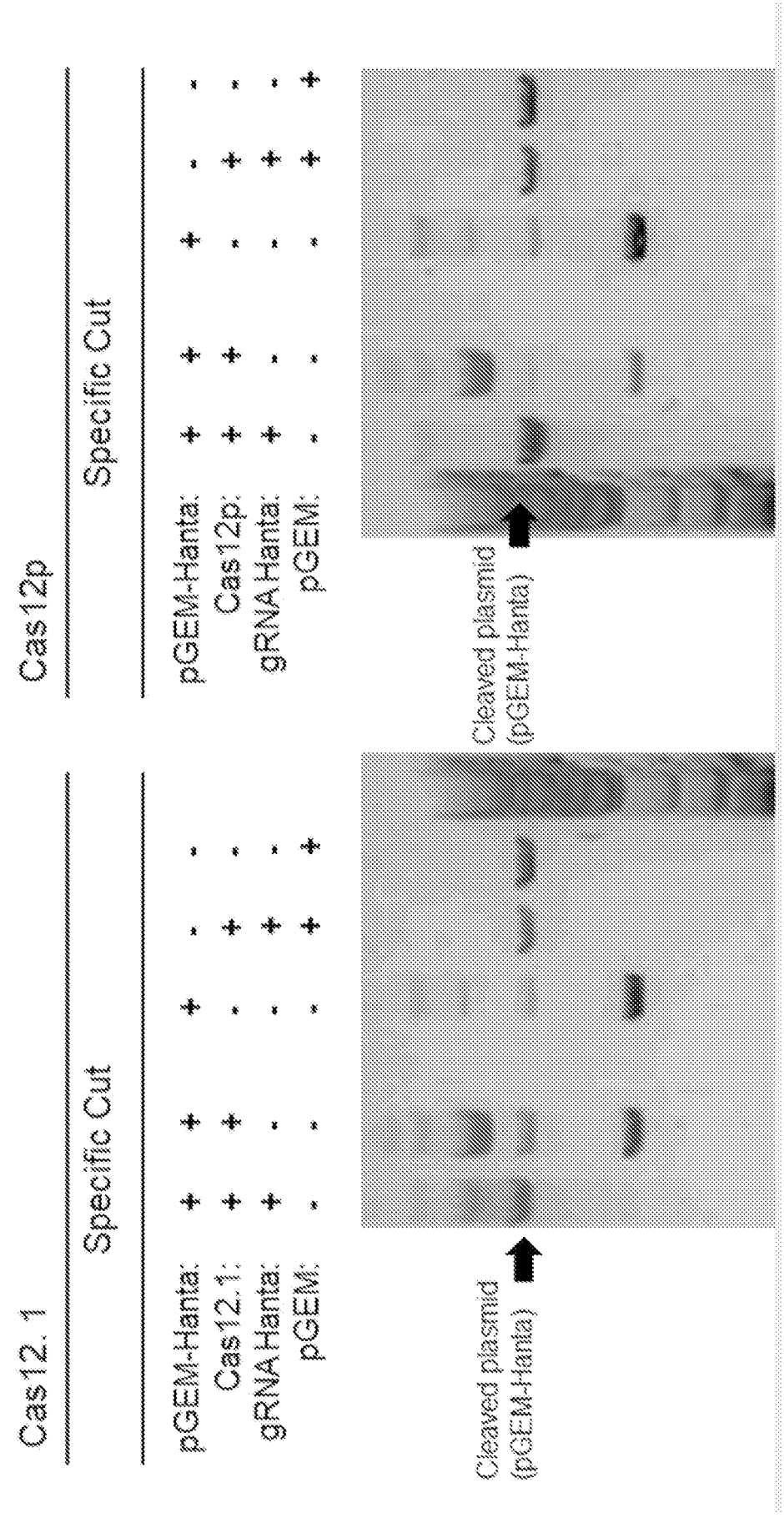

Cas12. 1
Specific Cut
pGEM-Hanta:
Cas12.1:
gRNA Hanta:
pGEM:
Cleaved plasmid
(pGEM-Hanta)
Cas12p
Specific Cut
pGEM-Hanta:
Cas12p:
gRNA Hanta:
pGEM:
Cleaved plasmid
(pGEM-Hanta)

FIG. 9B

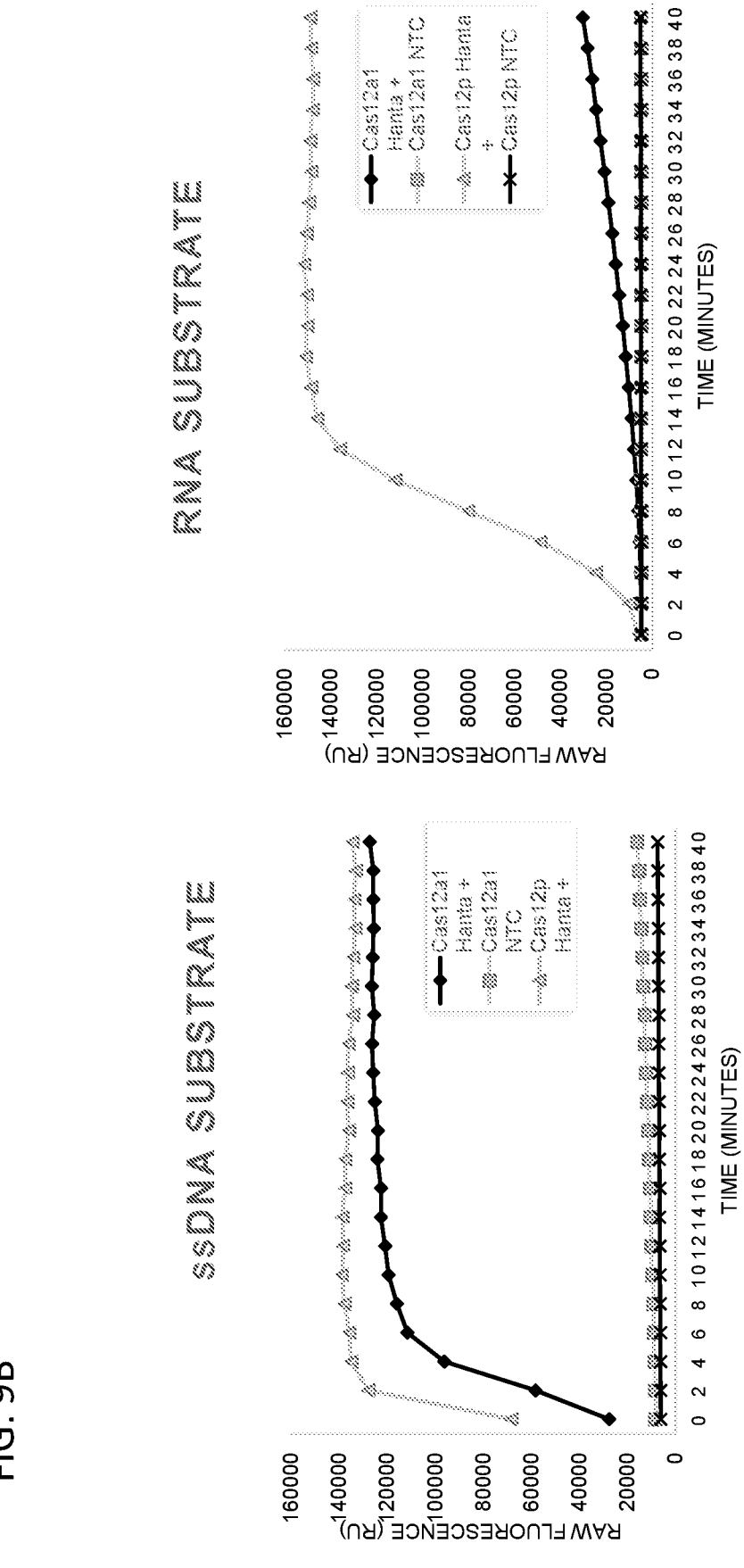
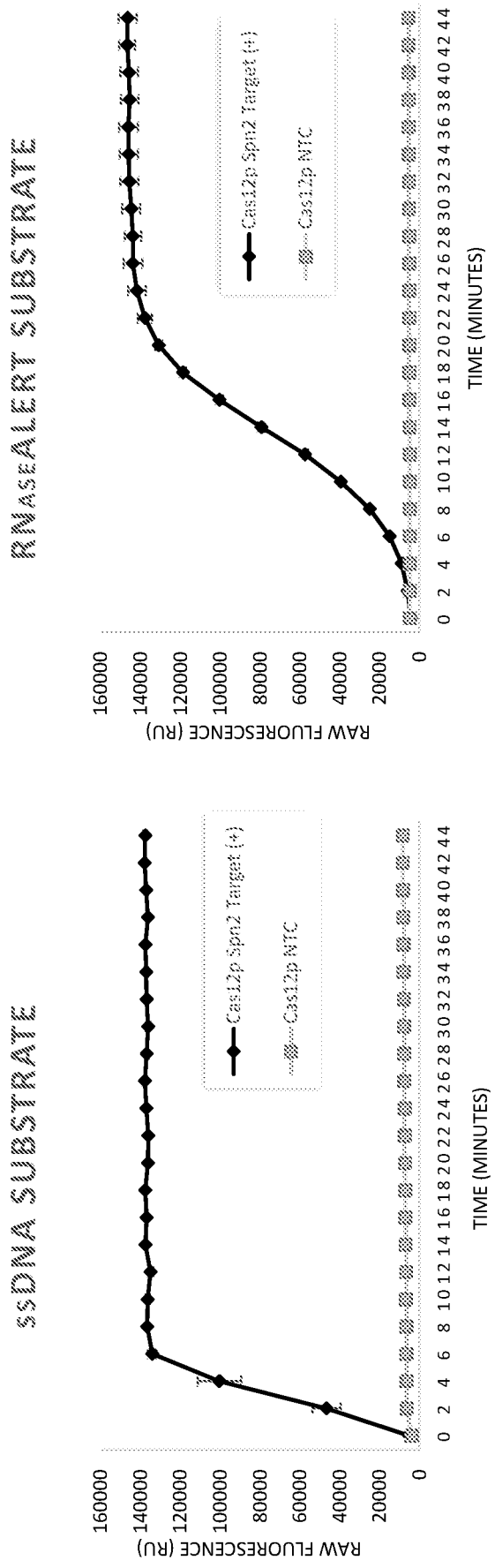

RNA SUBSTRATE
RAW FLUORESCENCE (RU)
TIME (MINUTES)
Cas12a1 Hanta +
Cas12a1 NTC
Cas12p Hanta +
Cas12p NTC
ssDNA SUBSTRATE
RAW FLUORESCENCE (RU)
TIME (MINUTES)
Cas12a1 Hanta +
Cas12a1 NTC
Cas12p Hanta +

FIG. 9C


RNaseAlert Substrate
RAW FLUORESCENCE (RU)
TIME (MINUTES)
Cas12p Spn2 Target (+)
Cas12p NTC
ssDNA SUBSTRATE
RAW FLUORESCENCE (RU)
TIME (MINUTES)
Cas12p Spn2 Target (+)
Cas12p NTC

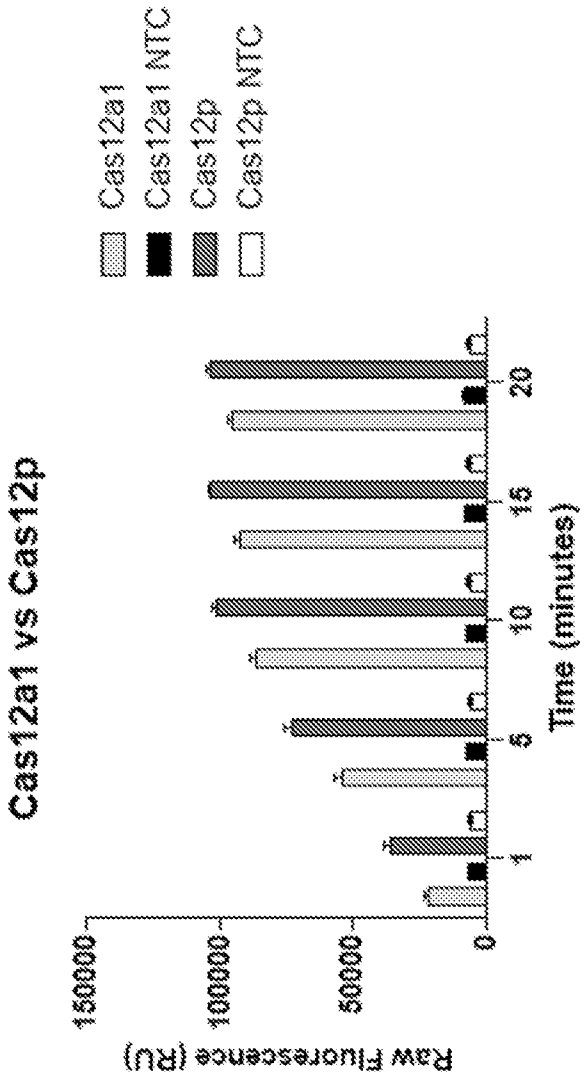
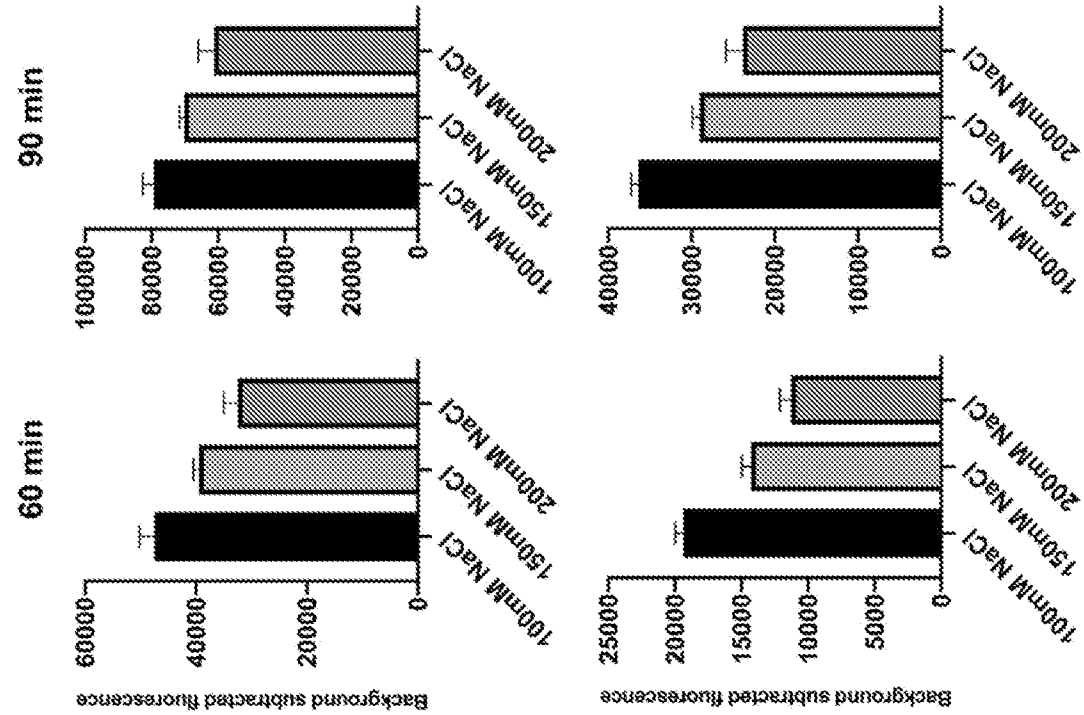

Cas12a1 vs Cas12p
Cas12a1
Cas12a1 NTC
Cas12p
Cas12p NTC
Raw Fluorescence (RU)
150000
100000
50000
0
1
5
10
15
20
Time (minutes)

FIG. 10



Cas12a.1 Cas12p

FIG. 11

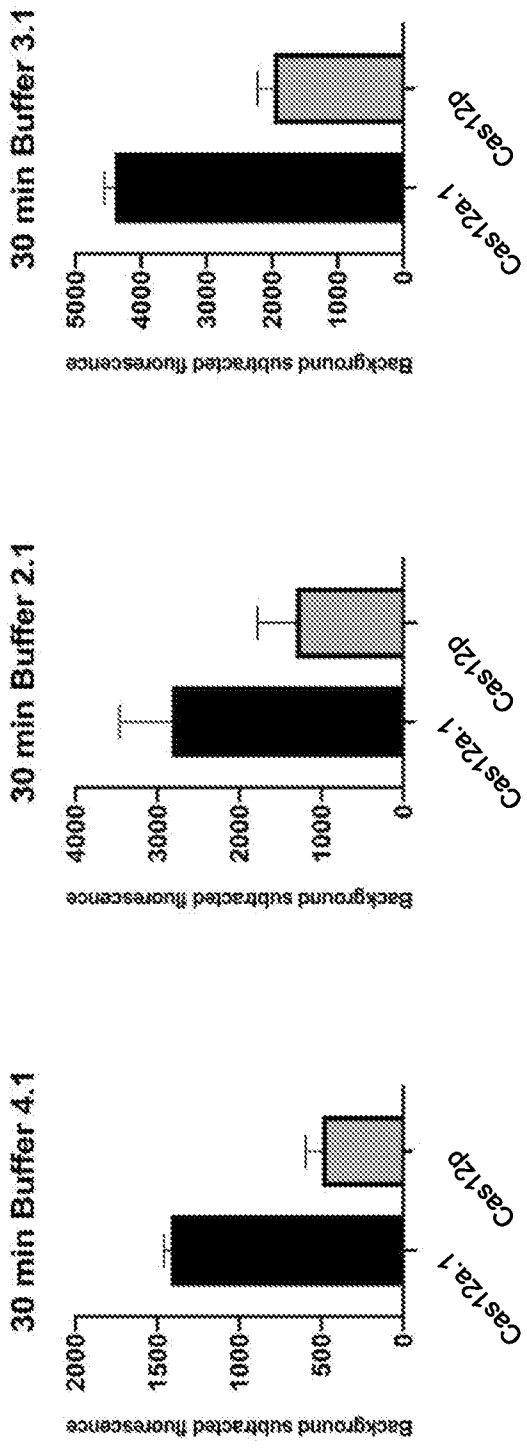

30 min Buffer 3.1
Background subtracted fluorescence
5000
4000
3000
2000
1000
0
Cas12a.1
Cas12p
30 min Buffer 2.1
Background subtracted fluorescence
4000
3000
2000
1000
0
Cas12a.1
Cas12p
30 min Buffer 4.1
Background subtracted fluorescence
2000
1500
1000
500
0
Cas12a.1
Cas12p

FIG. 12

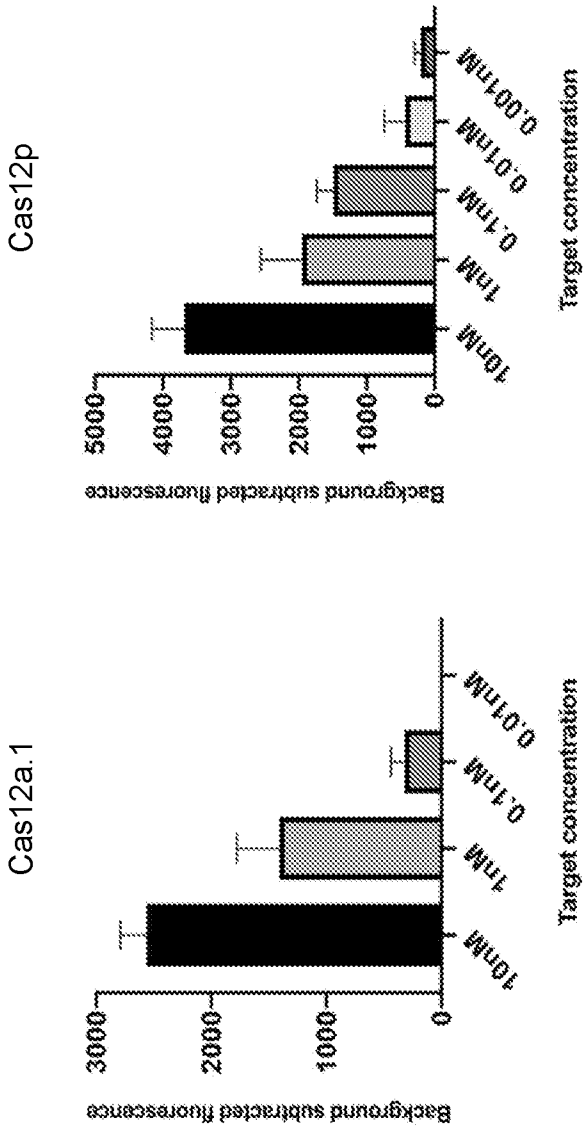

Cas12p
Background subtracted fluorescence
5000
4000
3000
2000
1000
0
10nM
1nM
0.1nM
0.01nM
0.001nM
Target concentration
Cas12a.1
Background subtracted fluorescence
3000
2000
1000
0
10nM
1nM
0.1nM
0.01nM
Target concentration

FIG. 13

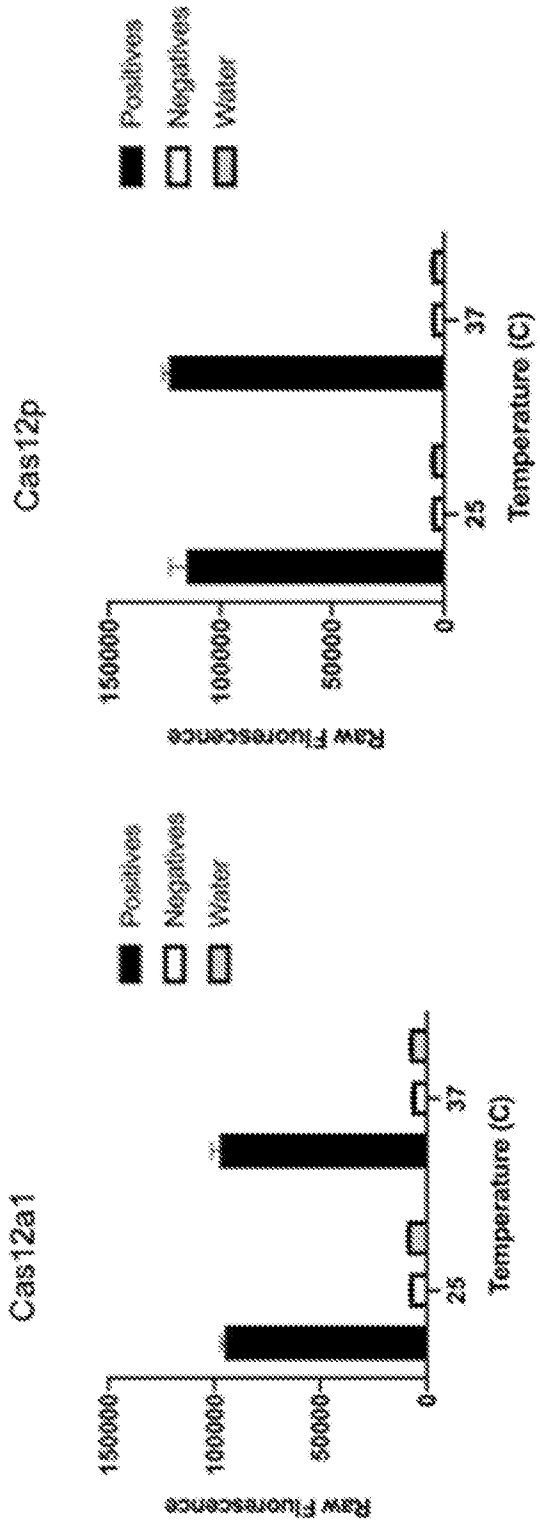
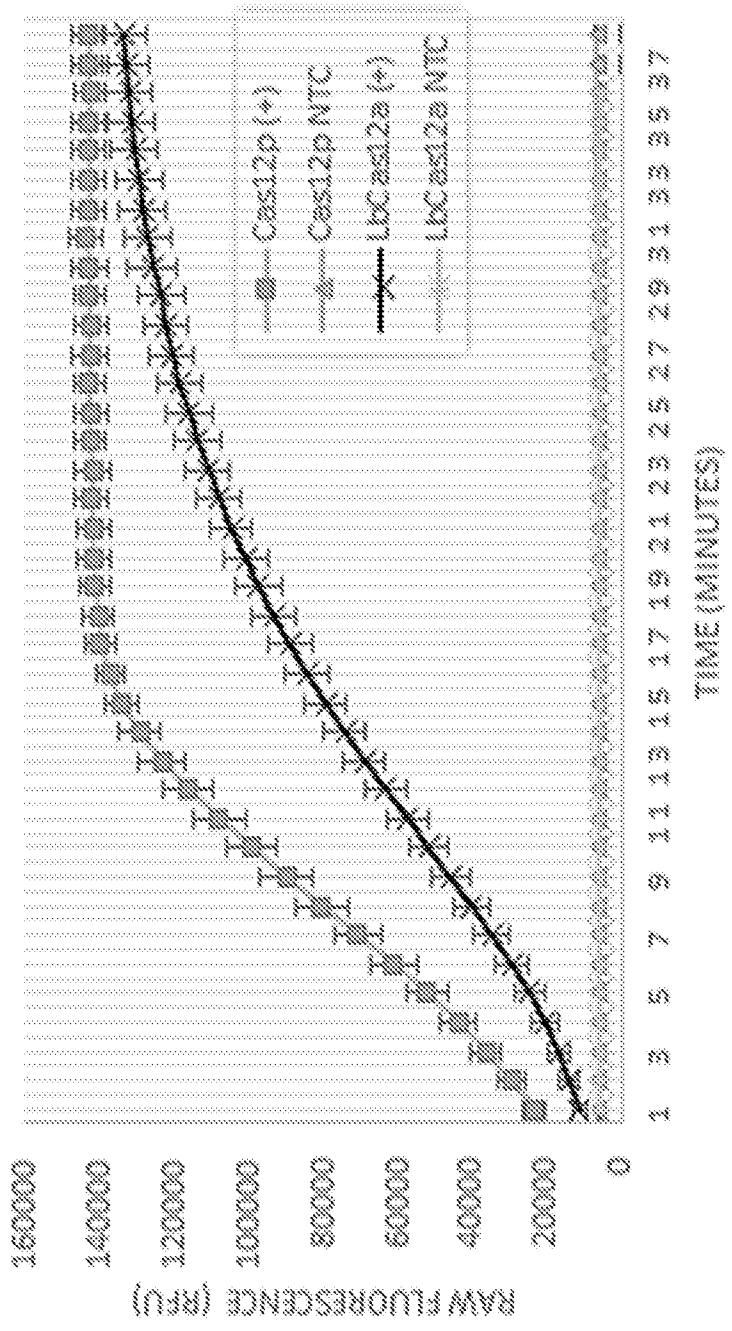

Cas12a1
Positives
Negatives
Water
Raw Fluorescence
150000
100000
50000
0
25
37
Temperature (C)
Cas12p
Positives
Negatives
Water
Raw Fluorescence
150000
100000
50000
0
25
37
Temperature (C)

FIG. 14

FIG. 15


RAW FLUORESCENCE (RFU)
160000
140000
120000
100000
80000
60000
40000
20000
0
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37
TIME (MINUTES)
Cas12p (+)
Cas12p NTC
LbCas12a (+)
LbCas12a NTC

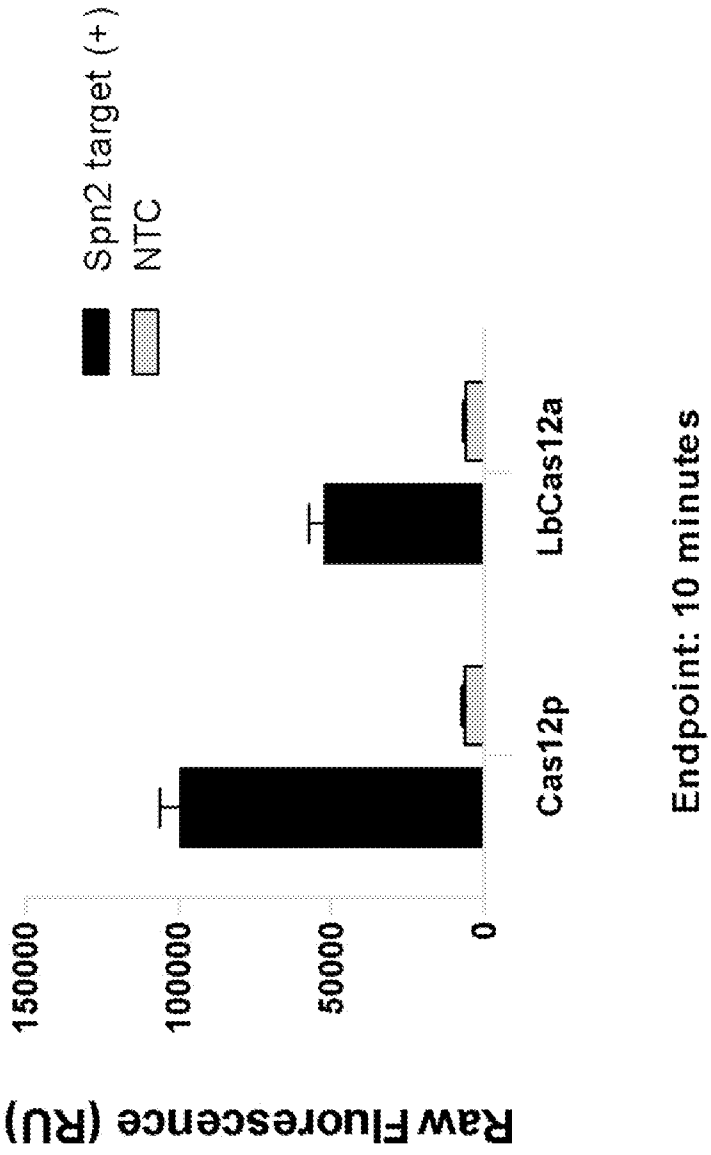

Spn2 target (+)
NTC
150000
100000
50000
0
Raw Fluorescence (RU)
Cas12p
LbCas12a
Endpoint: 10 minutes

FIG. 16

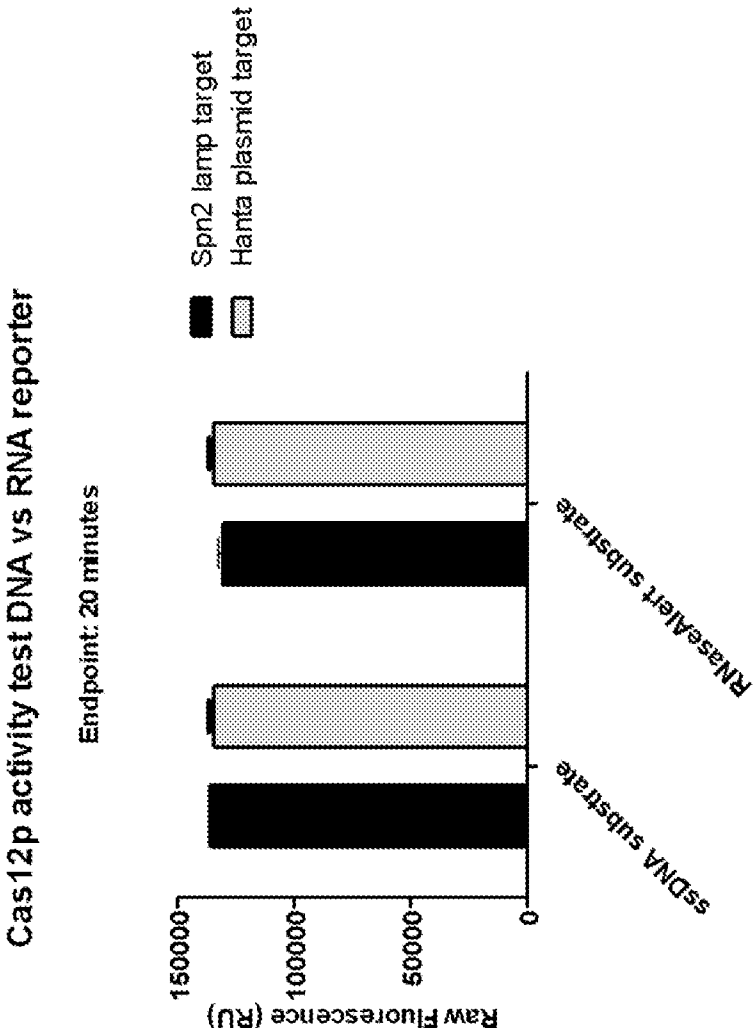

Cas12p activity test DNA vs RNA reporter
Endpoint: 20 minutes
Spn2 lamp target
Hanta plasmid target
Raw Fluorescence (RU)
150000
100000
50000
0
ssDNA substrate
RNaseAlert substrate

FIG. 17

FIG. 18

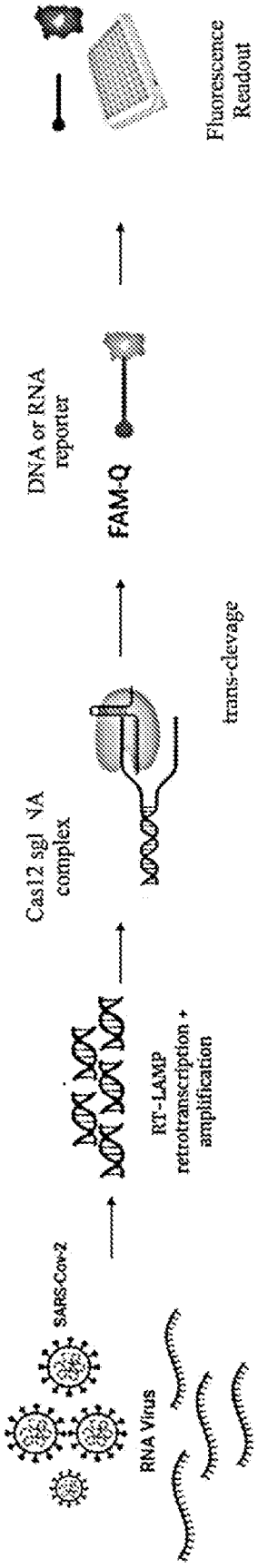
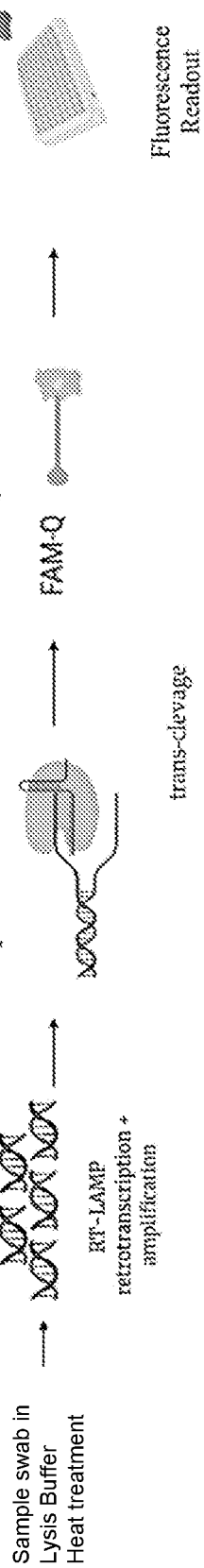

RNA Virus
SARS-Cov-2
RT-LAMP
retrotranscription +
amplification
Cas12 sgl VA
complex
trans-clevage
FAM-Q
DNA or RNA
reporter
Fluorescence
Readout

FIG. 19


Sample swab in
Lysis Buffer
Heat treatment
RT-LAMP
retrotranscription +
amplification
Cas12 sgRNA
complex
trans-clevage
DNA or RNA
reporter
FAM-Q
Fluorescence
Readout

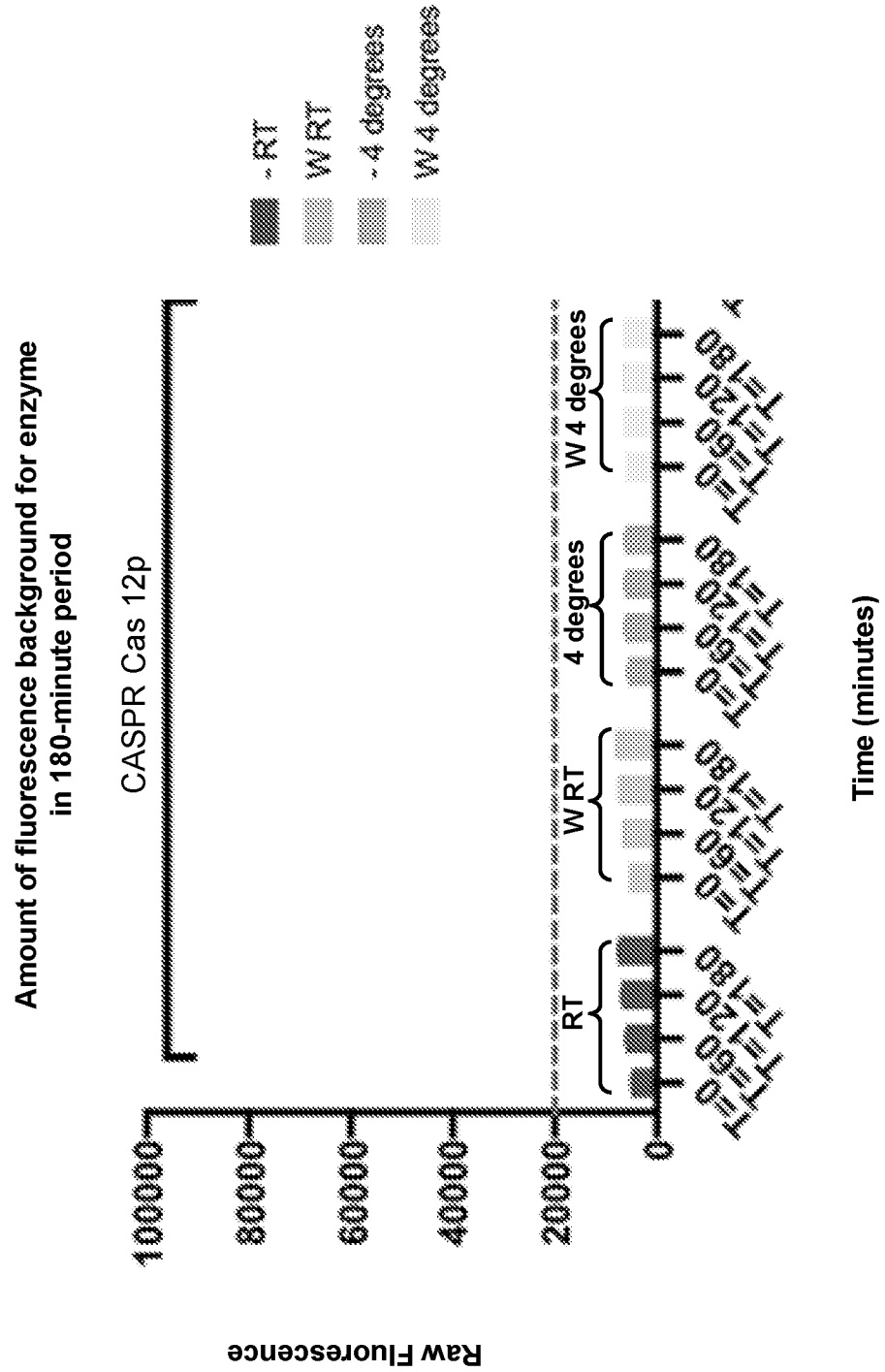

Amount of fluorescence background for enzyme in 180-minute period
CASPR Cas 12p
- RT
W RT
- 4 degrees
W 4 degrees
Raw Fluorescence
100000
80000
60000
40000
20000
0
RT
W RT
4 degrees
W 4 degrees
T=0
T=60
T=120
T=180
Time (minutes)

FIG. 20

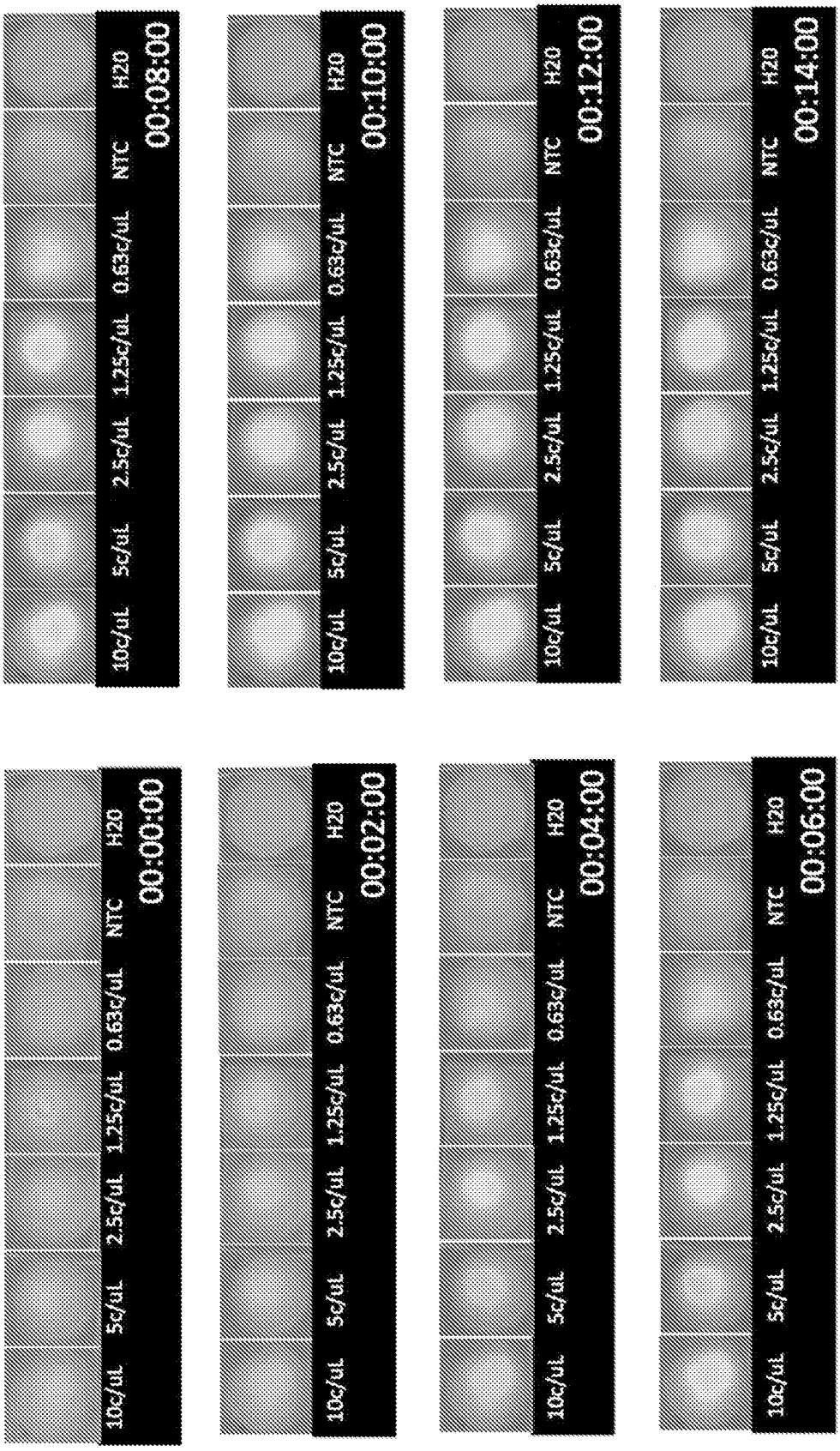
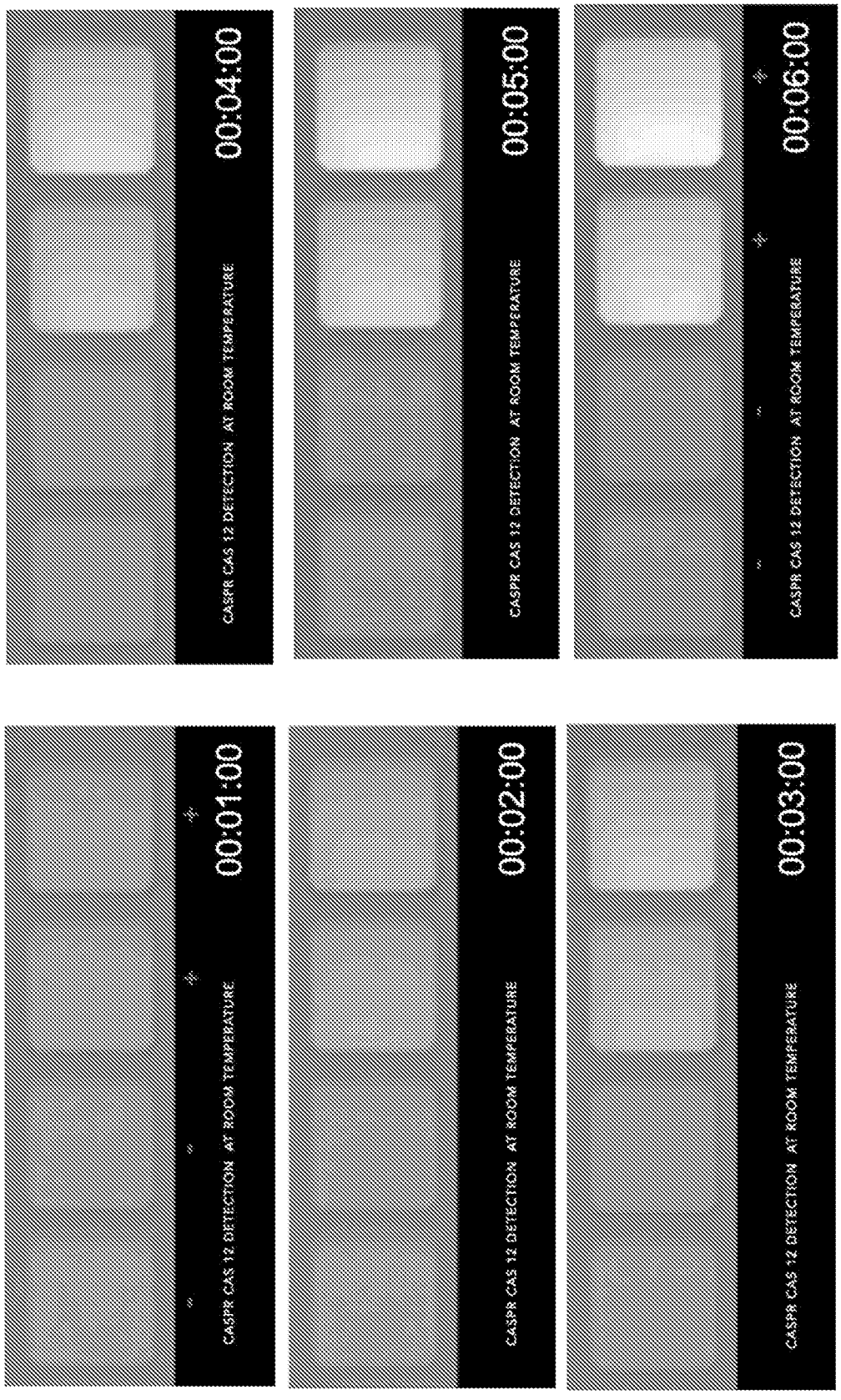

10c/uL 5c/uL 2.5c/uL 1.25c/uL 0.63c/uL NTC H20
00:00:00
10c/uL 5c/uL 2.5c/uL 1.25c/uL 0.63c/uL NTC H20
00:02:00
10c/uL 5c/uL 2.5c/uL 1.25c/uL 0.63c/uL NTC H20
00:04:00
10c/uL 5c/uL 2.5c/uL 1.25c/uL 0.63c/uL NTC H20
00:06:00
10c/uL 5c/uL 2.5c/uL 1.25c/uL 0.63c/uL NTC H20
00:08:00
10c/uL 5c/uL 2.5c/uL 1.25c/uL 0.63c/uL NTC H20
00:10:00
10c/uL 5c/uL 2.5c/uL 1.25c/uL 0.63c/uL NTC H20
00:12:00
10c/uL 5c/uL 2.5c/uL 1.25c/uL 0.63c/uL NTC H20
00:14:00

FIG. 21

FIG. 22


CASPR CAS 12 DETECTION AT ROOM TEMPERATURE
00:01:00
CASPR CAS 12 DETECTION AT ROOM TEMPERATURE
00:02:00
CASPR CAS 12 DETECTION AT ROOM TEMPERATURE
00:03:00
CASPR CAS 12 DETECTION AT ROOM TEMPERATURE
00:04:00
CASPR CAS 12 DETECTION AT ROOM TEMPERATURE
00:05:00
CASPR CAS 12 DETECTION AT ROOM TEMPERATURE
00:06:00

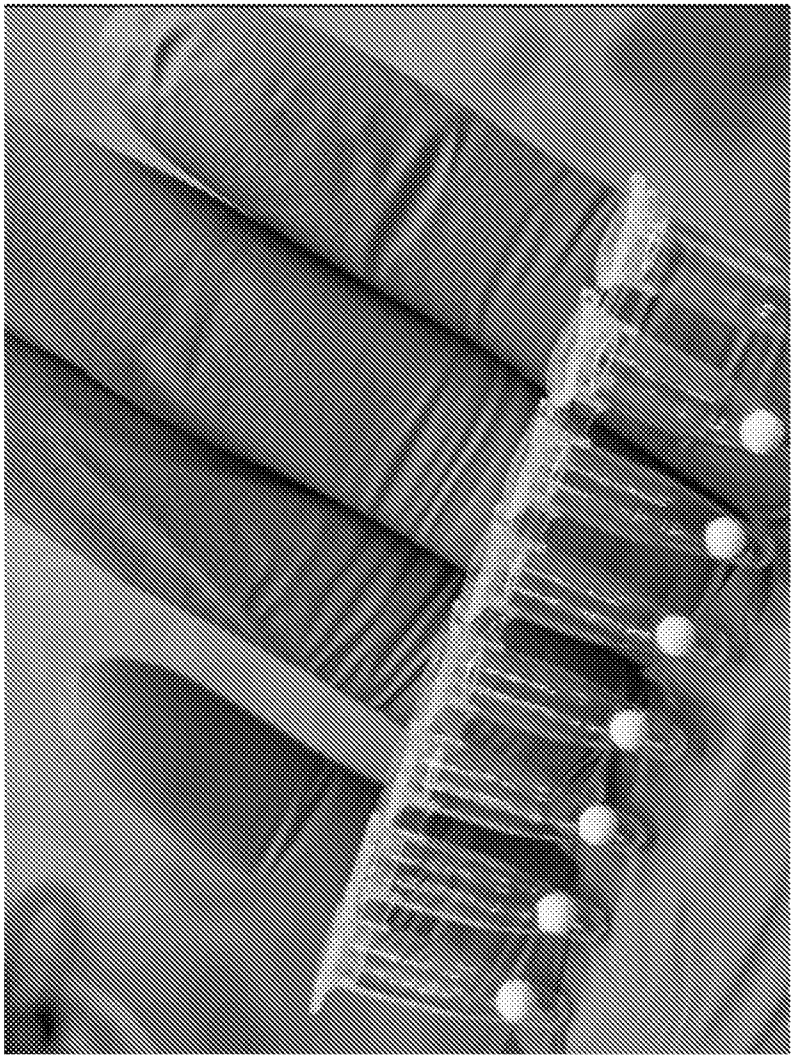


FIG. 23

FIG. 24

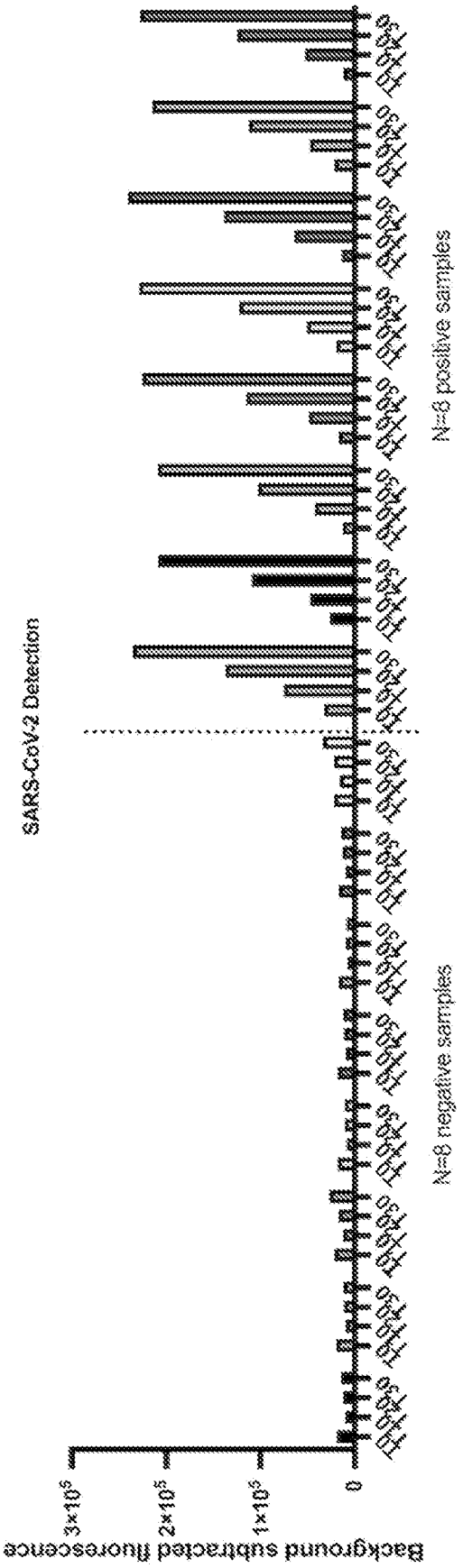

SARS-CoV-2 Detection
Background subtracted fluorescence
3×10^5
2×10^5
1×10^5
0
N=8 negative samples
N=8 positive samples

FIG. 25

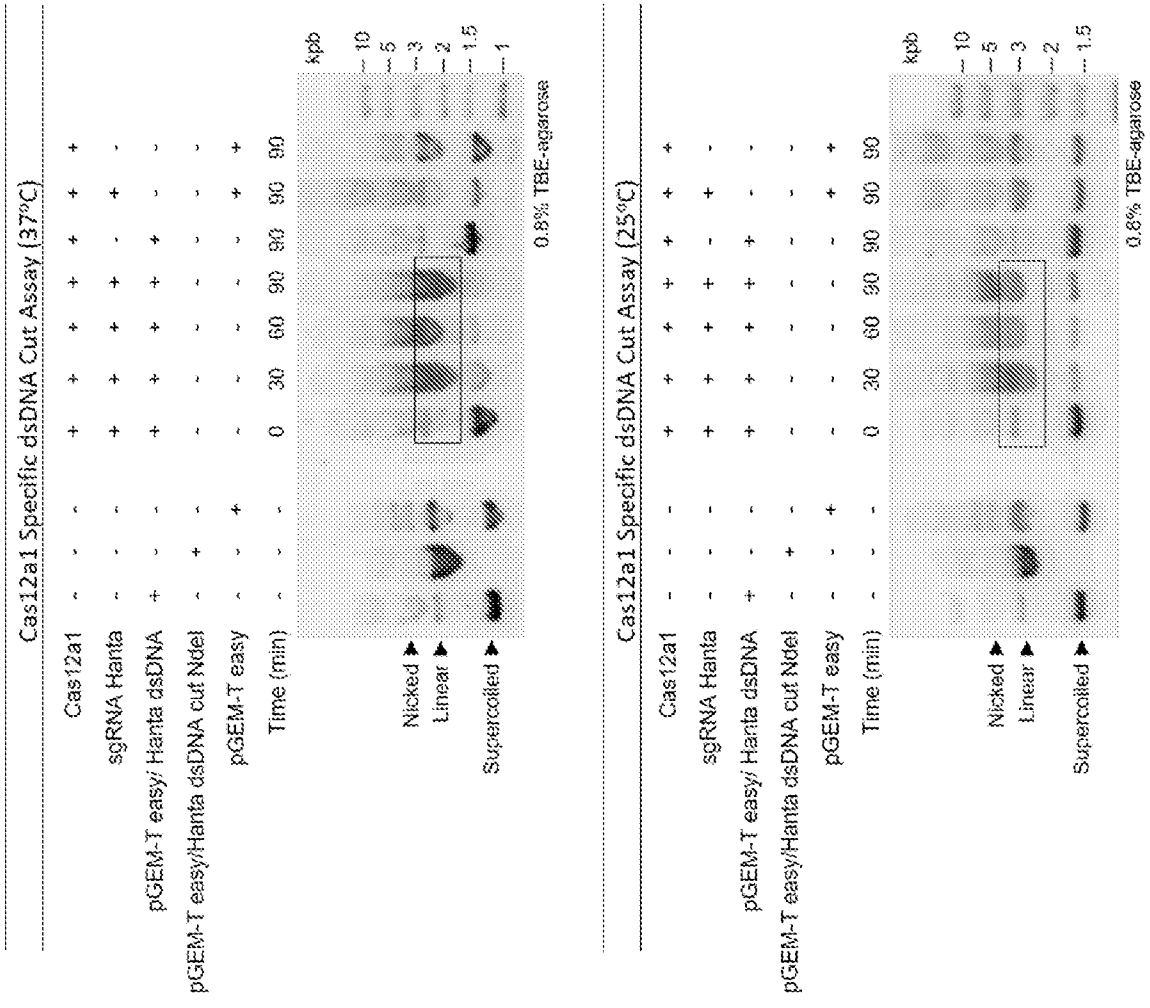

Cas12a1 Specific dsDNA Cut Assay (37°C)
Cas12a1
sgRNA Hanta
pGEM-T easy/ Hanta dsDNA
pGEM-T easy/Hanta dsDNA cut NdeI
pGEM-T easy
Time (min)
0 30 60 90 90 90 90
Nicked
Linear
Supercoiled
kpb
10
5
3
2
1.5
1
0.8% TBE-agarose
Cas12a1 Specific dsDNA Cut Assay (25°C)
Cas12a1
sgRNA Hanta
pGEM-T easy/ Hanta dsDNA
pGEM-T easy/Hanta dsDNA cut NdeI
pGEM-T easy
Time (min)
0 30 60 90 90 90 90
Nicked
Linear
Supercoiled
kpb
10
5
3
2
1.5
0.8% TBE-agarose

FIG. 25 continued

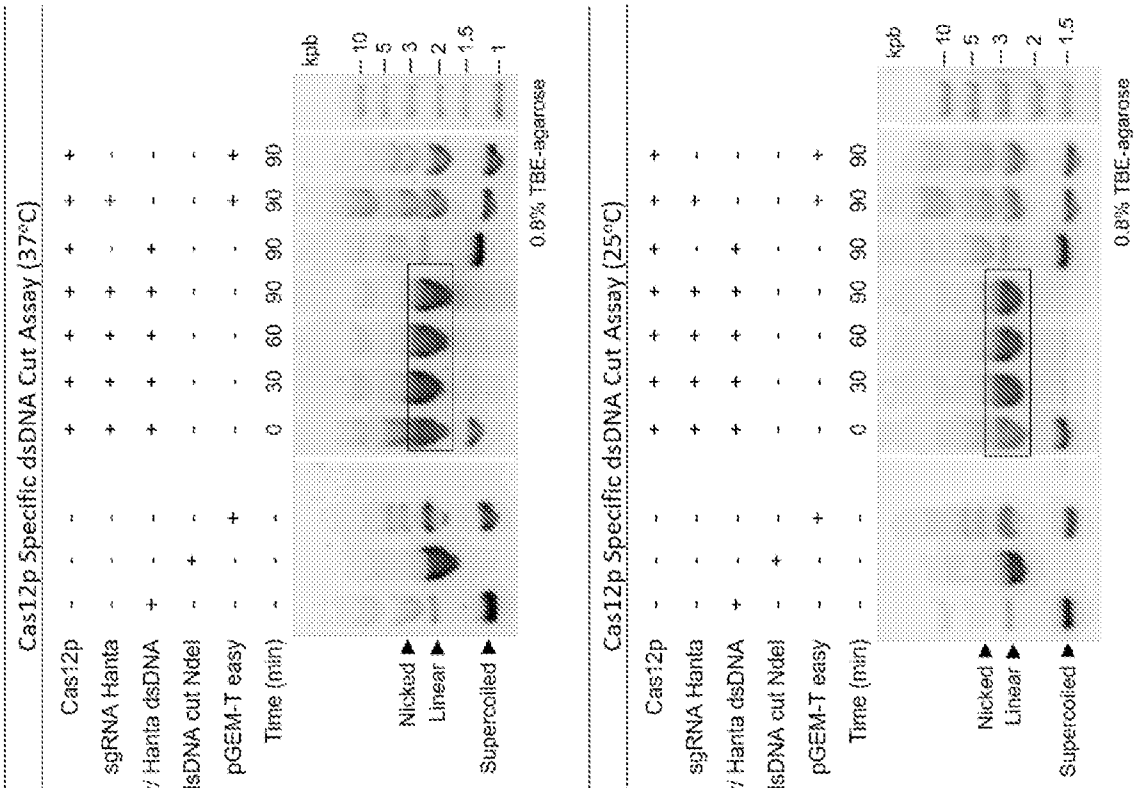


FIG. 26

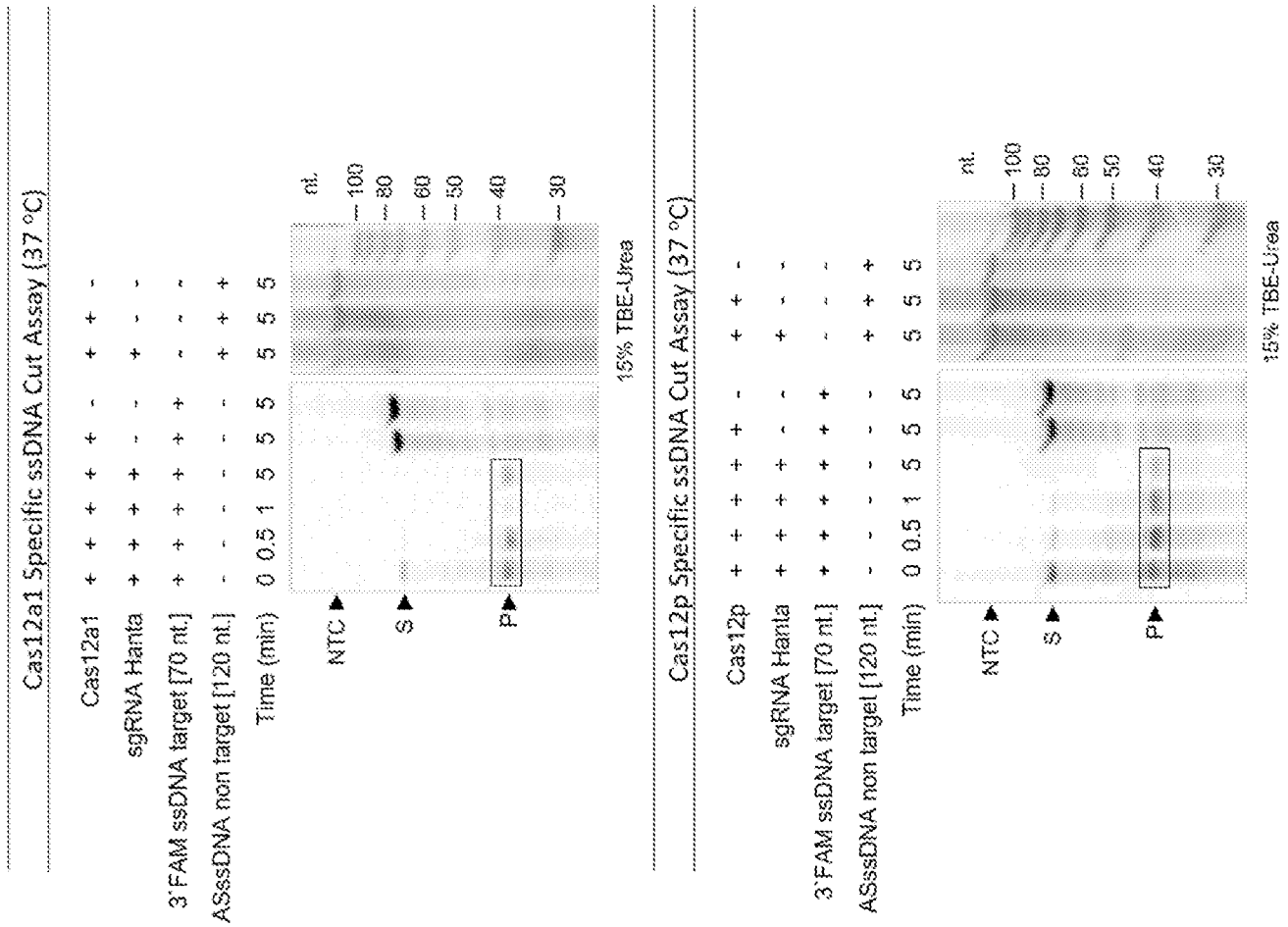

Cas12a1 Specific ssDNA Cut Assay (37 °C)
Cas12a1
sgRNA Hanta
3'FAM ssDNA target [70 nt.]
ASssDNA non target [120 nt.]
Time (min)
0 0.5 1 5 5 5 5 5 5
NTC
S
P
nt.
100
80
60
50
40
30
15% TBE-Urea
Cas12p Specific ssDNA Cut Assay (37 °C)
Cas12p
sgRNA Hanta
3'FAM ssDNA target [70 nt.]
ASssDNA non target [120 nt.]
Time (min)
0 0.5 1 5 5 5 5 5 5
NTC
S
P
nt.
100
80
60
50
40
30
15% TBE-Urea

FIG. 27

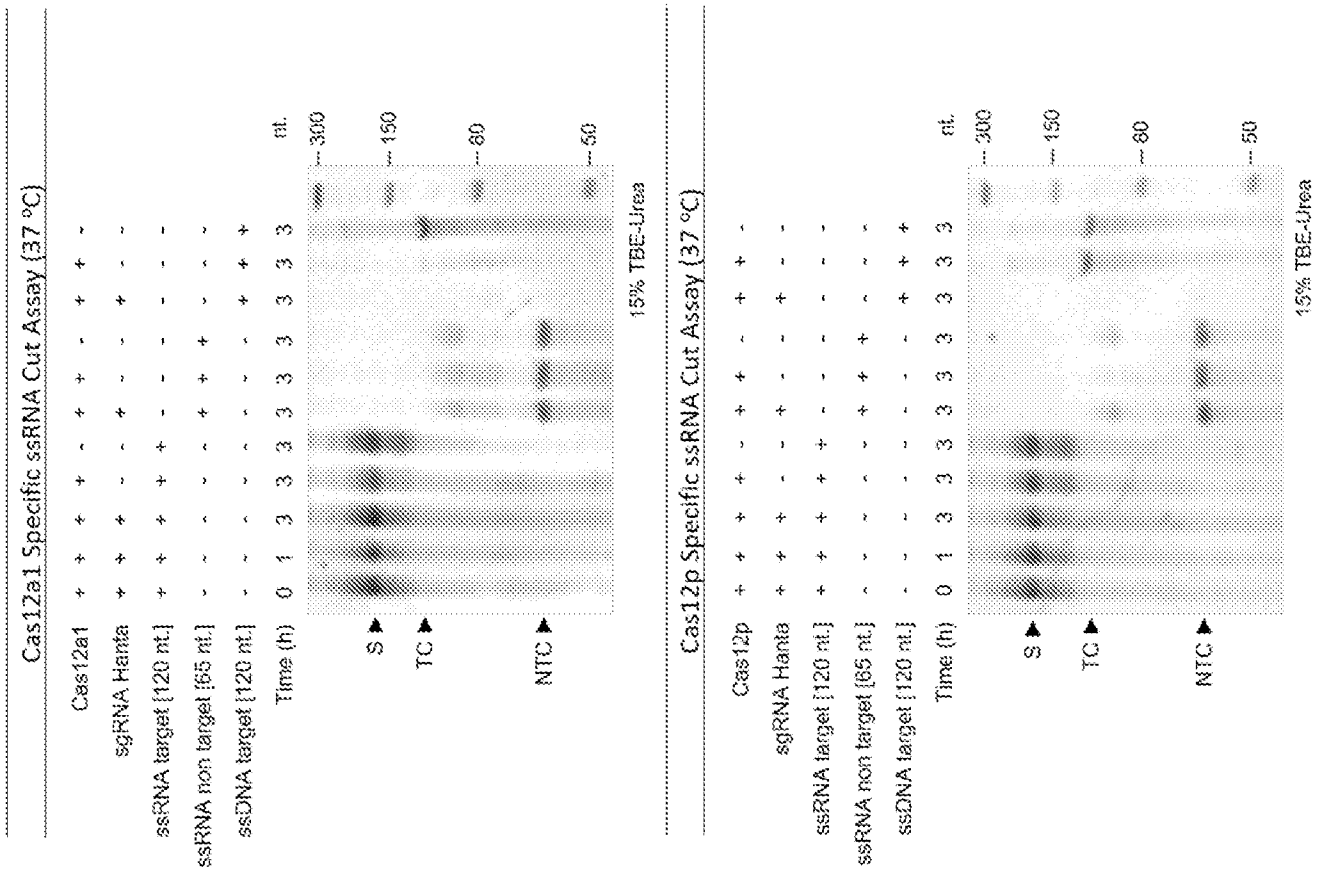

Cas12a1 Specific ssRNA Cut Assay (37 °C)
Cas12a1
sgRNA Hanta
ssRNA target (120 nt.)
ssRNA non target (65 nt.)
ssDNA target (120 nt.)
Time (h)
nt.
300
150
60
50
S
TC
NTC
15% TBE-Urea
Cas12p Specific ssRNA Cut Assay (37 °C)
Cas12p
sgRNA Hanta
ssRNA target (120 nt.)
ssRNA non target (65 nt.)
ssDNA target (120 nt.)
Time (h)
nt.
300
150
60
50
S
TC
NTC
15% TBE-Urea

FIG. 28

Cas12p DNA collateral hydrolysis products
T1

Observed products	Calculated m/z	Observed m/z	Relative Intensity
(Protection)TT	6038.65	6038.60	100.00

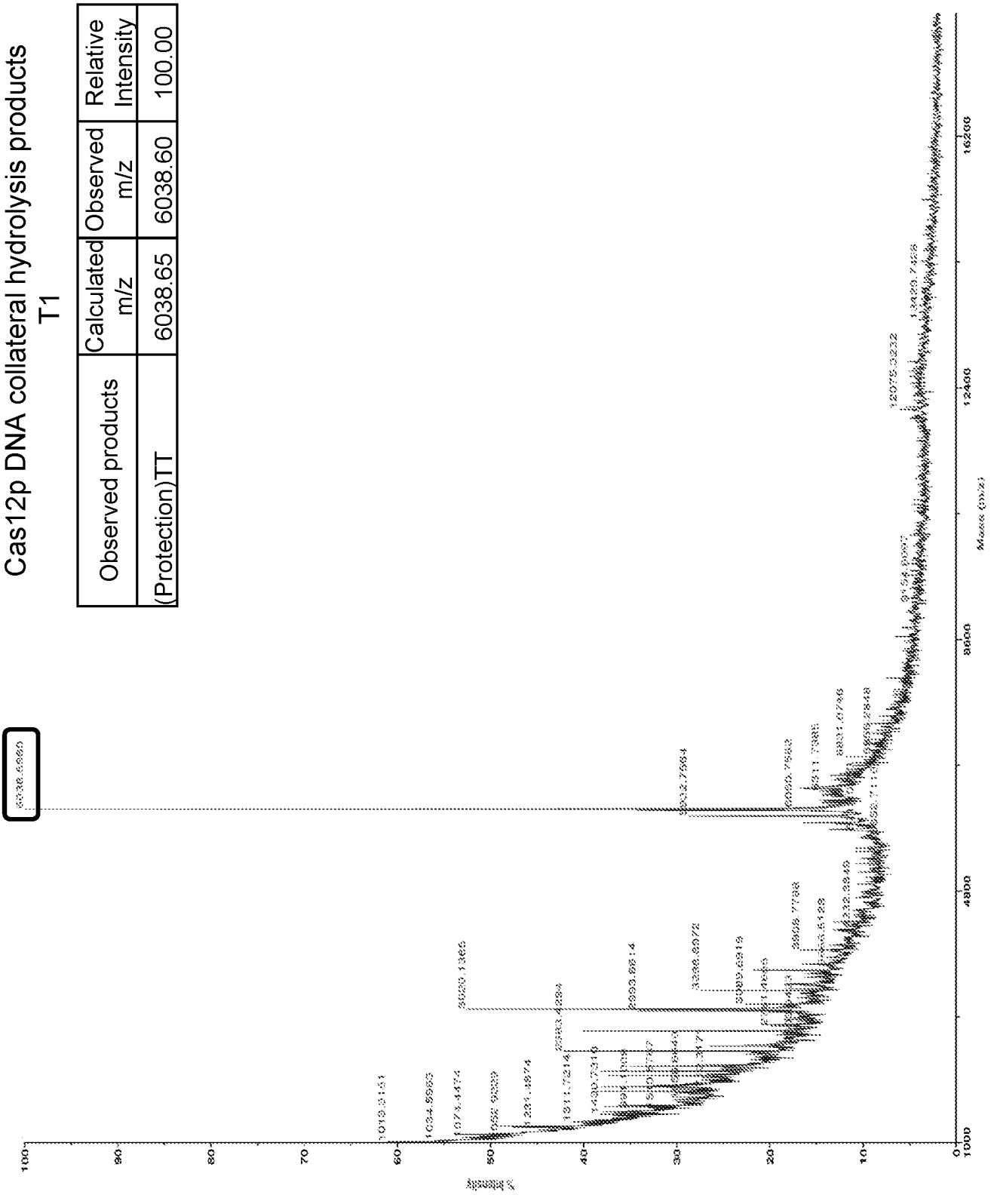


FIG. 29

Cas12p DNA collateral hydrolysis products
T0

Observed products	Calculated m/z	Observed m/z	Relative Intensity
(Protection)TTATT	6960.24	6960.29	100.00

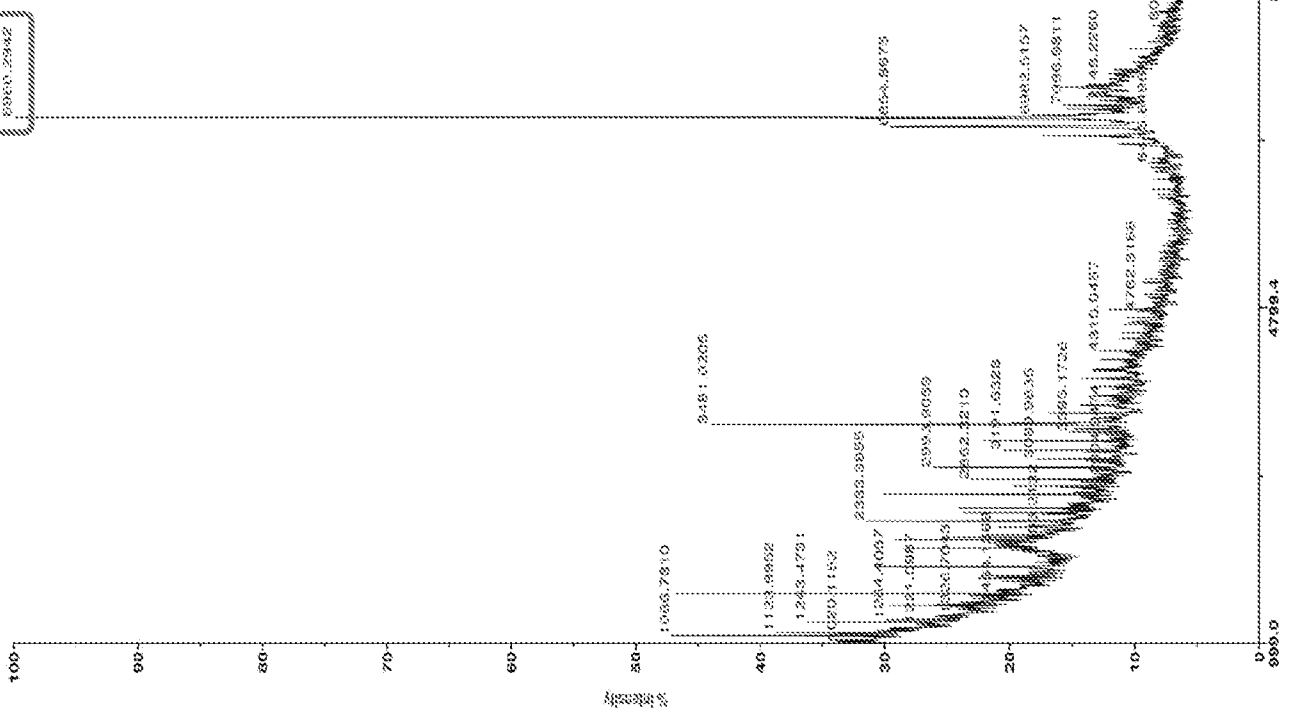


FIG. 30

Cas12p RNA collateral hydrolysis products
T1

Observed products	Calculated m/z	Observed m/z	Relative Intensity
(Protection)rU	6024.43	6025.89	66.63
(Protection)rUrU	6330.59	6331.96	100.00
(Protection)rUrU_3'Pho	6392.55	6993.54	16.03
(Protection)rUrUrA	6659.80	6660.73	38.03
(Protection)rUrUrArU	7272.12	7273.25	69.09

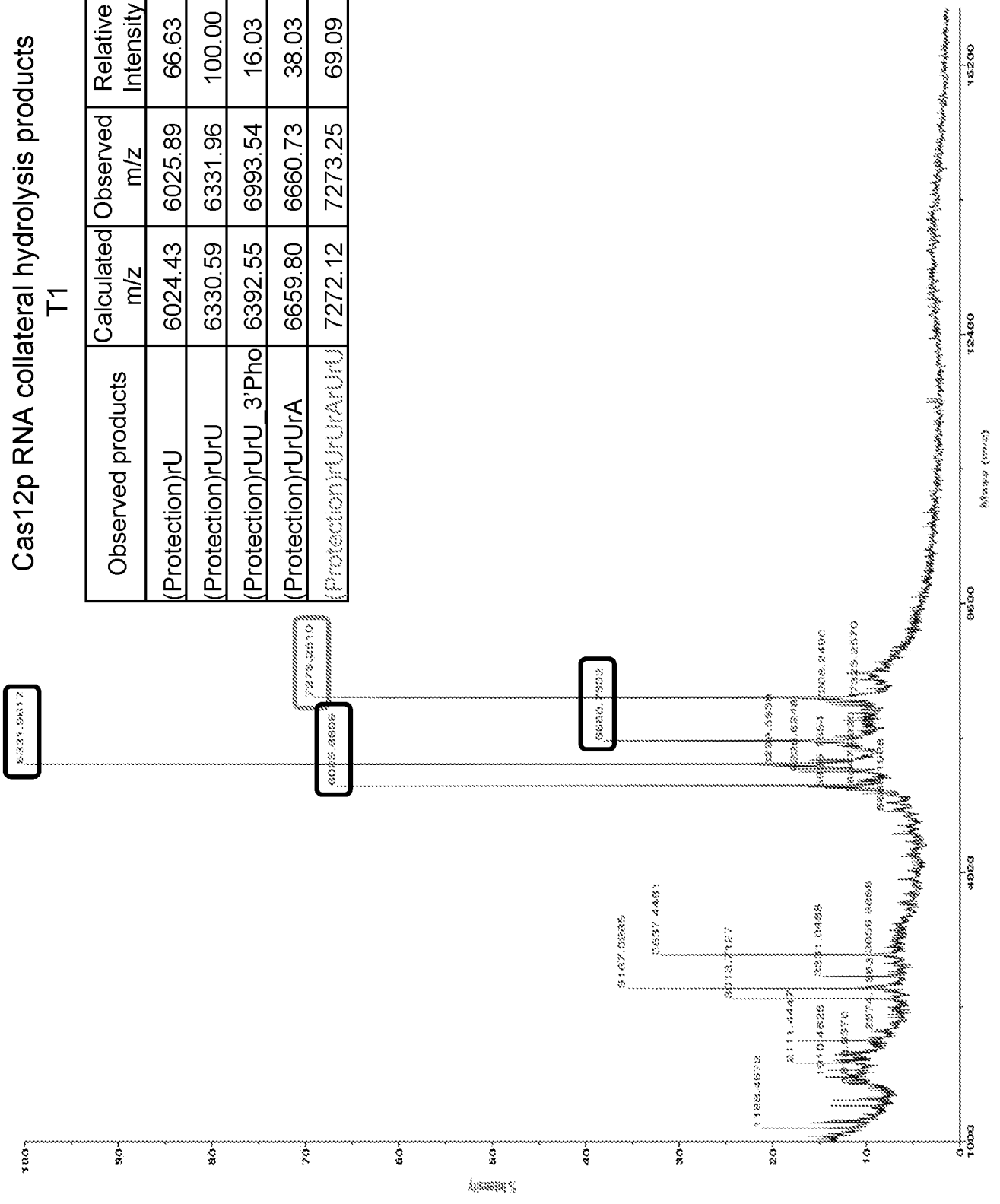
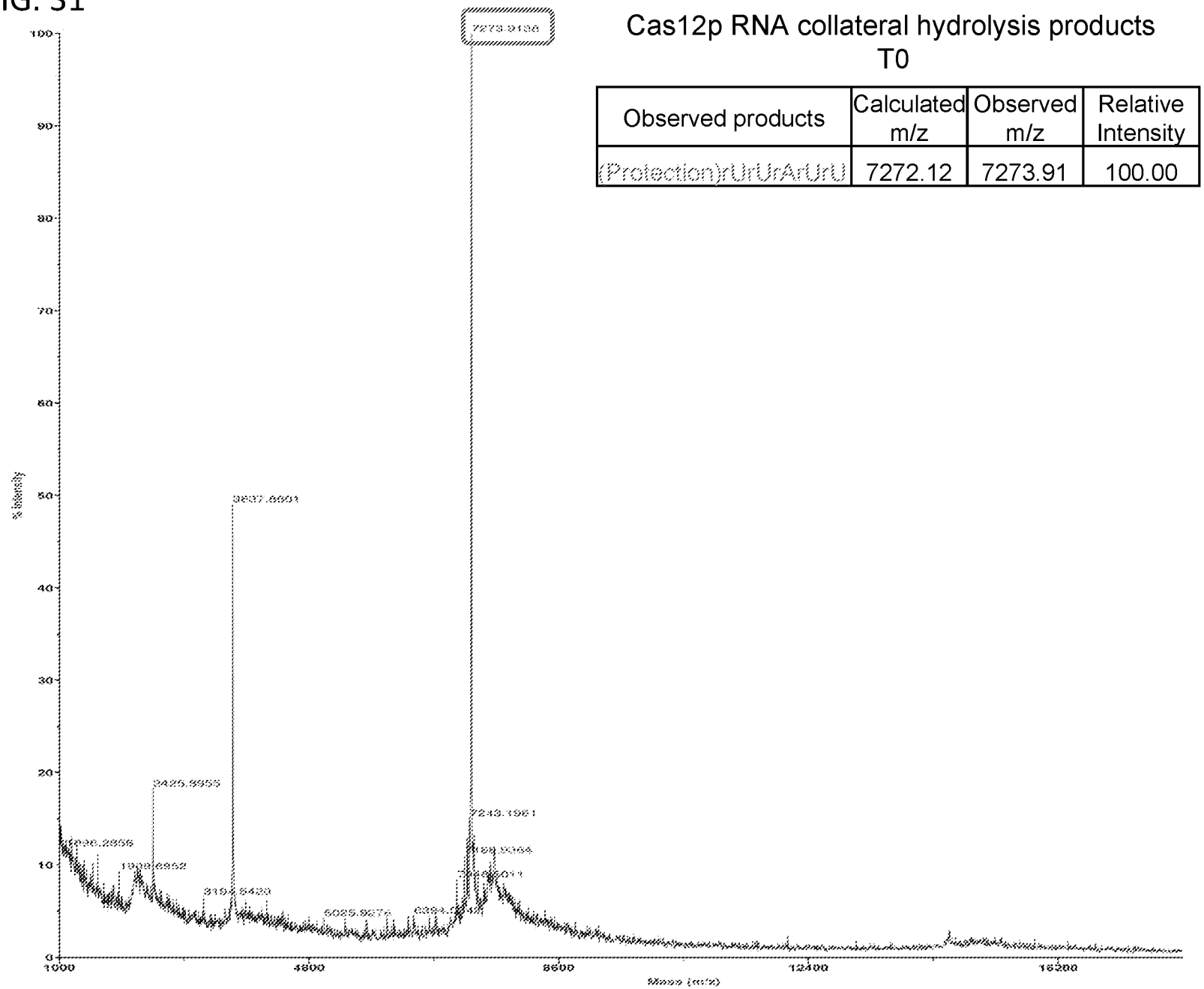


FIG. 31



Cas12p RNA collateral hydrolysis products
T0

Observed products	Calculated m/z	Observed m/z	Relative Intensity
(Protection)rUrUrArUrU	7272.12	7273.91	100.00

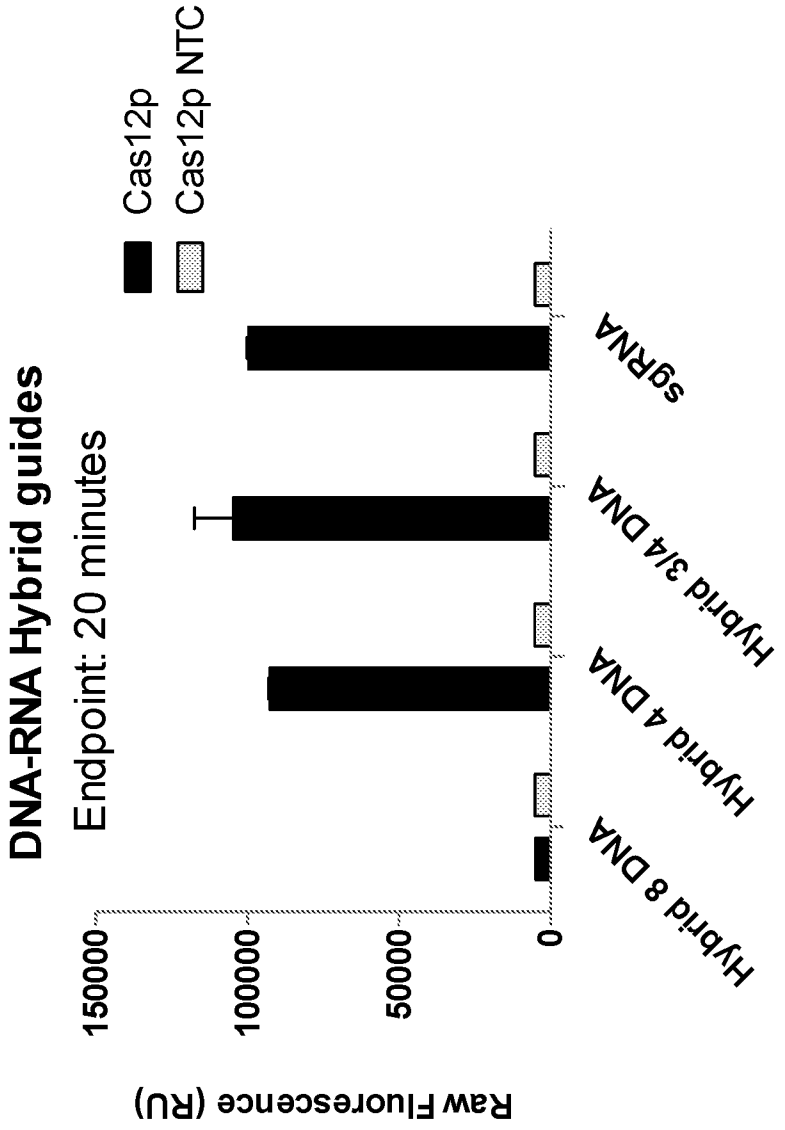

DNA-RNA Hybrid guides
Endpoint: 20 minutes
Cas12p
Cas12p NTC
Raw Fluorescence (RU)
150000
100000
50000
0
Hybrid 8 DNA
Hybrid 4 DNA
Hybrid 3/4 DNA
sgRNA

FIG. 32

FIG. 33

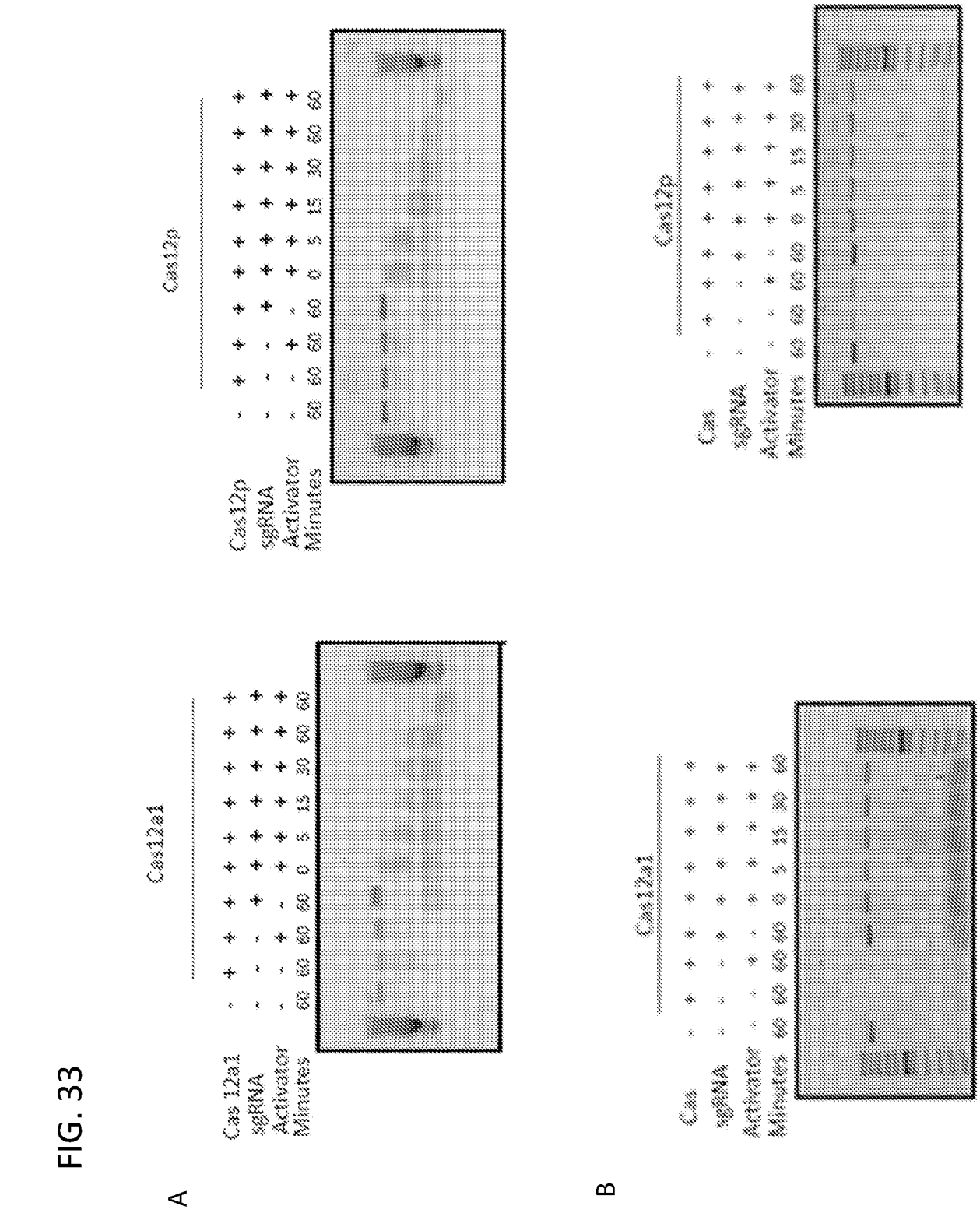
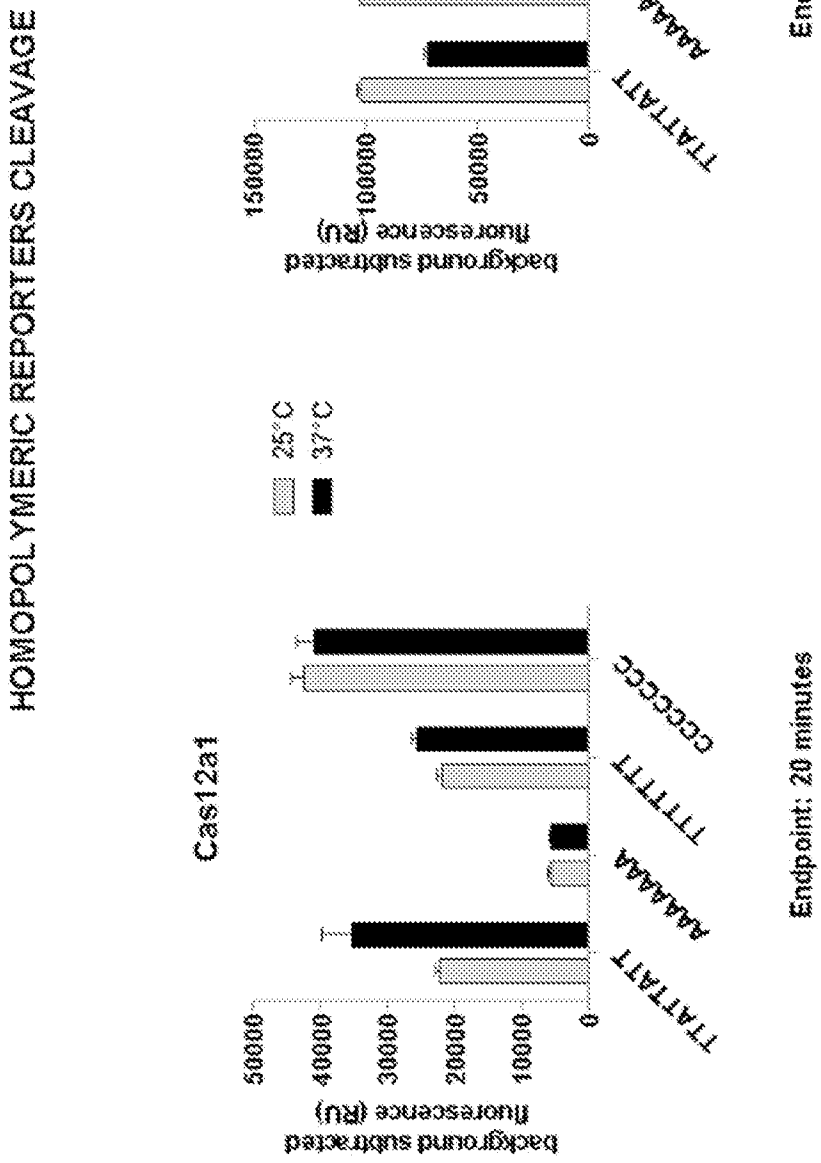

A
Cas12a1
Cas 12a1
sgRNA
Activator
Minutes
Cas12p
Cas12p
sgRNA
Activator
Minutes
B
Cas12a1
Cas
sgRNA
Activator
Minutes
Cas12p
Cas
sgRNA
Activator
Minutes

FIG. 34


HOMOPOLYMERIC REPORTERS CLEAVAGE
Cas12a1
25°C
37°C
background subtracted fluorescence (RU)
50000
40000
30000
20000
10000
0
TTATTATT
AAAAAA
TTTTTTT
CCCCCCC
Endpoint: 20 minutes
Cas12p
25°C
37°C
background subtracted fluorescence (RU)
150000
100000
50000
0
TTATTATT
AAAAAA
TTTTTTT
CCCCCCC
Endpoint: 20 minutes

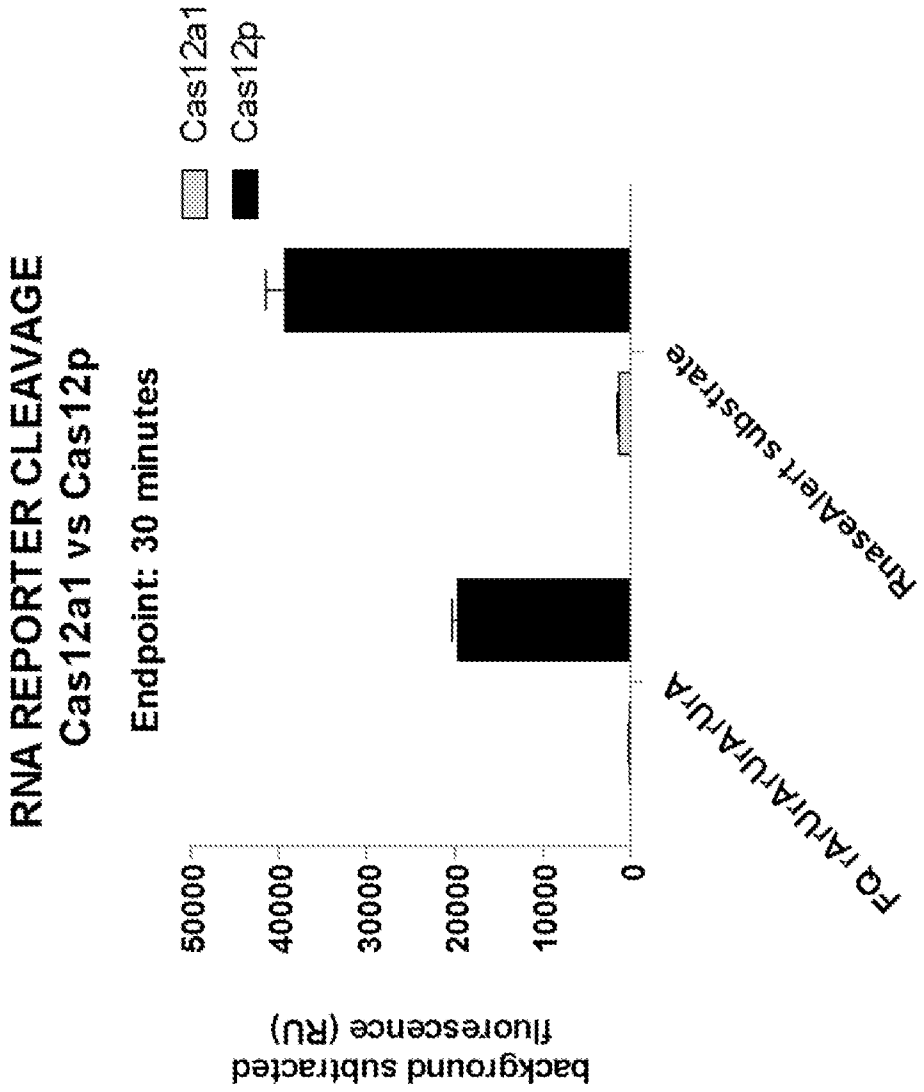

RNA REPORTER CLEAVAGE
Cas12a1 vs Cas12p
Endpoint: 30 minutes
Cas12a1
Cas12p
background subtracted fluorescence (RU)
50000
40000
30000
20000
10000
0
FQ TAMURAMURA
RnaseAlert substrate

FIG. 35

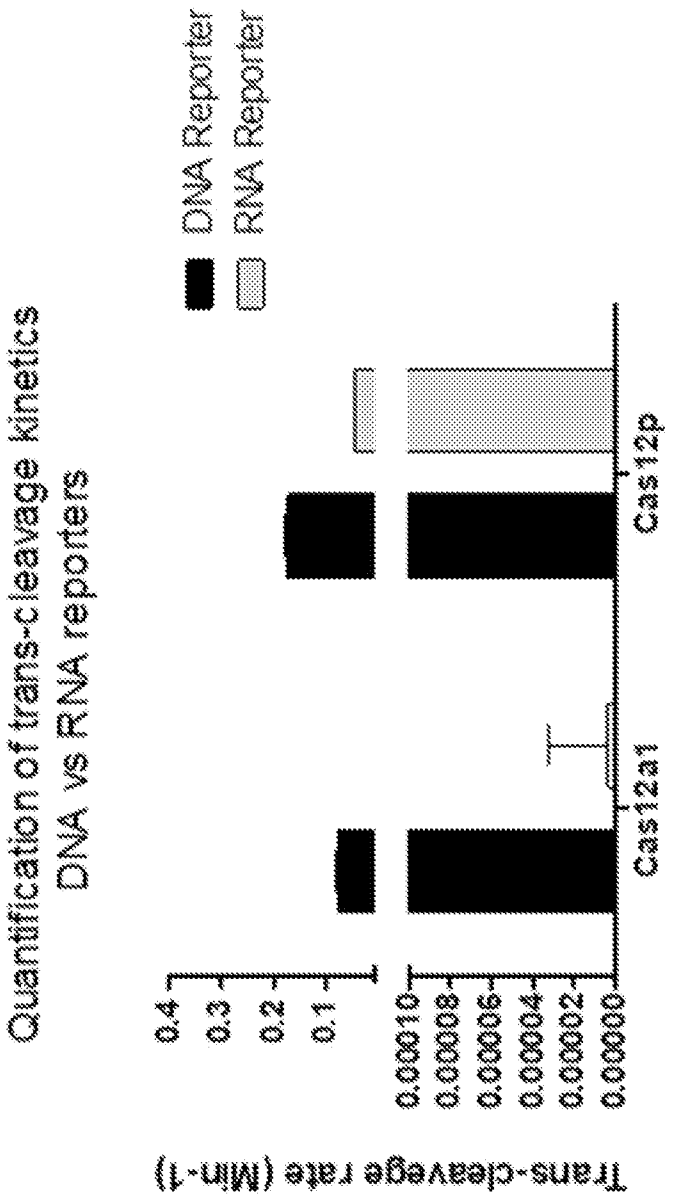

Quantification of trans-cleavage kinetics
DNA vs RNA reporters
DNA Reporter
RNA Reporter
Trans-cleavege rate (Min-1)
0.4
0.3
0.2
0.1
0.00010
0.00008
0.00006
0.00004
0.00002
0.00000
Cas12a1
Cas12p

FIG. 36

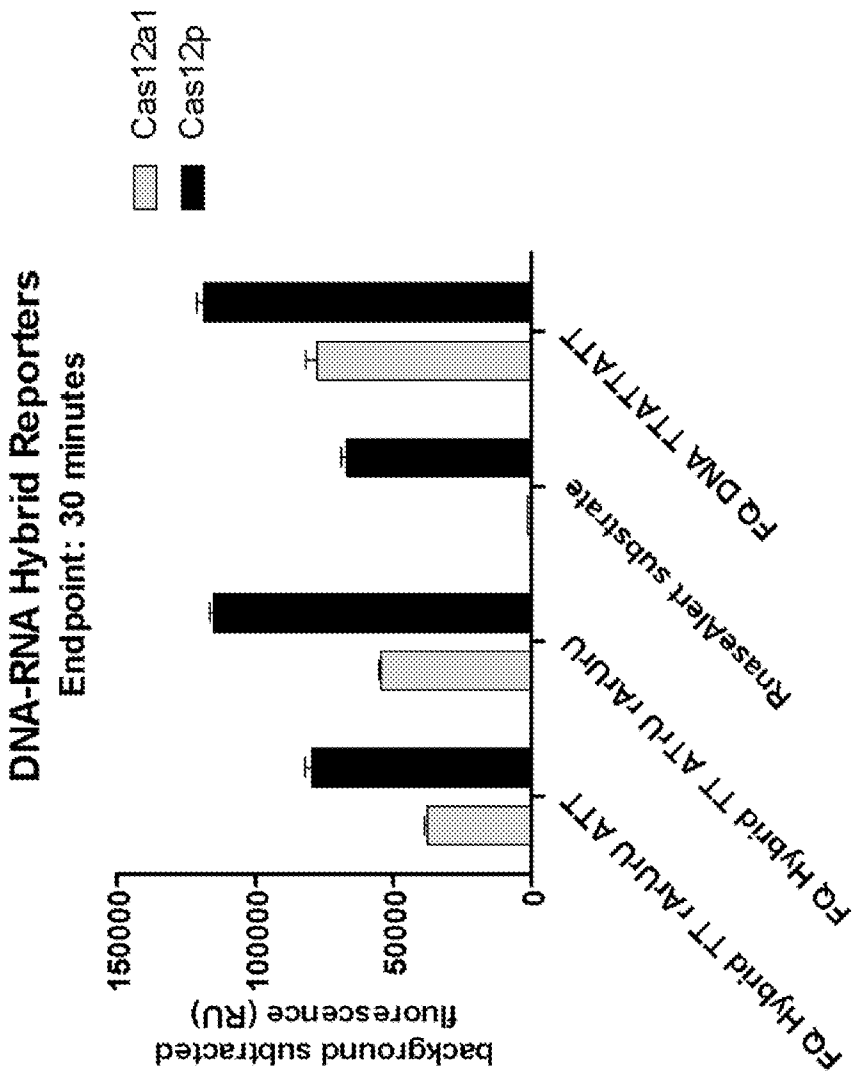

DNA-RNA Hybrid Reporters
Endpoint: 30 minutes
Cas12a1
Cas12p
background subtracted fluorescence (RU)
150000
100000
50000
0
FQ Hybrid TT rArUrU ATT
FQ Hybrid TT ArUrU rArUrU
RNaseAlert substrate
FQ DNA TTATTATT

FIG. 37

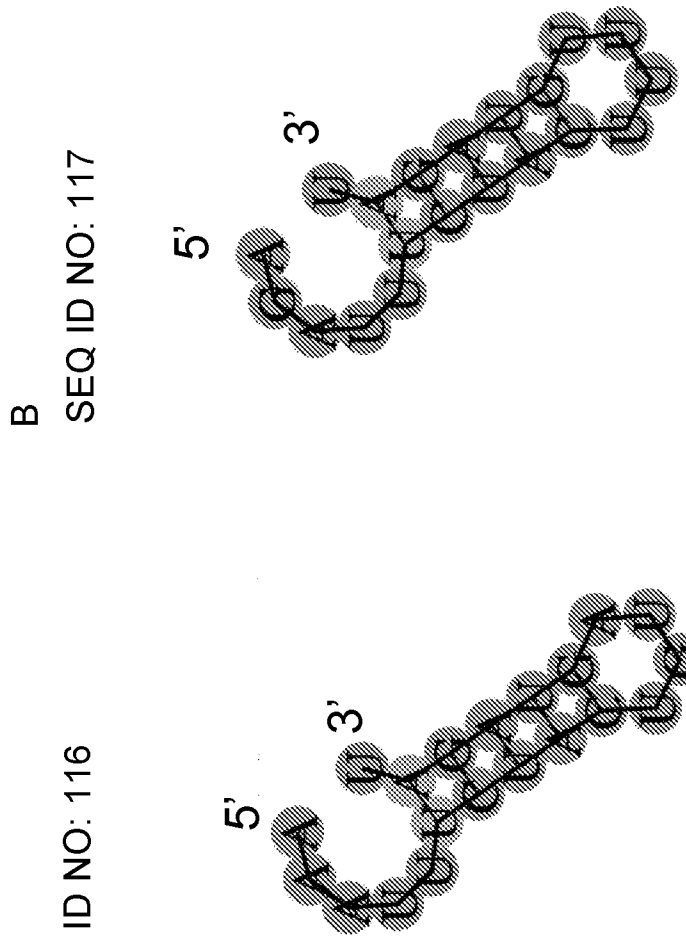


FIG. 38

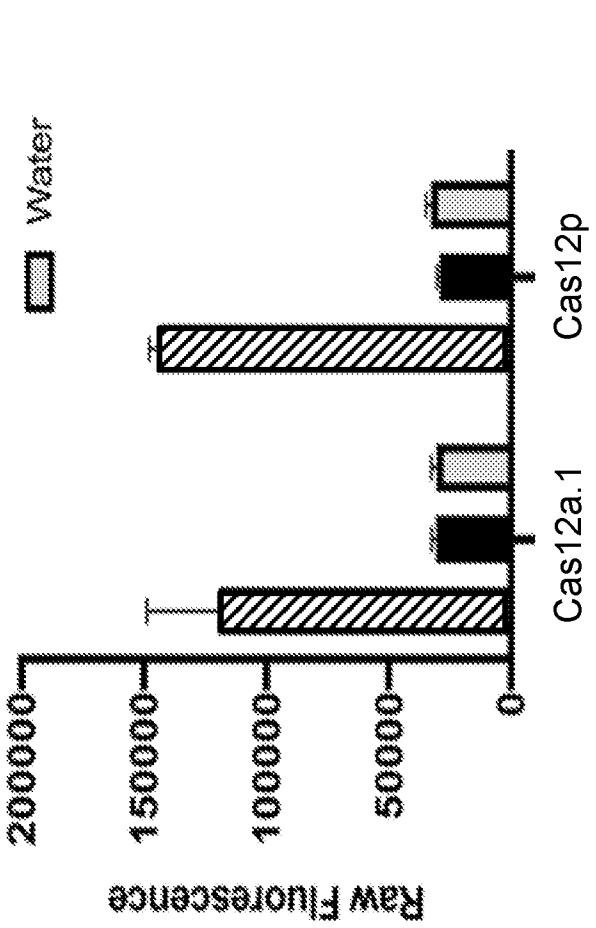

50 copies/ul
NTC
Water
Raw Fluorescence
200000
150000
100000
50000
0
Cas12a.1
Cas12p

FIG. 39

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2020/050237

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12N9/22 C12N15/113 C12Q1/682 C12Q1/70
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12N C12Q

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)
EPO-Internal

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2019/002889 A1 (CHENG DAVID R [US] ET AL) 3 January 2019 (2019-01-03) example 1	1
A	KOONIN EUGENE V ET AL: "Diversity, classification and evolution of CRISPR-Cas systems", CURRENT OPINION IN MICROBIOLOGY, vol. 37, 9 June 2017 (2017-06-09), pages 67-78, XP085276922, ISSN: 1369-5274, DOI: 10.1016/J.MIB.2017.05.008 page 71, left-hand column, last paragraph - page 71, right-hand column, paragraph 1	1

-/--

[X] Further documents are listed in the continuation of Box C.

[X] 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	Date of mailing of the international search report
13 November 2020	20/01/2021

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Knudsen, Henrik

1

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2020/050237

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	LI YI ET AL: "CRISPR/Cas Systems towards Next-Generation Biosensing", TRENDS IN BIOTECHNOLOGY, vol. 37, no. 7, July 2019 (2019-07), pages 730-743, XP085710951, ISSN: 0167-7799, DOI: 10.1016/J.TIBTECH.2018.12.005 pages 737-738; figures 1,4; table 1	1
X	WO 2013/176772 A1 (UNIV CALIFORNIA [US]; UNIV VIENNA [AT] ET AL.) 28 November 2013 (2013-11-28) sequence 51	1-6, 10-16, 40-48, 58-70, 101-119, 121,122

1

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2020/050237

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

6, 48(completely); 1-5, 10-16, 40-47, 58-70, 101-119, 121, 122(partially)

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. claims: 6, 48(completely); 1-5, 10-16, 40-47, 58-70, 101-119, 121, 122(partially)

   CRISPR/Cas9 system comprising Cas9.1, which is a protein having at least 70% identity to SEQ ID NO:1 and gRNA binding thereto.

   ---

2. claims: 7, 49(completely); 1-5, 10-16, 40-47, 58-70, 101-119, 121, 122(partially)

   CRISPR/Cas9 system comprising Cas9.2, which is a protein having at least 70% identity to SEQ ID NO:2 and gRNA binding thereto.

   ---

3. claims: 8, 50(completely); 1-5, 10-16, 40-47, 58-70, 101-119, 121, 122(partially)

   CRISPR/Cas9 system comprising Cas9.3, which is a protein having at least 70% identity to SEQ ID NO:10 and gRNA binding thereto.

   ---

4. claims: 9, 51(completely); 1-5, 10-16, 40-47, 58-70, 101-119, 121, 122(partially)

   CRISPR/Cas9 system comprising Cas9.4, which is a protein having at least 70% identity to SEQ ID NO:11 and gRNA binding thereto.

   ---

5. claims: 17-24, 29, 30, 88, 89, 120(completely); 25-27, 32-39, 52-86, 91-119, 121, 122(partially)

   Engineered system comprising a Class II Type V CRISPR-Cas endonuclease, which may be Cas12p with SEQ ID NO:4 or an endonuclease having at least 70% identity thereto, and a single guide RNA wherein the endonuclease possesses collateral activity and is capable of cleaving a single stranded polynucleotide comprising RNA without a tracrRNA. gRNA with SEQ ID NO:117.

   ---

6. claims: 28, 87(completely); 25-27, 32-39, 52-86, 91-119, 121, 122(partially)

   CRISPR/Cas9 system comprising Cas12a.1, which is a protein having at least 70% identity to SEQ ID NO:3, and gRNA binding thereto. gRNA with SEQ ID NO:116.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

7. claims: 31, 90(completely); 25-27, 32-39, 58-86, 91-119, 121, 122(partially)

CRISPR/Cas9 system comprising Cas12q, which is a protein having at least 70% identity to SEQ ID NO:5, and gRNA binding thereto.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2020/050237

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2019002889 A1	03-01-2019	AU 2018291496 A1	05-12-2019
		AU 2020201843 A1	02-04-2020
		CA 3068543 A1	03-01-2019
		DE 202018006334 U1	28-01-2020
		EP 3645722 A2	06-05-2020
		US 2019002875 A1	03-01-2019
		US 2019002889 A1	03-01-2019
		WO 2019006471 A2	03-01-2019
WO 2013176772 A1	28-11-2013	AU 2013266968 A1	20-11-2014
		AU 2017225060 A1	28-09-2017
		AU 2019201850 A1	11-04-2019
		BR 112014029441 A2	27-06-2017
		CA 2872241 A1	28-11-2013
		CL 2014003208 A1	19-06-2015
		CN 104854241 A	19-08-2015
		CN 107603976 A	19-01-2018
		CO 7151523 A2	29-12-2014
		CR 20140538 A	18-02-2015
		CY 1119282 T1	14-02-2018
		CY 1120291 T1	10-07-2019
		DE 202013012240 U1	02-02-2016
		DE 202013012241 U1	18-01-2016
		DE 202013012242 U1	02-02-2016
		DK 2800811 T3	17-07-2017
		DK 3241902 T3	07-05-2018
		DK 3401400 T3	03-06-2019
		EA 201401319 A1	29-05-2015
		EC SP14028704 A	30-09-2015
		EP 2800811 A1	12-11-2014
		EP 3241902 A1	08-11-2017
		EP 3401400 A1	14-11-2018
		EP 3597749 A1	22-01-2020
		ES 2636902 T3	10-10-2017
		ES 2670718 T3	31-05-2018
		ES 2728782 T3	28-10-2019
		GB 2518764 A	01-04-2015
		GB 2537000 A	05-10-2016
		HK 1204003 A1	06-11-2015
		HK 1207107 A1	22-01-2016
		HK 1211978 A1	03-06-2016
		HR P20171163 T1	06-10-2017
		HR P20180794 T1	13-07-2018
		HR P20190965 T1	04-10-2019
		HU E038850 T2	28-11-2018
		HU E043861 T2	30-09-2019
		IL 235461 A	31-01-2019
		IL 261567 A	30-11-2020
		JP 6343605 B2	13-06-2018
		JP 6692856 B2	13-05-2020
		JP 2015523856 A	20-08-2015
		JP 2018138054 A	06-09-2018
		JP 2020039351 A	19-03-2020
		KR 20150016588 A	12-02-2015
		KR 20170134766 A	06-12-2017
		LT 2800811 T	11-09-2017
		LT 3241902 T	25-06-2018
		LT 3401400 T	10-06-2019

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2020/050237

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
		MA 37663 A1	31-05-2017
		ME 02836 B	20-01-2018
		MX 349744 B	10-08-2017
		MX 362866 B	20-02-2019
		MX 369077 B	25-10-2019
		NZ 714353 A	26-05-2017
		NZ 728024 A	31-05-2019
		PE 20150336 A1	25-03-2015
		PE 20190842 A1	17-06-2019
		PE 20190843 A1	17-06-2019
		PE 20190844 A1	17-06-2019
		PH 12014502574 A1	21-01-2015
		PL 2800811 T3	30-11-2017
		PL 3241902 T3	28-09-2018
		PL 3401400 T3	31-12-2019
		PT 2800811 T	17-08-2017
		PT 3241902 T	28-05-2018
		PT 3401400 T	12-06-2019
		SG 10201701800Y A	27-04-2017
		SG 10201809817U A	28-12-2018
		SG 11201407702X A	30-12-2014
		SI 2800811 T1	30-10-2017
		SI 3241902 T1	31-07-2018
		SI 3401400 T1	30-10-2019
		TN 2014000493 A1	30-03-2016
		TR 201806812 T4	21-06-2018
		UA 118014 C2	12-11-2018
		US 10774344 B1	15-09-2020
		US 2014068797 A1	06-03-2014
		US 2016046961 A1	18-02-2016
		US 2016060653 A1	03-03-2016
		US 2016060654 A1	03-03-2016
		US 2016068864 A1	10-03-2016
		US 2016130608 A1	12-05-2016
		US 2016130609 A1	12-05-2016
		US 2016138008 A1	19-05-2016
		US 2017051310 A1	23-02-2017
		US 2017051312 A1	23-02-2017
		US 2017166893 A1	15-06-2017
		US 2018208931 A1	26-07-2018
		US 2018230495 A1	16-08-2018
		US 2018230496 A1	16-08-2018
		US 2018230497 A1	16-08-2018
		US 2018237801 A1	23-08-2018
		US 2018245100 A1	30-08-2018
		US 2018245101 A1	30-08-2018
		US 2018251791 A1	06-09-2018
		US 2018251793 A1	06-09-2018
		US 2018251794 A1	06-09-2018
		US 2018251795 A1	06-09-2018
		US 2018273981 A1	27-09-2018
		US 2018282764 A1	04-10-2018
		US 2018298406 A1	18-10-2018
		US 2018298407 A1	18-10-2018
		US 2018312874 A1	01-11-2018
		US 2018312875 A1	01-11-2018
		US 2018312876 A1	01-11-2018
		US 2019002921 A1	03-01-2019

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2020/050237

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
		US 2019002922 A1	03-01-2019
		US 2019002923 A1	03-01-2019
		US 2019010520 A1	10-01-2019
		US 2019062790 A1	28-02-2019
		US 2019093129 A1	28-03-2019
		US 2019106711 A1	11-04-2019
		US 2019106712 A1	11-04-2019
		US 2019106713 A1	11-04-2019
		US 2019106714 A1	11-04-2019
		US 2019106715 A1	11-04-2019
		US 2019169641 A1	06-06-2019
		US 2019169642 A1	06-06-2019
		US 2019169643 A1	06-06-2019
		US 2019169644 A1	06-06-2019
		US 2019169645 A1	06-06-2019
		US 2019169646 A1	06-06-2019
		US 2019169647 A1	06-06-2019
		US 2019169648 A1	06-06-2019
		US 2019169649 A1	06-06-2019
		US 2019256869 A1	22-08-2019
		US 2019256870 A1	22-08-2019
		US 2019256871 A1	22-08-2019
		US 2019264233 A1	29-08-2019
		US 2019264234 A1	29-08-2019
		US 2019264235 A1	29-08-2019
		US 2019264236 A1	29-08-2019
		US 2019271008 A1	05-09-2019
		US 2019284583 A1	19-09-2019
		US 2019330662 A1	31-10-2019
		US 2020071728 A1	05-03-2020
		US 2020080113 A1	12-03-2020
		US 2020190542 A1	18-06-2020
		US 2020277631 A1	03-09-2020
		US 2020299734 A1	24-09-2020
		US 2020308605 A1	01-10-2020
		US 2020308606 A1	01-10-2020
		US 2020308607 A1	01-10-2020
		US 2020325495 A1	15-10-2020
		US 2020354747 A1	12-11-2020
		US 2020354748 A1	12-11-2020
		US 2020354749 A1	12-11-2020
		US 2020362370 A1	19-11-2020
		WO 2013176772 A1	28-11-2013