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

(19) World Intellectual Property Organization
International Bureau

PCT

(43) International Publication Date
17 November 2011 (17.11.2011)

(10) International Publication Number
WO 2011/143124 A2

(51) International Patent Classification:
C12N 9/22 (2006.01) *C12Q 1/68* (2006.01)
C12N 15/55 (2006.01) *G01N 33/573* (2006.01)
C12Q 1/34 (2006.01)

(21) International Application Number:
PCT/US2011/035775

(22) International Filing Date:
9 May 2011 (09.05.2011)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/333,163 10 May 2010 (10.05.2010) US
61/365,627 19 July 2010 (19.07.2010) US
61/413,287 12 November 2010 (12.11.2010) US

(71) Applicant *(for all designated States except US)*: **THE REGENTS OF THE UNIVERSITY OF CALIFORNIA** [US/US]; 1111 Franklin Street, Twelfth Floor, Oakland, California 94607-5200 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **HAURWITZ, Rachel E.** [US/US]; 2350 Cedar St., Berkeley, California 94708 (US). **DOUDNA, Jennifer A.** [US/US]; 164 Vicente Rd., Berkeley, California 94705 (US). **WIEDENHEFT, Blake** [US/US]; 5004 A Manila, Oakland, California 94609 (US). **JINEK, Martin** [CZ/US]; 1846 Spruce Street, Berkeley, California 94709 (US).

(74) Agent: **BORDEN, Paula A.**; Bozicevic, Field & Francis LLP, 1900 University Avenue Suite 200, East Palo Alto, California 94303 (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, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

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

Published:

— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

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

WO 2011/143124 A2

(54) **Title**: ENDORIBONUCLEASE COMPOSITIONS AND METHODS OF USE THEREOF

(57) **Abstract**: The present disclosure provides variant Csy4 endoribonucleases, nucleic acids encoding the variant Csy4 endoribonucleases, and host cells genetically modified with the nucleic acids. The variant Csy4 endoribonucleases find use in a variety of applications, which are also provided. The present disclosure also provides methods of detecting a specific sequence in a target polyribonucleotide; and methods of regulating production of a target RNA in a eukaryotic cell.

ENDORIBONUCLEASE COMPOSITIONS AND METHODS OF USE THEREOF

CROSS-REFERENCE

[0001] This application claims the benefit of U.S. Provisional Patent Application No. 61/333,163, filed May 10, 2010, U.S. Provisional Patent Application No. 61/365,627, filed July 19, 2010, and U.S. Provisional Patent Application No. 61/413,287, filed November 12, 2010, each of which applications is incorporated herein by reference in its entirety.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH

[0002] This invention was made with government support under Grant No. T32 GM07232 awarded by the National Institutes of Health and Grant No. MCB-0950971 awarded by the National Science Foundation. The government has certain rights in the invention.

BACKGROUND

[0003] DNA restriction enzymes transformed molecular biology in the 1970s by making it possible to cleave specific DNA sequences at will. Sequencing of RNA molecules currently entails copying the RNA into a DNA strand that is then sequenced by conventional methods. This approach, also known as RNASeq, is robust and can yield many millions of sequence reads. However, the necessity of generating cDNA introduces inherent bias due to sequence-dependent efficiencies of individual steps.

Literature

[0004] Carte et al. (2008) *Genes Dev.* 22:3489; U.S. Patent Publication No. 2010/0093026.

SUMMARY OF THE INVENTION

[0005] The present disclosure provides variant Csy4 endoribonucleases, nucleic acids encoding the variant Csy4 endoribonucleases, and host cells genetically modified with the nucleic acids. The variant Csy4 endoribonucleases find use in a variety of applications, which are also provided. The present disclosure also provides methods of

detecting a specific sequence in a target polyribonucleotide; and methods of regulating production of a target RNA in a eukaryotic cell.

BRIEF DESCRIPTION OF THE DRAWINGS

[0006] Figures 1A-C depict specific recognition of a pre-crRNA substrate by Pa14Csy4. The nucleotide sequence depicted is 5'-GUUCACUGCCGUAUAGGCAGCUAAGAAA-3' (SEQ ID NO:1).

[0007] Figures 2A-C depict crystal structures of Csy4 bound to RNA substrate.

[0008] Figures 3A and 3B depict: a detailed view of the catalytic center of Csy4 (Figure 3A); and cleavage activity of Csy4 wild-type (WT) and mutants (Figure 3B).

[0009] Figure 4 depicts invariant amino acids among 12 Csy4 sequences. Pa (SEQ ID NO:8); Yp (SEQ ID NO:34); Ec89 (SEQ ID NO:39); Dn (SEQ ID NO:79); Ab (SEQ ID NO:84); MP1 (SEQ ID NO:2); MP01 (SEQ ID NO:3); SW (SEQ ID NO:4); Pm (SEQ ID NO:85); Pw (SEQ ID NO:13); and Dd (SEQ ID NO:10).

[0010] Figures 5A-5BD present an amino acid sequence alignment of various Csy4 polypeptides, as well as the nucleotide sequences of RNA sequences recognized by each Csy4 polypeptide.

[0011] Figure 6 depicts examples of amino acid sequences of enzymatically inactive, sequence-specific endoribonucleases.

[0012] Figure 7 depicts an example of a method for detecting a specific sequence in a target polyribonucleotide.

[0013] Figure 8 depicts the effect of imidazole on activation of various enzymatically inactive Csy4 variants.

[0014] Figure 9 depicts an exemplary method of isolating a target RNA. A Csy4 target stem-loop (SEQ ID NO:103) is shown.

[0015] Figure 10 depicts an exemplary method of regulating expression of a target RNA in a eukaryotic cell. A Csy4 RNA substrate sequence (SEQ ID NO:103) is shown.

DEFINITIONS

[0016] As used herein, "polyribonucleotide" refers to a polymeric form of ribonucleotides, and includes RNA, RNA containing deoxyribonucleotide(s), and DNA containing ribonucleotide(s). A polyribonucleotide can in some cases include one or

more modified nucleotides (e.g., deoxyinosine, deoxyuridine or hydroxymethyldeoxyuridine). In some cases, a polyribonucleotide consists of a ribonucleotides only (i.e., does not include any deoxyribonucleotides). In some cases, a polyribonucleotide comprises ribonucleotides, and one or more modified ribonucleotides, but does not include any deoxyribonucleotides. In other cases, a polyribonucleotide comprises ribonucleotides, and may comprise one or more modified ribonucleotides, and one or more deoxyribonucleotides (including modified deoxyribonucleotides). In some cases, where a polyribonucleotide comprises one or more deoxyribonucleotides, the deoxyribonucleotides comprise from about 50% to about 40%, from about 40% to about 30%, from about 30% to about 20%, from about 20% to about 10%, from about 10% to about 1%, or less than 1%, of the total nucleotides in the polyribonucleotide.

[0017] The terms “nucleic acid” and “polynucleotide” are used interchangeably and refer to a polymeric form of nucleotides of any length, either deoxyribonucleotides or ribonucleotides, or analogs thereof. Non-limiting examples of polynucleotides include linear and circular nucleic acids, messenger RNA (mRNA), cDNA, recombinant polynucleotides, vectors, probes, and primers.

[0018] A “biological sample” encompasses a variety of sample types obtained from a cell, extracellular matter, a tissue, or a multicellular organism. The definition 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. The definition also includes samples that have been manipulated in any way after their procurement, such as by treatment with reagents, solubilization, or enrichment for certain components, such as polynucleotides. The term “biological sample” encompasses a clinical sample, and also includes cells in culture, cell supernatants, cell lysates, serum, plasma, biological fluid (e.g., cerebrospinal fluid, bronchoalveolar lavage fluid, urine, blood, a blood fraction (e.g., plasma; serum), sputum, and the like), and tissue samples. In some cases, a biological sample comprises cells. In other cases, a biological sample is cell free.

[0019] The term “operably linked” refers to functional linkage between molecules to provide a desired function. For example, “operably linked” in the context of nucleic acids refers to a functional linkage between nucleic acids to provide a desired function

such as transcription, translation, and the like, e.g., a functional linkage between a nucleic acid expression control sequence (such as a promoter, signal sequence, or array of transcription factor binding sites) and a second polynucleotide, wherein the expression control sequence affects transcription and/or translation of the second polynucleotide. “Operably linked” in the context of a polypeptide refers to a functional linkage between amino acid sequences (e.g., of different domains) to provide for a described activity of the polypeptide.

[0020] “Isolated” refers to a protein or nucleic acid that, if naturally occurring, is in an environment different from that in which it may naturally occur. “Isolated” is meant to include proteins or nucleic acids that are within samples that are substantially enriched for the protein or nucleic acid of interest and/or in which the protein or nucleic acid of interest is partially or substantially purified. Where the protein or nucleic acid is not naturally occurring, “isolated” indicates the protein or nucleic acid has been separated from an environment in which it was made by either synthetic or recombinant means.

[0021] “Substantially pure” indicates that an entity (e.g., polypeptide or a nucleic acid) makes up greater than about 50% of the total content of the composition (e.g., total protein of the composition) and typically, greater than about 60% of the total protein content. In some embodiments, “substantially pure” refers to compositions in which at least 75%, at least 85%, at least 90% or more of the total composition is the entity of interest (e.g. 95%, of the total protein). In some embodiments, the protein or nucleic acid of interest will make up greater than about 90%, greater than about 95%, greater than about 98%, or greater than about 99%, of the total protein or nucleic acid in the composition.

[0022] Before the present invention is further described, it is to be understood that this invention is not limited to particular embodiments described, as such may, of course, vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to be limiting, since the scope of the present invention will be limited only by the appended claims.

[0023] 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.

[0024] 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. Although any methods and materials similar or equivalent to those described herein can also be used in the practice or testing of the present invention, the preferred methods and materials are now described. 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.

[0025] 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 site-specific endoribonuclease" includes a plurality of such site-specific endoribonucleases and reference to "the target polyribonucleotide" includes reference to one or more target polyribonucleotides 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.

[0026] The publications discussed herein are provided solely for their disclosure prior to the filing date of the present application. Nothing herein is to be construed as an admission that the present invention is not entitled to antedate such publication by virtue of prior invention. Further, the dates of publication provided may be different from the actual publication dates which may need to be independently confirmed.

DETAILED DESCRIPTION

[0027] The present disclosure provides variant Csy4 endoribonucleases, nucleic acids encoding the variant Csy4 endoribonucleases, and host cells genetically modified with the nucleic acids. The variant Csy4 endoribonucleases find use in a variety of

applications, which are also provided. The present disclosure also provides methods of detecting a specific sequence in a target polyribonucleotide; and methods of regulating production of a target RNA in a eukaryotic cell.

METHODS OF DETECTING A SEQUENCE IN A TARGET POLYRIBONUCLEOTIDE

[0028] The present disclosure provides a method of detecting a sequence in a target polyribonucleotide. The methods are useful for detecting the presence of a particular sequence in a polyribonucleotide, and can therefore be used to detect a polyribonucleotide comprising a particular sequence. For example, the method can be used to detect the presence of a polyribonucleotide of a pathogen in a sample (e.g., in a biological sample).

[0029] A subject method can detect as few as 100 copies, down to a single copy, of a target polyribonucleotide. Thus, e.g., a subject method can detect from 1 to about 5, from about 5 to about 10, from about 10 to about 50, or from about 50 to about 100, or more than 100, copies of a target polyribonucleotide in a sample (e.g., in a single cell, in a single embryo, or other biological sample). A subject method is thus useful for various forensic, research, and diagnostic applications.

[0030] In some embodiments, a subject method of detecting a specific sequence in a target polyribonucleotide comprises: a) contacting the target polyribonucleotide with a oligonucleotide probe comprising the specific sequence and an enzymatically active sequence-specific Csy4 endoribonuclease under conditions that favor duplex formation between the oligonucleotide probe and the target polyribonucleotide, wherein the duplex is cleaved by the Csy4 endoribonuclease; and b) detecting specific binding between the oligonucleotide probe and the target polyribonucleotide, wherein detection of duplex formation between the oligonucleotide probe and the target polyribonucleotide indicates the presence of the specific sequence in the target polyribonucleotide.

[0031] In some cases, the oligonucleotide probe is linked to a peptide, and the peptide is released upon cleavage of the duplex by the Csy4 endoribonuclease; in these cases, the detection step involves detection of the released peptide. For example, the released peptide is detected by binding to an antibody specific for the peptide, e.g., where the antibody is immobilized. In some embodiments, the target polyribonucleotide is immobilized on a solid support. Target polyribonucleotides include any of a variety of

polynucleotides, e.g., the target polyribonucleotide can be a polyribonucleotide of a pathogen.

[0032] As noted above, in some embodiments, the antibody or the target polynucleotide is immobilized on a solid support (insoluble support). Suitable insoluble supports include, but are not limited to agarose beads, magnetic beads, a test strip, a multi-well dish, and the like. The insoluble support can comprise a variety of substances (glass, polystyrene, polyvinyl chloride, polypropylene, polyethylene, polycarbonate, dextran, nylon, amylose, natural and modified celluloses, polyacrylamides, agaroses, and magnetite) and can be provided in a variety of forms, including, e.g., agarose beads, polystyrene beads, latex beads, magnetic beads, colloid metal particles, glass and/or silicon chips and surfaces, nitrocellulose strips, nylon membranes, sheets, wells of reaction trays (e.g., multi-well plates), plastic tubes, etc.

[0033] In some embodiments, the method generally involves: a) contacting a target polyribonucleotide with a sequence-specific endoribonuclease; and b) detecting cleavage fragments produced by site-specific cleavage of the target polyribonucleotide, where production of cleavage fragments expected upon cleavage at a specific sequence in the polyribonucleotide indicates the presence of the specific sequence.

[0034] In other embodiments, a subject method of detecting a sequence in a target polyribonucleotide involves: a) contacting a target polyribonucleotide with: i) a sequence-specific endoribonuclease; and ii) an oligonucleotide probe comprising a linked detection moiety, where the oligonucleotide probe comprises a specific, known nucleotide sequence; wherein the oligonucleotide probe forms a duplex with a complementary sequence in the target polyribonucleotide based on binding of the known nucleotide sequence present in the oligonucleotide probe to a complementary sequence in the target polyribonucleotide, and where the sequence-specific endoribonuclease cleaves the duplex in a sequence-specific manner, thereby releasing the detection moiety from the oligonucleotide probe; and b) detecting the released detection moiety, where release of the detection moiety indicates the presence of the specific sequence. In some embodiments, two or more different oligonucleotide probes are used, each comprising a different specific, known nucleotide sequence.

[0035] In some embodiments, the detection moiety is a polypeptide. The polypeptide can be detected using an immunological assay (e.g., an enzyme-linked immunosorbent

assay (ELISA); a radioimmunoassay (RIA); etc.), using an antibody specific for the polypeptide detection moiety. The antibody specific for the polypeptide detection moiety can comprise a detectable label. The immunological assay can be carried out on a test strip (e.g., in a lateral flow assay) or other suitable medium such as a multi-well plate.

[0036] In some embodiments, the detection moiety is a fluorescent protein, where suitable fluorescent proteins are as described herein. In other embodiments, the detection moiety is luciferin or other substrate for luciferase. Suitable luciferins or other luciferase substrates include, e.g., luciferin (e.g., a firefly luciferin); an aminoluciferin; coelenterazine; a modified coelenterazine as described in U.S. Patent No. 7,537,912; a coelenterazine analog as described in U.S. Patent Publication No. 2009/0081129 (e.g., a membrane permeant coelenterazine analog as described in U.S. Patent Publication No. 2009/0081129, e.g., one of Structures II, III, IV, V, and VI of U.S. Patent Publication No. 2009/0081129); aminoluciferin; dihydroluciferin; luciferin 6' methylether; or luciferin 6' chloroethylether. See, e.g., Branchini, B.R. et al. *Anal. Biochem.* 2010, 396, 290-296; and Mezzanotte, L. et al., In vivo bioluminescence imaging of murine xenograft cancer models with a red-shifted thermostable luciferase. *Mol. Imaging Biol.* (2009, Nov. 9, online; PubMed ID: 19937390).

[0037] A non-limiting example of a subject detection method is illustrated schematically in **Figure 7**. In the example depicted in Figure 7, small oligonucleotides that bind discrete regions of a target polynucleotide (e.g., a viral RNA) are contacted with the target polynucleotide, where the oligonucleotides comprise detectable moieties (e.g., ligands; peptides; etc.). An enzymatically active, sequence-specific restriction endonuclease (RRE) that targets the oligonucleotide/viral RNA duplex is added. The enzyme cleaves the oligonucleotide/viral RNA duplex; and ligands are released for detection. The enzyme cleaves further duplexes, thereby amplifying the signal. Released ligands are detected using a lateral flow (e.g., test strip) or an immunological based assay (e.g., ELISA).

[0038] A suitable sequence-specific endoribonuclease is an enzymatically active, sequence-specific endoribonuclease. Endoribonucleases that are suitable for use in a subject detection method include endoribonucleases that bind to and cleave a substrate polyribonucleotide in a sequence-specific manner include enzymatically active polypeptides and that have at least about 85%, at least about 90%, at least about 95%, at

least about 98%, at least about 99%, or 100%, amino acid sequence identity to an amino acid sequence set forth in **Figure 4** (Csy4 amino acid sequences).

[0039] Endoribonucleases that are suitable for use in a subject detection method include endoribonucleases that bind to and cleave a substrate polyribonucleotide in a sequence-specific manner include enzymatically active polypeptides and that have at least about 85%, at least about 90%, at least about 95%, at least about 98%, at least about 99%, or 100%, amino acid sequence identity to an amino acid sequence set forth in **Figure 5** (Csy4 amino acid sequences). Figure 5 provides sequences specifically bound by the various endoribonucleases. In some cases, a suitable enzymatically active sequence-specific Csy4 endoribonuclease can comprise an amino acid sequence of a Csy4 amino acid sequence depicted in Figure 5.

[0040] Endoribonucleases that are suitable for use in a subject detection method include endoribonucleases that bind to and cleave a substrate polyribonucleotide in a sequence-specific manner include enzymatically active polypeptides and that differ from an amino acid sequence set forth in any one of **Figures 4 or 5** by from 1 to 20 (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20) amino acid substitutions and/or insertions and/or deletions.

[0041] The target polyribonucleotide to be detected can be present in a sample, e.g., a biological sample such as blood, a blood product (e.g., plasma), urine, cerebrospinal fluid, bronchoalveolar lavage fluid, saliva, a tissue, cells, etc. The target polyribonucleotide can be isolated or purified. The target polyribonucleotide can be a messenger RNA (mRNA), a viral RNA, bacterial RNA, parasite RNA, or other RNA species. Viral RNAs include, but are not limited to, any member of the Flaviviridae, e.g., hepatitis C virus, Dengue virus, Yellow Fever Virus, West Nile Virus, etc.; any member of Retroviridae; an immunodeficiency virus (e.g., human immunodeficiency virus); etc.

[0042] The target polyribonucleotide to be detected can be present in a cell of a multicellular organism (or can be obtained from a cell of a multicellular organism).

[0043] The target polyribonucleotide to be detected can be present in or obtained from a cell or organism of any of the six kingdoms, e.g., Bacteria (e.g., Eubacteria); Archaebacteria; Protista; Fungi; Plantae; and Animalia. Suitable sources of target polyribonucleotides include plant-like members of the kingdom Protista, including, but not limited to, algae (e.g., green algae, red algae, glaucophytes, cyanobacteria); fungus-

like members of Protista, e.g., slime molds, water molds, etc.; animal-like members of Protista, e.g., flagellates (e.g., Euglena), amoeboids (e.g., amoeba), sporozoans (e.g, Apicomplexa, Myxozoa, Microsporidia), and ciliates (e.g., Paramecium). Suitable sources of target polyribonucleotides include members of the kingdom Fungi, including, but not limited to, members of any of the phyla: Basidiomycota (club fungi; e.g., members of Agaricus, Amanita, Boletus, Cantherellus, etc.); Ascomycota (sac fungi, including, e.g., Saccharomyces); Mycophycophyta (lichens); Zygomycota (conjugation fungi); and Deuteromycota. Suitable sources of target polyribonucleotides include members of the kingdom Plantae, including, but not limited to, members of any of the following divisions: Bryophyta (e.g., mosses), Anthocerotophyta (e.g., hornworts), Hepaticophyta (e.g., liverworts), Lycophyta (e.g., club mosses), Sphenophyta (e.g., horsetails), Psilophyta (e.g., whisk ferns), Ophioglossophyta, Pterophyta (e.g., ferns), Cycadophyta, Gingkophyta, Pinophyta, Gnetophyta, and Magnoliophyta (e.g., flowering plants). Suitable sources of target polyribonucleotides include members of the kingdom Animalia, including, but not limited to, members of any of the following phyla: Porifera (sponges); Placozoa; Orthonectida (parasites of marine invertebrates); Rhombozoa; Cnidaria (corals, anemones, jellyfish, sea pens, sea pansies, sea wasps); Ctenophora (comb jellies); Platyhelminthes (flatworms); Nemertina (ribbon worms); Ngathostomulida (jawed worms); Gastrotricha; Rotifera; Priapulida; Kinorhyncha; Loricifera; Acanthocephala; Entoprocta; Nemotoda; Nematomorpha; Cycliophora; Mollusca (mollusks); Sipuncula (peanut worms); Annelida (segmented worms); Tardigrada (water bears); Onychophora (velvet worms); Arthropoda (including the subphyla: Chelicerata, Myriapoda, Hexapoda, and Crustacea, where the Chelicerata include, e.g., arachnids, Merostomata, and Pycnogonida, where the Myriapoda include, e.g., Chilopoda (centipedes), Diplopoda (millipedes), Paropoda, and Symphyla, where the Hexapoda include insects, and where the Crustacea include shrimp, krill, barnacles, etc.; Phoronida; Ectoprocta (moss animals); Brachiopoda; Echinodermata (e.g. starfish, sea daisies, feather stars, sea urchins, sea cucumbers, brittle stars, brittle baskets, etc.); Chaetognatha (arrow worms); Hemichordata (acorn worms); and Chordata. Suitable members of Chordata include any member of the following subphyla: Urochordata (sea squirts; including Ascidiacea, Thaliacea, and Larvacea); Cephalochordata (lancelets); Myxini (hagfish); and Vertebrata, where members of Vertebrata include, e.g., members

of Petromyzontida (lampreys), Chondrichthyces (cartilaginous fish), Actinopterygii (ray-finned fish), Actinista (coelocanths), Dipnoi (lungfish), Reptilia (reptiles, e.g., snakes, alligators, crocodiles, lizards, etc.), Aves (birds); and Mammalian (mammals). Suitable plants include any monocotyledon and any dicotyledon.

[0044] Thus, e.g., a target polyribonucleotide can be present in or obtained from cells from organisms that include, but are not limited to, a protozoan, a plant, a fungus, an algal cell, a yeast, a reptile, an amphibian, a mammal, a marine microorganism, a marine invertebrate, an arthropod, an isopod, an insect, an arachnid, an archaebacterium, and a eubacterium.

[0045] A target polyribonucleotide can be present in or obtained from a non-human embryo, e.g., a *Drosophila* embryo; a zebrafish embryo; a mouse embryo; etc.

[0046] A target polyribonucleotide can be present in or obtained from a stem cell, e.g., an *in vitro* stem cell; a non-human stem cell; etc. Suitable stem cells include embryonic stem cells, adult stem cells, and induced pluripotent stem (iPS) cells.

[0047] In some embodiments, target polyribonucleotide will be isolated from a tissue 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, the target polyribonucleotide will in some embodiments be isolated from the xylem, the phloem, the cambium layer, leaves, roots, etc. Where the organism is an animal, the target polyribonucleotide will in some embodiments be isolated from a particular tissue (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.).

METHODS OF REGULATING PRODUCTION OF A TARGET RNA

[0048] The present disclosure provides a method of regulating production of a target RNA in a cell. The method generally involves contacting a genetically modified host cell with an agent that activates an inducible promoter, where the genetically modified host cell is genetically modified with a recombinant expression vector comprising a nucleotide sequence encoding an enzyme that catalyzes cleavage at a sequence-specific cleavage site in a substrate polyribonucleotide, where the enzyme-encoding nucleotide sequence is operably linked to the inducible promoter, and where, upon activation of the

inducible promoter, the enzyme is produced in the cell and cleaves said target RNA from a precursor RNA.

[0049] **Figure 10** provides a schematic depiction of an exemplary method of regulating production of a target RNA. In Figure 10, an endogenous target RNA is modified to include a Csy4 RNA substrate (e.g., GUUCACUGCCGUAUAGGCAG (SEQ ID NO:103); or SEQ ID NO:1) in the 3' untranslated region (3' UTR). Cys4 expression in the host cell leads to binding and cleavage of the RNA substrate. The cleaved RNA now lacks its polyA tail and will be degraded.

[0050] For example, in some embodiments, the present disclosure provides a method of regulating production of a target RNA in a eukaryotic cell, where the method involves contacting a genetically modified host cell with an agent that activates an inducible promoter, where the genetically modified host cell is genetically modified with a recombinant expression vector comprising a nucleotide sequence encoding an enzymatically active sequence-specific Csy4 endoribonuclease that catalyzes cleavage at a sequence-specific cleavage site in a substrate polyribonucleotide, where the enzyme-encoding nucleotide sequence is operably linked to the inducible promoter, and where, upon activation of the inducible promoter, the enzyme is produced in the cell and cleaves said target RNA from a precursor RNA. In some cases, the target RNA species is a regulatory RNA. In some cases, cleavage of said target RNA from a precursor RNA inactivates the precursor RNA.

[0051] A suitable sequence-specific endoribonuclease is an enzymatically active, sequence-specific endoribonuclease. Endoribonucleases that are suitable for use in a subject method of regulating production of a target RNA include endoribonucleases that bind to and cleave a substrate polyribonucleotide in a sequence-specific manner include enzymatically active polypeptides and that have at least about 85%, at least about 90%, at least about 95%, at least about 98%, at least about 99%, or 100%, amino acid sequence identity to an amino acid sequence set forth in **Figure 4** (Csy4 amino acid sequences).

[0052] Endoribonucleases that are suitable for use in a subject method of regulating production of a target RNA include endoribonucleases that bind to and cleave a substrate polyribonucleotide in a sequence-specific manner include enzymatically active polypeptides and that have at least about 85%, at least about 90%, at least about 95%, at

least about 98%, at least about 99%, or 100%, amino acid sequence identity to an amino acid sequence set forth in **Figure 5** (Csy4 amino acid sequences). Figure 5 provides sequences specifically bound by the various endoribonucleases.

[0053] Endoribonucleases that are suitable for use in a subject method of regulating production of a target RNA include endoribonucleases that bind to and cleave a substrate polyribonucleotide in a sequence-specific manner include enzymatically active polypeptides and that differ from an amino acid sequence set forth in any one of **Figures 4 or 5** by from 1 to 20 (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20) amino acid substitutions and/or insertions and/or deletions.

[0054] A suitable inducible promoter can include a promoter that is functional in a eukaryotic cell. Suitable inducible promoters are known in the art. For example, suitable inducible promoters include, but are not limited to, a GAL1 promoter, a GAL10 promoter, an ADH2 promoter, a PHO5 promoter, a CUP1 promoter, a GAL7 promoter, a MET25 promoter, a MET3 promoter, a CYC1 promoter, a HIS3 promoter, an ADH1 promoter, a PGK promoter, a GAPDH promoter, an ADC1 promoter, a TRP1 promoter, a URA3 promoter, a LEU2 promoter, an ENO promoter, a TP1 promoter, and AOX1. Suitable inducible promoters include tetracycline-inducible promoters; a metallothionein promoter; tetracycline-inducible promoters, methionine-inducible promoters; and galactose-inducible promoters, which promoters are all well known in the art. Other suitable promoters include the ADH2 alcohol dehydrogenase promoter (repressed in glucose, induced when glucose is exhausted and ethanol is made) and the CUP1 metallothionein promoter (induced in the presence of Cu^2+, Zn^2+).

[0055] Agents that induce any given inducible promoter are known in art. For example, tetracycline-regulatable promoters can be regulated by tetracycline or doxycycline; carbohydrates can be used to induce a carbohydrate-inducible promoter (e.g., galactose for a galactose-inducible promoter); methionine can be used to induce a methionine-inducible promoter; metals can be used to induce a metallothionein promoter.

[0056] The target RNA can be a regulatory RNA. Regulator RNAs are well known in the art and include, e.g., micro-RNAs, short hairpin RNAs (shRNAs), and the like.

[0057] In some embodiments, cleavage of the target RNA from a precursor RNA inactivates the precursor RNA.

[0058] The genetically modified host cell can be an *in vitro* cell, e.g., a prokaryotic cell, or a eukaryotic cell (e.g., a mammalian cell, including primary cells, transformed cell lines, and the like). The genetically modified host cell can be an *in vivo* cell. In some embodiments, the *in vivo* cell is a non-human cell.

[0059] The genetically modified host cell can be a cell of a multicellular organism (or can be obtained from a cell of a multicellular organism).

[0060] The genetically modified host cell can be a cell obtained from or present in an organism of any of the six kingdoms, e.g., Bacteria (e.g., Eubacteria); Archaebacteria; Protista; Fungi; Plantae; and Animalia. Suitable organisms include plant-like members of the kingdom Protista, including, but not limited to, algae (e.g., green algae, red algae, glaucophytes, cyanobacteria); fungus-like members of Protista, e.g., slime molds, water molds, etc.; animal-like members of Protista, e.g., flagellates (e.g., Euglena), amoeboids (e.g., amoeba), sporozoans (e.g, Apicomplexa, Myxozoa, Microsporidia), and ciliates (e.g., Paramecium). Suitable organisms include members of the kingdom Fungi, including, but not limited to, members of any of the phyla: Basidiomycota (club fungi; e.g., members of Agaricus, Amanita, Boletus, Cantherellus, etc.); Ascomycota (sac fungi, including, e.g., Saccharomyces); Mycophycophyta (lichens); Zygomycota (conjugation fungi); and Deuteromycota. Suitable organisms include members of the kingdom Plantae, including, but not limited to, members of any of the following divisions: Bryophyta (e.g., mosses), Anthocerotophyta (e.g., hornworts), Hepaticophyta (e.g., liverworts), Lycophyta (e.g., club mosses), Sphenophyta (e.g., horsetails), Psilophyta (e.g., whisk ferns), Ophioglossophyta, Pterophyta (e.g., ferns), Cycadophyta, Gingkophyta, Pinophyta, Gnetophyta, and Magnoliophyta (e.g., flowering plants). Suitable organisms include members of the kingdom Animalia, including, but not limited to, members of any of the following phyla: Porifera (sponges); Placozoa; Orthonectida (parasites of marine invertebrates); Rhombozoa; Cnidaria (corals, anemones, jellyfish, sea pens, sea pansies, sea wasps); Ctenophora (comb jellies); Platyhelminthes (flatworms); Nemertina (ribbon worms); Ngathostomulida (jawed worms); Gastrotricha; Rotifera; Priapulida; Kinorhyncha; Loricifera; Acanthocephala; Entoprocta; Nemotoda; Nematomorpha; Cycliophora; Mollusca (mollusks); Sipuncula (peanut worms); Annelida (segmented worms); Tardigrada (water bears); Onychophora (velvet worms); Arthropoda (including the subphyla: Chelicerata, Myriapoda, Hexapoda,

and Crustacea, where the Chelicerata include, e.g., arachnids, Merostomata, and Pycnogonida, where the Myriapoda include, e.g., Chilopoda (centipedes), Diplopoda (millipedes), Paropoda, and Symphyla, where the Hexapoda include insects, and where the Crustacea include shrimp, krill, barnacles, etc.; Phoronida; Ectoprocta (moss animals); Brachiopoda; Echinodermata (e.g. starfish, sea daisies, feather stars, sea urchins, sea cucumbers, brittle stars, brittle baskets, etc.); Chaetognatha (arrow worms); Hemichordata (acorn worms); and Chordata. Suitable members of Chordata include any member of the following subphyla: Urochordata (sea squirts; including Ascidiacea, Thaliacea, and Larvacea); Cephalochordata (lancelets); Myxini (hagfish); and Vertebrata, where members of Vertebrata include, e.g., members of Petromyzontida (lampreys), Chondrichthyces (cartilaginous fish), Actinopterygii (ray-finned fish), Actinista (coelocanths), Dipnoi (lungfish), Reptilia (reptiles, e.g., snakes, alligators, crocodiles, lizards, etc.), Aves (birds); and Mammalian (mammals). Suitable plants include any monocotyledon and any dicotyledon.

[0061] Thus, e.g., a genetically modified host cell can be a cell obtained from or present in a protozoan, a plant, a fungus, an algal cell, a yeast, a reptile, an amphibian, a mammal, a marine microorganism, a marine invertebrate, an arthropod, an isopod, an insect, an arachnid, an archaebacterium, and a eubacterium.

[0062] Suitable mammalian cells include primary cells and immortalized cell lines. Suitable mammalian cell lines include human cell lines, non-human primate cell lines, rodent (e.g., mouse, rat) cell lines, and the like. Suitable mammalian cell lines include, but are not limited to, HeLa cells (e.g., American Type Culture Collection (ATCC) No. CCL-2), CHO cells (e.g., ATCC Nos. CRL9618, CCL61, CRL9096), 293 cells (e.g., ATCC No. CRL-1573), Vero cells, NIH 3T3 cells (e.g., ATCC No. CRL-1658), Huh-7 cells, BHK cells (e.g., ATCC No. CCL10), PC12 cells (ATCC No. CRL1721), COS cells, COS-7 cells (ATCC No. CRL1651), RAT1 cells, mouse L cells (ATCC No. CCLI.3), human embryonic kidney (HEK) cells (ATCC No. CRL1573), HLHepG2 cells, and the like.

[0063] The genetically modified host cell can be a cell obtained from or present in a non-human embryo, e.g., a *Drosophila* embryo; a zebrafish embryo; a mouse embryo; etc.

[0064] The genetically modified host cell can be a stem cell, e.g., an *in vitro* stem cell; a non-human stem cell; etc. Suitable stem cells include embryonic stem cells, adult stem cells, and induced pluripotent stem (iPS) cells.

METHODS OF ISOLATING A TARGET NUCLEIC ACID

[0065] The present disclosure provides methods of isolating a target nucleic acid from a mixed population of nucleic acids. The methods generally involve: a) contacting a mixed population of nucleic acids with an immobilized sequence-specific, enzymatically inactive endoribonuclease, wherein the mixed population of nucleic acids includes a target nucleic acid comprising a "tag" (or "recognition") nucleotide sequence that is specifically bound by the immobilized sequence-specific, enzymatically inactive endoribonuclease, such that the target nucleic acid comprising the tag nucleotide sequence ("tagged target nucleic acid") binds to the immobilized sequence-specific, enzymatically inactive endoribonuclease, forming a tagged target nucleic acid/immobilized sequence-specific enzymatically active endoribonuclease complex, wherein the contacting step takes place in a liquid solution (a "binding solution"); and b) adding imidazole to the liquid solution to a final concentration of from about 100 mM to about 500 mM (e.g., from about 100 mM to about 150 mM, from about 150 mM to about 200 mM, from about 200 mM to about 250 mM, from about 250 mM to about 300 mM, from about 300 mM to about 350 mM, from about 350 mM to about 400 mM, from about 400 mM to about 450 mM, or from about 450 mM to about 500 mM), thereby forming a reactivation solution that enzymatically reactivates the enzymatically inactive endoribonuclease such that the endoribonuclease becomes enzymatically active and cleaves the target nucleic acid from the "tag" nucleotide sequence, thereby releasing the target nucleic acid. **Figure 9** is a schematic representation of an exemplary embodiment of a subject method for isolating a target RNA.

[0066] The method can further include one or more washing steps. For example, after step (a) and before step (b), the immobilized sequence-specific, enzymatically inactive endoribonuclease that comprises a bound target nucleic acid comprising a "tag" nucleotide sequence can be washed one or more times with the binding solution, such that the target nucleic acid remains bound to the sequence-specific, enzymatically inactive endoribonuclease, and any unbound nucleic acids are washed away.

[0067] The mixed population of nucleic acids can include RNA and DNA. The target nucleic acid is an RNA that comprises a "tag" or "recognition" nucleotide sequence that is specifically bound by the sequence-specific endoribonuclease. In its enzymatically inactive state ("uninduced" state), the endoribonuclease can bind, but cannot cleave, the tagged target RNA. In its enzymatically active state ("induced" state) (e.g., in the presence of imidazole in a concentration of from about 100 mM to about 500 mM), the endoribonuclease can both bind and cleave the recognition nucleotide sequence in the tagged target nucleic acid, thereby releasing the target nucleic acid from the tag.

[0068] The binding solution can include a buffer and a salt; and lacks imidazole. The reactivation solution can include imidazole in a final concentration of from about 100 mM to about 500 mM, e.g., from about 100 mM to about 150 mM, from about 150 mM to about 200 mM, from about 250 mM to about 350 mM, from about 350 mM to about 400 mM, or from about 400 mM to about 500 mM. The presence of imidazole reactivates the sequence-specific, enzymatically inactive endoribonuclease such that the endoribonuclease becomes enzymatically active, e.g., the endoribonuclease exhibits at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, at least about 95%, or more than 95%, of wild-type sequence-specific endoribonuclease (e.g., an amino acid sequence as depicted in **Figure 5** (e.g., SEQ ID NO:6, 8, or 9)). As one non-limiting example, the sequence-specific, enzymatically inactive endoribonuclease is an H29A mutant of Csy4 (as described below; and as depicted in **Figure 6**); contacting the Csy4(H29A) mutant with imidazole, as described above, reactivates the endoribonuclease such that it is capable of cleaving, in a sequence-specific manner, a recognition sequence in a target ribonucleic acid. Also suitable for use is an H29A, S50C double mutant of Csy4 (as described below). In some embodiments, the "tag" or recognition sequence comprises the nucleotide sequence 5'-GUUCACUGCCGUAUAGGCAGCUAAGAAA-3' (SEQ ID NO:1).

[0069] The "tag" or "recognition" nucleotide sequence can be introduced into a nucleic acid using standard recombinant methods. Thus, the tagged target nucleic acid will include a tag that is enzymatically cleaved, thereby releasing the target nucleic acid.

[0070] In some embodiments, the tagged target nucleic acid (RNA) will have one or more polypeptides bound thereto. A tagged target RNA that has one or more polypeptides bound thereto is referred to herein as a RNA protein complex. Thus, in

some embodiments, the target RNA that is isolated using a subject method is an RNA protein complex. In some embodiments, a subject method can further comprise analyzing the polypeptide(s) bound to the isolated target RNA.

[0071] A subject method provides for isolation of a target RNA (or RNA protein complex). In some embodiments, a subject method provides for purification of a target RNA (or RNA protein complex) such that the target RNA (or RNA protein complex) is at least about 50% pure, at least about 60% pure, at least about 70% pure, at least about 80% pure, at least about 90% pure, at least about 95% pure, at least about 98% pure, or greater than 98% pure.

[0072] In some embodiments, a protein bound to a target RNA in a target RNA/protein complex can be eluted from the RNA/protein complex. The eluted protein can be further characterized, e.g., by sequencing, enzymatic digestion, a functional assay, etc.

[0073] The mixed population of nucleic acids can be present in a cell lysate. For example, an expression vector comprising a nucleotide sequence encoding a tagged target RNA is introduced into a cell (e.g., *in vitro* or *in vivo*), such that the cell synthesizes the tagged target RNA. A lysate is made from the cell and the lysate (optionally subjected to one or more steps to enrich for nucleic acids) is applied to the immobilized sequence-specific enzymatically-inactive endoribonuclease.

[0074] The sequence-specific enzymatically-inactive endoribonuclease can be immobilized on any of a variety of insoluble support. Suitable insoluble supports include, but are not limited to agarose beads, magnetic beads, a test strip, a multi-well dish, and the like. The insoluble support can comprise a variety of substances (glass, polystyrene, polyvinyl chloride, polypropylene, polyethylene, polycarbonate, dextran, nylon, amylose, natural and modified celluloses, polyacrylamides, agaroses, and magnetite) and can be provided in a variety of forms, including, e.g., agarose beads, polystyrene beads, latex beads, magnetic beads, colloid metal particles, glass and/or silicon chips and surfaces, nitrocellulose strips, nylon membranes, sheets, wells of reaction trays (e.g., multi-well plates), plastic tubes, etc.

[0075] The present disclosure also provides a method of isolating a polypeptide that binds a target RNA, where the method comprises: a) contacting an immobilized complex with a liquid solution comprising a polypeptide that binds the target RNA, where the immobilized complex comprises the variant Csy4 endoribonuclease and a tagged target

RNA comprising a recognition nucleotide sequence that is specifically bound by the variant Csy4 endoribonuclease, where said contacting results in binding of the polypeptide to the target RNA, where said contacting is carried out in a binding solution lacking imidazole; and b) eluting the bound polypeptide.

ENDORIBONUCLEASES

[0076] The present disclosure provides a sequence-specific endoribonuclease. In some embodiments, the present disclosure provides a sequence-specific endoribonuclease that binds to a recognition sequence in a target polyribonucleotide, but that does not cleave the target polyribonucleotide, i.e., the sequence-specific endoribonuclease is enzymatically inactive in hydrolyzing the target polyribonucleotide. In some embodiments, the present disclosure provides a sequence-specific endoribonuclease that binds to a recognition sequence in a target polyribonucleotide, and cleaves the target polyribonucleotide within or near the recognition sequence, i.e., the sequence-specific endoribonuclease is enzymatically active in hydrolyzing the target polyribonucleotide.

[0077] In some embodiments, a subject sequence-specific endoribonuclease is immobilized on an insoluble substrate. Suitable insoluble substrates include, but are not limited to agarose beads, magnetic beads, a test strip, a multi-well dish, and the like. The insoluble substrate can comprise a variety of substances (glass, polystyrene, polyvinyl chloride, polypropylene, polyethylene, polycarbonate, dextran, nylon, amylose, natural and modified celluloses, polyacrylamides, agaroses, and magnetite) and can be provided in a variety of forms, including, e.g., agarose beads, polystyrene beads, latex beads, magnetic beads, colloid metal particles, glass and/or silicon chips and surfaces, nitrocellulose strips, nylon membranes, sheets, wells of reaction trays (e.g., multi-well plates), plastic tubes, etc.

Enzymatically inactive sequence-specific endoribonuclease

[0078] The present disclosure provides an enzymatically inactive, sequence-specific endoribonuclease, wherein the enzymatically inactive sequence-specific endoribonuclease binds to a target sequence in a polyribonucleotide in a sequence-specific manner. A subject enzymatically inactive, sequence-specific endoribonuclease binds a target polyribonucleotide in a sequence-specific manner, but does not cleave the target polyribonucleotide. A subject enzymatically inactive, sequence-specific

endoribonuclease is useful for isolating a target RNA from a mixed population of nucleic acids, as described above.

[0079] In some embodiments, a subject enzymatically inactive, sequence-specific endoribonuclease comprises one or more amino acid substitutions compared to a naturally-occurring, enzymatically active, Csy4, CasE, or Cas6 polypeptide.

[0080] In some embodiments, a subject enzymatically inactive, sequence-specific endoribonuclease comprises an amino acid substitution at His-29 of a Csy4 polypeptide, or at an equivalent position in a CasE or a Cas6 polypeptide. In some embodiments, a subject enzymatically inactive, sequence-specific endoribonuclease comprises an amino acid substitution at Ser-148 of a Csy4 polypeptide, or at an equivalent position in a CasE or a Cas6 polypeptide.

[0081] **Figure 6** depicts non-limiting examples of suitable enzymatically inactive, sequence-specific endoribonuclease amino acid sequences. In some embodiments, a subject enzymatically inactive, sequence-specific endoribonuclease comprises an amino acid sequence having at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 98%, at least about 99%, or 100%, amino acid sequence identity with an amino acid sequence depicted in **Figure 6**, where the amino acid sequence includes a substitution at His-29, Ser-50, or both His-29 and Ser-50. For example, the variant Csy4 endoribonuclease can include a H29A (His-29 to Ala-29) substitution, a S50C (Ser-50 to Cys-50) substitution, or both a H29A and a S50C substitution.

[0082] In some embodiments, a subject enzymatically inactive, sequence-specific endoribonuclease is a variant Csy4 endoribonuclease. In some cases, a subject variant Csy4 endoribonuclease comprises an amino acid sequence having at least about 95% amino acid sequence identity to the amino acid sequence set forth in **Figure 6**, where the endoribonuclease comprises an amino acid substitution at His-29, where the variant Csy4 endoribonuclease is enzymatically inactive in the absence of imidazole, and where the variant Csy4 endoribonuclease is activatable in the presence of imidazole. In some instances, the amino acid substitution is a His29 to Ala29 substitution. In some cases, variant Csy4 endoribonuclease also includes a Ser-50 substitution. In some instances, a subject variant Csy4 endoribonuclease binds an RNA substrate that comprises the

nucleotide sequence 5'-GUUCACUGCCGUAUAGGCAGCUAAGAAA-3' (SEQ ID NO:1).

[0083] A subject enzymatically inactive, sequence-specific endoribonuclease is "conditionally" enzymatically inactive, e.g., a subject enzymatically inactive, sequence-specific endoribonuclease (e.g., a subject variant Csy4 endoribonuclease) is enzymatically inactive in the absence of imidazole; and the enzymatically inactive, sequence-specific endoribonuclease (e.g., subject variant Csy4 endoribonuclease) is activatable by imidazole. For example, the enzymatically inactive, sequence-specific endoribonuclease (e.g., subject variant Csy4 endoribonuclease) can be enzymatically activated by contacting the endoribonuclease with imidazole at a concentration of from about 100 mM to about 500 mM.

[0084] The presence of imidazole (e.g., in a concentration range of from about 100 mM to about 500 mM) reactivates the sequence-specific, enzymatically inactive endoribonuclease such that the endoribonuclease becomes enzymatically active, e.g., the endoribonuclease exhibits at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, at least about 95%, or more than 95%, of wild-type sequence-specific endoribonuclease (e.g., an amino acid sequence as depicted in **Figure 5** (e.g., SEQ ID NO:6, 8, or 9)).

[0085] In some embodiments, a subject enzymatically inactive, sequence-specific endoribonuclease (e.g., a subject variant Csy4 endoribonuclease) comprises a detectable label, including a moiety that provides a detectable signal. Suitable detectable labels and/or moieties that provide a detectable signal include, but are not limited to, an enzyme, a radioisotope, a member of a FRET pair, a member of a specific binding pair; a fluorophore; a fluorescent protein; a quantum dot; and the like.

[0086] FRET pairs (donor/acceptor) suitable for use include, but are not limited to, EDANS/fluorescein, IAEDANS/fluorescein, fluorescein/tetramethylrhodamine, fluorescein/Cy 5, IEDANS/DABCYL, fluorescein/QSY-7, fluorescein/LC Red 640, fluorescein/Cy 5.5 and fluorescein/LC Red 705. In addition, a fluorophore/quantum dot donor/acceptor pair can be used.

[0087] Suitable fluorophores ("fluorescent label") include any molecule that may be detected via its inherent fluorescent properties, which include fluorescence detectable upon excitation. Suitable fluorescent labels include, but are not limited to, fluorescein,

rhodamine, tetramethylrhodamine, eosin, erythrosin, coumarin, methyl-coumarins, pyrene, Malacite green, stilbene, Lucifer Yellow, Cascade Blue™, Texas Red, IAEDANS, EDANS, BODIPY FL, LC Red 640, Cy 5, Cy 5.5, LC Red 705 and Oregon green. Suitable optical dyes are described in the 2002 Molecular Probes Handbook, 9th Ed., by Richard P. Haugland, hereby expressly incorporated by reference.

[0088] Suitable enzymes include, but are not limited to, horse radish peroxidase, luciferase, β-galactosidase, and the like.

[0089] Suitable fluorescent proteins include, but are not limited to, a green fluorescent protein (GFP), e.g., a GFP from *Aequoria victoria* or a mutant or derivative thereof e.g., as described in U.S. Patent No. 6,066,476; 6,020,192; 5,985,577; 5,976,796; 5,968,750; 5,968,738; 5,958,713; 5,919,445; 5,874,304; a red fluorescent protein; a yellow fluorescent protein; any of a variety of fluorescent and colored proteins from Anthozoan species, as described in, e.g., Matz et al. (1999) *Nature Biotechnol.* 17:969-973; and the like.

[0090] Suitable nanoparticles include, e.g., quantum dots (QDs), fluorescent or luminescent nanoparticles, and magnetic nanoparticles. Any optical or magnetic property or characteristic of the nanoparticle(s) can be detected.

[0091] QDs and methods for their synthesis are well known in the art (see, e.g., U.S. Pat. Nos. 6,322,901; 6,576,291; and 6,815,064). QDs can be rendered water soluble by applying coating layers comprising a variety of different materials (see, e.g., U.S. Pat. Nos. 6,423,551; 6,251,303; 6,319,426; 6,426,513; 6,444,143; and 6,649,138). For example, QDs can be solubilized using amphiphilic polymers. Exemplary polymers that have been employed include octylamine-modified low molecular weight polyacrylic acid, polyethylene-glycol (PEG)-derivatized phospholipids, polyanhydrides, block copolymers, etc. QDs can be conjugated to a polypeptide via any of a number of different functional groups or linking agents that can be directly or indirectly linked to a coating layer (see, e.g., U.S. Pat. Nos. 5,990,479; 6,207,392; 6,251,303; 6,306,610; 6,325,144; and 6,423,551).

[0092] QDs with a wide variety of absorption and emission spectra are commercially available, e.g., from Quantum Dot Corp. (Hayward Calif.; now owned by Invitrogen) or from Evident Technologies (Troy, N.Y.). For example, QDs having peak emission wavelengths of approximately 525, 535, 545, 565, 585, 605, 655, 705, and 800 nm are

available. Thus the QDs can have a range of different colors across the visible portion of the spectrum and in some cases even beyond.

[0093] Suitable radioisotopes include, but are not limited to ^{14}C, ^{3}H, ^{32}P, ^{33}P, ^{35}S, $^{125}1$, and ^{131}I. The use of radioisotopes as labels is well known in the art.

[0094] In some embodiments, a subject enzymatically inactive, sequence-specific endoribonuclease (e.g., a subject variant Csy4 endoribonuclease) is immobilized on an insoluble substrate. Suitable insoluble substrates include, but are not limited to agarose beads, magnetic beads, a test strip, a multi-well dish, and the like. The insoluble substrate can comprise a variety of substances (glass, polystyrene, polyvinyl chloride, polypropylene, polyethylene, polycarbonate, dextran, nylon, amylose, natural and modified celluloses, polyacrylamides, agaroses, and magnetite) and can be provided in a variety of forms, including, e.g., agarose beads, polystyrene beads, latex beads, magnetic beads, colloid metal particles, glass and/or silicon chips and surfaces, nitrocellulose strips, nylon membranes, sheets, wells of reaction trays (e.g., multi-well plates), plastic tubes, etc.

[0095] In some embodiments, a subject enzymatically inactive, sequence-specific endoribonuclease (e.g., a subject variant Csy4 endoribonuclease) is purified, e.g., is at least 80% pure, at least 85% pure, at least 90% pure, at least 95% pure, at least 98% pure, at least 99% pure, or greater than 99% pure.

<u>Compositions</u>

[0096] The present disclosure provides compositions comprising a subject sequence-specific, enzymatically inactive, endoribonuclease. A subject composition can comprise, in addition to a subject sequence-specific, enzymatically inactive, endoribonuclease, one or more of: a salt, e.g., NaCl, MgCl, KCl, $MgSO_4$, etc.; a buffering agent, e.g., a Tris buffer, N-(2-Hydroxyethyl)piperazine-N'-(2-ethanesulfonic acid) (HEPES), 2-(N-Morpholino)ethanesulfonic acid (MES), 2-(N-Morpholino)ethanesulfonic acid sodium salt (MES), 3-(N-Morpholino)propanesulfonic acid (MOPS), N-tris[Hydroxymethyl]methyl-3-aminopropanesulfonic acid (TAPS), etc.; a solubilizing agent; a detergent, e.g., a non-ionic detergent such as Tween-20, etc.; a protease inhibitor; and the like.

Enzymatically active sequence-specific endoribonuclease

[0097] In some embodiments, a subject enzymatically active sequence-specific endoribonuclease comprises a moiety that provides for detection. For example, a subject enzymatically active sequence-specific endoribonuclease can comprise a covalently or non-covalently linked moiety that provides for detection.

[0098] Suitable detectable labels include any composition detectable by spectroscopic, photochemical, biochemical, immunochemical, electrical, optical or chemical means. Moieties that provide for detection include, but are not limited to, a fluorescent molecule; a quantum dot; an enzyme (other than the endoribonuclease), where the enzyme catalyzes conversion of a substrate to a detectable product, where the product is directly detectable; a nanoparticle; and the like.

[0099] Suitable fluorescent proteins that can be linked to a subject enzymatically active sequence-specific endoribonuclease include, but are not limited to, a green fluorescent protein (GFP), e.g., a GFP from *Aequoria victoria* or a mutant or derivative thereof e.g., as described in U.S. Patent No. 6,066,476; 6,020,192; 5,985,577; 5,976,796; 5,968,750; 5,968,738; 5,958,713; 5,919,445; 5,874,304; a red fluorescent protein; a yellow fluorescent protein; any of a variety of fluorescent and colored proteins from Anthozoan species, as described in, e.g., Matz et al. (1999) *Nature Biotechnol.* 17:969-973; and the like.

[00100] Suitable nanoparticles include, e.g., quantum dots (QDs), fluorescent or luminescent nanoparticles, and magnetic nanoparticles. Any optical or magnetic property or characteristic of the nanoparticle(s) can be detected.

[00101] QDs and methods for their synthesis are well known in the art (see, e.g., U.S. Pat. Nos. 6,322,901; 6,576,291; and 6,815,064). QDs can be rendered water soluble by applying coating layers comprising a variety of different materials (see, e.g., U.S. Pat. Nos. 6,423,551; 6,251,303; 6,319,426; 6,426,513; 6,444,143; and 6,649,138). For example, QDs can be solubilized using amphiphilic polymers. Exemplary polymers that have been employed include octylamine-modified low molecular weight polyacrylic acid, polyethylene-glycol (PEG)-derivatized phospholipids, polyanhydrides, block copolymers, etc. QDs can be conjugated to a polypeptide via any of a number of different functional groups or linking agents that can be directly or indirectly linked to a

coating layer (see, e.g., U.S. Pat. Nos. 5,990,479; 6,207,392; 6,251,303; 6,306,610; 6,325,144; and 6,423,551).

[00102] QDs with a wide variety of absorption and emission spectra are commercially available, e.g., from Quantum Dot Corp. (Hayward Calif.; now owned by Invitrogen) or from Evident Technologies (Troy, N.Y.). For example, QDs having peak emission wavelengths of approximately 525, 535, 545, 565, 585, 605, 655, 705, and 800 nm are available. Thus the QDs can have a range of different colors across the visible portion of the spectrum and in some cases even beyond.

[00103] In some embodiments, a subject enzymatically active, sequence-specific endoribonuclease is purified, e.g., is at least 80% pure, at least 85% pure, at least 90% pure, at least 95% pure, at least 98% pure, at least 99% pure, or greater than 99% pure.

Compositions

[00104] The present disclosure provides compositions comprising a subject sequence-specific, enzymatically active endoribonuclease. A subject composition can comprise, in addition to a subject sequence-specific enzymatically active, endoribonuclease, one or more of: a salt, e.g., NaCl, MgCl, KCl, $MgSO_4$, etc.; a buffering agent, e.g., a Tris buffer, N-(2-Hydroxyethyl)piperazine-N'-(2-ethanesulfonic acid) (HEPES), 2-(N-Morpholino)ethanesulfonic acid (MES), 2-(N-Morpholino)ethanesulfonic acid sodium salt (MES), 3-(N-Morpholino)propanesulfonic acid (MOPS), N-tris[Hydroxymethyl]methyl-3-aminopropanesulfonic acid (TAPS), etc.; a solubilizing agent; a detergent, e.g., a non-ionic detergent such as Tween-20, etc.; a protease inhibitor; and the like.

[00105] The present disclosure provides compositions comprising a subject sequence-specific, enzymatically inactive endoribonuclease (e.g., a subject variant Csy4 endoribonuclease). A subject composition can comprise, in addition to a subject sequence-specific enzymatically inactive endoribonuclease (e.g., a subject variant Csy4 endoribonuclease), one or more of: a salt, e.g., NaCl, MgCl, KCl, $MgSO_4$, etc.; a buffering agent, e.g., a Tris buffer, N-(2-Hydroxyethyl)piperazine-N'-(2-ethanesulfonic acid) (HEPES), 2-(N-Morpholino)ethanesulfonic acid (MES), 2-(N-Morpholino)ethanesulfonic acid sodium salt (MES), 3-(N-Morpholino)propanesulfonic acid (MOPS), N-tris[Hydroxymethyl]methyl-3-aminopropanesulfonic acid (TAPS), etc.; a solubilizing agent; a detergent, e.g., a non-ionic detergent such as Tween-20, etc.; a

protease inhibitor; and the like. In some embodiments, the composition lacks imidazole. In some embodiments, the composition comprises imidazole in a concentration of from about 100 mM to about 500 mM.

Methods of producing a subject sequence-specific endoribonuclease

[00106] A subject sequence-specific endoribonuclease (e.g., a subject sequence-specific enzymatically active, endoribonuclease; a subject sequence-specific enzymatically inactive, endoribonuclease) can be produced by any known method, e.g., conventional synthetic methods for protein synthesis; recombinant DNA methods; etc.

[00107] Where a subject sequence-specific endoribonuclease is chemically synthesized, the synthesis may proceed via liquid-phase or solid-phase. Solid phase polypeptide synthesis (SPPS), in which the C-terminal amino acid of the sequence is attached to an insoluble support followed by sequential addition of the remaining amino acids in the sequence, is an example of a suitable method for the chemical synthesis of a subject sequence-specific endoribonuclease. Various forms of SPPS, such as Fmoc and Boc, are available for synthesizing a subject sequence-specific endoribonuclease. Techniques for solid phase synthesis are described by Barany and Merrifield, Solid-Phase Peptide Synthesis; pp. 3-284 in The Peptides: Analysis, Synthesis, Biology. Vol. 2: Special Methods in Peptide Synthesis, Part A., Merrifield, et al. J. Am. Chem. Soc., 85: 2149-2156 (1963); Stewart et al., Solid Phase Peptide Synthesis, 2nd ed. Pierce Chem. Co., Rockford, Ill. (1984); and Ganesan A. 2006 *Mini Rev. Med Chem.* 6:3-10 and Camarero JA et al. 2005 *Protein Pept Lett.* 12:723-8.

[00108] Standard recombinant methods can be used for production of a subject sequence-specific endoribonuclease. For example, nucleic acids encoding a subject sequence-specific endoribonuclease are inserted into expression vectors. The DNA segments encoding a subject sequence-specific endoribonuclease are operably linked to control sequences in the expression vector(s) that ensure the expression of the encoded polypeptides. Expression control sequences include, but are not limited to, promoters (e.g., naturally-associated or heterologous promoters), signal sequences, enhancer elements, and transcription termination sequences. The expression control sequences can be eukaryotic promoter systems in vectors capable of transforming or transfecting eukaryotic host cells (e.g., COS or CHO cells). Once the vector has been incorporated into the appropriate host, the host is maintained under conditions suitable for high level

expression of the nucleotide sequences, and the collection and purification of the endoribonuclease.

NUCLEIC ACIDS AND HOST CELLS

[00109] The present disclosure provides a nucleic acid comprising a nucleotide sequence encoding a subject sequence-specific endoribonuclease (e.g., a subject sequence-specific, enzymatically active endoribonuclease; a subject sequence-specific, enzymatically inactive endoribonuclease). In some embodiments, the nucleic acid is an expression vector, where the expression vector can provide for production of the sequence-specific endoribonuclease, e.g., in a cell.

[00110] A nucleotide sequence encoding a subject sequence-specific endoribonuclease (e.g., a subject sequence-specific, enzymatically active endoribonuclease; a subject sequence-specific, enzymatically inactive endoribonuclease) can be operably linked to one or more regulatory elements, such as a promoter and enhancer, that allow expression of the nucleotide sequence in the intended target cells (e.g., a cell that is genetically modified to synthesize the encoded endoribonuclease).

[00111] In some embodiments, a subject nucleic acid comprises a nucleotide sequence encoding a polypeptide having at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 98%, at least about 99%, or 100%, with an amino acid sequence set forth in Figure 4 or Figure 5. In some embodiments, a subject nucleic acid comprises a nucleotide sequence encoding a variant Csy4 polypeptide, as described above.

[00112] A nucleotide sequence encoding a subject sequence-specific endoribonuclease (e.g., a subject sequence-specific, enzymatically active endoribonuclease; a subject sequence-specific, enzymatically inactive endoribonuclease) can be operably linked to a transcription control element (e.g., a promoter, an enhancer, etc.). Suitable promoter and enhancer elements are known in the art. For expression in a bacterial cell, suitable promoters include, but are not limited to, lacI, lacZ, T3, T7, gpt, lambda P and trc. For expression in a eukaryotic cell, suitable promoters include, but are not limited to, cytomegalovirus immediate early promoter; herpes simplex virus thymidine kinase promoter; early and late SV40 promoters; promoter present in long terminal repeats from a retrovirus; mouse metallothionein-I promoter; and various art-known tissue specific promoters.

[00113] In some embodiments, e.g., for expression in a yeast cell, a suitable promoter is a constitutive promoter such as an ADH1 promoter, a PGK1 promoter, an ENO promoter, a PYK1 promoter and the like; or a regulatable promoter such as a GAL1 promoter, a GAL10 promoter, an ADH2 promoter, a PHO5 promoter, a CUP1 promoter, a GAL7 promoter, a MET25 promoter, a MET3 promoter, a CYC1 promoter, a HIS3 promoter, an ADH1 promoter, a PGK promoter, a GAPDH promoter, an ADC1 promoter, a TRP1 promoter, a URA3 promoter, a LEU2 promoter, an ENO promoter, a TP1 promoter, and AOX1 (e.g., for use in *Pichia*). Selection of the appropriate vector and promoter is well within the level of ordinary skill in the art.

[00114] Suitable promoters for use in prokaryotic host cells include, but are not limited to, a bacteriophage T7 RNA polymerase promoter; a trp promoter; a lac operon promoter; a hybrid promoter, e.g., a lac/tac hybrid promoter, a tac/trc hybrid promoter, a trp/lac promoter, a T7/lac promoter; a trc promoter; a tac promoter, and the like; an araBAD promoter; *in vivo* regulated promoters, such as an *ssaG* promoter or a related promoter (*see*, e.g., U.S. Patent Publication No. 20040131637), a *pagC* promoter (Pulkkinen and Miller, *J. Bacteriol*., 1991: 173(1): 86-93; Alpuche-Aranda et al., PNAS, 1992; 89(21): 10079-83), a *nirB* promoter (Harborne et al. (1992) *Mol. Micro.* 6:2805-2813), and the like (*see*, e.g., Dunstan et al. (1999) *Infect. Immun.* 67:5133-5141; McKelvie et al. (2004) *Vaccine* 22:3243-3255; and Chatfield et al. (1992) *Biotechnol.* 10:888-892); a sigma70 promoter, e.g., a consensus sigma70 promoter (see, e.g., GenBank Accession Nos. AX798980, AX798961, and AX798183); a stationary phase promoter, e.g., a *dps* promoter, an *spv* promoter, and the like; a promoter derived from the pathogenicity island SPI-2 (*see*, e.g., WO96/17951); an actA promoter (*see*, e.g., Shetron-Rama et al. (2002) *Infect. Immun.* 70:1087-1096); an rpsM promoter (*see*, e.g., Valdivia and Falkow (1996). *Mol. Microbiol.* 22:367); a tet promoter (*see*, e.g., Hillen,W. and Wissmann,A. (1989) In Saenger,W. and Heinemann,U. (eds), *Topics in Molecular and Structural Biology, Protein–Nucleic Acid Interaction.* Macmillan, London, UK, Vol. 10, pp. 143–162); an SP6 promoter (*see*, e.g., Melton et al. (1984) *Nucl. Acids Res.* 12:7035); and the like. Suitable strong promoters for use in prokaryotes such as *Escherichia coli* include, but are not limited to Trc, Tac, T5, T7, and $P_{Lambda.}$ Non-limiting examples of operators for use in bacterial host cells include a lactose promoter operator (LacI repressor protein changes conformation when contacted with

lactose, thereby preventing the LacI repressor protein from binding to the operator), a tryptophan promoter operator (when complexed with tryptophan, TrpR repressor protein has a conformation that binds the operator; in the absence of tryptophan, the TrpR repressor protein has a conformation that does not bind to the operator), and a tac promoter operator (see, for example, deBoer et al. (1983) Proc. Natl. Acad. Sci. U.S.A. 80:21-25).

[00115] A nucleotide sequence encoding a subject sequence-specific endoribonuclease (e.g., a subject sequence-specific, enzymatically active endoribonuclease; a subject sequence-specific, enzymatically inactive endoribonuclease) can be present in an expression vector and/or a cloning vector. An expression vector can include a selectable marker, an origin of replication, and other features that provide for replication and/or maintenance of the vector.

[00116] Large numbers of suitable vectors and promoters are known to those of skill in the art; many are commercially available for generating a subject recombinant construct. The following vectors are provided by way of example. Bacterial: pBs, phagescript, PsiX174, pBluescript SK, pBs KS, pNH8a, pNH16a, pNH18a, pNH46a (Stratagene, La Jolla, Calif., USA); pTrc99A, pKK223-3, pKK233-3, pDR540, and pRIT5 (Pharmacia, Uppsala, Sweden). Eukaryotic: pWLneo, pSV2cat, pOG44, PXR1, pSG (Stratagene) pSVK3, pBPV, pMSG and pSVL (Pharmacia).

[00117] Expression vectors generally have convenient restriction sites located near the promoter sequence to provide for the insertion of nucleic acid sequences encoding heterologous proteins. A selectable marker operative in the expression host may be present. Suitable expression vectors include, but are not limited to, viral vectors (e.g. viral vectors based on vaccinia virus; poliovirus; adenovirus (see, e.g., Li et al., Invest Opthalmol Vis Sci 35:2543 2549, 1994; Borras et al., Gene Ther 6:515 524, 1999; Li and Davidson, PNAS 92:7700 7704, 1995; Sakamoto et al., H Gene Ther 5:1088 1097, 1999; WO 94/12649, WO 93/03769; WO 93/19191; WO 94/28938; WO 95/11984 and WO 95/00655); adeno-associated virus (see, e.g., Ali et al., Hum Gene Ther 9:81 86, 1998, Flannery et al., PNAS 94:6916 6921, 1997; Bennett et al., Invest Opthalmol Vis Sci 38:2857 2863, 1997; Jomary et al., Gene Ther 4:683 690, 1997, Rolling et al., Hum Gene Ther 10:641 648, 1999; Ali et al., Hum Mol Genet 5:591 594, 1996; Srivastava in WO 93/09239, Samulski et al., J. Vir. (1989) 63:3822-3828; Mendelson et al., Virol.

(1988) 166:154-165; and Flotte et al., PNAS (1993) 90:10613-10617); SV40; herpes simplex virus; human immunodeficiency virus (see, e.g., Miyoshi et al., PNAS 94:10319 23, 1997; Takahashi et al., J Virol 73:7812 7816, 1999); a retroviral vector (e.g., Murine Leukemia Virus, spleen necrosis virus, and vectors derived from retroviruses such as Rous Sarcoma Virus, Harvey Sarcoma Virus, avian leukosis virus, human immunodeficiency virus, myeloproliferative sarcoma virus, and mammary tumor virus); and the like.

[00118] The present disclosure provides isolated genetically modified host cells (e.g., *in vitro* cells) that are genetically modified with a subject nucleic acid. In some embodiments, a subject isolated genetically modified host cell can produce a subject sequence-specific endoribonuclease (e.g., a subject sequence-specific, enzymatically active endoribonuclease; a subject sequence-specific, enzymatically inactive endoribonuclease).

[00119] Suitable host cells include eukaryotic host cells, such as a mammalian cell, an insect host cell, a yeast cell; and prokaryotic cells, such as a bacterial cell. Introduction of a subject nucleic acid into the host cell can be effected, for example by calcium phosphate precipitation, DEAE dextran mediated transfection, liposome-mediated transfection, electroporation, or other known method.

[00120] Suitable mammalian cells include primary cells and immortalized cell lines. Suitable mammalian cell lines include human cell lines, non-human primate cell lines, rodent (e.g., mouse, rat) cell lines, and the like. Suitable mammalian cell lines include, but are not limited to, HeLa cells (e.g., American Type Culture Collection (ATCC) No. CCL-2), CHO cells (e.g., ATCC Nos. CRL9618, CCL61, CRL9096), 293 cells (e.g., ATCC No. CRL-1573), Vero cells, NIH 3T3 cells (e.g., ATCC No. CRL-1658), Huh-7 cells, BHK cells (e.g., ATCC No. CCL10), PC12 cells (ATCC No. CRL1721), COS cells, COS-7 cells (ATCC No. CRL1651), RAT1 cells, mouse L cells (ATCC No. CCLI.3), human embryonic kidney (HEK) cells (ATCC No. CRL1573), HLHepG2 cells, and the like.

[00121] Suitable yeast cells include, but are not limited to, *Pichia pastoris*, *Pichia finlandica*, *Pichia trehalophila*, *Pichia koclamae*, *Pichia membranaefaciens*, *Pichia opuntiae*, *Pichia thermotolerans*, *Pichia salictaria*, *Pichia guercuum*, *Pichia pijperi*, *Pichia stiptis*, *Pichia methanolica*, *Pichia* sp., *Saccharomyces cerevisiae*,

Saccharomyces sp., *Hansenula polymorpha*, *Kluyveromyces* sp., *Kluyveromyces lactis*, *Candida albicans*, *Aspergillus nidulans*, *Aspergillus niger*, *Aspergillus oryzae*, *Trichoderma reesei*, *Chrysosporium lucknowense*, *Fusarium* sp., *Fusarium gramineum*, *Fusarium venenatum*, *Neurospora crassa*, *Chlamydomonas reinhardtii*, and the like.

[00122] Suitable prokaryotic cells include, but are not limited to, any of a variety of laboratory strains of *Escherichia coli*, *Lactobacillus* sp., *Salmonella* sp., *Shigella* sp., and the like. See, e.g., Carrier et al. (1992) *J. Immunol.* 148:1176-1181; U.S. Patent No. 6,447,784; and Sizemore et al. (1995) *Science* 270:299-302. Examples of Salmonella strains which can be employed in the present invention include, but are not limited to, *Salmonella typhi* and *S. typhimurium.* Suitable Shigella strains include, but are not limited to, *Shigella flexneri*, *Shigella sonnei*, and *Shigella disenteriae*. Typically, the laboratory strain is one that is non-pathogenic. Non-limiting examples of other suitable bacteria include, but are not limited to, *Bacillus subtilis*, *Pseudomonas pudita*, *Pseudomonas aeruginosa*, *Pseudomonas mevalonii*, *Rhodobacter sphaeroides*, *Rhodobacter capsulatus*, *Rhodospirillum rubrum*, *Rhodococcus* sp., and the like. In some embodiments, the host cell is *Escherichia coli.*

KITS

[00123] The present disclosure also provides kits for determining the nucleotide sequence of a target polyribonucleotide. The present disclosure provides kits for carrying out sequence-specific cleavage of a substrate polyribonucleotide. The present disclosure provides kits for carrying out detection of an RNA sequence in a target polyribonucleotide. The present disclosure provides kits for carrying out isolation of a target RNA. The present disclosure provides kits for carrying out isolation of a polypeptide that binds a target RNA.

Kits for carrying out direct sequencing of a polyribonucleotide

[00124] A subject kit for carrying out direct sequencing of a polyribonucleotide includes at least a subject sequence-specific, enzymatically inactive endoribonuclease, where the sequence-specific, enzymatically inactive endoribonuclease is purified. In some embodiments, the enzymatically inactive, sequence-specific endoribonuclease is linked to an acceptor molecule or a donor molecule, for FRET detection.

[00125] A subject kit for carrying out direct sequencing of a polyribonucleotide includes at least a subject sequence-specific, enzymatically inactive endoribonuclease; and can

include one or more additional components, where the one or more additional components can be: 1) a buffer; 2) a probe oligonucleotide comprising a defined sequence; 3) a probe oligonucleotide comprising a defined sequence, where the probe oligonucleotide is linked to an acceptor molecule or a donor molecule, for FRET detection; 4) an insoluble support, for linking to a target polyribonucleotide; 5) a positive control polyribonucleotide, where the positive control polyribonucleotide comprises a known nucleotide sequence; 6) a positive control probe oligonucleotide that binds to and forms a duplex with the known sequence of the positive control polyribonucleotide.

[00126] In addition to above-mentioned components, a subject kit can further include instructions for using the components of the kit to practice the subject methods. The instructions for practicing the subject methods are generally recorded on a suitable recording medium. For example, the instructions may be printed on a substrate, such as paper or plastic, etc. As such, the instructions may be present in the kits as a package insert, in the labeling of the container of the kit or components thereof (i.e., associated with the packaging or subpackaging) etc. In other embodiments, the instructions are present as an electronic storage data file present on a suitable computer readable storage medium, e.g. CD-ROM, diskette, etc. In yet other embodiments, the actual instructions are not present in the kit, but means for obtaining the instructions from a remote source, e.g. via the internet, are provided. An example of this embodiment is a kit that includes a web address where the instructions can be viewed and/or from which the instructions can be downloaded. As with the instructions, this means for obtaining the instructions is recorded on a suitable substrate.

Kits for carrying out sequence-specific cleavage of a substrate polyribonucleotide

[00127] A subject kit for carrying out sequence-specific cleavage of a substrate polyribonucleotide includes at least a purified sequence-specific endoribonuclease and/or a nucleic acid comprising a nucleotide sequence encoding the sequence-specific endoribonuclease. A subject kit for carrying out sequence-specific cleavage of a substrate polyribonucleotide can include, in addition to a purified sequence-specific endoribonuclease (and/or a nucleic acid comprising a nucleotide sequence encoding the sequence-specific endoribonuclease), one or more additional components. Suitable additional components include, e.g., a buffer; a polyribonucleotide substrate that serves as a positive control; polyribonucleotide size standards; a negative control substrate; and

the like. The components can each be in separate containers. The kit can further include one or more positive and negative controls.

[00128] In addition to above-mentioned components, a subject kit can further include instructions for using the components of the kit to practice the subject methods. The instructions for practicing the subject methods are generally recorded on a suitable recording medium. For example, the instructions may be printed on a substrate, such as paper or plastic, etc. As such, the instructions may be present in the kits as a package insert, in the labeling of the container of the kit or components thereof (i.e., associated with the packaging or subpackaging) etc. In other embodiments, the instructions are present as an electronic storage data file present on a suitable computer readable storage medium, e.g. CD-ROM, diskette, etc. In yet other embodiments, the actual instructions are not present in the kit, but means for obtaining the instructions from a remote source, e.g. via the internet, are provided. An example of this embodiment is a kit that includes a web address where the instructions can be viewed and/or from which the instructions can be downloaded. As with the instructions, this means for obtaining the instructions is recorded on a suitable substrate.

<u>Kits for carrying out detection of a sequence in a target polyribonucleotide</u>

[00129] A subject kit for carrying out detection of a sequence in a target polyribonucleotide (e.g., for carrying out detection of a polyribonucleotide) can include an oligonucleotide probe comprising a known sequence. In some embodiments, the kit will include an oligonucleotide probe comprising a known sequence and comprising a detectable moiety, e.g., a polypeptide that can be detected using an immunological assay; a fluorescent protein; a luciferin; etc. The kit can further include a positive control polyribonucleotide that comprises a nucleotide sequence capable of forming a duplex with the oligonucleotide probe. The kit can further include an enzymatically active, sequence-specific endoribonuclease that specifically detects and cleaves a duplex formed by the oligonucleotide probe and a target polyribonucleotide. The kit can further include one or more of a buffer; components for detecting the detectable moiety; a test strip; and the like. The kit can further include one or more positive and negative controls.

[00130] In addition to above-mentioned components, a subject kit can further include instructions for using the components of the kit to practice the subject methods. The instructions for practicing the subject methods are generally recorded on a suitable

recording medium. For example, the instructions may be printed on a substrate, such as paper or plastic, etc. As such, the instructions may be present in the kits as a package insert, in the labeling of the container of the kit or components thereof (i.e., associated with the packaging or subpackaging) etc. In other embodiments, the instructions are present as an electronic storage data file present on a suitable computer readable storage medium, e.g. CD-ROM, diskette, etc. In yet other embodiments, the actual instructions are not present in the kit, but means for obtaining the instructions from a remote source, e.g. via the internet, are provided. An example of this embodiment is a kit that includes a web address where the instructions can be viewed and/or from which the instructions can be downloaded. As with the instructions, this means for obtaining the instructions is recorded on a suitable substrate.

Kits for carrying out isolation of a target RNA

[00131] A subject kit for carrying out isolation (e.g., purification) of a target RNA can include one or more of: 1) a subject sequence-specific, enzymatically inactive endoribonuclease; 2) an expression construct comprising a “tag” nucleotide sequence, i.e., a nucleotide sequence that is specifically bound by the sequence-specific, enzymatically inactive endoribonuclease, where a nucleotide sequence encoding a target RNA of choice can be inserted 3’ of the “tag” nucleotide sequence; and 3) imidazole. The sequence-specific, enzymatically inactive endoribonuclease can be immobilized on an insoluble support. The kit can further include a liquid composition for contacting a mixed population of nucleic acids with the immobilized sequence-specific, enzymatically inactive endoribonuclease. The kit can further include a wash buffer. The kit can further include one or more positive and negative controls. A positive control could include an expression vector comprising a nucleotide sequence encoding a tagged target RNA, where the tag is specifically bound by the sequence-specific, enzymatically inactive endoribonuclease. The components can each be in separate containers.

[00132] For example, a subject kit can include a subject sequence-specific, enzymatically inactive endoribonuclease. A subject kit can further include a recombinant expression vector comprising, in order from 5’ to 3’ and in operable linkage: a) a nucleotide sequence encoding an RNA substrate that is specifically bound by a subject variant Csy4 endoribonuclease; and b) a multiple cloning site suitable for insertion of a nucleic acid encoding the target RNA. The nucleotide sequence encoding the RNA substrate can be

operably linked to a promoter. In some instances, the promoter is an inducible promoter. The RNA substrate can comprise the nucleotide sequence 5'-GUUCACUGCCGUAUAGGCAGCUAAGAAA-3' (SEQ ID NO:1). In some cases, the recombinant expression vector comprises, inserted into the multiple cloning site, a nucleotide sequence encoding the target RNA. The kit can further include a buffer that lacks imidazole. The kit can further include imidazole or an imidazole solution. The kit can further include one or more wash buffers. In some cases, the kit will include a positive control expression vector. The variant Csy4 endoribonuclease can be immobilized on an insoluble support, where suitable insoluble supports include, but are not limited to agarose beads, magnetic beads, a test strip, a multi-well dish, and the like. The insoluble support can comprise a variety of substances (glass, polystyrene, polyvinyl chloride, polypropylene, polyethylene, polycarbonate, dextran, nylon, amylose, natural and modified celluloses, polyacrylamides, agaroses, and magnetite) and can be provided in a variety of forms, including, e.g., agarose beads, polystyrene beads, latex beads, magnetic beads, colloid metal particles, glass and/or silicon chips and surfaces, nitrocellulose strips, nylon membranes, sheets, wells of reaction trays (e.g., multi-well plates), plastic tubes, etc..

[00133] In addition to above-mentioned components, a subject kit can further include instructions for using the components of the kit to practice the subject methods. The instructions for practicing the subject methods are generally recorded on a suitable recording medium. For example, the instructions may be printed on a substrate, such as paper or plastic, etc. As such, the instructions may be present in the kits as a package insert, in the labeling of the container of the kit or components thereof (i.e., associated with the packaging or subpackaging) etc. In other embodiments, the instructions are present as an electronic storage data file present on a suitable computer readable storage medium, e.g. CD-ROM, diskette, etc. In yet other embodiments, the actual instructions are not present in the kit, but means for obtaining the instructions from a remote source, e.g. via the internet, are provided. An example of this embodiment is a kit that includes a web address where the instructions can be viewed and/or from which the instructions can be downloaded. As with the instructions, this means for obtaining the instructions is recorded on a suitable substrate.

METHODS OF DIRECTLY SEQUENCING A TARGET POLYRIBONUCLEOTIDE

[00134] The present disclosure provides a method of directly determining the nucleotide sequence of a target polyribonucleotide. Thus, for example, the method does not require synthesis of a polydeoxyribonucleotide counterpart of a target polyribonucleotide in order to determine the nucleotide sequence of the target polyribonucleotide.

[00135] Viral diagnostics, personalized medicine, single-cell transcript analysis, and translational profiling are all fields in which direct RNA detection and sequencing find use. A subject polyribonucleotide sequencing method, and a subject method of detecting a specific sequence in a polyribonucleotide, find use in these various fields.

[00136] A subject polyribonucleotide sequencing method generally involves: a) contacting a target polyribonucleotide with an oligonucleotide probe comprising a specific known sequence and an enzymatically inactive sequence-specific endoribonuclease under conditions that favor duplex formation between the oligonucleotide probe and the target polyribonucleotide, wherein the enzymatically inactive sequence-specific endoribonuclease binds the specific sequence in the duplex; and b) detecting specific binding between the oligonucleotide probe and the target polyribonucleotide, where specific binding of the enzymatically inactive sequence-specific endoribonuclease to the duplex indicates the presence of the specific sequence in the target polyribonucleotide.

[00137] In some cases, the enzymatically inactive sequence-specific endoribonuclease is linked (covalently or non-covalently) to an emissive label. By "emissive label" is meant any molecule that may be detected via its inherent emission properties, which include emission detectable upon excitation. Suitable emissive labels include, but are not limited to, fluorescein, rhodamine, tetramethylrhodamine, eosin, erythrosin, coumarin, methyl-coumarins, pyrene, Malacite green, stilbene, Lucifer Yellow, Cascade Blue™, Texas Red, IAEDANS, EDANS, BODIPY FL, LC Red 640, Cy 5, Cy 5.5, LC Red 705 and Oregon green. Suitable optical dyes are described in the 2002 Molecular Probes Handbook, 9th Ed., by Richard P. Haugland.

[00138] In some instances, the oligonucleotide probe used in a subject polyribonucleotide sequencing method is linked to a donor molecule, the enzymatically inactive sequence-specific endoribonuclease is linked to an acceptor molecule, and detection of duplex

formation is by fluorescence resonance energy transfer (also referred to as "Förster resonance energy transfer" or "FRET").

[00139] Förster resonance energy transfer (FRET) is phenomenon known in the art wherein excitation of one emissive dye is transferred to another without emission of a photon. A FRET pair consists of a donor chromophore and an acceptor chromophore (where the acceptor chromophore may be a quencher molecule). The emission spectrum of the donor and the absorption spectrum of the acceptor must overlap, and the two molecules must be in close proximity. The distance between donor and acceptor at which 50% of donors are deactivated (transfer energy to the acceptor) is defined by the Förster radius, which is typically 10-100 angstroms. Changes in the emission spectrum comprising FRET pairs can be detected, indicating changes in the number of that are in close proximity (i.e., within 100 angstroms of each other). This will typically result from the binding or dissociation of two molecules, one of which is labeled with a FRET donor and the other of which is labeled with a FRET acceptor, wherein such binding brings the FRET pair in close proximity.

[00140] Binding of such molecules will result in an increased emission of the acceptor and/or quenching of the fluorescence emission of the donor. FRET pairs (donor/acceptor) suitable for use include, but are not limited to, EDANS/fluorescein, IAEDANS/fluorescein, fluorescein/tetramethylrhodamine, fluorescein/Cy 5, IEDANS/DABCYL, fluorescein/QSY-7, fluorescein/LC Red 640, fluorescein/Cy 5.5 and fluorescein/LC Red 705. In addition, a fluorophore/quantum dot donor/acceptor pair can be used. EDANS is (5-((2-Aminoethyl)amino)naphthalene-1-sulfonic acid); IAEDANS is 5-({2-[(iodoacetyl)amino]ethyl}amino)naphthalene-1-sulfonic acid); DABCYL is 4-(4-dimethylaminophenyl) diazenylbenzoic acid.

[00141] Cy3, Cy5, Cy 5.5, and the like, are cyanines. For example, Cy3 and Cy5 are reactive water-soluble fluorescent dyes of the cyanine dye family. Cy3 dyes are red (~550 nm excitation, ~570 nm emission and therefore appear green), while Cy5 is fluorescent in the red region (~650/670 nm) but absorbs in the orange region (~649 nm). Alexa Fluor dyes, Dylight, IRIS Dyes, Seta dyes, SeTau dyes, SRfluor dyes and Square dyes dyes can also be used.

[00142] In another aspect of FRET, an emissive donor molecule and a nonemissive acceptor molecule ("quencher") may be employed. In this application, emission of the

donor will increase when quencher is displaced from close proximity to the donor and emission will decrease when the quencher is brought into close proximity to the donor. Useful quenchers include, but are not limited to, DABCYL, QSY 7 and QSY 33. Useful fluorescent donor/quencher pairs include, but are not limited to EDANS/DABCYL, Texas Red/DABCYL, BODIPY/DABCYL, Lucifer yellow/DABCYL, coumarin/DABCYL and fluorescein/QSY 7 dye.

[00143] In some cases, the enzymatically inactive sequence-specific endoribonuclease is linked (covalently or non-covalently) to a label enzyme. By “label enzyme” is meant an enzyme which may be reacted in the presence of a label enzyme substrate which produces a detectable product. Suitable label enzymes also include optically detectable labels (e.g., in the case of horse radish peroxidase (HRP)). Suitable label enzymes include but are not limited to, HRP, alkaline phosphatase, luciferase, β-galactosidase, and glucose oxidase. Methods for the use of such substrates are well known in the art. The presence of the label enzyme is generally revealed through the enzyme’s catalysis of a reaction with a label enzyme substrate, producing an identifiable product. Such products may be opaque, such as the reaction of horseradish peroxidase with tetramethyl benzedine, and may have a variety of colors. Other label enzyme substrates, such as Luminol (available from Pierce Chemical Co.), have been developed that produce fluorescent reaction products. Methods for identifying label enzymes with label enzyme substrates are well known in the art and many commercial kits are available. Examples and methods for the use of various label enzymes are described in Savage et al., Previews 247:6-9 (1998), Young, J. Virol. Methods 24:227-236 (1989).

[00144] In some cases, the enzymatically inactive sequence-specific endoribonuclease comprises a radioisotope. By “radioisotope” is meant any radioactive molecule. Suitable radioisotopes for use in the invention include, but are not limited to ^{14}C, ^{3}H, ^{32}P, ^{33}P, ^{35}S, $^{125}1$, and ^{131}I. The use of radioisotopes as labels is well known in the art.

[00145] In some cases, the enzymatically inactive sequence-specific endoribonuclease is linked (covalently or non-covalently) to a member of a specific binding pair (“partner of a binding pair”). By “partner of a binding pair” or “member of a binding pair” is meant one of a first and a second moiety, wherein the first and the second moiety have a specific binding affinity for each other. Suitable binding pairs include, but are not limited to, antigen/antibodies (for example, digoxigenin/anti-digoxigenin, dinitrophenyl

(DNP)/anti-DNP, dansyl-X-anti-dansyl, fluorescein/anti-fluorescein, lucifer yellow/anti-lucifer yellow, and rhodamine anti-rhodamine), biotin/avidin (or biotin/streptavidin) and calmodulin binding protein (CBP)/calmodulin.

[00146] In some embodiments, the oligonucleotide probe comprises a modification that provides for increased resistance to non-specific hydrolysis. Such modifications are well known in the art and include, e.g., nuclease-resistant internucleosidic linkages, modified backbones, base modifications, base substitutions, sugar modifications, and the like.

[00147] Suitable modified oligonucleotide backbones containing a phosphorus atom therein include, for example, phosphorothioates, chiral phosphorothioates, phosphorodithioates, phosphotriesters, aminoalkylphosphotriesters, methyl and other alkyl phosphonates including 3'-alkylene phosphonates, 5'-alkylene phosphonates and chiral phosphonates, phosphinates, phosphoramidates including 3'-amino phosphoramidate and aminoalkylphosphoramidates, phosphorodiamidates, thionophosphoramidates, thionoalkylphosphonates, thionoalkylphosphotriesters, selenophosphates and boranophosphates having normal 3'-5' linkages, 2'-5' linked analogs of these, and those having inverted polarity wherein one or more internucleotide linkages is a 3' to 3', 5' to 5' or 2' to 2' linkage. Suitable oligonucleotides having inverted polarity comprise a single 3' to 3' linkage at the 3'-most internucleotide linkage i.e. a single inverted nucleoside residue which may be a basic (the nucleobase is missing or has a hydroxyl group in place thereof). Various salts (such as, for example, potassium or sodium), mixed salts and free acid forms are also included.

[00148] A modified oligonucleotide can comprise one or more phosphorothioate and/or heteroatom internucleoside linkages, in particular $-CH_2-NH-O-CH_2-$, $-CH_2-N(CH_3)-O-CH_2-$ (known as a methylene (methylimino) or MMI backbone), $-CH_2-O-N(CH_3)-CH_2-$, $-CH_2-N(CH_3)-N(CH_3)-CH_2-$ and $-O-N(CH_3)-CH_2-CH_2-$ (wherein the native phosphodiester internucleotide linkage is represented as $-O-P(=O)(OH)-O-CH_2-$). MMI type internucleoside linkages are disclosed in the above referenced U.S. Pat. No. 5,489,677. Suitable amide internucleoside linkages are disclosed in t U.S. Pat. No. 5,602,240.

[00149] A modified oligonucleotide can comprise one or more morpholino backbone structures as described in, e.g., U.S. Pat. No. 5,034,506. For example, in some embodiments, a modified oligonucleotide comprises a 6-membered morpholino ring in

place of a ribose ring. In some of these embodiments, a phosphorodiamidate or other non-phosphodiester internucleoside linkage replaces a phosphodiester linkage. Morpholino nucleic acids ("morpholinos") include bases bound to morpholine rings instead of deoxyribose rings; in addition, the phosphate backbone can include a non-phosphate group, e.g., a phosphorodiamidate group instead of phosphates. Summerton (1999) *Biochim. Biophys. Acta* 1489:141; Heasman (2002) *Dev. Biol.* 243:209; Summerton and Weller (1997) *Antisense & Nucl. Acid Drug Dev.* 7:187; Hudziak et al. (1996) *Antisense & Nucl. Acid Drug Dev.* 6:267; Partridge et al. (1996) *Antisense & Nucl. Acid Drug Dev.* 6:169; Amantana et al. (2007) *Bioconj. Chem.* 18:1325; Morcos et al. (2008) *BioTechniques* 45:616.

[00150] A modified oligonucleotide can comprise a modified backbone. Modified polynucleotide backbones that do not include a phosphorus atom therein have backbones that are formed by short chain alkyl or cycloalkyl internucleoside linkages, mixed heteroatom and alkyl or cycloalkyl internucleoside linkages, or one or more short chain heteroatomic or heterocyclic internucleoside linkages. These include those having morpholino linkages (formed in part from the sugar portion of a nucleoside); siloxane backbones; sulfide, sulfoxide and sulfone backbones; formacetyl and thioformacetyl backbones; methylene formacetyl and thioformacetyl backbones; riboacetyl backbones; alkene containing backbones; sulfamate backbones; methyleneimino and methylenehydrazino backbones; sulfonate and sulfonamide backbones; amide backbones; and others having mixed N, O, S and CH_2 component parts.

[00151] A modified oligonucleotide can comprise one or more substituted sugar moieties. Suitable oligonucleotides comprise a sugar substituent group selected from: OH; F; O-, S-, or N-alkyl; O-, S-, or N-alkenyl; O-, S- or N-alkynyl; or O-alkyl-O-alkyl, wherein the alkyl, alkenyl and alkynyl may be substituted or unsubstituted C_1 to C_{10} alkyl or C_2 to C_{10} alkenyl and alkynyl. Also suitable are $O((CH_2)_nO)_mCH_3$, $O(CH_2)_nOCH_3$, $O(CH_2)_nNH_2$, $O(CH_2)_nCH_3$, $O(CH_2)_nONH_2$, and $O(CH_2)_nON((CH_2)_nCH_3)_2$, where n and m are from 1 to about 10. Other suitable oligonucleotides comprise a sugar substituent group selected from: C_1 to C_{10} lower alkyl, substituted lower alkyl, alkenyl, alkynyl, alkaryl, aralkyl, O-alkaryl or O-aralkyl, SH, SCH_3, OCN, Cl, Br, CN, CF_3, OCF_3, $SOCH_3$, SO_2CH_3, ONO_2, NO_2, N_3, NH_2, heterocycloalkyl, heterocycloalkaryl, aminoalkylamino, polyalkylamino, substituted silyl, an RNA cleaving group, a reporter

group, an intercalator, and the like. A suitable modification includes 2'-methoxyethoxy (2'-O-CH_2 CH_2OCH_3, also known as 2'-O-(2-methoxyethyl) or 2'-MOE) (Martin et al., Helv. Chim. Acta, 1995, 78, 486-504) i.e., an alkoxyalkoxy group. A further suitable modification includes 2'-dimethylaminooxyethoxy, i.e., a $O(CH_2)_2ON(CH_3)_2$ group, also known as 2'-DMAOE, and 2'-dimethylaminoethoxyethoxy (also known in the art as 2'-O-dimethyl-amino-ethoxy-ethyl or 2'-DMAEOE), i.e., 2'-O-CH_2-O-CH_2-$N(CH_3)_2$.

[00152] A modified oligonucleotide can comprise one or more nucleobase (often referred to in the art simply as "base") modifications or substitutions. As used herein, "unmodified" or "natural" nucleobases include the purine bases adenine (A) and guanine (G), and the pyrimidine bases thymine (T), cytosine (C) and uracil (U). Modified nucleobases include other synthetic and natural nucleobases such as 5-methylcytosine (5-me-C), 5-hydroxymethyl cytosine, xanthine, hypoxanthine, 2-aminoadenine, 6-methyl and other alkyl derivatives of adenine and guanine, 2-propyl and other alkyl derivatives of adenine and guanine, 2-thiouracil, 2-thiothymine and 2-thiocytosine, 5-halouracil and cytosine, 5-propynyl (-C=C-CH_3) uracil and cytosine and other alkynyl derivatives of pyrimidine bases, 6-azo uracil, cytosine and thymine, 5-uracil (pseudouracil), 4-thiouracil, 8-halo, 8-amino, 8-thiol, 8-thioalkyl, 8-hydroxyl and other 8-substituted adenines and guanines, 5-halo particularly 5-bromo, 5-trifluoromethyl and other 5-substituted uracils and cytosines, 7-methylguanine and 7-methyladenine, 2-F-adenine, 2-amino-adenine, 8-azaguanine and 8-azaadenine, 7-deazaguanine and 7-deazaadenine and 3-deazaguanine and 3-deazaadenine. Further modified nucleobases include tricyclic pyrimidines such as phenoxazine cytidine(1H-pyrimido(5,4-b)(1,4)benzoxazin-2(3H)-one), phenothiazine cytidine (1H-pyrimido(5,4-b)(1,4)benzothiazin-2(3H)-one), G-clamps such as a substituted phenoxazine cytidine (e.g. 9-(2-aminoethoxy)-H-pyrimido(5,4-(b) (1,4)benzoxazin-2(3H)-one), carbazole cytidine (2H-pyrimido(4,5-b)indol-2-one), pyridoindole cytidine (H-pyrido(3',2':4,5)pyrrolo(2,3-d)pyrimidin-2-one).

[00153] Heterocyclic base moieties may also include those in which the purine or pyrimidine base is replaced with other heterocycles, for example 7-deaza-adenine, 7-deazaguanosine, 2-aminopyridine and 2-pyridone.

[00154] A suitable enzymatically inactive sequence-specific endoribonuclease includes an enzymatically inactive sequence-specific endoribonuclease described hereinbelow.

For example, an enzymatically inactive sequence-specific endoribonuclease as depicted in **Figure 6** can be used.

[00155] In some embodiments, the target polyribonucleotide is linked (covalently or non-covalently) to a solid support (an insoluble support). Suitable insoluble supports include, but are not limited to, beads, plates (e.g., multi-well plates), strips, etc., where the insoluble support can comprise various materials including, but not limited to, polystyrene, polypropylene, agarose, and the like.

[00156] Oligonucleotide probes ("detection oligonucleotide") can be RNA, DNA, or any chemically modified version of an RNA or DNA, e.g., peptide nucleic acids (PNAs), locked nucleic acids (LNAs), and the like.

[00157] A subject polyribonucleotide sequencing method can include one or more washing steps, e.g., to remove non-specifically bound components such as non-specifically bound oligonucleotide probes, any non-specifically bound detectable moieties, and the like.

[00158] A non-limiting example of how to carry out a subject polyribonucleotide sequencing method is as follows. A target polyribonucleotide bound to a solid support. The target polyribonucleotide is of unknown sequence and is the "RNA to be sequenced." Four oligonucleotide probes of four different known nucleotide sequences each comprise a different fluorophore (fluorophores 1-4). The fluorophores are members of FRET pairs. The counterpart members of the FRET pairs are quantum dots. The quantum dot is linked to an enzymatically inactive sequence-specific endoribonuclease. The enzymatically inactive sequence-specific endoribonuclease binds, but does not cleave, the duplex formed between an oligonucleotide probe and the target polyribonucleotide. Only one of the four oligonucleotide probes binds to and forms a duplex with the target polyribonucleotide. A washing step removes any unbound oligonucleotide probes. Binding of oligonucleotide probe-fluorophore2 results in duplex formation with the target polyribonucleotide. Fluorophore2 is thus brought into proximity to the quantum dot linked to the enzymatically inactive sequence-specific endoribonuclease, and fluorescence is quenched.

METHODS OF CLEAVING A POLYRIBONUCLEOTIDE

[00159] The present disclosure provides a method of cleaving a polyribonucleotide in a sequence-specific manner. The method generally involves contacting a substrate

polyribonucleotide with an enzymatically active sequence-specific endoribonuclease (e.g., a Csy4 endoribonuclease) under conditions that favor sequence-specific cleavage of the polyribonucleotide substrate. A subject method of cleaving a polyribonucleotide in a sequence-specific manner can be used to: 1) remove an affinity tag from a substrate polyribonucleotide; 2) to generate a population of product polyribonucleotides having homogeneity at the 5' end, e.g., where the substrate polyribonucleotides are *in vitro* transcribed mRNAs; and 3) to regulate gene expression in a cell *in vitro* or *in vivo*.

Substrate polyribonucleotides

[00160] The terms "substrate polyribonucleotide" and "target polyribonucleotide" are used interchangeably herein to refer to a polyribonucleotide that is bound by a sequence-specific endoribonuclease in a sequence-specific manner. A substrate polyribonucleotide can be single stranded. In some instances, a substrate polyribonucleotide is double stranded.

[00161] An endoribonuclease binds to and cleaves a substrate polyribonucleotide in a sequence-specific manner. Thus, for example, an endoribonuclease binds to and cleaves a substrate polyribonucleotide at a specific sequence, referred to herein as a "recognition sequence" or a "recognition site."

[00162] A recognition sequence can be a tetranucleotide sequence, a pentanucleotide sequence, a hexanucleotide sequence, a heptanucleotide sequence, an octanucleotide sequence, or longer than an octanucleotide. For example, in some embodiments, the recognition sequence is 9 ribonucleotides, 10 ribonucleotides, 11 ribonucleotides, 12 ribonucleotides, 13 ribonucleotides, 14 ribonucleotides, 15 ribonucleotides, 16 ribonucleotides, 17 ribonucleotides, 18 ribonucleotides, 19 ribonucleotides, or 20 ribonucleotides in length. In some embodiments, a sequence-specific endoribonuclease cleaves immediately 5' of a recognition sequence. In some embodiments, a sequence-specific endoribonuclease cleaves immediately 3' of a recognition sequence. In some embodiments, a sequence-specific endoribonuclease cleaves within a recognition sequence. In some cases, a recognition sequence is immediately 5' of a secondary structure. In some cases, a recognition sequence is located 5' of a secondary structure and within 1 nucleotide (nt), 2 nt, 3 nt, 4 nt, 5 nt, or 5 nt to 10 nt of the secondary structure. In some cases, a recognition sequence is immediately 3' of a secondary structure. In some cases, a recognition sequence is located 3' of a secondary structure

and within 1 nucleotide (nt), 2 nt, 3 nt, 4 nt, 5 nt, or 5 nt to 10 nt of the secondary structure.

[00163] In some embodiments, a substrate polyribonucleotide comprises the structure $X_xX_2X_3X_4X_5X_6X_7X_8X_9X_{10}X_{11}X_{12}X_{13}X_{14}X_{15}$, where nucleotides X_1-X_5 base pair with X_{11}-X_{15} such that X_1 and X_{15} form the base of a stem structure, and such that X_6, X_7, X_8, X_9, and X_{10} form a loop; the structure is a regular A-form helical structure.

[00164] In some embodiments, the substrate polyribonucleotide comprises an affinity tag; and a subject method provides for removal of the affinity tag from the substrate polyribonucleotide.

Sequence-specific endoribonucleases

[00165] Endoribonucleases that bind to and cleave a substrate polyribonucleotide in a sequence-specific manner include enzymatically active polypeptides that cleave (hydrolyze) a substrate polyribonucleotide in a metal ion-independent fashion.

[00166] Structural features of an endoribonuclease that binds to and cleaves a substrate polyribonucleotide in a sequence-specific and metal ion-independent manner can include one or more of the following: 1) a highly basic alpha helix for sequence non-specific recognition of the phosphate backbone of RNA through the RNA major groove, e.g., R114, R115, R118, R119, or equivalents thereof; 2) R102 and/or Q104, or equivalents thereof, making hydrogen bonding contacts with the major groove of the RNA stem; 3) and one or more of His29, Ser148, and Tyr176, or equivalents thereof, involved in catalysis; and 4) F155, or an equivalent thereof.

[00167] Endoribonucleases that bind to and cleave a substrate polyribonucleotide in a sequence-specific manner include enzymatically active polypeptides that have at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 98%, at least about 99%, or 100%, amino acid sequence identity to an amino acid sequence set forth in **Figure 4** (Csy4 amino acid sequences).

[00168] Endoribonucleases that bind to and cleave a substrate polyribonucleotide in a sequence-specific manner include enzymatically active polypeptides that have at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 98%, at least about 99%, or 100%, amino acid sequence identity to an amino acid sequence set forth in **Figure 5** (Csy4 amino acid sequences).

[00169] Endoribonucleases that bind to and cleave a substrate polyribonucleotide in a sequence-specific manner include enzymatically active polypeptides that have at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 98%, at least about 99%, or 100%, amino acid sequence identity to a Cas6 amino acid sequence.

[00170] Endoribonucleases that bind to and cleave a substrate polyribonucleotide in a sequence-specific manner include enzymatically active polypeptides that have at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 98%, at least about 99%, or 100%, amino acid sequence identity to a CasE amino acid sequence.

[00171] Endoribonucleases that bind to and cleave a substrate polyribonucleotide in a sequence-specific manner include enzymatically active polypeptides that differ from an amino acid sequence set forth in any one of **Figures 4 or 5** by from 1 to 20 (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20) amino acid substitutions and/or insertions and/or deletions.

Reaction conditions

[00172] A sequence-specific endoribonuclease can hydrolyze a substrate polyribonucleotide in a sequence-specific manner at a temperature in a range from about 15°C to about 100 °C, e.g., in a range of from about 15°C to about 17°C, from about 17°C to about 20°C, from about 20°C to about 25°C, from about 25°C to about 30°C, from about 30°C to about 40°C, from about 40°C to about 50°C, from about 50°C to about 60°C, from about 60°C to about 70°C, from about 70°C to about 80°C, from about 80°C to about 90°C, or from about 90°C to about 100°C.

[00173] A sequence-specific endoribonuclease can hydrolyze a substrate polyribonucleotide in a sequence-specific manner in a pH range of from about 4.0 to about 8.0, e.g., from about pH 4.0 to about 4.5, from about pH 4.5 to about 5.0, from about pH 5.0 to about 5.5, from about pH 5.5 to about 6.0, from about pH 6.0 to about 6.5, from about pH 6.5 to about 7.0, from about pH 7.0 to about 7.5, from about pH 6.5 to about 7.5, from about pH 7.5 to about 8.0, from about pH 6.5 to about 8.0, or from about pH 5.5 to about 7.5.

EXAMPLES

[00174] The following examples are put forth so as to provide those of ordinary skill in the art with a complete disclosure and description of how to make and use the present invention, and are not intended to limit the scope of what the inventors regard as their invention nor are they intended to represent that the experiments below are all or the only experiments performed. Efforts have been made to ensure accuracy with respect to numbers used (e.g. amounts, temperature, etc.) but some experimental errors and deviations should be accounted for. Unless indicated otherwise, parts are parts by weight, molecular weight is weight average molecular weight, temperature is in degrees Celsius, and pressure is at or near atmospheric. Standard abbreviations may be used, e.g., bp, base pair(s); kb, kilobase(s); pl, picoliter(s); s or sec, second(s); min, minute(s); h or hr, hour(s); aa, amino acid(s); kb, kilobase(s); bp, base pair(s); nt, nucleotide(s); i.m., intramuscular(ly); i.p., intraperitoneal(ly); s.c., subcutaneous(ly); and the like.

Example 1: Direct RNA detection and sequencing using Csy4 family proteins

MATERIALS AND METHODS

[00175] Wild-type Csy4, point mutants and selenomethionine (SeMet)-substituted Csy4 were expressed in Rosetta 2(DE3) cells as either a His_6-maltose binding protein (MBP) fusion or a His_6 fusion protein and purified by Ni-affinity chromatography, followed by proteolytic removal of the His(MBP) tag, a further Ni-affinity step, and size exclusion chromatography. The pre-crRNAs were transcribed *in vitro* with T7 polymerase and purified on a denaturing gel. The complex was formed by incubating RNA with Csy4 at a 2:1 ratio for 30 minutes at 30°C followed by size exclusion chromatography. The complex was crystallized using the hanging-drop method in 200mM sodium citrate pH 5.0, 100mM magnesium chloride, 20% (w/v) poly(ethylene glycol) (PEG)-4000 (wild-type (WT) complex) or 150mM sodium acetate pH 4.6, 17% PEG4000 or 160mM sodium acetate pH 4.6, 18% PEG4000 (S22C-containing complex). The structure of the WT Csy4-RNA complex was determined by the multiwavelength anomalous dispersion (MAD) method using SeMet-substituted crystals. The structure of the Csy4(S22C)-RNA complex was determined by molecular replacement.

[00176] **Gene annotation, cloning, protein expression and purification.** Comparative sequence analysis of Csy4 genes across species identified a conserved region 20 codons

upstream of the annotated start codon in the PA14 genome. Lee, *et al. Genome Biol* **7**, R90 (2006). The conserved Csy4 (PA14_33300) sequence was PCR amplified from *Pseudomonas aeruginosa* UCBPP-PA14 genomic DNA using Pa14Csy4_fwd: caccatggaccactacctcgacattcg and Pa14Csy4_rev: gaaccagggaacgaaacctcc. The polymerase chain reaction (PCR) product was cloned using the Gateway system into the pENTR/TEV/D-TOPO entry vector (Invitrogen), followed by site-specific recombination into expression vector pHGWA or pHMGWA. Busso, et al. *Analytical Biochemistry* **343**, 313-321, (2005). Point mutations were introduced into Csy4 using the QuikChange Site-Directed Mutagenesis Kit (Stratagene). The Pa14Csy4 expression plasmid was transform into *E. coli* Rosetta 2 (DE3) cells (Novagen) or co-transformed with a pMK vector expressing CRISPR RNA synthesized by Geneart (Regensburg, Germany). Rosetta 2 (DE3) cells were grown in Luria Broth (LB) supplemented with ampicillin and chloramphenicol. Protein expression was induced with 0.5mM isopropyl β-D-1-thiogalactopyranoside (IPTG) (Affymetrix) at a cell density of ~0.5OD followed by shaking at 18°C for 16 hours. Cells were pelleted and resuspended in lysis buffer (15.5mM disodium hydrogen phosphate, 4.5mM sodium dihydrogen phosphate, 500mM sodium chloride, 10mM imidazole, protease inhibitors, 5% glycerol, 0.01% Triton X-100, 100μ/ml DNaseI, 1mM Tris[2-carboxyethyl] phosphine hydrochloride (TCEP), 0.5mM phenylmethylsulfonyl fluoride, pH 7.4) and sonicated on ice for two minutes in 10 second bursts. Lysate was clarified by centrifugation (24,000 x g, 30 minutes) and incubated with nickel-nitrilotriacetic acid (Ni-NTA) affinity resin in batch (Qiagen). The bound protein was eluted with high imidazole buffer (15.5mM disodium hydrogen phosphate, 4.5mM sodium dihydrogen phosphate, 500mM sodium chloride, 300mM imidazole, 1mM TCEP, 5% glycerol, pH 7.4) and dialyzed overnight in dialysis buffer (elution buffer with only 20mM imidazole) in the presence of tobacco etch virus (TEV) to cleave the His_6 or His_6MBP tag. The protein was concentrated (Amicon) and purified on a nickel affinity column (GE) followed by tandem Sup75 (16/60) columns in gel filtration buffer (100mM HEPES pH7.5, 500mM KCl, 5% glycerol, 1mM TCEP). Sample was then dialyzed against gel filtration buffer containing only 150mM potassium chloride. A similar protocol was used for preparation of the selenomethionine (SeMet)-derivitized protein and the only notable difference was the expression media. Briefly, BL21(DE3) cells transformed with Csy4(pHGWA) expression vector were grown in M9

minimal media supplemented with ampicillin, as previously described. Wiedenheft, *et al. Structure* **17**, 904–912 (2009).

[00177] **Nuclease activity assays.** 75pmol of wild-type or mutant Csy4 were incubated with 5pmol *in vitro* transcribed Pa14 pre-crRNA (prepared as described; Wiedenheft (2009) *supra*) in 10μl reactions containing 20mM HEPES pH 7.5, 100mM potassium chloride buffer at 25°C for five minutes. Reactions were quenched with the addition of 50ul acid phenol-chloroform (Ambion). 10μl additional reaction buffer were added and samples were centrifuged (16,000 x g, 30 minutes) and 16μl aqueous sample was removed, mixed 1:1 with 2X formamide loading buffer, and separated on 15% denaturing polyacrylamide gel. RNA was visualized with SYBR Gold staining (Invitrogen).

[00178] **Crystallization.** All crystallization experiments were performed at 18 °C using the hanging drop vapour diffusion method by mixing equal volumes (1 μl + 1 μl) of the complex and reservoir solutions. Plate-shaped crystals of the wild-type Csy4-RNA complex were grown in 200mM sodium citrate pH 5.0, 100mM magnesium chloride, 20% (w/v) poly(ethylene glycol)-4000 (PEG4000). These crystals belonged to the space group *C*2, contained one copy of the complex in the asymmetric unit and diffracted to 2.3 Å resolution at synchrotron X-ray sources. Using complex reconstituted with the Csy4S22C point mutant, two additional crystal forms were obtained in 150mM sodium acetate pH 4.6, 17% (w/v) PEG4000 and 160mM sodium acetate pH 4.6, 18% PEG4000. Initially, hexagonal crystals appeared within 24 hr. These crystals diffracted to 2.6 Å resolution, belonged to space group $P6_1$ and contained one copy of the complex in the asymmetric unit. 48 hr later, the same crystallization condition yielded needle-shaped crystals that belonged to space group *P*212121, contained two copies of the complex and diffracted up to 1.8 Å resolution. For data collection, all crystal forms were cryoprotected by soaking in their respective mother liquor supplemented with 30% glycerol prior to flash cooling in liquid nitrogen.

[00179] **Structure determination.** All diffraction data we collected at 100 K on beamlines 8.2.2 and 8.3.1 of the Advanced Light Source (Lawrence Berkeley National Laboratory). Data were processed using XDS. Kabsch, *Acta Crystallogr D Biol Crystallogr* **66**, 125-132 (2010). Experimental phases were determined from a three-wavelength multiwavelength anomalous dispersion (MAD) experiment (peak, inflection

and remote data sets) using the monoclinic Csy4-RNA crystals containing selenomethionine-substituted wild-type Csy4. Two selenium sites were located using the Hybrid Substructure Search (HySS) module of the Phenix package. Grosse-Kunstleve, and Adams. *Acta Crystallogr D Biol Crystallogr* **59**, 1966-1973 (2003). Substructure refinement, phasing and density modification were performed using AutoSHARP. Vonrhein, et al. *Methods Mol Biol* **364**, 215-230 (2007). The resulting electron density map exhibited clear layers of density attributable to protein and RNA alternating along the *c*-axis, with the RNA layer made up of two coaxially-stacked RNA helices engaged in a "kissing loop" interaction. An initial atomic model for the Csy4 protein was obtained by automatic building using the Phenix AutoBuild module. Terwilliger, *et al. Acta Crystallogr D Biol Crystallogr* **64**, 61-69, (2008). The complex model was completed by iterative cycles of manual building in COOT (Emsley, and Cowtan, *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132 (2004)) and refinement using Phenix.refine[36] (Adams, *et al. Acta Crystallogr D Biol Crystallogr* **66**, 213-221 (2010)) against a native 2.33 Å resolution dataset, yielding a final model with a crystallographic R_{work} factor of 21.4% and a R_{free} factor of 26.4% (**Table 1**).

Table 1 Data collection, phasing and refinement statistics

	Native WT	Native S22C	Native S22C	SeMet WT		
Data collection						
Space group	*C*2	$P2_12_12_1$	$P6_1$		*C*2	
Cell dimensions						
a, b, c (Å)	62.37, 72.77, 86.82	40.1, 78.9, 145.9	39.25, 39.25, 297.37		62.33, 47.23, 87.26	
α, β, γ (°)	90.0, 108.2, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 120.0		90.0, 108.3, 90.0	
				Peak	*Inflection*	*Remote*
Wavelength (Å)	1.11159	0.99992	1.11588	0.97949	0.97971	0.97204
Resolution (Å)*	19.68-2.33 (2.50-2.33)	69.4-1.80 (1.90-1.80)	22.38-2.60 (2.70-2.60)	82.86-2.80 (2.90-2.80)	82.86-2.80 (2.90-2.80)	82.86-2.80 (2.90-2.80)
R_{sym} (%)*	5.8 (44.6)	7.0 (52.8)	3.3 (31.1)	9.4 (38.3)	8.9 (38.5)	9.0 (38.1)
$I/\sigma I$*	18.9 (3.35)	31.1 (3.1)	29.8 (3.8)	17.0 (4.4)	14.5 (3.7)	14.5 (3.6)
Completeness (%)*	96.6 (98.3)	98.7 (91.0)	99.4 (98.5)	99.6 (96.7)	99.5 (99.3)	99.3 (96.4)
Redundancy*	4.4 (4.4)	19.8 (6.5)	6.1 (5.4)	5.7 (5.3)	3.8 (3.7)	3.8 (3.7)
Refinement						
Resolution (Å)	19.70-2.33	69.4-1.80	19.60-2.60			
No. reflections	9974	43284	7798			
R_{work} / R_{free}	0.214 / 0.265	0.187/0.220	0.255 / 0.279			
No. atoms						
Protein	1273	2975	1364			
RNA	313	642	321			
Water/ligands	41	386	5			
B-factors						

Table 1 Data collection, phasing and refinement statistics

	Native WT	Native S22C	Native S22C	SeMet WT
Protein	47.7	29.1	101.5	
RNA	109.3	35.3	103.0	
Water/ligands	44.9	33.5	74.5	
R.m.s. deviations				
Bond lengths (Å)	0.007	0.011	0.002	
Bond angles (°)	1.0	1.5	0.7	

* Values in parentheses denote highest resolution shell

[00180] The model includes RNA nucleotides C1-G15 and the phosphate group of nucleotide C16 and protein residues 1-104, 109-120 and 139-187. Owing to the layered arrangement of protein and RNA in the crystal lattice and the lack of lateral crystal contacts within the RNA layer, the RNA exhibits significant disorder, as evidenced by markedly elevated temperature factors (>100 $Å^2$) and the absence of interpretable density for the nucleotide base of U9. The disorder is also evident in protein residues 109-120, corresponding to the arginine-rich helix inserted in the major groove of the RNA, for which only the polypeptide backbone could be built (except for residues Arg 115 and Arg 118).

[00181] The structures of the Csy4(S22C)-RNA complex in the hexagonal and orthorhombic crystal forms were determined by molecular replacement in Phaser (McCoy, *et al. J Appl Crystallogr* **40**, 658-674 (2007)), using the Csy4 protein (lacking the arginine-rich helix) and RNA models from the monoclinic crystal form as separate search ensembles. In both crystal forms, electron density for the arginine-rich helix and the linker region comprising Csy4 residues 105-108 was immediately noticeable in $2F_o$-F_c maps obtained from the molecular replacement solutions. The structure of the Csy4(S22C)-RNA complex in the hexagonal form was refined to an R_{work} factor of 25.5% and R_{free} of 27.9 at 2.6 Å resolution. The final model includes Csy4 residues 1-120 and 139-187 and RNA nucleotides C1-G15 plus the phosphate group of nucleotide C16. The orthorhombic crystal form of the Csy4(S22C)-RNA complex has been solved at 1.8 Å resolution and refined to an R_{work} factor of 18.7% and R_{free} of 22.0%, with excellent stereochemistry. Of the two complexes in the asymmetric unit, complex 1 (chains A and C) contains Csy4 residues 1-187 and RNA nucleotides C1-G15 plus the phosphate group of nucleotide C16, while the less ordered complex 2 (chains B and D) comprises Csy4 residues 1-187 with the exception of residues 13-15 and 135-138, which show no ordered electron density, and RNA nucleotides C1-G15 and the phosphate group of nucleotide C16. The two copies of Csy4 superpose with an rmsd of 1.15 Å over 179 $C\alpha$ atoms, the greatest differences coming from the slightly different positions of the arginine-rich helix. The two RNA molecules in the asymmetric unit superpose with an rmsd of 1.49 Å, the largest deviation being due to the bulged-out nucleotide U9, which assumes different conformations in the two RNAs. Our discussion and

illustrations throughout the manuscript are based on complex 1 of the orthorhombic crystal form. All structural illustrations were generated using Pymol (http://www(dot)pymol(dot)org).

RESULTS

[00182] CRISPR-mediated immunity is thought to occur in approximately 90% of archaeal and 40% of bacterial genomes based on the presence of CRISPR loci in sequenced genomes. Horvath and Barrangou, *Science* **327**, 167-170 (2010); Jansen, et al. *Molecular Microbiology* **43**, 1565-1575 (2002); Sorek, et al. *Nat Rev Microbiol* **6**, 181-186 (2008); Marraffini, and Sontheimer, *Nat Rev Genet* **11**, 181-190 (2010). CRISPR-associated (Cas) proteins belonging to the eight known CRISPR/Cas subtypes are highly divergent at the primary sequence level, obscuring identification of functional homologues. Haft, et al. *PLoS Comput Biol* **1**, e60 (2005); Makarova, et al. *Biology Direct* **1**, 1-26 (2006). *Pseudomonas aeruginosa* UCBPP-PA14 (hereafter Pa14), a Gram-negative opportunistic pathogen harboring a CRISPR/Cas system of the *Yersinia* subtype, contains six Cas genes flanked by two CRISPR elements (**Fig. 1A**). Although Cas1 is found universally among CRISPR-containing organisms, and Cas3 is evident in most subtypes, Csy1-4 are unique to the *Yersinia* subtype. Both CRISPR elements comprise a characteristic arrangement of 28-nucleotide repeats identical within both CRISPRs (save for one nucleotide) interspersed with ~32-nucleotide unique spacers, some of which match sequences found in bacteriophage or plasmids. Grissa, et al. *BMC Bioinformatics* **8**, 172 (2007). In many organisms it has been shown that CRISPR loci are transcribed as long single units and are post-transcriptionally processed to yield crRNAs that each contain one unique sequence flanked by sequences derived from the repeat element. Brouns *et al. Science* **321**, 960-964 (2008); Carte, et al. *Genes and Development* **22**, 3489-3496 (2008); Tang, *et al.. Proc. Natl. Acad. Sci. USA* **99**, 7536-7541 (2002); Lillestol, et al. *Archaea* **2**, 59–72 (2006); Lillestol,*et al. Mol Microbiol* **72**, 259-272 (2009); Tang, *et al. Molecular Microbiology* **55**, 469-481 (2005).

[00183] To identify the protein(s) responsible for producing crRNAs from long CRISPR transcripts (pre-crRNAs) in the *Yersinia* subtype, each of the six Cas proteins from Pa14 was recombinantly expressed, and the recombinantly expressed proteins were tested for endoribonucleolytic function using an *in vitro* transcribed pre-crRNA. Based on sequence-specific pre-crRNA processing activity, it was found that Csy4 is the

endoribonuclease responsible for crRNA biogenesis. As observed for crRNA processing within two other CRISPR/Cas subtypes (Brouns et al. (2008) *supra*; Carte et al. (2008) *supra*), CRISPR transcript cleavage is a rapid, metal ion-independent reaction. Csy4 cleaves pre-crRNA within the repeat element at the base of a predicted stem-loop structure, generating ~60 nucleotide crRNAs consisting of a 32-nucleotide unique (phage-derived) sequence flanked on the 5' and 3' ends by eight and 20 nucleotides, respectively, of repeat sequence (**Fig. 1A**).

[00184] For Csy4 to be effective, it was hypothesized that its RNA recognition mechanism must be highly specific in order to target only CRISPR-derived transcripts and not other cellular RNAs containing hairpins and/or related sequences. To test this, Csy4 was expressed in *E. coli* alone or co-expressed with a Pa14 CRISPR RNA. In spite of a high isoelectric point (PI=10.2), Csy4 does not associate with cellular nucleic acids; however, when co-expressed with a Pa14 CRISPR, the protein is associated with a crRNA (**Fig. 1B,C**). These observations underscored the specificity of Csy4 recognition, leading us to explore the protein/RNA interactions required for Csy4 substrate recognition and cleavage. Csy4 binding and activity assays were performed *in vitro* using RNA oligonucleotides corresponding to different regions of the 28-nucleotide Pa14 CRISPR repeat sequence. Using this approach, a minimal RNA fragment recognized by Csy4 consisting of the repeat-derived stem-loop and one downstream nucleotide was identified. Cleavage assays utilizing this minimal RNA as a substrate showed that Csy4 activity requires a 2'OH on the ribose immediately upstream of the cleavage site. A 2'-deoxyribose at this position completely abrogates cleavage, but does not disrupt Csy4 binding.

[00185] In order to obtain structural insights into crRNA recognition and cleavage, Csy4 was co-crystallized in complex with a minimal RNA substrate. To generate a stable complex for structural analysis, Csy4 was bound to the non-cleavable 16-nucleotide minimal RNA substrate described above in which the nucleotide preceding the cleavage site is a 2'-deoxynucleotide. Crystals of the complex were obtained in three unique space groups, each exhibiting different crystal packing; one contained wild-type Csy4 and two contained a Csy4 point mutant . The crystal structure of the Csy4-RNA complex was solved to a resolution of 1.8 Å (**Fig. 2A**, **Table 1**), revealing an unanticipated mechanism by which CRISPR RNA is recognized and processed for use by the

CRISPR-mediated silencing machinery. Csy4 makes sequence-specific contacts in the major groove of the stem-loop of the CRISPR repeat sequence and additional sequence non-specific contacts with the phosphate backbone of the RNA stem. The majority of characterized protein/RNA interactions are mediated via the minor groove of an RNA helix; the recognition of the RNA major groove by Csy4 is a highly unusual mechanism of protein/RNA interaction.

[00186] At the primary sequence level, Csy4 is highly dissimilar from the other known endoribonucleases involved in crRNA biogenesis (CasE from *Thermus thermophiles* (Ebihara, *et al. Protein Sci* **15**, 1494-1499 (2006)) and Cas6 from *Pyrococcus furiosus* Carte et al. (2008) *supra*), sharing only ~10% identity. The crystal structures of both CasE and Cas6 indicate that these proteins adopt tandem ferrodoxin-like folds. Notably, Csy4 shares this fold with these enzymes; in the Csy4-RNA complex, the N-terminal domain (residues 1-94) of Csy4 indeed adopts a ferredoxin-like fold. However, although the C-terminal domain (residues 95-187) shares the same secondary structure connectivity as a ferredoxin-like fold, but its conformation is markedly different. Strikingly, an arginine-rich helix (residues 108-120) from the putative C-terminal ferredoxin domain inserts into the major groove of the hairpin RNA. Structural superpositions using the DALI server (Holm, and Sander, *J Mol Biol* **233**, 123-138 (1993)) indicate that Csy4 in its RNA-binding conformation superposes with CasE and Cas6 with root-mean-square deviation (rmsd) of 3.8 Å (over 111 Ca atoms) and 3.9 Å (over 104 Ca atoms), respectively. Csy4, CasE and Cas6 could be descendants of a single ancestral endoribonuclease that has diverged markedly at the sequence level as it co-evolved with the repeat sequence of the CRISPR locus, while maintaining a similar protein fold.

[00187] The crRNA substrate forms a hairpin structure, as predicted for this subclass of crRNA repeats (Kunin, et al. *Genome Biol* **8**, R61 (2007)), with nucleotides 1-5 and 11-15 base pairing to produce a regular A-form helical stem. The GUAUA pentaloop contains a sheared G6-A10 base pair and a bulged-out nucleotide U9, its structure reminiscent of GNR(N)A pentaloops found in the yeast U6 small nuclear RNA intramolecular stem-loop (Huppler, et al. *Nat Struct Biol* **9**, 431-435 (2002)) and in bacteriophage lamda BoxB RNA (Legault, et al. *Cell* **93**, 289-299 (1998)). In the Csy4-RNA complex, the RNA stem-loop straddles the β-hairpin formed by strands β7-β8 of

Csy4, with the C1-G15 base-pair directly stacking onto the aromatic side chain of Phe155 (**Fig. 2B**). This anchors the RNA stem and orients it at the proper angle to permit sequence-specific interactions in the major groove.

[00188] Two residues in a linker segment connecting the body of Csy4 to the arginine-rich helix, Arg102 and Gln104, make hydrogen bonding contacts in the major groove of the RNA stem, sequence-specifically recognizing G15 and A14, respectively (**Fig. 2B**). The Csy4-crRNA interaction is further stabilized by the insertion of the arginine-rich helix into the major groove of the RNA hairpin in the proximity of the bulged-out nucleotide U9 (**Fig. 2C**). The side chains of Arg 114, Arg 115, Arg 118 and Arg119 contact the phosphate groups of nucleotides 2-6. Additionally, the sidechain of Arg115 engages the base of G6 as the only sequence-specific interaction between the arginine-rich helix and the RNA hairpin. Interestingly, this interaction is highly reminiscent of how certain viral proteins interact with the major groove of dsRNA molecules, for example the Tat/Tar interaction in human immunodeficiency virus (HIV)[23] and the lambda-N/boxB complex in lambdoid phages (Cai, *et al. Nature Structural Biology* **5**, 203-212 (1998)). In both cases, a highly basic α-helix is employed for sequence non-specific recognition with the phosphate backbone of RNA through the RNA major groove.

[00189] Csy4 recognizes the hairpin element of the CRISPR repeat sequence and cleaves immediately downstream of it. The structure described in this Example contains a substrate-mimic RNA, which is not competent for cleavage. In the active site, density was observed only for the phosphate group 3' of the penultimate nucleotide, but no density for the terminal sugar or base, presumably due to the flexibility of this nucleotide (**Fig. 3A**). The scissile phosphate binds in a pocket located between the β-turn of the β7-β8 hairpin on one side and helix α1 and a glycine-rich loop, previously identified in Cas6 and CasE, on the other. Three residues proximal to that phosphate group are likely to participate in catalysis, His29, Ser148 and Tyr176. These residues are invariant among 12 Csy4 sequences that were identified using a BLAST search (Altschul, *et al. Nucleic Acids Research* **25**, 3389-3402 (1997)) coupled with manual verification of a nearby CRISPR locus (Grissa, et al. *BMC Bioinformatics* **8**, 172 (2007)) (**Fig. 4**).

[00190] The structure suggests that several residues in Csy4 are important for mediating substrate recognition/binding and catalysis. Point mutants of each of these residues were

generated; their cleavage activity was tested biochemically (**Fig. 3B**). Mutation of putative catalytic site residues His29 or Ser148 abolishes cleavage activity. However, mutation of Tyr176 to phenylalanine does not disrupt activity, indicating that Tyr176 may play a crucial role in orienting His29, though it does not directly participate in catalysis. Mutation of Arg102 to alanine abolishes accumulation of crRNAs, whereas mutation of Gln104 to alanine does not significantly disrupt activity, suggesting that Arg102, which recognizes the terminal base pair, is important for properly orienting the RNA substrate, but that Gln104 is not required for *in vitro* activity. Phe155 appears to play a large role in appropriately orienting the RNA substrate, as an alanine mutation at this residue severely impairs crRNA biogenesis.

[00191] The identification of a serine involved in mediating RNA cleavage is unexpected. Although mutation of His29 to alanine results in a catalytically inactive Csy4, mutation to lysine partially restores activity, strongly suggesting that His29 acts as a proton donor, not to initiate cleavage via a nucleophilic attack.

[00192] CRISPRs are the genetic memory of a nucleic acid-based immune system that relies on small CRISPR-derived RNAs for guiding the immune system to cognate sequences associated with invading genetic elements. Phylogenetic analysis of CRISPR repeat sequences has identified distinct CRISPR categories (Kunin, et al. *Genome Biol* **8**, R61 (2007)) that correlate with a particular set of Cas genes. The co-variation of Cas genes with specific CRISPR repeat sequence types suggests that CRISPR repeats have co-evolved with the Cas genes that are responsible for CRISPR adaptation, the generation of crRNAs and the silencing of invading genetic elements. The structure described here details an unusual recognition mechanism that discriminates crRNA substrates based on both sequence- and structure-specificity, providing great insight into the ability of Csy4 and its homologues to readily distinguish substrate RNA from among all cellular RNAs.

[00193] **Figures 1A-C.** Pa14Csy4 specifically recognizes only its pre-crRNA substrate. **a,** Schematic of CRISPR/Cas locus in Pa14. The six Cas genes are flanked on both sides by CRISPR loci. Enlarged is a schematic showing the predicted stem-loop in the 28-nucleotide direct repeat (black lettering) separated by 32-nucleotide spacer sequences (blue). The red arrows note the bond cleaved by Csy4. **b,c** Comparison of protein (**b**) and RNA content (**c**) after Pa14Csy4 expression in *E. coli* with (+) and without (-) a

plasmid containing a Pa14 CRISPR locus. Purified Csy4 from both preparations was split into two pools. Half were resolved on SDS-PAGE and visualized with Coomassie blue staining; half were acid phenol-chloroform extracted, resolved on UREA-PAGE, and visualized with SYBR Gold (Invitrogen).

[00194] **Figures 2A-C.** The crystal structure of Csy4 bound to RNA substrate. **a,** Front and back views of the complex. Csy4 is colored in blue and the RNA backbone is colored in orange. **b,** Detailed interactions between residues R102 and Q104 and nucleotides A14 and G15. Hydrogen bonding is depicted with dashed lines. **c,** Detailed interactions between an arginine-rich alpha helix and the RNA backbone and G6.

[00195] **Figures 3A and 3B.** Putative active site. **a,** Detailed view of the catalytic center. **b,** Cleavage activity of Csy4. Wild-type (WT) Csy4 and a series of single point mutants were incubated with *in vitro* transcribed pre-crRNA for 5 minutes at 25 C. Products were acid phenol-chloroform extracted and resolved on UREA-PAGE and visualized by SYBR Gold staining.

<u>Example 2</u>: Direct RNA sequencing

[00196] An RNA can be sequenced at the single-molecule level using Förster Resonant Energy Transfer (FRET). The RNA to be sequenced will be attached to a solid surface through its 3' ribose. The RNA should be spaced far enough from neighboring RNA molecules on the surface to allow detection at the single-molecule level. The spacing is dictated by diffraction-limited methods, dependent on the wavelength of emitted light. Alternatively, the RNA spacing can be closer than the diffraction limit, if super-resolution imaging methods are used. In the first sequence detection step, a Csy4 family protein of known nucleic acid binding specificity is added to the RNA to be sequenced, along with a pool of detection oligonucleotides. The Csy4 protein will only bind to the RNA to be sequenced if one of the detection oligonucleotides can form a 4 base pair double helix with the RNA to be sequenced. In addition, the detection nucleotide must base pair with an additional 3 nucleotides 3' of the 4 base pair recognition sequence in the RNA to be sequenced, in order for the Csy4 protein to bind stably. The detection oligonucleotides will contain an extension of 3 nucleotides 3' of the 4-nucleotide recognition sequence. In the pool of detection oligonucleotides, the 3-nucleotide extension will have a defined 5' nucleotide followed by two random nucleotide positions; or a random nucleotide at the 5' position followed by a defined nucleotide and

a random nucleotide; or 2 random nucleotides at the 5' end, followed by a defined nucleotide. In any of these pools, the defined nucleotide is known based on an attached fluorescent molecule, the emission or excitation spectrum of which is defined by the nucleotide. The Csy4 protein will be attached to a quantum dot whose excitation spectrum overlaps with the emission spectrum of the fluorescent molecule attached to the detection oligonucleotide. After binding of detection oligonucleotides and Csy4, excess reagents will be washed away. A positive binding event is detected only if the detection nucleotide forms a 7-nucleotide double helix with the RNA to be sequenced. If binding occurs, the resulting ternary complex of RNA to be sequenced, detection oligonucleotide, and Csy4 protein can be detected by FRET from the fluorescent molecule attached to the detection oligonucleotide to the quantum dot attached to the Csy4 protein. After each cycle of binding, the Csy4 protein and detection oligonucleotides will be removed from the sample using chemical and/or heat denaturation and washing. In subsequent sequencing steps, other Csy4 proteins of different sequence specificity and their corresponding detection oligonucleotides will be incubated with the RNA to be sequenced, in a similar manner. Other variations of the 3-nucleotide extension on the detection oligonucleotide can be envisioned, such as extensions of different lengths, at either the 5' end or 3' end of the detection oligonucleotide. The detection oligonucleotide could be RNA, DNA, or any chemically modified version of these polymers, such as PNAs or LNAs.

Example 3: Inducible sequence-specific endoribonuclease

[00197] Via biochemical and structural techniques, point mutants of Csy4 that lack cleavage activity, while retaining substrate binding activity, have been generated. An example is the above-described Csy4(H29A) mutant. The otherwise catalytically inactive Csy4(H29A) mutant can be reactivated in the presence of exogenous imidazole. Addition of between 150mM and 300mM imidazole to the reaction buffer is sufficient to stimulate near-wild type cleavage activity. The results are shown in **Figure 8**. **Figure 8** shows a cleavage activity assay depicting the imidazole rescue. Csy4H29A is a catalytically inactive mutant of Csy4 that retains the ability to bind its substrate with a kd of < 1 nM.

[00198] Reaction details: Each 10 μl reaction contains 5pmol of the *in vitro* transcribed pre-crRNA substrate, 100pmol of Csy4 (WT or H29A, as indicated in **Figure 8**), 20mM

HEPES pH 7.5, 100mM KCl, and 150-300mM imidazole, as indicated. Reactions were carried out for 30 minutes at 25°C.) Products were acid phenol-chloroform extracted, separated on a 15% denaturing gel, and visualized with SYBR Gold. Biochemical characterization of Csy4(H29A) shows that it binds to its RNA substrate with <1 nM affinity.

[00199] Csy4(H29A) is useful for both *in vivo* and *in vitro* applications for which there is no current alternative approach.

[00200] Csy4(H29A) (also referred to herein as "inducible" Csy4), is useful for purifying a particular RNA/protein complex (RNP) from a complex mixture of RNAs and RNPs (RNA/protein complexes). For example, researchers may be interested in understanding which proteins bind to a particular RNA transcript. Using this system, the researchers could engineer an expression construct for their RNA of choice that would include a 5' tag consisting of the stem-loop Csy4 target sequence. The researchers would then transfect this expression construct into their cell type of choice, leading to the generation of many RNAs and RNPs. Cells would then be lysed and the lysate would be applied to a column that contains inducible Csy4 immobilized on agarose beads. RNAs or RNPs that have the Csy4 target sequence will bind. A subsequent wash step will remove non-specifically bound RNAs. A wash with imidazole (~300mM) will activate inducible Csy4, which will cleave the target sequence and release the bound RNA/RNP. This method is illustrated schematically in **Figure 9**.

[00201] A similar method could be useful for assembling RNPs *in vitro*. For example, an RNA of choice could be transcribed *in vitro* using a construct similar to the expression plasmid designed for the above experiment. (The construct must introduce the Csy4 stem-loop target sequence at the 5' end of the transcribed RNA.) This *in vitro* transcribed product could then be incubated with proteins known or suspected to bind the particular transcript. The inducible Csy4-containing column could be used to purify the *in vitro* formed RNPs away from free protein.

Example 4

[00202] The mechanism for specific substrate recognition by the endoribonuclease CasE, an essential component of the CRISPR immune system found in the majority of bacteria and archaea (van der Oost *et al.*, *Trends Biochem Sci.* **34**, 401-7 (2009)) has been determined. Using structural and biochemical methods, the minimal RNA sequence

required for optimal substrate cleavage, a 20 nucleotide sequence (5-24 of CRISPR repeat sequence) that includes a seven base-pair stem-loop followed by two unpaired nucleotides, was identified. The structure of this RNA bound to CasE from *Thermus thermophilus* was solved at 2.0Å resolution using X-ray crystallography. This structure reveals numerous sequence specific contacts between the protein and RNA, including several interactions in the major groove of the RNA. The terminal base-pair in the stem-loop in disrupted, with A22 flipped out of the helix and base-stacked with U23. This conformation is partially stabilized by interactions with S34 and E38, which also confer sequence specificity for substrate recognition. Further stabilization of the A22 and U23 conformation is achieved by positioning of the terminal nucleotide, G24, which flips back into register with the stem-loop, but resides well below the helix in a binding pocket made up of residues D18, E24, and K31, with R27 contacting the backbone between U23 and G24.

[00203] The positioning of A22 elongates the backbone of the RNA at the scissile phosphate, splaying it between two active site residues, Y23 and H26. Based on this observation, and the apparent stabilization of this RNA conformation by G24 binding, it was hypothesized that G24 may be required for positioning the RNA in a catalytic conformation. Consistent with this hypothesis, deletion or mutation of G24 significantly reduces cleavage activity, as does mutations of protein residues involved in G24 binding. To confirm the role of G24 in inducing the catalytic RNA conformation, the structure of CasE bound to a 19 nucleotide RNA that lacked the terminal G24 residue was determined. This complex crystallized in two different forms, which revealed two different RNA conformations at the active site of the protein. In one crystal form ($P2_1$), the 2.5 Å structure contained 8 molecules in the asymmetric unit. All 8 molecules revealed that A22 base stacks with G21, maintaining A-form geometry with the rest of the stem-loop. In addition to the changes in the RNA structure, the protein structure also differs from the catalytic conformation observed in the 2.0 Å structure. In that structure, a loop containing R158 and K160 is juxtaposed with the active site, suggesting that these residues may play a role in catalysis or in stabilization of a transition-state intermediate. In the 2.5 Å structure, this loop is distal from the active site and partially disordered, suggesting that the positioning of the loop is flexible. Interestingly, this loop is also

disordered in the apo structure of CasE (Ebihara *et al.*, *Protein Sci.* **15**, 1494-9 (2006)), suggesting that the correct RNA conformation is required for stabilization of this loop.

[00204] The second crystal form ($P2_12_12_1$) obtained for the CasE/19-nucleotide RNA complex was used to determine a 1.5 Å structure, which revealed the RNA bound in the catalytic conformation with A22 and U23 flipped out of the helix. However, the loop containing R158 and K160 remains disordered in this structure, suggesting that G24 binding may also be required for stabilization of this protein structure. The observation of two different RNA conformations for the same complex suggests that the RNA may sample several structural states, and that it may require G24 to lock it into the catalytically competent conformation.

[00205] While the present invention has been described with reference to the specific embodiments thereof, it should be understood by those skilled in the art that various changes may be made and equivalents may be substituted without departing from the true spirit and scope of the invention. In addition, many modifications may be made to adapt a particular situation, material, composition of matter, process, process step or steps, to the objective, spirit and scope of the present invention. All such modifications are intended to be within the scope of the claims appended hereto.

CLAIMS

What is claimed is:

1. A variant Csy4 endoribonuclease comprising an amino acid sequence having at least about 95% amino acid sequence identity to the amino acid sequence set forth in Figure 6, wherein the endoribonuclease comprises an amino acid substitution at His-29, wherein the variant Csy4 endoribonuclease is enzymatically inactive in the absence of imidazole, and wherein the variant Csy4 endoribonuclease is activatable in the presence of imidazole.

2. The variant Csy4 endoribonuclease of claim 1, wherein the amino acid substitution is a His29 to Ala29 substitution.

3. The variant Csy4 endoribonuclease of claim 1, wherein the variant Csy4 endoribonuclease comprises a moiety that provides a detectable signal.

4. The variant Csy4 endoribonuclease of claim 1, wherein the moiety that provides a detectable signal is a fluorophore, a quantum dot, an enzyme other than the endoribonuclease, or a nanoparticle.

5. The variant Csy4 endoribonuclease of claim 1, wherein the endoribonuclease is immobilized on an insoluble support.

6. The variant Csy4 endoribonuclease of claim 5, wherein the insoluble support is a bead.

7. The variant Csy4 endoribonuclease of claim 1, wherein, in the variant Csy4 endoribonuclease binds an RNA substrate comprising the nucleotide sequence 5'-GUUCACUGCCGUAUAGGCAGCUAAGAAA-3' (SEQ ID NO:1).

8. A nucleic acid comprising a nucleotide sequence encoding the variant Csy4 endoribonuclease of claim 1.

9. A recombinant expression vector comprising a nucleotide sequence encoding the variant Csy4 endoribonuclease of claim 1.

10. The recombinant expression vector of claim 9, wherein the nucleotide sequence encoding the variant Csy4 endoribonuclease is operably linked to a promoter.

11. The recombinant expression vector of claim 10, wherein the promoter is an inducible promoter.

12. An *in vitro* genetically modified host cell comprising the recombinant expression vector of claim 9.

13. A kit for purifying a target RNA present in a mixed population of nucleic acids, the kit comprising:

the variant Csy4 endoribonuclease of claim 1.

14. The kit of claim 13, further comprising a recombinant expression vector comprising, in order from 5' to 3' and in operable linkage:

a) a nucleotide sequence encoding an RNA substrate that is specifically bound by the variant Csy4 endoribonuclease of claim 1;

b) a multiple cloning site suitable for insertion of a nucleic acid encoding the target RNA.

15. The kit of claim 14, wherein the nucleotide sequence encoding the RNA substrate is operably linked to a promoter.

16. The kit of claim 15, wherein the promoter is an inducible promoter.

17. The kit of claim 14, wherein the RNA substrate comprises the nucleotide sequence 5'-GUUCACUGCCGUAUAGGCAGCUAAGAAA-3' (SEQ ID NO:1).

18. The kit of claim 14, wherein the recombinant expression vector comprises, inserted into the multiple cloning site, a nucleotide sequence encoding the target RNA.

19. The kit of claim 13, further comprising imidazole.

20. The kit of claim 13, further comprising one or more wash buffers.

21. The kit of claim 13, further comprising a positive control expression vector.

22. The kit of claim 13, wherein the variant Csy4 endoribonuclease is immobilized on an insoluble support.

23. A method of isolating a target RNA present in a mixed population of nucleic acids, the method comprising:

a) contacting a mixed population of nucleic acids with the variant Csy4 endoribonuclease of claim 1, where the variant Csy4 endoribonuclease is immobilized on an insoluble support, wherein the mixed population of nucleic acids comprises a tagged target RNA comprising a recognition nucleotide sequence that is specifically bound by the immobilized variant Csy4 endoribonuclease, forming a tagged target RNA-immobilized variant Csy4 endoribonuclease complex, wherein said contacting is carried out in a binding solution lacking imidazole;

b) adding imidazole to the binding solution to a final concentration of from about 100 mM to about 500 mM, thereby forming a reactivation solution that enzymatically reactivates the immobilized variant Csy4 endoribonuclease, wherein the reactivated immobilized variant Csy4 endoribonuclease cleaves the target RNA from the tag; and

c) collecting the released target RNA.

24. The method of claim 23, further comprising a wash step carried out after step (a) and before step (b).

25. A method of isolating a polypeptide that binds a target RNA, the method comprising:

a) contacting an immobilized complex with a liquid solution comprising a polypeptide that binds the target RNA, wherein the immobilized complex comprises the variant Csy4 endoribonuclease and a tagged target RNA comprising a recognition nucleotide sequence that is specifically bound by the variant Csy4 endoribonuclease, wherein said contacting results in binding of the polypeptide to the target RNA, wherein said contacting is carried out in a binding solution lacking imidazole; and

b) eluting the bound polypeptide.

26. A method of regulating production of a target RNA in a eukaryotic cell, the method comprising contacting a genetically modified host cell with an agent that activates an inducible promoter, wherein the genetically modified host cell is genetically modified with a recombinant expression vector comprising a nucleotide sequence encoding an enzymatically active sequence-specific Csy4 endoribonuclease that catalyzes cleavage at a sequence-specific cleavage site in a substrate polyribonucleotide, wherein the enzyme-encoding nucleotide sequence is operably linked to the inducible promoter, wherein, upon activation of the inducible promoter, the enzyme is produced in the cell and cleaves said target RNA from a precursor RNA.

27. The method of claim 26, wherein the target RNA species is a regulatory RNA.

28. The method of claim 26, wherein cleavage of said target RNA from a precursor RNA inactivates the precursor RNA.

29. A method of detecting a specific sequence in a target polyribonucleotide, the method comprising:

a) contacting the target polyribonucleotide with a oligonucleotide probe comprising the specific sequence and an enzymatically active sequence-specific Csy4 endoribonuclease under conditions that favor duplex formation between the oligonucleotide probe and the target polyribonucleotide, wherein the duplex is cleaved by the Csy4 endoribonuclease; and

b) detecting specific binding between the oligonucleotide probe and the target polyribonucleotide, wherein detection of duplex formation between the oligonucleotide probe

and the target polyribonucleotide indicates the presence of the specific sequence in the target polyribonucleotide.

30. The method of claim 29, wherein the oligonucleotide probe is linked to a peptide, wherein the peptide is released upon cleavage of the duplex by the Csy4 endoribonuclease, and where the detection step comprises detection of the released peptide.

31. The method of claim 29, wherein the released peptide is detected by binding to an antibody specific for the peptide.

32. The method of claim 31, wherein the antibody is immobilized.

33. The method of claim 29, wherein the target polyribonucleotide is immobilized on a solid support.

34. The method of claim 29, wherein the target polyribonucleotide is a polyribonucleotide of a pathogen.

35. The method of claim 29, wherein the enzymatically active Csy4 endoribonuclease comprises an amino acid sequence of a Csy4 amino acid sequence depicted in Figure 5.

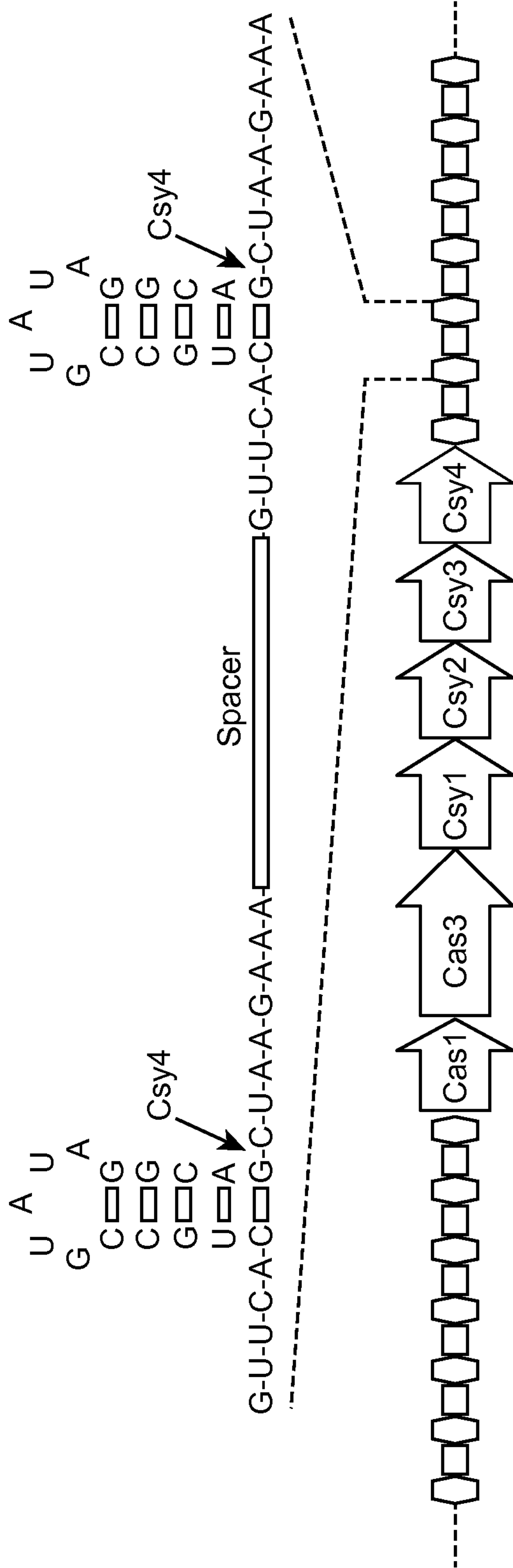

FIG. 1A

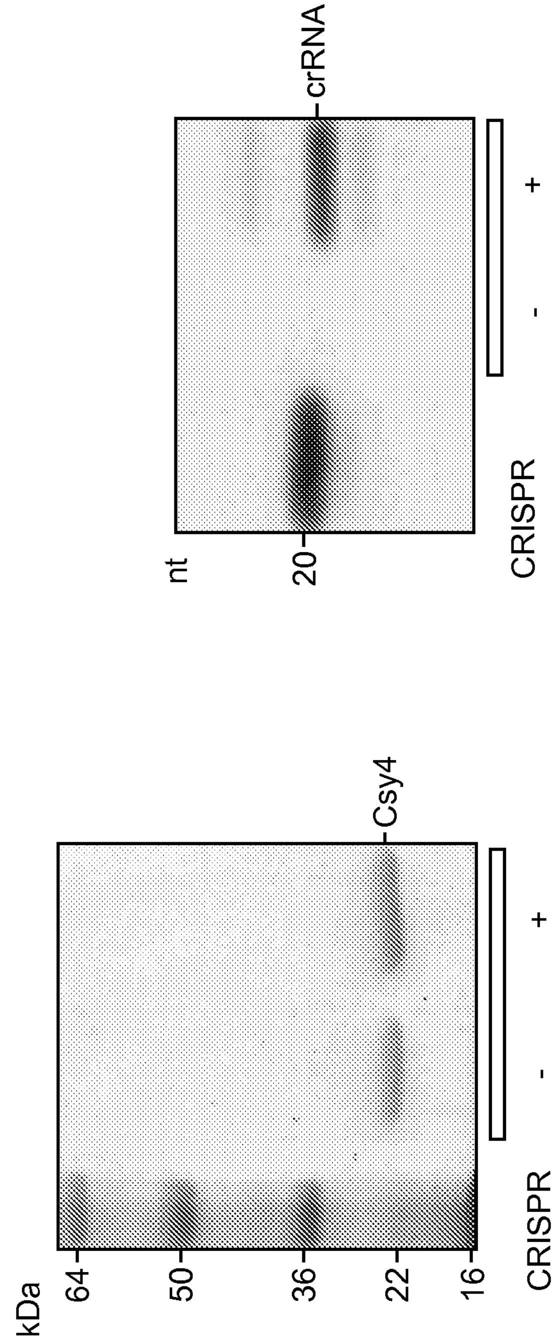

FIG. 1B

FIG. 1C

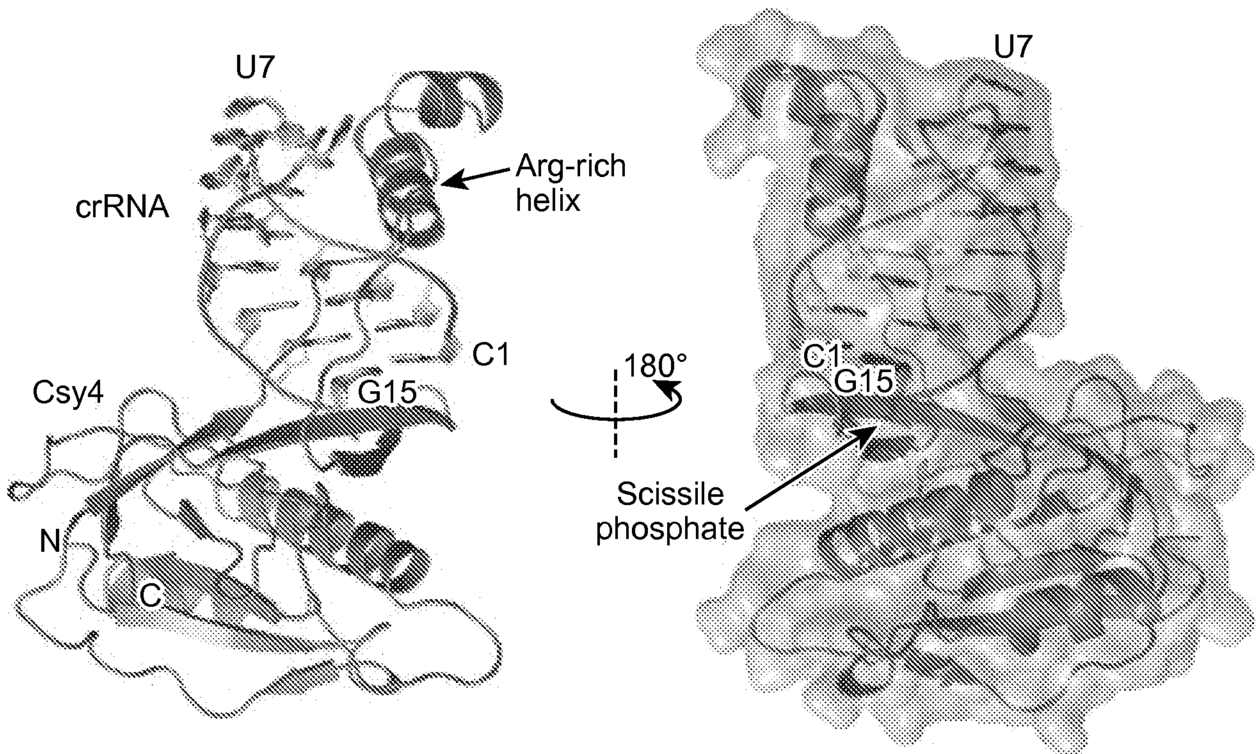

FIG. 2A

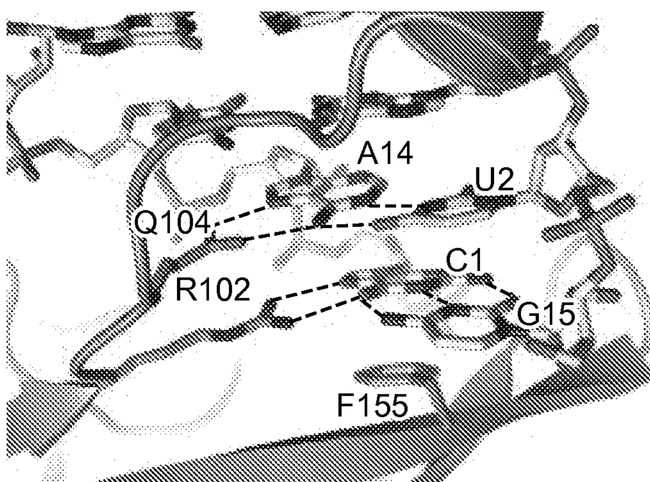

FIG. 2B

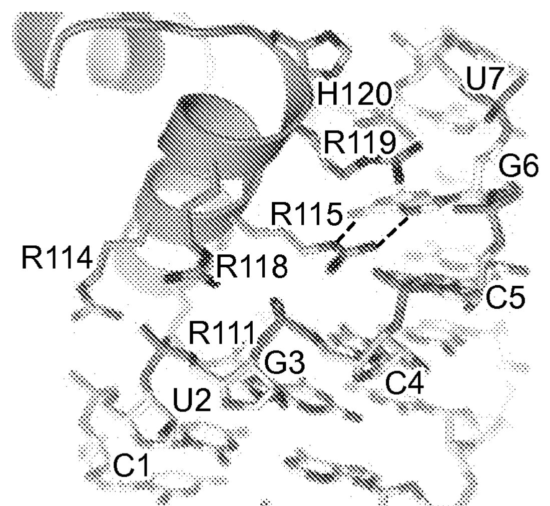

FIG. 2C

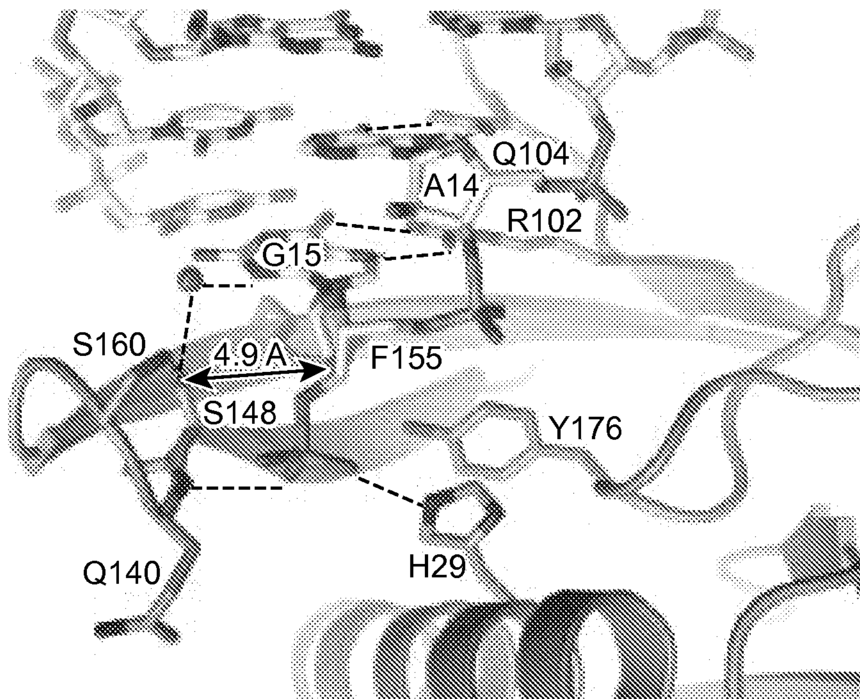

FIG. 3A

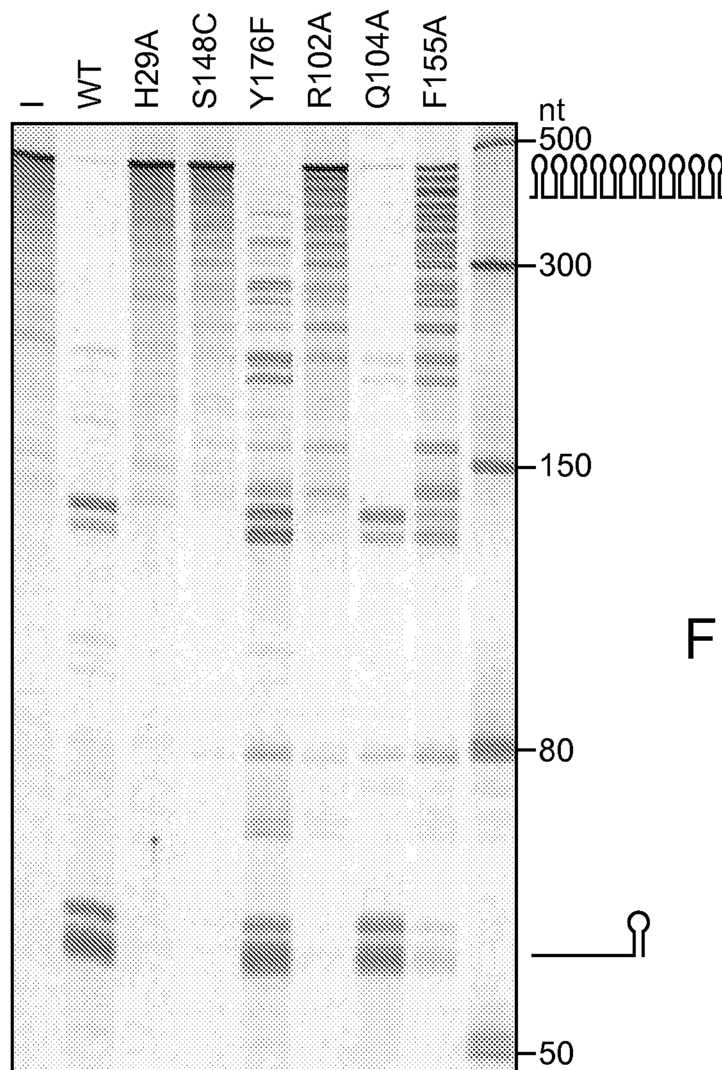

FIG. 3B

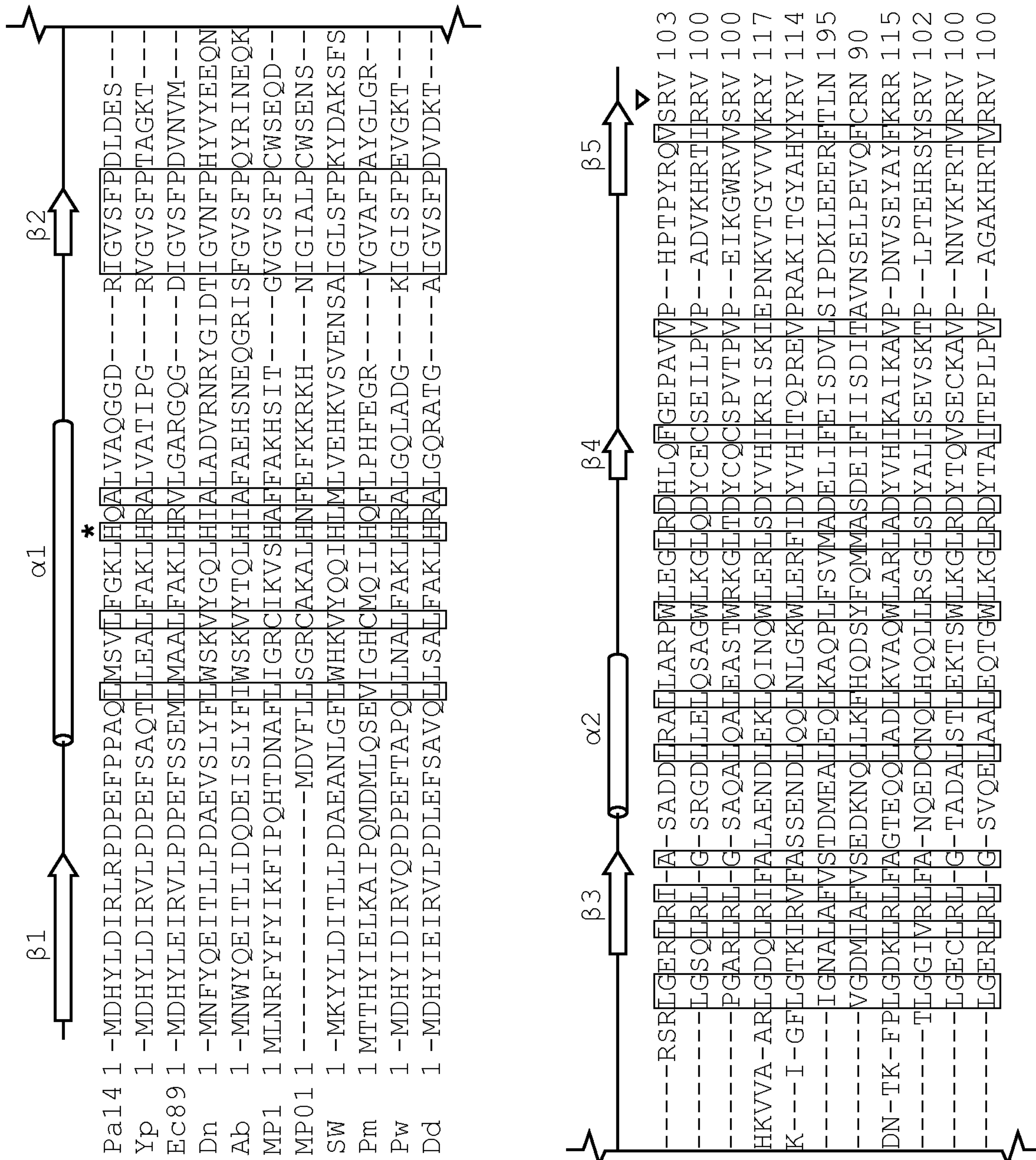

FIG. 4

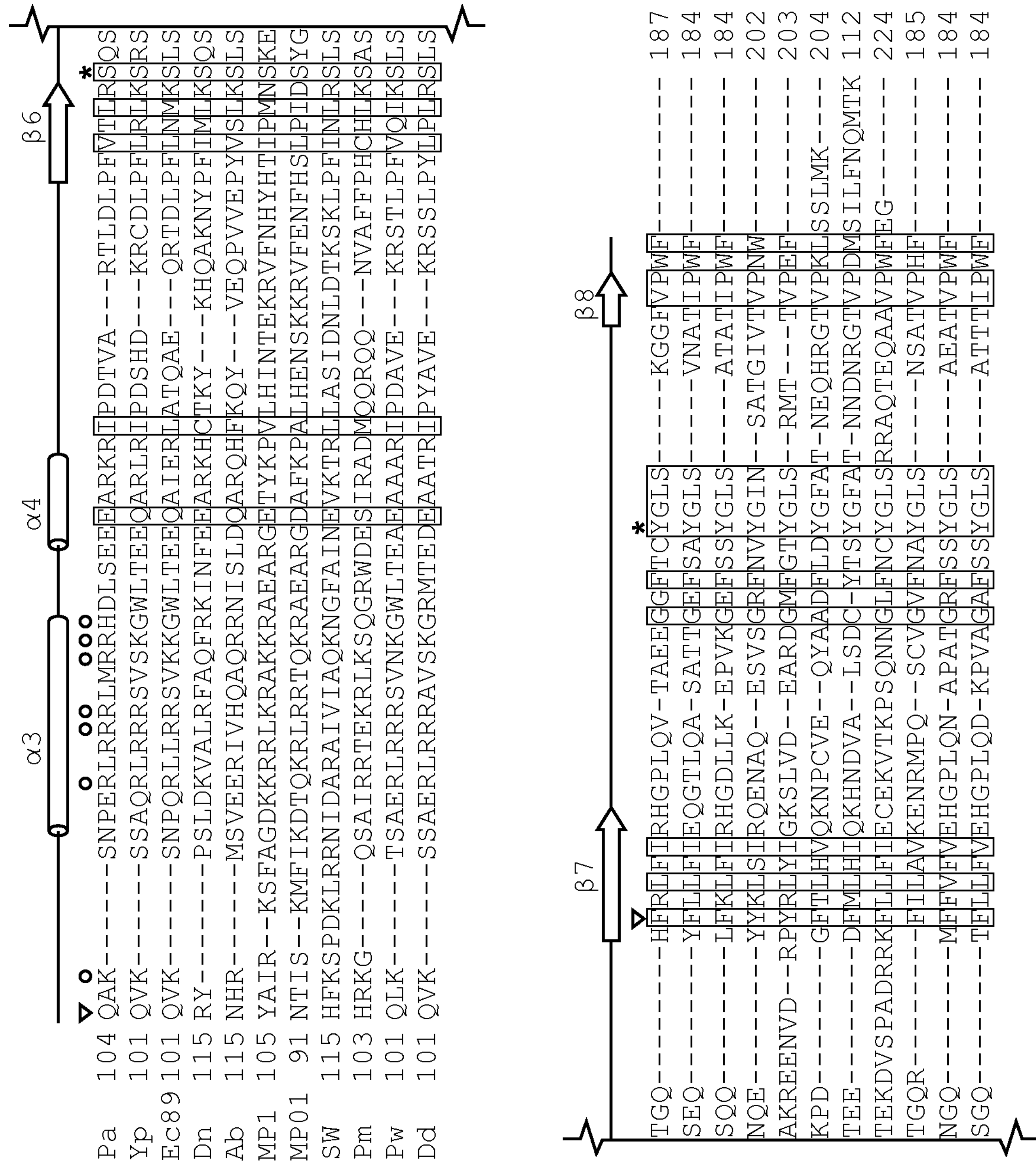

FIG. 4 (Cont.)

FIG. 5A

Csy4 sequence

>gi|116050369|ref|YP_790814.1| hypothetical protein PA14_33300 [Pseudomonas aeruginosa UCBPP-PA14]
MSVLFGKLHQALVAQGGDRIGVSFPDLDESRSRLGERLRIHASADDLRALLARPWLEGLRDHLQFGEPAVVPHPTPYRQVSRVQAKSNPERLRRRLMRRH
DLSEEEARKRIPDTVARALDLPFVTLRSQSTGQHFRLFIRHGPLQVTAEEGGFTCYGLSKGGFVPWF (SEQ ID NO:5)

RNA recognition sequence

GUUCACUGCCGUAUAGGCAGCUAAGAAA (SEQ ID NO:1)

FIG. 5B

Csy4 sequence

>gi|107101871|ref|ZP_01365789.1| hypothetical protein PaerPA_01002916 [Pseudomonas aeruginosa PACS2]
MDHYLDIRLRPDPEFPPAQLMSVLFGKLHQALVAQGGDRIGVSFPDLDESRSRLGERLRIHASADDLRALLARPWLEGLRDHLQFGEPAVVPHPTPYRQVSRVQVKSN
PERLRRRLMRRHDLSEEEARKRIPDTVARALDLPFVTLRSQSTGQHFRLFIRHGPLQVTAEEGGFTCYGLSKGGFVPWF (SEQ ID NO:6)

RNA recognition sequences

GUUCACUGCCGUGUAGGCAGCUAAGAAA (SEQ ID NO:7)
GUUCACUGCCGUAUAGGCAGCUAAGAAA (SEQ ID NO:1)

FIG. 5C

>gi|254235433|ref|ZP_04928756.1| hypothetical protein PACG_01340 [Pseudomonas aeruginosa C3719]
MDHYLDIRLRPDPEFPPAQLMSVLFGKLHQALVAQGGDRIGVSFPDLDESRSRLGERLRIHASADDLRAL
LARPWLEGLRDHLQFGEPAVVPHPTPYRQVSRVQAKSNPERLRRRLMRRHDLSEEEARKRIPDTVARTLD
LPFVTLRSQSTGQHFRLFIRHGPLQATAEEGGFTCYGLSKGGFVPWF
(SEQ ID NO:8)

RNA recognition sequences

GUUCACUGCCGUGUAGGCAGCUAAGAAA (SEQ ID NO:7)
GUUCACUGCCGUAUAGGCAGCUAAGAAA (SEQ ID NO:1)

FIG. 5D

>gi|254240857|ref|ZP_04934179.1| hypothetical protein PA2G_01531 [Pseudomonas aeruginosa 2192]
MDHYLDIRLRPDPEFPPAQLMSVLFGKLHQALVAQGGDRIGVSFPDLDESRSRLGERLRIHASADDLHAL
LARPWLEGLRDHLQFGEAAVVPHPTPYRQVSRVQAKSNPERLRRRLMRRHDLSEEEARKRIPDTVARTLD
LPFVTLRSQSTGQHFRLFIRHGPLQATAEEGGFTCYGLSKGGFVPWF (SEQ ID NO:9)

RNA recognition sequences
GUUCACUGCCGUGUAGGCAGCUAAGAAA (SEQ ID NO:7)
GUUCACUGCCGUAUAGGCAGCUAAGAAA (SEQ ID NO:1)

FIG. 5E

>gi|242238181|ref|YP_002986362.1| CRISPR-associated protein, Csy4 family [Dickeya dadantii Ech703]
MDHYIEIRVLPDLEFSAVQLLSALFAKLHRALGQRATGAIGVSFPDVDKTLGERLRLHGSVQELAALEQT
GWLKGLRDYTAITEPLPVPAGAKHRTVRRVQVKSSAERLRRRAVSKGRMTEDEAATRIPYAVEKRSSLPY
LPLRSLSSGQTFLLFVEHGPLQDKPVAGAFSSYGLSATTTIPWF (SEQ ID NO:10)

RNA recognition sequences
GUUCACUGCCGCGUAGGCAGCUUAGAAA (SEQ ID NO:11)
GUUCACUGCCGAGUAGGCAGCUUAGAAA (SEQ ID NO:12)

FIG. 5F

>gi|261822890|ref|YP_003260996.1| CRISPR-associated protein, Csy4 family [Pectobacterium wasabiae WPP163]
MDHYIDIRVQPDPEFTAPQLLNALFAKLHRALGQLADGKIGISFPEVGKTLGECLRLHGTADALSTLEKT
SWLKGLRDYTQVSECKAVPNNVKFRTVRRVQLKTSAERLRRRSVNKGWLTEAEAAARIPDAVEKRSTLPF
VQIKSLSNGQMFFVFVEHGPLQNAPATGRFSSYGLSAEATVPWF (SEQ ID NO:13)

RNA recognition sequence
GUUCACUGCCGUAUAGGCAGCUUAGAAA (SEQ ID NO:14)

FIG. 5G

>gi|253990195|ref|YP_003041551.1| hypothetical protein PAU_02718 [Photorhabdus asymbiotica subsp. asymbiotica ATCC 43949] MDYYFEILVLPDPEFSKQSLMEALFAKLHRALGQVGNGRIGVSFPCARKTLGDKLRIHGASEALNDLQAL
PWLKGLRDYTEIMDIQPVPQDTQYRRVSRVQVKSSAERLRRRSIKKGWLTEEQARQRIPISKEQRTHLPF
LLVKSLSSRQTFPLFIEQGPIEDKPTPGVFSSYGLSASATIPWF (SEQ ID NO:15)

RNA recognition sequences
GUUCACUGUCGUACAGGCAGCUUAGAAAA (SEQ ID NO:16)
GUGCACUGCCGUACAGGCAGCUUAGAAA (SEQ ID NO:17)
ACUGCCGUACAGGCAGUUUAGAAA (SEQ ID NO:18)
GUUCACUGCCGCACAGGCAGCUUAGAAA (SEQ ID NO:19)
GUGUACUGCCGUACAGGCAGCUUAGAAA (SEQ ID NO:20)
GUUCACUGCCGUACAGGCAGCUUAGAAA (SEQ ID NO:21)

FIG. 5H

>gi|307132482|ref|YP_003884498.1| hypothetical protein Dda3937_03453 [Dickeya dadantii 3937]
MDHYIEIRVLPDPEFSGVQLLSALFAKLHRALGQRATGAIGVSFPDAGKTLGERLRLHGSVQELAALEQT
GWLRGLRDYTAITEPLPVPAGVKHRTVRRVQVKSSAERLRRRAVNKGRMTVDEADARIPYTVEKRTSLPY
LPLRSLSNGQTFLLFVEHGPLQDKPVAGAFSSYGLSAVATIPWF (SEQ ID NO:22)

RNA recognition sequences
GUUCACUGCCGUGUAGGCAGCUUAGAAA (SEQ ID NO:23)
GUUCACUGCCGAGUAGGCAGCUUAGAAA (SEQ ID NO:12)

FIG. 5I

>gi|285019813|ref|YP_003377524.1| CRISPR-associated protein, csy4 family [Xanthomonas albilineans GPE PC73]
MQHYLDLHLRPDPELAPYQLLGALYARLHRSLVTLNTTRIGVSFPGHDNRVPTLGTHLRLHGDDSTLHHL
MATTWLHGVRDHVTITSIGAVPSEAVHRQVTRVQAKSSPERLRRRAMRRHGISEDLAVQRIPDSAAEQLR
LPFVVLGSRSTGQTAFPVFVRHGPVQQEPVPGDFSSYGLSRGATVPWF (SEQ ID NO:24)

RNA recognition sequences
GUUCACUGCCGUGUAGGCAGCUCAGAAA (SEQ ID NO:25)
UUCACUGCCGUGUAGGCAGCUCAGAAA (SEQ ID NO:26)
GUUCACUGCCGUAUAGGCAGCUCAGAAA (SEQ ID NO:27)

FIG. 5J

>gi|50122605|ref|YP_051772.1| hypothetical protein ECA3684 [Pectobacterium atrosepticum SCRI1043]
MDHYIDIRVQPDPEFTASQLLNALFAKLHRVLGQLANGKIGISFPEVGKTLGECLRLHGTEDALSTLEKT
SWLKGLRDYTQVSECKVVPNGVKFRTVRRVQLKSSAERLRRRSVSKGWLTAAEAAARIPDAVEKRSALPF
VQIKSLSNGQMFFVFVEHGPLQNAPTAGRFSSYGLSTEATVPWF (SEQ ID NO:28)

RNA recognition sequence
GUUCACUGCCGUACAGGCAGCUUAGAAA (SEQ ID NO:21)

FIG. 5K

>gi|37525730|ref|NP_929074.1| hypothetical protein plu1796 [Photorhabdus luminescens subsp. laumondii TTO1]
MDYYLEIRVLPDLEFSQQSLFEALFAKLHRALGQLSNGQVGVSFPCARKTLGDTLRIHGSSEALNDLQAL
PWLKGLRDYTEVIDIQPIPQETKYRCVSRVQVKSSAERLRRRAIKKGWLTGEQARQRIPISKEQRTHLPF
LFLKSLSSGQSFLLFVKQGPIQDKPTSGIFSSYGLSSSATIPWF (SEQ ID NO:29)

RNA recognition sequences
GUGCACUGCCGUACAGGCAGCUUAGAAA (SEQ ID NO:17)
GUUCACUGCCGUACAGGCAGCUUAGAAA (SEQ ID NO:21)

FIG. 5L

>gi|297569494|ref|YP_003690838.1| CRISPR-associated protein, Csy4 family [Desulfurivibrio alkaliphilus AHT2]
MVMAMDCYVEISLLPDPEFPDSILMNALFAKLHRALAENGKQEIGVSFPEFGKKLNSKLRIHGSEESLKR
LMDLNWIQGMKDYTRVSGIAKVPDSCQYRTVKRVQAKSSVDRLYRRSVKKGWLSEENAEQQKERAREGRL
KLPFVQLKSQTTGQQFRLFIQHGSLQEKPVTGRFSSYGLSNEATVPWF (SEQ ID NO:30)

RNA recognition sequence
GUUCACUGCCGCACAGGCAGCUCAGAAA (SEQ ID NO:31)

FIG. 5M

```
>gi|251788340|ref|YP_003003061.1| Csy4 family CRISPR-associated protein [Dickeya zeae Ech1591]
MDHYIEIRVLPDLEFSAVQLLSALFAKLHRALGQQATGAIGVSFPDVGKTLGERLRLHGSEQALTALEQT
GWRTGLRDYSTITDVLTVPTGAQYRTVRRVQVKSSAERLRRRAVSKGWLTADEAAARIPYAVEKRTSLPY
LPLRSLSSGQPFLLFVEHGPLQDKPVAGTFSSYGLSATATIPWF  (SEQ ID NO:32)
```

RNA recognition sequences

```
GUUCACUGCCGUGUAGGCAGCUUAGAAA (SEQ ID NO:23)
GUGCACUGCCGUAUAGGCAGCUUAGAAA (SEQ ID NO:33)
```

FIG. 5N

```
>gi|22125621|ref|NP_669044.1| hypothetical protein y1727 [Yersinia pestis KIM 10]
MDHYLDIRVLPDPEFSAQTLLEALFAKLHRALVATIPGRVGVSFPTAGKTLGSQLRLHGSRGDLLELQSA
GWLKGLQDYCECSEILPVPADVKHRTIRRVQVKSSAQRLRRRSVSKGWLTEEQARLRIPDSHDKRCDLPF
LRLKSRSSEQYFLLFIEQGTLQASATTGEFSAYGLSVNATIPWF   (SEQ ID NO:34)
```

RNA recognition sequences

```
GUUCACUGCCGCACAGGCAGCUUAGAAA (SEQ ID NO:19)
UGUUCACUGCCGCACAGGCAGCUUAGAAAA (SEQ ID NO:35)
```

FIG. 5O

```
>gi|271501952|ref|YP_003334978.1| Csy4 family CRISPR-associated protein [Dickeya dadantii Ech586]
MDHYIEIRVLPDPEFSAVQLLSALFAKLHRALGQRATGDIGVSFPDAGKTLGERLRLHGSVQALAALEQT
GWLKGLRDYSTITDVLTVPTGAQYRTVRRVQVKSSAERLRRRAVSKGRMTADEAAARIPYAAEKRTSLPY
LPLRSLSSGQTFLLFVEHGPLQEKPVAGVFSSYGLSAIATIPWF  (SEQ ID NO:36)
```

RNA recognition sequences

```
GUGAACUGCCGCAUAGGCAGCUUAGAAA (SEQ ID NO:37)
GUUCACUGCCGAGUAGGCAGCUUAGAAA (SEQ ID NO:12)
```

FIG. 5P

```
>gi|117623067|ref|YP_851980.1| hypothetical protein APECO1_1206 [Escherichia coli APEC O1]
MAVSLVRNRNKELPMDHYLEIRVLPDPEFSSEMLMAALFAKLHRVLGARGQGDIGVSFPDVNVMPGARLR
LHGSAQALQALEASTWRKGLTDYCQCSPVTPVPEIKGWRVVSRVQVKSNPQRLLRRSVKKGWLTEEQAIE
RLATQAEQRTDLPFLNMKSLSSQQLFKLFIRHGDLLKEPVKGEFSSYGLSATATIPWF (SEQ ID NO:38)
```

RNA recognition sequence
GUUCACUGCCGUACAGGCAGCUUAGAAA (SEQ ID NO:21)

FIG. 5Q

>gi|91209927|ref|YP_539913.1| hypothetical protein UTI89_C0896 [Escherichia coli UTI89]
MDHYLEIRVLPDPEFSSEMLMAALFAKLHRVLGARGQGDIGVSFPDVNVMPGARLRLLHGSAQALQALEAS
TWRKGLTDYCQCSPVTPVPEIKGWRVVSRVQVKSNPQRLLRRSVKKGWLTEEQAIERLATQAEQRTDLPF
LNMKSLSSQQLFKLFIRHGDLLKEPVKGEFSSYGLSATATIPWF (SEQ ID NO:39)

RNA recognition sequence
GUUCACUGCCGUACAGGCAGCUUAGAAA (SEQ ID NO:21)

FIG. 5R

>gi|237808124|ref|YP_002892564.1| CRISPR-associated protein, Csy4 family [Tolumonas auensis DSM 9187]
MDHYLDIRLLPEEPEVSESFLLNALFAKLHVRLGQQAQGRVGVSFPDHHKRLGDLLRLHGQRTDLQALMA
DDWLQGLKGYTQCSEVLPIPATVSYRAVKRVQAKSAHNKRQRSIAKGWLTESEAQIRIPDTQQKELHLPF
VQLKSRSNGQMMRVYVEHGPVLAVPVSGYFNAYGLSSIATIPWF (SEQ ID NO:40)

RNA recognition sequence
CUUCACUGCCGCACAGGCAGCUUAGAAA (SEQ ID NO:41)

FIG. 5S

>gi|259907505|ref|YP_002647861.1| CRISPR-associated protein Csy4 [Erwinia pyrifoliae Ep1/96]
MDHYQDIRVRVDEENGEAVLLAQVFMHLHQVLMRAANGRIGISFPNVKRTLGDRIRLHGTLDDLSALQQS
GWNKCLRDYIACSDIAPVPKGAAWRTVRRVQVKSSAERLRRRSVNKGWLSEQEAAERISVLNEQRSNLPF
LQIKSGSNGQAWRLFIEHGSLVSAPSDGSFSSYGLSAAATIPWF (SEQ ID NO:42)

RNA recognition sequences
GUUCACUGCCGCACAGGCAGCUUAGAAA (SEQ ID NO:19)
UUCACUGCCGUACAGGCAGCUUAGAAAA (SEQ ID NO:43)

FIG. 5T

>gi|218688670|ref|YP_002396882.1| hypothetical protein ECED1_0855 [Escherichia coli ED1a]
MAVSLVRNRNKELPMDHYLEIRVLPDPEFSSEMLMAALFAKLHRVLGARGQGDIGVSFPDVNVMPGTHLR
LHGSAQALQELEASTWRKGLTDYCQCSPVTPVPEIKGWRVVSRVQVKSNPQRLLRRSVKKGWLTEEQAIE
RLATQAEQRTDLPFLNMKSLSSQQQFKLFIRHGDLLKEPVKGEFSSYGLSATATIPWF (SEQ ID NO:44)

RNA recognition sequence
GUUCACUGCCGUACAGGCAGCUUAGAAA (SEQ ID NO:21)

FIG. 5U

>gi|121608426|ref|YP_996233.1| CRISPR-associated Csy4 family protein [Verminephrobacter eiseniae EF01-2]
MSTHYIDITLRPDPEFSPAHLLNALHAQLHLALVQLGTGDVGVSFPGFILRGEHSHLGTTLRLHGATSAL
QRLQALSWLRGMRDHVKTSEVAPVPTHTQHRVVRRVQAKSSPERSRRRLMRRLEIDEAQALQRIPDQEGR
RLALPYLRLQSASKGQVFRLFIEHGPLLDTPSPGSFGTYGLSTQATIPWF (SEQ ID NO:45)

RNA recognition sequences
GUUCACUGCCGGAUAGGCAGCUCAGAAA (SEQ ID NO:46)

FIG. 5V

```
>gi|34497206|ref|NP_901421.1| hypothetical protein CV_1751 [Chromobacterium violaceum ATCC 12472]
MDHYLDIRLLPDADFGPPVLMNALYAKLHRALAAQQRQDIGVSFPGYDPAPSSHDGKPLPPTLGLTLRLH
GSAAALDGLMARRWLSGFADHAIVGDIRPVPAGASAVSVRRRQAKSSPARARDRLMRRQGISAEEARRRI
PDETAQRLNLPYLTVDSASTGQCFRLFVEQQAAPSIAAGSFNAYGLSAAAALPAW  (SEQ ID NO:47)
```

RNA recognition sequence
GUUCACUGCCGGAUAGGCAGCUUAGAAA (SEQ ID NO:48)

FIG. 5W

```
>gi|188532992|ref|YP_001906789.1| hypothetical protein ETA_08450 [Erwinia tasmaniensis Et1/99]
MDRYQDIRVRVDAEMTAPVLLAQVFMRLHQVLMRAANGRIGISFPDVKLTLGDRIRLHGTLDDLSSLQQS
GWDKGLTDYIACSAIDPVPPGAAWRTVRRVQVKSSAERLRRRSVNKGWLNEAEAAERINVLSEQRSDLPY
LQIKSGSNGHAWRLFIEHGPLVSVPVNGGFSSYGLSATATVPWF  (SEQ ID NO:49)
```

RNA recognition sequences
GUUCACUGCCGCACAGGCAGCUUAGAAA (SEQ ID NO:19)
GUUCACUGCCGUACAGGCAGCUUAGAAG (SEQ ID NO:50)

FIG. 5X

```
>gi|160896663|ref|YP_001562245.1| CRISPR-associated Csy4 family protein [Delftia acidovorans SPH-1]
MAMTSHYIDTTLLPDPEFSHAHLLGALVAKLHRALVQLGSTDIGISFPGYSLRPRTLGTILRLHGSEAAL
RGLMEQPWLQGMRDHVHCTPPALVPEGAVPCLVQRRQFKTSPDRLRRRRMRRKGETAEQAAAAIPDSVER
TPDLPYVQLRSASTGQPFCLFVEQKAVQGTAGQEGFNTYGLSLGTAVPWF  (SEQ ID NO:51)
```

RNA recognition sequence
GUUCGCUGCCGCGUAGGCAGCUCAGAAA (SEQ ID NO:52)

FIG. 5Y

>gi|146311064|ref|YP_001176138.1| CRISPR-associated Csy4 family protein [Enterobacter sp. 638]
MDHYLEIRVLSDPEFSEETLMAALFAKLHRALGARGQGDIGVSFPRYSLKPGDTLRLHGSAQSLDELEKM
AWRKGLSDYCLCKGVLPAPDVNAWRCVSRVQVKSSPQRLMRRSVKKGWLTEEEAQQRLLNLQEARTDLPW
LNLQSLSTGQSFRLFIRHGDIVDMPMCGEFSSYGLSATATIPWF (SEQ ID NO:53)

RNA recognition sequence
GUUCACUGCCGUACAGGCAGCUUAGAAA (SEQ ID NO:21)

FIG. 5Z

>gi|289209612|ref|YP_003461678.1| CRISPR-associated protein, Csy4 family [Thioalkalivibrio sp. K90mix]
MDHYLDLRVMPDPEFKETTLLGALVSKLHRRLVSMSADDIGISLPDHEQEPPLGRRLRVHGTQGRLNLLM
QDEWLGGMQSLVDATPVQPVPDQVTYRPVRRRQYKTNAERLRRRRMRRHGESYEEARQHIPDTVERRVNT
PFLSVQSASTGQRFSLFIEHGPPQQHASPGRFNTYGLSQDATVPWF (SEQ ID NO:54)

RNA recognition sequence
GUUAGCUGCCGCACAGGCAGCUCAGAAA (SEQ ID NO:55)

FIG. 5AA

>gi|283856310|ref|YP_162420.2| Csy4 family CRISPR-associated protein [Zymomonas mobilis subsp. mobilis ZM4]
MLANPVDSYQDIYILPNQEIAPHIIMEKLFSLLHLELVRLGSQHIGISFPEHDNNKPCLGSRLRLHGTGA
DLHELALSGWITRLDDYLYCEDIKSVPEIRQYCVVSRVQAKSSPARLRRRAIRRHGFHDEEAKKVIPDTA
FERLELPFIMTGSCSTKQPRFPVFISHKIIQNKLMNGNFNSYGLSLGASVPWF (SEQ ID NO:56)

RNA recognition sequence
GUUCACUGCCGCACAGGCAGCUUAGAAA (SEQ ID NO:19)

FIG. 5AB

>gi|260752821|ref|YP_003225714.1| CRISPR-associated protein, Csy4 family [Zymomonas mobilis subsp. mobilis NCIMB 11163] MLANPVDSYQDIYILPNQEIAPHIIMEKLFSLLHLELVRLGSQHIGISFPEHDNNKPCLGSRLRLHGAGA DLHELALSGWITRLDDYLYCEDIKSVPEIRQYCVVSRVQAKSSPARLRRRAIRRHGFHDEEAKKVIPDTA FERLELPFIMTGSCSTKQPRFPVFISHKIIQDKLMNGNFNSYGLSLGASVPWF (SEQ ID NO:57)

RNA recognition sequence
GUUCACUGCCGCACAGGCAGCUUAGAAA (SEQ ID NO:19)

FIG. 5AC

>gi|121592915|ref|YP_984811.1| CRISPR-associated Csy4 family protein [Acidovorax sp. JS42]
MTTHYINITLLPDPEFSHAHLLGALVAKLHRALVQGHTTDIGVSYPQHVSQPLTKRTLGAVLRLHGTPEA
LQRLMEEDWLKGMRDHTQVGELLPVPANAQHRTVRRRQFKTNADRLRRRRMQRKGETAEQAAAAIPDTVE
RRPDLPFVQLRSSSTGQSFCLCVEHGPLQPLPVAGAFNAYGLGHDATVPWF (SEQ ID NO:58)

RNA recognition sequences
GUUCACUGCCGCAUAGGCAGCUCAGAAA (SEQ ID NO:59)

FIG. 5AD

>gi|317051103|ref|YP_004112219.1| Csy4 family CRISPR-associated protein [Desulfurispirillum indicum S5]
MDSYIEIRILPDQEFEATTLMSTVFAKLHRALVESGRSDIGVSFPEAGKTPGALLRLHGSLAALESIMTL
SWLTGLQDYTQTSGILQVPAQAAYVQVARVQSKMTASRIRRALKRGSLSEERALELLQSRDQLNQPFFRL
LSASTAQKFPLFIEQRNAEKAGKQSVYSAYGLSVGGSTVPWF (SEQ ID NO:60)

RNA recognition sequence
GUUCACUGCCGCAUAGGCAGCUCAGAAA (SEQ ID NO:59)

FIG. 5AE

>gi|54303643|ref|YP_133636.1| hypothetical protein PBPRB1991 [Photobacterium profundum SS9]
MMDSYVDIQLKPDAEMREAELSSKVFTKFHKALATLNTNKIGISFPQMNLKLGRLFRIHGNASLLKDLQG
IKWLGALAGYCQVGEITVVPDQVQYRVISVKRSNLSKAKLKRLIARGSIDKDGEKRYKVKMLSQGFDNPY
LDLFSSSTGQVYRKFFEFGDIQATSVSDEFDSYGLSNTATIPWF (SEQ ID NO:61)

RNA recognition sequence
GUUCACUGCCGCACAGGCAGCUUAGAAA (SEQ ID NO:19)

FIG. 5AF

>gi|54292953|ref|YP_122340.1| hypothetical protein plpl0047 [Legionella pneumophila str. Lens]
MDHYLDISILPDSEFTTPVLMNAIYTNLHKALHTLASTNIGVSFPKYSSTLGNLLRIHGKKEALQELQNL
NWIGGMIGYCEASLIKTVPADTKFRTVSRKQPTMSQSKLRRLIKRNSLTEDEIRQYKAKMFSKGLDNPYI
ELVSVSNGQRHRRYIEFGELFNEPIPGLFDQFGLSNSATVPWFD (SEQ ID NO:62)

RNA recognition sequences
GUUCACUGCCGUACAGGCAGCUUAGAAA (SEQ ID NO:21)
GUUCACUGCCGCACAGGCAGCUUAGAAA (SEQ ID NO:19)

FIG. 5AG

>gi|54295752|ref|YP_128167.1| hypothetical protein lpl2842 [Legionella pneumophila str. Lens]
MDHYLEISILPDSEFTTPVLMNAIYTNLHKALHTLASTSIGVSFPKYSSTLGNILRIHGKKEVLQDLQNL
NWIGGMIGYCEASLIKTVPADTKFRTVSRKQPTMSQSKLRRLIKRNTLTEDEIRQYKAKMFSKGLDNPYI
ELVSVSNGQRHRRYIEFGELFNEPSPGLFDQFGLSNSATVPWFD (SEQ ID NO:63)

RNA recognition sequences
GUUCACUGCCGUACAGGCAGCUUAGAAA (SEQ ID NO:21)
GUUCACUGCCGCACAGGCAGCUUAGAAA (SEQ ID NO:19)

FIG. 5AH

>gi|146292921|ref|YP_001183345.1| CRISPR-associated Csy4 family protein [Shewanella putrefaciens CN-32]
MNSYIDIRLKPDAEMREAELSSKVFTKFHKALVTLNSHKIGISFPQMKLSLGQLFRIHGDASLLHDLQGL
DWLGPLAGYCQVTAVSAVPDHVQYRIVSVKRSNLSKAKLKRLIARGSIDKDGEKRYKVKMLGQGFDNPYL
DLFSSSTGQVYRKFFEFSDIQAHPLDGEFDSYGLSKTATVPWF (SEQ ID NO:64)

RNA recognition sequence
GUUCACCGCCGCACAGGCGGCUUAGAAA (SEQ ID NO:65)

FIG. 5AI

>gi|296106567|ref|YP_003618267.1| hypothetical protein lpa_01476 [Legionella pneumophila 2300/99 Alcoy]
MDYYVDILIKPDSEKSLNFLLSTLYTKLHKVLHDMASTNIGVSFPKYNITLGNILRIHSKKVVLDELLGM
NFLSGINNYYEVSPIKSVPADSKFRIISRKQTTMSQSKMRRLFKRGSMTVGDIRQYKAKMFAKSIDNPYL
ELVSGSNGYRYRRYIEFGELLDQPVYGEFDRFGLSKTATVPWFD (SEQ ID NO:66)

RNA recognition sequence
GUUAACUGCCGCACAGGCAGCUUAGAAG (SEQ ID NO:67)

FIG. 5AJ

>gi|260772736|ref|ZP_05881652.1| hypothetical protein VIB_001192 [Vibrio metschnikovii CIP 69.14]
MDSYIEIRLQPDAEMPEAELSSKVFTKFHKALVILHSNQIGISFPEVNVKLGRLFRLHGEASFLHDLQGL
NWLGPLSGYCQVSEILAIPEQVQYRVISVKRSNLSQAKLRRLIARGSIDKEGEKRYKVKMLSQGFDNPYL
DLFSSSTKQVHRKFFEFGEIQPLPVSGKFDSYGLSHTTTVPWF (SEQ ID NO:68)

RNA recognition sequences
GUUCACUGCCGCACAGGCAGCUUAGAAA (SEQ ID NO:19)
GUUCACUGCCGCAUAGGCAGCUUAGAAA (SEQ ID NO:69)

FIG. 5AK

>gi|157146437|ref|YP_001453756.1| hypothetical protein CKO_02197 [Citrobacter koseri ATCC BAA-895]
MAITPVPAVKGWRTVSRVQVKSSPQRLLRRSVRKGWLTEEQAQLRLVESTEQHSDLPYLNVKSLSNQQQF
RVFIRHSELRSEPVSGTFTSYGLSSTATIPWF (SEQ ID NO:70)

RNA recognition sequence
UGUUCACUGCCGUACAGGCAGCUUAGAAA (SEQ ID NO:21)

FIG. 5AL

>gi|262402803|ref|ZP_06079364.1| hypothetical protein VOA_000785 [Vibrio sp. RC586]
MDAYIDIRLMPDAEMREAELSSKVFIKFHKALVKLRSNKIGISFPEANIKLGRLFRLHGEMSALHDLQGL
NWLGPLAGYCKITTVTHVPDQVQYRIISVKRSNLSKAKLTRLIARGSIDKDGEKRYKVKMLSQGFDNPYL
DLSSSSTGQVYRKFFEFSDIQADPVDGEFDSYGLSKTATVPWF (SEQ ID NO:71)

RNA recognition sequences
GUUCACUGCCGCACAGGCAGCUUAGAAA (SEQ ID NO:19)
AGUGUUCUGCCGAAUAGGCAGCUAAGAA (SEQ ID NO:72)

FIG. 5AM

>gi|229523353|ref|ZP_04412760.1| hypothetical protein VIF_000211 [Vibrio cholerae TM 11079-80]
MMDAYIDIRLMPDAEMREAELSSKVFIKFHKALVKLQSNKIGISFPEANIKLGRLFRLHGEVSALHDLQG
LNWLGPLAGYCKITTVTHVPDQVEYRIISVKRSNLSKAKLARLIARGSIDKDGEKRYKVKMLRQGFDNPY
LDLSSSSTGQVYRKFFEFSDIQAEPVDGEFDSYGLSKTATVPWF (SEQ ID NO:73)

RNA recognition sequence
GUUCACUGCCGCACAGGCAGCUUAGAAAU (SEQ ID NO:74)

FIG. 5AN

>gi|237748015|ref|ZP_04578495.1| crispr-associated protein [Oxalobacter formigenes OXCC13]
MKHYIEITLTGSPDFPLYHLWSKLYTQLHLALVENRDASDQVNIGVSFPEYYFNEEKGMGFLGTKLRLFA
EDETSLQKIDIQKWFVRLNDCIHITPVCRVPLNEITGYATFSRKHIKSNAERLARRQMKRHKDLSFHETV
QRYQKNLAKSPLPFIQLESLTNSHPFKLFIEKKPAINASLKVFTTYGLSAESTIPEF (SEQ ID NO:75)

RNA recognition sequence
GUUCACUGCCGUAUAGGCAGCUUAGAAG (SEQ ID NO:76)

FIG. 5AO

>gi|119945137|ref|YP_942817.1| CRISPR-associated Csy4 family protein [Psychromonas ingrahamii 37]
MKYYLDITLLPDIEIPLGFIWQKVFQQVHIALADNKVGENESDIALSLPNYGDKAFPLGNKLRLFSVSEQ
ALERLAITKWLKRFTDHTHITSVKAVPESANEYACFTRKQFNTNISRLARRRAKRHMETFEKALQYYDNF
AEEQTKLPFMNIKSLTNNAQFRIFIERSITKIPKQGTFNCYGLSQAIATVPWF (SEQ ID NO:77)

RNA recognition sequence
GUGUUCCCCGUGCCCACGGGGAUGAACCG (SEQ ID NO:78)

FIG. 5AP

>gi|146328647|ref|YP_001209099.1| hypothetical protein DNO_0170 [Dichelobacter nodosus VCS1703A]
MNFYQEITLLPDAEVSLYFLWSKVYGQLHIALADVRNRYGIDTIGVNFPHYVYEEQNHKVVAARLGDQLR
IFALAENDLEKLQINQWLERLSDYVHIKRISKIEPNKVTGYVVVKRYRYPSLDKVALRFAQFRKINFEEA
RKHCTKYKHQAKNYPFIMLKSQSNQEYYKLSIRQENAQESVSGRFNVYGINSATGIVTVPNW (SEQ ID NO:79)

RNA recognition sequence
GUUCACCGCCGCACAGGCGGCUUAGAAA (SEQ ID NO:65)

FIG. 5AQ

>gi|160876478|ref|YP_001555794.1| CRISPR-associated Csy4 family protein [Shewanella baltica OS195]
MNHYLDITLLPNEEVGHYFLWEKLYHQVHLALVEHKNRVGQFEIAAAFPQFNEMDNSLGSKLRLLATQPQ
HLEDLKVSNWLRHFTDYLHISSIRPVPEKIEVYVAYSRPAIRANKAREIARRMKRHNETLEQATAHFEGF
KPKKTKAPFVYMQSYTKDSRFPLFIQQTHSAVVKEGSVSFDSYGLSSRGYLPKF (SEQ ID NO:80)

RNA recognition sequence
GUUCACCGCCGCACAGGCGGCUUAGAAA (SEQ ID NO:65)

FIG. 5AR

>gi|153001745|ref|YP_001367426.1| CRISPR-associated Csy4 family protein [Shewanella baltica OS185]
MNHYLDITLLPNEEVGHYFLWEKLYHQMHLALVEHKNRVGQFEIAAAFPQFNEMDNNLGSKLRLLATQPQ
HLEDLKVSNWLRHFTDYLHISSIRPVPDKIEVYVAYSRPAIRANKAREIARRMKRHNETLVQATAHFEGF
KPKKTKAPFVYMQSYTKDSRFPLFIQQTHSAVVKEGNVSFDSYGLSSRGYLPKF (SEQ ID NO:81)

RNA recognition sequence
GUUCACCGCCGCACAGGCGGCUUAGAAA (SEQ ID NO:65)

FIG. 5AS

>gi|169795154|ref|YP_001712947.1| hypothetical protein ABAYE1000 [Acinetobacter baumannii AYE]
MMNWYQEITLIDQDEISLYFIWSKVYTQLHIAFAEHSNEQGRISFGVSFPQYRINEQKKIGFLGTKIRVF
ASSENDLQQLNLGKWLERFIDYVHITQPREVPRAKITGYAHYYRVNHRMSVEERIVHQAQRRNISLDQAR
QHFKQYVEQPVVEPYVSLKSLSAKREENVDRPYRLYIGKSLVDEARDGMFGTYGLSRMTTVPEF (SEQ ID NO:82)

RNA recognition sequence
GUUCAUGGCGGCAUACGCCAUUUAGAAA (SEQ ID NO:83)

FIG. 5AT

>gi|213158184|ref|YP_002320235.1| crispr-associated protein, Csy4 family [Acinetobacter baumannii AB0057]
MNWYQEITLIDQDEISLYFIWSKVYTQLHIAFAEHSNEQGRISFGVSFPQYRINEQKKIGFLGTKIRVFA
SSENDLQQLNLGKWLERFIDYVHITQPREVPRAKITGYAHYYRVNHRMSVEERIVHQAQRRNISLDQARQ
HFKQYVEQPVVEPYVSLKSLSAKREENVDRPYRLYIGKSLVDEARDGMFGTYGLSRMTTVPEF (SEQ ID NO:84)

RNA recognition sequence
GUUCAUGGCGGCAUACGCCAUUUAGAAA (SEQ ID NO:83)

FIG. 5AU

>gi|15602173|ref|NP_245245.1| hypothetical protein PM0308 [Pasteurella multocida subsp. multocida str. Pm70]
MTTHYIELKAIPQMDMLQSEVIGHCMQILHQFLPHFEGRVGVAFPAYGLGRTLGGIVRLFANQEDCNQLH
QQLLRSGLSDYALISEVSKTPLPTEHRSYSRVHRKGQSAIRRTEKRLKSQGRWDESIRADMQQRQQNVAF
FPHCHLKSASTGQRFILAVKENRMPQSCVGVFNAYGLSNSATVPHF (SEQ ID NO:85)

RNA recognition sequences
GUUCACCAUCGUGUAGAUGGCUUAGAAA (SEQ ID NO:86)
GUUAACUGCCGUAUAGGCAGCUUAGAAA (SEQ ID NO:87)

FIG. 5AV

>gi|293390434|ref|ZP_06634768.1| Csy4 family CRISPR-associated protein [Aggregatibacter actinomycetemcomitans D7S-1]
MTVQTHYIEIKAIPQVDMLQTEVIGFCLQKLHQILPHFEGRIGLAFPAYGNDKTLGGIIRLFGTENDCGF
IHFKLQSLRDYALISEVMPIPEKVRSYRIYQRIQPKGQSSIRRAEKRLTAQGKWNEEVLQNMLQKQATQR
IYPHAHLKSSSTKQQFILAIKSVHQTKAVEGVFSAYGLSQTTTVPHF (SEQ ID NO:88)

RNA recognition sequence
CUUCACUGCCGAAUAGGCAGCUUAGAAA (SEQ ID NO:89)

FIG. 5AW

>gi|152996699|ref|YP_001341534.1| CRISPR-associated Csy4 family protein [Marinomonas sp. MWYL1]
MKHYIDITLLPSDDIGVHFLWSKLMMQVHLALVEIQNEQKQVPVAVSFPKYQPRENEKLGFVGNKLRLFA
NDKTDLERLNFGKWLHRLEDYVHIKSIADVPNDVISYESFNRRSKSGSPDKHIKRRMQRHNETWEQAAAF
FKGYSMEKADKDLPFIRMKSLHSDNEFCMSIIRKEAAPSNKHIMFNTYGLSAEGVLPKF (SEQ ID NO:90)

RNA recognition sequence
GUUCGCCGCCGAGCACGCGGCUUAGAAA (SEQ ID NO:91)

FIG. 5AY

>gi|258645690|ref|ZP_05733159.1| CRISPR-associated protein, Csy4 family [Dialister invisus DSM 15470]
MEYYQEITLLPCAEVSLAFLWTKVFTQLHIAFADEKNKSGHNLYAVSFPEYRETGLGEKIRVFAEAQELE
RLNLSKVLGRLLDYVHCTSIRKVPERKLRGYAVYSRYQPEGSIWVKARRYAKRHPGVTIEEAARLLQGKR
KSVRLPYIQMKSLSRGGTFSLFIKKRVEKESALTECGTYGLSNNRTVPEF (SEQ ID NO:92)

RNA recognition sequences
GUUAACUGCCGCAUAGGUAGUUUAGAAA (SEQ ID NO:93)
GUUAUCUGCCGUAUAGGCAGCUUAGAAA (SEQ ID NO:94)

FIG. 5AZ

>gi|165975671|ref|YP_001651264.1| hypothetical protein APJL_0216 [Actinobacillus pleuropneumoniae serovar 3 str. JL03]
MSELTHYIELKAIPQVDILQTDVIAHGLQILHKFLPLYQGEIGLSFPAYGLGRTLGGIIRVFGNEQHCTQ
IKTQLIGEGLQDYVLITSVTPVPEEIVEYHRYQRVHRKGQSAIRRTEQFLVQQGKWTEEIRQEMLIHQQN
QKVFPYVKLKSGSTKQHFVLAIRQLRLAEPASGLFNAYGLSQAATVPHF (SEQ ID NO:95)

RNA recognition sequence
CUUCACUGCCGUAUAGGCAGCUUAGAAA (SEQ ID NO:96)

FIG. 5BA

>gi|190149486|ref|YP_001968011.1| hypothetical protein APP7_0217 [Actinobacillus pleuropneumoniae serovar 7 str. AP76]
MSELTHYIELKAIPQVDILQTDVIAHGLQILHKFLPLYQGEIGLSFPAYGLGRTLGGIIRVFGNEQHCTQ
IKTQLIGEGLQDYVLITSVTPVPEEIVEYHRYQRVHRKGQSAIRRTEQFLVQQGKWTEEIRQEMLIHQQN
QKVFPYVKLKSGSTKQHFVLAIRQLRLAEPVSGLFNAYGLSKIATVPHF (SEQ ID NO:97)

RNA recognition sequence
CUUCACUGCCGUAUAGGCAGCUUAGAAA (SEQ ID NO:96)

FIG. 5BB

>gi|303253029|ref|ZP_07339182.1| hypothetical protein APP2_1978 [Actinobacillus pleuropneumoniae serovar 2 str. 4226]
MSELTHYIELKAIPQVDILQTDVIAHGLQILHKFLPLYQGEIGLSFPAYGLGRTLGGIIRVFGNEQHCTQ
IKTQLIGEGLQDYVLITSVTPVPEEIVEYHRYQRVHRKGQSAIRRTEQFLVQQGKWTEEIRQEMLIHQQN
QKVFPHVKLKSGSTKQHFVLAIRQLRLAEPSFGLFNTYGLSKIATVPHF (SEQ ID NO:98)

FIG. 5BC

>gi|303251662|ref|ZP_07337835.1| hypothetical protein APP6_0866 [Actinobacillus pleuropneumoniae serovar 6 str. Femo]
MSELTHYIELKAIPQVDILQTDVIAHGLQILHKFLPLYQGEIGLSFPAYGLGRTLGGIIRVLGNEQHCTQ
IKTQLIGEGLQDYVLITSVTPVPEEIVEYHRYQRVHRKGQSAIRRTEQFLVQQGKWTEEIRQEMLIHQQN
QKVFPHVKLKSGSTKQHFVLAIRQLRLAEPSFGLFNTYGLSKIATVPHF (SEQ ID NO:99)

FIG. 5BD

>gi|116627507|ref|YP_820126.1| dephospho-CoA kinase [Streptococcus thermophilus LMD-9]
MSKTMIIGLTGGIASGKSTVVEIIKDAGYKVIDADQLVHDMQVKGGRLYQALLDWLGDGILLPNGELNRP
KLGQLIFSSEEMRYQSAEIQGKIIREELAAKRDCLAKEEDVFFMDIPLLFENDYQDWFDQIWLVAVSPQV
QGQRLMKRNHLSAEEAGMRIASQMPLAEKLPYASLVIDNNGNIDDLKKKVKGAIKDLANLV (SEQ ID NO:100)

FIG. 6

His29Ala

```
  1 mdhyldirlr pdpefppaql msvlfgklaq alvaqggdri gvsfpdldes rsrlgerlri
 61 hasaddlral larpwleglr dhlqfgepav vphptpyrqv srvqvksnpe rlrrrlmrrh
121 dlseeearkr ipdtvarald lpfvtlrsqs tgqhfrlfir hgplqataee ggftcyglsk
181 ggfvpwf  (SEQ ID NO:101)
```

His29Ala/Ser50Cys

```
  1 mdhyldirlr pdpefppaql msvlfgklaq alvaqggdri gvsfpdldec rsrlgerlri
 61 hasaddlral larpwleglr dhlqfgepav vphptpyrqv srvqvksnpe rlrrrlmrrh
121 dlseeearkr ipdtvarald lpfvtlrsqs tgqhfrlfir hgplqataee ggftcyglsk
181 ggfvpwf  (SEQ ID NO:102)
```

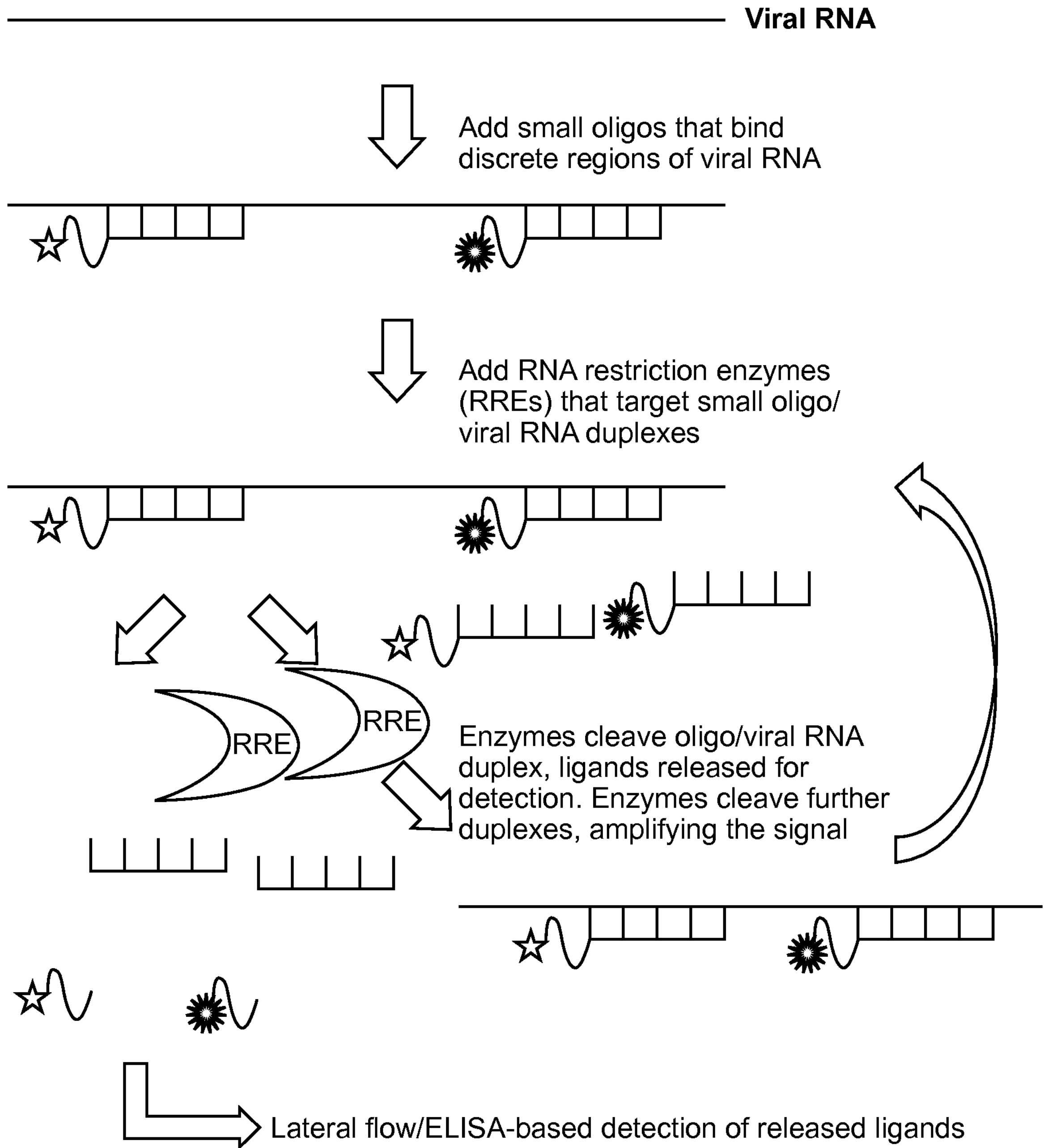

FIG. 7

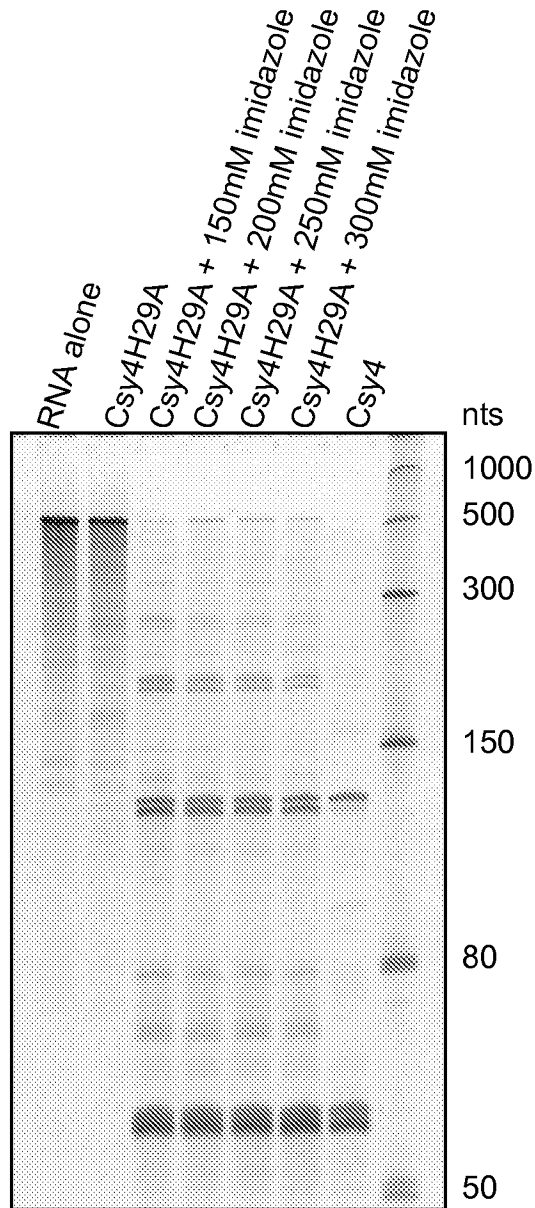

FIG. 8

G-U-U-C-A-C-U-G-C-C-G-U-A-U-A-G-G-C-A-G-NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNN

Csy4 target stem-loop

mRNA of choice

Cell lysate

Csy4 H29A

Csy4 H29A

Csy4 H29A

Csy4 H29A

Csy4 H29A

FIG. 9

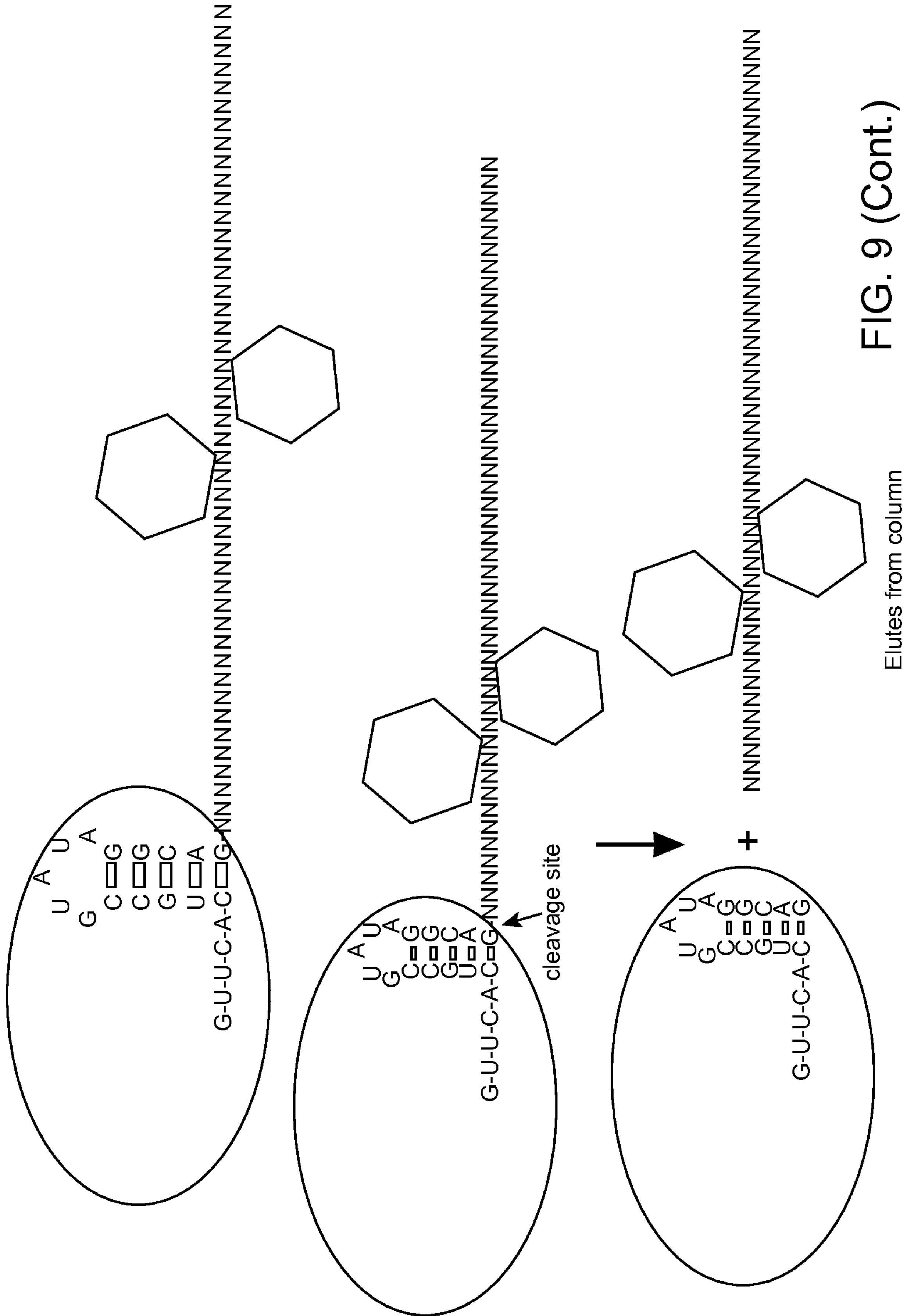

FIG. 9 (Cont.)

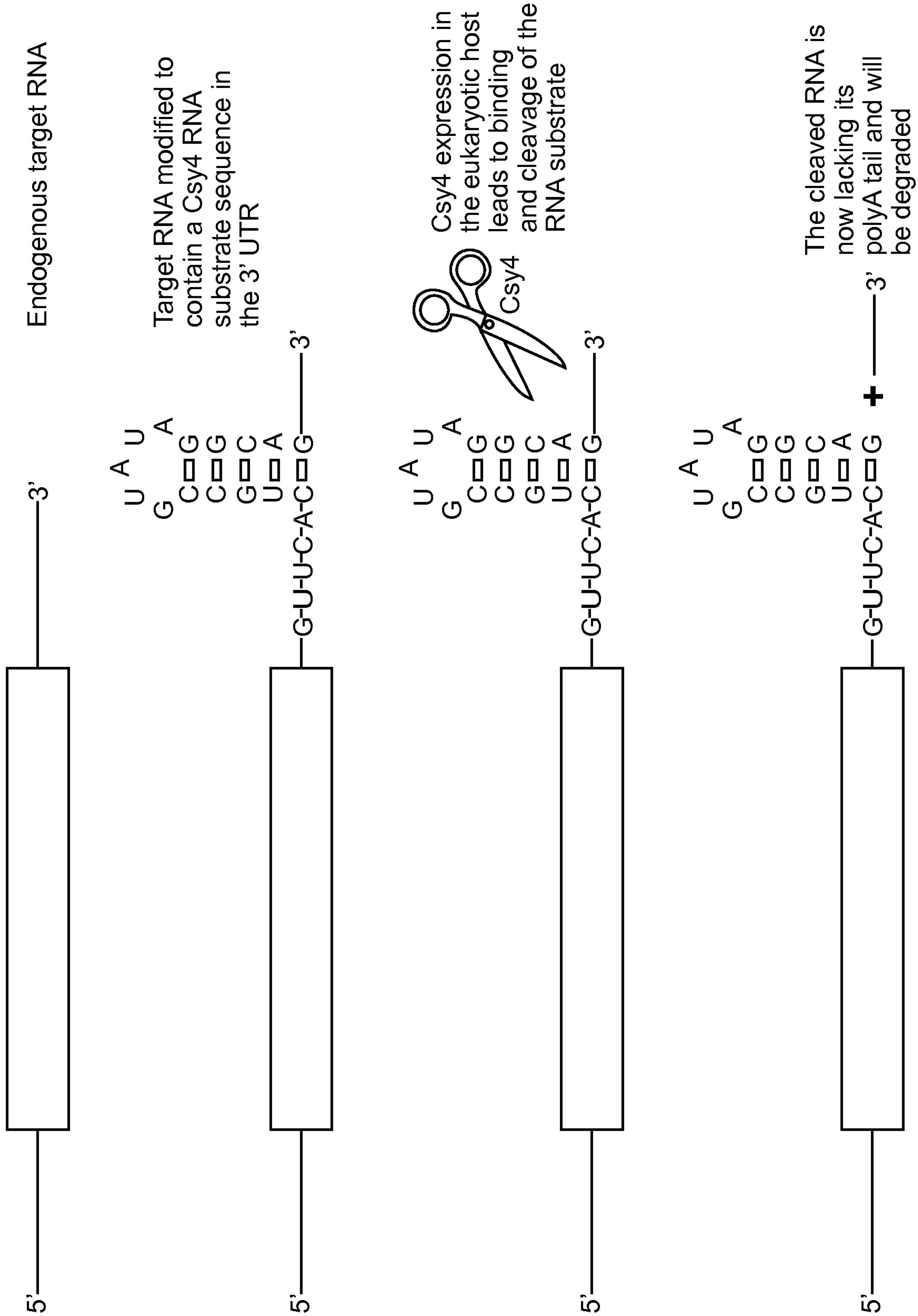

FIG. 10