



US 20250263803A1

(19) **United States**(12) **Patent Application Publication** (10) **Pub. No.: US 2025/0263803 A1**  
**FASCHING et al.** (43) **Pub. Date:** **Aug. 21, 2025**(54) **METHODS FOR QUANTITATION OF NUCLEIC ACID TARGETS**(71) Applicant: **Mammoth Biosciences, Inc.**, Brisbane, CA (US)(72) Inventors: **Clare Louise FASCHING**, Redwood City, CA (US); **Carley Gelenter HENDRIKS**, Burlingame, CA (US); **Bridget Ann Paine MCKAY**, Portola Valley, CA (US); **Janice Sha CHEN**, San Francisco, CA (US); **James Paul BROUGHTON**, South San Francisco, CA (US); **Vladyslava RATUSHNA**, Chula Vista, CA (US)(73) Assignee: **Mammoth Biosciences, Inc.**, Brisbane, CA (US)(21) Appl. No.: **19/203,260**(22) Filed: **May 9, 2025****Related U.S. Application Data**

(63) Continuation of application No. PCT/US2023/080111, filed on Nov. 16, 2023.

(60) Provisional application No. 63/384,212, filed on Nov. 17, 2022, provisional application No. 63/384,078, filed on Nov. 16, 2022.

**Publication Classification**(51) **Int. Cl.***C12Q 1/70* (2006.01)*C12Q 1/34* (2006.01)*C12Q 1/6806* (2018.01)*C12Q 1/6823* (2018.01)(52) **U.S. Cl.**CPC ..... *C12Q 1/70* (2013.01); *C12Q 1/34* (2013.01); *C12Q 1/6806* (2013.01); *C12Q 1/6823* (2013.01); *G01N 2333/922* (2013.01)

(57)

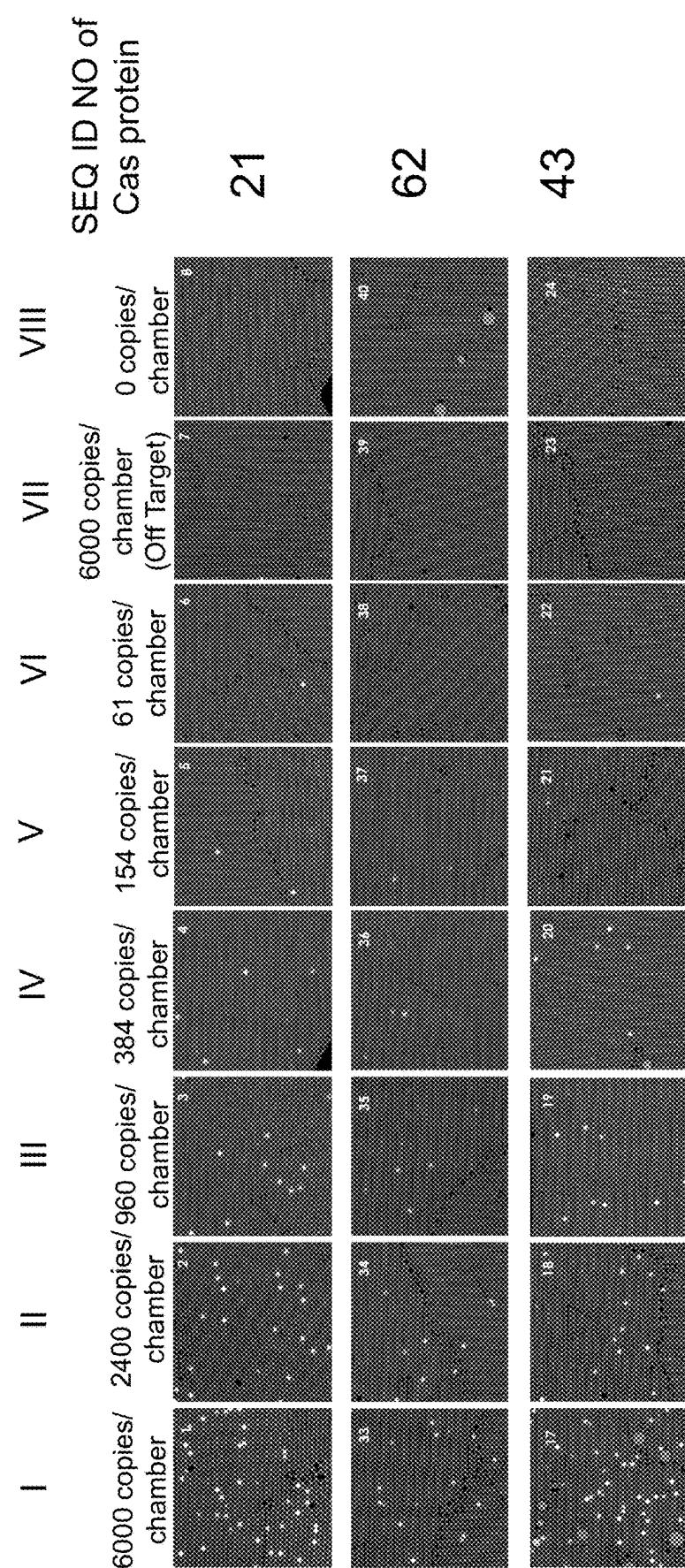
**ABSTRACT**

Provided herein are compositions, systems, and methods comprising effector proteins (e.g., CRISPR-associated (Cas) proteins), and uses thereof. Various compositions, systems, and methods of the present disclosure may leverage the activities of these effector proteins for the detection and quantitation of nucleic acids.

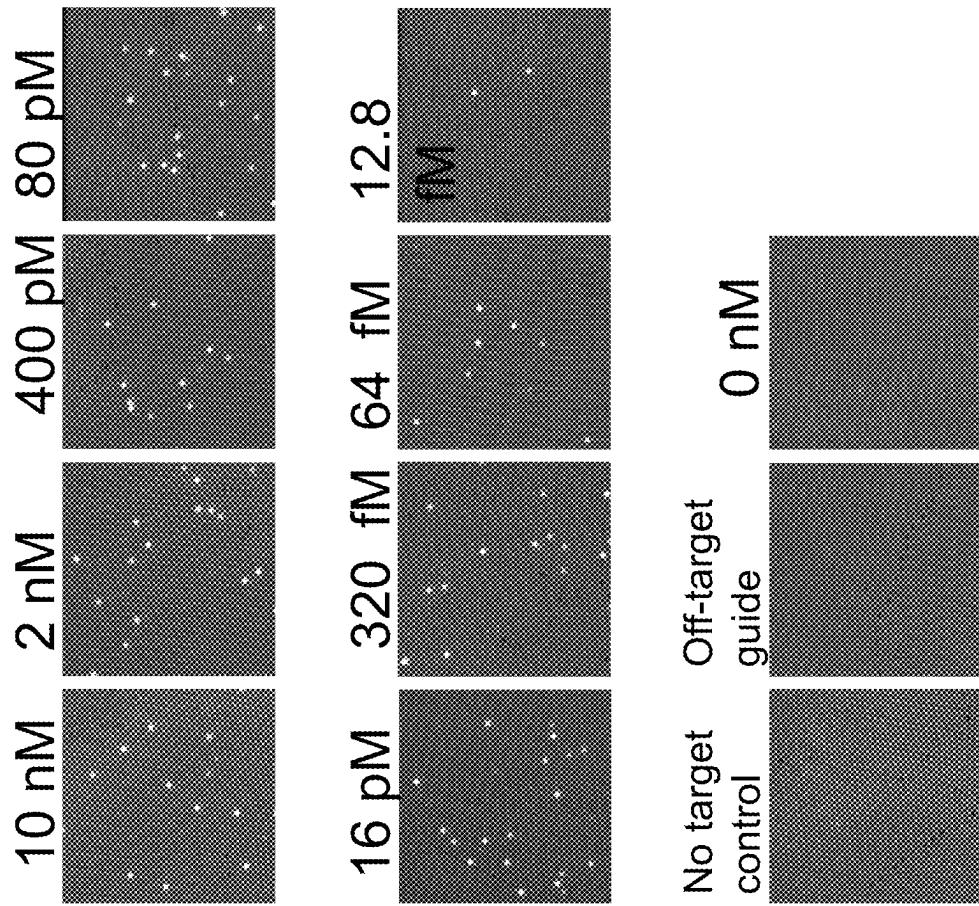
**Specification includes a Sequence Listing.**

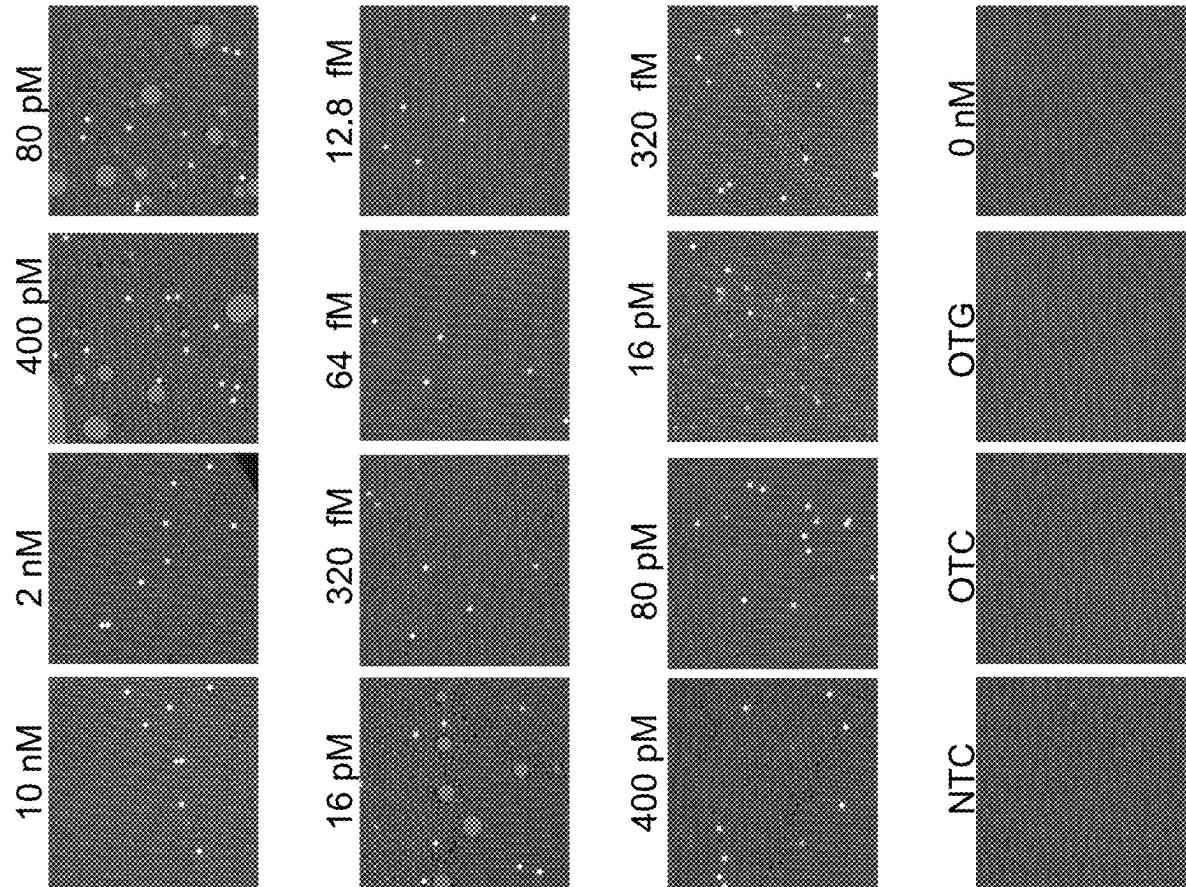
I	II	III	IV	V	VI	VII	VIII	SEQ ID NO of Cas protein
6000 copies/ chamber	2400 copies/ chamber	960 copies/ chamber	384 copies/ chamber	154 copies/ chamber	61 copies/ chamber	6000 copies/ chamber (Off Target)	0 copies/ chamber	21
								62
								43

**FIG. 1**

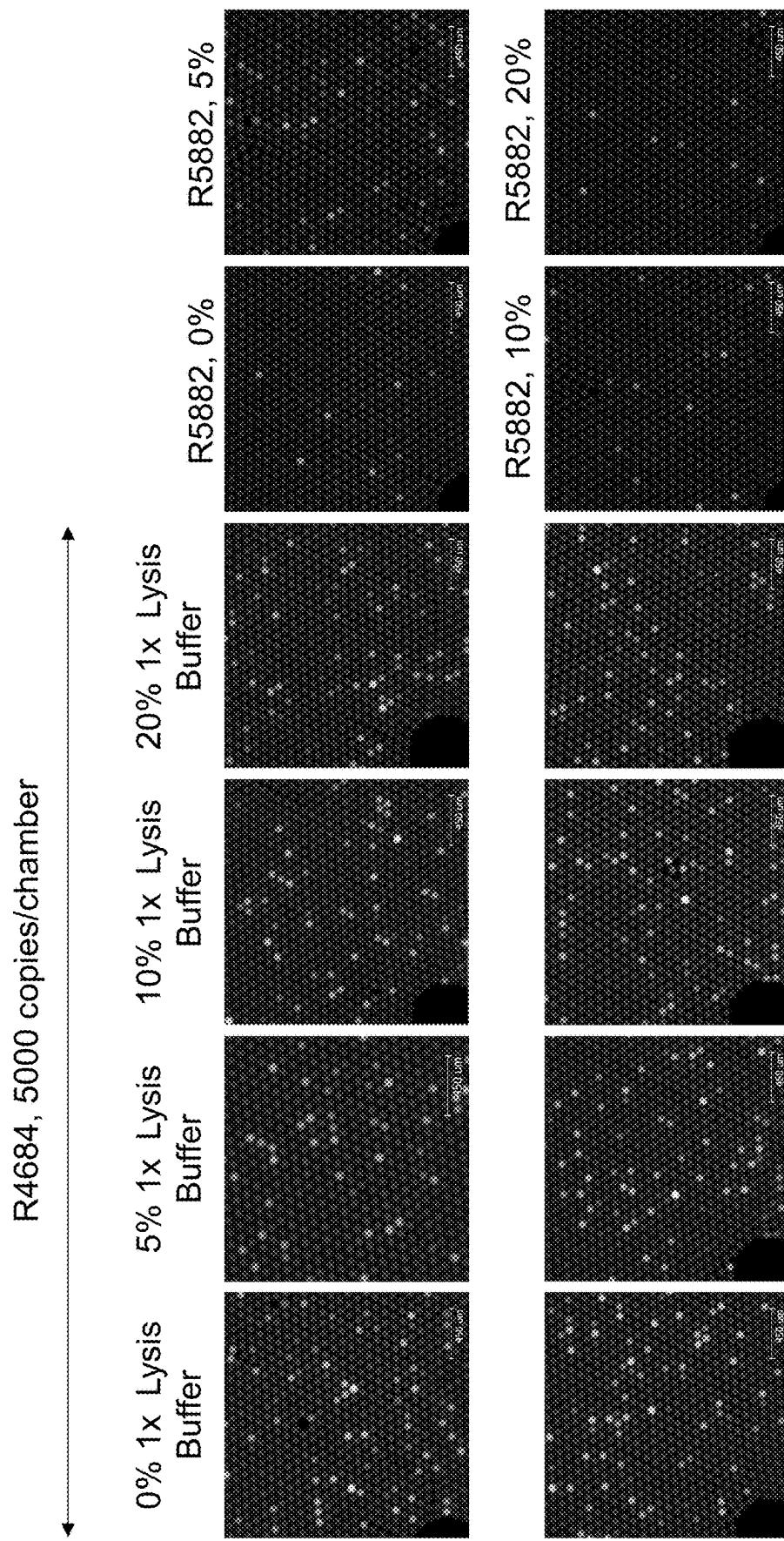


**FIG. 2**

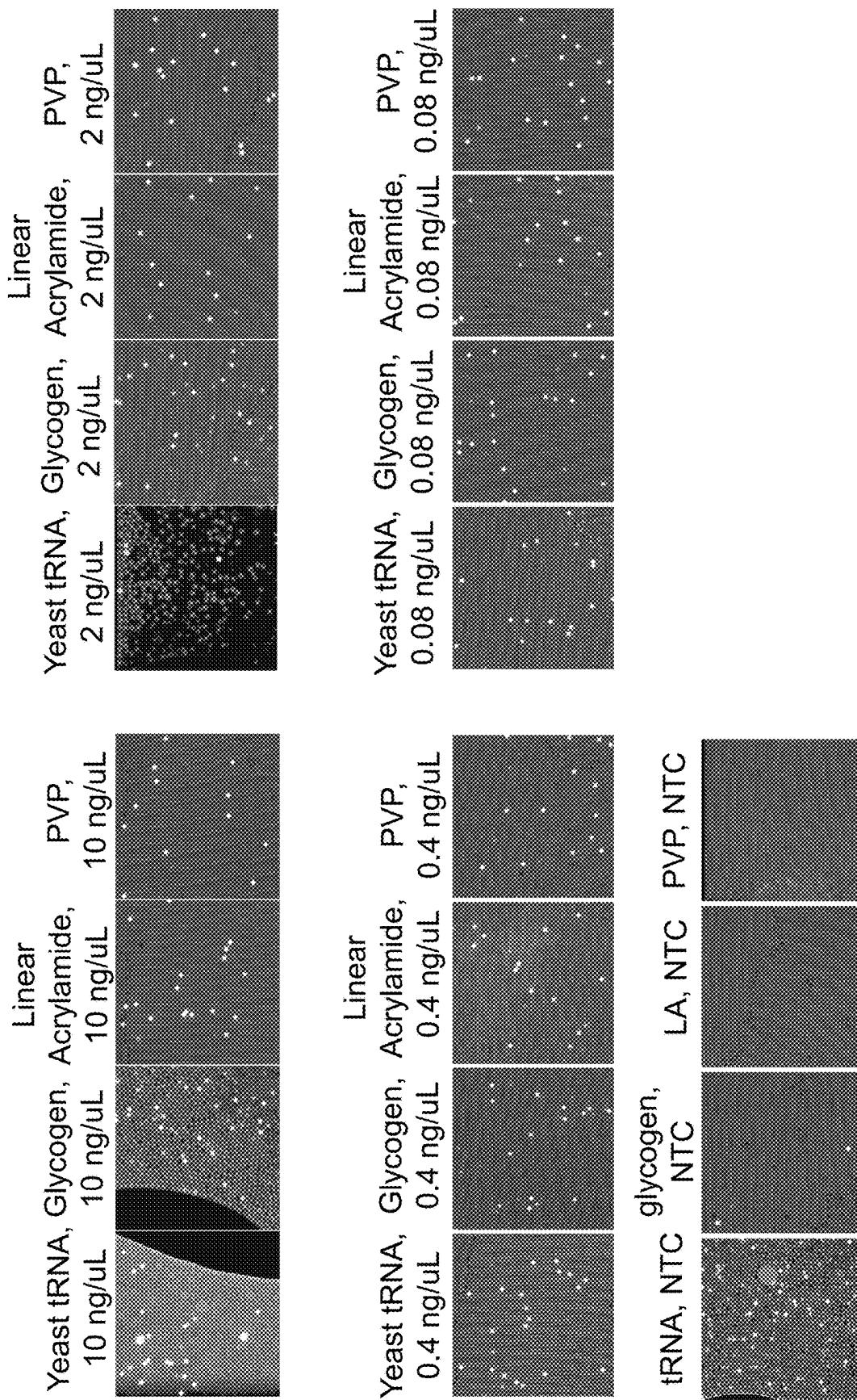




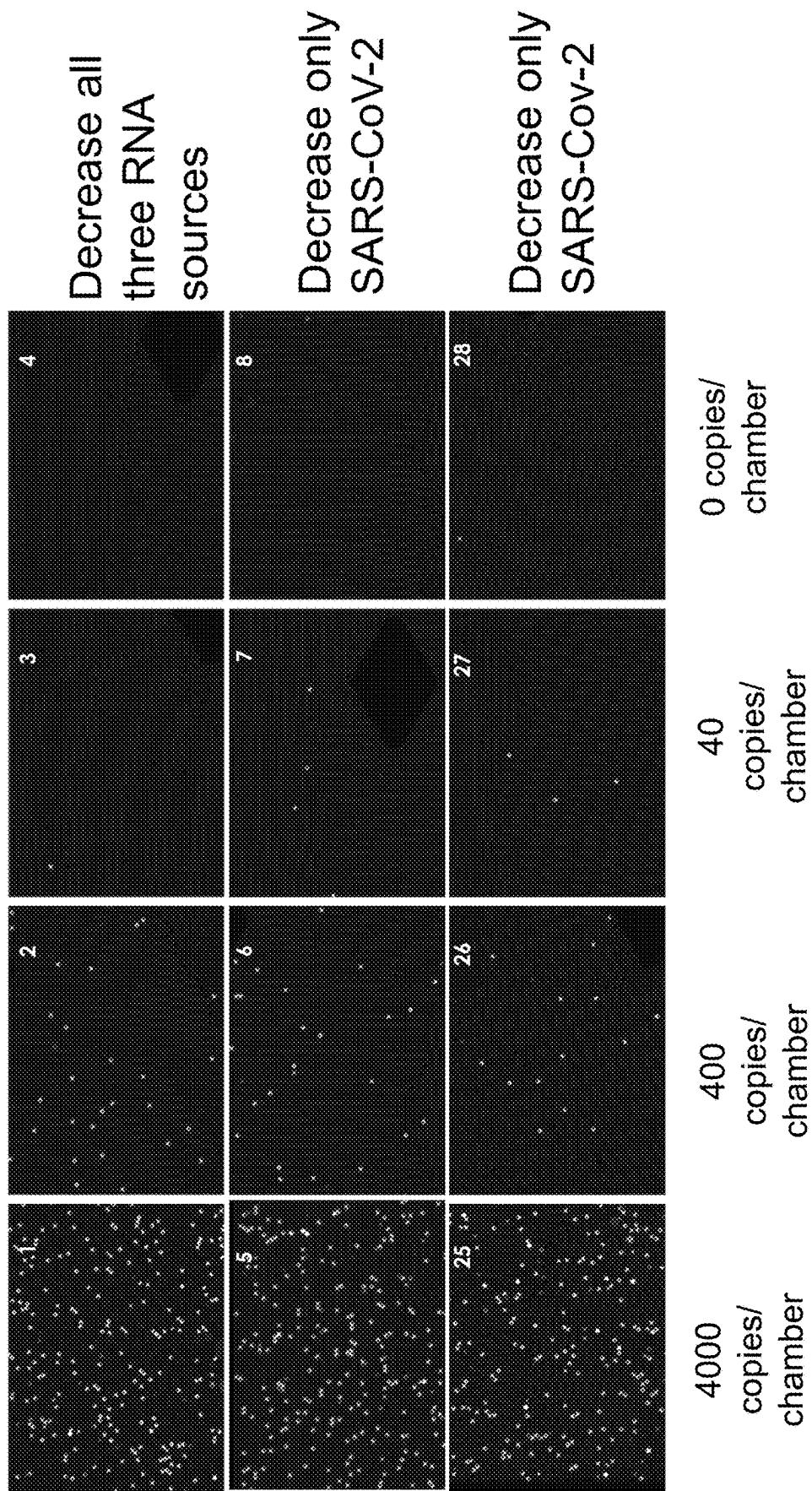
**FIG. 4**

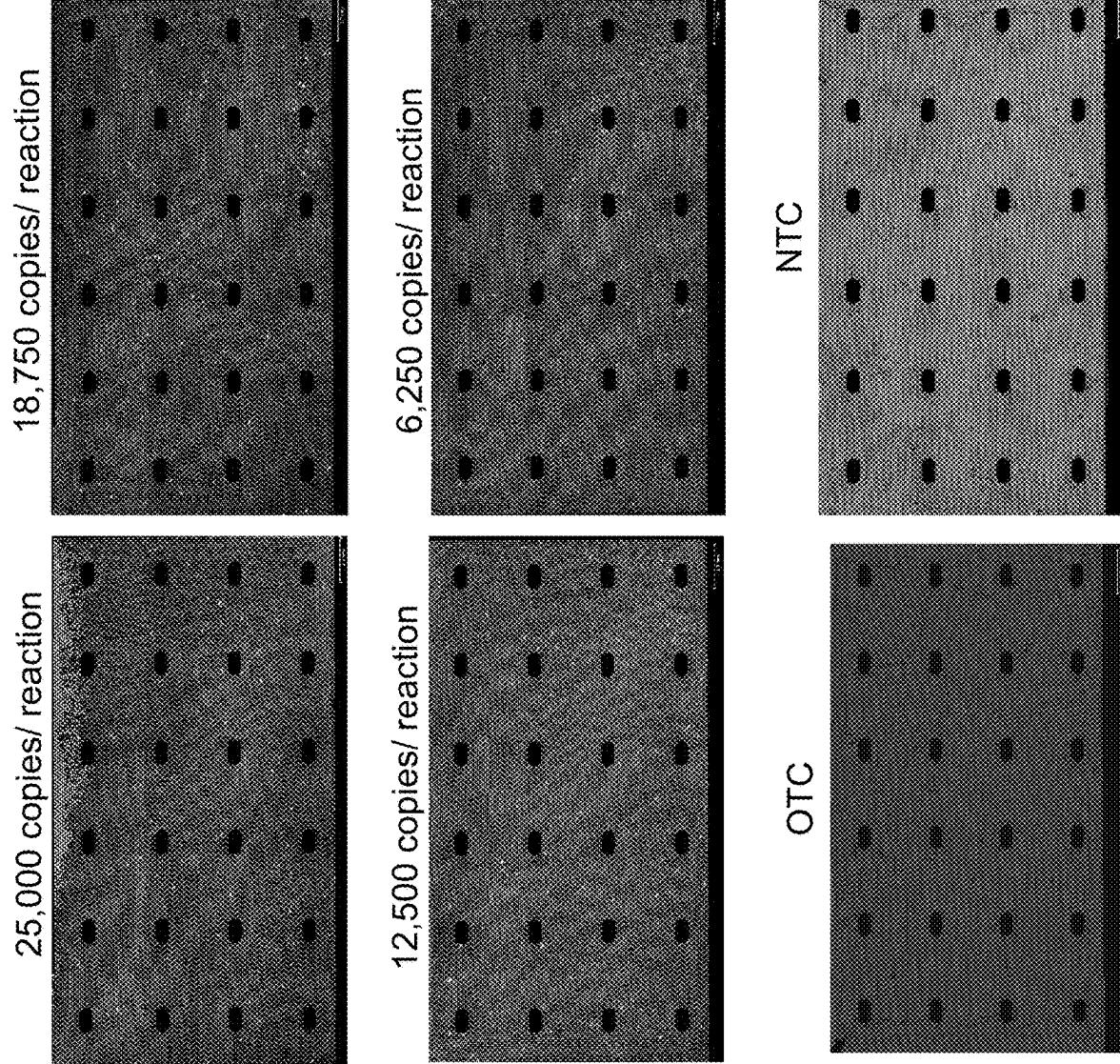


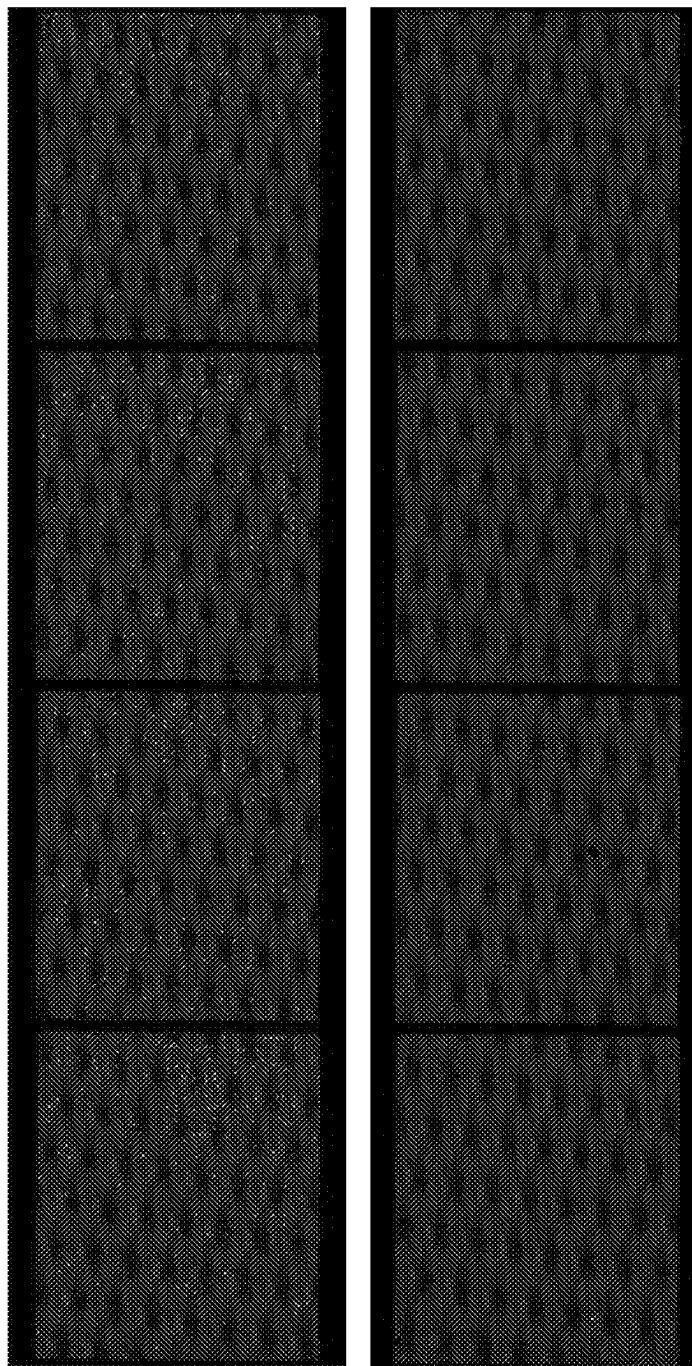
**FIG. 5**



**FIG. 6**





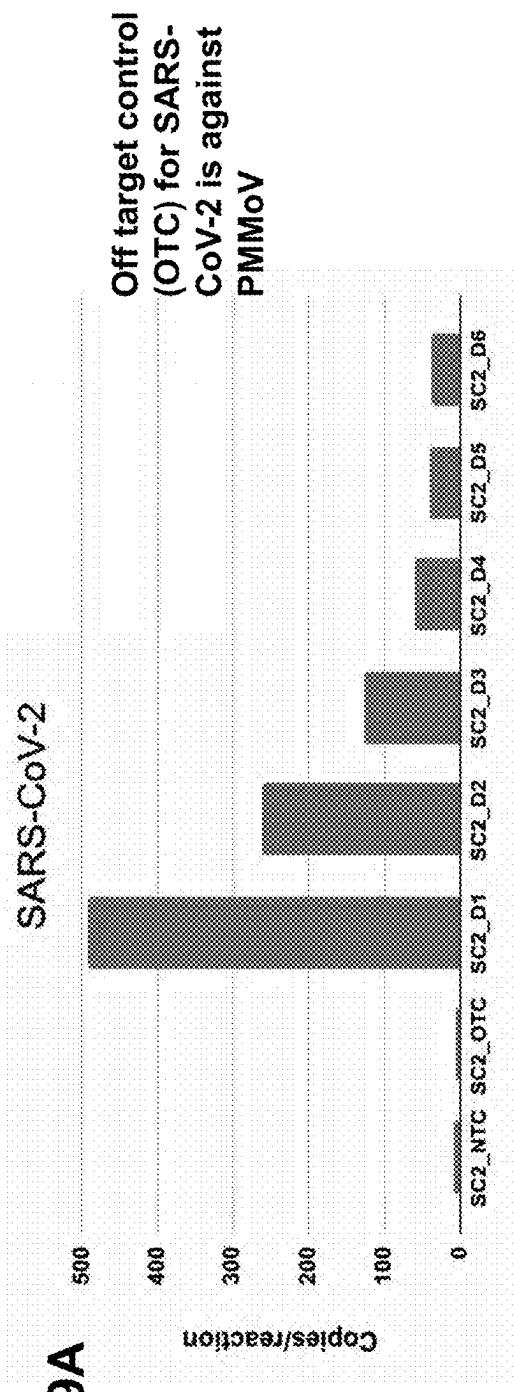


**FIG. 8**

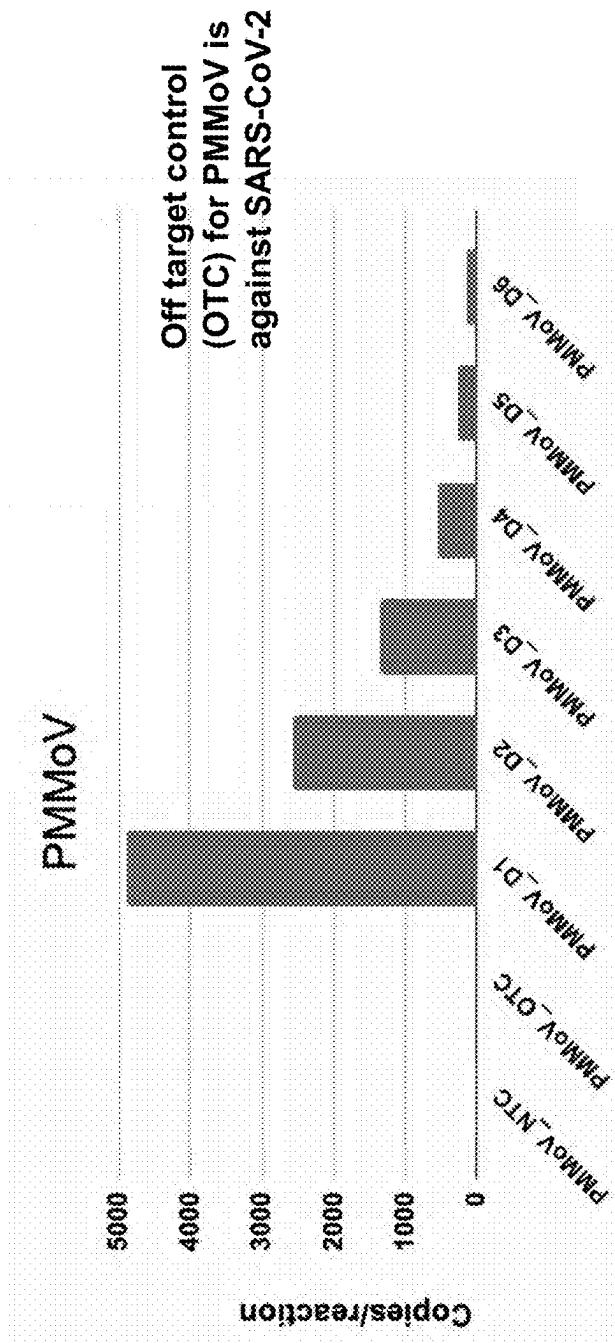
SARS-CoV-2

NTC

**FIG. 9A**



**FIG. 9B**



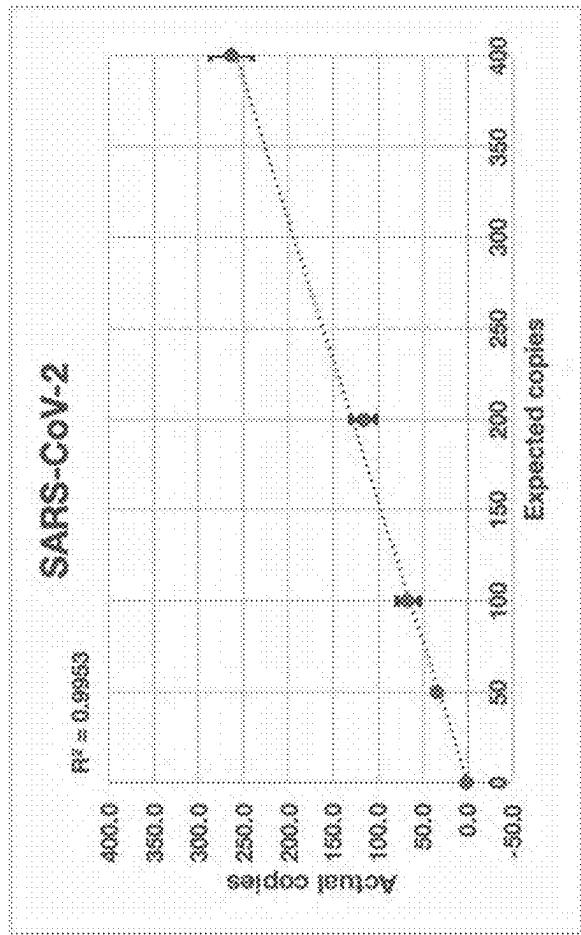


FIG. 10A

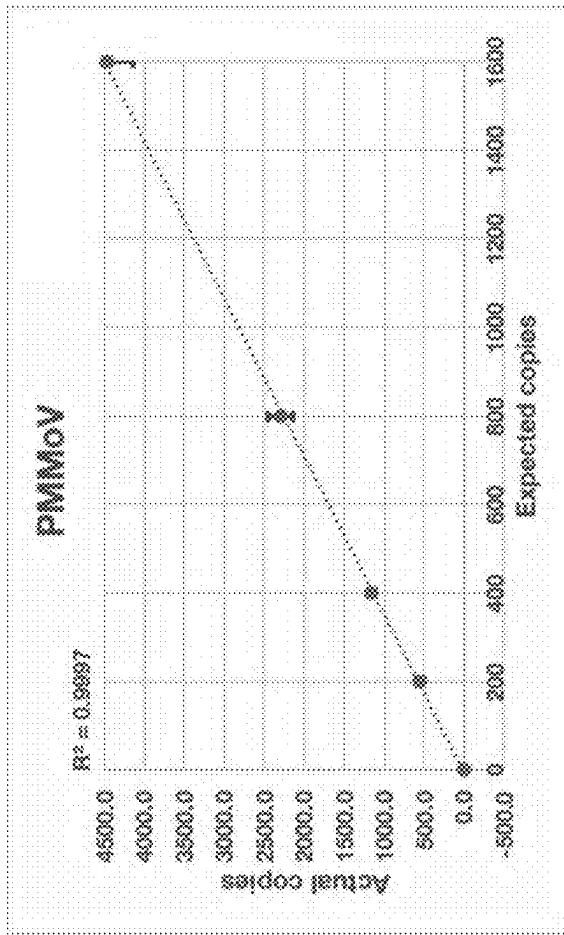
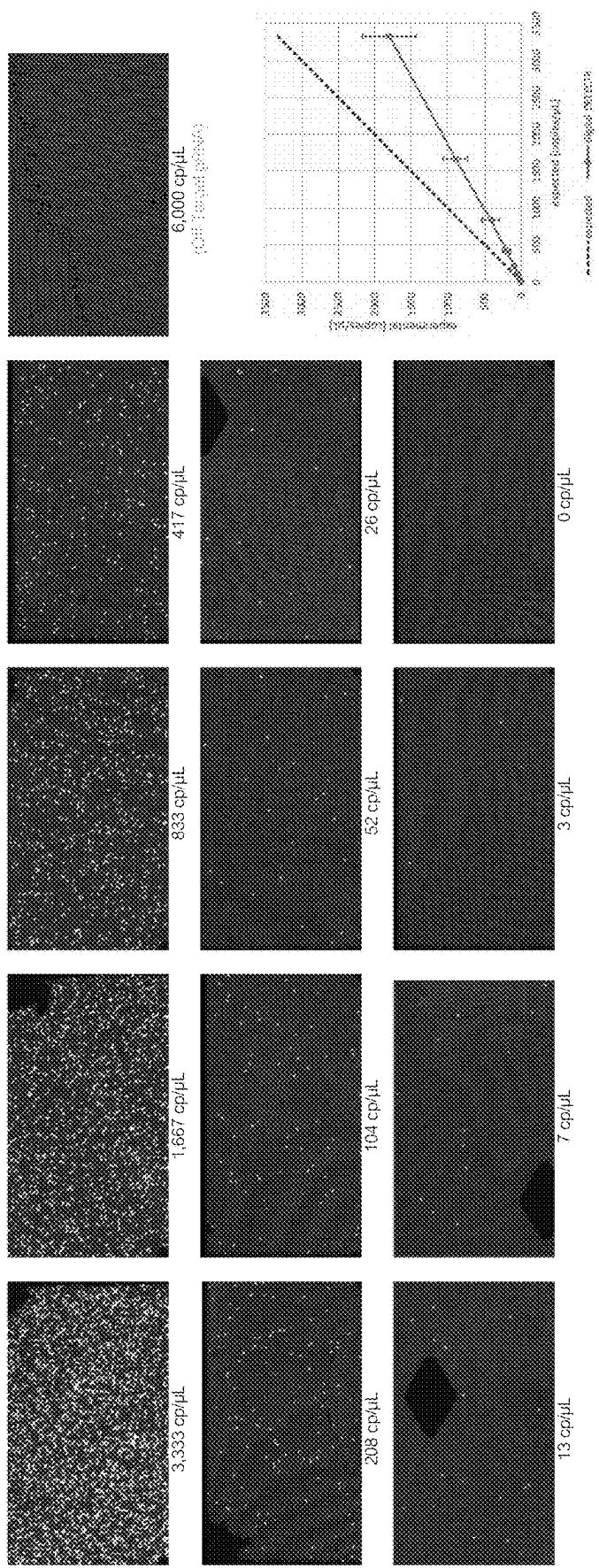


FIG. 10B



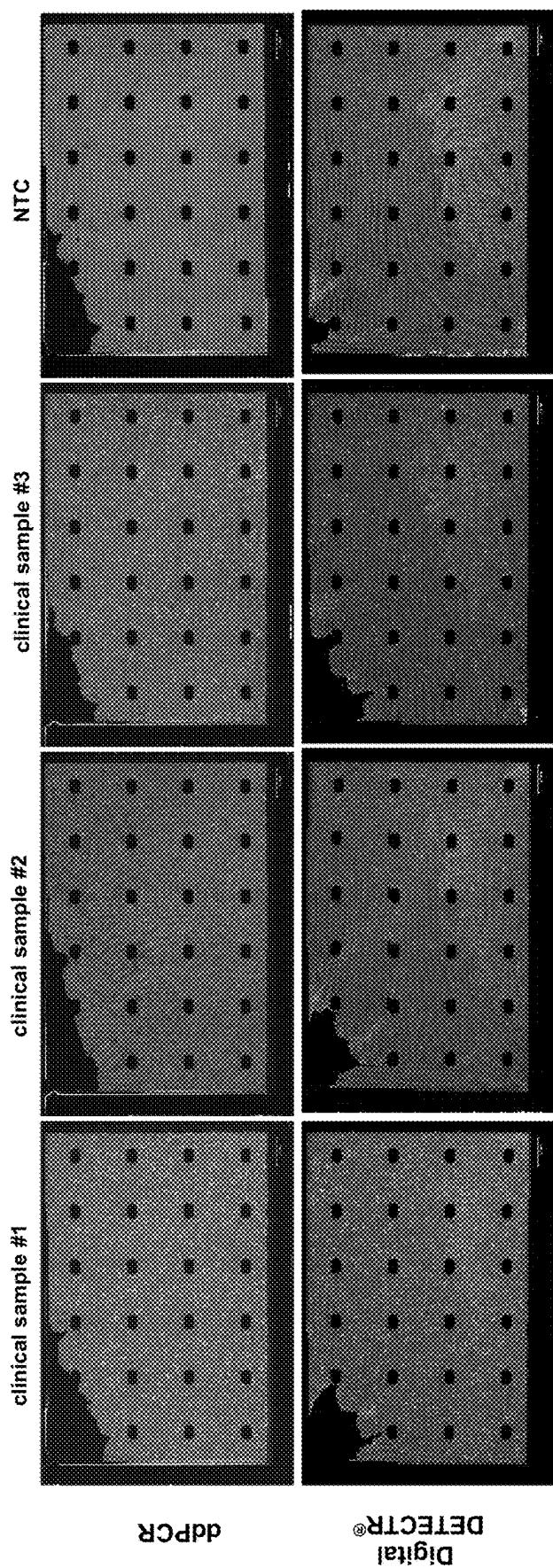


FIG. 12

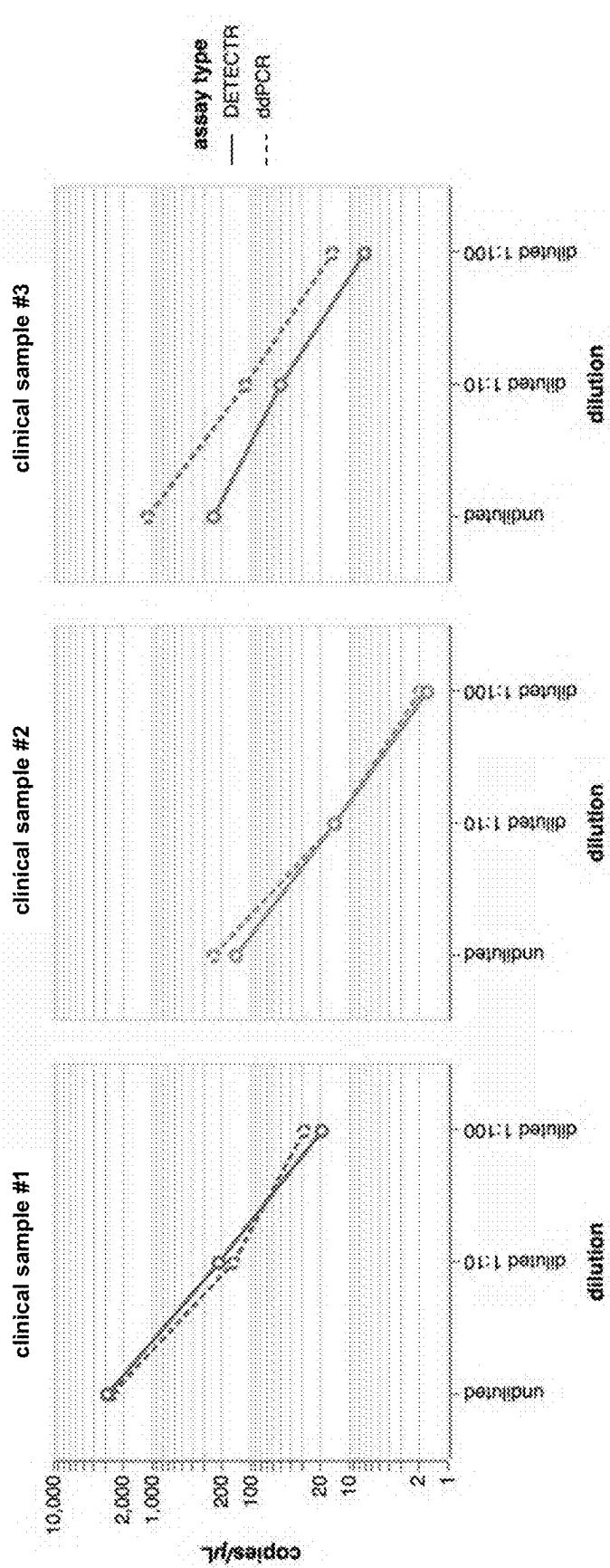
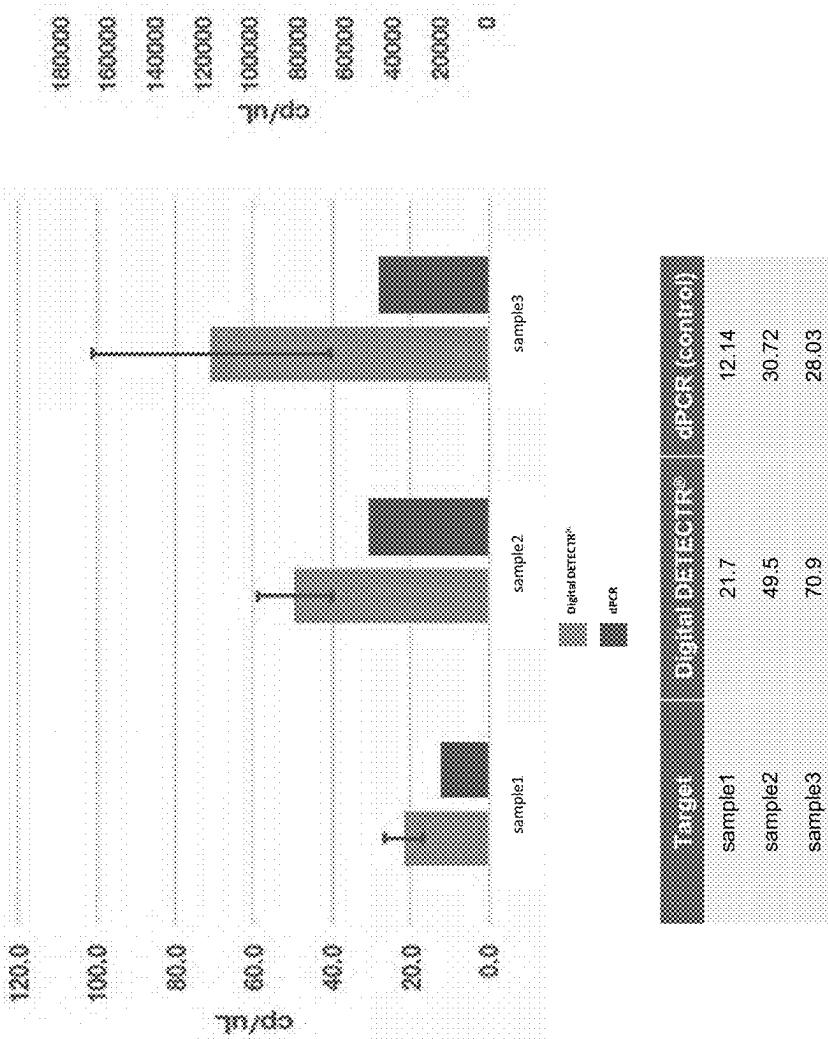


FIG. 13

SARS-CoV-2



PMMOV

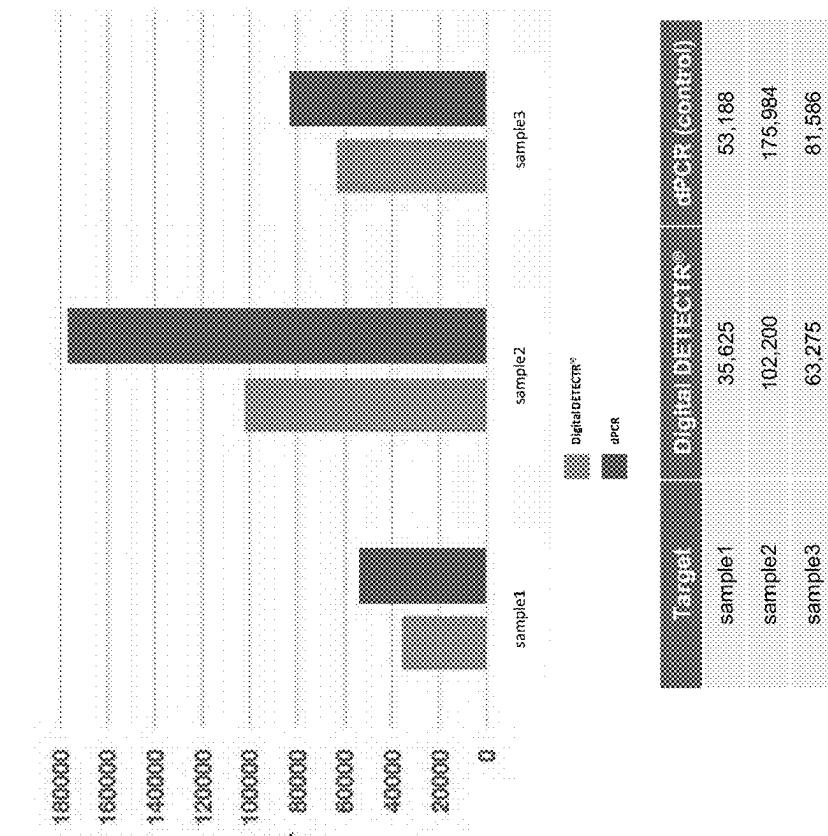


FIG. 14

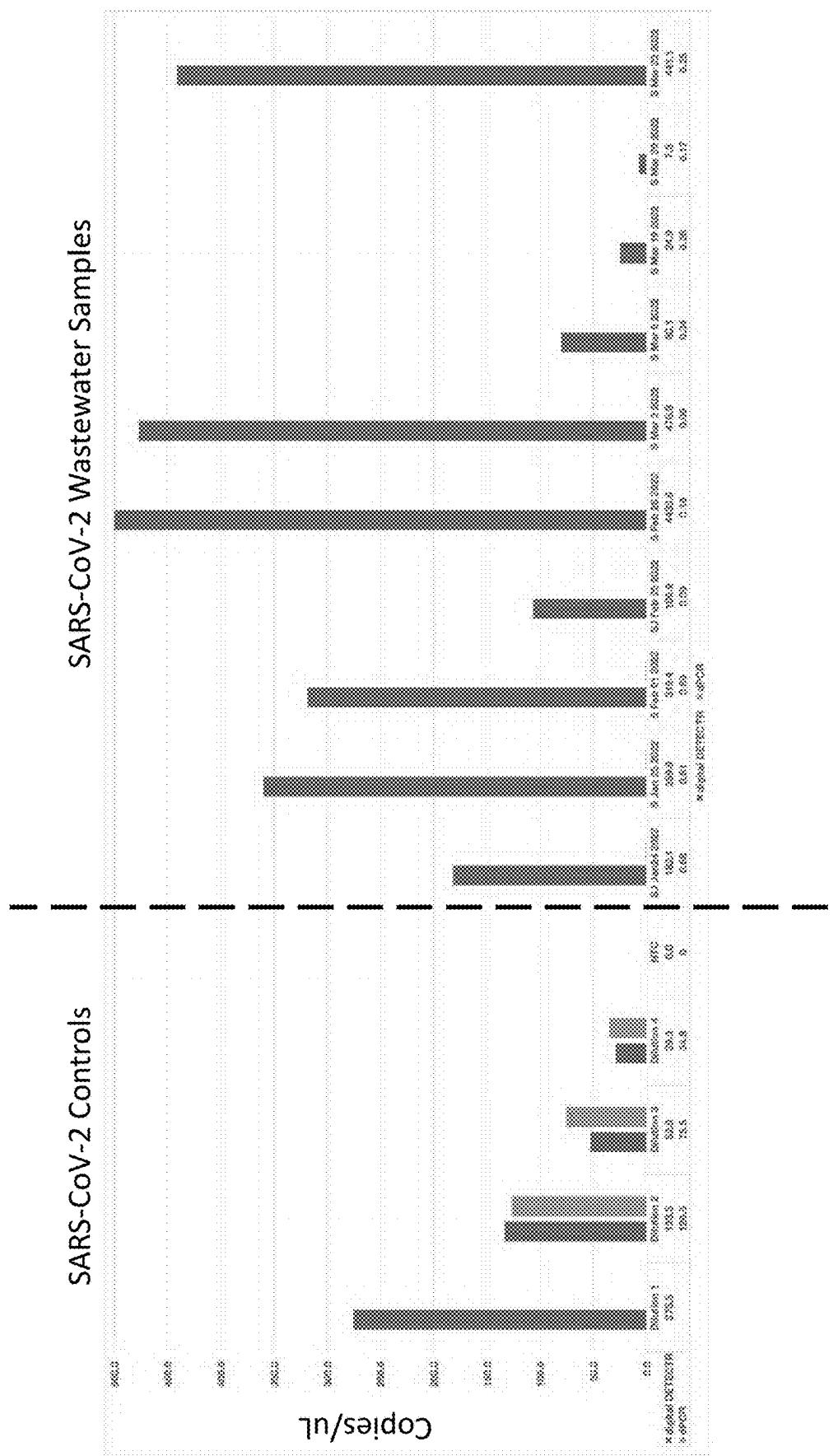


FIG. 15A

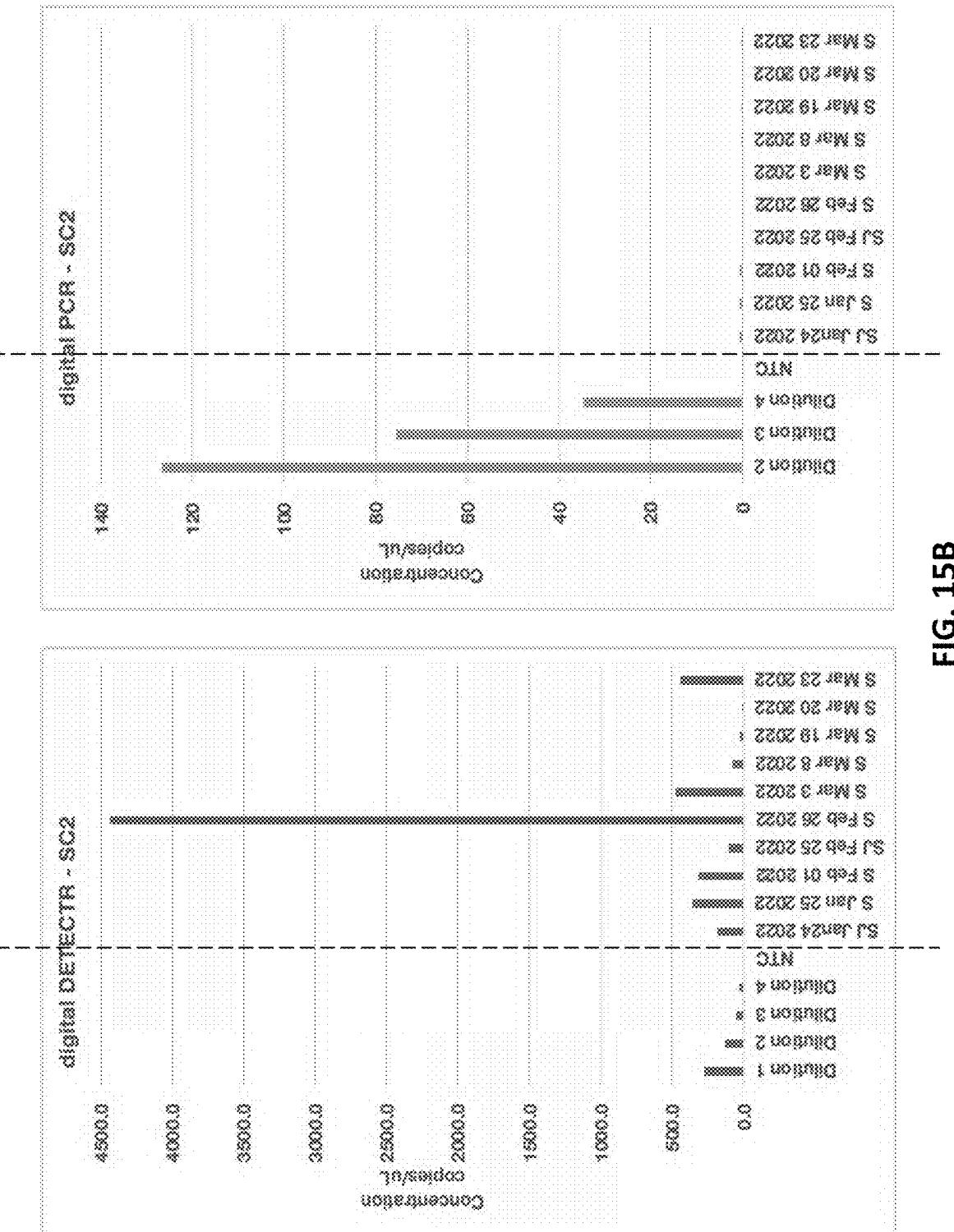
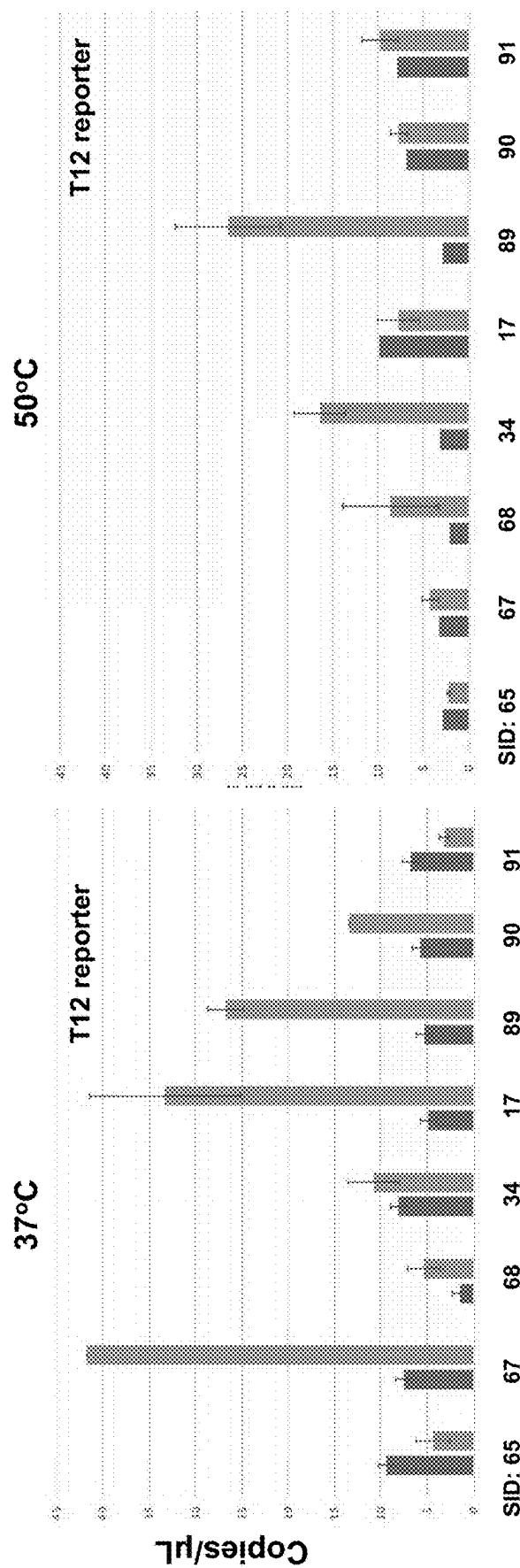


FIG. 15B



**FIG. 16**

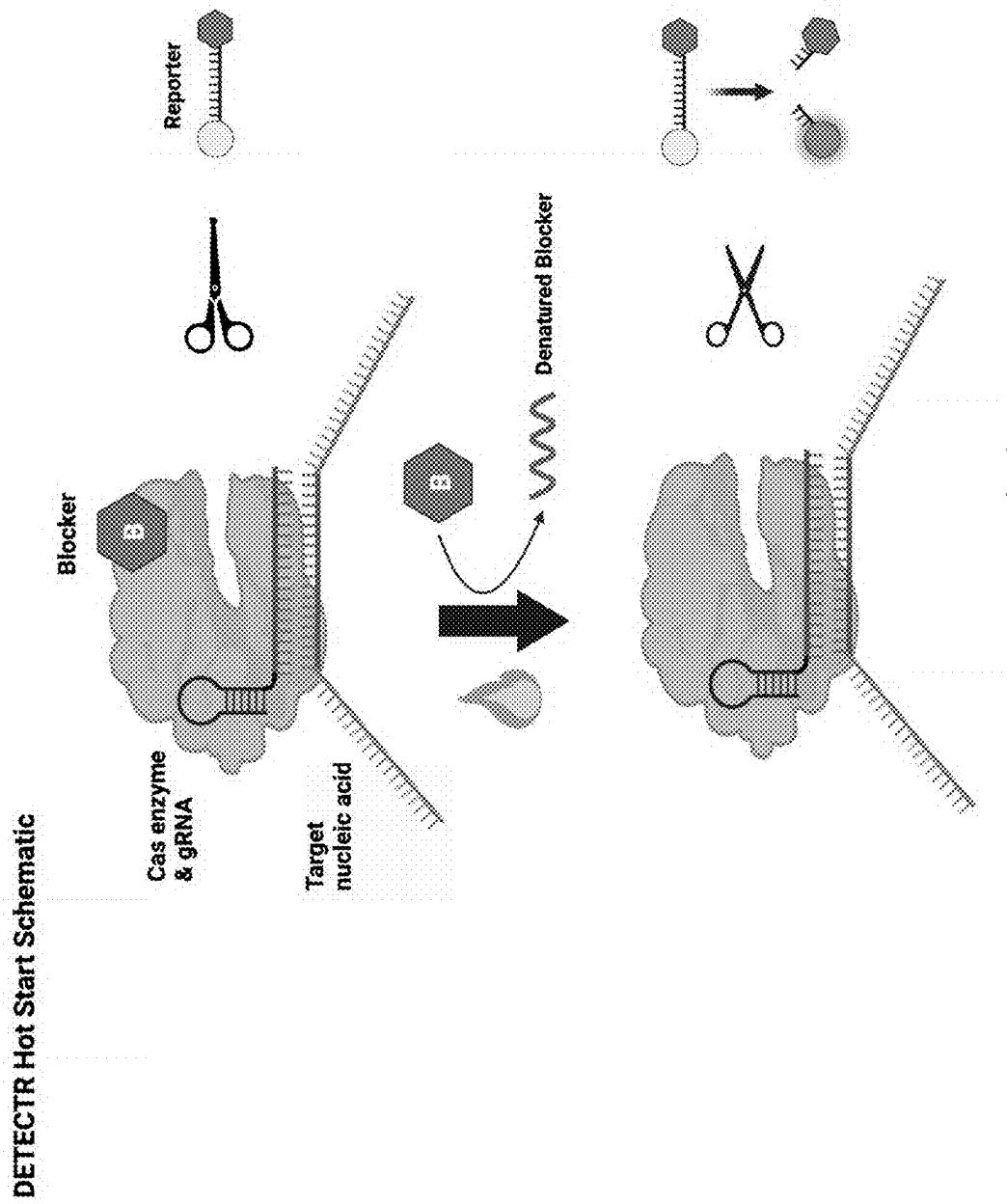


FIG. 17

## METHODS FOR QUANTITATION OF NUCLEIC ACID TARGETS

### CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application is a continuation of International PCT Application No. PCT/US2023/080111, filed Nov. 16, 2023, which claims priority to U.S. Provisional Patent Application No. 63/384,078, filed Nov. 16, 2022, and U.S. Provisional Patent Application No. 63/384,212, filed Nov. 17, 2022, each of which is incorporated by reference herein in its entirety for all purposes.

### REFERENCE TO AN ELECTRONIC SEQUENCE LISTING

[0002] The contents of the electronic sequence listing (MABI\_027\_02US\_SeqList\_ST26.xml; Size: 669,197 bytes; and Date of Creation: Apr. 28, 2025) are herein incorporated by reference in their entirety.

### BACKGROUND

[0003] The ability to quantitate the amount of target nucleic acid in a sample without a need to amplify the target nucleic acid prior to detection would be valuable in diagnostics, therapeutics, and epidemiology. For instance, quantitating the amount of viral nucleic acid in a patient sample over time can help with designing a treatment strategy, and quantitating the amount of viral nucleic acid in wastewater over time can be used to predict and prevent the spread of infectious diseases. Thus, there continues to be a need for assays for the quantitative detection of nucleic acids.

### SUMMARY

[0004] In a first aspect, the disclosure provides methods of quantitating a target nucleic acid in a test sample, the method comprising: (a) contacting the test sample with the following components (i) through (iii), resulting in at least two nanovolumes of reaction mixture: (i) a CRISPR/Cas effector protein, (ii) a guide RNA, comprising a region that is capable of binding to the CRISPR/Cas effector protein, and a guide sequence that is capable of hybridizing with the target nucleic acid, (iii) a reporter nucleic acid that is single stranded and does not hybridize with the guide sequence of the guide RNA; (b) measuring detectable signals detected from the at least two nanovolumes and generated by cleavage of the reporter nucleic acid by the CRISPR/Cas effector protein, and (c) quantitating the target nucleic acid in the test sample based on the measured signals from the at least two nanovolumes.

[0005] In some embodiments, the contacting step comprises sequentially adding each of the components (i) through (iii) to the test sample in the at least two nanovolumes. In some embodiments, the contacting step comprises: (a) adding each of the components (i) through (iii) to the test sample to generate a master reaction mixture, wherein the master reaction mixture has a volume of more than 1 nL, and (b) distributing the master reaction mixture into the at least two nanovolumes. In some embodiments, each of the at least two nanovolumes comprises no more than 1 molecule of the target nucleic acid. In some embodiments, the CRISPR/Cas effector protein and the guide RNA are incubated with each other prior to step (a). In some embodiments, the CRISPR/Cas effector protein and the guide RNA are not incubated with each other prior to step (a).

Cas effector protein and the guide RNA are not incubated with each other prior to step (a).

[0006] In some embodiments, the methods further comprise quantitating multiple target nucleic acids in the test sample, wherein the at least two nanovolumes of reaction mixture comprises one or more guide RNAs, wherein at least one of the one or more guide RNAs comprises a guide sequence that is capable of hybridizing with each of the multiple target nucleic acids. In some embodiments, each of the at least two nanovolumes of reaction mixture comprises at least one of the one or more guide RNAs comprising a guide sequence that is capable of hybridizing with each of the multiple target nucleic acids. In some embodiments, each of the at least two nanovolumes of reaction mixture comprises at least one of the one or more guide RNAs comprising a guide sequence that is capable of hybridizing with no more than one of the multiple target nucleic acids.

[0007] In some embodiments, the method does not comprise amplifying the target nucleic acid in the test sample. In some embodiments, the test sample comprises about 10,000 molecules to about 100,000 molecules of the target nucleic acid. In some embodiments, the test sample comprises about 50,000 molecules of the target nucleic acid. In some embodiments, the contacting step results in a number of nanovolumes in the range of about 5000 nanovolumes to about 100,000 nanovolumes. In some embodiments, step (b) comprises measuring a binary signal from each of the at least two nanovolumes. In some embodiments, the methods comprise contacting the test sample with a precursor guide RNA array, wherein the CRISPR/Cas effector protein cleaves the precursor guide RNA array to produce the guide RNA.

[0008] In some embodiments, the target nucleic acid is DNA or RNA. In some embodiments, the target nucleic acid is a viral nucleic acid or a bacterial nucleic acid. In some embodiments, the target nucleic acid is a viral nucleic acid. In some embodiments, the target nucleic acid is derived from a papovavirus, a human papillomavirus (HPV), a hepadnavirus, a Hepatitis B Virus (HBV), a herpesvirus, a varicella zoster virus (VZV), an Epstein Barr virus (EBV), a Kaposi's sarcoma-associated herpesvirus, an adenovirus, a poxvirus, a parvovirus, an influenza virus, a respiratory syncytial virus, or a coronavirus. In some embodiments, the target nucleic acid is derived from a severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2).

[0009] In some embodiments, the target nucleic acid is derived from a human cell. In some embodiments, the target nucleic acid is a human fetal nucleic acid or a cancer cell nucleic acid. In some embodiments, the target nucleic acid is single stranded. In some embodiments, the target nucleic acid is double stranded. In some embodiments, the test sample comprises DNA from a cell lysate. In some embodiments, the test sample comprises cells. In some embodiments, the test sample is a blood, serum, plasma, urine, aspirate, fecal or biopsy sample.

[0010] In some embodiments, the methods further comprise quantitating a positive control target nucleic acid in a positive control sample, the method comprising: (a) contacting the positive control sample with the following components (i) through (iii) resulting in at least two nanovolumes of reaction mixture: (i) a CRISPR/Cas effector protein; (ii) a positive control guide RNA, comprising a region that is capable of binding to the CRISPR/Cas effector protein, and a guide sequence that is capable of hybridizing with the positive control target nucleic acid; and (iii) a reporter

nucleic acid that is single stranded and does not hybridize with the guide sequence of the positive control guide RNA; (b) measuring detectable signals detected from the at least two nanovolumes and produced by cleavage of the reporter nucleic acid by the CRISPR/Cas effector protein, and (c) quantitating the amount of positive control target nucleic acid in the positive control sample based on the measured signals from the at least two nanovolumes.

[0011] In some embodiments, the methods further comprise quantitating the target nucleic acid in a positive control sample, the method comprising: (a) contacting the positive control sample with the following components (i) through (iii) resulting in at least two nanovolumes of reaction mixture: (i) a CRISPR/Cas effector protein; (ii) the guide RNA, comprising a region that is capable of binding to the CRISPR/Cas effector protein, and a guide sequence that is capable of hybridizing with the target nucleic acid; and (iii) a reporter nucleic acid that is single stranded and does not hybridize with the guide sequence of the guide RNA; (b) measuring detectable signals detected from the at least two nanovolumes produced by cleavage of the reporter nucleic acid by the CRISPR/Cas effector protein, and (c) quantitating the amount of target nucleic acid in the positive control sample based on the measured signals from the at least two nanovolumes.

[0012] In some embodiments, the method comprises generating a standard curve for the target nucleic acid in the positive control sample, and obtaining an absolute quantitation of the target nucleic acid in the test sample based on the standard curve. In some embodiments, the method comprises obtaining a relative quantitation of the target nucleic acid in the test sample based on the quantitation of the target nucleic acid in a positive control sample. In some embodiments, the detectable signal is detectable in less than 90 minutes. In some embodiments, the detectable signal is detectable in less than 30 minutes. In some embodiments, the CRISPR/Cas effector protein comprises an amino acid sequence with at least 90% identity to the amino acid sequence of any one of SEQ ID NOS: 1-72. In some embodiments, the CRISPR/Cas effector protein comprises the amino acid sequence of any one of SEQ ID NOS: 1-72.

[0013] In some embodiments, the target nucleic acid is an RNA and the CRISPR/Cas effector protein is an RNA-targeting CRISPR/Cas effector protein. In some embodiments, the RNA-targeting CRISPR/Cas effector protein comprises an amino acid sequence with at least 90% identity to the amino acid sequence of SEQ ID NO: 21, 62, 43, 41, or 42. In some embodiments, the RNA-targeting CRISPR/Cas effector protein comprises the amino acid sequence of SEQ ID NO: 21, 62, 43, 41, or 42.

[0014] In some embodiments, the target nucleic acid is a DNA and the CRISPR/Cas effector protein is a DNA-targeting CRISPR/Cas effector protein. In some embodiments, the DNA-targeting CRISPR/Cas effector protein comprises an amino acid sequence with at least 90% identity to the amino acid sequence of SEQ ID NO: 3, 34, 57, 36, 65, 67, 68, 89, 90, 91, or 17. In some embodiments, the DNA-targeting CRISPR/Cas effector protein comprises the amino acid sequence of SEQ ID NO: 3, 34, 57, 36, 65, 67, 68, 89, 90, 91, or 17.

[0015] In some embodiments, the reaction mixture comprises a buffer, wherein the buffer comprises tricine, MgOAc, BSA, TCEP, imidazole, KCl, MgCl<sub>2</sub>, BSA, Igepal Ca-630, glycerol, HEPES, KOAc, Triton-X 100, Tris-HCl,

(NH4)2SO<sub>4</sub>, Tween-20, TMAO, or any combination thereof. In some embodiments, the reporter nuclear acid is a RNA. In some embodiments, the reporter nuclear acid is a DNA. In some embodiments, the reporter nucleic acid comprises a modified nucleobase, a modified sugar moiety, and/or a modified nucleic acid linkage.

[0016] In another aspect, the present disclosure provides a method of assaying for a target nucleic acid in a sample, the method comprising a) amplifying the target nucleic acid using at least one amplification primer; b) contacting the sample to a reporter and a composition comprising a programmable nuclease and a guide nucleic acid that hybridizes to the target nucleic acid or an amplified product thereof, wherein the programmable nuclease cleaves the reporter upon hybridization of the guide nucleic acid to the target nucleic acid or the amplification product thereof; and c) assaying for a change in a signal, wherein the change in the signal is produced by cleavage of the reporter; wherein the target nucleic acid is a gene of a monkeypox virus or a segment thereof; and optionally wherein the at least one amplification primer comprises a nucleotide sequence at least 85%, at least 87%, at least 90%, at least 92%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to any one of SEQ ID NOS: 92-235.

[0017] In some embodiments, the gene of the monkeypox virus is selected from the group consisting of OPG123, OPG038, OPG094, OPG037, OPG151, OPG105, and OPG199.

[0018] In some embodiments, the at least one amplification primer comprises at least six amplification primers. The at least six amplification primers may comprise a FIP primer, a BIP primer, a B3 primer, a F3 primer, a LB primer, and a LF primer. In some embodiments, the FIP primer comprises a nucleotide sequence at least 85%, at least 87%, at least 90%, at least 92%, at least 94%, at least 95%, at least 96%, at least 98%, or 100% identical to any one of SEQ ID NOS: 94, 100, 106, 112, 118, 124, 130, 136, 142, 148, 154, 160, 166, 171, 178, 184, 190, 196, 202, 208, 214, 220, 226, or 232. In some embodiments, the BIP primer comprises a nucleotide sequence at least 85%, at least 87%, at least 90%, at least 92%, at least 94%, at least 95%, at least 96%, at least 98%, or 100% identical to any one of SEQ ID NOS: 95, 96, 101, 107, 113, 119, 125, 131, 137, 143, 149, 155, 161, 167, 172, 179, 185, 191, 197, 203, 209, 215, 221, 227, or 233. In some embodiments, the B3 primer comprises a nucleotide sequence at least 85%, at least 87%, at least 90%, at least 92%, at least 94%, at least 95%, at least 96%, at least 98%, or 100% identical to any one of SEQ ID NOS: 93, 99, 105, 111, 117, 123, 129, 135, 141, 147, 153, 159, 165, 170, 177, 183, 189, 195, 201, 207, 213, 219, 225, or 231. In some embodiments, the F3 primer comprises a nucleotide sequence at least 85%, at least 87%, at least 90%, at least 92%, at least 94%, at least 95%, at least 96%, at least 98%, or 100% identical to any one of SEQ ID NOS: 92, 98, 104, 110, 116, 122, 128, 134, 140, 146, 152, 158, 164, 170, 176, 182, 188, 194, 200, 206, 212, 218, 224, or 230. In some embodiments, the LB primer comprises a nucleotide sequence at least 85%, at least 87%, at least 90%, at least 92%, at least 94%, at least 95%, at least 96%, at least 98%, or 100% identical to any one of SEQ ID NOS: 6, 12, 18, 24, 30, 36, 42, 48, 54, 60, 66, 72, 78, 84, 90, 96, 102, 108, 114, 120, 126, 132, 138, or 144. In some embodiments, the LF primer comprises a nucleotide sequence at least 85%, at least 87%, at least 90%, at least 92%, at least 94%, at least 95%,

at least 96%, at least 98%, or 100% identical to any one of SEQ ID NOS: 96, 102, 108, 114, 120, 126, 132, 138, 144, 150, 156, 162, 168, 173, 180, 186, 192, 198, 204, 210, 216, 222, 228, or 234.

[0019] In some embodiments, the at least six amplification primers comprise six nucleotide sequences at least 85%, at least 87%, at least 90%, at least 92%, at least 94%, at least 95%, at least 96%, at least 98%, or 100% identical to SEQ ID NOS: 92-97, SEQ ID NOS: 98-103, SEQ ID NOS: 104-109, SEQ ID NOS: 110<sup>-115</sup>, SEQ ID NOS: 116-121, SEQ ID NOS: 122-127, SEQ ID NOS: 128-133, SEQ ID NOS: 134-139, SEQ ID NOS: 140-145, SEQ ID NOS: 146-151, SEQ ID NOS: 152-157, SEQ ID NOS: 158-163, SEQ ID NOS: 164-169, SEQ ID NOS: 170-175, SEQ ID NOS: 176-181, SEQ ID NOS: 182-187, SEQ ID NOS: 188-193, SEQ ID NOS: 194-199, SEQ ID NOS: 200-205, SEQ ID NOS: 206-211, SEQ ID NOS: 212-217, SEQ ID NOS: 218-223, SEQ ID NOS: 224-229, or SEQ ID NOS: 230-235.

[0020] In some embodiments, the at least one amplification primer comprises at least three amplification primers. The at least three amplification primers may comprise a BIP primer, a B3 primer, and a LB primer. The at least three amplification primers comprise a FIP primer, a F3 primer, and a LF primer.

[0021] In some embodiments, the amplifying comprises isothermal amplification. The amplifying may comprise helicase dependent amplification (HDA), circular helicase dependent amplification (cHDA), strand displacement amplification (SDA), loop mediated amplification (LAMP), exponential amplification reaction (EXPAR), rolling circle amplification (RCA), ligase chain reaction (LCR), single primer isothermal amplification (SPIA), multiple displacement amplification (MDA), nucleic acid sequence based amplification (NASBA), hinge-initiated primer-dependent amplification of nucleic acids (HIP), nicking enzyme amplification reaction (NEAR), or improved multiple displacement amplification (IMDA). The amplifying may comprise a thermal cycling amplification such as polymerase chain reaction (PCR).

[0022] In some embodiments, the amplifying comprises contacting the sample to reagents for amplification. In some embodiments, the contacting the sample to reagents for amplification occurs concurrent to the contacting the sample to the reporter and the composition comprising the programmable nuclease and the guide nucleic acid. In some embodiments, the reagents for amplification comprise a polymerase and dNTPs.

[0023] In some embodiments, the guide nucleic acid comprises a nucleobase sequence at least 85%, at least 87%, at least 90%, at least 92%, at least 94%, at least 95%, at least 96%, at least 98%, or 100% identical to any one of SEQ ID NOS: 236-247.

[0024] In some embodiments, the programmable nuclease is a type V CRISPR/Cas effector protein. In some embodiments, the type V CRISPR/Cas effector protein is a Cas14 protein. In some embodiments, the Cas 14 protein comprises a Cas 14a polypeptide, a Cas 14b polypeptide, a Cas14c polypeptide, a Cas14d polypeptide, a Cas14e polypeptide, a Cas 14f polypeptide, a Cas14g polypeptide, a Cas14h polypeptide, a Cas14i polypeptide, a Cas14j polypeptide, or a Cas14k polypeptide. In some embodiments, the Cas14 protein comprises an amino acid sequence at least 80%, 85%,

at least 87%, at least 90%, at least 92%, at least 94%, at least 95%, at least 96%, at least 98%, or 100% identical to Cas14a.

[0025] In some embodiments, the method further comprising lysing the sample. In some embodiments, the lysing comprises contacting the sample to a lysis buffer.

[0026] In some embodiments, the method further comprising multiplexed detection of one or more additional target nucleic acid(s). In some embodiments, the one or more additional target nucleic acid(s) is: a) a different gene of the monkeypox virus or a segment thereof, b) a different segment of the same gene of the monkeypox virus, and/or c) a gene of another Orthopoxvirus or a segment thereof. In some embodiments, the other Orthopoxvirus is selected from the group consisting of abatino macacapox virus, akhmeta virus, alaskapox virus, camelpox virus, cowpox virus, ectromelia virus, raccoonpox virus, skunkpox virus, taterapox virus, vaccinia virus, variola virus, and volepox virus.

[0027] In some embodiments, sample lysis, amplification, detection, or any combination thereof is carried out in a single volume.

[0028] In some embodiments, sample lysis, amplification, detection, or any combination thereof is carried out in separate volumes.

[0029] In some embodiments, the sample is a saliva sample or a wound swab sample.

[0030] In some embodiments, the method further comprising repeating one or more method steps to assay for a negative control nucleic acid not comprising a gene of the monkeypox virus or a segment thereof.

[0031] Another aspect of the present disclosure provides a composition comprising a non-naturally occurring guide nucleic acid comprising a nucleotide sequence at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to any one of SEQ ID NOS: 236-247.

[0032] Yet another aspect of the present disclosure provides a composition comprising an amplification primer comprising a nucleotide sequence at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to any one of SEQ ID NOS: 92-235.

[0033] In another aspect, the present disclosure provides a method of assaying for a target nucleic acid in a sample, the method comprising a) contacting the sample to a reporter and a composition comprising a programmable nuclease and a guide nucleic acid that hybridizes to the target nucleic acid or an amplified product thereof, wherein the programmable nuclease cleaves the reporter upon hybridization of the guide nucleic acid to the target nucleic acid or the amplification product thereof; and b) assaying for a change in a signal, wherein the change in the signal is produced by cleavage of the reporter; wherein the target nucleic acid is a gene of a monkeypox virus or a segment thereof, and optionally wherein the guide nucleic acid comprises a nucleobase sequence at least 85%, at least 87%, at least 90%, at least 92%, at least 94%, at least 95%, at least 96%, at least 98%, or 100% identical to any one of SEQ ID NOS: 236-247.

[0034] These and other embodiments are addressed in more detail in the detailed description set forth below.

## BRIEF DESCRIPTION OF THE DRAWINGS

[0035] The novel features of the disclosure are set forth with particularity in the appended claims. A better understanding of the features and advantages of the present disclosure will be obtained by reference to the following detailed description that sets forth illustrative embodiments, in which the principles of the disclosure are utilized, and the accompanying drawings of which:

[0036] FIG. 1 shows images of sections of partitioning chambers loaded with the reaction mixtures described in Example 1 to compare the use of different RNA-targeting Cas proteins, comprising the amino acid sequence of SEQ ID NO: 21, 62 or 43. Varying amounts of the target nucleic acid (columns I through VIII) and a SARS-CoV-2 N-gene targeting guide RNA (columns I through VI and VIII) or an off-target guide (OTG, column VII) were used in each of the chambers, as depicted. Each black square is an image of a section of one partitioning chamber, while the mini dots within the chamber images are individual nanovolumes. The presence of the target nucleic acid in a particular nanovolume in the chamber is detected by the presence of a positive signal in that nanovolume.

[0037] FIG. 2 shows images of sections of partitioning chambers loaded with the reaction mixtures described in Example 2, having varying concentrations (0 fM to 10 nM) of the Cas protein comprising the amino acid sequence of SEQ ID NO: 21, and a fixed amount (600 copies/chamber) of the Twist Synthetic SARS-CoV-2 Synthetic RNA Control target nucleic acid or a no-target control (NTC, 0 copies/chamber). Each mixture also comprised a SARS-CoV-2 N-gene targeting guide RNA or a control using an off-target guide RNA (OTG).

[0038] FIG. 3 shows images of sections of partitioning chambers loaded with the reaction mixtures described in Example 2, having varying concentrations (0 fM to 10 nM) of the Cas protein comprising the amino acid sequence of SEQ ID NO: 21, and a fixed amount (600 copies/chamber) of the ATCC® synthetic SARS-CoV-2 RNA target nucleic acid, a no-target control (NTC, 0 copies/chamber), or an off-target control nucleic acid (OTC, 600 copies/chamber). Each mixture also comprised a SARS-CoV-2 N-gene targeting guide RNA or a control using an off-target guide RNA (OTG).

[0039] FIG. 4 shows images of sections of partitioning chambers loaded with the reaction mixtures described in Example 3 but having different proportional volumes of a low pH crude lysis buffer (0%, 5%, 10% or 20%), 5000 copies/chamber of Twist Synthetic SARS-CoV-2 Synthetic RNA Control target nucleic acid, and the Cas protein comprising the amino acid sequence of SEQ ID NO: 21. Each mixture also comprised a SARS-CoV-2 N-gene targeting guide RNA (R4684) or an off-target guide (OTG, R5882) as a negative control.

[0040] FIG. 5 shows images of sections of partitioning chambers loaded with the reaction mixtures described in Example 3, but in the presence of different carrier molecules (yeast RNA, glycogen, linear acrylamide or PVP at different concentrations), Twist Synthetic SARS-CoV-2 Synthetic RNA Control target nucleic acid, and the Cas protein comprising the amino acid sequence of SEQ ID NO: 21. A no-target control (NTC) was used as a negative control.

[0041] FIG. 6 shows images of sections of partitioning chambers loaded with the reaction mixtures described in Example 4 and Table A, comprising three types of RNAs in

a mixture (the target SARS-CoV-2 RNA and the off-target Influenza A and B RNAs). Varying amounts of the target nucleic acid and/or off-target nucleic acids were used in each of the chambers, as depicted (4000 copies/chamber, 400 copies/chamber, 40 copies/chamber and 0 copies/chamber). Either decreasing the concentration of all the three RNAs (top row) or decreasing the concentration of just the SARS CoV-2 target nucleic acid (middle and bottom rows) resulted in a similar decrease in the number of positive nanovolumes.

[0042] FIG. 7 shows images of partitioning chambers loaded with the reaction mixtures described in Example 5 to test the function of a DNA-targeting Cas protein comprising the amino acid sequence of SEQ ID NO: 34. Varying amounts (6,250 copies/reaction, 12,500 copies, reaction, 18,750 copies/reaction or 25,000 copies/reaction) of the synthetic dsDNA target nucleic acid, a no-target control (NTC, 0 copies/reaction) or an off-target control (OTC) nucleic acid were used in each of the chambers, as depicted.

[0043] FIG. 8 shows images of partitioning chambers loaded with the reaction mixtures described in Example 5 to test the function of a DNA-targeting Cas protein comprising the amino acid sequence of SEQ ID NO: 3. The synthetic SARS-CoV-2 dsDNA target nucleic acid or a no-target control (NTC) was used, as indicated.

[0044] FIG. 9A is a bar graph depicting the number of copies/reaction of SARS-CoV-2 target nucleic acid or PMMoV target nucleic acid (off-target control, OTC) detected using the quantitative assays described herein employing a SARS-CoV-2 crRNA for various concentrations of synthetic SARS-CoV-2 target as an input (from dilution D1 down to dilution D6). FIG. 9B is a bar graph depicting the number of copies/reaction of PMMoV target nucleic acid or SARS-CoV-2 target nucleic acid (off-target control, OTC) detected using the quantitative assays described herein employing a PMMoV crRNA for various concentrations of synthetic PMMoV target as an input (from dilution D1 down to dilution D6). See Example 6.

[0045] FIG. 10A is a standard curve obtained using the quantitative assays described herein for the SARS-CoV-2 target nucleic acid, showing the relatively linear relationship between the number of target copies per chamber (x-axis) and the number of positive nanovolumes observed (y-axis). FIG. 10B is a standard curve obtained using the quantitative assays described herein for the PMMOV target nucleic acid, showing the relatively linear relationship between the number of target copies per chamber (x-axis) and the number of positive nanovolumes observed (y-axis).

[0046] FIG. 11 shows images of sections of partitioning chambers loaded with the reaction mixtures described in Example 7 to compare the use of an RNA-targeting Cas protein, comprising the amino acid sequence of SEQ ID NO: 21, over a range of target nucleic acid concentrations (3,333 copies/μL to 0 copies/μL) or an off-target guide (OTG), as depicted. Each black square is an image of a section of one partitioning chamber, while the mini dots within the chamber images are individual nanovolumes. The presence of the target nucleic acid in a particular nanovolume in the chamber is detected by the presence of a positive signal in that nanovolume.

[0047] FIG. 12 shows images of sections of partitioning chambers loaded with the reaction mixtures described in Example 8 to compare the use of an RNA-targeting Cas protein, comprising the amino acid sequence of SEQ ID NO: 21 in a digital DETECTR reaction with the gold standard

ddPCR assay on viral nucleic acid samples extracted from SARS-CoV-2-positive clinical samples. Each black square is an image of a section of one partitioning chamber, while the mini dots within the chamber images are individual nanovolumes. The presence of the target nucleic acid in a particular nanovolume in the chamber is detected by the presence of a positive signal in that nanovolume.

[0048] FIG. 13 shows a head-to-head comparison of data generated from ddPCR and digital DETECTR run on SARS-CoV-2-positive clinical samples as described in Example 8. Data collected showed high linearity across the range of extracted sample dilutions tested.

[0049] FIG. 14 shows a comparison of the number of copies per microliter (cp/μL) determined for three wastewater samples using digital DETECTR and ddPCR. These data show that digital DETECTR has a roughly equivalent dynamic range to digital PCR for viral quantification under the conditions tested for both SARS-CoV-2 and PMMOV RNA in wastewater samples.

[0050] FIG. 15A shows a head-to-head comparison of digital DETECTR and dPCR signals obtained from wastewater samples for SARS-CoV-2 quantitation, graphed with the same scaling along the Y-axis. FIG. 15B shows the digital DETECTR and dPCR signals graphed with different Y-axis scaling, in order to highlight the stark differences in signals obtained by each method.

[0051] FIG. 16 shows bar graphs depicting the number of copies/reaction of SARS-CoV-2 target dsDNA nucleic acid detected using the quantitative assays described herein employing a different DNA-targeting Cas proteins (comprising the amino acid sequence of SEQ ID NO: 65, 67, 68, 34, 17, 89, 90, or 91) at 37C or 50C reaction temperatures.

[0052] FIG. 17 shows a schematic of an exemplary warm-start strategy which may be employed to delay activation of the CRISPR/Cas complex until the complex has reached a predetermined reaction temperature.

#### DETAILED DESCRIPTION

[0053] It is to be understood that both the foregoing general description and the following detailed description are exemplary, and explanatory only, and are not restrictive of the disclosure.

[0054] The section headings used herein are for organizational purposes only and are not to be construed as limiting the subject matter described.

[0055] All documents, or portions of documents, cited in this application, including, but not limited to, patents, patent applications, articles, books, and treatises, are hereby expressly incorporated by reference in their entirety for any purpose.

#### Definitions

[0056] Unless otherwise indicated, all technical terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Unless otherwise indicated or obvious from context, the following terms have the following meanings:

[0057] The terms, "a," "an," and "the," as used herein, include plural references unless the context clearly dictates otherwise.

[0058] The terms, "or" and "and/or," as used herein, include any and all combinations of one or more of the associated listed items.

[0059] The terms, "including," "includes," "included," and other forms, are not limiting.

[0060] The terms, "comprise" and its grammatical equivalents, as used herein, specify the presence of stated features, integers, steps, operations, elements, and/or components, but do not preclude the presence or addition of one or more other features, integers, steps, operations, elements, components, and/or groups thereof.

[0061] The term, "about," as used herein in reference to a number or range of numbers, is understood to mean the stated number and numbers +/-10% thereof, or 10% below the lower listed limit and 10% above the higher listed limit for the values listed for a range.

[0062] The terms, "% identical," "% identity," "percent identity," and grammatical equivalents thereof, as used herein, in the context of an amino acid sequence or nucleotide sequence, refer to the percent of residues that are identical between respective positions of two sequences when the two sequences are aligned for maximum sequence identity. The % identity is calculated by dividing the total number of the aligned residues by the number of the residues that are identical between the respective positions of the at least two sequences and multiplying by 100. Generally, computer programs can be employed for such calculations. Illustrative programs that compare and align pairs of sequences, include ALIGN (Myers and Miller, Comput Appl Biosci. 1988 March; 4 (1): 11-7), FASTA (Pearson and Lipman, Proc Natl Acad Sci USA. 1988 Apr; 85 (8): 2444-8; Pearson, Methods Enzymol. 1990; 183: 63-98) and gapped BLAST (Altschul et al., Nucleic Acids Res. 1997 Sep. 1; 25 (17): 3389-40), BLASTP, BLASTN, or GCG (Devereux et al., Nucleic Acids Res. 1984 Jan. 11; 12 (1 Pt 1): 387-95).

[0063] The terms, "% complementary", "% complementarity", "percent complementary", "percent complementarity" and grammatical equivalents thereof, as used interchangeably herein, in the context of two or more nucleic acid molecules, refer to the percent of nucleotides in two nucleotide sequences in said nucleic acid molecules of equal length that can undergo cumulative base pairing at two or more individual corresponding positions in an antiparallel orientation. Accordingly, the terms include nucleic acid sequences that are not completely complementary over their entire length, which indicates that the two or more nucleic acid molecules include one or more mismatches. A "mismatch" is present at any position in the two opposed nucleotides that are not complementary. The % complementary is calculated by dividing the total number of the complementary residues by the total number of the nucleotides in one of the equal length sequences, and multiplying by 100. Complete or total complementarity describes nucleotide sequences in 100% of the residues of a nucleotide sequence are complementary to residues in a reference nucleotide sequence. "Partially complementarity" describes nucleotide sequences in which at least 20%, but less than 100%, of the residues of a nucleotide sequence are complementary to residues in a reference nucleotide sequence. In some instances, at least 50%, but less than 100%, of the residues of a nucleotide sequence are complementary to residues in a reference nucleotide sequence. In some instances, at least 70%, 80%, 90% or 95%, but less than 100%, of the residues of a nucleotide sequence are comple-

mentary to residues in a reference nucleotide sequence. "Noncomplementary" describes nucleotide sequences in which less than 20% of the residues of a nucleotide sequence are complementary to residues in a reference nucleotide sequence.

[0064] The terms, "amplification," "amplifying," and grammatical equivalents thereof, as used herein, refer to a process by which a nucleic acid molecule is enzymatically copied to generate a plurality of nucleic acid molecules containing the same sequence as the original nucleic acid molecule or a distinguishable portion thereof.

[0065] The terms, "bind," "binding," "interact" and "interacting," as used herein, refer to a non-covalent interaction between macromolecules (e.g., between two polypeptides, between a polypeptide and a nucleic acid; between a polypeptide/guide nucleic acid complex and a target nucleic acid; and the like). While in a state of noncovalent interaction, the macromolecules are said to be "associated" or "interacting" or "binding" (e.g., when a molecule X is said to interact with a molecule Y, it is meant the molecule X binds to molecule Y in a non-covalent manner). Non-limiting examples of non-covalent interactions are ionic bonds, hydrogen bonds, van der Waals and hydrophobic interactions. Not all components of a binding interaction need be sequence-specific (e.g., contacts with phosphate residues in a DNA backbone), but some portions of a binding interaction may be sequence-specific.

[0066] The term, "cis cleavage," as used herein refers to cleavage (hydrolysis of a phosphodiester bond) of a target nucleic acid by an effector protein complexed with a guide nucleic acid refers to cleavage of a target nucleic acid that is hybridized to a guide nucleic acid, wherein cleavage occurs within or directly adjacent to the region of the target nucleic acid that is hybridized to the guide nucleic acid.

[0067] The term, "codon optimized," as used herein, refers to a mutation of a nucleotide sequence encoding a polypeptide, such as a nucleotide sequence encoding an effector protein, to mimic the codon preferences of the intended host organism or cell while encoding the same polypeptide. Thus, the codons can be changed, but the encoded polypeptide remains unchanged. For example, if the intended target cell was a human cell, a human codon-optimized nucleotide sequence encoding an effector protein could be used. As another non-limiting example, if the intended host cell were a mouse cell, then a mouse codon-optimized nucleotide sequence encoding an effector protein could be generated. As another non-limiting example, if the intended host cell were a eukaryotic cell, then a eukaryote codon-optimized nucleotide sequence encoding an effector protein could be generated. As another non-limiting example, if the intended host cell were a prokaryotic cell, then a prokaryote codon-optimized nucleotide sequence encoding an effector protein could be generated. Codon usage tables are readily available, for example, at the "Codon Usage Database" available at [www.kazusa.or.jp/codon](http://www.kazusa.or.jp/codon).

[0068] The terms, "complementary" and "complementarity," as used herein, in the context of a nucleic acid molecule or nucleotide sequence, refer to the characteristic of a polynucleotide having nucleotides that can undergo cumulative base pairing with their Watson-Crick counterparts (C with G; or A with T) in a reference nucleic acid in antiparallel orientation. For example, when every nucleotide in a polynucleotide or a specified portion thereof forms a base pair with every nucleotide in an equal length sequence of a

reference nucleic acid, that polynucleotide is said to be 100% complementary to the sequence of the reference nucleic acid. In a double stranded DNA or RNA sequence, the upper (sense) strand sequence is, in general, understood as going in the direction from its 5'- to 3'-end, and the complementary sequence is thus understood as the sequence of the lower (antisense) strand in the same direction as the upper strand. Following the same logic, the reverse sequence is understood as the sequence of the upper strand in the direction from its 3'-to its 5'-end, while the "reverse complement" sequence or the "reverse complementary" sequence is understood as the sequence of the lower strand in the direction of its 5'-to its 3'-end. Each nucleotide in a double stranded DNA or RNA molecule that is paired with its Watson-Crick counterpart can be referred to as its complementary nucleotide. The complementarity of modified or artificial base pairs can be based on other types of hydrogen bonding and/or hydrophobicity of bases and/or shape complementarity between bases.

[0069] The term, "cleavage assay," as used herein, refers to an assay designed to visualize, quantitate or identify cleavage of a nucleic acid. In some instances, the cleavage activity may be cis-cleavage activity. In some instances, the cleavage activity may be trans-cleavage activity.

[0070] The terms, "cleave," "cleaving" and "cleavage," as used herein, in the context of a nucleic acid molecule or nuclease activity of an effector protein, refer to the hydrolysis of a phosphodiester bond of a nucleic acid molecule that results in breakage of that bond. The result of this breakage can be a nick (hydrolysis of a single phosphodiester bond on one side of a double-stranded molecule), single strand break (hydrolysis of a single phosphodiester bond on a single-stranded molecule) or double strand break (hydrolysis of two phosphodiester bonds on both sides of a double-stranded molecule) depending upon whether the nucleic acid molecule is single-stranded (e.g., ssDNA or ssRNA) or double-stranded (e.g., dsDNA) and the type of nuclease activity being catalyzed by the effector protein.

[0071] The term, "clustered regularly interspaced short palindromic repeats (CRISPR)," as used herein, refers to a segment of DNA found in the genomes of certain prokaryotic organisms, including some bacteria and archaea, that includes repeated short sequences of nucleotides interspersed at regular intervals between unique sequences of nucleotides derived from another organism.

[0072] The term, "conservative substitution," as used herein, refers to the replacement of one amino acid for another such that the replacement takes place within a family of amino acids that are related in their side chains. Conversely, the term "non-conservative substitution" as used herein refers to the replacement of one amino acid residue for another that does not have a related side chain. Genetically encoded amino acids can be divided into four families having related side chains: (1) acidic (negatively charged): Asp (D), Glu (E); (2) basic (positively charged): Lys (K), Arg (R), His (H); (3) non-polar (hydrophobic): Cys (C), Ala (A), Val (V), Leu (L), Ile (I), Pro (P), Phe (F), Met (M), Trp (W), Gly (G), Tyr (Y), with non-polar also being subdivided into: (i) strongly hydrophobic: Ala (A), Val (V), Leu (L), Ile (I), Met (M), Phe (F); and (ii) moderately hydrophobic: Gly (G), Pro (P), Cys (C), Tyr (Y), Trp (W); and (4) uncharged polar: Asn (N), Gln (Q), Ser(S), Thr (T). Amino acids may be related by aliphatic side chains: Gly (G), Ala (A), Val (V), Leu (L), Ile (I), Ser(S), Thr (T), with

Ser(S) and Thr (T) optionally being grouped separately as aliphatic-hydroxyl; Amino acids may be related by aromatic side chains: Phe (F), Tyr (Y), Trp (W). Amino acids may be related by amide side chains: Asn (N), Gln (Q). Amino acids may be related by sulfur-containing side chains: Cys (C) and Met (M).

[0073] The terms, "CRISPR RNA" and "crRNA," as used herein, refer to a type of guide nucleic acid that is RNA comprising a first sequence that is capable of hybridizing to a target sequence of a target nucleic acid and a second sequence that is capable of interacting with an effector protein either directly (by being bound by an effector protein) or indirectly (e.g., by hybridization with a second nucleic acid molecule that can be bound by an effector). The first sequence and the second sequence are directly connected to each other or by a linker.

[0074] The term, "detectable signal," as used herein, refers to a signal that can be detected using optical, fluorescent, chemiluminescent, electrochemical or other detection methods known in the art.

[0075] The term, "detecting a nucleic acid" and its grammatical equivalents, as used herein refers to detecting the presence or absence of the target nucleic acid in a sample that potentially contains the nucleic acid being detected.

[0076] The term, "detection moiety," as used herein refers to a molecule that can release a signal that can be detected using optical, fluorescent, chemiluminescent, electrochemical, calorimetric and other detection methods known in the art.

[0077] The term, "effector protein," as used herein, refers to a protein, polypeptide, or peptide that is capable of interacting with a nucleic acid, such as a guide nucleic acid, to form a complex (e.g., a RNP complex), wherein the complex interacts with a target nucleic acid.

[0078] The terms, "effector partner" and "partner polypeptide" as used herein, refer to a polypeptide that does not have 100% sequence identity with an effector protein described herein. In some instances, an effector partner described herein may be found in a homologous genome as an effector protein described herein.

[0079] The term, "engineered modification," as used herein, refers to a structural change of one or more nucleic acid residues of a nucleotide sequence or one or more amino acid residue of an amino acid sequence, such as chemical modification of one or more nucleobases; or a chemical change to the phosphate backbone, a nucleotide, a nucleobase, or a nucleoside. Such modifications can be made to an effector protein amino acid sequence or guide nucleic acid nucleotide sequence, or any sequence disclosed herein (e.g., a nucleic acid encoding an effector protein or a nucleic acid that encodes a guide nucleic acid). Methods of modifying a nucleic acid or amino acid sequence are known. One of ordinary skill in the art will appreciate that the engineered modification(s) may be located at any position(s) of a nucleic acid such that the function of the nucleic acid, protein, composition or system is not substantially decreased. Nucleic acids provided herein can be prepared according to any available technique including, but not limited to chemical synthesis, enzymatic synthesis, which is generally termed in vitro-transcription, cloning, enzymatic, or chemical cleavage, etc. In some instances, the nucleic acids provided herein are not uniformly modified along the entire length of the molecule. Different nucleotide modifi-

cations and/or backbone structures can exist at various positions within the nucleic acid.

[0080] The term, "guide nucleic acid," as used herein, refers to a nucleic acid that, when in a complex with one or more polypeptides described herein (e.g., an RNP complex) can impart sequence selectivity to the complex when the complex interacts with a target nucleic acid. A guide nucleic acid may be referred to interchangeably as a guide RNA, however it is understood that guide nucleic acids may comprise deoxyribonucleotides (DNA), ribonucleotides (RNA), a combination thereof (e.g., RNA with a thymine base), biochemically or chemically modified nucleobases (e.g., one or more engineered modifications described herein), or combinations thereof. In some embodiments, the guide nucleic acid comprises: a first nucleotide sequence that hybridizes to a target nucleic acid; and a second nucleotide sequence that is capable of being non-covalently bound by an effector protein, such as, Cas. The first sequence may be referred to herein as a spacer sequence. The second sequence may be referred to herein as a repeat sequence. In some embodiments, the first sequence is located 5' of the second nucleotide sequence. In some embodiments, the first sequence is located 3' of the second nucleotide sequence.

[0081] The term, "heterologous," as used herein, refers to at least two different polypeptide sequences that are not found similarly connected to one another in a native nucleic acid or protein. A protein that is heterologous to the effector protein is a protein that is not covalently linked by an amide bond to the effector protein in nature. In some instances, a heterologous protein is not encoded by a species that encodes the effector protein. A guide nucleic acid may comprise "heterologous" sequences, which means that it includes a first sequence and a second sequence, wherein the first sequence and the second sequence are not found covalently linked by a phosphodiester bond in nature. Thus, the first sequence is considered to be heterologous with the second sequence, and the guide nucleic acid may be referred to as a heterologous guide nucleic acid.

[0082] The terms, "hybridize," "hybridizable" and grammatical equivalents thereof, refer to a nucleotide sequence that is able to noncovalently interact, i.e., form Watson-Crick base pairs and/or G/U base pairs, or anneal, to another nucleotide sequence in a sequence-specific, antiparallel, manner (i.e., a nucleotide sequence specifically interacts to a complementary nucleotide sequence) under the appropriate in vitro and/or in vivo conditions of temperature and solution ionic strength. Standard Watson-Crick base-pairing includes: adenine (A) pairing with thymidine (T), adenine (A) pairing with uracil (U), and guanine (G) pairing with cytosine (C) for both DNA and RNA. In addition, for hybridization between two RNA molecules (e.g., dsRNA), and for hybridization of a DNA molecule with an RNA molecule (e.g., when a DNA target nucleic acid base pairs with a guide RNA, etc.); guanine (G) can also base pair with uracil (U). For example, G/U base-pairing is at least partially responsible for the degeneracy (i.e., redundancy) of the genetic code in the context of tRNA anti-codon base-pairing with codons in mRNA. Thus, a guanine (G) can be considered complementary to both an uracil (U) and to an adenine (A). Accordingly, when a G/U base-pair can be made at a given nucleotide position, the position is not considered to be non-complementary, but is instead considered to be complementary. While hybridization typically occurs between two nucleotide sequences that are complementary,

mismatches between bases are possible. It is understood that two nucleotide sequences need not be 100% complementary to be specifically hybridizable, hybridizable, partially hybridizable, or for hybridization to occur. Moreover, a nucleotide sequence may hybridize over one or more segments such that intervening or adjacent segments are not involved in the hybridization event (e.g., a bulge, a loop structure or hairpin structure, etc.). The conditions appropriate for hybridization between two nucleotide sequences depend on the length of the sequence and the degree of complementarity, variables which are well known in the art. For hybridizations between nucleic acids with short stretches of complementarity (e.g., complementarity over 35 or less, 30 or less, 25 or less, 22 or less, 20 or less, or 18 or less nucleotides) the position of mismatches may become important (see Sambrook et al., *supra*, 11.7-11.8). Typically, the length for a hybridizable nucleic acid is 8 nucleotides or more (e.g., 10 nucleotides or more, 12 nucleotides or more, 15 nucleotides or more, 20 nucleotides or more, 22 nucleotides or more, 25 nucleotides or more, or 30 nucleotides or more). Any suitable *in vitro* assay may be utilized to assess whether two sequences "hybridize". One such assay is a melting point analysis where the greater the degree of complementarity between two nucleotide sequences, the greater the value of the melting temperature (*T<sub>m</sub>*) for hybrids of nucleic acids having those sequences. The conditions of temperature and ionic strength determine the "stringency" of the hybridization. Temperature, wash solution salt concentration, and other conditions may be adjusted as necessary according to factors such as length of the region of complementation and the degree of complementation. Hybridization and washing conditions are well known and exemplified in Sambrook, J., Fritsch, E. F. and Maniatis, T. *Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor (1989), particularly Chapter 11 and Table 11.1 therein; and Sambrook, J. and Russell, W., *Molecular Cloning: A Laboratory Manual*, Third Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor (2001).

[0083] The term, "handle sequence," as used herein, refers to a sequence of nucleotides in a single guide RNA (sgRNA), that is: 1) capable of being non-covalently bound by an effector protein and 2) connects the portion of the sgRNA capable of being non-covalently bound by an effector protein to a nucleotide sequence that is hybridizable to a target nucleic acid. In general, the handle sequence comprises an intermediary RNA sequence, that is capable of being non-covalently bound by an effector protein. In some instances, the handle sequence further comprises a repeat sequence. In such instances, the intermediary RNA sequence or a combination of the intermediary RNA and the repeat sequence is capable of being non-covalently bound by an effector protein.

[0084] The terms, "intermediary RNA" and "intermediary RNA sequence," as used herein, in a context of a single nucleic acid system, refers to a nucleotide sequence in a handle sequence, wherein the intermediary RNA sequence is capable of, at least partially, being non-covalently bound to an effector protein to form a complex (e.g., an RNP complex). An intermediary RNA sequence is not a transactivating nucleic acid in systems, methods, and compositions described herein.

[0085] The term, "in vitro," as used herein, refers to describing something outside an organism. An *in vitro*

system, composition or method may take place in a container for holding laboratory reagents such that it is separated from the biological source from which a material in the container is obtained. *In vitro* assays can encompass cell-based assays in which living or dead cells are employed. *In vitro* assays can also encompass a cell-free assay in which no intact cells are employed. The term "in vivo" is used to describe an event that takes place within an organism. The term "ex vivo" is used to describe an event that takes place in a cell that has been obtained from an organism. An *ex vivo* assay is not performed on a subject. Rather, it is performed upon a sample separate from a subject.

[0086] The terms, "length" and "linked nucleosides," as used herein, refer to a nucleic acid (polynucleotide) or polypeptide, may be expressed as "kilobases" (kb) or "base pairs (bp)." Thus, a length of 1 kb refers to a length of 1000 linked nucleosides, and a length of 500 bp refers to a length of 500 linked nucleosides. Similarly, a protein having a length of 500 linked amino acids may also be simply described as having a length of 500 amino acids.

[0087] The term, "linker," as used herein, refers to a covalent bond or molecule that links a first polypeptide to a second polypeptide (e.g., by an amide bond) or a first nucleic acid to a second nucleic acid (e.g., by a phosphodiester bond).

[0088] The term, "mutation," as used herein, refers to an alteration that changes an amino acid residue or a nucleotide as described herein. Such an alteration can include, for example, deletions, insertions, and/or substitutions. The mutation can refer to a change in structure of an amino acid residue or nucleotide relative to the starting or reference residue or nucleotide. A mutation of an amino acid residue includes, for example, deletions, insertions and substituting one amino acid residue for a structurally different amino acid residue. Such substitutions can be a conservative substitution, a non-conservative substitution, a substitution to a specific sub-class of amino acids, or a combination thereof as described herein. A mutation of a nucleotide includes, for example, changing one naturally occurring base for a different naturally occurring base, such as changing an adenine to a thymine or a guanine to a cytosine or an adenine to a cytosine or a guanine to a thymine. A mutation of a nucleotide base may result in a structural and/or functional alteration of the encoding peptide, polypeptide or protein by changing the encoded amino acid residue of the peptide, polypeptide or protein. A mutation of a nucleotide base may not result in an alteration of the amino acid sequence or function of encoded peptide, polypeptide or protein, also known as a silent mutation. Methods of mutating an amino acid residue or a nucleotide are well known.

[0089] The terms, "non-naturally occurring" and "engineered," as used herein, refer to indicate involvement of the hand of man. The terms, when referring to a nucleic acid, nucleotide, protein, polypeptide, peptide or amino acid, refer to a molecule, such as but not limited to, a nucleic acid, nucleotide, protein, polypeptide, peptide or amino acid refers to a modification of that molecule (e.g., chemical modification, nucleotide sequence, or amino acid sequence) that is not present in the naturally molecule. The terms, when referring to a composition or system described herein, refer to a composition or system having at least one component that is not naturally associated with the other components of the composition or system. By way of a non-limiting example, a composition may include an effector protein and

a guide nucleic acid that do not naturally occur together. Conversely, and as a non-limiting further clarifying example, an effector protein or guide nucleic acid that is “natural,” “naturally-occurring,” or “found in nature” includes an effector protein and a guide nucleic acid from a cell or organism that have not been genetically modified by the hand of man.

[0090] The terms, “nuclease” and “endonuclease” as used herein, refer to an enzyme which possesses catalytic activity for nucleic acid cleavage.

[0091] The term, “nuclease activity,” as used herein, refers to catalytic activity that results in nucleic acid cleavage (e.g., ribonuclease activity (ribonucleic acid cleavage), or deoxyribonuclease activity (deoxyribonucleic acid cleavage), etc.).

[0092] The term, “nucleic acid,” as used herein, refers to a polymer of nucleotides. A nucleic acid may comprise ribonucleotides, deoxyribonucleotides, combinations thereof, and modified versions of the same. A nucleic acid may be single-stranded or double-stranded, unless specified. Non-limiting examples of nucleic acids are double stranded DNA (dsDNA), single stranded (ssDNA), messenger RNA, genomic DNA, cDNA, DNA-RNA hybrids, and a polymer comprising purine and pyrimidine bases or other natural, chemically or biochemically modified, non-natural, or derivatized nucleotide bases. Accordingly, nucleic acids as described herein may comprise one or more mutations, one or more engineered modifications, or both.

[0093] The term, “nucleic acid expression vector,” as used herein, refers to a plasmid that can be used to express a nucleic acid of interest.

[0094] The terms, “nucleotide(s)” and “nucleoside(s),” as used herein, in the context of a nucleic acid molecule having multiple residues, refer to describing the sugar and base of the residue contained in the nucleic acid molecule. Similarly, a skilled artisan could understand that linked nucleotides and/or linked nucleosides, as used in the context of a nucleic acid having multiple linked residues, are interchangeable and describe linked sugars and bases of residues contained in a nucleic acid molecule. When referring to a “nucleobase (s),” or linked nucleobase, as used in the context of a nucleic acid molecule, it can be understood as describing the base of the residue contained in the nucleic acid molecule, for example, the base of a nucleotide, nucleosides, or linked nucleotides or linked nucleosides. A person of ordinary skill in the art when referring to nucleotides, nucleosides, and/or nucleobases would also understand the differences between RNA and DNA (generally the exchange of uridine for thymidine or vice versa) and the presence of nucleoside analogs, such as modified uridines, do not contribute to differences in identity or complementarity among polynucleotides as long as the relevant nucleotides (such as thymidine, uridine, or modified uridine) have the same complement (e.g., adenosine for all of thymidine, uridine, or modified uridine; another example is cytosine and 5-methylcytosine, both of which have guanosine or modified guanosine as a complement). Thus, for example, the sequence 5'-AXG where X is any modified uridine, such as pseudouridine, N1-methyl pseudouridine, or 5-methoxyuridine, is considered 100% identical to AUG in that both are perfectly complementary to the same sequence (5'-CAU).

[0095] The term, “pharmaceutically acceptable excipient, carrier or diluent,” as used herein, refers to any substance formulated alongside the active ingredient of a pharmaceu-

tical composition that allows the active ingredient to retain biological activity and is non-reactive with the subject's immune system. Such a substance can be included for the purpose of long-term stabilization, bulking up solid formulations that contain potent active ingredients in small amounts, or to confer a therapeutic enhancement on the active ingredient in the final dosage form, such as facilitating absorption, reducing viscosity, or enhancing solubility. The selection of appropriate substance can depend upon the route of administration and the dosage form, as well as the active ingredient and other factors. Compositions having such substances can be formulated by suitable methods (see, e.g., Remington's Pharmaceutical Sciences, 18th edition, A. Gennaro, ed., Mack Publishing Co., Easton, Pa., 1990; and Remington, The Science and Practice of Pharmacy 21st Ed. Mack Publishing, 2005).

[0096] The terms, “polypeptide” and “protein,” as used herein, refer to a polymeric form of amino acids. A polypeptide may include coded and non-coded amino acids, chemically or biochemically modified or derivatized amino acids, and polypeptides having modified peptide backbones. Accordingly, polypeptides as described herein may comprise one or more mutations, one or more engineered modifications, or both. It is understood that when describing coding sequences of polypeptides described herein, said coding sequences do not necessarily require a codon encoding an N-terminal Methionine (M) or a Valine (V) as described for the effector proteins described herein. One skilled in the art would understand that a start codon could be replaced or substituted with a start codon that encodes for an amino acid residue sufficient for initiating translation in a host cell. In some embodiments, when a heterologous peptide, such as a fusion partner protein, protein tag or NLS, is located at the N terminus of the effector protein, a start codon for the heterologous peptide serves as a start codon for the effector protein as well. Thus, the natural start codon encoding an amino acid residue sufficient for initiating translation (e.g., Methionine (M) or a Valine (V)) of the effector protein may be removed or absent.

[0097] The terms, “promoter” and “promoter sequence,” as used herein, refer to a DNA regulatory region capable of binding RNA polymerase and initiating transcription of a downstream (3' direction) coding or non-coding sequence. A transcription initiation site, as well as protein binding domains responsible for the binding of RNA polymerase, can also be found in a promoter region. Eukaryotic promoters will often, but not always, contain “TATA” boxes and “CAT” boxes. Various promoters, including inducible promoters, may be used to drive expression by the various vectors of the present disclosure.

[0098] The terms, “protospacer adjacent motif” and “PAM,” as used herein, refer to a nucleotide sequence found in a target nucleic acid that directs an effector protein to edit the target nucleic acid at a specific location. In some instances, a PAM is required for a complex of an effector protein and a guide nucleic acid (e.g., an RNP complex) to hybridize to and edit the target nucleic acid. In some instances, the complex does not require a PAM to edit the target nucleic acid.

[0099] The term, “recombinant,” as used herein, in the context of proteins, polypeptides, peptides and nucleic acids, refers to proteins, polypeptides, peptides and nucleic acids that are products of various combinations of cloning, restriction, and/or ligation steps resulting in a construct having a

structural coding or non-coding sequence distinguishable from endogenous nucleic acids found in natural systems.

[0100] The term, “regulatory element,” used herein, refers to transcriptional and translational control sequences, such as promoters, enhancers, polyadenylation signals, terminators, protein degradation signals, and the like, that provide for and/or regulate transcription of a non-coding sequence (e.g., a guide nucleic acid) or a coding sequence (e.g., effector proteins, fusion proteins, and the like) and/or regulate translation of an encoded polypeptide.

[0101] The terms, “reporter” and “reporter nucleic acid,” as used herein, refer to a non-target nucleic acid molecule that can provide a detectable signal upon cleavage by an effector protein. Examples of detectable signals and detectable moieties that generate detectable signals are provided herein. In some embodiments, the reporter conveys the presence of a target nucleic acid.

[0102] The terms, “ribonucleotide protein complex” and “RNP” as used herein, refer to a complex of one or more nucleic acids and one or more polypeptides described herein. While the term utilizes “ribonucleotides” it is understood that the one or more nucleic acid may comprise deoxyribonucleotides (DNA), ribonucleotides (RNA), a combination thereof (e.g., RNA with a thymine base), biochemically or chemically modified nucleobases (e.g., one or more engineered modifications described herein), or combinations thereof.

[0103] The term, “sample,” as used herein generally refers to something comprising a target nucleic acid. In some embodiments, the sample is a biological sample, such as a saliva sample or a wound swab sample.

[0104] The terms, “single guide nucleic acid”, “single guide RNA” and “sgRNA,” as used herein, in the context of a single nucleic acid system, refers to a guide nucleic acid, wherein the guide nucleic acid is a single polynucleotide chain having all the required sequence for a functional complex with an effector protein (e.g., being bound by an effector protein, including in some instances activating the effector protein, and hybridizing to a target nucleic acid, without the need for a second nucleic acid molecule). For example, an sgRNA can have two or more linked guide nucleic acid components (e.g., an intermediary RNA sequence, a repeat sequence, a spacer sequence and optionally a linker, or a handle sequence and a spacer sequence). A handle sequence may comprise at least a portion of a tracrRNA sequence, at least a portion of a repeat sequence, or a combination thereof. A sgRNA does not comprise a tracrRNA.

[0105] The term, “spacer sequence,” as used herein, refers to a nucleotide sequence in a guide nucleic acid that is capable of, at least partially, hybridizing to an equal length portion of a sequence (e.g., a target sequence) of a target nucleic acid.

[0106] The term, “subject,” as used herein, refers to an animal. The subject may be a mammal. The subject may be a human. The subject may be diagnosed or at risk for a disease.

[0107] The term, “sufficiently complementary,” as used herein, refers to a first nucleotide sequence that is partially complementary to a second nucleotide sequence while still allowing the first nucleotide sequence to hybridize to the second nucleotide sequence with enough affinity to permit a biological activity to occur. Depending on the context, a biological activity may be the formation of a complex

between two or more components described herein, such as an effector protein and a guide nucleic acid. A biological activity may also be bringing one or more components described herein into proximity of another component, such as bringing an effector protein-guide nucleic acid complex into proximity of a target nucleic acid. A biological activity may additionally be permitting a component described herein to act on another component described herein, such as permitting an effector protein to cleave a target nucleic acid. In some instances, sequences are said to be sufficiently complementary when at least 85% of the residues of a nucleotide sequence are complementary to residues in a reference nucleotide sequence.

[0108] The term, “target nucleic acid,” as used herein, refers to a nucleic acid that is selected as the nucleic acid for editing, binding, hybridization or any other activity of or interaction with a nucleic acid, protein, polypeptide, or peptide described herein. A target nucleic acid may comprise RNA, DNA, or a combination thereof. A target nucleic acid may be single-stranded (e.g., single-stranded RNA or single-stranded DNA) or double-stranded (e.g., double-stranded DNA).

[0109] The term, “target sequence,” as used herein, in the context of a target nucleic acid, refers to a nucleotide sequence found within a target nucleic acid. Such a nucleotide sequence can, for example, hybridize to a respective length portion of a guide nucleic acid.

[0110] The terms, “trans-activating RNA”, “transactivating RNA” and “tracrRNA,” refer to a transactivating or transactivated nucleic acid in a dual nucleic acid system that is capable of hybridizing, at least partially, to a crRNA to form a tracrRNA-crRNA duplex, and of interacting with an effector protein to form a complex (e.g., an RNP complex).

[0111] The terms, “transactivating”, “trans-activating”, “trans-activated”, “transactivated” and grammatical equivalents thereof, as used herein, in the context of a dual nucleic acid system refers to an outcome of the system, wherein a polypeptide is enabled to have a binding and/or nuclease activity on a target nucleic acid, by a tracrRNA or a tracrRNA-crRNA duplex.

[0112] The term, “trans cleavage,” as used herein, in the context of cleavage (e.g., hydrolysis of a phosphodiester bond) of one or more target nucleic acids or non-target nucleic acids, or both, by an effector protein that is complexed with a guide nucleic acid and the target nucleic acid. Trans cleavage activity may be triggered by the hybridization of a guide nucleic acid to a target nucleic acid. The effector may cleave a target strand as well as non-target strand, wherein the target nucleic acid is a double stranded nucleic acid. Trans cleavage of the target nucleic acid may occur away from (e.g., not within or directly adjacent to) the portion of the target nucleic acid that is hybridized to the portion of the guide nucleic acid.

[0113] The terms, “treatment” and “treating,” as used herein, refer to a pharmaceutical or other intervention regimen for obtaining beneficial or desired results in the recipient. Beneficial or desired results include but are not limited to a therapeutic benefit and/or a prophylactic benefit. A therapeutic benefit may refer to eradication or amelioration of symptoms or of an underlying disorder being treated. Also, a therapeutic benefit can be achieved with the eradication or amelioration of one or more of the physiological symptoms associated with the underlying disorder such that an improvement is observed in the subject, notwithstanding

that the subject may still be afflicted with the underlying disorder. A prophylactic effect includes delaying, preventing, or eliminating the appearance of a disease or condition, delaying, or eliminating the onset of symptoms of a disease or condition, slowing, halting, or reversing the progression of a disease or condition, or any combination thereof. For prophylactic benefit, a subject at risk of developing a particular disease, or to a subject reporting one or more of the physiological symptoms of a disease may undergo treatment, even though a diagnosis of this disease may not have been made.

[0114] The term, "variant," as used herein, refers to a form or version of a protein that differs from the wild-type protein. A variant may have a different function or activity relative to the wild-type protein.

[0115] As used herein, a "reaction mixture" refers to a composition comprising an effector protein (e.g., a CRISPR/Cas effector protein), a guide nucleic acid (e.g., a guide RNA) and a reporter nucleic acid. In some embodiments, the reaction mixture comprises a target nucleic acid capable of binding to the guide nucleic acid.

[0116] As used herein, a "no target control" or "NTC" refers to a reaction mixture that may be used as a negative control in the methods described herein, which does not contain a target nucleic acid that is capable of binding to the guide nucleic acid.

[0117] As used herein, an "off-target control" or "OTC" refers to a nucleic acid that may be used as a negative control in the methods described herein having a sequence that is not capable of binding to the guide nucleic acid, or to a reaction mixture that may be used as a negative control in the methods described herein containing a negative control nucleic acid having a sequence that is not capable of binding to the guide nucleic acid.

[0118] As used herein, an "off-target guide" or "OTG" refers to a guide nucleic acid used as a negative control in the methods described herein, which is not capable of binding to the target nucleic acid molecule, or to a reaction mixture used as a negative control in the methods described herein which contains a guide nucleic acid that is not capable of binding to the target nucleic acid molecule.

[0119] As used herein, a "nanovolume" refers to an individual volume of less than or equal of 100 nL, preferably less than or equal to 1 nL. The form factor that can partition the nanovolumes from a larger input volume is not limited, and may be, for instance, a droplet, a well, a microwell, a nanowell, and the like.

[0120] As used herein, a "test sample" (or "sample") refers to a composition, comprising a target nucleic acid. In some instances, the test sample is a biological sample, such as a biological fluid or tissue sample. In some instances, the test sample is an environmental sample. The test sample may be a biological sample or environmental sample that is modified or manipulated. For instance, in some embodiments, the test samples may be modified or manipulated with purification techniques, heat, salts, buffers, or any combination thereof.

[0121] As used herein, a "positive control sample" refers to a sample in which the amount or concentration of a nucleic acid (e.g., target nucleic acid, or positive control target nucleic acid) is known prior to performing the methods for quantitation disclosed herein.

[0122] As used herein, a "positive control target nucleic acid" is a nucleic acid that is used as a control for the

reaction conditions of the methods disclosed herein, such as, the performance of the effector protein, the functionality of the reporter nucleic acid and/or the detection of the detectable signals.

#### Methods For Quantitation of Nucleic Acid Targets

[0123] Provided herein are compositions, systems, and methods of quantitating a target nucleic acid.

[0124] For instance, the disclosure provides methods for quantitating a target nucleic acid in a test sample, comprising (a) contacting the test sample with the following components (i) through (iii) resulting in at least two nanovolumes of a reaction mixture: (i) any one or more of the effector proteins described herein, (ii) a guide nucleic acid, comprising a region that is capable of binding to the effector protein, and (iii) a guide sequence that is capable of hybridizing with the target nucleic acid, and (iv) a reporter that is single stranded and does not hybridize with the guide sequence of the guide nucleic acid, (b) measuring signals detected from the at least two nanovolumes and generated by cleavage of the reporter by the effector protein, and (c) quantitating the target nucleic acid in the test sample based on the measured signals. The disclosure provides methods for quantitating a target nucleic acid that do not comprise amplifying the target nucleic acid in the test sample.

[0125] In some embodiments, the contacting step comprises sequentially adding each of the components (i) through (iii) to the test sample in at least two nanovolumes. The order of adding the components is not limited. In some embodiments, two or more of the components may be added together at the same time. In some embodiments, the contacting step comprises: (a) adding each of the components (i) through (iii) to the test sample to generate a master reaction mixture, wherein the master reaction mixture has a volume of more than 1 nL, and (b) distributing the master reaction mixture into at least two nanovolumes.

[0126] In some embodiments, the methods disclosed herein utilize an instrument that is capable of partitioning the reaction mixture into two or more nanovolumes (within two or more compartments or partitions), wherein each of the compartments comprises a nanovolume to be used in the methods described herein. Non-limiting examples are such instruments include a fluidic device (e.g., a Beckman Coulter Echo 525 Liquid Handler), or a digital PCR machine (e.g., QIAcuity Digital PCR System). In some embodiments, the methods disclosed herein utilize chambers comprising two or more compartments or partitions, wherein the two or more compartments comprise the two or more nanovolumes, as described herein.

[0127] In some embodiments, the contacting step comprises contacting the test sample with components (i) through (iii) at a first pre-determined temperature. The method further comprises changing the temperature from the first pre-determined temperature to a second pre-determined temperature after the reaction mixture is partitioned into the at least two nanovolumes. The first pre-determined temperature may be selected to reduce or prevent activity of and/or interaction between one or more of the components of the reaction mixture. The second pre-determined temperature may be selected to increase or enable activity of and/or interaction between one or more of the components of the reaction mixture.

[0128] In some embodiments, the first pre-determined temperature may be a temperature at which the effector

protein has reduced or no catalytic activity, such that the effector protein remains inactive during partitioning in order to delay the start of the detection reaction (and generation of detectable signals by cleavage of the reporter by the effector protein) until the at least two nanovolumes have been formed. The second pre-determined temperature may be a temperature at which the effector protein has improved catalytic activity relative to the first pre-determined temperature, such that changing the temperature from the first pre-determined temperature to a second pre-determined temperature activates the effector protein and begins the detection reaction. For example, an effector protein may be selected which has little to no catalytic activity at a partitioning temperature of 4C, room temperature, or 37C, but is highly active at a reaction temperature of 50C, 55C, or 60C. In another example, an effector protein may be selected which has some catalytic activity at a partitioning temperature of 4C, room temperature, or 37C, but is more active at a reaction temperature of 50C, 55C, or 60C, thus enabling the majority of reporter cleavage to occur after the reaction mixture has been partitioned into nanovolumes. In some embodiments, the effector protein may be engineered for specific levels (or not) of catalytic activity at specific temperatures as described herein. Such temperature shifting warm start strategies may be particularly useful when the chosen effector protein is highly active at its preferred operating temperature, which could result in high background within nanovolumes if the effector protein were active from the moment of contacting and throughout the partitioning process.

[0129] In some embodiments, the effector protein may be associated with one or more blockers (e.g., aptamers, antibodies, etc.) which prevent catalytic activity until disassociated therefrom. Blockers may be selected for their temperature sensitivity such that they remain associated with the effector protein at the first pre-determined temperature and disassociate from the effector protein (e.g., by denaturing) at the second pre-determined temperature. Upon disassociation, the effector protein is then active and free to begin the detection reaction within the nanovolumes. Such warm start strategies may be particularly useful when the chosen effector protein is highly active across a range of temperatures (e.g., at room temperature and at temperatures higher than the disassociation temperature of the blocker), which could result in high background within nanovolumes if the effector protein were active from the moment of contacting and throughout the partitioning process.

[0130] In some embodiments, the contacting step comprises contacting the test sample with one or more signal amplification reagents. For example, the one or more signal amplification reagents may comprise (i) a second effector protein and (ii) a second reporter that is single stranded and does not hybridize with the guide sequence of the guide nucleic acid. The second reporter may comprise a first nucleic acid section and a second nucleic acid section. The first nucleic acid section may act as an activator for the second effector protein when separated from the second nucleic acid section. The second nucleic acid section may act as a blocker nucleic acid and prevent the first nucleic acid from binding to and/or activating the second effector protein. In some embodiments, the second effector protein may comprise a Type III Cas protein such as NucC, Csm6, etc. and the second reporter is a capped activator thereof. Presence of the target nucleic acid may activate the first effector

protein, which may cleave the first nucleic acid section from the second nucleic acid section, thereby freeing the first nucleic acid section to bind to and activate the second effector protein. The activated second effector protein then cleaves the first reporter to generate the signal detected in the measuring step. In some embodiments, the first effector protein may cleave both the first reporter and the second reporter.

[0131] In some embodiments, each of the at least two nanovolumes comprises no more than 1 molecule of the target nucleic acid. In some embodiments, the contacting step results in a number of nanovolumes in the range of about 5000 nanovolumes to about 500,000 nanovolumes, for example, about 6000, about 7000, about 8000, about 9000, about 10,000, about 20,000, about 30,000, about 40,000, about 50,000, about 60,000, about 70,000, about 80,000, about 90,000, about 100,000 nanovolumes, about 200,000 nanovolumes, about 300,000 nanovolumes, about 400,000 nanovolumes, or about 500,000 nanovolumes. In some embodiments, step (b) comprises measuring a binary signal (e.g., positive or negative) from each of the at least two nanovolumes. In some embodiments, step (c) comprises counting the number of nanovolumes providing a signal indicative of the presence of the target nucleic acid therein.

[0132] In some embodiments, the methods disclosed herein comprise contacting the test sample with a precursor guide RNA array, wherein the effector protein cleaves the precursor guide RNA array to produce the guide RNA. In some embodiments, the effector protein and the guide RNA are incubated with each other prior to step (a). In some embodiments, the effector protein and the guide RNA are not incubated with each other prior to step (a).

[0133] In some embodiments, the methods further comprise quantitating multiple different target nucleic acids in the test sample, wherein the at least two nanovolumes of reaction mixture comprises one or more guide RNAs, wherein at least one guide RNA is capable of hybridizing with each of the multiple target nucleic acids. In some embodiments, each of the at least two nanovolumes of reaction mixture comprises at least one guide RNA that is capable of hybridizing with each of the multiple target nucleic acids. In some embodiments, no two of the one or more guide RNAs are capable of binding to the same target nucleic acid. In other words, in some embodiments, each of the one or more guide RNAs is capable of binding a different target nucleic acid among the multiple target nucleic acids in the test sample.

[0134] In some embodiments, methods of quantitating multiple target nucleic acids may comprise using one or more guide RNAs comprising a guide sequence that is capable of hybridizing with target nucleic acids (e.g., bacterial or viral target nucleic acids) in a patient sample to inform treatment strategy. For instance, detection of a bacterial target nucleic acid in the sample can guide the administration of antibiotics. In some embodiments, the methods disclosed herein may be used for pan-disease determination (e.g., pan-influenza testing with guide nucleic acids targeting multiple H/N subtypes), which may be especially useful when exact knowledge of the disease-causing agent or subtype is unnecessary to properly inform treatment.

[0135] In some embodiments, each of the at least two nanovolumes of reaction mixture comprises at least one of the one or more guide RNAs comprising a guide sequence that is capable of hybridizing with no more than one of the

multiple target nucleic acids. In some embodiments, each of the at least two nanovolumes of reaction mixture comprise different guide RNAs so as to detect different target nucleic acids.

[0136] In some embodiments, the methods further comprise quantitating a positive control target nucleic acid in a positive control sample, the method comprising: (a) contacting the positive control sample with the following components (i) through (iii) resulting in at least two nanovolumes of reaction mixture: (i) a effector protein; (ii) a positive control guide RNA, comprising a region that is capable of binding to the effector protein, and a guide sequence that is capable of hybridizing with the positive control target nucleic acid; and (iii) a reporter nucleic acid that is single stranded and does not hybridize with the guide sequence of the positive control guide RNA; (b) measuring signals detected from the at least two nanovolumes and produced by cleavage of the reporter nucleic acid by the effector protein, and (c) quantitating the amount of positive control target nucleic acid in the positive control sample based on the measured signals. In some embodiments, the at least two nanovolumes comprises more than one type of CRISPR/Cas effector protein, and one or more positive control guide RNAs capable of binding to each of the more than one type of CRISPR/Cas effector protein. In some embodiments, the at least two nanovolumes comprises a DNA-targeting effector protein (e.g., a Cas12 protein) and an RNA-targeting effector protein (e.g., a Cas13 protein), a positive control guide RNA comprising a region that is capable of binding to the DNA-targeting effector protein and a positive control guide RNA comprising a region that is capable of binding to the RNA-targeting effector protein.

[0137] In some embodiments, the methods further comprise quantitating the target nucleic acid in a positive control sample, the method comprising: (a) contacting the positive control sample with the following components (i) through (iii) resulting in at least two nanovolumes of reaction mixture: (i) a CRISPR/Cas effector protein; (ii) the guide RNA, comprising a region that is capable of binding to the CRISPR/Cas effector protein, and a guide sequence that is capable of hybridizing with the target nucleic acid; and (iii) a reporter nucleic acid that is single stranded and does not hybridize with the guide sequence of the guide RNA; (b) measuring signals detected from the at least two nanovolumes produced by cleavage of the reporter nucleic acid by the CRISPR/Cas effector protein, and (c) quantitating the amount of target nucleic acid in the positive control sample based on the measured signals.

[0138] In some embodiments, the methods comprise generating a standard curve for the target nucleic acid in the positive control sample, and obtaining an absolute quantitation of the target nucleic acid in the test sample based on the standard curve. In some embodiments, the method comprises obtaining a relative quantitation of the target nucleic acid in the test sample based on the quantitation of the target nucleic acid in a positive control sample.

[0139] The methods disclosed herein may be used for a wide range of applications, including, disease surveillance (e.g., wastewater-based epidemiology), cell and gene therapy, therapeutic selection, therapeutic monitoring, minimum residual disease (e.g., for oncology applications, etc.), carrier screening, oncology, cDNA conversion rates, and mRNA quality control (mRNA therapeutics, mRNA vaccines, mRNA reagents). For example, the methods disclosed

herein may be used for disease surveillance as described in Examples 6 and 9. In another example, the methods disclosed herein may be used for various cell and gene therapy applications, including quantifying mRNA and/or gRNA payloads of lipid nanoparticles (LNPs) or other nucleic acid delivery vehicles such as adeno-associated viruses (AAVs), quantifying gene silencing efficiency in treated tissue samples, quantifying nucleic acid delivery to tissues (e.g., pharmacokinetics), etc. Current methods for RNA quantitation are typically unable to discriminate between mRNA and gRNA, which is less than ideal for determining therapeutic dosages, manufacturing purity, etc. Similar challenges exist for distinguishing between gRNA and mRNA in tissues when quantifying gene silencing efficiencies. The methods disclosed herein may be used to quantify and distinguish between different RNA species within a single delivery vehicle or tissue sample, e.g., using a multiplexing and/or guide pooling strategy as described herein. In another example, the methods disclosed herein may be used to monitor nucleic acid targets within a cell or tissue which are associated with a particular disease state or severity, e.g., to monitor the efficacy of a therapeutic agent within a tissue by determining the amount of relevant biomarker present in the tissue. In another example, the methods disclosed herein may be used for mRNA quality control. The methods disclosed herein may be used to identify and/or quantify mRNA capping impurities, e.g., in vitro transcribed mRNA products which may be incorporated into delivery vehicles, such as for mRNA-based vaccines or gene therapies. Alternatively, or in combination, the methods disclosed herein may be used to identify and/or quantify unwanted nucleic acid impurities within a given samples, e.g., dsRNA contaminating in vitro transcribed mRNA products. A person of ordinary skill in the art will understand based on the disclosure herein that these and other applications are within the scope of the claims and embodiments described herein.

#### Effector Proteins

[0140] An effector protein provided herein interacts with a guide nucleic acid to form a complex. In some embodiments, the complex interacts with a target nucleic acid. In some embodiments, an interaction between the complex and a target nucleic acid comprises one or more of: recognition of a protospacer adjacent motif (PAM) sequence within the target nucleic acid by the effector protein, hybridization of the guide nucleic acid to the target nucleic acid, modification of the target nucleic acid by the effector protein, or combinations thereof. In some embodiments, recognition of a PAM sequence within a target nucleic acid may direct the modification activity of an effector protein.

[0141] In some embodiments, the effector protein is a CRISPR-associated (“Cas”) protein. An effector protein may function as a single protein, including a single protein that is capable of binding to a guide nucleic acid and editing a target nucleic acid. Alternatively, an effector protein may function as part of a multiprotein complex, including, for example, a complex having two or more effector proteins, including two or more of the same effector proteins (e.g., dimer or multimer). An effector protein, when functioning in a multiprotein complex, may have only one functional activity (e.g., binding to a guide nucleic acid), while other effector proteins present in the multiprotein complex are capable of the other functional activity (e.g., editing a target nucleic acid). In some embodiments, an effector protein,

when functioning in a multiprotein complex, may have differing and/or complementary functional activity to other effector proteins in the multiprotein complex. An effector protein may be a modified effector protein having increased modification activity and/or increased substrate binding activity (e.g., substrate selectivity, specificity, and/or affinity).

[0142] An effector protein may be small, which may be beneficial for nucleic acid detection (for example, the effector protein may be less likely to adsorb to a surface due to its small size). The smaller nature of these effector proteins may allow for them to be more readily incorporated as a reagent in an assay. In some embodiments, the length of the effector protein is at least about 100, about 200, about 300, about 400, about 500, about 600, about 700, about 800, about 900, about 1000, about 1100, about 1200, about 1300, about 1400, about 1500, or more linked amino acid residues.

[0143] TABLE 1 provides illustrative amino acid sequences of effector proteins that are useful in the compositions, systems and methods described herein.

[0144] In some embodiments, compositions, systems and methods described herein comprise an effector protein, or a nucleic acid encoding the effector protein, wherein the amino acid sequence of the effector protein comprises at least about 200 contiguous amino acids or more of any one of the sequences recited in Error! Reference source not found. In some embodiments, the amino acid sequence of an effector protein provided herein comprises at least about 200, at least about 220, at least about 240, at least about 260, at least about 280, at least about 300, at least about 320, at least about 340, at least about 360, at least about 380, at least about 400 contiguous amino acids, at least about 420 contiguous amino acids, at least about 440 contiguous amino acids, at least about 460 contiguous amino acids, at least about 480 contiguous amino acids, at least about 500 contiguous amino acids, at least about 520 contiguous amino acids, at least about 540 contiguous amino acids, at least about 560 contiguous amino acids, at least about 580 contiguous amino acids, at least about 600 contiguous amino acids, at least about 620 contiguous amino acids, at least about 640 contiguous amino acids, at least about 660 contiguous amino acids, at least about 680 contiguous amino acids, at least about 700, at least about 800, at least about 900, at least about 1000, at least about 1100, at least about 1200, at least about 1300, at least about 1400, at least about 1500, contiguous amino acids, or more of any one of the sequences of Error! Reference source not found.

[0145] In some embodiments, compositions, systems and methods described herein comprise an effector protein or a nucleic acid encoding the effector protein, wherein the effector protein comprises a portion of any one of the sequences recited in TABLE 1. In some embodiments, the effector protein comprises a portion of any one of the sequences recited in TABLE 1, wherein the portion does not comprise at least the first 10 amino acids, first 20 amino acids, 40 amino acids, 60 amino acids, 80 amino acids, 100 amino acids, 120 amino acids, 140 amino acids, 160 amino acids, 180 amino acids, or 200 amino acids of any one of the sequences recited in TABLE 1. In some embodiments, the effector protein comprises a portion of any one of the sequences recited in TABLE 1, wherein the portion does not comprise the last 10 amino acids, the last 20 amino acids, 40 amino acids, 60 amino acids, 80 amino acids, 100 amino acids, 120 amino acids, 140 amino acids, 160 amino acids,

180 amino acids, or 200 amino acids of any one of the sequences recited in TABLE 1.

[0146] In some embodiments, compositions, systems, and methods described herein comprise an effector protein, or a nucleic acid encoding the effector protein, wherein the effector protein comprises an amino acid sequence that is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identical to any one of the sequences as set forth in TABLE 1. In some embodiments, an effector protein provided herein comprises an amino acid sequence that is at least 65% identical to any one of the sequences as set forth in TABLE 1. In some embodiments, an effector protein provided herein comprises an amino acid sequence that is at least 70% identical to any one of the sequences as set forth in TABLE 1. In some embodiments, an effector protein provided herein comprises an amino acid sequence that is at least 75% identical to any one of the sequences as set forth in TABLE 1. In some embodiments, an effector protein provided herein comprises an amino acid sequence that is at least 80% identical to any one of the sequences as set forth in TABLE 1. In some embodiments, an effector protein provided herein comprises an amino acid sequence that is at least 85% identical to any one of the sequences as set forth in TABLE 1. In some embodiments, an effector protein provided herein comprises an amino acid sequence that is at least 90% identical to any one of the sequences as set forth in TABLE 1. In some embodiments, an effector protein provided herein comprises an amino acid sequence that is at least 95% identical to any one of the sequences as set forth in TABLE 1. In some embodiments, an effector protein provided herein comprises an amino acid sequence that is at least 97% identical to any one of the sequences as set forth in TABLE 1. In some embodiments, an effector protein provided herein comprises an amino acid sequence that is at least 98% identical to any one of the sequences as set forth in TABLE 1. In some embodiments, an effector protein provided herein comprises an amino acid sequence that is at least 99% identical to any one of the sequences as set forth in TABLE 1. In some embodiments, an effector protein provided herein comprises an amino acid sequence that is identical to any one of the sequences as set forth in TABLE 1.

[0147] In some embodiments, compositions, systems, and methods described herein comprise an effector protein, or a nucleic acid encoding the effector protein, wherein the effector protein comprises one or more amino acid alterations relative to any one of the sequences recited in TABLE 1. In some embodiments, the effector protein comprising one or more amino acid alterations is a variant of an effector protein described herein. It is understood that any reference to an effector protein herein also refers to an effector protein variant as described herein. In some embodiments, the one or more amino acid alterations comprises conservative substitutions, non-conservative substitutions, conservative deletions, non-conservative deletions, or combinations thereof. In some embodiments, an effector protein or a nucleic acid encoding the effector protein comprises 1 amino acid alteration, 2 amino acid alterations, 3 amino acid alterations, 4 amino acid alterations, 5 amino acid alterations, 6 amino acid alterations, 7 amino acid alterations, 8 amino acid alterations, 9 amino acid alterations, 10 amino acid alterations or more relative to any one of the sequences recited in TABLE 1.

**[0148]** In some embodiments, the one or more amino acid alterations may result in a change in activity of the effector protein relative to a naturally-occurring counterpart. For example, and as described in further detail below, the one or more amino acid alteration increases or decreases catalytic activity of the effector protein relative to a naturally-occurring counterpart. In some embodiments, the one or more amino acid alterations results in an effector protein variant that has lower catalytic activity at low temperatures (e.g., room temperature) compared to higher temperatures (e.g., 50C), relative to a naturally-occurring counterpart, e.g., in

order to facilitate a warm-start detection method. In some embodiments, the one or more amino acid alteration increases catalytic activity of the effector protein at elevated temperatures relative to a naturally-occurring counterpart. In some embodiments, the one or more amino acid alteration improves stability and/or manufacturability (e.g., expressibility, solubility, purification, etc.) of the effector protein relative to a naturally-occurring counterpart. In some embodiments, the one or more amino acid alteration improves efficiency of guide RNA complexing relative to a naturally-occurring counterpart.

TABLE 1

Amino Acid Sequences of Exemplary CRISPR/Cas effector proteins	
SEQ ID NO:	Programmable Nuclease Amino Acid Sequence
SEQ ID: 1	MADLSQFTHKYQVPKTLRFELIPQGKTLENLSAYGMVADDQKRSENYKKLKPVIDRIYK YFIEESLKNLNDWNPLYEAIERYRKEKTTATITNLKEQQDICRRAIASRFEKGKVPDKGD KSVKDFNKKQSLFKELFGKELFTDSVLEQLPGVLSDEDKALLKSFDKFTTYFVGFYD NRKVNFSDDDISTGIPHLVQENFPKFIDNCQFLVLAPELKEKAAEATKIFED VSLDEIFSIKFYRNLLQQNQIDQFNQLGGIAGAPGTPKIQGLNETLNLSMQQDKTLEQK LKSVPFRFSPLYQIILSDRSSLFPIESFSCDAEVLLAVQBYLDNLKTEHVIDLKEVENR LTTLSDLKHIVYVNSTKVTAFSQLFGDWNLRCRQLRVYKMSNGNEKITKKALGELESWL KNSDIAFTELQALADELPAVKNLKVQEAIISGLNEMQAKSLPKELKIPEEKEELKALLD AIQEYVYHTLEWFIVSDDVETDTDYVPLKTLQIIPQIPLYNKVRNFATQKPYSVKEFKL NFANPTLADGWDENKEQONCAVLFQKGNNYLLGILNPKNPKDPFDNVDTTEQGNCYQK MVKQFPDFSKMMPKCTTQLVKQHFEKGKDSDYLINNKNKFIKPLTI TREVYDLNNVL YDGKKKQFQIDYLRTKTEDGYYTALHTWIDPKFVASYKSTSIIYDTSTILPPEKEYEKL NEFYGAQDNLFYQIKFENIPEEIIDTVYEDGKLFLPQIYNKDPAAAGATGAPNLHTIYWKA VFDPENVKDVVVKLNGQAEFLYRPKSNMDVIRHKVGEKLVNRTLKDGSILTDELHKE YLYANGSLKKGLSEDAKIILDLKNLAVIYDVHHEIVKDRRTTDKFFFHVPLTNYKCDK NPVFKFNAEVQEYKLENKEDTYYIGIDRGERNLIIYAVVIDPKGRIVEQKSFNVINGFDYHGK LDQREKERVKARQAWTAVGKIELKQGQLSLLVHEISKMMVRYQAVVVLENLNVGFK RVRSGIAEKAVYQFCKMLINKLNLYMFKDAGGTTEPGSVLNAYQLTDRFESFAKMGLO TGFLFYIPAFTSKIDPATGFVDPFRWGAIKTLADKREFLSGFESLKFDSTTGNFILHFDS KNKNFQKLEGFPVDWIIIIEANKMTGKATYIAGKRIEPRDNNSQGHYEDYLPNC ALAETLRQCDIPYEEGKDILPLILEKNDSKLHSVFKVVRLLTQMRNSNAETGEDYISSPV EDVSGSCFDSRMENEKLPKDADANGAYHIALKGMLALERLRKDEKMAISNNDWLNYI QEKRA*
SEQ ID: 2	MAGKKKDVKINTLKVRIIRPRYSSDIEKEISDEKAKRQDGKTGELDRAFTSELKSRN PIDITNDLFPFLTEIQKNLTIYKNKISLILYMKLIVEEEEQGSTASALSAQPYKECKAREN YISLGLRKIQSNSFRRKELKGQFVSLPTAKSDRFPPIPFCHQVENGKGGFKVYETGDDFIFE VPLIKYTATNKKSTSGKNYTKVQLNNPPVMNVPLLLSTMRRQTKKGMOQNWDKDEGT NAELRRVMSGEYKVSYAEIIRRTRFGKHHDDWVNFSKFKNKTDELNQNVRGGIDIGVS NPLVCAVINGLDRYIVANNIDAFNERAMARRTLLRKNRKRSHGAKNLKEPITVL TEKNERFRKSILQRWAREVAEFFKRTSASVNMEDLSGITEREDFFSTKLRTTWNYRML QTIIENKLKEYGIAVNYISPKYTSQTCGHSCGRNDYPTFSYRENNYPPFECKECNKVKC NADFNAAKNIALKVVL
SEQ ID: 3	MAKNTITKTLKLRLIVRPYNSAEVEKIVADEKNREREKIALEKKNKDKVKEACSKHLVAA YCTTQVERNACILFCKARKLDDFKYQFLRGQFPAVFWQEISEIFRQLQKQAAEIYNQSL IELYYEIFIKGKGNASSVEHYLSDVCYTRAEEFKNAAIASGLRSKIKSNFRKLKEKLKNM KSGLPPTTSDNPFPIPLVKQKGQYTFGEPEISNHNSDFIICIPGRWQVKIEDKYPWEKF FEQVQKSPKPISSLSTQRRKRKNKGWSKDEGTEAEIKVMNGDYQTSYIEVKRGSKIGE KSAWMLNLSIDVPKIDKGVDPSIIGGIDVGVKSPLVCAINNFSRYSISDNDLFHFNKKM FARRRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLIERWACEIADEFIINKVGT VQMNLESMKRKEQDSYFNIRLGFWPYAEQMOKIEFKLKQYGIIEIRKVAPEWNTSKTCSK CGHLLNNYFNFEYRKKNKFPFKCEKCNFKENADYNAALNISNPKLKSTKEEP
SEQ ID: 4	MATLVSFTKQYQVQKTLRFELIPQGKTQANIDAKGFINDDLRDENYMKVKGVIDELH KNFIEQTLVNVVDYDWRSLATAIKNRYKDRSDTNKKNLEKTQEAAKKEIIIAWFEGKGN SAFKNNQKSFYGLFKKELFSEILRSDDLEYDEETQDAIAFCDFKFTTYFVGFHENRKNM YSTEAKSTSVAYRVVNENFSKFLSNCNEAFSVLEAVCPNVLVEAEQELHLHKAFDSLKLS DVPKVEAYNKYLSQTGIDYYNQIIGGIISSAEGVRKIRGVNEVNNNAIQQNDELKVALRN KQFTMVQLFKQILSDRSTLSPVSEQFTSDQEVITVVKQFNDIVNNKVLAVVKTLFENFN SYDLEKIIYINSKELASVSNALLKDWSKIRNAVLENKIIELGANPPKTKISAVEKEVKNKDF SIAELASYNDKYLDEGNDKEIICSIANVLEAVGALEMIAESLPAIDLKTLLENKNKVKGI LDAYENLLHLLNYFKVASVNDVLDLAFYGAPEKVVYDISHGMPYLYNKVRNYATKKPYS VEKEKFLNFMAMPFLADGWDKNCERDNGSIIILKKDGQYLLGVMPQNKPVIDNAVCNDA KGYQKMWYKMFPEISKMVTKCSTQLNNAVKAHFEDNTNDFVLDLDTDKFISDLTTITKEIYD LNNVLYDGGKKFQIDYLRTNGDFAGYHKALETWIDFVKEFLSKYRSTAIYDLTTLLPTN YYEKLDFVYSDVNNLCYKIDYENISVEQVNEWVEEGNLYLPKIJNKDFATGSTGKPNL

TABLE 1-continued

Amino Acid Sequences of Exemplary CRISPR/Cas effector proteins	
SEQ ID NO:	Programmable Nuclease Amino Acid Sequence
	HTMYWNAVFAEENLHDVVVKLNGGAEFLYRPKSNMPKVEHRVGEKLVNRKNVNGEPI ADSVHKETIYAYANGKIKSKSELENAQEELPLAIIDKDVKHNTKDKRYLSKYFFFVPITLN YKANGNPSAFNTKVQAFKLKNNPDVNIIGIDRGERNLLYVVVIDQQGNIIDKKQVSYNKV NGYDYYEKLNRREKERIEAROSWGAVGKIKELKEGYLSSLVREIADEMMVYKNAIVVM ENLNAGFKRVRGIAEKAQVQKFEMKLIDKLNYLVFKDVEAKEAGGVLNAYQLTDKF DSFEKMGNQSGLFVPAAYTSKIDPVTGFANVFSTKHITNTTEAKKEFICSFNSLRYDEA DKDFVLECDLNFKIVANSHIKNWKPIIGGKRIVYNSKNKTMEKYPCEDLKATLNASG IDFSSSIINLLKNVPAUREYKGKLFETYWAIMNTLQMRNSNALTGEDYIIISAVADDNEK VFDSSRTCGAELPKDADANGAYHIALKGLYLLQRIDISEEGEVKDLISKNEEWFKFVQQK EYAR*
SEQ ID: 5	MCKMKITKIDGISHKKYKEKGKLIKNNDTAKDIIIEERFNDIEKKTKELFQKTLDFYVKNYE KCQEQNKRERAKNYFSKVKILVDNKKITICNENTEKMEIEDPNEYDVRSRGKYFNVLN KILNGENYTEEDLEVFENDLQKRTGRIKSINKSLEENKAHFKKESINNNIIYDRVKGNNK KSLFYEYYRISSKHQBEYVNNIEFAFDKLYSNSHEAMNNLFSEITKDSKDRNIRKIREAYH EIILNKNTCEFGEELYKKIQDNRRNNFDKLLIEPEIKELTKSQIFYKYYIDKVNLDETSIKHC FCHLVEIEVNQLLKNVYYSKRNIKLENIFEYCKLKNLJKNKLVNKLNNYIRNCGKY NAYISNNDDVVNVNSEEKLSIERTKEAFLRSIIGVSSAYFSRLNILNTDNTQDITNKVDKEVD KLYQENKKIELEERLKLFNGNYFDINNNQOEIKVFLMNIDKIISIRRHEIIFHKMETNAQNF DENNVNLNGNTAKNFSNEINEEKKIPEKFKIFQQLNSANVFDYLSNKDITEYMDKVVSFTNR NVSFVPSFTKIQNVRQDANSLEIKWKWPDKSEGKDAQIYLLKNIYYGKFDELFLNEEN GIFISIKDKIIEELRNQNKRGTFCYKLEKFEKIEETNPCKYLEIIQSLYMINIEEIDSEGRNIFL DPIQKIFLKGFFEFIKNNYNNLLELKQDQKKNIPDSEMSEYIAGEKTLDEDIGEINEEIQQDIKI TEIDKILNQTDKINCFCVLLKLNNYKEITELKGNLKEQYQILSKTNVYEKELMLLNIVNLDN NKVKIENFKILAAEIGEIKEINIEEINKNNKIKTFEELRNFEKGENTGEYYNIYSDDKNIK NIRNLNYYKQGMDLLEKISEKTNYCICKKDLEFYSELRKQLEDEKTNFYKIQEYLHSK YQQPKKKILLNNKNDYEKYKKSIENIKEYVHLKNKIEFNELNLLQSLLLKILHRLVGFT SIWERDLRFRLIGEFPDLEDVDEDIFDHRKRYKGTGKGICKYDRFINTHTEYKNNNKME NVKFADNNPVPRVNYIAHFNYLPNPKYSILKMMEKLRLKLDYDRKLKNAVMSIKDILEE YGFKAIFIINSDEKIIILNLVKSVEIILHGKEDLKSRRNSEDLCKLKVAMLEYSK*
SEQ ID: 6	MEDKQFLERYKEFIGLNSLSKTLRNSLIPVGSTLKHQEQYEGILEEDSLRAQKREELKGIMD DYRNYIEMHLDVHDIDWNELFEALTEVKKNQTDAAKKRLEKIQEKKREIYQYLS DAVESEMFKLERMISGILPDFIRCNEGYSSEEKEELKLTVALFHRFTSSFNLNRKNVFT KEAIVTAIGYRVVHENAEIFLENMVAFQNIQKSAESQISIIERKNEHYFMEWKLISHIFTAD YYMMLMTQKAEHEYNECMGVNVQQMREYCQKEKKNWLYRMKRLHKQILSNASTSF KIPEKYENDAEYESVNSFLQNVMEKTVMERIAVLLKNSTDNFDLSKITYTAPYYEKISNY LCGSWNTITDCLTHYYEQQIAJGKARDQKVAAVADKWKSLSEIEQLLKEYARAEE VVKRPEYIAEIEIENIISLKEAHLLEYHPEVNLIENEYATEIKDVLNDNYMELFHWMKWF YIEEAVEKEVNFGYGEELDDLYEEIKDIPVLYNKVRNYVTQKPYSDTKIKLNGFTPTLANG WSKSKEYDYNAILLQDGKYYMGIFNPPIQKPEKEIIEGHSGQPLEGNEYKKMVYYLPSA NKMLPKVLLSKMGMEIYQPSEYIINGYKERRHISKEEKFQFDLQFCHDLIDYFKSGIERNSD WKVFGFDFSDTTDQDISGFYREVEDQGYKIDWTYIKEYADIDRNLNEEGKLYLFOIYNK FSEKSTGRENLHTMLKLNLFSEENVREQLKLNGEAEIFFRKSSVKKPIIHKGTMVNR TYMEEVNGNSVRNIPKEKEYOEIINYKHNHRLKGELSTEAKKYLEKAVCHETKKDVKD YRYSVDFKFFIHLPIITINYRASGKETLNSVAQRYIAHQNDMHVIGIDRGERNLIVYV OGEIKEQKSFNINEFNYKEKLERBQSRAARRNWKEIQQIKDLKEGYLGSVIEIAKM MIKYHAIAMEDLNYGFKRGRFKVERQVYQKFENMLIQKLNYLVFKDRPADEDGGVLR GYQLAYIPDSVKKMGRQCGMIYVPAFTSKIDPTTGFVDIFKHKVYTTTQAKREFILSF DEICYDVERQLKQFVYANFVTQNVTLARNNTWYTITNGTRAQKEFGNGRMRDKEDYN PKDKMVELLESIGIEFKSGKNLLPALKKVSNAKVFEELQKIVRFTVQLRNSKSEENDVD YDHVISPVNLNEEGNFFDSSKYKNKEEKESLLPVDADANGAYCIALKGLYIMQAIQKNW SEEKALSPDVLRNNNDWFYDQIQNKRYS*
SEQ ID: 7	MEEKKMSKIEKFIGKYKISKTLRFRAVPGKTDQNI EKKGI LEKDKKRS EDYEKVKAYL DSLH RDFIENTLKKVNLNEALCFSGTKDDGDKKMEKLEEKMRKTISNEFCNDE MYKKISSEKILSNEEDVSDIVSSYKGFTSLNGYVNNRKNLYVSDAKPTSIAYRCINE NLPKFLRNVECYKKVQVIPKEQIEYMSNNNLSPYRIEDCFNIDFFEPFCLSQQGIDLNT FIGGYSKKDGTKVQGINELVLYNQKNNKKDKEKYKLQPQFTPLFKQIILSDRTKFSIEKL ENIYEVEVLLVKKSYSDEMFDIETVFSNLNYYDASGIVYKNGPAITHISMNLTKDWATIR NNWNYEYDEKHTKKNNKI EKYEDTRNTMYKKIDSFTLEYISRLVKGKDI DELVKYFENE VANFVMDIKKFTSKLPLFDRCQKENFDISEDEVNDIKGYLDNVKLLESPMKSFTINGKE NNIDYVYFGKFTDDYDKLHEFDHYNKVRNYITTSRKPYKLKDQYLYFDNPQLLGGWD INKEKDYRTVMLTKDGKYYFAIIDKGEHPFDNIPKDYFDNNNGYKKIYRQIPNAAKYLS SKQIVPQNPPEEVKRLDKKKADSKSLTEEEKNIFIDYIKSDFLKNYKLLFDKNNNPYFNF AFRESSTYESLNEEFFDVERQAYSVRYENLPADYIDNLVNEGKYLFEIYSKDFSEYSKGT NNLHTMYFKALFDNDNLKNTVFKLGSNAEFLPFRASIKKDELVIPHKNQLLQNKPNLNP KKQSIPDYDLVLDKDRPFENQYMLHISIEINKNERDAKKIKNINEMVRKELKDSDDNNYIIGI DRGERNLILYCVVINSAGKIVEQMSLNEIINEYNGIKHTVDYQGLLDCKEKERNQAQRQSW KSIENIKELKDGYISQVVKHLQCLVEKYDAAIAMELNNGGFKRGRTKFEKQVYQFENK LINKMEYMADKKRKTTEGGILRGYQLTNGCINNSYQONGIFIYVPAWLTSKIDPTTGFV DLLKPKYTNVEE AHLWINKFNSITYDKLDMFAFNINYSQFPRADIDYRKIWTFTNGY

TABLE 1-continued

Amino Acid Sequences of Exemplary CRISPR/Cas effector proteins	
SEQ ID NO:	Programmable Nuclease Amino Acid Sequence
	R I E T F R N S E K N N E F D W K E V H L T S V I K K L L E E Y Q I N Y I S G K N I I D D L I Q I K D K P F W N S F I K Y I R L T L Q M R N S I T G R T D V D Y I I S P V I N N E G T F Y D S R K D L D E I T L P Q D A D A N G A Y N I A R K A L W I I E K L K E S P D E E L N K V K L A I T Q R E W L E Y A Q I N I *
SEQ ID: 8	M E K I K K P S N R N S I P S I I S D Y D A N K I K E I K V K Y L K L A R L D K I T I Q D M E I V D N I V E F K K I L N G V E H T I I D N Q K I E F D N Y E I T G C I K P S N K R R D G R I S Q A K Y V V T I T D K Y L R E N E K E K R F K S T E R E L P N N T L L S R Y K Q I S G F D T L T S K D I Y K I K R Y I D F K N E M L F Y F Q F I E F F N P L L P K G K N F Y D F N I E Q N K D K V A K F I V Y R L N D D F K N K S L N S Y I T D T C M I I N D F K K I Q K I L S D P R H A L A H F D P D F I Q K F F D D Q L D K N K L I N T I L L D Q K E E K N Y Q E K N N Y I D D D N I L T I F D E K G S K F S K L H N P Y T K I I S Q K K P A F F N K L I N S F L S Q D G V P N E E F K S Y L V T K K L D F F E D I H S N K E Y K K I Y I Q H K N L V I K K Q K E E S Q E K P D G Q K L K N Y N D E L Q O K L K D E M N T I T K Q N S L N R L E V K L R L A F G F I A N E Y N Y N F K N F N D E F T N D V K N E Q K I K A F K N S S N E K L K E Y F E S T F I E K R F F H F S V N F F N K K T K K E E T K Q K N I F N S I E T N D V K N E Q K I K A F K N S S N E K L K E Y F E S T F I E K R F F H F S V N F F N K K T K K E D E I S I B D A Q N S F S L K F K I L A K N L R G L Q L F H Y S L S H N T L Y N N K Q C F F Y E K G N R W Q S V Y K S F Q I S H N Q D E F D I H L V I P V I K Y Y I N L N K L M G D F E I Y A L L K Y A D K N S I T V K L S D I T S R D D L K Y N G H Y N F A T L L F K T F G I D T N Y Q K N V S I Q N I K K T R N N L A H O N I E N M L K A F E N S E I F A Q R E E I V N Y L Q T E H R M Q E V L H Y N P I N D F T M K T V Q Y L K S L S V H S Q E G K I A D I H K K E S L V P N D Y Y L I Y K L K A I E L L K Q K V I E V I G E S E D E K K I K N A I A K E E Q I K K G N N
SEQ ID: 35	M E K S L N D F I G L Y S V S K T L R F E L K P V S E T L E N I K K F H F L E E D K K K A N D Y K D V K K I I D N Y H K Y F I D D V L K N A S F N W K K L E E A I R E Y N K N K S D D S A L V A E Q K K L G D A I L K L F T S D K R Y K A L T A A T P K E L F E S I L P D W F C E Q C N Q D L N K A L K T F Q K F T S Y F T G F Q E N R K N V Y S A E A I P T A V P Y R I V N D M F P K F L Q N V L I F K T I I Q E K C P Q I I D E V E K E L S S Y L G K E K L A G I F T L E S F N K Y L G Q Q G K E N Q R G I D F Y N Q I I G G V V E K E G G I N L R G V N Q F L N L Y W Q Q H P D F T K E D R R I K M V P L Y K Q I L S D R S S L S F K I E S S E T L E C A D K L E K N D E K K S I F E E V C D L F S S V K N L D L S G I Y I N R K D I N S V R S I L T G D W S W L Q S R M N V Y A E E K F T K A E K A R W Q K S L D D E G E N K S K G F Y S L T D L N E V L E Y S E N V A E T D I R I T D Y F E H R C R Y Y V D K E T E M F V Q G S E L V A L S L Q E M C D D I L K K R K A M T V L E N L S S E N K L R E K T D D V A V I K E Y L D A V Q E L L H R I K P L K V N G V G D S T F Y S V Y D S I Y S A L E S V Y N V K T R N Y I T K K A A S P E K Y K L N F D N P T L A D G W D L N K E Q A N T S V I L R K D G M F Y G L I M M P K N K P K F A E K Y D C G N E S C Y E K M I Y K Q F D A T K Q I P K C S T Q K K E V Q K Y F L S G A T E P Y I L N D K K S F K S E L I I I T K D I W F M M N H V W D G E K F V P K R D N E T R P K K F Q I G Y F K Q T G D F D G Y K N A L S N W I S F C K N F L Q S Y L S A T V Y D Y N F K N S S E E Y E G L D E F Y N Y L N A T C Y K L N F I N I P E T E I N K M V S E G K L Y L F Q I Y N K D F A S G S T G M P N M H T L Y W K N L F S D E N L K N V C L K L N G E A E L P Y R A G I C K E P V Y K H E G S Y L V N R T T E D G E S I P E K I Y F E I Y K N A N G K L E K L S D E A Q N Y I S N H E V I I K K A G H E I I K D R H Y T E P E K F L F V P L T I N F K A S G N S Y S I N E V R K F L K N N P D V N I I G L D R G E R H L I Y L S I I N Q K G E I I K Q F T F N E V E R N K G R T I K V N Y H E K L D Q R E K E R D A A R K S W Q A I G K I A E L K E G Y L S A V I H Q L T K L M V E Y N A V V M E D L N F G F K R G R F H V E K Q V Y Q K F E H I I D K S N Y L V F K D R G L N E P G G V I L N G Y Q I A Q F E S P Q R L G K Q S G M L F V V P A G Y T S K I D P K T G F V S M M N F K D L T N V H K K R D F F S K F D N I H Y D E A N G S F V F T F D Y K K F D G K A K E E M K L T K W S V Y S R D K R I V Y F A K T K S Y E D V L P T E K L Q K I F E S N G I D Y K S G N N I Q D S V M A I G A D L K E G A K P S K E I S D F W D G L L S N F K L I I L Q M R N S N A R T G E D Y I I S P V M A D D G T F F D S R E E F K K G E D A K L P L D A D A N G A Y H I A L K G L S L I N K I N L S K D E E L K K F D M K I S N A D W F K F A Q E K N Y A K *
SEQ ID: 9	M E N Y G G F T G L Y P L Q K T L K F E L R P Q G R T M E H L V S S N F F E E D R D R A E K Y K I V K K V I D N Y H K D F I N E C L S K R S F D W T P L M K T S E K Y Y A S K E K N G K K Q D L D Q K I I P T I E N L S E K D R K E L E L B Q K R M R K E I V S V F K E D K R F K Y L F S E K L F S I L L K D E Y S K E K L T E K I I L A L K S F N K F S G Y F I G L H K N R A N F Y S E G D E S T A I A Y R I V N E N F P K F L S N L K Y K R E V C E K Y P E I I Q D A E Q S L A G L N I K M D D I F P M E N F N K V M T Q D G I D L Y N L A I G G K A Q A L G E K Q K G L N E F L N E V N Q S Y K K G N D R I R M T P L F Q I L S E R T S Y S Y I L D A F D D N S Q L S I T S I N G F T E V E K D K E G N T F D R A V G L I A S Y M K Y D L S R V Y I R K A D L N K V S M E I F G S W E R L G L L R I F K S E L Y G D V N A E K T S K K V D K W L N S G E F S I L S D V I N A I A G S K A E T F D E Y I L K M R V A R G E I D N A L E K I K C I N G N F S E D E N S K M I I K A I L D S V R L F H L F S S F Q R A D F S Q D G D F Y A E Y N E I Y E K L F A I V P L Y N R V R N Y L T K N N L S M K K I K L N F K N P A L A N G W D L N K E Y D N T A V I F L R E G K Y Y L G I M N P S K K K N I K F E E G S C T G P F Y K K M A Y K L L P D P N K M L P K V F F A K K N I N Y N P S D E I V K G Y Y A G K Y K K G E N F D I D F C H K L I D F F K E S I Q K N E D W R A F N Y L P S A T E S Y K D I S D F Y S E V E D Q G Y R M F L N V P V A N I D E Y V E K G D L F L F Q I Y N K D F A S G A K G N K D M H I I Y W N A A F S D E N L R N V V K L N G E A E L F Y R D K S I I E P I C H K K G E M L V N R T C F D K T P V P D K I I H K E L F D Y H N G R A K T L S I E A K G Y L D R V G V F Q A S Y E I I K D R R Y S E N K M Y F H V P L K L N F K A D G K K N L I N K M V I F K F L S D K D V H I I G I D R G E R N L L Y Y S V I D R R G N I I D Q D S L N I I D G F D Y Q K K L G Q R E I E R R E A R Q S W N S I G K I K D L K E G Y L S K A V H K V S K M V L E Y N A I V V L E D I N F G F K R G R F K V E K Q V Y Q K F E K M L I D K L N Y L V F K E V L D S R D A G G V L N A Y Q L T T Q L E S F N K L G Q S G I L F Y V P A A Y T S K I D P T T G F V S L E N T S R I E S D E K K D F L S G F D S I V Y S A K D G G I F P A F K D Y R N R N F Q R E K T D H K N I W T V Y T N G D R I K Y K G R M K G Y E I T S P T K R I K D V L S S G S I R Y D D G Q E L R D S I I Q S G N K V L I N E V Y N S F I D T L Q M R N S D G E Q D Y I I S P V K N R N G E F F R T D P D R R E L P V D A D A N G A Y H I A L R G E L L M Q K I A E D F D P K S D K F T M P K M E H K D W F E F M Q T R G D *
SEQ ID: 10	M E V Q K T V M K T L S L R I L R P L Y S Q E I E K E I K E E K R R K Q A G G T G E L D G G F Y K K L E K K H S E M F S F D R L N L L N L Q R E I A K V N H A I S E L Y I T A I Q C N K S N K H Y I S I I V Y N R A Y G Y F Y N A Y I A L G I C S K V E A N F R S N E L L T Q Q S A L P T A K S D N F P I V L H K Q K G A E G E D G G F R I S T E G S D L I F E I P I P F Y E Y N G E N R K E P Y K W V K K G G Q P V L K L I L S T F R R Q R N K G W A K D E G T D A E I R K V T E G K Y Q V S Q I E I N R G K K L G E H Q K W F A N F S I E Q P I Y E R K P N R S I V G G L D V I R S P L V C A I N N S F

TABLE 1-continued

Amino Acid Sequences of Exemplary CRISPR/Cas effector proteins	
SEQ ID NO:	Programmable Nuclease Amino Acid Sequence
	SRSYVSDNDVFVFSKQVFAPRRRLSLKNSLKRKGHGAHKLEPITEMTEKNDKFRKKIIE RWAKEVTNFFVKVNQGVIVQIBDLSTMKDREDHFFNQYLRGFWPYQMQTLIENKLKEY GIEVKRVQAKYTSQLCSCPNCNPNCRYWNNTNFNEYRKVNKFPKCEKCNLEISADYNAAR NLSTPDIEKFVAKATKGINLPEK*
SEQ ID: 11	MIIHNCYIGGSFMMKIDSFTNCYSLSKTLRFKLPIGATOSNFDLNKMLDEDKKRAENYS KAWSIIDKYHRFFIDKVLSVTENKAFDSFLLEDVRAYAELYRSNKKDSDKASMKTLES KMRKFIALQSDEGFKDLFGONLIIKKTLPFLESSTDKEIIAEFDGFSTYFTGFFNNRKN MYSADDPPVFLSQKGIEAYNSILGGYTNSDGSKIKGLNEYINLYNQKNEIHRIPKMKQL VPNVVDYPPVFLSQKGIEAYNSILGGYTNSDGSKIKGLNEYINLYNQKNEIHRIPKMKQL FKQILSERESVSFPEKFDSSDVLSINDYYLERDGKGVLSEIETKVEKIEKLFSAVTDYST DGIFVKNAAELTAVCSGAFGYWGTGVQNAWNNEYDALNGYKETEKYIDKRKKAYKSIE SFSLADIQKYADVSSESSETNAEVTEWLRLNEVQKDFLPEFLESSTDKEIIAEFDGFSTYFTGFFNNRKN NNDNAVELIKNALDVSKELENVLRLLLGTGEESKDENFYGEFLPCYERICEVDPDSLYDK VRNYMTQKLYKTDKIKLNQNPQFLGGWDRNKEADYSAVLLRRNSLYYIAIMPSGYKR VPEKIPAKADETVYEKVIYKLLPGPNKMLPKVFFSKKGIETFNPKEILEKYLELGTHKTG DGFNLDDCHALIDYFQSKALDVHSWDNSFGFRFSDTSTYKNIADFYNEVKNQGYKITFC VPQSYINELVDEGKLYLFQLYNKDFSEHSKGTPLNLHTLYFKMLFDERNLLENVVFNLNGE AEMFYREASISKDDMIVHPKNQPIKNNQEONSRKQSTFEYDIVKDRRTVDQFMLHIPIT LNFTAGCTNNINNEVRKALKDCDKNYVIGIDRGERNLLYICVVDSERGRIIEQYSLNEIINE YNGNTYSTDYHALLDKKEKERLESRKAWKTVENIKELKEGYISQVVKHICELVEKYDA V1VMEDLNLGFKQGRSGKFEKSVYQKEKMLIDKLNLYFADKKSPPEIGSVLNAYQLTN AFESFEKMGKQNGFIPYVPAYLTSKIDPTTGFDALLHPSSKQSKESMRDFVGRFDSITPN KTE NYEEFELDYNKFPRCTNDTQRKKWTVCTYGSRIKTFRNPEKNSEWDNKTVELTPAF MALFEKYSIDVNGDIKAQIMSVDKDFFVELIGLRLTLQMRNSETGKVDRDYLISPVK NSEGVFVNSDDYKGIENASLPKADANGAYNIARKGLWIEQIKACENDAELNKIRLAIS NAEWLEYAQKK*
SEQ ID: 12	MKDYIRKTLSSLRILRPYYGEEIEKEIAAAKKKSQAEGGDGALDNKFWRDLKAHEPEIIS REFYDLLDAIQRETTLYYNRAISKLYHSLIVEREQVSTAKALSAGPYHEPREKFNAYISLG LREKIQSNFRRKELAYRQVALPTAKSDTFPIPIYKGFDKNGKGGFVKREIENGDFVIDLPL MAYHRVGGKAGREYIELDRPPAVLNVPIVILSTSRRRANKTWRDEGTDAEIRRVMAGE YKVSUVEILQRKRGKPYGGWYVNFTIKYOPRDYGLDPKVKGDIGLSSPLVCATNS LARLTIRDNDLVAFNRKAMARRTLLRQNRYKRSGHGSANKLKPIEALTEKNELYRKAI MRRWARAEADFFRQHRAATVNMEDLTGIKDREDYFSQMLRCYWNYSQQLTMLENKL KEYGIAVKYIEPKDTSTKTCGHSCGHVNEYFDNFYRSAHKFPMFKCEKGCGVEGADYNA RNIAQA
SEQ ID: 13	MKEQFINRYPLSKTLRFLSLIPVGETENNFKNKNLLKKDKQRAENEYEVKCYIDRFHKEYI ESVLSKARIEKVNEYANLYWKSNSKDDSIKAMESLENDMRKQISKQLTSTEIYKKRLFG KELI CEDLPSFLTDKRETVECFRSFTTYFKGFTNTRENMYSSDGKSTAIAYRCINDNLP RFLDNVKSFQKVDFNLSDETI TKLNQDFNLSDETI FGRNIEDIFSVDYFEPVLTQSGIEIYNSMIC GYTCSDKTKI QGLNECINLYNQQVAKNEKS KKLPLMKPLYQKILSEKDSVSFPEKFN NEVLHAIDDYTYGHI GDFDLLTELLQSLNTYNANGI FVFKSGVAITDISNGAFNSWNVLRS AWNEKYEALHPVTSKTKIDKYIEKQDKIYKAIFSKSFLFELQSLGNENGNEITDWYISSINE SNSKIKEAYLQAQKLLNSDYEKSYNKRLYKNEKATELVKNLLDAIKEFQKLIKPLNGTG KEENKDELFYGKFTSYDSIADI DRLYDKVRNYITQCKPYSKDKIKLNFDPNQLLGWDK NKESEDYRTVLLHKDGLYLYLAVMDKSHSKAFVDAPEIT SDDKDYYEKM EYKLLPGPNK MLPKVFFASKNIDTFCPSDRLDIRKRESFKKGATF NKAKBCHFIDYF KDSI KKHDDWSQ PGFKFSPTESNDISEFYREISDQGYSVRFN KISK NYI DGLVNNGYI YLFQIYKNKDFSKYS KGTPNLLHLYFKMLFDERNLNSVQVYKLNGEAEMFYREASIGDKEKITHYANQPIKNNP DNEKKEVSEFYD IVKDKRFTKQFSLHLPITINFKAHGQEF LNYDVRKAVKYKDDNYVI GIDRGERNLIIYISVINSGNEI VEQMSLNEI ISDNGHKKV DYQKLLDTKEERDKARKNWTS VENI KEKLKEGYISQVVKICELVVKYDAVIAMEDLNF GFKRGRPFVKEQVYQKFENMLIS KLNLLIDKKAEPTE DGGLLRAYQLTNKFDGVNKAQNGI IFYVPAWDTSKIDPATGFVN LLKPKCMTSVPACKLFTETIDIKYANTDMFEFYIDYSKFP RCNSDFKKS WTVC TNSSR ILTFRKEKNNKWDNQKIVL TDEFKSLFNEPGIDYKGNLKDSI LSISNADFYRRLIKLLSL TLQMRNSITGSTGLPPEDDYLISP VANKSGE FYD SRS RNY GTNA ALPC DADANGAYNIARKA LWAINVLKDT PDDMLNKA KLSITNAEWLEYTQK*
SEQ ID: 14	MKEQFVNQYQPIKTLRFLSLIPVGETENNFKNKNLLKKDKQRAENEYEVKCYIDRFHKEYI ETALCNINFEGYEESSLYKCSKDDNDLKT MEDIEIKLRKQISKMTSHKLYKDLFGEN MIKTLNPLFSDDEEKSLSLEMPRGFYT YFSGFNTNRK NMYTEEAKSTSIA YRCI NDNLPK FL DNSKSFEKIKCALNKEELKAKN EEEFYEI FQIYATDIFNIDFFNFVLTQPGIDK YNGI IGG YTCSDGTVQQLNEI INLYNQ QIAKDDKS KRLPLK MLYKQILSD RETV SFPIPEKFSSDNE VLESIN MYFNSKVN VSNAIKSLKELFQGFEAYNMNGI FISSGVAITDLSNAVFGDWMAISTA WEKAYFETNPPKKNQSKQ EYEEELKANYKKIKSFSLDEI QRLG SIAKSPD SIGSVAE YYKI TVTEKIDNITEYDGSKE LLNCNYSESYD KKIKLNDT VIBKVKTLLDAVSKLEK I KPLV GTGKEDKDEL FVGTFLPLY TSLSA VDR LYDKVRN YATQK PYSKDKIKLNFNCSSFLSGW ATDYSSNGGLIFEKDG LLYLGIVVNKKFTTEEIDYLQ QNADENPAQ RIVYDFQKPDN KNT PRLFIRSKGTNSPSVKEYNLPV EEEV I VELYD KR YFT TEYRN KNPELYKASLVK LIDYF K GPTRHESYRHDFWKKSEEY NDISEFYKDVEIS CYSLKQEKINY NTLLNFVAEN RIYLF

TABLE 1-continued

Amino Acid Sequences of Exemplary CRISPR/Cas effector proteins	
SEQ ID NO:	Programmable Nuclease Amino Acid Sequence
	QI YNKDFSKYSKGTPNLHTRYFKALFDENNLSDVVFKLNGGSEMPFRKASIKDNEKVH PANQPIDNKNPDNSKKQSTFDYELIKDKRPTKHQFSIHIPITMNFKARGRDFINNDIIRKAI KSEYKPVIGIDRGERNLIVIYISVINNNGEIVEQMSLNDIISDNGYKVDYQRLLDRKEKERD NARKSWGTIENIKELKEGYISQVIHKICELVIKYDAVIAMEDLNFGFKRGRFVNVEKQVYQ KFENMLISKLNLYCDKKSEANSEGGLKAYQLTNKFDGVNKGKQNGIIFYVPWPLTSKI DPVTGFV DLLHPKIVSVEETHSLFEKLDDIRYNEFKDMFEPDIDYSKLPKCNADFKQKW TVCTNADRIMTFRNSEKNNEDNKRILLSDEFKRLFEEFGIDYCHNLKNIKLSISNKOFC YRFIKLFLALTQMQRNSITGSTNPEDDYLISPVRDENGVFYDSRNFIGSKAGLPIDADANG AYNIARKGLWAINAIKSTADDMLDKVDSLISNAKWLLEYVQK*
SEQ ID : 15	MKITKIDGILHKKYIKEGKLVKSTS EENKTDERLSELLTIRLDTYIKNP DNA SEEENRIRRE TLKEFFSNKVLYLKDSDI LYLDKDRREKNQLQNK NYSEEDISEYDLKKNNSFLVLKKILLNE DINSEELEIFRNDFEKKLDKINSLKYSLEENKANYQKINENNKKVEGSKRNI FYNYYK DSAKRNDYINNIQEAFD KLYKKEDIENLFFLIENSKKHEKYKIRECYHIIIGRKNDKENF ATIYEEI QVNVMKNEI KEVPNVSBLKKSQVYKYLKEKLNDENIKYVFCHFVEIEM SKLLKNVYVKKPSNISNDKVKRNLKMF SYDFNMNRKKEIEDFFSNIDEAISIRHGIVHFNLE ATSDFIVGNRQNEAFLRNIIIGVS STAYFSLRNILETENENDITGRIKGKTVKNNKGEKYIS GEIDKLYDNNKQNEVKKNLKMF SYDFNMNRKKEIEDFFSNIDEAISIRHGIVHFNLE EGKDIFTKNI VPSQISKKMFQNEINEKKLKLKIFRQLNSANVFRYLEKYKILNYLNRTRF EFVNKNIPFVPSFTKLYSRIDDLKNSLCIYWKIPKANDNNKTKETIDAOIYLLKNIYGEF LNYFMSNNGNPFEEI KELNNDKRNLLKTFGFKLQKFENLQEKTPKEYLANIQSFYMID AGNKDEEKD DAYIDFIQKIFLKGFM TYLANNGRSLSMYI GND EQINTSLAGKKQEFDFK LKKYEQNNNIEIPHEINEFVREIKLGKILKYTESLNMFYLILKLLNHKELTNLKGSLEKYQ SANKEEAFSDQLELINLNLDNNRVTEDFELEADEIGKFLDFNGNKGKDNKELKKFDTN KIVFDGENIIKHFYNIKKGMLNLKESDEAKYKISIEELKNYSNKKIEIEKNHTTQEN LHRKYARPRKDEKFLNDEKYYKEKTIRN IQQYTHLKNKVEFNE LNLQSLLLRLIHLRV GTYSIWERDLRFLRGFEPENQYIEEI IFNFDNSKVN KYKNGQIVEKYI SFYKELYKDDME KISIYSDKKVKELKKEKKDLYIRNYIAHFNYIPNAEVSLLEVLENRLKLLSYDRKLKNAI MKSIVDILKEYGFVUTFKIEKDKKIRIESLKEEVVHLLKLLKDNDKKPEIKTYRNSK ELCKLVKVMFEYKMKKEKSEN*
SEQ ID : 16	MKITKIDGISHKKYIKEGKLVKSTS EENKTDERLSELLTIRLDTYIKNP DNA SEEENRIRRE NLKEFFSNKVLYLKDSDI LYLDKDRREKNQLQNK NYSEEDISEYDLKKNNSFLVLKKILLNE EDINSEELEIFRKDVEAKLNKINSLKYSFEENKANYQKINENNVEKGGKSKRNIIYDYY RESAKRNDYINNIQEAFD KLYKKEDIENLFFLIENSKKHEKYKIRECYHIIIGRKNDKENF PAKIIYEEI QVNVMKNEI KEVPNVSBLKKSQVYKYLKEELNDKNIKYAFCHFVEIEM MSQLLKNVYVKKPSNISNDKVKRNLKMF SYDFNMNRKKEIEDFFSNIDEAISIRHGIVH EYKIPGVEDQIYYENQNEVKKNLKMF YGDMDM NKKEIEDFFSNIDEAISIRHGIVH FNLDLGD KDI PAFKNIPVPSI SKKMFQNEINEKKLKLKIFRQLNSANVFRYLEKYKILNYL KRTREEFVNKNIPFVPSFTKLYSRIDDLKNSLCIYWKPTKTNDDNKTKEIDAOIYLLKNI YYGEFLNYFMSNNGNPFEEI KELNNDKRNLLKTFGFKLQKFENLQEKTPKEYLANIQ SLYMINAGNQDEEKD DAYIDFIQKIFLKGFM TYLANNGRSLSMYI GND EQINTSLAGKK QEFDKFLKYYEQNNNIEIPHEINEFVREIKLGKILKYTESLNMFYLILKLLNHKELTNLK SLEKYQSANKEETFSDELEI INLNLDNNRVTEDFELEADEIGKFLDFNGNKGKDNKELKK KFDTKKIYFDGENIIKHFYNIKKGMLNLKESDEAKYKISIEELKNYSNKKIEIEKNHTTQEN NYTMQQNLHRKYARPKDEKFLNDEKYYKEKTIRN IQQYTHLKNKVEFNE LNLQSLLLQGL LLKILHRLVGYTSIWERDLRFLRGFEPENQYIEEI IFNFDNSKVN KYKNGQIVEKYI INFYK ELYKDNVEKRSIYSDKKVKKLQKEKKDLYIRNYIAHFNYIPHAEISLEVLENRLKLLSY DRKLKNAIMSVVDLILKEYGPVATFKIGADKKIGIQTLESEKIVHLKLNKLLKLM TDRN SEELCKLVKVMFEYKMEEKNLKTKKCKV*
SEQ ID : 17	MKKIDNFGVCYPSKTLRFKAIPIGKTQENIEKKRLVEEDEVRAKDYKAVKKLIDRYR EFIEGVLDNVKLDGLEEYYMLFNKS DREESDNKKIEIMEERFRRVISKSFKNNEEYKIFS KKIII EELIPNYIKDEEKKELVKGFKGFYTA FVGVAQNREN MYSDEKKSTAISYRIVNENM PRFITNIKVF EKAKSILDV DKINEI NEYI LNNDYVWDF FNIDFFN VLNQKGIDIYNAIIG GIVTGDRKIQGLNECIN QMNCNKKRKL P QFKPLYKQI LSESESMSFYIDEIESDDMLID MLKESLQIDSTINNAIDL KVL FNNI FDYD LSGI FIN GLPITT ISNDVYQWSTISDGWNE RYDVLSNAK DKESEK YFEKRRKEYKKVKSFSISDLQELGGKDLSICKKINEII SEMIDDYK SKIEEIQYLF DIKELEKPLVTDLNKIELIKNSLDGLKRIERYVIFL LGT GKEQNR DEVFGY FIKCIDAIEIDGVYVKNTRNYLT KKP YSKD KFKLYF ENPQ LMG WDRN KESDY RSTL R KNGKYYVAIIDKSSSNCMNIEEDENDNYE KINYKLLPGPNKMLPKVFFSKKNREYFAP SKEI ER IYSTGT FKKD TNFV KKD CENL ITFYK DSD LDR HEDWSK SFDSF KESSAY RDISEF YRDVEKQGYRVSFDL LS SNAVNTLVEEGKLYLFQ LYN KDFSEKSHGIPNLH TMYF RSLF DDNNKGNIIRLNGGAEMFMR RASLNKQDVTVHKANQPIK NKNL LNPKTTTLPYDVYK DKRFTEDQYEVHIPITMNKVPNNPYKINHMVR EQLV KDDNPYVIGIDRGERNLIYVVV DGQGHIVEQLSLNEIINENNGISIRT DYTLL DAKER DESRKQWQIENIKELKEGYIS QVVKHICELVEKYDAVIALEDLNSGFKNSRVKVEKQVYQKFEKMLITKLN YMVDKKK

TABLE 1-continued

Amino Acid Sequences of Exemplary CRISPR/Cas effector proteins	
SEQ ID NO:	Programmable Nuclease Amino Acid Sequence
	DYNKPGGVNLNGYQLTQFESFSKMGTQNGIMFYIPAWLTSKMDPTTGFV DLLPKKYKN KADAQKFQSQFD SIRYDQNQEDAFVFVKVNYTKFPRTDADYNEKEWEIYTNGERIRVFRNP KNNEYDYETVNNSERMKELFDSYDLLYDGKELKETI CEMEESKFFEE LKLFRRTLQMR NSISGR TDVDYLISPVKNSNGYFYNNSNDYKKEGAKYPKDADANGAYNIARKVLWAIEQ FKMADEDKLDTKISIKNQEWLEYAQTHCE
SEQ ID: 18	MKKIDSFVNYYPLSKTLRFLS LIPVGKTEDNFNAKLLLEDEDEKRAIEYEVKRYIDRYKH FIETVLANFHLD DVNEYXAE LYKKAGKDDKDLKYM EKLEGKMRKSISA AFKTDKKYKEI FGQEII KNLIPDNEESV KMFQGFTYFTGFDNDRNKRNM YTHEA QTTAISYRCINEN LPKFLDNVQSF AKI KESISSD IMNKLD EVCMDLYGVYA QDMFCTD YFSFVLSQSGIDRY NNIIGGGYVDDKGKVQI QG INYEINLYNQV D EKNKRLPLM KKLYKQ ILEKESIS FPIKEF SDNIVINA ISDYYHNNV ENLDDFNKL FNEFSE YDDNGI FVTSGLA VTD ISNA VFG SWNII SDSWNEEYKD SHPMKKT TNAEK YKEDM KKEY KKNL SFTIAELQ RLG EAGC NCDE CKGD KEYYKTTVAE KIENIKNAYEISK DLLASDYEKSND KKLCKN DSAISL LK NLLD SI K DLEK TIKP LLTGKEEN KDDV FYGKFTNLYEMISEID RLYDKVR NYVTQK PYSK D KI KLN FENP OHLGGWDK NKR D YSVL LKED K YL AIM DKSNNKA FIDFP DPDGE CYE KI EY KLL PG PNKMLPKV FFAS NSNIEYFAPS K KFLL EIR SRES FKKG DMF NLKDCHEFIDFFKESIKKHEDW SQFGFESPS TKE KYNDI S E FYNEV KI QG YSLK YKVN SKY K V D E L IEC GQOL YLF QI YN KDF SV YAKGNP NLH TMY FKML FDERN LANV VYQLNGGA EMF YRKASIKD SEK IVHHAN QPIK NKNADNVK KESV FEYDI IKDKRFTK RQFSI H IPI TLNFKA GQNF INND VRM ALK ADEN VYIGIDRGERNLLYI CVINSK GEI VE BOKS LNE II GDNGYRV DYH KLL D KKEA ERDE ARKS WGTIENIK ELK EGYLSQ I V BISL KV I K YDAVIA E DLS NGFK GRPF KVEK QVY QKF ENM LCTKLN YLV DKNADANE CGGLL KAY QLT NKEDG ANGR RQNGI I FSPV PAW LT SKID PVT GFADLLRPK YKVS ESEV FIS KIDN IRYNS KEDYF EFD IDY SKFP N ST ASYKKW T V CTY GERI INVRN KEK NN MWD NK TIV LT DE FKKL FAD FG VD GS F YD S RL V KEK TLP ENA DANGAY NIARK LLANTLQLRN SE VGNV D VDYL ISP VKG VD GS F YD S RL V KEK TLP ENA DANGAY NIARK ALWAIDV LQTK D EELK NANL SIKNA E WLEYVQK*
SEQ ID: 19	MKNQNTLPSNPTDILKDPFWA AFFN LARHN VYLT VN HINK LLD L EKLYN KDKH KIE F HED IFN ISDD VM ND VNS NGK KR KLD I KKI WAN L D LTR K YQL R E L I L K H P F I Q P A I G A QTKERT I D K DK RST S TS ND S L K P T G E G D I D P L S L S N V K S I F F R L L Q M L E Q L R N Y Y S H V KHSK SAT MPN F D E G L L K S M Y N I F D S V N K V K E D Y S S N S V I D P N T S F S H L I S K D E Q G E I K P C RYSFTS K DGS I N A S G L L F F V S L F L E K Q D S I W M Q K K I P G F K K T S E N Y M K M T N E V F C R N H I L LPKMR L E T V Y D K D W M L L D M L N E V V R C P L S L Y K R L A P D Q N K F K V P E K S S D N A N R Q E D DNP F S R I L V R H Q N R F P Y F A L R F D L N E V F T T L R F Q I N L G C Y H F A I C K K Q I G D K K E V H H L T R T L Y F R S R L Q F T Q N T R P E E W N T L V K T T E P S S G N D G K T V Q G V P L Y I S Y T I P H Y I E N E K I G K I F D G D T A V D T D I W P S V T E K Q L N K P D K Y L T P G F K A D V F L S V H E L L P M M F Y Q L L C E G M L K T D A G N A V E K V L I D T R N A I F N L Y D A F V Q E K I N T I T D L E N Y L Q D K P I L I G H L P K Q M I D L L K G H Q R D M L K A V E Q K K A M L I K D T E R R E L R N K Q P E Q K P N V A A K N T G T L L R N Q O I A D W L V K D M M R F Q P V K R D K E G N P I N C S K A N S T E Y Q M L Q R A F A Y T T D S Y R L P R Y F E Q L H L I N C D N S H L F L S R F E Y D K Q P N L I A F Y A A Y L E A K L E F L N E L Q P Q N W A S D N Y F L L R A P K N D R Q K L A E B G W K N G F N L P R G L F T E K I K T W N E H K T I V D I S C D I F K N R V G Q V A R L I P V F D K K P K D H S Q P F Y T Y N F V N G V N S K I T E A N Y L S K E K R E N L F K S Y Q N K F K N N I P A E K T K E Y R E Y K N F S S W K K F E R E L R L I K N Q D I L T W L M C N L F D E K I K P K K D I L E P R I A V S Y I K L D S L Q T N T S T A G S L N A L A K V V P M T L A I H D S P K P K G K A G N N E K E N K E F T V Y I K E E G T K L L K W G N F K T L L A D R R I K G L F S Y I E H D I N L E K Y P L T K Y Q V D S E L D L Y Q K Y R I D I F K Q T L L E A Q L L D K Y S D L N T D N F N Q M L S G W S E K E G I P R N I K Q D V A F L I G V R N G F S H N Q Y P D S K R I A F S R I K K E N P K T S S L Q E S E G L N I A K Q M Y E E A Q Q V N K I K N I E S F D *
SEQ ID: 20	MKVTKIDGISHKKF EDEGKLV KFTGH FN KIKN E M KER L K L K E L K L K L S N Y I K N P E N V K N K D K N K E K E T K S R R E N L K K Y F S E I I L R K K E E K Y L L K K T R K F K N I T E E I N Y D D I K K R E N Q Q K I F D V L K E L L E Q R I N E N D K E E T I L N F D S V K L K E A F E E D F I K K E L K I K A I E E S L E K N R A D Y R K D V E L E N E K Y E D V K G Q N D F K V N K P E N R E K F K E N I K Y A F E N L Y T E E N I K N L Y S E I K E I F E K V H L K S K V R Y F Y Q N E I I G E S E F S E K D E E G I S I L Y Q I O I I N S V E K K F I E F L Q K V K I K D L T R S Q I F Y K F L E N E L N D E N I K Y V F S Y F V E I E V N K L L K E N V Y K T K K F N E G N K Y R V K N I F N Y D K L K N L V Y Y K L E N K L N N V R N C G K Y N H M E N G D I A T S D I N K N R Q T E A F L R S I L G V S S F G Y F S L R M I L G V N D D F Y K I E K D E R K N E N F L K K A K E D F T S K N I F E K V D K S F E K K G I Y Q I K E N L K M F Y G N S F D K V D K K F F V N M L E A I T S V R H I V H Y N I N T S E N I F D F S N I E V S K L L K N I F K E I D T R E L K L K I F R Q L N S A G V F D Y W E S V I K K Y L E N V K F E F V N K N V P F V P S F K K L Y D R I D N L K G W N A L K G L N N I I P K R K E A K D S Q I Y L L K N I Y Y G E F V E K F V N D N K N F E K I V K E I I E I N R G A G T N K K T G F Y K L E K F E T L K A N T P T K Y L E K L Q S L H K I S Y D K E K I E B D K D V Y V D F V Q K I F L K G F V N Y L K K L D S L K S L N L N L R K D E T I T D K K S V H D E K L K L W E N S G S N L S K M P E E I Y E V V K K I K I S N I N Y N D R M S I F Y L L L K L I D Y R E L T N L R G N L E K Y E S M N K N K I Y S E E L T I I N L V N D N N K V R T N F S L E A E D I G K F L K S S I T I K N I A Q L N N F S K I F A D G E N V I K H R S F Y N I K K Y G I L D L L E K I V A K A D L K I T K E E I K K Y E N L Q N E L K R N D F Y K I Q E Q I H R N Y N Q K P F S I K K I E N K K D F E K Y K V V I E K I Q D Y T Q L K N K I E F N D L N L L Q S L I F R I L H R L A G Y T S L W E R D L Q F K L K G E F P E D K Y I D E I F N S D G N N N Q K Y K H G G I A D K Y A N F L E K K E E K S G E I L N K K Q R K K K I K E D L E I R N Y I A H F N Y L P N A E K S I L E I L E E L R E L L K H D R K L K N A V M K S I K D I F R E Y G F I V E F T I S H T K N G K K I K V C S V K S E K I K H L K N N E L I T T R N S E D L C E L V K I M L E H K E L Q K *

TABLE 1-continued

Amino Acid Sequences of Exemplary CRISPR/Cas effector proteins	
SEQ ID NO:	Programmable Nuclease Amino Acid Sequence
SEQ ID: 21	MKVTKVGGISHKKYTSEGRLVKSEEEENRTDERLSALLNMRDLYIKNPSSSTETKENQK RIGKLKKFFSNKVMVYLLKDNTLSSLKNGKKENIDREYSETDILESDVRDKKNFAVLKKIYL NENVNSELEEVFRNDIACKKLNKINSLKYSFEKNNKANYQKINENNIEKVEGSKRNIIYDY YRESAKRDAYSVNSKEAFNLKAKLVEIENLTKEKYKIREFYHEIIGRKNDKE NPAKIIYEEIQNVNNMKELIEKVPDMSELKKSQVFYKYYLDKEELNDKNIKYAFCHFVEI EMSQQLKKNYVYKRLSNI SNDKIKRIFEYQNLKKLIEENKLLNKLDTYVRNCGKYNYYLQD GEIATSDFIARNRQNKEAFLRNIIGVSSVAYFSLRNI LETENENDITGRMRGKTVKNNKGBE KYVSGEVDKIYNEKMFYSYDFNMNDKNEI EDFFFANIDEAIISSIRHGVHF NLELEGKDIAFKNIAPISEISKKMFQNEINEEKKLKLKIFRQLNSANVFVYLEKYKILNLYLK RTRFEFVNKNIPFVPSFTKLYSRIDLKNSLGIYWKTPKTNDDNKTKEIIDAQIYLLKNIY YGEFLNPFMSNNGNFPEISKEIIELNKNNDKRNLTGCFYKLQKFEDIQEKIPKEYLANIQL YMINAGNQDEEECDTYIDFQKIFLKGFMFTYLANNGRLSLIYIGSDEETNTSLAEKKQEF DKFLKKEQNNNIKIPYEINEFLREIKLGNILKYTERLNMFYLILKLNHKEELTNLKGSL KYQSANKEEAFSDQLELINLLNLDNNRVTEDEFELEADEIEGKFLDPNGNKVKDNKELKKF DTNKIYDGENI IKHRAFYNIKKYGMNLLEKIAKDAGYKISIEELKYSNKKNEIEKNH KMQEENLHRKYAPRKDEKFTDDEDYESYKQAIENIEEYTHLKNKVEFNEELNLLQGLLLRI LHRLVGYTSIWERDLRFLRGKGFPEHQYIEEIFNFENKKNVKYKGQGIVEKYIKFYKELH QNDEVKINKYSSANI KVLQKEKKDLYIRNYIAHFNYIPHAEISLLEVLENLRKLLSYDRK LKNAVMSVVDILKEYGVATFKIGADKKIGIQTLESEKIVHLKLNKKKLMTDRNSEE LCKLVKIMFEYKMEEEKSEN
SEQ ID: 22	MNELVKNCRKQTCTCQKLIPIGKTRETIEKYNLMEIDRKIAANKELMNKLFSLIAGKHI NDTLSKCTDLDPEPLTSLSSLNNAKENDRDLNLRREYDSVPEEKKTLAEEISSRLTAVKF AGKDFFTKNI PDPFLETYEGDDKNEMSELVSLVIENTVTAGYVKKLEKIDRSMEYRLVSG TVVVRVLTNDADIYEKNIKEAKDFDVGVLNIDEASQFTTTLVAKDYANYLTDGIAIYNV GIGKINLALNEYCQKNKEYSYNKLALLPLQKMLYGEKLSLFEKLEDFTSDEELINSYKF AKTVNESGLAEI IKA VPSYDEIVIKPNKISNYNSNSITGHWSLVNRMKDYLENNINGIKNAD KYMKGILTSEIGDALENNKIKHSDFISNLINDLGHYTBIKENKESLKKDESVMALIICK ELDMLLSILQNLVFDIDNEMFDTGFGIEVSKAIEILGYZGPPLYNKIRNYITTKKPDPKKF MTKFGSATIGTGTTSVEGSKKATFLKGDAVFLLLYNTAGCKANNVSVSNLADLISSN EIENSGKCYQKMIYQTPGDIKKQIPRPFVYKSEDDDLIKDFKAGLHKTDLSFLNGLRIPY LKEAFATHETYKNTYFSYRNSYESYDCEHMSEQAYILEWKWIDKKLIDDLVEDGSSL MFRVWNPMKMKKPSDENKISKHAKIVNELSFSDENAAIKLSSVFDIIFYRDQKQIDNPIVHK AGTTLYNKRTKDGEIVDVTMVKNKEKRPNVYTTTKYDIIKDRYTEBQFETHLHVN IGKEENKEKLETSKVINEKKNTLVVTRSNEHLLYVVIFDENDNILLKKS LNTVKGMNFKS KLEVVIEQKEMINMQS梧TVGSNQALMEGYLISFAIKEIADLVKEYDAI LVEQNSVGKNI LNERVYTRPKEMIITNLSDVDYENKDFYSTELEGKVASWRDCVTINGCICQVPSAYKY KDPTTSFSTI SMYAKTTAEKSKKLQIKSFKYNRERGLFELVIAKGVLGENNIVCDSGS RSI IENDISKEVSCSTLKI EKYLIDAGIEYNDKEV LKDLDTAAKTD AVHKAVTLLKCFNE SPDGRRYYISP CGEHTLCDAPEVLSAINYYI RSRYIREQIVEGVKKM EKKTILLAK*
SEQ ID: 23	MNGNRIIVYREFVGVTVAKTLRNLRLPIGHTQEHIIHNGL IQEDELROKSTELKNIMDD YYREYIDKSLSGVT DLDFTL FELMNVLQSSPSKDNK KALEKEQSKMREQICTHMQSDS NYKNIFNAFKLIEI LPDFIKNYNQYDADKAGKLET LALFNGFSTYFTDFTFEKRKNVFTK EAVSTSIA YRIVHENS LTFLANMTSYKKISEKALDEIEVIEKNNQDKMGDWELNQI FNPD FYNMVLIQSGIDFYNEICCGVVAHNMLYCQQT KNNYLFKMRKLHQQI LAYTSTSFEVP KMFEDDMSVYNAVNAFI DTEKEGN I GKLKD VNKYD EDELDEK I RYI SKDFYETLSCFMSG NWNLITGC VENFYDENI HAKGKSKEEKVKKAVKED KYKSINDVNDLVEKYIDEKERNE FKNSNAKQYIREISN I TDTETAHLEYDEHISL I SE EKA DEMKKR LDMMY MNHYWAKA FIVDEVLDRDEMFS DDDI YNI LENI VPYLN RVNRYT QPYNSKKI KLN FQSP TLANG WSQSKEFDNNAI I LIRD NKYYLA IFNAK NPKD KKII QGN SDK KND NDY KKM VY NLL PGA N KMLPKVFLSKKGK I EFTK PDSYI ISGVNAH KH KI TSEN FD I SF C RDL D I YF KNS I EK HAEW RKYEFKFSATD SYNDI E FYREVEN QGYR IDW TYI SEAD INKL DEEGK I YLFQ I YNK DPA ENSTGKENLHTM YPKN I FSEEN LKD II I KLN GQAE LFY RRA SVK NPV KHK DSV LVN K T YKNQLDNGDV VRIP I PDDI YNEI YK MNGY I KEND LSE AKEY LDK VEV R T A QKD I VKD YR YTVDFK YFIHTP ITI NTYKV TAR NN VND M A V KYI A QN DD I H VIG I D RGER NLI YI SVID SH GNI V KQKSY I LN NYN YK KKL VE K EK T RE A R K NW K S G N I K EL K E G Y I SG VV H E I A M L M V E Y N A I I A M E D L N Y G F K R G F P K V E R Q V Y Q K F E S M L I N K L N Y F A S K G K S V D E P G G L L K G Y Q L T V P D N I K N L G K Q C G V I F Y V P A A F T S K I D P S T G F I S A F N F K S I S T N A S R Q F F M Q D E I R Y C A B K D M F S F G F D Y N N F D T Y N I T M S K T Q W T V Y T N G E R L Q S E F N N A R R T G K T K S I N L T E T I K L L E D N E I N Y A D G H D V R I D M E K M D E D K N S E F F A Q O L L S L Y K L T V Q M R N S Y T E A E E Q E K G I S Y D K I I S P V I N D E G E F F D S D N Y K E S D D K E C K M P K D A D A N G Y C I A L K G L Y E V L K I KSEWTEDGFDRNCLKLPHAEWLDFIQNKRYE*
SEQ ID: 24	MNKDIRKNTDFVG I SEI QKTLRFILIPIGKTAQNI D KYNM FEDDEIRHEYYPILKEACDF YRNHIDQDFENLELDWSKLD E A S E D R D L I N E T R A T Y R Q V L F N R L K N S V D I K G D S K K N K T L S L E S S D K N L G K K K T K N T F Q Y N F D L F K A K L I K A I L P L Y I E Y I Y E G E K L E N A K K A L M Y N R F T S R L S N F W Q A R A N I F T D E I S T G S P Y R L V N D N F T I F R I N N S I Y T K N K P F I E E D I L E F E K K L K S K K I I K D F E S V D D Y F T V N A F N K L C T Q N G I D K Y N S I L G G F T T K E R E K V K G L N E F N L A Q Q S I N K G K G E Y R K N I R L G K L T K L K Q I L A I S D S T S F L I E Q I E D D Q D L Y N K I K D F F E L L K E E I E N E N I F T Q Y A N L Q K L I E Q A D L S K I Y I N A K H L N K I S H Q V T G K W D S L N K G I A L L E N I N I

TABLE 1-continued

Amino Acid Sequences of Exemplary CRISPR/Cas effector proteins	
SEQ ID NO:	Programmable Nuclease Amino Acid Sequence
	NEESKEKSEVISNGQTKDISSEAYKRYLQIQSEEKDIERLRTQIYFSLEDLEKALDLVLI NMDRSKDSKILSYVQSPLNVNFERDLTDLYSRIMKLEENNEKLLANHSADLIFLQLI MLRYSRWQILFCDSNYELDQTFYPIYDAVMEILSNIIRLYNLARNYLSRKPDRCMKKKIN FNNPNTLADGWSESKIPDNSSMLFIKDGMYYLGIIKNRAAYSELLEAESLQSSEKKKSENS SYERMNYHFLPDAFRSIKAMKAVKEHFEINQKTAIDLLDTCFSKPLRITKEIFDMQ YVDLHKNNKKYQVDYLRLDTGDKKGYRKALNTWLNFKCDFISKYGRNLFDYSKTAKA DHYETVNEFYNDVDKYSYHIFTSAETTVKEFKISEGKLYLFQQLYNKDFSPHSTGKPMLH TIYWRALFSEENLTSKNIKLNGQAEIFFRPKQIETPFTHKKGSLIVNRFDVNGNPIPINVYQ EIKGFKNNVWKWMDNKTQEGLENDQLYFESEFEIICDRRTYEDQLFFFHVPISENWDI GSNPKINDLATQYIVNSNDIHIIGIDRGENHLIYYSVIDLQGAIVEQGSINTITEYTNKFL NNKTNNLRKIPYKDLILQREDERADARIKWHAIDKIKDLKDGYLGQIVHFLAKLIIKYNA IVILEDNYGFGRGRFVKVERQVYQKFEMALMCKLNVLFKDYDIDEIGGPLKWPQLTRP IDSYERMGRQNGILFYVPAAYTSAVDPVTGFANFLYLNVNKSKEFKHFFSKFESIKYHSD QDMFSFADFDDNGFTTRINDLSKSQWQFTNHERSVWNNEKEKNVVTQNLTDLKLL RTYNEIEFKNNQNVLDSILKIENNTDKENFARELFRFLRFLTIQLRNNTVNENNTIELENLD YIISPVKDKGNMFDSRDELKNLPDNGDANGAYNIARKGLLYIEQLQESIKTGKLP TLSIS TLDWFNYIMK*
SEQ ID: 25	MNKGGYVIMEKMTKRNWENQFRITKTIKEEIPGTYTKVNLQRVNMLKREMERNEEDL KKMKCEI CDEYYRNMDIVSLRLEQVRLGWSLHCKYRMLNKEBEIKALEKEQEDLRK KISKGPGEKKAWTGEQFIKKILPQYLMHDHTYGELEEKLRIVKKPKGCTMFLSTFFKNRE NIFTDKP1HTAVGHRITSNAMLFAANTINYEKMESNTLEIERLQREFWRGINISEIFTD AYYVNVLUTQKQIEAYNKICGDINQHMNEYCQKQKLKFSEFRMRELKKQILAVVGEHFEI PEKIESTKEVYRELNNEYYESLKEHLHGQFEELKSVQLKYSQIYVQKKGYDRISRYIGGQW DLIQC ECKMDKASGKMGTKKHNDAKIEEEVAKVKYQSIIEHQKLVCTYEEDRRGHKVTD YVDEFIVSVCDLNGADHII TRDGERIELPLQYEPGT DLLKNDTINQRRLSDIKTILDWHD MLEWLKTFVLVNDLVIKDEEFYMAIEELNERMOCVVISVYNRIRNYVTQKGYEPEKIRICFD KG TILTGTWTGDNQYQSGFLMNRNDKYLYGIINTNEKSVRKILDGNGECKDENDYIRVG YHLINDASKQLPRIFVMPKAGKSEILMKDQCDYIWDGYCHNKHNESKBFMRELIDY YKRSIMYDKWEGFNCFSSTESYDNMQDFYKEVREQSYNISFSYINENVLEQLDKDG KIYLFQVNKDFAAGSTGTPNLHTMLQNLFSQNLLEKRLRLGGNAELFYRPCTEKDV THRKGSI LVDRTYVREEKDGEVRDTPVKEKEYLEIYRLNGQKGDLSESAKQWLKV HYREAPCDI IKDKRYAQEKYFLHFSEINPNAEGQTALNDNVRRLSSEBDIHIVIGIDRG ERNLIYVS LMDGPKQKSEONKDEKKKIDKPLETERGKLRAFTPIYEAEKGKLNKAGKEK IKDIKTYGLS YVHVHEI VEMAVERKAII VVMDLNYSFGRGRFKVERQVYQKFEENLINKL NYVVDKQLSVDEPGGLLRGYQLAIFI PDKKSSMRQNGIVFVYPAGYTSKIDPTTGVINIF KFPQFGKGD DNGKDGKDRKIRAFFGKFD ETRYECDEKVTAADNTREVKERYRFDFDYSK FETHLVHMKKTWTYVAEGERIKRKVGVNWTSEVISDIALRMSNTLNIAGIEYKDH NLVNEI CALRGKQAGI I LNELEI VRLTVQLRNNTTEGDVDERDEII SPVLTNEKYGCYHS TEYKQQNGDVLPKDADANGAYCIGLKGJYEIRQIKNWKEDMTKGEKGKALNEGMRISH DQWF EFi QNQNMKGE*
SEQ ID: 26	MNNP RGA FGGFTNLYSLSKTLRFELKPYLEIPEGEKGKLGPGDDKEYYKNCKTYTEYYLK KANKEYYDNEKV KNTDLQLVNFLHDERIEDAYQVLKPVFDLHEEFITDSLESAEAKI DGFNYGGYLYEOKQKSEONKDEKKKIDKPLETERGKLRAFTPIYEAEKGKLNKAGKEK KDKDILKESGFKVLI EAGILKYIKNNIDFADKKLKNNEGKEITKKDIDETALGAENIEGIFD GFTTYPSGFVNQNR ENVYSTEEKATAVASRIVDENLSKFCDNILLYRKNDYLKIFNFK NKGKDLKLKNSKFGKENEPEFIPAYDMKNDEKSFSVADFVNCLSQGEIEKYNKIANAN YLINLYQNKDGNSKLSMFKILYQIGCGEKKDIFTIKTDNAELKQILEKACEAGKKF IRGKSEDGGVSNIDQKSYNIVNSGNKEPVDTLQKLNKLGDAKV FNKNTGEDKADVKYKVPAQVMLS ELFAVFLDDNAGEDWREKGIFFFKASLFGEDQNKSEI IKNANRPSQALLKMI CDDMESLAKNFIDSGDKILKISDRDYQKDENQKQIKNWLNDNA WINQILKYFKVANKI KGDSIDARIDSGLDMVFSSDNPAEDYDMIRNYLTQKPODEINK LKLNFENS LLAGGWDEQKEDKDNSCIILKDEQDKQYLA VMKYENTKVFQKNSQIYIAD NAAWKMI YKLVPGASKTLPKVFFSKWWTA NRPTPSDIV EIYQKGSF KKENVDFNDKK EKDESREKEKNREK IIIAELQKTCWMDIRYNIDGKIESAKYV NKEKLAKLIDFYKENLKK PSEEESWDRLLAFAGFSDTYSKSIDQFYIEVDKQGYKLEFVTINKARLDYEVRDGKJYLF EIRSRDNNLNVNGEEKTSAKNLTQIYWNAAFGDDNPKLNGEAEI F YRPAIAENKLNNK DKNGKGEI IDGYRFSKEK FIPHCPCITLNFLKETKINDKLNA ALAKPENGQGVYPLGIDRG EKHLAYYSLVNQKG EILEQGTLNLPFLDKNGKRSRSIKVEKKSFEKDSNGI I KDKD GND KIKIEFVECWNYNDL DARA GRDRDYARKNWTITGKIELKDGYI S QVVRKIVVDSL IYKN TETKEFREMPA FIVLEDLNIGFKGRGRQKIEKQVYQKLELALAKLKNFLVDKKADIGEIGS VTKAIQLTPVNNF GDME NRKQFGNMLYI RADY TSQTDPATGWRKSIYLGSGSE SNVKE QIEKSFDIRYESG DYCFEYRDRHKGKMWQLYSSKNGVSLDRFHGERNNNSKVNWESEKQ PLNEMLDI LFDEKRFDSKSLYEQMFKGGVALTRLPEI NKDKPWAESLRFVII LIQQIR NTGKNGD DRNGDFIQSPVRDEKTGEHFDSRIYLDKEQKGEKADLPTSGDANGAYNIAR KGIVVVAEHIKRGFDKLYI SDEEDWDTWLAGDEIWDKWLKENRESLTKTRK*
SEQ ID: 27	MNYKTGLEDFIGKESL SKTLRNALI PTES T KIHMEEMGVIRDDELRAEKKQQLKEIMDDY YRTFIEEKLQGQI QGQI QWNSL FQKMEETMED ISVRKDLKQNEKRKEICCYFTSDKRFKD LFNAKLITDILPNF KDNKEYTEE EKAKEQTRVLFQRFATAFT NYFNQRRNNFSEDNIST AISFRIVNENSEIHLQNMRAFQRIEQQYPEEVCGMEEYKMDLQEWQM KHIYSVDFYDR

TABLE 1-continued

Amino Acid Sequences of Exemplary CRISPR/Cas effector proteins	
SEQ ID NO:	Programmable Nuclease Amino Acid Sequence
	ELTQPGIEYYNGICGKINENHMNQFCQKNRINKNDFRMKKKLHKQILCKKSSYYEI PFRFES DQEVDYDALNEPIKTMKKKEIIRRCVHLGQECDDYDLGKIVISSLNKYEQISNALYGSWDTI RKCIKEEYMDALPGKGEKKEEKAEEAAKKEEYRSIADIDKIISLYGSEMDRTISAKKCIT EICDMAGQISDPLVCNSDIKLQNKEKTTIEKTLDSFLHVVYQWGQTIVSDIIEKDSYFY SELEDVLEDPEGITLYNHVRSYVTQPKYSTVKFLHFGSPTLANGWSQSKEYDNNAIL LMRDQKPYLGIFVNVRNPDKQIIKGHEKEEKGDYKKMIVNLLPGPSKMLPKVFTTSRG QETYKPSKHLGDGYNEKRHIKSSPKFDLGYCWDLIDYYKECIEHKHPDWKNYDFHFSDTK DYEDISGPYREVEMQYQIKWTYIASADEIQLDKEKGQIIFLFQIYNKDFSVHSTGKDNLHT MYLKLNLFSEENKLIVKLNHGKGLSSEAQRYLDEGKIKSFTATKDIDVKNYRYCCDHYFLHLP SIEPEYYTEIINYLNHIGKGLSSEAQRYLDEGKIKSFTATKDIDVKNYRYCCDHYFLHLP TINFKAKSDEVAVNERTLAYIAKKEDIHIIIGIDRGERNLLYIISVVDVHGNIREQRSFNIVNG YDYYQQKLKDREKSRAARKNWEIEKIKELKEGYLSMVIHYIAQLVVKYNAAVVAAMED LNYGFKTRFKVERQVYQKFETMLIEKLHYLVFKDREVCCEGGVLRGRYQLTLYIPESLKK VGKQCGFIFVYPAGYTSKIDPTTGTVNLNFSPKNLNTRESQDFVGKFDEIRYDRDKKMFE FSFDYNNYIKKGTLASTKWKVYTNGTRLKKIVVNGKYTQSMEVELTDAMEKMLQRA GIYEYHDGDKLKGQIVEKGIEAEIIDIDFRLTVQMRNSRSESEDREYDRLISPVLNDKGEFFD TATADKTLPQDADANGAYCIALKGLEYEVKQIKAENWKENEQFPRNKLVQDNKTWFDFM QKKRYL*
SEQ ID: 28	MRISKTLSLRIVRPFYTPEVEAGIKAEKDKREAQGQTRSLSDAKFPNELKKHSEIILSSEF YSLLSVQRLPQDNLKQFPMMAVPTAKSDKPFPIYRQVGDGSKGPKISENDGKDFIVELPLVD OKIQSNFRRDLKQFPMMAVPTAKSDKPFPIYRQVGDGSKGPKISENDGKDFIVELPLVD VVAEEVKTAKGRFTKINISKPCKIKNIPVILSTLRRQSGQWFSDDGTAEIRVISGEYK VSWIEIVRRTRFGKHDDWFVNMFIVKYDKPEEGLDSKVGGIDVGVSPLVCALNNSLD RYFVKSSDIIAFNKRAMARRTLLRNQNKYKRSGGHGSKNNKLEPITVLTENRFKKSIMQ RWAKEVAEFFRGKGASVVRMELSLGLKEKDNNFSSYLRMWMNYGOLQQIIEENKLKEY GIKVNYVSPKDTSSKKCHSCTHINEFFTPEYRQKNNPFLFKCEKCGVECSADYNAAKNMA IA
SEQ ID: 29	MRTMVTFEDFTKQYQVSCTLRFELIPQGKTLENMKRDGIISVDRQRNEDYQKAKGILD LYKYILDFTMTVVIDALATAEFRKSCKDKTYEVQSKIRTALLEHVKKQKVGT DLFKGMFSSKIITGEVLAAPFEIRLSDEENLILEFKDFTTYFTGFFENRKVNFTDEALSTS FTYLVNDNFNIFKFFDCIVFKNVNVNISPMAKSLETCAASLGIIFPGVSLEEVSISFYNRLL TQTGIDQFNQLLGGISGKEGEHKQQGLNEIINLAMQONLEVKEVLKNAHRTPLFKQIL SDRSTMSFIPDAFADDDEVLSADEVAYRKYLSEKNIGDRAFQLISDMEAYSPELMRIGGK YVSVLMSQFLSYWSERDGVKAYKESLITGKTKKELENIDKEIKYGVTQLEIKEALPKK DIYEEVKKYAMSUVVKDHYAGLAEPLPEKIELTDERASIKHIMDSMLGTYRFLEYFSHDSI EDTDPVFGECLDLTDMMNETVPLYNKVRNFSTRKVYSTEKFKLNFNNSSLANGWDKN KEQANGAILLRKEGEYFLGIFNSKNPKLVSDDGGAGIYKEMIYKQFPDFKKMLPKCTIS LKDTKAHFQKSDEDFTLQTDKFEKSIVITKQIYDLGTQTVNGKKKFQVDYPRLTGDMEG YRAALKEWIDFGKEFQAYTSTAIDTSLFRDSSYPDLPSFYKDVDNICYKLTFEWI AVIDDCIDDGSLYLFKLHNKDFSSSGSGKPNLHTLYWKAFFEEENLSDVVVKLNGQAE FYRPKSLTRPVVHEEGEVINKTTSGLPVPPDDVYVELSFKVRNGKKGNLTDKAKNWLD KVTVRKMPHAITKDRTFTVDKFFFHVPIITLNYKADSSPYRENDFVRQYIKDCSDV DRGERNLIYAVVIDGKGNIIEQRSFNTVTGTYNYQEKLEKEKERQTARQDWATVTK LKKGYLSAVVHELSKMINVKYKAIAVLENLNVGFKMRGGIAERSVYQQFEBALIDKLN YLVFKEEQSGYGGVNLNAYQLTDKFESFSKMGQQTGFLFVPAAYTSKIDPLTFINPFS WKVKNRDPRRNFLNLSKLYDVNTDHFVLAYHHNSNQDKSYTIGKWNWEIADWDILIQ ENKEVFGKTGTPYCVGKRIIVMDDSTTGHNRMCAYPPHTELKLLSEYGIYEYTSQD LKIIQEFDDDKLVKGFLFYIIKAALQMRSNSNSETGEDYIISSPIEGRPGICFDSRAEADTLPYD ADANGAPHIAMKGLLTERIRNDDKLAIISNEEWLNYIQEMRG*
SEQ ID: 30	MSKLEKPTNCYSLSKTLRFKAIPVGKTOENIDNKRLLVEDEKRAEDYKGVKLLDRYYL SFINDVLHSIKLKNLNNYISLFRKKTRTEKENKELENLEINLRKEIAKFKGNEGKSLFK KDIETILPEFLDDKDEIALVNSFNGFTTAFTGFFDNRENMFSEEAKSTSIAFRCINENL YISNMIDIEFKVDAIFKHEVQEIKEKILNSDYDVEDFEEGEFFNVLTQREGIDVYNAIIGF VTESEGEKIKGLNEYINLQKTKQKLPKFKPLYQVLSDRESLSFYGEGETSDEEVLEV RNTLNKSEEIFSSIKKLEYLQKLPKFKPLYQVLSDRESLSFYGEGETSDEEVLEV YDDIHLKKKAVVTEKVEDDRRKSFKKIGSFSLEQLQEYADADLSVVEKLKEIIIQKVDEI YKVGGSSEKLFADFVLEKSLKKNDAVVAIMKDLLDSVKSFSFENYIKAFFGEGETNRDE SFYGFVLAYDILLKVDHIYDAIRNYVTQPKYSKDKFKLYFQNPQFMGGWDKDKE RATILRGSKYLAIMDKKAKCLQKIDKDVNGNYEKINYKLLPGPNKMLPKVFFSK KWMAYYNPSEDIQKIKYKNGTFKKGDMFNLNDCHKLIDFFKDSISRYPKWSNAYDFNFS ETEKYDIAFGYREVEEQGQYKVSFESASKKEVVDKLVEEGKLYMPQIYNNPQDFSDKSHGTP NLHTMYFKLLFDENNQHGNLRLSGGAELFMRASRSLKKEELVVPANSPIANKNPDNP KKT TTLSDYVVKDKEFSEDOYELHPIIAINKCPKNIFKINTEVRVVLLKHDNDP VIGIDRGERN LLYIVVVVDGKGNIVEQYSLNEIINNFNGIRIKTDYHSSL DKEKERFEARQNWTSIENIKE LKAGYISQVVKHICELVEKYDAVIALEDLNSGFKNSR VKEQVYQKFEKMLIDKLNY MVDKKSNCATGGALKGYQITNKFESFKSMSTQNGFIFYIPAWLTSKIDPSTGFV NLQ

TABLE 1-continued

Amino Acid Sequences of Exemplary CRISPR/Cas effector proteins	
SEQ ID NO:	Programmable Nuclease Amino Acid Sequence
	KYTSIADSKKF1SSPDRIMYVPEEIDLFEFALDYK NFSRTDADYIKKKWKLYSYGNRIRIFRN PKKNNVFDWEVECLTSAYKELFNKYGINYQCGDI RALLC EQSDKA FYSSPMALMSML QMRNSITGRTDVDFLISPVKNSDGI FYDSRN YEAQENAILPKNA DANGAYNIARKVLWA IGQFKKA EDEKL DVK V KIA ISNKEW LEYA QTSV KH
31	MTNFDNFTKKYVNSK TIRLEAPVGKTLKNIEKMGFIAADRQRD E DYQAKSVIDHIYK AFMDDCLKDLFLWDWPL YEA VVACWRERSPEGRQALQ I M QADYRKKIADRFRNHELY GSLFTKKI F DGSVVAQRLPD L E QSAE EKSLLS NF NKFT S YFRDFFD KRKR LFS DDEKHS AIA YRLIN E RLK FV ANCE KI QA QAD ALHS QALP I PENL KAL C A D G L K S M L D TV GLY RML QWFIV GD DNEK DSDF YFGLG KI LGSLDPV L VLYN RVN YI T KKP VSLT K PRL FN D NSQ L N GWD ENNLD TNCAS IF I KDGKYYLG ISNKNRNPQFDTVATSGKSGYQRMVYKQFANWGRDLP HSTT QMKV V KKH FSAS DAD YVLD GD KF IRPL I T KEI FDLN NVK PFG KKKL QVDYL RNT GD REGY THAL HTW FA KDFC A CYK STSI YDIS CLRP TDQYDNLMDFYADLGNL SHRI V WQ TIP EAD IN YV E QGQFL FQL YNKDFAPGADG KPNL H TLYW KAVFNP N LED VVVK LNGKAELF YR PRS NM DV VRH KV GEK L VNR KL KNG L TLPS RL HEE I YRV NGT LNK DLS ADAR SVLP LAV V RDV H E I I KDR RFT AD KFFF H ASL T FNF KSS D K P VGF NED VRE YL REH PDT YV VGD RGERN LI YIV V IDP QGN I V E QRS FNM ING IDY W SLL D QKE KER VEA KQAW ETVG KI KDLK CCG YLS FLI HE IT KII I KYH AVV ILEN SLG K R V RTG IA EK A VY QQ FER ML VT KLG YV V FK DRAG K A P G VLN AY QL TDN TRTA EN T G I QNGFL F VPA AFT SRV DPAT GFFDFY DWG KIK TAD KKNF I A G FNS VRY E RST GDF I VHV GAKN LAV RRA VED VR TEW DIV I EAN VRK M IDG NS Y S G K R I Y R S G E Q GH Q YEN H L P C Q E L I R A L Q O Y G I Q Y E T G K DIL PAIL L Q Q D D A K L T D T V F D V F R L A L Q M R N T S A E T G E D Y F N S V V R D R S G R C F D T R R A E A AMP K EAD AND A DAY HIAL KGL F VLE KLR GES I GIK NTE W L R Y V Q Q R H S *
32	MTPIFCNFVVYQIMLFNNNININVTMNKKHLSDFTNLFPVSKTLRFLRLEPOQKTMENIV KAQTIETDEERSHDYEKTKEYIDDYHRQFIDDTLDKFAKVESTGNNDSLQDYLDAYLS ANDNRTKQTEIQTNLRAKIAVS AF KM QPQFNLLFKKEMVKHLLPQFVTDKKRIVAKF NDFTTYFTGFFTNR EN MYS DEAKSTS I AY IRV NQN L I K F V E N M L T F K S H I L P I L P Q E Q L A T LYDDFKEYLN VASIA B E M F E L D H F S T V L T Q R Q I E V Y N S V I G R K D E N N K Q I K P G L N Q Y I N Q HNQAVKDKSAR L P L K L P F N Q I L S E K A G V S P L K Q F K S A E V V K S L N E A Y A E L S P V L A A I QDVVTNT D Y C D N G I F I K N D L G L T D I A Q R F Y G N Y D A V K R G L R N Q Y E L E T P M H N Q K A E K Y E Q V A K H L K S I E S V S L A Q I N Q V V T D G G D I C D Y F K A F G A T D D G D I Q R E N L L A S I N N A H T A I S P V L N K E N A N D P L E R K N T M L I K D L L D A I K R L Q F A K P L L G A G D E T N K D Q V F Y G K F E P L Y N Q L D E T I S P L Y D K V R S Y L T K K P Y S L D K F K I N F E K S N I L L G G W D P G A D R K Y Q Y N A V I L R K D N D F Y L G I M R D E A T S K R K C I Q V L D C N D E G L D E N F E K V E Y K Q I K P S Q N M P R C A F A K K E C E E N A D I M E L K R K K N A K S Y N T N K D D K N A L I R H Y Q R Y L D R T Y P E F G F V Y K D A D E Y D T V K A F T D M S D Q D Y K L S F L O V S E T G L N K L V D E G D L Y L F K I T N K D F S S Y A K G R P N L H T I Y W R M L F D P K N L A N V V V Y K L E G K A E V F F R R K S L A S T T H K A K Q A I K N K S R Y N E A V K P Q S T F D Y D I I K D R R F T A D K F E F H V P I K M N F K A G W N S T R L T N E V R E F I K S Q G V R H I G I D R G E R H L L Y L T M I D M D G N I V K Q C S L N A P A Q D N A R A S E V D Y H Q L L D S K E A D R L A A R R N W G T I E N I K E L K Q G Y L S Q V V H L L A T M M V D N D A I L V L E N L N A G F M R G R Q K V E K S V Y Q K P E K M L I D K L N Y I V D K G Q S P D K P T G A L H A V Q L T G L Y S D F N K S N M K R A N V R Q C G F V F Y I P A W N T S K I D P V T G F V N L F D T H L S L S G E I K A F F S K P D S I R Y N Q D K G W F E F K F D Y S R F T T R A E G C R T Q W T V C T Y G E R I W T H R S K N Q N N Q F V N D T V N V T Q Q M L Q L L Q D C G I D P N G N L K E A I A N D S K K S L E T L L H L F K L T V Q M R N S V T G S E V D Y M I S P V A D E R G H F F D S R E S D E H L P A N A D A N G A F N I A R K G L M V V R Q I M A T D D V S K I K F A V S N K D W L R F A Q H I D *
33	VKISKTLSLRIIRPYTTEPEVESAIAKAEDK KREAQGQTRNLD A K F F N E L K K H P Q I I L S G E F Y S L F E M Q R Q L T S I Y N R A M S L Y H K I I V E G E K T S T K A L S D I G Y D E C K S V F P S Y I A L G L R Q K I Q S N F R K E L K G F R M A V P T A K S D K F P I P I Y Q V D D G K G G F K I S E N K E G D F I V E L P L V E Y T A E D V K T A K G K P T K I N I S K P P K I K N I P V I L T L R K Q S G Q W F S D E G T N A E I R R V I S G E Y K V S W I E V V R R T R F G K H D D W F L N I V I K Y D K T E D G L D P E V V G G D I V G V S T P L V C A V N N S L D R Y F V K S S D I I A F K K R A M A R R T L L R Q N R F K R S G H G S K S K L E P I T I L T E K N E R F K K S I M Q R W A K E V A E F F K G E R A S V V Q M E E L S G L K E K D N F F G S Y L R M Y W N Y G Q L Q Q I I E N K L K E Y G I K V N V S P K D T S K K C H S C G Y I N E F F T F E R Q K N N F P L F K C K C G V E C M A D Y N A A K N I A I A
34	M L K S Y D Y F T K L Y S L Q K T L R F E L K P I G K T L E H I K N S G I I E S D E T L E E Q Y A I V K N I I D K L H R K H I D E A L S V D F T K H D L T L K T F Q E L Y L K R G K T D K E E L E K L S A D L R K L I V S Y L K G N V K E K T O H N L N P I K E R F E I L F G K E L P T N E E F F L L A E N E K E K K A I Q A F K G F T T Y F K G F Q E N R K N M Y S E E G N S T S I A Y R I I N E N L P L F I E N I A R F Q K V M S T I E K T T I K K L E Q N L K T E L K K H N L P G I F T I E Y F N N V L T Q E G I S R Y N T I I G G K T T H E G V K I Q G L N E I I N L Y N Q Q S K D V K L P I L K P L H K Q I L S E E Y S T S F K I K A F E N D N E V L K A I D T F W N E H I E K S I I H P V T G N K P N I L S K I E N L C D Q L Q K Y K D K D L E K L F I E R K N L S T V S H Q V Y G Q W N I I R D A L R M H L E M N K N K I E K D I D K Y L D N D A F S W K E I K D S I K I Y K E H V E D A K L N E N G I I K Y F S A M S I N E E D D E K E Y S I S L I K N I N E K Y N N V K S I L Q E D R T G K S D L H Q D K E K V G I I K E F L D S L K Q L Q W F L R L L Y Y T V P L D E K D Y F Y N E L E V Y Y A L L P L N S L Y N K V R N Y M T R K P Y S V E K F K L N F N S P T L L D G W D K N K E T A N L S I I L R K N G K Y Y L G I M N K E N N T I F E Y Y P G T K S N D Y Y E K M I Y K L L P G P N K M L P K V F F S K K G L E Y Y N P P K E I L N I Y E K G E F K K D K S G N F K K E S L H T L I D F Y K E A I A K N E D W E V F N P K F K N T K E Y E D I S Q F Y R D V E

TABLE 1-continued

Amino Acid Sequences of Exemplary CRISPR/Cas effector proteins	
SEQ ID NO:	Programmable Nuclease Amino Acid Sequence
	<p>EQGYLITFKEVUDANYVDKLVEKGKLYLFQIYNKDFSENKKSKGNPNLHTIYWKGLYDS      ENLKNVYVKLNGEAEVFYRKKSIDYPEEIYNHGHHEELLGKFNYPIIKDRRYTQDKFL      FHVPITMNFISKEEKRVNQLACEYLSATKEDVHIIGIDRGERHLLYLSSLIDKEGMNIKKQLS      LNTKNENYDKEIDYVRLDEKKRDEARKNWDVIENIKELKEGYSMSQVIIIAKMMV      EEKAIKIMEDLNLIGFKRGRFKVEQVYQKFEKMLIDLKNYLVLFKNKNPBLEPGGSLNAYQ      LTSKFDSPFKLGKQSGFIFYVPSAYTSKIDPTTGFYNFICQDVNPLEKGKEFFSKPEKIIYN      TKEDYFEEFHCKYGFVSEPKNKDNDRTKESLTYYNAIKDTVVVCSTNHERYKIVRN      KAGYYEHSVDPVTKNNLDIIFSQANINNEGKDIKPIIIIESNNAKLKSIAEQLKLILAMRY      NNGKHGDDKDYLILSPVKNKQGKFFCTLDGNQTLPINADANGYNAIKGLLLIEKIKK      QQGKIKDLYISNLEWPMFMMSR</p>
36	<p>M1KNPSNRHSPLPKVIISEVDHEKILEFKIKYEKLARLDRFEVKAMHYEGKEIVFDEVLVN      GGLIEVEYQDDNKTFLVKVGKGEKSYSIRGKKVGGKQRLLEDTRSSTKVKQLELSDGVVDN      KGMLRKSRTERELIVADNIKLYSQIVGREVTTTKEIYLVLKFLAYRSDSLFFYYSFVDNFF      KVAGNEKELWINKINFDDATSAQFMGYIIPFMVNDNLKNDNAYLKDYVRNDVQIKDDLKK      VOTIFSLARLHTLLHFNEYEFFKELNGEDVGPFDIIGFLNLIIENIDKLNIDAKKEFIDNEKI      RLFGENLSLAKVYRLSYEFKELNGEDVGPFDIIGFLNLIIENIDKLNIDAKKEFIDNEKI      DIHSNQBYKRIYNEHKKLVLKVSTLKDGQAIRRGNKISLKEQMKSMTKKNSLARLEC      KMRLAFLGFLYGEVNYYKAFKMNFDTNKINSQFDVNDVEKSKAYPLSTYERRKPRTREK      LEKVAKDIESLELKTVIANDTLLKFILLMFVFMPQELKGDFLGFVKKYYHDVHSIDDDT      KEQEEDEVVAMESTSLKLKILGRNIRSLSTLPKYALSSQVNYNSTDNIFYVEGNRYGKIKK      LGISHNQEEFDKTLVPLLRYYSSLFJKLMNDPEIYSLAKANPTAVSLQELVDDETSPYKQ      GNYFNFMKMLRDIIYGLTSDEIKSGQVVFMRNKIAHFDETEVLLSKPLLGOTKMNLRKDI      VSFIEARGDIKEELLGYDAINDFRMKVHLRTKMRVYSDLQTMMDLLRNAKTPNDFYN      VYKVKGVESINKHLLEVLAQTAEERTVEKQIRDGNEYKDL</p>
37	<p>LNSIEKIKKPSNRNSIPSIIISDYDENKIKEIKVYKLKLARDKTIQDMEIRDNIVEFKKILL      NGIEHTIKDNQKIEFDNYEITAYVRAVKQRDGGKTOQAKYVVTITDKYLRDNEKEKRFKS      TEREPLDTLMLRKYKISGFDLTSKDIYKIKRYIDFKNEMLFYQOFIEBFFSPPLPKGTNF      YSLNIEQNKDKVVKYIVYRLNDDFKNQSLNQFPIKKTDTIKYDFLKIQKILSDFRHALAHF      DFDFIQKFDDDELDKNRFDISTISL1KTMLOQEKEEYQQEKNNYIEDSDTTLTDFDEKESNF      SKIHNFYIKISQKKPAFKNLINSFLSKDGVNPNEELSKYLATKKIDFFEDIHSNKEYKIKYK      HKNLVVEBKQKEESQEKPGNGQQLKNNYDELQKLDEMNNKIKTQNSLNRLEVKLRLAFC      IANEYNYNFKNFNDKFTL DVKCKQKIKVFKVNSNEKLFYKEYFESTFIEKRFFHFCVKFVN      KTKKEETTKQKNIFLN1ENETLEELVKESPPLLQIITLLYLFIPKELQGEFVGFLIKIYHHTKNI      TNDTKEDEKSIEDTQNSFSLKLKILAKNLRLQFLFVNYSLSHNTLYNTKEHFFYEGGNRW      QSVYKSLEISHNQDEFIDHLVIPV1KYYINLNKLIGDFEYALLYADKNSITEKLSDITKR      DDLKFRGYYNFNSTLLFKTMINTNQEONQKSTQYIKQTRNDIAHQNIENMLKAPENNEIF      AQREEIVNVLQYKEMQEIHLVYNPINDFTMKTVCYQLKSLNHSQESKIADEHKKESLVP      NYKYYLKVIELLKQKVIAGETKDEEIKNIAKEEQIKKGYNK</p>
38	<p>MLKHKRKNKNSLARVVLNSYDNNIYEIKIKYEKLAKLDKINIIEMDYDADNNVMFKK      VLFNNKEIDLSHKDKTKINIELDNKKYNNISAKKQIGKTHLUVRNQTSKISRKKIQDTYY      RGKDFVFLDNNEI1LDKQTKDKFIVTLNDITNNKTTSTEALIDDTKD1FKKISAKKDLKS      SDIYKIKRFISIRSNFSFYTFVNDYFKIFHAKKDKNKEELYKIKFKDEINIKPYLENILDN      MKNKGNGILYNYANDRKVLNDLRRNIQYVFKEFRHKLAFHDYNFLDNFSNSVEEKYKQ      KVNEIKLLD1LLDN1DSLNVVPKQNYIEdETISVFDAKD1KLKRLYTTT1KLTINYPGPKKL      INSFFIQDGIENQELKEYINNKEKDTQVLELDNKAYMDISQYRKYKNIYNNKHELVSE      KELSSDGGKINSLNQKINKLKIEMKNITKPNALNRLIYRLRAFGFIYKEYATINNENKS      LFQDTKTRFENISQDQIKSYLDISYQDKGKFFVKSCKTFKNKTTVKYTFEDLDTLN      NEIITQDDIFVVKVIFLFSIFMPKELNGDFFGFVINMYHKMKNISYDTKDIIDMLDTISQNM      KKLKILEQNIKKTVFKYLLDLD1SIY5KLVQNIKITEDIDSCKYLYAKIFKYYQHLYKLISD      VEIYL LYKYNNSKENLSITIDKDELKHRGYYNFQSLLIKNNINKDAYWSIVNMRNNL      SHQNIDE LVGHFCKGCLRKSTTDIAELWLRKDILTITNEIINKIESFKD1K1LGYDCVNDFTQ      KVQKQKLKASNERLAKKIEEKQMQVVDENKKEELEKNILNMRNNLQKINRYILDIL</p>
39	<p>MLKHKRKNKNSLARVVLNSYDNNIYEIKIKYEKLAKLDKINIIEMDYDADNNVMFKK      VLFNNKEIDLSHKDKTKINIELDNKKYNNISAKKQIGKTHLUVRNQTSKISRKKIQDTYY      RGKDFVFLDNNEI1LDKQTKDKFIVTLNDITNNKTTSTEALIDDTKD1FKKISAKKDLKS      SDIYKIKRFISIRSNFSFYTFVNDYFKIFHAKKDKNKEELYKIKFKDEINIKPYLENILDN      MKNKGNGILYDYADDREKVLNDLKNQIYVFKEFRHKLAFHDYNFLDNFSNSVTDQYKQ      KVNEIKLLD1LLDN1DSLNVVPKQNYIEdETISVFDAKD1KLKRLYTTT1KLTINYPGPKKL      INSFFIQDGIENQELKEYINNKEKDTQVLELDNKAYMDISQYRKYKNIYNNKHELVSE      KELSSDGGKINSLNQKINKLKIEMKNITKPNALNRLIYRLRAFGFIYKEYATINNENKS      LFQDTKTRFENISQDQIKSYLDISYQDKGKFFVKSCKTFKNKTTVKYTFEDLDTLN      NEIITQDDIFVVKVIFLFSIFMPKELNGDFFGFVINMYHKMKNISYDTKDIIDMLDTISQNM      KKLKILEQNIKKTVFKYLLDLD1SIY5KLVQNIKITEDIDSCKYLYAKIFKYYQHLYKLISD      VEIYL LYKYNNSKENLSITIDKDELKHRGYYNFQSLLIKNNINKDAYWSIVNMRNNL      SHQNIDE LVGHFCKGCLRKSTTDIAELWLRKDILTITNEIINKIESFKD1K1LGYDCVNDFTQ      KVQKQKLKASNERLAKKIEEKQMQVVDENKKEELEKNILNMRNNLQKINRYILDIL</p>

TABLE 1-continued

Amino Acid Sequences of Exemplary CRISPR/Cas effector proteins	
SEQ ID NO:	Programmable Nuclease Amino Acid Sequence
SEQ ID: 40	MSQLKNPSNKNSLPRIIISDFNEIKINEIKIKYHKLDRLDKIIIVKEMEIINNKIFFKKILENNQIKDINSENIELENYILAGEVKPNSNTKIIILNRDGKEKSFIVYDGFTFKYKPNDKRISETKTNAKYIILTICKTRHRESSTQRDILKSSIIETYKQISGFENITSKDIYTIKRYIDFKNEMMFYTFIDDFFPITGKNCDDKKNFVNLYKIKENAKKFISLINVIRINDDFKNKGILYDYLNSKEEIIIIDNFIHQITQILKDVRHAIAHFNFDFOIQLFDNEQAFNSKFDGIEILNLIFNQOKQEKYFEAQTNYIEEETIKILDEKELSKFLHSFSQJCQKKPAFNKLINSFIIQDGIEENKEKDYISQKYNKFDDYYLHCTKIKYKDIYNQHKKFVADQKFLENQKTDGQKIKKLNDQINQLKTKMMNLTKKNSLRKLEIKFRALFGFIFTEYQTFKNFNERFIEDIKANKYSTKIELLDYGKIKEYISITHEEKRFFNYKTFNKTKNINKTIFQSLKETFENLVKNDNLKMMPLFQLLLPRELKGEFLGFILKIHYDLKNIDNDTTPDEKSLSLELNISTALKLKILVKVNIRQINLFNFTISNNTKYEEKEKRFYEEGNQWQDIIYKLYISHDFDIPDIHLIIPIIKYNINLYKLIGDFEVYLLLKYLERNTNYKTLDDKLIIEAEELKYKGYYNFTTLLSKAINIALNDKEYHNITHLRNNTSHQDIQNIISSFKNNKLQRENIIELISKESSLKKLHDPINDFTMKTLQLLKSLEVHSDKSEKIEENLLKKEPLPNVDVLLYKLKGIEFIKKELISNIGITKYEEKIQEKIAKGVEK
SEQ ID: 41	MVKNPANRHALPKVIISEVDDNNNILEFKIKYEKLARLDKVEVKSMHFDDNNKQVVFDEVVINGGLIIEPTYEDDKHKKLVVTAGEKSYSIVGQKVGGKPRLLLEDRTSPTKVQLELTNVVEDKEGKKRVSCTRERELIVADNIELYSQIVGREVTTKIEYLICKRPLEYRS DLLFYYGVFDNFVKVAGNGKELWKIDFTNSDSLHLIYEYFKFSINDNLKNDENYLNKNVSDNTKIEENDLVKCQNNFNSRLRHMDYDFFALFNGEDVGPDFDIEFLNIMIDKVDKLNIDTKEFIDDEEVTLFGEALSLLKLYGLFMSHIAINRVAFNKLINSFIIEDGIEENKEKDFFENNKKESQAYEIDHSNAEYKALYVQHKKLVMATSAMTDGDEIAKKNQEISSLKEKMKVITKENSARLEHKLRLAEGFPIYTEYKDYKTFKKHFDQDIKGAKYKGLNVEKLKEYEETTLKNSPKPTDEKLEDVAKKIDKSLKELKVLKDFVLLFIFMPQELKGDFLGFIKYVYHDKKHIDQDTDKDCTEIELSLSTGLKLVLDKNIRSLSLLKHSFSFQVKYRNKDNFVYEDGNLHGKFYKLLSISHNQEEFNMKSVAAPLFRYYSALEYKLINDFEYIYALAQHVHENETLADQVNKSQFQKSYFNFRKLLNDNTDISQSSSYNTLIVMRNDISHLSSYEPLFNYPPLDERKSYKKKTQKGVKTFHVELLYISRAKIIIELISLQTDMKLGYDAVNDFNMKVUHLRKRLSVYANKEESIRKMADAFTPNDFYNIYKVKGVESINQHLLKVIGVTEAEKSIEKQINEGNKKHNT
SEQ ID: 42	MTKKPSNRNSLPKVIINKVDESSILEFKIKYEKLARLDRFEVRSMRYDGDRIIFDEVVANAGLLVDYEDDNRTIVVKIENKAYNIYGKKVGGKEKLNGKISKAKVQLLTDSSTRKNAIDTHRHSLSLTEREIDLQKTYDFTTQKFLVCRFLAYRS DLLYYAFINHYRVNGNKKEFWKTEIDDKIIDYFIYTINDTLKNEGYLEYKIVDRDQIKKDLKEKIKQIFSHLRHKLMHYDFRFFFTDLDGKDVDIIVKDNISIQLKISELSDIEFLNIVIDKLEKLNIDAKKEFIDDEEKITLAFGKIDKSLKTLIQDDILLKFLFFFPLPQEIKGEFLGFIKYHDITSDEDTKDKDEITLKGIDKSLKTLIQDDILLKFLFFFPLPQEIKGEFLGFIKYHDITSDEDTKDKDEITELPRSLSLKLKIFSKNIRKLSLILKHSLSYQIKYVNKKESSYYEAGNVPNKMFKKQAI SHNLEEFGKSIYLPMLKSYQIKYKLINDFEIYALYKDMDTSETLSQVDKQEYKRNEYFNFETLLRKKFGNDIEKVLVTRYRNKIAHNDLDFNFLYDCKINKKISLYKSREKIVNYIKNHDIAQVLKYDAVNDFVMKVIQLRTKLKVYADKEQTIESMIQNTQNPNGFYNIYKVKAIVENIRHLLKVI GYTESEKAVEEKIRAGNTSKS
SEQ ID: 43	MICNPNRYALPKVIISKIDQNQILEFKIKYKKLSKLDIVVKVSMHYDDRAIIFDEVVNDGLIDVEYRDNHKTIFVKVGNKSYSISGQKVGGKERLLENRVSPTKVKQLELKDKATNRVSKTERELVDDNIKIYSQIVGRDVKTTKDIYLIKRFKFLAYRS DLLFYGVNNFFFVANNRSEFWKIDFDDNSNSKLIISDQFETTQKFTINDHLKNDENYLNQKTDYISDNEKLKNDLKVKNSEFKIRHALMHFDYDFVQLFNGEDVGLLELDIEFLDIMIDKLDKLNIDTKEFIDDEKITIFGEELS LAKLYRFYAHATAINRVAFNKLINSFIIENGVENQSLKEYFNNQQAGGIAYEIDIHQNEYKNLYNEHHKKLIVSRVLSISDQGQETAILNQKIAKLKDQMOKITKANSIKRLEYKLRLALGFIFTYENYENEFKNNDTPIKNGRTPKDNNDGKRAFDSSRELQKGYYEATIQTQKPKTDEKIEEVSKKIDRLSLKSLSIADDIILLKFLILLMFTMPQELKGEFLGFIKYVYHDTKHIDQDTISDSDDTIETLSIGLKLKILDKNIRSLSLIKHSLSFQTKYNKDRNYYEDGNIHGFKKLGISHNQEEFNKSVAAPLFRYYSALEYKLINDFEIYTLSLHVGSSETLTQVNKSQPLSGRYFNFETLLRLTQSYYHINNNNSTHSTFNAVINMRNDISHLSSYEPLFDCPLNGKKSYKRKIRNQFKTINI KPLVERSRKIIIDFTLQTDMKVULGYDAVNDFTMKIVQLRTRLKAYANEKEQTIQKMTIEAKTPNDFYNIYKVQGVSEEINKYLLLEVIGETOAEKIREKIERGNIANF
SEQ ID: 44	MKKSIFDQFVNQYALSKTLRFLPELKPGVGETGRMLLEEAKVFAKDETIKKKYEATKPFNNKLHREFVERALNEVELAGLPEYFEIIFKYWKRYKKFKEKDLQKKEKELRKSVVGFFNAQAKEWAKKYETLGVKKKDVGFLPEENVFAILKERYGNEEGSQIVDESTGKDVSI FDSWKGFTGYFIKFQETRKNFYKDDGTATALATRIIDQNLKRFCDNLLIFESIRDKIDFSEVBQTMGNSIDKVFSVIFYSSCLLQEGIDFYNCVLCVGGETLPNGEKRQGINELINLYRQKTSEKVPFLL DKQILSEKEKFMDEIENDEALLDTLKIFRKSAEEKTTLLKNIFGDFVMNQGKYDLAQIYISRESLNTISRKWTSETDI FEDSLYEVLKSKIVSASVKKDDGGYAFPEFTALIYVKSALQIPTEKFWKERYYKNIGDVLNKGPFLNGKBEGLQFLFIDFDFEPNSLFREREIIDENGDKVAGYNLFAKGFDLNNFQYDQKAKVVIKDFADEVLHIYQMGKYFAIEKKRSWLADYDIDSFYTDPEKGYLKFYENAYEEI IQVYNKLRNLYLTKPYS ESDKWKLNFENPTLADGWDKNK EADNSTVILKKDGRYYLGLMARGRKNKLFD DRNLPKILEGVENGKYEVYYKYFPDQA

TABLE 1-continued

Amino Acid Sequences of Exemplary CRISPR/Cas effector proteins	
SEQ ID NO:	Programmable Nuclease Amino Acid Sequence
	KMFPKVCFSTKGLFEEFPSEEVITIYKNSEFKKGTYFNVRSQRLIDFYKDCLVRYEGW QCYDFRNLRKTEDEYRNIEEFFSDVAMDGYKISFQDVSESYIKEKNQGDLYLFEIKNK DNEGANGKKNLHTIYFESLFSADNIAMNFPVKLNGQAEIFYRPRTEGLEKERITKGN VLEKGDKAFHRRYTFENKVFFHVPITLNRTKKNPQCFNAKINDFLAKNSDINVIGVDRG EKQLAYFSVISQRGKILDGSLSNVINGVNAYEAKLEEKARGREQARKDWQIEGIKDLKK GYISQVRKLADLAIQYNAIIVFEDLMRKFQIQRGGIEKSVYQQLEKALIDKTLFLVEKEB KDVEKAGHLLKAYQLAAPFETFQKMGKQTGIVFTYQAAATSRIIDPVTVGRPHLYLKYS SAEAKADLLKPKKIYKFDGRPEFETYDIKSFREQKEHPKATVWTVCSCVERFRWNRYL NSNKGGYDHYSVDTFLVFLPQEFQEQYGDIVGQIEVLETKGNEKKFKNFVFFFNLIC QIRNTNASELAKDKDGFILSPVEPFDSRNSEKFGEDLPKNGDDNGAPNIARKGLVIM DKITKFADENGGCEKMKGDLVYVSNUVEWDNFVANK
45	MENNFIKKYSLQKTLRFELKPVGETADYIEDFKSEYLKDVTLDKDEQRAKYQEIKTLIDD YHREYIBECLREPVDKKTGEILDFTQDLEDAFSYYQKLKENPTENRGWKEKEQESLRKK LVTSFVGNDGLFKKEFI TRDLPWEWLQKGLWGEYKD TVENFKKFTTYFSGFHENRKNM YTAEAQSTAIANRLMDNLPKFFNNYLAYQTKEKHDPLVFLRDLDALLQAAGVHEHLD AFQRPYFSRLLAQSGITAFENDLIGGRTTENGEKIQGLNEQINLYRQONPEKAKGFPRFMPL FKQILS DRETHSFLPDADFENDKELLQALRDYVDAATSEE GMISQLNKAMNQFVTADLKR VYIKAALTSLSQELFHFFGVISDAIAWYAEKRLSPKKAQESFLQKQEYVAIEELNQAVVG YIDQLEDQSELQQLLWDPDPQKPVSSFILTHWQKSQEPQLAQVIKVEPLFELELSKNK RAPKHDKDQGEGFQVQDAIKMDFMEVSHAIKPLYLVKGRKAIDMPDVDTGFYA DPAEAYSAYEQVTVSLYNKTTRNHLSSKPKPSKDKK1KINFADAPTLLNGWDLNKESDNKSIIL RKDGNYLAIMHPKHTKVFDCYSASEAAGCKYEKMYKLLSGANKMLPKVFFSKKGIE TFSPPQBEILDLYKNNNEHKKGATFKLESCHKLIDFFKRNIPKYKVHPTDNFGWDVPGFHFS PTFSYGLDLSGFYREVEAQGYLWFSDDSEAYINKCVEEGKLFLFOIYNNKDFSPNSTGKPN LHTLYWKGLFEPENLKVVLKLNGEAEIIFYRKHSIKHEDKTIHRAKDPIANKNADNPKK QSVPFDYDIKDKRTQDKFFHVPISLNFKSQGVVRENDKINGLLAAQDDHVIGIDRGE RHLLYYTUVNGKGEVVEQGSLNQVATDQGYVVDYQZQKLHAKEKERDQARKNWSTIE NIKELKAGYLSPQVHKLALQIYKVNHAIVCLEDLNFGFKRGRFKVEKQVYQKFEKALIDK LNYL VFKERGATQAGYLNAQLAAPFESFEKFLGKGTGILYVYVRSYDTSKIDPATGFD FLPKPYESMAKSKVFFESFERIQWNQAKGYFEEFDYKMKCPSRKFGDYTRVVCTF GDTRYQNRRNKSSQQWETETIDVTAQLKALFAAYGITYNQEDNIKDAIAAVKTYKFYK QLYWLLRLTLSLRHSVGTDEDFILSPVADENGVFFDSRKATDKQPKDADANGAYHIAL KGLWNLQQIRQHDWNVKEPKKLNLMAKNEEWFGFAQKKFRA
46	MIKNPNRYALPKVIISKIDQNQNILEFTKIKYKKLSSL DIVKVKSMHYDDRAIIFDEVIVND GLIDVEYRDNHHTKVFVGKNSYSISGQKVVGKERLLENRVSKTKVQLEKDKATNRVS KTERELIVDDNIKISQIVGRDVKTTKDYLKFLAYRS DLLFYYGFVNPFHVANNRSE FWKIDFDNSNSKLI BYFKFTINDHLKNDENYLKD YISDNEKLKNDL I KVKN SFKIRHA LMHFDYDFVFKLFNGEDVGELELDIEFLDIMIDKLDKLNIDTKKEFIDDEKITIFGEELSLA KLYRFYAHATAINRVAFNKLINSFIIENGVENQSLKEYFNCQAGGIAYEIDIHQNREYKNL YNEHKKLVSRVLSRNLISDQGQEAIA LNQKIAKLDQMKQITKANSIKRLEYKLRLALGFYIYTEY ENYEEFKNNFDTDNIKNGRFTPKDNDGKNAFDSRELEQLGYYEATI QTOKPKTDKIE EVSKKIDRLSLKSLIADDILLK FILLMFTFM P QELKG EFLGF I KYYHDTKHIDQDTISDSD DTIETLISGLKLKILDKNRSLSIKHSLSFQTKYMKD RNNYYEDGNIHGF KFFK L GISHN QEEFNKSVYAPLFRYYSLALYKINDFEIYTLSLHIVGSETL TDQVNKSQFLSGRYFNF RKL LTQSYHINNNSTHSTIFNAVINMRNDI SHL SYEPLFD C PLNGKKS YK RKIRNQFTINI KPL VESRKIIIDFITLQTDQMVKVLGYDAVN DFTMKIVQLRTRLKAYANKEQTIQKMI TEAKTP NDFYNIYKVQGV EEEINKYLLEVIGETQAEKIREKIERGNIANF
47	MIKNPNRSRHSPLKVI ISEVDHEKILEFKIKYKEKLARLDRPEVKAMHYEGKIEIVFDEVILVN GGLIEVEYQDDNKTFLVKVGKSYSIRGKKVGGKQRLLED RVS KTKVQLELSDGVVDN KG NLRK S TERELI VADNIKLYSQIVGREVTTK EYL V K RFLAYRS DLLF YYSFVDNFF KVAGNEKLW KINPQMSA QPMGYI P FMVNDNLKNDNAYLKD V VRNDVQIKD DLLK VQTI F S A LRHTL LH FN YEFF EKLF NGEDV GPFDI GFLN LIENIDKLNIDAKKEFIDNEKI RLFGENL S LAKVYR L YSDIC VN RVG F NKF IN SMLI K DGV EN QV LKA E FNR K F GGNAYTI DI HSNQBYK RIYNEHKKLV K V STL K DQ A I RRG N K K I S E L K E QM K S M T K K N S L A R L E C KMR L A FGFL Y GEYNNY KAF KNN FDT NI KNS QFD VND VE KSKAYPLSTYERRK P RTREK LEK VAK D I E S L E L K T V I A N D T L L K F I L L M F V F M P Q E L K G D P L G F V K Y Y H D V H S I D D D T KEQEE D V V E A M S T S L K L K I L G R N I R S L T L P K Y A L S Q V N Y N S T D N I F Y V E G N R Y G K I Y K K LGISHNQEEFDKTLVWPLLRYSSLFKLMNDFEIYSLAKANPTAVSLQELV D D E T S P Y K Q GNYFNFNMKMLRDIYGLTSDEIKSGQVVFMRNKIAHF DTEVLLSKPLL GQT K MN L Q R K D I VS FIE AR GD I K E L L G Y D A I N D P R M K V I H L R T K M R V Y S D K L Q T M M D L L R N A K T P N D F Y N VYKVKGVESINKHLLEVLAQTAEERTVEKQIRDGNEYKDL
48	MEEKMLKSYDYFTKL YSLQKTLRFELPKIGK TLEHI KNS GIIESDETLEE QYAI KNI IDKL HRKHIDEALS LVDFTKHD LDTL KTFQBL YL KRGKTDKEKEELEKLSADLRKLIVSYLKGN VKEK TQHNLNP K I KERPEI LP GKEL PTNEEFL LAE NEKEK KIAQFKGFTTYF KG FQ ENR K N M Y S E E G N S T I A Y R I I N E N L P L F I E N I A R F Q K V M S T I E K T T I K K L E Q N L K T E L K K H N L P G I F T I E Y F N N V L T Q E G I S R Y N T I I G G K T T H E G V K I Q O G L N E I I N LY N Q Q S K D V K L P I L K P L H K Q I L S E E Y S T S F K I K A F E N D N E V L K A I D T F W N E H I E K S I H P V T G N K F N I L S K I E N L C D Q L Q K Y K DKD L E K L F I E R K N L S T V S H Q V Y G Q W N I I R D A L R M H L E M N N K N I K E K D I D K Y L D N D A F S

TABLE 1-continued

Amino Acid Sequences of Exemplary CRISPR/Cas effector proteins	
SEQ ID NO:	Programmable Nuclease Amino Acid Sequence
	WKEIKDSIKIYKEHVEDAKELNENGIIKYFSAMSINEEDDEKEYSISLIKNINEKYNNVKSI LQEDRTGKSDLHQDKEKVGIIKEFLDSLKQLQWFRLLLYVTVPVLDEKDYEFYNELEVYY EALLPLNSLYNKVRNYMTRKPYSVEFKFLNFSNPTLLDGWDKNKETANLISILRKNGKY YLGIMKNNENNTIEFYYPGTKSNDYYEMIYKLLPGPNKMLPKVFFSKKGLEYYYNPKEI LNIYEKEFKDIEFVYKESLHTLIDFYKEIAKNEDEWEVFNFKFKNTKEYEDISQFYR DVEEGQYLITFEKVVDANYVDKLVKEGKLYLFQIYNAKDFSENKKSKGNPNLHTIYWKGL YDSENLNKVVYKLNGEAEVYFRKSIDYDPEEIYNHGHGKEELLGKFNPYIICKDRRTQD KFLFHVIITMFNIFSEKCRVNQLACEBYSATKEDVHIIGIDRGERHLLYSLIDKEGNIKK QLSLSNTIKNEYDKEIDYRVKLRDEARKNWDVNIENIKELEGYGMQSIVIHIIAK MMVEEKAIIMEDLNLIGFKRGRFKVEKQVYQKFEKMLIDKLNLVLFKNKNPLEPGSSL NAYQOLTSKFDASFKKLGKQSGFIFYVPSAYTISKIDPTTGFYNFIQVDVPMLEKGKEFFSKFE KIIYNTKEDYFEPHCKYGFVSEPKNNDRKTKESTLYYNAIKDTVVVVCSTNHERY KIVRNKAGGYYESHPDVTKVNLDFSQANINNEGKDIKPPIIESNNAKLLKSAEQLKLIL AMRYNNNGKHGDDEDKDYILSPVKNKQGKFFCTLDGNQTLPINADANGAYNIALKGLLLI EKIKKQQGKIKDLYISNLEWFMFMSR
SEQ ID : 49	MEKSLNDFIGLYSVSKTLRFLKPVSETLENIKKPHFLEEDKKKANDYKDVKKIIDNYHK YFIDDVLKNAFSNWKKLEEEAIRREYNKNKSDSALVAEQKLGDAILKLFTSDKRYKALT AATPKELFESIIPDWFGEQCNQDNLKAALKTFQKFTSYFTGFQENRKNVYSAEAPIAVP YRIVNDNPFLQNVNLIFKTIQEKCPCQIIDVEKELESSYLGKEKLAGIFTLESFNKYLQGQ GKENQRGIDFYQIIGGVVEKEGGINLRGVNQFLNLYWQQHPDFTKEDRRIKVMPLYK QILSDRSSLSFKIESIENDEELKNLALLECADKLELKNDKESKSIFEEVCDLFSSVKNLDLSGI YINRKDINSVSRILTDGSWSLQSRMNVYABEKFITKAEEKARWQSKSLDDEGENKSKGFY SLTDLNVELEYSENVAETDIRITDYFEHRCRYYVDKETEMFVQGSELVALSLQEMCDDI LKKRKAMANTVLENLSSENKLREKTDVAVIKEYLDAVQELLHR1KPLKVMNGVGDSTFY SVYDSIYSALEVISVYNNKTRNYITTKAAASPEKYLKNFDNPTLADGWDLNKEQANTSVIL RKDGMPYLGIMNPKNPKFAEKYDCGNESCYKMIYKQFDATKQIPKCSQTQKEVQKY FLSGATEPYIILNDKKSFKSELEIITKDIWFMNNHVWGDGEKFVPKRDNETRPKKFQIGYFKQ TGDFDGYKNALSNWISFCNKLQSYLSATVDYDNFKNSEEYEGLDEFYNYLNATACYKL NFINIPITEINKMVSEGKLYLFQIYNKDFASGSTGMNHTLYWKNLFSDENLKNVCLK LNGEAELFYRPGAIKEPVHKEGSYLVNRTTEDGESIPEKITYFEIYKANGKLEKLSDEAQ NYISNHEVVIKKAGHEIKDRHYTEPKFLFHVPLTINFKASGNSYSINENVRFKLNPNPDV NIIGLDRGERHLIYLSSNQGEIIKQFTNEVERNKGRTIKVNYHEKLDQREKERDAAR KSWQAIGKIAELKEGYSIHLQTLKLMVEYNAVVMEDLNFGKGRFRHVKEQVYQ KFEHILIDKSNYLVFKDRGLNEPGGVLINGYQIAGQFESFQKLKGQSGMLFYVPAGYTSKI DPKTGFVSMMNFKDLTNVHKRDRFFSKFDNIHYDEANGSFVFTFDYKFKDGAKEEM KLTKWSVYSDRKRIVYFAKTKSYEDVLPTEKLQKIFESNGIDYKSGNNIQSVMAIQAD LKEGAKPSKEISDPWDGLLSNFKLLQMRNSNARTGEDYIISPVMADGTFFDSREEFKK GEDAKLPLDADANGAYHIALKGSLSLINKINLSDKEELKKFDMKISNADWPKFAQEKN AK
SEQ ID : 50	MNTQKKEFNPKSFKDFTNLYSLNKTLRFLSLTPNKKTAIELEFNQKKEVKCFNSNDRKIAG AYQEIKKYLNLHQEFIQEAMKKFAFSEEEELKGFEKEYLNLLNFTDKDNPKKKNKRINE YEQERKILTIKATYFSKFKSEKYQSFNLANITGKKVFSILEQKYKEDKKTLLKIHFKYK TKDEKKGEAVMFSTYLTFGNENRKNFYKSEDKAGQFATRTIDNLAQFIKNKKLFEDKY QKNYSKIGILDEQIKINLDYFNNLFLQEGLDEYNGILGNNKGEENKSNEGQINQKINIFQ KEKARLKEKEKENFNKSDPFLPKELYQKIGSIRKENDVYVIEIKTDKELAFELNNFPKNEN YLDIQSFKTFFFEKLQNEEYELDKIYLPKSVGTYFSYIAFSDWNLAFIYKNGKNEKI KIVEGGDVNVYQRSNLVKNRIDEKLKEDNLFNFKPFIDKLKFNEAKKENNWNQFWFCI EYIINSQFIGEKNLVEIQLKEKNEYEILPFGSLKLKEKYYFEAVKKYKEKMVDTESGLTDE EKEIKETLKNYLDRIKEIERIAKYFDLKKSPFEEIKQEDLDSNFYGEYQKVVDKTNELKIYQ YYSEFRNYLTQNNSEVEEKIQLNFNSGLLLDGWDLNKEVKVFKSIIIFQENGKYYLGIINKEK DKTILDKDQHPEIIFTKNSDFRKMEYKLFPSPSKMLPKISFSETAKGGDEDVGWSBEIQKIK DEFAEQYKKSKDNWKDEFNRGKLNLKLDIYQVLEKHSEGYMNTYNFELKDSKK YKNLGEFNUDDIARQNYVKVFGVIDKNYIDEVANGELFLFQIYNAKDFSEDKKEGSTTNL ETIYFKELFSKENLENPVFKLSSGAEMFFRNKIEKKKEKKLKDQGKPMISKGEKVV KRRFSEKILFHLPBINYGKGKMPNFNKKLNEYISKNPENIKIIGIDRGEKHLIYYSIIDQN GNNIESMSLNAVDEFGNFVNPEKLEEEYEDNNGNGKERRWKYIVNDKEIKVTNYQRKLD ELEKERQPSQSWQNIKNIKLNKKGYISFVVKKIVDIAENNAIIILEDLNFGFKSFRQKIE KNVYQQFEKALIDKLGFFVVDQKQKQNRQFAPQLSAPFESFQKIGQQTGIVYVVLANTSK VCPSCWQIKNFYLKYEKNTIFNLQKNQQLKVFFFQEKRFRFEYQMSKEYISVSYSDVD RQRYDFTKNQNQGGYLYEKNNSNQKEIIDKDGVIQKQSiTLQLKELFKENHIDLEKEIILQ LDNKKEKNSSGYTGVYNKFYIYLFNLLQIRNAISFREKDYIQCPSCFDFTRKENYLKINDG DNGNGAYNIALRGLYLLKGKNGIINNLEKIKLIFSNNDYFQWAKKLKNKK
SEQ ID : 51	MQNKQSFADFTNLYSLSKTLRFLKPIGQTOAMLDENKIEFEDENRKAYDGTKPYFDR LHREFINESLSNAQLKGISEYFETFKQFRSNQNNKDLKELINKQQKFLRHQIVTLEFDENG KHWATTKYAHLLKIKKKNLDILFDEQVFYILKERYGSEKETQLOVLDKETGAVTSIPDNWKG FTGYFTKFFETRKNFYKSDGTSTALATRIIDQNLNRFDNLETFHKKDKIDVKEVEIFFK LKADNVSISDFYNQCLLQNGIDKYNDLFLGGQTLLENGEKQKGINEIINKYRQDNKQDQKL FLKFLDKQILSEKDRFINEIESKEEFFQVLTEFYQSATVKVIIKTLNDFVHNTDKYKLE KIYLTKEAFNTIANWTDETQIFEDNLDLVLKNKKITAKQDFIPLAYIKEALEVIEKDRK

TABLE 1-continued

Amino Acid Sequences of Exemplary CRISPR/Cas effector proteins	
SEQ ID NO:	Programmable Nuclease Amino Acid Sequence
	FKDRYYNDPQIGFPDPQSYWEQFLAILNFFFMTHFQRVAKDKitGKKIELGYFVFEKRIK ELLDSDPSLNSQSCKI I KEFADAEVLHIFQMAKYFALEKKREWKGDYQQLDDQFYHNHDY GFKDQFYENAYEKIVQPYNKIRNYLTCKPYSVDWKWLNGNPTLANGWDKNKEADNT AVILKKDGNYYLGVGMKGKNNKFSDONKEKYKAYNSAYYELVLYKLFPPDKMMPKV CFSKKGLNFFFQPSSEEILRKYKNNFEKKGNNTFSISSLMMQKLAIFYIDCLGLYEGWKHVEFKNI KDVRYQYKENIGFYADVAESEGYKLWPEKISSEYI TQKNOQGELFLFQIYKNDFAAKTTG RKNLHTIYFEELFSQTINDNNNPFKBLNGQAEFLYRPKSLEKIEEKRNFKRSIVNKKRTQN KIFFHRTITLNRTSENIGRFNVRVNNFLANNNSNVNIIVGVDGRGEKLNAYYSIIKQNGEVLK SGSLNIINGVDYHALLTDRARQREQERRRNQDVSIEKDLKRGRYISQVVELVSLAIKYN AIIVMDELNLNMRPKQIRGGIEKSTYQOLEKALIEKLNFLVNKEETDSNQAGNLLNAYQLT APFKTFKDMGKQTGI IFYTQASYTSKIDPLTGWRPNIYLRSNAKQAKADILMFTNIYFS EKKDRFETYDLEKIDKRDLPPIKTEWTVCNSVERFSWEKSLNNNNKGYYVHYPIQDSN GEESITSKLLKLFMDGIDLDTIKTQIESLDTNKKDNANFFRKFIYFQLICQIRNTQVNKS DDGNDNFIFSPVPEPFFDSRFADKFRKNLPGKNGDENGAAYNIARKGLIILHKISDYFVKEGSTD KISWKDLSISQTEWDNFTTDK
SEQ ID : 52	MKKEKEFKSFQGDFTNLYEISKTLRFLELKPVENTQMLDEADVFGKDKVIKDKYTKTKP IDKLHRFVDES LKDVSLSGLKKYSEVLENWKKNKKDKD1VKELKKEEERLRKEVVEFF DNTAKKWNANEKYKELGLKKKD1GIGILFESEVF DLLKEKYGEEQDSFLKKEEKGDFLKNEK GEKVSIPDEWKGFGVGYFPTKFQBTTRKNFYKNDGTETALATRIDQNLKRCFDNIDDFKKIK NKIRDSEVEKFNFKNTADVFSLDFYNQCLLQKGKGDSYNEFIGGKTLENGKGLKGVNELVN EYRQKNNKEKVSFLKLLDKQI SEKEKLSFGI ENDEQOLLVVLNSPYETABEKT KILRTL GDFVHEHENYDLDKTYISKVAFNTISHKWTMETHKFEEELLYGAMKEDKP IGLNYDKKE DSYKFPFDIALG YLKKCLNNLDCDTKFWKEKYYENNADKKDKDGFLTGQNAWQDF LQIFIFPENQNFNSEAFDNKGFQKDFEEI INQKDFKNDENL KIAIKNFADSVL WIYQMAKYFAIEKKRGWDDDFELSEFYTNPNSNGYSLFYDRAYEEIVQKYNDLRNYLT KPYKEDKWKLNFENPTLANGFDKNKESDNSTVILRKKRKYLYLGLMKGNNKIFEDRNL AEFIRNIESGAYEK MAYKYL PVDVAKMIPKC STQLNEAKNHFRNSADDLEIKKSFSNPLKI TKRIFDLNNI QDKTMNSK KI SGDNKG I KI FQKEYYKISGDFDVYKPSALNDWIDFCDFL SKYDSTKDFDSI LRKT KDSLDEFYDVAKITYKISFTPVSESYIDQKNGEYLFEI YMQDFAKGKMGAKNLHTLYFENVFS PENISKNFP I KLNQNAELFRRPKIESKKEKRNL REIVNKKRYSED KIFFHCP ITLNRETGSIYRFNNYVN NFLSENNINIIGVDRGEKHLAY VIDKNGVKGIGGGSFNEINKVDYAKKLEERAGERE QSRKD WQVVEG I KDLKGKYG ISQVV RELADLAIKHNAIIVL ELDLNMRFKQIRGGIEKSIYQOL EKALIDKLSFLVKEGKDPNQAG HILKAYQLOAAPFTSF KDMGKQTGIVFYTQASYTSKTC PNCGRKNNKPFYFENNIGKAQ DALKKLKTFEYDSENKC FGSYCLSDFANKEEVEKNKNNKRNNA PYS DIEKKDCFELST KDAVRYFWHDKNT ERGKTF FEYGESEVYEEKEI QOTKRLGVKEYDISKCLIGLFEKT LDYKQNLLDKINS GKF DGT F YK NL F NYL NL F EIRNSI SGT EIDYI SC PEC QF HTDKS NGDDNGSYNIARKGMI ILDKIKQFKKENGSLDKMGWGE LFD LEW DKFAQKNNN I ID K
SEQ ID : 53	MKYTDFTGIY SVSKTLRFL EI PQGSTVEN MKREG I LNNDMHRADS YKEMKKLIDEYHKA FIERCLSDFSLSKYDDTGKHDSLEEYFYYEQKRNDKTKKIFEDIQVALRKQISKRFTGDT AFKRLFKKELIKE DLP SFVKNDPVKTEL I KEFSDFTTYFQEFHKRNKNM YTS DAKSTAIA YRIIENLPKFDIDNINA FDI AVKVP EMQEH FKTIADELRSHLQVGNDIDIKMFNLQFFNKL TQS QLDVNV NAVIGKSEGNKKI QGNEYI NLYNQKHKA LPMLKLLYKQI LSDRVAIS WLQDEFNDQMDLTD I EAFYK NKLNSNETGVL GEGKLKQI LMGLDGYNLDG FVM RNDL QLSEVSQLCGGWNI I KDA MTS DLKRS VQKKKET DADFEERVSKLFSQA NSFSIAYIN QCLGQAGIRCKI QD YFA CLGAK EGENEAETTPD I FDQIAEAYHGAAPILNARPSSHNLAQ DIEKVKAI K ALL DALKLRLQRFVKPLLGRGDEGDKDNF FYGDFMP I WEVLDQ LTP LY NK VNRNRMTRKPKYSEK DALKLRLQRFVKPLLGRGDEGDKDNF FYGDFMP I WEVLDQ LTP LY NK FDANNVETIGDCYK MI YKLLPGPNKMLPKVFFSKSRVQEFSPSKKILEIWESKSF KGD NPFLDCCHALDFYK DSIAKHPDWNKPNFKPSDT QSYT NI SDFY RDVNQG GYSL SFTK SVDYVNRMVDEGKLYLQIYKDFSPSKGTPNMHTLWYRMLFDERNLHNV I YK LNG EAEV F YRKASLRC DRP THPAQ P ITC KNE ND SKRVC VFDY DII KN RYTV D KFMPH VPIT INYKCTGSDN IN QQVCDY LRSAGDDTHI I GIDRGERNLLYLVII DQHGT I KEQFSL NEIVN EYKGNTYCTN YHSLL EKEAGNKKARQDWQ TIES I KELKEGYLSQV I H KI SMLM QRYH AIVVLEDLN GSF KQPSGFLF YI PAWNTS KIDPV TGFVN LFDTRYCNA EAKAKE FFEKFD I S YN DERDW FEF SFDYRHF TNKPTGTR TQWTLCTQGTR VRT FNR PEKS NHWDNEEFD LTQAF KDLFNKG YGIDIAS GLKARI VNGQLT KETS AVKDFY ESS LLK LTL QMRN S V GTD IDYL VSPVADKG I FFDS RT CGS L PANADANG AFNIARKGMLL RQI QOS SIDA EK I QLAPI KN EDWLEFAQEKPYL
SEQ ID : 54	MEKEITELTKIRREFPNKKFSSTD MKKAGLLKAECPDA VRDFLNSCQE II GDFKPKVKT NIVSISRPFEEWPVSMVGRAI QEY YFPLT KEELESVHPG TS SEDHKSFFN ITG LSN YN YTS VQGLNLIFKNAKAIYDGT LVKANNK NKKLEK KFNEI NHKRSLEGL PI ITPD FEEPF DENG HLNNPPG I N RNIY GYQCAAKV FVPSKHM VSLP KEYEGY NRDPN LSLAG F RNR L E IPE GEPGHVWFQRM DIPEG QI GHVN KI QRFN FVHG KNSG KV FSD KTRGV KRYHHS KYKD ATKPYK FLEES KKV SALS D I L AI ITIG DDW VVF D I RGLYRN F YRE LAQ GLTA VQL LD FTGDPV IDPKKG VVT FS YKEGV VPV FS QKIV P R FK SR DT LEKL TS QGP VALL SVDL GQNE PVA ARVC S LK NIND K TLDN S CRISPL D DYK QI K DYR DS LDE LEI KIR L E AINS LETN QO

TABLE 1-continued

Amino Acid Sequences of Exemplary CRISPR/Cas effector proteins	
SEQ ID NO:	Programmable Nuclease Amino Acid Sequence
	VEIRDLDFVFSADRAKANTVDMFDIDPNLISWDMSMDARVSTQISDLYLKNGGDESRVYF EINNKRIKRSDYNISQLVRPKLSDSTRKNLNDSIWKLKRTSEEYLKLSKRKLELSRAVVN YTIRQSKLLSGINDIVIILEDLDVKKFNGRGRIRDIGWDNFSSRKENRWFIPAFHKAFSEL SSNRGLCVIEVNPAPWTSATCPDCGFCSKENRDGINFTCRKCGVSYHADIDVATLNIARV AVLGKPMSPGPADRERLGDTKKPRVARSRTKMRKDINSNSTVEAMVTA
55 SED ID :	MIKPTVSQFLTPGPFLRIRNHSRTAGLKLKNEGEACKFVRENEIPKDECPCPNFQGGPAIA NIIAKSREFTEWEIYQSLSIQEVIPTLPDKDLPPEIILKEEWRAOWLSEHGLDTVPYKEAA GLNLIIKNAVNTRYGVQKVVDNKNKNLAKINRKNEIAKLNQEAEISSEEIKAFDDKGY LLQKPSPKNSIYCQSVSPKFITSKHNVNLPEEYIGYYRKSNEPIVSPYQFDRLRPIGEPE GYVPKWQYTFLSKKENKRRKLSKRIKNVSPILGIICIKKDWCVFDMRGLRTNHWKKY HKPTDSINDLDFYFTGDPVIDTAKANVFRYKMENGIVNYKPVREKKGKELLENICDQN GSCKLATVDVGQNNPVAIGLFLKKVNGELTKTLISRHTPTIDFCNKITAYRERYDKLES SIKLDIAKQLTSEQKIEVDNYNNNTPQNTKQIVCSKLNINPNDL.PWDKMI5GTHFISEKA QVSNKSEIYFTSTDKGKTDVMKSODYKWFQDYKPKLSKEVRDALSDIEEWLRRESLEF NKLSKSREQDARQLANWISSCMDVIGIENLVKNNNFFGGSGKREPGWDNFYKPKKENR WWINAIIHALTELSONQKGRVILLPAMRTSITCPCKCYCDSKRNNGEKFNCLKCGIELN ADIDVATENLATVITAQSMPKPTCERSGDAKKPVRARKAKAPEFHDKLAPSYTVELRE AV
56 SED ID :	MEKEITELTKIRREFPPNKKPSSTDMKKAGKLLKAEGPDAVRDFLNSCQEIIGDFKPPVKT NIVSISRPFEWVPSMVGRAIQEYYFSLTKEELESVHPGTSSEDHSFFNTGLSNYYNTS VQGLNLIFKNAKAIYDGTIVKANNKMKLEKKFNEINHKSLEGLPIITPDFEEPFDEN HLNNPPGGINRNRYGYQGCAAKVFPVSKHMKMVLSPKEYEYEGYNRDPNLSLAGFRNRLEIPE GEPGHVWFQRMIDIPEGQIGHVNKIQRFNFWHGKNSGVFKFSDKTGRVCRYHHSKYKD ATKPYKPLEEESKVSALDSLAIITIGDDWVVFDIRGLYRNVFYRELAKGLTAVQLLDL FTGDPVIDPKKGVVTFSYKEGVVPVFSQKIVPRFKSRTDLEKLTSGQPVALLSVDLGQNE PVAARVCSLNINDKITLDNCRISPLDDYKKQIKDYRDSLDELEIKIRLEAINSLETNQQ VEIRDLDFVFSADRAKANTVDMFDIDPNLISWDMSMDARVSTQISDLYLKNGGDESRVYF EINNKRIKRSDYNISQLVRPKLSDSTRKNLNDSIWKLKRTSEEYLKLSKRKLELSRAVVN YTIRQSKLLSGINDIVIILEDLDVKKFNGRGRIRDIGWDNFSSRKENRWFIPAFHKAFSEL SSNRGLCVIEVNPAPWTSATCPDCGFCSKENRDGINFTCRKCGVSYHADIDVATLNIARV AVLGKPMSPGPADRERLGDTKKPRVARSRTKMRKDINSNSTVEAMVTA
57 SED ID :	VPDKKETPLVALCKKSFPGLRFKKHDSRQAGRILKSKEGAAVAFLEGKGTTQPNFKP PVKCNIVAMSRSRPLEEWPIYKASVVIQKYYQAQSYEEFKATDPGKSEAGLRAWLKATRV DTDGYFNVQGLNLIFDHTLGRPGINCSCVFGYQHMKLKVPGSIPGVTGYSRDPSTPIAACGVDR EIPETALDETGHLRHRCGQMLKPVPGSIPGVTGYSRDPSTPIAACGVDR EIPEGQGYVPPWDRENLSVKHHRRKRAASWRSRGGAIDDNMLLAVERVRAWDWALLD LRGLLRNTQYRKLLDRSVPVTEISLLNLTNDPTLSVVKPGKPVRTATLIYKQGVWP VVKAKVVKGSYVSKMLDDTTETFSLVGVDLGVNNLIAANALRIRPGKVERLQAFTL EQTVEDFRRFRKADQHQNENLRLAVERSLLAEQQAEVLAQDFTGPEQAKMQVCGHGL LSVDEVWDKVMSRSSILSDLAKERGVDDTLYMFFFFKGKGRKTEIRKRWDVNW QHFRPQLTSETRKLNEAKWEAERNSSKYHQLSIRKKEELSRTCNCVNYVIRTAEKRAQCG VIVAVEDLHHSFRRGKGSKRGSGWGGFAAKQEGRWLMDALFGAPCDLAVHRYGIV KVDPYNTSRTCPCEGCHCDKANRDRVNREAFTICVCCGYRGNAIDVAAVINAMAITGV SLRKAARASVASTPLESLAEE
58 SED ID :	MPKPAVESEFSKVLKKHFPGFRFRSSYMKRGKILAAQGEAAVAYLQGKSEEEPPNFQ PPAKCHVTVTSRDPWIMKASEAIQRYIYALSTTERAACKPGKSSESHAAWFAATGV SNHGYSHVQGLNLIFDHTLGRYDGVLKKVQLRNEKARARLESINASRADEGLPEIKAEE EEVATNETGHLLQPGPQINPSFYQYTISPQAYRPRDEIVLPPEYAGYVRDPNAPIPLGVVR NRCDIQKGCQYIPEWCREAGTAISPKTGKAVTVPGSPKKNKMRMRYWRSEKEKAQD ALLTVTRGTDWVVIDVRGLLRNARWRTIAPKDISLNALLDFTGDPVIDVRRNIVTFTY TLDACGTYARKWTLKGQTATLDKLTTATQTVALVAIDLQTNPISAGISRVTQENGAL LQ CEPLDRFTPLDPLLKDIISAYIARDRNEEELRARSVEALPEAQQAEVRALDGVSKETAR TQLCADFLGDPKRLPWDKMSNTTFISEALLNSNSVRDQVFFTPAPKKGAKKKAPVEVM RKDRTWARAYKPLSVEAQLKNEALWALKRTSPFEYLKLSRKKEELCRRSINYVIEKTR RRTQCQIVIPVIEDLNVRFFHGSGKRLPGWDNFFTAKKENRWFQGLHKAFLDLRHSF YVFEVRPERTSITCPKGCHCEVGNRGEAFQCLSCGKTCNADLDVATHNLTQVALTGK TMPKREEPRDAQGTAPARKTKKASKSKAPPAREDQTPAQEPSQTS
59 SED ID :	MSNKTTPPSPLSLLLRAHFPGKFESQDYKIAGKKLIRDGGPEAVISYLTGKGQAKLKDV KPPAKAVIAQSRPFIEDLVRVSRQIQLQKELIFGIPATKGRPQKQDGLSETAFNEAVASLEVD GKSKLNLNEETRAAFYEVGLGDAPSILHAQAQNALIKSAISIREGVLKVENRNEKNLSTK RRKEAGEEATFVEEKAHDERGYLIHPPGVNTQIPGYQAVVIKSCPSDFIGLPSGLAKES AEALTDTYPLHDRMTIPKGQPGYVPEWQHPLLNRRKNRNRDWDYASLNPKATCSKRS GTPNKRNSRTDQIQSGRFKGKAIQVLMRFQDEWVIIDIRGLLRNARYRKLLKEKSTIPDLLS LFTGDPSIDMRQGVCTFIYKAGQACSAKMVKTKNAPEILSELTKGSPVVLSIDLQQTNP IAAKVSRVTQLSDGQLSHETLRELLSDSSDGKEIARYRVASDRRLDNLANLAVERLSP EHKSEILRAKNDTPALCKARVCAALGLNPEMIAWDKMTPTEFLATAYLEKGDRKVA

TABLE 1-continued

Amino Acid Sequences of Exemplary CRISPR/Cas effector proteins	
SEQ ID NO:	Programmable Nuclease Amino Acid Sequence
	TLKPKNRPEMLRRDIKFKGTEGVRIEVSPPEAAEAYREAQNDLQRTSPEYRLSTWKQEL TKRILNQLRHKAAKSSCCEVVVMAFEDLNITKMMHGNKGWADGGWDASFIKKRERNRW FMQAFHKSLTELGAHKGVPTEVTPHRHTCTCKCGHCDKANRDGERFACQKCGFVAH ADLEIATDNIERVALTGKPMKPESERSGDAKSVGARKAAFKPDEEAEEAEE
SEQ ID: 60	MSKTKELNDYQEALARRLPGVRHQKSVRRAARLVYDROGEDAMVAFLDGKEVDEPYT LQPPAKCHILAVSRPIEWEPIARVTMAVQEHVYALPVHEVEKSRPETTEGSRSAWFKNS GVSNHGVTHAQTLNAILKNAYNVYNGVIKKVENRANKRDSLAAKNKSRERKGLPHF KADPPELATDEQGQPLVPLHREKLTSNKHRRMKLPKSLRAQGALPVCFRVFDWAVVDRGLL LAIPPGQPGYVPLHREKLTSNKHRRMKLPKSLRAQGALPVCFRVFDWAVVDRGLL RHAQYRLAPKNVSIAELLELYTGPVIDIKRNLMTFRFAEAVVEVTARKIVEKYHNK LLKLTEPKGKPVREIGLVSIDLNVQLRILAIYRVHQGTGESQLALSPCLHREILPAKGLGD FDKYGKSFNQLTEEILTAAVQTLTSAQQEYQRYVEESSHEAKADLCLKYSITPHELAW DKMTSSTQYISRWLDRDHGWNASDFTQITKGRKKVERLWSDRWAQELKPKLSNETRRK LEDAKHDLQRANPEWQLAKRKQEYSRHLANTVLSMAREYTAETCVVIAIENLPMKG GFVDGNGSRESGWDFNFTTHKKENRWMKDIHKALSDLAPNRGVHVLEVNPQYTSQTCP ECGHRDKANRDPIQRERFCCTHCGAQRHADLEVATHNIAMVATTGKSLTGKSLAPQRL QEAAE
SEQ ID: 61	VAFLDGKEVDEPYTLQPPAKCHILAVSRPIEWEPIARVTMAVQEHVYALPVHEVEKSRP ETTEGSRSAWFKNSGVSNHGVTTHAQLTNAILKNAYNVYNGVIKKVENRANKRDSL AKNKSRERKGLPHFKADPPELATDEQGQPLVPLHREKLTSNKHRRMKLPKSLRAQGALPVCFRV VVDPRSPIPSILPIDRLAPIPGQPGYVPLHREKLTSNKHRRMKLPKSLRAQGALPVCFRV FFDWAVDGRGLLRHQAQYRLAPKNVSIAELLELYTGPVIDIKRNLMTFRFAEAVVE VTARKIVEKYHNKYLKLTTEPKGKPVREIGLVSIDLNVQLRILAIYRVHQGTGESQLALS PCLHREILPAKGLGDFDKYKSKFNOLTEELTAAVQTLTSAQQEYQRYVEESSHEAKA DLCLKYSITPHELAWDKMTSSTQYISRWLDRDHGWNASDFTQITKGRKKVERLWSDRW AQELKPKLSNETRRKLEDAKHDLRQRANPEWQLAKRKQEYSRHLANTVLSMAREYTA CETVVIAIENLPMKGFFVGNGSRESGWDFNFTTHKKENRWMKDIHKALSDLAPNRGV HVLEVNPQYTSQTCPECGRDKANRDPIQRERFCCTHCGAQRHADLEVATHNIAMVAT TGKSLTGKSLAPQRLQ
SEQ ID: 62	MICKPSNRHALPKVIIISKVDNQNLIEFKIKYKQLSRLDRVEIKTMHYDDRAIVPDEVIIING GLIDVEYRNNDHKTIFVKVGDKSYSISGQKVGKGERLLENRISQTKVQLELKDEATNRVS KTERELIVDDNIKLYSCIVGRDVKTIDYIYLICRFLGYRS DLLFYGGFVNFFFHVANNRP EFWKIDFDNDNRNSKLIEYIFTINDHLKNDENYLKDYISDRGQIVDDLENIKHFSALRHG LMHFYDFFEALFGNGEDIDKMDNQNTQPLSSLNIFKFLDIMIDKLDKLNIDITKKEFIDAE KITIFGEEELSLAKLRYTAHTAINRVAFNKLINSFIENGVENQSLKEYFNQQAGGIAYEID IHQNREYKLNLYNEHKKLVSRVLISIDGQETATLNQKIVELKEQMKQITKINSIKLEYKL LAFGFIYTYEKYNEEFKNSFDTIDKNGRFTPKEDEGNKRAFDFSRELEHLKGYYKATLQT OKPQTDKEEMEEVKRVLKSLI GDDTLLKFIILLMFTMPQELKGEFLGFIKKYHDT KHIDQDTISDSDDTIEEGLSIGLKLKILDKNIRSLSIKHSLSFQTKYNKKDRSYYEDGINH GKFFKKLGISHNQEEFNKSVYAPLFYSSALYKLINDFEYTTLSLHIVGNETLSDQVNKP QFLSGRYFNFRKLLTQS YNISNNNSTHSVIFNAVINMRNDISHLSYEPLDCPLNGKKSYKR KIRNQFTINIKPLVRSRKMIDFITLQTDMQKVLCGCAVNDFTMKIVQLRTRLKAYANK EQTIEKMITEAKTPNDYNYIKVKGVEAINKYLLEVIGETQVEKEIREEIERGNIANS
SEQ ID: 63	MLKKPKSNSRYALPKVILSTVDHEKILEFKV KYEKLARLDRLVVERMHEDGESVVFDEVIA NSGDLIEAYQDHDHRKLLIQAAGKSYTTGKVKVGKRRKLEERISRAKIQLTLDQGQEDQ HRRIRATVTEKALPEKEDRDKIITSKESIYLVKRFLSYRSDLLFYFFFVDNFF KVGNNKQELWIKIKFQNQPELIEYFRFIINDRFKNAKNDKFDNYLKNDKAIQEDLEKIQK VFEKLRHALMHDYGFFEKLFGGEDQGF DLDIAFLDNFVKKIDKLNIDITKKEFVDEKI KIFGEDLNLADYLKLYASISINRVGPNRVMNEMIKDGIEKSELKRAFEKLLDKTYALDIH SDPSYKLLNLYNEHKRLTEVSTYTDGNKIKEGNQKIAKLYEMKETTKNALVRECKM RLAFLGLIYGRYDTHEAFKNGPDTDLKRGEFAQIGSEEAIYFNTT FEKS KPKSKEEIKKIA RQIDNLSLSTLIEDDPLMKFIVLMLPFLVPRELKGEGFLGFWRKYHDIHSIDSAKSDEMP DEVSLSLKLKILTRNIRRLNLFPEYLSSEKIKYSPKNTQFYTDKSPYQKVKYRKLISHNKEE FDKTLVPLFRKSMQARNDVIDFVLKYEKEIKAVL AHKVSQGDSKKHFGIRGEIAHINTKDLIYDPLFRKSMQARNDVIDFVLKYEKEIKAVL GYDAINDFRMKVQQLRTKLKVYSDKTQIEKLLNEVEAPDDFYVLYKVGVEAINKYL LEIVSVTQAEEEIERKIITGNKRYNT
SEQ ID: 64	MIKSYDDFTKLQLQKTLRFELPKIGKTLEHIKKSEIIESDETLEEQYAIKVNIIDRFHRKHI DEALSLVDFTKHLDTPKTQIPELYLKRKGTDREKKELEELSADLRKLIVSYLKGNVQKQT OHNLNPICKERFEILFGKELFTNEEFFTLAENKEEKKAQAFKGFTTYFKGFQENRKNMYS BEDKSTAIAYRIINENMPLFIENIARFQKVLVDVIEKTKLTELKQNLKTELKGHSVSDIPRIE YFNNVLTQEGISRYNTIIGGKTTTEGVKIQGLNEIILNLHNQSKDVKLPILKPLHKQILSEE YSTSFKIKAFENDNEVLKAIIDTFWNEHIEKSIHPVTGKRNINLLKIECNLCKKLEKYKDI EKLFIERKNLSTVSHQVYGOQWNIIRDALRMHLEMNNKNIKEKDIDKYLNDNAFAWKEI KDSIKIYKEHVEDAKELDENGIVKYFSSMSINEEDDEKEYSISLIKNINEKYNVNKSILEED RTGKSDLHQDKEKVAIIKEFLDSLKQLQWFLKLLYVTVPLDEKDYEFYNELEVYYEALL PLNSLYNKVNRNMTRKPYSEVKFLNFSYPTL LDGWDKNKETANLNSIILKKNGKYYLGI

TABLE 1-continued

Amino Acid Sequences of Exemplary CRISPR/Cas effector proteins	
SEQ ID NO:	Programmable Nuclease Amino Acid Sequence
	MNKENNNTIFENFPKSKSNDYYEKMIYKLPLPGPNKMLPKVFFSKKGLEYYKPSKEILRIYE KGEFKDKDSGNFKKESLHTLIDFYKEAIAKNEDWKIFKFKEPKNTREYEDIQSQFYRDVEEQ GYLIIFEFVKDANYVDKLVEEGEYLFPQIYNNKDFSENKSCKSGNPNLHTIYWESLFDNQNL KDVVYKLNGEAEVFYRKSIIDYPEEIYNNNGHHKEELNGKFNYPIIKDRRTYQDKFLFH PITMNFISKEEKVRNLQALACEYLSITKEDVHIIIGIDRGERHLLYLSLIDKEGNIKKQLSLNTI KNEYNDKEIDYRVKLDEKEKKRDEARKNWDVIENIKLKEGYMSOVIHTIAKMMAEKK AIIIMEDLNIGFKRGRFVKEKVQYQKFKEKMMIDKLNLYLVFKNKEPLEPGGSLNAYQLTS KPDFSKKKLGKQSGFIIVYVPSATSKIDPPTGFYNFIRVDVPNLEKGKEFFSKFEKIINYNTKE DYFEFHCKYKGVPPEPKNDNDRKTKESSLTYYNAIKDTVWVVCSTHERYKIVRNKAG YYESQPVDVTKNLRDIFSEANINYSQDGKD1KPIIESNNAKLLK5IAEQLKLILAMRYNN KHDDDEKYIILSPVKNQKGKFFCSLDGDQSLPINADANGAYNIALKGLLLIEKIKKQG KAKDLYISNLEWMFMMSR
SEQ ID: 65	MKSIFFDNFTGLYSLSKTLRFELRPVGQTLENIKNGHFLESDDKMMADDYQDVKKIIDNYH KFFIDDVLKGASFDWALLEKELTDFNKKNTDDSKVEAEQKQLREQIAKTLAGDKRFSL TASTPNDFNKKDQDFGWLQEQQSVKEIRKDALKDALTFKKFSSYFKGQENRKNVYSAADDIPT AVPYRIVNDNPKFLQNIISIPTQIEKCPQVIADVENELASYLGKEKLADIPTVQAFNKL CQGGKENQRGIDFYNQVNGGTAEEKGGVNLRGINQFLNLYWQQHDFAKENRRKIMVP LYKQILSDRSSLSPKIEITIDTDEELKTAISEYADKLESNSNDEKKSVLDVCVELFDSIKEQN LQEIQVNRKDINNISRLTGDWWSLQSRMNLIADEVFTTKAEKTRWQKSVDGDEGENK SKGAYSLAELNRVLEYASENAETDIRITDYFCHRNRFYYEKEESGLFKQGEELVALSIE SCEDIL-SKRKAMNEAFANISSENSLSDNSEDIAKTTKYLDSVQDLLHRIKPLKVNGLGDP SFYAVFDSIYSALEVISIYVNTKRNITYTRKAESPKEKYKLFDNPTLANGWDLNKEKDNTC VLLRKANGMYYLGIMNPDKPKPAEKYDCGTECYEMIYKLPLPGPNKMLPKVFFSTKG KKQYNPPENILHGKHTKQVAFDINFCHELIDWFKSAINQHEDWKKPGFKFSDTKS YKDIISDYREVTEQGYKLTFINIPESIISKMVSEGKLYLPQIYNNMDFAPGANGMPNMHTL YWKNLFSEENLKDVLVLLNGEAELFYRPAGIKEPVVHAKGSYLVNRTKDGEPPIPEKIH DEIYRNANGNLNNLSEAKEYKESHVKVIQASHEIIKDRHYTEPKFLFHPILTINFKAPS MKPTVTAIQNEVNVRFLKNNPVDNIIIGLDRGERNLILYSLINQKGEIIIQQFSFNDVREEQ NGQTVKVNHYHEKDLQREKERDAARKSWQAIGKIAELKEGYLSAVIHQLTKLMVEYNAI VVMEDLNFGFGRGRFHVKEQVYQKFEHMLIDKLNLYLVFKDRLGLNEPGGAINGYQLAG QFESPKQLGKQSGMLFVYPAAYTSKIDPKTGFVSMMFNKDLTNVHKRDFFSKFEDIHF DEATCSFVFTFDYKNINGKAKEEMKQTKWAVYSREKRIVYFSKTSSEDIMPTEKLRAL FESDGIEYKSGNNHDSVMAVGADLKEGAKPSKEIADFWDGLYNFKLILQMRNSNAKT GEDYIISPVMASDGTFFDSRVEAKKGDAKLPLDADANGAYHIALKGSLINKINLAGE DELKKFDMKISNEDWFNFAQEKKYAE
SEQ ID: 66	MTKQNKSVTQNTKRKNFGEFTNLYSLSKTLRFELPKVATKTIEVERKDRENFKKDRK IAKNYQKLKGILNELLQEFIQDVMRERFSFQKEIKEFEERYLEALNFKEDKNDYKKRTQLK NAYEKVAKKLAGKIATAFGKYNQEKYGVFKTPKKNLGENVFDILEGKYKGDKKILGIH TFKFKPTEKEEKKQGKEAVNFTYLGFFNQNRENFYKGEMKAGQFATRTIENLIOFLKNN KLFIDKYKQKCEKTRCKANGEKFNKSDYPIFKELYQIGS1KKNDVYVTEIKSDDELVNL QSLPEKTANTLREVQKFYENFFDKI FNDEFDLDDK1YLPKSVGTHFSHLAFSDWSKLA VFKRWRNEKVKIKEGEDVNVQSRSLADIKKRMEEILEMDGGSFGKTYCQKVGLEKEAR TIEDWVSGFWKI1QYHINSQF1GGEVFDKEKKDDKTEKIQTIDDLQEEYLQATEMYRE RMVESEGLNDEEKEIITKLKNYLDRIKDIERIARYFDLRLKHFDIDEASKDGFYIY QUELLQDISEAKINDHYNEIRNYLTAKANVVDDFKFLNFDGQTLSGWDLNKETEKFSLIF KRKVGDGVVEYYLGIINKEKNTI FD KKKHPEIFTENSEFEKMEYKLFPSKMLPKIAFTK NKEGERIIPVFLDENAGKEIAQ1KKEFALPQDAKKEDKNNWSDEFDRKKLNKLIDYYKL VLEKHPKYMOTFNFVFKASSAKYKNLGEFNDVVARQNYVTKFVSVSDKDYIDQKVESGE LYLFKIHKNKDWNLTKAGDTKKQSKKNLHTIYFEELFSEKNAIEPVFKLSSGAAEVFRDAI EKKKQKKKDKKGKEILEKFRFTKNNKILFHVPITINYGPBINSQGQFNQKINEPIADNSRS VNILGIDRGEKHLLYVSVDNNNGKIIKSGSLNEINGVDYHEKLDKAERQEARSKSWQK INQIKNLKAGYISQV1KIVD1AIENNAAIVLEDLNFGFKSFRQKIEKNVYQQFEKALIDKL GFVTDKEKLNRQAPQLSAPFESFEKMGKQTGIVFYVLAINTSCKVPCQCWKKNIFPHY STKKSIAENLQKQYKMKMRYWRENEENRFEFEYKGDGKDEFSSIFSNDVDRVRYDKRANN NQGGYVIYQIDSTTKEKDRNIEKEKSITNLLKELLLEKFEIDNLEGELLVKLSEKSPDVSK ETIKDFGFLNLSIINRNSMTDTEEDYIQCPCAGFDTRKENKIG1KNGDNGAYNIALGR FLIERIKKAKKEDKKPNLTSNNNDYFQOWREFVK
SEQ ID: 67	MRTTTSLSDAFTNRYALSSTLRFELKPIGNTQMMLEQNNVFAKDRAIREKYEKTKPWIDL LHREFVAESLQNAQLGNLDDYYAALQNVQKITKDTNAEDKKRWKKFKEQKRLRKE VVALFDKAHHIATQRYPQLKKKTKDFLFEEGVFHVLFARYGSAPDTTVKIVTSNPET GEVIDEREESIFKGWKGFTGYFDKPFETRKNFYKDNGTATAIATRAINQNLRRFAENMQ KLTDIKNNYPPELLAHTDFGDPDIAHAQSLDFYARTCLLQEGIDAYNKFKFVGVLKSAINE YQQQKNGKVRISYPKTLDNQILGERERRLFVIEDDRRELHDVFRAPVDDGTVFAAEMRQL AQAFSAQNGTYDYTQIYISKKGFETISRKYTHDRAWHALADVFKAKAKRIATTASG EKKFPAYIIVPVAITQALTLVQESEDTECTWKERYASITENKTLEEGFIAFADEPERLFWH MEATVQDTDYVVAEDKAKKLLSDGQITKNEQTTQIKEYADALLRIYQMAKYFAVEKK SMWDDAVAIDDTYETFKEIYGNTHSTIVASYNLLRNLYLTKKPWEVQWKLNFBENPT LLDGWDKNKEAANFGVILRGDKFYLGIIMRKGHNNIFANQHHSNPEGQGLQKMYKF

TABLE 1-continued

Amino Acid Sequences of Exemplary CRISPR/Cas effector proteins	
SEQ ID NO:	Programmable Nuclease Amino Acid Sequence
	PPDPKKMPKVCFSAKGMEFFPAPSEEIVRIYKNAEFKSGDTFNVESMQKLIDFYKNALQ KYDGWKIYDFKHLKDTAQYTSNIGEFPYDDVAKGGYQLGWQNISKEYVEEKANGELY LFQIKNKDWNDGATGRKNLHTLYFEYLFSEKNAAADFVFRLLNGGAEVFYRPAAIESKTE RRGNREVAKKRYTQDFKVLHPITLNRTAGDVTKTSASFNDAVNRFLAGNPDINIMGDIR GEKHLLAYSIIDQNGNRIVSGSFNTIGSKDYHALLTERQGAREEARKNWRQVEQIKDLK KGYISLVVRERADLAIAKHNAIAVLENLNMRPKQIRGIEKSVYQOLEKALIEKLNFLVNGK EVDATKAGHLLRAYQLAAPFTFEKMGNQGTGIFYTTASYSQVDPVTGWRPHVYLKY RNAQKTKDEDILRFIDIVENDEKQRPEFAVYRHNGWSWTVCSSVERHRWNRSNNAKGGG YDVFVPEGEGSITQRQLQEACAQRGIDTTRNLAQIDELDESASATVSFLRDLCFYFR利CQI RNTDDGADDINAQDFLMSPVPEFFD'TRNAQBEQYPQNGDENGAYNIARKGIIILQKITAW GRSQDTQRYYRPTDFVSQDEWDTFLTQHTT
SEQ ID : 68	MLKEKQFKTGFDTNLYELSKTLRFELRPTPETKDLLDKNKKIQTDKKIAENYQEIKKYF DKLHKPKIKEALSNTQIDFSDFCKLWQNKSDSGKIKDLSRKLRKSIAKQAFDKKGADWH KRYLEKGIGLKKKNLDILFEERVLDILKEEFKDDVVKLFESFKGFSTYPTNFHESRKNF YKDDGTTASAIATRIIDENLKRFCDCNIVKVKHSKKLISELNEREAKIFEADEFYNRCLLQOQI DDYQVIGDINKINNLRQKNTPTLKLQYKQILGQDVRQETEQDAFIEIKNNEEVPDF LQDFIKHSDENNKFKNLQYKPIEGKHSLDKIFLAKRFVNTISGKWFASWEVFGAElikk FGNKKDLPDFIPFAAVKDVQLQNCNIPANELPKEKIKIKNDEDKNYYDIFINLWKEEDSNLK KVEESKEVENMIAEDKVYSNKKEKRKNDNGEEIEIIEQEKIKNYADAAMNIIFRMK YPLLEKNGKTVEGMGEDNNFVNELNIVFKGGEIDGKVYEGVKTYLYNNFERNYLTKKP FNEEKTKLNFDCGQILSGWDKNKESKLGVLIRLKDKNKYLAIINKHHNKIFPDVKNSYA YIVGDNFYEKMEYKLPDAKRMIPKIAFKNNKEFGWTDEIQKIKNEYAEFQEGKKN DKNLWKDKNKMKELITYQQCLEKGGYKDIIYRNFRWSPDKYSGIGEFPNDEIDRQS YCLKFVVKDFVQLKSGEYLYLFQIYKNDPSDKADRAQKENIHTEYFKLLFDQRNLD NVVLKLGGAEIIFYRKTEGLPKKDNKNGEVNRHRYADDKYFLHLPIQLNFGRGNL SGGEFNSKINQYLSEQREIKIIGIDRGEKHLAYYSVINQDGKIBEIESLNTVNGIDYRKLD ELEKKRQEKRWSQSIKIKDLKKGYISHVIKKICDLAIEHNIAIVFEDLSGGFKNSRKKIE KQIYQNLERELATKLNNLTFKDKNFGESEGHYLNAQOLAPKIDNYQDIKMQTGIVFYTPA GTYSTSCTPQCGRKLTFKDFDTATSKAEDLIRGSKLNVFKEKEKNRFKINYLFPNPIEKKKK KIKENELFADAGAKNEFTIYSDVKKIKWHTGTRKLEEAEGERLLENKNRGRDKEYDI NKCLTRLFRENKIDVNGDIIGQITKIKSLKLYQDLYFLATLIRNNVSGSDIDYIQCPC HFHSDFGQQKQFNGDANGAYNIARKGILILKKIKQFAAQDKDMKNFGWKLTVNDINE WDKFTQK
SEQ ID : 69	MNKNFSNSTELEYTLSKTLRFELKPVQAQTKENIKKGKFLESDDKKADDYKDVKKIIDNYH KFFPDSLVLKNAKDFVLEKEMSDFNKSKADDKSKVEAEQKLLRQDIACKLTSDKRFKA LTASTPSDLFKDKDFIDWFTONSTKDINKEALETFKRFSSYFKGQENRKNVSAEPIPT AVPYRLVNDNFPKFLQNLALKFIIQPKCPQVISDVEKELASYLGEKEKLADIFTQAFNHY LCQGGKENQRGIDFYNNILGGIAEKEGGINLREGINQFLNLYWQHQDFAKQNKRKIMP LFKQIILSDRSSLFSKIESINTDQELLTSITEYADKLETSKNDEKKSVDLICSDLFASFIKAQNL QEIVYVNKRKDINSRGTWQSWLQSRMNVYADEVFTTAKTRWQKSIDGEENKSK GVFSLAELNSVLEYSENSETDVRITDFFDHRNRYYEKESGLFKQGDELVALSIRECE DILAKRKAMDEAFANVSENNLSRDNSEDVAKIKIYLDVCQVELLHRIKPLKVNGLGDPAF YAVFDTYVNSLSEVIISLYNKTRNYITKKAANPEKYKLNFDNPTLADGWDLNKEQANTS VLMRKDGMYLGQDKPKFAEKYECNGNEACYEKMUYKQFDATQKIPKCSTQVKE VKKHFQSGATDSIILNDSKSFKFLDVITKFWLNNHVVNGEKFVPKRESNETRPKKFQI GYYKQTGDLGGYKEALNIWISFCKTFLQSYISSLSSYIYDYDFKESNSYDSLDEFYNYLNATC YKLSFINIPEATISQMVSEGVYQKQYLNKDFAPGASGMPNMHTLYWKNLFSEENLKV VLKLNQEAEFLYRPGAKIPEVKLYLQFQIYKNDKDFAPGASGMPNMHTLYWKNLFSEENLKV VKLKLNGYLLFQKJLQYKQFQIYKNDKDFAPGASGMPNMHTLYWKNLFSEENLKV EATEYKASHKVVIEKAHDIIKDRHYTEPKPLFHVLPTINFKAQGNSYINENVRFLKNN PDVNViGLDRGERHLIYLSLINQKGEITIYKQFTFNEVERNKNGQVIKVNVHEKLDQREKVR GAARKSQAQIAKIAELKEGYLSAVIQLQKTLKLMVEYNAIVVMEDLNFGFKRGRFHVKEQ VYQKFEHMLIDKLNLYVFKDRGLTEAGGVNLNGYQLAGQFESFQKLQKGSQMLFVPA GTYSTSCTPQCGRKLTFKDFDTATSKAEDLIRGSKLNVFKEKEKNRFKINYLFPNPIEKKKK KEEMKRTKWSVSYSKDKRIVYLSKTSKSYEDVQPTEKIKASLESVGIEYMSGNNLIDSIMVI GAEALKDGAPEKSKIEADFWDRLLYNFKLIIQMRNSNAKTFGEYIISPVMADGGTFDSRE EFKKGENAKMPVADANGAYHIALKGLSLLKRFDAASENELKKFDMKISNVWDWFKFA QEKSYYAE
SEQ ID : 70	MKAKKSFYNQKRKFGKRGYRLHDERIAYSGGIGSMRSISIYELKDSYGIAGLRNRADIAT ISDNKWKLYGGINLNDYLEWRSSKTDQIEDGDRESSLLGFVLEALRLGFVFSKQSHAPN DFNETALQDLFETLDDDLKHVLDKWCDFIKIGTPKTNDQGRLKKQIKNLLKGNKRE EIEKTLNESDDELKEKINRIADVFAKNKSDDKTYTIPKLDKPNTEKYPRINDVQVAFFCHPDF EEITERDRTKTLQDLIINRFNKRYEITENKKDDKTSNRMALYSLNQGYIPRVLNDLFLFVK DNEDDFSQFLSDDLENPFSSFSNEQIKIIKERLKKLKKYAEPIPQGKQPLADKWDYASDFGG KLESWYSNRRIEKLKKIPESVSDLRNLLKEIRNVLKQNNASKELELSQKIIYEIRDYGVSF EKPEIIFKSWINKTQKQKVFYVAKMADREFIEKQLDLWADLRSQNLNEYNQDNKVSF KKKGKKIEELGVLDALNKAKKNKSTKNNENGWQQLSESITQSAPLFFGEGNRVRNEEV YNLKDLLFSEIKNVENILMSSEAEDLNKNIKIEYKEDGAKKGNVYLNVLARFYARFNEGD YGGWNKVKTVLENIAREAQDFSKYGNNNRNAGRFLYNGRERQVFTLIKFEKSITVE KILELVKLPSLLDEAYRDLVNNENKNHKLVDVIQSLSKTIMALVLVLSHSDKEQIGGNYIHSK

TABLE 1-continued

Amino Acid Sequences of Exemplary CRISPR/Cas effector proteins	
SEQ ID NO:	Programmable Nuclease Amino Acid Sequence
	LSGYNALISKRDFISRYSVQTNTGQCKLAIGKGKSKKGNEIDRYFYAFQFFKNDDSKIN LKVIKNNSHKNIDFNDNENKINALQVSSNYQIQFLDWFFEKHQGKKTSLLEVGGSFTIAE KSLTDIWGSNSPVRGVFKRSDEEKRVFSQPFTLIPDDEDKERRKERMIIKTKNRFIGIDIG EYGLAWSLIESIIVDNGDKNNRGIROLESGFITDNQQQVLKKNVKSWRQNQIRQTFSPDTK IARLRESYLSKQNLSEYEVSGFEVGGKRVAKIYDSIKRGSVRKKD NNSQNDQSWGKKGINSFETTAAGTSQFCTHCKBWSLAIVDIEYEYLKDYNNDLFK VKINDGEVRLLGKKGWRSGEKIKGKELFGPVKDAAMPNVDGLGMKIVKRKYLKLDR DWVSRYGNMAIFICPYVDCHHISHADQAAFNIAV
SEQ ID: 71	MKLTRRRISGNSVQDKITAAYFRDMSQGLLYYDSEDNDCTDKVIESMDFERSWGRILK NGEDDKNPFPYMFVKGGLVGSNDKIVCEPVIDVSDPDNLILINKNLGFRNLKAPDSND TLENLRKIQAGIPEEEVLPKKIKEMIQKDIIVNRKEQLLKSIIKNRIPFSLEGSKLVPSTK KMKWLFLKLIQVNPNTKNEKMLEKYWEIYDYDKLKANITNRLDKTDKARSISRAVSEE LREYHKNLRNTNVRFGSDRPAAGLDNGGSAYKNPDKEEFLFLKEVEQYFKKYFPVK SKHSNKSSDKLSDLVDKYKNYCYSYKVVKKEVNRSIINQLVAGLIQQGKLYYFYYNTDQ EDFLNSVGLSYIQVEEAFFKSVMTLSWGINRLTSFFIIDSNTVKFDIITKKAKEAIKESN YFNKLRTCSRQDHFKELAFYVPPVYVVKDKDRPDDIENLIVLKVNAIESVSYLRNRT PHFKLESSLELLKEDDKNSQNKIDYSVAAEFIKRDIENLYDVREQIRSLGIAEYYKAD MISDCFKTCGLEFALYSPKNSLMPAKVNKYKRGANLNKAYIRDKGPKETGDQGQNSYK ALEEYRLETWYIEVKNNDQSYNAYKNLQLIYYHAFLPEVRENEALITDFINRTKEWNR KETEERLNTKNNKKHKNFDENDDIITVNTYRYESIPDYQGESLDDYLKVLQRQKMARAK EVNEKEEGNNNYIQFIRDVVVWAFGAYLENKLKVNQELQPLSKENIGLNDTLKELFP EEKVKSFPNIKCRFSISTFDNKGKSTDNTSAEAVKTGKEDEKDKNIKRKDLLCFYLF LRLLDENEICKLQHOFIKYRCSLKERFPGNRTKLEKETELLAELELMELVRFTMPSIPE ISAKAESGYDITMKKVFDPFEEKKVFKNPKTSNLYHSDSKTPVTRKYMALLMMSAPLH LYKDIKFQKYYLTIKKECLEYIJKLNSIICKDQNSLNEHQLERIJKLKESEKQNGKDSLQYLDK KDFYKVKEYVENLEQVARYKHLQHKINFESLYRIFRIHVDIAARMVGYTQDWERMHF LFPKALVYNGVLEERRFEAIFFNNNDDNDGRIVKKIQNQNLNNKRELVSMLCWNNKLNK NEFGAIIWKRNPPIAHLNHFTQTEQNSKSSLESLINSRILLAYDRKRQNAVVTKTINDLLN DHYIRIKWEGRVDEGQIYFNICEKEDEIENEPIIHLKHLHKDCYIYKNSYMFQDKQEWIC NGIKEEVYDKSILKCIGNLKFEDYDKNKSSANPKHT
SEQ ID: 72	MLRRDKEVKKLYNVFCIQVGTCPKWWNDEKLSPPEENERRAQKNIKMKNYKWR ACSKYVVESSNIDVIFYYSRKAKNLRYMRKNEDILKKMQEAEKLSKFSGGKLEDV AYTLRKSLVVKYDQFDLSAAMVVFLECIGKNNISDHREIIVCKLLEIRKDFSKLD NVKGSQGANIVRSVRVNQNMIVQPGDRFLFPQVYAKENETVTNKVKEGLNEFLLNY ANLNDKRAESLRLKRLRIDLVFSAPNHYEKDMITLSDNIEKEKFVNWKHECGKKT GLFVDIIPDVLMEEAENIILKDAVVEKERKVNLDRVRKQNIICYRTTRAVVEKYN LFFENNAINQWVIIHIEINAVERILKNCAGLFLKLRKGYLAEKVWWDKDALNLISTKIALG KAVYNFALDDIWKDKKNELGIVDERIRNGITSFDEMIKAHENLQRELADIAFSV LARAVCDMSNLGNKESDFFLWKRNDIADKLKNKDDMASVASLVOFFGGKSSWDINIFK DAYKGKKKYYEVRFIDDLRKAIYCARNEFHKFTALVNDKEWNTLELFGKIFERETEFC LNVKEKDRFYSNNLYMPYQVSERLNMDHLYRSRSVRAAQPSYNSVIVRTAFPEYITNV LGYQKPSYDADTLGKWYSACYYLLKEIYYNSFLQSDRALQFLFEKSVKTLWSDDKKQQR AVDNFDRHDFSDIKSACTSLAQVCQIYMTNEYQNQNNQIKKVRSSNSDIFFDOPVYQHYKVL LKKAIANAFADELYKNNKDLFPGFIGKPFKANEIREIDKEQFLPDWTSRKYEALCIEVSGSQ ELQKWIYVGKFLNARSLNLMVGSMSRQIYQVTDIKRRAASIGNELHVSVHDVEKVEKW VQVIEVCSSLASRTSNQFEDYFNDKDDYARYLKSYVDSNVNDMPSEYSALVDFSNEQS DLYVDPKPNPKVNRNIVHSKLFADHILRDIVEPVSKDNIEEFYSQKAETAYCKIKGKEITA EEQKAVLKQYQKQVNLWVRELDLDRIVEGEIINELLGQLINWSFMRERDLYFQLGPHYDCLR NDSSKCPGQYKNIKVDENSISKDALYQIIGMYVNGVTVYAPEKDGDKLQCVKGGVGV KVSAFHRYSKYGLNLEKLTLYNAGLEIFEVVAEHEDIINLRNGIDHFKYYLGDYRMSLISY SEVFDRPTFYDICKYQKVNVLNLQNLILLRHNVIVEPILESGFKTIGEQTKPGAKLSIRSIS TFQYKVKGTLITDAKDERYLETIRKILYYAENEEDNLKSVVVTNAKYEKNKESDD QNKQKEKKNKDNKGKNEETKSDAEKNNNERLSYNPFANLNFKLSN
SEQ ID: 89	MISKSDFINHYAIQKTLRFELQPIGKTREHIQKNGIIEHDEALEQKYQIVKKIIDRFHRKH IDEALSVDFTKBLDTFTKTIQELYLKRKGKTDREKKELEELSADLRKRVVSFLEGKVEGD AFFAKVQORYGILFADKDKFESTACDDEIEKDAIEAFKRFATYFTGPHENRKNMYS DEESTATAYVINENLPRFLLENKARPEKIFKILDVIEKTKLTLKQNLKTELKGHSVSDIFRI FNNVLTQAGITYNTIILGGKTKENGEVKVQGLNEIINLFNQKKNKDMLPLKPLYQILSE EYSTSFKIKAPENDENVLAIDTFWNEHIEKSIHPTGKRFNILLKIENLCKKLEKYKDKE IEKLFIBERKNLSTVSHQVYQGWNIIRDALRMHLEMNNKNIKEKDIDKYLNDNAFAWKEI KDSIKIYKEHVEDAKELDENGIVKYFSSMSINEEDDEKEYSISLIKNINEKYNVKSILEED RTGKSDLHQDKEKVAIIKEFLDSLQKQLQWPLKLLYVTVPDKEKDYDFYNELEMYHTL LPLTTLYNKVNRNMTKRPVSEKEFKLTLFEKSTLLDGWDKNKERANLGVILRKGNYY GIMNKKYNDIFDSIPGLTTDYCEKMNQYKLLPGPNKMLPKVFFSKGVQFYKPSKEILRI YEKGEFKDKSGNFKKSLSLHTLIDFYKEAIAKNEDWKFIFKFKFKNTRYADISQFYKDV ERQGYKISFDKIDWELLVDEGKLFKLYNKDFSPYSGKGPNLHTIYWKNIIFSHDNL NNVVYKLNGEAEVYFVRKSIDYPEEITYNNGHKEELNGKFNYPITIKDKRYAEDKFLFH PITMNFISKEEKVNQLACEYLSTTKEVHIIIGDRGERHLLYLSLIDKEGNIKKQLSLNTI KNENYDKEIDYHAKLDEKEKKEKREARKNWDVIEENIKELKEGYLSQVHVQIAKLMVEYK

TABLE 1-continued

Amino Acid Sequences of Exemplary CRISPR/Cas effector proteins	
SEQ ID NO:	Programmable Nuclease Amino Acid Sequence
	A ILVMEIDLNTGPKRGRFVKVEQVYQKFEKMMIDKLNYLVLKDRQATQPGGSLKAYQL ASSLESFKKLGKQCGMI FYVPSAYTSKIDPTTGFYNFIRVDVPNLEKGKEFFSKPEKIIYNN TKEDYFEEFHCKYGFVPEPKNKDNDRKTESLTYYNAIKDTVVVCSTHHERYKIVRN KAGYYESQPVDVTKNLRDI FSEANINYSGDGDIKP III ESNNAKLLKSTAEQLKLILAMRY NNNGKHDDEKDYI LSPVVKNKQGKFFCSDLGDQS LPINADANGAYNIALKGLLLIEKIKK QQGKAQAKDLYI SNELEWFMFMSR
SEQ ID: 90	MIKS YDDFINHYAIQKTLRFELQPIGKTREHIQKNGII EHDEALEQKYQIVKK IDR FHRKH IDEALS LVDFTKHL DFTKFTI QELYLKR GKT DREK KEEL ELSADLRKLIVSYLKGNV KQKT QHNLNP IKER FEILFGKELFTNEEF TLAENKEEKKAIQAFKGFTTYFKGFQENRKNMYS EEDKSTA IAYRI INENMPLFIENIARFQKVLDVIEKTKLTELKQNLKTELKGHSVSDIFRIE YFNNVLTQEGISRYNTIIGGTTTEGVKIQGLNEIINLNHQNSKDVKLPLKPLYKQILSE EYSTSFT ISAFDNDV LQAI GSFCND CIFYAKNNVNGKAYNLLQTVQAF CNSIDTYNDN RLDGLH IERK NLATL SHQVYGEWNILR DALQI HYAEYEQKDNGNNN Nyles KTF SWKA LKD ALTT YKSL VEEA QDID ENGFIA YFKDMKFEEID GKTTSIDL IENI QTTRYK SIETI LQE DRNNKNNLHQKEEV KATV IKGFLDSV KYLQWPNL NM YIASPVDDKD YDFYNELEM YHD TLLPLTTLYNK VRN YMTRKP YSVEKPKL TLLDGWDK NKE RANLG VILRK GNNY YLGIMNKKYND IFD S I PGL TTTD YCEK MNK YLLP GP N KML P KVFF SKKG VQFY KPS QEI RLYNNKEF KKGD T F N K N S L H K L I N F Y K E S I A K T E D W S V F Q K F K N T N D Y A D I S Q F Y K D V ERQGYKISF DKIDW EYV L L L D E G K L F K I Y N K D F S P Y S K G K P N L H T I Y W K N I F S H D N L NNVY KLN GEA EVF Y R K S I E Y P E E I L Q K G H H V N E L K D K P K Y P I I K D K R Y A E D K F L F H V P ITM N L S K E E K R V N Q L A C E Y L S T T K E D V H I I G I D R G E R H L L Y L S L I D K E G N I K Q L S L N T I K N E N Y D K E I D Y R V K L D E K E K K R D E A R K N W D V I E N I K E L K E G Y M S Q V I H I I A K M M A E E K A I L I M D N I G F K R G R F V K E Q V Y Q K F E K M M I D K L N Y L V F K N K E P E P G G S L N A Y Q L T S K P D S F K K L G K Q G F I F Y V P S A Y T S K I D P T T G F Y N F I R V D V P N L E K G K E F F S K F E K I I Y N T K E D Y F E F H C K Y G K F P E P K N K D N D R K T K S E L T Y Y N A I K D T V V V C S T H E R Y K I V R N K A G Y Y E S Q P V D V T K N L R D I F S E A N I N Y S D G D I K P I I I E S N N A K L L K S I A E Q L K L I L A M R Y N G K H D D E K D Y I L S P V K N K Q G K F F C S L D G Q S L P I N A D A N G A Y N I A L K G L L I E K I K K Q Q G K A K D L Y I S N L E W F M F M S R
SEQ ID: 91	MIKS YDDFINHYAIQKTLRFELQPIGKTREHIQKNGII EHDEALEQKYQIVKK IDR FHRKH IDEALS LVDFTKHL DFTKFTI QELYLKR GKT DREK KEEL ELSADLRKLIVSYLKGNV KQKT QHNLNP IKER FEILFGKELFTNEEF TLAENKEEKKAIQAFKGFTTYFKGFQENRKNMYS EEDKSTA IAYRI INENMPLFIENIARFQKVLDVIEKTKLTELKQNLKTELKGHSVSDIFRIE YFNNVLTQEGISRYNTIIGGTTTEGVKIQGLNEIINLNHQNSKDVKLPLKPLYKQILSE YSTSFKI KAFENDNEV LKAID TFWN EHHIEKSIHPVTKR FNILLK IENL CKK LEKY K DKEI EKL FIE KRN L TS VSH VY G O W N I I RD AL RMH L E M M N K N I K E K I D K Y L D N D A F A W K E I KDS I K I Y K E H V E D A K E L D E N G I V K Y F S S M S I N E E D D E K E Y S I S I L K N I N E K Y N N V K S I L E D RTG K S D L H Q D K E K V A I I K E F L D S L K Q L Q W F L K L L V V T V P L D E K D Y E F Y N B E L V Y E A L L P L N S L Y M K V R N Y M T R K P Y S V E K F K L N F Y S P T L L D G W D K N K E T A N L S I L K K N G K Y Y L G I M N K E N N T I F E N P K S K S N D Y Y E K M I Y K L L P G P N K M L P K V F F S K K G L E Y Y Y K P S K E I L R I Y E K G E F K K D S G N F K Q D S H L T I D F Y K E A I K N E D W K I F K F K F K N T R E Y E D I S Q F Y R D V E E Q G Y L I I F K V D A N Y V D K L V E E G E L F L P Q I Y N K D F S E N K S K G N P N L H T I Y W E S L F D N Q N L K D V V Y K L N G E A E V F Y R K K S I D Y P E E I Y M N G H H K E E L N G K F N Y P I I K D R R Y T Q D K F L F H V P I T L N F L A K S D E K V N M V K N Y I A T T N E K I H I I G I D R G E R N L L Y L S L I D S N G N I V K Q Q S L N I I E L P K Y Q K Q I D Y H A K L N E K E K O R L A A R Q N W D V I E N I K E L K E G Y L S Q V I H Q I A R L M V D Y K A I L V M E D L N F G F K R G R F K V E K Q V Y Q K F E K M I D K L S Y L V F K E K N L C E P G G S L R A Y Q L S A P F K S F K A L G K Q S G M I F Y V P A Q Y T S K I D P T T G F Y N F L N I D V S N L A R S K E T F S K F D K I V Y N K K E D Y F E F Y C K M I N F E S A N Q L T K K S Q N K A N A E L K E F Q W I L C S T H H D R F K V E R K N N Q I N Y C K I N V N E E L K L L N S F E S A N Q L T K K S Q N K A N A E L K E F Q W I L C S T H H D R F K V E R K N N Q I N Y C K I Q D F I L S P V K N A S G K F F C T L D N N N T L P L D A D A N G A Y N I A L K G L M I V Q R V K A G G K L D L S I S K D D W I N F L I M N K K L P K

[0149] In some embodiments, target nucleic acid is an RNA and the CRISPR/Cas effector protein is an RNA-targeting CRISPR/Cas effector protein. In some embodiments, the RNA-targeting CRISPR/Cas effector protein comprises an amino acid sequence with at least 90% identity to the amino acid sequence of SEQ ID NO: 21, 62, 43, 41, or 42. In some embodiments, the RNA-targeting CRISPR/Cas effector protein comprises the amino acid sequence of SEQ ID NO: 21, 62, 43, 41, or 42.

[0150] In some embodiments, target nucleic acid is a DNA and the CRISPR/Cas effector protein is a DNA-targeting CRISPR/Cas effector protein. In some embodiments, the DNA-targeting CRISPR/Cas effector protein comprises an amino acid sequence with at least 90% identity to the amino

acid sequence of SEQ ID NO: 3, 34, 57, 36, 65, 67, 68, 89, 90, 91, or 17. In some embodiments, the DNA-targeting CRISPR/Cas effector protein comprises the amino acid sequence of SEQ ID NO: 3, 34, 57, 36, 65, 67, 68, 89, 90, 91, or 17.

[0151] In some embodiments, effector proteins described herein have been modified (also referred to as an engineered protein). In some embodiments, a modification of the effector proteins may include addition of one or more amino acids, deletion of one or more amino acids, substitution of one or more amino acids, or combinations thereof. In some embodiments, effector proteins disclosed herein are engineered proteins. Unless otherwise indicated, reference to effector proteins throughout the present disclosure include engineered proteins thereof.

[0152] In some embodiments, effector proteins described herein can be modified with the addition of one or more heterologous peptides or heterologous polypeptides (referred to collectively herein as a heterologous polypeptide). In some embodiments, an effector protein modified with the addition of one or more heterologous peptides or heterologous polypeptides may be referred to herein as a fusion protein.

[0153] In some embodiments, a heterologous peptide or heterologous polypeptide comprises a protein tag. In some embodiments, the protein tag is referred to as purification tag or a fluorescent protein. The protein tag may be detectable for use in detection of the effector protein and/or purification of the effector protein. Accordingly, in some embodiments, compositions, systems and methods comprise a protein tag or use thereof. Any suitable protein tag may be used depending on the purpose of its use. Non-limiting examples of protein tags include a fluorescent protein, a histidine tag, e.g., a 6XHis tag; a hemagglutinin (HA) tag; a FLAG tag; a Myc tag; and maltose binding protein (MBP). In some embodiments, the protein tag is a portion of MBP that can be detected and/or purified. Non-limiting examples of fluorescent proteins include green fluorescent protein (GFP), yellow fluorescent protein (YFP), red fluorescent protein (RFP), cyan fluorescent protein (CFP), mCherry, and tdTomato.

[0154] A heterologous polypeptide may be located at or near the amino terminus (N-terminus) of the effector protein disclosed herein. A heterologous polypeptide may be located at or near the carboxy terminus (C-terminus) of the effector proteins disclosed herein. In some embodiments, a heterologous polypeptide is located internally in an effector protein described herein (i.e., is not at the N- or C-terminus of an effector protein described herein) at a suitable insertion site.

[0155] In some embodiments, an effector protein described herein comprises 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more heterologous polypeptides at or near the N-terminus, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more heterologous polypeptides at or near the C-terminus, or a combination of these (e.g., one or more heterologous polypeptides at the amino-terminus and one or more heterologous polypeptides at the carboxy terminus). When more than one heterologous polypeptide is present, each may be selected independently of the others, such that a single heterologous polypeptide may be present in more than one copy and/or in combination with one or more other heterologous polypeptides present in one or more copies. In some embodiments, a heterologous polypeptide is considered near the N- or C-terminus when the nearest amino acid of the heterologous polypeptide is within about 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 40, 50, or more amino acids along the polypeptide chain from the N- or C-terminus.

[0156] In some embodiments of the methods disclosed herein, certain conditions are employed that enhance trans cleavage activity of an effector protein. In some embodiments, under certain conditions, trans cleavage occurs at a rate of at least 0.005 mmol/min, at least 0.01 mmol/min, at least 0.05 mmol/min, at least 0.1 mmol/min, at least 0.2 mmol/min, at least 0.5 mmol/min, or at least 1 mmol/min. In some embodiments of the methods disclosed herein, certain conditions are employed that enhance cis cleavage activity of an effector protein.

[0157] Certain conditions that may enhance the activity of an effector protein include a certain salt presence or salt concentration of the solution in which the activity occurs. For example, cis-cleavage activity of an effector protein may be inhibited or halted by a high salt concentration. The salt may be a sodium salt, a potassium salt, or a magnesium salt. In some embodiments, the salt is NaCl, KCl, or MgCl<sub>2</sub>. In some embodiments, the salt is KNO<sub>3</sub>. In some embodiments, the salt concentration is less than 150 mM, less than 125 mM, less than 100 mM, less than 75 mM, less than 50 mM, less than 25 mM, less than 10 mM, or less than 5 mM.

[0158] Certain conditions that may enhance the activity of an effector protein include the pH of a solution in which the activity. For example, increasing pH may enhance trans cleavage activity. For example, the rate of trans cleavage activity may increase with increase in pH up to pH 9 in at least some settings. In some embodiments, the pH is about 7, about 7.1, about 7.2, about 7.3, about 7.4, about 7.5, about 7.6, about 7.7, about 7.8, about 7.9, about 8, about 8.1, about 8.2, about 8.3, about 8.4, about 8.5, about 8.6, about 8.7, about 8.8, about 8.9, or about 9. In some embodiments, the pH is 7 to 7.5, 7.5 to 8, 8 to 8.5, 8.5 to 9, or 7 to 8.5. In some embodiments, the pH is less than 7. In some embodiments, the pH is greater than 7.

[0159] Certain conditions that may enhance the activity of an effector protein includes the temperature at which the activity is performed. In some embodiments, the temperature is about 25° C. to about 70° C. In some embodiments, the temperature is about 20° C. to about 40° C., about 30° C. to about 50° C., or about 40° C. to about 60° C. In some embodiments, the temperature is about 25° C., about 30° C., about 35° C., about 40° C., about 45° C., about 50° C., about 55° C., about 60° C., about 65° C., or about 70° C.

[0160] In some embodiments, effector proteins described herein have been modified (also referred to as an engineered protein or engineered polypeptide). In some embodiments, a modification of the effector proteins may include addition of one or more amino acids, deletion of one or more amino acids, substitution of one or more amino acids, or combinations thereof. In some embodiments, effector proteins disclosed herein are engineered proteins. Unless otherwise indicated, reference to effector proteins throughout the present disclosure include engineered proteins thereof.

[0161] In some embodiments, engineered effector proteins as described herein comprise one or more amino acid modifications relative to cognate effector protein (e.g., a modification as exemplified when comparing to an effector protein having any one of the amino acid sequences recited in TABLE 1 to the cognate effector protein), and wherein the engineered effector protein exhibits one or more improved characteristics compared to the cognate effector protein (e.g., a naturally occurring effector protein). In some embodiments, a cognate effector protein, as described herein, refers to a naturally occurring effector protein that may be used as the parental effector protein sequence for protein engineering. In some embodiments, the naturally occurring effector protein comprises certain characteristics (e.g., structure and/or activity) that may be of interest for protein engineering.

[0162] In some embodiments, the complex comprising the engineered polypeptide and an engineered guide nucleic acid comprises increased stability as compared to a complex comprising the cognate effector protein and an engineered guide nucleic acid. In some embodiments, the one or more

improved characteristics of the engineered effector protein compared to the cognate effector protein are selected from: increased catalytic activity at a temperature above 37° C.; decreased catalytic activity at a temperature below 37° C.; increased catalytic activity at a defined salt concentration; increased cleavage rate of target nucleic acid; increased trans cleavage rate; increased formation of a complex comprising the engineered polypeptide and an engineered guide nucleic acid; increased solubility; increased manufacturability (e.g., increased expressibility, solubility, purification, etc.), and increased stability.

[0163] Effector proteins of the present disclosure can be engineered, using any suitable protein engineering method known in the art. Examples of suitable protein engineering methods are described herein. Suitable protein engineering methods can include a method of using mutagenesis to generate a novel nucleic acid encoding a novel effector protein or novel polypeptide, which novel effector protein is itself a modified biological molecule and/or contributes to the generation of another modified biological molecule as compared to wild-type equivalents. Protein engineering methods can be geared towards maintaining certain existing protein functions while modifying others (e.g., maintaining binding activity to a guide nucleic acid, while modifying nuclease activity or specificity), increasing existing protein function, gaining a novel protein function, improving the stability of a protein under certain conditions, improving function in different environments, such as, for example, high temperature and/or high salt, or combinations thereof. Suitable protein engineering methods can include, but are not limited to, random mutagenesis, focused mutagenesis, or methods that integrate both random and focused mutagenesis. In some embodiments, effector proteins can be engineered *in vitro* or *in vivo* by eukaryotic cells or by prokaryotic cells.

[0164] Random mutagenesis engineering methods can generate random point mutations at codons corresponding to specific structurally characterized residues (e.g., protein residues involved in binding or catalysis, such as, for example, catalytic residues in RuvC nuclease active site). Although protein engineering by methods such as directed evolution via repeated random mutagenesis (e.g., random chemical or error prone PCR (epPCR)) and selection can yield engineered proteins with desirable characteristics, some protein engineering efforts require more specificity. For example, protein engineering methods which require mutation of more than one nucleotide relative to a non-modified codon can require focused mutagenesis, which can introduce specific amino acid substitutions at positions corresponding to targeted nucleotide(s) or targeted residue(s). Focused mutagenesis can employ a synthetic nucleic acid, such as a synthetic DNA oligonucleotide comprising one or more modifications, which can also be referred to as a mutagenic oligonucleotide. The mutagenic oligonucleotide, which can be incorporated into a gene library as a mutagenic cassette, can comprise modified/degenerate codons corresponding to targeted residues. Focused mutagenesis can also yield more functional variations, beneficial mutations, or modifications resulting in the desired engineered protein activity while minimizing neutral or deleterious mutations.

[0165] Effector proteins can be engineered *in vitro* or *in vivo* by focused and/or random mutagenesis methods, such as chemical mutagenesis, combinatorial libraries, computational strategies for high-quality library design, homologous

recombination, non-homologous recombination, recombination based methods such as DNA shuffling (e.g., molecular breeding), directed evolution, deletion mutagenesis, error prone PCR (epPCR), insertion mutagenesis, random mutagenesis, scanning mutagenesis, site-directed mutagenesis (SDM) (and similar methods such as: site-specific mutagenesis, oligonucleotide-directed mutagenesis, site-saturation mutagenesis (SSM)), use of mutator strain, assembly PCR, sexual PCR mutagenesis, cassette mutagenesis, recursive ensemble mutagenesis, exponential ensemble mutagenesis, site-specific mutagenesis, gene reassembly, gene site saturation mutagenesis (GSSM), synthetic ligation reassembly (SLR), recombination, replacing codon(s) encoding the same amino acid, recursive sequence recombination, phosphothioate-modified DNA mutagenesis, uracil-containing template mutagenesis, gapped duplex mutagenesis, point mismatch repair mutagenesis, repair-deficient host strain mutagenesis, chemical mutagenesis, radiogenic mutagenesis, deletion mutagenesis, restriction-selection mutagenesis, restriction-purification mutagenesis, artificial gene synthesis, ensemble mutagenesis, chimeric nucleic acid multimer creation, other mutagenesis methods described herein, or combinations thereof (See, e.g., Packer, M., et al., *Nature Reviews Genetics*, 16 (7): 379-94 (2015)).

[0166] *In vivo* mutagenesis methods can be focused, random, or combinations thereof. *In vivo* focused mutagenesis methods can comprise selectively introducing localized DNA damage into a genome, such as, for example, targeting a pathway requiring long-range resection so as to form a single-stranded region during biasing repair and selectively mutating said single-stranded region. In some embodiments, *in vivo* focused mutagenesis methods comprise delivering a nucleic acid encoding an effector protein and a guide nucleic acid to a cell, and contacting the cell with a mutator compound or mutator enzyme. In some embodiments, *in vivo* focused mutagenesis methods comprise selectively introducing localized DNA damage in a preselected region of an organism's DNA *in vivo*, biasing repair of the localized DNA damage by targeting a pathway requiring long-range resection of the localized DNA damage, wherein the DNA forms a single-stranded region during the biasing repair, and selectively mutating the single stranded region to cause targeted mutagenesis, optionally wherein the organism is an eukaryotic organism. In some embodiments, localized DNA damage is a double stranded break (DSB). In some embodiments, a DSB is introduced by a DNA mutator enzyme domain (e.g., DNA glycosylase, 3-methyladenine glycosylase (e.g., Magl), DNA nuclease, FokI). In some embodiments, biasing repair of the DSB involves contacting the cell with a compound that elicits DNA damage checkpoint activation. In some embodiments, the compound that elicits DNA damage checkpoint activation is a chemical checkpoint activator (e.g., methyl methanesulphonate (MMS)), or an enzymatic checkpoint activator (e.g., Magi). *In vivo* random mutagenesis methods (i.e., traditional genetic screens) can randomly damage DNA via chemical and/or physical agents such as, for example, alkylating compounds (e.g., ethyl methanesulfonate (EMS)), deaminating compounds (e.g., nitrous acid), base analogues (e.g., 2-aminopurine), radiation (e.g., ultraviolet irradiation), bisulfite, or combinations thereof. In some embodiments, random chemical mutagenesis can facilitate dose-dependent modification or mutation of DNA. In some embodiments, random chemical mutagenesis, which has a broad muta-

tional spectrum, can be used to randomly deactivate genes for a genome-wide screen *in vivo* or *in vitro*. In some embodiments, random mutagenesis enhances the error rate during DNA replication, which can lead to off-target mutations and/or deleterious genome mutations.

[0167] Random mutator strain mutagenesis, an *in vivo* random mutagenesis method, can produce randomly mutagenized plasmid libraries upon propagation of the genes cloned in plasmids through a mutator strain, like *Escherichia coli* XL1-red. In brief, random mutator strain mutagenesis is a method for introducing random point mutations throughout a gene encoding a protein of interest with the use of a plasmid. The method involves transformation and propagation of a plasmid containing the target gene into a mutator strain, isolating the resulting randomly mutagenized plasmid library, transforming the library into a strain comprising the mutant target gene, and screening the mutant target gene phenotype. In some embodiments, the method elicits random mutagenesis via phage-assisted continuous evolution (PACE), a method which harnesses the phage virus bacterial infection cycle to generate multiple rounds of DNA sequence mutations, selecting for DNA mutations in a mutant target gene encoding a protein that result in a desired protein structure or activity. In some embodiments, random mutagenesis involves yeast orthogonal replication. Although the methods generally offer ease of use, host intolerance to a high degree of genomic mutation(s) can place an upper limit on *in vivo* mutagenesis rates. *In vitro* random mutagenesis methods generally offer protein engineering methods with higher target mutation rates as compared to most *in vivo* random mutagenesis methods.

[0168] Homologous recombination, a random mutagenesis method which can be carried out *in vivo* or *in vitro*, can lead to DNA modification, damage, or repair upon DNA shuffling, family shuffling, staggered extension process (StEP), random chimeragenesis on transient templates (RACHITT), nucleotide exchange and excision technology (NexT), heritable recombination, assembly of designed oligonucleotides (ADO), synthetic shuffling, or combinations thereof. For example, StEP is a modified PCR that uses highly abbreviated annealing and extension steps to generate staggered DNA fragments and promote crossover events along the full length of the template sequence(s), such that most of the resulting polypeptides comprise sequence information from different template sequence(s). RACHITT performs molecular mutagenesis at a high recombination rate by aligning parental gene fragments on a full-length DNA template, which are then stabilized on the template by a single long annealing step at a relatively high ionic strength. RACHITT can yield a considerable number of crossovers per gene in a single annealing step. NexT is also a modified PCR that uses uridine triphosphate (dUTP) as a DNA fragmentation defining exchange nucleotide with thymidine. In NexT, the exchange nucleotides are removed enzymatically, followed by chemical cleavage of the DNA backbone. Finally, the oligonucleotide pool is reassembled into full-length genes by internal primer extension, and the recombinant gene library is amplified by standard PCR. Another modified PCR, ADO, is a two-step reaction involving an overlap extension PCR step using synthetic oligonucleotides followed by a PCR amplification step using outer primers, resulting in double-stranded DNA assembled with engineered gene fragments. In some embodiments, homologous recombination (HR) methods can lead to DNA modification

comprising mutations generated by knocking out or removing one or more nucleotides. In some embodiments, HR methods repair gene function by identifying sequence homology and replicating the functional version of the target gene. In some embodiments, knock out mutations result in functional modifications to the protein encoded by the modified nucleic acid sequence. HR can prove advantageous in its ability to identify beneficial mutation combinations, eliminate passenger mutations, shuffle functional sequences of orthologous proteins, or combinations thereof.

[0169] Error prone PCR (epPCR) mutagenesis, an *in vitro* random mutagenesis method, can result in the modification/damage of DNA via PCR amplification involving supplemental mixture components such as, for example, proprietary enzyme mixes (e.g., Mutazyme), Taq supplemented with Mg<sup>2+</sup>, Taq supplemented with Mn<sup>2+</sup> and/or unequal dNTPs, or combinations thereof. EpPCR involves the modification of DNA or creation of a mutation during PCR amplification of a target gene, a fragment of a target gene, a target sequence, a DNA sequence, or combinations thereof. In some embodiments, the low fidelity of DNA polymerases under certain conditions generates point mutations during PCR amplification of a gene of interest. In some embodiments, the base-pairing fidelity of DNA polymerases can be reduced with increased magnesium concentrations (e.g., Taq supplemented with Mg<sup>2+</sup>), supplementation with manganese (e.g., Taq supplemented with Mn<sup>2+</sup>), the use of mutagenic dNTP analogues (e.g., unequal/unbalanced dNTPs), or the use of proprietary enzyme mixes (e.g., Mutazyme) to increase mutation rates (e.g., 10<sup>-4</sup>–10<sup>-3</sup> per replicated base). Given that each cycle of PCR amplification leads to the accumulation of mutations, high mutation rates (e.g., high number mutations per clone) can be achieved by increasing the number of PCR amplification cycles. EpPCR offers advantages, such as, for example, its tendency for high mutation rates and/or a relatively even mutation spectrum, as well as easy to use commercial formulations. Optionally, a more ideal nucleotide mutational spectrum can be achieved via sequence saturation mutagenesis (SeSaM), a mutagenesis method that randomizes a target sequence at every single nucleotide position. Briefly, SeSaM is a chemo-enzymatic random mutagenesis method which involves the enzymatic insertion of a base, such as the universal base deoxyinosine (2'-deoxyInosine (dI)), throughout the target gene.

[0170] Suitable applications of epPCR include, but are not limited to, the generation of neutral drift libraries, which can be used to identify an evolvable starting point for protein engineering (e.g., the directed evolution of a target protein of interest). Generating a neutral drift library can involve exploring accessible sequence space by repeated rounds of mutagenesis and selection for the accumulation of mutations that are largely neutral and compatible with maintaining wild-type function. Mutations that are largely neutral for the wild-type protein function accumulate, while mutations detrimental to the wild-type protein function are purged, yielding a library of high diversity and quality. Specifically, a target gene is mutagenized by epPCR, fused to a reporter nucleic acid (e.g., GFP reporter), and the mutagenized gene variants are then screened for target protein expression. After multiple rounds of mutagenesis and screening, the resulting neutral drift library exhibits sequence diversity that does not destabilize protein structure or protein function.

Screening for target protein expression ensures the resulting neutral drift library mostly lacks non-target deleterious mutations.

[0171] Another *in vitro* method for generating high-quality libraries is site-directed saturation mutagenesis (SDSM). SDSM and similar methods such as site-directed mutagenesis (SDM), site-saturation mutagenesis (SSM), site-specific mutagenesis, or oligonucleotide-directed mutagenesis, are *in vitro* focused mutagenesis methods, capable of fully sampling the amino acid repertoire, focusing on functionally relevant residues, and/or increasing library quality. In some embodiments, SDSM involves NNK and NNS codons (where N can be any of the four nucleotides, K can be G or T, and S can be G or C) on mutagenic primers. SDM, which is commonly applied to study the function of a single amino acid in relation to the rest of the protein, involves the substitution of a single amino acid for another, usually an alanine. In some embodiments, site-directed mutagenesis is performed via means that are synthetic, where the design of the engineered/desirable/target/progeny polynucleotide(s) is derived by analysis of a wild-type/parental set of proteins and/or of the polypeptides correspondingly encoded by the wild-type/parental proteins. SSM, which is a similar method to SDM, involves the substitution of a single amino acid for another, usually for any of the other 19 standard amino acid substituents other than alanine. Thus, the SSM mutagenesis product is a collection of clones, each having a different codon in the targeted position (i.e., saturated), yielding all possible substitutions. Analysis of the SSM mutagenesis product can indicate the relationship between the targeted amino acid positions and protein function. In some embodiments, site-specific protein engineering methods, such as SSM, target the diversification of functionally relevant residues, some of which may not be comprised in the protein's primary structure. In some embodiments, simultaneous SSM of, for example, multiple target residues, can result in combinations of mutations that can exhibit synergistic or epistatic interactions. Combinations of mutations exhibiting epistatic interactions (e.g., sign epistasis, a type of interaction in which mutations can be individually non-desirable/deleterious, but confer gain-of-function in combination) can be selected for with the use of simultaneous SSM. In some embodiments, simultaneous SSM targets combinations of mutations exhibiting synergistic interactions (e.g., a type of interaction in which mutations in combination have a greater effect as compared to the sum of the effects of each individual mutation) with desirable/target effects. Overall, a site-saturation library can result from sequential enrichment of epistatic mutation combinations, sequential enrichment of synergistic mutation combinations, sequential enrichment of functionally relevant mutations, sequential enrichment of functionally relevant residues, or combinations thereof. Site-specific mutagenesis or oligonucleotide-directed mutagenesis involves the modification of DNA or creation of an intentional mutation at a specific location on the oligonucleotide sequence. Modification of DNA or creation of an intentional mutation can involve insertional mutagenesis and/or deletion mutagenesis. Insertional mutagenesis can involve the incorporation of a mutation into a target gene via the incorporation of a few nucleotides (e.g., insertional mutagenesis via conventional PCR, nested PCR, or similar techniques). Deletion mutagenesis can involve the removal of a target gene, a fragment of a target gene, a target sequence, a DNA sequence, a few nucleotides, or combina-

tions thereof (e.g., deletion mutagenesis via inverse PCR, or a similar technique). Site-specific mutagenesis or oligonucleotide-directed mutagenesis can involve amplifying a gene of interest via PCR with the use of a synthetic primer possessing a specific mutation or a target mutation, which can result in a deletion, insertion, or single nucleotide polymorphism (SNP), as confirmed by sequencing. In some embodiments, oligonucleotide-directed mutagenesis involves the replacement of a short sequence with a synthetically mutagenized oligonucleotide. In brief, a synthetically mutagenized oligonucleotide, can comprise one or more modifications, such as, for example, modified codon(s) corresponding to targeted residue(s). Mutagenesis with synthetic oligonucleotides requires sequencing of individual clones after each selection round, grouping individual clones into families, arbitrarily choosing a single family, and reducing the chosen family to a consensus motif. The consensus motif is resynthesized and reinserted into a single gene for additional selection. Oligonucleotide-directed mutagenesis can be best suited for fine-tuning sequence areas of comparatively low information content. Cassette mutagenesis, a type of SDM, uses a short, double-stranded oligonucleotide sequence (e.g., a gene cassette) to replace a fragment of target DNA such that, a sequence block of a single template is typically replaced by a (partially) randomized sequence (e.g., a mutagenic cassette, which can be a mutagenic oligonucleotide).

[0172] Computational strategies, an *in vitro* focused mutagenesis method for high-quality library design, can involve one or more of Rosetta design, computationally guided libraries, incorporating synthetic oligonucleotides via gene reassembly (ISOR), consensus design, reconstructed evolutionary adaptive path (REAP) analysis, and SCHEMA algorithm(s). The method offers an advantage in the form of creating small libraries pre-enriched for functional variation by natural selection and/or *in silico* filtering. Consensus design (a method which involves the identification of common ancestral mutations (i.e., evolutionary history) by aligning all sequences and identifying the most frequently observed amino acid(s) at each position in the sequence alignment) can lead to the introduction of consensus mutations or significantly distinct/divergent mutations, yielding engineered proteins with improved thermostability, catalytic stability, enzymatic efficiency, or combinations thereof. In contrast, reconstructed evolutionary adaptive path (REAP) analysis provides a method for the identification of significant mutational divergence, which can (i) comprise mutational signatures related to known protein function(s) or protein pathway characteristics, or which can (ii) be used to predict changes in protein function(s) as related to, for example, structural proximity to an active site. In some embodiments, a protein engineering method, incorporating synthetic oligonucleotides via gene reassembly (ISOR), can be used to predict desirable protein engineering outcomes, such as, for example, the introduction of mutations that can improve protein stability and/or protein folding. ISOR, a versatile combinatorial method for the partial diversification of large sets of protein residues or targeted protein positions, offers a method to select target engineered proteins capable of desirable/target activity/properties. As compared to site-specific methods of diversification, ISOR can prove more efficient in identifying target protein positions related to target protein activity, while building a reasonably sized protein library for protein engineering. Briefly, ISOR incor-

porates synthetic oligonucleotides comprising randomized codons flanked by wild-type sequences to wild-type gene fragments via assembly PCR. The resulting reassembled gene comprises randomized cassettes (e.g., mutagenic cassettes) at target sites. As a factor of oligonucleotide concentration, the resulting reassembled gene comprises semi-randomly introduced mutations, such that resulting variants can comprise a different quantity and/or combination of mutated positions. In some embodiments, randomly introduced mutations can comprise a random subset of the resulting mutations. In some embodiments, ISOR is used to create libraries focused on the randomization of individual positions of interest, on the identification of proteins comprising combinations of mutated residues while maintaining, upregulating, or downregulating wild-type protein function, and/or on the identification of proteins comprising combinations of mutated residues while gaining a desirable protein function. In some embodiments, ISOR is used to create libraries characterizing protein function as related to insertions and/or deletions in sequence positions surrounding an active site of interest.

[0173] Computational strategies or computational modeling, as described herein, can facilitate the identification of specific amino acid substitution/modification as related to desired/target engineered protein activity/function. Computational strategies for high-quality library design, can involve, for example, the use of computational algorithms such as SCHEMA and/or Rosetta. Briefly, SCHEMA provides a method for identifying protein fragments and designing novel proteins by recombination of homologous sequences. For example, SCHEMA identifies interacting amino acid residue pairs via structural information, accounting for amino acid residue pair interactions that are broken upon recombination, and predicting which elements in homologous sequences/proteins can be swapped without disturbing the integrity of the protein structure. Briefly, Rosetta is a computational modeling software comprising algorithms which can be used to design methods for protein engineering based on protein structure analysis, such as, for example, protein structure prediction, protein structure refinement, protein conformation, protein docking, functional protein design, and combinations thereof. Rosetta models can be employed to adapt protein engineering methods to specific applications, such as, for example, protein-protein docking interaction/activity of engineered protein(s). Rosetta models can also be employed to consider protein folding, translation, rotation, association, amino acid sequence design, molecular structure interactions, degrees of freedom (DOFs), electrostatic interactions, hydrogen bonding, hydrophobic interactions, or combinations thereof. In some embodiments, Rosetta models can facilitate the design of a protein engineering method to optimize protein sequences (including, for instance, suggesting a single base change) for engineering protein(s) capable of a target protein conformation. In some embodiments, Rosetta models are geared towards maintaining existing protein function, increasing existing protein function, gaining a novel protein function, improving the stability of protein function, improving function in different environments, such as, for example, high temperature and/or high salt, or combinations thereof. In some embodiments, Rosetta's design models can be employed to identify mutations that improve engineered protein stability and binding affinity.

[0174] Non-homologous recombination is an *in vitro* focused mutagenesis method which can lead to DNA modification, damage, or repair upon incremental truncation for the creation of hybrid enzymes (ITCHY), sequence homology-independent protein recombination (SHIPREC), non-homologous random recombination (NRR), sequence-independent site-directed chimeragenesis (SISDC) and overlap extension PCR. For example, ITCHY is a recombination method capable of generating a single-crossover hybrid library based on generation of N- or C-terminal fragment libraries of two genes by progressive truncation of the coding sequences by an exonuclease followed by ligation. Thus, ITCHY allows the creation of hybrid libraries between fragments of genes without any sequence dependency. SHIPREC is a recombination method capable of generating single-crossover hybrid libraries of unrelated or distantly related proteins by maintaining sequence alignment between the parent sequences and introducing crossovers mainly at structurally related sites distributed over the aligned sequences. NRR is a recombination method that enables nucleic acid or DNA fragments to randomly recombine in a length-controlled manner at sites where there is little or no sequence homology. SISDC is a recombination method that enables the recombination of distantly related (or unrelated) proteins at multiple discrete sites, such as sites related to protein function. In some embodiments, non-homologous recombination (NHR) can lead to the recombination of portions of nucleic acid(s) at sites with low or no sequence homology. Thus, NHR can increase the frequency at which novel modified nucleic acid sequences are generated, yielding a more efficient and/or complete exploration of nucleic acid or protein diversity, as compared to HR. NHR can prove advantageous in its capacity to shuffle distantly related sequences, rearrange gene order, rearrange nucleic acids comprising low information content, or combinations thereof.

[0175] In some embodiments, the methods for protein engineering can comprise generating a nucleic acid encoding a polypeptide comprising a mutation or modification (e.g., deleting or adding one or more nucleotides, or a combination thereof) wherein the methods for introducing the mutation or modification comprise any of the protein engineering methods disclosed herein. In some embodiments, the method for protein engineering further comprising expressing nucleic acid comprising a mutation or modification to generate a polypeptide comprising a mutation or modification. In some embodiments, the methods described herein comprise repeating the method for protein engineering until the desired modification or mutation is achieved.

[0176] In some embodiments, the methods for protein engineering can further comprise a screening step, an assaying step, an isolation step, a purification step, or combinations thereof. In some embodiments, the engineered effector proteins can be further processed by unfolding (e.g., heat denaturation, dithiothreitol reduction, etc.) and can be further refolded, using any suitable method.

[0177] In some embodiments, effector proteins may be engineered to improve thermostability, catalytic stability, enzymatic efficiency, or combinations thereof.

#### Guide Nucleic Acids

[0178] In the methods disclosed herein, a guide nucleic acid, as well as any components thereof (e.g., spacer sequence, repeat sequence, linker nucleotide sequence,

handle sequence, intermediary sequence etc.) may comprise one or more deoxyribonucleotides, ribonucleotides, biochemically or chemically modified nucleotides (e.g., one or more engineered modifications as described herein), or any combinations thereof. Such nucleotide sequences described herein may be described as a nucleotide sequence of either DNA or RNA, however, no matter the form the sequence is described, it is readily understood that such nucleotide sequences can be revised to be RNA or DNA, as needed, for describing a sequence within a guide nucleic acid itself or the sequence that encodes a guide nucleic acid, such as a nucleotide sequence described herein for a vector. Similarly, disclosure of the nucleotide sequences described herein also discloses the complementary nucleotide sequence, the reverse nucleotide sequence, and the reverse complement nucleotide sequence, any one of which can be a nucleotide sequence for use in a guide nucleic acid as described herein.

[0179] A guide nucleic acid may comprise a naturally occurring sequence. A guide nucleic acid may comprise a non-naturally occurring sequence, wherein the sequence of the guide nucleic acid, or any portion thereof, may be different from the sequence of a naturally occurring guide nucleic acid. A guide nucleic acid of the present disclosure comprises one or more of the following: a) a single nucleic acid molecule; b) a DNA base; c) an RNA base; d) a modified base; e) a modified sugar; f) a modified backbone; and the like. Modifications are described herein and throughout the present disclosure (e.g., in the section entitled "Engineered Modifications"). A guide nucleic acid may be chemically synthesized or recombinantly produced by any suitable methods. Guide nucleic acids and portions thereof may be found in or identified from a CRISPR array present in the genome of a host organism or cell.

[0180] In general, a guide nucleic acid comprises a first region that is not complementary to a target nucleic acid (FR) and a second region is complementary to the target nucleic acid (SR). In some embodiments, FR is located 5' to SR (FR-SR). In some embodiments, SR is located 5' to FR (SR-FR). In some embodiments, the FR comprises one or more repeat sequences or intermediary sequence. In some embodiments, an effector protein binds to at least a portion of the FR. In some embodiments, the SR comprises a spacer sequence, wherein the spacer sequence can interact in a sequence-specific manner with (e.g., has complementarity with, or can hybridize to a target sequence in) a target nucleic acid.

[0181] The guide nucleic acid may also form complexes as described through herein. For example, a guide nucleic acid may hybridize to another nucleic acid, such as target nucleic acid, or a portion thereof. In another example, a guide nucleic acid may complex with an effector protein. In such embodiments, a guide nucleic acid-effector protein complex may be described herein as an RNP. In some embodiments, when in a complex, at least a portion of the complex may bind, recognize, and/or hybridize to a target nucleic acid. For example, when a guide nucleic acid and an effector protein are complexed to form an RNP, at least a portion of the guide nucleic acid hybridizes to a target sequence in a target nucleic acid. Those skilled in the art in reading the below specific examples of guide nucleic acids as used in RNPs described herein, will understand that in some embodiments, a RNP may hybridize to one or more target sequences in a target nucleic acid, thereby allowing the RNP to modify and/or recognize a target nucleic acid or sequence contained

therein (e.g., PAM) or to modify and/or recognize non-target sequences depending on the guide nucleic acid, and in some embodiments, the effector protein, used.

[0182] In some embodiments, a guide nucleic acid may comprise or form intramolecular secondary structure (e.g., hairpins, stem-loops, etc.). In some embodiments, a guide nucleic acid comprises a stem-loop structure comprising a stem region and a loop region. In some embodiments, the stem region is 4 to 8 linked nucleotides in length. In some embodiments, the stem region is 5 to 6 linked nucleotides in length. In some embodiments, the stem region is 4 to 5 linked nucleotides in length. In some embodiments, the guide nucleic acid comprises a pseudoknot (e.g., a secondary structure comprising a stem, at least partially, hybridized to a second stem or half-stem secondary structure). An effector protein may recognize a guide nucleic acid comprising multiple stem regions. In some embodiments, the nucleotide sequences of the multiple stem regions are identical to one another. In some embodiments, the nucleotide sequences of at least one of the multiple stem regions is not identical to those of the others. In some embodiments, the guide nucleic acid comprises at least 2, at least 3, at least 4, or at least 5 stem regions.

[0183] In some embodiments, the compositions, systems, and methods of the present disclosure comprise two or more guide nucleic acids (e.g., 2, 3, 4, 5, 6, 7, 9, 10 or more guide nucleic acids), and/or uses thereof. Multiple guide nucleic acids may target an effector protein to different locations in the target nucleic acid by hybridizing to different target sequences. In some embodiments, a first guide nucleic acid may hybridize within a location of the target nucleic acid that is different from where a second guide nucleic acid may hybridize the target nucleic acid. In some embodiments, the first loci and the second loci of the target nucleic acid may be located at least 1, at least 10, at least 20, at least 30, at least 40, at least 50, at least 60, at least 70, at least 80, at least 90 or at least 100 nucleotides apart. In some embodiments, the first loci and the second loci of the target nucleic acid may be located between 100 and 200, 200 and 300, 300 and 400, 400 and 500, 500 and 600, 600 and 700, 700 and 800, 800 and 900 or 900 and 1000 nucleotides apart.

[0184] In some embodiments, the first loci and/or the second loci of the target nucleic acid are located in an intron of a gene. In some embodiments, the first loci and/or the second loci of the target nucleic acid are located in an exon of a gene. In some embodiments, the first loci and/or the second loci of the target nucleic acid span an exon-intron junction of a gene. In some embodiments, the first portion and/or the second portion of the target nucleic acid are located on either side of an exon and cutting at both sites results in deletion of the exon. In some embodiments, composition, systems, and methods comprise a donor nucleic acid that may be inserted in replacement of a deleted or cleaved sequence of the target nucleic acid. In some embodiments, compositions, systems, and methods comprising multiple guide nucleic acids or uses thereof comprise multiple effector proteins, wherein the effector proteins may be identical, non-identical, or combinations thereof.

[0185] In some embodiments, a guide nucleic acid comprises about: 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, or 60 linked nucleotides. In general, a guide nucleic acid comprises at least: 10, 15, 20, 25, 30,

35, 40, 45, 50, 55, 60 linked nucleotides. In some embodiments, the guide nucleic acid has about 10 to about 60, about 20 to about 50, or about 30 to about 40 linked nucleotides.

[0186] In some embodiments, a guide nucleic acid comprises at least 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, or 30 contiguous nucleotides that are complementary to a eukaryotic sequence. Such a eukaryotic sequence is a nucleotide sequence that is present in a host eukaryotic cell. Such a nucleotide sequence is distinguished from nucleotide sequences present in other host cells, such as prokaryotic cells, or viruses. Said sequences present in a eukaryotic cell can be located in a gene, an exon, an intron, a non-coding (e.g., promoter or enhancer) region, a selectable marker, tag, signal, and the like. In some embodiments, a target sequence is a eukaryotic sequence.

[0187] In some embodiments, a length of a guide nucleic acid is about 30 to about 120 linked nucleotides. In some embodiments, the length of a guide nucleic acid is about 40 to about 100, about 40 to about 90, about 40 to about 80, about 40 to about 70, about 40 to about 60, about 40 to about 50, about 50 to about 90, about 50 to about 80, about 50 to about 70, or about 50 to about 60 linked nucleotides. In some embodiments, the length of a guide nucleic acid is about 40, about 45, about 50, about 55, about 60, about 65, about 70 or about 75 linked nucleotides. In some embodiments, the length of a guide nucleic acid is greater than about 15, about 20, about 25, about 30, about 35, about 40, about 45, about 50, about 55, about 60, about 65, about 70 or about 75 linked nucleotides. In some embodiments, the length of a guide nucleic acid is not greater than about 40, about 45, about 50, about 55, about 60, about 65, about 70, about 75, about 80, about 85, about 90, about 95, about 100, about 105, about 110, about 115, about 120, or about 125 linked nucleotides.

[0188] In some embodiments, guide nucleic acids comprise additional elements that contribute additional functionality (e.g., stability, heat resistance, etc.) to the guide nucleic acid. Such elements may be one or more nucleotide alterations, nucleotide sequences, intermolecular secondary structures, or intramolecular secondary structures (e.g., one or more hair pin regions, one or more bulges, etc.).

[0189] In some embodiments, guide nucleic acids comprise one or more linkers connecting different nucleotide sequences as described herein. A linker may comprise 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more nucleotides. A linker may be any suitable linker, examples of which are described herein.

[0190] Guide nucleic acids described herein may comprise one or more spacer sequences. In some embodiments, a spacer sequence is capable of hybridizing to a target sequence of a target nucleic acid. In some embodiments, a spacer sequence comprises a nucleotide sequence that is, at least partially, hybridizable to an equal length of a sequence (e.g., a target sequence) of a target nucleic acid. Exemplary hybridization conditions are described herein. In some embodiments, the spacer sequence may function to direct an RNP complex comprising the guide nucleic acid to the target nucleic acid for detection and/or modification. The spacer sequence may function to direct a RNP to the target nucleic acid for detection and/or modification. A spacer sequence may be complementary to a target sequence that is adjacent to a PAM that is recognizable by an effector protein described herein.

[0191] In some embodiments, a spacer sequence comprises at least 5 to about 50 contiguous nucleotides that are

complementary to a target sequence in a target nucleic acid. In some embodiments, a spacer sequence comprises at least 5 to about 50 linked nucleotides. In some embodiments, a spacer sequence comprises at least 5 to about 25, at least about 10 to at least about 25, or at least about 15 to about 25 linked nucleotides. In some embodiments, the spacer sequence comprises 15-28 linked nucleotides. In some embodiments, a spacer sequence comprises 15-26, 15-24, 15-22, 15-20, 15-18, 16-28, 16-26, 16-24, 16-22, 16-20, 16-18, 17-26, 17-24, 17-22, 17-20, 17-18, 18-26, 18-24, or 18-22 linked nucleotides. In some embodiments, the spacer sequence comprises 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, or more nucleotides.

[0192] In some embodiments, a spacer sequence comprises a nucleotide sequence that is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95% or 100% complementary to a target sequence of a target nucleic acid. A spacer sequence is capable of hybridizing to an equal length portion of a target nucleic acid (e.g., a target sequence). In some embodiments, the spacer sequence comprises at least 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, or 30 contiguous nucleotides that are capable of hybridizing to the target sequence. In some embodiments, the spacer sequence comprises at least 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, or 30 contiguous nucleotides that are complementary to the target sequence.

[0193] It is understood that the spacer sequence of a spacer sequence need not be 100% complementary to that of a target sequence of a target nucleic acid to hybridize or hybridize specifically to the target sequence. For example, the spacer sequence may comprise at least one alteration, such as a substituted or modified nucleotide, that is not complementary to the corresponding nucleotide of the target sequence.

[0194] In some embodiments, a guide nucleic acid for use with compositions, systems, and methods described herein comprises one or more linkers, or a nucleic acid encoding one or more linkers. In some embodiments, the guide nucleic acid comprises at least one, at least two, at least three, at least four, at least five, at least six, at least seven, at least eight, at least nine, or at least ten linkers. In some embodiments, the guide nucleic acid comprises one, two, three, four, five, six, seven, eight, nine, or ten linkers. In some embodiments, the guide nucleic acid comprises more than one linker. In some embodiments, at least two of the more than one linker are the same. In some embodiments, at least two of the more than one linker are not same.

[0195] In some embodiments, a linker comprises one to ten, one to seven, one to five, one to three, two to ten, two to eight, two to six, two to four, three to ten, three to seven, three to five, four to ten, four to eight, four to six, five to ten, five to seven, six to ten, six to eight, seven to ten, or eight to ten linked nucleotides. In some embodiments, the linker comprises one, two, three, four, five, six, seven, eight, nine, or ten linked nucleotides. In some embodiments, a linker comprises a nucleotide sequence of 5'-GAAA-3'.

[0196] In some embodiments, a guide nucleic acid comprises one or more linkers connecting one or more repeat sequences. In some embodiments, the guide nucleic acid comprises one or more linkers connecting one or more repeat sequences and one or more spacer sequences. In some

embodiments, the guide nucleic acid comprises at least two repeat sequences connected by a linker.

[0197] Guide nucleic acids described herein may comprise one or more intermediary sequences. In general, an intermediary sequence used in the present disclosure is not transactivated or transactivating. An intermediary sequence may also be referred to as an intermediary RNA, although it may comprise deoxyribonucleotides instead of or in addition to ribonucleotides, and/or modified bases. In general, the intermediary sequence non-covalently binds to an effector protein. In some embodiments, the intermediary sequence forms a secondary structure, for example in a cell, and an effector protein binds the secondary structure.

[0198] In some embodiments, a length of the intermediary RNA sequence is at least 30, 50, 70, 90, 110, 130, 150, 170, 190, or 210 linked nucleotides. In some embodiments, a length of the intermediary RNA sequence is not greater than 30, 50, 70, 90, 110, 130, 150, 170, 190, or 210 linked nucleotides. In some embodiments, the length of the intermediary RNA sequence is about 30 to about 210, about 60 to about 210, about 90 to about 210, about 120 to about 210, about 150 to about 210, about 180 to about 210, about 30 to about 180, about 60 to about 180, about 90 to about 180, about 120 to about 180, or about 150 to about 180 linked nucleotides.

[0199] An intermediary sequence may also comprise or form a secondary structure (e.g., one or more hairpin loops) that facilitates the binding of an effector protein to a guide nucleic acid and/or modification activity of an effector protein on a target nucleic acid (e.g., a hairpin region). An intermediary sequence may comprise from 5' to 3', a 5' region, a hairpin region, and a 3' region. In some embodiments, the 5' region may hybridize to the 3' region. In some embodiments, the 5' region of the intermediary sequence does not hybridize to the 3' region.

[0200] In some embodiments, the hairpin region may comprise a first sequence, a second sequence that is reverse complementary to the first sequence, and a stem-loop linking the first sequence and the second sequence. In some embodiments, an intermediary sequence comprises a stem-loop structure comprising a stem region and a loop region. In some embodiments, the stem region is 4 to 8 linked nucleotides in length. In some embodiments, the stem region is 5 to 6 linked nucleotides in length. In some embodiments, the stem region is 4 to 5 linked nucleotides in length. In some embodiments, an intermediary sequence comprises a pseudoknot (e.g., a secondary structure comprising a stem at least partially hybridized to a second stem or half-stem secondary structure). An effector protein may interact with an intermediary sequence comprising a single stem region or multiple stem regions. In some embodiments, the nucleotide sequences of the multiple stem regions are identical to one another. In some embodiments, the nucleotide sequences of at least one of the multiple stem regions is not identical to those of the others. In some embodiments, an intermediary sequence comprises 1, 2, 3, 4, 5 or more stem regions.

[0201] Guide nucleic acids described herein may comprise one or more handle sequences. In some embodiments, the handle sequence comprises an intermediary sequence. In such instances, at least a portion of an intermediary sequence non-covalently bonds with an effector protein. In some embodiments, the intermediary sequence is at the 3'-end of the handle sequence. In some embodiments, the intermediary sequence is at the 5'-end of the handle sequence.

Additionally, or alternatively, in some embodiments, the handle sequence further comprises one or more of linkers and repeat sequences. In such instances, at least a portion of an intermediary sequence, or both of at least a portion of the intermediary sequence and at least a portion of repeat sequence, non-covalently interacts with an effector protein. In some embodiments, an intermediary sequence and repeat sequence are directly linked (e.g., covalently linked, such as through a phosphodiester bond). In some embodiments, the intermediary sequence and repeat sequence are linked by a suitable linker, examples of which are provided herein. In some embodiments, the linker comprises a sequence of 5'-GAAA-3'.

[0202] In some embodiments, the intermediary sequence is 5' to the repeat sequence. In some embodiments, the intermediary sequence is 5' to the linker. In some embodiments, the intermediary sequence is 3' to the repeat sequence. In some embodiments, the intermediary sequence is 3' to the linker. In some embodiments, the repeat sequence is 3' to the linker. In some embodiments, the repeat sequence is 5' to the linker. In general, a single guide nucleic acid, also referred to as a single guide RNA (sgRNA), comprises a handle sequence comprising an intermediary sequence, and optionally one or more of a repeat sequence and a linker.

[0203] A handle sequence may comprise or form a secondary structure (e.g., one or more hairpin loops) that facilitates the binding of an effector protein to a guide nucleic acid and/or modification activity of an effector protein on a target nucleic acid (e.g., a hairpin region). In some embodiments, handle sequences comprise a stem-loop structure comprising a stem region and a loop region. In some embodiments, the stem region is 4 to 8 linked nucleotides in length. In some embodiments, the stem region is 5 to 6 linked nucleotides in length. In some embodiments, the stem region is 4 to 5 linked nucleotides in length. In some embodiments, the handle sequence comprises a pseudoknot (e.g., a secondary structure comprising a stem at least partially hybridized to a second stem or half-stem secondary structure). An effector protein may recognize a handle sequence comprising multiple stem regions. In some embodiments, the nucleotide sequences of the multiple stem regions are identical to one another. In some embodiments, the nucleotide sequences of at least one of the multiple stem regions is not identical to those of the others. In some embodiments, the handle sequence comprises at least 2, at least 3, at least 4, or at least 5 stem regions.

[0204] In some embodiments, a length of the handle sequence is at least 30, 50, 70, 90, 110, 130, 150, 170, 190, or 210 linked nucleotides. In some embodiments, a length of the handle sequence is not greater than 30, 50, 70, 90, 110, 130, 150, 170, 190, or 210 linked nucleotides. In some embodiments, the length of the handle sequence is about 30 to about 210, about 60 to about 210, about 90 to about 210, about 120 to about 210, about 150 to about 210, about 180 to about 210, about 30 to about 180, about 60 to about 180, about 90 to about 180, about 120 to about 180, or about 150 to about 180 linked nucleotides.

[0205] In some embodiments, compositions, systems and methods described herein comprise a single nucleic acid system comprising a guide nucleic acid or a nucleotide sequence encoding the guide nucleic acid, and one or more effector proteins or a nucleotide sequence encoding the one or more effector proteins. In some embodiments, a first region (FR) of the guide nucleic acid non-covalently inter-

acts with the one or more polypeptides described herein. In some embodiments, a second region (SR) of the guide nucleic acid hybridizes with a target sequence of the target nucleic acid. In the single nucleic acid system having a complex of the guide nucleic acid and the effector protein, the effector protein is not transactivated by the guide nucleic acid. In other words, activity of effector protein does not require binding to a second non-target nucleic acid molecule. An exemplary guide nucleic acid for a single nucleic acid system is a crRNA or a sgRNA.

[0206] In some embodiments, a guide nucleic acid comprises a crRNA. In some embodiments, the guide nucleic acid is the crRNA. In general, a crRNA comprises a first region (FR) and a second region (SR), wherein the FR of the crRNA comprises a repeat sequence, and the SR of the crRNA comprises a spacer sequence. In some embodiments, the repeat sequence and the spacer sequences are directly connected to each other (e.g., covalent bond (phosphodiester bond)). In some embodiments, the repeat sequence and the spacer sequence are connected by a linker.

[0207] In some embodiments, a crRNA is useful as a single nucleic acid system for compositions, methods, and systems described herein or as part of a single nucleic acid system for compositions, methods, and systems described herein. In some embodiments, a crRNA is useful as part of a single nucleic acid system for compositions, methods, and systems described herein. In such embodiments, a single nucleic acid system comprises a guide nucleic acid comprising a crRNA wherein, a repeat sequence of a crRNA is capable of connecting a crRNA to an effector protein. In some embodiments, a single nucleic acid system comprises a guide nucleic acid comprising a crRNA linked to another nucleotide sequence that is capable of being non-covalently bond by an effector protein. In such embodiments, a repeat sequence of a crRNA can be linked to an intermediary RNA. In some embodiments, a single nucleic acid system comprises a guide nucleic acid comprising a crRNA and an intermediary RNA.

[0208] A crRNA may include deoxyribonucleosides, ribonucleosides, chemically modified nucleosides, or any combination thereof. In some embodiments, a crRNA comprises about: 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, or 60 linked nucleotides. In some embodiments, a crRNA comprises at least: 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60 linked nucleotides. In some embodiments, the length of the crRNA is about 20 to about 120 linked nucleotides. In some embodiments, the length of a crRNA is about 20 to about 100, about 30 to about 100, about 40 to about 100, about 40 to about 90, about 40 to about 80, about 40 to about 70, about 40 to about 60, about 40 to about 50, about 50 to about 90, about 50 to about 80, about 50 to about 70, or about 50 to about 60 linked nucleotides. In some embodiments, the length of a crRNA is about 20, about 25, about 30, about 35, about 40, about 45, about 50, about 55, about 60, about 65, about 70 or about 75 linked nucleotides.

[0209] In some embodiments, a guide nucleic acid comprises a sgRNA. In some embodiments, a guide nucleic acid is a sgRNA. In some embodiments, a sgRNA comprises a first region (FR) and a second region (SR), wherein the FR comprises a handle sequence and the SR comprises a spacer sequence. In some embodiments, the handle sequence and the spacer sequences are directly connected to each other

(e.g., covalent bond (phosphodiester bond)). In some embodiments, the handle sequence and the spacer sequence are connected by a linker.

[0210] In some embodiments, a sgRNA comprises one or more of one or more of a handle sequence, an intermediary sequence, a crRNA, a repeat sequence, a spacer sequence, a linker, or combinations thereof. For example, a sgRNA comprises a handle sequence and a spacer sequence; an intermediary sequence and an crRNA; an intermediary sequence, a repeat sequence and a spacer sequence; and the like.

[0211] In some embodiments, a sgRNA comprises an intermediary sequence and an crRNA. In some embodiments, an intermediary sequence is 5' to a crRNA in an sgRNA. In some embodiments, a sgRNA comprises a linked intermediary sequence and crRNA. In some embodiments, an intermediary sequence and a crRNA are linked in an sgRNA directly (e.g., covalently linked, such as through a phosphodiester bond) In some embodiments, an intermediary sequence and a crRNA are linked in an sgRNA by any suitable linker, examples of which are provided herein.

[0212] In some embodiments, a sgRNA comprises a handle sequence and a spacer sequence. In some embodiments, a handle sequence is 5' to a spacer sequence in an sgRNA. In some embodiments, a sgRNA comprises a linked handle sequence and spacer sequence. In some embodiments, a handle sequence and a spacer sequence are linked in an sgRNA directly (e.g., covalently linked, such as through a phosphodiester bond) In some embodiments, a handle sequence and a spacer sequence are linked in an sgRNA by any suitable linker, examples of which are provided herein.

[0213] In some embodiments, a sgRNA comprises an intermediary sequence, a repeat sequence, and a spacer sequence. In some embodiments, an intermediary sequence is 5' to a repeat sequence in an sgRNA. In some embodiments, a sgRNA comprises a linked intermediary sequence and repeat sequence. In some embodiments, an intermediary sequence and a repeat sequence are linked in an sgRNA directly (e.g., covalently linked, such as through a phosphodiester bond) In some embodiments, an intermediary sequence and a repeat sequence are linked in an sgRNA by any suitable linker, examples of which are provided herein. In some embodiments, a repeat sequence is 5' to a spacer sequence in an sgRNA. In some embodiments, a sgRNA comprises a linked repeat sequence and spacer sequence. In some embodiments, a repeat sequence and a spacer sequence are linked in an sgRNA directly (e.g., covalently linked, such as through a phosphodiester bond) In some embodiments, a repeat sequence and a spacer sequence are linked in an sgRNA by any suitable linker, examples of which are provided herein.

[0214] In some embodiments, compositions, systems and methods described herein comprise a dual nucleic acid system comprising a crRNA or a nucleotide sequence encoding the crRNA, a tracrRNA or a nucleotide sequence encoding the tracrRNA, and one or more effector protein or a nucleotide sequence encoding the one or more effector protein, wherein the crRNA and the tracrRNA are separate, unlinked molecules, wherein a repeat hybridization region of the tracrRNA is capable of hybridizing with an equal length portion of the crRNA to form a tracrRNA-crRNA duplex, wherein the equal length portion of the crRNA does not include a spacer sequence of the crRNA, and wherein the

spacer sequence is capable of hybridizing to a target sequence of the target nucleic acid. In the dual nucleic acid system having a complex of the guide nucleic acid, tracrRNA, and the effector protein, the effector protein is transactivated by the tracrRNA. In other words, activity of effector protein requires binding to a tracrRNA molecule.

[0215] In some embodiments, a repeat hybridization sequence is at the 3' end of a tracrRNA. In some embodiments, a repeat hybridization sequence may have a length of about 1, about 2, about 3, about 4, about 5, about 6, about 7, about 8, about 9, about 10, about 12, about 14, about 16, about 18, or about 20 linked nucleotides. In some embodiments, the length of the repeat hybridization sequence is 1 to 20 linked nucleotides.

[0216] A tracrRNA and/or tracrRNA-crRNA duplex may form a secondary structure that facilitates the binding of an effector protein to a tracrRNA or a tracrRNA-crRNA. In some embodiments, the secondary structure modifies activity of the effector protein on a target nucleic acid. In some embodiments, the secondary structure comprises a stem-loop structure comprising a stem region and a loop region. In some embodiments, the stem region is 4 to 8 linked nucleotides in length. In some embodiments, the stem region is 5 to 6 linked nucleotides in length. In some embodiments, the stem region is 4 to 5 linked nucleotides in length. In some embodiments, the secondary structure comprises a pseudoknot (e.g., a secondary structure comprising a stem at least partially hybridized to a second stem or half-stem secondary structure). An effector protein may recognize a secondary structure comprising multiple stem regions. In some embodiments, nucleotide sequences of the multiple stem regions are identical to one another. In some embodiments, the nucleotide sequences of at least one of the multiple stem regions is not identical to those of the others. In some embodiments, the secondary structure comprises at least two, at least three, at least four, or at least five stem regions. In some embodiments, the secondary structure comprises one or more loops. In some embodiments, the secondary structure comprises at least one, at least two, at least three, at least four, or at least five loops.

[0217] Polypeptides (e.g., effector proteins) and nucleic acids (e.g., engineered guide nucleic acids or reporters) can be further modified as described herein. Examples are modifications that do not alter the primary sequence of the polypeptides or nucleic acids, such as chemical derivatization of polypeptides (e.g., acylation, acetylation, carboxylation, amidation, etc.), or modifications that do alter the primary sequence of the polypeptide or nucleic acid. Also included are polypeptides that have a modified glycosylation pattern (e.g., those made by: modifying the glycosylation patterns of a polypeptide during its synthesis and processing or in further processing steps; by exposing the polypeptide to enzymes which affect glycosylation, such as mammalian glycosylating or deglycosylating enzymes). Also embraced are polypeptides that have phosphorylated amino acid residues (e.g., phosphotyrosine, phosphoserine, or phosphothreonine).

[0218] Modifications disclosed herein can also include modification of described polypeptides and/or guide nucleic acids through any suitable method, such as molecular biological techniques and/or synthetic chemistry, to improve their resistance to proteolytic degradation, to change the target sequence specificity, to optimize solubility properties, to alter protein activity (e.g., transcription modulatory activ-

ity, enzymatic activity, etc.) or to render them more suitable for their intended purpose (e.g., in vivo administration, in vitro methods, or ex vivo applications). Analogs of such polypeptides include those containing residues other than naturally occurring L-amino acids, e.g., D-amino acids or non-naturally occurring synthetic amino acids. D-amino acids may be substituted for some or all of the amino acid residues. Modifications can also include modifications with non-naturally occurring unnatural amino acids. The particular sequence and the manner of preparation will be determined by convenience, economics, purity required, and the like.

[0219] Modifications can further include the introduction of various groups to polypeptides and/or guide nucleic acids described herein. For example, groups can be introduced during synthesis or during expression of a polypeptide (e.g., an effector protein), which allow for linking to other molecules or to a surface. Thus, e.g., cysteines may be used to make thioethers, histidines for linking to a metal ion complex, carboxyl groups for forming amides or esters, amino groups for forming amides, and the like.

[0220] Modifications can further include changing of nucleic acids described herein (e.g., engineered guide nucleic acids) to provide the nucleic acid with a new or enhanced feature, such as improved stability. Such modifications of a nucleic acid include a base editing, a base modification, a backbone modification, a sugar modification, or combinations thereof. In some embodiments, the modifications can be of one or more nucleotides, nucleosides, or nucleobases in a nucleic acid.

[0221] In some embodiments, nucleic acids (e.g., nucleic acids encoding effector proteins, engineered guide nucleic acids, nucleic acids encoding engineered guide nucleic acids, or reporters) described herein comprise one or more modifications comprising: 2'-O-methyl modified nucleotides, 2' fluoro modified nucleotides; locked nucleic acid (LNA) modified nucleotides; peptide nucleic acid (PNA) modified nucleotides; nucleotides with phosphorothioate linkages; a 5' cap (e.g., a 7-methylguanylate cap (m7G)), 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 aminoalkyl phosphoramidates, phosphorodiamidates, thionophosphor amidates, thionoalkylphosphonates, thionoalkylphosphotriesters, selenophosphates and borano-phosphates 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; phosphorothioate and/or heteroatom internucleoside linkages, such as —CH<sub>2</sub>—NH—O—CH<sub>2</sub>—, -CH<sub>2</sub>—N(CH<sub>3</sub>)—O—CH<sub>2</sub>— (known as a methylene (methylimino) or MMI backbone), —CH<sub>2</sub>—O—N(CH<sub>3</sub>)—CH<sub>2</sub>—, —CH<sub>2</sub>—N(CH<sub>3</sub>)—N(CH<sub>3</sub>)—CH<sub>2</sub>—and—O—N(CH<sub>3</sub>)—CH<sub>2</sub>—CH<sub>2</sub>— (wherein the native phosphodiester internucleotide linkage is represented as —O—P(=O)(OH)—O—CH<sub>2</sub>—); morpholino linkages (formed in part from the sugar portion of a nucleoside); morpholino backbones; phosphorodiamidate or other non-phosphodiester internucleoside linkages; siloxane backbones; sulfide, sulfoxide and sulfone backbones; formacetyl and thioformacetyl backbones; methylene formacetyl and thioformacetyl backbones; riboacetyl backbones; alkene containing backbones; sulfamate backbones; meth-

yleneimino and methylenehydrazino backbones; sulfonate and sulfonamide backbones; amide backbones; other backbone modifications having mixed N, O, S and CH<sub>2</sub> component parts; and combinations thereof.

[0222] In some embodiments, guide nucleic acids described herein can be selected from a group of non-naturally occurring guide nucleic acids that have been tiled against the nucleic acid sequence of a strain of an infection or genomic locus of interest. Often, guide nucleic acids that are tiled against the nucleic acid of a strain of an infection or genomic locus of interest can be pooled for use in a method described herein. Often, these guide nucleic acids are pooled for detecting a target nucleic acid or segment thereof in a single assay. The pooling of guide nucleic acids that are tiled against a single target nucleic acid or segment thereof can enhance the detection of the target nucleic acid using the methods described herein. The pooling of guide nucleic acids that are tiled against a single target nucleic acid or segment thereof can ensure broad coverage of the target nucleic acid or segment thereof within a single reaction using the methods described herein. In some embodiments, the tiling is sequential along the target nucleic acid or segment thereof. In some embodiments, the tiling is overlapping along the target nucleic acid or segment thereof. In some embodiments, the tiling comprises gaps between the tiled non-naturally occurring guide nucleic acids along the target nucleic acid or segment thereof. In some embodiments, the tiling of the guide nucleic acids is non-sequential. Often, a method for detecting a target nucleic acid comprises contacting a target nucleic acid to a pool of guide nucleic acids and a programmable nuclease, wherein a guide nucleic acid of the pool of guide nucleic acids has a sequence selected from a group of tiled guide nucleic acid that correspond to nucleic acids of a target nucleic acid or segment thereof; and assaying for a signal produced by cleavage of at least some reporters of a population of reporters. Pooling of guide nucleic acids can ensure broad spectrum identification, or broad coverage, of a target species within a single reaction. In some embodiments, the guide pooling comprises guide nucleic acids that produce the best signal in a DETECTR reaction (e.g., top 10 gRNAs). In some embodiments, there is an increased signal to noise ratio as the number of pooled gRNAs increases (e.g., signal to noise for 1 gRNA <2 pooled gRNAs <3 pooled gRNAs<4 pooled gRNAs<5 pooled gRNAs<6 pooled gRNAs<7 pooled gRNAs<8 pooled gRNAs<9 pooled gRNAs<10 pooled gRNAs).

#### Target Nucleic Acids

[0223] In some embodiments, the target nucleic acid is a double stranded nucleic acid. In some embodiments, the target nucleic acid is a single stranded nucleic acid. Alternatively, or in combination, the target nucleic acid is a double stranded nucleic acid and is prepared into single stranded nucleic acids before or upon contacting an RNP. In some embodiments, the single stranded nucleic acid comprises a RNA, wherein the RNA comprises a mRNA, a rRNA, a tRNA, a non-coding RNA, a long non-coding RNA, a microRNA (miRNA), and a single-stranded RNA (ssRNA). In some embodiments, the target nucleic acid is a dsRNA or a crRNA. In some embodiments, the target nucleic acid is complementary DNA (cDNA) synthesized from a single-stranded RNA template in a reaction catalyzed by a reverse transcriptase. In some embodiments, the target

nucleic acid is or mRNA. In some embodiments, the target nucleic acid is from a virus, a parasite, or a bacterium described herein.

[0224] In some embodiments, a target nucleic acid comprising a target sequence comprises a PAM sequence. In some embodiments, the PAM sequence is 3' to the target sequence. In some embodiments, the PAM sequence is directly 3' to the target sequence. In some embodiments, the PAM sequence 5' to the target sequence. In some embodiments, the PAM sequence is directly 5' to the target sequence. In some embodiments, the target nucleic acid as described in the methods herein does not initially comprise a PAM sequence. However, any target nucleic acid of interest may be generated using the methods described herein to comprise a PAM sequence, and thus be a PAM target nucleic acid. A PAM target nucleic acid, as used herein, refers to a target nucleic acid that has been amplified to insert a PAM sequence that is recognized by an effector protein system.

[0225] In some embodiments, a target nucleic acid comprises 5 to 100, 5 to 90, 5 to 80, 5 to 70, 5 to 60, 5 to 50, 5 to 40, 5 to 30, 5 to 25, 5 to 20, 5 to 15, or 5 to 10 linked nucleotides.

[0226] In some embodiments, the target nucleic acid comprises 10 to 90, 20 to 80, 30 to 70, or 40 to 60 linked nucleotides. In some embodiments, the target nucleic acid comprises 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 45, 50, 60, 70, 80, 90, or 100 linked nucleotides. In some embodiments, the target nucleic acid comprises at least 10, at least 20, at least 30, at least 40, at least 50, at least 60, at least 70, at least 80, at least 90, or at least 100 linked nucleotides.

[0227] In some embodiments, compositions, systems, and methods described herein comprise a target nucleic acid may be responsible for a disease, contain a mutation (e.g., single strand polymorphism, point mutation, insertion, or deletion), be contained in an amplicon, or be uniquely identifiable from the surrounding nucleic acids (e.g., contain a unique sequence of nucleotides). In some embodiments, the target nucleic acid has undergone a modification (e.g., an editing) after contacting with an RNP. In some embodiments, the editing is a change in the sequence of the target nucleic acid. In some embodiments, the change comprises an insertion, deletion, or substitution of one or more nucleotides compared to the target nucleic acid that has not undergone any modification.

[0228] In some embodiments, the target nucleic acid comprises a nucleic acid sequence from a pathogen responsible for a disease. Non-limiting examples of pathogens are bacteria, a virus and a fungus. The target nucleic acid, in some embodiments, is a portion of a nucleic acid from a sexually transmitted infection or a contagious disease. In some embodiments, the target nucleic acid is a portion of a nucleic acid from a genomic locus, or any DNA amplicon, such as a reverse transcribed mRNA or a cDNA from a gene locus, a transcribed mRNA, or a reverse transcribed cDNA from a gene locus in at least one of: human immunodeficiency virus (HIV), human papillomavirus (HPV), chlamydia, gonorrhea, syphilis, trichomoniasis, sexually transmitted infection, malaria, Dengue fever, Ebola, chikungunya, and leishmaniasis. Pathogens include viruses, fungi, helminths, protozoa, malarial parasites, Plasmodium parasites, Toxoplasma parasites, and Schistosoma parasites.

Helminths include roundworms, heartworms, and phytophagous nematodes, flukes, Acanthocephala, and tapeworms. Protozoan infections include infections from *Giardia* spp., *Trichomonas* spp., African trypanosomiasis, amoebic dysentery, babesiosis, balantidial dysentery, Chaga's disease, coccidioides, malaria and toxoplasmosis. Examples of pathogens such as parasitic/protozoan pathogens include, but are not limited to: *Plasmodium falciparum*, *P. vivax*, *Trypanosoma cruzi* and *Toxoplasma gondii*. Fungal pathogens include, but are not limited to *Cryptococcus neoformans*, *Histoplasma capsulatum*, *Coccidioides immitis*, *Blastomyces dermatitidis*, *Chlamydia trachomatis*, and *Candida albicans*. Pathogenic viruses include but are not limited to coronavirus (e.g., SARS-CoV-2); immunodeficiency virus (e.g., HIV); influenza virus; dengue; West Nile virus; herpes virus; yellow fever virus; Hepatitis Virus C; Hepatitis Virus A; Hepatitis Virus B; papillomavirus; and the like. Pathogens include, e.g., HIV virus, *Mycobacterium tuberculosis*, *Streptococcus agalactiae*, methicillin-resistant *Staphylococcus aureus*, *Legionella pneumophila*, *Streptococcus pyogenes*, *Escherichia coli*, *Neisseria gonorrhoeae*, *Neisseria meningitidis*, *Pneumococcus*, *Cryptococcus neoformans*, *Histoplasma capsulatum*, *Hemophilus influenzae* B, *Treponema pallidum*, Lyme disease spirochetes, *Pseudomonas aeruginosa*, *Mycobacterium leprae*, *Brucella abortus*, rabies virus, influenza virus, cytomegalovirus, herpes simplex virus I, herpes simplex virus II, human serum parvo-like virus, respiratory syncytial virus (RSV), *M. genitalium*, *T. vaginalis*, varicella-zoster virus, hepatitis B virus, hepatitis C virus, measles virus, adenovirus, human T-cell leukemia viruses, Epstein-Barr virus, murine leukemia virus, mumps virus, vesicular stomatitis virus, Sindbis virus, lymphocytic choriomeningitis virus, wart virus, blue tongue virus, Sendai virus, feline leukemia virus, Reovirus, polio virus, simian virus 40, mouse mammary tumor virus, dengue virus, rubella virus, West Nile virus, *Plasmodium falciparum*, *Plasmodium vivax*, *Toxoplasma gondii*, *Trypanosoma rangeli*, *Trypanosoma cruzi*, *Trypanosoma rhodesiense*, *Trypanosoma brucei*, *Schistosoma mansoni*, *Schistosoma japonicum*, *Babesia bovis*, *Eimeria tenella*, *Onchocerca volvulus*, *Leishmania tropica*, *Mycobacterium tuberculosis*, *Trichinella spiralis*, *Theileria parva*, *Taenia hydatigena*, *Taenia ovis*, *Taenia saginata*, *Echinococcus granulosus*, *Mesocestoides corti*, *Mycoplasma arthritidis*, *M. hyorhinis*, *M. orale*, *M. arginini*, *Acholeplasma laidlawii*, *M. salivarium* and *M. pneumoniae*. In some embodiments, the target sequence is a portion of a nucleic acid from a genomic locus, a transcribed mRNA, or a reverse transcribed cDNA from a gene locus of bacterium or other agents responsible for a disease in the sample comprising a mutation that confers resistance to a treatment, such as a single nucleotide mutation that confers resistance to antibiotic treatment.

**[0229]** In some embodiments, the target nucleic acid comprises a nucleic acid sequence of a virus, a bacterium, or other pathogen responsible for a disease in a plant (e.g., a crop). Methods and compositions of the disclosure may be used to treat or detect a disease in a plant. For example, the methods of the disclosure may be used to target a viral nucleic acid sequence in a plant. An effector protein of the disclosure may cleave the viral nucleic acid. In some embodiments, the target nucleic acid comprises a nucleic acid sequence of a virus or a bacterium or other agents (e.g., any pathogen) responsible for a disease in the plant (e.g., a crop).

In some embodiments, the target nucleic acid comprises RNA. The target nucleic acid, in some embodiments, is a portion of a nucleic acid from a virus or a bacterium or other agents responsible for a disease in the plant (e.g., a crop). In some embodiments, the target nucleic acid is a portion of a nucleic acid from a genomic locus, or any NA amplicon, such as a reverse transcribed mRNA or a cDNA from a gene locus, a transcribed mRNA, or a reverse transcribed cDNA from a gene locus in at a virus or a bacterium or other agents (e.g., any pathogen) responsible for a disease in the plant (e.g., a crop). A virus infecting the plant may be an RNA virus. A virus infecting the plant may be a DNA virus. Non-limiting examples of viruses that may be targeted with the disclosure include Tobacco mosaic virus (TMV), Tomato spotted wilt virus (TSWV), Cucumber mosaic virus (CMV), Potato virus Y (PVY), Cauliflower mosaic virus (CaMV) (RT virus), Plum pox virus (PPV), Brome mosaic virus (BMV) and Potato virus X (PVX).

**[0230]** In some embodiments, a target nucleic acid may be a cancer gene or gene associated with a genetic disorder, or an amplicon thereof, as described herein. In some embodiments, the target nucleic acid is a viral nucleic acid or a bacterial nucleic acid. In some embodiments, the target nucleic acid is a viral nucleic acid. In some embodiments, the target nucleic acid is derived from a papovavirus, a human papillomavirus (HPV), a hepadnavirus, a Hepatitis B Virus (HBV), a herpesvirus, a varicella zoster virus (VZV), an Epstein Barr virus (EBV), a Kaposi's sarcoma-associated herpesvirus, an adenovirus, a poxvirus, a parvovirus, an influenza virus, a respiratory syncytial virus, or a coronaviruses. In some embodiments, the target nucleic acid is derived from a severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). In some embodiments, the target nucleic acid is derived from a human cell. In some embodiments, the target nucleic acid is a human fetal nucleic acid or a cancer cell nucleic acid.

**[0231]** The target nucleic acid disclosed herein can be from a genomic locus. The target nucleic acid disclosed herein can be a gene of a monkeypox virus or a segment thereof. In some embodiments, the gene of the monkeypox virus is conserved in monkeypox virus and does not exist in other Orthopoxviruses.

**[0232]** In some embodiments, the gene of the monkeypox virus is OPG123. In some embodiments, the gene of the monkeypox virus is OPG038. In some embodiments, the gene of the monkeypox virus is OPG094. In some embodiments, the gene of the monkeypox virus is OPG037. In some embodiments, the gene of the monkeypox virus is OPG151. In some embodiments, the gene of the monkeypox virus is OPG105. In some embodiments, the gene of the monkeypox virus is OPG199.

**[0233]** In some embodiments, the target nucleic acid is found in the genome of a monkeypox virus belonging to Clade I. In some embodiments, the target nucleic acid is found in the genome of a monkeypox virus belonging to Clade II. In some embodiments, the target nucleic acid is found in the genome of a monkeypox virus belonging to Clade IIa. In some embodiments, the target nucleic acid is found in the genome of a monkeypox virus belonging to Clade IIb.

**[0234]** In some embodiments, at least a portion of at least one target sequence is within about 1, about 5 or more, about 10 or more, about 15 or more, about 20 or more, about 25 or more, about 30 or more, about 35 or more, about 40 or

more, about 45 or more, about 50 or more, about 55 or more, about 60 or more, about 65 or more, about 70 or more, about 75 or more, about 80 or more, about 85 or more, about 90 or more, about 95 or more, about 100 or more, about 105 or more, about 110 or more, about 115 or more, about 120 or more, about 125 or more, about 130 or more, about 135 or more, about 140 or more, about 145 or more, or about 150 to about 300 nucleotides adjacent to: the 5' end of an exon; the 3' end of an exon; the 5' end of an intron; the 3' end of an intron; one or more signaling element comprising a 5'SS, a 3'SS, a premature stop codon, U1 binding sequence, U2 binding sequence, a BS, a PYT, ESE, an ISE, an ESS, an ISS; a 5' UTR; a 3' UTR; more than one of the foregoing, or any combination thereof. In some embodiments, the target nucleic acid comprises a target locus. In some embodiments, the target nucleic acid comprises more than one target loci. In some embodiments, the target nucleic acid comprises two target loci. Accordingly, in some embodiments, the target nucleic acid can comprise one or more target sequences.

[0235] In some embodiments, target nucleic acids described herein comprise a mutation. In some embodiments, a composition, system or method described herein can be used to edit a target nucleic acid comprising a mutation such that the mutation is edited to be the wild-type nucleotide or nucleotide sequence. In some embodiments, a composition, system or method described herein can be used to detect a target nucleic acid comprising a mutation. A mutation may result in the insertion of at least one amino acid in a protein encoded by the target nucleic acid. A mutation may result in the deletion of at least one amino acid in a protein encoded by the target nucleic acid. A mutation may result in the substitution of at least one amino acid in a protein encoded by the target nucleic acid. A mutation that results in the deletion, insertion, or substitution of one or more amino acids of a protein encoded by the target nucleic acid may result in misfolding of a protein encoded by the target nucleic acid. A mutation may result in a premature stop codon, thereby resulting in a truncation of the encoded protein.

[0236] Non-limiting examples of mutations are insertion-deletion (indel), a point mutation, single nucleotide polymorphism (SNP), a chromosomal mutation, a copy number mutation or variation, and frameshift mutations. In some embodiments, an indel mutation is an insertion or deletion of one or more nucleotides. In some embodiments, a point mutation comprises a substitution, insertion, or deletion. In some embodiments, a frameshift mutation occurs when the number of nucleotides in the insertion/deletion is not divisible by three, and it occurs in a protein coding region. In some embodiments, a chromosomal mutation can comprise an inversion, a deletion, a duplication, or a translocation of one or more nucleotides. In some embodiments, a copy number variation can comprise a gene amplification or an expanding trinucleotide repeat. In some embodiments, an SNP is associated with a phenotype of the sample or a phenotype of the organism from which the sample was taken. In some embodiments, an SNP is associated with altered phenotype from wild type phenotype. In some embodiments, the SNP is a synonymous substitution or a nonsynonymous substitution. In some embodiments, the nonsynonymous substitution is a missense substitution or a nonsense point mutation. In some embodiments, the synonymous substitution is a silent substitution.

[0237] In some embodiments, a target nucleic acid described herein comprises a mutation of one or more nucleotides. In some embodiments, the one or more nucleotides comprises 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or more nucleotides. In some embodiments, the mutation comprises a deletion, insertion, and/or substitution of about 5, about 10, about 15, about 20, about 25, about 30, about 35, about 40, about 45, about 50, about 55, about 60, about 65, about 70, about 75, about 80, about 85, about 90, about 95, about 100, about 200, about 300, about 400, about 500, about 600, about 700, about 800, about 900, or about 1000 nucleotides. In some embodiments, the mutation comprises a deletion, insertion, and/or substitution of 1 to 5, 5 to 10, 10 to 15, 15 to 20, 20 to 25, 25 to 30, 30 to 35, 35 to 40, 40 to 45, 45 to 50, 50 to 55, 55 to 60, 60 to 65, 65 to 70, 70 to 75, 75 to 80, 80 to 85, 85 to 90, 90 to 95, 95 to 100, 100 to 200, 200 to 300, 300 to 400, 400 to 500, 500 to 600, 600 to 700, 700 to 800, 800 to 900, 900 to 1000, 1 to 50, 1 to 100, 25 to 50, 25 to 100, 50 to 100, 100 to 500, 100 to 1000, or 500 to 1000 nucleotides. The mutation may be located in a non-coding region or a coding region of a gene, wherein the gene is a target nucleic acid. A mutation may be in an open reading frame of a target nucleic acid. In some embodiments, guide nucleic acids described herein hybridize to a portion of the target nucleic acid comprising or adjacent to the mutation.

[0238] In some embodiments, target nucleic acids comprise a mutation, wherein the mutation is a SNP. In some embodiments, the single nucleotide mutation or SNP is associated with a phenotype of the sample or a phenotype of the organism from which the sample was taken. In some embodiments, the SNP is associated with altered phenotype from wild type phenotype. In some embodiments, a single nucleotide mutation, SNP, or deletion described herein is associated with a disease, such as a genetic disease. In some embodiments, the SNP is a synonymous substitution or a nonsynonymous substitution. In some embodiments, the nonsynonymous substitution is a missense substitution or a nonsense point mutation. In some embodiments, the synonymous substitution is a silent substitution. In some embodiments, the mutation is a deletion of one or more nucleotides. In some embodiments, the single nucleotide mutation, SNP, or deletion is associated with a disease such as a genetic disorder. In some embodiments, the mutation, such as a single nucleotide mutation, a SNP, or a deletion, may be encoded in the sequence of a target nucleic acid from the germline of an organism or may be encoded in a target nucleic acid from a diseased cell.

[0239] In some embodiments, the mutation is associated with a disease, such as a genetic disorder. In some embodiments, the mutation may be encoded in the sequence of a target nucleic acid from the germline of an organism or may be encoded in a target nucleic acid from a diseased cell. In some embodiments, a target nucleic acid described herein comprises a mutation associated with a disease. In some examples, a mutation associated with a disease refers to a mutation whose presence in a subject indicates that the subject is susceptible to or suffers from, a disease, disorder, condition, or syndrome. In some examples, a mutation associated with a disease refers to a mutation which causes, contributes to the development of, or indicates the existence of the disease, disorder, condition, or syndrome. A mutation associated with a disease may also refer to any mutation which generates transcription or translation products at an

abnormal level, or in an abnormal form, in cells affected by a disease relative to a control without the disease. In some examples, a mutation associated with a disease refers to a mutation whose presence in a subject indicates that the subject is susceptible to, or suffers from, a disease, disorder, or pathological state. In some embodiments, a mutation associated with a disease, comprises the co-occurrence of a mutation and the phenotype of a disease. The mutation may occur in a gene, wherein transcription or translation products from the gene occur at a significantly abnormal level or in an abnormal form in a cell or subject harboring the mutation as compared to a non-disease control subject not having the mutation.

[0240] In some embodiments, a target nucleic acid is in a cell. In some embodiments, the cell is a single-cell eukaryotic organism; a plant cell; an algal cell; a fungal cell; an animal cell; a cell of an invertebrate animal; a cell of a vertebrate animal such as fish, amphibian, reptile, bird, and mammal; or a cell of a mammal such as a human, a non-human primate, an ungulate, a feline, a bovine, an ovine, and a caprine. In some embodiments, the cell is a eukaryotic cell. In some embodiments, the cell is a mammalian cell, a human cell, or a plant cell. In some embodiments, the cell is a human cell. In some embodiments, the human cell is a: muscle cell, liver cell, lung cell, cardiac cell, visceral cell, cardiac muscle cell, smooth muscle cell, cardiomyocyte, nodal cardiac muscle cell, smooth muscle cell, visceral muscle cell, skeletal muscle cell, myocyte, red (or slow) skeletal muscle cell, white (fast) skeletal muscle cell, intermediate skeletal muscle, muscle satellite cell, muscle stem cell, myoblast, muscle progenitor cell, induced pluripotent stem cell (iPS), or a cell derived from an iPS cell, modified to have its gene edited and differentiated into myoblasts, muscle progenitor cells, muscle satellite cells, muscle stem cells, skeletal muscle cells, cardiac muscle cells or smooth muscle cells.

[0241] In some embodiments, an effector protein-guide nucleic acid complex may comprise high selectivity for a target sequence. In some embodiments, an RNP comprise a selectivity of at least 200:1, 100:1, 50:1, 20:1, 10:1, or 5:1 for a target nucleic acid over a single nucleotide variant of the target nucleic acid. In some embodiments, an RNP may comprise a selectivity of at least 5:1 for a target nucleic acid over a single nucleotide variant of the target nucleic acid.

[0242] By leveraging such effector protein selectivity, some methods described herein may detect a target nucleic acid present in the sample in various concentrations or amounts as a target nucleic acid population. In some embodiments, the method detects at least 2 target nucleic acid populations. In some embodiments, the method detects at least 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, or 50 target nucleic acid populations. In some embodiments, the method detects 3 to 50, 5 to 40, or 10 to 25 target nucleic acid populations. In some embodiments, the method detects at least 2 individual target nucleic acids. In some embodiments, the method detects at least 3, 5, 10, 20, 30, 40, 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, or 10000 individual target nucleic acids. In some embodiments, the method detects 1 to 10,000, 100 to 8000, 400 to 6000, 500 to 5000, 1000 to 4000, or 2000 to 3000 individual target nucleic acids. In some embodiments, the method detects target nucleic acid present at least at one copy per 10 non-target nucleic acids, 102 non-target nucleic acids, 103 non-target nucleic acids, 104

non-target nucleic acids, 105 non-target nucleic acids, 106 non-target nucleic acids, 107 non-target nucleic acids, 108 non-target nucleic acids, 109 non-target nucleic acids, or 1010 non-target nucleic acids.

#### Reporter Nucleic Acid

[0243] As used herein, a reporter comprises a nucleic acid (e.g., RNA and/or DNA). In some embodiments, a reporter is double-stranded. In some embodiments, a reporter is single-stranded. In some embodiments, a reporter comprises a protein that generates a detectable signal or signal. In some embodiments, a reporter is operably linked to the protein that generates a signal. In some embodiments, a signal is a calorimetric, potentiometric, amperometric, optical (e.g., fluorescent, colorimetric, etc.), or piezo-electric signal. In some embodiments, the reporter comprises a detection moiety. In some embodiments, the reporter is configured to release a detection moiety or generate a signal indicative of a presence or absence of the target nucleic acid. For example, the signal can indicate a presence of the target nucleic acid in the sample, and an absence of the signal can indicate an absence of the target nucleic acid in the sample. In some embodiments, suitable detectable labels and/or moieties provide a signal. In some embodiments, non-limiting example of a suitable detectable label and/or moiety comprises an enzyme, a radioisotope, a member of a specific binding pair; a fluorophore; a fluorescent protein; and a quantum dot.

[0244] In some embodiments, methods described herein utilize a reporter. By way of non-limiting and illustrative example, a reporter may comprise a nucleic acid and a detection moiety (e.g., a labeled single stranded RNA reporter), wherein the nucleic acid is capable of being cleaved by an effector protein (e.g., a CRISPR/Cas protein as disclosed herein) or a multimeric complex thereof, releasing the detection moiety, and generating a detectable signal.

[0245] In some embodiments, the methods described herein utilize the indiscriminate trans-cleavage of the reporter (e.g., comprising a reporter nucleic acid and a detectable moiety) catalyzed by the effector protein (activated upon hybridization of a guide nucleic acid to a target nucleic acid) to cause the release of the detection moiety and hence the generation of the detectable signal, indicating the presence of the target nucleic acid. Cleaving the “reporter” may be referred to herein as cleaving the “reporter nucleic acid,” the “reporter molecule,” or the “nucleic acid of the reporter.” Reporters may comprise RNA. Reporters may comprise DNA. Reporters may be double-stranded. Reporters may be single-stranded. In some embodiments, the reporter has one or more regions of single strandedness, and/or one or more regions of double strandedness.

[0246] In some embodiments, reporters comprise a protein capable of generating a signal.

[0247] A signal may be a calorimetric, potentiometric, amperometric, optical (e.g., fluorescent, colorimetric, etc.), or piezo-electric signal. In some embodiments, the reporter comprises a detection moiety. Suitable detectable labels and/or moieties that may provide a signal include, but are not limited to, an enzyme, an enzyme-substrate, a radioisotope, a member of a specific binding pair (e.g., biotin or avidin), a fluorophore, a fluorescent protein, a quantum dot, and the like.

[0248] In some embodiments, the reporter comprises a detection moiety and a quenching moiety. In some embodi-

ments, the reporter comprises a cleavage site, wherein the detection moiety is located at a first site on the reporter and the quenching moiety is located at a second site on the reporter, wherein the first site and the second site are separated by the cleavage site. Sometimes the quenching moiety is a fluorescence quenching moiety. In some embodiments, the quenching moiety is 5' to the cleavage site and the detection moiety is 3' to the cleavage site. In some embodiments, the detection moiety is 5' to the cleavage site and the quenching moiety is 3' to the cleavage site. Sometimes the quenching moiety is at the 5' terminus of the nucleic acid of a reporter. Sometimes the detection moiety is at the 3' terminus of the nucleic acid of a reporter. In some embodiments, the detection moiety is at the 5' terminus of the nucleic acid of a reporter. In some embodiments, the quenching moiety is at the 3' terminus of the nucleic acid of a reporter.

**[0249]** Suitable fluorescent proteins include, but are not limited to, green fluorescent protein (GFP) or variants thereof, blue fluorescent variant of GFP (BFP), cyan fluorescent variant of GFP (CFP), yellow fluorescent variant of GFP (YFP), enhanced GFP (EGFP), enhanced CFP (ECFP), enhanced YFP (EYFP), GFPs65T, Emerald, Topaz (TYFP), Venus, Citrine, mCitrine, GFPuv, destabilised EGFP (dEGFP), destabilised ECFP (dECFP), destabilised EYFP (dEYFP), mCFPm, Cerulean, T-Sapphire, CyPet, YPet, mKO, HcRed, t-HcRed, DsRed, DsRed2, DsRed-monomer, J-Red, dimer2, t-dimer2 (12), mRFP1, picoiloporin, Renilla GFP, Monster GFP, paGFP, Kaede protein and kindling protein, Phycobiliproteins and Phycobiliprotein conjugates including B-Phycoerythrin, R-Phycoerythrin and Allophycocyanin. Suitable enzymes include, but are not limited to, horseradish peroxidase (HRP), alkaline phosphatase (AP), beta-galactosidase (GAL), glucose-6-phosphate dehydrogenase, beta-N-acetylglucosaminidase, beta-glucuronidase, invertase, Xanthine Oxidase, firefly luciferase, and glucose oxidase (GO).

**[0250]** In some embodiments, the detection moiety comprises an invertase. The substrate of the invertase may be sucrose. A DNS reagent may be included in the system to produce a colorimetric change when the invertase converts sucrose to glucose. In some embodiments, the reporter nucleic acid and invertase are conjugated using a heterobifunctional linker by sulfo-SMCC chemistry.

**[0251]** Suitable fluorophores may provide a detectable fluorescence signal in the same range as 6-Fluorescein (Integrated DNA Technologies), IRDye 700 (Integrated DNA Technologies), TYE 665 (Integrated DNA Technologies), Alex Fluor 594 (Integrated DNA Technologies), or ATTO TM 633 (NHS Ester) (Integrated DNA Technologies). Non-limiting examples of fluorophores are fluorescein amidite, 6-Fluorescein, IRDye 700, TYE 665, Alex Fluor 594, or ATTO TM 633 (NHS Ester). The fluorophore may be

an infrared fluorophore. The fluorophore may emit fluorescence in the range of 500 nm and 720 nm. In some embodiments, the fluorophore emits fluorescence at a wavelength of 700 nm or higher. In other embodiments, the fluorophore emits fluorescence at about 665 nm. In some embodiments, the fluorophore emits fluorescence in the range of 500 nm to 520 nm, 500 nm to 540 nm, 500 nm to 590 nm, 590 nm to 600 nm, 600 nm to 610 nm, 610 nm to 620 nm, 620 nm to 630 nm, 630 nm to 640 nm, 640 nm to 650 nm, 650 nm to 660 nm, 660 nm to 670 nm, 670 nm to 680 nm, 690 nm to 690 nm, 690 nm to 700 nm, 700 nm to 710 nm, 710 nm to 720 nm, or 720 nm to 730 nm. In some embodiments, the fluorophore emits fluorescence in the range 450 nm to 750 nm, 500 nm to 650 nm, or 550 to 650 nm.

**[0252]** In some embodiments, the reporter comprises a quenching moiety. A quenching moiety may be chosen based on its ability to quench the detection moiety. A quenching moiety may be a non-fluorescent fluorescence quencher. A quenching moiety may quench a detection moiety that emits fluorescence in the range of 500 nm and 720 nm. A quenching moiety may quench a detection moiety that emits fluorescence in the range of 500 nm and 720 nm. In some embodiments, the quenching moiety quenches a detection moiety that emits fluorescence at a wavelength of 700 nm or higher. In other embodiments, the quenching moiety quenches a detection moiety that emits fluorescence at about 660 nm or about 670 nm. In some embodiments, the quenching moiety quenches a detection moiety that emits fluorescence in the range of 500 to 520, 500 to 540, 500 to 590, 590 to 600, 600 to 610, 610 to 620, 620 to 630, 630 to 640, 640 to 650, 650 to 660, 660 to 670, 670 to 680, 690 to 690, 690 to 700, 700 to 710, 710 to 720, or 720 to 730 nm. In some embodiments, the quenching moiety quenches a detection moiety that emits fluorescence in the range 450 nm to 750 nm, 500 nm to 650 nm, or 550 to 650 nm. A quenching moiety may quench fluorescein amidite, 6-Fluorescein, IRDye 700, TYE 665, Alex Fluor 594, or ATTO TM 633 (NHS Ester). A quenching moiety may be Iowa Black RQ, Iowa Black FQ or IRDye QC-1 Quencher. A quenching moiety may quench fluorescein amidite, 6-Fluorescein (Integrated DNA Technologies), IRDye 700 (Integrated DNA Technologies), TYE 665 (Integrated DNA Technologies), Alex Fluor 594 (Integrated DNA Technologies), or ATTO TM 633 (NHS Ester) (Integrated DNA Technologies). A quenching moiety may be Iowa Black RQ (Integrated DNA Technologies), Iowa Black FQ (Integrated DNA Technologies) or IRDye QC-1 Quencher (LiCor). Any of the quenching moieties described herein may be from any commercially available source, may be an alternative with a similar function, a generic, or a non-tradename of the quenching moieties listed. Table 2 provides exemplary single stranded reporter nucleic acids that may be used in the methods disclosed herein.

TABLE 2

Non-limiting list of exemplary single stranded reporter nucleic acids that may be used in the methods disclosed herein.		
5' Detection Moiety*	Sequence (SEQ ID NO.)	3' Quencher*
/5'-FAM/	rUrUrUrUrU	/3'IABKFQ/
/5'TYE665/	rUrUrUrUrU	/3'IAbRQSp/

TABLE 2-continued

Non-limiting list of exemplary single stranded reporter nucleic acids that may be used in the methods disclosed herein.		
5' Detection Moiety*	Sequence (SEQ ID NO:)	3' Quencher*
/5Alex594N/	rUrUrUrUrU	/3IAbRQSp/
/5ATTO633N/	rUrUrUrUrU	/3IAbRQSp/
/56-FAM/	rUrUrUrUrUrUrUrU	/3IABKFQ/
/5IRD700/	rUrUrUrUrUrUrUrU	/3IRQC1N/
/5TYE665/	rUrUrUrUrUrUrUrU	/3IAbRQSp/
/5Alex594N/	rUrUrUrUrUrUrUrU	/3IAbRQSp/
/5ATTO633N/	rUrUrUrUrUrUrUrU	/3IAbRQSp/
/56-FAM/	rUrUrUrUrUrUrUrU (SEQ ID NO: 73)	/3IABKFQ/
/5IRD700/	rUrUrUrUrUrUrUrUrU (SEQ ID NO: 74)	/3IRQC1N/
/5TYE665/	rUrUrUrUrUrUrUrUrU (SEQ ID NO: 75)	/3IAbRQSp/
/5Alex594N/	rUrUrUrUrUrUrUrUrU (SEQ ID NO: 76)	/3IAbRQSp/
/5ATTO633N/	rUrUrUrUrUrUrUrUrU (SEQ ID NO: 77)	/3IAbRQSp/
/56-FAM/	TTTTrUrUTTTT (SEQ ID NO: 78)	/3IABKFQ/
/5IRD700/	TTTTrUrUTTTT (SEQ ID NO: 79)	/3IRQC1N/
/5TYE665/	TTTTrUrUTTTT (SEQ ID NO: 80)	/3IAbRQSp/
/5Alex594N/	TTTTrUrUTTTT (SEQ ID NO: 81)	/3IAbRQSp/
/5ATTO633N/	TTTTrUrUTTTT (SEQ ID NO: 82)	/3IAbRQSp/
/56-FAM/	TTrUrUTT	/3IABKFQ/
/5IRD700/	TTrUrUTT	/3IRQC1N/
/5TYE665/	TTrUrUTT	/3IAbRQSp/
/5Alex594N/	TTrUrUTT	/3IAbRQSp/
/5ATTO633N/	TTrUrUTT	/3IAbRQSp/
/56-FAM/	TArArUGC	/3IABkFQ/
/5IRD700/	TArArUGC	/3IRQC1N/
/5TYE665/	TArArUGC	/3IAbRQSp/
/5Alex594N/	TArArUGC	/3IAbRQSp/
/5ATTO633N/	TArArUGC	/3IAbRQSp/
/56-FAM/	TArUrGGC	/3IABKFQ/
/5IRD700/	TArUrGGC	/3IRQC1N/
/5TYE665/	TArUrGGC	/3IAbRQSp/
/5Alex594N/	TArUrGGC	/3IAbRQSp/
/5ATTO633N/	TArUrGGC	/3IAbRQSp/
/56-FAM/	rUrUrUrUrU	/3IABKFQ/
/5IRD700/	rUrUrUrUrU	/3IRQC1N/

TABLE 2-continued

Non-limiting list of exemplary single stranded reporter nucleic acids that may be used in the methods disclosed herein.

5' Detection Moiety*	Sequence (SEQ ID NO:)	3' Quencher*
/5TYE665/	rUrUrUrUrU	/3IAbRQSp/
/5Alex594N/	rUrUrUrUrU	/3IAbRQSp/
/5ATTO633N/	rUrUrUrUrU	/3IAbRQSp/
/56-FAM/	TTATTATT	/3IABKFQ/
/56-FAM/	TTATTATT	/3IABKFQ/
/5IRD700/	TTATTATT	/3IRQC1N/
/5TYE665/	TTATTATT	/3IAbRQSp/
/5Alex594N/	TTATTATT	/3IAbRQSp/
/5ATTO633N/	TTATTATT	/3IAbRQSp/
/56-FAM/	TTTTTT	/3IABKFQ/
/56-FAM/	TTTTTTT	/3IABKFQ/
/56-FAM/	TTTTTTTTT (SEQ ID NO: 83)	/3IABKFQ/
/56-FAM/	TTTTTTTTTTT (SEQ ID NO: 84)	/3IABKFQ/
/56-FAM/	TTTTTTTTTTTTT (SEQ ID NO: 85)	/3IABKFQ/
/56-FAM/	AAAAAA	/3IABKFQ/
/56-FAM/	CCCCCC	/3IABKFQ/
/56-FAM/	GGGGGG	/3IABKFQ/
/56-FAM/	TTATTATT	/3IABKFQ/

/56-FAM/: 5' 6-Fluorescein (Integrated DNA Technologies)  
 /3IABKFQ/: 3' Iowa Black FQ (Integrated DNA Technologies)  
 /5IRD700/: 5' IRDye 700 (Integrated DNA Technologies)  
 /5TYE665/: 5' TYE 665 (Integrated DNA Technologies)  
 /5Alex594N/: 5' Alexa Fluor 594 (NHS Ester) (Integrated DNA Technologies)  
 /5Alex488N/: 5' Alexa Fluor 488 (NHS Ester) (Integrated DNA Technologies)  
 /5ATTO633N/: 5' ATTO TM 633 (NHS Ester) (Integrated DNA Technologies)  
 /3IRQCIN/: 3' IRDye QC-1 Quencher (Li-Cor)  
 /3IAbRQSp/: 3' Iowa Black RQ (Integrated DNA Technologies)  
 rU: uracil ribonucleotide  
 rG: guanine ribonucleotide

\*This Table refers to the detection moiety and quencher moiety as their tradenames and their source is identified. However, alternatives, generics, or non-tradename moieties with similar function from other sources can also be used.

**[0253]** The generation of the detectable signal from the release of the detection moiety may indicate that cleavage by the effector protein has occurred and that the sample contains the target nucleic acid. In some embodiments, the detection moiety comprises a fluorescent dye.

**[0254]** Sometimes the detection moiety comprises a fluorescence resonance energy transfer (FRET) pair. In some embodiments, the detection moiety comprises an infrared (IR) dye. In some embodiments, the detection moiety comprises an ultraviolet (UV) dye. Alternatively, or in combination, the detection moiety comprises a protein. Sometimes the detection moiety comprises a biotin. Sometimes the detection moiety comprises at least one of avidin or streptavidin. In some embodiments, the detection moiety comprises a polysaccharide, a polymer, or a nanoparticle. In some embodiments, the detection moiety comprises a gold nanoparticle or a latex nanoparticle.

**[0255]** In some embodiments, a detection moiety comprises any moiety that generates a detectable product or detectable signal upon cleavage of the reporter by the effector protein. In some embodiments, the detectable product comprises a detectable unit generated from the detectable moiety and that emits a detectable signal as described herein. In some embodiments, the detectable product further comprises a detectable label, a fluorophore, a reporter, or a combination thereof. In some embodiments, the detectable product comprises RNA, DNA, or both. In some embodiments, the detectable product is configured to generate a signal indicative of the presence or absence of the target nucleic acid in, for instance, a cell or a sample.

**[0256]** A detection moiety may be any moiety capable of generating a calorimetric, potentiometric, amperometric, optical (e.g., fluorescent, colorimetric, etc.), or piezo-electric signal. A nucleic acid of a reporter, sometimes, is protein-

nucleic acid that is capable of generating a calorimetric, potentiometric, amperometric, optical (e.g., fluorescent, colorimetric, etc.), or piezo-electric signal upon cleavage of the nucleic acid. Often a calorimetric signal is heat produced after cleavage of the nucleic acids of a reporter. Sometimes, a calorimetric signal is heat absorbed after cleavage of the nucleic acids of a reporter. A potentiometric signal, for example, is electrical potential produced after cleavage of the nucleic acids of a reporter. An amperometric signal may be movement of electrons produced after the cleavage of nucleic acid of a reporter. Often, the signal is an optical signal, such as a colorimetric signal or a fluorescence signal. An optical signal is, for example, a light output produced after the cleavage of the nucleic acids of a reporter. Sometimes, an optical signal is a change in light absorbance between before and after the cleavage of nucleic acids of a reporter. Often, a piezo-electric signal is a change in mass between before and after the cleavage of the nucleic acid of a reporter.

[0257] The detectable signal may be a colorimetric signal or a signal visible by eye. In some embodiments, the detectable signal may be fluorescent, electrical, chemical, electrochemical, or magnetic. In some embodiments, the first detection signal may be generated by interaction of the detection moiety to the capture molecule in the detection region, where the first detection signal indicates that the sample contained the target nucleic acid. Sometimes systems are capable of detecting more than one type of target nucleic acid, wherein the system comprises more than one type of guide nucleic acid and more than one type of reporter nucleic acid. In some embodiments, the detectable signal may be generated directly by the cleavage event. Alternatively, or in combination, the detectable signal may be generated indirectly by the signal event. Sometimes the detectable signal is not a fluorescent signal. In some embodiments, the detectable signal may be a colorimetric or color-based signal. In some embodiments, the detected target nucleic acid may be identified based on its spatial location on the detection region of the support medium. In some embodiments, the second detectable signal may be generated in a spatially distinct location than the first generated signal.

[0258] In some embodiments, the reporter nucleic acid is a single-stranded nucleic acid sequence comprising ribonucleotides. The nucleic acid of a reporter may be a single-stranded nucleic acid sequence comprising at least one ribonucleotide. In some embodiments, the nucleic acid of a reporter is a single-stranded nucleic acid comprising at least one ribonucleotide residue at an internal position that functions as a cleavage site. In some embodiments, the nucleic acid of a reporter comprises at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, or at least 10 ribonucleotide residues at an internal position. In some embodiments, the nucleic acid of a reporter comprises from 2 to 10, from 3 to 9, from 4 to 8, or from 5 to 7 ribonucleotide residues at an internal position. Sometimes the ribonucleotide residues are continuous. Alternatively, the ribonucleotide residues are interspersed in between non-ribonucleotide residues. In some embodiments, the nucleic acid of a reporter has only ribonucleotide residues. In some embodiments, the nucleic acid of a reporter has only DNA residues. In some embodiments, the nucleic acid comprises nucleotides resistant to cleavage by the effector protein described herein. In some embodiments, the nucleic acid of a reporter

comprises synthetic nucleotides. In some embodiments, the nucleic acid of a reporter comprises at least one ribonucleotide residue and at least one non-ribonucleotide residue.

[0259] In some embodiments, the nucleic acid of a reporter comprises at least one uracil ribonucleotide. In some embodiments, the nucleic acid of a reporter comprises at least two uracil ribonucleotides. Sometimes the nucleic acid of a reporter has only uracil ribonucleotides. In some embodiments, the nucleic acid of a reporter comprises at least one adenine ribonucleotide. In some embodiments, the nucleic acid of a reporter comprises at least two adenine ribonucleotides. In some embodiments, the nucleic acid of a reporter has only adenine ribonucleotides. In some embodiments, the nucleic acid of a reporter comprises at least one cytosine ribonucleotide. In some embodiments, the nucleic acid of a reporter comprises at least two cytosine ribonucleotides. In some embodiments, the nucleic acid of a reporter comprises at least one guanine ribonucleotide. In some embodiments, the nucleic acid of a reporter comprises at least two guanine ribonucleotides. In some embodiments, a nucleic acid of a reporter comprises a single unmodified ribonucleotide. In some embodiments, a nucleic acid of a reporter comprises only unmodified DNAs.

[0260] In some embodiments, the nucleic acid of a reporter is 5 to 20, 5 to 15, 5 to 10, 7 to 20, 7 to 15, or 7 to 10 nucleotides in length. In some embodiments, the nucleic acid of a reporter is 3 to 20, 4 to 10, 5 to 10, or 5 to 8 nucleotides in length. In some embodiments, the nucleic acid of a reporter is 5 to 12 nucleotides in length. In some embodiments, the reporter nucleic acid is at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least 16, at least 17, at least 18, at least 19, at least 20, at least 21, at least 22, at least 23, at least 24, at least 25, at least 26, at least 27, at least 28, at least 29, or at least 30 nucleotides in length. In some embodiments, the reporter nucleic acid is 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, or 30 nucleotides in length.

[0261] In some embodiments, the methods disclosed herein utilize a plurality of reporters. The plurality of reporters may comprise a plurality of signals. In some embodiments, systems comprise at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least 20, at least 30, at least 40, or at least 50 reporters. In some embodiments, there are 2 to 50, 3 to 40, 4 to 30, 5 to 20, or 6 to 10 different reporters.

[0262] In some embodiments, the methods disclosed herein utilize an effector protein and a reporter nucleic acid configured to undergo trans cleavage by the effector protein. Trans cleavage of the reporter may generate a signal from the reporter or alter a signal from the reporter. In some embodiments, the signal is an optical signal, such as a fluorescence signal or absorbance band. Trans cleavage of the reporter may alter the wavelength, intensity, or polarization of the optical signal. For example, the reporter may comprise a fluorophore and a quencher, such that trans cleavage of the reporter separates the fluorophore and the quencher thereby increasing a fluorescence signal from the fluorophore. Herein, detection of reporter cleavage to determine the presence of a target nucleic acid may be referred to as ‘DETECTR’. Further, the quantitation based on the number of positive nanovolumes as detected by reporter

cleavage to determine the amount or quantity of a target nucleic acid using the methods disclosed herein may be referred to as ‘digital DETECTR’.

[0263] In the presence of a large amount of non-target nucleic acids, an activity of an effector protein (e.g., an effector protein as disclosed herein) may be inhibited. This is because the activated effector proteins collaterally cleave any nucleic acid. If total nucleic acids are present in large amounts, they may outcompete reporters for the effector proteins. In some embodiments, the methods disclosed herein utilize an excess of reporter(s), such that the concentration of the reporter in the reaction mixture is greater than the concentration of the target nucleic acid.

[0264] In some embodiments, the concentration of the reporter is greater than the concentration of target nucleic acids and non-target nucleic acids. The non-target nucleic acids may be from the original sample, either lysed or unlysed. In some embodiments, systems comprise a reporter wherein the concentration of the reporter in a solution 1.5 fold, at least 2 fold, at least 3 fold, at least 4 fold, at least 5 fold, at least 6 fold, at least 7 fold, at least 8 fold, at least 9 fold, at least 10 fold, at least 11 fold, at least 12 fold, at least 13 fold, at least 14 fold, at least 15 fold, at least 16 fold, at least 17 fold, at least 18 fold, at least 19 fold, at least 20 fold, at least 30 fold, at least 40 fold, at least 50 fold, at least 60 fold, at least 70 fold, at least 80 fold, at least 90 fold, at least 100 fold excess of total nucleic acids.

[0265] In some embodiments, target nucleic acids may activate an effector protein to initiate sequence-independent cleavage of a nucleic acid-based reporter (e.g., a reporter comprising an RNA sequence, or a reporter comprising DNA and RNA). For example, an effector protein of the present disclosure is activated by a target nucleic acid to cleave reporters having an RNA (also referred to herein as an “RNA reporter”). Alternatively, an effector protein of the present disclosure is activated by a target nucleic acid to cleave reporters having an RNA. Alternatively, an effector protein of the present disclosure is activated by a target RNA to cleave reporters having an RNA (also referred to herein as a “RNA reporter”). The RNA reporter may comprise a single-stranded RNA labelled with a detection moiety or may be any RNA reporter as disclosed herein.

[0266] Further description of editing or detecting a target nucleic acid in a gene of interest can be found in more detail in Kim et al., “Enhancement of target specificity of CRISPR-Cas12a by using a chimeric DNA-RNA guide”, Nucleic Acids Res. 2020 Sep. 4; 48 (15): 8601-8616; Wang et al., “Specificity profiling of CRISPR system reveals greatly enhanced off-target gene editing”, Scientific Reports volume 10, Article number: 2269 (2020); Tuladhar et al., “CRISPR-Cas9-based mutagenesis frequently provokes on-target mRNA misregulation”, Nature Communications volume 10, Article number: 4056 (2019); Dong et al., “Genome-Wide Off-Target Analysis in CRISPR-Cas9 Modified Mice and Their Offspring”, G3, Volume 9, Issue 11, 1 Nov. 2019, Pages 3645-3651; Winter et al., “Genome-wide CRISPR screen reveals novel host factors required for *Staphylococcus aureus* a-hemolysin-mediated toxicity”, Scientific Reports volume 6, Article number: 24242 (2016); and Ma et al., “A CRISPR-Based Screen Identifies Genes Essential for West-Nile-Virus-Induced Cell Death”, Cell Rep. 2015 Jul. 28; 12 (4): 673-83, which are hereby incorporated by reference in their entirety.

[0267] In some embodiments, the reporter nucleic acid comprises ribonucleotides. In some embodiments, the reporter nucleic acid comprises deoxyribonucleotides. In some embodiments, the reporter nucleic acid comprises both ribonucleotides and deoxyribonucleotides. In some embodiments, the reporter nucleic acid comprises a single stranded nucleic acid. In some embodiments, the reporter nucleic acid comprises a double stranded nucleic acid. In some embodiments, the reporter nucleic acid comprises one or more regions of single stranded nucleic acid and one or more regions of double stranded nucleic acid. In some embodiments, the reporter nucleic acid comprises a modified nucleobase, a modified sugar moiety, and/or a modified nucleic acid linkage.

[0268] In some embodiments, the reporter nucleic acid comprises the nucleic acid sequence of rUrUrUrUrU. In some embodiments, the reporter nucleic acid comprises the nucleic acid sequence of TTATTATT. In some embodiments, the reporter nucleic acid comprises the nucleic acid sequence of TTTTTTTTTT (SEQ ID NO: 86). In some embodiments, the 5' end of the reporter nucleic acid is bound to a fluorescent dye, such as, Alexa Fluor 647. Non-limiting examples of fluorescent dyes include Alexa Fluor 350, Alexa Fluor 405, Alexa Fluor 488, Alexa Fluor 532, Alexa Fluor 546, Alexa Fluor 555, Alexa Fluor 561, Alexa Fluor 568, Alexa Fluor 594, Alexa Fluor 647, Alexa Fluor 660, Alexa Fluor 680, Alexa Fluor 700, and Alexa Fluor 750. In some embodiments, the 3' end of the reporter nucleic acid is bound to a dark quencher, such as, for example, 3' Iowa Black® RQ (“3IAbRQSp”).

[0269] In some embodiments, systems and methods are employed under certain conditions that enhance an activity of the effector protein relative to alternative conditions, as measured by a detectable signal released from cleavage of a reporter in the presence of the target nucleic acid. The detectable signal may be generated at about the rate of trans cleavage of a reporter nucleic acid. In some embodiments, the reporter nucleic acid is a homopolymeric reporter nucleic acid comprising 5 to 20 consecutive adenines, 5 to 20 consecutive thymines, 5 to 20 consecutive cytosines, or 5 to 20 consecutive guanines. In some embodiments, the reporter is an RNA-FQ reporter.

[0270] In some embodiments, effector proteins disclosed herein recognize, bind, or are activated by, different target nucleic acids having different sequences, but are active toward the same reporter nucleic acid, allowing for facile multiplexing in a single assay having a single ssRNA-FQ reporter.

[0271] In some embodiments, methods of detecting target nucleic acids in a sample have a threshold, wherein the threshold corresponds to a minimal amount of target nucleic acid that must be present in the sample, prior to partitioning into nanovolumes, in order for detection to occur. For example, in some embodiments, when a threshold of detection is 10 nM, then a signal can be detected when a target nucleic acid is present in the sample at a concentration of 10 nM or more. In such embodiments, the methods are not capable of detecting target nucleic acids that are present in a sample at a concentration less than 10 nM. In some embodiments, the threshold is less than or equal to 5 nM, 1 nM, 0.5 nM, 0.1 nM, 0.05 nM, 0.01 nM, 0.005 nM, 0.001 nM, 0.0005 nM, 0.0001 nM, 0.00005 nM, 0.00001 nM, 10 pM, 1 pM, 500 fM, 250 fM, 100 fM, 50 fM, 10 fM, 5 fM, 1 fM, 500 attomole (aM), 100 aM, 50 aM, 10 aM, or 1 aM.

In some embodiments, the threshold is in a range of from 1 aM to 1 nM, 1 aM to 500 pM, 1 aM to 200 pM, 1 aM to 100 pM, 1 aM to 10 pM, 1 aM to 1 pM, 1 aM to 500 fM, 1 aM to 100 fM, 1 aM to 1 fM, 1 aM to 500 aM, 1 aM to 100 aM, 1 aM to 50 aM, 1 aM to 10 aM, 10 aM to 1 nM, 10 aM to 500 pM, 10 aM to 200 pM, 10 aM to 100 pM, 10 aM to 1 pM, 10 aM to 500 fM, 10 aM to 100 fM, 10 aM to 1 fM, 10 aM to 500 aM, 10 aM to 100 aM, 10 aM to 50 aM, 100 aM to 1 nM, 100 aM to 500 pM, 100 pM to 200 pM, 100 aM to 100 pM, 100 aM to 10 pM, 100 aM to 1 pM, 100 aM to 500 fM, 100 aM to 100 fM, 100 aM to 1 fM, 100 aM to 500 aM, 500 aM to 1 nM, 500 aM to 500 pM, 500 aM to 200 pM, 500 aM to 100 pM, 500 aM to 1 pM, 500 aM to 500 fM, 500 aM to 100 fM, 500 aM to 1 fM, 1 FM to 1 nM, 1 FM to 500 pM, 1 FM to 200 pM, 1 FM to 100 pM, 1 FM to 10 pM, 1 FM to 1 pM, 500 fM to 1 nM, 500 fM to 500 pM, 500 fM to 200 pM, 500 fM to 100 pM, 500 fM to 10 pM, 500 fM to 1 pM, 800 fM to 1 nM, 800 fM to 500 pM, 800 fM to 200 pM, 800 fM to 100 pM, 800 fM to 10 pM, 800 fM to 1 pM, 1 pM to 1 nM, 1 pM to 500 pM, 1 pM to 200 pM, 1 pM to 100 pM, or 1 pM to 10 pM. In some embodiments, the threshold of detection is in a range of from 800 fM to 100 pM, 1 pM to 10 pM, 10 fM to 500 fM, 10 fM to 50 fM, 50 fM to 100 fM, 100 fM to 250 fM, or 250 fM to 500 fM. In some embodiments, the threshold is in a range of from 2 aM to 100 pM, from 20 aM to 50 pM, from 50 aM to 20 pM, from 200 aM to 5 pM, or from 500 aM to 2 pM.

[0272] In some embodiments, a minimum concentration at which the methods detect a target nucleic acid a sample is in a range of from 1 zeptomolar (zM) to 1 nM, 1 aM to 1 nM, 10 aM to 1 nM, 100 aM to 1 nM, 500 aM to 1 nM, 1 fM to 1 nM, 1 fM to 500 pM, 1 fM to 200 pM, 1 fM to 100 pM, 1 fM to 10 pM, 1 fM to 1 pM, 10 fM to 1 nM, 10 fM to 500 pM, 10 fM to 200 pM, 10 fM to 100 pM, 10 fM to 1 pM, 500 fM to 1 nM, 500 fM to 500 pM, 500 fM to 200 pM, 500 fM to 100 pM, 500 fM to 1 pM, 800 fM to 1 nM, 800 fM to 500 pM, 800 fM to 200 pM, 800 fM to 100 pM, 800 fM to 10 pM, 800 fM to 1 pM, 1 pM to 1 nM, 1 pM to 500 pM, from 1 pM to 200 pM, 1 pM to 100 pM, or 1 pM to 10 pM. In some embodiments, a minimum concentration at which the methods detect in a sample is in a range of from 2 aM to 100 pM, from 20 aM to 50 pM, from 50 aM to 20 pM, from 200 aM to 5 pM, or from 500 aM to 2 pM. In some embodiments, a minimum concentration at which the methods detect a single stranded target nucleic acid in a sample is in a range of from 1 zM to 100 pM. In some embodiments, a minimum concentration at which the methods detect a target nucleic acid in a sample is in a range of from 1 fM to 100 pM. In some embodiments, a minimum concentration at which the methods detect a single stranded target nucleic acid in a sample is in a range of from 10 fM to 100 pM. In some embodiments, a minimum concentration at which the methods detect a single stranded target nucleic acid in a sample is in a range of from 800 fM to 100 pM. In some embodiments, a minimum concentration at which the methods detect a single stranded target nucleic acid in a sample is in a range of from 1 pM to 10 pM. In some embodiments, the devices, systems, fluidic devices, kits, and methods described herein detect a single stranded target nucleic acid in a sample comprising a plurality of nucleic acids such as a plurality of non-target

nucleic acids, where the target single-stranded nucleic acid is present at a concentration as low as 1 aM, 10 aM, 100 aM, 500 aM, 1 fM, 10 fM, 500 fM, 800 fM, 1 pM, 10 pM, 100 pM, or 1 pM.

[0273] In some embodiments, a minimum concentration at which the methods detect a target nucleic acid at a concentration of about 10 nM, about 20 nM, about 30 nM, about 40 nM, about 50 nM, about 60 nM, about 70 nM, about 80 nM, about 90 nM, about 100 nM, about 200 nM, about 300 nM, about 400 nM, about 500 nM, about 600 nM, about 700 nM, about 800 nM, about 900 nM, about 1  $\mu$ M, about 10  $\mu$ M, or about 100  $\mu$ M. In some embodiments, a minimum concentration at which the methods detect a target nucleic acid at a concentration of from 10 nM to 20 nM, from 20 nM to 30 nM, from 30 nM to 40 nM, from 40 nM to 50 nM, from 50 nM to 60 nM, from 60 nM to 70 nM, from 70 nM to 80 nM, from 80 nM to 90 nM, from 90 nM to 100 nM, from 100 nM to 200 nM, from 200 nM to 300 nM, from 300 nM to 400 nM, from 400 nM to 500 nM, from 500 nM to 600 nM, from 600 nM to 700 nM, from 700 nM to 800 nM, from 800 nM to 900 nM, from 900 nM to 1  $\mu$ M, from 1  $\mu$ M to 10  $\mu$ M, from 10  $\mu$ M to 100  $\mu$ M, from 10 nM to 100 nM, from 10 nM to 1  $\mu$ M, from 10 nM to 10  $\mu$ M, from 10 nM to 100  $\mu$ M, from 100 nM to 1  $\mu$ M, from 100 nM to 10  $\mu$ M, from 100 nM to 100  $\mu$ M, or from 1  $\mu$ M to 100  $\mu$ M. In some embodiments, a minimum concentration at which the methods detect a target nucleic acid at a concentration of from 20 nM to 50  $\mu$ M, from 50 nM to 20  $\mu$ M, or from 200 nM to 5  $\mu$ M.

[0274] In some embodiments, the methods disclosed herein detect a target nucleic acid in less than 60 minutes. In some embodiments, methods detect a target nucleic acid in less than about 120 minutes, less than about 110 minutes, less than about 100 minutes, less than about 90 minutes, less than about 80 minutes, less than about 70 minutes, less than about 60 minutes, less than about 55 minutes, less than about 50 minutes, less than about 45 minutes, less than about 40 minutes, less than about 35 minutes, less than about 30 minutes, less than about 25 minutes, less than about 20 minutes, less than about 15 minutes, less than about 10 minutes, less than about 5 minutes, less than about 4 minutes, less than about 3 minutes, less than about 2 minutes, or less than about 1 minute.

[0275] In some embodiments, the detectable signal is detectable in a time period in the range of about 10 min to about 24 hours, for example, about 15 min, about 30 min, about 45 min, about 1 hour, about 90 min, about 2 hours, about 5 hours, about 10 hours, about 15 hours, about 20 hours, or about 24 hours, including all values and subranges that lie therebetween. In some embodiments, the detectable signal is detectable in less than 90 minutes. In some embodiments, the detectable signal is detectable in less than 30 minutes. In some embodiments, the detectable signal is detectable in a time period in the range of about 15 min to about 24 hours.

[0276] In some embodiments, the detectable signal is detectable in at least about 120 minutes, at least about 110 minutes, at least about 100 minutes, at least about 90 minutes, at least about 80 minutes, at least about 70 minutes, at least about 60 minutes, at least about 55 minutes, at least about 50 minutes, at least about 45 minutes, at least about 40 minutes, at least about 35 minutes, at least about 30 minutes, at least about 25 minutes, at least about 20 minutes, at least about 15 minutes, at least about 10 minutes, or at least about 5 minutes. In some embodiments, the sample is contacted

with the reagents for from 5 minutes to 120 minutes, from 5 minutes to 100 minutes, from 10 minutes to 90 minutes, from 15 minutes to 45 minutes, or from 20 minutes to 35 minutes.

[0277] In some embodiments, methods of detecting are performed in less than 10 hours, less than 9 hours, less than 8 hours, less than 7 hours, less than 6 hours, less than 5 hours, less than 4 hours, less than 3 hours, less than 2 hours, less than 1 hour, less than 50 minutes, less than 45 minutes, less than 40 minutes, less than 35 minutes, less than 30 minutes, less than 25 minutes, less than 20 minutes, less than 15 minutes, less than 10 minutes, less than 9 minutes, less than 8 minutes, less than 7 minutes, less than 6 minutes, or less than 5 minutes. In some embodiments, methods of detecting are performed in about 5 minutes to about 10 hours, about 10 minutes to about 8 hours, about 15 minutes to about 6 hours, about 20 minutes to about 5 hours, about 30 minutes to about 2 hours, or about 45 minutes to about 1 hour.

[0278] In some embodiments, the detection occurs within 5 minutes of contacting a sample and/or a target nucleic acid with a composition described herein. In some embodiments, the detection occurs within 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 70, 80, 90, 100, 110, or 120 minutes of contacting the target nucleic acid. In some embodiments, the detection occurs within 1 to 120, 5 to 100, 10 to 90, 15 to 80, 20 to 60, or 30 to 45 minutes of contacting the target nucleic acid.

#### Test Samples

[0279] Various sample types comprising a target nucleic acid of interest are consistent with the present disclosure. These samples may comprise a target nucleic acid for detection. In some embodiments, the detection of the target nucleic indicates an ailment, such as a disease, cancer, or genetic disorder, or genetic information, such as for phenotyping, genotyping, or determining ancestry and are compatible with the reagents and support mediums as described herein. Generally, a sample from an individual or an animal or an environmental sample may be obtained to test for presence of a disease, cancer, genetic disorder, or any mutation of interest.

[0280] In some embodiments, the test sample comprises DNA from a cell lysate. In some embodiments, the test sample comprises cells. In some embodiments, the test sample is a blood, serum, plasma, urine, aspirate, fecal or biopsy sample. In some embodiments, the test sample comprises, or derived from, wastewater.

[0281] In some embodiments, a sample comprises a target nucleic acid from 0.001% to 20% of total nucleic acids in the sample. In some embodiments, a sample comprises a target nucleic acid from 0.005% to 20% of total nucleic acids in the sample. In some embodiments, a sample comprises a target nucleic acid from 0.05% to 20% of total nucleic acids in the sample. In some embodiments, the target nucleic acid is 0.1% to 10% of the total nucleic acids in the sample. In some embodiments, the target nucleic acid is 0.1% to 5% of the total nucleic acids in the sample. In some embodiments, the target nucleic acid is 0.1% to 1% of the total nucleic acids in the sample. In some embodiments, the target nucleic acid is in any amount less than 100% of the total nucleic acids in the sample. In some embodiments, the target nucleic acid is 100% of the total nucleic acids in the sample. In some embodiments, the sample comprises a portion of the target

nucleic acid and at least one nucleic acid comprising less than 100% sequence identity to the portion of the target nucleic acid but no less than 50% sequence identity to the portion of the target nucleic acid. For example, the portion of the target nucleic acid comprises a mutation as compared to at least one nucleic acid comprising less than 100% sequence identity to the portion of the target nucleic acid but no less than 50% sequence identity to the portion of the target nucleic acid. In some embodiments, the portion of the target nucleic acid comprises a single nucleotide mutation as compared to at least one nucleic acid comprising less than 100% sequence identity to the portion of the target nucleic acid but no less than 50% sequence identity to the portion of the target nucleic acid.

[0282] In some embodiments, a sample comprises target nucleic acid populations at different concentrations or amounts. In some embodiments, the sample has at least 2 target nucleic acid populations. In some embodiments, the sample has at least 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, or 50 target nucleic acid populations. In some embodiments, the sample has 3 to 50, 5 to 40, or 10 to 25 target nucleic acid populations.

[0283] In some embodiments, a sample has at least 2 individual target nucleic acids. In some embodiments, the sample has at least 3, 5, 10, 20, 30, 40, 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, or 10000 individual target nucleic acids. In some embodiments, the sample comprises 1 to 10,000, 100 to 8000, 400 to 6000, 500 to 5000, 1000 to 4000, or 2000 to 3000 individual target nucleic acids.

[0284] In some embodiments, the test sample comprises about 10,000 molecules to about 100,000 molecules of the target nucleic acid. In some embodiments, the test sample comprises about 50,000 molecules of the target nucleic acid.

[0285] In some embodiments, a sample comprises one copy of target nucleic acid per 10 non-target nucleic acids, 102 non-target nucleic acids, 103 non-target nucleic acids, 104 non-target nucleic acids, 105 non-target nucleic acids, 106 non-target nucleic acids, 107 non-target nucleic acids, 108 non-target nucleic acids, 109 non-target nucleic acids, or 1010 non-target nucleic acids.

[0286] In some embodiments, samples comprise a target nucleic acid at a concentration of less than 1 nM, less than 2 nM, less than 3 nM, less than 4 nM, less than 5 nM, less than 6 nM, less than 7 nM, less than 8 nM, less than 9 nM, less than 10 nM, less than 20 nM, less than 30 nM, less than 40 nM, less than 50 nM, less than 60 nM, less than 70 nM, less than 80 nM, less than 90 nM, less than 100 nM, less than 200 nM, less than 300 nM, less than 400 nM, less than 500 nM, less than 600 nM, less than 700 nM, less than 800 nM, less than 900 nM, less than 1 µM, less than 2 µM, less than 3 µM, less than 4 µM, less than 5 µM, less than 6 µM, less than 7 µM, less than 8 µM, less than 9 µM, less than 10 µM, less than 100 M, or less than 1 mM. In some embodiments, the sample comprises a target nucleic acid at a concentration of 1 nM to 2 nM, 2 nM to 3 nM, 3 nM to 4 nM, 4 nM to 5 nM, 5 nM to 6 nM, 6 nM to 7 nM, 7 nM to 8 nM, 8 nM to 9 nM, 9 nM to 10 nM, 10 nM to 20 nM, 20 nM to 30 nM, 30 nM to 40 nM, 40 nM to 50 nM, 50 nM to 60 nM, 60 nM to 70 nM, 70 nM to 80 nM, 80 nM to 90 nM, 90 nM to 100 nM, 100 nM to 200 nM, 200 nM to 300 nM, 300 nM to 400 nM, 400 nM to 500 nM, 500 nM to 600 nM, 600 nM to 700 nM, 700 nM to 800 nM, 800 nM to 900 nM, 900 nM to 1 µM, 1 µM to 2 µM, 2 µM to 3 µM, 3 µM to 4 µM, 4 µM to

5  $\mu\text{M}$ , 5  $\mu\text{M}$  to 6  $\mu\text{M}$ , 6  $\mu\text{M}$  to 7  $\mu\text{M}$ , 7  $\mu\text{M}$  to 8  $\mu\text{M}$ , 8  $\mu\text{M}$  to 9  $\mu\text{M}$ , 9  $\mu\text{M}$  to 10  $\mu\text{M}$ , 10  $\mu\text{M}$  to 100  $\mu\text{M}$ , 100  $\mu\text{M}$  to 1 mM, 1 nM to 10 nM, 1 nM to 100 nM, 1 nM to 1  $\mu\text{M}$ , 1 nM to 10  $\mu\text{M}$ , 1 nM to 100  $\mu\text{M}$ , 1 nM to 1 mM, 10 nM to 100 nM, 10 nM to 1  $\mu\text{M}$ , 10 nM to 10  $\mu\text{M}$ , 10 nM to 100  $\mu\text{M}$ , 10 nM to 1 mM, 100 nM to 1  $\mu\text{M}$ , 100 nM to 10  $\mu\text{M}$ , 100 nM to 100  $\mu\text{M}$ , 100 nM to 1 mM, 100 nM to 10  $\mu\text{M}$ , 1  $\mu\text{M}$  to 100  $\mu\text{M}$ , 1  $\mu\text{M}$  to 1 mM, 10  $\mu\text{M}$  to 100  $\mu\text{M}$ , 10  $\mu\text{M}$  to 1 mM, or 100  $\mu\text{M}$  to 1 mM. In some embodiments, the sample comprises a target nucleic acid at a concentration of 20 nM to 200  $\mu\text{M}$ , 50 nM to 100  $\mu\text{M}$ , 200 nM to 50  $\mu\text{M}$ , 500 nM to 20  $\mu\text{M}$ , or 2  $\mu\text{M}$  to 10  $\mu\text{M}$ . In some embodiments, the target nucleic acid is not present in the sample.

[0287] In some embodiments, samples comprise fewer than 10 copies, fewer than 100 copies, fewer than 1000 copies, fewer than 10,000 copies, fewer than 100,000 copies, or fewer than 1,000,000 copies of a target nucleic acid. In some embodiments, the sample comprises 10 copies to 100 copies, 100 copies to 1000 copies, 1000 copies to 10,000 copies, 10,000 copies to 100,000 copies, 100,000 copies to 1,000,000 copies, 10 copies to 1000 copies, 10 copies to 100,000 copies, 10 copies to 1,000,000 copies, 100 copies to 10,000 copies, 100 copies to 100,000 copies, 1000 copies to 100,000 copies, or 1,000 copies to 1,000,000 copies of a target nucleic acid. In some embodiments, the sample comprises 10 copies to 500,000 copies, 200 copies to 200,000 copies, 500 copies to 100,000 copies, 1000 copies to 50,000 copies, 2000 copies to 20,000 copies, 3000 copies to 10,000 copies, or 4000 copies to 8000 copies. In some embodiments, the target nucleic acid is not present in the sample.

[0288] In some embodiments, the sample is a biological sample, an environmental sample, or a combination thereof. Non-limiting examples of biological samples are blood, serum, plasma, saliva, urine, mucosal sample, peritoneal sample, cerebrospinal fluid, gastric secretions, nasal secretions, sputum, pharyngeal exudates, urethral or vaginal secretions, an exudate, an effusion, and a tissue sample (e.g., a biopsy sample). In some embodiments, a biological sample is saliva. In some embodiments, the biological sample is taken as a swab from a wound or a lesion. A tissue sample from a subject may be dissociated or liquified prior to application to detection system of the present disclosure. Non-limiting examples of environmental samples are soil, air, or water. In some embodiments, an environmental sample is taken as a swab from a surface of interest or taken directly from the surface of interest. In some embodiments, a sample is a wastewater sample. Wastewater surveillance for a pathogen (e.g., SARS-CoV-2, monkeypox virus) can serve as an early indicator that a pathogen is spreading in a community.

[0289] In some embodiments, the sample is a raw (unprocessed, unedited, unmodified) sample. Raw samples may be applied to a system for detecting or editing a target nucleic acid, such as those described herein. In some embodiments, the sample is diluted with a buffer or a fluid or concentrated prior to its application to the system or be applied neat to the detection system. Sometimes, the sample contains no more 20  $\mu\text{l}$  of buffer or fluid. The sample, in some embodiments, is contained in no more than 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 90, 100, 200, 300, 400, 500  $\mu\text{l}$ , or any of value 1  $\mu\text{l}$  to 500  $\mu\text{l}$ , preferably 10  $\mu\text{L}$  to 200

$\mu\text{L}$ , or more preferably 50  $\mu\text{L}$  to 100  $\mu\text{L}$  of buffer or fluid. Sometimes, the sample is contained in more than 500  $\mu\text{l}$ .

[0290] In some embodiments, the sample is taken from a single-cell eukaryotic organism; a plant or a plant cell; an algal cell; a fungal cell; an animal cell, tissue, or organ; a cell, tissue, or organ from an invertebrate animal; a cell, tissue, fluid, or organ from a vertebrate animal such as fish, amphibian, reptile, bird, and mammal; a cell, tissue, fluid, or organ from a mammal such as a human, a non-human primate, an ungulate, a feline, a bovine, an ovine, and a caprine.

[0291] In some embodiments, the sample is taken from nematodes, protozoans, helminths, or malarial parasites. In some embodiments, the sample comprises nucleic acids from a cell lysate from a eukaryotic cell, a mammalian cell, a human cell, a prokaryotic cell, or a plant cell. In some embodiments, the sample comprises nucleic acids expressed from a cell.

[0292] In some embodiments, samples are used for diagnosing a disease. In some embodiments the disease is cancer. The sample used for cancer testing may comprise at least one target nucleic acid that may hybridize to a guide nucleic acid of the reagents described herein. The target nucleic acid, in some embodiments, comprises a portion of a gene comprising a mutation associated with a disease, such as cancer, a gene whose overexpression is associated with cancer, a tumor suppressor gene, an oncogene, a checkpoint inhibitor gene, a gene associated with cellular growth, a gene associated with cellular metabolism, or a gene associated with cell cycle. Sometimes, the target nucleic acid encodes a cancer biomarker. In some embodiments, the assay may be used to detect "hotspots" in target nucleic acids that may be predictive of a cancer. In some embodiments, the target nucleic acid comprises a portion of a nucleic acid that is associated with a cancer. In some embodiments, the target nucleic acid is a portion of a nucleic acid from a genomic locus, any DNA amplicon of, a reverse transcribed mRNA, or a cDNA from a locus of at least one of a gene. Any region of the aforementioned gene loci may be probed for a mutation or deletion using the compositions and methods disclosed herein. For example, in the EGFR gene locus, the compositions and methods for detection disclosed herein may be used to detect a single nucleotide polymorphism or a deletion.

[0293] The methods disclosed herein can be used as a companion diagnostic with medicaments used to treat a disease or an infection, or can be used in reagent kits, point-of-care diagnostics, or over-the-counter diagnostics. The methods may be used as a point of care diagnostic or as a lab test for detection of a target nucleic acid and, thereby, detection of a condition in a subject from which the sample was taken. The methods may be used in various sites or locations, such as in laboratories, in hospitals, in physician offices/laboratories (POLs), in clinics, at remote sites, or at home.

[0294] In some embodiments, samples are used to diagnose a genetic disorder, also referred to as genetic disorder testing. The sample used for genetic disorder testing may comprise at least one target nucleic acid that may hybridize to a guide nucleic acid of the reagents described herein. The target nucleic acid, in some embodiments, is from a gene with a mutation associated with a genetic disorder, from a gene whose overexpression is associated with a genetic disorder, from a gene associated with abnormal cellular

growth resulting in a genetic disorder, or from a gene associated with abnormal cellular metabolism resulting in a genetic disorder. In some embodiments, the target nucleic acid is a nucleic acid from a genomic locus, a transcribed mRNA, or a reverse transcribed mRNA, a DNA amplicon of or a cDNA from a locus of a gene.

[0295] A sample used for phenotyping testing may comprise at least one target nucleic acid that may hybridize to a guide nucleic acid of the reagents described herein. The target nucleic acid, in some embodiments, is a nucleic acid encoding a sequence associated with a phenotypic trait. A sample used for genotyping testing may comprise at least one target nucleic acid that may hybridize to a guide nucleic acid of the reagents described herein. A target nucleic acid, in some embodiments, is a nucleic acid encoding a sequence associated with a genotype of interest. A sample used for ancestral testing may comprise at least one target nucleic acid that may hybridize to a guide nucleic acid of the reagents described herein. A target nucleic acid, in some embodiments, is a nucleic acid encoding a sequence associated with a geographic region of origin or ethnic group. A sample may be used for identifying a disease status. For example, a sample is any sample described herein, and is obtained from a subject for use in identifying a disease status (e.g., infected with monkeypox virus or uninfected) of a subject. Sometimes, a method comprises obtaining a saliva sample or a wound swab sample from a subject; and identifying a disease status of the subject. In some embodiments, the disease is cancer. In some embodiments, the disease is a genetic disorder. In some embodiments, a method comprises obtaining a serum sample from a subject; and identifying a disease status of the subject.

#### Reaction Mixtures

[0296] In some embodiments, the reaction mixtures described herein comprise a solution or buffer; a reagent; a support medium; other components or appurtenances as described herein; or combinations thereof.

[0297] In general, the reaction mixtures comprise a solution in which the activity of an effector protein occurs. Often, the solution comprises or consists essentially of a buffer. The solution or buffer may comprise a buffering agent, a salt, a crowding agent, a detergent, a reducing agent, a competitor, or a combination thereof. Often the buffer is the primary component or the basis for the solution in which the activity occurs. Thus, concentrations for components of buffers described herein (e.g., buffering agents, salts, crowding agents, detergents, reducing agents, and competitors) are the same or essentially the same as the concentration of these components in the solution in which the activity occurs. In some embodiments, a buffer is required for cell lysis activity or viral lysis activity.

[0298] In some embodiments, the reaction mixtures comprise a buffer, wherein the buffer comprise at least one buffering agent. Exemplary buffering agents include HEPES, TRIS, MES, ADA, PIPES, ACES, MOPS, BIS-TRIS propane, BES, MOPS, TES, DISO, Trizma, TRICINE, GLY-GLY, HEPPS, BICINE, TAPS, A MPD, A MPSO, CHES, CAPSO, AMP, CAPS, phosphate, citrate, acetate, imidazole, or any combination thereof. In some embodiments, the concentration of the buffering agent in the buffer is 1 mM to 200 mM. A buffer compatible with an effector protein may comprise a buffering agent at a concentration of 10 mM to 30 mM. A buffer compatible with an effector

protein may comprise a buffering agent at a concentration of about 20 mM. A buffering agent may provide a pH for the buffer or the solution in which the activity of the effector protein occurs. The pH may be 3 to 4, 3.5 to 4.5, 4 to 5, 4.5 to 5.5, 5 to 6, 5.5 to 6.5, 6 to 7, 6.5 to 7.5, 7 to 8, 7.5 to 8.5, 8 to 9, 8.5 to 9.5, 9 to 10, or 9.5 to 10.5.

[0299] In some embodiments, the reaction mixtures comprise a solution, wherein the solution comprises at least one salt. In some embodiments, the at least one salt is selected from potassium acetate, magnesium acetate, sodium chloride, potassium chloride, magnesium chloride, calcium chloride, and any combination thereof. In some embodiments, the concentration of the at least one salt in the solution is 5 mM to 100 mM, 5 mM to 10 mM, 1 mM to 60 mM, or 1 mM to 10 mM. In some embodiments, the concentration of the at least one salt is about 105 mM. In some embodiments, the concentration of the at least one salt is about 55 mM. In some embodiments, the concentration of the at least one salt is about 7 mM. In some embodiments, the solution comprises potassium acetate and magnesium acetate. In some embodiments, the solution comprises sodium chloride and magnesium chloride. In some embodiments, the solution comprises potassium chloride and magnesium chloride. In some embodiments, the salt is a magnesium salt and the concentration of magnesium in the solution is at least 5 mM, 7 mM, at least 9 mM, at least 11 mM, at least 13 mM, or at least 15 mM. In some embodiments, the concentration of magnesium is less than 20 mM, less than 18 mM, or less than 16 mM.

[0300] In some embodiments, the reaction mixtures comprise a solution, wherein the solution comprises at least one crowding agent. A crowding agent may reduce the volume of solvent available for other molecules in the solution, thereby increasing the effective concentrations of said molecules. Exemplary crowding agents include glycerol and bovine serum albumin. In some embodiments, the crowding agent is glycerol. In some embodiments, the concentration of the crowding agent in the solution is 0.01% (v/v) to 10% (v/v). In some embodiments, the concentration of the crowding agent in the solution is 0.5% (v/v) to 10% (v/v).

[0301] In some embodiments, the reaction mixtures comprise a solution, wherein the solution comprises at least one detergent. Exemplary detergents include Tween, Triton-X, and IGEPAL. A solution may comprise Tween, Triton-X, or any combination thereof. A solution may comprise Triton-X. A solution may comprise IGEPAL CA-630. In some embodiments, the concentration of the detergent in the solution is 2% (v/v) or less. In some embodiments, the concentration of the detergent in the solution is 1% (v/v) or less. In some embodiments, the concentration of the detergent in the solution is 0.00001% (v/v) to 0.01% (v/v). In some embodiments, the concentration of the detergent in the solution is about 0.01% (v/v).

[0302] In some embodiments, the reaction mixtures comprise a solution, wherein the solution comprises at least one reducing agent. Exemplary reducing agents comprise dithiothreitol (DTT),  $\beta$ -mercaptoethanol (BME), or tris(2-carboxyethyl) phosphine (TCEP). In some embodiments, the reducing agent is DTT. In some embodiments, the concentration of the reducing agent in the solution is 0.01 mM to 100 mM. In some embodiments, the concentration of the reducing agent in the solution is 0.1 mM to 10 mM. In some embodiments, the concentration of the reducing agent in the solution is 0.5 mM to 2 mM. In some embodiments, the

concentration of the reducing agent in the solution is 0.01 mM to 100 mM. In some embodiments, the concentration of the reducing agent in the solution is 0.1 mM to 10 mM. In some embodiments, the concentration of the reducing agent in the solution is about 1 mM.

[0303] In some embodiments, the reaction mixtures comprise a solution, wherein the solution comprises a competitor. In general, competitors compete with the target nucleic acid or the reporter nucleic acid for cleavage by the effector protein or a dimer thereof. Exemplary competitors include heparin, and imidazole, and salmon sperm DNA. In some embodiments, the concentration of the competitor in the solution is 1 µg/mL to 100 µg/mL. In some embodiments, the concentration of the competitor in the solution is 40 µg/mL to 60 µg/mL.

[0304] In some embodiments, the reaction mixtures comprise a solution, wherein the solution comprises a co-factor. In some embodiments, the co-factor allows an effector protein or a multimeric complex thereof to perform a function, including pre-crRNA processing and/or target nucleic acid cleavage. The suitability of a cofactor for an effector protein or a multimeric complex thereof may be assessed, such as by methods based on those described by Sundaresan et al. (*Cell Rep.* 2017 Dec. 26; 21 (13): 3728-3739). In some embodiments, an effector or a multimeric complex thereof forms a complex with a co-factor. In some embodiments, the co-factor is a divalent metal ion. In some embodiments, the divalent metal ion is selected from Mg<sup>2+</sup>, Mn<sup>2+</sup>, Zn<sup>2+</sup>, Ca<sup>2+</sup>, Cu<sup>2+</sup>. In some embodiments, the divalent metal ion is Mg<sup>2+</sup>. In some embodiments, the co-factor is Mg<sup>2+</sup>.

[0305] In some embodiments, the reaction mixture comprises a buffer, wherein the buffer comprises tricine, MgOAc, BSA, TCEP, imidazole, KCl, MgCl<sub>2</sub>, BSA, Igepal Ca-630, glycerol, HEPES, KOAc, Triton-X 100, Tris-HCl, (NH4)2SO<sub>4</sub>, Tween-20, TMAO, or any combination thereof.

### Multiplexing

[0306] The methods, systems, compositions, and kits described herein can be multiplexed in a number of ways. These methods of multiplexing are, for example, consistent with methods, systems, compositions, and kits disclosed herein for detection of a target nucleic acid within the sample.

[0307] Multiplexing can be spatial multiplexing wherein multiple different target nucleic acids are detected at the same time, but the reactions are spatially separated. In some embodiments, the multiple target nucleic acids or segments thereof are detected using the same programmable nuclease, but different non-naturally occurring guide nucleic acids. In some embodiments, the multiple target nucleic acids are detected using the different programmable nucleases. Alternatively, multiplexing can be single reaction multiplexing wherein multiple different target acids are detected in a single reaction volume (e.g., in a single nanovolume). In some embodiments, a single population of programmable nucleases is used in single reaction multiplexing. In some embodiments, at least two different populations of programmable nucleases are used in single reaction multiplexing.

[0308] Furthermore, signals from multiplexing can be quantified. For example, a method of quantification for a disease panel comprises assaying for a plurality of unique target nucleic acids in a plurality of aliquots from a sample,

assaying for a control nucleic acid in a second aliquot of the sample, and quantifying a plurality of signals of the plurality of unique target nucleic acids by measuring signals produced by cleavage of reporters compared to the signal produced in the second aliquot. In some embodiments, the plurality of unique target nucleic acids or segments thereof are from a plurality of viruses in the sample. In some embodiments, the quantification of a signal of the plurality correlates with a concentration of a unique target nucleic acid or segment thereof of the plurality.

[0309] In some embodiments, the reagents for multiplexed assays comprise multiple non-naturally occurring guide nucleic acids, multiple programmable nucleases, and multiple single stranded reporters, where a combination of one of the non-naturally occurring guide nucleic acids, one of the programmable nucleases, and one of the single stranded reporters detects one target nucleic acid or segment thereof and can provide a detection spot on the detection region. In some embodiments, the combination of a non-naturally occurring guide nucleic acid, a programmable nuclease, and a single stranded reporter configured to detect one target nucleic acid or segment thereof is mixed with at least one other combination in a single reagent chamber. In some embodiments, the combination of a non-naturally occurring guide nucleic acid, a programmable nuclease, and a single stranded reporter configured to detect one target nucleic acid or segment thereof is mixed with at least one other combination on a single support medium. When these combinations of reagents are contacted with the sample, the reaction for the multiple target nucleic acids or segments thereof occurs simultaneously in the same medium or reagent chamber (e.g., in the same nanovolume).

[0310] In some embodiments, the combination of a non-naturally occurring guide nucleic acid, a programmable nuclease, and a single stranded reporter configured to detect one target nucleic acid or segment thereof is provided in its own reagent chamber or its own support medium (e.g., in its own nanovolume). In this case, multiple reagent chambers or support mediums are provided in the device, kit, or system, where one reagent chamber is designed to detect one target nucleic acid or segment thereof. In this case, multiple support mediums are used to detect the panel of viral infections, or other diseases of interest.

[0311] Multiplexing of a DNA-activated programmable DNA nuclease, such as a Type V CRISPR-Cas protein, with a DNA-activated programmable RNA nuclease, such as a Type VI protein, with a DNA reporter and an RNA reporter, can enable multiplexed detection of target ssDNAs or a combination of a target dsDNA and a target ssDNA, respectively. Multiplexing of different RNA-activated programmable RNA nucleases that have distinct RNA reporter cleavage preferences can enable additional multiplexing.

[0312] In some embodiments, multiplexing enables detections of different segments of the same gene at the same time. In some embodiments, multiplexing enables detections of at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100 or more different segments of the same gene at the same time. In some embodiments, multiplexing enables detections of 2-500, 2-400, 2-300, 2-200, 2-100, 2-90, 2-80, 2-70, 2-60, 2-50, 2-45, 2-40, 2-35, 2-30, 2-25, 2-20, 2-15, 2-10, or 2-5 different segments of the same gene at the same time. In some embodiments, the gene is a monkeypox gene.

**[0313]** In some embodiments, multiplexing enables detections of different genes at the same time. In some embodiments, multiplexing enables detections of at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100 or more different genes at the same time. In some embodiments, multiplexing enables detections of 2-500, 2-400, 2-300, 2-200, 2-100, 2-90, 2-80, 2-70, 2-60, 2-50, 2-45, 2-40, 2-35, 2-30, 2-25, 2-20, 2-15, 2-10, or 2-5 different genes at the same time. In some embodiments, at least of the genes is a monkeypox gene.

**[0314]** In some embodiments, multiplexing enables detections of both different segments of the same gene and different genes at the same time. In some embodiments, there are at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100 or more different segments of the same gene. In some embodiments, there are 2-500, 2-400, 2-300, 2-200, 2-100, 2-90, 2-80, 2-70, 2-60, 2-50, 2-45, 2-40, 2-35, 2-30, 2-25, 2-20, 2-15, 2-10, or 2-5 different segments of the same gene. In some embodiments, there are at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100 or more different genes at the same time. In some embodiments, there are 2-500, 2-400, 2-300, 2-200, 2-100, 2-90, 2-80, 2-70, 2-60, 2-50, 2-45, 2-40, 2-35, 2-30, 2-25, 2-20, 2-15, 2-10, or 2-5 different genes at the same time. In some embodiments, at least of the genes is a monkeypox gene.

#### Methods For Detecting and Assaying for Monkeypox Virus

**[0315]** Provided herein are compositions, systems, kits, and methods of detecting and assaying for monkeypox virus. The detection assays disclosed herein may provide low cost, portable, and accurate detection of monkeypox virus and may be performed using commercially available reagents. Such assays may be referred to herein as monkeypox virus DNA Endonuclease-Targeted CRISPR (clustered regularly interspaced short palindromic repeats) Trans Reporter (DETECTR) assays.

**[0316]** In particular, the various methods, compositions, and kits disclosed herein use a programmable nuclease complexed with a non-naturally occurring guide nucleic acid that hybridizes to a target sequence of a target nucleic acid from monkeypox virus. The complex can be contacted to a sample from a subject. The subject may or may not be infected with monkeypox virus. The target nucleic acid in the sample can be amplified by thermal amplification (e.g., PCR) or isothermal amplification (e.g., LAMP). If the subject is infected with monkeypox virus, the non-naturally occurring guide nucleic acid hybridizes to the target nucleic acid leading to activation of programmable nuclease. Upon

activation, the programmable nuclease can cleave a reporter, wherein the reporter optionally comprises a detectable label attached to a polynucleotide (e.g., polydeoxyribonucleotide or polyribonucleotide). Upon cleavage of the polynucleotide, the detectable label emits a detectable signal, which is then detected and quantified (e.g., the detectable label may be a fluorophore and the detectable signal may be fluorescence). Upon detection of a detectable label, it can be determined that the sample from the subject contained target nucleic acids from monkeypox virus. A patient may be diagnosed with monkey pox if the presence of monkeypox virus is detected in a sample from the patient. In some embodiments, a DETECTR assay may detect multiple target nucleic acids or amplicons. In some embodiments, a DETECTR assay may detect multiple target nucleic acids that are specific to monkeypox virus. In some embodiments, a DETECTR assay may detect a combination of one or more target nucleic acid(s) specific to monkeypox virus and one or more target nucleic acid(s) present in other Orthopoxviruses to distinguish or discriminate between monkeypox virus and other Orthopoxviruses in patient samples.

**[0317]** A number of reagents are consistent with the methods, compositions, and kits disclosed herein. The reagents described herein may be used in methods of assaying for a target nucleic acid in a sample, the method comprising a) amplifying the target nucleic acid using at least one amplification primer; b) contacting the sample to a reporter and a composition comprising a programmable nuclease and a guide nucleic acid that hybridizes to the target nucleic acid or an amplified product thereof, wherein the programmable nuclease cleaves the reporter upon hybridization of the guide nucleic acid to the target nucleic acid or the amplification product thereof; and c) assaying for a change in a signal, wherein the change in the signal is produced by cleavage of the reporter; wherein the target nucleic acid is a gene of a monkeypox virus or a segment thereof; and optionally wherein the at least one amplification primer comprises a nucleotide sequence at least 85%, at least 87%, at least 90%, at least 92%, at least 94%, at least 95%, at least 96%, at least 98%, or 100% identical to any one of SEQ ID 5 NOS: 92-235.

#### Amplification Primers

**[0318]** The reagents described herein may comprise amplification primers.

**[0319]** In some embodiments, an amplification primer comprises a nucleotide sequence at least 80%, at least 85%, at least 87%, at least 90%, at least 92%, at least 94%, at least 95%, at least 96%, at least 98%, or 100% to any one of SEQ ID NOS: 92-235, as provided in Table 3 below.

TABLE 3

Exemplary Amplification Primer Sequences					
SEQ ID NO	Primer Name	Sequence	Primer Set	Associated sgRNA	
92.	M44982 monkeypox-set1-F3	TTTTCATCGACTAG ACGT	MPXV-set1	MPXV-1 or MPXV-3	
93.	M44983 monkeypox-set1-B3	CATTGATGAATGTCA TAACTTC	MPXV-set1	MPXV-1 or MPXV-3	
94.	M44984 monkeypox-set1-FIP	GCAAGAATTCAACCAT GTTGGTTAAATAGCG ATTGGTGTTGT	MPXV-set1	MPXV-1 or MPXV-3	

TABLE 3-continued

Exemplary Amplification Primer Sequences				
SEQ ID NO	Primer Name	Sequence	Primer Set	Associated sgRNA
95.	M44985 monkeypox-set1-BIP	CTTATGGTTTTAA CGCAATGGTCGAAG ATGGTAAGATCCGTC	MPXV-set1	MPXV-1 or MPXV-3
96.	M44986 monkeypox-set1-LF	ACTTACTGCGACCAG G	MPXV-set1	MPXV-1 or MPXV-3
97.	M44987 monkeypox-set1-LB	ATATACTGACCGAGT AGG	MPXV-set1	MPXV-1 or MPXV-3
98.	M44988 monkeypox-set2-F3	GTTTATTCTCAAATA GCGATTG	MPXV-set2	MPXV-1 or MPXV-3
99.	M44989 monkeypox-set2-B3	CATTGATGAATGTCA TAACCTCA	MPXV-set2	MPXV-1 or MPXV-3
100.	M44990 monkeypox-set2-FIP	ATCGTTAATAGTGTG CAAGAATTCACTGTT GTAAGGATCCTGG	MPXV-set2	MPXV-1 or MPXV-3
101.	M44991 monkeypox-set2-BIP	CTTATGGTTTTAA CGCAATGGTCGAAG ATGGTAAGATCCGTC	MPXV-set2	MPXV-1 or MPXV-3
102.	M44992 monkeypox-set2-LF	TACTGCGACCAGGAT C	MPXV-set2	MPXV-1 or MPXV-3
103.	M44993 monkeypox-set2-LB	TATACTGACCGAGTA G	MPXV-set2	MPXV-1 or MPXV-3
104.	M44994 monkeypox-set3-F3	TTTTTCATCGACTAG ACGT	MPXV-set3	MPXV-1 or MPXV-3
105.	M44995 monkeypox-set3-B3	CATTGATGAATGTCA TAACCTCA	MPXV-set3	MPXV-1 or MPXV-3
106.	M44996 monkeypox-set3-FIP	GCAAGAACCAT GTGGAATAGCGATT GGTGTGTT	MPXV-set3	MPXV-1 or MPXV-3
107.	M44997 monkeypox-set3-BIP	TGGTTTTAAACGCA ATGGTCGAAGATGGT AAGATCCGTC	MPXV-set3	MPXV-1 or MPXV-3
108.	M44998 monkeypox-set3-LF	ACTTACTGCGACCAG G	MPXV-set3	MPXV-1 or MPXV-3
109.	M44999 monkeypox-set3-LB	ATATACTGACCGAGT AGG	MPXV-set3	MPXV-1 or MPXV-3
110.	M45000 monkeypox-set4-F3	GTTTATTCTCAAATA GCGATTG	MPXV-set4	MPXV-1 or MPXV-3
111.	M45001 monkeypox-set4-B3	CATTGATGAATGTCA TAACCTCA	MPXV-set4	MPXV-1 or MPXV-3
112.	M45002 monkeypox-set4-FIP	ATCGTTAATAGTGTG CAAGAATTCACTGTT GTAAGGATCCTGG	MPXV-set4	MPXV-1 or MPXV-3
113.	M45003 monkeypox-set4-BIP	CTTATGGTTTTAA CGCAATGGTCGAAG ATGGTAAGATCCGTC	MPXV-set4	MPXV-1 or MPXV-3
114.	M45004 monkeypox-set4-LF	CATGTTGGTTAACCT ACTGC	MPXV-set4	MPXV-1 or MPXV-3
115.	M45005 monkeypox-set4-LB	ATATACTGACCGAGT AGG	MPXV-set4	MPXV-1 or MPXV-3

TABLE 3-continued

Exemplary Amplification Primer Sequences				
SEQ ID NO	Primer Name	Sequence	Primer Set	Associated sgRNA
116.	M45006 monkeypox-set5-F3	GTTTATTCTCAAATA GCGATTG	MPXV-set5	MPXV-1 or MPXV-3
117.	M45007 monkeypox-set5-B3	CATTGATGAATGTCA TAACATTCA	MPXV-set5	MPXV-1 or MPXV-3
118.	M45008 monkeypox-set5-FIP	ATCGTTAATAGTGTG CAAGAATTCACTGTT GTAAGGATCCTGG	MPXV-set5	MPXV-1 or MPXV-3
119.	M45009 monkeypox-set5-BIP	CTTATGGTTTTTAA CGCAATGGTCGAAG ATGGTAAGATCCGTC	MPXV-set5	MPXV-1 or MPXV-3
120.	M45010 monkeypox-set5-LF	GTTGGTTAACCTACT GCCA	MPXV-set5	MPXV-1 or MPXV-3
121.	M45011 monkeypox-set5-LB	TATACTGACCGAGTA G	MPXV-set5	MPXV-1 or MPXV-3
122.	M45012 monkeypox-set6-F3	GTTTATTCTCAAATA GCGATTG	MPXV-set6	MPXV-1 or MPXV-3
123.	M45013 monkeypox-set6-B3	CATTGATGAATGTCA TAACATTCA	MPXV-set6	MPXV-1 or MPXV-3
124.	M45014 monkeypox-set6-FIP	CGTTAATAGTGTGCA AGAATTTCAGTGTGTT AAGGATCCTGG	MPXV-set6	MPXV-1 or MPXV-3
125.	M45015 monkeypox-set6-BIP	TGGTTTTAACGCA ATGGTCGAAGATGGT AAGATCCGTC	MPXV-set6	MPXV-1 or MPXV-3
126.	M45016 monkeypox-set6-LF	CATGTTGGTTAACCT ACTGC	MPXV-set6	MPXV-1 or MPXV-3
127.	M45017 monkeypox-set6-LB	ATATACTGACCGAGT AGG	MPXV-set6	MPXV-1 or MPXV-3
128.	M45018 monkeypox-set7-F3	GTTTACTTTAGTCC GTATC	MPXV-set7	MPXV-2
129.	M45019 monkeypox-set7-B3	TTACTCTTTGTATC GCA	MPXV-set7	MPXV-2
130.	M45020 monkeypox-set7-FIP	ACAATTATTGGTAG TGTCATATGACAGTC AACACTATGTTAGC	MPXV-set7	MPXV-2
131.	M45021 monkeypox-set7-BIP	CGGTAATCTTGTGCA TGAGTCTCTCGGGTA TTGGTAG	MPXV-set7	MPXV-2
132.	M45022 monkeypox-set7-LF	GTGAAAGCTATATCG ACAG	MPXV-set7	MPXV-2
133.	M45023 monkeypox-set7-LB	GACATATAGTATTCT TGTATTC	MPXV-set7	MPXV-2
134.	M45024 monkeypox-set8-F3	TCCGTATCCAGTCAA CAC	MPXV-set8	MPXV-2
135.	M45025 monkeypox-set8-B3	TCCGTATCCAGTCAA CAC	MPXV-set8	MPXV-2
136.	M45026 monkeypox-set8-FIP	TCCGTATCCAGTCAA CAC	MPXV-set8	MPXV-2
137.	M45027 monkeypox-set8-BIP	TCCGTATCCAGTCAA CAC	MPXV-set8	MPXV-2

TABLE 3-continued

Exemplary Amplification Primer Sequences				
SEQ ID NO	Primer Name	Sequence	Primer Set	Associated sgRNA
138.	M45028 monkeypox-set8-LF	GTGAAAGCTATATCG ACAG	MPXV-set8	MPXV-2
139.	M45029 monkeypox-set8-LB	GACATATAGTATTCT TGTATTTC	MPXV-set8	MPXV-2
140.	M45030 monkeypox-set9-F3	GTTTACTTTTAGTCC GTATC	MPXV-set9	MPXV-2
141.	M45031 monkeypox-set9-B3	TTACTCTTTGTATC GCA	MPXV-set9	MPXV-2
142.	M45032 monkeypox-set9-FIP	ATTGGTGAGTGTCAT ATGACAGTCACACT ATGTTAGC	MPXV-set9	MPXV-2
143.	M45033 monkeypox-set9-BIP	CGGTAATCTTGTCGA TGAGTCCTCGGGTA TTCGGTAG	MPXV-set9	MPXV-2
144.	M45034 monkeypox-set9-LF	GTGAAAGCTATATCG ACAG	MPXV-set9	MPXV-2
145.	M45035 monkeypox-set9-LB	GACATATAGTATTCT TGTATTTC	MPXV-set9	MPXV-2
146.	M45036 monkeypox-set10-F3	TCCGTATCCAGTCAA CAC	MPXV-set10	MPXV-2
147.	M45037 monkeypox-set10-B3	TTACTCTTTGTATC GCA	MPXV-set10	MPXV-2
148.	M45038 monkeypox-set10-FIP	AATTCTACAATTATT GGTGAGTGTATGTTA GCATTTCTGTCG	MPXV-set10	MPXV-2
149.	M45039 monkeypox-set10-BIP	CGGTAATCTTGTCGA TGAGTCCTCGGGTA TTCGGTAG	MPXV-set10	MPXV-2
150.	M45040 monkeypox-set10-LF	GTGAAAGCTATATCG ACAG	MPXV-set10	MPXV-2
151.	M45041 monkeypox-set10-LB	GACATATAGTATTCT TGTATTTC	MPXV-set10	MPXV-2
152.	M45042 monkeypox-set11-F3	TCCGTATCCAGTCAA CAC	MPXV-set11	MPXV-2
153.	M45043 monkeypox-set11-B3	TTACTCTTTGTATC GCA	MPXV-set11	MPXV-2
154.	M45044 monkeypox-set11-FIP	AATTCTACAATTATT GGTGAGTGTATGTTA GCATTTCTGTCG	MPXV-set11	MPXV-2
155.	M45045 monkeypox-set11-BIP	CGGTAATCTTGTCGA TGAGTCCTCGGGTA TTCGGTAG	MPXV-set11	MPXV-2
156.	M45046 monkeypox-set11-LF	GTGAAAGCTATATCG ACAG	MPXV-set11	MPXV-2
157.	M45047 monkeypox-set11-LB	GACATATAGTATTCT TGTATTTC	MPXV-set11	MPXV-2
158.	M45048 monkeypox-set12-F3	TCCGTATCCAGTCAA CAC	MPXV-set12	MPXV-2
159.	M45049 monkeypox-set12-B3	TTACTCTTTGTATC GCA	MPXV-set12	MPXV-2

TABLE 3-continued

Exemplary Amplification Primer Sequences				
SEQ ID NO	Primer Name	Sequence	Primer Set	Associated sgRNA
160.	M45050 monkeypox-set12-FIP	CTACAATTATTGGTG AGTGTATGTTAGCAT TTCTGTCG	MPXV-set12	MPXV-2
161.	M45051 monkeypox-set12-BIP	CGGTAATCTTGTGCA TGAGTCTCTCGGGTA TTCGGTAG	MPXV-set12	MPXV-2
162.	M45052 monkeypox-set12-LF	GTGAAAGCTATATCG ACAG	MPXV-set12	MPXV-2
163.	M45053 monkeypox-set12-LB	GACATATAAGTATTCT TGTATT	MPXV-set12	MPXV-2
164.	M45054 monkeypox-set13-F3	TATTCGACTGGTGTC AGG	MPXV-set13	MPXV-4
165.	M45055 monkeypox-set13-B3	TGTATTGTGTGCGCG TAA	MPXV-set13	MPXV-4
166.	M45056 monkeypox-set13-FIP	CTATCACTCCTATTA AAGGCCTGGATACGTT CGATATAAACATATG C	MPXV-set13	MPXV-4
167.	M45057 monkeypox-set13-BIP	TCTGTATAATAAGAT GCAAAGGCAAAAAT AATTCACATATTGG	MPXV-set13	MPXV-4
168.	M45058 monkeypox-set13-LF	TGAACCGATCCACTG ATG	MPXV-set13	MPXV-4
169.	M45059 monkeypox-set13-LB	GTAGTAAAGATGCTA GTG	MPXV-set13	MPXV-4
170.	M45060 monkeypox-set14-F3	TATTCGACTGGTGTC AGG	MPXV-set14	MPXV-4
171.	M45061 monkeypox-set14-B3	TGTATTGTGTGCGCG TAA	MPXV-set14	MPXV-4
172.	M45062 monkeypox-set14-FIP	TCACTCCTATTAAAG GCTGCGTTCGATATA AACATATGC	MPXV-set14	MPXV-4
173.	M45063 monkeypox-set14-BIP	TCTGTATAATAAGAT GCAAAGGCAAAAAT AATTCACATATTGG	MPXV-set14	MPXV-4
174.	M45064 monkeypox-set14-LF	TGAACCGATCCACTG ATG	MPXV-set14	MPXV-4
175.	M45065 monkeypox-set14-LB	GTAGTAAAGATGCTA GTG	MPXV-set14	MPXV-4
176.	M45066 monkeypox-set15-F3	AATACATGTCTTAGA TGTTC	MPXV-set15	MPXV-6
177.	M45067 monkeypox-set15-B3	TCTTGATAATCTTGA TGAGT	MPXV-set15	MPXV-6
178.	M45068 monkeypox-set15-FIP	AGAAATACTTTAGAT ACGTGCGTAAACATT AATCTCTCTCCT	MPXV-set15	MPXV-6
179.	M45069 monkeypox-set15-BIP	AGTTAAATTAGATTT CGAACGAAGGTATA CCAGAAAAGACGGT	MPXV-set15	MPXV-6
180.	M45070 monkeypox-set15-LF	GGTATCAATTTTGT TAAGAG	MPXV-set15	MPXV-6

TABLE 3-continued

Exemplary Amplification Primer Sequences				
SEQ ID NO	Primer Name	Sequence	Primer Set	Associated sgRNA
181.	M45071 monkeypox-set15-LB	TCAAATGAACGCCA GAGGCCT	MPXV-set15	MPXV-6
182.	M45072 monkeypox-set16-F3	CAATACATGTCTTAG ATGTTTC	MPXV-set16	MPXV-6
183.	M45073 monkeypox-set16-B3	TCTTGATAATCTTGA TGAGT	MPXV-set16	MPXV-6
184.	M45074 monkeypox-set16-FIP	GAAATACTTTAGATA CGTGCACAAACTTAAT CTCTCTCC	MPXV-set16	MPXV-6
185.	M45075 monkeypox-set16-BIP	GTAAATTAGATTTC GAACGAAGGTATAC CAGAAAAGACGGT	MPXV-set16	MPXV-6
186.	M45076 monkeypox-set16-LF	GGTATCAATTTTGT TAAGAG	MPXV-set16	MPXV-6
187.	M45077 monkeypox-set16-LB	TCAAATGAACGCCA GAGGCCT	MPXV-set16	MPXV-6
188.	M45078 monkeypox-set17-F3	TTCAGGGAAATCGCA TCT	MPXV-set17	MPXV-8
189.	M45079 monkeypox-set17-B3	CCCATATACTTTATT CATGG	MPXV-set17	MPXV-8
190.	M45080 monkeypox-set17-FIP	CCATTAGCTCCATAA TACAGTCTATGAAA GGAAAGAATGT	MPXV-set17	MPXV-8
191.	M45081 monkeypox-set17-BIP	AAAATATGTAGAA AAGGAGGTTGAAT GAGATATTCTGAG	MPXV-set17	MPXV-8
192.	M45082 monkeypox-set17-LF	TTGACGCTGGAGAA ATGA	MPXV-set17	MPXV-8
193.	M45083 monkeypox-set17-LB	AGAACATGGATAAG GTTAGC	MPXV-set17	MPXV-8
194.	M45084 monkeypox-set18-F3	TTCAGGGAAATCGCA TCT	MPXV-set18	MPXV-8
195.	M45085 monkeypox-set18-B3	CCCATATACTTTATT CATGG	MPXV-set18	MPXV-8
196.	M45086 monkeypox-set18-FIP	CCATTAGCTCCATAA TACAGTCTATGAAAAG GAAAGAATG	MPXV-set18	MPXV-8
197.	M45087 monkeypox-set18-BIP	AAAATATGTAGAA AAGGAGGTTGAATGA GATATTCTGAG	MPXV-set18	MPXV-8
198.	M45088 monkeypox-set18-LF	TGACGCTGGAGAAA TGA	MPXV-set18	MPXV-8
199.	M45089 monkeypox-set18-LB	AGAACATGGATAAG GTTAGC	MPXV-set18	MPXV-8
200.	M45090 monkeypox-set19-F3	TGCGCATCTTCTATGA AAG	MPXV-set19	MPXV-8
201.	M45091 monkeypox-set19-B3	CCCATATACTTTATT CATGG	MPXV-set19	MPXV-8
202.	M45092 monkeypox-set19-FIP	CCATTAGCTCCATAA TACAGTGAAGAAT GTATTCTTCTCCA	MPXV-set19	MPXV-8

TABLE 3-continued

Exemplary Amplification Primer Sequences				
SEQ ID NO	Primer Name	Sequence	Primer Set	Associated sgRNA
203.	M45093 monkeypox-set19-BIP	CAAAATATGTAGAA AAGGAGGTTTGAAT GAGATATTCTGAG	MPXV-set19	MPXV-8
204.	M45094 monkeypox-set19-LF	ACTGACGAGATTGAC GC	MPXV-set19	MPXV-8
205.	M45095 monkeypox-set19-LB	AGAACATGGATAAG GTTAGC	MPXV-set19	MPXV-8
206.	M45096 monkeypox-set20-F3	TCGCATCTTCTATGA AAG	MPXV-set20	MPXV-8
207.	M45097 monkeypox-set20-B3	CCCATATACTTTATT CATGG	MPXV-set20	MPXV-8
208.	M45098 monkeypox-set20-FIP	CATTAGCTCCATAAT ACAGAGAATGTATTTC ATTCTCC	MPXV-set20	MPXV-8
209.	M45099 monkeypox-set20-BIP	TATGTAGAAAAGGA GGTGATGAGATATT CTGAG	MPXV-set20	MPXV-8
210.	M45100 monkeypox-set20-LF	ACTGACGAGATTGAC GC	MPXV-set20	MPXV-8
211.	M45101 monkeypox-set20-LB	GAAACATGGATAAGG TTAGC	MPXV-set20	MPXV-8
212.	M45102 monkeypox-set21-F3	TGTTATCCAATTGCA CAA	MPXV-set21	MPXV-7
213.	M45103 monkeypox-set21-B3	ATACTCGAGTCTCTG CTG	MPXV-set21	MPXV-7
214.	M45104 monkeypox-set21-FIP	AGGTTATTATCTGGA TCATCTATCACTTAA TATCAGAGAGATAG AAGA	MPXV-set21	MPXV-7
215.	M45105 monkeypox-set21-BIP	GTAATCCCACAGA ACTAATGGTTGCCCA TCTATCCTCT	MPXV-set21	MPXV-7
216.	M45106 monkeypox-set21-LF	GCGTTTGTCATAT GT	MPXV-set21	MPXV-7
217.	M45107 monkeypox-set21-LB	TCTAGGTACTTATGG ACAAAC	MPXV-set21	MPXV-7
218.	M45108 monkeypox-set22-F3	AGTTCCATTATCTAA AGCT	MPXV-set22	MPXV-7
219.	M45109 monkeypox-set22-B3	ATACTCGAGTCTCTG CTG	MPXV-set22	MPXV-7
220.	M45110 monkeypox-set22-FIP	GTTTGTCTCATATGT TCTTCTATATCCATG TTATCCAACCTT	MPXV-set22	MPXV-7
221.	M45111 monkeypox-set22-BIP	GTAATCCCACAGA ACTAATGGTCGCCAT CTATCCTCTG	MPXV-set22	MPXV-7
222.	M45112 monkeypox-set22-LF	CTCTGATATTAAGAT TTGTC	MPXV-set22	MPXV-7
223.	M45113 monkeypox-set22-LB	TCTAGGTACTTATGG ACAAAC	MPXV-set22	MPXV-7

TABLE 3-continued

Exemplary Amplification Primer Sequences				
SEQ ID NO	Primer Name	Sequence	Primer Set	Associated sgRNA
224.	M45114 monkeypox-set23-F3	GGAATCCATGTTATC CAA	MPXV-set23	MPXV-7
225.	M45115 monkeypox-set23-B3	CCAATACTCGAGTCT CTG	MPXV-set23	MPXV-7
226.	M45116 monkeypox-set23-FIP	TCATCTATCAGCGTT TGTCTCTTGACAAAT CTTAATATCAGAG	MPXV-set23	MPXV-7
227.	M45117 monkeypox-set23-BIP	GTAAATCCCACAGA ACTAATGTTGCCCAT CTATCCTCTG	MPXV-set23	MPXV-7
228.	M45118 monkeypox-set23-LF	CATATGTTCTTCTAT CT	MPXV-set23	MPXV-7
229.	M45119 monkeypox-set23-LB	TCTAGGTACTTATGG ACAAAC	MPXV-set23	MPXV-7
230.	M45120 monkeypox-set24-F3	TGTTATCCAACATTGA CAA	MPXV-set24	MPXV-7
231.	M45121 monkeypox-set24-B3	AATACTCGAGTCTCT GCT	MPXV-set24	MPXV-7
232.	M45122 monkeypox-set24-FIP	TCAGGAGGTTATTAT CTGGAGAGAGATAG AGAACATATGAGA	MPXV-set24	MPXV-7
233.	M45123 monkeypox-set24-BIP	GTAAATCCCACAGAC TAATGTTGCCCATCT ATCCTCTG	MPXV-set24	MPXV-7
234.	M45124 monkeypox-set24-LF	TCATCTATCAGCGTT TG	MPXV-set24	MPXV-7
235.	M45125 monkeypox-set24-LB	TCTAGGTACTTATGG ACAAAC	MPXV-set24	MPXV-7

**[0320]** In some embodiments, amplifying a target nucleic acid comprises at least one amplification primer. In some embodiments, amplifying a target nucleic acid comprises at least two amplification primers. In some embodiments, amplifying a target nucleic acid comprises at least three amplification primers. In some embodiments, amplifying a target nucleic acid comprises at least four amplification primers. In some embodiments, amplifying a target nucleic acid comprises at least five amplification primers. In some embodiments, amplifying a target nucleic acid comprises at least six amplification primers. In some embodiments, amplifying a target nucleic acid comprises three amplification primers. In some embodiments, amplifying a target nucleic acid comprises six amplification primers.

**[0321]** In some embodiments, an amplification primer is a FIP primer, a BIP primer, a B3 primer, a F3 primer, a LB primer, or a LF primer. In some embodiments, amplifying a target nucleic acid comprises at least one primer selected from the group consisting of a FIP primer, a BIP primer, a B3 primer, a F3 primer, a LB primer, and a LF primer. In some embodiments, amplifying a target nucleic acid comprises at least two primers selected from the group consisting of a FIP primer, a BIP primer, a B3 primer, a F3 primer, a LB primer, and a LF primer. In some embodiments, amplifying a target nucleic acid comprises at least three primers selected from the group consisting of a FIP primer, a BIP primer, a B3 primer, a F3 primer, a LB primer, and a LF primer. In some embodiments, amplifying a target nucleic acid comprises at least four primers selected from the group consisting of a FIP primer, a BIP primer, a B3 primer, a F3 primer, a LB primer, and a LF primer. In some embodiments, amplifying a target nucleic acid comprises at least five primers selected from the group consisting of a FIP primer, a BIP primer, a B3 primer, a F3 primer, a LB primer, and a LF primer. In some embodiments, amplifying a target nucleic acid comprises at least six primers selected from the group consisting of a FIP primer, a BIP primer, a B3 primer, a F3 primer, a LB primer, and a LF primer. In some embodiments, amplifying a target nucleic acid comprises at least seven primers selected from the group consisting of a FIP primer, a BIP primer, a B3 primer, a F3 primer, a LB primer, and a LF primer. In some embodiments, amplifying a target nucleic acid comprises at least eight primers selected from the group consisting of a FIP primer, a BIP primer, a B3 primer, a F3 primer, a LB primer, and a LF primer. In some embodiments, amplifying a target nucleic acid comprises at least nine primers selected from the group consisting of a FIP primer, a BIP primer, a B3 primer, a F3 primer, a LB primer, and a LF primer. In some embodiments, amplifying a target nucleic acid comprises at least ten primers selected from the group consisting of a FIP primer, a BIP primer, a B3 primer, a F3 primer, a LB primer, and a LF primer.

amplifying a target nucleic acid comprises at least three primers selected from the group consisting of a FIP primer, a BIP primer, a B3 primer, a F3 primer, a LB primer, and a LF primer. In some embodiments, amplifying a target nucleic acid comprises at least four primers selected from the group consisting of a FIP primer, a BIP primer, a B3 primer, a F3 primer, a LB primer, and a LF primer. In some embodiments, amplifying a target nucleic acid comprises at least five primers selected from the group consisting of a FIP primer, a BIP primer, a B3 primer, a F3 primer, a LB primer, and a LF primer. In some embodiments, amplifying a target nucleic acid comprises at least six primers selected from the group consisting of a FIP primer, a BIP primer, a B3 primer, a F3 primer, a LB primer, and a LF primer. In some embodiments, amplifying a target nucleic acid comprises at least seven primers selected from the group consisting of a FIP primer, a BIP primer, a B3 primer, a F3 primer, a LB primer, and a LF primer. In some embodiments, amplifying a target nucleic acid comprises at least eight primers selected from the group consisting of a FIP primer, a BIP primer, a B3 primer, a F3 primer, a LB primer, and a LF primer. In some embodiments, amplifying a target nucleic acid comprises at least nine primers selected from the group consisting of a FIP primer, a BIP primer, a B3 primer, a F3 primer, a LB primer, and a LF primer. In some embodiments, amplifying a target nucleic acid comprises at least ten primers selected from the group consisting of a FIP primer, a BIP primer, a B3 primer, a F3 primer, a LB primer, and a LF primer.

**[0322]** In some embodiments, a FIP primer comprises a nucleotide sequence at least 80%, at least 85%, at least 87%, at least 90%, at least 92%, at least 94%, at least 95%, at least

96%, at least 98%, or 100% identical to any one of SEQ ID NOs 94, 100, 106, 112, 118, 124, 130, 136, 142, 148, 154, 160, 166, 171, 178, 184, 190, 196, 202, 208, 214, 220, 226, or 232. In some embodiments, a BIP primer comprises a nucleotide sequence at least 80%, at least 85%, at least 87%, at least 90%, at least 92%, at least 94%, at least 95%, at least 96%, at least 98%, or 100% identical to any one of SEQ ID NOs 4, 10, 16, 22, 28, 34, 40, 46, 52, 58, 64, 70, 76, 82, 88, 94, 100, 106, 112, 118, 124, 130, 136, or 14295, 101, 107, 113, 119, 125, 131, 137, 143, 149, 155, 161, 167, 172, 179, 185, 191, 197, 203, 209, 215, 221, 227, or 233. In some embodiments, a B3 primer comprises a nucleotide sequence at least 80%, at least 85%, at least 87%, at least 90%, at least 92%, at least 94%, at least 95%, at least 96%, at least 98%, or 100% identical to any one of SEQ ID NOs 93, 99, 105, 111, 117, 123, 129, 135, 141, 147, 153, 159, 165, 170, 177, 183, 189, 195, 201, 207, 213, 219, 225, or 231. In some embodiments, a F3 primer comprises a nucleotide sequence at least 80%, at least 85%, at least 87%, at least 90%, at least 92%, at least 94%, at least 95%, at least 96%, at least 98%, or 100% identical to any one of SEQ ID NOs 92, 98, 104, 110, 116, 122, 128, 134, 140, 146, 152, 158, 164, 170, 176, 182, 188, 194, 200, 206, 212, 218, 224, or 230. In some embodiments, a LF primer comprises a nucleotide sequence at least 80%, at least 85%, at least 87%, at least 90%, at least 92%, at least 94%, at least 95%, at least 96%, at least 98%, or 100% identical to any one of SEQ ID NOs 96, 102, 108, 114, 120, 126, 132, 138, 144, 150, 156, 162, 168, 173, 180, 186, 192, 198, 204, 210, 216, 222, 228, or 234. In some embodiments, a LB primer comprises a nucleotide sequence at least 80%, at least 85%, at least 87%, at least 90%, at least 92%, at least 94%, at least 95%, at least 96%, at least 98%, or 100% identical to any one of SEQ ID NOs 97, 103, 109, 115, 121, 127, 133, 139, 145, 151, 157, 163, 169, 174, 181, 187, 193, 199, 205, 211, 217, 223, 229, or 235.

**[0323]** In some embodiments, the amplifying comprises isothermal amplification. In some embodiments, the amplifying comprises helicase dependent amplification (HDA), circular helicase dependent amplification (cHDA), strand displacement amplification (SDA), loop mediated amplification (LAMP), exponential amplification reaction (EX-PAR), rolling circle amplification (RCA), ligase chain reaction (LCR), single primer isothermal amplification (SPIA), multiple displacement amplification (MDA), nucleic acid sequence based amplification (NASBA), hinge-initiated primer-dependent amplification of nucleic acids (HIP), nicking enzyme amplification reaction (NEAR), or improved multiple displacement amplification (IMDA). In some embodiments, the amplifying comprises loop mediated amplification (LAMP).

**[0324]** In some embodiments, the amplifying comprises a thermal cycling amplification. In some embodiments, the amplifying comprises polymerase chain reaction (PCR).

**[0325]** The amplifying can improve at least one of sensitivity, specificity, or accuracy of the detection the target nucleic acid. The reagents for nucleic acid amplification can comprise a recombinase, an oligonucleotide primer, a single-stranded DNA binding (SSB) protein, and a polymerase. The nucleic acid amplification can be performed for no greater than 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 40, 50, or 60 minutes. In some embodiments, the nucleic acid amplification reaction is performed at a temperature of around 20-65° C. In some embodiments, the nucleic acid amplification reaction is performed at a temperature no greater than 20° C., 25° C., 30° C., 35° C., 37° C., 40° C., 45° C., 50° C., 55° C., 60° C., 65° C. In some embodiments, the nucleic acid amplification reaction is performed at a temperature of at least 20° C., 25° C., 30° C., 35° C., 37° C., 40° C., 45° C., 50° C., 55° C., 60° C., of 65° C.

**[0326]** In some embodiments, the amplifying comprises contacting the sample to reagents for amplification. In some embodiments, the contacting the sample to reagents for amplification occurs concurrent to the contacting the sample to the reporter and the composition comprising the programmable nuclease and the guide nucleic acid. In some embodiments, the contacting the sample to reagents for amplification occurs prior to the contacting the sample to the reporter and the composition comprising the programmable nuclease and the guide nucleic acid. In some embodiments, the reagents for amplification comprise a polymerase and dNTPs.

#### Guide Nucleic Acids

**[0327]** The reagents described herein may comprise guide nucleic acids, as described herein, designed to target at least a segment of a target nucleic acid of monkeypox virus. The reagents described herein comprise multiple guide nucleic acids. Each guide nucleic acid comprises a sequence (e.g., a spacer sequence) that is reverse complementary to a segment of a target nucleic acid of monkeypox virus. Each guide nucleic acid specifically binds to the segment of the target nucleic acid of monkeypox virus. In some embodiments, each guide nucleic acid is able to distinguish between two target nucleic acids, wherein the two target nucleic acids comprise a difference in the nucleotide sequences between each other. In some embodiments, the guide nucleic acid is non-naturally occurring and made by artificial combination of otherwise separate segments of sequence. In some embodiments, the artificial combination is performed by chemical synthesis, by genetic engineering techniques, or by the artificial manipulation of isolated segments of nucleic acids. The non-naturally occurring guide nucleic acid or segment thereof can be designed and made to provide desired functions.

**[0328]** In some embodiments, a guide nucleic acid comprises a nucleotide sequence at least 80%, at least 85%, at least 87%, at least 90%, at least 92%, at least 94%, at least 95%, at least 96%, at least 98%, or 100% identical to any one of SEQ ID NOs 236-247, as provided in Table below.

TABLE 4

<u>Exemplary Non-Naturally Occurring Guide Nucleic Acid Sequences</u>		
SEQ ID NO	Name	Guide Nucleic Acid Sequence*
236.	R13932 MPXV-1	gACCGCUUCACCAAGUGCUGUCCCUUAGGGGAAUAG CACUUGAGUGAAGGUGGGCUGCUUGCAUCAGCCUAA UGUCGAGAAGUGCUUUCCUUCGGAAAGUAACCCUCGA ACAAAAUCAUUGaaaGAAUGAAGGAAUGCAACUCGG CUACACCUCGUUAA
237.	R13933 MPXV-2	gACCGCUUCACCAAGUGCUGUCCCUUAGGGGAAUAG CACUUGAGUGAAGGUGGGCUGCUUGCAUCAGCCUAA UGUCGAGAAGUGCUUUCCUUCGGAAAGUAACCCUCGA ACAAAAUCAUUGaaaGAAUGAAGGAAUGCAACACCA AUAGUGAGUUCGGCGA
238.	R13934 MPXV-3	gACCGCUUCACCAAGUGCUGUCCCUUAGGGGAAUAG CACUUGAGUGAAGGUGGGCUGCUUGCAUCAGCCUAA UGUCGAGAAGUGCUUUCCUUCGGAAAGUAACCCUCGA ACAAAAUCAUUGaaaGAAUGAAGGAAUGCAACCGAU AGGUGUAGCCGAUAAA
239.	R13935 MPXV-4	gACCGCUUCACCAAGUGCUGUCCCUUAGGGGAAUAG CACUUGAGUGAAGGUGGGCUGCUUGCAUCAGCCUAA UGUCGAGAAGUGCUUUCCUUCGGAAAGUAACCCUCGA ACAAAAUCAUUGaaaGAAUGAAGGAAUGCAACUUAA CGAUUGUCGACCCUCU
240.	R13936 MPXV-5	gACCGCUUCACCAAGUGCUGUCCCUUAGGGGAAUAG CACUUGAGUGAAGGUGGGCUGCUUGCAUCAGCCUAA UGUCGAGAAGUGCUUUCCUUCGGAAAGUAACCCUCGA ACAAAAUCAUUGaaaGAAUGAAGGAAUGCAACUGAA GAUGCCAUGUACUACG
241.	R13937 MPXV-6	gACCGCUUCACCAAGUGCUGUCCCUUAGGGGAAUAG CACUUGAGUGAAGGUGGGCUGCUUGCAUCAGCCUAA UGUCGAGAAGUGCUUUCCUUCGGAAAGUAACCCUCGA ACAAAAUCAUUGaaaGAAUGAAGGAAUGCAACACUU UUACGCCUCUGCGUU
242.	R13938 MPXV-7	gACCGCUUCACCAAGUGCUGUCCCUUAGGGGAAUAG CACUUGAGUGAAGGUGGGCUGCUUGCAUCAGCCUAA UGUCGAGAAGUGCUUUCCUUCGGAAAGUAACCCUCGA ACAAAAUCAUUGaaaGAAUGAAGGAAUGCAACAAAAA UGGCCAAAGCGGGUUAA
243.	R13939 MPXV-8	gACCGCUUCACCAAGUGCUGUCCCUUAGGGGAAUAG CACUUGAGUGAAGGUGGGCUGCUUGCAUCAGCCUAA UGUCGAGAAGUGCUUUCCUUCGGAAAGUAACCCUCGA ACAAAAUCAUUGaaaGAAUGAAGGAAUGCAACGAUC CACUGCUGAACAGCUA
244.	R13977 MPXV-3 (fixed)	gACCGCUUCACCAAGUGCUGUCCCUUAGGGGAAUAG CACUUGAGUGAAGGUGGGCUGCUUGCAUCAGCCUAA UGUCGAGAAGUGCUUUCCUUCGGAAAGUAACCCUCGA ACAAAAUCAUUGaaaGAAUGAAGGAAUGCAACUUAA UCGGCUACACCUAUUCG
245.	R13978 MPXV-7 (fixed)	gACCGCUUCACCAAGUGCUGUCCCUUAGGGGAAUAG CACUUGAGUGAAGGUGGGCUGCUUGCAUCAGCCUAA UGUCGAGAAGUGCUUUCCUUCGGAAAGUAACCCUCGA ACAAAAUCAUUGaaaGAAUGAAGGAAUGCAACUAC CCGCUUUGGCCAUUUU
246.	R13979 MPXV-8 (fixed)	gACCGCUUCACCAAGUGCUGUCCCUUAGGGGAAUAG CACUUGAGUGAAGGUGGGCUGCUUGCAUCAGCCUAA UGUCGAGAAGUGCUUUCCUUCGGAAAGUAACCCUCGA ACAAAAUCAUUGaaaGAAUGAAGGAAUGCAACAUAG CUGUUCAGCAGUGGAU

TABLE 4-continued

<u>Exemplary Non-Naturally Occurring Guide Nucleic Acid Sequences</u>		
SEQ ID NO	Name	Guide Nucleic Acid Sequence*
247.	R13980 MPXV-2 (fixed)	gACCGCUUCACCAAGUGCUGGUCCUUAGGGGAAUAG CACUUGAGUGAAGGUGGGCUGCUUGCAUCAGCCUAA UGUCGAGAAGUGCUUUUCUUCGAAAGUAACCCUCGAA ACAAAAAUCAUJaaaaGAAUGAAGGAAUGCAACUCGC CGAACUCACU2AUUGGU

\*\*The first lowercase base (g) is the starting nucleotide from T7. The middle lowercase bases (gaaa) are a linker between (i) the portion of the handle that comprises at least a portion of a tracrRNA sequence and (ii) the portion of the handle that contains at least a portion of a repeat sequence of the sgRNA backbone.

[0329] The target site sequence of each non-naturally occurring guide nucleic acid is disclosed in Table below.

TABLE 5

<u>Exemplary Target Site Sequences of Guide Nucleic Acids</u>		
SEQ ID NO	Name	Target Site Sequence
248.	R13932 MPXV-1	TGGCTACACCTATCGTTAA
249.	R13933 MPXV-2	ACCAATAGTGAGTTGGCGA
250.	R13934 MPXV-3	CGATAGGTGTAGCCGATAAA
251.	R13935 MPXV-4	TTAACGATTGTCGACCCCTCT
252.	R13936 MPXV-5	TGAAGATGCCATGTACTACG
253.	R13937 MPXV-6	ACTTTACGCCCTCTGGCGTT
254.	R13938 MPXV-7	AAAATGGCAAAGCGGGTTA
255.	R13939 MPXV-8	GATCCACTGCTGAACAGCTA
256.	R13977 MPXV-3 (fixed)	TTTATCGGCTACACCTATCG
257.	R13978 MPXV-7 (fixed)	TAACCCGCTTGGCCATT
258.	R13979 MPXV-8 (fixed)	ATAGCTGTTCAGCAGTGGAT
259.	R13980 MPXV-2 (fixed)	TCGCCGAACTCACATTGGT

#### Effector Proteins

[0330] The reagents described herein comprise a programmable nuclease (also referred to, interchangeably, as an effector protein) capable of being activated when complexed with the guide nucleic acid and the target nucleic acid or segment thereof segment, as described herein. The reagents for monkeypox detection may comprise any of the effector proteins described herein. For example, in some embodiments, the Type V CRISPR/Cas enzyme is a programmable Cas14 nuclease. In some embodiments, the Cas14 nuclease is Cas14a.1.

#### Reporter Nucleic Acids

[0331] The reagents described herein comprise one or more reporters as described herein. The one or more report-

ers may comprise any of the detection moieties described herein. In some embodiments, the one or more reporters may be in solution. In some embodiments, the one or more reporters may be immobilized on a solid support or other support medium.

[0332] In some embodiments, the reporter comprises a nucleic acid conjugated to an affinity molecule and the affinity molecule conjugated to the fluorophore (e.g., nucleic acid-affinity molecule-fluorophore) or the nucleic acid conjugated to the fluorophore and the fluorophore conjugated to the affinity molecule (e.g., nucleic acid-fluorophore-affinity molecule). In some embodiments, a linker conjugates the nucleic acid to the affinity molecule. In some embodiments, a linker conjugates the affinity molecule to the fluorophore. In some embodiments, a linker conjugates the nucleic acid to the fluorophore. A linker can be any suitable linker known in the art. In some embodiments, the nucleic acid of the reporter can be directly conjugated to the affinity molecule and the affinity molecule can be directly conjugated to the fluorophore or the nucleic acid can be directly conjugated to the fluorophore and the fluorophore can be directly conjugated to the affinity molecule. In this context, “directly conjugated” indicated that no intervening molecules, poly-peptides, proteins, or other moieties are present between the two moieties directly conjugated to each other. For example, if a reporter comprises a nucleic acid directly conjugated to an affinity molecule and an affinity molecule directly conjugated to a fluorophore-no intervening moiety is present between the nucleic acid and the affinity molecule and no intervening moiety is present between the affinity molecule and the fluorophore. The affinity molecule can be biotin, avidin, streptavidin, or any similar molecule.

[0333] In some embodiments, the reporter comprises a substrate-nucleic acid. The substrate may be sequestered from its cognate enzyme when present as in the substrate-nucleic acid, but then is released from the nucleic acid upon cleavage, wherein the released substrate can contact the cognate enzyme to produce a detectable signal. In some embodiments, the reporter comprises an enzyme-nucleic acid. The enzyme may be sterically hindered when present as in the enzyme-nucleic acid, but then functional upon cleavage from the nucleic acid.

[0334] In some embodiments, an enzyme is invertase. In some embodiments, the substrate of invertase is sucrose or DNS reagent.

[0335] A protein-nucleic acid may be attached to a solid support. The solid support, for example, is a surface. A surface can be an electrode. Sometimes the solid support is

a bead. Often the bead is a magnetic bead. Upon cleavage, the protein is liberated from the solid support and interacts with other mixtures. For example, the protein is an enzyme, and upon cleavage of the nucleic acid of the enzyme-nucleic acid, the enzyme flows through a chamber into a mixture comprising the substrate. When the enzyme meets the enzyme substrate, a reaction occurs, such as a colorimetric reaction, which is then detected. As another example, the protein is an enzyme substrate, and upon cleavage of the nucleic acid of the enzyme substrate-nucleic acid, the enzyme flows through a chamber into a mixture comprising the enzyme. When the enzyme substrate meets the enzyme, a reaction occurs, such as a calorimetric reaction, which is then detected.

[0336] In some embodiments, the signal is present prior to reporter cleavage and changes upon reporter cleavage. In some embodiments, the signal is absent prior to reporter cleavage and is present upon reporter cleavage. In some embodiments, the detectable signal is generated directly by the cleavage event. Alternatively, or in combination, the detectable signal is generated indirectly by the signal event. In some embodiments, the detected target nucleic acid or segment thereof is identified based on its spatial location on the detection region of the support medium. In some embodiments, the second detectable signal is generated in a spatially distinct location than the first generated signal. In some embodiments, the detectable signal can be detectable within 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 70, 80, 90, 100, 110, or 120 minutes of contacting the sample.

#### Buffers

[0337] The reagents described herein comprise buffers. These buffers are compatible with the other reagents and samples as described herein for detection of a target nucleic acid.

[0338] In some embodiments, a buffer comprises 20 mM HEPES pH 6.8, 50 mM KCl, 5 mM MgCl<sub>2</sub>, and 5% glycerol. In some embodiments, the buffer comprises from 0 to 100, 0 to 75, 0 to 50, 0 to 25, 0 to 20, 0 to 10, 0 to 5, 5 to 10.5 to 15, 5 to 20, 5 to 25, to 30, 5 to 40, 5 to 50, 5 to 75, 5 to 100, 10 to 20, 10 to 30, 10 to 40, 10 to 50, 15 to 20, 15 to 25, 15 to 30, 15 to 4, 15 to 50, 20 to 25, 20 to 30, 20 to 40, or 20 to 50 mM HEPES pH 6.8. In some embodiments, the buffer comprises 0 to 500, 0 to 400, 0 to 300, 0 to 250, 0 to 200, 0 to 150, 0 to 100, 0 to 75, 0 to 50, 0 to 25, 0 to 20, 0 to 10, 0 to 5, 5 to 10, 5 to 15, 5 to 20, 5 to 25, to 30, 5 to 40, 5 to 50, 5 to 75, 5 to 100, 10 to 20, 10 to 30, 10 to 40, 10 to 50, 15 to 20, 15 to 25, 15 to 30, 15 to 4, 15 to 50, 20 to 25, 20 to 30, 20 to 40, or 20 to 50 mM MgCl<sub>2</sub>. In some embodiments, the buffer can comprise 0 to 25, 0 to 20, 0 to 10, 0 to 5, 5 to 10, 5 to 15, 5 to 20, 5 to 25, to 30, 5 to 40, 5 to 50, 5 to 75, 5 to 100, 10 to 20, 10 to 30, 10 to 40, 10 to 50, 15 to 20, 15 to 25, 15 to 30, 15 to 4, 15 to 50, 20 to 25, 20 to 30, 20 to 40, or 20 to 50 mM KCl. In some embodiments, the buffer comprises 0 to 100, 0 to 75, 0 to 50, 0 to 25, 0 to 20, 0 to 10, 0 to 5, 5 to 10, 5 to 15, 5 to 20, 5 to 25, to 30, 5 to 40, 5 to 50, 5 to 75, 5 to 100, 10 to 20, 10 to 30, 10 to 40, 10 to 50, 15 to 20, 15 to 25, 15 to 30, 15 to 4, 15 to 50, 20 to 25, 20 to 30, 20 to 40, or 20 to 50 mM Tris-HCl, phosphate, or HEPES, a reducing agent (e.g., N-Acetyl Cysteine (NAC), Dithiothreitol (DTT), β-mercaptoethanol (BME), or tris(2-carboxyethyl) phosphine (TCEP)), a chelating agent (e.g., EDTA or EGTA), a detergent (e.g., deoxycholate, NP-40 (Ipgal), Triton X-100, or Tween 20), a salt (e.g., ammonium acetate, magnesium acetate, manganese acetate, potassium acetate, sodium acetate, ammonium chloride, potassium chloride, magnesium chloride, manganese chloride, sodium chloride, ammonium sulfate, magnesium sulfate, manganese sulfate, potassium sulfate, or sodium sulfate), or a combination thereof. For example, a viral lysis buffer may comprise a buffer and a reducing agent, or a viral lysis buffer may comprise a buffer and a chelating agent. The viral lysis buffer may be formulated at a low pH. For example, the viral lysis buffer may be

0 to 25, 0 to 20, 0 to 10, 0 to 5, 5 to 10, 5 to 15, 5 to 20, 5 to 25, to 30, 5 to 40, 5 to 50, 5 to 75, 5 to 100, 5 to 150, 5 to 200, 5 to 250, 5 to 300, 5 to 400, 5 to 500, 25 to 50, 25 to 75, 25 to 100, 50 to 100, 50 to 150, 50 to 200, 50 to 250, 50 to 300, 100 to 200, 100 to 250, 100 to 300, or 150 to 250 mM Imidazole pH 7.5. In some embodiments, the buffer comprises 0 to 500, 0 to 400, 0 to 300, 0 to 250, 0 to 200, 0 to 150, 0 to 100, 0 to 75, 0 to 50, 0 to 25, 0 to 20, 0 to 10, 0 to 5, 5 to 10, 5 to 15, 5 to 20, 5 to 25, to 30, 5 to 40, 5 to 50, 5 to 75, 5 to 100, 5 to 150, 5 to 200, 5 to 250, 5 to 300, 5 to 400, 5 to 500, 25 to 50, 25 to 75, 25 to 100, 50 to 100, 50 to 150, 50 to 200, 50 to 250, 50 to 300, 100 to 200, 100 to 250, 100 to 300, or 150 to 250 mM KCl. In some embodiments, the buffer comprises 0 to 100, 0 to 75, 0 to 50, 0 to 25, 0 to 20, 0 to 10, 0 to 5, 5 to 10, 5 to 15, 5 to 20, 5 to 25, to 30, 5 to 40, 5 to 50, 5 to 75, 5 to 100, 10 to 20, 10 to 30, 10 to 40, 10 to 50, 15 to 20, 15 to 25, 15 to 30, 15 to 4, 15 to 50, 20 to 25, 20 to 30, 20 to 40, or 20 to 50 mM MgCl<sub>2</sub>. In some embodiments, the buffer comprises 0 to 100, 0 to 75, 0 to 50, 0 to 25, 0 to 20, 0 to 10, 0 to 5, 5 to 10, 5 to 15, 5 to 20, 5 to 25, to 30, 5 to 40, 5 to 50, 5 to 75, 5 to 100, 10 to 20, 10 to 30, 10 to 40, 10 to 50, 15 to 20, 15 to 25, 15 to 30, 15 to 4, 15 to 50, 20 to 25, 20 to 30, 20 to 40, or 20 to 50 µg/mL BSA. In some embodiments, the buffer comprises 0 to 1, 0 to 0.5, 0 to 0.25, 0 to 0.01, 0 to 0.05, 0 to 0.025, 0 to 0.01, 0.01 to 0.025, 0.01 to 0.05, 0.01 to 0.1, 0.01 to 0.25, 0.01 to 0.5, 0.01 to 1, 0.025 to 0.05, 0.025 to 0.1, 0.025, to 0.5, 0.025 to 1, 0.05 to 0.1, 0.05 to 0.25, 0.05 to 0.5, 0.05 to 0.75, 0.05 to 1, 0.1 to 0.25, 0.1 to 0.5, or 0.1 to 1% Igepal Ca-630. The buffer can comprise 0 to 25, 0 to 20, 0 to 10, 0 to 5, 5 to 10, 5 to 15, 5 to 20, 5 to 25, 5 to 30% glycerol.

[0340] A buffer of the present disclosure may comprise a viral lysis buffer. In some embodiments, the methods provided herein further comprising lysing the sample. In some embodiments, the lysing comprises contacting the sample to a lysis buffer. In some embodiments, a viral lysis buffer lyses a monkeypox virus capsid in a viral sample (e.g., a sample collected from an individual suspected of having a monkeypox virus infection), releasing a viral genome. In some embodiments, the viral lysis buffer is compatible with amplification (e.g., RT-LAMP amplification) of a target region of the viral genome. In some embodiments, the viral lysis buffer is compatible with detection (e.g., a DETECTR reaction disclosed herein). In some embodiments, a sample is prepared in a one-step sample preparation method comprising suspending the sample in a viral lysis buffer compatible with amplification, detection (e.g., a DETECTR reaction), or both. A viral lysis buffer compatible with amplification (e.g., LAMP amplification), detection (e.g., DETECTR), or both, may comprise a buffer (e.g., Tris-HCl, phosphate, or HEPES), a reducing agent (e.g., N-Acetyl Cysteine (NAC), Dithiothreitol (DTT), β-mercaptoethanol (BME), or tris(2-carboxyethyl) phosphine (TCEP)), a chelating agent (e.g., EDTA or EGTA), a detergent (e.g., deoxycholate, NP-40 (Ipgal), Triton X-100, or Tween 20), a salt (e.g., ammonium acetate, magnesium acetate, manganese acetate, potassium acetate, sodium acetate, ammonium chloride, potassium chloride, magnesium chloride, manganese chloride, sodium chloride, ammonium sulfate, magnesium sulfate, manganese sulfate, potassium sulfate, or sodium sulfate), or a combination thereof. For example, a viral lysis buffer may comprise a buffer and a reducing agent, or a viral lysis buffer may comprise a buffer and a chelating agent. The viral lysis buffer may be formulated at a low pH. For example, the viral lysis buffer may be

formulated at a pH of from about pH 4 to about pH 5. In some embodiments, the viral lysis buffer is formulated at a pH of from about pH 4 to about pH 9. The viral lysis buffer may further comprise a preservative (e.g., ProClin 150). In some embodiments, the viral lysis buffer comprises an activator of the amplification reaction. For example, the buffer may comprise primers, dNTPs, or magnesium (e.g., MgSO<sub>4</sub>, MgCl<sub>2</sub> or MgOAc), or a combination thereof, to activate the amplification reaction. In some embodiments, an activator (e.g., primers, dNTPs, or magnesium) may be added to the buffer following lysis of the monkeypox virus to initiate the amplification reaction.

[0341] A viral lysis buffer may comprise a pH of about 3.5, about 3.6, about 3.7, about 3.8, about 3.9, about 4, about 4.1, about 4.2, about 4.3, about 4.4, about 4.5, about 4.6, about 4.7, about 4.8, about 4.9, about 5, about 5.1, about 5.2, about 5.3, about 5.4, about 5.5, about 6, about 6.5, about 7, about 7.5, about 8, about 8.5, or about 9. In some embodiments, a viral lysis buffer may comprise a pH of from 3.5 to 4.5, from 4 to 5, from 4.5 to 5.5, from 3.5 to 4, from 4 to 4.5, from 4.5 to 5, from 5 to 5.5, from 5 to 6, from 6 to 7, from 7 to 8, or from 8 to 9.

[0342] A viral lysis buffer may comprise a magnesium concentration of about 0 mM, about 2 mM, about 4 mM, about 5 mM, about 6 mM, about 8 mM, about 10 mM, about 12 mM, about 13 mM, about 14 mM, about 15 mM, about 20 mM, about 25 mM, about 30 mM, about 35 mM, about 40 mM, about 45 mM, about 50 mM, about 55 mM, or about 60 mM of magnesium (e.g., MgSO<sub>4</sub>, MgCl<sub>2</sub> or MgOAc). A viral lysis buffer may comprise a magnesium concentration of from 0 mM to 5 mM, from 5 mM to 10 mM, from 10 mM to 15 mM, from 15 mM to 20 mM, from 20 mM to 25 mM, from 25 mM to 30 mM, from 30 mM to 40 mM, from 40 mM to 50 mM, or from 50 mM to 60 mM of magnesium (e.g., MgSO<sub>4</sub>, MgCl<sub>2</sub> or MgOAc). In some embodiments, the magnesium may be added after viral lysis to activate an amplification reaction.

[0343] A viral lysis buffer may comprise a reducing agent (e.g., NAC, DTT, BME, or TCEP) at a concentration of about 1 mM, about 2 mM, about 3 mM, about 4 mM, about 5 mM, about 6 mM, about 7 mM, about 8 mM, about 10 mM, about 12 mM, about 15 mM, about 20 mM, about 25 mM, about 30 mM, about 40 mM, about 50 mM, about 60 mM, about 7 mM, about 80 mM, about 90 mM, about 100 mM, or about 120 mM. A viral lysis buffer may comprise a reducing agent (e.g., NAC, DTT, BME, or TCEP) at a concentration of from 1 mM to 5 mM, from 5 mM to 10 mM, from 10 mM to 15 mM, from 15 mM to 20 mM, from 20 mM to 25 mM, from 25 mM to 30 mM, from 30 mM to 40 mM, from 40 mM to 50 mM, from 50 mM to 60 mM, from 60 mM to 70 mM, from 70 mM to 80 mM, or from 80 mM to 90 mM, from 90 mM to 100 mM, or from 100 mM to 120 mM. A viral lysis buffer may comprise a chelator (e.g., EDTA or EGTA) at a concentration of about 0.1 mM, about 0.2 mM, about 0.3 mM, about 0.4 mM, about 0.5 mM, about 0.6 mM, about 0.7 mM, about 0.8 mM, about 0.9 mM, about 1 mM, about 2 mM, about 3 mM, about 4 mM, about 5 mM, about 6 mM, about 7 mM, about 8 mM, about 10 mM, about 12 mM, about 15 mM, about 20 mM, about 25 mM, or about 30 mM. A viral lysis buffer may comprise a chelator (e.g., EDTA or EGTA) at a concentration of from 0.1 mM to 0.5 mM, from 0.25 mM to 0.5 mM, from 0.4 mM to 0.6 mM, from 0.5 mM to 1 mM, from 1 mM to 5 mM, from 5 mM

to 10 mM, from 10 mM to 15 mM, from 15 mM to 20 mM, from 20 mM to 25 mM, or from 25 mM to 30 mM.

[0344] A viral lysis buffer may comprise a salt (e.g., ammonium acetate ((NH<sub>4</sub>)<sub>2</sub>OAc), magnesium acetate (MgOAc), manganese acetate (MnOAc), potassium acetate (K<sub>2</sub>OAc), sodium acetate (Na<sub>2</sub>OAc), ammonium chloride (NH<sub>4</sub>Cl), potassium chloride (KCl), magnesium chloride (MgCl<sub>2</sub>), manganese chloride (MnCl<sub>2</sub>), sodium chloride (NaCl), ammonium sulfate ((NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>), magnesium sulfate (MgSO<sub>4</sub>), manganese sulfate (MnSO<sub>4</sub>), potassium sulfate (K<sub>2</sub>SO<sub>4</sub>), or sodium sulfate (Na<sub>2</sub>SO<sub>4</sub>)) at a concentration of about 1 mM, about 5 mM, about 10 mM, about 15 mM, about 20 mM, about 25 mM, about 30 mM, about 35 mM, about 40 mM, about 45 mM, about 50 mM, about 55 mM, about 60 mM, about 70 mM, about 80 mM, about 90 mM, or about 100 mM. A viral lysis buffer may comprise a salt (e.g., (NH<sub>4</sub>)<sub>2</sub>OAc, MgOAc, MnOAc, K<sub>2</sub>OAc, Na<sub>2</sub>OAc, NH<sub>4</sub>Cl, KCl, MgCl<sub>2</sub>, MnCl<sub>2</sub>, NaCl, (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, MgSO<sub>4</sub>, MnSO<sub>4</sub>, K<sub>2</sub>SO<sub>4</sub>, or Na<sub>2</sub>SO<sub>4</sub>) at a concentration of from 1 mM to 5 mM, from 1 mM to 10 mM, from 5 mM to 10 mM, from 10 mM to 15 mM, from 15 mM to 20 mM, from 20 mM to 25 mM, from 25 mM to 30 mM, from 30 mM to 35 mM, from 35 mM to 40 mM, from 40 mM to 45 mM, from 45 mM to 50 mM, from 50 mM to 55 mM, from 55 mM to 60 mM, from 60 mM to 70 mM, from 70 mM to 80 mM, from 80 mM to 90 mM, or from 90 mM to 100 mM.

[0345] A viral lysis buffer may comprise a detergent (e.g., deoxycholate, NP-40 (Ipgal), Triton X-100, or Tween 20) at a concentration of about 0.01%, about 0.05%, about 0.10%, about 0.15%, about 0.20%, about 0.25%, about 0.30%, about 0.35%, about 0.40%, about 0.45%, about 0.50%, about 0.55%, about 0.60%, about 0.65%, about 0.70%, about 0.75%, about 0.80%, about 0.85%, about 0.90%, about 0.95%, about 1.00%, about 1.10%, about 1.20%, about 1.30%, about 1.40%, about 1.50%, about 2.00%, about 2.50%, about 3.00%, about 3.50%, about 4.00%, about 4.50%, or about 5.00%. A viral lysis buffer may comprise a detergent (e.g., deoxycholate, NP-40 (Ipgal), Triton X-100, or Tween 20) at a concentration of from 0.01% to 0.10%, from 0.05% to 0.15%, from 0.10% to 0.20%, from 0.15% to 0.25%, from 0.20% to 0.30%, from 0.25% to 0.35%, from 0.30% to 0.40%, from 0.35% to 0.45%, from 0.40% to 0.50%, from 0.45% to 0.55%, from 0.50% to 0.60%, from 0.55% to 0.65%, from 0.60% to 0.70%, from 0.65% to 0.75%, from 0.70% to 0.80%, from 0.75% to 0.85%, from 0.80% to 0.90%, from 0.85% to 0.95%, from 0.90% to 1.00%, from 0.95% to 1.10%, from 1.00% to 1.20%, from 1.10% to 1.30%, from 1.20% to 1.40%, from 1.30% to 1.50%, from 1.40% to 1.60%, from 1.50% to 2.00%, from 2.00% to 2.50%, from 2.50% to 3.00%, from 3.00% to 3.50%, from 3.50% to 4.00%, from 4.00% to 4.50%, or from 4.50% to 5.00%.

[0346] A lysis reaction may be performed at a range of temperatures. In some embodiments, a lysis reaction may be performed at about room temperature. In some embodiments, a lysis reaction may be performed at about 95° C. In some embodiments, a lysis reaction may be performed at from 1° C. to 10° C., from 4° C. to 8° C., from 10° C. to 20° C., from 15° C. to 25° C., from 15° C. to 20° C., from 18° C. to 25° C., from 18° C. to 95° C., from 20° C. to 37° C., from 25° C. to 40° C., from 35° C. to 45° C., from 40° C. to 60° C., from 50° C. to 70° C., from 60° C. to 80° C., from 70° C. to 90° C., from 80° C. to 95° C., or from 90° C. to 99° C. In some embodiments, a lysis reaction may be

performed for about 5 minutes, about 15 minutes, or about 30 minutes. In some embodiments, a lysis reaction may be performed for from 2 minutes to 5 minutes, from 3 minutes to 8 minutes, from 5 minutes to 15 minutes, from 10 minutes to 20 minutes, from 15 minutes to 25 minutes, from 20 minutes to 30 minutes, from 25 minutes to 35 minutes, from 30 minutes to 40 minutes, from 35 minutes to 45 minutes, from 40 minutes to 50 minutes, from 45 minutes to 55 minutes, from 50 minutes to 60 minutes, from 55 minutes to 65 minutes, from 60 minutes to 70 minutes, from 65 minutes to 75 minutes, from 70 minutes to 80 minutes, from 75 minutes to 85 minutes, or from 80 minutes to 90 minutes.

[0347] In some embodiments, the methods, compositions, and kits described herein detect a target nucleic acid (e.g., a nucleic acid from a monkeypox virus) in a sample where the sample is contacted with the reagents for a predetermined length of time sufficient for the trans cleavage to occur or cleavage reaction to reach completion. In some embodiments, the methods, compositions, and kits described herein detect a target nucleic acid in a sample where the sample is contacted with the reagents for no greater than 60 minutes. Sometimes the sample is contacted with the reagents for no greater than 120 minutes, 110 minutes, 100 minutes, 90 minutes, 80 minutes, 70 minutes, 60 minutes, 55 minutes, 50 minutes, 45 minutes, 40 minutes, 35 minutes, 30 minutes, 25 minutes, 20 minutes, 15 minutes, 10 minutes, 5 minutes, 4 minutes, 3 minutes, 2 minutes, or 1 minute. Sometimes the sample is contacted with the reagents for at least 120 minutes, 110 minutes, 100 minutes, 90 minutes, 80 minutes, 70 minutes, 60 minutes, 55 minutes, 50 minutes, 45 minutes, 40 minutes, 35 minutes, 30 minutes, 25 minutes, 20 minutes, 15 minutes, 10 minutes, or 5 minutes. In some embodiments, the methods, compositions, and kits described herein can detect a target nucleic acid or segment thereof in a sample in less than 10 hours, less than 9 hours, less than 8 hours, less than 7 hours, less than 6 hours, less than 5 hours, less than 4 hours, less than 3 hours, less than 2 hours, less than 1 hour, less than 50 minutes, less than 45 minutes, less than 40 minutes, less than 35 minutes, less than 30 minutes, less than 25 minutes, less than 20 minutes, less than 15 minutes, less than 10 minutes, less than 9 minutes, less than 8 minutes, less than 7 minutes, less than 6 minutes, or less than 5 minutes.

[0348] In some embodiments, sample lysis, amplification, detection, or any combination thereof is carried out in a single volume. In some embodiments, sample lysis, amplification, detection, or any combination thereof is carried out in separate volumes.

#### Devices

[0349] A number of devices are consistent with the methods, compositions, and kits disclosed herein.

[0350] The method described herein can be carried out in various ways. In some embodiments, the method is carried out on a lateral flow strip. In some embodiments, the method is carried out on a high throughput workflow. In some embodiments, the method is carried out on a handheld device. In some embodiments, the method is carried out on a cartridge-based workflow.

[0351] The results from a completed assay can be detected and analyzed in various ways. In some embodiments, the device measures or detects a calorimetric, potentiometric, amperometric, optical (e.g., fluorescent, colorimetric, etc.), piezo-electric, piezo-electric, chemical, electrochemical, or

magnetic signal. In some embodiments, the results are visible to the eye and can be read directly by the user without the use of another measurement or detection device. In some embodiments, a positive control spot and a detection spot in the detection region of a lateral flow strip is visible by eye, and the results can be read by the user. In some embodiments, the positive control spot and the detection spot in the detection region is visualized by an imaging device or other device depending on the type of signal. Often, the imaging device is a digital camera, such a digital camera on a mobile device. The mobile device may have a software program or a mobile application that can capture an image of the support medium, identify the assay being performed, detect the detection region and the detection spot, provide image properties of the detection spot, analyze the image properties of the detection spot, and provide a result. Alternatively, or in combination, the imaging device can capture fluorescence, ultraviolet (UV), infrared (IR), or visible wavelength signals. The imaging device may have an excitation source to provide the excitation energy and captures the emitted signals. In some embodiments, the excitation source can be a camera flash and optionally a filter. In some embodiments, the imaging device is used together with an imaging box that is placed over the support medium to create a dark room to improve imaging. The imaging box can be a cardboard box that the imaging device can fit into before imaging. In some embodiments, the imaging box has optical lenses, mirrors, filters, or other optical elements to aid in generating a more focused excitation signal or to capture a more focused emission signal. Often, the imaging box and the imaging device are small, handheld, and portable to facilitate the transport and use of the assay in remote or low resource settings.

[0352] The assay described herein can be visualized and analyzed by a mobile application (app) or a software program. Using the graphic user interface (GUI) of the app or program, an individual can take an image of the support medium, including the detection region, barcode, reference color scale, and fiduciary markers on the housing, using a camera on a mobile device. The program or app reads the barcode or identifiable label for the test type, locate the fiduciary marker to orient the sample, and read the detectable signals, compare against the reference color grid, and determine the presence or absence of the target nucleic acid or segment thereof, which indicates the presence of the gene, virus, or the agent responsible for the disease. The mobile application can present the results of the test to the individual. The mobile application can store the test results in the mobile application. The mobile application can communicate with a remote device and transfer the data of the test results. The test results can be viewable remotely from the remote device by another individual, including a healthcare professional. A remote user can access the results and use the information to recommend action for treatment, intervention, cleanup of an environment.

#### Multiplexing

[0353] The methods, compositions, and kits described herein can be multiplexed in a number of ways, as described herein.

[0354] In some embodiments, multiplexing can be enabled by immobilization of multiple categories of reporters within a fluidic system, to enable detection of multiple target nucleic acids or segments thereof within a single sample.

[0355] The methods, compositions, and kits described herein can be multiplexed by various configurations of the reagents and the support medium. In some embodiments, the kit or system is designed to have multiple support mediums encased in a single housing. In some embodiments, the multiple support mediums housed in a single housing share a single sample pad. The single sample pad may be connected to the support mediums in various designs such as a branching or a radial formation. In some embodiments, each of the multiple support mediums has its own sample pad. In some embodiments, the kit or system is designed to have a single support medium encased in a housing, where the support medium comprises multiple detection spots for detecting multiple target nucleic acids or segments thereof.

[0356] In some embodiments, this reacted sample is applied to the multiplexed support medium described herein. In some embodiments, the methods, compositions, and kits described herein can be multiplexed in a configuration lacking a support medium.

[0357] In some embodiments, multiplexing enables detections of different segments of the same gene of the monkeypox virus at the same time. In some embodiments, multiplexing enables detections of different genes of the monkeypox virus at the same time. In some embodiments, multiplexing enables distinguish between a gene of the monkeypox virus or a segment thereof and a gene of another Orthopoxvirus or a segment thereof. In some embodiments, the other Orthopoxvirus is abatino macacapox virus. In some embodiments, the other Orthopoxvirus is akhmeta virus. In some embodiments, the other Orthopoxvirus is alaskapox virus. In some embodiments, the other Orthopoxvirus is camelpox virus. In some embodiments, the other Orthopoxvirus is cowpox virus. In some embodiments, the other Orthopoxvirus is ectromelia virus. In some embodiments, the other Orthopoxvirus is raccoonpox virus. In some embodiments, the other Orthopoxvirus is skunkpox virus. In some embodiments, the other Orthopoxvirus is taterapox virus. In some embodiments, the other Orthopoxvirus is vaccinia virus. In some embodiments, the other Orthopoxvirus is variola virus. In some embodiments, the other Orthopoxvirus is volepox virus.

#### Detection of Disease as a Research Tool, Point-of-Care, or Over-the-Counter

[0358] Disclosed herein are methods of assaying for one or more target nucleic acid(s) that can be used for disease detection. The disease can be monkeypox. The various methods, compositions, and kits disclosed herein can be used as a companion diagnostic with medicaments used to treat monkeypox. In some embodiments, the methods, compositions, and kits disclosed herein specifically target and assay for a target nucleic acid of monkeypox virus.

[0359] Also disclosed herein are methods of assaying for one or more target nucleic acid(s) that can be used as a research tool, point-of-care, or over-the-counter. In some embodiments, one or more target nucleic acid(s) are from monkeypox virus.

[0360] For example, a method of assaying for a target nucleic acid (e.g., from a monkeypox virus) in a sample comprises contacting the sample to a complex comprising a non-naturally occurring guide nucleic acid comprising a segment that is reverse complementary to a segment of the target nucleic acid and a programmable nuclease that exhibits sequence independent cleavage upon forming a complex

comprising the segment of the non-naturally occurring guide nucleic acid binding to the segment of the target nucleic acid; and assaying for a signal indicating cleavage of at least some protein-nucleic acids of a population of protein-nucleic acids, wherein the signal indicates a presence of the target nucleic acid or segment thereof in the sample and wherein absence of the signal indicates an absence of the target nucleic acid or segment thereof in the sample.

[0361] In yet another example, a method of assaying for a plurality of target nucleic acid in a sample comprise amplifying the target nucleic acid of monkeypox virus using at least one amplification primer; contacting the sample to a plurality of complexes comprising a non-naturally occurring guide nucleic acid comprising a segment that is reverse complementary to a segment of a target nucleic acid of the plurality of target nucleic acids and a programmable nuclease that exhibits sequence independent cleavage upon forming a complex comprising the segment of the non-naturally occurring guide nucleic acid binding to the segment of the target nucleic acid; and assaying for a signal indicating cleavage of at least some protein-nucleic acids of a population of protein-nucleic acids, wherein the signal indicates a presence of the target nucleic acid in the sample and wherein absence of the signal indicates an absence of the target nucleic acid in the sample. The plurality of complexes may comprise programmable nucleases complexes with non-naturally occurring guide nucleic acids directed to different target nucleic acids.

[0362] In some embodiments, a method of assaying for a target nucleic acid in a sample comprising a) amplifying the target nucleic acid using at least one amplification primer; b) contacting the sample to a reporter and a composition comprising a programmable nuclease and a guide nucleic acid that hybridizes to the target nucleic acid or an amplified product thereof, wherein the programmable nuclease cleaves the reporter upon hybridization of the guide nucleic acid to the target nucleic acid or the amplification product thereof; and c) assaying for a change in a signal, wherein the change in the signal is produced by cleavage of the reporter; wherein the target nucleic acid is a gene of a monkeypox virus or a segment thereof; and optionally wherein the at least one amplification primer comprises a nucleotide sequence at least 85%, at least 87%, at least 90%, at least 92%, at least 94%, at least 95%, at least 96%, at least 98%, or 100% identical to any one of SEQ ID NOs: 92-235.

[0363] In some embodiments, any of the reagents (e.g., amplification primers, guide nucleic acids, programmable nucleases, reporters), target nucleic acids, devices, assay formats, or other items disclosed herein can be used as research tools. The research tools can be used to detect any number of target nucleic acids or segments thereof, mutations, or other indications disclosed herein in a laboratory setting. Reagent kits can be provided as reagent packs for open box instrumentation. The high sensitivity lab tests can be performed in a single assay.

[0364] In some embodiments, any of the reagents (e.g., amplification primers, guide nucleic acids, programmable nucleases, reporters), target nucleic acids, devices, assay formats, or other items disclosed herein can be used in a point-of-care (POC) test, which can be carried out at a decentralized location such as a hospital, POL, or clinic. These point-of-care tests can be used to diagnose any of the indications disclosed herein, such as monkeypox. POC tests can be provided as small instruments with a consumable test

card, wherein the test card is any of the assay formats (e.g., a lateral flow assay) disclosed herein. This may be valuable in detecting diseases in a developing country and as a global healthcare tool to detect the spread of a disease or efficacy of a treatment or provide early detection of a disease.

[0365] In some embodiments, any of the reagents (e.g., amplification primers, guide nucleic acids, programmable nucleases, reporters), target nucleic acids, devices, assay formats, or other items disclosed herein can be used in an over-the-counter (OTC), readerless format, which can be used at remote sites or at home to diagnose a range of indications, such as monkeypox. OTC products can include a consumable test card, wherein the test card is any of the assay formats (e.g., a lateral flow assay) disclosed herein. In an OTC product, the test card can be interpreted visually or using a mobile phone.

[0366] Assays that deliver results in under an hour, for example, in 15 to 60 minutes, are particularly desirable for at home testing for many reasons. For example, antivirals can be most effective when administered within the first 48 hours after disease exposure. Thus, the methods disclosed herein, which are capable of delivering results in under an hour, may allow for the delivery of anti-viral therapy during the first 48 hours after infection. Additionally, the systems and assays provided herein, which are capable of delivering quick diagnoses and results, can help keep or send a patient at home, improve comprehensive disease surveillance, and prevent the spread of an infection.

### Compositions

[0367] Disclosed herein are compositions for use to detect a target nucleic acid. The target nucleic acid can be from a monkeypox virus.

[0368] In some embodiments, a composition comprises a non-naturally occurring guide nucleic acid comprising a nucleotide sequence at least 80%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to any one of SEQ ID NOs: 236-247. In some embodiments, a composition comprises an amplification primer comprising a nucleotide sequence at least 80%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to any one of SEQ ID NOs: 92-235.

[0369] In some embodiments, the composition further comprises one or more programmable nuclease(s) described herein. In some embodiments, the composition further comprises one or more reporter(s) described herein. In some embodiments, the composition further comprises one or more reagent(s) for amplification described herein. In some embodiments, the composition further comprises the lysis buffer described herein. In some embodiments, the composition further comprises the negative control nucleic acid described herein.

### Kits

[0370] Disclosed herein are kits, reagents, and systems for use to detect a target nucleic acid. The target nucleic acid can be from a monkeypox virus.

[0371] In some embodiments, the kit comprises the reagents and a support medium. The reagent may be provided in a reagent chamber or on the support medium. Alternatively, the reagent may be placed into the reagent chamber or the support medium by the individual using the kit. Optionally, the kit further comprises a buffer and a dropper. The reagent chamber can be a test well or container. The opening of the reagent chamber may be large enough to accommodate the support medium. The buffer may be provided in a dropper bottle for ease of dispensing. The dropper can be disposable and transfer a fixed volume. The dropper can be used to place a sample into the reagent chamber or on the support medium.

[0372] In some embodiments, a kit for detecting a target nucleic acid comprising a support medium; a non-naturally occurring guide nucleic acid targeting a segment of the target nucleic acid; a programmable nuclease capable of being activated when complexed with the non-naturally occurring guide nucleic acid and the target nucleic acid segment; and a single stranded reporter comprising a detection moiety, wherein the reporter is capable of being cleaved by the activated nuclease, thereby generating a detectable signal.

[0373] In some embodiments, a kit for detecting a target nucleic acid or segment thereof comprising a PCR plate; a non-naturally occurring guide nucleic acid targeting a segment of the target nucleic acid; a programmable nuclease capable of being activated when complexed with the non-naturally occurring guide nucleic acid and the target nucleic acid segment; and a single stranded reporter comprising a detection moiety, wherein the reporter is capable of being cleaved by the activated nuclease, thereby generating a detectable signal. The wells of the PCR plate can be pre-aliquoted with the non-naturally occurring guide nucleic acid targeting a segment of the target nucleic acid, a programmable nuclease capable of being activated when complexed with the non-naturally occurring guide nucleic acid and the target sequence, and at least one population of a single stranded reporter comprising a detection moiety. A user can thus add the biological sample of interest to a well of the pre-aliquoted PCR plate and measure for the detectable signal with a fluorescent light reader or a visible light reader.

[0374] In some embodiments, such kits may include a package, carrier, or container that is compartmentalized to receive one or more containers such as vials, tubes, and the like, each of the container(s) comprising one of the separate elements to be used in a method described herein. Suitable containers include, for example, test wells, bottles, vials, and test tubes. In one embodiment, the containers are formed from a variety of materials such as glass, plastic, or polymers.

[0375] The kit or systems described herein contain packaging materials. Examples of packaging materials include, but are not limited to, pouches, blister packs, bottles, tubes, bags, containers, bottles, and any packaging material suitable for intended mode of use.

[0376] A kit typically includes labels listing contents and/or instructions for use, and package inserts with instructions for use. A set of instructions will also typically be included. In one embodiment, a label is on or associated with the container. In some embodiments, a label is on a container when letters, numbers or other characters forming the label are attached, molded or etched into the container itself; a

label is associated with a container when it is present within a receptacle or carrier that also holds the container, e.g., as a package insert. In one embodiment, a label is used to indicate that the contents are to be used for a specific therapeutic application. The label also indicates directions for use of the contents, such as in the methods described herein.

[0377] After packaging the formed product and wrapping or boxing to maintain a sterile barrier, the product may be terminally sterilized by heat sterilization, gas sterilization, gamma irradiation, or by electron beam sterilization. Alternatively, the product may be prepared and packaged by aseptic processing.

#### INCORPORATION BY REFERENCE

[0378] All publications, patents, and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication, patent, or patent application was specifically and individually indicated to be incorporated by reference.

#### Examples

[0379] The following examples, which are included herein for illustration purposes only, are not intended to be limiting.

##### Example 1: Detection of a Target Nucleic Acid Using Different RNA-Targeting CRISPR/Cas Effector Proteins

[0380] To evaluate the use of different RNA-targeting CRISPR/Cas proteins in a method of quantitating a target nucleic acid, the following experiment was performed. A reagent mixture was generated that comprised the following components: 10 nM of a Cas protein (comprising the amino acid sequence of SEQ ID NO: 21, 62 or 43), 10 nM of a SARS-CoV-2 N gene guide RNA (R4684) or an off-target guide RNA, 1 μM reporter (5Alex647N/rUrUrUrUrU/3IAbRQSp), and buffer (20 mM imidazole pH 7.5, 50 mM KCl, 5 mM MgCl<sub>2</sub>, 10 μg/mL BSA, 0.01% Igepal Ca-630, 5% glycerol). The reagent mixture was mixed with different amounts of the target nucleic acid (synthetic SARS-CoV-2 N gene)—6000 copies/chamber, 2400 copies/chamber, 960 copies/chamber, 384 copies/chamber, 154 copies/chamber, 61 copies/chamber, and 0 copies/chamber. 10 μL of each of the resulting master reaction mixtures was loaded on a partitioning chamber of the digital PCR chip, which comprised about 20,000 nanovolumes. The master reaction mixture was distributed into 20,000 nanovolumes during the partitioning step at 4° C. for 15 min. Subsequently, the nanovolumes were incubated at 37° C. for 30 min, followed by depressurization at ambient temperature for 40 min. Thereafter, the partitioning chambers were scanned to detect fluorescent signals from the cleaved reporter at 150 millisecond exposure using a red laser.

[0381] As shown in FIG. 1, the presence of the target nucleic acid in a particular nanovolume in the chamber was detected by the presence of a positive signal in that nanovolume. Furthermore, the number of positive nanovolumes (that is, the nanovolumes in which a positive signal reflected the presence of the target nucleic acid) was proportional to the number of copies of target nucleic acid added to the chamber, indicating the quantitative nature of this assay. Moreover, when the chamber was loaded with a reaction mixture comprising an off-target guide nucleic acid ("Off

Target"), no nanovolumes in that chamber were seen to be positive, indicating the specificity of this assay in detecting the presence of the target nucleic acid only when all components of the reaction mixture were present. Finally, FIG. 1 also demonstrates that this quantitation assay may be used with multiple different RNA targeting CRISPR/Cas protein enzymes, such as a Cas protein comprising the amino acid sequence of SEQ ID NO: 21, 62 or 43.

##### Example 2: Detection of a Target Nucleic Acid Using Varying Concentrations of an RNA-Targeting CRISPR/Cas Effector Protein

[0382] To try to understand the effects of reducing the CRISPR/Cas effector protein concentration on this quantitation assay, similar reaction mixtures as described in Example 1 were generated using different concentrations of CRISPR/Cas effector protein (0 fM, 12.8 fM, 64 fM, 320 fM, 16 pM, 80 pM, 400 pM, 2 nM, or 10 nM), 10 nM SARS-CoV-2 N-gene targeting guide RNA or an off-target guide RNA (OTG), and 600 or 0 copies per chamber of the target nucleic acid (Twist Synthetic SARS-CoV-2 Synthetic RNA Control) and loaded on the partitioning chamber as described in Example 1. Subsequently, as described in Example 1, the nanovolumes were incubated at 37° C. for 30 min, followed by depressurization at ambient temperature for 40 min and scanning to detect signals from the cleaved reporter. As shown in FIG. 2, while the negative control chambers (no-target control, NTC (0 copies/chamber), with 10 nM CRISPR/Cas protein; off-target guide, OTG, with 10 nM CRISPR/Cas protein; and "0 nM", the reaction mixture with no CRISPR/Cas protein) did not show any fluorescence as expected, assay sensitivity was maintained even at extremely low concentrations of the CRISPR/Cas protein, e.g., in the range of fM.

[0383] A similar experiment was performed with 600 copies per chamber of a different source of SARS-CoV-2 target nucleic acid (ATCC® synthetic SARS-CoV-2 RNA), which further confirmed that extremely low concentrations of the CRISPR/Cas protein may be used for quantitation in the assay disclosed here with a sensitivity of detection that is comparable to using higher concentrations of the CRISPR/Cas protein (FIG. 3).

##### Example 3: Detection of a Target Nucleic Acid in the Presence of Other Reagents

[0384] To evaluate whether low pH crude lysis buffer has any inhibitory effects on the quantitation assays described herein, different volumes (0%, 5%, 10%, or 20%) of pH lysis buffer were added to the reaction mixture prepared as described in Example 1, using 10 nM Cas protein comprising the amino acid sequence of SEQ ID No: 21 and 5000 copies/chamber of Twist Synthetic SARS-CoV-2 Synthetic RNA Control target nucleic acid before distributing the reaction mixture into nanovolumes within the chamber. Each mixture also comprised a SARS-CoV-2 N-gene targeting guide RNA (R4684) or an off-target guide (OTG, R5882) as a negative control.

[0385] As shown in FIG. 4, positive nanovolumes were reliably obtained even when 10% lysis buffer was used, indicating that the Cas protein comprising the amino acid sequence of SEQ ID No: 21 is tolerant of crude lysis buffer up to 10% of the total assay volume. 20% lysis buffer started to show some dropout of positive nanovolumes under the

conditions tested, though additional optimization could be done to improve performance with higher concentrations of lysis buffer if desired.

[0386] To further evaluate the effect of other reagents on the quantitative assays described herein, carrier molecules (e.g., yeast tRNA, glycogen, polyvinylpyrrolidone, PVP)

Influenza B, as listed in Table A. 10  $\mu$ L of each of the resulting master reaction mixtures was loaded on a partitioning chamber of the digital PCR chip, which comprised about 20,000 nanovolumes. The master reaction mixture was processed, and the signals were detected as described in Example 1.

TABLE A

Amounts of the synthetic SARS-CoV-2 target nucleic acid and two off-target nucleic acids from Influenza A and Influenza B used in the study				
Chamber number in FIG. 6	Guide	Amount of SARS-CoV2 target nucleic acid (copies/chamber)	Influenza A Virus nucleic acid (copies/chamber)	Influenza B Virus nucleic acid (copies/chamber)
1	R4684	4000	4000	4000
2	R4684	400	400	400
3	R4684	40	40	40
4	R4684	0	0	0
5	R4684	4000	4000	4000
6	R4684	400	4000	4000
7	R4684	40	4000	4000
8	R4684	0	4000	4000
25	R4684	4000	4000	4000
26	R4684	400	4000	4000
27	R4684	40	4000	4000
28	R4684	0	4000	4000

were tested at different concentrations for their ability to help stabilize RNA during the distribution of the reaction mixture into nanovolumes. Reaction mixtures containing the Cas protein comprising the amino acid sequence of SEQ ID NO: 21 and Twist Synthetic SARS-CoV-2 Synthetic RNA Control target nucleic acid were prepared as described above, and different concentrations of carrier molecules (yeast tRNA, glycogen, linear acrylamide, and PVP) were added before distribution into nanovolumes. The chambers were processed and imaged as described in Example 1. FIG. 5 shows that the addition of tRNA resulted in high background signal under the conditions tested even in the presence of the negative non-targeting control (NTC) nucleic acid. However, the use of the other carrier molecules tested, such as, glycogen, linear acrylamide and PVP did not interfere with the assay reactions and signal detection under the conditions tested.

#### Example 4: Specific Detection of a Target Nucleic Acid in a Mixture of Viral RNAs

[0387] To evaluate the use of the quantitative assays disclosed herein to specifically detect a viral target nucleic acid in a mixture of RNAs derived from three different viruses, the following experiment was performed.

[0388] A reagent mixture was generated that comprises the following components: 10 nM Cas protein comprising the amino acid sequence of SEQ ID NO: 21, R4684 guide RNA (GCCACCCCAAAAUGAAGGGGAC-UAAAACAAGAACAUUCAGAUUUUUA (SEQ ID NO: 87)), 1  $\mu$ M reporter (5Alex647N/rUrUrUrUrU/3IAbRQSp), and buffer (20 mM imidazole pH 7.5, 50 mM KCl, 5 mM MgCl<sub>2</sub>, 10  $\mu$ g/mL BSA, 0.01% Igepal Ca-630, 5% glycerol). The reagent mixture was mixed with different amounts of the synthetic SARS-CoV-2 target nucleic acid and two off-target nucleic acids from Influenza A and

[0389] As shown in FIG. 6, when the concentration of all the three RNAs (the target SARS-CoV-2 RNA and the off-target Influenza A and B RNAs) was decreased, a corresponding decrease in the number of positive nanovolumes was seen (top row). Notably, a similar decrease in the number of positive nanovolumes was seen when decreasing the concentration of just the SARS-CoV-2 target nucleic acid (duplicates shown in middle and bottom rows). These results demonstrate that the positive nanovolumes were a result of specific detection of the target SARS-CoV-2 RNA. Accordingly, the quantitative assays described herein are capable of specifically detecting and quantitating a target nucleic acid in a mixture of multiple different nucleic acids, suggesting that the assays described herein may be effectively used to quantitate nucleic acids in complex samples, such as clinical patient samples, for diagnostic purposes. Furthermore, given the specificity of the detection indicated by these data, the quantitative assays described herein may be multiplexed to detect multiple different target nucleic acids in the sample.

#### Example 5: Detection of a Target Nucleic Acid Using Different DNA-Targeting CRISPR/Cas Effector Proteins

[0390] To evaluate the use of different DNA-targeting CRISPR/Cas proteins in the quantitation assays disclosed herein, the following experiment was performed. A reagent mixture was generated that comprised the following components: 40 nM of the Cas protein comprising the amino acid sequence of SEQ ID No: 34, 62.5 nM of a SARS-CoV-2 N gene crRNA (GCCACCCCAAAAUGAAGGGGAC-UAAAACAAGAACAUUCAGAUUUUUA (SEQ ID NO: 87)), 1.5  $\mu$ M of a ssDNA reporter (5Alex647N/TTATTATT/3IAbRQSp, or 5Alex647N/TTTTTTTTTTT/3IAbRQSp (SEQ ID NO: 88)), and buffer (20 mM Tris-HCl, pH 8.8, 2 mM KOAc, 0.02 mg/mL BSA, 15 mM MgOAc). [0391] The reagent mixture was mixed, either with different amounts of the target nucleic acid (a synthetic SARS-

CoV-2 N gene dsDNA), for example, 25,000 copies/chamber, 18,750 copies/chamber, 12,500 copies/chamber/chamber, or 6,250 copies/chamber; or with a no-target control (NTC; 0 copies/reaction) or an off-target control (OTC). 25  $\mu$ L of each of the resulting master reaction mixtures was loaded on a partitioning chamber of the digital PCR chip, which comprised about 30,000 nanovolumes. The master reaction mixture was distributed into nanovolumes during the partitioning step at 4° C. for 15 min. Subsequently, the nanovolumes were incubated at 60° C. for 90 min, followed by depressurization at ambient temperature for 40 min. Thereafter, the chambers were scanned to detect signals from the cleaved reporter at 50 millisecond exposure using a red laser.

[0392] As shown in FIG. 7, the presence of the target DNA in a particular nanovolume in the chamber was detected by the presence of a positive signal in that nanovolume. Notably, when the chamber was loaded with a reaction mixture comprising an off-target DNA or no-target DNA, no nanovolumes in that chamber were seen to be positive, indicating the specificity of this assay in detecting the presence of only the target nucleic acid. FIG. 7 further demonstrates that the quantitation assays described herein may be used not only with RNA-targeting Cas proteins, but also with DNA-targeting Cas proteins, highlighting the versatility of this assay.

[0393] The ability of the assays described herein to detect and quantitate target DNA was further investigated using another DNA-targeting Cas protein comprising the amino acid sequence of SEQ ID NO: 3. A similar reagent mixture was generated as described above, and mixed with either the target nucleic acid (a synthetic SARS-CoV-2 N gene dsDNA) or a no-target control (NTC). 25  $\mu$ L of each of the resulting master reaction mixtures was loaded on a chamber of the digital PCR chip in a QIAcuity Digital PCR System. The master reaction mixture was distributed into nanovolumes during the partitioning step at 23° C. for 15 min. Subsequently, the nanovolumes were incubated at 55° C. for 45 min. Thereafter, the chambers were scanned to detect signals from the cleaved reporter.

[0394] As shown in FIG. 8, the presence of the target DNA in a particular nanovolume in the chamber was detected by the presence of a positive signal in that nanovolume. Notably, when the chamber was loaded with a reaction mixture comprising a non-targeting DNA, no nanovolumes in that chamber were seen to be positive, indicating the specificity of this assay in detecting the presence of only the target nucleic acid. FIG. 8 further establishes that the quantitation assays described herein may be used with multiple different DNA targeting Cas proteins.

#### Example 6: Detection of Synthetic Standards Representing Wastewater Targets

[0395] To evaluate whether the quantitative assays described herein may be used to detect and quantitate target nucleic acids in wastewater samples, the following experiments were performed. Nucleic acids derived from SARS-CoV2 and PMMOV can be used as targets for their detection in wastewater.

[0396] A reagent mixture was generated that comprised the following components: Cas protein, SARS-CoV-2 N gene crRNA or PMMOV crRNA, and 1.5  $\mu$ M reporter. The reagent mixture was mixed with: (i) different concentrations of the target nucleic acid from dilution D1 down to dilution

D6 (a synthetic SARS-CoV-2 N gene or a synthetic nucleic acid derived from pepper mild mottle virus (PMMoV)); or (ii) a non-targeting control (NTC); or (iii) an off-targeting control (OTC), as indicated in FIGS. 9A and 9B. The reaction mixtures were processed and signals detected as described in Example 1.

[0397] FIG. 9A shows that the quantitative assays described herein employing the SARS-CoV-2 guide RNA specifically detected SARS-CoV-2 target nucleic acid, while emitting minimal or no signal upon exposure to PMMOV nucleic acid (OTC in FIG. 9A). FIG. 9B, on the other hand, shows that the quantitative assays described herein employing the PMMOV guide RNA specifically detected PMMOV target nucleic acid, while emitting minimal or no signal upon exposure to SARS-CoV-2 nucleic acid (OTC in FIG. 9B). These results illustrate not only the specificity of the quantitative assays described herein, but also that these assays can be used to effectively detect and quantitate target nucleic acids that are found in a complex, real-world sample, such as, wastewater.

[0398] Moreover, FIGS. 10A and 10B show standard curves generated using the detection of SARS-CoV-2 target nucleic acid or PMMOV target nucleic acid, which show a relatively linear relationship between the number of positive nanovolumes detected (y-axis) and the number of copies added per chamber (x-axis). The curves highlight that the quantitative assays described herein may be used for absolute or relative quantitation of a target nucleic acid in a wastewater sample, as compared to control samples comprising defined amounts of the target nucleic acid.

[0399] In sum, the experiments described herein demonstrate the wide range of applications for the quantitative assays described herein, including in quantitating target RNA using RNA-targeting CRISPR-Cas proteins, quantitating target DNA using DNA-targeting CRISPR-Cas proteins, quantitating a single target nucleic acid specifically among a mixture of different nucleic acids, and quantitating a target nucleic acid in a complex sample, such as, wastewater.

#### Example 7: Detection of Different Concentrations of Target Nucleic Acid Using an RNA-Targeting CRISPR/Cas Effector Protein

[0400] To evaluate the threshold of detection of an RNA-targeting CRISPR/Cas protein comprising the amino acid sequence of SEQ ID NO: 21 in a method of quantitating a target nucleic acid, the following experiment was performed. A reagent mixture was generated that comprised the following components: 10 nM of a Cas protein (comprising the amino acid sequence of SEQ ID NO: 21), 10 nM of a SARS-CoV-2 N gene guide RNA (R4684) or an off-target guide RNA, and buffer (20 mM imidazole pH 7.5, 50 mM KCl, 5 mM MgCl<sub>2</sub>, 10  $\mu$ g/mL BSA, 0.01% Igepal Ca-630, 5% glycerol). The reagent mixture was incubated at 37° C. for 30 minutes to pre-complex the Cas protein and guide RNA, then 1  $\mu$ M reporter (5Alex647N/rUrUrUrUrU/3IAbRQSp) was added to the mix. The reagent mixture was mixed with different amounts of the target nucleic acid (synthetic SARS-CoV-2 N gene)-10,000 copies/chamber (3,333 copies/ $\mu$ L), 5,000 copies/chamber (1,667 copies/ $\mu$ L), 2,500 copies/chamber (833 copies/ $\mu$ L), 1,250 copies/chamber (417 copies/ $\mu$ L), 625 copies/chamber (208 copies/ $\mu$ L), 313 copies/chamber (104 copies/ $\mu$ L), 157 copies/chamber (52 copies/ $\mu$ L), 78 copies/chamber (26 copies/ $\mu$ L), 39 copies/chamber (13 copies/ $\mu$ L), 20 copies/chamber (7 copies/

$\mu\text{L}$ ), 10 copies/chamber (3 copies/ $\mu\text{L}$ ), and 0 copies/chamber (10 copies/ $\mu\text{L}$ ). 10  $\mu\text{L}$  of each of the resulting master reaction mixtures was loaded on a partitioning chamber of the digital PCR chip, which comprised about 20,000 nanovolumes. The master reaction mixture was distributed into 20,000 nanovolumes during the partitioning step at 4° C. for 15 min. Subsequently, the nanovolumes were incubated at 37° C. for 30 min, followed by depressurization at ambient temperature for 40 min. Thereafter, the partitioning chambers were scanned to detect fluorescent signals from the cleaved reporter at 150 millisecond exposure using a red laser.

[0401] As shown in FIG. 11, the presence of the target nucleic acid in a particular nanovolume in the chamber was detected by the presence of a positive signal in that nanovolume. Furthermore, the number of positive nanovolumes (that is, the nanovolumes in which a positive signal reflected the presence of the target nucleic acid) was proportional to the number of copies of target nucleic acid added to the chamber, indicating the quantitative nature of this assay. Moreover, when the chamber was loaded with a reaction mixture comprising an off-target guide nucleic acid ("Off Target"), no nanovolumes in that chamber were seen to be positive, indicating the specificity of this assay in detecting the presence of the target nucleic acid only when all components of the reaction mixture were present. Finally, FIG. 11 also demonstrates that this quantitation assay is relatively linear over a range of at least 0 copies/ $\mu\text{L}$  to 3,333 copies/ $\mu\text{L}$  under the conditions tested. Interestingly, the slope of the experimental digital DETECTR copies detected to expected copies detected was not 1:1, indicating that the system was not operating at 100% efficiency under the conditions tested. Without being bound by any particular theory, it is hypothesized that one possible reason for this observed inefficiency is due to the efficiency of the gRNA (and/or CRISPR/Cas enzyme and/or complex) in binding to the target nucleic acid. Such guide efficiency can be adjusted for as described herein in order to better correlate the observed number of positive nanovolumes with the actual copies per  $\mu\text{L}$  in a sample of interest.

#### Example 8: Detection of a Target Nucleic Acid in Clinical Samples

[0402] To evaluate the use of an RNA-targeting CRISPR/Cas protein comprising the amino acid sequence of SEQ ID NO: 21 in a method of quantitating a target nucleic acid extracted from a clinical sample, the following experiment was performed. A digital DETECTR reagent mixture was generated that comprised the following components: 10 nM

of a Cas protein (comprising the amino acid sequence of SEQ ID NO: 21), 10 nM of a SARS-CoV-2 N gene guide RNA (R4684) or an off-target guide RNA, 1  $\mu\text{M}$  reporter (5Alex647N/rUrUrUrUrU/3IAbRQSp), and buffer (20 mM imidazole pH 7.5, 50 mM KCl, 5 mM MgCl<sub>2</sub>, 10  $\mu\text{g}/\text{mL}$  BSA, 0.01% Igepal Ca-630, 5% glycerol). A digital droplet PCR (ddPCR) reagent mixture was generated using the 3-Color Crystal Digital PCR™ kit for detection of COVID-19 from ApexBio. Viral nucleic acids were purified from SARS-CoV-2-positive clinical respiratory samples in saline using the EZ1&2 Virus Mini Kit v2.0 from Qiagen. The digital DETECTR reagent mixture or ddPCR reagent mixture was mixed with a 1:100 dilution, 1:10 dilution, or undiluted sample of the extracted viral nucleic acids from each clinical sample or water (non-template control, NTC). 25  $\mu\text{L}$  of each of the resulting ddPCR or digital DETECTR reaction mixtures was loaded on a partitioning chamber of the digital PCR chip, which comprised about 20,000 nanovolumes. The ddPCR reaction mixture was distributed into 20,000 nanovolumes during the partitioning step at 40° C. for 12 min. Subsequently, the ddPCR nanovolumes were incubated at 50° C. for 10 min to allow for cDNA synthesis to occur before an initial denaturation step at 95° C. for 1 min, and then thermocycled for 45 cycles (denaturation 95° C. for 10 seconds, annealing/extension at 55° C. for 30 seconds). The partitioning chamber was then depressurized at ambient temperature for 33 min. Thereafter, the ddPCR partitioning chambers were scanned to detect fluorescent signals from the ddPCR nanovolumes. The digital DETECTR reaction mixture was distributed into 20,000 nanovolumes during the partitioning step at 4° C. for 15 min. Subsequently, the digital DETECTR nanovolumes were incubated at 37° C. for 30 min, followed by depressurization at ambient temperature for 40 min. Thereafter, the digital DETECTR partitioning chambers were scanned to detect fluorescent signals from the cleaved reporter at 150 millisecond exposure using a red laser.

[0403] As shown in FIG. 12, the presence of the target nucleic acid in a particular nanovolume in the chamber was detected by the presence of a positive signal in that nanovolume. Furthermore, the number of positive nanovolumes (that is, the nanovolumes in which a positive signal reflected the presence of the target nucleic acid) was proportional to the number of copies of target nucleic acid added to the chamber for both ddPCR and digital DETECTR assayed clinical samples. FIG. 13 and TABLE B below show side-by-side comparisons of the number of copies detected by ddPCR or digital DETECTR for three clinical samples, indicating the quantitative, linear nature of this assay under the conditions tested.

TABLE B

SARS-CoV-2 Clinical Sample Comparison of ddPCR and digital DETECTR			
	Undiluted (copies/uL)	1/10 Dilution (copies/uL)	1/100 Dilution (copies/uL)
clinical sample #1 (ddPCR)	2543.0	147.7	28.2
clinical sample #1 (DETECTR)	2809.8	208.5	18.7
clinical sample #2 (ddPCR)	235.2	13.7	1.9
clinical sample #2 (DETECTR)	139.4	13.9	1.6
clinical sample #3 (ddPCR)	1092.2	111.7	14.6
clinical sample #3 (DETECTR)	235.5	49.1	6.9

**Example 9: Detection of a Target Nucleic Acid in Wastewater Samples**

**[0404]** To evaluate whether the quantitative assays described herein may be used to detect and quantitate target nucleic acids in wastewater samples, the following experiments were performed. Wastewater samples were collected over a period of two months and viral RNA was extracted therefrom. Digital DETECTR reagent mixtures for SARS-CoV-2 N-gene or PMMoV were generated and processed, and signals detected, as described in Example 6. Digital PCR reagent mixtures for SARS-CoV-2 N-gene were generated and processed, and signals detected, as described in Example 8. Digital PCR reagent mixtures for PMMoV generated using the UltraPlex 1-Step ToughMix® qPCR kit from Quanta BioSciences and PMMoV primers obtained from the SARS-CoV-2 RT-qPCR Kit for Wastewater from Promega. Digital PCR reagent mixtures for PMMoV were processed, and signals detected, as described in Example 8 with respect to SARS-CoV-2. Standard curves for SARS-CoV-2 and PMMoV were generated as described in Example 6 and used to adjust the digital DETECTR copy number values based on gRNA efficiency. Briefly, standard curves generated with actual vs. expected copies as described in Example 6, and the slope of the curve used as an “efficiency coefficient” to transform measured DETECTR copy numbers into actual copy numbers for comparison against ddPCR.

**[0405]** FIG. 14 shows a comparison of the number of copies per microliter (cp/μL) determined for three wastewater samples using digital DETECTR and ddPCR. These data show that digital DETECTR has a roughly equivalent dynamic range to digital PCR for viral quantification under the conditions tested for both SARS-CoV-2 and PMMOV RNA in wastewater samples. FIGS. 15A and 15B show the difference in signal obtained for digital DETECTR-assayed wastewater samples compared to dPCR-assayed wastewater samples.

**[0406]** While the number of copies/μL obtained using positive control samples with known concentrations (dilutions 1–4 or NTC) for standard curve generation was relatively similar for both dPCR and digital DETECTR, the signal generated by dPCR was significantly lower than that of the digital DETECTR signal for corresponding wastewater samples. FIG. 15A shows a head-to-head comparison of the digital DETECTR and dPCR signals graphed with the same scaling along the Y-axis while FIG. 15B shows the digital DETECTR and dPCR signals graphed with different Y-axis scaling, in order to highlight the stark differences in signals obtained by each method. One possible reason for this observed difference in signal strength is that dPCR may be more sensitive to inhibitors (e.g., salts) left in the wastewater samples following sample prep and viral RNA extraction than digital DETECTR. These results illustrate not only the specificity of the digital DETECTR quantitative assays described herein, but also that these assays can be used to effectively detect and quantitate target nucleic acids that are found in a complex, real-world sample, such as, wastewater.

**Example 10: Detection of a Target Nucleic Acid Using Different DNA-Targeting CRISPR/Cas Effector Proteins at Different Reaction Temperatures**

**[0407]** To evaluate the use of different DNA-targeting CRISPR/Cas proteins at different reaction temperatures in

the quantitation assays disclosed herein, the following experiment was performed. A reagent mixture was generated that comprised the following components: 40 nM of the Cas protein comprising the amino acid sequence of SEQ ID No: 65, 67, 68, 34, 17, 89, 90, or 91, 40 nM of a SARS-CoV-2 N gene crRNA, 1.5 μM of a ssDNA reporter (5Alex647N/TTATTTATT/3IAbRQSp, or 5Alex647N/TTTTTTTTTTTT/3IAbRQSp (SEQ ID NO: 88)), and buffer (20 mM Tris-HCl, pH 8.8, 2 mM KOAc, 0.02 mg/mL BSA, 15 mM MgOAc). The reagent mixture was incubated at 37° C. for 30 minutes to pre-complex the Cas protein and guide RNA. The reagent mixture was then mixed with 1000 copies/microliter of the target nucleic acid (a synthetic SARS-CoV-2 N gene dsDNA) or with a no-target control (NTC; 0 copies/reaction) or an off-target control (OTC). 40 μL of each of the resulting master reaction mixtures was loaded on a partitioning chamber of the digital PCR chip, which comprised about 30,000 nanovolumes. The master reaction mixture was distributed into nanovolumes during the partitioning step at room temperature for 15 min. Subsequently, the nanovolumes were incubated at 37° C. or 50° C. for 45 minutes, followed by depressurization at ambient temperature for 40 min. Thereafter, the chambers were scanned to detect signals from the cleaved reporter at 50 millisecond exposure using a red laser.

**[0408]** As shown in FIG. 16, the presence of the target DNA in a particular nanovolume in the chamber was detected by the presence of a positive signal in that nanovolume. FIG. 16 further establishes that the quantitation assays described herein may be used with multiple different DNA-targeting Cas proteins and that different DNA-targeting Cas proteins may be operative at different temperatures, which may be particularly useful for warm-start methods such as those described in Example 11.

**Example 11: Warm-Start Detection of a Target Nucleic Acid**

**[0409]** To evaluate the use of delaying tactics to improve signal quality and quantitative range, a warm-start strategy is employed to reduce and/or prevent activation of the CRISPR/Cas complex during partitioning until the complex has reached a predetermined reaction temperature. A reagent mixture is generated at room temperature that comprises the following components: 40 nM of a Cas protein comprising the amino acid sequence of SEQ ID No: 34, 62.5 of nM a SARS-CoV-2 N gene crRNA (GCCACCC-CAAAAUGAAGGGGAC-UAAAACAAGAAAGAUUCAGAUUUUA (SEQ ID NO: 87)), 1.5 μM of a ssDNA reporter (5Alex647N/TTATTTATT/3IAbRQSp, or 5Alex647N/TTTTTTTTTTTT/3IAbRQSp (SEQ ID NO: 88)), and buffer (20 mM Tris-HCl, pH 8.8, 2 mM KOAc, 0.02 mg/mL BSA, 15 mM MgOAc). The Cas protein is associated with a blocking thermosensitive aptamer which prevents catalytic activity of the Cas protein at or below room temperature. The blocking thermosensitive aptamer is configured to denature at temperatures about 50° C.

**[0410]** The reagent mixture is mixed, either with different amounts of the target nucleic acid (a synthetic SARS-CoV-2 N gene dsDNA), for example, 25,000 copies/chamber, 18,750 copies/chamber, 12,500 copies/chamber/chamber, or 6,250 copies/chamber; or with a no-target control (NTC; 0 copies/reaction) or an off-target control (OTC). 25 μL of each of the resulting master reaction mixtures is loaded on

a partitioning chamber of the digital PCR chip, which comprises about 30,000 nanovolumes. The master reaction mixture was distributed into nanovolumes during the partitioning step at 4° C. for 15 min. Subsequently, the nanovolumes are incubated at 60° C. for 90 min (at which point the aptamer is denatured and the Cas protein is activated), followed by depressurization at ambient temperature for 40 min. Thereafter, the chambers are scanned to detect signals from the cleaved reporter at 50 millisecond exposure using a red laser. The presence of the target DNA in a particular nanovolume in the chamber is detected by the presence of a positive signal in that nanovolume. When the chamber is loaded with a reaction mixture comprising an off-target DNA or no-target DNA, no nanovolumes in that chamber will be seen to be positive.

[0411] FIG. 17 shows a schematic of the exemplary warm-start strategy which may be employed to delay activation of the CRISPR/Cas complex until the complex has reached a predetermined reaction temperature.

**Example 12: Two-Pot LAMP DETECTR Reactions for Detection of Monkeypox Virus**

[0412] This example describes LAMP DETECTR reactions for the detection of monkeypox virus in a two-step, two-pot reaction.

[0413] SgRNA-Cas14a.1 complexes are prepared by pre-incubating Cas 14a.1 with sgRNA MPXV-1 (SEQ ID NO: 236) for 30 minutes at 37° C. After formation of the sgRNA-Cas14a.1 complexes, a labeled ssDNA reporter is added to the reaction.

[0414] Samples of patients either positive or negative for monkeypox virus are collected. Viral DNA targets in the samples are amplified by loop mediated amplification (LAMP) using six amplification primers: M44982 monkeypox-set1-F3 (SEQ ID NO: 92), M44983 monkeypox-set1-B3 (SEQ ID NO: 93), M44984 monkeypox-set1-FIP (SEQ ID NO: 94), M44985 monkeypox-set1-BIP (SEQ ID NO: 95), M44986 monkeypox-set1-LF (SEQ ID NO: 96), M44987 monkeypox-set1-LB (SEQ ID NO: 97).

[0415] After completion of the pre-amplification step, 2 μL of amplicon is combined with 18 μL of sgRNA-Cas14a.1 complexes and 80 μL of 1X reaction buffer. The 100 μL Cas14a.1 trans-cleavage assay is allowed to proceed for 10 minutes at 37° C.

[0416] A lateral flow strip (Milenia HybriDetect 1, TwistDx) is then added to the reaction tube and a result is visualized after approximately 2-3 minutes. A single band, close to the sample application pad indicates a negative result, whereas a single band close to the top of the strip or two bands indicates a positive result.

**Example 13: One-Pot LAMP DETECTR Reactions for Detection of Monkeypox Virus**

[0417] This example describes LAMP DETECTR reactions for the detection of monkeypox virus in a one-step, one-pot reaction.

[0418] SgRNA-Cas14a.1 complexes are prepared by pre-incubating Cas 14a.1 with sgRNA MPXV-1 (SEQ ID NO: 236) for 30 minutes at 37° C. After formation of the sgRNA-Cas14a.1 complexes, a fluorophore-quencher labeled ssDNA reporter, amplification primers, dNTPs, and a polymerase are added to the reaction in a 1X reaction buffer.

[0419] Samples of patients either positive or negative for monkeypox virus are collected and added to the reaction and the one-pot amplification/trans-cleavage assay is allowed to proceed for 30 minutes at 37° C. Viral DNA targets in the samples are amplified by loop mediated amplification (LAMP) using six amplification primers: M44982 monkeypox-set1-F3 (SEQ ID NO: 92), M44983 monkeypox-set1-B3 (SEQ ID NO: 93), M44984 monkeypox-set1-FIP (SEQ ID NO: 94), M44985 monkeypox-set1-BIP (SEQ ID NO: 95), M44986 monkeypox-set1-LF (SEQ ID NO: 96), M44987 monkeypox-set1-LB (SEQ ID NO: 97). Amplification occurs concurrently with detection of the viral DNA targets and/or amplicons. Presence of the target DNA, or an amplicon thereof, within the sample activates the sgRNA-Cas14a.1 complex to trans-cleave the reporter, thereby separating the fluorophore and the quencher from one another and resulting in a detectable fluorescent signal. Absence of the target DNA, or an amplicon thereof, within the sample results in no trans-cleavage of the reporter, and thus no separation of the fluorophore and the quencher from one another and no detectable fluorescent signal.

[0420] The foregoing is illustrative of the present disclosure, and is not to be construed as limiting thereof. The invention is defined by the following claims, with equivalents of the claims to be included therein.

---

**SEQUENCE LISTING**

```
Sequence total quantity: 478
SEQ ID NO: 1      moltype = AA  length = 1307
FEATURE          Location/Qualifiers
source           1..1307
mol_type = protein
organism = synthetic construct
SEQUENCE: 1
MADLSQFTHK YQVPKTLRFE LIPQGKTLEN LSAYGMVADD KQRSENYKKL KPVIDRIYKY 60
FIEESLKNTN LDWNPLYEQA REYRKETTA TITNLKEQQD ICRRAIASRF EGKVPDKGDK 120
SVKDFNKKQS KLFKELEFGKE LFTDSVLEQL PGVSLSDEDK ALLKSFDKFT TYFVGFYDNR 180
KNVFSSDDIS TGIPHRLVQE NFPKFIDNCD DYKRLVLVAP ELKEKLEKAA EATKIFEDVS 240
LDEIFSIKFY NRLLQQNQID QFNQLLGGIA GAPGTPKIQG LNETLNLSMQ QDKTLEQKLK 300
SVPHRSPLV KQILSDRSSL SFIPESFSCD AEVLLAVQEV LDNLKTEHVI EDLKEVFNRL 360
TTLDLKHIYV NSTKVTAFSQ ALFGDWNLCR EQLRVYKMSN GNEKITKKAL GELESWLKNS 420
DIAFTTELQEA LADEALPAKV NLKVQEAIISG LNEQMAKSLV KELKIPREEKE ELKALLDAIQ 480
EYVHTLEWFI VSDDVETDTD FVYPLKETLQ IIQPIIPLYN KVRNFATQKP YSVEKFKLFN 540
ANPTLADGW ENKEQQNCAV LFQKGNNYYL GILNPKNKPD FDNVDTEKQG NCYQKMVYKQ 600
FPDFSKMMMPK CTTQLKEVKQ HFEGKDSDYI LNNKNFIKPL TITREVYDLN NVLYDGKKF 660
QIDYLRKTKD EDGYYTALHT WIDFAKKPVA SYKSTSITYD STILPPEKYE KLNEFYGALD 720
```

-continued

---

NLFYQIKFEN	IPEEIIDTYV	EDGKLFLFQI	YNKDFAAGAT	GAPNLHTIYW	KAVFDPENVK	780
DVVVKLNGQA	ELFYRPKSNN	DVIRHKVGEEK	LVNRTLKDGS	ILTDELHKEL	YLYANGSLKK	840
GLSEDAKIIL	DKNLAVIYDV	HHEIVKDRRF	TTDKFFFHVP	LTLNYKCDKN	PVKFNAEVQE	900
YLKENPDTYV	IGIDGERNL	IYAVVIDPKG	RIVEQKSFNV	INGFDYHGL	DQREKERVKA	960
RQAWTAVGKI	KELKQGYLSL	VVHEISKMMV	RYQAVVVLEN	LNVGFKRVRS	GIAEKAVYQQ	1020
FEKMLINKLN	YLMFKDAGGT	EPGVLNAYQ	LTDTRFESFAK	MGLQTGFLFY	IPAFTSKID	1080
PATGFVDPFR	WGAIKTLADK	REFLSGFESL	KFDSTTGNF1	LHFDVSKNN	FQKKLEGFPV	1140
DWIIIEANK	MKTGKGATYI	AGKRIEFVRL	NNSQGHYEDY	LPCNALAETL	RQCDIPYEEG	1200
KDILPLILEK	NDSKLLHSVF	KVVRLLQLMQR	NSNAETGEDY	ISSPVVEDVSG	SCFDSRMENE	1260
KLKPDA	DADANG	AYHIALKGML	ALERLRKDEK	MAISNNDWLN	YIQEKRA	1307

SEQ ID NO: 2	moltype = AA	length = 492				
FEATURE	Location/Qualifiers					
source	1..492					
	mol_type = protein					
	organism = synthetic construct					
SEQUENCE: 2						
MAKKKKD KV	INKT LSVRII	RPRYSDDIEK	EISDEKAKRK	QDGK TGE LDR	AFFSEL KSRN	60
PDIITND ELP	PLFTEI QK NL	TEI YNKS ISL	LYMKL IVEE	GG STASAL SA	G PYKE C KAR F	120
NSYISL GLR Q	KIQ S N P R R K E	LKG FQ VSL PT	AKS DRF P I P F	CHQ VENG KGG	F KV YET G DDF	180
IFEVPLI KYT	ATN KK STSG K	NYTK VQL NN P	P VP MNV PLL	ST MRR RQ T K K	GM QWN K D E G T	240
NAEL RRV MSG	EYKV SYAE II	R RT RFG K H D D	WFVN FS I K F K	NK TD E LN Q N V	R GG ID I G V S N	300
PLVCA T NGL	D RY I V A N N D I	MA F NER A M A R	R RT L L R K N R F	K RSG H G A K N K	L E P I T V L T E K	360
NEFR RKS IL Q	R WARE VAE FF	KRT S A S V V M M	E D L S G I T E R E	D P F S T K L R T T	W NY R L M Q T T I	420
ENKL KEY GIA	V NY I S P K Y T S	Q T C H S C G K R N	D Y F T F S Y R S E	N N Y P P F E C K E	C N K V K C N A D F	480
NAAK NIA L K V	VL					492
SEQ ID NO: 3	moltype = AA	length = 529				
FEATURE	Location/Qualifiers					
source	1..529					
	mol_type = protein					
	organism = synthetic construct					
SEQUENCE: 3						
MAKNTIT KTL	KL R I V R P Y N S	A E V E K I V A D E	K N N R E K I A L E	K N K D K V K E A C	S K H L K V A A Y C	60
T T Q V E R N A C L	F C K A R K L D D K	F Y Q K L R Q G Q F P	D A V F W Q E I S E	I F R Q L Q K Q A A	E I Y N Q S L I E L	120
Y Y E I F I K G K G	I A N A S S V E H Y	L S D V C Y T R A A	E L P K N A A I A S	G L R S K I K S N F	R L K E L K N M K S	180
G L P T T K S D N F	P I P L V K Q K G G	Q Y T G F E I S N H	N S D F I I K I P F	G R W Q V K K E I D	K Y R P W E K F D F	240
E Q V Q K S P K P I	S L L L S T Q R R K	R N K G W S K D E G	T E A E I K K V M N	G D Y Q T S Y I E V	K R G S K I G E K S	300
A W M L N L S I D V	P K I D K G V D P	I I G G I D V G V K	S P L V C A I N N P	F S R Y S I S D N D	L F H F N K K M F A	360
R R R I L L K K N R	H K R A G H G A K N	K L K P I T I L T E	K S E R F R K K L I I	E R W A C E I A D F	F I K N K V G T V Q	420
M E N L E S M K R K	E D S Y F N I R L R	G F W P Y A E M Q N	K I E F K L K Q Y G	I E I R K V A P N N	T S K T C S K C G H	480
L N N Y F N F E Y R	K K N K F P H F K C	E K C N F K E N A D	Y N A A L N I S N P	K L K S T K E E P		529
SEQ ID NO: 4	moltype = AA	length = 1297				
FEATURE	Location/Qualifiers					
source	1..1297					
	mol_type = protein					
	organism = synthetic construct					
SEQUENCE: 4						
MAT L V S F T K Q	Y Q V Q K T L R F E	L I P Q G K T Q A N	I D A K G F I N D D	L K R D E N Y M K V	K G V I D E L H K N	60
F I E Q T L V N V D	Y D W R S L A T A I	K N Y R K D R S T	N K K N L E K T Q E	A A R K E T I A W F	E G K R G N S A F K	120
N N Q K S F Y G K L	F K K E L F D K I	R S D D L E Y D E E	T Q D A I A C F D K	F T T Y F V G F H E	N R K N M Y S T E A	180
K S T S V A Y R V V	N E N F S K F L S N	C E A F S V L E A V	C P N V L V E A E Q	E L H L H K A F S D	L K L S D V F K V E	240
A Y N K Y L S Q T G	I D Y Y N Q I I G G	I S S A E G V R K I	R G V N E V V N A	I Q Q N D E L K V A	L R N K Q F T M V Q	300
L F K Q I L S D R S	T L S F V S E Q F T	S D Q E V I T V V K	Q F N D D I V N N K	V L A V V K T L F E	N F N S Y D L E K I	360
Y I N S K E L A S V	S N A L L K D W S K	I R N A V L E N K I	I E L G A N P P K T	K I S A V E K V K	N K D F S I A E L A	420
S Y N D K Y L D K E	C G N D K E I C S I A	N V V L E A V G A L	E I M L A E S L P A	D L K T L E N K N K	V K G I L D A Y E N	480
L L H L L N Y F K V	S A V N D V D L A F	Y G A F E K V Y V D	I S G V M P L Y N K	V R N Y A T K K P Y	S V E K F K L N F A	540
M P T L A D G W D K	N K E R D N G S I I	L L K D Q Y Y L G	V M N P Q N K P V I	D N A V C N D A K G	Y Q K M V Y K M F P	600
E I S K M V T K C S	T Q L N A V K A H F	E D N T N D F V L D	T D T K F I S D L T	I T K E I Y D L N N	V L Y D G K K F Q	660
I D Y L R N T G D F	A G Y H K A L E T W	I D F V K E F L S K	Y R S T A I Y D L T	T L L P T N Y Y E K	L D V F Y S D V N N	720
L C Y K I D Y E N I	S V E Q V N E W V E	E G N L Y L F K I Y	N K D F A T G S T G	K P N L H T M Y W N	A V F A E N L H D	780
V V V K L N G G E E	L F Y R P K S N M P	K V E H R V G E K L	V N R K N V N G E P	I A D S V H K E I Y	A Y A N G K I S K S	840
E L S E N A Q E E L	P L A I I K D V K H	N I T K D K R Y L S	D K Y F F H V P I T	L N Y K A N G N P S	A F N T K V Q A F L	900
K N N P D V N I I G	I D R G E R N L Y	V V R E I A D M M V	I D K K Q V S Y N N	V N G Y D Y Y E K L	N Q R E K E R I E A	960
R Q S W G A V G K I	K E L K E G Y L S L	V V R E I A D M M V	K Y N A I V V M E N	L N A G F K R V R G	G I A E K A V Y Q K	1020
F E K M L I D K L N	Y L V F K D V E A K	E A G G V L N A Y Q	L T D K F D S F E K	M G N Q S G F L F Y	V P A A Y T S K I D	1080
P V T G F A N V F S	T K H I T N T E A K	K E F I C S F N S L	R Y D E A K D K F V	L E C D L N K F K I	V A N S H I K N W K	1140
F I I G G K R I V Y	N S K N K T Y M E K	Y P C E D L K A T L	N A S G I D F S S S	E I I N L L K N V P	A N R E Y G K L F D	1200
E T Y W A I M N T L	Q M R N S N A L T G	E D Y I I S A V A D	D N E K V F D S R T	C G A E L P K D A D	A N G A Y H I A L K	1260
G L Y L L Q R I D I	S E E G E K V D L S	I K N K E E W F K V	Q Q K E Y A R			1297
SEQ ID NO: 5	moltype = AA	length = 1145				
FEATURE	Location/Qualifiers					
source	1..1145					
	mol_type = protein					

---

-continued

---

```

SEQUENCE: 5
organism = synthetic construct
MCMKITKIDG ISHKKYKEKG KLIKNNNTAK DIIERFNDI EKKTKELFQK TLDFYVKNYE 60
KCKEQNKERR EKAKNYFSKV KILVDNKKIT ICMENTEKME IEDFNEYDVR SGKYPNVLNK 120
ILNGENYTEE DLEVFDNLQ KRTGRIKSIC NSLEENKAHF KKESINNNII YDRVKGNNKK 180
SLFYEYRRIS SKHQEVNNI FEAFDKLYSN SHEAMNNLFS EITKDKSKDR IRKIREAYHE 240
ILNKNKTEFG EELYKKIQDN RNNFDKLLEI EPEIKELTKS QIFYKYYIDK VNLDSETSikh 300
CFCHLVIEV NQLLKNVYVS KRNINKEKLE NIPEYCKLKN LIKKNLUVNKL NNYIRNCGKY 360
NAYISNNDDV VNSEKISEIR TKEAFLRSII GVSAYFSL RNILNTDNTQ DITNKVDKEV 420
DLKYQENKCI ELEERLKLFV GNYFIDINNO EIKVFLMNID KIISSTIRHET IHFKMETNAQ 480
NIFDFNNVNI CNTAKNIFSN EINEEKPLNSANV FDYLSNNDT EYMDKVVFSF 540
TNRNVSVFVPS PTKIYNRVQD LANSLEIKKW KIPDKSEGKD AQIYLLKNIY YGKFLDEFLN 600
EENGIFISIK DKIIIELNRNQ NKRTGFYKLE KFEKIEETNP KKYLEIIQL YMINIEEDS 660
EGKNLDFI QKIFLKGFFE FIKNNGNNYL ELKKIQDKNN IFDSEMSEYI AGEKTLEDIG 720
EINEIIQDIDK ITEIDKLINQ TDKINCFYLL LKLLNLYKEIT ELKGNLEKYQ ILSKTNVYEK 780
ELMLLNIVNL DNNKVKIENF KILAEIGKF IEKINIEEBIN KNKKIKTFEE LRNFKGENT 840
GEYNNIYSDD KNIKNIRNL NIKKYGMLDL LEKISEKTNY CIKKKDLEY SELRKQLEDE 900
KTMFYKIQEY LHSKYQOKPK KILLKNNKND YEKYKKSIEI IEKYVHLKNU IEFNELNLLQ 960
SLLLKILHRL VGFTSIWERD LRFRFLIGE DELDVEDIFD HRKRYKGTKG GICKKYDRFI 1020
NTTHTEYKNNN KMENVKPFDN NPVRNHYIAHF NYLPNPVKSYI LKMMEKLRLKL LDYDRKLKNA 1080
VMKSIKIDILE EYGFKAEIFII NSDKEIILNL VKSVEIIHLG KEDLKSRRNS EDLCLKVAKM 1140
LEYSK 1145

SEQ ID NO: 6 moltype = AA length = 1271
FEATURE Location/Qualifiers
source 1..1271
mol_type = protein
organism = synthetic construct
SEQUENCE: 6
MEDKQFLERY KEFIGLNSLS KTLRNSLIPV GSTLKHIQEY GILEEDSLRA QKREELKGIM 60
DDYRNYIEM HLRDVHDIDW NELFEALTEV KKNQTDDAKK RLEKIQEKKR KEIYQYLSDD 120
AVPSEMFKEK MISGILPDKFI RCNEGYSEEPE KEEKLKTVL FHRFTSSFND FFLNRKVNFT 180
KEAIVTAIGY RVVHENAEIF LENMVAFQNI QKSAESQISI IERKNEHYFM EWKLISHIFTA 240
DYYMLMLMTQK AIEHYNEMCG VVNQQMREYC QKEKKNNLNY RMKRLHKQIL SNASTSFKIP 300
EKYENDAEVY ESVNSPLQNV MEKTVMERIA VLKNSTDNF DLSKIVITAPY YEKISNYLCG 360
SWNTITDCLT HYEQQIAKG GARKDQVKVA AVKADWKSL SEIEQLLKEY ARAEEVKRKP 420
EEYIAEIEI VSLKEAHLLE YHPEVNLIEN EKYATEIKDV LDNYMELFW MKWFYIEAV 480
EKEVNFYQEL DDLYEEIKDI VPPLYKVRNY VTQKPYSDTK IKLNFGTPTL ANGWSKSKEY 540
DYNAIQLQKD GKYYMGIFPN IOKPEKEIIE GHQSOPLEGNE YKKMVVYYLP SANKMLPKVL 600
LSKKGMEIYQ PSEYIINGYK ERRHICSEEK FDLQFCHDLI DYFKSGIERN SDWKVFGFDF 660
SDTDTYQDIS GFYREVEDQG YKIDWTYIKE ADIDRLNEEG KLYLFQIYNK DFSEKSTGRE 720
NLHTMYLKLNL FSEENVREQV LKLNGEAEIF FRKSSVKKPI IHKKGTMLVN RTYMBEVNGN 780
SVRNRIPEKE YQEIINYKHN RLKGELSTEA KKYLEKAVCH ETKKDIVKDY RYSVDKFFIH 840
LPITINYRAS GKTLENSVAQ RIYAHQNDMH VIGIDRGERN LIYVSVINMQ GEIKEQKSFN 900
IINEFNYKEK LKEREOSRGA ARRNWKEIQC IKDLKEGYLS GVIHETAKMM IKYHAIAME 960
DLNYGFKRGR FKVERQVYQK FENMLIQKLN YLVFKDRPAD EDGGVLRGYQ LAYIPDSVKK 1020
MGRQCGMIFY VPAAFTSKID PTTGFDVDFIK HKVYTTEQRAQ KEFGNGRMRD KEDYNPKDKM CYDVERQLFR 1080
FTPDYANFVT QNVTLARNNW TIYTNGTRAQ KEFGNGRMRD KEDYNPKDKM VELLESEGIE 1140
FKSGKNLLPA LKKVSMNAKVF EELQKIVRPT VQLRNSKSEE NDVVDYDHVIS PVLNNEGNFF 1200
DSSKYKNKEE KKESSLPVDA DANGAYCIAL KGLYIMQAIQ KNWSEEKALS PDVLRNNND 1260
WFDYIQNKRY R 1271

SEQ ID NO: 7 moltype = AA length = 1229
FEATURE Location/Qualifiers
source 1..1229
mol_type = protein
organism = synthetic construct
SEQUENCE: 7
MEKKKMSKIE KFIGKYKISE TLRFRAPVG KTODNIEKKG ILEKDKKRSE DYEVKAYLD 60
SLHRDFKLNELN EYACLFPSGT KDDGDKKKME KLEEKMRKT SNEFCNDEMY 120
KIFSEKILS ENNEEDVSDI VSSYKGGFTS LNGYVNNRKN LYVSDAKPTS IAYRCINENL 180
PKFLRNVECY KKVVQVIPKE QIEYMSNNLN LSPYRIEDCF NIDFFEFCLS QGGIDLYNTF 240
IGGYSKKDGT KVQGINIEVN LYNQKNNKKD EKYKLPQFTP LFKQILSDRD TKSFSIEKLE 300
NIYEVVELVK KSYSLTMEDDD IETVFSNLNY YDASGIYVN GPAITHISM LTWDATIRN 360
NWNYEYRAS STKKNKNIEK YEDTRNTMYK KIDSFTLEYI SRLVGKDIDE LVKYFENEVA 420
NFVMDIJKTY SKLTPLFDRQ QKENFDISIDE EVNDIKGYLD NVKLLESFMK SFTINGKENN 480
IDYVYFGKFT DDYDKLHEFD HIYNKVRNYI TTSRKPYKLD KYKLYFDNPQ LLGGWDINKE 540
KDYRTVMLTK DGKYYFAIID KGEPFDNIP KDYFDNNGGY KKIIYRQIPN AAKYLSSKQI 600
VPQNPPEEVK RILDKKADS KSLTEEEKNI FIDYIKSDFL KNYKLLFDKQ NNPFYFNFAFR 660
ESSTYESLNE FFEDVERQAY SVRYENLPAD YIDNLVNEGK IYLFEIYSKD FSEYSKGTTNN 720
LHMYFKALF DNDNLKNTV RLSGNAELFI RPASIKKDEL VIHPKNQLLQ NKNPLNPKKQ 780
SIFDYDLVKD KRFFENQYML HISIEINKNE RDAKKIKNI EMVRKELKDS DDNYIIGIDR 840
GERNLLYVCV INSAGKIVEQ MSLNEIINEY NGIKHTVDQ GLLDKCEKER NAQRQSWKSI 900
ENIKELKDGY ISQVVKLCLQ LVEKYDAITA MENLNGGFKR GRTKFEKQVY QKFNKLINK 960
MEYMADKKRK TTENGGLRLG YQLTNGCINN SYQNGFIFVY PAWLTSKIDP TTGFV DLLKP 1020
KYTNVEEABL WINKFNSITY DKKLDMFAPN INYSQFPRAD IDYRKIWTY TNGYRIETFR 1080

```

-continued

---

NSEKNNEFDW KEVHLTSVIK KLLEEYQINY ISGKNIIDDL IQIKDKPFWN SFIKYIRLTL	1140
QMRSNITGRT DVVDIISPVI NNEGTFYDSR KDLDEITLPQ DADANGAYNI ARKALWIEK	1200
LKESPDEELN KVKLAITQRE WLEYAQINI	1229

SEQ ID NO: 8                  moltype = AA length = 835  
 FEATURE                  Location/Qualifiers  
 source                  1..835  
 mol\_type = protein  
 organism = synthetic construct

SEQUENCE: 8  
 MEKIKKPSNR NSIPSIIISD YDANKIKEIK VKYLKLARLD KITIQDMEIV DNIVEFKIL 60  
 LNGVEHTIID NQKIEPDNYE ITGCKPKSNK RRDGRISQAK YVVTITDKYL RENEKEKRFK 120  
 STERELPNNT LLSRYKQISG FDTLTSKDIY KIKRYIDFKN EMLFYFQFIE EFFNPPLPKG 180  
 KNFYDNLNIEQ NDKDVAKFVY YRLNNDFKNK SLNSYITDTC MIINDFKKIQ KILSDFRHAL 240  
 AHFDFDFDIQK FDINTISLIE TLLDQKEEKN YQEKNNYIDD NDILTIFDEK 300  
 GSFKFSKLHNF YTAKISQKPA FNKLINSLFLS QDGVPNEEFK SYLVTKKLDF FEDIHSNKEY 360  
 KKIYIQQHKNL VIKKQKEESQ EKPDGQKLKN YNDELQKLKD EMNTITKQNS LNRLEVKLRL 420  
 AFGFIANEYN YNFKNPDEF TNDVKNEQK1 KAFKNSSNEK LKEYFESTFI EKRFFFHFSVN 480  
 FFVNKKTKEEKF TQKKNFNSI ENETLEELVK ESPPLLQIITQ EKRFIPRELQ GEFVGFFILKI 540  
 YHHHTKNTKSD TKEDEISIED AQNSFSLKPK ILAKNLRLQL LFHYHSLSHNT LYNNKQCFYY 600  
 EKGNRWQSVY KSFQISHNQD EFDIHLVIP IKYYINLNKL MGDFEIYALL KYADKNSITV 660  
 KLSDITSRDD LKYNGHYNFA TLLFKTFGID TNYKQNKVSI QNIKKTRNNL AHQNIENMLK 720  
 AFENSEIFQAQ REEIVNVLQT EHRMQUEVLHY NPINDFTMKT VQYLSLHSV SQKEGKIAIDI 780  
 HKKESLVPND YYLIYKLKAI ELLKQKVIEV IGESEDEKKI KNAIAKEEQI KKGNN 835

SEQ ID NO: 9                  moltype = AA length = 1264  
 FEATURE                  Location/Qualifiers  
 source                  1..1264  
 mol\_type = protein  
 organism = synthetic construct

SEQUENCE: 9  
 MENYGGFTGL YPLQKTLKFE LRPQGRTMEH LVSSNFFEED RDRAEKYKIV KKVIDNYHHD 60  
 FINECLSKRS PDWTPLMKTS EKYYASKEKN GKKQDLDQK IIPTIENLSE KDRKELELEQ 120  
 KMRKREIVSV FKEDKRFKYL FSEKLFSSL KDEDYSKEKL TEKEILALKS FNKFSGYFIG 180  
 LHNKRANFYS EGDESTAIAY RIVNENFPKF LSNLKKYREV CEKYPEIIQD AEQSLAGLNI 240  
 KMDDIFPMEN FNKVMTQDGI DLYNLAIGGK AQAQGEKQKG LNEFLNEVNQ SYKKGNDRIR 300  
 MTPFLFKQILS ERTSYSYILD AFDDNSQLIT SINGFFTEVE KDKEGNTFDR AVGLIASYMK 360  
 YDLSRVYIRK ADLNKVSMEI FGWSWERLGGL LRIFKSELYG DVNAEKTSSK VDKWLNSGEF 420  
 SLSDVINAIA GSKSAAETKSD YILKMRVARG EIDNALEKIK CINGNPSEDD NSKMIKAIL 480  
 DSVQRLFHLF SSFQVRADFS QDGDFYAEYN EIYEKLFIAV PLYNVRNYL TKNNLMSKKI 540  
 KLNFKNPALPA NGWDLNKEYD NTAVIFLREG KYLGIMNPS KKKNIKFEEG SGTGPFYKKM 600  
 AYKLLPDPNK MLPKVFAKK NINYYNPSDE IVKGYKAGKY KKGENFDIDF CHKLIDFFKE 660  
 SJQKNEWDRA FNYLFSATES YKDQDFYSE VEQGQYRMY LNVPVANIDE YVEKGDLFLF 720  
 QIYNKDFASG AKGNKDMHTI YWNAAFSDEN LRNVVVKLNG EAELFYRDKS IIEPICHKKG 780  
 EMLVNRTCFD KTPVPDKIH ELFDYHNRA KTLSIEAKGY LDRVGVFQAS YEIIKDRRYS 840  
 ENKMYFHVPL KLNFKADGKK NLNMVIEKF LSDKDVKHIG IDRGERNLLY YSVIDRRGNI 900  
 IDQDSLNIID GFDFYQKLLQG REIERREARQ SWNSIGKIKL LKEGYLSKAV HKVSKMVLEY 960  
 NAIVVLEDLN FGPKRGRFKV EKQVYQKFEK MLIDKLNLYLV FKEVLDSDRA GGVLNAYQLT 1020  
 TQLESFNKLQ KGSGILFYVP AAYTSKIDPT TGTVSLFNITS RIESDSEKKD FLSGPDFSIVY 1080  
 SAKDGGIFAF KFDYRNRFQ REKTDHKNIW TVYTNGDIRK YKGRMKGYEI TSPTKRIKDV 1140  
 LSSSGIRYDD QOELRDSIIQ SGNKVLINEV YNSFIDTLQM RNSDGEQDYI ISPVKNRNGE 1200  
 FFRTDPDRRE LPVDADANGA YHIALRGELL MQKIAEDFPD KSDKFTMPKM EHKDWFEFMQ 1260  
 TRGD 1264

SEQ ID NO: 10                  moltype = AA length = 500  
 FEATURE                  Location/Qualifiers  
 source                  1..500  
 mol\_type = protein  
 organism = synthetic construct

SEQUENCE: 10  
 MEVQKTVMKT LSLRILRPLY SQEIEKEIKE EKERRKQAGG TGELDGGFYK KLEKKHSEMF 60  
 SFDRLLNLLN QLQREIAKVV NHAISELYIA TIAQGNKSNK HYISSIVNR AYGFYVNAYI 120  
 ALGICSKVEA NFRSNELLTO QSALETTAKSD NFPIVLHKQK GAEGEDGGRF ISTEGSDLIF 180  
 EIPIPPEYEN GENRKEPYKV VKKGQKPV LKLLSTFRRQ RNKGWAKDEG TDAEIRKVT 240  
 GKYQVSQIEI NRGKKLGEHQ KWFANFSIEQ PIYERKPNNRS IVGGGLDVGIR SPLVCAINNS 300  
 FSRYSVDSND VFKFSKQVFA FRRRLLSKNS LKRGKGHAAH KLEPITEMTE KNDKFRKKII 360  
 ERWAKEVTNF FVKNQVQIVQ IEDLSTMKDR EDHFFNQYLR GFWPYQQMOT LIENKLKEYG 420  
 IEVKRVQAKY TSQLCSNPNC RYWNYYFNFE YRKVNKFPKF KCEKCNLEIS ADYNAARNL 480  
 TP DIEKEFVAK ATKGINLPEK 500

SEQ ID NO: 11                  moltype = AA length = 1261  
 FEATURE                  Location/Qualifiers  
 source                  1..1261  
 mol\_type = protein  
 organism = synthetic construct

SEQUENCE: 11

-continued

---

MIIHNCYIGG	SFMKKIDSFT	NCYSLSKTLR	FKLPIGATQ	SNFDLNKMLD	EDKKRAENYS	60
AKAISIIDYKH	RFFIDKVLS	VTEKNAFDSF	LEDVRAYAEL	YYRSNKKDSD	KASMTKLESK	120
MRKFIALALQ	SDEGFKDMLFG	QNLIKKTLP	FLESDTDKEI	IAEFDGFSTY	FTGFFNNRKN	180
MYSAADDQPTA	ISYRCINDNL	PKFLDNVRTF	KNSDVASILN	DNLKILNEDF	DGIYGTSaed	240
VFNVDYFPFV	LSQKGIEAYN	SILGGYTNSD	GSKIKGLNEY	INLYNQKNEN	IHRIPMKQL	300
FIQIILSERES	VSFPIPEKFDS	DDDVLSSIND	YYLERDGCKV	LSIEKTVEKI	EKLFSAVTDY	360
STDGIFVKNA	AELTAVCSGA	FGYWGTQVNA	WNNEYDALNG	YKETEKYIDK	RKKAYKSIES	420
FSLADIOQKYA	DVSESETNA	EVTIEWLRNEI	KEKCNLAVOG	PYTESKDLISK	EVTSKLLFN	480
NDNAVELIKN	ALDSVKELEN	VLRLLLGTTGK	EESKDENFYG	EFLPCYERIC	EVDSLWYDKVR	540
NYMTQKLYKT	DKIKLNFQNP	QFLGGWDRNK	EADYSAVLLR	RNSLYYIAIM	PSGYKRVFEK	600
IPAPKADETV	YEKVIYKLLP	GPNKMLPKVF	FSKKGIEITN	PPKEILEKYE	LGHHTGDGF	660
NLDDCHALID	YFKSALDVHS	DWSNFGFRFS	DTSTYKNIAD	FYNEVKNQGQ	KITFCDVPOQ	720
YINELVDEKG	LYLFQLYNKD	FSEHSKGTPN	LHTLYFKMLF	DERNLENVVF	KLNGEAEMFY	780
REASISKDDM	IVHPKNQPIK	NKNEONSRSQ	STFELYDIVKD	RRYTVDQFML	HIPITLNFTA	840
NGGTTINNEV	RALKLKDCKN	YVIGIDTRVENI	NLLYICVVDS	EGRRIEQYSL	NEIINEYNGN	900
TYSTDYHALL	DKKEKERLES	RKAWKTVENI	KELKEGYISO	VVHKICELVE	KYDAVIMMED	960
LNLGFKQGRS	GKFEKSVYQK	FEKMLIDKLN	YFADKKKSPE	EIGSVLNAYQ	LTNAFESFEK	1020
MGKQNGFIFY	VPAVYLTSKID	PTTGFAIDLH	PSSKQSKESM	RDFVGRFDI	TFNKTENYFE	1080
FELDYNKFPR	CNTDYRKWWT	VCTYGSRIK	FRNPEKNSEW	DNKTVELTPA	FMALFEKYSI	1140
DVNGDIKAQI	MSVDDKKDFV	ELIGLLRLTL	QMRNSETGKV	DRDYLISPVK	NSEGVFYNSD	1200
DYKGRIENASL	PKDADANGAY	NIARKGLWII	EQIKACENDA	ELNKIRLAIS	NAEWLEYAQK	1260
K						1261

---

SEQ ID NO: 12	moltype = AA length = 478
FEATURE	Location/Qualifiers
source	1..478
	mol_type = protein
	organism = synthetic construct

SEQUENCE: 12						
MDYDYIRKTL	LRLRPPYYGE	EIEKEIAAAK	KKSQAEGGDG	ALDNKFWDR	KAHEPEIIS	60
REFYDLLDAI	QRETTLYYNR	AISKLYHSLI	VEREQVSTAK	ALSAGPYHEF	REKFNAYISL	120
GLREKIQSNF	RRKELARYQV	ALPTAKSDTF	PIPIYKGFDK	NGKGGFKVRE	IENGDFVIDL	180
PLMAYHRVGG	KAGREYIELD	RPPAVLNVP	ILSTSRRAN	KTWFRDEGT	AEIRRVMAGE	240
YKVSWEILQ	RKRFGKPYGG	WVYNFTIKYQ	PRDYGLDPK	KGGIDIGLSS	PLVCATVNSL	300
ARLTIRDNDL	VAFNRKAMAR	RRTLLRQNR	KRGSGHSANK	LKPIEALTEK	NELYRKAIMR	360
RWAREAADFF	RQHRAATVN	EDLTGKDR	DYFSQMLRCY	WNYSQLQML	ENKLKEYGIA	420
VKVIIEPKDTS	KTCHSCGHVN	EYFDENYRSA	HKPFPMFKCEK	CGVECGADYN	AARNIAQQA	478

SEQ ID NO: 13	moltype = AA length = 1229
FEATURE	Location/Qualifiers
source	1..1229
	mol_type = protein
	organism = synthetic construct

SEQUENCE: 13							
MKEQFINRYP	LSKTLRFSLI	PVGETENNFN	KNLLLKKDKQ	RAENYEKVKC	YIDRPHKEYI	60	
ESVLSKARIE	KVNEYANLYW	KSNKDDSDIK	AMESLENDMR	KQISKQLTST	EIYKKRLFGK	120	
ELICEDLPSF	LTDKDERETV	ECFPLSFTTYF	KGFNTNRENN	YSSDGKSTAI	AYRCINDNLP	180	
RFLDNVKSFQ	KVFDNLSDET	ITKLNTDLYN	IFGRNIEDIF	SVDYFEPVLT	QSGIEIYNSM	240	
IGGYTCSDKT	KIQGLNECIN	LYNQQVAKNE	KSKKLPLMKP	LYKQILSEKD	SVSFIEPEKFN	300	
SDNEVLHAID	DYYTGHIGDF	DLLTELLQSL	NTYNANGIVF	KSGVAITDIS	NGAFNSWNVL	360	
RSAWNEKYEA	LHPVTSKTKI	DKYIEKQDKI	YKAIKSFSLF	ELQSLGNENG	NEITDWYISS	420	
INESNSNKE	AYLQAQKLLN	SDYEKSYNK	LYKNEKATEL	VKNLLDAAIKE	FQKLKPLMNG	480	
TGKEENDEL	FGYKFTSYDD	SIADIIDR	KVNRNITYTOPK	YSKDKIKLNL	DNPQLLGWD	540	
KNKESDYRTV	LLHKDGLYYL	AVMDKSHSKA	FVDAPEITS	DKDYYEKMEY	KLLPGPNKML	600	
PKVFFASKNI	DTFQPSDRIL	DIRKRESFKK	GATFNMKA	EFIDYFKDSI	KKHDDWSQFG	660	
FKFSPTESYN	DISEFYREIS	DQGYSVPTEN	ISKNYI	DNNGYIYLFQI	YNKDFSKYSK	720	
GTPNLHLYF	KMLFDERNLS	NVYVQLNGEA	EMFYREASIG	DKEKITHYAN	QPIKKNPNDN	780	
EKKESVFEYD	IVKDKRFTKR	QFSLHLPTI	NFKAHGQEFL	NYDVRKAVKY	KDDNYVIGID	840	
RGERRNLYIS	VINSNGEIVE	QMSLNEII	NGHKV	YDQKL	LDTKEKERDK	ARKNWTSVEN	900
IKELKEGYI	AKELKEGYI	IKYDIAVIME	DLMNGFKRGR	FPVEKQVYQK	FENMLISKLN	960	
LILDKKAEP	EDGGLLRAYQ	LTNPKDGVNK	AKONGII	YDYSKFP	PCDADANGAY	1020	
KCNTSVPEAK	KLFETIDDIK	YNANTDMFEF	YIDYSKFP	NSDFKSWTV	CTNSSRILTF	1080	
RNKEKNKKWD	NKQIVLTDDEF	KSLFNEF	YKGNLKD	SISNADFYRR	LIKLLSLTQ	1140	
MRNSITGSTL	PEDDYLISPV	ANKSGEFYDS	RNYKG	TAAL	PCDADANGAY	1200	
NVLKDTPDDM	LNKAKLSSITN	AEWLEYTQK				1229	

SEQ ID NO: 14	moltype = AA length = 1242
FEATURE	Location/Qualifiers
source	1..1242
	mol_type = protein
	organism = synthetic construct

SEQUENCE: 14						
MKBQFVNQYP	ISKTLRFSLI	PIGKTEENFN	KNLLLKEDEK	KAEEYQKVKG	YIDRYHKFFI	60
ETALCNINFE	GFEEYSLYY	KCSKDDNDLK	TMEDIEIKLR	KQISKMTSH	KLYKDLFGEN	120
MIKITILPNFL	DSDEEKN	MFRGFYTYFS	GFNTNRKNM	TEEAKSTSIA	YRCINDNLPK	180
FLDNSKSFEK	IKCALNKEEL	KAKNEEFYEI	FQIYATDIFN	IDFFNFVLTQ	PGIDKYNGII	240

-continued

---

GGYTCSDGTK	VQGLNEIIINL	YNQQIAKDDK	SKRLPLLKML	YKQILSDET	VSFIPEKFSS	300
DNEVLESINN	YFSKNVSNAL	KSLKELFQGF	EAYNMNGIFI	SSGVAITDLS	NAVFGDWNAI	360
STAWEKAYFE	TNPPKKNKSQ	EKYEEELKAN	YKKIKSFSLD	EIQRLGSIAK	SPDSIGSVAE	420
YYKITVTEKI	DNITELYDGS	KELLNCNYSE	SYDKKLKIKND	TVIEKVTKLL	DAVKSLEKLI	480
KPLVGTGKED	KDELFYGTFL	PLYTSLSAVD	RLYDKVNRNYA	TQKPYSKDKI	KLFNFNCSSFL	540
SGWATDYSN	GGLIFEKDGL	YYLGINVNKKF	TTEEIDYLQO	NADENPAQR	VYDFQKPDNK	600
NTPLRFIRSK	GTNYSPSVKE	YNLPVEEIVE	LYDKRYFTTE	YRNKNPELYK	ASLVKLIDYF	660
KLGFTTRHESY	RHYDFWKWKS	EYENDISEFY	KDVEISCYSL	KQEKENYNTL	LNFVAENRIY	720
LFQIYINKDFS	KYSKGTPNLH	TRYFKALFDE	NNLSDVVFKL	NGGSEMFFRK	ASIKDNEKVV	780
H PANQPIDNK	NPDNNSKKQST	FDELKLKDR	FTKHQFSIHI	PITMNFKARG	RDFINNDIRK	840
AIKSEYKPYV	IGIDRGERNL	IYISVINVNG	EIVEQMSLND	IISDNGYKVD	YQRLLDRKEK	900
ERDNARKSWG	TNIENIKELKE	GYIISOVIHKI	CELVIKYDAV	IAMEDLNFGI	KRGRPNVEKQ	960
VYQKFENMLI	SKLNYLCDKK	SEANSEGGLL	KAYQLTNKFD	GVNKGKQNGI	IFYVPawlTS	1020
KIDPVTGFVD	LLHPKYISVE	ETHSLFEKLD	DIRYNFEKDM	FEFIDYISKL	PKCNADFQKQ	1080
WTVCCTNADRI	MTRFRSEKNN	EWDNKRILL	DEFKRLFEEF	GIDYCHNLKN	SISIISNKDF	1140
CYRFIFLFL	TMQMRNSITG	STNPEDDYLI	SPVRDENGVF	YDSRNRFIGSK	AGLPIDADAN	1200
GAYNIARKGL	WAIAKSTA	DDMLDKVDSL	ISNAKWLEYV	QK		1242

---

SEQ ID NO: 15	moltype = AA length = 1168
FEATURE	Location/Qualifiers
source	1..1168
	mol_type = protein
	organism = synthetic construct

SEQUENCE: 15						
MKITKIDGIL	HKKYIKEGKL	VKSTSEENKT	DERLSELLTI	RLDTYIKNPD	NASEEENRIR	60
RETLKEFFSN	KVLYLKDSIL	YLDKDRREKNQ	LQNKNYSEBD	ISEYDLKNKN	SFLVLKKILL	120
NEDINSEELE	I FRNDFEKKL	DKINSLKYSL	EENKANYQKI	NENNKKVEG	KSKRNIFYNN	180
YKDSAKRNDY	INNIQBAFDK	LYKKEDIENL	FFLIELNSKKH	EKYKIRECYH	KIIGRKNDKE	240
NFATIYIEEI	QNVNNMKELI	EKVPNVSELK	KSQVFYKYYL	NKEKLNDENI	KYVPCHFVEI	300
EMSKLLKNVV	YKKPSNISL	VKVRIFEYQS	LKKLIENKLL	NKLDTYVRNC	GKYSFYLQDG	360
EIATSDFIGV	NRQNEAFLRN	IIGVSVSTAYF	SLRNILETEN	ENDITGRIKG	KTVKNKGEE	420
KVYSGEIDKL	YDNNKQNEVK	KNLKMFSYD	FNMRNKKIE	DPFSNIDEAI	SSIRHGIVHF	480
NLELEGKDF	TFKNIVPSQI	SKKMFQNEIN	EKKLKLKIFR	QLNSANVFRY	LEKYKILNYL	540
NRTTRFEFVNK	NIPFVPSFTK	LYSRIDDLKN	SLCIYWKPK	ANDNNKTKKEI	TDAOQYLLKN	600
IYYGEFLNYF	MSNNNGNFFE	IKEEIELNK	DKRNLKTCGY	KLQKFENLQE	KTPKEYLANI	660
QSFSYIMADGN	KDEEEKDAYI	DFIQKIFLKG	FMTYLANGR	LSLMYIGNDE	QINTSLAGKK	720
QEFDKFLKKY	BQNNNIEIPH	EINEFVREIK	LGKILKYTES	LNMFYLILKL	LNHKELTNLK	780
GSLEKYQSAN	KEEAFSQDLE	LINLLNLDNN	RVTEDFELEA	DEIGKFLDFN	GNVKVDNKL	840
KKFDTNKIYF	DGENIIKHRA	FYNKKYGL	NLLEKISDEA	KYKISIEELK	NYSNKKIEIE	900
KNHTTQENLH	RKYARPRKDE	KFNDEDYKKY	EKTIRNIQYQ	THLKNKVEFN	ELNLLQSSL	960
RILHRLVGYT	SIWERDLRFR	LKGEEFPENQY	IEEIFNFDNS	KNVKYKNGQI	VEKYISFYKE	1020
LYKDDMEKIS	IYSDKKVKEL	KKEKKDLYIR	NYIAHFNYIP	NAEVSLLEV	ENLRKLLSYD	1080
RKLKNAIMKS	IVDILKEYGF	VVTFKIEKDK	KIRIESLSE	EVVHLKLLK	KDNDKKKEPI	1140
KTYRNSKELC	KLVKVMFEYK	MKEKKSEN				1168

SEQ ID NO: 16	moltype = AA length = 1168
FEATURE	Location/Qualifiers
source	1..1168
	mol_type = protein
	organism = synthetic construct

SEQUENCE: 16						
MKITKIDGIS	HKKYIKEGKL	VKSTSEENKT	DERLSELLTI	RLDTYIKNPD	NASEEENRIR	60
RENLKEFFSN	KVLYLKDGIL	YLDKDRREKNQ	LQNKNYSEBD	ISEYDLKNKN	SFLVLKKILL	120
NEDINSEELE	I FRKDVEAKL	NKINSLKYSF	EENKANYQKI	NENNVEKVG	KSKRNIIYDY	180
YRESAKRNDY	INNVQBAFDL	LYKKEDIEKL	FFLIELNSKKH	EKYKIRECYH	KIIGRKNDKE	240
NFAKIIYEEI	QNVNNMKELI	EKVPDMSELK	KSQVFYKYYL	DKEELNDKNI	KYAFCHFVEI	300
EMSQLLKNVV	YKRLSNISL	VKRIKFEYQN	LKKLIENKLL	NKLDTYVRNC	GKYNLYLQDG	360
EIATSDFIG	NRQNEAFLRN	IIGVSVAYF	SLRNILETEN	KDDITGKMRG	KTRIDSKTGE	420
EKYIPGEVDQ	IYYENKQNEV	KNLKMFMGY	DFMDMNKEI	EDFFANIDEA	ISSIRHGIVH	480
FNLDLDDGDI	FAFKNIVPSE	ISKLLKMFQNEI	NEKKLKLKIF	RQLNSANVFR	YLEKYKILNY	540
LKTRTRFEFV	KNIPFVPSFT	KLYSRIDLK	NSLGIYWKTP	KTNDNNKTKKE	IIDAOIYLLK	600
NIYYGEFLNY	FMSNNGNFFE	ISREIIELNK	NDKRLNLTGF	YKLQKFEDIQ	EKTPKKYLAN	660
IQSLSYMINAG	NQDEEEKDTY	IDFIQKIFLKG	GMFTYLANNG	RLSLMLYIGND	EQINTSLAGK	720
KQEFDFKFLKK	YEQNNNIEIP	HEINEFLREI	KLGKILKYTE	SLNMFYLILK	LLNHKELTNL	780
KGSLEKYQSA	NKEETFSDEL	ELINLLNLDN	NRVTEDFELE	ANEIGKFLDF	NGNKKIKDRKE	840
LKKFDTCKKI	FDGENIICKH	AFYNIKKYGM	LNLLEKIAADK	AKYKISLKE	KEYSNKKNEI	900
EKNYTMQQNL	HRKYARPKKD	EKFNDDEDYKE	YEKAIGNIQ	YTHLKNKVEF	NELNLLQGL	960
LKLHRLVGY	TSIWERDLRF	RLKGEFPENQY	YIEEIFNFDNS	SKNVKYKSGQ	IVEKYINFYK	1020
EIYKDNVEKR	SIYSDKKVKK	LKQEKKDLYI	RNYIAHFNYI	PHAEISLLEV	LENLRKLLSY	1080
DRKLKNAIMK	SVVDILKEYG	FVATFKIGAD	KKIGIQTLES	EKIVHLKNLK	KKKLMTDRNS	1140
EELCKLVKVM	FEYKMEEKNL	TTKKCKV				1168

SEQ ID NO: 17	moltype = AA length = 1224
FEATURE	Location/Qualifiers
source	1..1224
	mol_type = protein

---

-continued

---

SEQUENCE: 17		organism = synthetic construct
MKKIDNFVGC	YPVSKTLRFK	AIPIGKTQEN IEKKRLVEED EVRAKDYKAV KKLIDRYHRE 60
FIEGVLDNVK	LDGLEEYYML	FNKSREESD NKKIEIMEER FRRVISKSFK NNEEYKKIFS 120
KKIIIEILPN	YIKDEEEKEL	VKGFKGFYTA FVGYAQNREN MYSDEKKSTA ISYRIVNEM 180
PRFITNIKVF	EKAKSILDV	KINEINEYL NNDYYVDDDF NIDFFNYVLM QKGIDIYNAI 240
IGGIVTGDGR	KIQGLNECIN	LYNQENKKIR LPQFKPLYKQ ILSESESMSF YIDEIESDDM 300
LIDMLKESLQ	IDSTINNAID	DLKVLFNNIF DYDLSGIFIN NGLPITTISN DVYGOWTIS 360
DGWNERYDVL	SNAKDESEK	YFEKRRKEYK KVKSFSISDL QELGGKDSL CKKINEIISE 420
MIDDYKSKIE	EIQYFLDFIK	LEKPLVTDLN KIELIKNSLD GLKRIERYVI PFLTGKEQN 480
RDEVFYGYFI	KCIDAIEKD	GVYNKPLDFK TKKPYSKDFK KLYFENPQLM GGWRNKESED 540
YRSTLLRKNG	KYYVAIDKS	SSNCMMNIEE DENDNYEKIN YKLLPGPNKM LPKVFSKKN 600
REYFAPSKEI	ERIYSTGTFK	KDTNFVKKDC ENLITFYKDS LDRHEDWSKS FDGSFKESSA 660
YRDISEFYRD	VEKQGYZRV	DLLSSNAVLT LVEEGKLYLF QLYNKDFSEV SHGPNLHTM 720
YFRSLFDDNN	GKNIRLNGGA	EMFMRASLNL KQDVTVHKAN QPIKKNLNL PKTTTLPYD 780
VYKDKRFTED	QEYHVIPITM	NHVMVREQLVK DDNPVYIGID RGERNLIYVV 840
VVDGQGHIVE	QLSLNEIINE	NNGISIRTDY HTLDAKERE RDESRKQWQ IENIKELKEG 900
YISQVVKHIC	ELVEKYDAV	ALEDLNSGFK NSRVRVKEQV YQKFEMKLIT KLNYMVDKKK 960
DYNKPQGVNF	GYQLTTQFES	FSKMGMTQNGI MFYIPAWLTS KMDPTGFDV LLKPKYKNA 1020
DAQKFFSQFD	SIRYDNQEDA	FVFKVNYTKE PRTDADYNE WEIYTNGERI RVFRNPKN 1080
EYDYETVNVS	ERMKELFDSY	DLLYDKGEKL ETICEMEESK FFEELIKLFR LTLQMRNSIS 1140
GRTDVDYLIS	PVKNSNGYFY	NSNDYKKEGA KYPKDADANG AYNIARKVWL AIEQPKMADE 1200
DKLDKTAKI	KNQEWLEYAQ	THCE 1224
SEQ ID NO: 18		moltype = AA length = 1225
FEATURE		Location/Qualifiers
source	1..1225	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 18		
MKKIDSFVNY	YPLSKTLRFS	LIPVGKTEDN FNAKLLLEED EKRAIEYEKV KRYIDRYHKH 60
FJETVLANFH	LDDVNEYAEL	YYKAGKDDKD LKYMELLEGK MRKSISAFT KDKKYKEIFG 120
QEIIKNILPE	FLENEDEKES	VKMFQGFPY FTGFNDNRKN MYTHEAQFTA ISYRCINENL 180
PKFLDNQSF	AKIKESISSD	IMNKLDEVCM DLYGVYAOQM FCTDYFSFVL SQSGIDRYNN 240
IIGGYDWDKG	VKIQGQNEYI	NLYNQOVDFK NKRLPLMKKL YKQILIEKES ISFPIKEFES 300
DNIVINAISD	YHHNNVNEILF	DDFNKLFNEF SEYDDNGIFV TSGLAVTDIS NAVFGSWNII 360
SDSWNEEYKD	SHPMKTTNA	EKYYEDMKKE YKKNLSFTIA ELQRGLGEAGC NDECKGDIKE 420
YKTTTVAEKI	ENIKNAYEIS	KDLLASDYEK SNDKKLCKND SAISLLKNLL DSIKDLEKTI 480
KPLLGTGKEE	NKDDDVYFGK	TNLVEMISEI DRLYDKVNRN VTQKPYSKDK IKLNPENPQH 540
LGGWDKNNKER	DYRSVLLKKF	DQYKYLAMDK SNNKAFIDFP DDGECYKIE YKLLPGPNKM 600
LPKVFFAASN	IEYFAPSKKI	LEIRSRESFK KGDMFNLKDC HEFIDFFKES IKKHEDWSQF 660
GFFESPTEKY	NDISEFYNEV	KIQGYSLKYK NVSKKYIDEL IECGQLYLFQ IYNKDFSVYA 720
KGNPNLHTMY	FKMLFDERNL	ANVYVQLNGG AEMFYRKASI KDSEKIVHHA NQPIKKNAD 780
NVKKESVSEF	NKDDKFRFTK	RQFSIHIPIT LNFKAQGQNF INNDVRMALK KADENYVIGI 840
DRGERNLILYI	CVINSKGEIV	EQKSLNNEIIG DNGYRVDYHK LLDKKEAERD EARKSWGTTIE 900
NIKELKEGYL	SQIVHEISKL	VIKYDAVIAI EDLNSGFKKG RPKVEKQVYQ KFENMLCTKL 960
NYLVDKNADA	NECGGLLKAY	QLTNKEDGAN RGRQNGIIFS VPAWLTSKID PVTGPADLLR 1020
PKYKSVSSESV	EPISKIDNIR	YNSKEDYFEP DIDYSKFPNS TASYKKWTV CTYGERIINV 1080
RNKEKNMMWD	NKTIIVLTDEF	KKLFDADFGVD VSKNIFKESV AIDSKDFYR FINLLANTLQ 1140
LRNSEVGNVD	VDYLISPVKG	VDGSFYDSRL VKEKTLPENA DANGAYNIAR KALWAIDVLIK 1200
QTKEELKN	NLSIKNAEWL	EYVQK 1225
SEQ ID NO: 19		moltype = AA length = 1165
FEATURE		Location/Qualifiers
source	1..1165	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 19		
MKNQNTLPSN	PTDILDKDPF	WAAFFNLRH NVYLTVNHNIN KLLDLKLYN KDKHKEIFEH 60
EDIFNISDDV	MNDVNSNGKK	RKLDKIDVA NLDTDLTRK YLQRELILKHF PFIQPAIIGA 120
QTKERTTIDK	DKRSTSTSND	SLKPTGEGDI NDPLSLSNVK SIFFFFQLML EQLRNYYSHV 180
KHSKSATMPN	FDEGLLKSMSY	NIFIDSVNKV KEDYSSNSVI DPNTSFSHLI SKDEQGEIKP 240
CRYSFSTSKG	SINASGLLFF	VSLFLEKQDS IWMQKKP1GF KKTSENYMKM TNEVPCRNHI 300
LIPKMRLETW	YDKDWMLLDM	LNEVVRCPLES LYKRLAPADQ NKFVKEPEKSS DNANRQEDDN 360
PFSRILVRHQ	NRFPYFLARF	FDLNEVFTTL RFQINLGCYH FAICKVQKIGD KKEVHHLTRT 420
LYGFSRLQNF	TQNTTRPEEWN	TLVKTTEPSS GNDGKTVQGV PLPYISYTIP HYQIENEKIG 480
IKIFDGDGTA	DTDIWPSVST	EKQLNPKDY TLTPGFKADV FLSVHELLPM MFYYQLLCE 540
GMLKTDAGNA	VEKVLIDTRN	AIFNLYDAFV QEKINTITDL ENYLQDKPIL IGHLPKQMD 600
LILGHQRDML	KAVEQKAMU	IKDTERRLER LNKQPEQKPN VAAKNTGTL RNGQIADWLV 660
KDMMRFQPVK	RDKEGNPINC	SKANSTEYQM LQRAFAFYTT DSYRLPRYFE QLHLINCDNS 720
HFLLSRFYED	KQPNLIAFYA	AYLEAKLEFL NELQPQNWAS DNYFLLLRAP KNDRKQLAEG 780
WKNGFNLPLRG	LTTEKIKTWF	NEHKTIVDIS DCDIFKNRVG QVARLIPVFF DKKFKDHSPQ 840
FYTYTNFNGVN	VSKITEANLY	SKEKRENLFK SYQNKFKNNI PAEKTKEYRE YKNFSSWWKF 900
EREELRLIKNQ	DILTWLCKN	LFDEKIKPKK DILEPRIAVS YIKLDSLQTN TSTAGSLNAL 960
AKVVPMTLAI	HIDSPKPKGK	AGNNEKENKE FTVYIKEEGT KLLKGWNFKT LLADRRIKGL 1020
FSYIEHDDIN	LEKYPLTKYQ	VDSELDLYQK YRIDIFKQTL DLEAQQLDKY SDLNTDFNQ 1080

-continued

---

MLSGWSEKEG IPRNIKQDVA FLIGVRNGFS HNQYPDSKRI AFSRIKKFNP KTSSLQESEG	1140
LNIAKQMYEE AQQVNVNIKRN IESFD	1165
SEQ ID NO: 20	moltype = AA length = 1194
FEATURE	Location/Qualifiers
source	1..1194
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 20	
MKVTKIDGIS HKKFEDEGKL VKFTGHFNK NEMKERLEKL KELKLSNYIK NPENVKNDK	60
NKEKETKSRR ENLKKYFSEI IIRKKEEYKL LKKTRKFKNI TEEINYDDIK KRENQKIFD	120
VLKELLEQRI NENDKBEILN FDSDKLKEAF EEDFIKKBLK IKAIEESLEY NRADYRKDV	180
ELENEKYEDV KGQNKRSLVF EYYKNPENRE KFKENIKYAF ENLYTEENIK NLYSEIKEIF	240
EKVHLKSKVR YYFQNBIIGE SEFSEKDEEG ISILYKOII SVEKEKFIE FLQKVKIKDL	300
TRSQIFYKFL LENELENNDEN IKYVFSYFVE IEVNKLNN VYKTKKFNEG NKYRVKNIFN	360
YDKLKNLVVY KLENKLNNYU RNCGKYNYHM ENGDIATSDI NMKNRQTEAF LRSILGVSSF	420
GYFSLRNLG VNDDDFYKIE KDERKNENFI LKKAKEDFTS KNIFEKVVDK SFEKKGIYQI	480
KENLKMFGN SFDKVVDKDEL KKFFVNMLEA ITSFRHRIVH YNINTNSENI FDFSNIEVSK	540
LLKNIFEKEI DTRELKLKIF RQLNSKLFV DSQIYLLKNI YYGEFVEKFV NDNKNFEKIV	600
KLYDRIDNLN GWNALKLGNN INIPKRKEAK DSQIYLLKNI YYGEFVEKFV NDNKNFEKIV	660
KEIEEINRGA GTNKKTGFYK LEKFETLKN TPTKYLEKLQ SLHKISYDKE KIEEDKDVVY	720
DFVQKIFLKG PVNVLKLLDS LKSLNLLNLK KDETITDKKS VHDEKLKLWE NSGSNLSKMP	780
EEIYEYVKKI KISRNINYNDR MSIFYLLKL IDYRELTNLR GNLEKEYESMN KNKIYSEELT	840
IINLVNLNDN KVRTNFSLEA EDIGKFLKSS ITIKNIAOLN NFSKIFADGE NVIKHRSFYN	900
IKKYGILDLL EKIVAKADLK ITKEEIKKYE NLQNELKRND FYKIQEIQIHR NYNQKPSIK	960
KIENKKDPEK YKKVIEKIQD YTQLKNKIEP NDNLNLQSLI FRILHRLAGY TSLWERDLQF	1020
KLKGEFPEDK YKIDEIFNSDG NNNQKVKHHG IADKYANFL EIKEEKSGEI LNKKQRKKI	1080
KEDLEIRNYI AHFNYLPNAE KSILEHDKRL KNAVMSKSIKD IFREYGFIVE	1140
FTISHTKNGK KIKVCSVKSE KIKHLKNNEL ITTRNSEDLC ELVKIMLEHK ELQK	1194
SEQ ID NO: 21	moltype = AA length = 1159
FEATURE	Location/Qualifiers
source	1..1159
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 21	
MVKTVGGIS HKKYTSEGRL VKSESEENRT DERLSALLNM RLDMYIKNPS STETKENQKR	60
IGKLKKFFSN KMVYLKDNTL SLKNGKKENI DREYSETDIL ESDVDRDKNF AVLKKIYLNE	120
NVNSEEVEF RNDIKKLNK INSLKYSFEN NKANYQKINE NNIEKVEGKS KRNIYDYYR	180
ESAKRDAYVS NVKEAFDKLY KEDIKAFLV EBIENLTGLEK YKIREFYHEI IGRKNDKENF	240
AKIYEEIQN VNNMKELIEK VPDMSELKKS QVFYKYYLDK EELNDKNIKY AFCHPVEIEM	300
SQLKNVYVK RLSNISNDKI KRIFEYQNLK KLIENKLNLK LDTYVRNCGK YNYYLQDGEI	360
ATSDFIARNR QNEAFLRNII GVSSVAYFSL RNILETENEN DITGMRGKGT VKNNKGEKY	420
VSGEVDKLYN ENKKNEVDFN KLMFYSQDFN MDNKNEIEDF FANIDEAIS IRHGIVHFNL	480
ELEGKDIIFP KNIAPSEISE KMFQNEINEK KLKLKIFRQL NSANVFRYLE KYKILNLYLKR	540
TRFEFVNKNI PFVPSFTKLY SRIDLKNSL GIYWKTPKTN DNDNKTKEIID AQIYLLKNIY	600
YGFGLNYFMS MNGNPFEEISE EIIELNKNDK RNLKTGFYKL QKFEDIQEKI PKEYLANIQS	660
LYMINAGNQD EEEKDTYIDF IQKIPKGPW TYLANNGRLS LIYIGSDEET NTSLAEKKQE	720
FDKFLKKYEQ MNNIKIPYEI NEFLREIKLG NILKYTERLN MFYLLKLNL HKELTNLKGS	780
LEKYQSANKE EAFSDQLELI NLLNLDNNNRV TEDFELEADE IGKFLDFNQN KVVDNKLK	840
FDTNKKYFPG ENIICKHRAFY NIKKYGMLNL LEKIADKAGY KISIEELKKY SNKNEIEKN	900
HKMQENLHRK YARPCRDEKF TDEDYSEYQK AIENIEEYQK LKNKVEFNEI NLLQGLLR	960
LHLRLVGYSITI WERDLRFRLG GEFFENQYIE EIFNFENKKN VKYKGQOIVE KYIKPYKELH	1020
QNDEVKINKY SSANIKVLKQ EKKDLYIRNY IAHFNYIPHA EISLLEVLEN LRKLLSYDRK	1080
LKNAVMSKSV DILKEYGFVA TFKIGADKKI GIQTLSEKTI VHLKLNKKKK LMTRDNSEEL	1140
CKLVKIMFEY KMEEKKSEN	1159
SEQ ID NO: 22	moltype = AA length = 1134
FEATURE	Location/Qualifiers
source	1..1134
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 22	
MNELVKNRCK QTKTICQKLI PIGKTRETIE KYNLMEIDRK IAANKELMNK LFSLIAGKHI	60
NDTLSKCTDL DFEPLLTSLS SLNNAKENDR DNLREYYDSV FEKKTLAEE ISSRLTAVKF	120
AGKDFFTKNI PDFLETYEGD DKIEMSELVS LVIENTVTAG YVKKLEKIDR SMEYRLVSGT	180
VVKRVLTDNA DIYEKNEKA KDFDYGVNLN DEASQFTTLL AKDYANYLTA DGIAIYNVGI	240
GKINLALNEY CQKNKEYSYN KLALLPLQKM LYGEKLSLFE KLEDFTSDEE LINSYNKFAK	300
TVNESGLAEI IKKAVFSYDE IVIKPNKISN YSMSITGHWS LVNRIMKDYL ENNGIKNADK	360
YMEKGLTLSE IGDALENKNI KHSDFISNLI NDLGHTYTEI KENKESLKKD ESVNALIICK	420
ELDMLLSILQ NLKVFIDNE MFDTGFGIEV SKAIEILGYG VPLYNKIRNY ITKKDPKKK	480
FMTKFGSATI GTGITTSEVG SKKATFLKD DAVFLLLYNT AGCKANNVSV SNLADLINSS	540
LEIENSGKCY QKMIYQTPGD IKKQIPRWFV YKSEDDDLIK DFKAQGLHKTQ LSFLNGLRIP	600
YLKEAFATHE TYKNTFSYR NSYESYDEFN EHMSEQAYIL EWKWDIKKLI DDLVEDGSLL	660
MFRVWNRFMK KKEGKISKHA KIVNELEFSDE NASNAAIKLL SVFDIFYRDQ QIDNPIVHKA	720
GTTLYNKRTK DGEVIVDYTT MVKNEKEKRPN VYTTTKYDI IKDRRYTEEQ FEIHLHNIG	780

---

-continued

---

KEENKEKLET SKVINEKKNT	LVVTRSNEHL LYVVIFDEND	NILLKKSLNT VKGMNFKSKL	840
EVVEIQKKEN MQSWKTVGSG	QALMEGYLSF AIKEIADLVK	EYDAILVLEQ NSVGKNILNE	900
RVYTRFKEML ITNLSSLDVDY	ENKDFYSYTE LGGKVASWRD	CVTNGICIQV PSAYKYKDPT	960
TSFSTISMYA KTTAEKSKKL	KQIKSFKYNR ERLLFELVIA	KGVGLENNIV CDSFGSRSI	1020
ENDISKEVSC TLKIEKYLID	AGIEYNDEKE VLKDLDATAK	TDAVHKAVTL LLKCFNESPD	1080
GRYYISPCGE HFTLCDAPEV	LSAINYYIRS RYIREQIVEG	VKKMEYKKTI LLAK	1134

SEQ ID NO: 23	moltype = AA length = 1282	
FEATURE	Location/Qualifiers	
source	1..1282	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 23		
MNGNRIIIVYR EFVGVTVAK	TLRNELRPIG HTQEHHIIHNG LIQEDELRQE KSTELKNIMD	60
DYYREYIDKS LSGVTLDFT	LLFELMNLVQ SSPSKDNKKA LEKEQSKMRE QICTHMQSDS	120
NYKNIFNAFK LKEILPDFIK	NYNQYDAKDK AGKLETALAF NGFSTYFTDF FEKRKNVFTK	180
EAVSTSIAYR IVHENSITFL	ANMTSYKKIS EKALDEIEVI EKNNQDKMGD WELNQIFNPD	240
FYMMVLIQSG IDFYNEICGV	VNAHMMLNLYCQ QTKNNYNLFK MRKLHKQILA YTSTSFEVPK	300
MFEDDMSVYN AVNAFIDETE	KGNIIGKLKD IVNKYDELDE KRIYISKDFY ETLSCFMMSGN	360
WNLITGCVEN PYDENIHAKG	KSKEEKVKKA VKEKDKYKSIN DVNDLVEKYI DEKERNEFKN	420
SNAKQYIREI SNIITDTETA	HLEYDEHISL IESEEKADEM KKRLDMYMMNM YHWAKAFIVD	480
EVLDRDMEFY SDIDDIYNLN	ENIVPFLYRN RNVVTQKPYN SKKIKLNFQS PTLANGWSQS	540
KEFDNNAIIL DRDNKYYLAI	FNAKNDPKKK IIQGNNSDKKK DNDIHKMVYN LLPGANKMLP	600
KVPLSKKGIE TFKPSDYIIS	GYNAHKHKT SENFDISPCR DLIDYFKNSI EKHAERWKYE	660
FKFSATDSYN DISEFYREVE	MQGYRIDWTY ISEADINKLD EEGKIYLFQI YNKDFAENST	720
GKENLHTMYF KNIFSEENLK	DIIIKLNQGA ELYFRRASVK NPVKHKKDSV LVNKTYKNQL	780
DNGDVVRIPPI PDDIYNEIY	MYNGYIPI LSEAACEYLD KVEVRTAQKD IVKDYRTVD	840
KYPIHTPTI NYKVTLRNNV	NDMAVKYIAQ NDDIHVIGID RGERNLIIYIS VIDSHGNIVK	900
QKSYNILNNY DYKKTLLVEKE	KTREYARKNW KSIGNIKEKL EGYISGVVHE IAMLMVEYNA	960
IIAMIODELNYG FKRGFRKVER	QVYQKFESML INKLNYFASK GKSVDDEPGGL LKGYQLTYVP	1020
DNIKNLGKQC GVNFDTYNTM	TSKIDPSTGF ISAFNPKFSIS TNASRKQFPM QFDEQRYCAE	1080
KDMPSFGPDY NFNTDVTYNTM	SKTQWTVYTN GERLQSEFNN ARRTGKTSI NLTEFIKLLL	1140
EDNEINYADG HDVRIDMEKM	DEDKNSEFFA QLLSLYKLTV QMRNSYTEAE EQEKGSYDK	1200
IISPVINDEG EFFTSDNYKE	SDDKECKMPK DADANGAYCI ALKGLYEVLK IKSEWTEDGF	1260
DRNCLKLPHA EWLDFIQNKR YE		1282

SEQ ID NO: 24	moltype = AA length = 1399	
FEATURE	Location/Qualifiers	
source	1..1399	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 24		
MNKDIRKNT DFVGISEIQQ	TLRFILIPIG KTAQNIDKYN MFEDEDEIRHE YYPILKEACD	60
DFYRNHIDQQ FENLELDWSK	LDEALASEDR DLINETRATTY RQVLFNRLKN SVDIKGDSKK	120
NKTLSSLSSD KNLGKKKTKN	TFQYNFNDLF KAKLIKAIILP LYIEIYIEGE KLENAKKALK	180
MYNRFTSRLS NFWQARANIF	TDDEISTGSP YRLVNDNFTI FRINNNSIYT KNPFIEDIL	240
EFKKGLKSKK I KDKFESVND	YFTVNAFNLK CTQNGIDKYN SILGGFTTKE REKVKGLENL	300
FNLAAQOSK GKGEYRKNI	RLGKLTKLKK QLILAISDSTS FLIEQIEDDQ DLYNKIKDFF	360
ELLLKKEEIN ENIFTOYANL	QKLIEQADLS KIYINAKHLN KISHQVIGKW DSLNKGIAALL	420
LENININEES KEKSEVISNG	QTQDISSEAY KRYLQIQSEE KDIERLRTQI YFSLEDLEKA	480
LDLVLIDENM DRSDKILSY	VQSPDPLNVNF ERDLTDLYSR IMKLEENNEK LLANHSAILD	540
IKEFLDLIML RYSRWQILFC	DSNYLDQTF YPIYDAVMEI LSNIIRLYNL ANYLSRKPD	600
RMKKKKINFN NPTLADGWSE	SKIPDNSSML FIKDGMYLIG IIKNRRAAYSE LLEABSLQSS	660
EKKKSENSSY ERMNYHFLPD	AFRSIPKSSI AMKAVKEHFE INQKTADELL DTDKFSKPLR	720
ITKEIFDMQY VDLHKNKKY	QVDYLDRDGD KKGYRKALNT WLNFCKDFIS KYKGRNLFDY	780
SKIKDADHYE TVNEFYNDV	KYSYHIFFFS VAETTVEK SEGKLYLFQI YNKDFSPHST	840
GKPNLHTIYW RALFSEENLT	SKNKLNGQA EIFFRPKQIE TPFTHKKGSI LVNRPDVNGN	900
PIPINVYQEI KGFKNNVIKW	DDLNKTTQEG LENDQYLYFE SEFEIIKDRR YTEDQLFHV	960
PISFNWDIGS NPKINDLATQ	YIVNSNDIHI IGIDRGENHL IYYSVIDLQG AIVEQGSLNT	1020
ITEYETENKML MNKTNNLRKI	PYKDILOQRE DERADARIK HAIDKLKDLK DGYLGQIVHF	1080
LAKLIIKYNA IVILEDNYG	MKKLKNVLVFK DYDIDEIGGP	1140
LKPWQLTRPI DSYERMGRQN	GILFYVPAAY TSAVDPTVGF ANLFYLNWK NSEKFHFPSK	1200
FESIKYHSQ DMFSFAFDYN	NFGTTTRIND LSKSKWQVFT NHERSVWNK EKNVYTQNL	1260
DLIKLLRKY NIEFKNNQNV	LDSILKIEENN TDKENFAREL FRLFRLTIQL RNTTVNENNT	1320
EITENELDYI ISPVKDKGN	FFDSRDELKN LPDNGDANGA YNIARKGLY IEQLQESIKT	1380
GKLPTLSIST LDWFNYIMK		1399

SEQ ID NO: 25	moltype = AA length = 1313	
FEATURE	Location/Qualifiers	
source	1..1313	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 25		
MNKGGYVIME KMTEKNRWN	QFRITKTIKE EIIPPTGYTKV NLQRVNMLKR EMERNEDLKK	60
MKEICDEYYR NMIDVSLRL	EQRTLGWESL IHKYRMLNKD EKEIKALEKE QEDLRKKISK	120
GFGEKKAWTG EQFIIKKILPQ	YLMDDHYTGEE LEEKLIRIVKK FKGCMTFLST FFKNRENIFT	180

- continued

---

DKPIHTAVGH RITSENAMLF AANINTYEKM ESNVTLEIER LQREFWRGI NISEIFTDAY	240
YVNVNLKQKI EAYNKICGD1 NHQHMEYCQK QKLKFSEFRM RELKKQILAV VGEHPEIPEK	300
IESTKEVYRE LNEYYESLKE LHGQFEELKS VQLKYSQIYV QKKGYDRISR YIGGQWDLIQ	360
ECMKKDCASG MKGTKKNDA KIEEEEAKVK YQSIEHIQKL VCTYEEEDRHH KVTDVDEFI	420
VSCVCDLLGAD HIITRDGERI ELPLQYEPGT DLLKNDTINQ RRLSDIKTIL DWHMDMLEWL	480
KTFLVNDLVI KDEEFYMAIE ELNERMQCVI SVYNRIRNYV TQKGYYPEK1 RICFDKGTL	540
TGWTTGDNQY YSGFLLMRND KYYLGIINTN EKSVRKILDG NEECKDENDY IRVGYHLIND	600
ASKQLPRIVF MPKAGKKS1 LMKDQECDYI WDGYCHNKH1 ESKEFMRELI DYYKRSIMNY	660
DKWEGYCFK1 SSTEYSDNMQ DFYKEVREQS YNISFSYINE NVLEQLDKG KIYLFQVYNK	720
DFAAGSTGTP NLHTMYLQNL FSSQNLLEKR LRLGGNAELF YRPGETKDVT HRKGSILVDR	780
TVYREEKDGI EVRDTVPEKE LLEYIYRNG1 KQKGDLSES1 KQWLDKVHYR EAPCDIIKDK	840
RYAQEKYFLH FSVEINPNAE GQTALNDNVR RWLSEEDIH VIGIDRGERN LIYVSLMDGK	900
GRIKDQKSYN IVNSGNKEPV DYLAULKVRE KERDEARRNW KAIGKIKDIK TGYLSYVH	960
IVEMAREKA IIVMEDLNQY FKRGFRKVER QVYQKFEEM1 INKLNVYVVDK QLSVDEPGGL	1020
LRGYQLAFIP DKKKSMRQN GIVFYVPGY TSKIDPTTG1 VNIFKFPQFG KGDDDGNGKD	1080
YDKIRAFGGK PDEIREYCDE KVTADNTREV KERYRFDPDY SKFETHLVHM KKTWTVYAE	1140
GERIKRKVVG NYWTSEVISD IALRMSNTLN IAGIEYKDHG NLVNEICALR GKQAGIILNE	1200
LLEIVRLTVQ LRNSTTEGDV DERDEIISPV LNEKYGCFYH STEYKQONGD VLPKDADANG	1260
AYCIGLKG1 EIRQIKNKWK EDMTKGEGKA LNEGMRISHD QWFEFIQNMN KGE	1313

SEQ ID NO: 26                  moltype = AA length = 1470  
 FEATURE                        Location/Qualifiers  
 source                        1..1470  
 mol\_type = protein  
 organism = synthetic construct

SEQUENCE: 26  
 MNINPRGAGGG FTNLNSLSKT1 LRFELKPYLE IPEGEKGKLF GDDKEYYKNC KTYTEYYLKK 60  
 ANKEYYDNEK VKNTDLQLVN FLHDERIEDA YQVLKPVFDT LHEEFITDSL ESAEAKKIDF 120  
 GNYYGLYEKO KSEQQNKEKK KIDKPLETER GKLRKAFTPI YEAEGKNLKN KAGKEKKDKD 180  
 ILKESGFKVL IEGAILKYIK NNIDEFADKK LKNNEGKEIT KKDIETALGA ENIEGIFDFG 240  
 FTYFSGFNQN RENYYSSTTAT AVASRIVD ENLSKFCND1 LLYRKNEENDY LKIFNFNLLKNK 300  
 GKDLSKLSK PFGKENEPF1 PAYDMKNDK SF5VADFVNC LSQGEIEKYN AKIANANYLI 360  
 NLINQNQKDG1 SSKLMSMFKIL YKQIGCGEKK DFIKTIKD1 ELKQILEKAC EAGKKYFIRG 420  
 KSEDGGVSNI FDFTDYIQSH ENYKGVYWSD KAIMTISGKY FANWDTLKNN LGDAKVFNKN 480  
 TGEDKADVKY KVPQAVMLSE LFAVLDDNAG EDWREKGIF1 KASLFFEGDQN KSEI1KNNANR 540  
 PSQALLKMIC DDMESLAKNF1 IDSGDKILK1 SDRDYQKDEN1 QKQIKNWLDN ALWINQILKY 600  
 FVKKANK1KG DSIDARIDSG LDMLVFFSSDN PAEDYDMIRN YLTQKQPDEI NKLKLNFENS 660  
 SLAGGWDENK EKDNC1IILK DEQDKQYLV MKYENTKVF1 QKNSQLYIAD NAAWKKMIYK 720  
 LVPGASKTLP WMMFSKKWTA NRPTPSD1E IYQKGFSKKE NVDWFNDKKE DESRKEKNRE 780  
 KIIAELQKTC WMDIRYNDG KIESAKVY1NK EKLAKSLIDFY1 KENLKYKPS1 EESWDRLF1AF 840  
 GFSDTCKSYKS IDQFYIEVDK QGYKLEFVT1 NKARLDBEYR DGK1YLFEIR SRDNNLVNGE 900  
 EKTSAKN1Q1T IYWNAFGGD DNPKPLNGEA EIFYRPAIAE NKLNNKKDDKN GKEIIDGYRF 960  
 SKEKF1FHCP ITLNFCLKET1 KINDKLNAA1 AKPENGQGVY FLGIDRGEKH LAYYSLVNQK 1020  
 GEILEQGTLN1 LPFLDKNGKS1 RSIKVEKSF1 EKDSNGGIK1 DKDGNNDK1 EFVECWN1YND 1080  
 LLDARAGDRD YARKNWT1T1 TIKE1KDG1Y1 SOVVRK1V1L SIYKNTETKE FREMPAF1V1L 1140  
 EDLNIGFKRG RQKIEKQVYQ KLELALAKKL1 NFLVDDK1ADI GEIGSVTKAI QLTPPVNNFG 1200  
 DMENRKQFGN MLYI1RADYT5 QTDPATGW1K STY1LKSGS1ES NVKCEQ1EKS1 FDIRYESG1DY 1260  
 CFEYRDRHRGK MWQLYSSK1 VSLDRPH1GER NNSKNW1WESE KQPLNEMLD1 LFDEKRPDKS 1320  
 KSLYEQMFKG GVALTRLPKE1 INKKD1KPAWE1 SLRFV11L1Q QIRNTGKNGD1 DRNGDF1QSP 1380  
 VRDEKTGEHF DSRIYLDKEQ1 KGEKADLPTS1 GDANGAYNIA1 RKGIVVAEHI1 KRGFDKLY1S 1440  
 DEEWDTWLAG DEIWDKWLKE1 NRESLT1KTRK1 1470

SEQ ID NO: 27                  moltype = AA length = 1259  
 FEATURE                        Location/Qualifiers  
 source                        1..1259  
 mol\_type = protein  
 organism = synthetic construct

SEQUENCE: 27  
 MYKKTGLEDF1 IGKESLSKTL1 RNAL1PTTEST1 KIHMEEMGV1 RDDELRAEKQ1 QELKEIMDDY 60  
 YRTF1EKKLG1 Q1QGQ1WNSL1 FQKMEETMED1 ISVRKD1DK1 QNEKRKE1CC1 YFTSDKRFKD 120  
 LFNNAKL1T1D1 LPNFIKDNKE1 YT1EE1KAEK1 QTRVL1FQRFA1 TAFTY1FNQ1 RNNFSE1D1N1S 180  
 TAISFRV1NE1 NSE1IHLQ1NMR1 AFQRIEQQY1P1 EEVCGMEEY1 KDM1QEWQMK1 H1YSVDFY1DR 240  
 EL1TQPG1IEY1 NG1C1GK1NEH1 MNQFQ1QK1N1I1 NKND1FRM1KK1 HKQ1LCK1SS1 YYE1P1FRF1ES 300  
 DQEYV1DALN1E1 PIKTM1KK1E1 IRR1CV1H1L1Q1E1 CDDY1DLG1K1Y1 ISSN1K1YE1Q1S1 NALY1GS1W1DT1 360  
 RKCIKEEY1MD1 ALPGK1GEK1E1 EKAAEAAK1E1 EYRS1IAD1K1 I1S1L1Y1G1SEM1 RT1S1AK1C1T1 420  
 E1CD1M1C1Q1S1 D1P1L1V1C1N1S1 K1LL1Q1K1E1T1 E1K1T1L1D1S1F1 H1V1Y1Q1W1Q1T1F1 V1S1D1I1E1K1D1S1Y1 480  
 FYSE1LED1V1L1E1 D1F1E1G1T1L1Y1N1 H1V1R1S1Y1T1Q1K1P1 Y1S1T1V1F1K1L1H1F1 G1S1P1T1L1A1G1W1S1 540  
 L1L1R1D1Q1K1Y1 G1F1V1R1N1K1P1D1 K1Q1I1K1G1H1E1 K1E1K1D1Y1K1M1 Y1N1L1P1G1P1S1M1 K1P1V1F1T1S1R1G1 600  
 QETY1K1P1K1H1 I1D1G1Y1N1E1K1R1H1 K1S1P1K1F1D1G1Y1 C1W1D1L1Y1K1 E1C1H1K1H1P1D1W1K1 Y1D1F1H1P1S1T1K1D1 660  
 YED1S1G1F1Y1R1 E1V1M1Q1G1Y1Q1K1W1 TY1I1S1A1D1E1Q1K1 L1D1E1K1G1Q1F1L1F1 R1T1I1S1A1K1C1T1 720  
 Y1L1K1N1L1F1S1E1N1 L1K1D1V1L1K1L1N1G1 E1A1E1L1F1R1K1S1 A1K1T1P1I1V1H1K1 G1S1V1L1N1R1S1Y1T1 Q1V1G1N1K1E1R1V1S1 780  
 I1P1E1E1Y1Y1T1 E1Y1N1L1H1G1K1G1K1 LS1S1E1A1Q1Y1L1D1 E1G1K1K1S1F1T1 A1K1T1V1N1Y1R1Y1C1 D1C1H1F1Y1L1H1P1 840  
 T1N1F1K1A1K1S1 D1V1N1E1R1T1L1A1Y1 I1K1K1E1H1I1H1 I1D1R1G1E1R1N1L1Y1 I1S1V1D1V1H1V1G1N1I1 R1E1Q1R1S1F1N1I1V1N1 900  
 GYD1Y1Q1Q1K1L1D1 R1E1K1S1R1D1A1R1K1 N1W1E1E1I1K1 E1L1K1E1G1Y1L1S1M1 V1H1Y1A1Q1L1V1V1V1K1 Y1N1A1V1V1A1M1 D1L1N1E1D1L1N1 960  
 YGCF1KT1G1R1F1K1 V1E1R1Q1V1Y1K1 F1E1R1D1R1E1V1C1 E1G1L1K1H1Y1L1V1Y1 F1K1D1R1E1V1C1 E1G1L1R1G1Y1Q1L1T1 1020  
 QCGF1F1Y1V1P1A1 G1Y1T1S1K1D1P1 T1G1F1V1N1S1F1 S1K1T1N1R1S1Q1D1 F1G1V1F1N1R1S1Q1D1 F1G1V1F1N1R1S1Q1D1 1080

-continued

---

DYNYYIKKGT I LASTKWKVY TNGTRLKKIV VNGKYTSQSM EVELTDAMEK MLQRAGIEYH	1140
DGKDLKQCOIV EKGIEAEIID IFLRLTVQMRN SRSESEDREY DRLISPVLND KGEFFDTATA	1200
DKTLPQDADA NGAYCIALKG LYEVKQIEN WKENEQFPRN KLVQDNKTWF DFMQKKRYL	1259
 SEQ ID NO: 28	moltype = AA length = 477
FEATURE	Location/Qualifiers
source	1..477
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 28	
MRIKSTLSSLR IVRPFYTPEV EAGIKAEDK REAQGQTRSL DAKFFNELKK KHSEIILSSE	60
FYSLLSEVQR QLTISIYNHAM SNLYHKIIIVE GEKTSTSALK SNIGYDECKA IFPSYMALGL	120
RQKIQSNSR RDLKNFRMMAV PTAKSDFKPI PIYRQVDGSK GGFKKISENDG KDFIVEPLV	180
DYVAEVEVKA KGRFTKINIS KPPKTKNIPV ILSTLRRRQG GQWFSDDGTN AEIRRVISGE	240
YKVSWIEIVR RTRFGKHDW FVNVMVICKY PEEGLDSKVV GGIDVGVSSP LVCALNNSLD	300
RYFVKSSDII AFNKRAMAR RTLLRQNPKY RSGHGSKNKL EPITVLTEKU ERFKKSIMQR	360
WAKEVAEFFR GKGASVVRME ELSGLKEKD FFSSYLRMYW NYGQLQQIE NKLKEYGIKV	420
NVVSPKDTSK KCHSCTHINE FFTFEYRQKN NFPLFKCEKC GVECSADYNA AKNMAIA	477
 SEQ ID NO: 29	moltype = AA length = 1293
FEATURE	Location/Qualifiers
source	1..1293
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 29	
MRTMVTPEDF TKQYQVSKTL RFELIPQGKT LENMKRDGII SVDRQRNEDY QKAKGILDKL	60
YKYLDFITME TVVIDWEALA TATEEFRSK DKKTYEKVQS KIRTALLEHV KKQKVGTEDL	120
FKGMFSSKII TGEVLAAFPE IRLSDEENLI LEKFKDFTTY FTGFFENRKN VFTDEALSTS	180
FTYRLVMDNF IKFFDNCIVF KNVNVNISPHM AKSLETCASD LGIFPGVSLE EVFSISFYNR	240
LITQTGIDQF NQLLGGISGK EGEHKQQGLN EIINLAMQON LEVKEVLKNNK AHRFTPLFKQ	300
IILSDRSTMSF ICPDAFADDDE VLSAVDAYER YLSEKNIGDR AFQLISDMEA YSPELMRIGG	360
KVVSVLSQLL FYWSWEIRDG VKAYKESLIT GKTTKKELEN IDKEIKYGV LQEIKEALPK	420
KDIYEEVKY AMSVVKDYHA GLAEPPLPEKI ETDDERASIK HIMDSMLGLY RFLEYFSHDS	480
IETDTPVGE CLDTILDDMN ETVPYLYNKVR NFSTRKYST EKFKLNFNNS SLANGWDKKN	540
EQANGAILL RKEGYFLGIF NSKPNPKLVS DGGAGIGYEK MIYKQFPDFK KMLPKCTISL	600
KDTKAHFQKS DEDFTLQTDK PEKSIVITQK IYDLGQTQVN GKKKFQVDPY RLTGDMEGYR	660
AALKEWIDFG KEFIQAYTST AIYDTSLPFD SSDYPDLPDF YKDVDNICYK LTFEWIPDAV	720
IDDCICDDGSL YLFKLHNKDF SSGSIGKPNL HTLYWKALFE EENLSDVVVK LNGQAELFYR	780
PKSLTRPVVH EGEVEIINKT TSTGFLPVPDD VYVELSKFVR NGKKGNLTDK AKNWLDDKTV	840
RKMPHAITKD RRFTVDKFFF HVPIVTLNYKA DSSPYRNFDFD VRQYIKDCSD VKİIGIDRGE	900
RNLIYAVVID GKGNIIEQRS FNTVGTYNYQ EKLEQKEKER QTARQDWATV TKIKDLKKGY	960
LSAVVHESLX MIVKYKAIVA LENLNVGFKR MRGGIAERSV YQQFEKALID KLNLYVFKDE	1020
EQSGYGVNLN AYQLTDFKES FSKMCGQQTGF LFYVPAAYTS KIDPLTGFIN PFSWKHVKNR	1080
EDRRNFLNLF SKLYYDVNTN DFVFLAYHHSN DKSQYTIKGN WEIADWDLI QENKEVFGKT	1140
GTPYCVGKRI VYMDDSTGH NRMCCAYYPT ELKQLLSSEYG IEYTSQDLL KIIQEFDDDK	1200
LVKGFLFYIIK AALQMRNSNS ETGEDYIISSP IEGRPGICFD SRAEADTLPY DADANGAFHI	1260
AMKGLLLTER IRNDDKLAIS NEEWLNYIQE MRG	1293
 SEQ ID NO: 30	moltype = AA length = 1228
FEATURE	Location/Qualifiers
source	1..1228
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 30	
MSKLEKFTNC YSLSKTLRFK AIPVGKTOEN IDNKRLLVED EKRAEDYKGV KKLLDRYLYS	60
FINDVLHSIK LKNLNNYIISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYSKSLFK	120
KDIETILPE PLDDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEEKSTS IAFRCINENL	180
TRYISNMDF EKVDAIFDKH EVQEIKEKIL NSDYDVDFDF EGEEFFNFVLT QEGIDVYNAY	240
IGGFVTESEGE KIKGLNEYIN LYNGKTKQKL PKFKPLYQV LSDRESLSFY GEGYTSDEEV	300
LEVFRNTLNK MSEIFKSIKK LEKLFKNPDE YSSAGIFVKN GPAISTISKD IFGEWNVIRD	360
KWNAEYDDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLSQL QEYADADLSV VEKLKEIIIQ	420
KVDEIYKVYG SSEKLFDAF VLEKSLKKND AVVAIMKDLL DSVKSFENYI KAFFGEKGKET	480
NRDESFGYDF VLAYDILLKV DHIYDAIRNY VTQKPYSDKK FKLYFQNQPF MGGWDKDKET	540
DYRATILRYG SKYALIADM KYAKCLQKID KDDVNGNYEK INYKLLPGPN KMLPKVFFSK	600
KWMAYYNPSE DIQKIQYKNGT DFKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET	660
EKYKDIAGFY REVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNNKDFS DKSHTCPNLH	720
TMYFKLLFDE NNHGQIRLSG GAELFMRAS LKKEELVVP ANSPIANKNP DNPKKTTTLS	780
YDVYKDKRFS EDQYELHIPI AINKCPKNIK KINTEVRVLL KHDDNPYVIIG IDRGERNLLY	840
IVVVDGKGNV VEQYSLNEII NNFNCGIRIKT DYHSLLDKKE KERFEARQNW TSIENIKELK	900
AGYISQVVKH ICELVEKYDA VIALEDLNSG FKNSRVRKVEK QVYQKFEKML IDKLNMYMVDK	960
KSNCPCATGGA LKGYQITNPK ESFKSMSTON GFIFYIPAWL TSKIDPSTGE VNLLKTKYTS	1020
IADSKKFISS FDRIMYVPEE DLFEFALDYK NFSRTDADYI KKWKLISYGN RIRIFRNPKK	1080
NNVFDWEVEC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSMLQMRNS	1140
ITGRTDVDFL ISPVKNSDGI FYDSRNYEAQ ENAILPKNAID ANGAYNIARK VLWAIGQFKK	1200
AEDEKLDKVK IAISNKEWLE YAQTSVKH	1228

-continued

---

SEQ ID NO: 31	moltype = AA length = 1286
FEATURE	Location/Qualifiers
source	1..1286
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 31	
MTNFDNFTKK YVNSKTIRLE AIPVGKTLKN IEKMGFIAAD RQRDEDYQKA KSVIDHIYKA	60
FMDDCLKDLF LDWDPLYEAV VACWRERSPE GRQALQIMQA DYRKKTADRF RNHELYGSLF	120
TKKIFDGSAV QRLPDLEQSA EEKSSLNSFN KFTSYFRDF DKRKRLFSDD EKHSAIAYRL	180
INENFLKFVA NCEAFRRMTE RVPELREKLQ NTGSLQVYNG LALDEVFSAD FYNOLIVQKQ	240
IDLYNLQIIG IAGEPGPTNI QGLNATINLQ LGQDSSLHEK LAGIPHRFNP LYKQILSDVS	300
TLSFVPMSAQ SDGEMLAAVE GFVKVQLESGR VLQNVRRLFN GLETEADLSR VYVNNNSKLA	360
FSSMFFGRWN LCSDALFAWK KGKQKKITNK KLTEIKWLK NSDIAIAEIQ EAEGEDFPRG	420
KINEKIAQAA DALHSQLALP IPENKLALCA KDGKLSMLDT VLGLYRMLQW FIVGDDNEKD	480
SDFYFGLGKI LGSLDPVLVY YNRVRNYITK KPYSLTKPFQ NFDNSQLLNG WDENNLDTNC	540
ASIFIKDQKY YLGISNKNNP PQFDTVATSG KSGYQRMVYK OFANWGRDLP HSTTMQMKVK	600
KHFSASDADY VLDGDKFIRP LIITKEIFDL NNVFKFNGKKK LQVDYLRTNG DREGYTHALH	660
TWINFLKDFC ACYKSTSIYD ISCLRPSTDQY DNLMDFYADL GNLSHRIWQ TIPEEAIADNY	720
VEQGQLFLFQ ZLNKDFAPGA DGKPMLHLYT WKAVENTPNEL EDVVFVNLNGK AELFYRPRSN	780
MDVVRHKVGE KLVNRLKLNG LTLPRLHEE IYRYVNGTLN KDLSADARSV LPLAVRVDQ	840
HEIIKDRRTF ADKFFFHASL TFMFKSSDKP VGFNEDVREY LREHPDTYVV GVDRGERNL	900
YIVVIDPQGN IVEQRFSNMI NGIDYWSLLD QKEKERVEAK QAWETVGKIK DLKCGYLSFL	960
IHEITKIIK TDNTRTAENT SLGFKVRTG IAEKAVYQQF ERMLVTKLGY VVFKDRAGKA	1020
PGGVLNAYQL TDNTRTAENT GIQNGFLFVY PAAFTSRVDPD ATGFFDFYDW GKIKTATDKK	1080
NFIAGFNSVR YERSTGDFIV HVGAKNLAVR RVAEDVRTEW DIVIEANVRK MGIDGNSYIS	1140
GKIRIYRSGE QGHGQYENHL PCQELIRALQO QYGIQYETGK DILPAILQOD DAKLTDVFD	1200
VFRALALQMNR TSAETGEDYF NSVVRDRSGR CFDTRRAEAA MPKEADANDA YHIALKGFLV	1260
LEKLRKGESI GIKNTEWLRY VQQRHS	1286
SEQ ID NO: 32	moltype = AA length = 1272
FEATURE	Location/Qualifiers
source	1..1272
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 32	
MTPIFCNFVV YQIMLFNNNI NINVKTMNNK HLSDFTNLFP VSktLRFRLV PQGKTMENIV	60
KAQTIEDEE RSHDYEKTK YIDDYHRQPI DDTLDKFAFK VESTGNNDSL QDYLDAYLSA	120
NDNRTKQTEE IQTNLRAKIV SAFKMQPQFN LLFHKEMVKH LLPQFVDTDD KKRIVAKFND	180
FTTYFTGFFT NRENMSYDEA KSTSIAYRN NONLIKFVEN MLTFKSHILP ILPOEQLATL	240
YDDFKEYLNV ASIAEMFELD HFSIVLTLQRQ IEVYNSVIGG RKDENNKQIK PGGLNQYINQH	300
NQAVKDKSAR LPLLKPLFNQ ILSEKAGVSF LPKQFKSASE VVKSLENEYA ELSPVLAQI	360
DVVTNITYD CNGIFIKNDL GLTDIAQRFY GNYDAVKRGL RNQYELETPM HNGQKAEKYE	420
EQVAKHLKSI ESVSLAQINQ VVTDDGGDID YFKAFGATDD GDIQRENLIA SINNAHTAIS	480
PVLNKENLAND NELRKNTMLI QWFAKPLLGA GDETNKDQVF YGKFEPLYNQ	540
LDETISPLYD KVRSYLTKKP YSLDKFKINF EKSNLLLGGWD PGADRKYQYN AVILRKDNDF	600
YLGIMRDEAT SKRKCIVQLD CNDEGLDENF EKVEYKQIKP SQNMPRCFAA KKECEENADI	660
MELKRKKNQ SYNTNKKDDK ALIRHYQRL DRTYPEFGFV YKDADEYDTV KAFTDSMDSQ	720
DYKLSFLQVS ETGLNLKLVDE GDYLFLKTN KDFSSYAKGR PNLTHTYWRM LFDPKNLANV	780
VYKLEGKAEV FFRRKSLAST TTHAKQAIK NKSRYNEAVK PQSTFDYDII KDRRPTADKF	840
EFHVPIKMNF KAAGWNSTRL TNEVREFIKS QGVRHIIGID RGERHLLYLT MIDMDGNIVK	900
QCCLNAPQAQD NARASEVDYH QLQDSKEADR LAARRNNGTI ENIKELKQGY LSQVVHLLAT	960
MMVDNDAILV LENLNAGFMR GRQKVESVY RQCFKMLIDK LNYIVDKGQS PDKPTGALHA	1020
VQLTGLYSDF NKSNSMKRANV RQCGFVFYIP AWNTSKIDPV TGTVNLFDTH LSSMGEIKAF	1080
FSKFDISRYN QDKGWFEFKF DYSRFTTRAEG CRCTQWTVCT YGERIWTHRS KNQNNQFVND	1140
TWVNTQQLMQ LLQDCGIDPN GNLLKEAIANI DSKKSLETLL HLFKLTQVMR NSVTGSEVDY	1200
MISPVADERG HFFDSRESDE HLPANADANG AFNIARKGLM VVRQIMATDD VSKIKFAVSN	1260
KDWLRLFAQHI DD	1272
SEQ ID NO: 33	moltype = AA length = 477
FEATURE	Location/Qualifiers
source	1..477
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 33	
VKISKTLSLR IIRPYYTPEV ESAIKAEDK REAQGQTRNL DAKFFNELKK KHPQIILSGE	60
FYSLLFEMQR QLTSIYNRAM SSLYHKIIVE GEKTSTSALK SDIGYDECKS VFPSYIALGL	120
RQKIQSNFRR KELKGFRMAV PTAKSDKFPI PIYKQVDDGK GGFKISENKE GDFIVEPLV	180
EYTAEDVKTA KGKFTKINIS KPPKIKNIPV ILSTLRRKQS QWQFSDEGTN AEIRRVISGE	240
YKVSWIEVVR RTRFGKHDDW FLNIVIKYDK TEDGLDPEVV GGIDGVGVSTP LVCAVNNSLD	300
RYFVKSSDII AFKKRAMEARR RTLLRQNRFK RSGHGSKSKL EPITILTEKN ERFKKSIMQR	360
WAKEVAEEFK GERAHSVQME ELSGLKEKDQ FFGSYLRLMWY NYGQLQQIE NKLKEYGIKV	420
NYVSPKDTSK KCHSCGYINE FFTFEFRQKN NFPLFKCKK GVECNAODYNA AKNIAIA	477
SEQ ID NO: 34	moltype = AA length = 1281
FEATURE	Location/Qualifiers
source	1..1281

---

-continued

---

```

mol_type = protein
organism = synthetic construct

SEQUENCE: 34
MLKSYDYFTK LYSLOQKTLRF ELKPIGKTL EHIKNNSGII EDETLEEQQYAI VKNIIDKLHR 60
KHIDEALSLV DFTKHLDTLK TFQEELYLKR KTDKEKEEL KLSADLRKLI VSYIKGKVNKE 120
KTQHNLNP KERFEILFGK LFNTNEEFFLL AENEKEKKAI QAFKGFTTYE KGFQENRKNM 180
YSEEGNNTSI AYRIIENNL PFIENIARFQ KVMSSTIEKTT IKKLEQNLKT ELKKHNLPGI 240
FTIEYFNVL TQEGRISRYNT IIGGGTTTHEG VKIQGLNEII NLYNQOSKDV KLPILKPPLHK 300
QILSEEYSTS FKIKAFFENDN EVLKAIDTFW NEHIEKSIHP VTGNKFNILS KIENLCDQLQ 360
KYKDKDLEKL PIERKNLSTV SHQVYQWNI IRDALRMHLE MNKKNIKEKQ IDKYLDNDAF 420
SWKEIKDSIK IYKEHEVEDAK ELNENGIIKQ FSAMSINEED DEKEYSISLI KNINEKYNV 480
KSILQEDRTG KSDLHQDKEE VGIKEFLDS LKQLQWFRLR LYVTVPPLDEY DYEFYNELEV 540
YYEALLPLNS LYNKVRNYMT RKPYSVEKFK LNFNSPTLLD GDWKKNETAN LSIILRKNGK 600
YLYLGIMKEN NTIFFEYPGT KSNDYYEKMI YKLPGPNKMK LPKVFFSKKG LEYYNPPKEI 660
LNIYEKGEFK KDKSGNFKKE SLHTLIDFYK EAIAKNEDEW VFNFKFKNTK EYEDISQFYR 720
DVEEQGYLIT PEKVDANYVD KLVKEGKLYL FOIYQNKFDFE NKKSKGPNL HTIYWKGLYD 780
SENLKVNYYK LNGAEAVFYR KKSIDYPEEI YNHGHHHKEEL LGKFNYPIK DRRTYQDKFL 840
FHVPITMNF1 SKEEKRVNQL ACEYLTSATKE DVHIIIGIDRG ERHLLYLSLI DKEGNIKKQL 900
SLNTIKNENY DKEKEKRDEA RKNWDVIEEN KELKEGMSQ VIHIIAKMMV 960
EKKALIMED LNIGFKRGRF VKEVQVYQKF EKMLIDKLYN LVFKKNPYLE PGGSLNAYQL 1020
TSKFDSFKKL GKQSGFIFYV PSAYTISKIDP TTGFYNFQIV DVPNLEKGKE FFSKFEKIIY 1080
NTKEDYFEFH CKYKGPFVSEPN KNKDNDRKT KESLTYYNAIK DTVWVVCSTN HERYKIVRNK 1140
AGYYYESHPVD VTKNLKDIFS QANINYNEGK DIKPIIIIESN NAKLLKSIAE QLKLILAMRY 1200
NNNGKHGDEK DYILSPVKNK QGKFFCTLDG NQTLPINADA NGAYNIALKG LLLIEKIKKQ 1260
QGKIKDLYIS NLEWFMFMMS R 1281

SEQ ID NO: 35 moltype = AA length = 1367
FEATURE Location/Qualifiers
source 1..1367
mol_type = protein
organism = synthetic construct

SEQUENCE: 35
MEKSLNDFIG LYSVSKTLRF ELKPVSETLE NIKKFHFLEE DKKKANDYKD VKKIIDNYHK 60
YFIDDVLKNA SFNWKKLLEEA IREYNKNKSD DSALVAEOKK LGDAILKLFT SDKRYKALTA 120
ATPKELFESI LPDNPFKQCN QDLNNAALKT FOKFTSYPTFQ FQENRKNVY AEEAIPATVPY 180
RIVNDNPFK LQNVLIFKTI QEKCPOQALIDE VEKEBLSSYLG KEKLAGIFTL ESFNKYLQGQ 240
GKENQRGIDF YNQIIGGVVE KEGGINLRRV NOFLNLYWQO HPDFTKEDRR IKMVPLYKQI 300
LSDRSSLFSK IESIENDEL KNALLECADC LELKNDEKKS IFEEVCDLFS SVKNLDSLGI 360
YINRKDINSV SRILTGDFWSL LQSRMNMYAE EKPTTAKAE RRWQKSLDDEG ENKSKGFYSL 420
TDLNEVLEYS SENVATDIR ITDYDVKR YYVDKETEMF VQGSELVALS LQEMCDDILK 480
KRKAMNTVLE NLSSENKLRE KTDDAVAVIKE YLDAVQELLH RIKPLKVNGV GDSTFYSVYD 540
SIYALSEV1 SVYNKTRNYI TKAASAPEKY KLNFDNPTLA DGWDLNKEQA NTSVILRKDG 600
MFYLGIMNPK NPKPFAEKYD CGNESCYEM IYKQFDATQK IPKCSCTQKKE VQKYFLSGAT 660
EPYILNDKKS PKSELIITKD IWPMNNHNV GEKFVPKRDN ETRPKKFOIG YFKQTGDFDG 720
YKNALSNWIS FCKNFLQSYL SATVDYDNPK NSBEYYEGLDE FYNYLNATCY KLNFINIPET 780
EINKMVSEGK LYLFQIYKND FASGSTGMVN MHTLYWKNLF SDENLKNVCL KLNGEAELFY 840
RPAIGIKEPV1 HKEGSGYLVN TTEDGESIPE KIYFEIYKNA NGKLEKLSDE AQNYIISNHEV 900
VIKKAGHEII KDRHYTEPKF LFHVPLTINF KASGNNSYIN ENVRFKLN PDVNIIGLDR 960
GERHHLIYLSL INQKGEIIKQ FTFNEVERNK NGRTIKVNYH EKLDQREKER DAARKSQAII 1020
GKIAELKEGY LSAVIHQLTK LMVEYNAVVA MEDLNGFGK GRFHVKEQVY QKFEHILIDK 1080
SNYLVFKDRG LNEPGVGVLNQ YQIAQFESF QKLGKQSGML FYVPAGYTSK IDPKTGFVSM 1140
MNFKDLTNVH KKRDFFSKFD NIHDEANGS FVFTFDYKFK DGKAKEEMKL TKWPSVSRDK 1200
RIVYFAKTKS YEDVLPTEKL QKIFESNGID YKSGNNIODS VMAIGADLKL GAKPSKEISD 1260
FWDGLLSNFK LILOQMRNSNA RTGEDYIISP VMADDGTFFD SREEFKKGED AKLPLDADAN 1320
GAYHIALKGL SLINKINLSK DEELKKFDMK ISNADWFKFA QEKNYAK 1367

SEQ ID NO: 36 moltype = AA length = 809
FEATURE Location/Qualifiers
source 1..809
mol_type = protein
organism = synthetic construct

SEQUENCE: 36
MIKNPSNRHS LPKVIISEVD HEKILEFKIK YEKLARLDRF EVKAMHYEGK EIVFDEVVLN 60
GGLIEVEYQD DNKTLFVKVG EKSYSIRGKK VGGKQRLLED RVSKTKVQLE LSDGVVDNKG 120
NLRKSRTERE LIVADNLKLY SQIVGREVTT TKEIYLVKRF LAYRS DLLFY YSFVDNFFKV 180
AGNEKEWLKI NFDDATSAQF MGYIIPFMVND NLKNDNAYLK DYVVRNDVQIK DDLKKVQTIF 240
SALRHTLLHF NYEFFEKL FN GEDVGFDFDI GFLNLLIENI DKLNNIDAKKE FIDNEKIRLF 300
GENLSSLAKVY RLYSDICVNR VGFNKFINSM LIKDGVENQV LKAEFRNRKFG GNAYTIDIHS 360
NQEYKRIYNE HKKLVIKVST LKDQQAIRRG NKKISELKEQ MKSMTKKNSL ARLECKMRLA 420
FGFLYGEYNN YKAFKNNFDT NIKNSQFDVN DVEKSKAYFL STYERRKPT REKLEKVAKD 480
IESLELKTVI ANDTLLKFL LMVFVMPQEL KGDFLGFVKK YYHDVHSIDD DTKEQEEDVV 540
EAMSTSLKLK ILGRNRISLT LFKYALSSQV NYNSTDNIFY VEGNRYGKIV KKLGISHNQE 600
EFDKTLVPL LRYYSSLFKL MNDFEIYSLA KANPTAVSLQ ELVDDETSPY KQGNYFNFNK 660
MLRDIYGLTS DEIKSGQVVF MRNKIAHFDT EVLLSKPLLG QTKMNLQRKD IVSFIEARGD 720
IKELLGYDAI NDFRMKVHL RTKMRVYSDK LQTMMDLLRN AKTPNDFYNV YKVKGVESIN 780
KHLLEVLAQT AEERTVEKQI RDGNEKYD 809

```

---

-continued

---

```

SEQ ID NO: 37      moltype = AA length = 839
FEATURE
source          Location/Qualifiers
1..839
mol_type = protein
organism = synthetic construct

SEQUENCE: 37
LNSIEKIKKP SNRNSIPSII ISDYDENIK EIKVKYLKLA RLDKITIQDM EIRDNIVEFK 60
KILLNGIEHT IKDNQKIEFD NYEITAYVRA SKQRDRGKIT QAKYVVTITD KYLRDNEKEK 120
RFKSTERELP NDTLLMRYKQ ISGFDTLTSK DIYKIKRYID FKNEMLFYFQ FIEEFFSPLL 180
PKGTNFNLSN IEQNNDKVVN IYIVYRLNDDF KNQSLNQPIK KTDTIKYDFL KIQKILSDFR 240
HALAHFDDF 1QKFFDDELL KNRFDISTIS LIKTMQLQBEK EKYYQEKNYIED SDSLTLTF 300
DEKESENFSKI HNFYIKISQK KPAFNKLINS FLSKDGVPN ELKSYLATKK IDFFEDIHSN 360
KEYKKIYIKH KNLVVEKQKESQEKPPNGQK LKNYNDDELQK LKDEMKNITK QNSLNRLLEV 420
LRLAEGFIAN EYNVNFKNFN DKFTLVDVKKE QKIKVFKNSN NEKKEYFES TFIEKRFHF 480
CVKFFNKKTK KEETQKONIF NLINENETLEE LVKESPPLLQI ITLLYLFIPIK ELQGEFGVGFI
LKIYHHTKNI TNDTKEDEKS IEDTQNSFSL KLKILAKNLR GLQLFNYSLS HNTLYNTKEH 600
FFYEKGNRWQ SVYKSLEISH NQDEFDIHLV IPVIKYIYIN NKLIGDFEIY ALLTYADKNS 660
ITEKLSDTIK RDDLKFRGYY NFSTLNFPTF MINTNYEQNQ KSTQYIKQTR NDIAHQNIEN 720
MLKAFNNEEI FAQREBIVNVY LQKEHKMQUEI LHYNPINFT MTKVQLYKSL NIHSOKESKI 780
ADIHKKESLV PNDYYLIYKL KVIELLQKV IEAIGETKDE EKIKNAIAKE EQIKKGYNK 839

SEQ ID NO: 38      moltype = AA length = 840
FEATURE
source          Location/Qualifiers
1..840
mol_type = protein
organism = synthetic construct

SEQUENCE: 38
MLKHKRKNKN SLARVVLNSNY DSNNIYEIKI KYEKLAKLDK INIIEMDYDA DNNVMFKVL 60
FNNKEIDLH KDKTCKINEL DNKKYNNISAK KQIGKTHLUV RNKQTSKISR IKKIQDTYYR 120
GKDVFILDNN ILLDQQKQT DKFIVTLNDI TNDKTTSTEALI IDDDTKDIF KKISAKKDLK 180
SSDIYKIKRF ISIRSNFSFY YTFVDNYFKI FHAKKDKNKE ELYKIKFKDE INIKPYLENI 240
LDNMKRNNGI LYNYANDRKV VLNDLNRNIQY VFKEFRHKL HFDYNFLDNF FSNSVEEKYK 300
QKNEIKLLED ILLDNQDSLN VVPKQNYIED ETISVFDAKD IKLKRLYTTT YKLTIYPGF 360
KKLINSFFIQ DGIENQELKE YINNKEKDQV VLKELDNKAY YMDISQYRKY KNIYNKHKEL 420
VSEKELSSDG KKINSLNQKQI NKLKIDMKNI TKPNALNRLI YRLRVAFGFI YKEYATINNF 480
NKSFLQDTKT KRFENISQDQ IKSYLDISYQ DGKFFFVKS KTFKNKTTVK YTfedLDTL 540
NEIITQDDIF VKVIFLFSIF MPKELNGDFP GFINMYYHMK KNISYDTKDI DMLDTISQNM 600
KLKILEQNIK KTYVFKYYLD LDSSIYSKLV QNIKITEDSKYKLYAKIF KYYQHLYKLI 660
SDVEIYLLYK YNSKENLSIT IDKDELKHRG YYNFQSLLLIK NNINKDDAYW SIVNMRNNL 720
HQNIDELVGH FCKGCLRKST TDIAELWLRK DILITITNEII NKIESFKDIK ITLGYDCVND 780
FTQKVKQYKQ KLKASNERLA KKIIEKQNV VDEKNKEELE KNILNMKNIQ KINRYILDIL 840

SEQ ID NO: 39      moltype = AA length = 840
FEATURE
source          Location/Qualifiers
1..840
mol_type = protein
organism = synthetic construct

SEQUENCE: 39
MLKHKRKNKN SLARVVLNSNY DSNNIYEIKI KYEKLAKLDK INIIEMDYDA DNNVMFKVL 60
FNNKEIDLH KDKTCKINEL DNKKYNNISAK KQIGKTHLUV RDQTSKISR IKKIQDTYYR 120
GKDVFILDNN ILLDQQKQT DKFIVTLNDI TNDKTTSTEALI IDDDTKDIF KKISAKKDLK 180
SSDIYKIKRF ISIRSNFSFY YTFVDNYFKI FHAKKDKNKE ELYKIKFKDE INIKPYLENI 240
LDNMKRNNGI LYDYADDREK VLNDLNRNIQY VFTEFRHKL HFDYNFLDNF FSNSVTDQYK 300
QKNEIKLLED ILLDNQDSLN VVPKQNYIED ETISVFDAKD IKLKRLYTTT YKLTIYPGF 360
KKLINSFFIQ DGIENQELKE YINNKEKDQV VLKELDNKAY YMDISQYRKY KNIYNKHKEL 420
VSEKELSSDG QKINSLNQKQI NKLKIDMKNI TKPNALNRLI YRLRVAFGFI YKEYATINNF 480
NKSFLQDTKT KRFENISQDQ IKNYLDISYQ DGKFFFVKS KTFKNKTTIK YTfedLDTL 540
NEIITQDDIF VKVIFLFSIF MPKELNGDFP GFINMYYHMK KNISYDTKDI DMLDTISQNM 600
KLKILEQNIK KTYVFKYYLD LDSSIYSKLV QNIKITEDSKYKLYAKIF KYYQHLYKLI 660
SDVEIYLLYK YNSKENLSIT IDKDELKHRG YYNFQSLLLIK NNINKDDAYW SIVNMRNNL 720
HQNIDELVGH FCKGCLRKST TDIAELWLRK DILITITNEII NKIESFKDIK ITLGYDCVND 780
FTQKVKQYKQ KLKASNERLA KKIIEKQNV VDEKNKEELE KKILNMKNIQ KINRYILDIL 840

SEQ ID NO: 40      moltype = AA length = 849
FEATURE
source          Location/Qualifiers
1..849
mol_type = protein
organism = synthetic construct

SEQUENCE: 40
MSQLKNPSNK NSLPRIIISD FNEIKINEIK IKYHKLDRLL KIIVKEMEII NNKIFFKKIL 60
FNNQNIKDIIS ENIELENYIL AGEVKPSNTK IIILNRDGKEK SFIVYDGFTF KYKPNDKRIS 120
ETKTKTNAKYIL TIKDKTRHRE SSTQRDILKS SIETYQKIS GFENITSKDI YTICKYIDFK 180
NEMMFYYTFTI DFFFFPITGK NKQDKKNNFY NYKIKENAKK FISLINYRIN DDFKNKNGIL 240
YDYLNSKEEI IINDFIHQI 1LKDVRAIA HFNFDFIQKL FDNEQAFNSK FDGIEILNL 300
FNQKQEKYFE AQTNYIEEET IJKILDEKELS FKKLHSFYSQ ICQKKPAFNK LINSPIQDG 360

```

-continued

---

IENKELKDYI	SQKYNSKFDY	YLDIHTCKIY	KDIYNQHKKF	VADKQFLENQ	KTDGQKIKKL	420
NDQINQLTKT	MNNLTCKNSL	KRLEIKFRLA	FGRIFTEVQF	FKNFNERFIE	DIKANKYSTK	480
IELLDYKGK	EYISITHEEK	RFFNYKTFNK	CTKNKNINKTI	FQSLEKEETFE	NLVKNDNLIK	540
MMPLFQLLLP	RELGEBFLGF	ILKTYHDLKN	IDNDTKPDEK	SLSELNISTA	LKLKLVLVNI	600
RQINLFNTYI	SNNTKYEEKE	KRFYEEGNQW	KDIYKKLYIS	HDFDIFDIHL	IIPIIKYNIN	660
LYKLIBDFEV	YLLLKYLERN	TNYKTLDKLI	EAEELKYKGY	YNFTTLLSKA	INITALNDKEY	720
HNITHLRNNT	SHQDIQNIIS	SFKNNKLLEQ	RENIIELISK	ESLKKKLHFD	PINDFTMKT	780
QLLKSLEVHS	DKSEKIEENLL	KKEPLLPNDV	YLLYKLKGIE	FIKKELISNI	GITKYEEKIQ	840
EKIAKGVEK						849

SEQ ID NO: 41            moltype = AA    length = 822

FEATURE                    Location/Qualifiers  
source                    1..822  
                          mol\_type = protein  
                          organism = synthetic construct

SEQUENCE: 41

MVKNPANRHA	LPKVIISEVD	NNNILEFKIK	YEKLARLDKV	EVKSMHFDDN	KQVFDEVVI	60
NGGLIEPTYE	DKHKKLTVTA	GEKSYSIVCG	KVGGKPRLL	DRVSKTKVQL	ELTNYVEDKE	120
GKKRVSKTER	ELIVADNIEL	YSQIVGREVK	TTKEIYLIK	FLEYRSDDLF	YYGFVDNFFK	180
VAGNGKELWK	IDFTNSDSLH	LIEYFKFSIN	DNLKNDENYL	KNYVSDNTKI	ENDLVKCQNN	240
FNSLRHALMH	FDYDFFEKF	NGEDVGPFDF	IEFLNIMIDK	VDKLNIDTKK	EFIDDEETVL	300
FGBALSKKL	YGLFSHIAIN	RVAFNKLINS	FIIEDGIENK	ELKDFFNNKK	ESQAYEIDIH	360
SNAEYKALYV	QHKKLVMATS	AMTDGDEIAK	KNQEISDLKE	KMKVITKENS	LARLEHKRL	420
AFGFIYTYKEV	DYKTFKKHF	QDIKGAKYKG	LNVKEKLKEYY	ETTLKNSKPK	TDEKLEDVAK	480
KIDKLSLKEL	IDDDTLLK	LLLFFMPQ	LKGDFLGFIK	KYYHDKHHID	QDTKDKDTEI	540
EELSTGLKLK	VLDKNIRSLS	ILKHSFSFQV	KYNRKDKNFY	EDGNLHGKFY	KKLSISHNQE	600
EFNKSVYAPL	PRYYSALYKL	INDFEYALALA	QHVNENHETLA	DQVNKSQFQ	KSYFNFRRKLL	660
DNTDSISQSS	SYNTLIVMRN	DISHLSYEP	FNYPLDERKS	YKKKTQGVK	TFHVHELLYIS	720
RAKIELISL	QTDMKLLGY	DAVNDFNMKV	VHLRKRLSVY	ANKEESIRKM	QADAKTPNDF	780
YNIYKVKGVE	SINQHLLKVI	GVTEAEKSIE	KQINEGNKKH	NT		822

SEQ ID NO: 42            moltype = AA    length = 806

FEATURE                    Location/Qualifiers  
source                    1..806  
                          mol\_type = protein  
                          organism = synthetic construct

SEQUENCE: 42

MTKKPSNRNS	LPKVIINKVD	ESSILEFKIK	YEKLARLDRF	EVRSRMRYDGD	GRIIFDEVVA	60
NAGLLDWDY	DDNRTIVVKI	ENKAYNIYKG	KVGGEKRLNG	KISKAKVQLI	LTDSIRKNA	120
DTHRHSLLTER	ELINKNEVVL	YSKIAKRE	TTKDIYLVKR	FLAYRSDDLL	YYAFINHYVR	180
VINGNKKFWK	TEIDDDKIIDY	FIYTINDTLK	NKEGYLEKYI	VDRDQIKKDL	EKIKQIFSHL	240
RHKLMHYDFR	FTTDLFDGKD	VDIKVDNSIQ	KISELLDIEF	LNIVIDKLEK	LNIDAKKEFI	300
DDEKITLFGQ	EIELKKLYSL	YAHTSINRA	FNLKINSFLI	KDGVENKELK	EYFNNAHNQGK	360
ESYYIDIHQN	QEYKKLYIEH	KNLVAKLSSAT	TDGKEIAKIN	RELADKKEQM	KQITKANSLK	420
RLEYKLRLAF	GFIYTYKEV	ERFKNSFDT	TKKKKFDAID	NAKIIHEYFA	TNSKAKKIEKL	480
EEILKGIDKL	SLKTLIQQDI	LLKFLLLFFT	FLPQEIKGFE	LGFIIKKYHD	ITSLDEDTKD	540
KDEDEITELPR	SLKLKIFFSK	IRKLSILKHS	LSYQIKYMKH	ESSYYEAGNV	FNKMPKKQAI	600
SHNLEEFGKS	IYLPMLKYY	ALYKLINDE	IYALYKDMDT	SETLSQVDK	QEYKRNEYFN	660
FETLLRKFGF	NDIEKVLVTY	RNKAHLDFN	FLYDKPINKF	ISLYKSREKI	VNYIKNHDIQ	720
AVLVYDAVND	FVMKVIQLRT	KLKVYADKEQ	TIESMIQNTQ	NPNGFYNIYK	VKAVENINRH	780
LLKVIYGTE	EKAVEEKIRA	GNTSKS				806

SEQ ID NO: 43            moltype = AA    length = 832

FEATURE                    Location/Qualifiers  
source                    1..832  
                          mol\_type = protein  
                          organism = synthetic construct

SEQUENCE: 43

MIKNPNSRNYA	LPKVIISKID	NQNILEFKIK	YKKLSKLDIV	KVKSMDYDDR	AIIFDEVIVN	60
DGLIDVEYRD	NHKTIFVKVG	NHKTIFVKVG	VGGKERLLEN	RVSXTKVLQE	LKDKAATNRVS	120
KTRRELIVDD	NIKIIYSQIVG	RDVKTTKDIY	LIKRFLAYRS	FLFYYGFVN	NFFFHVANNRS	180
EFWKIDFNDS	NNSKLIEYFK	FTINDHKLND	ENYLKDYISD	NEKLKNDLIK	VKNNSFEKIRH	240
ALMHFDYDF	VKLFNGEDVG	LELDIEFLDI	MIDKLDKLNI	DTKKEFIDE	KITIFGEELS	300
LAALKYRFYAH	TAINRVAFNK	LINSFIIENG	VENQSLKEYF	NQOAGGIAYY	IDIHQNREYK	360
NLYNEHKKLV	SRVLSLSDQG	EIAILNQKIA	KLKDQMKQIT	KANSIKRLEY	KLRLALGFIY	420
TEYENYHEEFK	MNFDTDIKNG	RFTPDKNDGN	KRAFDSRELE	QLKGYYEATI	QTQKPKTDEK	480
IEEVSKKIDR	LSLKLSLIADD	ILLKFILLMF	TFMPQELKGE	FLGFIKKYH	DTKHIDQDTI	540
SDSDDTIELT	SIGLKLKILD	KNIRSLSILK	HSLSFQTKYN	KKDRNYYEDG	NIHGKFKKL	600
GISHNQEEFN	KSVYAPLFRY	YSALYKLIND	FEIYTLSLHI	VGSETLTDQV	NKSQFLSGRY	660
FNFRKLLTQS	YHINNNNSTHS	TIFNAVINMR	NDISHLSYEP	LFDCPLNGKK	SYKRKIRNQF	720
KTINIKPLVE	SRKIIIDFIT	LQTMQKVLG	YDAVNDFTMK	IVQLRTRLKA	YANKEQTIQK	780
MITEAKTPND	FYNIYKVQGV	EEINKYLLEV	IGETQAECIE	REKIERGNIA	NF	832

SEQ ID NO: 44            moltype = AA    length = 1281

FEATURE                    Location/Qualifiers  
source                    1..1281

-continued

---

```

mol_type = protein
organism = synthetic construct

SEQUENCE: 44
MKKSIFDQFV NQYALSKTLR FELKPVGETG RMLEEAKVFA KDETIKKKYE ATKPPFNKLH 60
REFVEEALNE VELAGLPEYF EIFKYWKRYK KKFEKDLQKQ EKELRKSVVG FFNAQAKEWA 120
KRYETLGVKK KDVGGLFEEEN VFAILKERYG NEEGSQIVDE STGKDVSIFD SWKGPTGYFI 180
KFQETRKNFY KDDGTATALA TRIIDQNLKR FCDNLLIFES IRDKIDFSEV EQTMGNSIDK 240
VFSVIFYFSC LLQEGIDFYN CVLGGETLPN GEKRQGINEL INLYRQKTSF KVPFLKLLDK 300
QILSEKEKFM DEIENDEALL DTLKIFRKSA EKETTLLKNI FGDFVMNQGK YDLAQIYISR 360
ESLNTSRKWW TSETDIFEDS LYEVLLKKSII VSASVKKKDQ GYAFPEFIAL IYVKSALEQI 420
PTEKFWKERY YKNIGDVLNK GFLNGKEGVW LQFLLIFDPE FNSLFEREII DENGDKKVAG 480
YNLFAKGFD LLNNNFKYDQE AKVVIKDFAD EVLHIYQMGK YFAIEKKRSW LADYDIDSFY 540
TDPEKGYLKF YENAYEEEIQ VYNKLRNYLT KKPYSEDWKW LNENPNTLAD GWDKNKEADN 600
STVILKDDGR YVLGLMARGE NKLKFLDRNLK KILEGVENGK YEKVVVYKYPF DQAKMFPKVC 660
FSTKGLEFFQ PSEEVITIYK NSEFKDGRN VNRSMQRLLI FYKDCLVRYE GWQCYDFRNL 720
RKTEDYRKNI EEEFFSDVAMD GYKISFQDVS ESYIKEKNQN GDLYLFEDIKU KDWNEGANGK 780
KMLHTIYFES LFSADNIAMN FPVKLNGQAE IFYPRPRTEGL EKERIITKKG NVLEKGDKAF 840
HKRRTENKV FFHVPIITLNK TCKNPQFQNA KINDFLAKNS DINVIGVDRG EKQLAYFSVI 900
SQRGKILDRC SLNVINGVN YAEKLEKARG REQARKDWQO IEGKDLKKG YISQVVRKLA 960
DLAIQYNAII VFEDLNMRFF QIRGGIEKSV YQQLEKALID KLTFLVEKEEE KDVEKAGHLL 1020
KAYQLAAPFE TFQKMGKQTG IVFYTQAAYT SRIDPVTGWR PHLYLKSSA EKAKADLKF 1080
KKIKFVDGRF EFTYDIKSFR EQKEHPKATV WTVCSVERF RWNRYLNSNK GGYDHYSVDT 1140
KFLVELFQEY GIDFERGDIV GQIEVLETKG NEKFFKNFVF FFNLICQIRN TNASELAKKD 1200
GKDDFILSPV EPFFDSRNSE KFGEDLPKNG DDNGAFNIAR KGLVIMDKIT KFADENGGE 1260
KMKGWGDLYVS NVEWDNFVAN K 1281

SEQ ID NO: 45 moltype = AA length = 1280
FEATURE Location/Qualifiers
source 1..1280
mol_type = protein
organism = synthetic construct

SEQUENCE: 45
MFNNFIKKYS LQKTLRFELK PVGETADYIE DPKSEYLNKDT VLKDEQRAKD YQEIKTLIDD 60
YHREYIEECL REPVDKKTGE ILDFTQDLED AFSYYQKLKE NPTENRVGWE KEQESLRKKL 120
VTSFVGNDGL CKEKFEITRDL PEWLQKKGLW GEYKDTVENF KKFTTYFSGF HENRKNMYTA 180
EAQSTAIAAN LMNDNLPKFF NNLYLAYQTIK EKHPLDLVRL DALLQAAVG EHLDEAFQPR 240
YFSRLFAQSG ITAFNELIGG RTTENGEKIQ GLNEQINLYR QQNPEKAKGF PRFMPFLFKQI 300
LSDRETHSFL PDAFENDKEL LQALRDYVA ATSEEGMISQ LNKAMMNQFVT ADLKEVYIKS 360
AALTSLSQEL PHFFGVISDA IAUYAEKRLS PKKAQESFLPK QEVYIAEELN QAVVGYIDQL 420
EDQSELQQLL VDLPDPQPKV SSFILTHWQK SQEPLQAVIA KVEPLFELEE LSKNKRAPKH 480
DKDQGGEFGQ QVDAIKNMLD AFMVEWSAIK PLYLVKGKRA IDMPDVDTGF YADPAEAYSA 540
YEQVTVSLYN KTRNHLSSKKP FSKDKIKINF DAPTLNNGWD LNKESEDNPKSI ILRKDGDNFYL 600
AIMHPKHTKV FDCYASASEAA GKCYKEKMNPK LLGANKMPL KVFFSKKGIE TFSPPQEILD 660
LYKNNEHKKG ATFKLESCHK LIDFFKRNPI KYKVHPTDNF GWDVFGFHES PTSSYGDLSG 720
FYREVEAQGY KLWFSDVSEA YINKCVEEGK LFLFQIYNDK FSPNSTGKPN LHTLYWKGLF 780
EPENLKDVLV KLNGEAEIFY RKHSIKHEDK TIHRAKDPIA NKNADNPKKQ SVFDYDIKID 840
KRYTQDKFFF HVPISLNFKS QGVVRFNDKI NGLLAAQDQDV HVIGIDRGER HLLYYTVVNG 900
KGEVVEQGSQ NQVATDQGYV DVYQOKLHAK EKERDQARKN WSTIENIKEL KAGYLSQVWH 960
KLAOLIVKHN AIVCLEDLNF GFKGRGRFKEV KQVYQKFEKA LIDKLNLYLVF KERGATQAGG 1020
YLNAYQLAAP FESFEKLKGQ TGILYYVRSR YTSKIDPATG FVDFELPKYE SMAKSKVFFE 1080
SFERIOWNQA KGYFEPEFDY KKMCPSRKPG DYTRTRWVCT FGDTRYQNRN NKSSQGWETE 1140
TIDVTQALKA LFAAYGITYN QEDNIKDAKA AVKYTKFVQK LYWLRLRTL LRHSVTGTDE 1200
DFILSPVADE NGVFFDSRKA TDQPKDADA NGAYHIALKG LWNLQQQIRQH DWNVKEPKKL 1260
NLAMKNEEFW GFAQKKFRA 1280

SEQ ID NO: 46 moltype = AA length = 832
FEATURE Location/Qualifiers
source 1..832
mol_type = protein
organism = synthetic construct

SEQUENCE: 46
MIKNPSNSRYA LPKVIISKID NQNLIEFKIK YKKLSKLDIV KVKSMDYDDR AIIFDEVIVN 60
DGLIDVEYRD NHKTIFVKVG NKSYSISGOK VGGKERLLEN RVSKTKVQLE LKDKATNRVS 120
KTERELIVDD NIKIYSQIVG RDVKTTKDIIY LIKRFPLAYS DLLFYYGFVN NFFHVANNRS 180
EFWKIDFNDS NNIKISLIEYFK FTINDHLKND ENYLKDYISD NEKLKNDLIK VKNSFEKIRH 240
ALMHFDYDFV VKLFLNGEDVG LEDLIEFLDI MIDKLDKLN1 DTKKEFIDE KITIFGEELS 300
LAKLYRFAHY TAINRVAFNK LINSFIIENG VENQSLKEYF NQQAGGIAYE IDIHQNREYK 360
NLVNEHKKLV SRVLSISDGQ EIAILNQKTA KLKDQMVKQIT KANSIKRLEY KLRLALGFIY 420
TEYENYEEFK MNFDTDIKNG RFTPDKNDGN KRAFTDSRELE QLKGYYEATI QTQKPKTDEK 480
IEEVSKKIDR LSLKSLSIADD ILLKFLILLMF TFMPQELKGE FLGFIKKYH DTKHIDQDTI 540
SDSDDTIELT SIGLKLKILL KNIRSLSILK HSLSFQTKYN KKDRNYYEDG NIHGKFFKLL 600
GISHNQEENF KSVYAPLFRY YSALYKLIND FEIYTLSLHI VGSETLTDQV NKSQFLSGRY 660
FNPRLKLLTQS YHINNNNSTHS TIFNAVINMR NDISHLSYEP LFDCPLNGKK SYKRKIRNQF 720
KTINIKPLVE SRKIIIDFIT LQTDMQKVLG YDAVNDFTMK IVQLRTRLKA YANKEQTIQK 780
MITEAKTPND FYNIYKVQGV EEINKYLLEV IGETQAEKEI REKIERGNIA NF 832

```

-continued

---

SEQ ID NO: 47	moltype = AA length = 809
FEATURE	Location/Qualifiers
source	1..809
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 47	
MIKNPSNRHS LPKVIISEVD HEKILEFKIK YEKLARLDRF EVKAMHYEGK EIVFDEVLVN 60	
GGLIEVEYQD DNKTLFVKG EKSYSIRGKK VGGKQRLLLED RVSKTKVQLE LSDGVVDNKG 120	
NLRKSRTERE LIVADNIKLY SQIVGREVTT TKEIYLVKRF LAYRS DLLFY YSFVDNFVKV 180	
AGNEKEWLKI NFEDDATSAQF MGYFPMVND NLKNDNAYLK DYVRNDVQIK DDLKKVQTIF 240	
SALRHTLLHF NYEFFEKLFN GEDVGDFDII GFLNLLIENI DKLNIDAKKE FIDNEKIRLF 300	
GENLSLAKVY RLYSDICVNP VGPNKFINSM LIKDGVENVQ LKAEOFNRKFG GNAYTIDHS 360	
NQEYKRIYNE HKKLVIKVST LKDQGAIIRR NKKISELKEQ MKSMTKNSL ARLECKMRLA 420	
FGFLYGEYNN YKAFKNNFDT NIKNQSDFVN DVEKSKAYFL STYERRKPERT REKLEKVAKD 480	
IESLELKTVT LMFVFMQPEL KGDFLGFKV YYHDVHSIDD DTKEQEEDVV 540	
EAMSTSLLKLK ILGRNIRSLT LFKYALSSQV NYNSTDNIFY VEGNRGKLY KKLGI SHNQE 600	
EFDKTLVVP LRYYSSLFKL MNDFEIYSLA KANPTAVSLO ELVDDETSPY KQGNYFNFK 660	
MIRDYGLTS DEIKSGQVVF MRNKIAHFTD EVLLSKPLLG QTKMNQLRKD IVSFIEARGD 720	
IKELLYGDAI NDFRMKVHL RTKMRVYSDK LQTMMDLLRN AKTPNDFYNV YKVKGVESIN 780	
KHLLEVLAQT AEERTVEKQI RDGNEKYD 809	
SEQ ID NO: 48	moltype = AA length = 1285
FEATURE	Location/Qualifiers
source	1..1285
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 48	
MEEKMLKSYD YPTKLYSLQK TLRFELKPIG KTLEHIKNSS IIESTDETLEE QYAIVKNIID 60	
KLHRKHIDEA LSLVDFTKHL DTLKTFQELY LKRKTDKEK EELEKLSADL RKLIVSYLKG 120	
NVKEKTQHNL NPIKERFEIL FGKELTFTNEE FFLLAENEKE KKAIQAFKGF TTYFKGQFQN 180	
RKNMYSSEGN STSIAYRIIN ENLPLFIENI ARFQKVMSTI EKTTIKKLEQ NLKTELKHN 240	
LPGIFTIEYF NNVLTQEGIS RYNTIIGGKT THEGVKIQGL NEIIINLYNQQ SKDVKLPILK 300	
PLHKQILSEE YSTSFKIKAF ENDNEVLKAI DTPWNEHIEK SIHPVTGNKF NILSKIENLC 360	
DQLQKYKDQD LEKLFLIERKN LSTVSHQVVG QWNIIIRDALR MHLEMNNKNI KEKDIDKYLD 420	
NDAFSWEKIE DSIIKIYKEH EDAKELNENG IIKYFSAMSII NEEDDEKEYS ISLIKNINEK 480	
YNNVKSILQD DRTGKSDLHQ DKEVKGIIKE FLDSLKQLQW FLRLLYXVTVP LDEKDYE FYN 540	
ELEVYYEALL PLNSLYNKVR NYMTRKPYSV EKPKLNFNSP TLLDGWDK NK ETANLSIILR 600	
KNGKYYLGIM NKENNTIFYEY YPGTKSNDYY EKMIYKLLPG PNKMLPKVFF SKKGLEYYNP 660	
PKEILNIIYEK GEFKKDKSGN FKKEKSHTLII DFYKEAIAKN EDWEVFNFKF KNTKEYEDIS 720	
QFYRDVEEQG YLITFVKVDA NYVDKLVSEG KLYLFQIYMK DFSENKSKKG NPNIHLTIYWK 780	
GLYDSENLNK VVYKLNGEAE VFYRRKKSIDY PEIYNHGHG KEELLGKFNY PIICKDRRTQ 840	
DKPLFHVPIT MNFISKEEKR VNQOLACEYLS ATKEDVHIIG IDRGERHLLY LSLIDKEGNI 900	
KKQLSLNTIK NENYDKEIDY RVKLEDEKEKK RDEARKNWDV IENIKELEG YMSQVIHITA 960	
KMMVEEKAIL IMEDLNIGFK RGRFKVKEQV YQKFEKMLID DGFLVFKNK NPLEPGGSLN 1020	
AQYOLTSKFD S PKKLKGQSGF IFYVPSAYS T KIDPTTGFPY FIQVDVPNLE KGKEFFSKFE 1080	
KIIYNTKEDY FEFHCKYKGKF VSEPKNKDND RKTKESTYY NAIKDVTWWV CSTNHERYKI 1140	
VRNKAGGYES HPVDVTKNL DIFSQANINY NEGKDIKPII IESNNAKLLK SIAEQLKLIL 1200	
AMRYNNNGKG DDEKDYLSP VKNQKGKFFC TLDGNQTLPI NADANGAYNI ALKGLLLIEK 1260	
IKKQQGKIKD LYISNLEWFM FMMR 1285	
SEQ ID NO: 49	moltype = AA length = 1367
FEATURE	Location/Qualifiers
source	1..1367
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 49	
MEKSLNDFIG LYSVSKTLRF ELKPVSETLE NIKKFHFLEE DKKKANDYKD VKKIIDNYHK 60	
YFIDDVLKNA SFNWKKLEEA IREYNKNKSD DSALVAEQQK LGDAILKLFT SDKRYKALTA 120	
ATPKELFESI LPDWFGEQCN QDLNKAALKT FQKFTSYFTG QOENRKNVYS AEA IPTAVPY 180	
RIVNDNFPDFK LQNVLIFKTI QEKCPLQIIDE VEKELSSYLG KEKLAGIFTL ESFNKYLQGQ 240	
GKENQRGQDFK YNQIIGGVVE KEGGINLRLVQ NQFLNLYWQO HPDFTKEDRR IKMVPLYKQI 300	
LSDRSSLSFK IESIENDEEL KNALLECADK LELKNDDEKKS I FEEVCDLFS SVKNLDSLGI 360	
YINRKDINSV SRILTGDWSW LSQSRMNVYAE EKPTTAKAEGA RWQKSLDDEG ENKSKGFYSL 420	
TDLNEVLEYS SENVAETDII ITDYFEHCR YYVDKETEM VQGSELVALS LQEMCDDILK 480	
KRKAMNTVLL NLSSENKLRE TTDVVAVIKE YLDAVQELLH RIKPLKVNGV GDSTFYSVYD 540	
SIYSALSEVI SVYNKTRNYI T KKAASPEKY KLNFDNPTLA DGWDLNKEQA NTSVILRKDG 600	
MFYLGIMNPK NPKPFAEKYD CGNESCYEK IYKQFDATQK IPKCSQKKE VQKYFLSGAT 660	
EPIYILNDKKS FKSELJITKD IWFMMNHVWD GEKFVPKRDN ETRPKKEQIG YFKQTGDFDG 720	
YKNALSNWIS FCKNPLQSYL SATVYDYNFK NSEEEYEGLDE FYNYLNATCV KLNFINIPET 780	
EINKMVSEGK LYLFQIYKND FASGSTGMN MHTLYWKNLF SDENLKNVCL KLNGEAELFY 840	
RPAGIKEPVI HKEGSYLVNR TTEDGESIPE KIYFEIYKNA NGKLEKLSDE AQNYIISNEV 900	
VIKKAGHEII KDRHYTEPKF LFHVPLTINF KASGNSYIN ENVRKFLKNN PDVNIIGLDR 960	
GERHHLIYSL INQKGRIIKQ FTFNEVERNK NGRTIKVNY EKLDQREKER DAARKSWQAI 1020	
GKIAELKEGY LSAVIHQLTK LMVEYNAAVY MEDLNFGFKR GRFHVEKQVY QKFEHILIDK 1080	
SYLVLFKDRG LNEPGGVNLNG YQIAQGFESF QKLKGQSGML FYVPAGYTSK IDPKTGFSVSM 1140	
MNPKDLTNVH KKRDFFSKFD NIHDEANGS FVPTFDYKKF DGKAKEEMKL TKWSVYSRDK 1200	

-continued

---

RIVYFAKTKS YEDVLPTEKL QKIFESNGID YKSGNNIQDS VMAIGADLKE GAKPSKEISD	1260
FWDGLLSNFK LILQMNRNSNA RTGEDIYISP VMADDGTFDD SREEFKKGED AKLPLDADAN	1320
GAYHIALKGL SLINKINLSK DEELKKFDMK ISNADWFKFA QEKNYAK	1367
 SEQ ID NO: 50	
FEATURE	moltype = AA length = 1366
source	Location/Qualifiers
	1..1366
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 50	
MNTQKKEFPN KSFKDFTNL Y QNLSKTLRFLS TPNKKTABEL EFNKQKEVKC FSNDRKIAGA	60
YQBIKKYLNK LHQEPIQEA KFFAFSEEEL KGPFKEYLNL LNFTDKDNFV KKNKIRNEYE	120
QERKILTIKI ATYFSKFSE KYQSFNLANI TGKKVFSILE QKYKEDKKT KIIHIFKYKP	180
TDKEKKEGEA VNFSTYLTG NENRKNFYKS EDKAGQFATR TIDNLAQFIK NKKLFEDKYQ	240
KNYSKIGILD EQIKIFNLDY FNNLFLQEGL DEYNGILGNM KGEENKSNEG INQKINIFKQ	300
KEKARLKKEK ENFNKSDFPL FKELYKQIGS IRKENDVYVE IKTDKELVVE LNNFPKNVEN	360
YLKDIQSFYK TFFEKLQNEE YELDKIYLPK SVGTYFSYIA FSDWNKLAIFI YNKRYKNEKI	420
KIVEGGDVNV QYRSLEVLNK RIDELKDEDN LNPNKFFFIDK LKFNEAKKEN NWQNFWFCIE	480
YYINSQFIGG EKNILNKEKN EYEILPFGSL KELKEKYFEA VKKYKEKMD TESGLTDDEE	540
KEIKEKLKNY LDRLIKEIERI AKYFDLKKSF EEEIKQEDLDS NYFGEQKVV DKTNELKIQ	600
YYSEFRNYLT QNNSVEEKIK LNFNSGLLLD GWDLNKEKVK FSIIFQENGK YYLGIINKEK	660
DKTILDKDH PEIFTKNSDFP RKMELYKLFPS PSKMLPKISF SETAKKGDED VGWSSEEIQL	720
KDEFAEFQDYE KKKSKDNWKD EFNPKLNLK IDYYKQVLEK HSEGMNTYN FELKDSKSYK	780
NLGEFNDNA RQNYKVKFVG IDKNYIDEKV ANGELFLFQI YNKDFSEDKK EGSTNNLETI	840
YFKELFSKEN LENPVFKLSG GAEMFFRNKI EKKKEKKKLD KDGKPMISKK GEKVVNDKRR	900
SENKILPHP LEINYKGKGM PNFNKRNKINEY ISKNPENIKI IGIDRGEKHL LYYSIIDQNG	960
NNIESMSLNA VDEFGNFVNPE EKLEEEYIDN NGKKERRWYK IVNDKEIKVT NYQRKLDELE	1020
KERQKSRSOW QNINKIKNLK KGYSIVFVVK IVDLAIENNA IIILEDLNFG FKSFQRKIEK	1080
NVYQQFEKAL IDKLGFFVVDK QKQNQRFAPQ LSAPFESFQK IGKQTGIVYY VLAMNTSKVC	1140
PSCQWIKNFY LKYEKKTNTF NLQKNQNLKV FFEQEKNRFR FEYQMSKEYI SVYSDVDRQR	1200
YDKTKNQNKG GLEYKNSNQ KEIIDKDGVL QKQSITLQLK ELFKENHIDL EKEILKQLDN	1260
KKEKNSGYTG VYNKFIYLFN LILQIRNAIS FREKDYIQCP SCHFDTKREN YLKINDGDGN	1320
GAYNIALRGL YLLKGNGII NNLEKIKLIF SNNDYFQWAK KLKNKK	1366
 SEQ ID NO: 51	
FEATURE	moltype = AA length = 1275
source	Location/Qualifiers
	1..1275
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 51	
MQNQKQSFADF TNLYSLSKTL RFELKPIGQT QAMLDENKIF EVDENRKAY DKTYPFDRL	60
HREFINESLS NAQLKGISEY FETFKQFRSN QNNKDLKELI NKQQKFLRHQ IVTLPFDENGK	120
HWATTKYAHL KIKKKNLDL FDEQVFYILK ERYGSEKETQ LVDKETGAVT SIFDNWKGFT	180
GYFTKFFETR KNFYKSDGTS TALATRIIDQ NLRNFFFDNL ETPHKIKDKID VKEVEIFFKL	240
KADNVFSIDF YNQCLLQNGI DKYNDFLGGC TLENKEQKQG INEIINKYRQ DNKQKLPLF	300
KKLDKQILSE KDRFINEIES KEEFFQVLT E FYQSATVKV IIKTLNDVF HNTDKYKLEK	360
IYLTKEAFNT IANKWTDETQ IFEDNLDLVL KNKCIITAKQD FIPLAYIKEA LEVIEKDRKF	420
FKDRYYNDPQ IGFFPDQSYW EQFLAIIINFE FMTHFQVRAK DKITGKIEL GYFVFEKRIK	480
ELLDSDPSLN SQSKIIKEF ADEVLHIFOM AKYFALEKKR EWKGDYYQLD DQFYHNIDYG	540
FKDQFYENAY EKIVQPYNKI RNYLTKKPYS DVWKLNLNGN PTLANGWDKN KEADNTAVIL	600
KKDGNYLLGV MKKGKKNIFS DQNKEKYKAY NSAYYEKLVY KLFPPDPSKM PKVCFPSKGL	660
NFFQPSEEL RIYKNNEFKK GNTFSISMSQ KLIAYFIDCL GLYEWKHYE FKNIKDVRQY	720
KENIGEYFAD VAESGYKLW EKISEEYITQ KNOLGELFLF QIYNKDFAKK TTGRKNLHTI	780
YFEELFSQTN IDNNNPFKLN GQAELFYRPK SLEKIEEKRN FKRSIVNKCR YTQNKIFPHV	840
PTTLNRTSEN IGRFNVRVNN FLANNNSNVNI VGDRGKELN AYYSIIKQNG EVLKSGSLNI	900
INGVDYHALL TDRAQRREQE RRNWQDVESI KDLKRGYISQ VVHELVS LAI KYNAAIVMED	960
LNNRFQKIRG GIEKSTYQQL EKALIEKLN LVNKEETDSN QAGNLLNAYQ LTAPPKTFKD	1020
MGKQTGIFYF TQASYTSKID PLTGWRPNIY LYRYSNAKQAK ADILMFTNIY FSEKKDRFEF	1080
TYDLEKIDDK RKDLPIKTEW TVCSNVERFS WEKSLNNNNKG GYVHYPIQDS NGEEISITSKL	1140
KKLFMDFGID LTDIKTQIES LDTNKKDNAN FFRKFIFYFQ LICQIRNTQV NKSDDGNDFI	1200
FSPVEPFDFDS RFADKFRKNL PKNGDENGAY NIARKGLIL HKISDYFVKE GSTDKISWKD	1260
LSISQTEWDN FTTDK	1275
 SEQ ID NO: 52	
FEATURE	moltype = AA length = 1368
source	Location/Qualifiers
	1..1368
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 52	
MKKEKEFKSF GDFTNLYEIS KTLRFLKPV ENTQTMLEDA DVFGKDKVIK DKYTCKPFI	60
DLKLHREFVDE SLKDVSLSGL KKYSEVLENW KKNKKDKDIV KELKKEEERL RKEVVEFDN	120
TAKKWAENEY KELGLKKKDI GILFEESVFD LLKEKYGEEQ DSFLKEEKGD FLKNEKGEKV	180
SIDFEWKGFV GFYTFKQETR KNFYKNDGTE TALATRIIDQ NLKRFCDNID DFKKKIKNID	240
FSEVEKFNK TADVFSLDFY NQCLLQKGID SYNEFIGGKT LENGKKLKGV NELVNEYRQK	300
NKNEKVSFLK LLDKQILSEK EKLSFGIEND EQLVVVLNSF YETAEEKTKI LRTLFGDFVE	360
HNENYDLDKT YISKVAFNTI SHKWTNETHK FEELLYGAMK EDKPIGLNYD KKEDSYKFPD	420

---

-continued

---

FIALGYLKKC	LNNLDCTKF	WKEKYIYENNA	DKKDCKDGFL	TGGQNAWDQF	LQIFIFEFNQ	480
LFNSEAEDNK	GKEIKIGYDNK	FRKDFEEIIN	QKDFKNDENL	KIAIKNFADS	VLWIYQMAKY	540
FAIEKKRQWD	DDFELSEFYT	NPSNGYSLFY	DRAYEEIVQK	YNDLRLNYLTK	KPYKEDWKWL	600
NFENPTLANG	FDKNKESDNS	TVILRKCRKY	YLGKMKGNM	KIFEDRNKAE	FIRNIESGAY	660
EKMAYKYLPD	VAKMIPKCST	QLNEAKNHFR	NSADDLEIJK	SFSNPLKITK	RIFDLNNIQY	720
DKTNVSKKIS	GDNKGIKIFQ	KEYYKISGDF	DVYKSALNDW	IDFCCKDFLSK	YDSTKDFDFS	780
ILRKTCKYKS	LDEFYVDAVK	ITYKISFTPV	SESYIDQKMN	NGELYLFEIY	NQDFAKGKMG	840
AKNLHTLYFE	NVFSPENISK	NFPKILNGNA	ELFPRPKSIE	SKKEKRNFVR	EIVNKNGRYSE	900
DKIFFHCPIT	LNRETGSIYR	FNNYVNNFLS	ENNINIIGVD	RGEKHAYYS	VIDKNGVKIG	960
GGSFNEINKV	FAKCKLEERA	GEREQSRKDW	QVVEGIKDLK	KGYISQVRE	LADLAIKHNA	1020
IIVLEDLNMR	FQKIRGKIEK	SIYQOLEKAL	IDKLSPFLVEK	GEKDPNQAGH	ILKAYQLAAP	1080
FTSFKDGMQKQ	TGIVFYTQAS	YTSKTCPCNC	FRKNNNNKFYF	ENNIGKAQDA	LKKLKTFEYD	1140
SENKCFGLSY	CLSDFANKEE	VEKKNKNKRN	NAPYSDIEKK	DCFELSTKDA	VRYRWHDKNT	1200
ERGKTFEGFE	SVYEKEEKEE	IGQTKRGLVK	EYDISKCLIG	LFEKTKGLDYK	QNLLDKINSG	1260
KFDGTFYKNL	PNYLNLLEFI	RNSISGTEID	YISCPECQFH	TDKSCTIKNG	DDNGSYNIAR	1320
KGMIIILDKIK	QFKKENGSLD	KMGWGEFLID	LEEWDKFAOK	KNNNIIDK		1368
SEQ ID NO: 53		moltype = AA	length = 1257			
FEATURE		Location/Qualifiers				
source		1..1257				
		mol_type = protein				
		organism = synthetic construct				
SEQUENCE: 53						
MYKTDFTGIY	SVSKTLRFEL	IPQGSTVENM	KREGILNNDM	HRADSYKEMK	KLIDEYHKAF	60
IERCLSDLFSL	KYDDTGKHD	LEEEFFYYEQ	KRNDKTKKIF	EDIQVALRKQ	ISKRFTGDTA	120
FKRLFKKELI	KEDLPSFVKN	DPVKTELKE	FSDFTTYQOE	PHKNRKNMVT	SDAKSTAIAY	180
RIINENLPKF	IDNINAIDIV	AKVPEMQUEHF	KTIADELRSH	LQVGNDIDKM	FNLQFFNKVL	240
TQSQQLDVYNA	VIGGKSEGK	KIQQGNEYIN	LYNQQHKKAR	LPMLKLLYKQ	ILSRVAISW	300
LQDEFDNDQD	MLDTIEAFYN	KLNSNETGVL	GEGKLKQILM	GLDGYNLDGV	FMRNDLQLSE	360
VSQRLLCGGW	IJKDAMTSNL	KRSVQKKKE	TDADFEERS	KLFSAQNSFS	IAYINQCLGQ	420
AGIRCKIODY	FAKCKLEERA	NEAETPPDW	DQIAEAYHGA	APILNARPSS	HNLAQDIEKV	480
KAIAKALLDAL	KRLQRFVKPL	LGRDGEKGDK	NFFYGDMPFI	WEVLDQLTPL	YNKVRNRMTR	540
KPYSQEIKIKL	NFENSTLLNG	WDLNKEHDNT	SVILRRREGY	YLGIMKNYN	KIFDANNVET	600
IGDCYERKMI	KLLPGPNKML	PKVFFSKSRV	QEFSPSKKIL	EIWESKSFKK	GDNFNLDCH	660
ALIDFVYKDSI	AKHPDWKFN	FKPSDTSQSYT	NISDFYRDVN	QQGYSLSFTK	VSVDYVNRMV	720
DEGKLYLFQI	YNKDFSPQSK	GTPNMHTLYW	RMLFDERNLH	NVIYKLNGEA	EVFYRKASLR	780
CDRPPTPAHQ	PITCKENDS	KRVCVFDYDI	IKNRRYTVDK	FMFHVPITIN	YKCTGSDNIN	840
QQVCDYLRSA	GDDTHIGID	RGERNLLYLV	IIDQHGTIKE	QFSLNEIVNE	YKGNTYCTNY	900
HSLLEEEAKG	NKKARQDWQT	IESIQLKQH	YLSQVIHKIS	MLMQRHYAIV	VLEDLNGSF	960
RSRKVKEQV	YQKFEHMLIN	KLNYLVNKQY	DATEPGGLLH	ALQLTSRMDS	FKKLKGQSGF	1020
LFYIPAWNTS	KIDPVTFGVN	LFDTRYCNEA	KAKEFFEKF	DISYNDERDW	FEFSFDYRH	1080
TNKPTGTRTQ	WTLCTQGTRV	RTFRNPEKS	HWDNEEFDLT	QAFKDLFNKY	GIDIASGLKA	1140
RIVNGQLTKE	TSAVKDFYES	LLKLLKLTQ	MRSVTGTDI	DYLVSPVADK	DGIFFDSTRC	1200
GSLLPANADA	NGAFNIARKG	LMLLRQIQQS	SIDAEKIQLA	PIKNEDWLEF	AQEKPYL	1257
SEQ ID NO: 54		moltype = AA	length = 766			
FEATURE		Location/Qualifiers				
source		1..766				
		mol_type = protein				
		organism = synthetic construct				
SEQUENCE: 54						
MEKEITELTK	IRREFPNKKF	SSTDMMKAGK	LLKAEGPDAV	RDFLNSCQEI	IGDFKPPVKT	60
NIVSISRPFE	EWPVSMVGRA	IQEYFFSLTK	EELESVHPGT	SEEDHKSFFN	ITGLSNYYT	120
SVQGLNLIFK	NAKAIYDGT	VKANNKNKL	EKKFNEINHK	RSLEGLPLIT	PDFEEPFDEN	180
GHLLNNPPG	RNIYGYQGCA	AKVPPVSKHK	MVSLPKKEYEG	YNRDPNLSLA	GFRNRLEIPE	240
GEPGHNPWFQ	RMDIPEQGQ	VHNKIQRDFN	VHGKNSGKVK	FSDKTGRVKR	YHHSKYKDAT	300
KPYKFLEESK	KVSALESLA	IITIGDDWWV	FDIRGLYRN	FYRELAKGL	TAVQLLDLFT	360
GDPVIDPKKG	VVTFSYKEGV	VPVFSQKIVP	RFKSRDTLEK	LTSQGPVALL	SVDLGQNPEV	420
AARVCSSLKNI	NDKITLDNSC	RISFLDDYKK	QIKDYRDSL	ELEIKIRLEA	INSLETNQOV	480
EIRDLDVFS	DRAKANTVDM	FDIDPNLISW	DSMSDARVST	QISDLYLKNG	GDESRVYFEI	540
NNKRIKRSYD	NISQLVRPKL	SDSTRKLNLD	SIWKLKRTSE	EYLKLSKRKL	ELSRAVVNYT	600
IRQSKLLSGI	NDIVIILEDL	DVKKKFNGRG	IRDIGWDNF	SSRKENRWFI	PAFKAFSEL	660
SSNRGLCVIE	VNPATWSATC	PDCGFCSKEN	RDGINFTRCK	CGVSYHADID	VATLNIARVA	720
VLGKPMMSGPA	DRERLGDTKK	PRVARSRKTM	KRKDISNSTV	EAMVTA		766
SEQ ID NO: 55		moltype = AA	length = 717			
FEATURE		Location/Qualifiers				
source		1..717				
		mol_type = protein				
		organism = synthetic construct				
SEQUENCE: 55						
MIKPTVSQFL	TPGFKLIRNH	SRTAGLKLKN	EGEEACKKFV	RENEIPKDEC	PNFQGGPAIA	60
NIIAKSREFT	EWEIYQSSLA	IQEVIFTLPK	DKLPEPILKE	EWRAQWLSEH	GLDTVPYKEA	120
AGLNLIKNA	VNTYKGQVK	VDNKNKNLA	KINRKNEIAK	LNGEQEISFE	EIKAFDDKGY	180
LLQKPSPNKS	IYCYQSVPK	PFITSKYHNV	NLPEEYIGYY	RKSNEPIVSP	YQFDRLRIP	240
GEPGYVPWKQ	YTFLSKKENK	VSPILGIICI	KKDWCVPDMR	GLLRTNHWKK		300

---

-continued

---

YHKPTDSIND	LFDYFTGDPV	IDTKANVVRF	RYKMENGIVN	YKPVREKKGK	ELLENICDQN	360
GSCKLATAVDV	CQNNPVAIGL	FELKVKVNGEL	TKTTLISRHP	PIDFCNKITA	YRERYDKLES	420
SIKLDIAIKQL	TSEQKIEVDN	YNNNFTPQNT	KQIVCSKLNI	NPNDLPWDKM	ISGTHFISEK	480
AQVSNKSEIY	FTSTDKGKTH	DVMKSODYWF	QDYKPKLSE	VRDALSDIEW	RLRRESLEFN	540
KLSKSRSQDA	RQLANWISSM	CDVIGIENLV	KKNNFFGGSG	KREP瓜DNFY	KPKKENRWI	600
NAIHKALET	SONKGKRVIL	LPAMRTSITC	PKCKYCDSKN	RNGEKFNCNK	CGIELNADID	660
VATENLATVA	ITAQSMPKPT	CERSGDAKPK	VRARKAKAPE	FHDKLAPSYT	VVLREAV	717

SEQ ID NO: 56	moltype = AA	length = 766				
FEATURE	Location/Qualifiers					
source	1..766					
	mol_type = protein					
	organism = synthetic construct					
SEQUENCE: 56						
MEKEITELTK	IRREFPNKKF	SSTDMMKKAGK	LLKAEGPDAV	RDFLNSCQEI	IGDFKPPVKT	60
NIVSISRPFE	EWPVSMVGRA	IQEYFFSLTK	EELESVHPGT	SSEDHKSFFN	ITGLSNYYNT	120
SVQGLNLIFK	NAKAIYDGTI	VKANNKNKKL	EKKFNEINHK	RSLEGLPLIT	PDFEEPFDEN	180
GHLLNNPPG	RNIYGYQGC	AKVFVPSKHK	MVSLPKKEYVG	YNRDPNLSLA	GFRNRLEIPE	240
GEPEHNVPPWF	RMDIPEGQIG	HVNKIQRNFN	VHGKNSGKVK	FSDKTGRVKR	YHHSKYKDAT	300
KPYKFLEESK	KVSAldSILA	IITIGDDWWV	FDIRGLYRN	FYRELAKGL	TAVGQLLDFT	360
GDPVIDPKKG	VVTFSYKEGV	VPVFSQKIVP	RFKSRDTLEK	LTSQGPVALL	SVDLGQNNEPV	420
AARVCSSLKNI	NDKITLDNSC	RISFLDDYKK	QIKDYRDSL	ELEIKIRLEA	INSLETNQQV	480
EIRDLDVFS	DRAKANTVDM	FDIDPNLISW	DSMSDARVST	QISDLYLKNG	GDESRVYFEI	540
NNKRIKRSDY	NISQLVRPKL	SDSTRKLNLD	SIWKLKRTSE	EISLRAVNVNT		600
IRQSKLLSGI	NDIVIILEDL	DVKKKFNGRG	IRDIGWDNF	SSRKENRWFI	PAFHKTFS	660
SSNRGLCVIE	VNPAAWTSATC	PDCGFCSKEN	RDGINFTRCRK	CGVSYHADID	VATLNIARVA	720
VLGKPMSPGA	DRERLGDTKK	PRVARSRKTM	KRKDISNSTV	EAMVTA		766

SEQ ID NO: 57	moltype = AA	length = 718				
FEATURE	Location/Qualifiers					
source	1..718					
	mol_type = protein					
	organism = synthetic construct					
SEQUENCE: 57						
VPDKKETPLV	ALCKKSFPGL	RFKKHDSRQA	GRILKSKGEG	AAVAFLEGKG	GTTQPNFKPP	60
VKCNIVAMS	PLEEWPIYKA	SVVIQKYVVA	QSYEEFKATD	PGKSEAGLRA	WLKATRVDTD	120
GYFNVQGLNL	IFQNARATYE	GVLKKVENRN	SKKVAKIEQR	NEHRAERGLP	LLTLDDEPETA	180
LDTGHLRHR	PGINCSVFGY	QHMLKLKPVWP	GSIPGVTGYS	RDPSTPIAAC	GVDRLEIPEG	240
QPGYVPPWDR	ENLSVKKHHR	KRASWARSRG	GAIDDNMLLA	VVRVADDWAL	LLDRGLLRNT	300
QYRKLLDRSV	PVTIESLLNL	VTNDPLTSV	KKPGKPVRYT	ATLIYKQGVV	PVVKAKVVKG	360
SYVKMlldt	TETFSLVGV	LGVNLLIAAN	ALRIRPGKCV	ERLQAFTLPE	QTVEDFFRRF	420
KAYDKHQENL	RIAAVRSLTA	EQQAEVLA	TFGPQEAKMQ	VCGHGLSVD	EVPWDKVNSR	480
SSILSLDAKE	RGVDDTLYM	PFFKGKGK	KTEIRKRW	NWAQHFRPQL	TSETRKALNE	540
AKWEAERNSS	YHQLSIRKK	ELSRHCVN	IRTAEKRAQC	GKVIVADEL	HHSFRGGKG	600
SRKSGWGFFF	AAKQEGRWL	DALEGAFCDL	AVHRCYRVIK	VDPYGNNTSRTC	PECGHCDKAN	660
RDRVNREAFI	CVCCGYRGNA	DIDVAAYNIA	MVAITGVSLR	KAARASVAST	PLESLAAE	718

SEQ ID NO: 58	moltype = AA	length = 757				
FEATURE	Location/Qualifiers					
source	1..757					
	mol_type = protein					
	organism = synthetic construct					
SEQUENCE: 58						
MPKPAVESEF	SKVLLKKHFPG	ERFRSSYMKR	GGKILAAQGE	EAVVAYLQGK	SEEPPNFQP	60
PAKCHVUTKS	RFDAEWPI	ASEAIQRYIY	ALSTTERAAC	KPGKSSESHA	AWFAATGVSN	120
HGYSHVQGLN	LIFDHDTLGRY	DGVLKVKVQ	NEKARLAR	INASRADEGL	PEIKAEEEV	180
ATNETGHLLQ	PPGINPSFYV	YQTISPQAYR	PRDEIVLP	YAGYVARDPNA	PIPLGVVRNR	240
CDIQKGCPG	IPEWQREAGT	AISPKTGKAV	TVPGLSPKKN	KRMRRYWSE	KEKAQDALLV	300
TVRIGTDWV	IDVVRGLLRNA	RWRTIAPKDI	SLNALLDLFT	GDPVIDVRRN	IVTFYTTLDA	360
CGTYARKWTL	KGKQTKATLD	KLTATQTVAL	VAIDLQGTQ	ISAGISRTQ	ENGALQCEPL	420
DRFTLPPDLL	KDISAYRIAW	DRNEEALRAR	SVEALPEAQO	AEVRALDGV	KETARTQLCA	480
DFGLDPKRLP	WDKMSSNTTF	ISEALLNSV	SRDVQFFTPA	PKKGAKKKAP	VEVMRKDRTW	540
ARAYKPLRSV	EAQKLKNEAL	WALKRTSPEY	LKLSRRKEEL	CRRSINYVIE	KTRRRTQCQI	600
VIPVIEDLNV	RFFHGSGK	PGWDNFFTAK	KENRWFQGL	HKAFTSLRTH	RSFYVFEVRP	660
ERTSITCPKC	GHCEVGNRDG	EAFQCLSCGK	TCNADLVDVAT	HNLTQVALTG	KTMPKREEPR	720
DAQGTAPARK	TKKASKSKAP	PAEREDQTPA	QEPSQTS			757

SEQ ID NO: 59	moltype = AA	length = 761				
FEATURE	Location/Qualifiers					
source	1..761					
	mol_type = protein					
	organism = synthetic construct					
SEQUENCE: 59						
MSNKTTPSP	LSLLLRAHFP	GLKFESQDYK	IAGKLLRDGG	PEAVISYL	TG KGQAKLKDVK	60
PPAKAFVIAQ	SRPFIEWDLV	RVSRIQKEI	FGIPATKGR	KQDGLSETAF	NEAVASLEV	120
GKSKLNEETR	AAFYEVGLD	APSLHAQAOQN	ALIKSAISIR	EGVLKKVENR	NEKNLNSKTKR	180

-continued

---

RKEAGEEATF VEEKAHDERG YLIHPPGVNQ TIPGYQAVVI KSCPSDFIGL PSGCLAKESA	240
EALTDYLPH RMTIPKGQPP YVPEWQHPPLL NRRKNRRRKP ATCSKRSGTP	300
NRKNSRTDQI QSGRFKGAIP VLMRFQDEWW IIDIRGLLN ARYRKLLKEK STIPDILSLF	360
TGDPSIDMRQ GVCTFIYKAG QACSAKMVKT KNAPEILSEL TKSGPVLVLS IDLGOTNPIA	420
AKVSRVTQLS DGQLSHETLL RELLSNDSSE GKEIARYRVA SDRLRDKLAN LAVERLSPHE	480
KSEILRAKND TPALCKARVC AALGLNPEMI AWDKMTPYTE FLATAYLEKG GDRKVATLKP	540
KNRPEMLRRD I KFKGTEGVR IEVSPEAAEA YREAQWDLQR EDLNIKMMHG NGKWADGGWD AFFIKKREN R WFMQAFHKSL	600
NQLRHKAKS SQCEVVVMFAF EDLNIKMMHG NGKWADGGWD AFFIKKREN R WFMQAFHKSL	660
TELGAHKGPV TIEVTPHRTS ITCTKCGHCD KANRDGERFA CQKCGFVAHA DLEIATDNIE	720
RVALTGKMPK KPESERSGDA KKGSKGARKAA FKPEEDEAA E	761

SEQ ID NO: 60	moltype = AA length = 708
FEATURE	Location/Qualifiers
source	1..708
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 60	
MSKTKELNDY QEAALARLPG VRHQKSVRRA ARLVYDROGE DAMVAFLDGK EVDEPYTLQP	60
PAKCHILAVS RPIEEWPIAR VTMAVQEHVY ALPVHEVEKS RPETTEGSR S AWFKNSGVSN	120
HGVTHAQTLN AILKNAYNVY NGVIKKVENR NAKKRDLSAA KNKSRERKGL PHFKADPPEL	180
ATDEQGYLLQ PPSPNSSVYL VQQHLRTPQI DLPSGYTGPV VDPRSPIPSL IPIIDLAIAPP	240
GQPGYVPLHD REKLTNSNKH RMLKLPKSLRA QGALPVCFRV FDDWAVVDGDR GLLRHAQYRR	300
LAPKNVSI AILELYTGDGV IDIKRNLMFT RFAEAVVEVT ARKIVEKYHN KYLLKLTEPK	360
GKPVREIGLV SIDLNVQRLLI ALAIYRVHQT GEQSLALSPC LHREILPAKG LGDFDKYKSK	420
FNQLTEEILT AAVQTLTSQAQ QEYQRYVEE SSHEAKADLC LKYSITPHEL AWDKMTSTQ	480
YISRWLRDHG WNASDPTQIT KGRKKVERLW SDSRWAQELK PKLSNETRTRK LEDAKHDLQR	540
ANPEWQLRALK RQKEYSRHLA NTVLSPRQTAQ TACETVVIAI ENLPMKGGFV DGNGSRESGW	600
DNPFTTHKKEN RWMKDIHKA LSDLAPNRRV HVLEVNPQYT SQTCPCEGHR DKANRDPIQR	660
ERFCCTHCGA QRHADLEVAT HNIAMVATTG KSLTGKSLAP QRLQEEAAE	708

SEQ ID NO: 61	moltype = AA length = 661
FEATURE	Location/Qualifiers
source	1..661
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 61	
VAFLDGKEVD EPYTLQPPAK CHILAVSRPI EEWPIARVTM AVQEHVYALP VHEVEKSRPE	60
TTBGSRSRAWF KNSGVSNHGKV THAQTLNAIL KNAYNVYNGV IKKVENRRAK KRDSLAAKNK	120
SERKGLPHF KADPPELATD EQGPPQPS PNSSVYLVQQ HLRTQPIDLP SGYTGPVVDP	180
RSPPIPSSLPI DRLAIPPGQP GYVPLHDRE LTSNKHRRMKL LPKSLRAQGA LPVCPRVFDD	240
WAVVDDGGLL RHAQYRRLAP KNVSIAELLE LYTGDPVIDI KRNLMTFRFA EAVVEVTARK	300
IVEKYHNYKL LKLTEPKGKP VREIGLVSID LNVQRLIALA IYRVHOTGES QLALSPCLHR	360
EILPAKGLGD FDKYKSFKFNQ LTEELTAAV QTLTSAQEE QEYQVEESSH EAKADLCLKY	420
SITPHELAWD KMTSSTQYIS RWLRDHGWN A SDFTQITKGR KKVERLWSDS RWAQELKPKL	480
SNETRRKLED AKHDLQRANP EWQRQLAKRKQ EYSRHLANTV LSMAREYTA ETVVIAIENL	540
PMKGFFDVGNS GSRESGWDFN FTHKKENRWM IKDIHKALSD LAPNRGVHVL EVNPQYTSQT	600
CPECGHGRDKA NRDPIQRERF CCTHCAGAQRH ADLEVATHNI AMVATTGKSL TGKSLAPQRL	660
Q	661

SEQ ID NO: 62	moltype = AA length = 845
FEATURE	Location/Qualifiers
source	1..845
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 62	
MIKKPSNRHA LPKVIISKVD NQNILEFKIK YKKLSRLDRV EIKTMHYDDR AIVFDEVII	60
GGLIDVBYRD NHKTIFVKVG DKSYSISGOK VGGKERLLEN RISQTKVQLE LKDEATNRVS	120
KTERELIVDD NIKLYSQIVG RDVKTTKDIY LIKRLFLGYRS DLLFYYYGFVN NFFHVANNRP	180
EFWKIDFDNN RNSKLIEYFI FTINDHKLND ENYLKDYISD RGQIVDDLEN IKHIFSALRH	240
GLMHFDYDFF EKMDNFGEDID KIMDNQGNTQ PLSSLNKFL DIMDILKDLN NIDTKKEFID	300
AEKITIFGEE LSLAKLYRFV AHTAINRVAF NKLINSLII NGVENOSLKE YFNQOAGGIA	360
YEIDIHQNRE YKNLYNEHKK LVSRLVLSISD GQEATLNQK IVELKEQMKQ ITKINSIKRL	420
EYKLRLAEGF IYTEYKNYEE FKNSFDTDIK NGRFTPDKED GNKRAFDSE LEHLKGYYKA	480
TLQTQKBQTD EKMEEVSKRV DRLLSLKSLIG DDTLLKFILL MFTFMPQELK GEFGLGFIKKY	540
YHDTKHDQD TISDSDDTIE EGLSIGLKLK ILDKNIRSLS ILKHSLSFQZ KYNKKDRSY	600
EDGNIHKGFF KKLIGISHNQE EFNKSVYAPI FRYYSALYKL INDFEIYTLS LHIVGNETLS	660
DQVNKPQFLS GRYFNFRKLL TQSYNISNN S THSVIFNAVI NMRNDISHLS YEPLLDCPLN	720
GKKSYKRKIR NQFRTINIKP LVESRKMIID FITLQTDMOK VLGCDAVNDF TMKIVQLRTR	780
LKAYANKEQT IEKMITAEKT PNDFYNIYKV KGVEAINKYL LEVIGETQVE KEIREEIERS	840
NIANS	845

SEQ ID NO: 63	moltype = AA length = 802
FEATURE	Location/Qualifiers
source	1..802
	mol_type = protein
	organism = synthetic construct

---

-continued

---

SEQUENCE: 63

MLKPKPSNRYA LPKVILSTVD HEKILEFKVK YEKLARLDRL VVERMHDGEE SVVFDEVIAN	60
SGDLEIAYQD DHRKLIIQAA GKSYTITGKK VGGKKRKLEE RISRALKQLT LTDGQEDQHR	120
RIRATVTEKA LLEPKBDRDI YSKISDRKIK TSKEIYLVKR FLSYRSDDLIF YYFFVDNNFK	180
VGNNKQELWK IKFQNQPELI EYFRFIINDR FKNAKNDKF NYLKNDKAIQ EDLEKIQKVF	240
EKLRRHLMHY DYGFEEKLFGE GEDQGFDDI AFLDNFVKKI DKLNIIDTKK FVDDEKIKIF	300
GEDLNLADLY KLYASISINR VGFNRVVNEM IIKGIEKLN KLFRAEKKLD KTYALDIHSD	360
PSYKKLYNEH KRLVTEVSTY TDGNIKEGK QKIAKLLKYM KEITKKNALV RLECKMRLAF	420
GLIYGRYDTH EAFKNGFDTD LKRGEFAQIG SSEEAGYFNT TFEKSKPKSK EEIKKiarqi	480
DNLSSLTLIE DDPLMKFIVL MFLFLVPRELK GEPLGFWRKY YHDIHSIDSS AKSDEMPEDEV	540
SLSLKLKILT RNIRRLNFE YSLSKIKYS PKNTOFYTDK SPYQKVYKRL KISHNKEFD	600
KTLLVPLFLRY YSILFKLIND FEIYSLAKAN PDASSLSELT KTKHGFRGHY NFTTLLMMDAH	660
KVSQGDSKKH FGIRGEIAHI NTKDLIYDPL FRSKMMAQQR NDVIDFVLKY EKEIKAVLGY	720
DAINDFRMKV VQLRTKLKVY SDKTQTTIEKL LNEVEAPPDF YVLYKVGVE AINKYLLEIV	780
SEQ ID NO: 64	802

SEQ ID NO: 64                    moltype = AA length = 1281

FEATURE                         Location/Qualifiers

source                         1..1281

mol\_type = protein

organism = synthetic construct

SEQUENCE: 64

MIKSYDDFTK LYSLQKTLRF ELKPICKTLE HIKKSEIIIES DETLEEQYAI VKNIIDRFHR	60
KHIDEALSLV DPTKHLDFTK TIQEYLKRG KTDRKKBLE ELSADRKLII VSYLGKVNQ	120
KTQHNLNPPIK ERFEILFGKE LFTNNEEFTL AENKEEKKAI QAFKGFTTF KGFQENRKNM	180
YSEEDKSTAI AYRIINENMP LFIENIARPO KVLDVIEKTK LTELQONLKT ELKGHSVSDI	240
FRIEYFNNVL TQEGLSRYNT IIGGKTTTEG VKIQGLNEIT NLHNQSQSKDV KLPILKPLHK	300
QILSEEYTS FPKIAPENDN EVLKAIDFTW NEHIEKSIHP VTGKRFNILL KIENLCKKLE	360
KYKDKEIEKL FIERKNLSTV SHQVYQGWNI IRDALRMHLE MNKNKNIKEKD IDKYLDNDAF	420
AWKEIKDSIK IYKEHVEDAK ELDENGIVKY FSSMSINEED DEKEYSISLII KNINEKYNV	480
KSILEEDRTK KSDLHQDKKEK VAIIFLDS LKQLQWFLKL LYVTVPDKEK DYEFYNELEV	540
YYPALPLNS LYNNVRENMT RKPYSVEFKF LNPFYSPTLID GDWDKNKETAN LSILKKNKG	600
YYLGIMNKEN NTIFENFPKS KSNDYYEKM YKLLPGPNKML LPKVFPSKKG LEYYKPSKEI	660
LRIYEKGEFK KDKSGNFKKE SLHTLIDFYK EAIAKNEDWK IFKFKFKNTR EYEDISQFYR	720
DVEEQGYLII PEKVDANYVD KLVEGEELYF FOIYNKDFSE NKKSGNPNL HTIYWESLPD	780
NQNLKDVVYK LNGEAEVYFYR KKSIDYPEEI YNNHHKBEEL NGKFVYPIK DRRYTQDKFL	840
FHVPITMNF1 SKEEKRVNQL ACEYLSTTKE DVIIIGIDRG ERHLLYLSLI DKEGNIKKQL	900
SLNTIKNENY DKEIDYRVKL DEKEKCRDEA RKNWDVBIENI KELKEGYMSQ VIHIIAKMMA	960
EEKAILMED LNIGFKRGRF KVEKQVYQKF EKMMIDKLNY LVFKNKEPLE PGGSLNAYQL	1020
TSKFDSFKKL GKQSGFIFYV PAYSATSKID TTGFYNNFIRV DVPNLEKGKE FFSKFEKIIY	1080
NTKEDYFEFH CKYGFVPEP KNKDNDRRTK ESLTYYNAIK DTWVVCSTH HERYKIVRNK	1140
AGYYESQPVDF VTKNLRDIFS EANINYSDGK DIKPIIIBSN NAKLLKSIAE QLKLILAMRY	1200
NGKHDDEK DYILSPVKNK QGKFFCSLDG DQSLPINADA NGAYNIALKG LLLIEKIKQ	1260
QGKAKDLYIS NLEWMFMFMMS R	1281

SEQ ID NO: 65                    moltype = AA length = 1332

FEATURE                         Location/Qualifiers

source                         1..1332

mol\_type = protein

organism = synthetic construct

SEQUENCE: 65

MKSIFDNFTG LYSLSKTLRF ELRPVGQTLE NIKNGHFLES DKKMADDYQD VKKIIDNYHK	60
FFIDDVLKGA SFDWALLEKE LTDFNKNKT DSKVVAEQQKK LREQIAKTLA GDKRFKSLTA	120
STPNDFLNFNQD KDFIGWLEQS SVKEIRKDAL DTFKKFSSYF KGFQENRKNV YSADDIPTAV	180
PYRIVNDNFP KFLQNISIFF TIQEKKCPOVI ADVENELASY LGKEKLADIF TVQAFPNKYL	240
QGGKENQRGI DFYNQVIGGI AEKEGGVNLN QINQFLNLYW QQHPDFAKEN RRIKMPVLYK	300
QILSDRSSLS FKIETIDTDE ELKTAISEYA DKLESKSND EKSVLDCVCE LFDSIKEQNL	360
QEIYVNRKD1 NNIISRLTGD WSWLQSRMNL YADEVFTTKA EKTRWQKSVD GDEGENSKG	420
AYSLAELNRV LEYASENVAE TDIRITDYFC HRNRFYYKEE SGLFKQGEEL VALSIKESCE	480
DILSKRKAMN EAFANISESN SLRDNSEIA KTKTYLDSVQ DLLHRIKPLK VNGLGDPSPFY	540
AIVEDSIYAL SEVISIYNTK RNYITRKAES PEKYKLNPDN PTLANGWDLN KEKDNTCVLL	600
RKNGMYYLGI MNPKDKPKFA EKYDCGTESC YEKMIYKLPP GPNKMLPKVF FSTKGKKQY	660
PPENILHGYE QGKHKKGVAF DINFCHELID WFKSAINOHE DWKKFGFKFS DTKSYKDID	720
FYREVTEBQGY KLTFINIPES EISKMWSEKG LYLFQIYNMDF PAPGANGMPP MHTLYWKNL	780
SEENLKDVLVQ LKTFINIPES EISKMWSEKG LYLFQIYNMDF PAPGANGMPP MHTLYWKNL	840
QSGMLFVYPA ATTSKIDPKT GFVSMNNFKD LTNVHKKRDF FSFKEDIHFD EATCSFVFTF	900
DXKNFNGKAK EEMQTKWAV YSREKRIVYF SKTKSSEDIM PTEKLRALFE SDGIBYKSGN	960
NIHDHSVMAVG ADLKEGAKPS KEIADFWDGL LYNFKLILQW RNSNAKTGED YIISPVMASD	1260
GTFFDSSRVEA KKGKDAKLPL DADANGAYHI ALKGSLSLINK INLAGEDELK KFDMKISNED	1320
WFNFAQEKKY AE	1332

SEQ ID NO: 66                    moltype = AA length = 1345

-continued

---

FEATURE source	Location/Qualifiers	
	1..1345	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 66		
MTKQNKSVTQ MTKRKNFGEF TNLYSLSKTL RFELKPKAT KTIILEVERKD RENFKKKDRKI	60	
AKNYQKLKG1 LNELHQEFIQ DVMREFSFQE KEIKEFEERY LEALNFKED NYKKRTQLKN	120	
AYEKVAKKLA GKIATAFGKY NQEKEYGVKPT KKNLTGENVF DILEGKYKGD KKILGIHTF	180	
KFKPTKEKK QGKEAVNFST YLGGFNQNRE NFYKgemkAG QFATRTIENL IQFLKNKLF	240	
IDKYKDNYQK LGFSQBOVEI FNLYNFFNNLF LOEGLDVYNG ILGAKKGEKU TENDGLNQKI	300	
NLFQKQEKT R CKANGEFKNK SDYPIKFKE KQJGSIKKDN DVYVEIKSD E LVNVLQSLP	360	
EKTANTLREV QKFYENFFD1 IFNDEFDLDK IYLPKSVGTH FSHLAFSDWS KLAFLVFNKRW	420	
RNEVKVIKEG EDVNQVSRS1 ADIKKRMEI LEMDGGVSFG KTYCQKVGLE KEARTIEDVW	480	
SGFWKIIQYH INSQFIGGEE EVFDEKEKKDD KTEKIQTIDD LQEEYLQATE MYRERMVESE	540	
EGLNDGEKEE I KTKLNLYD RIKDIERA YFDLRLKHFDD IDEASKDGF YFIYQELLQD	600	
ISEAKINDHY NEIRNLTKA NVVDDKFKLN FNDGQTLSGW DLNKETEKF5 LIFKRKVVDGG	660	
VEYYLGIINK EKNKTIFDKK KHPEIFTENS EFEKMEYKLF PSPSKMLPKI AFTKNKEGER	720	
IKPVFLDENA GKEIAQIKKE FALFQDAKKE DKNKWSDEFD RKKLKNLIDY YKLVLLEKHPE	780	
KYMQTFNPFV KSSAKYKNLG EFNDQDVARQ1 YVTKFVSVKD DYIDQKVESG ELYLFKIHNK	840	
DWNLTAGDT KKQSKKNLHT IYFEELFSEK NIAEPVFKLS GGAEVFRDA IEKKQQKKKK	900	
DKKGKEILEK PRFTKKNKILF HVPITINYGK PSINQGQFNQ KINEFIADNS RSVNLIGDR	960	
GEKHLLYSSL VDNNGKIIKS GSLNEINGVD YHEKLDKAECR ERQEARKSWQ KINQIKNLKA	1020	
GYISQVKK1 VDLAINEKK1 IVLEDLNFGF KSFRQKIEKVN VYQQFEKALI DKLGFVTDKE	1080	
KLNHRQAPQL SAPFESFEK M GKQTGIVFVY LATNTSKVCP QCQWKKNIFF HYSTKKSIAE	1140	
NLQKQYKMKM YWRENENRFE FEYKGDGDE FSSIIFSNVDR VRYDKRANNN QGGVVIYQID	1200	
STTKEKDGRN I KEKSITNLN KELLLEKFEI DNLEGELLVK LSEKSPDVSK ETIKDFFGLL	1260	
NSILNIRNSM TDEEYIQC PACGFDTRKE NKIGIKNGDD NGAYNIALRG RFLIERIKKA	1320	
KBEDKKPNLT FSNNNDYFQWV REFVK	1345	
SEQ ID NO: 67	moltype = AA length = 1265	
FEATURE source	Location/Qualifiers	
	1 .. 1265	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 67		
MRTTTSLSDF TNRYALSKTL RFELKPIGN1 QMMLEQNNVF AKDRAIREKY EKTPWPIDL	60	
HREFVAESLQ NAQLGNLDDY YAALQNVQKI TKDTNAEDKK RWKKGPEKQE KRLRKEVVAL	120	
FDKAAHIWAT QRYPQLKKKT KDFLFLFEEGVF EHVLFARYGS APDTTVKIVT SNPETGEVID	180	
EREESIFPKGW KGFTGTYFDK FETRQFVYKD NGTATAIATR AINQNLRRFA ENMOKLTDIK	240	
NNYPEELAHT DFGDFDIAHA QSLDFYARTC LLQEGIDAYN KKFVGVLKSA INEYQQKNG	300	
VRISYPKTLQ NQILGERERR LFVDIEDDR LHDVFRFVFD DGTVFAAEMR QLAQAFSAQN	360	
GTYDYTQIYI SKKGKFETSR KYTHDTRAWH DALADVFKAK AKKRIATTAS GEKKPPAYIP	420	
VAYITQALTL VQESEDETECT WKERYASITE NKTLEEGFFA IFADEPERL1 VHMEATVQDT	480	
DYVVAEDKAK KLLSDGQITK NEQTTQIIKE YADALLRIYQ MAKYFAVEKK SMWDDAAVID	540	
DTFYETFKEI YGNTHTSTIVA SYNLLRNLYT KKPWFEDVQKWL KLFNFENPTLL DGWDKNKEAA	600	
NFGVILRQGD KFYLGIMRK HNNIFANQHH SNFEGQQGLQK MVYKFFPDPK KMFPKVCFS	660	
KGMEEFAPSE EIVRUYKNAF FKSGDTFNVE SMQKLIDPYFQ NALQKYDGWK IYDFKHLKD	720	
AQYTSNIGEY YDDVAKGGYQ LGWQNIKEY VEEKNANGEL YLFQIKNKD W NDGATGRKNL	780	
HTLYFEYLFS EKNAAADEFVF RLNGGAEVY RPAAIESKTE RRGNRVEAAK KRYTQDKVFL	840	
HVPITLNRTA GDVKTSAFND AVNRFLAGNP DINIMGIDRG EKHLAYYSII DQNGNIRVSG	900	
SFTNTGSKDY HALLTERQQA REEARKEKNWQ VEQIKDLKKG YISLUVREIA DLAIKHNAII	960	
VLENLNMRK QIRGGIEKSV YQOLEKARIE KLNFLVNKG VDATKAGHLL RAYQLAAPFE	1020	
TFEKMGNQTC IIFYTTASYT SQVDPVTGWR PHVYLKYRNA QTKTEDILRI FDDIVFNDEK	1080	
QRFEFAYRH GVSHTVCSSV ERHRWNRSNN AGKGGYDVFP VEGEGSITQR LQEACAQRGI	1140	
DTRRNILAQI DELDESASAT VSFLRDLFCY FRLLICQIRNT DDGADDINAQ DFLMSPVEPF	1200	
FDTRNAQEY PQNGDENGAY NIARKGIIIL QKITAWGRSQ DTQRYPDTF VSQDEWDTFL	1260	
TQHTT	1265	
SEQ ID NO: 68	moltype = AA length = 1259	
FEATURE source	Location/Qualifiers	
	1..1259	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 68		
MLKEKQFKTF GDFTNLYELS KTLRFELRPT PETKDLLDKN KIIQTDKIA ENYQEIKYF	60	
DLKLHKKFIKE ALSNTQIDFS DFCKLWEQNS KDSGKIKDLS RKLRLKS1KQA FDKGADWHK	120	
RYLEKG1KLK KKNLDILFEE RVLDILKEEF KDDVDVULKF SFKGFSFTYFT NFHESRKNFY	180	
KDDGTASAIA TRIIDENLKR FCDNIKVKKH SKKLISELNE REAKIFEADF YNRCLLQQGI	240	
DDYNQVIGDI NKKINNLRQ1 KIENTPTLK1 LYKQILGDVR RQETEQAIFI EIKNNNEEVFD	300	
FLQDFIKHSD ENNKYFKNLF YKFIKEGHSL DKIFLAKRFL NTISGKWFAS WEVFGAELIK	360	
KFGNKKLDPL PIPFAAVKDV LQNCNIPANE LFKEKIKNDE DKNIYDIFIN LWKEEFDMSL	420	
KKVEESKKEV ENMIAEDKVV SNKKEKRND NGEEIEIEIQ KEKIKNAYADA AMNIFRMK	480	
FILLEKNGKTV EGMGEDNNFY NELNIVFKGG EIDGKVYEGV KTYLYYNEFR NYLTKKPFNE	540	
EKTKLNFDGQ QILSGWDKNK ESEKLGVLIR KDNKYYLAI1 NKKHNK1FDV KKNSYAYIVG	600	
DNFYEKMEYK LFPDAKRMIP KIAFAKNNKE KFGWTDEIQK IKNEYAEFQE GKNDKNLWK	660	
DKPNKKNMKEK LITYYQNCLE KGGYKDIYNF RWKSPDKVKG IGEOFNDEIDR QSYCLKFV	720	

-continued

---

DFNYVFEVK	SGELYLFQIY	NKDFSDKADR	AQKENIHTEY	FKLLFDQRNL	DNVVLKLSGG	780
AEIFYRPKTE	GLPKKKDNKG	NEVRHRHRYA	DDKYFLHLPI	QLNFGRGNLNS	GGEFNSKINQ	840
YLSEQREIKI	IGIDRGEKHL	AYYSVINQDG	KIEEIESLNT	VNGIDYRKKL	DELEKKREQE	900
RKSWQSISKI	KDLKKGYISH	VIKKICDLAI	EHNAAIVFED	LSGGFKNSRK	KIEQIYQNL	960
ELALATKLNY	LTFKDKNFGE	SGHYLNAYQL	APKIDNYQDI	KMQTGIVFT	PAGYTSSTCP	1020
QCGFRKTLKF	DYTATISKAE	DLIRGSKLN	VFEKEKNRFK	INYLFNPIEK	KKKKIKENEL	1080
FADAGAKNEF	TIYSDVKRIK	WHNTGTRKLE	EAEGERLLEN	KNSRGRDKEY	DINKCLTRLF	1140
RENKIDVNGD	IIGQITKIKS	LKLYQDLFY	LFLATLIRNN	VSGSDIDYIQ	CPSCHFHSDG	1200
GFQKQKFNGD	ANGAYNIARK	GILILKKIKQ	FAAQDKDMKN	FGWKHLLTVDI	NEWDKFTQK	1259

SEQ ID NO:	69	moltype = AA	length = 1371			
FEATURE		Location/Qualifiers				
source		1..1371				
		mol_type = protein				
		organism = synthetic construct				
SEQUENCE:	69					
MNKNFSNFT	LYTLSKTLRF	ELKPVQAQTKE	NIKKGKFLES	DKKKADDYKD	VKKIIDNYHK	60
FFIDDVLKNA	SFDWTWLEKE	MSDFNKSAD	DSKVEAEQKK	LRDQIAKKLT	SDKRKPALTA	120
STPSDLFNKD	KDFIDWTFQN	STKDINKEAL	ETFKRFSSYF	LGQFQENRKNV	YSAEPIPTAV	180
PYRLVNDNFP	KFLQNLIALFK	IIQEKCPQVI	SDVEKEELASY	LGKEKLADIF	TVQAFPNKYLC	240
QGGKENQRGI	DFYNNILGGI	AEKEGGINLR	GINQFLNLYW	QQHQDFAKQN	KRIKMIPLFK	300
QILSDRSSSL	PKIESINTQD	ELLTSITEYA	DKLETKSNDE	KKSVDLICSD	LFASIKAQNL	360
QEIVYNRKDI	NSISRILTGD	WSWLQSRNMV	YADEVFPTKA	EKTRWQKSID	GDEGENKSKG	420
VFLSALELNSV	LEYSSENVSE	TDVRLTDFPD	HRNRFYYKEE	SGLFKQGDEL	VALSIRESC	480
DILAKRKAMD	EAFAVNSENN	SLRDNSEDVA	KIKIYLDCCVQ	ELLHRIKPLK	VNLGDPAFY	540
AVPDTVYNSL	SEVISLYNKI	RNYTITKKAAN	PEKYKLNFDN	PTLADGWDLN	KEQANTSVLM	600
RKDGMYLGI	MNPKDCKPFL	EKYECGNEAC	YEKMIYQFD	ATKQIPKCST	QVKEVKKHFQ	660
SGATDSIILN	DKSFKFLDLV	ITKEIWFLNN	HVNNGEKEFVP	KRESNETRPK	KFQIGYYKQT	720
GDLGGYKEAL	NIWISFCKTF	LQSYISSSIY	DYDFKESSNY	DSLDEFNYL	NATCYKLSFI	780
NIPEATISQM	VSEGKLYLRFQ	IYNKDFAPGA	SGMPNMHTLY	WKNLFSEENL	KNVVLKLNGE	840
AELFYRPAGI	KEVPIHAKGS	YLVNDRITKDG	EPIPEKIHD	IYRNANGKLE	SLSKEATEYK	900
ASHKVVIEA	KHDIIKDRHY	TEPKPLFHPV	LTINFKASGN	SYINENVRRF	LKNNDPDVNVI	960
GLDRGERHLI	YLSLINQKGE	IIKQFTFNEV	ERNKNGQVIK	VNYHEKLDQR	EKVRGAARKS	1020
WQAIKGIAEL	KEGYLMSAVI	QLTKLMVVEYN	AIVVMEDLNF	GFKGRGRHV	KQVYQKFEHM	1080
LIDKLNLYLVF	KDRGLLTEAGG	VLNGYQLAGQ	FESFQKLGKQ	SGMLFYVPAG	YTSKIDPKTG	1140
FAFSMSNFKDL	TNVHKRAFF	SKFDIHFDD	ATGSFVFTFD	YKNFDGKAKE	EMKRTKWSVY	1200
SDKRIVYLS	KTKSYEDVQP	TEKIKASLES	VGIEYMSGNN	LIDSIMVIGA	ELKDGAJKPSK	1260
EIADFWDRLL	YNFKLILQMR	NSNAKTGEDY	IISPVMADDG	TFFDSDREEFK	KGENAKMPVD	1320
ADANGAYHIA	LKGSLLLKRF	DAASENELKK	FDMKISNVWD	PKFAQEKS	Y	1371

SEQ ID NO:	70	moltype = AA	length = 1160			
FEATURE		Location/Qualifiers				
source		1..1160				
		mol_type = protein				
		organism = synthetic construct				
SEQUENCE:	70					
MKAKKSFYNQ	KRKFGKRGYR	LHDERIAYSG	GIGSMRSIKY	ELKDSYGIAG	LRNRIADATI	60
SDNKWLWGN	NLNDYLEWRS	SKTDQIEDG	DRESSLLGFW	LEALRLGFVF	SKQSHAPNDF	120
NETALQDLFE	TLDSDLKHVL	DRKKWCDIFK	IGTPKTNQDG	RLKKQIKNLL	KGNKREEIEK	180
TLNESDDELK	EKINRIADV	AKNKSDDKTYI	FKLDPKNTPEK	YPRINDVQVA	FFCHPDPFEEI	240
TERDRTKTDL	LIINRPNKRY	EITENKKDDK	TSNRMALYSL	NQGYIPRVLN	DLFLFVKDNE	300
DDFSQFLSDL	ENFFFSFSNEQ	IKIIKERLKK	LKKYAEPIPG	PKQLADKWDD	YASDFGGKLE	360
SWYSNRKIEL	KKIPESVSDL	RNNLKEIRNV	LKKQNNASKI	LELSQKIIYEY	IRDYGVSF	420
PEIJKFSWIN	KTKDGQKKV	YVAKMADREF	IEKLDLWMAD	LRSQNLNEYINQ	DNKVSFKKKG	480
KKIEELGVLD	FALNKAKKN	STKNGNENGWQ	KLSESIIQSAP	LFFGEGNVR	NEEVYNLKD	540
LFSEIKNVEN	ILMSSEAEDL	KNKIEYKED	GAKKGNYVNLN	VLARFYARFN	EDGYGGWNKV	600
KTVLENITARE	AGTDFSKYGN	NNRNAGRFY	LNGRERQVFT	LIKFEKSITV	EKILELVLKLP	660
SLLDEAYRDL	VNENKNHKL	DVIQLSKTIM	ALVLSHSDKE	KQIGGGYIHS	KLSGYNALIS	720
KRDFISRYSV	QTNTNGTQCKL	AIGKGKSKKG	NEIDRYYAF	QFFKNDSSK	NLKVIKNNSH	780
KNIDFNDNEN	KINALQVYSS	NYQIQFLDNN	FEKHQGKTS	LEVGGSFTIA	EKSSTIDWWS	840
SNPRVGFKRS	DTEEKRVFVS	QPFTLIPDDE	DKERRKERM	TKTKNRFIGID	IGEYGLAWSL	900
I EVDNGDKNN	RGIRQLESGF	ITDNQQVNLK	KNVKSWRQNQ	IRQTFTSPDT	KIARLRESLI	960
GSYKNQLES	MVAKKANLSF	EYEVSGFEGV	GKRVAKIYDS	IKRGSRVRKKD	NNSQNDQSWG	1020
KKGWRSGEKI	KGKELFGPVK	DAMRPNVDGL	GMKIVKRKYL	KLDLRLDWVSR	YGNMAFICP	1140
YVDCHHISHA	DKQAAFPNIAV					1160

SEQ ID NO:	71	moltype = AA	length = 1285			
FEATURE		Location/Qualifiers				
source		1..1285				
		mol_type = protein				
		organism = synthetic construct				
SEQUENCE:	71					
MKLTRRRISG	NSVDQKITAA	FYRDMQSQGLL	YYDSEDNDCT	DKVIESMDFE	RSWRGRILKN	60
GEDDKNPFYM	FVKGLVGSND	KIVCEPIDVD	SDPDNLIDLI	NKNLTGFGRN	LKAPDSNDTL	120
ENLIRKIQAG	IPEEEVLP	KKIKEMIQKD	IVNRKEQLLK	SIKNNRIPFS	LEGSKLVPST	180

-continued

---

KKMKWLFKLI	DVPNKTTFNEK	MLEKYWEIYD	YDKLKANITN	RLDKTDDKKAR	SISRRAVSEEL	240
REYHKNLRTN	YNRFVSGDRP	AAAGLDNGGS	KYNPDKEEF	LFLKEVEQYF	KKYFPVKSKH	300
SNKSCKDSL	DKYKNYCSYK	VVKKEVNRSI	INQLVAGLIIQ	QGKLYYYFFYY	NDTWQEDFLN	360
SYGLSYIQVE	EAFKKSVMTS	LSWGINRLTS	FFIDDSNTVK	FDDITTKAK	EAIESNYFNK	420
LRTCSRMDH	FKEKLAFFYP	YVVKDKKDRP	DDDIENLIVL	VKAIESVSY	LRNRTFHFK	480
SSLLELLKEL	DDKNSQNKI	DYSVAEAEFIK	RDIENLYDVF	REQIRSLGIA	EYYKADMISD	540
CFKTCGLEFA	LYSPKNSLMP	AFKNVYKRG	NLNKAYIRD	GPKETGDQGQ	NSYKALEEYR	600
ELTWYIEVKN	NDQSYNAYKN	LLQLIYYHAF	LPREVRENAL	ITDFINRTK	WNRKTEERL	660
NTKNNKKHKN	FDENDDITVN	TYRIESIPDY	QGESLDDYLYK	VLORKQMARA	KEVNEKEGN	720
NNYIQFIRDV	VVWAFFGAYLE	NKLKNYKNE	QPPLSKENI	LNDTLKELFY	EEKVKSPFNI	780
KCRFSISTFI	DNKGKSTDNT	SAEAVTDGK	EDEBKDKKNIK	RKDLLCFYL	LRLLDENEIC	840
KLQHQFIKYR	CSLKERRFP	NRTKLEKETE	LLAELEELME	LVRFTMPSIP	EISAKAESGY	900
DTMIKKYFKD	FIEKKVFKNP	KTSNLYYHSD	SKTPVTRKYM	ALLMRSAPLH	LYKDFIKGYY	960
LTKKCECLEY	IKLSNIKI	QNSNLNHEQ	LERIKLKESEK	QNGKDSL	KKDFYKVKEY	1020
VENLEQVARY	KHLQHFKINF	SLYRIFRIHV	DIAARMVGYT	QDWERDMHFL	FKALVYNGVL	1080
EERFEAIFN	MNDDNNNDGR	VKKIQQNNLN	KNRELVMSLC	WNKKLNKNF	GAIIWKRNP	1140
AHLNHFTQTE	QNSKSSLES	INSLRILLAY	DRKRQNAVTK	TINDLLNDY	HIRIKWEGRV	1200
DEQOIQYFNK	EKEDIENEPI	IHLKHLHKKD	CYIYKNSYMF	DKQKEWICNG	IKEEVYDKSI	1260
LKCIGNLKF	DYEDKKNSSA	NPKHT				1285

SEQ ID NO: 72            moltype = AA    length = 1344

FEATURE                    Location/Qualifiers  
source                    1..1344  
                          mol\_type = protein  
                          organism = synthetic construct

SEQUENCE: 72

MLRRDKEVKK	LYNVPNQIQV	GTPKPKWWND	EKLSPEENER	RAQQKNIKMK	NYKWREACSK	60
YVESSQRIIN	DVIFYSYRKA	KNKLRYMRKN	EDILKKMQE	EKLSKFGGK	LEDFTVAYTLR	120
KSLVVSKYDT	QEFDLSAAMV	VFLECIGKNN	ISDHHEREIVC	KLLELIRKDF	SKLDPNVKGS	180
QGANIVRSVR	NQNMMIVQPQG	DRFLFPQVYA	KENETVTNNK	VEKEGLNEFL	LNYANLDEK	240
RAESLRKLRR	ILDVAFVSPN	HYECKMDITL	SDNIEKEKFN	VWEKHECGKK	ETGLFVDIPD	300
VIMEAEAAERN	KLDLVVKEKRE	RKVLNDRVKE	QNIICYRTR	AVVEKYSNNE	PLFFENNAIN	360
QYWIHHIENA	VERILKNCKA	GKLFKLRLKG	LAEKVWKDAI	NLISIKYIAL	GKAVYNFALD	420
DIWKDKKNKE	LGIVDERIRN	GITSFDYEMI	KAHENLQREL	AVDIAFSVNN	LARAVCDMSN	480
LGNKESDFLL	WKRKNLADKL	KNKDDMASVS	AVLQFFGGKS	SWDINIFKDA	YKGKKKYYNE	540
VRFIDDLRKA	IYCARNENFH	FKTALVNDEK	WNTLELFGKIF	ERETEFCCLNV	EKDRFYSNNL	600
YMFYQVSELN	NMMLDHLYSRS	VSRAAQVPSY	NSVIVRTA	EYITNVLGQ	KPSYDADTLG	660
KWYSACYLL	KEIYYSFLQ	SDRALQLFEK	SVKTLSWDDK	KQQRADVNFK	DHFSDIKSAC	720
TSLAQVCQIY	MTEYNNQNNQ	IKKVRSSNTS	IDFOPVYQHY	KVLLKKAIAN	AFADYLKNNK	780
DLFFIGKPF	KANEIREIDK	EQFLPDWTSR	KYEALCIEVS	GSOELQKWIYI	VGKFLNARSL	840
NLMVGSMSRY	IQYVTDIKRR	AASIGNELHV	SVHDEVEKVE	WVQVIEVC	LASRTSNQFE	900
DYFNDKDDY	RYLKSYVDFS	NVDMPSSEYSA	LVDPSNEEQS	DLYVDPKNPK	VNRNIVHSKL	960
FAADHILRDI	VEPVSKDNIE	EFYSQKAETA	YCKIKGKETI	AEEQKAVLKY	QKLKNRVELR	1020
DIVEYGEIIN	ELGQLINWS	FMRERDLYF	QLGFHYDCLR	NDSKHPGEGYK	NIKVDENSIK	1080
DAILYQIIGM	YVNGVTVYAP	EKDGDKLKBQ	CVKGGVGKV	SAFHRYSKYL	GLNEKTLYNA	1140
GLEIFEVVAE	HEDIINLRNG	IDHFKYYLGD	YRSMLSLIYE	VFDRFFTYDI	KYQKVNLL	1200
QNIILLRNHVI	VEPILESGFF	TIGEQTKPGA	KLSIRS1KSD	TFQYKVKG	GT LITDAKDERY	1260
LETIRKILYY	AENEEDNLKK	SVVVTNADKY	EKNKESDDQ	KQKEKKNIDN	KGKKNEETKS	1320
DAEKNNNNERL	SYNPFANLNF	KLSN				1344

SEQ ID NO: 73            moltype = RNA    length = 10

FEATURE                    Location/Qualifiers  
source                    1..10  
                          mol\_type = other RNA  
                          organism = synthetic construct  
misc\_feature            1  
                          note = 6-Fluorescein moiety attached  
misc\_feature            10  
                          note = Iowa Black FQ moiety attached

SEQUENCE: 73

ttttttttt

                          10

SEQ ID NO: 74            moltype = RNA    length = 10

FEATURE                    Location/Qualifiers  
source                    1..10  
                          mol\_type = other RNA  
                          organism = synthetic construct  
misc\_feature            1  
                          note = IRDye 700 moiety attached  
misc\_feature            10  
                          note = IRDye QC-1 Quencher moiety attached

SEQUENCE: 74

ttttttttt

                          10

SEQ ID NO: 75            moltype = RNA    length = 10

FEATURE                    Location/Qualifiers

---

-continued

---

```

source          1..10
               mol_type = other RNA
               organism = synthetic construct
misc_feature   1
               note = TYE 665 moiety attached
misc_feature   10
               note = Iowa Black RQ moiety attached
SEQUENCE: 75
ttttttttt          10

SEQ ID NO: 76
FEATURE
source          1..10
               mol_type = other RNA
               organism = synthetic construct
misc_feature   1
               note = Alexa Fluor 594 (NHS Ester) moiety attached
misc_feature   10
               note = Iowa Black RQ moiety attached
SEQUENCE: 76
ttttttttt          10

SEQ ID NO: 77
FEATURE
source          1..10
               mol_type = other RNA
               organism = synthetic construct
misc_feature   1
               note = ATTO TM 633 (NHS Ester) moiety attached
misc_feature   10
               note = Iowa Black RQ moiety attached
SEQUENCE: 77
ttttttttt          10

SEQ ID NO: 78
FEATURE
source          1..10
               mol_type = other DNA
               organism = synthetic construct
misc_feature   1
               note = 6-Fluorescein moiety attached
misc_feature   10
               note = Iowa Black FQ moiety attached
misc_feature   1..4
               note = DNA fragment
misc_feature   5..6
               note = RNA fragment
misc_feature   7..10
               note = DNA fragment
SEQUENCE: 78
ttttttttt          10

SEQ ID NO: 79
FEATURE
source          1..10
               mol_type = other DNA
               organism = synthetic construct
misc_feature   1
               note = IRDye 700 moiety attached
misc_feature   10
               note = IRDye QC-1 Quencher moiety attached
misc_feature   1..4
               note = DNA fragment
misc_feature   5..6
               note = RNA fragment
misc_feature   7..10
               note = DNA fragment
SEQUENCE: 79
ttttttttt          10

SEQ ID NO: 80
FEATURE
source          1..10
               mol_type = other DNA
               organism = synthetic construct
misc_feature   1

```

---

-continued

---

```

misc_feature          note = TYE 665 moiety attached
10
misc_feature          note = Iowa Black RQ moiety attached
1..4
misc_feature          note = DNA fragment
5..6
misc_feature          note = RNA fragment
7..10
misc_feature          note = DNA fragment
SEQUENCE: 80
ttttttttt                                         10

SEQ ID NO: 81      moltype = DNA length = 10
FEATURE
source
1..10
mol_type = other DNA
organism = synthetic construct
1
note = Alexa Fluor 594 (NHS Ester) moiety attached
1..4
note = DNA fragment
5..6
note = RNA fragment
7..10
note = DNA fragment
10
note = Iowa Black RQ moiety attached
SEQUENCE: 81
ttttttttt                                         10

SEQ ID NO: 82      moltype = DNA length = 10
FEATURE
source
1..10
mol_type = other DNA
organism = synthetic construct
1
note = ATTO TM 633 (NHS Ester) moiety attached
1..4
note = DNA fragment
5..6
note = RNA fragment
7..10
note = DNA fragment
10
note = Iowa Black RQ moiety attached
SEQUENCE: 82
ttttttttt                                         10

SEQ ID NO: 83      moltype = DNA length = 10
FEATURE
source
1..10
mol_type = other DNA
organism = synthetic construct
1
note = 6-Fluorescein moiety attached
10
note = Iowa Black FQ moiety attached
SEQUENCE: 83
ttttttttt                                         10

SEQ ID NO: 84      moltype = DNA length = 12
FEATURE
source
1..12
mol_type = other DNA
organism = synthetic construct
1
note = 6-Fluorescein moiety attached
10
note = Iowa Black FQ moiety attached
SEQUENCE: 84
ttttttttt tt                                         12

SEQ ID NO: 85      moltype = DNA length = 14
FEATURE
source
1..14
mol_type = other DNA

```

-continued

---

misc_feature	organism = synthetic construct	
	1	
misc_feature	note = 6-Fluorescein moiety attached	
	10	
SEQUENCE: 85	note = Iowa Black FQ moiety attached	
ttttttttt tttt		14
 SEQ ID NO: 86	moltype = DNA length = 12	
FEATURE	Location/Qualifiers	
source	1..12	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 86		
ttttttttt tt		12
 SEQ ID NO: 87	moltype = RNA length = 50	
FEATURE	Location/Qualifiers	
source	1..50	
	mol_type = other RNA	
	organism = synthetic construct	
SEQUENCE: 87		
gcaccccca aatgaaggg gactaaaaca agaagaattc agattttaa		50
 SEQ ID NO: 88	moltype = DNA length = 12	
FEATURE	Location/Qualifiers	
source	1..12	
	mol_type = other DNA	
	organism = synthetic construct	
misc_feature	1	
misc_feature	note = Alexa Fluor 647 (NHS Ester) moiety attached	
	10	
SEQUENCE: 88	note = Iowa Black RQ moiety attached	
ttttttttt tt		12
 SEQ ID NO: 89	moltype = AA length = 1277	
FEATURE	Location/Qualifiers	
source	1..1277	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 89		
MIKSYDDFIN HYAIQKTLRF ELQPIGKTRE HIQKNGIIEH DEALEQKYQI VKKIIDRFHR 60		
KHIDEALSLV DFTKHLDTFK TIQEYLKRG KTDREKKELE ELSADLRKRV VSFLEGKVEG 120		
DARFAVKVQR YGILFDAKIK DKDFESTAC DDIEKDAIEA FKRFATYFTG FHENRKNMYS 180		
ADEESTAIAY RVINENLPRF LENKARFEKI LDVIEKTKLT ELKQNLKTEL GHGSVSDIFR 240		
IEYFNVLITQ AGITYNTIL GGKTKENGK VQGLNEIINL FNQKNDTML PLLKPLYKQI 300		
LSEEEYSTSFK LKAIDTFWNE HIEKSIPHVT GKRENILLKI ENLCKKLEKY 360		
KDKEIEKLFI ERKNLSTVSH QVYQWNIIR DALRMHLEMN NKNIKEKDID KYLDNDAFAW 420		
KEIKDSIKIY KEHVEDAKEL DENGIVKYFS SMSINEEDDE KEYSISLIKN INEKYNNVKS 480		
ILEEDRTGKS DLHQDKEKVIA IIKEFLDSLQ QLQWFLKLLY VTVPLDEKDY DFYNELEMYH 540		
DTLLPLTTLY NKVRNYMTRK PYSVEKFKL FEKSTLLDGW DKNKERANLG VILRKGNNYY 600		
LGIMMNKKYND IFDSIPGLTT TDYCEKMNYK LLPGPNKMLP KVFFSKKGVQ FYKPSKEILR 660		
IYEKGEFKKD KSGNFKKESL HTLIDFYKEA IAKNEDWKIF KFKFKNTREY ADISQFYKDV 720		
EROGYKISFD KIDWEYILLV VDEGKLFLFK IYNKDFSPS Y KGKPNLHTIY WKNIFSHDNL 780		
NNVVYKLNNG AEVFYRKNSI DYPEEYINNG HHKEELNGKF YNPIIKDKRY AEDKFLFHVP 840		
ITMFNFISKEE KRVNQLACEY LSTTKEDVHI IGIDRGERHL LYLSLIDKEG NIKKQLSLNT 900		
IKNENYDKEI DYHAKLDEKE KKREEARKNW DVNIKELK EGYLSQVHQ IAKLMVEYKA 960		
ILVMEDLNTG FKGRGRFKVKEI QVYQKFEKMM IDKLNYLVLK DRQATOQPGGS LKAYGLASSL 1020		
ESFKKLGKQC GMIFYVPPSAY TSKIDPTTGF YNFIRVDVPN LEKGKEFFSK FEKIIYNTKE 1080		
DYPEFHCKYG KFVPEPKNKA NDRKTKESLT YYNAIKDTWV VVCSTHHERY KIVRNKAGYY 1140		
ESQPVDVTKN LRDIIFSEANI NYSDGKDIKP IIIESNNNAKL LKSIAEQLKL ILAMRYNNKG 1200		
HDDDEKDYIL SPVKNQKGKF FCSDLGDQSL PINADANGAY NIALKGLLI EKIKKQQGKA 1260		
KDLYISNLEW FMFMMRS 1277		
 SEQ ID NO: 90	moltype = AA length = 1276	
FEATURE	Location/Qualifiers	
source	1..1276	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 90		
MIKSYDDFIN HYAIQKTLRF ELQPIGKTRE HIQKNGIIEH DEALEQKYQI VKKIIDRFHR 60		
KHIDEALSLV DFTKHLDTFK TIQEYLKRG KTDREKKELE ELSADLRKLI VSYLUKNVQK 120		
KTQHNLNPBK ERFEILFGKE LFTNEEFFTL AENKEEKKAI QAFKGFTTYF KGFQENRKNM 180		
YSEEDKSTAI AYRIINENMP LFIENIARFQ KVLDVIEKTK LTELKQNLKT ELKGHSVSDI 240		
FRIEYFNNVL TQEGISRYNT IIGGGKTTEG VKIQGLNEII NLHNQQSKDV KLPLLLKPLYK 300		

-continued

---

QILSEEYSTS	FTISAFEKDN	DVLQAIGSFC	NDCIFYAKNN	VNGKAYNLQ	TVQAFNSID	360
TYNDNRDGL	HIERKNLATL	SHQVYGEWINI	LRLDALQIHVE	AYEQKDNGNN	NNYLESKTFS	420
WKALKDALTT	YKSLVVEAQD	IDENGFIAYF	KDMKFKEEID	GKTTSDILIE	NIQTRYKSIE	480
TILQEDRNNK	MNLHQKEKV	ATIKGFLDSV	KYLQWFNLNM	YIASPVDKD	YDFYNELEMV	540
HDTLLPLTLT	YNKVRNMYTR	KPYSVEKFKL	TFEKSTLLDG	WDKNKERANL	GVLRKGNNY	600
YLGIMMNKKY	DIFDSIPGLT	TTDYCEKMNY	KLLPGPNKML	PKVFFSKKGV	QFYKPSQEII	660
RLYNNKEFKK	GDTFNKNSLH	KLINFYKESI	AKTEDWSVFPQ	FKFKNTNDYA	DISQFYKDVE	720
RGQYKISFDK	IDWEYILLLV	DEGKLFLFKI	YNKDFSPYSK	GKPNLHTIYW	KNIFSHDNLN	780
NVVYKLNGEA	EVFYRKKSIE	YPEEILQKGH	HVNELKDKFK	YPIIKDKRYA	EDKFLFHVP	840
TMFNLSKEEK	RVNQLACEYI	STTKEDVHII	GIDRGERHHL	YLSLIDKEGN	IKKQLSLNTI	900
KNENYDKEID	RVVKLACEKEK	KRDEARKNW	VIENIKELKE	GYMSQVIHII	AKMMAEAKAI	960
LIMEDLNIGF	KRGRFRKVEKQ	VYQKFEKMMI	DKLNYLVPKN	KEPLEPGSSL	NAYQLTSKFD	1020
SFKKLGKQSG	FIFYVPSAYT	SKIDPTTGFY	NFIRVDVPNL	EKGKEFFSKF	EKIIYNTKED	1080
YFFFHCKYKG	FVPEPEKNKD	DRKTESLTY	YNAIKDTWV	VCSTHHERYK	IVRNKAGYYE	1140
SQPVDVTKNL	RDIFSEANIN	YSDGKDIP	IIESNNAKL	KSIAEQLKLI	LAMRYNNNGKH	1200
DDEKDYDIL	PVKNQGKFF	CSDLGDQSPL	INADANGAYN	IALKGLLLIE	KIKKQOGKAK	1260
DLYISNLEWF	MFMMRS					1276

SEQ ID NO: 91	moltype = AA	length = 1278				
FEATURE	Location/Qualifiers					
source	1..1278					
	mol_type = protein					
	organism = synthetic construct					
SEQUENCE: 91						
MIKSYDDFIN	HYAIQKTLRF	ELQPIGKTR	HIQKNGIIEH	DEALEQKYQI	VKKIIDRFHR	60
KHIDEALSLV	DFTKHLDTFK	TIQEYLKRG	KTDREKKEBLE	ELSADRKL	VSYLKGNVQ	120
KTQHNLNPKI	BRFEILFGKE	LFTNEEFTL	AENKEEKKAI	QAFKGFTTYF	KGFQENRKNM	180
YSEEDKSTAI	AYRIIENMP	LFIENIARFQ	KVLDVIEKTK	LTELQNLKT	ELKGHSVSDI	240
FRIEYFNNVL	TQEGISRYNT	IIGGGTTTEG	VKIQGLNEII	NLHNQQSKDV	KLPILKPLHK	300
QILSEEYSTS	FIKIAPENDV	EVLKAIDFW	NEHIEKSIHP	VTGKRFNILL	KIENLCKKLE	360
KYKDKEIEK	PIERKLNSTV	SHQVYQWNI	IRDALRMHLE	MNNKNIKEKD	IDKYLDNDAAF	420
AWKEIKDSIK	IYKEHVEDAK	ELDENGIVK	FSSMSINEED	DEKEYSISL	KNINEKNNV	480
KSILEEDRTG	KSDLHQDKEK	VAIIKEFLDS	LKQLQWFLKL	LYVTVPDKE	DYEFYNELEV	540
YYBALLPLNS	LYNKVNRNYMT	RKPYSVEKPK	LNFYSPTLLD	GWDKNKETAN	LSIILKKNGK	600
YYLGMNKEN	NTIFENFPKS	KSNDYYEKMI	YKLLPGPNKML	LPKVFVFSKKG	LEYYKPSKEI	660
LRUYEKGMFK	KDKSGNFKKE	SLHTLIDFYK	EAIAKNEDWK	IFPKFKFNTR	EYEDISQFYR	720
DVEEQGYLII	PEKVDANYVD	KLVEEGELFL	FQIYNKDFSE	NKKSKGNPNL	HTIYWESLFD	780
NQNLKDVVYK	LNGEAEVFYR	KKSIDYPEEI	YNNGHHKEL	NGKFNYPIK	DRRYTQDKFL	840
FHVPIITNLN	AKSDEKVNM	VKNYIAATNE	KIHIIGIDRG	ERNLLYLSLI	DSNGNIVKQQ	900
SLNIIELPKY	QKQIDYHAKL	NEKEKQRLAA	RQNWDVBIENI	KELKEGYLSQ	VIHQIARLMV	960
DYKAILVMED	LNFGFKRGRF	KVEQVYQKF	EKMLIDKLSY	LVFKEKNLCE	PGGSRLRAYQL	1020
SAPFKSFKAL	GKQSGMIFYV	PAQYTSKIDP	TTGFYNFLNI	DVSNLARSKE	TFSKFDKIVY	1080
NIKEDYPEFY	CKMINFESAN	QLTKSQNKA	NABELKEFWI	LCSTHHDRFK	VERKNNQINY	1140
CKINVNEELK	KLLNSKGINY	EKSNDLKSEI	LNIDESCFFK	ELGYLLKILV	SLRYNNNGKKG	1200
SEEQDFILSP	VKNASGKFFC	TLDNNNTLPL	DADANGAYNI	ALKGLMIVQR	VKAGGKLDSL	1260
ISKDDWINFL	IMNKKLPK					1278

SEQ ID NO: 92	moltype = DNA	length = 19			
FEATURE	Location/Qualifiers				
source	1..19				
	mol_type = other DNA				
	organism = synthetic construct				
SEQUENCE: 92					
tttttcatcg	actagacgt			19	

SEQ ID NO: 93	moltype = DNA	length = 23			
FEATURE	Location/Qualifiers				
source	1..23				
	mol_type = other DNA				
	organism = synthetic construct				
SEQUENCE: 93					
cattgatcaa	tgtcataact	tca		23	

SEQ ID NO: 94	moltype = DNA	length = 41			
FEATURE	Location/Qualifiers				
source	1..41				
	mol_type = other DNA				
	organism = synthetic construct				
SEQUENCE: 94					
gcaagaatcc	accatgttgg	ttaaatagcg	attggtgttg	t	41

SEQ ID NO: 95	moltype = DNA	length = 44
FEATURE	Location/Qualifiers	
source	1..44	
	mol_type = other DNA	
	organism = synthetic construct	

---

-continued

---

```

SEQUENCE: 95
cttatggttt tttaacgcaa tggtcgaaga tggtaagatc cgtc          44

SEQ ID NO: 96      moltype = DNA  length = 16
FEATURE           Location/Qualifiers
source            1..16
                  mol_type = other DNA
                  organism = synthetic construct

SEQUENCE: 96
acttactgcg accagg                                16

SEQ ID NO: 97      moltype = DNA  length = 18
FEATURE           Location/Qualifiers
source            1..18
                  mol_type = other DNA
                  organism = synthetic construct

SEQUENCE: 97
ataatactgac cgagtagg                                18

SEQ ID NO: 98      moltype = DNA  length = 22
FEATURE           Location/Qualifiers
source            1..22
                  mol_type = other DNA
                  organism = synthetic construct

SEQUENCE: 98
gtttatttc aaatagcgat tg                                22

SEQ ID NO: 99      moltype = DNA  length = 23
FEATURE           Location/Qualifiers
source            1..23
                  mol_type = other DNA
                  organism = synthetic construct

SEQUENCE: 99
cattgatgaa tgtcataact tca                                23

SEQ ID NO: 100     moltype = DNA  length = 43
FEATURE           Location/Qualifiers
source            1..43
                  mol_type = other DNA
                  organism = synthetic construct

SEQUENCE: 100
atcgttaata gtgtgcaaga attcaagtgtt gtaaggatcc tgg          43

SEQ ID NO: 101     moltype = DNA  length = 44
FEATURE           Location/Qualifiers
source            1..44
                  mol_type = other DNA
                  organism = synthetic construct

SEQUENCE: 101
cttatggttt tttaacgcaa tggtcgaaga tggtaagatc cgtc          44

SEQ ID NO: 102     moltype = DNA  length = 16
FEATURE           Location/Qualifiers
source            1..16
                  mol_type = other DNA
                  organism = synthetic construct

SEQUENCE: 102
tactgcgacc aggatc                                16

SEQ ID NO: 103     moltype = DNA  length = 16
FEATURE           Location/Qualifiers
source            1..16
                  mol_type = other DNA
                  organism = synthetic construct

SEQUENCE: 103
tatactgacc gagtag                                16

SEQ ID NO: 104     moltype = DNA  length = 19
FEATURE           Location/Qualifiers
source            1..19
                  mol_type = other DNA
                  organism = synthetic construct

SEQUENCE: 104
tttttcatcg actagacgt                                19

SEQ ID NO: 105     moltype = DNA  length = 23

```

---

-continued

---

FEATURE	Location/Qualifiers
source	1..23
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 105	
cattgatgaa tgtcataact tca	23
SEQ ID NO: 106	moltype = DNA length = 38
FEATURE	Location/Qualifiers
source	1..38
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 106	
gcaagaatcc accatgttgg aatagcgatt ggtgttgt	38
SEQ ID NO: 107	moltype = DNA length = 40
FEATURE	Location/Qualifiers
source	1..40
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 107	
tggttttta acgcaatggt cgaagatggt aagatccgtc	40
SEQ ID NO: 108	moltype = DNA length = 16
FEATURE	Location/Qualifiers
source	1..16
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 108	
acttactgcg accagg	16
SEQ ID NO: 109	moltype = DNA length = 18
FEATURE	Location/Qualifiers
source	1..18
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 109	
ataatactgac cgagtagg	18
SEQ ID NO: 110	moltype = DNA length = 22
FEATURE	Location/Qualifiers
source	1..22
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 110	
gttttattctc aaatagcgat tg	22
SEQ ID NO: 111	moltype = DNA length = 23
FEATURE	Location/Qualifiers
source	1..23
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 111	
cattgatgaa tgtcataact tca	23
SEQ ID NO: 112	moltype = DNA length = 43
FEATURE	Location/Qualifiers
source	1..43
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 112	
atcgttaata gtgtgcaaga attcagtggtt gtaaggatcc tgg	43
SEQ ID NO: 113	moltype = DNA length = 44
FEATURE	Location/Qualifiers
source	1..44
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 113	
cttatggttt tttaacgcaa tggtcgaaga tggtaagatc cgtc	44
SEQ ID NO: 114	moltype = DNA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = other DNA
	organism = synthetic construct

---

-continued

---

```

SEQUENCE: 114
catgttggtt aacttactgc                                         20

SEQ ID NO: 115      moltype = DNA  length = 18
FEATURE          Location/Qualifiers
source           1..18
                  mol_type = other DNA
                  organism = synthetic construct

SEQUENCE: 115
atactactgac cgagtagg                                         18

SEQ ID NO: 116      moltype = DNA  length = 22
FEATURE          Location/Qualifiers
source           1..22
                  mol_type = other DNA
                  organism = synthetic construct

SEQUENCE: 116
gtttatttc tcaatagcgat tg                                         22

SEQ ID NO: 117      moltype = DNA  length = 23
FEATURE          Location/Qualifiers
source           1..23
                  mol_type = other DNA
                  organism = synthetic construct

SEQUENCE: 117
cattgatgaa tgtcataact tca                                         23

SEQ ID NO: 118      moltype = DNA  length = 43
FEATURE          Location/Qualifiers
source           1..43
                  mol_type = other DNA
                  organism = synthetic construct

SEQUENCE: 118
atcgttaata gtgtgcaaga attcgttgg ttaaggatcc tgg                                         43

SEQ ID NO: 119      moltype = DNA  length = 44
FEATURE          Location/Qualifiers
source           1..44
                  mol_type = other DNA
                  organism = synthetic construct

SEQUENCE: 119
cttatggttt tttaacgcaa tggtcgaaga tggtaagatc cgtc                                         44

SEQ ID NO: 120      moltype = DNA  length = 19
FEATURE          Location/Qualifiers
source           1..19
                  mol_type = other DNA
                  organism = synthetic construct

SEQUENCE: 120
gttgggttaac ttactgcga                                         19

SEQ ID NO: 121      moltype = DNA  length = 16
FEATURE          Location/Qualifiers
source           1..16
                  mol_type = other DNA
                  organism = synthetic construct

SEQUENCE: 121
tatactgacc gagtag                                         16

SEQ ID NO: 122      moltype = DNA  length = 22
FEATURE          Location/Qualifiers
source           1..22
                  mol_type = other DNA
                  organism = synthetic construct

SEQUENCE: 122
gtttatttc tcaatagcgat tg                                         22

SEQ ID NO: 123      moltype = DNA  length = 23
FEATURE          Location/Qualifiers
source           1..23
                  mol_type = other DNA
                  organism = synthetic construct

SEQUENCE: 123
cattgatgaa tgtcataact tca                                         23

SEQ ID NO: 124      moltype = DNA  length = 41

```

---

-continued

---

FEATURE	Location/Qualifiers
source	1..41
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 124	
cgttaatagt gtgcagaat tcagtggt aaggatctg g	41
SEQ ID NO: 125	moltype = DNA length = 40
FEATURE	Location/Qualifiers
source	1..40
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 125	
tggttttta acgcaatggt cgaagatggt aagatccgtc	40
SEQ ID NO: 126	moltype = DNA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 126	
catgttgtt aacttactgc	20
SEQ ID NO: 127	moltype = DNA length = 18
FEATURE	Location/Qualifiers
source	1..18
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 127	
ataatactgac cgagtagg	18
SEQ ID NO: 128	moltype = DNA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 128	
gtttaactttt agtccgtatc	20
SEQ ID NO: 129	moltype = DNA length = 18
FEATURE	Location/Qualifiers
source	1..18
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 129	
ttactctttt gtatcgca	18
SEQ ID NO: 130	moltype = DNA length = 44
FEATURE	Location/Qualifiers
source	1..44
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 130	
acaattattt gtgagtgta tatgacagtc aacactatgt tagc	44
SEQ ID NO: 131	moltype = DNA length = 38
FEATURE	Location/Qualifiers
source	1..38
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 131	
cggtaatctt gtcgatgagt ctctcggtta ttccggtag	38
SEQ ID NO: 132	moltype = DNA length = 19
FEATURE	Location/Qualifiers
source	1..19
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 132	
gtgaaagcta tatcgacag	19
SEQ ID NO: 133	moltype = DNA length = 22
FEATURE	Location/Qualifiers
source	1..22
	mol_type = other DNA
	organism = synthetic construct

---

-continued

---

SEQUENCE: 133 gacatatagtt attcttgat tc	22
SEQ ID NO: 134 FEATURE source moltype = DNA length = 18 Location/Qualifiers 1..18 mol_type = other DNA organism = synthetic construct	
SEQUENCE: 134 tccgttatcca gtcaaacac	18
SEQ ID NO: 135 FEATURE source moltype = DNA length = 18 Location/Qualifiers 1..18 mol_type = other DNA organism = synthetic construct	
SEQUENCE: 135 tccgttatcca gtcaaacac	18
SEQ ID NO: 136 FEATURE source moltype = DNA length = 18 Location/Qualifiers 1..18 mol_type = other DNA organism = synthetic construct	
SEQUENCE: 136 tccgttatcca gtcaaacac	18
SEQ ID NO: 137 FEATURE source moltype = DNA length = 18 Location/Qualifiers 1..18 mol_type = other DNA organism = synthetic construct	
SEQUENCE: 137 tccgttatcca gtcaaacac	18
SEQ ID NO: 138 FEATURE source moltype = DNA length = 19 Location/Qualifiers 1..19 mol_type = other DNA organism = synthetic construct	
SEQUENCE: 138 gtgaaagcta tatcgacag	19
SEQ ID NO: 139 FEATURE source moltype = DNA length = 22 Location/Qualifiers 1..22 mol_type = other DNA organism = synthetic construct	
SEQUENCE: 139 gacatatagtt attcttgat tc	22
SEQ ID NO: 140 FEATURE source moltype = DNA length = 20 Location/Qualifiers 1..20 mol_type = other DNA organism = synthetic construct	
SEQUENCE: 140 gtttactttt agtccgtatc	20
SEQ ID NO: 141 FEATURE source moltype = DNA length = 18 Location/Qualifiers 1..18 mol_type = other DNA organism = synthetic construct	
SEQUENCE: 141 ttactctttt gtatcgca	18
SEQ ID NO: 142 FEATURE source moltype = DNA length = 38 Location/Qualifiers 1..38 mol_type = other DNA organism = synthetic construct	
SEQUENCE: 142 attgggtgagt gtcataatgac agtcaaacact atgttagc	38
SEQ ID NO: 143 moltype = DNA length = 38	

---

-continued

---

FEATURE	Location/Qualifiers
source	1..38
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 143	
cggtaatctt gtcgatgagt	ctctcgggta ttccggtag
	38
SEQ ID NO: 144	moltype = DNA length = 19
FEATURE	Location/Qualifiers
source	1..19
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 144	
gtgaaagcta tatcgacag	
	19
SEQ ID NO: 145	moltype = DNA length = 22
FEATURE	Location/Qualifiers
source	1..22
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 145	
gacatatatgtt attcttgat	tc
	22
SEQ ID NO: 146	moltype = DNA length = 18
FEATURE	Location/Qualifiers
source	1..18
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 146	
tccgtatcca gtcaaacac	
	18
SEQ ID NO: 147	moltype = DNA length = 18
FEATURE	Location/Qualifiers
source	1..18
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 147	
ttactctttt gtatcgca	
	18
SEQ ID NO: 148	moltype = DNA length = 42
FEATURE	Location/Qualifiers
source	1..42
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 148	
aattctaca ttattggta gtgtatgtta gcatttctgt cg	
	42
SEQ ID NO: 149	moltype = DNA length = 38
FEATURE	Location/Qualifiers
source	1..38
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 149	
cggtaatctt gtcgatgagt	ctctcgggta ttccggtag
	38
SEQ ID NO: 150	moltype = DNA length = 19
FEATURE	Location/Qualifiers
source	1..19
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 150	
gtgaaagcta tatcgacag	
	19
SEQ ID NO: 151	moltype = DNA length = 22
FEATURE	Location/Qualifiers
source	1..22
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 151	
gacatatatgtt attcttgat	tc
	22
SEQ ID NO: 152	moltype = DNA length = 18
FEATURE	Location/Qualifiers
source	1..18
	mol_type = other DNA
	organism = synthetic construct

---

-continued

---

SEQUENCE: 152	tccgtatcca gtcaacac	18
SEQ ID NO: 153	moltype = DNA length = 18	
FEATURE	Location/Qualifiers	
source	1..18	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 153		
ttactctttt gtatcgca		18
SEQ ID NO: 154	moltype = DNA length = 42	
FEATURE	Location/Qualifiers	
source	1..42	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 154		
aattctaca ttatggta gtgtatgtt gcatttctgt cg		42
SEQ ID NO: 155	moltype = DNA length = 38	
FEATURE	Location/Qualifiers	
source	1..38	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 155		
cggtaatctt gtcgatgagt ctctcggta ttcggtag		38
SEQ ID NO: 156	moltype = DNA length = 19	
FEATURE	Location/Qualifiers	
source	1..19	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 156		
tgaaagcta tatcgacag		19
SEQ ID NO: 157	moltype = DNA length = 22	
FEATURE	Location/Qualifiers	
source	1..22	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 157		
gacatatagt attcttgat tc		22
SEQ ID NO: 158	moltype = DNA length = 18	
FEATURE	Location/Qualifiers	
source	1..18	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 158		
tccgtatcca gtcaacac		18
SEQ ID NO: 159	moltype = DNA length = 18	
FEATURE	Location/Qualifiers	
source	1..18	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 159		
ttactctttt gtatcgca		18
SEQ ID NO: 160	moltype = DNA length = 38	
FEATURE	Location/Qualifiers	
source	1..38	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 160		
ctacaattat tggtgagtgt atgttagcat ttctgtcg		38
SEQ ID NO: 161	moltype = DNA length = 38	
FEATURE	Location/Qualifiers	
source	1..38	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 161		
cggtaatctt gtcgatgagt ctctcggta ttcggtag		38
SEQ ID NO: 162	moltype = DNA length = 19	

---

-continued

---

FEATURE	Location/Qualifiers
source	1..19
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 162	
gtgaaagcta tatcgacag	19
SEQ ID NO: 163	moltype = DNA length = 22
FEATURE	Location/Qualifiers
source	1..22
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 163	
gacatatagt attcttgat tc	22
SEQ ID NO: 164	moltype = DNA length = 18
FEATURE	Location/Qualifiers
source	1..18
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 164	
tattcgaactg gtgtcagg	18
SEQ ID NO: 165	moltype = DNA length = 18
FEATURE	Location/Qualifiers
source	1..18
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 165	
tgtatttgtt ggcgcgtaa	18
SEQ ID NO: 166	moltype = DNA length = 46
FEATURE	Location/Qualifiers
source	1..46
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 166	
ctatcaactcc tattaaaggc tggatacggtt cgatataaac atatgc	46
SEQ ID NO: 167	moltype = DNA length = 43
FEATURE	Location/Qualifiers
source	1..43
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 167	
tctgtataat aagatgcaaa ggcaaaaata attcacatat tgg	43
SEQ ID NO: 168	moltype = DNA length = 18
FEATURE	Location/Qualifiers
source	1..18
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 168	
tgaaccgatc cactgatg	18
SEQ ID NO: 169	moltype = DNA length = 18
FEATURE	Location/Qualifiers
source	1..18
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 169	
gtagtaaaga tgcttagtg	18
SEQ ID NO: 170	moltype = DNA length = 18
FEATURE	Location/Qualifiers
source	1..18
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 170	
tattcgaactg gtgtcagg	18
SEQ ID NO: 171	moltype = DNA length = 18
FEATURE	Location/Qualifiers
source	1..18
	mol_type = other DNA
	organism = synthetic construct

---

-continued

---

SEQUENCE: 171	tgtattgtgt ggcgttaa	18
SEQ ID NO: 172	moltype = DNA length = 39	
FEATURE	Location/Qualifiers	
source	1..39	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 172	tcactcctat taaaggctgc gttcgatata aacatatgc	39
SEQ ID NO: 173	moltype = DNA length = 43	
FEATURE	Location/Qualifiers	
source	1..43	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 173	tctgtataat aagatgaaa ggcaaaaata attcacatata tgg	43
SEQ ID NO: 174	moltype = DNA length = 18	
FEATURE	Location/Qualifiers	
source	1..18	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 174	tgaaccgatc cactgatg	18
SEQ ID NO: 175	moltype = DNA length = 18	
FEATURE	Location/Qualifiers	
source	1..18	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 175	tgtagtaaaga tgcttagtg	18
SEQ ID NO: 176	moltype = DNA length = 20	
FEATURE	Location/Qualifiers	
source	1..20	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 176	aatacatgtc ttagatgttc	20
SEQ ID NO: 177	moltype = DNA length = 20	
FEATURE	Location/Qualifiers	
source	1..20	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 177	tcttgataat cttagatgagt	20
SEQ ID NO: 178	moltype = DNA length = 42	
FEATURE	Location/Qualifiers	
source	1..42	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 178	agaaataactt tagatacggt cgtaaaactt aatctctctc ct	42
SEQ ID NO: 179	moltype = DNA length = 43	
FEATURE	Location/Qualifiers	
source	1..43	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 179	agttaaatta gatttcgaaac gaaggatatac cagaaaaagac ggt	43
SEQ ID NO: 180	moltype = DNA length = 21	
FEATURE	Location/Qualifiers	
source	1..21	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 180	ggtatcaatt ttgtttaaga g	21
SEQ ID NO: 181	moltype = DNA length = 21	

---

-continued

---

FEATURE	Location/Qualifiers
source	1..21
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 181	
tcaaataaacccggcg t	21
SEQ ID NO: 182	moltype = DNA length = 21
FEATURE	Location/Qualifiers
source	1..21
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 182	
caatacatgtt ctttagatgtt c	21
SEQ ID NO: 183	moltype = DNA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 183	
tcttgataat cttgatgagt	20
SEQ ID NO: 184	moltype = DNA length = 38
FEATURE	Location/Qualifiers
source	1..38
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 184	
gaaaataactt agatacgtgc caaaacttaat ctctctcc	38
SEQ ID NO: 185	moltype = DNA length = 42
FEATURE	Location/Qualifiers
source	1..42
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 185	
gtttaatttag atttcgaacg aaggataacc agaaaagacg gt	42
SEQ ID NO: 186	moltype = DNA length = 21
FEATURE	Location/Qualifiers
source	1..21
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 186	
ggtatcaatttttgtttaaga g	21
SEQ ID NO: 187	moltype = DNA length = 21
FEATURE	Location/Qualifiers
source	1..21
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 187	
tcaaataaacccggcg t	21
SEQ ID NO: 188	moltype = DNA length = 18
FEATURE	Location/Qualifiers
source	1..18
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 188	
ttcaggaaatccgcatct	18
SEQ ID NO: 189	moltype = DNA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 189	
ccccataact ttattcatgg	20
SEQ ID NO: 190	moltype = DNA length = 41
FEATURE	Location/Qualifiers
source	1..41
	mol_type = other DNA
	organism = synthetic construct

---

-continued

---

```

SEQUENCE: 190
ccattagtc cataatacag ttctatgaaa ggaaagaatg t           41

SEQ ID NO: 191      moltype = DNA  length = 41
FEATURE          Location/Qualifiers
source           1..41
                  mol_type = other DNA
                  organism = synthetic construct

SEQUENCE: 191
caaaatatgt agaaaaggag gtttgaatga gatattctga g           41

SEQ ID NO: 192      moltype = DNA  length = 18
FEATURE          Location/Qualifiers
source           1..18
                  mol_type = other DNA
                  organism = synthetic construct

SEQUENCE: 192
ttgacgctgg agaaatga                                     18

SEQ ID NO: 193      moltype = DNA  length = 20
FEATURE          Location/Qualifiers
source           1..20
                  mol_type = other DNA
                  organism = synthetic construct

SEQUENCE: 193
agaacatgga taaggtagc                                     20

SEQ ID NO: 194      moltype = DNA  length = 18
FEATURE          Location/Qualifiers
source           1..18
                  mol_type = other DNA
                  organism = synthetic construct

SEQUENCE: 194
ttcagggaaa tcgcatct                                     18

SEQ ID NO: 195      moltype = DNA  length = 20
FEATURE          Location/Qualifiers
source           1..20
                  mol_type = other DNA
                  organism = synthetic construct

SEQUENCE: 195
cccatatact ttattcatgg                                     20

SEQ ID NO: 196      moltype = DNA  length = 39
FEATURE          Location/Qualifiers
source           1..39
                  mol_type = other DNA
                  organism = synthetic construct

SEQUENCE: 196
ccattagtc cataatacag tctatgaaa gaaagaatg               39

SEQ ID NO: 197      moltype = DNA  length = 39
FEATURE          Location/Qualifiers
source           1..39
                  mol_type = other DNA
                  organism = synthetic construct

SEQUENCE: 197
caaaatatgt agaaaaggag gtgaatgaga tattctgag               39

SEQ ID NO: 198      moltype = DNA  length = 17
FEATURE          Location/Qualifiers
source           1..17
                  mol_type = other DNA
                  organism = synthetic construct

SEQUENCE: 198
tgacgctgg agaardtga                                     17

SEQ ID NO: 199      moltype = DNA  length = 20
FEATURE          Location/Qualifiers
source           1..20
                  mol_type = other DNA
                  organism = synthetic construct

SEQUENCE: 199
agaacatgga taaggtagc                                     20

SEQ ID NO: 200      moltype = DNA  length = 18

```

---

-continued

---

FEATURE	Location/Qualifiers
source	1..18
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 200	
tcgcatctc tatgaaag	18
SEQ ID NO: 201	moltype = DNA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 201	
cccatatact ttattcatgg	20
SEQ ID NO: 202	moltype = DNA length = 44
FEATURE	Location/Qualifiers
source	1..44
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 202	
ccattagctc cataatacag tgaaaagaatg tattcatttc tcca	44
SEQ ID NO: 203	moltype = DNA length = 41
FEATURE	Location/Qualifiers
source	1..41
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 203	
caaaaatatgt agaaaaggag gtttgaatga gatattctga g	41
SEQ ID NO: 204	moltype = DNA length = 17
FEATURE	Location/Qualifiers
source	1..17
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 204	
actgacgaga ttgacgc	17
SEQ ID NO: 205	moltype = DNA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 205	
agaacatgga taaggtagc	20
SEQ ID NO: 206	moltype = DNA length = 18
FEATURE	Location/Qualifiers
source	1..18
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 206	
tcgcatctc tatgaaag	18
SEQ ID NO: 207	moltype = DNA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 207	
cccatatact ttattcatgg	20
SEQ ID NO: 208	moltype = DNA length = 38
FEATURE	Location/Qualifiers
source	1..38
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 208	
cattagctcc ataatacaga gaatgtattc atttctcc	38
SEQ ID NO: 209	moltype = DNA length = 34
FEATURE	Location/Qualifiers
source	1..34
	mol_type = other DNA
	organism = synthetic construct

---

-continued

---

SEQUENCE: 209		
tatgtagaaa aggaggatgaa tgagatattc tgag		34
SEQ ID NO: 210	moltype = DNA length = 17	
FEATURE	Location/Qualifiers	
source	1..17	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 210		
actgacgaga ttgacgc		17
SEQ ID NO: 211	moltype = DNA length = 19	
FEATURE	Location/Qualifiers	
source	1..19	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 211		
gaacatggat aaggttagc		19
SEQ ID NO: 212	moltype = DNA length = 18	
FEATURE	Location/Qualifiers	
source	1..18	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 212		
tgttatccaa cttgacaa		18
SEQ ID NO: 213	moltype = DNA length = 18	
FEATURE	Location/Qualifiers	
source	1..18	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 213		
atactcgagt ctctgctg		18
SEQ ID NO: 214	moltype = DNA length = 48	
FEATURE	Location/Qualifiers	
source	1..48	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 214		
aggttattat ctggatcatc tatcaactaa tatcagagag atagaaga		48
SEQ ID NO: 215	moltype = DNA length = 39	
FEATURE	Location/Qualifiers	
source	1..39	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 215		
gtaaatccca cagaactaat ggttcgccat ctatcctct		39
SEQ ID NO: 216	moltype = DNA length = 17	
FEATURE	Location/Qualifiers	
source	1..17	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 216		
gcgtttgtct catatgt		17
SEQ ID NO: 217	moltype = DNA length = 20	
FEATURE	Location/Qualifiers	
source	1..20	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 217		
tctaggact tatggacaac		20
SEQ ID NO: 218	moltype = DNA length = 19	
FEATURE	Location/Qualifiers	
source	1..19	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 218		
agttccatta tctaagct		19
SEQ ID NO: 219	moltype = DNA length = 18	

---

-continued

---

FEATURE	Location/Qualifiers
source	1..18
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 219	
atactcgagt ctctgctg	18
SEQ ID NO: 220	moltype = DNA length = 41
FEATURE	Location/Qualifiers
source	1..41
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 220	
gttttgtca tatgttcttc tataccatg ttatccaact t	41
SEQ ID NO: 221	moltype = DNA length = 39
FEATURE	Location/Qualifiers
source	1..39
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 221	
gttaaatccca cagaactaat gttcgccatc tattcctctg	39
SEQ ID NO: 222	moltype = DNA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 222	
ctctgatatt aagatttgtc	20
SEQ ID NO: 223	moltype = DNA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 223	
tcttagtact tatggacaac	20
SEQ ID NO: 224	moltype = DNA length = 18
FEATURE	Location/Qualifiers
source	1..18
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 224	
ggaatccatg ttatccaa	18
SEQ ID NO: 225	moltype = DNA length = 18
FEATURE	Location/Qualifiers
source	1..18
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 225	
ccaaatactcg agtctctg	18
SEQ ID NO: 226	moltype = DNA length = 43
FEATURE	Location/Qualifiers
source	1..43
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 226	
tcatctatca gcgtttgtct cttgacaaat cttatatca gag	43
SEQ ID NO: 227	moltype = DNA length = 39
FEATURE	Location/Qualifiers
source	1..39
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 227	
gttaaatccca cagaactaat gttcgccatc tattcctctg	39
SEQ ID NO: 228	moltype = DNA length = 17
FEATURE	Location/Qualifiers
source	1..17
	mol_type = other DNA
	organism = synthetic construct

---

-continued

---

SEQUENCE: 228 catatgttct tcttatct	17
SEQ ID NO: 229 FEATURE source moltype = DNA length = 20 Location/Qualifiers 1..20 mol_type = other DNA organism = synthetic construct	
SEQUENCE: 229 tctaggtaact tatggacaac	20
SEQ ID NO: 230 FEATURE source moltype = DNA length = 18 Location/Qualifiers 1..18 mol_type = other DNA organism = synthetic construct	
SEQUENCE: 230 tgtttatccaa cttgacaa	18
SEQ ID NO: 231 FEATURE source moltype = DNA length = 18 Location/Qualifiers 1..18 mol_type = other DNA organism = synthetic construct	
SEQUENCE: 231 aataactcgag tctctgct	18
SEQ ID NO: 232 FEATURE source moltype = DNA length = 43 Location/Qualifiers 1..43 mol_type = other DNA organism = synthetic construct	
SEQUENCE: 232 tcaggaggtt attatctgga gagagataga agaacatatg aga	43
SEQ ID NO: 233 FEATURE source moltype = DNA length = 38 Location/Qualifiers 1..38 mol_type = other DNA organism = synthetic construct	
SEQUENCE: 233 gtaaatccca cagactaattc ttgcggatct atcctctg	38
SEQ ID NO: 234 FEATURE source moltype = DNA length = 17 Location/Qualifiers 1..17 mol_type = other DNA organism = synthetic construct	
SEQUENCE: 234 tcatctatca gcgtttg	17
SEQ ID NO: 235 FEATURE source moltype = DNA length = 20 Location/Qualifiers 1..20 mol_type = other DNA organism = synthetic construct	
SEQUENCE: 235 tctaggtaact tatggacaac	20
SEQ ID NO: 236 FEATURE source moltype = RNA length = 161 Location/Qualifiers 1..161 mol_type = other RNA organism = synthetic construct	
SEQUENCE: 236 gaccgcgttca ccaagtgcgtg tcccttaggg gattagcact tgagtgaagg tgggctgctt gcatcagcct aatgtcgaga agtgcttct tcggaaatgaa accctcgaaa caaattcatt gaaagaatgaa aggaatgcaa ctgcgtaca cctatcgta a	60 120 161
SEQ ID NO: 237 FEATURE source moltype = RNA length = 161 Location/Qualifiers 1..161 mol_type = other RNA organism = synthetic construct	
SEQUENCE: 237 gaccgcgttca ccaagtgcgtg tcccttaggg gattagcact tgagtgaagg tgggctgctt	60

---

-continued

---

```

gcatcagcct aatgtcgaga agtgcgttc tcggaaagta accctcgaaa caaattcatt 120
gaaaatga aggaatgcaa caccaatagt gagttcgccg a 161

SEQ ID NO: 238      moltype = RNA length = 161
FEATURE           Location/Qualifiers
source            1..161
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 238
gaccgcgttca ccaagtgcgt tcccttaggg gattagcact tgagtgaagg tgggctgctt 60
gcatcagcct aatgtcgaga agtgcgttc tcggaaagta accctcgaaa caaattcatt 120
gaaaatga aggaatgcaa ccgttaggtg tagccgataa a 161

SEQ ID NO: 239      moltype = RNA length = 161
FEATURE           Location/Qualifiers
source            1..161
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 239
gaccgcgttca ccaagtgcgt tcccttaggg gattagcact tgagtgaagg tgggctgctt 60
gcatcagcct aatgtcgaga agtgcgttc tcggaaagta accctcgaaa caaattcatt 120
gaaaatga aggaatgcaa cttaaacgatt gtcgaccctc t 161

SEQ ID NO: 240      moltype = RNA length = 161
FEATURE           Location/Qualifiers
source            1..161
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 240
gaccgcgttca ccaagtgcgt tcccttaggg gattagcact tgagtgaagg tgggctgctt 60
gcatcagcct aatgtcgaga agtgcgttc tcggaaagta accctcgaaa caaattcatt 120
gaaaatga aggaatgcaa ctgaatgc catgtactac g 161

SEQ ID NO: 241      moltype = RNA length = 161
FEATURE           Location/Qualifiers
source            1..161
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 241
gaccgcgttca ccaagtgcgt tcccttaggg gattagcact tgagtgaagg tgggctgctt 60
gcatcagcct aatgtcgaga agtgcgttc tcggaaagta accctcgaaa caaattcatt 120
gaaaatga aggaatgcaa cactttacg cctctggcgt t 161

SEQ ID NO: 242      moltype = RNA length = 161
FEATURE           Location/Qualifiers
source            1..161
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 242
gaccgcgttca ccaagtgcgt tcccttaggg gattagcact tgagtgaagg tgggctgctt 60
gcatcagcct aatgtcgaga agtgcgttc tcggaaagta accctcgaaa caaattcatt 120
gaaaatga aggaatgcaa caaatggcc aaagcgccgtt a 161

SEQ ID NO: 243      moltype = RNA length = 161
FEATURE           Location/Qualifiers
source            1..161
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 243
gaccgcgttca ccaagtgcgt tcccttaggg gattagcact tgagtgaagg tgggctgctt 60
gcatcagcct aatgtcgaga agtgcgttc tcggaaagta accctcgaaa caaattcatt 120
gaaaatga aggaatgcaa cgatccactg ctgaacagat a 161

SEQ ID NO: 244      moltype = RNA length = 161
FEATURE           Location/Qualifiers
source            1..161
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 244
gaccgcgttca ccaagtgcgt tcccttaggg gattagcact tgagtgaagg tgggctgctt 60
gcatcagcct aatgtcgaga agtgcgttc tcggaaagta accctcgaaa caaattcatt 120
gaaaatga aggaatgcaa ctttatcgcc tacacctatc g 161

SEQ ID NO: 245      moltype = RNA length = 161
FEATURE           Location/Qualifiers
source            1..161

```

---

-continued

---

```

mol_type = other RNA
organism = synthetic construct
SEQUENCE: 245
gaccgcctca ccaagtgcgt tcccttaggg gattagcact tgagtgaagg tgggctgctt 60
gcatcagcct aatgtcgaga agtgcgttct tcggaaagta accctcgaaa caaattcatt 120
gaaaagaatga aggaatgcaa ctaaccgcgt ttggccattt t 161

SEQ ID NO: 246      moltype = RNA length = 161
FEATURE           Location/Qualifiers
source            1..161
mol_type = other RNA
organism = synthetic construct
SEQUENCE: 246
gaccgcctca ccaagtgcgt tcccttaggg gattagcact tgagtgaagg tgggctgctt 60
gcatcagcct aatgtcgaga agtgcgttct tcggaaagta accctcgaaa caaattcatt 120
gaaaagaatga aggaatgcaa catagctgtt cagcagtggat t 161

SEQ ID NO: 247      moltype = RNA length = 161
FEATURE           Location/Qualifiers
source            1..161
mol_type = other RNA
organism = synthetic construct
SEQUENCE: 247
gaccgcctca ccaagtgcgt tcccttaggg gattagcact tgagtgaagg tgggctgctt 60
gcatcagcct aatgtcgaga agtgcgttct tcggaaagta accctcgaaa caaattcatt 120
gaaaagaatga aggaatgcaa ctcgcgaaac tcactattgg t 161

SEQ ID NO: 248      moltype = DNA length = 20
FEATURE           Location/Qualifiers
source            1..20
mol_type = genomic DNA
organism = Monkeypox virus
SEQUENCE: 248
tcggctacac ctatcgtaa 20

SEQ ID NO: 249      moltype = DNA length = 20
FEATURE           Location/Qualifiers
source            1..20
mol_type = genomic DNA
organism = Monkeypox virus
SEQUENCE: 249
accaataatgtt agttcgccgaa 20

SEQ ID NO: 250      moltype = DNA length = 20
FEATURE           Location/Qualifiers
source            1..20
mol_type = genomic DNA
organism = Monkeypox virus
SEQUENCE: 250
cgataggtgtt agcccgataaa 20

SEQ ID NO: 251      moltype = DNA length = 20
FEATURE           Location/Qualifiers
source            1..20
mol_type = genomic DNA
organism = Monkeypox virus
SEQUENCE: 251
ttaacgatttgc tcgaccctct 20

SEQ ID NO: 252      moltype = DNA length = 20
FEATURE           Location/Qualifiers
source            1..20
mol_type = genomic DNA
organism = Monkeypox virus
SEQUENCE: 252
tgaagatgcc atgtactacg 20

SEQ ID NO: 253      moltype = DNA length = 20
FEATURE           Location/Qualifiers
source            1..20
mol_type = genomic DNA
organism = Monkeypox virus
SEQUENCE: 253
acttttacgc ctctggcggtt 20

SEQ ID NO: 254      moltype = DNA length = 20

```

-continued

---

FEATURE source	Location/Qualifiers 1..20 mol_type = genomic DNA organism = Monkeypox virus	
SEQUENCE: 254 aaaaatggcca aaggcggtta		20
SEQ ID NO: 255	moltype = DNA length = 20	
FEATURE source	Location/Qualifiers 1..20 mol_type = genomic DNA organism = Monkeypox virus	
SEQUENCE: 255 gatccactgc tgaacagacta		20
SEQ ID NO: 256	moltype = DNA length = 20	
FEATURE source	Location/Qualifiers 1..20 mol_type = genomic DNA organism = Monkeypox virus	
SEQUENCE: 256 tttatcgctt acacctatcg		20
SEQ ID NO: 257	moltype = DNA length = 20	
FEATURE source	Location/Qualifiers 1..20 mol_type = genomic DNA organism = Monkeypox virus	
SEQUENCE: 257 taaccccgctt tggccatttt		20
SEQ ID NO: 258	moltype = DNA length = 20	
FEATURE source	Location/Qualifiers 1..20 mol_type = genomic DNA organism = Monkeypox virus	
SEQUENCE: 258 atagctgttc agcagtggat		20
SEQ ID NO: 259	moltype = DNA length = 20	
FEATURE source	Location/Qualifiers 1..20 mol_type = genomic DNA organism = Monkeypox virus	
SEQUENCE: 259 tcgcccgaact cactatttgt		20
SEQ ID NO: 260	moltype = AA length = 1228	
FEATURE source	Location/Qualifiers 1..1228 mol_type = protein organism = unidentified	
SEQUENCE: 260 MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLVED EKRAEDYKGV KKLDRYYLS 60 FINDVLHSIK LKNLNNYISI FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYSKSLFK 120 KDIETILDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKS TSIAFRCINENL 180 TRYISNMDIF EKVDAIFDKH EVQEIKEKIL NSDYDVEDFV EGEFFNFVLT QEGIDVYNAY 240 IGGFVTESGE KIKGLNEYIN LYNAQTKQKL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV 300 LEVFRNTLNK NSEIFSSIKE LEKLFKNFDE YSSAGIFVKN GPAISTISKD IFGEVNVIIRD 360 KWNAEYDDIH LKKKAVTPEK YEDDRRSKTF KIGSFSLEQL QEYADADLSV VEKLKEIIQ 420 KVDIEYKVYG SSEKLFADDF VLEKSLKND AVVAIMKDLL DSVKSFENYI KAFFFEGKET 480 NRDESFYGDF VLAYDILLKV DHIYDAIRNY VTQKPYSDKD FKLYFQNPQF MGGWDKDDET 540 DYRATILRYG SKYLYLAIMDK KYAKCLQKID KDDVNGNYEK INYKLLPGPN KMLPKVFESK 600 KMMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET 660 EKYKDIAGFY KVFESASQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH 720 TMYFKLFLDE MNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKTTTLS 780 YDVYKDKRFS EDQYELHIPI AINKCPKNIF KINTEVRL KHDDNPYVIG IDRGERNLLY 840 IVVVDGKGNI VEQYSLNEII NNFNGIRIKT DYHSLLDKKE KERFEARQNW TSIENIKELK 900 AGYISQVVKH ICELVEKYDA VIALEDLNSG FKNSRVKVEK QVYQKFEMK IDKLNMYMDK 960 KSNPCATGGA LKGYQITNKF ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLLKTKYTS 1020 IADSKKFISS PDRIMVPEE DLFEFALDYK NFSRTDADYI KKWKLYSYGN RIRIFRNPKK 1080 NNVFDWEEVC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS 1140 ITGRRTDVDFL ISPVKNSDGI FYDSRNYEAQ ENAILPKNAD ANGAYNIARK VLWAIGQFKK 1200 AEDEKLDKVK IAISNKEWLE YAQTSVHK 1228		
SEQ ID NO: 261	moltype = AA length = 1307	

-continued

---

FEATURE	Location/Qualifiers
source	1..1307
	mol_type = protein
	organism = Acidaminococcus sp.
SEQUENCE: 261	
MTQFEGPTNL YQVSKTLRFF LIPOGKTLKH IQBQGFIBED KARNDHYKEL KPIIDRIYKT 60	
YADQCLQLVQ LDWENLSAAI DSYRKETKEE TRNALIEEQAA TYRNNAIHDF IGRTDNLTDA 120	
INKRHAEIYK GLFKAELFNG KVLQQLGTVT TTEHENALLR SFDKFITYFS GFYENRKNVF 180	
SAEDISTAIP HRIVQDNFPK FKENCHIFTR LITAVPSLRE HFENVKKAIG IFVSTSIEEV 240	
FSFPFYNLQQ TQTQIDLYNQ LLGGISREAG TEKIKGLNEV LNLAIQKNDE TAIIIASLPH 300	
RFIPLFKQIL SDRNTLKSIL EEFKSDEEVII QSFCKYKTLRN RENENVLETAE ALFNELNISID 360	
LTHIFISHKK LETISSLALCD HWDTLRLNAYL ERRISELTGK ITKSAKEKVQ RSLKHEDINL 420	
QEIIASAAGE LSEAFQKQTS EILSHAHAAAL DQPLPTTLKK QEEKEILKSQ LDSLLGLYHL 480	
LDWFVADESN EVDPEFSRAL TGKLEMESP LSFYNNKARNY ATTKKPYSVEK FKLNFQMPML 540	
ASGWDVNKEK NNGAILFVNK GLYYLIMPK QKGRYKALSF EPTEKTSEGF DKMYDYFPD 600	
AAKMPIPKST QLKAVTAHFQ THTPTILLSN NFIEPLEITK EIYDLNNPEK EPKKQATAYA 660	
KKTGDQKGYR EALCKWIDFT RDPLSKYTKT TSIDLSSLRP SSQYKDLGEY YAELNPLLYH 720	
ISFQRIAEKE IMDAVETGKL YLFQIYNNKDF AKGHHGKPNL HTLYWTGLFS PENLAKTSIK 780	
LNGQAELFYR PKSRSMKRMH RLGEKMLNKK LKDQKTPIPD TLYQELYDYY NHRLSHDLSD 840	
EARALLPNVI TKEVSHEIIK DRRFTSDKFF FHVPITLNQY AANSPSKFNQ RVNAYLKEHP 900	
ETPIIGIDRG ERNLIYITVI DSTGKILEQR SLNTIQFDY QKKLDNREKE RVAARQAWSV 960	
VGTIKAQKQG YLSQVIHEIV DLMIHYQAVV VLENLNFGFK SKRTGIAEKA VYQQFEKMLI 1020	
DKLNCLVLKD HESRKHFKLEG FDFLHYDVKT GDFILHFKMN RNLSFQRGLP GFMPAWDIVF 1080	
DPPFWKTIKN HESRKHFKLEG FDFLHYDVKT GDFILHFKMN RNLSFQRGLP GFMPAWDIVF 1140	
EKNETQFDAK GTPFIAGKRI VPVIENHRFT GRYRDLYPAN ELIALLEEKKG IVFRDGSNIL 1200	
PKLLENDSSH AIDTMVALIR SVLQMRNSNA ATGEDYINSP VRDLNGVCFD SRFQNPEWPM 1260	
DADANGAYHI ALKGQLLNNH LKESKDLKLQ NGISNQDWLA YIQELRN 1307	
SEQ ID NO: 262	moltype = AA length = 1300
FEATURE	Location/Qualifiers
source	1..1300
	mol_type = protein
	organism = Francisella tularensis
SEQUENCE: 262	
MSIYQEfvNK YSLSKTLRFE LIPQGKTLLEN IKARGLILDD EKRAKDYKKA KQIIDKYHQF 60	
FIEBILSSVC ISEDLLQNYS DVYFKLKKSD DDNLQKDPKS AKDTIKKQIS EYIKDSEKFK 120	
NLFNQNLLIDA KKGQESDLIL WLKQSKDNGI ELFKANSDT DIDEALEIJK SFKGWTTYFK 180	
GHFHENRKNVY SSNDIPTSII YRIVDDNLPK FLENKAKYES LKDKAPEAIN YEQIKKDLAE 240	
ELTFDIDYKT SEVNQRVFSL DEVPEIANP NYLNQSGITF PNTIIGGKFV NGENTKRKGI 300	
NEYINLYSQO INDKTLLKYYK MSVLFQKILS DTESKSFVID KLEDDSDVVT TMQSFYEQIA 360	
AFKTVEEKSI KETLSSLFDD LKAQKLDLSK IYFKNDKSLT DLSQQVFDDY SVIGTAVLEY 420	
ITQQIAPKNL DNPSKKEQEL IAKKTEKAKY LSLETIYL EEFNKHRDID KQCRFEEILA 480	
NFAAIPMIFD BIAQNKDNL QISIYKQNCQ KKDLLQASAE DDVKAIKDLL DQTNNLLHLK 540	
KIFHISQSED KANILDKDEH FYLVFEECYK ELANIVPLYN KIRNYITQKP YSDEKFKLNF 600	
ENSTLANGWD KMKEDPNTAI LFIKDDKYYL GMVNKKNNKI FDDKAIKENK GEKYKIVYK 660	
LLPGANKMLP KVFFSAKSIIK FYNPSEDILR IRNHSTHTKN GSPQKGYEKF EFNIEDCRKF 720	
IDFYKQSIK HPEWKDFGFI FSDTQROINSI DEFYREVENQ GYKLTFENIS ESYIDSVNVQ 780	
GKLYLFQIYN KDPFSAYSKGR PLNHTLYWKA LFDERNLQDVY VKLNGEAEI FYRKQSIPKK 840	
IITHPAKEAIA NKNKDNPKKE SVFHEYDLIKD KRPTEDKFFF HCPIITINFKS SGANKFNDEI 900	
NLLLKEKAND VHILSIDRGE RHLAYYTLVD GKGNIIQDFT FNIIGNDRMK TNYHDKLAAI 960	
EKDRDSARKD WKKINNIKEM KEGYLSQVUVH EIAKLVIEYN AIVVFEDELNG GFKGRGFVKE 1020	
KQVYQKLEKMQVYKLVF DKNFEDKDTQV VLRAYQLTAP FETFKKMKGQ TGIIYYPAG 1080	
FTSKICPVVTG FVNQLYPKYE SVSKSQEFFS KFDKICYNLDD KGYFEFSFDY KNFGDKAAKG 1140	
KWTIASFGSR LINFRNSDKN HNWDTREVYP TKELEKLKD YSIEYGHGEC IKAACIGESD 1200	
KKPFKAQLTSV LNTILQMRNS KTGTEDLYLI SPVADVNCGF FDSRQAPKNM PQDADANGAY 1260	
HIGLKGLMLL GRIKNNQEGK KLNVLVKNNE YFEFVQNRNN 1300	
SEQ ID NO: 263	moltype = AA length = 1246
FEATURE	Location/Qualifiers
source	1..1246
	mol_type = protein
	organism = Porphyromonas macacae
SEQUENCE: 263	
MKTQHFFEDF TSLYSLSKTI RFELKPIGKT LENIKKNGLI RRDEQRLLDY EKLKKVIDEY 60	
HEDFIANILS SFSFSEELQ SYIQNLSESE ARAKIEKTM RDTLAKAFSED ERYKSIFKKE 120	
LVKKDIPWVC PAYKSLCKKF DNFTTSLVPF HENRKNLNTS NEITASIPYR IVHVNLPKFI 180	
QNIIEALCELQ KKGMDALYLE MMENLRNWP SFVKTTPDLC NLKTYNHLMV QSSISEYNRF 240	
VGGYSTEDGT KHQGINEWIN IYRQRNKEMR LPGLVFLHKQ ILAKVDSSSF ISDTLENDQ 300	
VFCVLRQFRK LFVNNTVSSKE DDAASLKDLF CGLSGYDPEA IYVSDAHLAT ISKNIFDRWN 360	
YISDAIRRKT EVLMPRKES VERYAEKISK QIKKRQSYSL AELEDDLLAHY SEESLPGAFS 420	
LLSYFTSLGG QKYLVSDGEV ILYEEGGSNIW DEVLIARFDL QVILDKDFTE KKLKGDEAV 480	
SVIKKALDSA LRLRKFFDLS SGTGAEIRRD SSFYALYTD MDKLKGLLKM YDKVRNLYLTK 540	
KPYSIEKPKL HDNDPSSLG WDKNKELNNL SVIFRQNQYY YLGIMTPKGK NLFKTLPKLG 600	
AEEMFYEKME YKQIAEPMLM LPKVFVFPKKT KPAFAPDQSV VDIYNKTFK TGQKGFNKKD 660	
LYRLIDFYKE ALTVHEWKLW NFSFSPTEQY RNIGEFFDEV REQAYKVSMV NVPASYIDEA 720	
VENGKLYLFQ IYNKDFSPYS KGIPNLHTLY WKALFSEQNQ SRVYKLCGGG ELFYRKASLH 780	

---

-continued

---

MQDTTVHPKG	ISIHKKNLNK	KGETSLFNYD	LVKDKRFTED	KFFFHVPISI	NYKNKKITNV	840
NQMRVDRYIAQ	NDDLQIIGID	RGERNLLYIS	RIDTRGNLLE	QFSLNVIEST	KGDLRTDYQK	900
ILGDREQERL	RRRQEWSIE	SIKDLKDGYM	SQVVKICNM	VVEHKAIVVL	ENLNLSFMKG	960
RKKVEKSVYE	KFERMLVDKL	NYLVDDKKNL	SNEPGGLYAA	YQLTNPLFSF	EELHYPQSG	1020
ILFFVDPWNT	SLTDPSTGFV	NLLGRINYTN	VGDARKFFDR	FNAIRYDGKG	NILFDLDLSR	1080
FDRVETQRK	LWTLLTFGSR	IAKSKKSGKW	MVERIENLSSL	CFLELFQQFN	IGYRVEKDLK	1140
KAILSQRKE	FYVRLIYLFN	LMMQIRNSDG	EEDYILSPAL	NEKNLQFDSR	LIEAKDLPVD	1200
ADANGAYRNA	RKGGLMVVQRI	KRGDHESIHR	IGRAQWLRYV	QEGIVE		1246

SEQ ID NO: 264	moltype = AA	length = 1373				
FEATURE	Location/Qualifiers					
source	1..1373					
	mol_type = protein					
	organism = Moraxella bovoculi					
SEQUENCE: 264						
MFQDFTHLY PLSKTVRFEL	KPIDRTLEHI	HAKNFLSQDE	TMADMHQVKV	VILODDYHRDF	60	
IADMMGEVKL	TKLAEFYDVY	LKFRKNPKDD	ELQKQLKDLQ	AVLRKEIVKP	IGNGGKYKAG	120
YDRLFGAKLF	KDGKEGLDLA	KFVIAQEGES	SPKLAHLAHF	EKFSTYFTG	HDNRKNMYS	180
EDKHTAIAYR	LIHENLPRFI	DNLQIYTTIK	QKHSALYDQI	INELTASGLD	VSLASHLDGY	240
HKLLTQEGIT	AYNTLLGGIS	GEAGSPKIQG	INELINSHHN	QHCHKSERIA	KLRPLHKQIL	300
SDGMSVSFLP	SKFADDSEMC	QAVNEFYRH	ADVFAKVQSL	FDGFDDHQKD	GIYVEHKNLN	360
ELSKQAFGDF	ALLGRVLDGY	YVDVNPEN	ERPAKAKTDN	AKAKLTKEKD	KFIKGVHSLA	420
SLEQAIIEHYT	ARHDDESVAQ	GKLGYFVKHG	LAGVDNPQIK	IHNNHSTIKG	FLERERPAGE	480
RALPKIISKGK	NPEMQLRLQ	KELLDNALNV	AHPAKLTTK	TTLDNQDGNF	YGEFGVLYDE	540
LAKIPTLYNK	VRDYLSQLPF	STEKYKLNFG	NPTLLNGWDL	NKEKDNGFVI	LQKDGCYLA	600
LILDKAHKKVF	DNAPNTGKSI	YQKMIYKYLE	VRKQFPKVVF	SKEAIAINYH	PSKELVEIKD	660
KGRQRSDDER	LKLYRFILEC	LKIHPKYDKK	FEGAIGDIQL	FKKDKKGKREV	PISEKDLFDK	720
INGIFSSKPK	LEMEDPFFIGE	FKRKNPQSQD	VQDQNIYKKI	DSNDNRKKEN	FYNNHPKFKK	780
DLVRYYYYESM	CKHEEWEESF	EFSKQLQDIG	CYVDVNELFT	EIETRRLNYK	ISFCNINADY	840
IDELVEGQQL	YLFQIYNKDF	SPKAHGKPNL	HTLYFKALFS	EDNLADPIYK	LNGEAQIFYR	900
KASLDMNETT	IHTRAGEVLEN	KNPDNPKKRQ	YTQDKFMLHV	PITMNFGVQG	960	
MTIKEFNKVK	NQSIQOYDEV	NVIGIDRGER	HLLYLTWINS	KGEILEQCSL	NDITTASANG	1020
TQMTTPYHKI	LDKREIERLN	ARVGWGEIET	IKELKSGYLS	HVVHQISQLM	LKYNAIVVLE	1080
DLNFGFKRGR	FKVEKQIYQN	FENALIKKLN	HVLVLDKADD	EIGSYKNALQ	LTNNTFTDLKS	1140
IGKQTGFLFY	VPAWNTSKID	PETGFVDSL	PRYENIAQSQ	AFFGKFDKIC	YNADKDYFEF	1200
HIDYAKFTDK	AKNSRQIWTI	CSHGDKRYVY	DKTANQNKGA	AKGINVNDL	KSLFARHHIN	1260
EKQPNLVMDI	CQNNDKEFHK	SLMYLLKTL	ALRYSNASSD	EDFILSPVAN	DEGVFPNSAL	1320
ADDTQPQNAD	ANGAYHIALK	GLWLLNELKN	SDDLNKVKLA	IDNQWTWNFA	QNR	1373

SEQ ID NO: 265	moltype = AA	length = 1259				
FEATURE	Location/Qualifiers					
source	1..1259					
	mol_type = protein					
	organism = Moraxella bovoculi					
SEQUENCE: 265						
MGIHGVPAA	PQDFTHLYPL	SKTVRFELKP	IGRTLEHIHA	KNFLSQDETM	ADMYQKVVI	60
LDDYHRDFIA	DMMGEVKLT	LAEFYDVYLK	FRKNPKDDGL	QKQLKDLQAV	LRKEIVKPIG	120
SGGKYKTYGYD	RLFGAKLFKD	GKEGLDLAKF	VIAQESESSP	KLAHLAHFEK	FSTYFTGFHD	180
NRKNMYSDED	KHTATAYRLI	HENLPRFIDN	LQIYTTIKQK	HSALYDQIIN	ELTASGLDVS	240
LASHLDGYHK	LLTQEGITAY	NRIIGEVNGY	TNKHQNCICH	SERIAKLRPL	HQKILSDGMG	300
VSPFLPSKFAD	DSEMCOAVNE	FYRHYTVDVFA	KVQSLFDGFD	DHQKDGIVYE	HKNLNELSQ	360
AFGDFALLGR	LDGYYVDDV	NPEFNFRAK	AKTDNAKAKL	TEKDKFVKG	VHSLASLEQA	420
IIEHTARHDD	ESVQAGKLGQ	YFKHGLAGVD	NPIQKIHNNH	STIKGFLER	RPAGERALPK	480
IKSGKNPEMT	QLRQLKELL	NALNVAHFAK	LLTTKTTLDN	QDGNFYGEFG	VLYDELAKIP	540
TLYNKVDRYD	SQKPFSTEKY	KLNPNCPNTL	NGWDLNKEKD	NFGVILQKDG	CYYLALLDKA	600
HKKVFDRNAP	TGKNVYQKVM	YKLLPGNPKM	LPKVNFFAKS	YNGVNPNSAE	LDKYAKGTHK	660
KGDNFNLKDC	HALIDFFKAG	INKHPEWQHF	GFKFSPSTSSY	RDLSDFYREV	EPQGYQVKF	720
DINADYIDEL	VEQGKLYLFQ	IYNKDFSPKA	HGKPNLHTLY	FKALFSEDNL	ADPIYKLN	780
AQIFYRKASL	DMNETTIHRA	GEVLENKNDP	NPKKRQFVYD	IIKDKRYTQD	KFMLHVPITM	840
NFGVQGMFTIK	EFNKKVQNSI	QOYDEVNVIG	IDRGERHLLY	LTVINSKGEI	LEQRSLNNDIT	900
TASANGTQVT	TPYHKILDKE	EIERLNRARV	WGBIETIKEL	KSGYLSHVH	QINOLMLKYN	960
AIIVLEDLNF	GFGRGRFKVE	KQIYQNPENA	LIKKNLHVL	KDKADDEIGS	YKNALQLTNN	1020
FTDLKSIGKQ	TGFLFYWPW	NTSKIDPETY	FV DLLKPRYE	NIAQSQAFGG	KFDKICYNTD	1080
KGYFEFHIDY	AKFTDKAKNS	RQKWAICSHG	DKRYVYDKTA	NQNKGAAKGI	NVNDELKSLF	1140
ARYHINDKQP	NLVMDICQNN	DKEFHKSLSMC	LLKTLALLRY	SNASSDEDFI	LSPVANDEGV	1200
FFNSALADDT	QPNQADANGA	YHIALKGLWL	LNELKNSDDL	NKVKLAI	DNQ	1259

SEQ ID NO: 266	moltype = AA	length = 1269				
FEATURE	Location/Qualifiers					
source	1..1269					
	mol_type = protein					
	organism = Moraxella bovoculi					
SEQUENCE: 266						
MGIHGVPAA	PQDFTHLYPL	SKTVRFELKP	IGKLEHIHA	KNFLNQDETM	ADMYQKVKA	60
LDDYHRDFIA	DMMGEVKLT	LAEFYDVYLK	FRKNPKDDGL	QKQLKDLQAV	LRKEIVKPIG	120
NGGKYKAGYD	RLFGAKLFKD	GKEGLDLAKF	VIAQESESSP	KLAHLAHFEK	FSTYFTGFHD	180

-continued

NRKNMYSDED	KHTAIAYRLI	HENLPRFIDN	LQILATIKQK	HSALYDQIIN	ELTASGLDVS	240
LASHLDGYHK	LLTQEGITAY	NTLLGGISGE	AGSRKIQGIN	ELINSHHNQH	CHKSRIAKL	300
RPLHKQILSD	GMGVSPFLPSK	FADDSEVCQA	VNEFYRHYAD	VFAKVQSLFD	GFDDYQKDGI	360
YVEYKNLNEL	SKQAFGDFAL	LGRVLDGYVV	DVVNPENFER	FAKAKTDNAK	AKLTKEDKF	420
IKGVHSLASL	EQAIEHYTAR	HDDESVQAGK	LQQYFKHGLA	GVDNPIQKI	NNHSTIKGFL	480
ERERPAGERA	LPKIKSDKSP	EIROQLKELLD	NALNVVAHFAK	LLTTKTTLHH	QDGNFYGEFG	540
ALYDELAKIA	TLYNKVRDYL	SQKPFSSTEKY	KLNFGNPPLL	NGWDLNKEKD	NFGVILQKDQ	600
CYLLALLDKA	HKKVFDNAPN	TGKSVYQKMI	YKLLPGPNNK	LPKVFFAKSN	LDYYNPSAEL	660
LDKYAQGTHK	KGDNFNLKDC	HALIDFFKAG	INKHPEWQHF	GFKFESPTSSY	QDLSDFYREV	720
EPOGYQVFV	DINADYINEL	VEQGOLYLFC	IYNKDFSPKA	HGKPNLHTLY	FKALFSEDNL	780
VNPYIYKLNGE	AIFYFRKASL	DMNETTIHRA	GEVLENKNPD	NPKKRQFVYD	IIKDKRYTQD	840
KFMLHVPITM	NFGVQGMTI	EFNKKVNQSI	QOYDEVNVIG	IDRGERHLIY	LTVINSKGEI	900
LEQRSLNDIT	TASANGTQMT	TPYHKILDKR	EIERLNARVG	WGEIETIKEL	KSGYLSHVHH	960
QISQLMLKYN	AIVVLEDLNF	GFKFRFRKVE	KOIQYQNFEN	LIKKLNHHLVL	KDKADDEIGS	1020
YKNALQLTNN	FTDLKSGIKQ	TGFLFVPPAN	NTSKIDPETG	FVDLKPRTYE	NIAQSQAFFG	1080
KFDKICYNAD	RGYFEPHIDY	AKFNDKAKNS	RQIWKICSHG	DKRYVYDFTA	NQNKGATIGV	1140
NVNDELKSLF	TRYHINDKQP	NLVMDICQNN	DKEFHKSMLY	LLKTLLALRY	SNASSDEDFI	1200
LSPVANDEGV	FFNSALADDT	QPQNADANGA	YHIALKGLWL	LNELKNSDDL	NKVLAIDNQ	1260
TWLNFQAQNR						1269

SEQ ID NO: 267                  moltype = AA    length = 1306  
 FEATURE                          Location/Qualifiers  
 source                          1..1306  
 mol\_type = protein  
 organism = Thiomicrospira sp.

SEQUENCE: 267  
 MGJHGVPAAAT KTFDSEFFNL YSLQKTVRFE LKPVGETASF VEDFKNEGLK RVVSEDERRA 60  
 VDVQKVKEII DDYHRDFIEE SLNYFPEQVS KDALEQAPHL YQKLKAAKVE EREKALKWE 120  
 ALQKKLREKV VKCFSDSNKA RFSRIDKKEL IKEDLINWLW AQNREDDIPT VETFNNFTTY 180  
 FTGFHENRKN IYSKDDHATA ISFRLIHENL PKFEDNVISF NKLKEGFPTEL KFDKVKEDLE 240  
 VDYDLKHAFE IEYFVNFTQ AGIDQYNNYL GGKTLLEDGTY KQGMNEQINL FKQQQTRDKA 300  
 RQIPKLIPLF KOILSERTES QSFIPKQFES DQRLFDSLQK LHNNCQDKFT VLQQAILGLA 360  
 EADLKKVFIK TSDLNALSNT IFGNYSVPSD ALNLYKESLK TTKAQAEFEK LPAHSIHDLI 420  
 QYLEQFNSL DAEKQOSTDT VLNVYFIKTE LYSRFIKSTS EAFTQVQPLF ELEALSSKRR 480  
 PPESEDEGAK QOEGFPEQIKR IKAYDLTLM EAVHFAKPLYL VKGRKMIEGL DKDQSFYEAF 540  
 EMAYQELESI IPIYLNKARS YLSRKPKFAD KFKLINFDDNTT LLSGWWDANKE TANASILFKK 600  
 DGLYLYLGIMP KGKTFLFDFY VSSEDSEKLK QRQRQTAEEA LAQDGESYFE KIRYKLLPGA 660  
 SKMLPKVFFS NKNIGFYNPS DDILRIRNTA SHTKNGTPQK GHSKVFPNLN DCHKWMIDFFK 720  
 SSIQKHPWEW SFGFTFSN DDFEADSMAFV EVENQGYVIS FDKIKETYIQ SQVEQGNLYL 780  
 FQIYNKDFSP YSKGKPNLHT LYWKALFEEA NLNNNVAKLNLN GEAEIIFRRH SIKASDKVWH 840  
 PANQAIIDNKN PHTEKTQSTF EYDLVKDKRY TDQKFFFHVPI ISLNPKAQGV SKFNDKVNGF 900  
 LKGNPVDVNII GIDRGERHLL YFTVNVQKGE ILVQESLNTL MSDKGHVNDY QQKLDKKEQE 960  
 RDARKSWTT VENIKELKEG YLSHARKSWTT VENIKELKEG YLSHARKSWTT VENIKELKEG 1020  
 YQKFEKALID KLNLYLVPKEK ELGEVQKALID AYQLTAPFES FKKLKGQSGI LFYVPADYTS 1080  
 KIDPTTGFWN PLDLRQYQSVK KAKQLLSDFN AIRFNSVQNY FEFEIDYKKL TPKRKVGQTQ 1140  
 KWVICTYGDV RYQNRNRNQKG HWETEEVNT EKLKALFASD SKTTTVIDYA NDDNLIDVIL 1200  
 EQDKASFFKE LLWLLKLMTM LRHSSKIKS EDLILSPVKNQ QGEFYDSRKA GEVWPKDADA 1260  
 NGAYHIALKG LWNLQQINQW EKGKTLNLAI KNQDWFSFIQ EKPYQE 1306

SEQ ID NO: 268                  moltype = AA    length = 1214  
 FEATURE                          Location/Qualifiers  
 source                          1..1214  
 mol\_type = protein  
 organism = Butyriribrio sp.

SEQUENCE: 268  
 MGIHGVPAA YQNLTKKYPV SKTIRNELIP IGKTLENIRK NNILES DVKR KQDYEHVKGI 60  
 MDEYHKQOLIN EALDNYMLPS LNQAAEIYLIK KHVDVEDREE FKKTQDLLRR EVTGRLKHEE 120  
 NYTKIGKKDI LDLLEKLPSI SEEDYNALES FRNFYTYFTS YNKVRENLYS DEEKSTVAY 180  
 RILINENLPKF LDNIKSYAFV KAAGVLA DCI EEEEQDALFM VETFNM TLQ EGIDMYNQI 240  
 GKVNSAINLY NQKNNHKVEEF KKIPKMKVLY KQILSDREEV FIGEFKDDET LLSSIGAYGN 300  
 VLMITYLKEK INIFFDALRE SEGKNVYVKN DLSTKTTMSNI VFGWSAFDE LLNQBEYDLAN 360  
 ENKKKDDKYF EKRQKELKKN KSYTLEQMSN LSKEDISPIE NYIERISEDI EKICIYNGEF 420  
 EKIVVVNEHDS SRKLSKNKA VKVVKDYLDS IKELEHDIKL INGSQGLEK NLVVVVGQEE 480  
 ALEQLRPVDS LYNLTRNLYL KKPFS TEKVK LNFNKSTLLN GWDKNKETDN LGILFFKDQK 540  
 YYLGIMNTTA NKA FVN PAA KTENVFKV YKLLPGSNKLPK VFFAKSN IGYYNPSTEL 600  
 YSNYKKGTHK KGPSFSIDDC HNLIDFFKES IKKHEDWSKF GFEFSDTADY RDISEFYREV 660  
 EKQGYKLTFT DIDESYINDL IEKNELYLFQ IYNKDFSEYI KGKLNHLHTLY FMMLFDQRNL 720  
 DNVVYKLNGE AEVYRYPASI AENELVIHKA GEGIKKNPN RAKVKTSTF SYDIVKDKRY 780  
 SKYKFTLHIP ITMNPFGVDEV RRFNDVINNA LRTDDNVNVI GIDRGERNLL YVVVINSSEGK 840  
 ILEQISLNSI INKEYDIETN YHALLDERED DRNKARKDWN TIENIKELKT GYLSQVNV 900  
 AKLVLKYNAYI ICLEDLNFGF KRGQRKVEQ VYQKFEKMLI EKLNLVIDE SREQVSPEKM 960  
 GGALNALQLT SKFKSFAELG KQSGI IYYVP AYLTSKIDPT TGFLVNLFYIK YENIEKAKQF 1020  
 FDGFDFIRFN KKDDMPEFSF DYKSF TQKAC GIRSKWIVYT NGERIIKYPN PEKNNLFDEK 1080  
 VINVTDEIKG LFKQYRIPYE NGEDIKEIII SKAEADFYKR LFRLLHQTLQ MRNSTSDGTR 1140  
 DYIISPVKND RGEFFCSSEFS EGTMPKDADA NGAYNIARKG LWVLEQIRQK DEGEKVNLSM 1200  
 TNAEWLKYAQ LHLL 1214

---

-continued

---

SEQ ID NO: 269	moltype = AA length = 1129
FEATURE	Location/Qualifiers
source	1..1129
	mol_type = protein
	organism = Alicyclobacillus acidoterrestris
SEQUENCE: 269	
MAVKSIKVKL RLDDMPEIRA GLWKLHKEVN AGVRYYTEWL SLLRQEENLYR RSPNGDGEQE 60	
CDKTAEECKA ELLERLRARQ VENGHGRPAG SDDELLQLAR QLYELLVPQA IGAKGDAQQI 120	
ARKFLSPLAD KDAVGGGLGIA KAGNPKPRWR MREAGEPGWE EEKEKAETRK SADRTADVLR 180	
ALADFGLKPL MRVYTDDSEMS SVEWKPLRKQ QAVRTWDRDM FQQAIERMMS WESWNQRVGQ 240	
EYAKLVBQKN RFEQKNFVGQ EHHLVHLVNLQ QODMKEASPQ LESKEQTAAH VTGRALRGSD 300	
KVFEKGKLA PDAPFDLYDA EIKNVQRRNT RRGFSHDLFA KLAEPYEQAL WREDASFLTR 360	
YAVYNSILRK LNHAKMFATP TLPDATAHP WTRFDKLGNN LHQYTFLLFNN FGERRHAIRF 420	
HKLLKVENGK AREVDDVTVP ISMSEQDNLV LPRDPNEPIA LYFRDYGAEQ HFTGEFGGAK 480	
IQCRRDQLAH MHRRRGARDV YLNVSVRVQS QEARGERRP PYAAVFRVLG DHNRAFVHF 540	
KLSDYLAEHP DDGKLGSEGL LSGLRVMVSD LGLRTSASIS VFRVARKDEL KPNSKGRVPF 600	
FFPIKGNNDL VAVHERSQQL KLPGETESKD LRRAIREEROR TLRQLRTQLA YLRLLLVRCGS 660	
EDVGRRERSW AKLIEQPVDA ANHMPDWRE AFENELQKLK SLHGKCSDE WMDAVYESVR 720	
RWVRHMGKSW RDWRKDVRSG ERPKIRGYAK DVVGGSNIEQ IEYLERQYKF LKSWSFFGKV 780	
SGQVIRAEKG SRFAITLREH IDHAKEDRLK KLADRIIMEA LGYVYALDER GKGKWKVAKYP 840	
PCQLLILEEL SEYQFNNDRP PSENQNLQM SHRGVFQBLI NQAQVHDLIV GTMYAASSR 900	
FDARTGAPGI FDARTGAPGI QEHNPEFPW WLNAKFVVEHT LDACPLRADD LIPTGEGEIF 960	
VSPFSAEEGD PHQIHADLNA AQNLQQRLLWS DFDISQIRLR CWDGEVDGEL VLIPRLTGKR 1020	
TADSYSNKVF YTNTGVYYE RERGKKRRKV FAQEKLSEE AELLVEADEA REKSVVLMRD 1080	
PSGIINRGNW TRQKEFWSMV NQRIEGYLVK QIRSVPQLQ SACENTGDI 1129	
SEQ ID NO: 270	moltype = AA length = 1224
FEATURE	Location/Qualifiers
source	1..1224
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 270	
MKKIDNFVGC YPVSKTLRFK AIPIGKTQEN IEKKRLVEED EVRAKDYKAV KKLIDRYHRE 60	
FIEGVLDNVK LDGLEEYML FNKNSDREESD NKKIEIMEER FRRVISKSFK NNEEYKKIFS 120	
KKIIIEELPN YIKDEEKEKL VKGFKGFYTA FVGYAQNREN MYSDEKKSTA ISYRIVVNEM 180	
PRFITNIKVF EAKAKILDVD KINEINEYL NNDYYVDDF NIDFFNYVLN QKGIDIYNAI 240	
IGGIVTGDGR KIQGLNECIN LYNQENKKIR LPQFKPLYKQ ILSESESMSF YIDEIESDDM 300	
LIDMLKESLQ IDSTINNAID DLKVLFNNLF DYDLSGIFIN NGLPITTISN DVYQGWSTIS 360	
DGWNERYDV LNAKDKSEK YFEKRLKEYK KVKSFSISDL QELGGKDSL CIKKINEIISE 420	
MIDDYKSKIE EIQLFLDIKE LEKPLVTDLN KTELKLNKSL GLKRIERYVI PFLGTGKEQN 480	
RDEVFVYGYFI KCIDAIEKD GVYNKTRNYL TKKPYSKDFK KLYFENPQLM GGWDRNKESD 540	
YRSTLLRKNG KYVVAIDKS SSNCMMNTE DENDNYEKIN YKLLPGPNKM LPKVFFSKKN 600	
REYFAPSKEI ERIYSTGTFK KDTNFKVKKC ENLITFYKDS LDRHEDWSKS FDPSFKESSA 660	
YRDISEFYRD VEKQGYRVSF DLLSSNAVNT LVEEGKLYLF QLYNKFSEK SHGIPNLHMT 720	
YFRSLFDNNN KGNIRLNNGA EMFMRRASLN KQDVTVHKAN QPIKKNLNL PKKTTLPHYD 780	
VYKDKRFTED QEYEVHIPITM NKVPPNPNYKI NHMVREQLVK DDNPYVIGID RGERNLIVVV 840	
VVDGQGHIVE QLSLNEIINE NNGISIRTDY HTLILDAKERE RDESRKQWKQ IENIKELKEG 900	
YISQVVKHIC ELVEKYDAV ALEDLNSGFK NSRVKVEQV YQKFEKMLIT KLNYMVDKKK 960	
DYNKPGGVNL GYQLTTQFES FSKMGQTQNGI MFYIPAWLTS KMDPTTGFD LLKPKYKNKA 1020	
DAQKFFSQFD SIRYDNQEDA FVFKVNYTKF PRTDADYNE WEIYTNGER RVFRNPKKNN 1080	
EYDYETVNVS ERMKELFDSY DLLYDKGEEL ETICEMEESK FFEIKEKLF RLTLMQRNSIS 1140	
GRTDVDYLIS PVKNSNGYFY NSNDYKKEGA KYPKDADANG AYNIARKVLW AIEQPKMADE 1200	
DKLDKTKISI KNQEWLEYAQ THCE 1224	
SEQ ID NO: 271	moltype = AA length = 1297
FEATURE	Location/Qualifiers
source	1..1297
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 271	
MATLVSFTQK YQVQKTLRFE LIPQGKTQAN IDAKGFINDD LKRDENYMKV KGVIDELHKN 60	
FIEQTLVNVD YDWRSLATAI KNYRKDRSDT NKKNLEKTQE AARKEIIAWF EGKRGNSAFK 120	
NNQKSFYGKL FPKKELFSEI RSDDLEYDEE TDQDAIACFDK FTTYFVGFHNE NRKNMYSSTE 180	
KSTSVAYRVV NENFSKFLSN CEAFSVLEAV CPNVLVEAQEL ELHLHKAFSD LKLSDFVKVE 240	
AYNKYLQSTQG IDYYNQIIGG ISSAEGVRKI RGVNEVNMNA IQQNDELKVA LRNQKFTMVQ 300	
LFKQILSDRS TLSFVSEQFT SDQEVTIVVK QFNDDIVNNK VLAVVKTLEF NFNSYDLEKI 360	
YINSKELASV SNALLKDWSK IRNAVLENKI IELGANPpkt KISAVEKEVK NKDFSIAELA 420	
SYNDKYLDKE GNDKEICSLA NVVLEAVGAL EIMLAESLPA DLKTLENKNC VKGILDAYEN 480	
LLHLLNRYFKV SAVNDVDLAF YGAFEKVYD ISGVMPLYNK VRNYATKPKY SVEFKFLNFA 540	
MPTLADGWDK NKERDNGSII LLKDCQYYLG VMNPQNKPVN DNAVCNDAKG YQKMYKMF 600	
EISKMTVKCS TQLNAVKAHF EDNTNDFVLD DTDKFISDLT ITKEIYDLNN VLYDGKKFQ 660	
IDYLRNTGDF AGYHKALETW IDFVKEFLSK YRSTAIIYDLT TLLPTNYYEK LDVFYSDVNN 720	
LCYKIDYENI SVEQVNEWPE EGNLYLFKIY NKDFATGSTG KPNLHTMYWN AVFAEENLHD 780	
VVKLNGGAE LFYRPKSNMP KVEHRVGEKL VNRKVNNGEP IADSVHKEIY AYANGKISK 840	
ELSENAQEEL PLAIKDKVHK NITKDKRYLS DKYFFHVPI LNYKANGNPS AFNTKQVQFL 900	

-continued

KNNPDVNIIG IDRGERNLLY VVVIDQQGNI IDKKQVSYNK VNGYDYYEKL NQREKERIEA	960
RQSWGAVGKI KELKEGYLSS VVREIADMMV KYNAIVVMEN LNAGFKRVRG GIAEKAVYQK	1020
FEKMLIDKLN YLVFKDVEAK EAGGVLNAYQ LTDKFDSFEK MGNQSGFLFY VPAAYTSKID	1080
PVIGGFANFES TKHITNTPEAK KEFICSFNSL RYDEAKDKFV LECDLNKFKI VANSHIKNWK	1140
FIIGGKRIVY NSKNKTYMEK YPCEDLKATL NASGIDFSSS EIINLLKNVP ANREYGKLF	1200
ETYWAIMNTL QMRNSNALTG EDYIISAVAD DNEKVFDSSRT CGAELPKDAD ANGAYHIALK	1260
GLYLLQRIDI SEEGEKVDSL IKNEEWFKFV QQKEYAR	1297

SEQ ID NO: 272	moltype = AA length = 1229
FEATURE	Location/Qualifiers
source	1..1229
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 272	
MKEQFINRYP LSKTLRFSLI PVGETENNPF KNLLLKKDKQ RAENEYEVKC YIDRFHKYEI	60
ESVLSKARIE KVNEYANLYW KSNKDDSDIK AMESLENDMR KOISKQLTST EYKKRLFGK	120
ELICEDLPSF LTDKDERETV ECFRSFTTYF KGFNTNREMM YSSDGKSTAI AYRCINDNLP	180
RFLDNVKSQF KVFDNLSDET ITKLNTDLYN IFGRNIEDIF SVDYFEEVLT QSGIEIYNSM	240
IGGYTCSDKT KIQGLNNECIN LYNNQVAKNE KSKKLPLMKP LYKQILSEKD SVSFIPEKFN	300
SDENEVLAHD DYYTGHIGDF DLLTELLQSL NTYANGIFV KSGVAITDIS NGAFNSWNVL	360
RSAWNEKYEA LHPVTSKTI DKYIEKQDKI YKAIFSLSL ELQSLGNENG NEITDWYISS	420
INESNSKIKE AYLQAQKLLN SDYEFKSYNR LYKNEKATEL VKNLLDAIKE FQKLICKPLNG	480
TGKEENRPLF TGKFTSYD SIADIDRLY KVRNYITQKP YSKDKIKLNF DNPQLLGGWD	540
KNKESDRYTV LLHKDGLYYL AVMDKSHSKA FVDAPEITSD KDYDYYEKMEEY KLLPGPNKML	600
PKVFFASKNI DTFQPDSRIL DIRKRESPKK GATFNKAECB EFIDYFKDSI KKHDWSQFG	660
FPKSPTESYN DISEFYREIS DQGYSVRFNK I SKNYIDGLV NNGYIYLFQI YNKDFSKYSK	720
GTPNLHTLYP KMLFLDERSL NVVYKLNGEA EMFYREASING DKEKITHYAN QPIKKNPDPN	780
EKKESVFEYD IVKDKRFTKR QFSLHLPITI NFKAHGQFEL NYDVRKAVKY KDDNYVIGID	840
RGERNLIYIS VINSNGEIVE QMSLNEIISD NGHKVVDYQKL LDTKEKERDK ARKNWTSTVEN	900
IKELKEGYIS QVVKHICELV IKYDAVIAME DLNFGFKRGR FPVEKQVYQK FENMLISKLN	960
LLIDKPTTAK EDGGILLRAYQ LTNKFDGVNK AKQNGIIYFV PAWDTSKIDP ATGFVNLLKP	1020
KCNTSVPEAK KLFETIDDIK YNANTDMFEE YIDYSKFPRC NSDFKKSWTV CTNSSRILTF	1080
RNKEKNKKWD NKQIVLTDEF KSLFNEFGID YKGNLKDSDL SISNADFYRR LIKLLSLTLQ	1140
MRNSITGSTL PEDDYLISPV ANKSGEFYDS RNYKGTAAL PCDADANGAY NIARKALWAI	1200
NVLKDTPDDM LNKAKLSSITN AEWLEYTQK	1229

SEQ ID NO: 273	moltype = AA length = 1470
FEATURE	Location/Qualifiers
source	1..1470
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 273	
MNNPRGAFGG PTNLYLSKTI LRFLKPYLE IPEGEKGKLF GDDKEYYKNC KTYTEYYLKK	60
ANKEYYDNEK VKNTDLQLVN FLHDERIEDA YQVLKPVFDT LHEEFITDSL ESEAACKIDF	120
GNYYGLYEKQ KSEQNQKDEKK KIDKPLETER GKLRLKAFTPI YEAEKGKLUK KAGKEKKDKD	180
ILKESGFKVL IEAGILKYIK NNIDEFADKK LKNNEGKEIT KKDIETALGA ENIEGIFDFG	240
FTYFSGFNQN RENEYSTEER ATAVASRIVD ENLSKFCDDN LLYRKRNENDY LKIFNFLKKNK	300
GDKDLKLKNSK PGKENEPEFI PAYDMKNDK SFPSVADFVNC LSQGEIEKYN AKIANANYLI	360
NYLNQNQKDGN SSKLMSMFKIL YKQIGCGEKK DFIFTIKDNA ELKQILEKAC EAGKKYFIRG	420
KSEDGGVSNI FDFTDYIQSH ENYKGVYWSD KAINTISGKY FANWDTLKNK LGDAKVFNKN	480
TGEDKDADVYK VKPQAVMLSE LFAVLDDNAG EDWREKGIFF KASLFEQDQK KSEIINKANR	540
PSQALLKMIC DDMESLAKNF IDSGDKILK SDRDYQKDEN KQKIKNLWDN ALWINQILKY	600
FKVKANKIKG DSIDARIDSC LDMLVFSSDN PABDYDMIRN YLTQKPODEI NKLKLNFENS	660
SLAGGWDENK EKDNCIIILK DEQDKQYLA MKYENTKVFE QKNSQLYIAD NAAWKKMIYK	720
LPGASKTLP KVFFSKKWTN NRPTPSDIVE IYQKGFSKKE NVDFFNDKKEK DESRKEKNRE	780
KIAELQKTC WMDIRYNDG KIESAKVYVKN EKLAKLIDFY KENLKKYPSE EESWDRLLFAF	840
GFSDTCKSYKS IDQFYIEVDE QGYKLEFVTI NKARLDEVYR DGKYLFEIR SRDNNLVNGE	900
EKTSAKNLTQ IYWNAAFGGD DNPKPLNGEA EIFYRPAAIE NKLNNKKDKN GKEIIDGYRF	960
SKEKFIHFPC ITLNFLKET KINDEKLNAA AKPENGQGCVY FLGIDRGEKH LAYYSLVNQK	1020
GEILEQGTLN LPFLDKNGKS RSIKVKCSF EKDSNGGIKI DDKDNGDKKI EFVCEWNYND	1080
LIDARAGDRD YARKNWTIG TIKELKDGYI SQVVRKIVDL SIYKNTETKE FREMPAFIVL	1140
EDLNIGFKRG RQKIEKQVYQ KLELALAKKL NFLVDDKADI GEIGSVTKAI QLTPPNNF	1200
DMENRKQFGN MLYIRADYTS QTDPATGWRK SIYLKSGSES NVKEQIEKSF FDIREYESGDY	1260
CFEYRDRHKG MWQLYSKNG VSLDRFHGER NNSKNVWESE KQPLNEMLDI LFDEKRFDKS	1320
KSLYEQMFKG GVALTRLPKE INKKDPAWE SLRFVIIILQ QIRNTGKNGD DRNGDFIQSP	1380
VRDEKTGEHF DSRIYLDKEQ KGEKADLPTS GDANGAYNIA RKGIVVAAEH KRGFDFKLYIS	1440
DEEWDTWLAG DEIWDWLKE NRESLTKTRK	1470

SEQ ID NO: 274	moltype = AA length = 1282
FEATURE	Location/Qualifiers
source	1..1282
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 274	
MNGNRIIVYR EFVGVTPVAK TLRNELRPIG HTQEHHIHNG LIQEDELRQE KSTELKNIMD	60
DYYREYIDKS LSGVTDLDFT LLFELMNLVQ SSPSKDNKKA LEKEQSKMRE QICTHMQSDS	120

-continued

---

NYKNIFNAKF	LKEILPDFIK	NYNQYDAKDK	AGKLETLALF	NGFSTYFTDF	FEKRKNVFTK	180
EAVSTSIAYR	IvhensltfL	ANMITSYKKIS	EKALDEI EVI	EKNNQDKMGD	WELNQIFNPD	240
FYNMVLIQSG	IDFYNEICGV	VNAHMNLHYCQ	QTKNNYNLFK	MRKLHKQILA	YTSTSFEVPK	300
MFEDDMSVYN	AVNAFIDETE	KGNIIKGKLD	IVNKYDELDE	KRIYISKDFY	ETLSCFMMSGN	360
WNLITGCVEN	FYDENIHAKG	KSKEEKVKKA	VKEKYKSIIN	DVNDLVEKYI	DEKERNEFKN	420
SNAKQYIREI	SNIITDTETA	HLEYDEHISL	IIESEEKADEM	KKRLDMYMMN	YHWAKAFIVD	480
EVLRDREMFY	SDIDDIYNIL	ENIVPLYNRV	RNYVTQPKYN	SKKIKLNFQS	PTLANGWSQS	540
KEFDNNAILL	IRDNKYYLAJ	FNAKNKPDKK	IIQGNSDKNN	NDNYKMMVY	LLPGANKMLP	600
KVFLSKKGIE	TFKPSDYIIS	GYNAHKHICKT	SENFDISFCR	DLIDYFKNSI	EKHAERWKYE	660
FKPSATDSYN	DISEFYREVME	MQGYRIDWTY	ISEADINKLDA	EEGKIYLFQI	YNKDFAENST	720
GKENLHTHMYF	KNIFSEENLK	DIKKLNQGA	ELFYRRASVK	NPVKHKDSDV	LVNKTYKNQL	780
DNGDVVRIPY	PDDIYNEIYK	MYNGYIKEND	LSEAAKEYLD	KVEVRTAQKD	IVKDYRYTVD	840
KYFIHTPITI	NYKVARTNNV	NDMAVKYIAQ	NDDIHVGIGD	GERNLNLYIS	VIDSHGNIVK	900
QKSYNLLNRY	DYKKKLVEK	KTRYEARKNW	KSIGNIKEBLK	EYGIVSGVHIE	IAMLMVEYNA	960
IIAMEDLNLYG	PKFLRNQSF	QVYQKFESML	INKLNYFASK	GKSDEPGL	LKGYQLTYVP	1020
DNIKNLGKQC	CVIFVYPAAF	TSKIDPSTGF	ISAFNFKSIS	TNASRKQFPM	QFDEIERYCAE	1080
KDMFSFGFDY	NNFTDTNITM	SKTQWTVTYTN	GERLQSEFNM	ARRTGKTKSI	NLTETIKLLL	1140
EDNEINYADG	HDRVRIDMEKM	DEDKNSEFFA	QLLSLYKLTV	QMRNSYTEAE	EQEKGISYDK	1200
ISPVINDEG	EFFDSDSNYKE	SDDKECKMPK	DADANGAYCI	ALKGLYEVLK	IKSEWTEDGF	1260
DRNCLKLPHA	EWLDFIQNKR	YE				1282

---

SEQ ID NO: 275	moltype = AA	length = 1225
FEATURE	Location/Qualifiers	
source	1..1225	
	mol_type = protein	
	organism = synthetic construct	

SEQUENCE: 275					
MIKKIDSFVN YPLSKTLRFS	LIPVGKTEDN	FNAKLLLEED	EKRAIEYEKV	KRYIDRYHKH	60
FIETVLANFH LDDVNEAYEL	YYKAGKDDKD	LKYMEKLEGK	MRKSISAFT	KDKKYKEIFG	120
QEIIKNILPE FLENEDEKES	VKMFQGFFTY	FTGFNDNRKN	MYTHEAQTTA	ISYRCINENL	180
PKFLDNQSF	IMMNKLDEVCM	DLYGVYQAQDM	FCTDYFSFVL	SQSGIDRYNN	240
IIGGYVDDKG VKİQGİNEYİ	NLYNQVQDĘK	NKRLPLMKKL	YKQILIEKES	ISFİPEKFES	300
DNIVINAISD YYHHNNVENLF	DDFNKLNFNEF	SEYDDNGIFV	TSGLAVTDIS	NAVFGSWNII	360
SDSWNEEYKD SHPMKTTNA	EKYYEDMKE	YKKNLSFTIA	ELQRGEAGC	NDECKGDIKE	420
YYKTTVAEKI ENIKNAYEIS	KDLLASDYKE	SNDKKLCKND	SAISLLKNLL	DSIKDLEKTI	480
KPLLGTGKEE NKDDDVGFKE	TNLYMEISEI	DRLYDKRVMY	VTQKPYSKDK	IKLNFENPQH	540
LGGWDKNKER DYRSVLLKKE	DKYYLAIMDK	SNNKAFIDFP	DDGECYEKIE	YKLLPGPNKM	600
LPKVFFASSN IEYFAPSJKI	LEIRSRESFK	KGDMFNLKDC	HEFIDFFKES	IKKHEDWSQF	660
GFEFSPTKYE NDISEFVNEY	KIQYRKYK	NVSKKYIDEL	IECGQLYLFQ	IYNKDFSTVY	720
KGNPNLHTMY KPLMFDERNL	ANVVYQLNGG	AEMFYRKASI	KDSEKIVHHA	NQPIKNKNAD	780
NVKKESVFEE DIKDKRFTK	RQFSIHIPIT	LNFKAKGONF	INNDVRMALK	KADENYVIGI	840
DRGERNLNLYI CVINSKGEIV	EQKSLNNEIIG	DNGYRVDYHK	LDDKKEAERD	EARKSWGTIE	900
NIKELKEGYL SQIVHEISKL	VIKYDAVIAI	EDLNSGFKKG	RFKVEQVYQ	KFENMLCTKL	960
NYLVDKNADA NECGGLLKAY	QLTNKEDGAN	RGRQNGLIIFS	VPAWLTSKID	PTVGFADLLR	1020
PIYKSVSESV EFISKIDNIR	YNSKEDYFEEF	DIDYSKFPNS	TASYKKWTV	CTYGRINV	1080
RNKEKNNNMWD NKTIVLTD	DEF KKLFADFGVD	VSKNIKESVL	AIDSKDFFYR	FINLLANTLQ	1140
LRNSEVGNVD VDYLISPVKG	VGDSFYDSRL	VKEKTLPPENA	DANGAYNIAR	KALWAIDVLK	1200
QTKDEELKNA NLSIKNAEWL	EYVQK				1225

SEQ ID NO: 276	moltype = AA	length = 1293
FEATURE	Location/Qualifiers	
source	1..1293	
	mol_type = protein	
	organism = synthetic construct	

SEQUENCE: 276						
MRTMVTFEDF TKQYQVSKTL	RFELIPQGKT	LENMKRDGII	SVDRQRNEDY	QAKGILDKL	60	
YKYILDFTME TVVIDWEALA	TATEEFRKSK	DKKTYEKVQS	KIRTALLEHV	KKQKVGTEDL	120	
FKGMFSSKII TGEVLAAPF	IRLSDEENLI	LEFKDFDTTY	FTGFFENRKN	VFTDEALSTS	180	
FTYRLVNDNF IKFFDNCIVF	KNVNISPBM	AKSLETCAST	LGIFPGVSLE	EVFSISFYR	240	
LLTQTGIDQF NQLLGGISGK	EGEHQHQGLN	EIIINLAMQON	LEVKEVLKNN	AHRFTPLFKQ	300	
IILSDRSTMSF IPDAFADDD	VLSAVDAYRK	YLSEKNIGDR	AQFLISDMEA	YSPPELMRIGG	360	
KYVSVLSQLL FYSWSEIRDG	VKAYKESLIT	GKKTKELEN	IDKEIKYGV	LQEIKEALPK	420	
KDIYEEVKY AMSVVKDYHA	GLAEPPLPEKI	ETDDEERASIK	HIMDSMLGLY	RFLEYFSHDS	480	
IETDTPVFGE CLDTILDDNF	ETVPLYNKVR	NFSTRKVYST	EKFKLNFNN	SLANGWDKKN	540	
EQANGAILLR KEGEYFLGIF	NSKNPKLVS	DGGAGIGYEK	MIYKQFPDFK	KMLPKCTISL	600	
KDTKAHQKQS DEDFTLQTD	FEKSIVITKQ	IYDLGTQTVN	GKKKFQVDPY	RLTGDMEGYR	660	
AALKEWIDFG KEFIQAYTST	AIYDTSLFRD	SSDYPDLPSP	YKDVDNICYK	LTFEWIPDAV	720	
IDDCIDDSGL YLFKLHNKDF	SSGSIGKPNL	HTLYWKALFE	EENLSDVVK	LNGQAELFYR	780	
PKSLTRPVH EEEVIIINKT	TSTGLPVPPD	VYVELSKFVR	NGKKGNLTDK	AKNWLKDVT	840	
RKMPHAITKD RRFTVDKFFF	HVPITLNYKA	DSSPYRFNDF	VRQYIKDCSD	VKIIGIDRGE	900	
RNLIYAVVID GKGNIIQE	FNTVGTNYQ	EKLEQKEKER	QTARQDWATV	TKIKDLKKG	960	
LSAVVHESL MIVKYKAIVA	LENLNVGFKR	MRGGIAERSV	YQQFEKALID	KLNLYLFKDE	1020	
EQSGGYGGVNL	AYQLTDKFES	FSKMGQQTGF	LFYVPAAYTS	KIDPLTGFIN	PFSWKHVKNR	1080
EDRRNFLNLF SKLYYDVNTH	DFVLAYHHSN	KDSKYTIKGN	WEIADWILII	QENKEVFGKT	1140	
GTPYCVGKRI VYMDSTTG	NRMCAYYPH	ELKKLLSEY	IEYTSQDQ	LL KIIQEFD	1200	
LVKGFLFYIK AALQMRNSNS	ETGEDYI	SSP	IEGRGPICFD	SRAEADTLPY	DADANGAFHI	1260

-continued

---

AMKGLLLTER IRNDDKLAIS NEEWLNYIQE MRG	1293
SEQ ID NO: 277	moltype = AA length = 1399
FEATURE	Location/Qualifiers
source	1..1399
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 277	
MNKDIRKNFT DFVGISEIQK TLRFLIPIG KTAQNIDKYN MFEDDEIRHE YYPILKEACD	60
DFYRNHIDQQ FENLEDDWSK LDEALASEDR DLINETRATRY RQVLFNRLLKN SVDIKGDSKK	120
NKTLTSLESSD KNLGKKTKN TFQYNTFNDLF KAKLIKAILP LYIEYYIYEGER KLENAAKALK	180
MYNRTFTSRSL NFWQARANIF TDDEISTGS P YRLVNDNFTI FRINNSIYTTR NKPFIEDIL	240
EFEKKLKSKK IIKDFESVDD YFTVNAPNLK CTQNGIDKYN SILGGFTTKE REKVKGELN	300
FNLAQQSINK GKKGEYRKNI RLGLTKLKK QILAISDSTS FLIEQIEDDQ DLYNKIKDFF	360
ELLKKEEIJN ENITQYANL QKLIQADALS KIYINAKHLN KISHQVTGKW DSLNKGIAALL	420
LENININEES KEKSEVINSQ QTQDISSEAY KRYLQIQCSEE KDIERLRTQI YFSLEDLEKA	480
LDLVLIIDENM DRSDKSILSY VQSPDLNVNF ERDLTDLYSR IMKLEENNEK LLANHSAIL	540
IKEFLDLIML RYSRWNQILFC DSNEYLDQTF YPIYDAVMEI LSNIIRLYNL ARNYLSRKPD	600
RMKKKKINFN NPTLADGWSE SKIPDNSSML FIKDGMYYLG IIKRNRAAYSE LLEAESLQSS	660
EKKKSENNSY ERMNYHFLPD AFRSIPKSSI AMKAVKEHFE INQKTAIDL DTDKPSKPLR	720
ITKEIFDMQY VDLHKNNKKY QVDYLRTGD KKGYRKALNT WLNFKCDFIS KYKGRNLFYD	780
SKIKDADHYE TVNEFYNDV KYSYHIFTTS VAETTVEKPI SEGKLYLFQL YNKDFSPHST	840
GKPNLHTIYW RALFSEENLT SKNPKLNGQA EIFFRPKQIE TPFTHKKGSI LVNRFDVNGN	900
PIPINVQEI KGFKNNVIKW DDLNKTTCQEG LENQDQYLIPE SEFEIICKDRR YTEDQLFHV	960
PISFNWDIGS NPKINDLATQ YIVNSNDIHI IGIDRGENHL IYYSVIDLQG AIVEQGSLNT	1020
ITBYTENKFL NNKTNNLRKI PYKDILQORE DERADARIKW HAIDKIKDLK DGYLQIVHF	1080
LAKLIIKYNIA IVILEDLYNG FKGRGRFKVER QVYQKFEMAL MKKLNVLVFK DYDIDEIGGP	1140
LKPWQLTRPI DSYERMRQN GILFVVPAAV TSAVDPVTFG ANLFYLNVNK NSEKPFHFSK	1200
FESIKYHSDQ DMFSFAFDYN NFGTTTRIND LSKSKWQVF NHERSVWNK EKNYVTQNL	1260
DILIKLRLTY NIEFKNNQNQV LDSILKIENN TDKENFAREL FRLFRLTIQL RNTTVNENNT	1320
EITENELDYI ISPVKDKGNM FFDSRDELKN LPNGDANGA YNIARKGLLY IEQLQESIKT	1380
GKLPPLSIST LDWFNYIMK	1399
SEQ ID NO: 278	moltype = AA length = 1272
FEATURE	Location/Qualifiers
source	1..1272
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 278	
MTPIFCNFVV YQIMLFNNNI NINVKTMNKK HLDSDFTNLLPV VS KTLRFRLE PQGKTMENIV	60
KAQTIETDEE RSHDYETKTE YIDDYHRQPI DDTLDKFAFK VESTGNNDSL QDYLDAYLSA	120
NDNRTKQTEE IQTNLRKAIV SAFKMQPQFN LLEFKEMVKH LLPQFVDTDD KKRVAKFND	180
FITYTGTGFT NRENMYSDA KSTSTAYRV NQNLIKFVEN MLTFKSHILP ILPQBLATL	240
YDDFKEYLNV ASIAEMFLN HFSIVLTLQRQ IEVYNSVIGG RKDENNKQIK PGQLNQYINQH	300
NQAVKDKSAR LPLLKPLFNFQ ILSEKAGVSF LPKQFKSASE VVKSLENEAYA ELSPVLAQI	360
DVVTNITDYD CNGIFIKNDL GLTDIAQRFY GNYDAVKRGQ RNQYELETPM HNGQKAEKYE	420
EQVAKHLKSI ESVSLAQINQ VVTDGGDIDC YFKAFGATDD GDIQRENLIA SINNAAHTAIS	480
PVLNKENAND NELRKTMLI KDLDAIKRL QWFPAKPLLG AGETNQDQVF YGKFEPLYNQ	540
LDTETISPLYD KVRSYLTKKP YSLDKFKINF EKSNLGGWD PGADRKYQYQ AVILRKDNDF	600
YLGIMRDEAT SKRKCIVQLD CNDEGLDENF EKVEYKQIKP SQNMPPRCFAA KKECEENADI	660
MELKRKKNAK SYNTNKKDKN ALIRHYQYRL DRTYPEFGFV YKDADEYDTV KAFTDSMSQ	720
DYKLSFLQVS ETGLLNKLVDE GDLYLKFITN KDFSSYAKGR PNLTHTYWRM LFDPKNLANV	780
VYKLEGKAFF FRRRKSLAST TTHAKQAIK NKSRYNEAVK PQSTFDYDII KDRRPTADKF	840
EFHVIKMFN KAAGWNSTR LNEVREFI KS QGVRHHIIGD RGERHLLYLT MIDMDGNIVK	900
QCSLNAQPAQD NARASEVDYI QLQLDSKEADR LAARRNNNTI ENIKELKQGY LSQVWHLAT	960
MMVDNDAILY LENLNAGFM GRQKVNCSVY QKFEKMLIDK NYIVDKGQS PDKPTGALHA	1020
VQLTGLYSDF NKSNNMKRANV RQCGFVFYIP AWNTSKIDPV TGFBVNLFDTH LSSMGEIKAF	1080
FSKFDSIRYQ QDKGWFEFKF DYSRFTTRAEG CRCTQWTVCT YGERIWTHRS KNQNNQFVND	1140
TWNVTQOMLQ LLQDCGIDPN GNLLKEAIANI DSKKSLETLL HLFKLTQVOMR NSVTGSEVDY	1200
MISPVAADERG HFFDSRESDE HLPANADANG AFNIARKGLM VVRQIMATDD VSKIKFAVSN	1260
KDWLRLFAQHI DD	1272
SEQ ID NO: 279	moltype = AA length = 1313
FEATURE	Location/Qualifiers
source	1..1313
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 279	
MNKGGYVIME KMTEKNRWEQ QFRITKTIKE EIPTGTYKV NLQRVNMLKR EMERNEDLKK	60
MKEICDEYYR NMIDVSLRLE QVRTLGWESL IHKYRMLNKD EKEIKALEKE QEDLRKISK	120
GFGEKKAWTG EQFIKKILPQ YLMDHYTGE LEEKLRLIVKK FKGCMLFLST FFKRNENIFT	180
DKPIHTAVGH RITSENAMLF AANINTYEM ESNVTLEIER LQREFWRRGI NISEIFTDAY	240
YVNVLTQKQI EAYNKICGDI NQHMNEYCQK QKLKFSEFRM RELKKQIILAV VGEHFEIPEK	300
IESTKEVYRE LNEYYESLKE LHGQFEELKS VQLKYSQIYV QKKGYDRISR YIGGQWDLIQ	360
ECMKKDCASG MKGTKKNHDA KIEEEVAKVK YQSIEHIQKL VCTYEEEDRGH KVTDYDEFI	420
VSVCDLLGAD HIITRDGERI ELPLQYEPGT DLLKNDTINQ RRLSDIKTIL DWHMDMLEWL	480

---

-continued

---

KTFLVNLDLVI	KDEEFYMAIE	ELNERMQCVI	SVYNRIRNYV	TQKGYEPEKI	RICFDKGTLI	540
TGWTTGDNYQ	YSGFLLMRND	KYYLGIINTN	EKSVRKILLD	NEECKDENDY	IRVGYHLIND	600
ASKQLPRTFV	MPKAGKKSEI	LMKDEQCDYI	WDGYCHNKHN	ESKEFMRELI	DYYKRSIMNY	660
DKWEGYCFKF	SSTESYDNMQ	DFYKEVREQS	YNISFSYINE	NVLEQLDKDG	KIYLQVYINK	720
DFAAGSTGTG	NLHTMYLQNL	FSSQNLELKR	LRLGGNAELF	YRPGTEKDVT	HRKGSLVDR	780
TVYREEKDGI	EVRODTVPEKE	YLETYRYLNG	KQKGDLSESA	KQWLKDVKHYR	EAPCDIICKD	840
RYAQEKYFLH	FSVVEINPNAE	GQTAALNDNR	RWLSEEEDIH	VIGIDGERN	LIYVSLMDGK	900
GRIKDQKSYN	IVNSGNKEPV	DYLAULKVRE	KERDEARRNW	KAIGKIKDIK	TGYSLSYVH	960
IVEMAVREKA	IIVMEDLNHG	FKRGRFKVER	QVYQKFEEML	INKLNYVVDK	QLSVDEPGGL	1020
LRGYQAFAP	KDKKSSMRQN	GIVFVVPAGY	TSKIDPTTGF	VNFKFPQFG	KGDDDDNGKGD	1080
YDKIRAFPK	PDEIRYCECD	KVTAADNTREV	KERYRFDPDFY	SKFETHLVHM	KTKWTVYAE	1140
GERIKRKVVG	NYWTSEVISD	IALRMSNTLN	IAGIEYKDGH	NLVNEICALR	GKQAGIIILNE	1200
LLEIVRLTVQ	LRNSTTEGDV	DERDEIISPV	LNEKYGCFYH	STEYKQQNGD	VLPKDADANG	1260
AYCIGLKGYI	EIRQIKNWK	EDMTKGECKA	LNEGMRISHD	QWFEFIQNMN	KGE	1313

SEQ ID NO: 280	moltype = AA	length = 1134
FEATURE	Location/Qualifiers	
source	1..1134	
	mol_type = protein	
	organism = synthetic construct	

SEQUENCE: 280						
MNELVKNRCK	QTKTICQKL	PIGKTRTIE	KYNLMEIDRK	IAANKELMNK	LFSLIAGKHI	60
NDTLSKCTDL	DPEPLLTSLS	SLNNAKENDR	DNLREYYDSV	FEKKTLAEE	ISSRLTAVKF	120
AGKDFFTKNI	PDFLETYEGD	DKNEMSELVS	LVIENTVTAG	VVKKLEKIDR	SMEYRLVSGT	180
VVKRVLTDNA	DIYEKNIEKA	KDFDYGVLNI	DEASQFTTLV	AKDYANYLTA	DGIAIYNVG	240
GKINLALNEY	COKNKEYSYN	KLALLPLQKM	LYGEKSLSLFE	KLEDFTSDEE	LINSYNKFAK	300
TVNESGLAEE	IKKAVPSYDE	IIVKPNKISN	YSNSITGHWS	LVNRMKDYL	ENNGIKNADK	360
YMEKGLTLSE	IGDALENKNI	KHSDFISNL	NDLGHTYETI	KENKESLKKD	ESVNALIHK	420
ELDMLLSILQ	NLKVPFDIDNE	MFDTGFGIEV	SKAIEILGYG	VPLYNKIRNY	ITKKPDPKKK	480
FMTKFGSATI	GTGITTTSVEG	SKKATFLKD	DAVFLLLNT	AGCKANNVSV	SNLADLINSS	540
LEIENSGKCY	QKMIYQTPGD	IKKQIPRVCY	YKSEDDDLIK	DFKAGLHKTD	LSFLNGLRIP	600
YIKEAFATHE	TYKNTYFSYR	NSYESYDEF	EHMSEQAYIL	EWKWDKCLI	DDLVEDGSLL	660
MFRVWNRFMK	KKEGKISKHA	KIVNLFSD	NASNAAIKLL	SVFDIFYRD	QIDNPIVHKA	720
GTTLYNKRKT	DGEVIVDYT	MVKNEKEPRN	VYTTTKYDI	IKDRRYTEEQ	FEIHLHVNI	780
KEENKEKLET	SKVINEKNT	LVVTSRNEHL	LYVVFEDEND	NILLKKSLSNT	VKGMMFKSKL	840
EVVEIQKKEN	MQSWKTVGSN	QALMEGYLSF	AIKEIADLVK	EYDAILVLEQ	NSVGNILNE	900
RVYTRFKEML	ITNLSLDVDY	ENKDFSYTE	LGGKVASWRD	CVTNGICIQV	PSAYKVDPT	960
TSFSTISMYA	KTTAEKSKKL	KQIKSFKNR	ERGLFELVIA	KGVGLENNIV	CDSFGSRSI	1020
ENDISKEVSC	TLKIEKYLID	AGIEYNDEKE	VLKDLDATAK	TDAVHKAVTL	LLKCFNESPD	1080
GRYYISPCGE	HFTLCDAPEV	LSAINYYIRS	RYIREQIVEG	VKKMEYKKTI	LLAK	1134

SEQ ID NO: 281	moltype = AA	length = 1259
FEATURE	Location/Qualifiers	
source	1..1259	
	mol_type = protein	
	organism = synthetic construct	

SEQUENCE: 281						
MNYKTGLEDF	IGKESLSKTL	RNALIPTEST	KIHMEEMGV	RDDDELRAEKQ	QELKEIMDDY	60
YRTFIEEKLG	QIOQGIQWNSL	FQKMEETMED	I SVRKDLDKI	QNEKRKEIICC	YFTSDKRFKD	120
LFNAKLITDI	LPNFIKDNKE	YTEEEKAKE	QTRVLQFRQA	TAFTNYFNQR	RNNFSEDNIS	180
TAISFRVNE	NSEIHLQNMB	AQFQRIEQQY	EEVCGMEEY	KDMLQEWQM	HIYSVDFYDR	240
ELTQPGIEYY	NGICGKINR	MNQFCQKNI	IKNDFRMKML	HKQILCKKSS	YYEIPFRFES	300
DQEVDYDALNE	FIKTMKKEI	IRRCVHLQGE	CDDYDLGKIY	ISSNKYEQIS	NALYGSWTD	360
RKCIKEEYMD	ALPGKGEKKE	EKAEEAAKKE	EYRSIADIDK	IISLYGSEMD	RTISAKKCIT	420
EICDMAQCS	IDPLVCSNDSI	KLLDVKKEKTT	EIKTILDSTF	HVYQWQOTFI	VSDIIEKDSDY	480
FYSELEVDLE	DFEGITTLYN	HVRSYVQTKQP	YSTVKFKLHF	GSPTLANGWS	QSKEYDNNAI	540
LIMRDQKFYL	GIFVNVRNKP	KQIIKGHEK	EKGDYKKMI	NLLPGPSKML	PKVFFTSRSG	600
QETYKPSKHI	LDGYNEKRHI	KSSPKFDLGY	CWDLIDYYKE	CIHKHPDWKN	YDFHFSDTKD	660
YEDISGFYRE	VEMQGYQIKW	TYIISADEI	QIYNGQI	LDEKGQIPLF	QIYNKDFSVH	720
YLNKNLFSSEEN	LKDIVLKLNG	EAELFFRKAS	IKTPIVHKKG	SVLVNRNSYTQ	TVGNKEIRVS	780
IPPEYTYIY	NYLNHHIGKGE	LSSEAQRYLAS	EGKIKSFTAT	KDIVKNYRC	CDHYFLHLPI	840
TINFKAKSVD	AVNERTLAYI	AKKEDIHII	IDRGERNL	ISVVDVHGN	REQRSFNIVN	900
GYDYQQKLKD	REKSDRAARK	NWEEEIKKE	LKEGYLSMVI	HYIAQLVV	NAVVA	960
YGFKTRFKV	ERQVYQKFET	MLEI	KLHYL	FKDREVCEEG	GVLRGYQLT	1020
QCGFIFYVPA	GYTSKIDPTT	GTVNLFSFKN	LTNRESRQDF	RDGFDEIRYD	RDKKMFEFSF	1080
DYNNYIKKGT	I LASTKWKVY	TNGTRLKKIV	TNGKYTSQSM	EVELTDAMEK	MLQRAGIEYH	1140
DGKDLKGQIV	EKGIEAEIID	I FRLTVQMRN	SRSESEDREY	DRLISPVLND	KGEFFDTATA	1200
DKTLPQDADA	NGAYCIALKG	LYEVKQI	WKENEQFPRN	KLVQDNKTWF	DFMQKKR	1259

SEQ ID NO: 282	moltype = AA	length = 1271
FEATURE	Location/Qualifiers	
source	1..1271	
	mol_type = protein	
	organism = synthetic construct	

SEQUENCE: 282						
MEDKQFLERY	KEFIGLNSLS	KTLRNSLIPV	GSTLKHIQEY	GILEEDSLRA	QKREELKGIM	60

- continued

DDYYRNYIEM	HLRDVHDIDW	NELFEALTEV	KKNQTDDAKK	RLEKIQEKKR	KEIYQYLSDD	120
AVFSEMPFEK	MISGILPDFI	RCNEGYSSEE	KEEKLKTVAL	FHRFTSSFND	FFLNRKNVFT	180
KEAIVTAIGY	RVVENAEIF	LENMVAFQNI	QKSAESQISI	IERKNEHYFM	EWKLSHIFTA	240
DYYMMMLMTQK	AIEHYNEMCG	VVNQMREYC	QBEKKWNWLY	RMKRLHKQIL	SNASTSFKIP	300
EKYENDAEVY	ESVNSFLQNV	MEKTVMERIA	VLKNSTDNFN	LSKIYITAPY	YEKISNYLCG	360
SWNTTIDCLT	HYEQOIIAGE	GARKDQKVKA	AVKADKWKSL	SEIEQLLKEY	ARAEEVKRKP	420
EEYIAEIEINI	VSLKEAHLL	YHPEVNLIEN	EKYATEIKDV	LDNYMELFWH	MKWFYIEAV	480
EKEVNFGYEL	DDLYEEIKDI	VPLYNKVRNY	VTQKPYSDTK	IKLNFGTPTL	ANGWSKSKEY	540
DYNAILLQKD	GKYYMGIFNP	IQKPEKEII	EHSQPLEGNE	YKKMVYYLP	SANKMLPKVL	600
LSKGGMIEIQ	PSEYIINGYQ	ERRKNESEK	FDLQFCHDLI	DYFKSGIERP	SDWKVFGFDF	660
SDTDTYQDIS	GYREVEDQG	YKIDWTYIKE	ADIDRNLNEEG	KLYLFQIYNK	DFSEKSTGRE	720
NLHTMYLKNL	FSEENVREQV	LKLNGEAEIF	FRKSSVKPPI	IHKKGTMVLU	RTYMBEVNGN	780
SVRRNIPEKE	YQEIYNYKNH	RLKGELSTEA	KKYLEKAVCH	ETKKDIVKDY	RYSVDKFIIH	840
LPITINYRAS	GKETLNSVAQ	RYIAHQNDMH	VIGIDRGERIN	LIIVSVINMQ	GEIKEBQKSFN	900
IINEFNYKKE	LKEREKSRGA	ARRNWKEIGQ	IKDLKKEGYLS	GVIHEIAKMM	IKYHAIIAM	960
DLYNYGFGRGR	PKVEROYQ	FENMLIQLKLN	YLVFKDRPAD	EDGGVLRGYQ	LAYIPDSVKK	1020
MGRQCGMIFY	VPAFTSKID	PTTGFVDIFK	HKVYTTEQAK	REFILSFDEI	CYDVERQLFR	1080
FTFDYANFVT	QNVTILARNNW	TIYTNGTRAQ	KEPGNGRMRD	KEDYNPKDKM	VELLESEGIE	1140
FKSGKNNLPPA	LKKVSNAKVF	EELQKIVRFT	VQLRNSKSEE	NDVDYDHVIS	PVLNEEGNFF	1200
DSSKYKRNKEE	KKESLLPVDA	DANGAYCIAL	KGLYIMQIAQ	KNWSEEKALS	PDVLRNNND	1260
WFDYIQNKR	R					1271

SEQ ID NO: 283                  moltype = AA length = 1367  
 FEATURE                  Location/Qualifiers  
 source                  1..1367  
 mol\_type = protein  
 organism = synthetic construct

SEQUENCE: 283  
 MEKSLNDFIG LYSVSKTLRF ELKPVSETLE NIKKFHFLEE DKKKANDYKD VKKIIDNYHK 60  
 YFIDDVLKNA SFNWKKLEEA IREYNKNKSD DSALVAEOKK LGDAILKLT SDKRYKALTA 120  
 ATPKLFESI LPPDFGECQN QDLNKAALKT FQKFTSYFTPQ FQENRKNVYS AEA IPTAVPY 180  
 RIVNDNPFPK LQNVLIFKTI QEKCQPNPDE VEKELSSYLG KEKLAGIFTL ESFNKYLGQG 240  
 GKENQRGIDF YNQIIGGVVE KEGGINLRLGV NOFLNLYWQO HPDFTKEDRR IKMVPPLYQKI 300  
 LSDRSSLFSK IESIENDEL KNALLECADK LELKNDEKK IFEEVCDLFS SVKNLDSLGI 360  
 YINRKDINSV SRILTGDSWSL LQSRSNVMYAE EKFTTAKAEGA RWQKSLDDEG ENKSKGFYSL 420  
 TDLNEVLEY SENAVENTDIR ITDYEFEHCR YYVDKETEMF VQGSELVALS LQEMCDDILK 480  
 KRKAMNTVLE NLSSENKLRE KTDDAVAIKE YLDAVQELLH RIKPLKVNGV GDSTFYSVYD 540  
 SIYSALSEVI SVYNKTRNYI TTKAASPEKY KLNFDNPTLA DGWDLNKEQA NTSVILRKDG 600  
 MFYLGIMNPK NPKPFAEKY CGNESCYEKM IYKQFDATQ IPKCSCTQKKE VQKYFLSGAT 660  
 EPYILNDKKS FKSELIITKD IWFPMNNHWVD GEKFVPKRDN ETRPKKFQIG YFKQTGDFDG 720  
 YKNALSNWIS FCKNFLQSYL SATVYDYNPK NSEEYEGLDE FYNYLNATCY KLNFINIPET 780  
 EINKMVSEGK LYLFQIYKND FASGSTGMPN MHTLYWKNLF SDENLKNVCL KLNGBAEELFY 840  
 RPAGIKBPEV HKEGSYLVN TTEDGESPE KIYFEIYKNA NGKLEKLSDE AQNYISNHEV 900  
 VIKKAGHEII KDRHYTEPKF LFHVPLTINF KASGNNSYIN ENVRKFLKNN PDVNIIGLDR 960  
 GERHLLIVLSL INQKGIIKQ FTFNEVERNK NGRTIKVNMH EKLDQREKER DAARKSWQAI 1020  
 GKIAELKEGY LSAVIHQLTK LMVEYNAAVV MEDLNGFGPKR GRFHVEKQVY QKFEHILIDK 1080  
 SNYLVFKDRG LNEPGVGLNG YQIAQGFESF QKLGKQSGML FYVPGAYTSK IDPKTGFVSM 1140  
 MNPKDLTNVH KNRDFFSKFD NIHDEANGS FVPTPDFYKFF DGKAKEEMKL TKWSVYSSRD 1200  
 RIVYFAKTKS YEDVLPTEKL QKIFESNGID YKSGNNIQDS VMAIGADLKE GAKPSKEISD 1260  
 FWGDGLSNFK LILQMRNSNA RTGEDYIISP VMADDGTFDD SREEFKKGED AKLPLDADAN 1320  
 GAYHIALKGL SLINKINLSK DEELKKFDMK ISNADWFKFA QEKNYAK 1367

SEQ ID NO: 284                  moltype = AA length = 1229  
 FEATURE                  Location/Qualifiers  
 source                  1..1229  
 mol\_type = protein  
 organism = synthetic construct

SEQUENCE: 284  
 MEEKKMSKIE KFIGKYKISE TLRFRAPVG KTODNIEKKG ILEKDKKRSE DYEKVKAYLD 60  
 SLHRDVKLNLN EYACLFFSGT KDDGDKKKM KLEEKMRKTI SNEFCNDEMY 120  
 KKFSEKILS ENNEEDVSDI VSSYKGFPTS LNGYVNNRKN LVSADAKPTS IAYRCINENL 180  
 PKFLRNVECY KKVVQVIPKE QIEYMSNNLN LSPYRIEDCF NIDFFEFCLS QGGIDLYNTF 240  
 IGGYSKKDGT KVQGINIEVN LYNQKNKKDK EKYKLPQFTP LFKQILSDRD TKSFSIEKLE 300  
 NIYEVVVELVK KSYSLDMEMDD IETVFSNLNY YDASGIYVN GPAITHISMN LTKDWTIRN 360  
 NWNYEYEDKHH STKKNKNIEK YEDTRNTMYK KIDSFTLEYI SRLVGKDIDE LVKYFENEVA 420  
 NFVMDIJKTY SKLTPLFDRQ QKENFDISED EVNDIKGYLD NVKLLESFMK SFTINGKENN 480  
 IDYVYFGKFT DDYDKLHEFD HIYNKVRNYI TTSRKPYKLD KYKLYFDNPQ LLGGWDINKE 540  
 KDYRTVMLTK DGKYYFAIID KGEPFDNIP KDYFDNNGGY KKIIYRQIPN AAKYLSSKQI 600  
 VPQNPPEEVK RILDKKKADS KSLTEEEEKNI FIDYIKSDFL KNYKLLFDKQ NNPFYFNFAFR 660  
 ESSTYYESLNE FFEDVERQAY SVRYENLPAD YIDNLVNEGK IYLFEIYSKD FSEYSKGTTNN 720  
 LHTMYFKALF DNDNLKNTVF KLSGNAELFI RPASIKKDEL VIHPKNQLLQ NKNPLNPKKQ 780  
 SIFDYDLVKD KRFFENQYML HISIEINKNE RDAKKKINN EMVRKELKDS DDNYIIGIDR 840  
 GERNLLYVCV INSAGKIVEQ MSLNNEIINEY NGIKHTVDQY GLLDKCEKER NAQRQSWKSI 900  
 ENIKELKDGY ISQVVKLQC LVEKYDAITA MENLNGGFKR GRTKFEKQVY QKFNKLINK 960  
 MEYMADKKRK TTENGGLLRG YQLTNGCINN SYQNGFIFVY PAWLTSKIDP TTGFV DLLKP 1020  
 KYTNVEE AHL WINKFNSITY DKKLMDFAPN INYSQFPRAD IDYRKIWTY TNGYRIETFR 1080

-continued

---

NSEKNNEFDW KEVHLTSVIK	KLLEEQYQINY ISGKNIIDDL	IQIKDKPFWN SFIKYIRLTL	1140
QMRSNITGRT DVVDIISPVI	NNEGTFYDSR KDLDEITLPQ	DADANGAYNI ARKALWIEK	1200
LKESPDEELN KVKLAITQRE	WLEYAQINI		1229

SEQ ID NO: 285	moltype = AA length = 1261
FEATURE	Location/Qualifiers
source	1..1261
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 285	
MIIHNCYIGG SFMKKIDSFT NCYSLSKTLR FKLIPIGATO SNFDLNKMLD EDKKRAENYS	60
KAKSIIDKYH RFFIDKVLSS VTEENKAQDFSF LEDVRAYAEL YYRSNKKDSD KASMTKLESK	120
MRKFIALALQ SDEGFKDLFG QNLKIKTLP ELESDTDKEI IAEFDGFSTY FTGFFNNRKN	180
MYSAADDQPTA ISYRC1NDNL PKFLDNVRFT KNSDVASILIN DNLKILNEDY DGIYGTSAED	240
VFNVDYPPFV LSQKGIEAYN SILGGYTNSD GSKIKGLNEY INLYNQKNE N IHRIPKMQL	300
FQKILSERES VSFPIPEKFDS DDDVLSSIND YYLERDGKGV LSIEKTVEK EKLFSAVTDY	360
STDGIFVKNA AELTAVCAGA FGWGTQVQNA WNNEYDALNG YKETEKYIDK RKKAYKSIES	420
FSLADIOQYKA DVSESETNA EVTEWLRNEI KEKCNLAVQG YESSKDLISK PYTESKKLFN	480
NDNAVELLNK ALDSVKELEN VLRLLLGTGK EESKDENEYG EFLPCYERIC EVDSLSDKVR	540
NYMTQKLKYK DKIKLNFQNP QFLGGWDNRK EADYSAVLLR RNLSLYYIAIM PSGYKRVFEK	600
IPAPKADET YEKVIYKLLP GPNKMLPKVF FSKKGIETFN PPKEILEKYE LGTHKTGDGF	660
NLDDCHALID YFKSALDVH DWSNFGFRFS DTSTYKNIAD FYNEVKNQG KITFCDPVQS	720
YINELVDEGK FSEHQLYNKD FSEHQLYNKD LHTLYFKMLF DERNLENVF KLNGEAEMFY	780
REASISKDDM IVHPKNQPIK NKNQEONSRSQK STPEYDIVKD RRYTVQFML HIPITLNFTA	840
NGGTNNINNEV RKALKDCDKN YVIGIDRGER NLLYICVVDs EGRIIEQYSL NEIINEYNGN	900
TYSTDYHALL DKKEKERLES RKAWTVENTI KELKEGYISQ VVHKICELVE KYDAVIVMED	960
LNLGFKQGRS YKFEKSVYQK FEKMLADKLN YFADKKKSPE EIGSVLNAYQ LTNAFESFEK	1020
MGKQNGFIFY VPAYLTSKID PTTGFPADLH PSSKQSKESM RDFVGRFDSDI TFKNKTENYFE	1080
FELDYNKFPR CNTDYRKWT VCTYGSRIKT FRNPEKNSEW DNKTVELTPA FMALFEKYSI	1140
DVNGDIQAQI MSVDKDDFFV ELIGLRLRTL QMRNSETGKV DRDYLISPVK NSEGVFYNSD	1200
DYKGIGENASL PKDADANGAY NIARKGLWII EQIKACENDA ELNKIRLAIS NAEWLEYAQK	1260
K	1261

SEQ ID NO: 286	moltype = AA length = 1242
FEATURE	Location/Qualifiers
source	1..1242
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 286	
MKBQFVNQYP ISKTLRFSLI PIGKTEENFN KNLLLKEDEK KAAEYQVKVG YIDRYHKFFI	60
ETALCNINF GFEELYSLYY KCSKDDNDLK TMEDIEIKLR KQISKTMTH KLYKDLFGEN	120
MIKTIPLNFL DSDEEKNSL MFRGFYTYFS GFNTNRKMY TEEAKSTSIA YRCINDNLPK	180
FLDNSKSFEK IKCALNKEEL KAKNEEYKBI FOIYATDIFN IDFFFNVLTQ PGIDKYNGII	240
GGYTCSDGK VQGLNLYEINL YNQQIAKDDK SKRPLLKML YKQILSDET VFPIPEKFSS	300
DNEVLESINN YFSKVNNSAI KSLKELFQGF EAYNMNGIFI SSGVAITDLS NAVFGDWNAI	360
STAWEKAYFE TNPPKKNKSQ EKYEEELKAN YKKIKSFSLD EIQRGLGSIAK SPDSIGSVAE	420
YYKITVTEKI DNITELYDGS KELLNLCNSE SYDKKLTKND TVIEKVKTLL DAVSLEKLJ	480
KPLVGTGKED KDELFGYGT PLYTLSAVID RLYDKVNRYA TQKPYSKDKI KLNFCNSFL	540
SGWATDYSNN CGGLIFEKDGL YYLIVVNKKF TTBEIDYLQO NADENPAQRI VYDFOKPDNK	600
NTPRLFIRSK GTNYSPSVKE YNLPVEEIVE LYDKRFTTE YRNKNPELYK ASLVKLIDYF	660
KLGFTTRHESY RHYDFKWWKES EYENDISEFY KDVETSCYSL KQEKENYNTL LNFAVENRIY	720
LFQIYMKDVS YKSKGTPNLK TRYFKALFDE NNLSDDVVFKL NGGSEMFRRK ASIKDNEKVV	780
H PANQPIDNK NPDNSKKQST FDTELKDKR FTKHQFSIHI PITMNNFKARG RDFINNDIRK	840
AIKSEYKPYV IGIDRGERNL IYISVINNNG EIVEQMSLND IIISDNGYKVD YQRLLDRKEK	900
ERDNARKSWG TIENIKEKL GYISQVIHKI CELVICKYDAV IAMEDLNFGE KRGGRPNVEQ	960
VYQKFNENMLI SKNQLCDKK SEANSEGGLL KAYQLTNKFD GVNGKQNGI IFYVPAWLTS	1020
KIDPVTFGVDF LLHPKYISVE ETHSLFKEKL DIRYNFEKDM FEFIDYSKL PKCNADFKQK	1080
WTVCCTNADRI MTFRNSEKNN EWDNKRILLS DEFKRLFEEF GIDYCHNLKN KILSISNKDF	1140
CYRFIKLFL TMQMRNSITG STNPEDDYLI SPVRDENGVF YDSRNFIGSK AGLPIDADAN	1200
GAYNIARKGL WAINAIKSTA DDMLDKVDSL ISNAKWLEYVQ KQ	1242

SEQ ID NO: 287	moltype = AA length = 1307
FEATURE	Location/Qualifiers
source	1..1307
	mol_type = protein
	organism = synthetic construct

SEQUENCE: 287	
MADLSQFTHK YQVPKTLRFE LIPQGKTLLEN LSAYGMVADD KQRSENYKKL KPVIDRIYK	60
FIEESLKNTN LDWNPLYEAI REYRKEKTTA TITNLKEQOD ICRRATASRF EGKVPDKGDK	120
SVKDFNKKQS KLFKELFGKE LFTDSVLEQL PGVSLSDDEK ALLKSFDKFT TYFVGFYDNR	180
KNFSSDDIS TGIPHLVQE NFPKPIDNCD DYKRLVLVAP ELKEKLEKAA EATKIFEDVS	240
LDEIFSIKFY NRLLQQNQID QFNQLLGGIA GAPGTPKIQG LNETLNLSMQ QDKTLEQKLK	300
SVPHRFSPLY KQILSDRSSL SFPIESFSCD AEVLLAVQEY LDNLKTEHVI EDLKEVFNRL	360
TTLDDLKHIVV NSTKVTAFSQ ALFGDWNLCR EQLRVYKMSN GNEKITKKAL GELESWLKNS	420
DIAFTELQEA LADEALPAKV NLKVQEAIISG LNEQMAKSLP KELKIPREEKE ELKALLDAIQ	480
EYVHTLEWFI VSDDVETDTD FVYPLKETLQ IIQPIIPLYN KVRNFATQKP YSVEKFKLFN	540

-continued

ANPTLADGWD	ENKEQQNCAV	LFQKGNNYLY	GILNPKNKPD	FDNVDTKEKQG	NCYQKMVYKQ	600
FPDFSKMMPK	CTTQLKEVKQ	HFEFGKDSYI	LNNKNIKFPL	TITREVYDNL	NVLYDGKKKP	660
QIDYLRLKTD	EDGYYTALHT	WIDFAKKFVA	SYKSTSIYDT	STILPPEKEYE	KLNEFYGALD	720
NLFYQIKFEN	IPEEE11DTYV	KBLGLFLFQYI	YNDKDFAGAT	GAPNLHITYW	KADEFPEVNK	780
DVVVKLNGQA	ELFYRPKSNM	DVRHKVGKEK	LVNRTLKDGS	ILTDELHKEL	YLYANGSLKK	840
GLSEDAKIIL	DKNLAVIYDV	HHEIVKDERRF	TTDKFFHVP	LTLNYKCDKN	PVKPNAEVQE	900
YLNKEPNPTYV	IGIDGRERNL	IYAVVDPDKG	RIVEQKSFSN	INGPDFYHGKL	DQECKERVK	960
RQAWATVKGK	KELKQGYLSL	VVHEISKMMV	RYQAVVWLEN	LNVGFKRVR	GIAEKAVYQO	1020
FEKMLINKLN	YLMFKDAGGT	EPGVSFLNAYQ	LTDTRFESFAK	MGLQTGFLLFV	IPAAFTSKID	1080
PATGFVDPFR	WGAIKTLADK	REFLSGFESL	KFDSTTGNNFI	LHFDVSKNKN	FQKKLEGFVP	1140
DWDIIIEBANK	MKTGKGATYY	AGKRIEFVRL	NNNSQGHYEDY	LPCNALAETL	RQCDIPYEEG	1200
DNDLPLILEM	NDSKLLHSVF	VKVRLTLQMR	NSNAETGEDY	ISSPVDEVSG	SCFDSSRMENE	1260
KLPKDADANG	AHYIALKGML	ALERLRKDE	MAISNNDWLW	YIOEKRA		1307

SEQUENCE :	288	
M	MTNFDNFTKK YVNSKTIIRLE AIPVGKTLKN IEKMGFIAAD RQRDEDYQKA KSVIDHIYKA	60
F	FMDDCLKDLF LDWDPLYEAV VACWRERSPE GRQALQIMQA DYRKKIADRF RNHELYGSLF	120
T	TKKIFDGSVA QRLPDLQEQA EFKSLLSNRQ KFTTSYFRDFD DKRKRLFLSDD EKHSAIAYL	180
I	INENFLKGVA NCEAFRRMTE RVPELREKLQ NTGSLQVYNG LALDEVFSAD FYNQLIVQKQ	240
D	IDLYNQLIGG IAGEPGTPNI QGLNATINLA LQGDSSLHEK LAGIPHRFNP LYKQILSDVS	300
L	TLSFVPSAQS SDGEMLAAVR GFKVQLESGR VLQNVRRLFN GLETEADLSR VYVUNSKLAA	360
S	FSSMMFFGRWN LCSDDAFLWK KGKQKKITNK KLTEIKWLK NSDIAIAEIQ EAFGEDFPRG	420
K	KINEKIQAQAA DALHSQALP DALHNKALCA GLDKLMSMLDT VFLGLYRMLQW FIVGDDNEKD	480
F	SDFYFGLGKI LGSLDPVLV YNRVRNYITK KPYSLTKFRL NFDNSQLLNG WDENNLDTNC	540
G	ASIFIKDGKY YLGISKNKNR PQFDTWTATSG KSGYQRMVYK QFANWGRDLP HSTTQMVKVK	600
H	KHFSASADY VLDGDKFIRP LIITKEIFDL NNVKPENGKK LQDVYLRNTG DREGYTHALH	660
W	TWINFAKDFC ACYKSTSIYD ISCLRPTDQY DLNDMLFYADL GNLSHRIWVQ TIPEEAIDNY	720
B	VEQQLQLFLFQ LYNKDFAPGA TGNPKLNLTH WKAWFNPENL DVUVVKLNGK AELFYPRRSN	780
O	MDVVRHKVG KLVNRKLKG LTLPSSLRHLHEE IYRYVNGLTN KDLSDARSLV LPLAVVRDVQ	840
Q	HEIIKDRRTF ADKFFFHASL TTFNFKSSDKP VGFNEDVREY LREHPDTYVV GVDGRERNLII	900
P	YIVVWDPGQN IVEQRFSFNMJ NGIDYWSLLD QKEKERVEAK QAWEVTGKIK DLKCGYLSFL	960
V	THEITKIIKK YHAVVILENL SGLFKVRVTQ IAEKAVYQQF ERMLVTKLGY VVFKDRAGKA	1020
G	PGGVLNAYQL TDNTRTAENT GIQNGFLFYV PAAFTSRVDP ATGFFDFYDW GKIKTATDKK	1080
G	NFIAGFNSVR YERSTGDFIV HVGAKNLAVR RVAEDVRTEW DIVIEANVRK MGIDGNSYIS	1140
R	GKRIRYRSGE QGHGQYENHL PCQCELIRALQ QYGIQYETGK DILPAILOQD DAKLTDVTVD	1200
R	VFRRLAQMRN TSAETGDEYD NSFVNDRDSGR CFDTRRAEEAA MPKEADANDA YHIALKGFLV	1260
S	LEKLKRKGESI GIKNTEWLRY VQQRHS	1286

SEQ ID NO: 289 moltype = AA length = 1264  
FEATURE Location/Qualifiers  
source 1..1264  
mol\_type = protein  
organism = synthetic construct

```
SEQ ID NO: 290      moltype = AA  length = 1232
FEATURE          Location/Qualifiers
source          1..1232
                 mol type = protein
```

-continued

SEQUENCE:	organism	= synthetic construct
MLHAFTNQYQ	LSKTLRLFGAT	LKEDEKKCKS HEELKGFVDI SYENMKSSAT IAESLNENEL
VVKCERCYSE	IVKFHNDAWEK	IYYRTDQIAV YKDIFYRQLSR KARFDAGKQN SQLITLASLC
GMYQGAKLRS	YITNYWKDNI	TRQKSFLKDF SQQLHQYTRA LEKSDKAHTK PNLINFNKTF
MVLANLVNEI	VIPLSNGAIS	FPNISKLEDG EESHLLIEFAL NDYSQSLSEL GELKDAIATN
GGYTPFKAVT	LNHYTAEQKP	HFVFKNDIDAK IRELKIGLIV ETLKGKSEQ IEEYFSNLDK
NRDNQRNS	VIVRTOCFKY	KPIPLFLVKHQ LAKYISEPENG DEDAWAVAKVL DAVAIGRSPA
HDYANNQEGF	DLNHYPIKVA	FDYAWEQLAN SLYTTVTFPQ EMCEKYLNSI YGCEVSKEPV
FKFYADLLYI	RKNLAVLEHK	NNLPSNQEEF ICKINNTFEN IVLPYKISQF ETYKKDILAW
INDGHDHKKY	TDAKQQLGF	RGGLKGRIKA EEVSKDKYK KIKSYSENPY TKLTNEFKQJ
SSTYGKTFAE	LRDKFKEKNE	ITKTHIFGII IEDKNDRDRYL LASELKHEQI NHVSTILNLK
DKSSEFITYQ	VKSLSKTLI	KLIKNHTKK GAISPYADFH TSKTGFNKNE IEKWDNYKR
EQLVLEVVKID	CLTDSTMKA	NQWAEFGWNF EKCNSYEDIE HEIDQKSYLL QSDTISKQSI
ASLVEGGGL	LPIINQDTS	KERKDKNQFS KDWNHIFEQGS FKLRLHPEFA VSYRTPIEGY
PVQKRYGRLO	FVCAFHAIHV	PQNGEFINL KQIENFNDED VQKRNVTEFVN KKVNHALSDK
EYVVGIDRQ	LKQLATLVC	DKRGKILGDF EIYKKEFVRA EKRSESHWEH TQAETRHILD
LSNLRVETTI	EGKKVVLVDQS	LTLVKKNRDT PDEEATEENK QKIKLQLSY IRKLQHKMOT
NEQDVLDLN	NEPSDEEFFK	RIEGLSSPG EGQKYADLPI NTMREMISDL QGVIARGNNQ
TEKNNKIELD	AADDNLKQGV	AMIGIVINYI FAKYSYKAYI SLEDSRAYG GAKSGYGDRY
LPSTSQDEDV	DFKEQQNQML	AGLGTYQFFF MQLLKKLQKI QSDNTVLRV PAFKSADNYR
NILRLEETKY	KSKPFGVVHF	IDPKFTSKKC PVCSKTNVYR DKDILVCKE CGFRSDSQLK
ERENNINHIYH	NGDDNGAYHI	ALKSVENLIQ MK

```
SEQ ID NO: 291          moltype = AA  length = 1067
FEATURE                 Location/Qualifiers
source                  1..1067
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 291
```

SEQ ID NO: 292 moltype = AA length = 1309  
FEATURE Location/Qualifiers  
source 1..1309  
mol\_type = protein  
organism = synthetic construct

SEQUENCE :	292	01341620	01400010	00111100		
MAKETKEFKT	FDDFTNLYEV	QKTLRPELEA	VPETEIVLEN	RGIWYKRDKK	RADEKPIVKF	60
YMDILHREFT	DEALEKIKES	GVLNLNSGYFK	LFEELRRLLQN	HGANTKEEKK	LKLBEEIRAKK	120
REISNELSQI	RRVFSVRGFD	WTSSDWWKKY	TIEGKKIKND	KSKTYLILSE	NILNFLENRF	180
TSKEVERLRS	IDKKHVEDYD	NNVNSCGENI	FATFKGGFFG	FDLSLKRNERS	FYETTDKGAGR	240
VATRSVDENL	NFFAENLHIF	STDLPKALKD	DLSDTQKAIF	ERSYYKNCLL	QDKISYNLI	300
IGGDINKEINK	HRQQRDRTKIK	FLNTLFLQKIL	SIEEKQYKH	IEINNDEDLI	RAIRDIFISLN	360
ESKISEGTINK	FNQFIQRCRQL	KEDLGQIYLP	KDSVNTIARH	IFPKWPDEIMA	LFDRKYFVSL	420
EEIKDLTESS	VWKERVLEES	KTKSLIFKD	HIHTIISGQE	IFSNFLKEYN	KEYKNQFSGF	480
ISETRRGKAA	FVGYDESLSNK	LRATIKWFEG	KNLKLSETEK	WEVIKAIKY	ADAALRIFOM	540
TKYLWLPVVG	DEEDKDYLRI	KAEIDQLTKD	NDFYNKINAF	IDGYKPEPF	YRSSFQEYLT	600
RRPFSTDKFK	INFENSRLLD	GWDKDMIDDR	MGILLQRDGD	YFLGLLNKED	RHCLDNLVDV	660
KSEDKNSYAL	MQFKQLTGLY	RQLPRAFPK	KKQPVLLEAN	EIKKIKEDFD	FLQKQKKERE	720
VNVNVVFDPNK	KLNNLHNHYA	EFLKPNMAYK	CKYDFPSLLNK	EKVYESLSDF	YADVKITYS	780
LSFQIVS1DQ	LIKTKGKILLF	RLRKNKDLKLG	SLGQNKNLHT	YYHFALERE	NLSQGRIRLG	840
AQAEIIFRPA	SIEKEKDKNR	SNALKSPKT	RYVKEILKNA	RYSEDKVFLH	LPIQLNADAY	900
DLPNSINQNVF	EFIGNRQEKV	KIIGIDRGEK	NLAYYSVISQ	NSNGKIKIEE	PPRDLNLGYL	960
EPLDELENKR	QDERKAWQSI	SEIKSKRDGY	ISYAVSKIVE	LMLKYQAIIV	LEDLSGKFKR	1020
SRMKFEKAPY	QOLELALIIK	LNYLVKKNSK	SGKPGHYLSA	YQLTEPVGSY	KEMGKQTGII	1080
FYTQAGYTSR	TCPTCWWKRK	VQGLYYKDR	SAQRFPDKT	GVKIFYDSVN	DRFVFOYHPV	1140
YEQKELKEWD	KEIYSDVTI	RWNNEEKKNN	EYRKGDITLK	K1KRLFRDRGI	DLSRNINNEQL	1200
VNVGDASFWE	ELINLLRLIT	EIRNIIDNNEEN	RDFIECPHCH	FQSENGFHGV	AWNGDANGAY	1260

-continued

---

NIARKGLLIT KAVCDPEKVN GDIRWSDLKV DMKDWDAAATD EWAKKNPEK	1309
--	------

---

SEQ ID NO: 293	moltype = AA length = 1456
FEATURE	Location/Qualifiers
source	1..1456
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 293	
MENEKIFSDL TNRYQVVKTL PFELKPVPRV LLGLDNPV KGEIFSKDRE RAENFTIICK YIDRLHSLFI NESLKCADID FSNFYKQYK NINTKNNKNI DDDNDINDE KEDSENDNLK KYRQEIANLF NKSJKYKSWVN VLGDKLISL LRTHFSDNLM EDIEIPELFS NNKKIKDTKRL KEIINSFGKD GKDGQNFPTTY FSFSFHNNRK NYYKSDGKMG RVSTRIVDEN LERFCCKNIYL YKEIIGKNEI KEIFSGNWDI YLQKPKPNFSN DKTYKLLDEF KNDKYDWEMI FRDVNSYNKY FLQSDIEFYV YIRGKLNQDI NEYNGKKRS DEKINSQFEN LRNOVHGEKK NYDDDFEINE DNNIQFINEI FVRHNQNKMR FSEKLFSDFI DLLMVDNGDK LDKVYFSQKA VENAIARYVE VEETTNEGREGPLLSLLQAGKDRKKLSSN KPIKLGDIKE VLDOQANNKP EDIFKNRYVL SESNNNDGIIN ANDKNHWANL LRLIKKDFYV HKDNLIKSQD KLAETKYN GSDEGEROIE TIKNFAESAL AILRFLKFDL LRNGVIONV IGGKDPHHEE VDKYFDGDVL SGEESCRNFIK YKDALRNFIK KAWASADKII LNFDCCSEFLG GWDRSQEQQK RGIIILHRDG DEERRYLAFL GKGKQYFEN RTLFKGCESS DWQKIEYNVI QKPHMSLPKN LITPFKKDK ITNERFIDRS KKGAKALIEI DINPSDEFLN NYNLGKHTKE NLDSKFLCDY FKYLMDAIAK YYKGEFNPNF PDVSNPNTQ PFYFSKPKNA YSIKYFGISS KEIEKLIADC YYKEDVYLFQ IYCKDKEIDP KIGKAKYGNF FRTKAEIRKS KGEAEAGNENL NTYKFKLFD EKNLKNQNGI VYKLLNGAKM FYRPPSIKKD EKIDGKWRYK EDKYSNLNITI TCNFSSKKDD LSIDKDINKK IAEVNANSDF RIISIDRGEK NLAYCCVMDE NANILDIKSL NRITTRYDKNG KAIKEKNMFH EVKDGKLYCG EPVYDFYKDY QNLNDEREIK RLVNRRSWNVN IEDIKNLKKG YVALLINYIC KAVVIAINEG KYPIIIVLES DKGMHLHNRVK IEKQIYRGVE EGLVRLKLYF VDKKTDNVLN AWOLLAKPET VGSSLDRKKQ LGIIFYVDPG YTSLTCPCGG FRQRKYIKAEE RAEENFKEIK IKFDGKRYSF AYDYRCIDDN GKEKSKEII YSNVKRLLRS GRNGRAVQIE DVTDELTNLF KKNINIEQD INEQLAGKDN KFWKQOLLWWF NAIEQIRNTQ SLRRKFNTTE NKLEI LENND CDFILCPHCY PDSNNDKFQN KIWNGDANGA FNIGRKGIID IFEIKKHQRM LSDFMEQWGI DKLPKANGN QAVIEIVKND KKYNLCILNN KKIPYYCLRI GKEKIDSIAID DRKCNCQLPDL MVNWKKWDMW LDKWGK 1440 1456	

SEQ ID NO: 294	moltype = AA length = 1486
FEATURE	Location/Qualifiers
source	1..1486
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 294	
MPREVKNVFQD PTNLYELSKT LRFEKPVPE TEKILELNAA KTKKFPKDLY RAENFEIICK YTDDELHRTYI RETLNNVNID YLKFELEIFRI NGKKKNEMTD ENEESDENNE KDDIQKIKKE LRSKIGNLFN KWNNNDKDNKF KDWVKIDVGK KEKEVSGDLF GKELITILKK YFKNKLDSKV NVPMLFFNEQ EKINGEAKKQ RKLAEVAFENF DKFTTYFTDS FYNNRKNYK TEGRVGVQVAT RILDENLPRF CSNLTFAFNEV VSLYSTLLNN FDLGWKEYLN EKKINQTWVE KFELSNDW ALFNDVNYYN QCLLQEGIDK YNYIIKKLNK DINEYQTQNKY KSVEKGNNNN PDINFFQKLH KQIHGERDFK LIEIDIDENK IFTKLPEPI LHSDMKLMLKT IDEEVGVVEI VGAERIIKIF IIKQELKMLEK TYLSRRAIET ISAKWFHSWE TLKDLILGYL NKDLLESKKR KKVPDFVDFN IIKIVLENNK DDYKDLFKRK YFEADKNEF DWIDSSGGTK KLEFGGENWI NFLNVFYEYF GTLLTEYKKN KNALLYLIDK KIDYDKNNEV GQTAIAKNA DSALGIFRMV SYFALRKKG MVEPKNGKDE IFYAFVDRYL DGDDNDREBQ NKIVQYYNTL RNFVTOQAWS IDKVRLCFDC GEFLKGWDKD KIHERGLL RNNNKFYLG LNKHNHQIFI KIKSHDNNNF YYVIYDYKQL NNVYRQTPRL APPSRSVKKG DAYMLRAIQE RKKKFFLEDE EFIELQEIKA EYDKIGNDLS KEKLTKLIEY YKKVVISNYS SLYNVSNLNN KKFNSINEFN QYVENLMSL IPRTRSPDFI KEKISKGELEY LFQIYNKDFD LDESIGKEKF GEDFAPVIMD GKNNLHTEYY KLLFNDSNLK NPNGVVFVFKLS GGAKMFYRPA TESLNGKDR DGNIIKKNPENVIVGQRYK EDKYFLHLPI IILNFVNKGKN YSINDMVNKA ITNASDDQDK FRIIGLDRGE KHLVYVSYVIN ERQEIEIGS LNNISRKDNK GEIIIEKNWY HDKFGNIEKE PTKEYHKDHY NLLDQREIER LKSQRQSWEKI ENIKELEGY ISAVINKICN LVIAKAIKEN IPIVALENLN SGMKRGRKI DKQIYQKLEL KLAKKLNFLV DKKEKNLVA WQFTPQIETF SGDIEKKNQV GIIFYVDPAF TSATCPNCGF RRIKMDPQN AKKKIKDMEI TYENGIYKFD YPIENGENDV YVSDVERLKW DNEKKKVIKT KNVSDDFGKL FEDIKDKNNL KKELLSIGEE NKEFWKEFSS CFNLLLRIRN SKLIKRLND DTGKVEIIAD DLLADRDRDF IYCPQCHFHS EGGDVFGFV KKKYLGKDHF EFNGDANGAY NIARKTIIAV NIKIDYQLGL NHFIEKYRIS ELPNNNGDKK NIFYNNNSYI LSFFEVOQDEK FRKVKVYGLK KDGDROTIQK KEMWYRRYPD IFVNNKEWDK FVQNKS 1440 1486	

SEQ ID NO: 295	moltype = AA length = 1403
FEATURE	Location/Qualifiers
source	1..1403
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 295	
MIFFMSTDIT NKPREKGVFD NFTNLYEFPSK TLTFGLIPLK WDDNKKMIVE DEDPSVLRKY GVIEEDKRIA ESIKIAKFYL NILHRELIGK VLGLSLKFEKK NLENYDRLLG EIEKNNKNEN ISEDKKKEIR KNFKKELSIA QDILLKKVGE VFESNGSGIL SSKNCLDELT KRFTRQEVDK LRRENKRDIGV EYPDVAYREK DGKEETKSFF AMDVGYLDDF HKNRKQLYSV KGKKNSLGRR 60 120 180 240	

-continued

ILDNFEIFCK NKKLYEKYKN LDIDFSEIER NFNLTLEKVF DFDNYNERLT QEGLDEYAKI	300
LGGESVNRQER TANIHLNQI INLYIQKKQS EQKAEQKETG KKKIKFNKKD YPTFTCLQKQ	360
IILSQVFRKEI IIESDRDLIR ELKFFVVEES EKVDKARGII EFLLNHEEND IDLAMVYLPK	420
SKINSFVVKV FKEPQDFLSV FQDGASNLDV VSPDKIKTHL ENNKLTYKIF FKTLKENHD	480
FESFLILLQQ EIDLILLGGE TVTLLGGKES ITSLDEKKNR LKEKLWFGF KVRENEMKMD	540
EEEGEFCSTV LAYSQAVLN1 TKRAEIFWL1 EKQDAKVGED NKDMIFYKK DEFADGFAP	600
FFYFDKFGNY LKRRSRNTTK EIKLHFGNNN LLEGWDMNKE PEYWFILRD RNQYYLIGIK	660
KDGEIFHKKL GNSVEAVKEA YELENEADFY EKIDYKQLNI DRFEGTAFPK KTKTEAFAFQ	720
VCKKRADEF1 GGDTYEFKIL LAIKKEYDDF KARQKEKDW DSKFSKEKMS KLIEYYITCL	780
GKRDDWKRFN LNRFQPKYE1 DRSDFVRHIQ RQAYWIDPRK VSXKDYDKKV AEGEMLFKV	840
HNKDFYDPER KSEDKKNHTA NLFTQYDPER FSCENIKNIK SKDLIESIFE LDGKAEIRFR	900
PXTDDVLLKI YQKKKGKDVTY ADKRDGNKEK EVIYHRRPAK DALTLHLKIR LNFGKHVNLF	960
DFNKLVNTL FAKPVKILG MDRGENNL1Y YCFLDEHGEI ENGKCGSLNR VGEQIITLED	1020
DKKVDFVYDQF QOLLWDREGQ RDWEQKRNQK MTRIKDLKKA YLGNVVWSIS KEMLSGIKEG	1080
VVTIGVLEDL NSNFKRTRFF RERQVYQGF E KALVNLKGYL VDJKKYDNYRN VYQFAPIVDS	1140
VEEMEKNQKI CTLVYVPASY TSKICPHPK C WRERLYMKN SASKEIVGL LKSDGKISY	1200
DQKNDRFYF YQWEQEHKSD GKKKYSQVD KVFSNVSRMR WDVEQKKSID FVDGTDGSIT	1260
NKLKSSLKGK GIELDNINQQ IVNQQKELGV EFFQSIIFYF NLIMQIRNYD KEKSGSEADY	1320
IQCPSCLFDS RKPEMNGKLS AITNGDANGA YNIARKGFMQ LCRIRENPQE PMKLITNREW	1380
DEAVREWDIY SAAQKIPVLS EEN	1403

SEQ ID NO: 296	moltype = AA length = 1356
FEATURE	Location/Qualifiers
source	1..1356
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 296	
MTIKKHKPFT NFECLTPVQK TLRFRLLIPVG RTTEFVKCRN IIeadRKRSE MYPLLKELAD	60
RFYREFMTDQ LSNLLFDWSP LVEALLLARN NTDPRENQRI ASLVRDEQKK YRTLLLKRLS	120
GOVDRNGTPL PKNTASVNKK YYDDLFKARF VTETLTPAYLE HLKNKPDGRI SDELFDAYKD	180
ALDSYQKFTS ALDTNFWQARK NIFTDEDIAT GFAYRIVHEI VPDYLFNRRV YEQHKLDFPE	240
PDLLETELK KKNLINDAES LDALFTIPAI NRLLTQKGVD LHNAVIGFFF TDDHTKVQGF	300
NELANLNQNT LKNVSDNSEI KPVGKMTLKC KHLSISEST SFLFEQIESD DDLLARIIEF	360
NNLTSEPPID GLSIADINDQ LYNIMTGVDP STILVHARNL NKLHSHEASLS WNRLRDGLYQ	420
MATESPYRED EERFKRYIDAS EERDLKSLK NDIYFSLQEL QFALDQSIDL EEEATPTEDI	480
FLPFEPFGMD LKSELTVFLR SIEQOLISSET KLIIGNPDAIA TIKKYLDAIM ARYSIWNLLS	540
CEAELQDQL FYPEYDRVMG SLSNIILLYN LARNYLSRK SSKEKFRLNF DKPLADGWS	600
ESKVPDNFSV LLRKDDLFYL GILKDRKAYR VLSYENCDET AKNIKGYYER MIYHPSDAY	660
RMIPKSTAR EKDVKKHFGQE GETTYGTLYP GASNFVKPFT IPYEIYRLQT ELVNDKRYQ	720
ADYLKQTEDE EGYRQAVTAW IDFCKSYLES YEGTSTFDYS HLLKSEDYED VNQFYADVDR	780
ASYSIYFEKV SVDLIHTMVD RGDLYLFOLY NKDFSPHSTG KPNLHTMYWR ALFSNDNLQN	840
NTIKLNGQAE LFYRPKQVEQ PTVHLOQGSYL LNRFDKHDV IPAGLYCEIY NHINERHPEG	900
YTLLSEEAATQG LLDGRFVYRE APFELVKDKR YTEDQDPLFLHV PLEFNTWASA NVPFENLANE	960
YIKKDSLDHI IIGIDRGERNL LYYSVINLQ DIVKQGSLNT LIQQTTLKGE TVERQIPYQS	1020
MLKQREDERA EARQNWQSID RIKDLKEGYL SHVIYKLSRL IIKYHAIVM ENLNVGFKRG	1080
RFKVERQVYQ KFEVALINKL NALSFKEYEP NELGGVMPRW QLARRVVSPE DTRSQNGIVF	1140
YVPASYTSIV DPVTGPANL YLNRIRNKKL NSFYGHFQEI RYDHEFDRFI FRFNYADFGV	1200
FCRIKVNPSR TWNLVSGERK AFNPKRRMIE KRDTTDEIKL ALEAHGIAYQ NEQNLLPLLL	1260
ENENLLARIH RSFRLLVLQLR NSDSDRDDIV SPALDKENNNT FDSGQQOPYES SLPINADANG	1320
AYNIARKGLL LVDKVNDKR AVLSNREWFE YLMAEE	1356

SEQ ID NO: 297	moltype = AA length = 1296
FEATURE	Location/Qualifiers
source	1..1296
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 297	
MENKDYSLSR FTKQYQNSKT VRFALTPIGR TEEYIIQNQY IEAARRKNQA YKIVKPIIDE	60
KFRSMDDVL THCEKQDW1 LDKLILQYQ1N NKCRENMDAL AEQQEEIRKN ISEEETKSDE	120
YKNNFGKEDS KKLFLKFLPE YLNQINASES DKEAVNEFQK DFKTFSNFL1 VRADIFKADN	180
KHNTIPYRIV NENFMIFAGN KRTQFSNII1R1 IPIQNEEIAK DMGMKKEEWSF YNIQNVDSWF	240
EPDSFQCMCS QKG1QKYNFI IGLVNSYINL YTQQNPQATE VKRSRLKLM LHKQILSDRV	300
NPSWLPQEFK EGEEGEKGQIY EAILALENDL IKNCFDKKYD LWIQSIDIQN PRIYIAASEM	360
ARVSSALHMG WNGLNDVRK1 ILLKSDKKQ1A KVEKILKQDV SLKDLSDT1L RYADIYKEEQ	420
IPSLYQYIEY GSELLQDCAI TRKEYHDL1N GNSNTLSSLNQ NEKLEIGLKA YLDSYQAIHV	480
FLNMFIVGDE LDKDQDFYAE LDGLVESLSE IVPFLYNKVRN YITRKVYSLD KMRIMFERSD	540
FLGGWGQSF1 TKEALLFQKD NLYYIGIIEK KYTNMDVEYL HEGIKEGRNA IRFIYNFQKA	600
DNKNIPRTFI RSKG1TNYAPA VRKYNLP1ES IIDIYDVGF KTNYKKINEK EYYESLEKLI	660
DYFKD1L1K1N ENYKKFHFMW KPSNEYENIN EFYNDTNMAC FLLEKEEINY DHLKBOANQG	720
KIYLEQISSK DFNEGSKGTP NLQTMWREL FSNQNCKDG1V IKLCGGASIY MRDASIKQPV	780
VRHKNAWLIN KWYKVNGQNV VIPDNTTYV1F1 TK1AQERMNE DELTPQERQL WNSGLIQQKK	840
ATHDIMKDR1 FTKKQYMLHA PLTINYKQOD SPRYFNEKVR SFLKDNP1IN IIGDIRGEKN	900
LIYITI1DQK CNILKGMQKS FNQ1EEKGKE GRTIDYVSKL ESVEARHDA1 RKNWQ1G1T1	960
RELKEGYLSQ VVHEITQ1MI QYNAVIVMEN LNMGFKKGRM KVEKSVYQKF EKML1DKMNY	1020
LAFKRDQMGN AIDPYEVGGV MNGYQLTDRF TSFADMGSQN GFIFYVPAAY TSVIDPVTGF	1080
VNVFQKTEFK TNDFLHFRFDS ISWNDKEQSF VFTFDYQNF1K CNGTCYQNKW SLYADVDR1E	1140

-continued

---

TIKKNNQVDR IEPCNPQKL IDFFDKKGII YRDGHNIVDD LEKYDSKTIS EIIHNFKLIL	1200
QLRNSMNRNP TGEIIDYIAS PVMHNEERFD SRKRNPELPQ DADANGAYHI ALKGLMFLQK	1260
INEYADSDGN MDNRKLKITN EEWFKYMQTR KEHTYF	1296
 SEQ ID NO: 298	
FEATURE	moltype = AA length = 1250
source	Location/Qualifiers
	1..1250
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 298	
MNSKTSSITT TNKLSYTGFH NNGQSKTLM FELKPIGRTT EHLDRKGYLA DDIDRAESYK	60
TFKEIADNFH KNLIEEESLAT FTFSDTLKDY FDLWLSPVRT NEDTPKLRKM EAALKRKELE	120
ALKQHPSFAA TSSGKRRLIDE ALYPNASDKE RQCLDRFKGR SSYLDTSYEV RSFIYTDLCK	180
HNTIAYRVN ENLKYLLENL LAYEKLQMTA VNGKLETVKF MFPHDLYPTFS MDISIFFFTSY	240
GFDYCLSLQNA ITRYNLLGG WSDDDNGIHHK GLNNYINEYQ QTVPRNKRLP KLNKLQKML	300
SEENNSMSFII DKFENDVDA NAIRYWLNKC QFDALNLLW TLDVHYNLDE IHFKNDNQGK	360
NISDLSQALF KNHHVIRDW DYDYDIVNAK AKSRQKPERY AEKRDKAFKK INSFSLSYLA	420
NILSQYDNQY ANFVAQFKTR ISVHJQNVOQ MIADKTLDMR LDPLMLLKSII SSDTAKLVEDI	480
KRVLDSLQNA QRMLLPLGE GTEPFEPLMNY VDTLTPLYNK VRNYITTKPY	540
STKKTSLYFG ASNFGSGFDV TKLPVSHTII MRDKGCYLA VIDNNKLIDK LYDHNDNDGY	600
EYMYVKQIPS PIKYFSLKNI LPQDPDPDIR QLLEDRKNGA KWSHDDETFR IDYIVNEFLP	660
TYPPIHDKNG NPYFSWKFKN PDEYEVSLNEF FDDVSKQAYQ TSFRFVSRDF VDDAVENGDI	720
FFFQIYINQDF SPASHGKPSP HTLWFRALFS DVNLKETDRL LKGNNATAYFR PASIFYTDEK	780
WRKGHHYEQL KNKFKPPIIK DKRYALDKEF FHITLEINCN ATVEKYFNRR VNEEIRKADR	840
YNILAINRGE RNLLYAVVMD QDGITLEQKS FNIKSELPN KTVKETDWK KLHAREKERD	900
TARKSWKSIE CIKDLKKGYL SYVVKTTIDM MFEYNAVLLM ENLDIEMKRS RQKIEKNVYA	960
QFQNAAIQKL SMYVNKIDL HIARTAPGGT LNPYQOLTYIP ASRRTKTPQN GFVFFLNPWN	1020
ITIEDPTGK VDLFQTCFRK KNEYKDFFAK FKDIRYNEAQ GWFEFDTDYT YFRDKEKAGK	1080
RTRWNICSYG TRLRRFRNPD KNYAEDAMTV YPTQMLKDFL DEYNIPYAPA SAKSTSISIK	1140
DDIQIDKLD FYKKLLYILK LIVQLRNTSP SSTEQEDDYI ISPVINEDTN WFYDSRDYNE	1200
ESLLPCNTDA NGAYSLALKC NMVIDRIKNT IPGEPVDMYI SNADWLDARQ	1250
 SEQ ID NO: 299	
FEATURE	moltype = AA length = 1328
source	Location/Qualifiers
	1..1328
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 299	
MNSKTSIFDF SNIFGRDTL RFKLTPVTIN SKGEVKDANG ADPYRPYLSA DEELQEQQYEL	60
LKTAIDAYHQ MYIDKKLKH LCLPLTEKGK DGVEHDTAKS KFVKSLAYI KDYGKDKKR	120
QTADLRTFIS RVFADDNII LPPYKVKSDF ITKTLRQWLE QPDTKVEKKE AILDLIEKNG	180
SKLYANCQGLLEARQRLYEK DGKSTSVPYR CIDRNLPRFS KDYHLFEKIL GDCSDVDFDE	240
QLDKDFSEEL KGIAKARLSGIR VESPREVFQP LIILAYLNLQD GQYQLNTIIG TKKEKGTSAL	300
GLNEYINQYN QKQGKIKKKD GIPMLNKLNN QILFGDEVFI ETLAEHKEAI PVIKKVVSSL	360
GKLGAFDGEC HENKLYQFLL SLSSYAGNIY VNTKVVQAIS SSLWDGYSIL YDAVKHDKNG	420
RLIQKSVTLG ELNEKIERLK LEDNRDAFEY FRRSQVKDVH HGSSNVGVFE QLKNCYNDFV	480
EKKILKCSFF SEDQVLVIQR LFDSILSLSRQ IFKVFPCPSLY EVDSDGLFVA KFSDWVNVLR	540
GFDKDYDLLR NLFKRKPYST DDKRIVHFGLS NLMDGFVDSW TDKDKGTQY NGYILRQAHs	600
FVDENTSKEL QEFQRYNNYL VISGNVRFLR EKGNALCEK KKEKLVASDE FSGFERFDYY	660
QSSINNFNRN FKRLTGRDRK SFTDEILQNE GKKELKSTYI ENLIKVAKSM KRLTALQNLV	720
SDEKVRKYSE NLDYETLSAE IGQILATGRE RKYVPVSTNE MKNLLKSSKU NKGEEVRTFM	780
FRISNKDLSY AETMQKGERK SHGAENMHTM YFRALLDTLQ NTFDIFTGTV YFRKASDKRK	840
MYKDEKNPTF RKGDELAFKN PYNKGKKSVF FGYDLIKDR YTKDLSYFLH SITQNYQKKG	900
NAEIDLNAVMR DYIRTQEDLR VIGIDRGERN LLYATMIDGE GHILAQKSFN VIGYQGTAS	960
GESFQVETDY HOLLNEKAER MRSLSQREWE MDKIQDMKDG YLSVVVHELA KMVVENNAII	1020
VMEDLNMGFM ESRQSQLANV YQKFEELRN KLQFYVDKRK RNDPEPSGLYH ALQLAGTETK	1080
DNONGFIFYI PAWNTSKIDS TGTGFVNLFNL KYTNIKDAKA FFSTFEKIIEK NVETGHYDFT	1140
FSYSSMARKK MAKRMDGTRD SWTISTHGSR IVREQKGNWY EYREIESLTS EFDALFEKYS	1200
IDTRCRKLEA IDKCGEAEFF KELIRLMWT LQLRNYDDR GNDYIVSPVCY RGNEYCCSLD	1260
YDNEEGMCIS KIPCQMPKDA DANGAFNIAR KGLMLCERLK KGEKIGVIKG TEWLQYVQNM	1320
SERYVGMV	1328
 SEQ ID NO: 300	
FEATURE	moltype = AA length = 1434
source	Location/Qualifiers
	1..1434
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 300	
MINTMEQPKK SIWDEFTNLV SLQKTLRFEL KPQGKTKELV RTLFINPEEH HHKLISDDLE	60
LSKNYKKVKK LIDCMHRNII NNVLSKHQFT GEELKKLDKN SNAEDNDTET DNADKKDPFA	120
KIRERLTKAL NEESKIMFDN KLLNPKKGKKN KGECELKKWM DKAEDKYFEL GNNEKIDKEA	180
VKADMERLEG FFTYFGGFNK NRENVYSSKK IATAIPFRRI HDNFPFIFKKN IENYKKITEK	240
HPLAFLKLLNE KGANEIIFQLE HFNCNLTDQG IDVYNNKEGLG IIAKEQGKEQ DKGINQLINE	300
YAQKKNKEIK ENAKGGEKPK KIKIAVFDSL KKQILSISKI KSFQFEEVFD TSDIINGINK	360
RYTFLTEAKE GMSIVDEIKK IIGSVGDEKY SLDEIYLKEK FISTLSKLF NYSRYIEVAL	420
EKWyDDDRYDD KINKSGTDKR KFISAKQFSI TSIQDAINYY LEKYEKDEEL SKKYTGKNII	480

---

-continued

---

VDYFKNPTIT	IEHKQKEEV	SEEKDLFKEL	EVRRNVIQHI	LNGDYKKDLK	EJKQQGDGSE	540
KVKAFLDALL	EFNYILNPFI	IKDKNLRKEQ	EKDEEFYNEI	KKLQESIFEA	EILDLYNQTR	600
NYITKKPYKL	DKFKLTFGSG	YFLGSWSNDM	EEREGSILIK	YNEDRSKNYY	LIIMAKPLTD	660
DDKQLFSDN	GTHSKICIYE	FQKMDMKNFP	RMPINSKGNS	PAPAIKEYNL	PIKTIWADYQ	720
KYKNLNQKGK	DKFLEENPDF	RHNLLIGYFKI	CAEKHESLAP	FKHQFSSIWK	PTKEYENLAQ	780
FYKDTLEACY	NLKFEVNFD	NISQLVSSGK	LHLFKIHKNKI	FNPGSTGKCK	LHTLYWEMLF	840
DEKNLQDVIF	KLSGGAAELFY	REASILKNI	IHKIGEVLK	KFFKLPDGKL	EPVPAESIKN	900
LSAYFRKELP	EHELTEIDRK	YIDNYSIIKG	KDDKLGMIMKD	ERFTVKDQF	HCPITINFKS	960
KNKNFINDDV	LEYLHKRDDV	IIIGLDRGER	HLIYLTMIN	DGKIVDNMQF	SLNELQRYYK	1020
INGNEEIQKI	NTSDSNGNKN	VSRTEARRNW	QTENIKNKL	EGYLSLIVHQ	LAKLMIKNA	1080
IVVMENINYC	PKDSRARVEK	QIYQKFESIL	IKKLQYLVMD	KNNLYDSCGV	LSAYQLTNQE	1140
VPAYKYISKQ	NGFLFYVPPD	YTSKIDPETG	FINLLDTRYY	SRKNAVALN	KFDKIVYDRD	1200
NKYFRFDFD	NSTDSNGNKN	FDKLRVDISE	LTRTKWSVC	HAKRSITVQ	INNKWVRQPI	1260
NDVTDKLIKL	FEDKQJGYES	GKCLKDEIL	VEDAKFFEDL	LYRLSVLLAL	RHTYTENGVE	1320
YDLIISSEVFVS	APGSNEFFVS	GKDNNLPPANA	DANGAYNIAR	KGLWLLRKLD	EIDNQELAIK	1380
KFNEELHAKKE	IKKNGEESKE	DKGRDKRKKK	WVSQWCNPKE	WLAFQASMQD	VSEK	1434

SEQ ID NO: 301	moltype = AA	length = 1263				
FEATURE	Location/Qualifiers					
source	1..1263					
	mol_type = protein					
	organism = synthetic construct					
SEQUENCE: 301						
MNNGTNNFQN	FIGISSLQKT	LRNALIPTET	TQQFIVKNGI	IKEDELREGN	RQILKDIMDD	60
YYRGFISETL	SSIDDIDWTS	LFEKMEIQLK	NGDNKDTLIK	EQTEYRKAH	KKFANDDRFK	120
NMPSAKLISD	ILPEFVIHNN	NYSASEKEEK	TQVIKLFSRF	ATSFKDYFKN	RANCFSADDI	180
SSSSCHRIVN	DNAEIFFSNA	LVYRRIVKSL	SNDDINKISG	DMKDSLKEMS	LEEIYSYEKY	240
GEPITQECIS	FYNDICGKVN	SFMNLYCQCN	KENKNLYKLQ	KLHKQILCIA	DTSYEPVYKF	300
ESDEEVYQSV	MGFLDNISSK	HIVERLRLKIG	DNYNGYNLDK	IYIVSKFYES	VSQKTYRDWE	360
TINTALEIHY	NNILPGNGKS	KADKVKA	NDLQKSITEI	NELVSNYKLC	SDDNIKAETY	420
IHEISHILNN	FEAQELKYNP	EIHLVSELYK	ASELKVNLDV	IMNAFHWC	CSV FMTEELVDKD	480
NNPYAELEEI	YDEIYPIVSL	YNLVRNEYVTC	KPYSTKKIKL	NFGIPTLADG	WSKSKEYSNN	540
AIILMRDNLN	YLGIFNAKNK	PDKKIEGNT	SENKGDYKKM	IYNLLPGPNK	MIPKVFLSSK	600
TGVETYKPSA	YLEGYKQNK	HIKSSKDFDI	TFCHDLIDYF	KNCIAIHPEW	KNFGDFDS	660
STYEDISGFY	REVELQGYKI	DWTYI	SEKDI DLLQEKGQLY	LFQIYNKDFS	KKSTGNDNLH	720
TMYLKNLFE	ENLKDV	NGEAEIFFRK	SSIKNPIIH	KGSILVNR	TY EAEKKDQFGN	780
IQIVRKNIPE	NIYQELYK	NDKSDKELSD	EAAKLKNVVG	HHEAATNIVK	DYRTYDKYF	840
LHMPITINFK	ANKTGFINDR	ILOQYIAKEKD	LHVIGIDRGE	RNLIYVSVID	TCGNIVEQKS	900
FNIVNGYDYQ	IKLKQOEGAR	QIARCKE	GKIKEIKEGY	LSLVHIEISK	MVIKYNAAIA	960
MEDLSYGFKK	GRFKVERQV	QKFETMLINK	LNLYL	VFKD	ITENGGLKG YQLTYIPD	1020
KNVGHQCGCI	FYVPAAYTSK	IDPTTGFVN	FKPKDLTVDA	KREFIKKFD	IRYDSEKNLF	1080
CFTFDYNNFI	TQNTVMSKSS	WSVYTYGVRI	KRRFVN	GFRS	NESDTIDITK DMEKTLEM	1140
INWRDGHDLR	QDIIDYEIVQ	HIFEI	FRLTV	QMRNSLSEL	DRDYDRLISP VLNNENI	1200
SAKAGDALPK	DADANGAYCI	ALKGLYEIKQ	ITENWKEDGK	FSRD	KLKISN KDWDFDIQNK	1260
RYL						1263

SEQ ID NO: 302	moltype = AA	length = 1283				
FEATURE	Location/Qualifiers					
source	1..1283					
	mol_type = protein					
	organism = synthetic construct					
SEQUENCE: 302						
MSNLNTFISP	EFTGKIKMTK	SLKVSMIPIG	ETEHWIAKHH	VFEKDRELFD	KNLKARPILD	60
EPIKYTVSRA	LPNLLDFEA	YYLVKDKRTK	ARAFEKELAK	TVTDLILKEM	DELKASASLID	120
SADFVKTLK	KFAGTHDIPG	LSRIBAIESL	EAASKLTA	LNKFNTSRIA	INTLIPKR	180
ENFDIYLSNM	EKIRNVYESG	EFGFLMPEAN	DTLLFMEPAN	YRTVCSP	EDYNRFISGY	240
GDSTESTWIKG	PNQELSEASN	SSKSSNGGVR	RYSLIKPLHK	QHLFETKFFF	TFASISSDDD	300
VRELINSVKG	STEDACLNAL	AFFSSSDPKT	LFVKGSYLH	LSAFLYGSAN	SYILPERIKE	360
GEKARLTAEY	DSVAKKTKA	TTRYNNVAMMN	ISKKINEKIF	SLADIDAYCC	DISKR	420
ILLGIMQM	AAVYGENGKW	SNIEAEAVL	SKTKIWKAKN	GAVAKAVNDY	LTAILEIRKF	480
IRPFALRMEE	LEELGLDTSS	ALDAGEITNT	LFEEAVRAQKL	VHAYL	TRNDA DIALSTQVYF	540
GGTQKAAASW	WNYETGDIQN	RQIALAKKD	MYYFIGTDFE	RGSYSIEPAS	PGEDYYEMLD	600
VKKGQDANKQ	IKKVLFSNKA	IREEHPADSSN	DYVITTKVNS	PITVREIFD	KYQAGEFKLT	660
SQKIRKGDLV	GEKEMTYYR	YMDLIFQMAK	GTYEYSRFNM	DTLLPIEYD	TENDLDDVN	720
TNTIDYRWVR	ISAACIDGV	RNGDIFVFR	QTSSMYGKRE	NKKGTYTGLFL	ELVSDENLLV	780
TRGMSLNSAM	SIYIYRAKVHD	AITVHKKGDV	LVNKFTNARE	RIPENSYKAI	CAFYNSGKSI	840
EELTIEDRDW	LAKATTRICS	GEI	IKDRRY	KNQYSISI	Y NINRSVNRK RV	900
TASAGRIISV	TRGTKDLVY	TVIDDGGSVI	EARSLN	VING INYAKMLAQI	SEERHDSN	960
FDIPKRVETI	KEAYCAF	EIIISAAKHN	ALIVVEL	AIKD	KYSLLD NQVFLKF	1020
LKNCLMSVKV	KGARGM	EPGS	ISNPLQLCNA	DDKSFRNG	GIL YQIPSSYINI CPVTGYADII	1080
DYNNIVSAGD	IRNFFVR	EN	IVYNEKE	ARF EFSFDLKNIP	I KLECPDRT KWT	1140
TTYDPLTKSN	HYVFDAQML	AETVSK	GLD PCANIVEH	ELSAATLKKM FNT	FRNIAKG	1200
IVSECDEV	SYKSPV	IDE	ADIKNKSLDN	KSISEIKCYN	LAKSSSDGEN	1260
KNRYVSSTAI	EWLNYI	QEKR THE				1283

SEQ ID NO: 303	moltype = AA	length = 500
----------------	--------------	--------------

-continued

---

FEATURE	Location/Qualifiers
source	1..500 mol_type = protein organism = unidentified
SEQUENCE: 303	
MEVQKTVMMKT LSLRLIRPLV SQEIEKEIKE EKERRKQAGG TGELDGGFYR KLEKKHSEMF 60	
SFDRLNLLLN QLQREIAKVN NHAISELYIA TIAQGNKSNK HYISSIVYNR AYGYFYNAYI 120	
ALGICSKVEA NFRSNEELLQ QSALPTAKSD NFPFIVLHKQK GAEGEEDGGRF ISTEGSDLIF 180	
EIPIPFYEYN GENRKPEPYKW VKKGGQKPVL KLILSTFRRQ RNKGWAKDEG TDAEIRKVTE 240	
GKYQVSQIEI NRGKKLGHEHQ KWFVNFNSIBQ PIYERKPNRS IVGGLDVGIR SPLVCAINNS 300	
FSPRYSVDSND FPKFSKQAFKA FRRRLLSKNS LKRKGHGAAH KLEPITEMTE KNDKFRKKII 360	
ERWAKEVTINF FVKNQVGIVQ IEDLSTMKDR EDHFFNQYLR GFWPYQQMQT LIENKLKEYG 420	
IEVKRVQAKY TSQLC SNPNC RYWNNYFNFE YRKVNKF PKF KCEKC NLEIS ADYNAARNLS 480	
TPDIEKFVAK ATKGINLPK 500	
SEQ ID NO: 304	moltype = AA length = 507
FEATURE	Location/Qualifiers
source	1..507 mol_type = protein organism = unidentified
SEQUENCE: 304	
MEAAKTVMSKT LSLRLIRPLV SAETIEKEIKE EKERRKQAGG SGELDSCFYK KLEKKHTQMF 60	
GWDKLNLMLS NLQRQIARVF NQSISEMLNS TVIYGKKSNN HTYSKIVYNR AYSVFYNAYL 120	
ALGITSKVEA NFRSNEELLQ KSSLPTAKSD NFPIILLHKQK GVEGEEGGFK ISADGNLIF 180	
EIPIPFYEYD SANKKEPFKW IKKGGQKPTI KLILSTFRRQ RNKGWAKDEG TDAEIRKVIE 240	
GKYQVSHIEI NRGKKLGHDQ KWFVNFTIBQ PIYERKLDKN IIGGIDVGIK SPLVCAINNS 300	
FARYSVDSND VLKFSKQAFKA FRRRLLSKNS LKRSGHGSKN KLDPITRMT E KNDRFRKKII 360	
ERWAKEVTINF FIKNQVGTVQ IEDLSTMKDR QDNFFNQYLR GFWPYQQMQN LIENKLKEYG 420	
IETKRIKARY TSQLC SNPSC RHWNSYFSFD HRKTNNP KPKF KCEKCALEIS ADYNAARNIS 480	
TPDIEKFVAK ATKGINLPDK NENVILE 507	
SEQ ID NO: 305	moltype = AA length = 529
FEATURE	Location/Qualifiers
source	1..529 mol_type = protein organism = unidentified
SEQUENCE: 305	
MAKNTITKTL KLRIVRPYNS AEVEKIVADE KNKDKVKEAC SKHLKVAAYC 60	
TTQVERNACL PCKARKLDDK FYQLRLRGQPP DAVFWQEISE IFRQLQKQAA EIYNQSLIEL 120	
YYEIFIKGKG IANASSVEHY LSDVCYTRAEL PKNAIAAS GLRSKIKSNF RLKELKNMKS 180	
GLPTTTSNDN PIPLVKQKGG QYTGFEEISNH NSDFIPIKIP GRWQVKKEID KYRPWEKFDF 240	
EQVQKSPKPI SLLLSTQRRK RNKGWSKDEG TEAEIKKVMN GDYQTSEIEV KRGSKIGEKS 300	
AWMLNLSDDEV PKIDKGVDPD IIQGDVGVK SPLVCAINNA FSRSYISDND LFHPNKKMFA 360	
RRRILLKKNR HKRAGHGAKN KLKPITVTE KSERFRKKLI ERWACEIADF FIKNKVGTQ 420	
MENLESMRKR EDSYFNIRLR GFWPYAEMON KIEFKLQYQ IEIRKVPANN TSKTC SKCGH 480	
LNNYFNFEYR KKNKFPHFKC EKCNFKENAD YNAALNISNP KLKSTKEEP 529	
SEQ ID NO: 306	moltype = AA length = 726
FEATURE	Location/Qualifiers
source	1..726 mol_type = protein organism = unidentified
SEQUENCE: 306	
MERQKVPQIR KIVRVVPLRI LRPKYSVDIE NALKKFKEKG DDTNTNDFWR AIRDRDTEFF 60	
RKEBLNFSSEDE INQLERDPLF RVGFLDRNRLF SYPDFLQBEKL MKDYNKIISK LFINRQSKSS 120	
FENDLTDEEV BELIEKDVTP FYGAYIKGKI KSVIKSNLGG KFIKSVKIDR ETKKVTKLTA 180	
INIIGLMGLPV AKSDTTPKI IKTNPDYITF QKSTKENLQK IDEYETGIEY GDLLVQITIP 240	
WFKNENKDFS LIKTKEAIEY YKLNGVGKKD LLNINLVLTT YHIRKKWSQ IDGSSQSLVR 300	
EMANGELEEK WKSFFDTFIK KYGDEGKSAL VKRVRVNKKSR AKGEKGRELN LDERIKRLYD 360	
SIKAKSFPS E INLIPYKWW KLHNSIEIPP MVNDIDSMLY GGIDFGEQNI ATLCVKNIEK 420	
DYDFLTIYD NDLLKHAQAS YARRRLMRVQ DEYKARGHKGK SRKTKAQEDY SERMOKLQRK 480	
ITERLVKQIS DFFLWRNKFH MAVCSLRYED LNTLYKGESV KAKRMRQFIN KQQLFNGIER 540	
KLKDYNSEIY VNSRYPHTS RLC SKCGKL N LYPDFFLKPR KNIIIRKNPD GSEIKYMPFF 600	
ICEFCGWKQA GDKNASANIA DKYQDQLNK EKEFCNIRKP KSKKEDIGEE NEEERDYSRR 660	
FNRNSFIYNS LKKDNKLQN E KLFDEWKNQL KRKIDGRNKF EPKEYKDRFS YLFAYYQEII 720	
KNESES 726	
SEQ ID NO: 307	moltype = AA length = 517
FEATURE	Location/Qualifiers
source	1..517 mol_type = protein organism = unidentified
SEQUENCE: 307	
MVPTELITKT LQLRVIRPLV FEEIEKELA E LKEQKEKEFE ETNSLLLESK KIDAKSLK 60	
KRKARSSAAV EFWKIAKEKY PDILTKPEME FIFSEMOKMM ARFYNKSMTN IFIEMNNDEK 120	
VNPLSLISKA STEANQVIK C SSISGLNRK IAGSINKTF KQVRDGLISL PTARTETFPI 180	

-continued

---

SFYKSTANKD	EIPISKINLP	SEEADLTIT	LPFPFFFEIKK	EKKGQKAYSY	FNIIEKSGRS	240
NNKIDLLSLST	HRQRKKGW	EEGGTSAEIR	RLMEGEFDK	WEIYLGEAEK	SEKAKNDLIK	300
NMTRGKLSKD	IKEQLEDIQV	YKFSDNNVES	WNDSLSEKQKQ	ELSKLRKKV	EELKDWHVK	360
EILKTRAKIG	WVELRGKQR	RDRNKFVN	TITRPPFINK	ELDDDTKEFGGI	DLGVKVPFVC	420
AvgHSPARLI	IKENEILQFN	KMVSARNRQI	TKDSEQRKGR	GKKNKFIKE	IFNERNELFR	480
KKIIERWANQ	IVKFFEDQKC	ATVQIENLES	FDRTSYK			517

SEQ ID NO: 308	moltype = AA	length = 481				
FEATURE	Location/Qualifiers					
source	1..481					
	mol_type = protein					
	organism = unidentified					
SEQUENCE: 308						
MKSDTKDKKI	I IHQTKTSL	RIVKPQSIPM	EEPTDLVRYH	QMIIFPVYNN	GAIDLYKKLF	60
KAKIQKGNEA	RAIKYFMNKI	YAPIANTVK	NSYIALGYST	KMSSFSGKR	LWDLRFGEAT	120
PPTIKADFPL	PFYNGSGFKV	SSENCEIFIIG	IPPGQYTAKTT	VSDIEKKT	AWDKFTLEDT	180
TKKTLIELL	STKTRKMNEG	WKNNEGTEAE	IKRVMGTYQ	VTSLEILQRD	DSWFVNPNIA	240
YDSLKKPDR	DKIAGIHMG	TRPLTAVIN	NKYRALSIIYP	NTVMHLTQKQ	LARIKEQRTN	300
SKYATGGHGR	NAKVTGTDL	SEAYRQRKK	IIEDWIASIV	KFAINNEIGT	IYLEDISNTN	360
SFFAAREQKL	IYLEDISNTN	SFLSTYKPYI	SAISDTLOHK	LEEKAIQVIR	KKAYVNVQIC	420
SLCGHYNKG	TYQFRRKNKF	PKMKCQGCLE	ATSTEFNAAA	NVANPDYEKL	LIKHGLLQLK	480
K						481

SEQ ID NO: 309	moltype = AA	length = 358				
FEATURE	Location/Qualifiers					
source	1..358					
	mol_type = protein					
	organism = unidentified					
SEQUENCE: 309						
MSTITRQVRL	SPTPEQSRLL	MAHCQQYIST	VNLVAAFDS	EVLTGKVSTK	DFRAALPSAV	60
KNQALRDAQS	VFKRSVELGC	LPVLKKPHCQ	WNNQNWRVEG	DQLILPICKD	GKTQERFR	120
AAVALEGKAG	ILRIKKRKGK	WIADLTVTQE	DAPESSGSAI	MGVDLGIKVP	AVAHIGHGKGT	180
RRFFGNGRSQR	SMRRRFYARR	KTLQKAKKLR	AVRKSKGKEA	RWMKTINHQ	SRQIVNHABA	240
LGVTIKIEA	LQGIRKGTR	KSRGAAARKN	NRMTNTWSFS	QLTLCITYKA	QRQGITVEQV	300
DPAYTSQDCP	ACRARNQAQD	RTYVCSECW	RGHRTDTVGA	NISRAGL	SG HRRGATGA	358

SEQ ID NO: 310	moltype = AA	length = 507				
FEATURE	Location/Qualifiers					
source	1..507					
	mol_type = protein					
	organism = unidentified					
SEQUENCE: 310						
MTAQKTIKIK	LNPTKEQIIK	LNSIIIEYIK	VSNFTAKKIA	EIQESPTDSG	LTQGTCSECG	60
KEKTYRKYHL	LKKDNKLF	TCYKRKYSQF	TLQKVEFQN	TGLRNVAKLP	KTYYTNAIRF	120
ASDTFSGFDE	I IKKKQNRLN	SIQNRLNFWK	ELLYNPSNRM	EIKIKVVKYA	PKTDTREPH	180
YYSEAEIKGR	I KRLEKQLKK	FKMPKYPEFT	SETISLQREL	YSWKNPDELK	ISSITDKNES	240
MNYYGKEYLK	RYIDLINSQ	PQILLKEENN	SFYLCFPITK	NIEMPKIDDT	FEPVGIDWGI	300
TRNIAVVSIL	DSKTKKPKFV	KFYSAGYILG	KRKHYKSLRK	HFGQKKRQDK	INKLGTKEDR	360
FIDSNIHLA	PLIVKEIRNH	SNKPIILMEN	ITDNREEAAK	SMRQNILLHS	VKSRLQNYIA	420
YKALWNNIPT	NLVKPEHTSQ	ICNRGHHQDR	ENRPKGSKLF	KCVKCNYMSN	ADFNASINIA	480
RKFYIGEYEP	FYKDNEKMKM	GVNISM				507

SEQ ID NO: 311	moltype = AA	length = 534
FEATURE	Location/Qualifiers	
source	1..534	
	mol_type = protein	
	organism = unidentified	
SEQUENCE: 311		

LKLSEQENIT	TGVVKFLKLKD	KETSEGLNDY	FDEYGKAINF	AIKVIOKELA	EDRFAGKVRL	60
DENKKPLLNE	DGKKIWDFPN	EFCSCGKQVN	RYVNGKSLCQ	ECYKVNKFTEY	GIRKRMYSAK	120
GRKAEQDINI	KNSTNPKISK	HFNAYAIREAF	I LDKSISKKQ	KERFRLRREM	KKKLQEFIEI	180
RDGNKILCPK	I EKQRVERYI	HPSWINKEKK	LEDFRGYSMS	NVLGKIKILD	RNIKREEKSL	240
KEKGQINPKA	RRLLMDLKV	FLNDNKISPT	ISKNLPKEYE	LDLPEKEKRL	NWLKEKIKII	300
KNQKPKYAYL	LRKDDNFY	YTLTEFNLK	EDYSGIVGID	RGVSHIAVYT	FVHNNGKNER	360
PLFLNSSEIL	RKLNQKLERD	FLRFLRNHNK	KKIQLILHNY	SKQIVDFAKN	420	
KNAFIVFEKL	EKKPKNRSKM	SKKSQYKLSQ	FTP KKLSDLV	DYKAKREGIK	VLYISPEYTS	480
KECSHCGEVK	NTQRPFNGNS	SLFKCNKCGV	ELNADYNASI	NIAKGLN	IL NSTN	534

SEQ ID NO: 312	moltype = AA	length = 537
FEATURE	Location/Qualifiers	
source	1..537	
	mol_type = protein	
	organism = unidentified	
SEQUENCE: 312		

MEESIITGVK	FKLRIDKET	KKLNEYFDEY	GAINFOAVKI	I QKELADDRF	AGKAKLQNK	60
NPILDENGKK	IYEFPDEFCS	CGKQVNKYVN	NKPFQCECYK	IRFTENGIRK	RMYSAKGRKA	120

-continued

---

EHKINILNST	NKISKTHFNY	AIREAFILDK	SIKKQRKKRN	ERLRESKRL	QQFIDMRDGK	180
REICPTIKGQ	VKDRFIHPSW	ITKDKKLEDF	RGYTLSINTS	KIKIDRNIK	REEKSLKEKG	240
QIIFKAKRML	LDKSIRFVGD	RKVLFTISKT	LPKEYEILDLP	SKEKRLNWLK	EKIEIIKNQK	300
PIKYAYLLRK	I ESEKKPNYE	YYLQYTLEIK	PELKDFYDGA	IGIDRGINH	AVCTFISNDG	360
KVTPPKFFFSS	GEILRLKNLQ	KERDRFLLRK	HNKNRKKGNM	RVIENKINLI	LHRYSKQIVD	420
MAKKLNASIV	FEELGRIGKS	RTKMKKSQRY	KLSLFIFKLL	SDLVDYKSRR	EGIRVTYVPP	480
ETYTSKECSHC	GEKVNTQRPF	NGNYSLFKCN	KCGIQLNSDY	NASINIAKKG	LKIPNST	537

SEQ ID NO: 313      moltype = AA length = 540  
 FEATURE                Location/Qualifiers  
 source                1..540  
 mol\_type = protein  
 organism = unidentified

SEQUENCE: 313  
 LWTVTIVGDFI EMPKQDLVTT GIKFKLDVVK ETRKKLDDYF DEYGKAINFO VKIIQKNLKE 60  
 DRPFAGKIALG EDKKPPLDKD GKKIYNYPNE SCSCGNQVR YVNAKPFCDV CYKLKFTENG 120  
 IIRKRMYSARG RKADSDINIK NSTNKISKTH FNYAIREGFI LDKSLKQRS KRIKLLLELK 180  
 RKLQEFIDIR QGQMVLCPKI KNQRVDFKFIH PSWLKRDKKL EEFRGRYSLSV VEGKIKIFNR 240  
 NILREEDSLR QRGHVNFKAN RIMLDSKVSF NKGLPKEYLL DLPKKENKLS 300  
 WLNEKISLQ LQKPKYAYLL RREGSFFIOY TIEENVPKTF5 DYLGAIIGIDR GISHIAVCTF 360  
 VSKNGVNKP VFFSSGEILK LKSLQKQRLD FLRGKHKNIR KKSNNMRNIIDN KINLILHKYS 420  
 RNIVNLAKSE KAFIVPEKLE KIKKSFRKMS KSLQYKLSQF TFKKLSDLVE YKAKIEGIKV 480  
 DYVPPEYTSK ECSVHCGEKVD TQRPFNGNSS LFKCNKCRVQ LNADYNASIN IAKKSLNISN 540

SEQ ID NO: 314      moltype = AA length = 542  
 FEATURE                Location/Qualifiers  
 source                1..542  
 mol\_type = protein  
 organism = unidentified

SEQUENCE: 314  
 MSKTTISVKL KIIDLSSSEKK EFLDNYFNEY AKATTFCQLR IRRLRNTHW LGKKEKSSKK 60  
 WIFESGICDL CGENKELVNE DRNSCEPAKI CKRCYNGRYG NQMIRKLFVS TKKREVQENM 120  
 DIRRVAKLNN THYHRIPEEA FDMIKAADTA EKRRKKNVEY DKKRQMEFIE MFNDEKKRAA 180  
 RPCKPNERET RVVHISKLES PSKGYTLNGI KRKIDGMGKK IERAEGLSR KKIFGQYQGNR 240  
 IKLDSNWRFL DLAEESEITIP SLFKEMKLRI TGPTNVHSKS GQIYFAWE RINKQPNNYC 300  
 YLIRKTSSNG KYEYQYLQTY EAEVEANKEY AGCLGVDIGC SKLAAAVYYD SKNKKAQKPI 360  
 EIFTNPIKKI KMRREKLIKLSR LSRVKVRHRR RKLMLQSKTE PIIDYTCCHKT ARKIVEMANT 420  
 AKAFISMENL ETGIKQKQQA RETKKQKFYR NMPLFRKLSK LIEYKALLKG IKIVYVKPDY 480  
 TSQTCSSCGA DKEKTERPSQ AIFRCLNPTC RYYQRDINAD FNAAVNIACK ALNNTEVVTT 540  
 LL

SEQ ID NO: 315      moltype = AA length = 564  
 FEATURE                Location/Qualifiers  
 source                1..564  
 mol\_type = protein  
 organism = unidentified

SEQUENCE: 315  
 MARAKNQPYQ KLTTTTGIGKF KLDLSEESEGK RFDEYFSEYA KAVNFCAKVI YQLRKNLKFA 60  
 GKKELAAKEW KFEISNCDFC NKQKEIYYYKN IANGQKVCKG CHRTNFSDNA IRKKMIPVKG 120  
 RKVESKFNH NTTKKISGTH RHWAFFEDAAD IIESMDKQRK EKQKRLRREK RKL SYFFELF 180  
 GDPAKRYELP VKGKQVRPVR LHKIIIDKDSL TKKRGYSLSY IKNKIKISER NIERDEKSLR 240  
 KASPIAFGAR KIKMSKLDPK RAFDLENNFF KIPGKVIKGQ YKFFGTNVAN EHGGKKFYKDR 300  
 ISKILAGKPK YYFYLRLKKVA ESDGNPIFFEY YVQWSIDDET PAITSYDNIL GIDAGITNLA 360  
 TTVLIPKMLS AEHCSHCGNN HVKPIFTKFF SGKELKAIFI KSRKQKYFLR GKHNLVVKIK 420  
 RIRPIEQKVD GYCHVVSKQI VEMAKERNSC IALEKLEKPK KSKFRQRRE KYAVSMFVFK 480  
 KLATFIKYKA AREGIEIIPV EPEGTTSYTC HCKNAQNQNR PYFKPNSKKS WTSMFKGK 540  
 GIELNSDYN A FNIAQKALN MTS A

SEQ ID NO: 316      moltype = AA length = 610  
 FEATURE                Location/Qualifiers  
 source                1..610  
 mol\_type = protein  
 organism = unidentified

SEQUENCE: 316  
 MDEKHFFCSY CNKELKISKN LINKISKGSI REDEAVSKAI SIHNKKEHSL ILGIKFKLF 60  
 ENKLDKKKLN EYFDNYSKAV TFAARIFDKI RSPYKFGLK DKNTKKWTFF KAKCVFCLEE 120  
 KEVAYANEKD NSKICTECYL KEPFGENGIRK KIYSTRGRKV EPKYNIFNST KELSSTHYNY 180  
 AIRDAFQLLD ALKKQRQKKL KSIFNQKLLR KEPEDIFSDP QKRIELSLKP HQREKRYIHL 240  
 SKSGQESINR GYTLRFVRGK IKSLTRNTER EKSLRKKTP IHFKGNRLMI FPAGIKFDA 300  
 SNKVKISISK NLPNEFNFSG TNVKNEHGKS FFKSRIELIK TQKPKYAYVL RKIKREYSKL 360  
 RNYEIEKIRL ENPNADLCDF YLQYTIETES RNNEEINGII GIDRGITNLA CLVLLKGDK 420  
 KPSGVKFYKG NKILGMKIA YKHLYLLKGK RNKLRQKQRI RAIEPKINLI LHQISKDIVK 480  
 IAKEKNFAIA LEQLEKPKKA RFAQRKKEKY KLA LFTFKNL STLIEYKSKR EGIPVIYVPP 540  
 EKTSQMCSC AINGDEHVDT QRPYKKPNAQ KPSYSLFKCN KCGIELNADY NAAPNIAQKG 600  
 LKTLMLNHSH 610

-continued

---

SEQ ID NO: 317	moltype = AA length = 369
FEATURE	Location/Qualifiers
source	1..369
	mol_type = protein
	organism = unidentified
SEQUENCE: 317	
MLQTLVVKLD PSKEQYKMLY ETMERFNEAC NQIAETVFAI HSANKIEVQK TVYYPIREKF 60	
GLSAQLTILA IRKVCBAYKR DKS1KPEFRL DGALVYDQRV LSWKGGLDKVS LVTLQGRQII 120	
PIKFGDYQKA RMDRIRGQAD LILVKGVFYL CVVVEVSEES PYDPKGVLGV DLGIKNLAVD 180	
SDGEVHSGEQ TTNTTRERLDS LKARLQSKGT KSAKRHLKLL SGRMAKFSKD VNHC1ISKLV 240	
AKAKGTLMSI ALEDLQGIRD RVTVRKAQQR NLHTWNFGLL RMFVDYKAKI AGVPLVFDV 300	
RNTSRTCPSC GHVAKANRPT RDEFRCVSCG FAGAADHIAA MNIAFRAEVS QPIVTRFFVQ 360	
SQAPSFRVG	369
SEQ ID NO: 318	moltype = AA length = 552
FEATURE	Location/Qualifiers
source	1..552
	mol_type = protein
	organism = unidentified
SEQUENCE: 318	
MDEEPDASAEP NLAPISVKLK LVKLDGEKLA ALNDYFNEYA KAVNFCELKM QKIRKNLVNI 60	
RGTYLKKEKKA WINQTGECCCI CKKIDELCRE DNPDINGKIKI CKKCYNGRYG NQMIRKLFLVS 120	
TNKRKAVPKSL DIRKVARLHN THYHRIPPEA ADIKAIAETE ERKRRNRILF DERRYNELKD 180	
ALENEEKRVA RPKPKPEREV RYVIPSKKDT PPKGTMNAL VRKVSGMAKK IERAKRNLNK 240	
RKKIEYLGRR ILLDKNWVRF DFDFSEISIP TMKEFFGEMR FEITGPSNV SPNGREYFTK 300	
WFDRKIQPD NYCYLLRKES EDETDFYLOV TWRPDAHPKK DYTGCLGIDI GGSKLASAVY 360	
FDADKRNRAKQ VPIQIFSNPIG KWTKTRQKVI KVLSKAAVRH KTKKLESLRN IEPRIDVHCH 420	
RIARKIVGMA LAANAPISME NLEGGIKEQ KAKETKKQKF SRNMVFVRKL SKLIEYKALM 480	
EGVKVVYIVP DYTSQLCSSC GTNNTRKRPQ AIFMCQNTEC RYFGKNINAD FNAAINIAKK 540	
ALNRKDIVRE LS	552
SEQ ID NO: 319	moltype = AA length = 534
FEATURE	Location/Qualifiers
source	1..534
	mol_type = protein
	organism = unidentified
SEQUENCE: 319	
MEKNNSEQTS ITTGKFKKLK LDKETKEKLN NYFDEYGKAI NFAVRIIOMQ LNDDRLAGKY 60	
KRDEKGKPIL GEDGKIKILET PNDFCSCGNQ VNHYVNGVSE CQECKYKREFS ENGIRKRMYS 120	
AKGRKAEQDI NIKNSTNKIS KTHFNQYAIR EAFNLDKSIKK QREKRFKKLK DMKRKLQEFQ 180	
EIRDGKRVIC PKIEKQKVER YIHPSWINKE KKLLEFRRGYS LSIVNSKIKS FDRNIOREEK 240	
SLIKEKGQINF KAQRQLMDKIS VKFLKDNKVS FTISKELPKT FELDLPKKEK KLNWLNEKLE 300	
IICKNQPKYA YLLRKENNIF LQYTFDLSPE IHSEYSGAVG IDRGVSHIAV YTFLDKDGKN 360	
ERPPFFLSSSG ILLRKLQKRE RDKFLRKHHN KIRKKGNMRN IEQKINLILH EYSKQIVNFA 420	
KDKNAFIVFE LLEKPKKSRE RMSKKIQYKL SQPTFKKLSD LVDYKAKREG IKVIVYEPAY 480	
TSKDCSHCGE RVNTQRPFNG NFSLFKCNKC GIVLNSDYNA SLNIARKGLN ISAN	534
SEQ ID NO: 320	moltype = AA length = 577
FEATURE	Location/Qualifiers
source	1..577
	mol_type = protein
	organism = unidentified
SEQUENCE: 320	
MAEKKFFFCE KCNKDIKIPK NYINKQGAEE KARAKHEHRV HALILGIKFK IYPKKEDISK 60	
LNDYFDEYAK AVTFTAKIVD KLKLFPLFAG KRDKDTSKKK WVFPVDKCSCE CKEKTEINYR 120	
TKQGKNICNS CYLTEFGEQG LLEKITYATKG RKVSSSFNLN NSTKKTGTH NNYVVKESLQ 180	
LILDALKKQRS KRLKQLLSNTK RKLQFEEIMP EKBDKRFOLP LKEKQRELRF IHVSKQDRAT 240	
EFKGYTMNKI KSKIKVLRN IEREQRSLNR KSPVFFRGR TIRLSPSVQFD DKDNKIKLTL 300	
SKELPKYEF SGLNVANEHG RKFFAEKLLK IKENKSKYAY LLRROVNKNN KKPIYDYYLQ 360	
YTVEFLPNII TNYNGILGD RGINTLACIV LLENKKEPKS FVKKFSGKGI LNLKNKRRKQ 420	
LYFLKGVHNK YRKQQKIRP1 EPRIDQILHD ISKQOIDLAK EKRAVISLEQ LEKPKPKF 480	
QSRKAKYKLS QFNFKTLSNY IDYKAKKEGI RVIYIAPEMT SQNCSCAMK NDLHVNTQRP 540	
YKNTSSLFKC NKCGVELNAD YNAAFNIAQK GLKILNS	577
SEQ ID NO: 321	moltype = AA length = 613
FEATURE	Location/Qualifiers
source	1..613
	mol_type = protein
	organism = unidentified
SEQUENCE: 321	
MISLKLKLLP DEEQKKLLDE MFWKVASICT RVGFGRADKE DLKPPKDAEG VWFSLTQLNQ 60	
ANTDINDLRE AMKHQKHRLE YEKNRLEAQR DDTQDALKNP DRREISTKRK DLFRPKASVE 120	
KGPLKLKYHQ ERYWVRRLKE INKLIERKTK TLTKIEKGRI KFKATRITH QGSFKIRFGD 180	
KPAFLIKALS GKNQIDAPFV VVPEQPICGS VVNSKKYLDE ITTNFLAYSV NAMLFGLSRS 240	
EEMLLKAKRP EKIKKKEEKL AKKQS AFENK KKELOKLLGR ELTQQEEAII EETRNQFQD 300	
FEVKITKQYS ELLSKIANEL KQKNDFLKVN KYPILLRKPL KKA KS KINN LSPSEWKYYL	360

-continued

---

QFGVKPLLKQ	KSRKSRNVL	GIDRGLKHLL	AVTVLEPDKK	TFVWNKLYPN	PITGWKWRRL	420
KLLRSLRLRK	RRIKSQKHET	IHENQTRKKL	KSLQGRIDL	LHNISRIV	TAKEYDAVIV	480
VEDLQSMRQH	GRSKGNRLKT	LNYALSLFDY	ANVMQLIKYK	AGIEGIQIYD	VKPAGTSQNC	540
ACYLLAQRD	HEYKRSQENS	KIGVCLNPNC	QNHKKQIDAD	LNAARVIASC	YALKINDSQP	600
FGTRKRFKRR	TTN					613
SEQ ID NO: 322		moltype = AA	length = 615			
FEATURE		Location/Qualifiers				
source		1..615				
		mol_type = protein				
		organism = unidentified				
SEQUENCE: 322						
METLSLKLKL	NPSKEQLLLV	DKMFWKWASI	CTRLGLKKA	MSDLEPPKDA	EGVWFSKTQL	60
NQANTDVNDL	RKAMHQGKR	IYEELDKVEN	RRNEIQEML	KPDRRDISP	RKDLFPRPKAA	120
VEKGYKLKY	HKLGYWSKEL	KTANKLIERK	RKTLAKIDAG	KMKFKPTRIS	LHTNSFRIKF	180
GEERP KIALST	TSKHEKIELP	LITSLQRPLK	TSCAKKSKEY	LDAAILNFLA	YSTNAALFGL	240
SRSEEMLLKA	KKPEKIEKRD	RKLATKRESF	DKKLKTLEKL	LERKLSEKEK	SFVFRKQTEF	300
FDKFCITLDE	TYVEALHRIA	EELVSKNKYL	EIKKYPVLLR	KPESRIRS	KK LKNLKPEDWT	360
YYIQFGQPQ	DLTPKPIKT	TVLGIDRGR	HLLAVSIFD	RTKTFTFNRL	YSNPIVDWKW	420
RRRKLLRSIK	RLKRLRKSEK	HVLHENOFK	AKLRSLEGRI	EDHFHNLSKE	IVDLAKENN	480
VIVVENLGGM	RQHGRGRGK	LKALNAYALSH	FDYAKVQMLI	KYKAELAGVF	VYDVAPAGTS	540
INCAYCLLND	KDASNYTRGK	VINGKKNTKI	GECKTCKKEF	DADLNALARVI	ALCYEKRLND	600
PQPFGTGKQF	KPKKP					615
SEQ ID NO: 323		moltype = AA	length = 775			
FEATURE		Location/Qualifiers				
source		1..775				
		mol_type = protein				
		organism = unidentified				
SEQUENCE: 323						
MKALKLQLIP	TRKQYKILDE	MFWKwaslan	RVSQKGESKE	TLAPPKDIQK	IQFNATQLNQ	60
IEKDIKDLRG	AMKEQQKQKE	RLLLQIQERR	STISEMLND	NNKERDPHRP	LNFRPKGWRK	120
FHTSKHWVGE	LSKILRQEDR	VKKTIERIVA	GKISFKPKRI	GIWSSNYKIN	FFKRKISINP	180
LNSKGKFELTL	MTEPTQDLDL	KNGGKSVLN	KRYLDDDSIKS	LLMFALHSRF	FGLNNNTDTYL	240
LGGKINPSLV	KYYKKNQDMG	EFGREIVEKF	ERKLKQEINE	QQKKIIMSQI	KEQYSNRDSA	300
FNKDYLGLIN	EFSEVFNQRK	SERAELYLDS	FEDKIKQIKQ	EIGESLNISD	WDFLIDEAKK	360
AYGYEEGFT	YVYSKRYLET	LNKKIVKAVLI	TDIYFDLRKY	PILLRKL	DK IKKISNLKPD	420
EWSSYYIQFGY	DSINPVQLMS	TDKFLGIDRG	LTHLLAYSVF	DKEKKEFIIN	QLEPNPIMGW	480
KWKLRLVKRS	LQHLERR	QKVMKLPENQ	MKKKLKSIEP	KIEVHYHNIS	RKIVNLAKDY	540
NASIVRKVSE	GGGLKQHGK	RKARNRSLN	ALSLFDYGIKI	ASLIKYKADL	EGVPMYEVLP	600
AYTSQQCAKC	VLEKGSFVDP	IIIGYVEDIG	IKGSLLDSL	EGTELSSIQV	LKKIKNIEL	660
SARDNHNKEI	NLILKYNFKG	LVIVRGQDK	EIAEHPIKEI	NGKFAILDFV	YKRGKEKVKG	720
KGNQKVRYTG	NKKVGYCSKH	GQVDADLN	ASR VIALCKY	LD INDPILFGEQ	RKSFK	775
SEQ ID NO: 324		moltype = AA	length = 777			
FEATURE		Location/Qualifiers				
source		1..777				
		mol_type = protein				
		organism = unidentified				
SEQUENCE: 324						
MVTRAIKLKL	DPTKNQYKLL	NEMFWKwasl	ANRFSQKGAS	KETLAPKDTG	QKIQFNATQL	60
NQIKKDVDDL	RGAMEQKQKQ	KERLLIQIQE	RLLTISEILR	DDSKEKDPH	RPQNFRPFGW	120
RRFHTSAYWS	SEASKLTRQV	DRV RVTIERI	KAGKINFPKP	RIGLWSSTYK	INFLKKKINI	180
SPLKSKSFEL	DLITEPQQKI	IGKEGGKSA	NSKKYLD	KSLLIFAIKS	RLFGLNNKDK	240
PLFENIITPLN	LVRYHKKGQE	QENFKKEV	KFENKLKKEI	SQKQKEIIFS	QIERQYENRD	300
ATFSEDYLRA	ISEFSEIFNQ	RKKERAKELL	NSFNEKIRQL	KVEVGNGNISE	EDLKILEVEA	360
EKAYNYENG	IWEWEYSEQFL	GVLEKIA	RAV LISDN	YFLK KYPILIRKPT	NKSKKITNLK	420
PEEWDDYYIQF	GYGLINSPMK	IETKFMGID	RGLTHLL	AYS IFDRDSEKFT	IN QLELPNIK	480
GWKWLRLVKV	RSLQHLERR	RAQKGVKLP	EQMKKRLKSI	EPKIESYHNN	LSRKIVNLAK	540
ANNASIVVLES	LEGGGLQHGQH	RKKNRSHRAL	NYALSLFDY	KIASLIKYKS	DLEGVPMYEV	600
LPAYTSQQCA	CVCLKKGSFV	EPEIIGYIEB	IGFKENL	LTLFEDTGLSSV	QVLKSKNKM	660
TLSARDKEGK	MVDLVLKYNF	KGLVISQEK	KEEIVEFPIK	EIDGKFAVLD	SAYKRGKERI	720
SKKGNQKLVY	TGNKKVGYCS	VHGQVADLN	ASR VIALCKY	LG INEP IVFG	EQRKSFK	777
SEQ ID NO: 325		moltype = AA	length = 610			
FEATURE		Location/Qualifiers				
source		1..610				
		mol_type = protein				
		organism = unidentified				
SEQUENCE: 325						
LDLITEPIQP	HKSSSLRSKE	FLEYQISDFL	NFSLHSLFFF	LASNEGPLVD	F KIYDKIVIP	60
KPEERFPKKE	SEEGKKLDSF	DKRVEEYSD	KLEKKIERKL	NTEEKVN	DR EKTRI WGEVN	120
KLEEIRSIID	EINEIKKQKH	I SEKS KLLGE	KWKKVNNIQE	TLLSQEV	VSL ISNLSDEL	180
KKKELLAKKY	SKFDDKIKKI	KEDYGLEFDE	NTIKKEGEKA	FLNPDKFSKY	QFSSSYLKL	240
GEIARSLITY	KGFLDLNKYP	IIFRK PINKV	KKIHNL	EPD WPE	WKYYIQFGYE	300
NILGIDRGLT	HILAYS FEP	RSSKPI	LNKL	EPNPIEGW	KLRLKRRS	360

-continued

---

NVKLPENQMK	KNLRSIDEDKV	ENLYHNLSRK	IVDLAKEKNA	CIVFEKLEGQ	GMKQHGRKKS	420
DRLRLGLNYKL	SLFIDYGIKAK	LIKYKAEIIG	IPIYRIDSAV	TSQNCAKCVL	ESRRFAQPEE	480
ISCLDDDFKEG	DNLDKRILEG	TGLVEAKIYK	KLLKEKKEDF	EIEEDIAMFD	TKKVIKENKE	540
KTVILDVYVT	RRKEIIIGTNH	KKNIKGIAKY	TGNTKIGYCM	KHGQVDADLN	ASRTIALCKN	600
FDINNPETWK						610
SEQ ID NO: 326			moltype = AA	length = 632		
FEATURE			Location/Qualifiers			
source			1..632			
			mol_type = protein			
			organism = unidentified			
SEQUENCE: 326						
MSDESLVSSE	DKLAIKIKIV	PNAEQAKMLD	EMFKKWSSIC	NRISRGKEDI	ETLRPDEGKE	60
LQFNSTQQLNS	ATMDVSDLKK	AMARQGERLE	AEVSKLRLRQE	ETIDASLRDP	SRRHTNPQKP	120
SSFYPSDWI	SGRLLTPRFHT	ARHYSTELRK	LKAKEDKMLK	TINKINGKGI	VFKPKRITLW	180
PSSVNMAFKG	SRLLLKPFAN	GFEMELPIV	SPOKTADGKS	QKASABYMRN	ALLGLAGYSI	240
NQLLFGMNRS	QKMLANAKKP	EKVEKFLEQM	KNKDANFDKK	IKALEGKWL	DRKLKESEKS	300
STAVVRLTFF	KSGKVELNED	YLKLLKHMAN	EILERDGFWN	LNKYPILSRK	PMKRYKQKNI	360
DNLKPNMWKY	YIQFGYEPIF	ERKASGKPKN	IMGDRLGTLH	LLAVALFSPD	QQKFLFNHLE	420
SNPIMHWKWK	LRKIRRSIQH	MERRIRAEKN	KH1HEAQQLKK	RLGSIEEKTE	QHYHVSSKI	480
INWAIEYEAA	IVLESLSHMK	QRGGKKSVRT	RALNYALSFL	DYEKVARLIT	YKARIRGIPV	540
YDVLPGMTSK	TCATCLLNGS	QGAVVRGLET	TKAAGKATKR	KNMKIGKCMV	CNSSENSMID	600
ADLNAARVIA	ICKYKNLNDP	QPAGSRKVFK	RF			632
SEQ ID NO: 327			moltype = AA	length = 625		
FEATURE			Location/Qualifiers			
source			1..625			
			mol_type = protein			
			organism = unidentified			
SEQUENCE: 327						
MLALKLIMP	TEKQAEILDA	MFWKVASICS	RIAKMKKKVS	VKENKKELES	KIPSNSDIWF	60
SKTQLCQAEV	DVGDHKKALK	NFEKRQESLL	DELKYKVKA	NEVINDESKR	EIDPNNPSKF	120
RIKDSTKKG	LNSPKFTL	KWQKILQENE	KRIKKKESTI	EKLKRGNIFF	NPTKISLHEE	180
EYSINFGSSK	LLLNCFYKYN	KKSGINSDQL	ENKFNEFQNG	LNIICSPLOP	IRGSSKRSFE	240
FIRNSINDN	MYSLYAKLFG	IPRSVAKLMK	SNKDENKLKL	EEKLKMKKSS	FNKTVKEFEK	300
MIGRKLSDNE	SKILNDESKK	FFEIIKSNNK	YIPSEEEYLKL	LKDISEEIYN	SNIDFKPYKY	360
SILIRKPLSK	PKSKKLYNLK	PTDYKYYLQL	SYEPFSKQLI	ATKTIKGIDR	GLKHLLAVSV	420
FDSQNKEFVY	NKLIKNPVFK	WKKRYHDLR	SIRNRERRIR	ALTGVHICHEN	QLIKKLKSMK	480
NKINVLYHNV	SKNIVDLAKK	YESTIVLERL	ENLKQHGRSK	GKRYKLLNYV	LSNPDYKKIE	540
SLISYKAKKE	GVPVSNINPK	YTSKTCAKCL	LEVNLSELK	NEYNRDSKNS	KIGICNIHGQ	600
IDADLNARV	IACLYSKNLN	EPHPK				625
SEQ ID NO: 328			moltype = AA	length = 517		
FEATURE			Location/Qualifiers			
source			1..517			
			mol_type = protein			
			organism = unidentified			
SEQUENCE: 328						
VINLFGYKFA	LYPNKTQEEEL	LNKHLGECCW	LYNKAIEQNE	YYKADSNIEE	AQKKFELLPD	60
KNSDEAKVLR	GNISKDNVYV	RTLVLKKKSE	INVQIRKAVV	LRLPAETIRNL	AKVKKKGSLV	120
GRILKFIPIRE	WDVLPFKQSD	QIRLENLYI	LEPYGRLKFK	MHRPLLGKPK	TFCIKRTATD	180
RWTISFSTEY	DDSNMRKNDG	GQVGIDVGLK	THLRLSNENP	DEDPRYPNPK	IWKRYDRRLT	240
ILQRRIKSJK	KLGKRNTRLRL	LRLSRLWEKI	RNSRADLQJN	ETYEILSENK	LIAIBDLNVK	300
GMQEKKDKKG	RKGRTTRAQEK	GLHRSISDAA	FSEFRRVLEY	KAKRFGSEVK	PVSAIDSKE	360
CHNCGNKKG	PLESRIYEC	KCGKLKIDRL	NSAKVILARA	TGVRPGSNAR	ADTKISATAG	420
ASVQTEGTVS	EDFRQQMETS	DQKPMQGEGS	KEPPMNPEHK	SSGRGSKHVN	IGCKNKVGLY	480
NEDENSRSTE	KQIMDENRST	TEDMVEIGAL	HSPVLT			517
SEQ ID NO: 329			moltype = AA	length = 410		
FEATURE			Location/Qualifiers			
source			1..410			
			mol_type = protein			
			organism = unidentified			
SEQUENCE: 329						
MIASIDYEAV	SQALIVFEFK	AKGKDSQYQA	IDEAIRSYRF	IRNSCLRYWM	DNKKVGKYDL	60
NYCKVLAQK	YPFANKLNSQ	ARQSAEAECSW	SAISRFYDNC	KRKVSGKGF	PKFKKHARSV	120
EYKTSGWKL	ENRKAIFTD	KNGIGKLKL	GTYDLHFSQL	EDMKRVLVR	RADGYVQFC	180
ISVDDVKVETE	PTGKAIGLDV	GIKYFLADSS	GNTIENPQFY	RKAEEKKLNR	NRRKSKKYIR	240
GVKPQSKNYH	KARCRYARKH	LRVSRQRKEY	CKRVAYCVIH	SNDVVAYEDL	NVKGMVKNRH	300
LAKSISDVAW	STFRHWLEYF	AIKYGKLTIP	VAPHNTSQNC	SNCDKKVPKS	LSTRTHICHHH	360
CGYSEDRDVN	AAKNILKKAL	STVGQTGSLK	LGEIEPLLVL	EQSCTRKFDL		410
SEQ ID NO: 330			moltype = AA	length = 486		
FEATURE			Location/Qualifiers			
source			1..486			
			mol_type = protein			

---

-continued

---

SEQUENCE: 330    organism = unidentified

LAAEENTLHLT	LAMSLPLNDL	PENRTRSELW	RRQWLPQKQL	SLLLGVNQSV	RKAAADCLRW	60
FPEPYQEELLWW	EPTDPDGKKL	LDKEGRPIKR	TAGHMRVLRK	LEEIAPFRGY	QLGSAVKNGL	120
RHKVADLLS	YAKRKLDPQF	TDKTSYPSIG	DQFPIVWTGA	FVCYEQSITG	QLYLYLPLFP	180
RGSHQEDITN	NYDPDRGPAL	QVFGEKETAR	LSRSTSGLLL	PLQFDKWGEA	TFIRGENNPP	240
TWKATHRRSD	KKWLSEVLLR	EKDFQPKRVE	LLVRNNGRIFV	NVACEIPTKP	LLEVENFMGV	300
SFGLEHLTVT	VVIINRDGNVV	HQRQEPARRY	EKTYFARLER	LRRRGPPFSQ	ELETPHYRQV	360
AQIVEEALRF	KSVPAVEQVG	NIPKGYRNPR	LNLRLSYWPF	GKLADLTSYK	AVKEGLPKPY	420
SVYSATAKML	CSTCGAANKE	GDQPISLKG	TVYCGNCGTR	HNTGFNTALN	LARRAQELFV	480
KGVVAR						486

SEQ ID NO: 331    moltype = AA length = 602

FEATURE    Location/Qualifiers

source    1..602

mol\_type = protein

organism = unidentified

SEQUENCE: 331

MSQSLWKWD	MAGRDKDASR	SLQKSAVEGV	LLHLTASHRV	ALEMLEKSVS	QTAVATMEAA	60
QQRLVIVLED	DPTKATSRKR	VISADLQFTR	EEPGSLPNWA	QKLASTCPEI	ATKYADKHIN	120
SIRIAWVAK	ESTNGDAVEQ	KLQWQIRLLD	VTMFLQQQLVL	QLADKALLEQ	IPSSIRGGIG	180
QEVAQQVTSH	IQLLDSGTVL	KAELEPTISDR	NSELARKQWE	DAIQTVCTYA	LPFSRERARI	240
LDPGKYAAED	PRGDRLINID	PMWARVLPKG	TVKSLPLLFV	SGSSIRIVKL	TLPRKHAAGH	300
KHTFTATYLV	LPVSREWINS	LPGTVQEKVQ	WWKHPDVILAT	QELLVKGAL	KKSANTLVIP	360
ISAGKKRFFN	HILPALQRGF	PLQWQRRIVGR	SYRRPATHRK	WFAQLTIGYT	NPSSLPEMAL	420
GIHFGMKDIL	WVALADKQGN	ILKDGSIQPN	SILDFSLQEK	GKIERQOKAG	KNVAGKKYKG	480
SLLNATYRVV	NGVLEFSKGI	SAEHASQPIG	LGLETIRFVD	KASGSSPVNA	RHSNWNYGQL	540
SGIFANKAGP	AGFSVTEITL	KKAQRDLSDA	EQARVLAIEA	TKRFASRIKR	LATKRKDSDL	600
FV						602

SEQ ID NO: 332    moltype = AA length = 494

FEATURE    Location/Qualifiers

source    1..494

mol\_type = protein

organism = unidentified

SEQUENCE: 332

VEPVEKERFY	YRTYTFRLDG	QPRTQNLTTQ	SGWGLLTAV	LDNTKHYWEI	VHHARIANQP	60
IVFENPVIDE	QGNPKLNKLG	QPRFWKRPIS	DIVNQLRALF	ENQNPyQLGS	SLIOGTYWDV	120
AENLASWYAL	MKEYLAGTAT	WGEPSYAL	PLTEINQWMP	LTFSSGKVR	LLKNASGRYF	180
IGLPILGENN	PCYRMRTIEK	LIPCDGKGRV	TSGSLLPLP	VGIYAQQHRR	MTDICESIRT	240
EKGKLAWAQV	SIDYREVVDK	RRRMRTRKS	QGWIQGPWQE	VFILRLVLAH	KAPKLYKPRC	300
FAGISLGPKT	LASCVILDQD	ERVVEKQWNS	GSELLSLIHQ	GEERLRSLRE	QSKPTWNAAAY	360
RQLQSLINT	QVFTIVTFLR	ERGAARVLES	IARVRKSTPA	PPVNFLLSHW	AYRQITERLK	420
DLAIRNGMPL	THSNGSYGVR	FTCSQCGATN	QGIKDPTKYG	VDIESETFLC	SICSHREIAA	480
VNTATATNLAKQ	LLDE					494

SEQ ID NO: 333    moltype = AA length = 526

FEATURE    Location/Qualifiers

source    1..526

mol\_type = protein

organism = unidentified

SEQUENCE: 333

MNDTETSETL	TSVRTVCAHL	HVVGETGSLP	RLVEAAALAEI	ITLNGRATQA	LLSLAKNGLV	60
LRRDKEENLI	AAEATLPCRK	NKYADVAAKA	GEPILATRIN	NKGKLVTKKW	YGEGNNSYHIV	120
RFTPETGMFT	VRFVFDYAFF	EELLHLHSEV	VFGSDLPKGI	KAKTDSLPA	FLQAVFTSFL	180
ELPFQGFDPDI	VVKPAMKQAA	EQLLSYVQLE	AGENQQAEYP	DTNERDPELR	LVEWQKSLHE	240
LSVRTEPEF	VRARDIDYYA	ETDRRGNRFV	NITPEWTKFA	ESPFARRLPL	KIPPEFCILL	300
RRKTEGHAKI	PNIYIYGLQI	FDGVTDPSTL	GVLATAEDGK	LFWWHDHLDE	FSNLEGKPEP	360
KLKNKPQLLM	VSLEYDREQR	FEESVGGDRK	ICLVTLKETR	NFRRGNGR	LGIHFQHNPV	420
ITWALMDHDA	EVLEKGFIGE	NAFLKGALDK	QALNEYLQKG	GKWVGDRSFG	NKLKGITHL	480
ASLIVRLARE	KDAWIALEEI	SWVQKQSADS	VANHEIVEQP	HHSLTR		526

SEQ ID NO: 334    moltype = AA length = 649

FEATURE    Location/Qualifiers

source    1..649

mol\_type = protein

organism = unidentified

SEQUENCE: 334

MNDTETSETL	TSVRTVCAHL	HVVGETGSLP	RLVEAAALAEI	ITLNGRATQA	LLSLAKNGLV	60
LRRDKEENLI	AAEATLPCRK	NKYADVAAKA	GEPILATRIN	NKGKLVTKKW	YGEGNNSYHIV	120
RFTPETGMFT	VRFVFDYAFF	EELLHLHSEV	VFGSDLPKGI	KAKTDSLPA	FLQAVFTSFL	180
ELPFQGFDPDI	VVKPAMKQAA	EQLLSYVQLE	AGENQQAEYP	DTNERDPELR	LVEWQKSLHE	240
LSVRTEPEF	VRARDIDYYA	ETDRRGNRFV	NITPEWTKFA	ESPFARRLPL	KIPPEFCILL	300
RRKTEGHAKI	PNIYIYGLQI	FDGVTDPSTL	GVLATAEDGK	LFWWHDHLDE	FSNLEGKPEP	360
KLKNKPQLLM	VSLEYDREQR	FEESVGGDRK	ICLVTLKETR	NFRRGNGR	LGIHFQHNPV	420
LWRADFATSA	EVAAPKWNGR	ILGIHFQHNP	VITWALMDHD	AEVLEKGFIGE	GNAFLGKALD	480

-continued

---

KQALNEYLQK GGKWVGDRSF GNKLKGITHT LASLIVRLAR EKDIAWIALEE ISWVQKQSAD 540
SVANRRFSMW NYSRLATLIE WLGTIDATRD CGTAAPLAHK VSDYLTHFTC PECGACRKAG 600
QKKEIADTVR AGDILTCRKC GFSGPIPNDN IAEFVAKKAL ERMLKKPKV 649
SEQ ID NO: 335 moltype = AA length = 414
FEATURE Location/Qualifiers
source 1..414
mol_type = protein
organism = unidentified
SEQUENCE: 335
MAKRNFGEKS EALYRAVRFE VRPSKEELSI LLAVSEVLRM LFNSALAERQ QVFTEFIASL 60
YAEELKSASVP EEISEIRKKL REAYKEHSIS LFDQINALTA RRVEDEAFAS VTRNWQEETL 120
DALDGGAYKSF LSLRRKGDYD AHSPRSRDSG FFQKIPGRSG FKIGEGRIAL SCGAGRKLSF 180
PIPDYQQGRL AETTKLKKFE LYRDPQPNLAK SGRFWISVYY ELPKPEATT C QSEQVAFVAL 240
GASSIGVVSQ RGEEVIALWR SDKHWPKIE AVEERMKRRV KGSRGWLRLL NSGKRRMHMI 300
SSRQHVQDER EIVDYLVRNH GSHFVVTELV VRSGEGLAD SSKPERGSSL GLNWAAQNTG 360
SLSRLVRLQLE EKVKEHGGSV RKHKLTLTEA PPARGAENKL WMARKLRESF LKEV 414
SEQ ID NO: 336 moltype = AA length = 413
FEATURE Location/Qualifiers
source 1..413
mol_type = protein
organism = unidentified
SEQUENCE: 336
LAKNDEKELL YQSVKFEIYP DESKIRVLTR VSNILVLVWN SALGERRARF ELYIAPLYEE 60
LKKFPPRKSAA SNALRQKIRE GYKEHPIPTF DQLKKLLTTP RKEDPALLGS VPRAYQETL 120
NTLNNGSFVVF MTLRRNNMDM AKPPKGRAED RFHEISGRSG FKIDGSEFVL STKEQKLRFP 180
IPPNYQLEKLK EAKQIKKFTL YQSDRFRWI SIAYEIELPD QRPFNPVEVI YIAFGASSIG 240
VISPEGEVKI DFWRPDKHWK PKIKEVENRM RSCKKGSRRAW KKRAAARRKM YAMTQRQQL 300
NHREIVASLL RLGFHFVVTY YTVRSKPGKL ADGSNPKRGG APQGFNWSAQ NTGSFGEFIL 360
WLQKVKVEQG GTVQTFRLVL GQSERPEKRG RDNIKIEMVRL LREKYLESTQ IVV 413
SEQ ID NO: 337 moltype = AA length = 449
FEATURE Location/Qualifiers
source 1..449
mol_type = protein
organism = unidentified
SEQUENCE: 337
MAKGKKKEKG PLYRAVRFEI FPTSDQITLF LRVSKNLQQV WNEAWQERQS CYEOFQGSYI 60
ERIGQAKKRA QEAGFSEVWE NEAKKGLNNK LRQQEISMQL VSEKESLLQE LSIAFQEHSV 120
TLYDQINGLT ARRIIGEFAL IPRNWQETL DSDLGSFKSF LALRKNGDPD AKPPRQRVSE 180
NSFYKIPGRS GFVKVSNQIY LSFGKIGQTL TSVIPEFQLK RLETAIKLKK FELCRDERDM 240
AKPGRFWISV AYEIPKPEK PVVSKQITYL AIGASRLGVV SPKGEFCNLN PRSDYHWKPQ 300
INALQERLEG VVKGSRKWKK RMAACTRMRFA KLGHQQKQHG QEYEVKKLLR HGVHFVVTEL 360
KVRSKPGALA DASKDRKGS PTGPNWSAQN TGNIARLIQK LTDKASEHGG TVIKRNPLL 420
SLEERQLPDA QRKIFIAKKL REEFLADQK 449
SEQ ID NO: 338 moltype = AA length = 711
FEATURE Location/Qualifiers
source 1..711
mol_type = protein
organism = unidentified
SEQUENCE: 338
MAKREKKDDV VLRGTMRIY PTDRQVTLMD MWRRRCISLW NLLLNLTTAA YGAKNTRSKL 60
GWRSIARVEEN HAKALIV YQHKGCKKDQ SFVLKRDCTV KHPFRERFPG DRKILLGLFD 120
ALRHTLDKGA KCKCNVNQPY ALTRAWLDET GHGARTADII AWLKDFKGEK DCTAISTAAK 180
YCPAPPTAEL LTKIKRAAPA DDLPVWDQAIL LDLFGALRGG LKQKECDHHT ARTVAYFEKH 240
ELAGRAEIDL AWLIAHGGTC DCKIVEEAAN HCPGPRFLIW EHESLAMIMAR LKAEPRTWEI 300
GDLPSHAAQT VVKDLVVKALQ TMLKERAKAA AGDESARKTG FPKFKKQAYA AGSVYFPNTT 360
MFFDVAAQGRV QLPNGCKGSMR CEIPRQLVLL ERNRLKPKLG VIGAQLGLG GRIWRQGDRW 420
YILSCQWERPQ PTLLPKTGRV AGVKIAASIV FTYYDNRQGT KEYPMPPPADK KLTAVHLVAG 480
KQNSRALEAQ KEKEKKLKR KERLRLGKLE KGHDPNALKP LKRPRVRRSK LFYKSAARLA 540
ACEAIERDRR DGFLHRVTNE IVHKEPDAVSV QKMSVAPMMR RQKQKEKQIE SKKNREAKED 600
NGAAKKPRNL KPVRKLLRHV AMARGRQFLE YKYNDLRGPG SVLIADRLEP EVQECSRCGT 660
KNPQMKDGR LLRCIGVLPD GTDCDAVLPR NRNAARNAEK RLRKHREAHN A 711
SEQ ID NO: 339 moltype = AA length = 574
FEATURE Location/Qualifiers
source 1..574
mol_type = protein
organism = unidentified
SEQUENCE: 339
MNEVLPPIAV GEDAADTIMR GSKMRIYPSV RQAATMDLWR RRCIQLWNLL LELEQAAAYS 60
ENRRTQIGWR SIWATVVEDS HAEAVRVAE GKKRKDGTFR KAPSGKEIPP LDPAMLAKIQ 120
RQMNGAVDVD PKTGEVTPAQ PRLFMWEHEL QKIMARLKQA PRTHWIDDLP SHAQSVVKD 180
LIKALQAMLR ERKKRASGIG GRDTGFPKFK KNRYAAGSVY FANTQLRFEA KRGKAGDPDA 240

-continued

---

VRGEFARVKL	PNGVGWMECR	MPRHINAABA	YAQATLMGGR	IWRQGENWYL	SCQNKMPPKA	300
PLPRAGRATAA	IKIAAAIPIT	TVDNRGQTR	YAMPPIDRER	IAAHAAAGRA	QSRALEARKR	360
RAKKREAYAK	KRHAKKLERG	IAAKPPGRAR	IKLSPGFYAA	AAKLAKEAE	DANAREAWLH	420
ELTTQIVRN	DVIAVPRMVE	AKLMKKPPEPP	EEKEEQQVKAP	WQGKRRLSLKA	ARVMMRRTAM	480
ALIQTTLKYK	AVDLRGQPQAY	EEIAPLDVTA	AACSGCGVLK	PEWKMARAKG	REIMRCQEPL	540
PGGKTCNTVL	TYTRNSARVI	GRELAVALAE	RQKA			574

SEQ ID NO: 340	moltype = AA	length = 400				
FEATURE	Location/Qualifiers					
source	1..400					
	mol_type = protein					
	organism = unidentified					
SEQUENCE: 340						
MTTQKTYNFC	FYDQRFFELS	KEAGEVYSRS	LEEFWKIYDE	TGVWLSKFDL	QKHMNRNLER	60
KLLHSSFLG	AMQQVHANLA	SWKQAKKVP	DACPFRPKPF	LQAILFKKSQ	IKYKNGFLRL	120
TLGTEKEFLY	LKWDINIPLP	IYGSVTYSKT	RGWKINLCLE	TEVEQKNLSE	NKYLISIDLGV	180
KRVATIFDGE	NTITLSGKKF	MGLMHYRNKL	NGKTQSRLSH	KKGSNNYKK	IQRAKRKTTD	240
RILLNIQKEML	HKYSSFIVNY	AIRNDIGNII	IGDNSSTHDS	PNMRGKTNQK	ISQNPFEQKLK	300
NYIKYKFESI	SGRVDIVPEP	YTSRKCPHCK	NIKKSSPKGR	TYKCKKGFI	FDRDGVGAIN	360
IYNENVSFGQ	IIISPGRIRSL	TEPIGMKFHN	EIYFKSYVA			400

SEQ ID NO: 341	moltype = AA	length = 743				
FEATURE	Location/Qualifiers					
source	1..743					
	mol_type = protein					
	organism = unidentified					
SEQUENCE: 341						
MSVRSFQARV	ECDKQTMHEHL	WRTHKVFNER	LPEIIKILFK	MKRGECCQND	KQKSLYKSI	60
QSILEANAQN	ADYLLNSVSI	KGWKPGTAKK	YRNASFTWAD	DAAKLSSQGI	HVYDKKQVLG	120
DILGMMMSQM	CRQSVEAISG	HIELTKKWEK	EHNEWLKEKE	KWESEDEHKK	YLDLREKFEQ	180
FEQSISGGKIT	KRRGRWHLL	KWLSNDNPFA	AWRGNKAVIN	PLSEKAQIRI	NKAKPNKNS	240
VERDEFKKP	PEMKALDNLH	GYYERNFVRR	RKTKKNPDPF	DHKPTFTLPH	PTIHPRFWFV	300
NKPKTNPEGY	RKLILPKKAG	DLGSLLEMRL	TGEKNKNGNP	DDWISVKFKA	DPRLSLIRPV	360
KGRRVVORKGK	EQQGQTETDS	YEFPDKHLLK	WRPAKLSGVK	LIFPDKTPKA	AYLYFTCDIP	420
DEPLTETATT	IQWLETGDVT	KKGKRRKKV	LPHGLVSCAV	DLSMRGRTTG	FATLCRYENG	480
KIHILRSRNL	WVGYKEGKGC	HPYRWTGPD	LGHIAKHRE	IRILRSKRGK	PVKGEESHID	540
LQKHIDYMG	DRFKKAARTI	VNFALNTENA	ASKNGFYPR	DVLLLENLEG	LIPDAEKERG	600
INRALAGWNR	RHLVERVIEM	AKDAGFKRRV	FEIIPPYGTQS	VCSKCGALGR	RYSIIRENNR	660
REIRFGYVEK	LFACPNCGYC	ANADHNASVN	LNRRFLIEDS	FKSYDDWKRL	SEKKQKEEIE	720
TIESKLMDKL	CAMHKISRG	ISK				743

SEQ ID NO: 342	moltype = AA	length = 769				
FEATURE	Location/Qualifiers					
source	1..769					
	mol_type = protein					
	organism = unidentified					
SEQUENCE: 342						
MHLWRTHCVC	NQRLPALLKR	LFAMRRGEVG	GNEAQQRQVYQ	RVAQFVLARD	AKDSVDLLNA	60
VSLRKRSANS	AFKKKATISC	NGQAREVTGE	EVPFAEAVALA	SKGVFAYDKD	DMRAGLPDSL	120
FQPLTRDAV	CMRSHEELVA	TWKKEYREWR	DRKSEWEAEP	EHALYLNLRP	KFEEGEAARG	180
GRFRKRAERD	HAYLWDLEAN	PQLAAMRRAKA	PPAVVPIDEA	GKRRRIARAKA	WKQASVRAEE	240
FWKRKNPELHA	LHKIHVQYLR	EFVRPDRTRR	NKRREGFKQR	PTFTMPDPVR	HPRWCLFNAP	300
QTSPOQYRLL	RLPQSRRTVG	SVELRLLTGP	SDGAGFPDAW	VNVRFKADPR	LAQLRPVKV	360
RTVTRGKNG	AKVEADGFRY	YDDQLLIERD	AQVSGVKLLF	RDIRMAPFAD	KPIEDRLLSA	420
TPYLVFAVEI	KDEARTERAI	AIRFDSETSEL	TKSGKKRKT	PAGLVSVAVD	LDTRGVGF	480
RAVIGVPEIQ	QTHHGVRLLQ	SRVAVVGQE	ARASGEAEWS	PGPDLAHIAR	HKREIRRLLQ	540
LRGKPVKGER	SHVRLQAHID	RMGEDRFKKA	ARKIVNEALR	GSNPAAAGDPY	TRADVLLYES	600
LETLLPDAER	ERGINRALLR	WNRAKLIEHL	KRM CDDLAGIR	HFPVSPFGTS	QVCSKCGALG	660
RRYSLARENG	RAVIRFGWVE	RLFACPNPEC	PGRRPDRPDR	PFTCNSDHNA	SVNLHRVFA	720
GDQAVAAFA	LAPRDS	LAVKRVEDTL	RPQLMRVHKL	ADAGVDSPP		769

SEQ ID NO: 343	moltype = AA	length = 666
FEATURE	Location/Qualifiers	
source	1..666	
	mol_type = protein	
	organism = unidentified	
SEQUENCE: 343		

MATLVYRYGV	RAHGSARQDQ	AVVSDPAMLE	QLRLGHELRN	ALVGVQHRYE	DGKRAWWSGF	60
ASVAAADHRV	TTGETAVAEI	EKQARAEGSA	DRTAATRQGT	AESLKAARRAA	VKQARADRKA	120
AMAAVAEQAK	PKIQALGDDR	DAEIKDLYRR	FCQDGVLPLR	CGRCAGDLRS	DGDCTDCGAA	180
HEPRKLYWAT	YNAIREDHQ	AVKLVEAKRK	AGOPARLFR	RWTGDTLTV	QLQRMHGPAC	240
RCVTCAEKL	RRARKTDPQA	PAVAADPAPY	PTDPPRDPAL	LASGQGKWRN	VLQLGTWIPP	300
GEWSAMSRAE	RRRVGRSHIG	WQLGGGRQLT	LPVOLHQRMP	ADADVMAQL	TRVRVGRHR	360
MSVALTAKLP	DPPQVQLPP	VALHLGWRQR	PDGSLRVATW	ACPQPLDPP	AVADVVVSHG	420
GRWGEVIMPA	RWLADAEPV	RLLGRDRKAM	EPVLEALADW	LEAHTEACTA	RMTPALVRRW	480
RSQGRLAGLT	NRWRGQPPTG	SAEILTYLEA	WRIQDKLLWE	RESHLRRRLA	ARRDDAWRRV	540

-continued

---

ASWLARHAGV LVVDDADIAE LRRRDPADT DPTMPASAAQ AARARAALAA PGRLRHATI	600
TATRDGLGVH TVASAGLTRL HRKCGHQACP DPRYAASAVV TCPGCGNGYD QDYNAAMML	660
DRQQQP	666
 SEQ ID NO: 344 moltype = AA length = 564	
FEATURE Location/Qualifiers	
source 1..564	
mol_type = protein	
organism = unidentified	
 SEQUENCE: 344	
MSRVELHRAY KFRLYPTPAQ VAELEAEWERQ LRRLYNLNLAHS QRLAAMQRHV RPKSPGVLK 60	
ECLSCGAVAV AEIGTDGKA KTVKHAVGC VLECRSCGGS PDAEGRTAHT AACSFVDYYP 120	
QGREMTQLLE EDDQLARVVC SARQETLRDL EKAWQRWHKM PGFGKPHFKK RIDSCRIFYFS 180	
TPKSWAVDGL YLSFTGVASS VGRKIRQDR VWPGDAKPSS CHVVRDVDEW YAVFPLTFTK 240	
EIEPKKGAV GINRGAHVAI ADSTGRVVD S PKFYARSLGV IRHRARLLDR KVPFGRAVKP 300	
SPTKYHGLPK ADIDAAAARV NASPGRLVYE ARARGSTAAEA HAHALALVLP APROTSQLPS 360	
EGRNRERARR FLALAHQRVR RQREWFLHNE SAHYAQSYTK IAIEDWSTKE MTSSEPRDAE 420	
EMKRVTRARN RSILDVGWYE LGRQIAYKSE ATGAEFAKVD PGLRETETHV PEAIVRERDV 480	
DVSGMLRGEA GISGTCSRCG GLLRASASGH ADAECEVCLH VEVDVNAAAV NVLKRAMFP 540	
AAPPSKEKAK VTIGIKGRKK KRAA 564	
 SEQ ID NO: 345 moltype = AA length = 565	
FEATURE Location/Qualifiers	
source 1..565	
mol_type = protein	
organism = unidentified	
 SEQUENCE: 345	
MSRVELHRAY KFRLYPTPVQ VAEELSEWERQ LRRLYNLHGHE QRLLTTLTRHL RPKSPGVLK 60	
ECLSCDSTQV QEVGADGRPK TTVRHAEQCP TLACRSCGAL RDAEGRTAHT VACAFVDYYP 120	
QGREMTLLA ADDQLARVVC SARQEVLRDL DKAWQRWRKM PGFGKPRFKR RTDSCRIFYFS 180	
TPKAWKLEGG HLSFTGAAATT VGAIKMRQDR NWPASVQFSS CHVVRDVDEW YAVFPLTFVA 240	
EVARPKGAV GINRGAHVAI ADSTGRVVD S PRYARALGV IRHRARLFDR KVPSPGHAVKP 300	
SPTKYRGLSA IEVDRVARAT GFTPGRVVT ALNRGGVAYA ECALAAIAVL GHGPERPLTS 360	
DGRNRERARK FLALAHQRVR RQREWFLHNE SAHYARTYSK IAIEDWSTKE MTASEPQGEE 420	
TRRVTRSRRN RSILDVGWYE GRQLAYKTEA TGAEFAQVDP GLKETETNVP KAIADARDVD 480	
VSGMLRGEAG ISGTCSKCGG LLRAPASGHA DAECEICLNV EVGDVNAAVN VLKRAMFP 540	
APPASGEKPK VSIGIKGRQK KKAA 565	
 SEQ ID NO: 346 moltype = AA length = 499	
FEATURE Location/Qualifiers	
source 1..499	
mol_type = protein	
organism = unidentified	
 SEQUENCE: 346	
MEAIATGMSP ERRVELGILP GSVELKRAYK FRLYPMKVQQ AELSEWERQL RRLYNLAHEQ 60	
RLAALLRYRD WDFQKGACPS CRVAVPGVHT AACDHVDYFR QAREMTQLLE VDAQLSRVIC 120	
CARQEVLRLD DKAWQRWRKRR LGGRPRFKRR TDSCRIVLST PKHWEITAGRY LRLSGLASSV 180	
GEIRIEQDRA PEGEALLSSC SIVRDVDEWY ACLPLTFTOP IERAPHRSVG LNRCGVVHALA 240	
DSDGRVVDSP KFFERALATV QKRSRDLARK VSGSRNAHKA RIKLAKAHQR VRRQRAAFLH 300	
QESAYYSKGF DLVALEDMVS RKMTATAGEA PEMGRGAQDR LNRGILDVGW YELARQIDYK 360	
RLAHGGELLR VDPGQTTPLA CVTEEQPARQ ISSACAVCGI PLARPASGNA RMRCTACGSS 420	
QVGDVNAAEN VLTRALSSAP SGPKSPKASI KIKGRQKRLG TPANRAGEAS GGDPPVRGPV 480	
EGGTLAYVVE PVSEQSQDT 499	
 SEQ ID NO: 347 moltype = AA length = 560	
FEATURE Location/Qualifiers	
source 1..560	
mol_type = protein	
organism = unidentified	
 SEQUENCE: 347	
MVTRTYKYRA YPTPEQAEAL TSWLRFASQL YNAALEHRKN AWGRHDAHGR GFRFWDGAA 60	
PRKKSDPPGR WVYRGGGGAH ISKNDQGKLL TEFRREHAEL LPPGMPALVQ HEVLARLERS 120	
MAAFFQORATK GOKAGYPRWR SEHRYDSLTF GLTSPSKERF DPETGESLGR GKTVGAGTYH 180	
NGDRLRLTGLG ELRILEHRRJ PMGAIPKSVI VRSGKRWFW SIAMEMPSV PAASGRPAVG 240	
LDMGVVTWGT AFTADTSAAA ALVALDRLRRM TDPSDCRRLR ELEREEAAQLS EVLAHCRARG 300	
LDPARPRRCP KELTKLYRRS LHRIGELDRA CARIRRLQIA AHDIAPVVD EAGSAVLIEG 360	
SNAGMRHARR VARTQRVAR RTRAGHAHSN RRKKAVQAYA RAKERERSAR GDHRHKVSRA 420	
LVRQFEEISV EALDIKQLTV APEHNPDQP DLPAHVQRRL NRGELDAAWG AFFAALDYKA 480	
ADAGGRVARK PAPHTTQEC A RCGTLPVKPI SLRVHRCPAC GYTAPRTVNS ARNVLQRPLE 540	
EPGRAGPSGA NGRGVPHAVA 560	
 SEQ ID NO: 348 moltype = AA length = 404	
FEATURE Location/Qualifiers	
source 1..404	
mol_type = protein	
organism = unidentified	

---

-continued

---

SEQUENCE: 348  
 MNCRYRYYI PTPGQRQSLA RLFGCVRVWW NDALFLCRQS EKLPKNSELQ KLCITQAKKT 60  
 EARGWLGVQS AIPLQQSVAD LGVAFKNFFQ SRSGKRKGKK VNPPRVKRRN NRQGARFTRG 120  
 GFKVKTSKVY LARIGDIKIK WSRPLPSEPS SVTIVIKDCAG QYFLSFVVEV KPEIKPPKPN 180  
 SIGIDGLGLKT FASCNSNGEKI DSPDYSRLYR KLKRCQRRRA KRQRGSKRRE RMRVKVAKLN 240  
 AQIRDKRKDF LHKLSTKVVN ENQVIALEDL NVGGMLKNQR LSRAISQAGW YEFRSLCEGK 300  
 AEKHNRDPRV ISRWEPTSQV CSECGYRWGK IDLSVRSSIVC INCVGHEHRS DNASVNIEQA 360  
 GLKVGVGHTH DSKRTGSACK TSNGAVCVP STHREYVQLT LFDW 404

SEQ ID NO: 349 moltype = AA length = 392  
 FEATURE Location/Qualifiers  
 source 1..392  
 mol\_type = protein  
 organism = unidentified

SEQUENCE: 349  
 MKSRWTFRCY PTPEQEBOHLA RTFGCVRVFW NWALRARTDA FRAGERIGYP ATDKALTLLK 60  
 QQPETVWLNE VSSVCLQQAL RDLQVAFSNF FDKRAAHPSF KRKEARQSAN YTERGFSFDH 120  
 ERRILKLAKI GAIKVKWSRKA AIPHPSSIRL IRTASGKYFV SLVVETQPAP MPETGESVGV 180  
 DFGVARLATL SNGERISNPK HGAKWQRRLA FYQKRLARAT KGSKRRMRIK RHVARIHEKI 240  
 GNSRSDTLHK LSTDLVTRFL LICVEDLNLR GMVKNHSLAR SLHDASIGSA IRMIEKAER 300  
 YGKNNVVKIDR WFPSSKTCSD CGHIVEQLPL NVREWTCPCE GTTHDRDANA AANILAVGQT 360  
 VSAGGTVRR SRAKASERKS QRSAQRGVN RA 392

SEQ ID NO: 350 moltype = AA length = 500  
 FEATURE Location/Qualifiers  
 source 1..500  
 mol\_type = protein  
 organism = unidentified

SEQUENCE: 350  
 KEPLNICKTA KAVFKEIDPT SLNRAANYDA SIELNCHECK FKPFKNVCRY EFNFYNNWYR 60  
 CNPNSQLQST YKAQVRKVEI GYEKLNEIL TQMQYYPWFG RLYQNFFFHDE RDKMTSLDEI 120  
 QVIGVQNQVF PNTVEKAWRE IIKKRFKDNN ETMTEPIELK HAAGHGKRKL SNKSSLRRRF 180  
 AFVQKSKFV DNSDVSYRSF SNNIACVLPS RIGVDLGGV SRNPKREYIP QEISFNAFWK 240  
 QHEGLKKGRN IEIQSVQYKG ETVKRIEADT GEDKAWGKMR QRRFTSLILK LVPKQGGKKV 300  
 WKYPEKRNEG NYEYPPPIE FILDSGETSI RFGGDEGEAG KQKHLVIPFN DSKATPLASQ 360  
 QTLLENSRFN AEVKSCIGLA IYANYFYGYA RNVISSIYH KNSKNGQAIT AIYLESIAHN 420  
 YVKAIEROLQ NLLLNLNRDFS FMESHKKELK KYFGGDLEGTT GGAQKRREKE EKIEKEIEQS 480  
 YLPLRLIRLSSL TKMVTQVEM 500

SEQ ID NO: 351 moltype = AA length = 507  
 FEATURE Location/Qualifiers  
 source 1..507  
 mol\_type = protein  
 organism = unidentified

SEQUENCE: 351  
 ELIVNENKDP LNIGKTAKAV FKEIDPTTSIN RAANYDASIE LACKECKFKP FNNTKRHDFTS 60  
 FYSNWHRCSP NSCLQSTYRA KIRKTEIGYE KLKNEILNMQ QYYPWFGRLY QNFFNDQRDK 120  
 MTSLDEIQVT GVQNKFIFNT VEKAWRREIIK KRPFRDNKETM RTIPDLKNKS GHGSRKLSNK 180  
 SLLRRRFAFA QKSFKLVDNS DVSYRAFSNN VACVLPSKIG VDIGGIINKD LKREYIPOEI 240  
 TFPNFWQKHD GLKKGRNIEI HSVQYKGEIV KRIEADTGED KAWGKRNQRR FTSLILKITP 300  
 KQGKKIWKF PEKKNSADYE YFPPIPIEF DNGDASTIKFG GEEGEVGKQK HLLIPFNDSK 360  
 ATPLSSKQML LETSRFNAEV KSTGIALYA NYFVSYARNY KVKSTYHKNs KKGQIVTEIY 420  
 LESISQNFVR AIQRQLQSLM LNLDWGFMQ THKKELKFYF GSDLEGSKGG QRREKEEKI 480  
 EKEIEASYLP RLIRSLTLKS VTKAEM 507

SEQ ID NO: 352 moltype = AA length = 529  
 FEATURE Location/Qualifiers  
 source 1..529  
 mol\_type = protein  
 organism = unidentified

SEQUENCE: 352  
 PEEKTSKLKP NSINLAANYD ANEKFNCKEC KFHPFKNKRR YEFNFYNNLH GCKSCTKSTN 60  
 NPAVKRIEIG YOKLKFEIKN QMEAYPWGR LRINFYSEDEK RKMSELNEMQ VTGVKNKIFF 120  
 DAEICAWREI LKKRFFRESK TLITIPKLKN KAGHGARKHR NKKLLIRRRA FMKKNFHFLD 180  
 NDSISYRSFA NNIACVLPSK VGDVGDIPKVR DISLNLMWAS KEGIKSGRKV 240  
 EYISTQYDGN MVKKIBAETG EDKSWGKRNK RRQTSLLLSI PKPSKQVQEF DFKEWPRYKD 300  
 IEKKVQWGRF PIKIIIFDSNH NSIEFGTYQG GKQKVLPIPF NDSKTTPLGS KMNKLEKLRF 360  
 NSKIKSRGLS AIAANKFLEA ARTYCVDSLW HEVSSANAIG KGKIFIEYYL EILSQNYIEA 420  
 AQQLQRFIE SIEQWFVADP FGQRLKQYFK DDLKRAKCFL CANREVQTC YAAVKLHKSC 480  
 AEKVKDKNKE LAIKERNNKE DAVIKEVEAS NYPRVIRLKL TKTITNPK 529

SEQ ID NO: 353 moltype = AA length = 726  
 FEATURE Location/Qualifiers  
 source 1..726  
 mol\_type = protein  
 organism = unidentified

---

-continued

---

SEQUENCE: 353  
 SESENKLIQ YAFLYSFRD KYEKPEFKNR GDIKRKLQNK WEDFLKEQNL KNDKKLSNYI 60  
 FSNRNFRRSY DREENEEGI DEKKSKPKRI NCFEKEKNLK DQYDKDAINA SANKDGAQKW 120  
 GCFECIFPPM YKIESGDPNK RIIINKTRFK LFDYLNLLKG CKSCLRSTYH PYRSNVYIES 180  
 NYDKLKREIG NFLQQKNIFQ RMRKAKVSEG KYLTLNDEYR LSCVAMHFKN RWLFFDSIQK 240  
 VIRETICKQLR KQMRESYDEQ AKTRSKKGHG RAKYEDQVRM IRRRAYSQAQ HKLLDNGYIT 300  
 LFDFDDKEIN KVCLTAINQE GFDIGGYLNS DIDNVMPPIE ISFHLWKYN EPILNIESPF 360  
 SKAKISDYL R KIREDLNLER GKEGKARSKK NVRRKVLASK GEDGYKKIFT DFFSKWKEEL 420  
 EGNAMERVLs QSSGDIQWSK KKRIHYTTLV LNINLLDKKG VGNLKYYEIA EKTKILSFDK 480  
 NEENKFNPITI QVLLDGYEIG TEYDEIKOLN EKTSKQFTIY DPNTKIIKIP FTDSKAVPLG 540  
 MILGINIPATL TVKKTERDIK VSKIFKGGLN SKIVSKIGK TYAGYFPPTVD KEILEEEVED 600  
 TLDNEFSSKS QRNIFLKSII KNYDKMLKBQ LFDFYSFLVR NDLGVRFLLTD RELQNIEDES 660  
 FNLEKRFFFET DRDRIARWFD NTNTDDGKEK FKKLANEIVD SYKPRLIRLP VVRVIKRIQP 720  
 VKQREM 726

SEQ ID NO: 354 moltype = AA length = 517  
 FEATURE Location/Qualifiers  
 source 1..517  
 mol\_type = protein  
 organism = unidentified

SEQUENCE: 354  
 KYSTRDFSEL NEIQTACKQ DEFFKVIQNA WREIIKKRFL ENRENPIEKK IFKNKGRGK 60  
 RQESDKTIQR NRASVMKNFQ LIENEKQIPL APSGHVACVF PVKVGLDIGG FKTDDLEKNI 120  
 FPPRTITINV FWKVNDRQRK GRKLEVWGIK ARTKLIKEVHKW KWDKLEEVKK KRLKSLEQKQ 180  
 EKSLDNWSEV NNDSFYKVQI DELQEKEIKS LKGRTMNKIL DNKAKESEA EGLYIEWEKD 240  
 FEGERMLRRIE ASTGGEKEKG KRRQRHSTS LLDIKNNNSRG SKEIINFYSY AKQGKKEKKI 300  
 EFFFPPLTIT LDAAEESPLN IKSIPIEDLN ATSKYFSIPF TETRATPLSI LGDRVQKFKT 360  
 KNISGAIKRN LGSSISSECKI VQNAETSAKS ILSLPNVKED NNMEIFINTM SKNYFRAMMK 420  
 QMESFIFEME PKTLIDPYKE KAIKWFEVA SSRAKRKLKK LSKADIKKSE LLLSNTEEEF 480  
 KEKQEKLEAL EKEIEEFYLP RIVRLQLTKT ILETPVM 517

SEQ ID NO: 355 moltype = AA length = 481  
 FEATURE Location/Qualifiers  
 source 1..481  
 mol\_type = protein  
 organism = unidentified

SEQUENCE: 355  
 KKLQLLGHKI LLKEYDPNAV NAAANFETST AELCGQCKMK PFKNKRRFQY TFGKNYHGCL 60  
 SCIQNVYAK KRIVQIAKEE LKHQTLDSIA SIPYKYTSLQ SNTNSIDEYL ILKOERAAFF 120  
 SNTNSIDEYL ITGNIANIAF KVISAWEDEI IKRRQRVYE SLSTDGTVKA NRGHGGTAYK 180  
 SNTRQEKRIRA LQKQTLHMTV NPYISLARYK NNYIVATLPR TIGMHGAIK DRDPQKKLSD 240  
 YAINFNIEWS DDRQLIELST VQYTGDMVRK IEAETGENNK WGENMRTKKT SLLLEILTKK 300  
 TTDELTFKDWA AFSTKKEIDS VTKKTYQGFP IGIIFEGNES SVKFGSQNYF PLPFDAKITP 360  
 PTAEGFRLDW LRKGFSSSQM KTSYGLAIYS NKVTNAIPAY VIKNMFYKIA RAENGKQIKA 420  
 KFLKKYDIA GNNYVPFIIM QHYRVLDTFE EMPISQPKV RLSLTKTQHI IIKKDKTDISK 480  
 M 481

SEQ ID NO: 356 moltype = AA length = 534  
 FEATURE Location/Qualifiers  
 source 1..534  
 mol\_type = protein  
 organism = unidentified

SEQUENCE: 356  
 TTSNLINLGK KAINISANYD ANLEVGCKNC KFLSSNGNFP RQTNVKEGCH SCEKSTYEPS 60  
 IYLVKIGERK AKYDVLDLSK KFTFQSLVKYQ SKKSMKSRNK KPKELEKEFVI FANKNKAFFD 120  
 IQKSYNHLIL QIKREKINRMN SKKRKNHHR RLFRDRKEQL NKLRLIESNN LFLPRENKG 180  
 NHVFTYVAIH SVGRDIGVIG SYDEKLNFET ELTYQLYPNF DKRLLYAYKP KQNKIIKIKE 240  
 KLWNLRKEKE PLDLEYEKPL NKSITFSIKN DNLFKVSKDL MLRRAKFNIQ GKEKLSKEER 300  
 KINRDLIKIK GLVNSMSYGR FDELKKEKNI WSPhiYREVR QKEIKPCLIK NGDRIEIFEQ 360  
 LKKKMERLRR FREKRKKIS KDLIFAEIRIA YNFHTKSIKN TSNKINIDQE AKRGKASYMR 420  
 KRIGYETFKN KYCEQCLSKG VYRNVQKGC SCFENPFDWI KKGDENLILPK KNEDLRVKGA 480  
 FRIGEALEKQI VKIAFNIAKG YEDFYDNLGE STEKDLKLKF KVGTINEQE SLKL 534

SEQ ID NO: 357 moltype = AA length = 537  
 FEATURE Location/Qualifiers  
 source 1..537  
 mol\_type = protein  
 organism = unidentified

SEQUENCE: 357  
 TSNPIKLGKK AINISANYDS NLQIGCKNC FLSYNGGNFP QTNVKEGCHS CEKSTYEPVV 60  
 YTWRIGERRS KYDVLDLSK K FIFLSQLKYQ SKKMKTTSKG IRGLEEFVIS ANLKKAMDVI 120  
 QKSYRHLILN IKNIEIVRMNG KKRKNHHR LFRDRKEQLN KLRLIEGSSF FKPPTVKGDN 180  
 SIFTCAVIAHN IGRDIGIAGD YFDKLEPLKIE LTYQLYYEVN PKKESEINKR LLYAYPKQ 240  
 KIEIKEKLW NLRKEKSPLD LEYEKPLTKS ITFLVKRDGV FRISKDLMLR KAKFIIQGKE 300  
 KLSKEERKIN RDLIKIKSNI ISLYTGRFDE LKKDKTIWSP HIFRDVKQGK ITPCIERKGD 360  
 RMDIFQQLRK KSERLRENRK KRQKKISKDL IFAERIAYNF HTKSICKNTSN LINIKHEAKR 420

-continued

---

GKASYSMRKRI GNETFRIKYC EQCFFPKNNVY KNVQKGCSF EDPFEYIKKG NEDLIPNKNQ	480
DLAKAGAFRD DALEQIIKV AFNIAKGYED FYENLKKTE KDIRLKFKVG TIISEEM	537

```
SEQ ID NO: 358      moltype = AA length = 541
FEATURE          Location/Qualifiers
source           1..541
                  mol_type = protein
                  organism = unidentified
```

```
SEQUENCE: 358
NSINLSSKKA INISANYDAN LQVRCKNCKF LSSNGNFPQR TDVKEGCHSC EKSTYEPPVY 60
DVKIGEIKAK YEVLDSLKKF TFQSLKYQLS KSMKFRSKKI KELKEFVIFA KESKALNVIN 120
RSYKHLILNI KNDINRMNSK KRIKNHKGRJ FLDRKQQLSK LKLIEGSSFF VPAKVNKNK 180
VFTCVAIHSI GRDIGIAGLY DSFTKPVNEI TYQIFFSGER RLLYAYPKQ LKILSIKENL 240
WSLKNEKKPL DLLYEKPLGE NLNPNVKGGD LFRRVSCKLMI RNAKFNVHGR QRLSDEERLI 300
NRNFPIKIGE VSVLSYGRFE ELKKDKLWS PHIFKDVRQN KIKPCLVMQG QRIDIFEQLK 360
RKLLELLKKR KSRQKKLSD LIFGERIAYN FHTKSIKNTS NKİNİDSDAK RGRASYSMRK 420
IGNETFKLKY CDVCFFPKANV YRRVQNGCSC SENPYNYIKK GDKDLPKKD EGLAIKGAFR 480
DEKLNKQIIK VAFNIAKGYE DFYDDLKKRT EKDVDLKPKI GTTVLDQKPM EIFDGIVITW 540
L                                         541
```

```
SEQ ID NO: 359      moltype = AA length = 542
FEATURE          Location/Qualifiers
source           1..542
                  mol_type = protein
                  organism = unidentified
```

```
SEQUENCE: 359
LTITVVETNN LAKKAINVAA NF DANIDRQY YRCTPNLCRF IAQSPRETKE KDAGCSSCTQ 60
STYDPKVVI KIGKLLAKYE ILKSLKRFLF MNRYFKQKKT ERAQQKQKIG TELNEMSIFA 120
KATNAMEVIK RATKHCTYDI IPETKSLQML KRRRHRVKVR SLLKILKERR MKIKKIPNTF 180
IEIPKQAKKN KSDYYVAAL KSCGIDVGLC GAYEKNAEVE AEYTYQLYYE YKGNNSTKRI 240
LYCYNNPQKN IREFKWAEFYI QGSKSHVNTP GTIRLKMEKF LSPTIESSEA LDFRVWNSDL 300
KIRNGQYCFI KKRSLGKEAR RKTGNLTYKG SPSELEKSIHV YTTERENPKK 360
PRAARKKEDN FMEIFEMQRK KDYEVNKRR KEATDAKIM DFAEEPIRHY HTNNLKAVRR 420
IDMNEQVERK KTSVFLKRM QNGYRGNYCR KC1KAPEGSN RDENVLEKNE GCLDCIGSEF 480
IWKKSSKEKK GLWHTNRLLR RIRLQCFTTA KAYENFYNDL FEKKESSLDI IKLKVSITTK 540
SM                                         542
```

```
SEQ ID NO: 360      moltype = AA length = 564
FEATURE          Location/Qualifiers
source           1..564
                  mol_type = protein
                  organism = unidentified
```

```
SEQUENCE: 360
ASTMNLAKQA INFAMYDSN LEIGCKGCKF MSTWSKKSNP KFYPRQNNQA NKCHSCTYST 60
GEPEVPPIEI GERAAKYKIF TALKKFVFMV VAYKERRRQF FKSKKPKELK ELAICSNREK 120
AMEVIQKSVV HCYGDVQKQE PRIRKIVLK NHKGRLFYKQ KR SKIKI LAKL EGGSFFKTFI 180
PKVHNNGCHS CHEASLNKTI LVTTALNTIG ADIGLINDTS TIA PTTEDIS WQVYYEFIPN 240
GDSEAVKRL IKSIRDKYFK KGHENAVNTG FFKYQGKIVK GPIKFVNNE 300
DFARKPDLKs MKIKRAGFAI PSAKRLSKED REINRESIKI KNKIYSLSYG RKKTLSDKDI 360
IKHLYRPVRQ KGKVPLEYRK APDGFLFEEY SLKRKERRLR KQKEKRQKDM SEIIDAADEF 420
AWHRHRTGSIK KTTNHINFKS EVKRGKVPIM KKR IANDSFN TRHCGKCVKQ GNAINKYIY 480
KQKNCFCDCNS EEFKWEKAAL EKKGAKFLNK RLQYIVKACF NVAKAYESFY EDFRKGEES 540
LDLKFKIGTT TTLKQYPQNE ARAM                                         564
```

```
SEQ ID NO: 361      moltype = AA length = 610
FEATURE          Location/Qualifiers
source           1..610
                  mol_type = protein
                  organism = unidentified
```

```
SEQUENCE: 361
HSHNLMLTKL GKQAINFAAN YDANLEIGCK NCKFLSYSPK QANPKKYPRQ TDVHEDGNIA 60
CHSCMQSTKE PPVYIVPIGE RKSKYEILTS LNKFTFLALK YKEKKRQAFR AKKPKELQEL 120
AIAFNKEKAI KVIDKSIQHL ILNPKPETAR IQRQKRLKMR KGKLLYLHKR YAIKMLIKN 180
GHYFKVGSPK KDGKLLVLVCA ALNTYGRDIG IIGNIEENN SETEITYQLY FDCLDANPNE 240
LRIKEIEYNR LKSYERKIKR LVYAYPKQQT KILEIRSKFV SKGHENKVNT GSFNFENPLN 300
KSISIKVKNs AFDFKIGAPF IMLRNGKFHI PTKKRLSKEE REINRTLISKI KGRVFRLTYG 360
RNISEQGSKS LHIYRKERQH PKL SLEIRKQ PDSFIDEFEK LRLKQNFISK LKKQRQKLA 420
DILLQFADRIA YNYHTSSLEK TSNFINYKPE VKRGRTSYIK KRIGNEGFEK LYCETCIKSN 480
DKENAYAVEK EELCFVCKAK PFTWKTKTNKD KLGIFKYPSR IKDFIRAFT VAKSYNDFYE 540
NLKKKDLKNE IFLKFKIGLI LSHEKKNHS IAKSVAEDER ISGKSIKNIL NKSIKLEKNC 600
YSCFFHKEDM                                         610
```

```
SEQ ID NO: 362      moltype = AA length = 552
FEATURE          Location/Qualifiers
source           1..552
                  mol_type = protein
```

---

-continued

---

SEQUENCE: 362    organism = unidentified

SLERVIDKRN	LAKKAINIAA	NFDANINKGF	YRCETNQCMF	IAQKPRKTNN	TGCSSCLQST	60
YDPVIYVVKV	GEMLAKYEIL	KSLKRFVFMN	RSPFKQKKTEK	AKQKERIGGE	LNEMSIFANA	120
ALAMGVKRA	IRHCHVDIRP	EINRLSELKK	TKHRVAAKSL	VKIVKQRKTK	WKGIPNSFIQ	180
IPOKARNKDA	DFYVASALKS	GGIDIGLCGT	YDKKPHADPR	WTYQLYFDTT	DESEKRLILYC	240
YNDPQAKIRD	FWKTFYERGN	PSMVNSPGTI	EFRMEGFFEK	MTPISIESKD	FDFRVNWKDL	300
LIRRGLYEIK	KRKNLNRKAR	EIKKAMGSVK	RVLANMTYGGK	SPTDKKSIPV	YRVEREKPKK	360
PRAVRKEENE	LADKLENYRR	EDFLIRNRKR	REATEIAKII	DAAEPPIRHY	HTNHRLAVKR	420
IDLSKPVARL	NTSVFLKRIM	QNGYRGNYCK	KC1KGNIQPN	KDECRLDEDI	KC1CCEGTQN	480
IWAKKEKLYT	GRINVLNKRI	KQMKLECFNV	AKAYENFYDN	LAALKEGDLK	VLKLKVSIPIA	540
LNPEASDPEE	DM					552

SEQ ID NO: 363    moltype = AA length = 534  
 FEATURE    Location/Qualifiers  
 source    1..534  
 mol\_type = protein  
 organism = unidentified

SEQUENCE: 363    organism = unidentified

NASINLGKRA	INLSANYDSN	LVIGCKNCKF	LSFNGNFPQR	TNVREGCHSC	DKSTYAPEVY	60
IVKIGERAK	YDVLDLSKKF	TFQSLKYQIK	KSMRERSKKP	KELLEVFIFA	NKDKAFNVIQ	120
KSYEHLLNLI	KQEINRMNGK	KRIKNHKKRL	FKDREKQLNK	LRLIGSSLF	FPRENKGDKD	180
LFTYVAIHSV	GRDIGVAGSY	ESHIPEISDL	TYQLFINNEK	RLLYAYKPKQ	NKIELKENL	240
WNLKKEVAKPL	DLEFTKPLEK	SITFSVKNDK	LFKVSKDML	ROAKFNIQGK	EKLSKEERQI	300
NRDFSKIKSN	VISLSYGRFE	ELKKEKNIWS	PHIYREVVKQ	EIKPCIVRKG	DRIELFEQLK	360
RKMDKLKFR	KEROKKISKD	LNFNAERIAYN	FHTKSIKNTS	NKINIDQEAK	RGKASYMRKR	420
IGNESFRKKY	C EQCFSVGNV	YHNVQNGCSC	FDNPIELIKK	GDEGLIPKGK	EDRKYKGALR	480
DDNLQMGIIR	VAFNIAKGYE	DFYNNLKEKT	EKDLKLKPKI	GTTISTQESN	NKEM	534

SEQ ID NO: 364    moltype = AA length = 577  
 FEATURE    Location/Qualifiers  
 source    1 .. 577  
 mol\_type = protein  
 organism = unidentified

SEQUENCE: 364    organism = unidentified

SNLIKLGKQA	INFAANYDAN	LEVGCCKNCF	LSSTTNKYPRQ	TNVHLDNKMA	CRSCNQSTM	60
PAIYIVRIGE	KKAKYDIYNS	LTKFNFQSLK	YAKRSQRFK	PKQPKELOEL	SIAVRKEKAL	120
DIIQKSIDHL	IQDIIRPEIPR	IKQQKRYKNNH	VGKLFYQLQKR	RKNKLNLIQK	GSFFKVFSPK	180
EKKNELLVIC	ALTNIGRDIQ	LIGNYNTIIN	PLFETVYQLY	YDYIPKKNNK	NVQRLLYAY	240
KSKNEKILKL	KEAFFKRGHE	NAVNLQPKV	EKPLEKSLT	KIKNDKDFQ	VSPSLRIRTG	300
RRFFVPSKRNL	S RQEREINRR	LVKIKSKIKN	M TYGKFETAR	DKQSVHIFRL	ERQKEKLPLQ	360
FRKDEKEFME	E FQKLKRRTN	SLKQLRKSRQ	KKLADLLQLS	EKVYVNNHTG	TLKKTTSNFLN	420
FSSSVKRKGK	A YIKELLGQE	GFETLYCSNC	INKGQKTRYN	IETKEKFCSC	KDVPFWKKK	480
STDKDRKGAF	LFPAAKLKDVI	KATFTVAKAY	EDFYDNLKSI	DEKKPYIKFK	IGLILAHVRH	540
EHKARAKEEA	C QKNIYNKPI	KIDNCNECF	FFKEEAM			577

SEQ ID NO: 365    moltype = AA length = 613  
 FEATURE    Location/Qualifiers  
 source    1 .. 613  
 mol\_type = protein  
 organism = unidentified

SEQUENCE: 365    organism = unidentified

NTTRKKFRKR	TGFPQSDNIK	LAYCSAIVRA	ANLDADIQKK	HNQCNPNLCV	GIKSNEQSRK	60
YEHSDRQALL	CYACNQSTGA	PKVODYIQIGE	IGAKYKILQM	VNAYDFLSLA	YNLTKLRNGK	120
SRGHQRMQLS	DEVVIVADYE	KATEVIKRSI	NHLLDDIRGQ	LSKLKRTQN	EHITEHKQSK	180
IRRKLRLKSLR	LLKRRRLWKW	TIPNPYLNK	VFTKKDPELV	TVALLHKLGR	DIGLVNRSKR	240
RSKQKLKLPV	GFQLYKKWES	PSLNNIKKSK	AKKLPKRLII	PYKVNKLFDN	KQKLENNAIKS	300
LLESYQKTIK	VEFDQFFQNR	TEEIIAEEQQ	TLERGLLQL	EKKKNEFASQ	KKALKEEKKK	360
IEKPRKAKLL	MEESRSLGFL	M ANVSYALFN	TTIEDLYKKS	NVVS GCI PQE	PVVVPADIQ	420
NKGS LAKILF	APKDGFRIKF	SGQH LTIRTA	KFKIRGKEIK	I LT KTKREIL	KNIEKLRRVW	480
YRBQHYKLKL	FGKEVSAKPR	F LDKRKTSIE	RRDPNKLADQ	T DDRQAE LRN	KEY ERLRHKQH	540
KMAERLDNID	TNAQNLQTLS	F FWVGEADKPP	KLDEKDARGF	G VRTCISAWK	WFMEDLLKKQ	600
EEDPLLLKLL	SIM					613

SEQ ID NO: 366    moltype = AA length = 615  
 FEATURE    Location/Qualifiers  
 source    1 .. 615  
 mol\_type = protein  
 organism = unidentified

SEQUENCE: 366    organism = unidentified

PKKPKFQKRT	GFPQPDNLK	EYCLAIVRAA	NLDADFEKK	TKCEGIKTNK	KGNIVKGRTY	60
NSADKDNL	LC YACNISTGAP	AVDYYFVGAL	EAKYKILQMV	KAYDFHSLAY	NLA KLWKG GRG	120
RGHQRMGGLN	EVVIVSNNEK	ALD VIEKSLN	H FHD EIRGEL	SRLKAKFQNE	HLHVHKE SKL	180
RRKLRKISRL	LKRRRWKWDV	IPNSYLRNFT	F T KTR PDF FIS	V ALL H RV GRD	I GLVTKTKIP	240
KPTDLPQFG	F QI YYTWDEP	K L N K L K K S RL	R S E P K R L L V P	Y K K I E LY K NK	S V L E E A I R H L	300
A E V Y T E D L T I	C F K D F F E T Q K	R K F V S K E K E S	L K R E L L K E L T	K L K D F S E R K	T A L K R D R K E I	360

-continued

---

KEPKAKKLLM	EESRSLGFLA	ANTSYALFNL	IAADLYTKSK	KACSTKLPRQ	LSTILPLEIK	420
EHKSTTSLAI	KPEEGFKIR	SNTHLSIRTP	KFKMKGADI	ALTKRKREIL	KNATKLEKSW	480
YGLKHYKLKL	YGKEVAAKPR	FLDKRNPSID	RRDPKELMEQ	IENRRNEVKD	LEYEIRKGQH	540
QMAKRLDNVD	TNAQNLQTKS	FWVGEADKPP	ELDSMEA	KKL GLRTCISAWK	WFMKDLVLLQ	600
EKSPNLKLKL	SLTEM					615

SEQ ID NO: 367	moltype = AA	length = 775					
FEATURE	Location/Qualifiers						
source	1..775						
	mol_type = protein						
	organism = unidentified						
SEQUENCE: 367							
KFSKRQEGLF	IPDNIDLYKC	LAI	VRSANLD	ADVQGHKSCY	GVKKNGTYRV	KQNGKKGVKE	60
KGRKYVFDL	IAFKGNIKEIP	HEA	IEEKDOG	RVIVL	GKPNY	KLILNIEKHN	120
KIKKLVQISS	LETGEFLSDL	LSGKIG	DEFDEV	YGIIE	PDVFS	GKELVCKACQ	180
MPVGELDAKY	KILSAIKGYD	FLSLAYNLSR	NRANKKRQHQ	KLGGG	ELSEV	VISANYDKAL	240
NVIKRSINHY	HVEIKPEISK	LKKKM	QNENPL	KVMKQARIRR	ELHQLSRKV	RLWKWGMIP	300
NPELQNIIFE	KKEKD	PFVSYA	LLHTL	GRDIG	LFKDTSMLQV	PNISDYCFQI	360
SIKKIKDLPK	RLLIPYKRLD	FYIDT	SDILYR	VIKNLIELYR	KSYVYETFGE	EYGYAKKAED	420
ILPDWDSSINL	SEGI	EQKIQ	IKDE	RSRDN	VESFRN	YDKNFASDRN	480
SYQEKIQS	MII	KKQQENIEQ	KLKREF	KEV	ERGFEGMDQ	KKYYKVLPN	540
NNLGFFRS	HL AFM	LLS	KISD	DLYRKNNLVS	KGGNKG	ILDQ	600
KRKFFNPK	SSWIGIRKPK	FSIKGAVIRE	ITKKV	RDEQR	LIKSLEGW	VKH	660
PRPNLPRHPD	REKNNNDNL	ESITSRREQI	QLLLRE	KQKQ	QEKMAGR	LDK IDKEI	720
ANFQIKQIDK	KPALTEKSEG	KQSVRNALSA	WKWF	MEDLIK	YQKRTPI	LQL KLA	775

SEQ ID NO: 368	moltype = AA	length = 777					
FEATURE	Location/Qualifiers						
source	1..777						
	mol_type = protein						
	organism = unidentified						
SEQUENCE: 368							
KFSKRQEGLF	IPENIGLYKC	LAI	VRSANLD	ADVQGHVSCY	GVKKNGTYVL	KQNGKKSIRE	60
KGRKYASDLV	AFKGDI	EKIP	FEVIE	EKKKE	QSIVL	GKPNY	120
KSKKLVQVSS	LGTDEF	LLTL	LNEKFG	YI	PEPVFS	GKELVCKACQ	180
MPVGELDSKY	KILSAIKGYD	FLSLAYNLR	HRSNK	KRQHQ	KLGGG	ELSEV	240
NVIKRSLNHY	YESEIKPEISK	LRK	MQNENPL	KVGKQARMR	ELHQLSRKV	RLWKWGMIP	300
NLELQ	NTFK ESDRDF	ISY	A	MFNKTE	I	MFN	360
TIKKS	KNTPK RILIPY	KK	FYND	SILVAR	AIKELV	GVLFQ	420
ELIKL	DEESI NGNVE	KLQR	I	KENFNS	LL	ESE	480
EYOREI	QSFIEK I	QKQ	EKF	NEQBOQ	KKH	YRVLNPT	540
NNLGFLRS	SKI	AFILL	SKISD	DLYKKSNA	KG	GEKG	600
KKKLFN	YKTYT SSWL	GI	YK	AKTR	TLK	SAESSWY	660
PRFNQPRHPD	KEKKSDD	R	ESIT	LLRE	Q	ILLRE	720
ANFQIKQIDK	KPALTEKSEG	KQSVRNALSA	WKWF	MENLLK	YQ	QKRTPI	777

SEQ ID NO: 369	moltype = AA	length = 610					
FEATURE	Location/Qualifiers						
source	1..610						
	mol_type = protein						
	organism = unidentified						
SEQUENCE: 369							
KWIEPNNIDF	NKCLAITRSA	NLDAD	VQGHK	MCYGIKTNGT	YKAIGKINKK	HNTGIIIEKRR	60
TYVYDLIVTK	EKNEKIVKKT	DFMA	IDE	EIE	FDE	KEKEKL	120
GEKFDDLC	CSI EEPQAF	PRSE	LV	CACNQ	ST	YASD	180
LYKNLGR	LRD SKKR	GHQKM	QGEL	EVFC	ANKE	KDVL	240
KM	NEPLKVN DQAR	NRRELN	QIS	RLKLR	WKW	GEIPNPE	300
TLGR	DIGLIN ETEL	KPNNIQ	EYGF	QIY	EDPE	LNHIK	360
YTIL	SR	YK	YK	YK	YDF	ELGYDE	420
YKK	KALLE	KKK NTLED	LSV	YQCSLL	EQIN	NVKKW	480
DIISRIE	EL NVEG	WIRT	KE	TI	RDIV	KEETN	540
EKKP	FREEPK PIV	YD	DV	LG	PE	GEN	600
PQIPETILD	L	L	SA	LF	LS	AL	610

SEQ ID NO: 370	moltype = AA	length = 632					
FEATURE	Location/Qualifiers						
source	1..632						
	mol_type = protein						
	organism = unidentified						
SEQUENCE: 370							
FRKFV	KRS	GA	PQP	DNL	NKYK	CIAIVRAANL	60
TT	ELGR	VYA	GSQ	GNLL	CTK	STMGPLV	120
ARTR	VSK	GG	QAF	Q	ST	DYVPIGRIRA	180
HKN	KEAR	IR	Q	Q	Y	KY	TYT
MINK	PKG	SAK	Q	Q	Q	TY	ILR
IENAM	HKL	LYDEN	EV	Q	Q	YEL	180

-continued

---

NKMQELFKEV	KEPKKANALM	KQSRNMGFLL	QNISYGALGL	LANRMYEASA	KQSKGDATAKQ	420
PSIVIPLEME	FGNAFFKLII	RSGKFAMNVS	SPWLTIRKPK	FVIKGNIKIN	ITKLMKDEKA	480
KLKRLETSYH	RATHFRPTLR	GSIDWDSPYF	SSPKQPNTTHR	RSPDRLSADI	TEYRGRLKSV	540
EAEELREQORA	MAKKLDVDM	TASNLTQTSNF	QLEKGEDPRL	TEIDEKGRSI	RNCISSLKKF	600
MEDLMKAQEAA	NPVIKIKIAL	KDESSVLS	SM			632

SEQ ID NO: 371	moltype = AA	length = 625
FEATURE	Location/Qualifiers	
source	1..625	
	mol_type = protein	
	organism = unidentified	

SEQUENCE: 371						
KFHPELNKNS	YCLAI	VRAAN	LDAD	IQGHIN	CIGIKSNKSD	60
TKSTYKPNIN	SVPV	GEKKAK	YISL	SEIEKKY	DFNSL	120
ITSEYKKALD	VINK	SVNHYL	VNIK	MNSKSL	KKLQNEHIIH	180
YRKWWKFPVN	KILKN	VFKN	QSPDF	VVAL	LHKLGRDIGL	240
KYDTPKLNLYL	KKS	KFKSLPK	RILISY	KYPK	FDINSNYIEE	300
IEIFFKKSED	NIK	ISENDLII	KRGIM	KEFEK	SIDKLLKLYE	360
VSRRPIGFLKA	YLSY	MLFNII	SNRIF	FSRK	ESPIYKNNSK	420
IGSKKNNKYF	CNLL	LLKSSGF	NISYEE	HLS	QFPE	480
I	KWKKLTLFF	KPSNLNGKKT	SDKIR	FKSPN	TFNPF	540
RKEFNKLAKK	HDGV	DVEAQC	LQT	KSF	ING RKL	600
WKWMADLIE	AQKETPMI	KL	KLALM			625

SEQ ID NO: 372	moltype = AA	length = 517
FEATURE	Location/Qualifiers	
source	1..517	
	mol_type = protein	
	organism = unidentified	

SEQUENCE: 372						
TTLVPSHLAG	I	EVMDETTSR	NEDMIQKETS	RSNEDENYL	G VKNKG	60
PNMPEKSGE	QOMP	KQDSTE	MQR	RFDES	VAT ASIKTDARAN	120
LIVKASNLD	DI	KLGC	KPC	EYIR	SELPMGK KNGCNHCEKS	180
VERFESFAAD	SISRH	LGEQ	ARTR	GRK	SDIASVPKVE SGFRKAKYEL	240
LDARSNIKE	TRN	SLRRL	DKKEQM	GKVN LDEIA	ILKNE SLIEY	300
NSLRLHTKLG	V	DIGVQGGDN	SIRQL	ITLRLR	TRKICFTKPK GLLPRHM	360
LRGYPELILY	NEEL	RIQDSQ	KFPL	DWERT	PRV	420
KRIQVNIESK	KKV	VLTRYVY	PIP	KLRGV	YVVA	480
EIAK	YLN	WG	KLGV	SVKALNR	ITEAPR	517
WGC	EGL	HNKL	QTKNP	YAFK	LVVA	

SEQ ID NO: 373	moltype = AA	length = 410
FEATURE	Location/Qualifiers	
source	1..410	
	mol_type = protein	
	organism = unidentified	

SEQUENCE: 373						
LDFKRTCSQE	L	VLLPEIEGL	KLSGTQGVT	S LAKLINKAA	NVRD ESYGC HHC	60
SKPVKKDCNS	CNQ	STNHPAV	PITL	KGYKIA FYELWHR	PT SAVSAGT ASIKTDARAN	120
LDEYAVVDNS	HIV	CYAVRKC	YEK	RQSRVRL HKRAYRCRAK	YHNSQPKVG RIYKKSKRRN	180
ARNLKKEAKR	YFQP	NEITNG	SSDAL	FYKIG VDLGI	AKTP ETEVKVDVSI CFQVYYGDAR	240
RVLRV	RKMD	EQSFHLDYT	KLKL	GIGNK DFTPI	AKRN SLKGSTKYE VSRAHK	300
FGKKGSV	WSCEA	NRKMSR	QSNL	KNAPFY QKALV	KCYKN LDYKG	360
WYRLCSNRI	PRYS	RAEDI	QYQ	SDKGKA KFEFV	ILAQS VAEYDISAIM	410

SEQ ID NO: 374	moltype = AA	length = 602
FEATURE	Location/Qualifiers	
source	1..602	
	mol_type = protein	
	organism = unidentified	

SEQUENCE: 374						
VFLTDDKRKT	ALRK	I	KIRSAFR	KTAEIALVRA	QEADSLDRQA	60
GSLQGYNWN	HRAN	V	PSSGS	AKDVFRIT	EL GLGIPQSAHE ASIGKS	120
LSKGYKKGAV	NKG	AQRE	KGK	QLSFDL ISNGPIS	SDK LINGQKDALA WWLIDKMG	180
IGLAMEPLSS	PNTY	GT	FWK	RHTAPR YSRGV	YQWQ LFFGRQLAPL IHNFFRK	240
SIPIVL	TNAS	KKLAG	KG	EQTALVDPKK W	WQVKEQVTG PLSN	300
HHGAAHKRP	L	T	WV	WV	WIERSV PLVLYTATFT	360
DLIRARER	PLAY	CTV	QI	MPD	INILRDG RPDEA	420
EQGIGGR	PIQ	ELL	Q	AKL	VGKGP	480
ISNIHKDAYK	TAIE	P	TF	VTG	SIT PLE A	540
QQAAEMTV	Q	S	Q	Q	ADGNT SEKA	600
TQSVSK	V	S	Q	Q	EVADGNT SEKA	602

SEQ ID NO: 375	moltype = AA	length = 494
FEATURE	Location/Qualifiers	
source	1..494	
	mol_type = protein	

---

-continued

---

SEQUENCE: 375	organism = unidentified
EDLLQKALNT ATNVAAIERH SCISCLFTES EIDVKYKTPD KIGQNTAGCQ SCTFRVGYSQ	60
NSHTLPMGNR IALDKLRETI QRYAWHSLLF NVPAPTSKR VRAISELRVA AGRERLFTVI	120
TFVQTNLSK LQKRYAANWT PKSQRRLSRL REEGQHILSL LESGSWQQKE VVREDQDLIV	180
CSALTCKPLS IGAFCRPKYL KPAKHALVL R LIFVEQWPGQ IWGQSXRTRR MRRRKDVERV	240
YDISVQAWAL KGKETRISEC IDTMRRHQQA YIGVLPFLIS SGSTVRKGKD CPILKEITRM	300
RYCPNNNEGLI PLGIFYRGSA NKLLRUVVKGS SFTLPMWNQNI ETLPHPEPFS PEGWTATGAL	360
YEKNLAYWSA LNEAVDWYTG QILSSGLQYP NQNEFLARLQ NVIDSIPRKW FRPQGLKNLK	420
PNGQEDIVPN EFVIPQNAIR AHHVIEWYHK TNLDLVAKLTL GWGSQTLINQ TRPQGDLRFT	480
YTRYYYFREKE VPEV	494
 SEQ ID NO: 376	moltype = AA length = 649
FEATURE Location/Qualifiers	
source 1..649	
mol_type = protein	
organism = unidentified	
SEQUENCE: 376	
VPKKKLMREL AKKAVFEAIF NDPIPFGSCG KRCTLIDGAR VTDAIEKKQG AKRCAGCEPC	60
TFHTLYDSVK HALPAATGCD RTAIDTGLWE ILTALRSYMW MSFRRRAVSD ASQKQWWSIE	120
ELAIWADKER ALRVILSALT HTIGKLKNGF SRDGVWKGGK QLYENLAQKD LAKGLFANGE	180
IFGKELVEAD HDMLAWTIVE NHQPHIGLIR GNWKPAAVEA STAFDARWLT NGAPLRTDRT	240
HGHGRGRFFNR TEKLTVCILK RDGGVSEEFR QERDYELSVN LLQPKNKLKP EPKGELNSFE	300
DLHDHWWFLK GDEATALVGL TSDPTVGDPI QLGLYIRNP1 KAHGETKRL LICFEPPIKL	360
PLRRAFPSEA PTKTWEPTINV FRNNGRRDEA YYDIDRARVF EFPEPTRVSLE HLSKQWEVLR	420
LEPDRENTDP YEAQQNEGAE LVQYSSLQEA AQKMAPKVII DPFGQFPLEL FSTFVAQLFN	480
APLSDTKAKI GKPLDSGFV ESHLHLLEED FAYRDFVRVT FMGTEPTFRRV IHYSNGEGYW	540
KKTVLKGNN IRTALIPEGA KAAVDAYKNK RCPLTLEAAI LNEEKDRRLV LGNKALSLLA	600
QTARGNLTL EALAAEVLRP LSGTEGVVHL HACVTRHSTL TESTETDNM	649
 SEQ ID NO: 377	moltype = AA length = 414
FEATURE Location/Qualifiers	
source 1..414	
mol_type = protein	
organism = unidentified	
SEQUENCE: 377	
VEKLFSERLK RAMWLKNEAG RAPPAETLTL KHKRVSGGHE KVKEELQRLV RSLSGTNQAA	60
WNLGLSGGRE PKSSDALKGE KSRVVLLETWV FHSGHNRVLY DVIEREDQVH QRSSIMMMRR	120
KGSNLLRLWG RSGKVRKMR EEVAVIKPVW HKDSRWLAIIV EEGRQSVVGI SSAGLAVFAV	180
QESQCTTAEP KPLEYVWSIW PRGSKALNPQ DRYLEFKKLK TTEALRCQQY DPIPFSLKRG	240
AGCSSLAIRGE GIKFGSRGPI KQFFGSDRSR PSHADYDGKR RLSLFSKYAG DLADLTEEQW	300
NRTVUSAFAED EVRRATLANI QDFLSSHEK YAERLKKRIE SIEEPVSASK LEAYLSAIFE	360
TFVQQREALA SNFLMRLVES VALLISLEEK SPRVEFRVAR YLAESKEGFN RKAM	414
 SEQ ID NO: 378	moltype = AA length = 413
FEATURE Location/Qualifiers	
source 1..413	
mol_type = protein	
organism = unidentified	
SEQUENCE: 378	
VVITQSELYK ERLLRVMEIK NDRGRKEPRE SQGLVLRFTQ VTGGQEKVKQ KLWLIFEGFS	60
GTNQASWNFG QPAGGRKPN S GDALKGPKSR VTYETVVVFH GLRLLSAVIE RHNLKQQRQ	120
MAYMKRRAAA RKKWARSGKK CSRMRNNEVEK IKPKWHKDPR WFDIVKEGEP SIVGISSAGF	180
AIYIVEEPNF PRQDPLEIEY AISIWFRRDR SQYLTFKKIQ KAEKLKELQY NPIPFRLKQE	240
KTSLVFESGD IKFGSRGSIE HFRDEARGKPK PKADMNNR LTMFSVFSGN LTNLITEEQYA	300
RPVSGLLAPD EKRMPTLLKK LQDFFTPIHE KYGERIKQRL ANSEASKRPF KKLEEYLPAI	360
YLEFRARREG LASNWVLVLI NSVRTLVRIK SEDPYIEFPKV SQYLLEKEDN KAL	413
 SEQ ID NO: 379	moltype = AA length = 449
FEATURE Location/Qualifiers	
source 1..449	
mol_type = protein	
organism = unidentified	
SEQUENCE: 379	
KQDALFEERL KKAIFIKRQA DPLQREELSL LPPNRKIVTG GHESAKDTLK QILRAINGTN	60
QASWNPGTPS GKRDKSKSADA LAGPKSRVKL ETVVFHVGHF LLKKVVVEYQG HQKQOHGLKA	120
FMRTCAAMRK KWKRSGKVVG ELREQLANIQ PKWHYDSRPL NLCFEGKPSV VGLRSAGIAL	180
YTIQKSVPV KEPKPIEYAV SIWFRGPKAM DREDRCLEFK KLKIATELRK LQFEPIVSTL	240
TQGIKGFSLY IQGNSVKFGS RGPIKYFSNE SVRQRPPKAD PDGNKRLALF SKFSGDLSLD	300
TEEQWNRPIL AFEGIIRRAT LGNIQDYLTV GHEQFAISLE QLLSEKESVL QMSIEQQRLLK	360
KNLGKKAENE WVESFGAEQA RKKAQGIREY ISGFFQEYCS QREQWAENWV QQLNKSVRFL	420
LTIQDSTPFI EFRVARYLPK GEKKKGKAM	449
 SEQ ID NO: 380	moltype = AA length = 711
FEATURE Location/Qualifiers	
source 1..711	

---

-continued

---

```

mol_type = protein
organism = unidentified

SEQUENCE: 380
ANHAERHKRL RKEANRAANR NRPLVADCDT GDPLVGICRL LRRGDKMQPN KTGRSCEQV 60
EPELRDAILV SGPGRLDNYK YELFQRGRAM AVHRLLKRVP KLNRPKKAAG NDEKKAENKK 120
SEIQKEQKQ RRMMPAVSMK QVSVADFHV IENTVRHLFG DRRDRTEAEC AALRAASKYF 180
LKSRRVRPRK LPKLANPDHG KELKGLRLRE KRAKLLKEKE KQAEALARSNQ KGAVLHVATL 240
KJDAPPMPYE KTQGRNDYTT FVISAIAKVG ATRGTKPLLT POPPREQCSL YWRDGQRWIR 300
GGLLGLQAGI VLGPKLNREL LEAVLQRPPIE CRMMSGCGNPL QVRGAADVFFF MTTNPFYVSG 360
AZAYAQKFKP FGTKRASEDG AAAKAREKLM TQLAKVLDKV VTQAAHSPLD GIWETRPEAK 420
LRAMIMALEH EWIFLRPGPC HNAAEVIK CDTGCHAILW ALIDEARGAL EHKEFYAVTR 480
AHTHDCEKQK LGGRLAGFLD LLIQDQVPLD DAPAARKIKT LLEATPPAPC YKAATSIATC 540
DCEGKFDKLW AIIDATRAGH GTEDLWARTL AYPQNVNCCK KAGKDLTHRL ADFLGLLIKR 600
DGPFRRERPH KVTGDRKLVF SGDKKCKGHQ YVILAKAHNE EVVRAWISRW GLKSRTNKAG 660
YAATELNLLL NWLSICRRRW MDMLTVQRDT PYIRMKTGRL VVDDKKERKA M 711

SEQ ID NO: 381      moltype = AA length = 574
FEATURE             Location/Qualifiers
source              1..574
mol_type = protein
organism = unidentified

SEQUENCE: 381
AKQREALRVA LERGIVRASN RTYTLLVTNCT KGGPLPEQCR MIERGKARAM KWEPKLVGCG 60
SCAAATVDLP AIEEYAQPGR LDVAKYKLTT QILAMATRMM MVRAAKLSRR KGQWPQAKVQE 120
EKEEPPEPKK MLKAVEMRPV AIVDFNRVIQ TTIEHLWAER ANADEAELKA LKAAAAYFGP 180
SLKIRARGPP KAAIGRELKK AHRKKAYAER KKARRKRAEL ARSQARGAAA HAAIRERDIP 240
PMAYERTQGR NDVTITPIAA AIKIAATRGA RPLPAKPKPMK WQCSLYWNEG QRWIRGMLT 300
AQAYAHAAANI HRPMRCEMWG VGNPLKVRAP EGRVADPDA GKRKAERFLQ TNAYFVSGAA 360
YRNKKFKPFG TDRGGIGSAR KKRERLMAQL AKILDKVVSQ AAHSPLDDIW HTRPAQKLRA 420
MIKOLEHEWM FLRPQAPTVE GTKPDVDAVG NMQRQIKALM APDLPIEKG SPAKRTGDK 480
RKKGERAVR AEAHSDEVVT AWISRWTQI RRNEGGSYAAQ ELELLLNWLQ ICRRRQLDMT 540
AAQRVSPYIR MKSGRMITDA ADEGVAPIPL VENM 574

SEQ ID NO: 382      moltype = AA length = 743
FEATURE             Location/Qualifiers
source              1..743
mol_type = protein
organism = unidentified

SEQUENCE: 382
KSISGRSIKH MACLKDMILKS EITEIEEKQK KESLRKWDYY SKFSDEILFR RNLNVSANHD 60
ANACYGCNPAC AFLKEVYGFRI IERRNNERII SYRRLLAGCK SCVQSTGYPP IEFVRRKFGA 120
DKAMEIVREV LHRRNWGALA RNIGREKEAD PILGELNELL LVDARPVFGN KSAANETNLA 180
FNVITRAAKK PRDEGMYDIH KQLDIHSEEG KVPKGRKSRL IRIERKHKAI HGLDPGETWR 240
YPHCGKGEKY GVWLNRSLRI HIKGNEYRCL TAFGTTGRRM SLDVACSVLG HPLVKKRKK 300
GKKTVDGTEL WQIKKATETL PEDPIDCTFY LYAAKPTKDP FILVKVSLKA PRWKLLHKDF 360
FEYSDTEKTQ QEKKGKRVVR RGKVPRLSL RPDAKFVSI WDDPYNGKKNK EGTLRLMELS 420
GLDGAKKPLI LKRYGEPNTK PKNFKFWRPH ITPHPLTFPT KHDFGDPNKK TKRRRVFNRE 480
YVGHNLNDLAK MEPNAKFFED REVSNKKNPK AKNIRIQAKE SLPNIVAKNG RWAAPDPNDS 540
LWKLYLHWRG RRKTIKGGIS QEOFQEFKERL DLYKKHEDES EWKEKEKLWE NHEKEWKTL 600
EIHGSIAEVS QRCVMQSMMG PLDGLVQKKD YVHIGQSSLK ADDAWTFSA NRYKKATGPK 660
WGKISVSNLL YDANQANAEI ISQSIISKYLs KQKDNQGCEG RKMKFlikii EPLRENFVKH 720
TRWLHEMTQK DCEVRAQFSR VSM 743

SEQ ID NO: 383      moltype = AA length = 769
FEATURE             Location/Qualifiers
source              1..769
mol_type = protein
organism = unidentified

SEQUENCE: 383
FPSDVGADAL KHVRMLQPRL TDEVRKVALT RAPSDRPALA RFAAVAQDGL AFVRHLNVSA 60
NHDSNCTPDR DPRDPRGPG EPNPAFRLB VWGFRIVARG NERALSYRRG LAGCKSCVQS 120
TGFPSPVPHR IGADDCMRKL HEILKARNWR LLARNIGRER EADPLLTLES EYLLVDARTY 180
PDGAAPNSGR LAENVIKRAA KKFRDEGMRD IHACLRLVHSR EGKVPKGRLQ RLRIERKHR 240
AIHALDGPWS WEAEGSARAE VQGVAVYRSQ LLRVGHHTQ IEPVGIVART LFGVGRTDLD 300
VAWSVLGAPL TKRKKGSKTL ESTEDVLAQK ARETRAEDKI EVAFVLYPTA SLLRDEIPKD 360
AFFPAMRIDRF LLKVGSVQAD REILLQDDYY RFQDAEVKAG KNKGRTVTRP VKVPRQLQALR 420
PDAKFRVNW ADPFGAGDSP GTLLRLEVSG VTRRSQPLRL LRYGQPSTQP ANFLCWRPHR 480
VPPDMTFTPR QKFGERRKRN RTRRPRVFER LYQVHikhla HLEPNRKWFE EARVSAQKWA 540
KARAIRKGA EDIPVVAAPPA KRRWAALQPN AELWDLYAHD REARKRFRGG RAAECEEFKP 600
RLNLYLAHEP EAEWESKDR WERYEKKWTQ VLEEHSRMCA VADRTLQFL SDPLGARMDD 660
KDYAFVGKSA LAVAEEFVEE GTVERAQGNIC SITAKKKPAS NASRKRLSVA NLLDVSDKAD 720
RALVFQAVRQ YVQRQAENGG VEGRRMAFLR KLLAPLQRNF VCHTRWLHM 769

SEQ ID NO: 384      moltype = AA length = 564
FEATURE             Location/Qualifiers
source              1..564

```

-continued

---

```

mol_type = protein
organism = unidentified

SEQUENCE: 384
AARKKKRGK1 GITVKAKEKS PPAAGPFMAR KLVNVAANVD GVEVHLCVEC EADAHGSASA 60
RLLGGCRSCT GSIGAEGLRM GSVDVDRERV IAEPVHTETE RLGPDVKAFE AGTAESKYAI 120
QRGLEYGVGD LISRNRARTV RKMEEDARPE SSTMEKTISWD EIAIKTYSQA YHASENHLFW 180
ERQRRVRQHA LALFRRARER NRGEPLQST QRPAPLVLIA LHAEEAAISG RARAEVYLRG 240
PSANVRAAAA DIDAKPLGHY KTPSPKVARG FPVKRDLRLRA RHRIVGLSRA YFKPDSVVRG 300
TSDAIAHVAG RNIGVAGGKP KEIEKFTFLP FVAYWEDVDR VVHCSSFKAD GPWVRDQRIK 360
IRGVSSAVGT FSLYGLDVAW SKPTSFYIRC SDIRKKFHPKF GFGPMKHWRQ WAKELDRLTE 420
QRASCVVRAL QDDEELLQTM ERGQRYYDVF SCAATHATRG EADPSGGCSR CELVSCGVAH 480
KVTKKAKGDT GIEAVAVAGC SLCESKLVGP SKPRVHRQMA ALRQSHALNY LRRRLQREWEA 540
LEAVQAPTPY LRFKYARHLE VRSM 564

SEQ ID NO: 385      moltype = AA length = 565
FEATURE           Location/Qualifiers
source            1..565
mol_type = protein
organism = unidentified

SEQUENCE: 385
AAKKKKQRGK1 IGISVVPKPEG SAPPADGPFM ARKLVNVAAN VDGVEVNLCI ECEADAHGSA 60
PARLLGGCKS CTGSIGAEGR LMGSVVDVRA DAIAKPVNTTE TEKLGPDVQA FEAGTAETKY 120
ALQRGLEYWG VDLISRNRSR TVRRTEEGQP ESATMEKTISWD DEIAIKSYTR AHASENHLF 180
WERQRRVRQH ALALFKRAKE RNRGDSTLPR EPGHGLVAIA ALACEAYAVG GRNLAEVVVR 240
GPTFGTARAV RDVEIASLGR YKTPSPKVVAH GSPVKRDPLR ARHRIVGLAR AYYRPSDVRV 300
GTSDAIAHVA GRNIGVAGGK PRAVEAVFTL PFVAYWEDVDR RVVHCSSFQV SAPWNRDQRM 360
KJAGVTTAAC TFSLHGGELK WAKPTSFYIR CSDTRRKFRP KGFGPMKRWR QWAQDDLRLV 420
EQRASCVVRAL QDDAALLET MERGORYYDV FACAVTHATR GEADRLAGCS RCALTPCQEAE 480
HRVTTKPRGD AGVEQVQTSD CSLCEGKLVG PSKPRLHRTL TLLRQEHEGLN YLRRRLQREWEA 540
SLEAVQVPTP YLRFKYARHLE EVRSM 565

SEQ ID NO: 386      moltype = AA length = 499
FEATURE           Location/Qualifiers
source            1..499
mol_type = protein
organism = unidentified

SEQUENCE: 386
TDSQSESVPE VVYALTGGEV PGRVPDPGGS AEGARNAPTG LRKQRGKIKI SAKPSKPGSP 60
ASSLATTLVN EAANVDPGVQS SGCATCRMRA NGSAPRALPI GCVACASSIG RAPQBEETVCA 120
LPTTQGPDVR LLEGGHALRK D1DQRALEYW GVDLIGRNLD RQAGRGMPEA EGATATMKRV 180
SMDELAVLDF GKSYYASEQH LFAARQRRVR QHAKALKIRA KHANRSGSVK RALDRSRKQV 240
TALAREFFKP SDVVRGDSDA LAHVVGGRNLG VSRHHPARIP QTFTLPLCAY WEDVDRVISC 300
SSLLAGEPFRA RDQEIRIEGV SSALGSLRQL RGAIIEWHKPT SLYIRCSDTR RKFRPRGGLK 360
KRWRQWAKDL DRLVEQRACC IVRSLQADVE LLQTMERAQR FYDVHDCAAT HVGPVAVRCS 420
PCAGKQFDWD RYRLLAALRQ EHANLYLRLQ QREWESLEAQ QVKMPYLRFK YARKLEVSGP 480
LIGLEVRREP SMGTAIAEM 499

SEQ ID NO: 387      moltype = AA length = 358
FEATURE           Location/Qualifiers
source            1..358
mol_type = protein
organism = unidentified

SEQUENCE: 387
AGTAGRRHGS LGARRSINIA GTVDRHGRWG CESCVYTRDQ AGNRARCAPC DQSTYAPDVQ 60
EVITIGQRAK YTIFLTQSQE SWTNTMRNNK RAAAGRKSRT TGKRIQOLAE IKITGVGLAH 120
AHNVIQRSLO HNITLTKR QKAKQLTKRR AFYRRRMSRQ SRGNQFFRTG 180
KGIGIHAVAPV K1GLDVGMIA SGSSEPADBQ TVTLDAIWKG RKKKIRLIGA KGELAVAACR 240
FREQQTGDK CIPLILQDGE VRWNQNNWQC HPKKLVPCLC LEVSRKFVSQ ADRLAQNKVA 300
SPLAARFDKT SVKGTLVESD FAAVLVNTS IYQQCHAMLL RSQEPTPSLR VQRTITSM 358

SEQ ID NO: 388      moltype = AA length = 369
FEATURE           Location/Qualifiers
source            1..369
mol_type = protein
organism = unidentified

SEQUENCE: 388
GVRFSPAQSQ VFFFRTVIPQS VEARFAINMA AIHDAAGAFG CSVCRFEDRT PRNAKAVHGC 60
SPCTRSTNRP DVFVLPGAI KAKYDVFML LGPNWTHLNR RQAKRVTVRD RIGQLDELA 120
SMLTGKAKAV LKKSICHNVD KSFKAMRGSL KKLHRKASKT GKSQRLAKLS DLRERTNTTQ 180
EGSHVGEQDSD VALNKIGLDV GLVGKPDYPS EESVEVVVCL YFVGKVLILD AQGRIRDMDRA 240
KQYDGFKIP1 IQRGQLTCLS VKDLDGKWSLV RQDYVLAGDL RFEPKISKDR KYAECVKRIA 300
LITLQASLGF KERIPYYVTK QVEIKNASHI AFVTEAIQNC AENFREMTEY LMKYQEKSPD 360
LKVLTTQLM 369

SEQ ID NO: 389      moltype = AA length = 486
FEATURE           Location/Qualifiers

```

-continued

---

```

source          1..486
               mol_type = protein
               organism = unidentified

SEQUENCE: 389
RAVGKVFLE QARRALNLAT NFGTNHRTGC NGCYVTPGKL SIPQDGKNA AGCTSCLMKA 60
TASYVSYKP LGEKVAKYST LDALKGFPWY SLRLNLRPNA RGKPINGVQE VAPVSKFRLA 120
EEVIQAVQRY HFTELEQSFP GCRRLRRELH AFYTKERYRN PEGRPQHVNG DRNIVVVTVL 180
HELGFSVGMF NEVELLPKTP IECAVNVFIR GNRVLLEVRK PQFDKERLLV ESLWKDSRF 240
HTAKWTTPNN EGRIFTAEGW KDFQLPLLLG STSRSLRAIE KEGFVQLAPG RDPDYNNTID 300
EQHSGRPFPLP LYLYLQGTIS QEYCVFAGTW VIPFQDGISP YSTKDTFQPD LKRKAYSLLL 360
DAVKHRLGNK VASGLQYGRF PAIEELKRLV RMHGATRKIP RGEKDLLKG DPDTPEWLL 420
EQYPEFWRLC DAAAKRVSQN VGLLLSLKKQ PLWQRWRWLES RTRNEPLDNL PLSMALTLHL 480
TNEEAL                                         486

SEQ ID NO: 390      moltype = AA length = 400
FEATURE
source          1..400
               mol_type = protein
               organism = unidentified

SEQUENCE: 390
AAVYSKFYIE NHFKMGIPET LSRIRGPSII QGFSVNENYI NIAGVGDRDF IFGCKKCYT 60
RGKPSSKKIN KCHPCKRSTY PEPVIDVRGS ISEFKYKIYN KLKQEPNQSI KQNTKGRMNP 120
SDHTSSNDGI IINGIDNRIA YNVIFQSYKH LMEKQINLLR DTTKRKARQI KKYNNSGKKK 180
HSLRSQTKGK LKNRYHMLGM FKKGSLTITN EGDFITAVRK VGLDILSYKN ESLNKQEVE 240
ELCLNIKWGR TKSYTSGYI PLPINIDWKL YLFEKETGLT LRLFGNKYKI QSKKFLIAQL 300
FKPKRPPCAD PVVKAQKWS ALNAHVAQOMA GLFSDSHLLK RELKNRMHKQ LDFKSLWVGT 360
EDYIKWFEEL SRSYVEGAEK SLEFFRQDYF CFNYTKQTTM                                         400

SEQ ID NO: 391      moltype = AA length = 666
FEATURE
source          1..666
               mol_type = protein
               organism = unidentified

SEQUENCE: 391
PQQQRDLMLM AANYDQDYGN GCGPCTVVAS AAYRPDPQAO HGCKRHLRTL GASAVTHVGL 60
GDRTATITAL HRLRGPAALA ARARAAQAAS APMPDPTDAP DRRRLAEAD ADDVVLVGAH 120
RALWSAVRRW ADDRRAALRR RLHSEREWLL KDQIRWAELY TLIEASGTPP QGRWRNTLGA 180
LRQOSRWRV LAPTMRATCA ETHAELWDAL AELVPEMAKD RRGLLRRPVE ADALWRAPMI 240
VEGWRGGHSV VUDAVAPPD LPQPCAWTAV RLSGDPQRW GLHLAVPPLG QVQPPDPPLKA 300
TLAVSMRHG CDRVRLQAM AVADAPMQR HLQVPLTLQR GGGLQNGIHS RGVRREARS 360
MASWEGPPI TGLQLVNRWK QGGSALLAPD RPPDTPPYAP DAAVAPAQPD TKRARRTLKE 420
ACTVCRCAPG HMRQLQVTLT GDGTWRRFRL RAPQGAKRKA EVLKVATQHD ERIANYTAWY 480
LRPPEHAAGC DTCGDGSRLD GACRGCRLP VGDQCFCRRL DKIEADRDG LAQIKPKAQE 540
AVAAMAAKRD ARAQKVAARA AKLSEATGQR TAATRDASHE ARAQKELEAV ATEGTTVRHD 600
AAAASFAGSW VARKGDEYRH QVGVLANLRE HGLRLQELMA PDSVVADQQR ASGHARVGYR 660
YVLTAM                                         666

SEQ ID NO: 392      moltype = AA length = 560
FEATURE
source          1..560
               mol_type = protein
               organism = unidentified

SEQUENCE: 392
AVAHPVGRGN AGSPGARGPE ELPRQLVNRA SNVTRPATYG CAPCRHVRLS IPKPVLTGCR 60
ACEQTTHPAP KRAVPGGADA AKYDLAFFA GWAADLEGRN RRRQVHAPLD PQPDPNHEPA 120
VTLQKIDLAE VSIEEFQRLV ARSVKHLRQD RASREREAQ AYAQVAKRR NSHAHGARTR 180
RAVRRQTRAV RRAHMRGANS GEILVASGAE DPVPEAIDHA AQLRRRIRAC ARDLEGLRHL 240
SRRYLKTLEK PCRRRAPDL GRARCHALVE SLQAAERELE ELRRCDSPT AMRRRLDAVLA 300
AAASTDATFA TGWTUVGMDL GVAPGRGSAAP EVSPMEMAIS VFWRKGSRV IVSKPIAGMP 360
IRRHELIRLE GLGTLRLDGN HYTGAGVTKG RGLSETEPD FREKSPSTLG FTLSDYRHES 420
RWRPYGAKQG KTARQFFAAM SRELRALVHB QVLAPMGPLL LEAHERRFET LLKGQDNKSI 480
HAGGGGRYVW KGPPDSKKRP AadGDWFRFG RGHADHRGWA NKRHELAANY LQSAFRLWST 540
LAEAQEPY ARYKYTRVTM                                         560

SEQ ID NO: 393      moltype = AA length = 404
FEATURE
source          1..404
               mol_type = protein
               organism = unidentified

SEQUENCE: 393
WDFLTLQVYE RHTSPEVCVA GNSTKCASGT RKSDHHTHGVG VKLGAQEINV SANDDRDHEV 60
GCNICVISRV SLDIKGWRYG CESCVQSTPE WRSIVRFDRN HKEAKGECLS RFYWGAQSI 120
ARSLKRNLKM GGVNLDELAI VQNENVVKTS LKHLFDKRD RIQANLAKVK VRMRERRKSG 180
RQKALRRCQ RKLKRYLRSY DPSDIKEGNS CSAFTKLGLD IGISPKNPK IEPKVEVVF 240
LFYQGACDKI VTVSSPESPL PRSWKIKID IRALYVKSTK VKFGGRTFRA GQRNNRRKVR 300
PPNVKKGKRK GSRSQFFNKF AVGLDAVSQQ LPIASVQGLW GRAETKKAQT ICLKQLESNK 360

```

-continued

---

PLKESQRCLF LADNWVVRVC GFLRLALSQRQ GPTPYIRYRY RCNM	404
SEQ ID NO: 394 moltype = AA length = 392	
FEATURE Location/Qualifiers	
source 1..392	
mol_type = protein	
organism = unidentified	
 SEQUENCE: 394	
ARNVGQRNAS RQSKEAKA RSRRVTGGHA SVTQGVALIN AAANADRHT TGCEPCTWER 60	
VNLPLQEVIY GCDSCTKSSP FWDRVKVVNK GYREAKEEIM RIASGISADH LSRALSHNKV 120	
MGRLLNLDEVC ILDFRTVLDT SLKHLTDERS NGIKEHIRAV HRKIRMRRKS GKTARALRKQ 180	
YFALRROWKA GHKPNSIREG NSLTALRAVG FDVGSETE PMPAPQTEVV LSVFYKGSA 240	
RILRISSPHP IAKRSWKVKI AGIKALKLIR REHDFSGRE TYNASQRAEK RKFSPPAARK 300	
DFFNSFAVQL DRLAQQLCVS SVENLWVTEP QQKLTLAKD TAPYGIREGA RFADTRARLA 360	
WNWVFRVCGF TRALHQEQEP TPYCRFTWRS KM 392	
 SEQ ID NO: 395 moltype = AA length = 707	
FEATURE Location/Qualifiers	
source 1..707	
mol_type = protein	
organism = unidentified	
 SEQUENCE: 395	
MADPTPLTFTQ FLRHHPGQR FRKDILKQAG RILANKGEDA TIAFLRGKSE ESPPDFQPPV 60	
KCPIIACSRP LTEWPIQAS VAIQGYVYQQ SLAEFEASDP GCSDKGLLGW FDKGVCVTDY 120	
FSVQGLNLIF QNARKRYIGV QTAKVNRNEK RHKKLKRINA KRIAEGLPEL TSDEPESALD 180	
ETGHLIDPPG LNTNIYCYQQ VSPKPLASE VNQLPTAYG YSTSGDPIQ PMVTKDRLSI 240	
SKGQPGYIPE HRRMRGYGLK ARALLVIVRI QDDWAVIDL SLLRNAYWRR 300	
IVQTKEPSTI TKLLKLVTGD PVLDATRMA TFTYKPGIVQ VRSAKCLKNK QGSKLFSE 360	
LNETVSVTSI DLGSNNLVAV ATYRLVNGNT PELLQRFTLP SHLVKDFERY KQAHDTLEDS 420	
IQKTAVASLP QQQQTEIRMW SMYGFREAOE RVQCQELGLAD GSIPWNVMTA TSTILTDLFL 480	
ARGGDPKKM FTSEPKKKN SKQVLYKIRD RAWAKMYRTL LSKETREAWN KALWGLKRG 540	
PDYARLSKRK EELARCRVN TIASTEKRAQ CGRTIVALED LNIGFFHGRG KQEPGWVGLF 600	
TRKKENRWLM QALHKAFEL AHHRGYHVIE VNPPAYTSQTC PVCRHCDPDN RDQHNREAFH 660	
CIGCGFRGNA DLDVATHNIA MVAITGESLK RARGSVASKT PQPLAAE 707	
 SEQ ID NO: 396 moltype = AA length = 757	
FEATURE Location/Qualifiers	
source 1..757	
mol_type = protein	
organism = unidentified	
 SEQUENCE: 396	
MPKPAVESEF SKVLKKHFPG ERFRSSYMKR GGKILAAQGE EAVVAYLQGK SEEPPNFQP 60	
PAKCHVVTKS RDFAEWPIK MASEIQRYIY ALSTTERAAC KPGKSSESNA AWFAATGVSN 120	
HGYSHVQGLN LIFDHITLGRY DGVLKKVQLR NEKARARLES INASRADEGL PEIKAEEEV 180	
ATMETGHLLQ PPGINPSFYV YTQISPQAYR PRDEIVLPPE YAGYVRDPNA PIPLGVVRNR 240	
CDIQKGCPGY IPEWQREAGT AISPPTGKAV TVPGLSPKKN KRMRRYWRSE KEKAQDALLV 300	
TVRIGTDWVV IDVRGILLRNA RWRTIAPKDI SLNALLDLFT GDPVIDVRRN IVTFYTTLDA 360	
CGTYARKWTI KGKQTKATLD KLTATQTVAL VAIIDLQTNP ISAGISRVTQ ENGALQCEPL 420	
DRFTLPDDLL KDISAYRIAW DRNEEELRAR SVEALPEAQO AEVRALDGVS KETARTQLCA 480	
DFGLDPKRLP WDKMSSNTTF ISEALLNSV SRDQVFPTA PKKGAKKKAP VEVMRKDRTW 540	
ARAYKPRLSV EAQKLKNEAL WALKRTSPE LKLSRRKEEL CRRSINYVIE KTRRRTQCQI 600	
VIPVIEDLNV RFFHGSGKRL PGWDNFFTAK KENRWFQIQL HKAESDLRTH RSFYVFEVRP 660	
ERTSITCPKC GHCEVGNRDC EAFQCLSCGK TCNAADLVAT HNLTVQVALTG KTMPKREEP 720	
DAQGTAPARK TKKASKSKAP PAEREDQTPA QEPSQTS 757	
 SEQ ID NO: 397 moltype = AA length = 765	
FEATURE Location/Qualifiers	
source 1..765	
mol_type = protein	
organism = unidentified	
 SEQUENCE: 397	
MYILEMDLK SEPSLLAKLL RDRFPKGKYL PKYWKLAEKK RLTGGEAAC EYMADKQLDS 60	
PPPNFRPPAR CVILAKSRPF EDWPVHRVAS KAQSFGVIGLS EQQGFAALRAA PPSTADARRD 120	
WLRSHGASED DLMALEAQSL ETIMGNAISL HGGVLKKIDN ANVKAAKRLS GRNEARLINKG 180	
LQELPPQEG SAYGADGLLV NPPGLNLNYY CRKSCCPKPV KNTARFVGHY PGYLRDSDSI 240	
LISGTMDRLT IIEGMGPCHIP AWQREQGLVK PGGRRRRLSG SESNMRQKVD PSTGPRRSTR 300	
SGTVNRSNQR TGRNGDPLLW EIRMKEDWVL LDARGLRLRN RWRESKRLS CDHEDLSLSG 360	
LIALFSGDPV IDPVVRNEVVF LYGEGIIPVR STKEPVGTRQS KKLLERQASM GPLTLISCDL 420	
GQTNLIAGRA SAISLTHGSL GVRSSVRIEL DPEIIKSFER LRKDADRLET EILTAAKETL 480	
SDEQRGEVNS HEKDSPQTAK ASLCRELGLH PPSLPWGQMG PSTTFIADM1 ISHGRDDDAF 540	
LSHGEFTPLE KRKKFDKRCF LESRPLLSE TRKALNESLW EVKRTSSEYA RLSQRKKE 600	
RRAVNFVVEI SRRKGTLSNV IVNIEDLNRV IFHGGGKQAP GDWGFRRPKS ENRWFQIAH 660	
KAFSDLAAHH GIPVIESDPQ RTSMTCPEGC HCDSKNRNGV RFLCKGCGAS MDADFDAACR 720	
NLERVALTGU PMPKPSTSCE RLLSATTGKV CSDHSLSHDA IEKAS 765	
 SEQ ID NO: 398 moltype = AA length = 766	

-continued

---

FEATURE	Location/Qualifiers
source	1..766
	mol_type = protein
	organism = unidentified
SEQUENCE: 398	
MEKEITELTK IRREFPNKKP SSTDMKKAGK LLKAEGPDAV RDFLNSCQEIGDFKPPVKT 60	
NIVSISRPFE EWPVSMVGRA IQEYYSFLTK EELESVHPGT SSEDHKSFN ITGLSNYYNT 120	
SVQGLNLIFK NAKAIYDGTL VKANNKNKKL EKKFNEINHK RSLEGLPIIT PDFEPPFDEN 180	
GHLNNPPGIN RNIYGYQGCA AKVFVPSKHK MVSLPKEYEG YNRDPNLSLA GFRNRLEIPE 240	
GEPGHVPWFQ RMDIPEGQIG HVNKIQRNFN VHGNNSGKVK FSDKTGRVKR YHHSKYKDAT 300	
KPYKFLEESK KVSAALDSILA IITGDDWVV FDIRGLYRNV FYRELAQKGL TAVQLDLFT 360	
GDPVIDPKKG VVTFSYKEGV VPVFSQKIVP RFKSRTDTEK LTSQGPVALV SVDLQNEPV 420	
AARVCSLKNI NDKITLDNSC RISFLDDYKK QIKDYRDSL ELEIKIRLEA INSLETNQQV 480	
EIRDLLDVSFA DRAKANTVDM FDIDPDLNISW DSMSDARVST QISDLYLKNG GDESRVYFEI 540	
NNKRIKRSDY NISQLVRPLK SDSTRKLNLN SIWKLRKTSE EYLKLSKRKL ELSRAVVNYT 600	
IROSKLLESGI NDIVIILEDL DFKKKFFNGRG IRDIGWDNFV SSRKENRWFI PAFHKAFSEL 660	
SSNRGLCVIE VNPWTSATC PDCGFCSKEN RDGINFTCRK CGVSYHADID VATLNIARVA 720	
VLGKPMMSGPA DRERLGDTKK PRVARSRKTM KRKDINSNSTV EAMVTA 766	
SEQ ID NO: 399	moltype = AA length = 812
FEATURE	Location/Qualifiers
source	1..812
	mol_type = protein
	organism = unidentified
SEQUENCE: 399	
MDMLDTETNY ATETPAQQQD YSPKPPKKAQ RAPKGFSKKA RPEKKPPKPI TLFTQKHFG 60	
VRFPLKRVIRD ASKILKLSSES RTITFLEQAI ERDGSAPPDV TPPVHNTIMA VTRPFEEWPE 120	
VILSKALQKH CYALTTKKIKI KTWPKGPK KCLAAWSART KIPLIPGQVQ ATNGLFDRIG 180	
SIYDGVEKKV TNRNANKKLE YDEAIKEGRN PAVPEYETAY NIDGTLINKP GYNPNLYITQ 240	
SRTPRLITEA DRPLVEKILW QMVEKKTQSR NQARRARLEK AAHLQGLPVP KFVPEKVDRS 300	
QKIEIRIIDL DLDKIEPYMPQ DRMAIKASRD GHVPYWQRPF LSKRRNRVVR AGWGKQVSSI 360	
QAWLTGALLV IVRLGNEAFL ADIRGALRNA QWRKLLKPDA TYQSLFNLFV GDPVVNRTTN 420	
HILT MAYREGV VNIVKSRFSK GRQTREHLLT LLGQGKTVAG VSFDLGQKHA AGLLAAHFGL 480	
GEDGNPVPFTP IQACFLPQRY LDSLTNYNR YDALTLDMRR QSLLALTPAQ QQEFADAQRD 540	
PGGQAKRACC LKLNLPNDEI RWDLVSGIST MISDLYIERG GDPRDVHQVQ ETKPKGKRKS 600	
EIRILKIRDG KWADYDFRPKI ADETRKAQRE QLWKLQKASS EFERLRSYKI NIARAIANWA 660	
LQWGRELSGC DIVIPVLEDL NVGSKFFDGK GKWLWGNDNR FTPKKENRWF IKVHLHKVAE 720	
LAPHRGVPVY EVMPHRTSMT CPACHYCHPT NREGDRFECQ SCHVVKNTRD DVAPYNILRV 780	
AVEGKTLDRW QAEKKPQAEP DRPMILIDNQ ES 812	
SEQ ID NO: 400	moltype = AA length = 812
FEATURE	Location/Qualifiers
source	1..812
	mol_type = protein
	organism = unidentified
SEQUENCE: 400	
MDMLDTETNY ATETPAQQQD YSPKPPKKAQ RAPKGFSKKA RPEKKPPKPI TLFTQKHFG 60	
VRFPLKRVIRD ASKILKLSSES RTITFLEQAI ERDGSAPPDV TPPVHNTIMA VTRPFEEWPE 120	
VILSKALQKH CYALTTKKIKI KTWPKGPK KCLAAWSART KIPLIPGQVQ ATNGLFDRIG 180	
SIYDGVEKKV TNRNANKKLE YDEAIKEGRN PAVPEYETAY NIDGTLINKP GYNPNLYITQ 240	
SRTPRLITEA DRPLVEKILW QMVEKKTQSR NQARRARLEK AAHLQGLPVP KFVPEKVDRS 300	
QKIEIRIIDL DLDKIEPYMPQ DRMAIKASRD GHVPYWQRPF LSKRRNRVVR AGWGKQVSSI 360	
QAWLTGALLV IVRLGNEAFL ADIRGALRNA QWRKLLKPDA TYQSLFNLFV GDPVVNRTTN 420	
HILT MAYREGV VDIVKSRFSK GRQTREHLLT LLGQGKTVAG VSFDLGQKHA AGLLAAHFGL 480	
GEDGNPVPFTP IQACFLPQRY LDSLTNYNR YDALTLDMRR QSLLALTPAQ QQEFADAQRD 540	
PGGQAKRACC LKLNLPNDEI RWDLVSGIST MISDLYIERG GDPRDVHQVQ ETKPKGKRKS 600	
EIRILKIRDG KWADYDFRPKI ADETRKAQRE QLWKLQKASS EFERLRSYKI NIARAIANWA 660	
LQWGRELSGC DIVIPVLEDL NVGSKFFDGK GKWLWGNDNR FTPKKENRWF IKVHLHKVAE 720	
LAPHRGVPVY EVMPHRTSMT CPACHYCHPT NREGDRFECQ SCHVVKNTRD DVAPYNILRV 780	
AVEGKTLDRW QAEKKPQAEP DRPMILIDNQ ES 812	
SEQ ID NO: 401	moltype = AA length = 793
FEATURE	Location/Qualifiers
source	1..793
	mol_type = protein
	organism = unidentified
SEQUENCE: 401	
MSSLPTPLEL LKQKHADLFK GLQFSSKDNK MAGKVLKKDG EEAAL AFLSE RGVS RGE LPN 60	
FRP PAKTLV V AQS RPFEEFP IYRVSEAIQL YYV SLSV KEL ETVPSGSSTK KEH QRF FQDS 120	
SVP DFGYTSV QGLN KI FGLA RGI YLG VITR GEN QLQ KAKS KHE ALN KRR ASGE AE TEF D 180	
PTPY EYMTPE RKLAKPVGVN HSIMCYVDIS VDEF DFRNP D GIVLP SEYAG YCRE INTAIE 240	
KGT VDRLGHL KGGPGYI PGH QRKE STTEGP KIN FRK GRI R SYT ALYAKR DS RRV RQG KGL 300	
ALPS YRH HMM RL NSNA ESAI LAVI FFG KDW VVFDL RGLL R NVR WRN LFVD GST PSL TLL GM 360	
FGDP VIDPKR GVVAFCYKEQ IVPV VS KSI T KMV KAPE LLL KLYLK SEDPL VL VAI DL GQ T 420	
NPV GVG VY RV MNAS LDYEV V TRFA LESELL REIES YR QRT NAFE AQIR AE TFD AMT SEE Q 480	
EEIT RVRA FS ASKA ENV CH RFG MPV DAVD WAT MG SNTIH IAKW VMR HG D PSL VEV LEY R 540	

---

---

-continued

---

KDNEIKLDKN	GVPKKVLT	KRIANLTSIR	LRFSQETSKH	YNDTMWELLR	KHPVYQKLSK	600
SKADFSRRVV	NSIIIRRVNHL	VPRARIVFII	EDLKNLGKF	HGSCKRELGW	DSYFEPKSEN	660
RWFICVLHKA	FSETGKHGY	YIIIECWPWT	SCTCPKCSCC	DSENRHGEVF	RCLACGTYTCN	720
TDFGTAPDNL	VKIATTGKGL	PGPKRKCKGS	SKGKNPKIAR	SSETGVSVTE	SGAPKVKKSS	780
PTQTSQSSSQ	SAP					793

---

SEQ ID NO: 402	moltype = AA	length = 441				
FEATURE	Location/Qualifiers					
source	1..441					
	mol_type = protein					
	organism = unidentified					
SEQUENCE: 402						
MNKIEKEKTP	LAKLNMENFA	GLRFPFAIIK	QAGKKLLKEG	ELKTIEYMTG	KGSIEPLPNF	60
KPPVKCLIVA	KRRDLKYFPI	CKASCEIQSY	VYSLNLYKDFM	DYFSTPMTSQ	KQHEEFFKKS	120
GLNIEYQNVNA	GLNLIFNNVVA	NTYNGVILK	KNRNEKLKK	AIKNNYEFEFEE	IKTFFNDGCL	180
INKPGINNVNA	YCFQSISPKI	LKNITHLPK	YNDYDCSVDR	NIIQKVVSRL	DIPESQPGHV	240
PEWQRKLPEF	NNNNPDRRRR	KWYSNGRNIS	KGYSVDQVNQ	AKIEDSLLAQ	IKIGEDWIIL	300
DIRGLLRLDN	RRELISYKNA	LTIKDVLGF	SDYPIIDIKK	NLVTFCYKEG	VIQVVSQKSI	360
GNKKSQQLLE	KLIENKPIAL	VSIDLQQTNP	VSVKISKLNK	INNKISIESF	TYRFLNEEIL	420
KEIEKYRKDY	DKLELKLINE	A				441

---

SEQ ID NO: 403	moltype = AA	length = 812				
FEATURE	Location/Qualifiers					
source	1..812					
	mol_type = protein					
	organism = unidentified					
SEQUENCE: 403						
MDMLDTETNY	ATETPSQQD	YSPKPPKKDR	RAPKGFSKKA	RPEKKPPKPI	TLFTQKHFSG	60
VRFLKRVIRD	ASKILKLSSES	RTITFLEQAI	ERDGSAAPPDV	TPPVHNTIMA	VTRPFEWPE	120
VILSKALQKH	CYALTKKIKI	KTWPKGPGK	KCLAAWSART	KIPLIPGQVQ	ATNGLFDRIG	180
SIYDGVEKKV	TNRNANKKLE	YDEAIKEGRN	PAVPEYETAY	NIDGTLINKP	GYNPNLYITQ	240
SRTPRLITEA	DRPLVEKILW	QMVEKKTQSR	NQARRARLEK	AAHLQGLPVP	KFVPKEVDRS	300
QKIEIRIIDP	LDKIEPYMPQ	DRMAIKASQD	GHVPYWQRPF	LSKRRNRRVR	AGWGKQVSSI	360
QAWLTGALLV	IVRLGNEAFL	ADIRGALRNA	QWRKLLKPD	TYQSLFNLF	GDPVVNTRTN	420
HILTMYREGV	VDIVKSRSFK	GRQTREHLLT	LLGQGKTVAG	VSFDLGQKHA	AGLLAAHFGL	480
GEDGNPVPFTP	IQACFLPQRY	LDSLTNYRNR	YDALTLDMR	QSLLALTPAQ	QQEFDADAQRD	540
PGQQAKRACC	LKLNLPNDEI	RWDLVSGIST	MISDLYIERG	GDPRDVHQQV	ETKPKGKRKS	600
EIRILKIRDG	KWAYDPRPKI	ADETRKAQRE	QLWKLQKASS	EFERLSRYKI	NIARAIANWA	660
LQWGRELSGC	DIVIPVLEDL	NVGSKFNDGK	GKWLWGWDNR	FTPCKKENRWF	IKVLUHKAVAE	720
LAPHRGVPVY	EVMPHRSTMT	CPACHYCHPT	NREGDRFECQ	SCHVVKNTDR	DVAPYNILRV	780
AVEGKTLDRW	QAEKKPQAEP	DRPMILIDNQ	ES			812

---

SEQ ID NO: 404	moltype = AA	length = 812				
FEATURE	Location/Qualifiers					
source	1..812					
	mol_type = protein					
	organism = unidentified					
SEQUENCE: 404						
MDMLDTETNY	ATETPSQQD	YSPKPPKKDR	RAPKGFSKKA	RPEKKPPKPI	TLFTQKHFSG	60
VRFLKRVIRD	ASKILKLSSES	RTITFLEQAI	ERDGSAAPPDV	TPPVHNTIMA	VTRPFEWPE	120
VILSKALQKH	CYALTKKIKI	KTWPKGPGK	KCLAAWSART	KIPLIPGQVQ	ATNGLFDRIG	180
SIYDGVEKKV	TNRNANKKLE	YDEAIKEGRN	PAVPEYETAY	NIDGTLINKP	GYNPNLYITQ	240
SRTPRLITEA	DRPLVEKILW	QMVEKKTQSR	NQARRARLEK	AAHLQGLPVP	KFVPKEVDRS	300
QKIEIRIIDP	LDKIEPYMPQ	DRMAIKASQD	GHVPYWQRPF	LSKRRNRRVR	AGWGKQVSSI	360
QAWLTGALLV	IVRLGNEAFL	ADIRGALRNA	QWRKLLKPD	TYQSLFNLF	GDPVVNTRTN	420
HILTMYREGV	VIVIKSRSFK	GRQTREHLLT	LLGQGKTVAG	VSFDLGQKHA	AGLLAAHFGL	480
GEDGNPVPFTP	IQACFLPQRY	LDSLTNYRNR	YDALTLDMR	QSLLALTPAQ	QQEFDADAQRD	540
PGQQAKRACC	LKLNLPNDEI	RWDLVSGIST	MISDLYIERG	GDPRDVHQQV	ETKPKGKRKS	600
EIRILKIRDG	KWAYDPRPKI	ADETRKAQRE	QLWKLQKASS	EFERLSRYKI	NIARAIANWA	660
LQWGRELSGC	DIVIPVLEDL	NVGSKFNDGK	GKWLWGWDNR	FTPCKKENRWF	IKVLUHKAVAE	720
LAPHRGVPVY	EVMPHRSTMT	CPACHYCHPT	NREGDRFECQ	SCHVVKNTDR	DVAPYNILRV	780
AVEGKTLDRW	QAEKKPQAEP	DRPMILIDNQ	ES			812

---

SEQ ID NO: 405	moltype = AA	length = 761				
FEATURE	Location/Qualifiers					
source	1..761					
	mol_type = protein					
	organism = unidentified					
SEQUENCE: 405						
MSNKTTPPSL	LSLLLRAHFP	GLKFESQDYK	IAGKLLRDGG	PEAVISYLTG	KGQAKLKDVK	60
PPAKAFVIAQ	SRPFIEWDLV	RVSQIQEKI	FGIPATKGRP	KQDGLSETAF	NEAVASLEVD	120
GKSCLNEETR	AAFYEVLGLD	APSLHQAQAN	ALIKSAISIR	EGVLKKVENR	NEKNLSKTKR	180
RKAGEEATF	VEEKAHDERG	YLIHPPGPNQ	TIPGYQAVVI	KSCPDSFIGL	PSGCLAKESA	240
EALTDYLPHD	RMTIPKGQPG	YVPEWQHPLL	NRRKNRRRRD	WYSASLNPK	ATCSKRSGTP	300
NRKNSRTDQI	QSGRFKGA	VLMRFQDEWV	IIDIRGLLRRN	ARYRKLLKEK	STIPDILSLF	360
TGDPSIDMRQ	GVCTFIYKAG	QACSAKMVK	KNAPEILSEL	TKSGPVVLVS	IDLGQTNPIA	420

-continued

---

AKVSRVTQLS DGQLSHETLL RELLSNDSSD GKEIARYRVA SDRLRDKLAN LAVERLSPHEH	480
KSEILRKAND TPALCKARVC AALGLNPEMI AWDKMTPYTF FLATAYLEKG GDRKVATLKP	540
KNRPEMLRRD IKFKGTEGVR IEVSPEAAEA YREAQWDLQR TSPEYLRSLST WKQELTKRIL	600
NQLRHKAAKS SQCEVVVMAT EDLNKIMMHG NGKWDGGWD AFFIKKREN R WFMQAFHKSL	660
TELGAHKGPV TIEVTPHRTS ITCTKCGHCD KANRDGERFA CQKCGFVAHA DLEIATDNIE	720
RVALTGKMPM KPESERSGDA KKSVGARKAA FKPEEDEAEEA E	761

SEQ ID NO: 406 moltype = AA length = 717  
 FEATURE Location/Qualifiers  
 source 1..717  
 mol\_type = protein  
 organism = unidentified

SEQUENCE: 406  
 MIKPTVSQFL TPFGKLIRNH SRTAGLKLKN EGEEACKKFV RENEIPKDEC PNFQGGPAIA 60  
 NIIAKSREFT EWEIYQSSLA IQEVIFTLPK DKLPEPILKE EWRAQWLSEH GLDTVPYKEA 120  
 AGLNLIKNA VNTYKGVQVK VDNKNKNNLA KINRKNEIAK LNGEQEISFE EIKAFDDKGY 180  
 LLQKPSPNKS IYCYQSVPK PFITSKYHNV NLPEEYIGYY RKSNEPIVSP YQFDRLRIP 240  
 GEPGVVPKWQ YTFLSKKENK RRKLSKRIKN VSPILGICI KKDWCVFDMR GLLRTNHWKK 300  
 YHKPTDSIND IDTAKANWRV RYKMENGIVK FELDYFTGDPV 360  
 GSCKLATVDV QONNPVAIGL TKTTLISRHP PIDFCNKITA YRERYDKLES 420  
 SIKLDAIKQL TSEQKIEVDN YNNNNFTPQNT KQIVCSKLNI NPNDLPWDKM ISGTHFISEK 480  
 AQVSNKSEIY FTSTDKGKTF DVMKSDYKWF QDYKPKLSKE VRDALSDIEW RLRRESLEFN 540  
 KLSKSREQDA RQLANWISSM CDVIGLKKRN VKNNNFFGGSGK KREPWNDNFY KPKKENRWI 600  
 NAIIHKALTEL SQNKGKRVIL LPAMRTSITC PKCCKYCDSKN RNGEKFNCLK CGIELNADID 660  
 VATENLATVA ITAQSMPKPT CERSGDAKXP VRARKAKAPE FHDKLAPSYT VVLREAV 717

SEQ ID NO: 407 moltype = AA length = 781  
 FEATURE Location/Qualifiers  
 source 1..781  
 mol\_type = protein  
 organism = unidentified

SEQUENCE: 407  
 MRQPAEKTAQ QVFRQEVIGT QKLSSGGDAKT AGRLYKQGKM EAAREWLKG ARDDVPPNFQ 60  
 PPAKCLVVAV SHPFEEWDIS KTNHDVQAYI YAQPLQAEGH LNGLSEKWED TSADQHKLWF 120  
 EKTGVPDRGL PVQAINKIAK AAVNRAFGVV RKVENRNEKRR RSRDNRIAEH NRREGLTEVV 180  
 REAPEVATNA DGFLLHPPGI DPSILSYASV SPVPYNSSKH SFVRLPEEYQ AYNVEPDAPI 240  
 PQFVVEDRFA IPPGQPGYVP EWQRKLCSTN KHRRMRQWSN QDYKPKAGR AKPLEFQAH 300  
 TRERAKGALL VVMRIKEDWV VFDFVRGLLRN VEWRKVLSSEE AREKLTALKL LDLFTGDPVI 360  
 DTKRGIVTFL YKAEITKILS KRTVTKNAR DLLRLRTEPG EDGLRREVGL VAVDLGQTHP 420  
 IAAAIYRIGR TSAGALESTV LHRQQLREDQ KEKLKEYRKR HTALDSRLRK EAFETLSVEQ 480  
 QEKEIVTVSGS GAQITKDKVC NYLGVDPSTL PWEKMGSYTH FISDDFLRNG GDPNVHFDR 540  
 QPKGGKVSKK SRQIKRSDSQ WVGRMRPRRLS QETAKARMEA DWAAQNEENEE YKRLRSKQE 600  
 LARWCNTLQ QNTRCITQCD EIVVVIEDLN VKSLHGKGAR EPGWDNFFT P KTENRWFQI 660  
 LHKTFSELPK HRGEHVIEGC PLRTSITCPA CSYCDKNSRN GEKFVCVACG ATFHADFEVA 720  
 TYNLVRLATT GMPPMPKSLER QGGGEKAGGA RKARKKAKQV EKIVVQANAN VTMNGASLHS 780  
 P 781

SEQ ID NO: 408 moltype = AA length = 793  
 FEATURE Location/Qualifiers  
 source 1..793  
 mol\_type = protein  
 organism = unidentified

SEQUENCE: 408  
 MSSLPTPLEL LKQKHADLFK GLQFSSKDNK MAGKVLKKDG EEAALALFSE RGVSRGELPN 60  
 FRPPAKTLVV AQSRRPFEFP IYRVSEAIQL YVYSLSVKEL ETVPSSGSTK KEHQRFQFQDS 120  
 SVPDFGYMTSV QGLNKGFLA RGIYLGVITR GENQLQKAKS KHEALNKRR ASGEAETEFD 180  
 PTPYEYMTPE RKLAKPGVN HSIMCYVDIS VDEFDFRNPD GIVLPSEYAG YCREINTAIE 240  
 KGTVDRLGLH KGGPGYIPGH QRKESTTEGP KINFRKGRIR RSYTALYAKR DSRRVRQGKL 300  
 ALPSYRHHMM RLNSNAESA LAVIFFGKD VVFDLRLGLLR NVRWRNLFVFD GSTPSTLLGM 360  
 FGDPVIDPKR GVAVFCYKEQ IVPVVSKSIT KMVKAPELLN KLYKLKSEDPL VLVAIDLQQT 420  
 NPVGVGVYRV MNASLDYEVV TRFALESELL REIESYRQRT NAFEAQJRAE TFDAMTSEEQ 480  
 EEITRVRRAFS ASKAKENVCH RFGMPVDAVD WATMGSNTIH IAKWVMRHGD PSLVEVLEYR 540  
 KDNIEIKLDKN GPVKVVKLTD KRIANLTSIR LRFSQETSKH YNDTMWELLR KHPVYQKLSK 600  
 SKADFSRVRV NSIIRRVNHL VPRARIVFII EDLKNLKGKFV HSGSKRELGW DSYFEPKSEN 660  
 RWFIQVLHKA PSETGKHKGY YIIECWPNWT SCTCPKCSC DSENRHGEVF RCLACGYTCN 720  
 TDFTGTAPEPDNL VKIATTGKGL PGPKRCKGS SKGKNPKIAR SSETGVSUTE SGAPKVKSS 780  
 PTQTSQSSSQ SAP 793

SEQ ID NO: 409 moltype = AA length = 717  
 FEATURE Location/Qualifiers  
 source 1..717  
 mol\_type = protein  
 organism = unidentified

SEQUENCE: 409  
 MIKPTVSQFL TPFGKLIRNH SRTAGLKLKN EGEEACKKFV RENEIPKDEC PNFQGGPAIA 60  
 NIIAKSREFT EWEIYQSSLA IQEVIFTLPK DKLPEPILKE EWRAQWLSEH GLDTVPYKEA 120

-continued

---

AGLNLIKNA	VNTYKGQVK	VDNKNKNL	KINRKNEIAK	LNGEQEISFE	EIKAFDDKGY	180
LIQKPSPKS	IYCYQSVP	PFITSKYHNV	NLPEEYIGY	RKSNEPIVSP	YQFDRLRIPI	240
GEFGYVPWKQ	YTFLSKKENK	VSPILGICCI	KKDWCVFDMR	GLLRTNHWKK		300
YHKPTDSIND	LFDFYFTGDPV	IDTKANVVF	RYKMENGIVN	YKPVREKGK	ELLENICDQN	360
GSCKLATVDV	GQNNPVAIGL	FELKKVNGL	TKTLISRHT	PIDFCNKITA	YRERYDKLES	420
SIKLDAIKQL	TSEQKIEVDN	YNNNFTPQNT	KQIVCSKLNI	NPNDLPWDKM	ISGTHFISEK	480
AQVSNKSEIY	FTSTDKGKTK	DVMKSDYKWF	QDYPKPLSKE	VRDALSDIEW	RLLRRESLEFN	540
KLSSKSRQDQA	RQLANWISSM	CDVIGIENLV	KKNNFFGGSG	KREPWNDNFY	KPKKENRWI	600
NAIHKALTEL	SQNKGKRVIL	LPAMRTSITC	PKCKYCDSKN	RNGEKFNLCK	CGIELNADID	660
VATENLATVA	ITAQSMPKPT	CERSGDAKJP	VRARKAKAPE	FHDKLAPSYT	VVLREAV	717

SEQ ID NO: 410	moltype = AA	length = 761				
FEATURE	Location/Qualifiers					
source	1..761					
	mol_type = protein					
	organism = unidentified					
SEQUENCE: 410						
MSNKTTPPPS	LSLLLRAHFP	GLKFESQDYK	IAGKQLRDGG	PEAVISYLTG	KGQAKLKDVK	60
PPAKAFVIAQ	SRPFIEWDLV	RVSRIQZEKI	FGIPATKGRP	KQDGLSETAF	NEAVASLEV	120
GKSCLNEETR	AAFYEVGLD	APSLHAQAOQN	ALIKSAISIR	EGVLKKVENR	NEKNLNSKTKR	180
RKEAGEEATF	VEEKAHDERG	YLIHPPGVNQ	TIPGYQAVVI	KSCPSDFIGL	PSGCLAKESA	240
EALTDYLPHD	RMTIPKGQPG	YVPEWQHPLL	NRRKNRRRRD	WYSASLNKPK	ATCSKRSGTP	300
NRKNSRTDQI	QSGRFPKAIP	VLMRFDEWW	IIDIRGLLRN	ARYRKLKEK	STIPDLLSLF	360
TGDPSIDMRQ	GVCTFIYKAG	QACSAKMVKT	KNAPEILSEL	TKSGPVVLLS	IDLGQTNPRIA	420
AKVSRVTQLS	DGQLSHETLL	RELLSNDSSD	GKEIARYRVA	SDRLRDKLAN	LAVERLSPHE	480
KSEILRAKND	TPALCKARVC	AALGLNPENMI	AWDKMTPYTE	FLATAYLEKG	GDRKVATLKP	540
KNRPEMPLRRD	IKFKGTEGV	IEVSPEAAEA	YREAQWDLQR	TSPEYRLST	WKQELTKRIL	600
NQLRHKAAKS	SQCEVVVMMAF	EDLNKMMHG	NGKWADGGWD	AFFIKKREN	WFMQAFHKSL	660
TELGAHKGPV	TIEVTPHRTS	ITCTKCGHCD	KANRDGERFA	CQKCGFVAHA	DLEIATDNIE	720
RVALTGKPM	KPESERSGDA	KKSVGARKAA	FKPPEDEAEE		E	761

SEQ ID NO: 411	moltype = AA	length = 765				
FEATURE	Location/Qualifiers					
source	1..765					
	mol_type = protein					
	organism = unidentified					
SEQUENCE: 411						
MSYLEMDLK	SEPSLLAKLL	RDRFPGKYWL	PKYWKLAEKK	RLTGGEAAC	EYMADKQLDS	60
PPPNFRPPAR	CVILAKSRPF	EDWPVHRVA	KAQSFVIGLS	EQGFAALRRAA	PPSTADARRD	120
WLRSHGASED	DLMALEAQLL	ETIMGNAISL	HGGVLKKIDN	ANVKAARLS	GRNEARLNKG	180
LQLPPRQEKG	SAYGADGLL	NPPGLNLNIY	CRKSCCPKPV	KNTARFVGHY	PGYLRDSDSI	240
LISGTMDRLT	IIEGMPGHIP	AWQREQGLVK	PGGRRRRLSG	SESNMRQKVD	PSTGPRRSTR	300
SGTVNRSNQR	TGRNGDPLL	EIRMKEDWV	LDARGLLERN	RWRESKRGLS	CDHEDLSSLG	360
LLALFSGDPV	IDPVRNEVVF	LYGEGIIPV	STKPGVTRQS	KKLLEHQASM	GPLTLISCDL	420
GQTNLIAGRA	SAISLTHGSL	GVRSSVRIEL	DPEIIKSFER	LRKDADRLET	EILTAAKETL	480
SDEQRGEVNS	HEKDSPQTAK	ASLCRELGLH	PPSLPWGQMG	PSTTFIADML	ISHGRDDDAF	540
LSHGEFTTLE	KRKFDKRF	LESRPLLSE	TRKALNESLW	EVKRTSSEYYA	RLSQRKKEMA	600
RRAVNFVVEI	SRKRTGLSNV	IVNIEDLNVR	IFHGGGKQAP	GWDGFFRPKS	ENRWFIQAIH	660
KAFSDLAAHH	GIPVIESDPQ	RTSMTCPCEG	HCDSKNRNGV	RFLCKCCGAS	MDADPDAACR	720
NLERVALTGK	PMPKPSTSCE	RLLSATTGKV	CSDHSLSHDA	IEKAS		765

SEQ ID NO: 412	moltype = AA	length = 766				
FEATURE	Location/Qualifiers					
source	1..766					
	mol_type = protein					
	organism = unidentified					
SEQUENCE: 412						
MEKEITELT	IRREFPNKKF	SSTMKAGK	LLKAEGPDAV	RDFLNQCQE	IGDFKPPVKT	60
NIVSISRPF	EWPVSMVGRA	IQEYFFSLTK	EELESVHPGT	SSEDHKSFFN	ITGLSNYYNT	120
SVQGLNLIFK	NAKAIYDGTI	EKKFNEINHK	RSLEGLPIT	PDPEEPFDEN		180
GHNNNPNGIN	RNIYGYQGC	AKVFVPSKH	MVSLPKKEYG	YNRDPNLSLA	GFRNRLIEI	240
GEPGHVPWFQ	RMDIPEQGQ	HVNKIQRFNF	VHGKNSGKV	FSDKTRVVR	YHHSKYKDAT	300
KPYKFLEESK	KVSALDSILA	IITIGDDWV	FDIRGLYRN	FYRELAQKGL	TAQQLLDLFT	360
GDPVIDPKG	VVTFSYKEGV	VPVFSQKIVP	RFKSRDTLEK	LTSQGPVALL	SVDLGQNEPV	420
AARVCSLKN	NDKITLDNSC	RISFLDDYKK	QIKDYRDSLV	ELEIKIRLEA	INSLETNQQV	480
EIRDLDVFA	DRAKANTVDM	FDIDPNLISW	DSMSDARVST	QISDLYLKNG	GDESRVYFEI	540
NNKRIKRSDY	NISQLVRPKL	SDSTRKLNND	SIWKLKRTSE	EYLKLSKRL	ELSRAVVNYT	600
IROSQKLSSGI	NDIVIILEDL	DVKKKFNGRG	IRDIGWDNFF	SSRKENRWF	PAFKHTFSEL	660
SSNRGLCVIE	VNPATWSATC	PDCGFCSKEN	RDGINFCTR	CGVSYHADID	VATLNIARVA	720
VLGKPMMSGPA	DRERLGDTKK	PRVARSRKTM	KRKDISNSTV	EAMVTA		766

SEQ ID NO: 413	moltype = AA	length = 870
FEATURE	Location/Qualifiers	
source	1..870	
	mol_type = protein	
	organism = unidentified	

---

-continued

---

**SEQUENCE: 413**

MLVRTSTLVQ	DNKNSRSASR	AFLKKPKMPK	NKHIKEPETEL	AKLIRELFPQ	QRFTRAINTQ	60
AGKILKHKG	DEVVEFLKNK	GIDKEQFMDF	RPPTKARIVA	TSGAIEEFSY	LRVSMAIQEC	120
CFGKYKPFKE	KVNGKLVLLET	VGLTKEELDD	FLPKKYYENK	KSRDRFFLKT	GICDYGYTYA	180
QGLNEIFRNT	RAIYEGVFTK	VNNRNKRRE	KKDKNYEERR	SKGLSEEPYD	EDESATDESG	240
HЛИNPPGVNL	NIWTCCEGFCK	GPYVTKLSGT	PGYEVILPKV	FDGYNRDPNE	IISCGITDRF	300
AIPEGEPEGHI	PWHQRLEIPE	GQPGYVPGHQ	RFADTGQNNS	GKANPNKKG	MRKYYGHGTTK	360
YTOPGEQEY	FRKGHREGNK	RRYWEEDFRS	EAHDCLILYVI	HIGDDVVVCD	LRGPLRDAYR	420
RGLVPKEGIT	TQECLCNLFSG	DPIVIDPKHG	VTFCYKNGLV	RAQKTISAGK	KSRELLGALT	480
SQGPIALIGV	DLGQTEPVGA	RAFIVNQARG	SLSLPLTKGS	FLLTAENSSS	WNVFKGEIKA	540
YRKAIDDIAI	RLKKEAVDAA	SVEQQTEIES	YEAFSAEADAK	QLAACEKFVGD	SSFILWEDMT	600
PYHTGPATYY	FAKQFLKKNG	GNKSLIEYIP	YQKKKSKKTP	KAVLRSODYNI	ACCVRPKLLP	660
ETRKALNEAI	RIVQKNSDEY	QRLSKRKLEF	CRRVNVYLVR	KAKKLTGLER	VIIAIEDLKS	720
LEKFFTGSGG	RDNGWNSNFFF	PKKRNWRWFIP	AFHKAFSELAA	PNRGFVVIIEC	NPARTSITDP	780
DCGYCDGDNR	DGKFWCEKKC	GAKHHTDLD	APLNIAIVAV	TGRPMPPKTVS	NKSKRERSGG	840
EKSVGASRKR	NHRKSKANQE	MLDATSSAE				870

**SEQ ID NO: 414**

FEATURE	moltype = AA	length = 830
source	Location/Qualifiers	
	1..830	
	mol_type = protein	
	organism = unidentified	

**SEQUENCE: 414**

MPKIKKPTEI	SLLRKEVFVD	LHFAKDRMRA	ASLVLKNEGR	EEAIEYLRVN	HEDKPPNFMP	60
PAKPTYVALS	RPLEQWPIAQ	ASIAIQKYIF	GLTKDEFSAT	KKLLYGDKST	PNTESRKRWF	120
EVITGVNPFGY	MSAQGLNAIF	SGALARVEGV	VQVKVENRNKK	RPEKLSEKNQ	LLIEEGQPVK	180
DVVPDTAYHT	PETLQLKAEN	NHVRVEDLGD	MIDRLVHPPG	IHSRIYGYQQ	VPPFAYDPDN	240
PKGIILPKAY	AGYTRKPHDI	IEAMPNRLNI	PEQGAGYIPE	HQRDKLKKGG	RVKRLRTTRV	300
RVDATEVTRA	KAEALNAEKA	RLRGKEAILA	VFQIEEDWAL	IDMRGRLRNV	YMRKLIAAGE	360
LTPTTLLGYF	TETLTLDPDR	TEATFCYHLR	SEGALHAEVY	RHGKNTRELL	LDLTKDNEKI	420
ALVTIDLGQR	NPLAAAIFRV	GRDASGDLTE	NSLEPVSRML	LPQAYLDQIK	AYRDAYDSFR	480
QNIWDTALAS	LTPEQQRQIL	AYEAYTPDDS	KENVLRLLLG	GNVMPDDLPW	EDMTKNTHYI	540
SDRYLADGGD	PSKVWFPVGP	RKRKKNAPPL	KKPPKPRELV	KRSDHNISHL	SEFRPQLLKE	600
TRDAFEKAKI	ETERGHVGYQ	KLSTRKDQLC	KEILNWLEAE	AVRLTRCKTM	VLGLEDLNGP	660
PFNQGKGVHD	GWVSFPFRQKQ	ENRIVWNGFRR	KNALARAHDK	GKYILELWPS	WTSCQCPKCK	720
HVHADNRHGD	DFVCLQCGAR	LHADADEVATW	NLAIVAIQGH	SLPGPVREKS	NDRKKSGSAR	780
KSKKANESGK	VVGAWAAQAT	PKRATSKKET	GTARNPVYNP	LETQASCAPAP		830

**SEQ ID NO: 415**

FEATURE	moltype = AA	length = 790
source	Location/Qualifiers	
	1..790	
	mol_type = protein	
	organism = unidentified	

**SEQUENCE: 415**

MTSPQIARL	VETPLAALK	AHHPGKKFRS	DYLKKAGKIL	KDQGVAAAMA	HLDGKDQAEQ	60
PNFKPPAKCR	IVARSREFSE	WPIVKASVEI	QKYIYGLTLE	ERKACDPGKS	SASHKAWFAK	120
TGVNTFGYSS	VQGFNLIFGH	TLGRYDGVLV	KTEENLNKKRA	EKNERFRAKA	LAEGRAEPVC	180
PPLVTATNDT	QODVTLLEDG	VVRPGQLLQP	PGINPNIYAY	QQVSPKAYVP	GIIELPEEFQ	240
GYSRDPNAVI	LPLVPRDRRLS	IPKGQPGYVP	EPHREGLTGR	KDRRMRYYE	TERGTLKLKP	300
PLTAKGRADK	ANEALLVVVR	IDSDWVVMMDV	RGLLNRNARWR	RLVSKEGITL	NGLLDLFTGD	360
PVLNPDKCSV	SRDTGDPVND	PRHGVVTFCY	KLGVDVCSK	DRPIKGFRTK	EVLERLTSSG	420
TVGMVSIDLG	QTNPVAAAVS	RVTKGQAEQ	LETFTLPDDL	LGKVRAYRAK	TDRMEEGFRR	480
NALRKLTAEQ	QAEITRYNDA	TEQQAKALVC	STYGIGPEEV	PWERMTSNTT	YISDHILDHG	540
GDPDTVFFMA	TKRGQNKPTL	HKRKDCAWQG	KFRPAISVET	RLARQAAEWE	LRRASLEFQK	600
LSVWKTELCR	QAVNYVMERT	KKRTQCDVII	PVIEDLPVPL	FHGSGKRDPG	WANFFVHKRE	660
NRWFIDGLHK	AFSELGKHG	IYVFEVCPQR	TSITCPKGH	CDPDNRGEK	FVCLSCQATL	720
NADLDVATTN	LVRVALTGKV	MPRSERSGDA	QTPGPARKAR	TGKIKGSKPT	SAPQGATQTD	780
AKAHLSQTGV						790

**SEQ ID NO: 416**

FEATURE	moltype = AA	length = 790
source	Location/Qualifiers	
	1..790	
	mol_type = protein	
	organism = unidentified	

**SEQUENCE: 416**

MTSPQIARL	VETPLAALK	AHHPGKKFRS	DYLKKAGKIL	KDQGVAAAMA	HLDGKDQAEQ	60
PNFKPPAKCR	IVARSREFSE	WPIVKASVEI	QKYIYGLTLE	ERKACDPGKS	SASHKAWFAK	120
TGVNTFGYSS	VQGFNLIFGH	TLGRYDGVLV	KTEENLNKKRA	EKNERFRAKA	LAEGRAEPVC	180
PPLVTATNDT	QODVTLLEDG	VVRPGQLLQP	PGINPNIYAY	QQVSPKAYVP	GIIELPEEFQ	240
GYSRDPNAVI	LPLVPRDRRLS	IPKGQPGYVP	EPHREGLTGR	KDRRMRYYE	TERGTLKLKP	300
PLTAKGRADK	ANEALLVVVR	IDSDWVVMMDV	RGLLNRNARWR	RLVSKEGITL	NGLLDLFTGD	360
PVLNPDKCSV	SRDTGDPVND	PRHGVVTFCY	KLGVDVCSK	DRPIKGFRTK	EVLERLTSSG	420
TVGMVSIDLG	QTNPVAAAVS	RVTKGQAEQ	LETFTLPDDL	LGKVRAYRAK	TDRMEEGFRR	480
NALRKLTAEQ	QAEITRYNDA	TEQQAKALVC	STYGIGPEEV	PWERMTSNTT	YISDHILDHG	540
GDPDTVFFMA	TKRGQNKPTL	HKRKDCAWQG	KFRPAISVET	RLARQAAEWE	LRRASLEFQK	600
LSVWKTELCR	QAVNYVMERT	KKRTQCDVII	PVIEDLPVPL	FHGSGKRDPG	WANFFVHKRE	660

-continued

NRWFIDGLHK	AFSELGKHG	IYVFVCPQR	TSITCPKGH	CDPDNRDGEK	FVCLSCQATL	720
HADLDWATTN	LVRVALTGKV	MPRSERSGDA	QTPGPARKAR	TGKIKGSKPT	SAPQGATQTD	780
AKAHLSTQTV						790

SEQ ID NO: 417 moltype = AA length = 782  
FEATURE Location/Qualifiers  
source 1..782  
mol\_type = protein  
organism = unidentified

SEQ ID NO: 418 moltype = AA length = 774  
FEATURE Location/Qualifiers  
source 1..774  
mol\_type = protein  
organism = unidentified

```

SEQUENCE: 418
VYNPDMDKPKN NIRRIREEHF EGLCFGKDVL TKAGKUYEKD GEEAAIDFLM GKDEEDPPNF 60
KPPAKTTVA QSRPFDQWPI YQVSQAVQER VFAYTEEEFN ASKEALFSGD ISSKSRSDFWF 120
TKTNNSDQGI GAQQLNTILS HAFRSYSGVI KKVENRKKR LKLLSKKNQL KIEEGLEILE 180
FKPDSAFNEN GLLAQPPGIN PNYIYGYQAVT PFVFDPDPNG DVILPKQYEG YSRKDDII 240
KGPSRLDIKP GQPGYVPEHQ RKLNLKKGRV RLYRRTPPKT KALASILAVL QIGKDWWLF 300
MRGLLRSVYMI REAAATPGQIS AKDLDDFTTG CPVLNTRGTE FTFCYKLRS GALHARKYTI 360
KGETRTLTLTS LTSENNITAL VTVDLGQRNP AAIMISLRSR KEELSEKDIQ PVSRRLPDR 420
YLNELKRYRD AYDAFPRQEVE DEATDSLCP EHQBVQVOYEA LTPEKAKNLV HKLFFGHTDP 480
DLPWDMMTSN THYIANLYLE RGDPDSKVFF TRPLKKDSKS KKPRKPTKRT DASISRLPEI 540
RPKMPEDARK AFEKAKWEIY TGHEKFPKLA KRVNQLCREI ANWIEKEAKR LTLCDDTVVG 600
IEDLSSLPPKQ GKGKQETWQ GFRRQKFENR WVIDTLLKAI QNRAHDGKYV VLGLAPYWT 660
QRCPACGFIH KSNRNGDHFK CLKCEALFH ADEVATWNLA LVAVLGKIGT NPDSDKPSQG 720
KKTGTTRKKQ IKGKNGKGET VNVPTPTQEW EDIAIAFFKDET VRNPVYK PTGT 774

```

SEQ ID NO: 419 moltype = AA length = 769  
FEATURE Location/Qualifiers  
source 1..769  
mol\_type = protein  
organism = unidentified

SEQ ID NO: 420 moltype = AA length = 767  
FEATURE Location/Qualifiers  
source 1..767  
mol\_type = protein  
protein\_id = 420

```

organism = unidentified
SEQUENCE: 420
VIKTHPAGR FRKDHQKTAG KKLKHGEEEA CVEYLRMKVS DYPPNPKPPA KGTIVAQSRP 60
FSEWPIVRAS EAIQKYVYGL TVAELDVFS P GTSKPASHAEW FAKTGVENYG YRQVQLNTI 120
FQNTVNRFKG VLKKVNRKEN KSLSKRQEAN RRRVVEALPE VPVTVESATD DEGRLLQPPR 180
VNPSIIGYQG VAPRPTCDLQ GFSGMSVDPKA GYTRDPDAVL EWSLPGRLS IPKGERRGYVP 240
EWORDEERNK EPLREGSSRRO RKWNSYACHK PGKGRSTSQYD PEALKKASAS DALLVISIG 300

```

-continued

---

EDWAIIDVRG	LLRDARRGF	TPEEGLSLNS	LLGLFTEYPV	FDVQRGLITF	TYKLGQDVH	360
SRKTVPTRS	RALLESVAK	EEIALVSVDL	GOTNPASMKV	SRVRAQEGLA	VAEPVHRMFL	420
SDVLLGELOSS	YRKRMDFED	AIRAQAFETM	TPEQQAEITR	VCDVSVEVAR	RRVCEKYSIS	480
PQDVPWGEMT	GHSTFIVDAV	LRKGDESILV	YFKNKEGETL	KFRDLRISRM	EGVRPLTKD	540
TRDALNKAVL	DLKRAHPTFA	KLAKQKLELA	RRCVNFIERE	AKRYTQCERV	VFVIEDLNVG	600
FFHGKGKRDR	GWDAAFTAKK	ENRWIQLAH	KAFSDLGLHR	GSYVIEVTPO	RTSMTCPRCG	660
HCDKGNRNGE	KFVCLQCGAT	LHADLEVATD	NIERVALTGK	AMPKPPVRER	SGDVQKAGTA	720
RKARKPLPKP	QKTEPSVQEG	SSDDGVDKSP	GDASRNVPY	PSDTLSI		767

SEQ ID NO: 421	moltype = AA	length = 763				
FEATURE	Location/Qualifiers					
source	1..763					
	mol_type = protein					
	organism = unidentified					
SEQUENCE: 421						
MAKAKTLAAL	LRELLPGQHL	APHHRWVANK	LLMTSGAAA	FVIGKSVSDP	VRGSFRKDVI	60
TKAGRIFKKD	GPDAAAFLD	GKWEDEPPNF	OPPAKAAIVA	ISRSFDEWPI	VKVSCAIQQY	120
LYALPVQEFE	SSVPEARAQHA	HAAPQDFTGV	DDCNFKSTQG	LNAIFNHGKR	TYEGVLKAO	180
NRNDKKNLRL	ERINAKRAEA	GQAPLVAGPD	ESPDTDAGCL	LHPPGINANI	YCYQQVSPRP	240
YEQSCGIQLP	PEYAGYNRLS	NVAIPPMNR	LDIPQGQPGY	VPEHHHRHGIK	KFGVRKRKYRG	300
VVPGNRNDAD	GKRTRQVLTE	AGAAKARDS	VLA VIRIGD	WTVVDLRGLL	RNAQWRKLV	360
DGGITVQGLL	DLFTGDPVID	PRRGVTFIY	KADSVGTHSE	KVCRGQSKN	LLERLCAMP	420
KSSTRLDCAR	QAVALVSVDL	GQRNPVAAFR	SRVSLAEGQL	QAQLVSAQFL	DDAMVAMIRS	480
YREYEYDRFPES	LVREQAKAAL	SPEQLSEIVR	HEADSAESVK	SCVCVAKFGID	PAGLSDWKMT	540
SGTWRIADHV	QAAGGDVEWF	FFKTCGKGKE	IKTVRRSDFN	VAKFQFRLRS	PETRKDWND	600
IWELKRGNPA	YVSFSKRKSE	FARRVNDLV	HRARRAVRCD	EVVFAIEDLN	ISFFFHGKGQR	660
QMJDWDAFFEV	KQENRWFQIA	LHKAFVERAT	HKGGYVLEVA	PARTSTTCPE	CRHCDPESRR	720
GEQFCCIKCR	HTCHADLEVA	TFNIEQVALT	GVSLPKRLSS	TLL		763

SEQ ID NO: 422	moltype = AA	length = 761				
FEATURE	Location/Qualifiers					
source	1..761					
	mol_type = protein					
	organism = unidentified					
SEQUENCE: 422						
MSKEKTPPSA	YAILKAKHFP	DLDFEKKHKM	MAGRMFKNGA	SEQEVVQYLQ	GKGSESLMDV	60
KPPAKSPILA	QSRPFDEWEM	VRTSRLIQET	IFGIPKRGSI	PKRDGLSETQ	FNELVASLEV	120
GKPKMLNKQT	RAIFYGLGI	KPPTFHAMAQ	NILIDLAINI	RKGVLKKVDN	LNEKNRKVK	180
RIRDAGEQDV	MVPAEVTAHD	DRGVLNHPGG	VNPTIPGYQG	VVIPFPEGFE	GLPSGMTPVD	240
WSHVLVLDQPG	HDRLSIPKGS	GPIYIPEWQRP	LLNRHKGRRH	RSWYANSLINK	PRKSRTTEAK	300
DRQNAGKRTA	LIEAEERLKGV	LPVLMRKFED	WLII DARGLL	RNARYRGVLP	EGSTLGNLID	360
LFSDSPRVDT	RRGICFLYR	KGRAYSTKPV	KRKESKETLL	KLTEKSTIAL	VSIDLGQTNP	420
LTAKLKVRO	VDGCLVAEPV	LRKLIDDNASE	DGKEIARYRV	AHDLLRARIL	EDAIDLGLIY	480
KDEVVRARSD	TPDLCKERC	RFLGLDQSQI	DWDRMTPYD	FIAQAFVAKG	GDPKVVTIKP	540
NGKPKMPFRKD	RSIKNMKGIR	LDSIKEASSA	YREAQWAIQR	ESPDFQRLAV	WQSQLTKRIV	600
NQLVAWAKKC	TQCDTVVLA	EDLNIGMMHG	SGKWANGGW	ALFLHKQENR	WFMQAFHKAL	660
TELSAHHGIP	TIEVLPHRTS	ITCTQCGHCH	PGNRDGERFK	CLKCEFLANT	DLEIATDNE	720
RALVATGLPMP	KGERSSAKRK	PGGTRKTKS	KHSGNNSPLAA	E		761

SEQ ID NO: 423	moltype = AA	length = 746				
FEATURE	Location/Qualifiers					
source	1..746					
	mol_type = protein					
	organism = unidentified					
SEQUENCE: 423						
MEKAGPTSP	SVLIHKNFEG	CRFQIDHLKI	AGRKLAREGE	AAAIEYLLDK	KCEGLPPNFQ	60
PPAKGNVIAQ	SRPFTEWAPY	RASVAIQKYI	YSLSVDERKV	CDPGSSSDSH	EKWFQQTGVQ	120
NYGYTHVQGL	NLIFKHALAR	YDGVLKKVDN	RNEKNRKKEA	RVNSFRREEG	LPEEVFEEEK	180
ATDETGHLLQ	PPGVNHSIYC	YQSVRPKPN	PRKPGGISLP	EAYSGYSLKP	QDELPIGSLD	240
RLSIPPPQPG	YVPEWQRSQL	TTQKRRKLSR	WYSAQKWKPR	TGRTSTFDPD	RLNCARAQGA	300
II LAVVRIHED	WVVFDRVRL	RNALWRELAG	KGLTVRDLLD	FFTGDPPVDT	KRGVVTFTYK	360
LGKVDVHSLR	TVRGKRSKKV	LEDLTLSSDV	GLVTIDLQG	NVLAADYSKV	TRSENGELLA	420
VPLSKSFLPK	HLLHEVTAYR	TSYDQMEEGF	RRKALLTLTE	DQQVEVTLVR	DFSVESSTKT	480
LQLQGVDVTS	LPWEKMSNT	TYISDQLLQQ	GADPASLFFD	GERDGKPCRH	KKKDRTWAYL	540
VRPKVSPETR	KALNEALWAL	KNTSPEFESL	SKRKIQFSR	CNMYYLNEAK	RISGCGQVVF	600
VIEDDLNVRVH	HGRGKRAIGW	DNFFPKPKREN	RWPMQALHKA	ASELAIHGRM	HIIIEACPARS	660
SITCPKCGHC	DPENRCSSDR	EKFCLVKCGA	AFHADLEVAT	FNLRKVALTG	TALPKSIDHS	720
RDGLIPKGAR	NRKLKEPQAN	DEKACA				746

SEQ ID NO: 424	moltype = AA	length = 735				
FEATURE	Location/Qualifiers					
source	1..735					
	mol_type = protein					
	organism = unidentified					
SEQUENCE: 424						
MKEQSPPLSSV	LKSNNFPGKKF	LSADIRVAGR	KLAQLGEAAA	VEYLSPRQRD	SVPNFRPPAF	60

-continued

---

CTVVAKSRPF	EEWPIYKASV	LLQEIQIYGMT	GQEFEERCGS	IPTSLSGLRQ	WASSVGLGAA	120
MEGLHVQGMN	LMVKNAINRY	KGVLVKVENR	NKKLVEANNE	KNSSREERGL	PPLRPPELGS	180
AFGPDPGLVN	PPGIDKSIRL	YQGVSPVPVV	KTTGRPTVHR	LDIPAGEKGH	VPLWQREAGL	240
VEKGPRRRRM	WYSNSNLKRS	RKDRAEASE	ARKADSVVR	VSVKEDVDI	DVRGLLRNVA	300
WRGIERAGES	TEDLLSLFSG	DPVVDPSRDS	VVFLYKEGVV	DVLSSKKVVGA	GKSRKQLEKM	360
VSEGPVALVS	CDLGQTNYVA	ARVSVLDESL	SPVRSFRVDP	REFPSADGSQ	GVVGSLDRIR	420
ADSDRLEAKL	LSEAEASLPE	PVRAEIEFLR	SERPSAVAGR	LCLKLGDPR	SIPWEKGST	480
TSFISEALSA	KGSPALAHDC	APIKDSRFAH	AARGRLSPES	RKALNEALWE	RKSSSREYGV	540
ISRRKSEASR	RMANAVLSES	RRITGLAVVA	VNLEDLNVMV	KFFHGRRKRA	PWGAGFTP	600
MENRWFTIRSI	HKAMCDLSKH	RGITVIESRP	ERTSISCP	GHCDPENRSG	ERFSCKSCGV	660
SIHADFEVAT	RNLERVALTG	KPMPPRRENLH	SPEGATASRK	TRKKPREATA	STFLDLRSVL	720
SSAENEKGSP	AARAG					735

SEQ ID NO: 425            moltype = AA length = 725  
 FEATURE                    Location/Qualifiers  
 source                    1..725  
 mol\_type = protein  
 organism = unidentified

SEQUENCE: 425  
 MLPPSNKIGK SMSLKEFINK RNFKSSIIQK AGKILKKEGE EAVKKYLDNN YVEGYKKRDF 60  
 PITAKCNIVA SNRKIEDFDI SKFSSFIQNY VFNLNKDNF EFSKIKYNRK SFDELYKKIA 120  
 NEIGLEKPNY ENIQGEIAVI RNAINIYNGV LKKVENRNKK IQEKNQSKDP PKLLSAFDDN 180  
 GFLAERPGIN ETIYGYQSVR LRHLDVEKDK DIIVQLPDIY QKYNKKSTDK ISVKKRLNKY 240  
 NVDEYGLKIS KRKRERINKD DAILCVSNFG DDWIIFLDARG LLRQTYRYKL KKKGLCIKDL 300  
 LNLFTGDPPI NPTKTDLKEA LSLSFKDGI NNRTLKVKNY KKCPELISEL IRDKGVAMI 360  
 SIDLGQNTPI SYRLSKPTAN NVAYIENGVI SEDDIVKMKK WREKSDKLEN LIKEEAIASL 420  
 SDDEQREVRV YENDIADNTK KKILEKFNR EEDLDFSNSM NNTYFIRDCL KNKNIDESEF 480  
 TFEKNGKLLD PTDACPAREY KNKLSELTRK KINEKIEWIK KNSKEYHKIS IYKKETIRYI 540  
 VNKLIKQSKE KSECDDIIVN IEKLQIGGGN FGGRGKRDPG WNNFFLPKEE NRWFNACHK 600  
 AFSELAPHKG IIVIESDPAY TSQTCPKCEN CDKENRNGEK FKCKKCNYEA NADIDVATEN 660  
 LEKIAKNGRR LIKNFQDQLGE RLPGAEMPGG ARKRKPSKSL PKNGRGGAGVG SEPELINQSP 720  
 SQVIA                    725

SEQ ID NO: 426            moltype = AA length = 718  
 FEATURE                    Location/Qualifiers  
 source                    1..718  
 mol\_type = protein  
 organism = unidentified

SEQUENCE: 426  
 VPDKKETPLV ALCKKSFPGL RFKKHDSRQA GRILKSKGEG AAVAFLEGKG GTTOPNFKPP 60  
 VKCNIVAMSR PLEEWPIYKA SVVIQKYVYA QSYEEFKATD PGKSEAGLRA WLKATRVDTD 120  
 GYFNVQGLNL IFQONARATY GVLKKVENRN SKKVAKIEQR NEHRAERGLP LLTLDEPETA 180  
 LDETGHLRHR PGINSVCFGY QHMKLKPVYP GSIPGVGTGYS RDPSTPIAAC GVDRLEIPEG 240  
 QPGYVPPWDR ENLSVKKHRR KRASWARSDR GAIDDNNMLLA VVRVADDWAL LDLRGLLRNT 300  
 QYRKLLDRSV PTIESLLNL VTNDPLTSVV KKPGBKPYRTT ATLIYQOGVV PVVKAKVVKG 360  
 SYVKMDDT TETEFLVGVD LGVNNLIAAN ALRIRPGKCV ERLQAFTLPE QTVEDFFRFR 420  
 KAYDKHQENL RLAARVSLTA EQQAEVLA LD TFGPEQAKM VCGHGLLSVSD EVPWDKVNSR 480  
 SSILSLDAKE RGVDDTLYMF PFFKGKGKKE KTEIRKRWDV NWAQHFRPQL TSETRKALNE 540  
 AKWEARNSS KYHQLSIRKK ELSRHCNVYV IRTAEKRAQC GKVVAVEDL HHSFRGGKG 600  
 SRKSGWGGFF AAKQEGRWLM DALFGAFCDL AVHRYGRVIK VDPYNTSRIC PECGHCDKAN 660  
 RDRVNREAFI CVCCGYRGNA DIDVAAYNIA MVAITGVSLR KAARASVAST PLESIAAE 718

SEQ ID NO: 427            moltype = AA length = 708  
 FEATURE                    Location/Qualifiers  
 source                    1..708  
 mol\_type = protein  
 organism = unidentified

SEQUENCE: 427  
 MSKTKELNDY QEAALARLPG VRHKQSVRA ARLVYDROGE DAMVAFLDGK EVDEPYTILQP 60  
 PAKCHILAVS RPIEEWPIAR VTMAVQEHVY ALPVHEVEKS RPETTEGSRSS AWFKNSGVSN 120  
 HGVTQTLN AILKNAYNVY NGVVKVENR NAKKRDSLAA KNKSREKGL PHFKADPPEL 180  
 ATDEQGYLLQ PSPPNSSVYL VQQHLRTPQI DLPSGYTGPV VDPRSPIPSL IPIIDLAIAPP 240  
 GQPGYVPLHD REKLTSNKHS RMKLPKSLRA QGALPVCFRV FDDWAVVDGR GLLRHQAQYRR 300  
 LAPKNVIAE LLELYTGDPV IDIKRNLMTF RFAEAAVEVT ARKIVEKYHH KYLLKLTEPK 360  
 GKPVREIGLV SIDLNVQRLL ALAIYRVHQQT GESQLALSPC LHREILPAKG LGDFDKYKSK 420  
 FNQLTEEILT AAVQTLTSAQ QEEYQRYVEE SSHEAKADLC LKYSITPHEL AWDKMTSSTQ 480  
 YISRWLRDHG WNASDFTQIT KGRKKVERLW SDSRWAQELK PKLSNETRK LEDAKHDLQR 540  
 ANPEWQRLAK RKQEYSRHLA NTVLMSAREY TACETVVIAI ENLPMKGFFV DGNGSRESGW 600  
 DNFPTHKKEN RWMIKDIHKA LSDLAPPNRGV HVLEVNPQYT SQTCPECGHR DKANRDPPIQR 660  
 ERFCCTHCGA QRHADLEVAT HNIAMVATTG KSLTGKSLAP QRLQEEAE 708

SEQ ID NO: 428            moltype = AA length = 491  
 FEATURE                    Location/Qualifiers  
 source                    1..491  
 mol\_type = protein  
 organism = unidentified

-continued

---

SEQUENCE: 428  
VLLSDRIQYT DPSAPIPAMT VVDRRKIKKG EPGYVPPPFMR KNLSTNKHRR MRLSRGQKEA 60  
CALPVGLRLP DGKDGWDFII FDGRALLRAC RRLRLEVTSR DDVLDFKFTGD PRIQLSPAGE 120  
TIVTCMLKPQ HTGVIQOKLI TGKMKDRLVQ LTAEAPITAML TVDLGEHNLV ACGAYTVGQR 180  
RGKLQSERLE AFLLPEKVLA DFEGYRRDSD EHSETLRHEA LKALSKRQQR EVLDMRLRTGA 240  
DQARESCLCYK YGLDLQALPW DKMSSNSTPI AQHLMMSLGFG ESATHVRYRP KRKAERTIL 300  
KYDSRFAAEE KIKLTDETRR AWNEAIWECQ RASQEFRCLS VRKLQLARAA VNWLTLTQAKQ 360  
RSRCPRVVV VEDLNVRFMH GGGKRQEGWA GFFKARSEKR WFIQALHKAY TELPTNREGIH 420  
VMEVNPARTS ITCTKCGYCD PENRYGEDFH CRNPKCKVRG GHVANADLDI ATENLARVAL 480  
SGMPMPKAPKL K 491

SEQ ID NO: 429 moltype = AA length = 848  
FEATURE Location/Qualifiers  
source 1..848  
mol\_type = protein  
organism = unidentified  
SEQUENCE: 429  
MTPSFGYQMI IVTPIHHASG AWATLRLFL NPCTSGVMLG MTKTKSAFAL MREEVFPGLL 60  
FKSADLKMG RPKFAKEGRE AIEYLRKGDE ERPANFKPPA KGDIIAQSRP FDQWPIVQVS 120  
QAIQKYIFGL TKAEDATKT LLYGEGNHPT TESRRRWPEA TGVPDFGFTS AQGLNAIFSS 180  
ALARYEGLIQ KVENERNEKRL KKLSEKNQRL VEEGHAVEAY VPETAFHTLE SLKALSEKSL 240  
VPLDDLMDK1 DRLAQPPGIN PCLYGYQQVA PYIYDPENPR GVVLPLDYLG YCRKPDDPIT 300  
ACPNRRLDIPK GQPGYIPEHQ RGQLKKHGVR RRFRYTNPQA KARAKAQTAI LAVLRIDEDW 360  
VVMDLRGLLR NVYFREVAAP GELTARTLL TFTGCPVNLN RSNNVTFCYD IESKGALHAE 420  
YVRKGWATRN KLLDLTKDGQ SVALLSVDLG QRHPVAVMIS RLKRDDKGDL SEKSIQVSR 480  
TFAQDYVDKL KRYRVDYDAL RKEYDAALV SLPPEQQAEI RAYEAFAPGD AKANVLSVMF 540  
QGEVSPDDELW WKDMNTNTHY ISDLYLRGG DPSRVFFVPO PSTPKKNAKK PPAPRKPVKR 600  
TDENVSHMPE PRPHLSNETR EAFOKAKWTM ERGNVRYAQL SRFLNQIVRE ANNWLVSSEAK 660  
KLTQCQTVVW AIEDLHVPFF HGKGYKHTW DGFFRQKKED RWFVNVFHKA ISERAPNKG 720  
YVMEVAPYRT SQRCPCVCGFV DADNRHGDHF KCLRCGVELH ADLEVATWNI ALVAVQGHGI 780  
AGPPREQSCG GETAGTARKG KNIKKNKGLA DAVTVEAQDS EGGSKKDAGT ARNPVYIPSE 840  
SQVNCPAP 848

SEQ ID NO: 430 moltype = AA length = 781  
FEATURE Location/Qualifiers  
source 1..781  
mol\_type = protein  
organism = unidentified  
SEQUENCE: 430  
MKPKTPKPK TPVAALIDKH FPGKFRFRASY LKSVGKKLKN QGEDVAVRFL TGKDEERPPN 60  
FQPPAKSNIV AQSRRPIEWP IHKVSVAQVE YYVGLTVAEK EACSDAGESS SSHAAWFAKT 120  
GVENFGYTSV QGLNKIFPPPT FNRFDGVVIK VENRNEKKRQ KATRINEAKR NKGQSEDPP 180  
AEVKATDDAG YLLQPPGINH SVYGYQSTTL CPYTAEKPTT IKLPEEYAGY HSNPDAPIPA 240  
GVPDRLAIEP GQPGYIPEEH RAGLTLKHKR VRVQWYAMAN WKPKPKRTSK PDYDRLAKAR 300  
AQGALLIVIR IDEDWVVDA RGLLRNVRWR SLGKREITPN ELLDLFTGDP VLIDLKRGVVT 360  
FTYAEGVNVN CSRSTTKGKQ TKVLLDAMTA PRDGKQRQIG MVAVDLGQTN PIAAEYSRVG 420  
KNAAGTLEAT PLRSRSTLPD LLREITALYRK AHDRLLEAQLR EEAVALKLTAE QQAENARYVE 480  
TSEEGAKLAL ANLGVDTSTL PWDAMTGWST CISDHЛИННГ GDTSAVFQQT IRKGTKLET 540  
IKRKDSSWAD IVPRPLTKET REALNDFLWE LKRSHEGYEK LSKRLEELAR RAVNHVQEV 600  
KWLTCQCDIV IVIEDLNRN FHGGGKRGGG WSNNFTVKE NRWMQMALK AFSDLAAHRG 660  
IPVLEVYPAR TSITCLGCGH CDPENRDGEA FVCQQCQATF HADLEVATRN IARVALTGEA 720  
MPKAPAREQP GGAKKRGTSR RRKLTEVAVK SAEPTIHQAK NQQLNGTSRD P VYKGSELPA 780  
L 781

SEQ ID NO: 431 moltype = AA length = 767  
FEATURE Location/Qualifiers  
source 1..767  
mol\_type = protein  
organism = unidentified  
SEQUENCE: 431  
MSEITDLLKA NPKGKTFKSA DMRRMAGRILK KSGAQAVIKY LSDKGAVDPP DFRPPAKCNI 60  
IAQSRPFDEW PICKASMAIQ QHIYGLTKNE FDESSPGTSS ASHEQWFAKT GVDTHGFTHV 120  
QGLNLNIFQHA KRYEGLVIKK VENYNEKERK KFEGINERRS KEGMPILLEPR LRTAPGDDGK 180  
FAEKPGVNPNS IYLYQQTSPR PYDGTKHYPV HAFFELKEIT TIPTQDDRKL IPFGAPGHVP 240  
EKRHSQSLRAWYA DGSKGRSSRNK DLSADLKAASL ADAIPLVSRV 300  
GFDWWVVIDGR CLLRNLRWRK LAHEGMTVVEE MLGFFSGDPV IDPDRRNVATF IYKABHATVK 360  
SRKPIGGAKR AREELLKATA SSDGVIRQVG LISVLDLGQTN PVAYEISRMH QANGELVAEH 420  
LEYGLLNDEQ VNSIQRYRAA WDSMNESFRQ KAIIESLSMEA QDEIMQASTG AAKRTREAVL 480  
TMFGPNATLW WSRMSSNTTC ISDALIEVGK EEEETNFVTSN GPRKRTDAQW AAYLPRPRVNP 540  
ETRALLNQAV WDLMKRSDEY ERLSKRKLEM ARQCVNFVVA RAEKLTQCNN IGIVLENLVV 600  
RNPHGSGRRE SGWEGFFEPK RENRWMQVLL HKAFSDLQAH RGVMVFEVHP AYSSQTCPAC 660  
RYVDPKNRSS EDRERFKCLK CGRSFNADRE VATFNIREIA RTGVLGLPKD CERSRGVQTT 720  
GTARNPGRSL KSNKNPSEPK RVLQSKTRKK ITSTETQNEP LATDLKT 767

SEQ ID NO: 432 moltype = AA length = 760  
FEATURE Location/Qualifiers

-continued

---

source	1..760
	mol_type = protein
	organism = unidentified
<b>SEQUENCE: 432</b>	
MPKTESPLS ALCKKHFPKG RFRNTNYLDA GKILKKHGED AVVAFLSDKQ EDEPANFCPP	60
AKVHILAQSRR PFEWDWPINLA SKAIQTYVVG LTADERKTCE PGTSKESHDR WFKETGVDH	120
GFTSVQGLNL I FKHTLNRYD GVIKKVETRN EKRRSSVVRI NEKKAAEGLP LIAAEAEETA	180
FGEEDGRLLQP PGVNHSIYCF QVQSPQPYSS KKHPQVVLPH AVQGVPDPAP IPVGRPNRLD	240
I PKGQPGYVP EWQRPHLSMK CKRVRMWYAR ANWRRKPGRR SVLNEARLKE ASAKGALPIV	300
LVIGDDWLVM DARGLLRSVW WRRVKPGPLS LSELLNVPTP GLFSGDPVID PKRLGLVTFTS	360
KLGVVAVHSL KPTRGKSKD LLLKMTKPTD DGMPRHVGVM AIDLQTNPVA AEYSRVQ	420
DAGTLKQEPV SRGVLPDDLL KDVAHYRRAY DLTEESTRQE AIALLSEGHRAEVTKLDQTT	480
ANETKRLLV D RGVS E SLPWE KMSSNTTYIS DCLVALGKTD DVFFVPKAKK GKKETGIAVK	540
RKDHWGSKLL RPRTSPEARE ALNENQWAVK RASPEYERLS RRKLELGRRC VNHIQETKR	600
WTQCEDIVVV LEDLNVGFHH GSGKRPDGWD NFFFVKRENR WFIVQLHKA F GDLATHRGTH	660
VIEVHPARTS ITCIKCGHD AGNRDGESFV CLASACGDR HADLEVATRN VARVATGER	720
MPPSEQARDV QKAGGARKR K PSARNVKSSY PAVEPAPASP	760
 SEQ ID NO: 433	moltype = AA length = 752
FEATURE	Location/Qualifiers
source	1..752
	mol_type = protein
	organism = unidentified
<b>SEQUENCE: 433</b>	
MSDNKMKLKS KEKPPLTPLQ ILIRKYIDKS QYPGSGFKTTI IKQAGVRIKS VKSEQDEINL	60
ANWIISKYDP TYIKRDFNPS AKCQJIATSR SVADF DIVKM SNKVQEIFFA SSHLDKNVFD	120
I GKS KSDHDS WFERNNVDRG IYTYSNQGM NLIFSNNTKNT YLGVAVKAQN KFSSKMKRIQ	180
DINNFNFRINH QSPLIPIDEI KYIDDAGFLL NPDPGVNPNIY GYQSCLLKPL ENKEIISKTS	240
FPEYSRLPAD MIEVNYKISN RLKFSNDQKG FIQFKDKLNL FKINSQELFS KRRRLSGQPI	300
LIVASFGDW VVLDGRGLLQV QVYYRGIACP GSITISELLG FFTGDPIVDP IRGVVSLGF	360
PGVLSQETL TS S TSARIFAEK LPNVLVNNNV GLMSIDLGQT NPVSYRLSEI TSNMSVSEHIC	420
SDPLSQDQIS SIEKAKTSLD NLEEEIAIKA VDHLSDEDKI NFANFSKLNL PEDTROSILFE	480
KYPELIGSKL DFGSMGSGTS YIADELIKPE NKDAFYPSKG KKFDLSFSDR LRKQLSDETR	540
KSYNDALFLE KRTNDKYLKN AKRRQCIVRT VANSLVSKIE ELGLTPVINI ENLAMSGGFF	600
DGRGKREKGW DNFFKVKKEN RWLVMKDFHKA FSELSPHGV IVIESPPYCT SVTCTKCNFC	660
DKKNRNGHKF TCQRCGLDAN ADLIDATENL EKVAISGKRM PGSERSSDER KVAVARKAKS	720
PKGKAIKGVK CTITDEPALL SANSDQCSQS TS	752
 SEQ ID NO: 434	moltype = AA length = 747
FEATURE	Location/Qualifiers
source	1..747
	mol_type = protein
	organism = unidentified
<b>SEQUENCE: 434</b>	
MALS LAEVBRE RHF KGLRFRS SYLK RAGKIL KKEGEAACVA YLTGKDEESP PNFKPPAKCD	60
VVAQSRPFEE WPIVQASVAV QSYVYGLTKE AFEAFNP GTT KQSHEACLAA TGIDTCGYSN	120
VQGLNLIFRQ AKNRYEGVIT KVENRNNKKK KKLTRKNEWR QKNGHSELPE APEEELTFNDE	180
GRLLQPGPGN PSLTYQQIS ITPWSPPKDSS TQKKFMHPG L STRKNKMRML PRSVR SAPL G ALLVTIHLGE	240
GC PGYIPEWM RTAGEK TNP TQKKFMHPG L STRKNKMRML PRSVR SAPL G ALLVTIHLGE	300
DWLVLVDVRGL LRN ARW RGV A PKDISTQG L NLFTGDPVID TRRGVFTFTY KPETVGIHSR	360
TWLYKGQTK EVLEKLTQDQ TVALVAIDLG QTNPVSAAS RVRSRGENLS IETVDRF FLP	420
DEL I KELRLY RMAH DR L R I REESTLALT EAQQA E V R AL EKKA F N L D A	480
ASLPWDQM TS NTY SEAIL AQGVSRDQVF F TPNPKKGSK EPF VEV MRKDR AWVYAFKA K L	540
SEETRKAKNE ALWALKRASP DYARLSKRRE ELCRRSVNM INRAK KRTQC QVVIPVLED L	600
NIGFFHGSGK RLPGNDNFFV AKKENR WLMN GLHKSFSDSL A VRG FV V FEV MPH R TSITCP	660
ACGHCDSENR DGE AFV CLSC KRTYHADL DV ATHNL TQVAG TGLPMP PEREH PGGT KKP GGS	720
RKPESPQTHA PILHRTDYSE SADRLGS	747
 SEQ ID NO: 435	moltype = AA length = 733
FEATURE	Location/Qualifiers
source	1..733
	mol_type = protein
	organism = unidentified
<b>SEQUENCE: 435</b>	
QAVIKYLSDK GAVDPPDFRP PAKC NIIAQS RPF DEW PICK ASMAI QQH Y GLT KNEFDES	60
S PGTSSASHE QWFAK TGVDT HGF THVQGLN L IFQHAKK RY EG VIKK VENY NEK EKK FEG	120
INERRSKEGM PLLEPLRLTA FG DGDG KFAEK PGVNP SIYLY QQTSPR PYDK TKHPYV HAP F	180
E LKE ITTI TPT QDDR LK I PFG AP GHV P E K H R SQL SMA K H K R R RAW YAL SQN K P R P K D G S K	240
GRR S V R D L A D I A L P L V S R V G F D W V V I D G R G L L R N L R W R K L A H E G M T V E E M L G F	300
FSGDPVIDPR RN VAT F I Y K A EH AT V K S R K P I G G A K R A R E E L L K K A T A S S D G V I R Q V G L I S V	360
D LG Q T N P V A Y E I S R M H Q A N G E L V A E H L E Y G L L N D E Q V N S I Q R Y R A W D S M N E S F R Q K A I E	420
SLSMEA Q D E I M Q A S T G A A K R T E A V L T M F G P N A T L P W S R M S S N T T C I S D A L I E V G K E E T	480
N F V T S N G P R K R T D A Q W A A Y L R P R V N P E T R A L L N Q A V W D L M K R S D E Y E R L S K R K L E M A R Q C	540
V N F V V A R A E K L T Q C N N I G I V L E N L V V R N F H G S G R R E S G W E G F F E P K R E N R W F M Q V L H K A F	600
S D L A Q H R G V M V F E V H P A Y S S Q T C P A C R Y V D P K N R S S E D R E R F K C L K C G R S F N A D R E V A T F	660
N I R E I A R T G V G L P K P D C E R S R D V Q T P G T A R K S G R S L K S Q D N L S E P K R V L Q S K T R K K I T S T	720

-continued

---

ETQNEPLATD LKT	733
SEQ ID NO: 436	moltype = AA length = 702
FEATURE	Location/Qualifiers
source	1..702
	mol_type = protein
	organism = unidentified
SEQUENCE: 436	
MIKEQSELSK LIEKYYPGKK FYSNDLKQAG KHLKKSEHLT AKESEELTVE FLKSCKEKL	60
DFRPPAKALI ISTSRPFEW PIYKASESIQ KYIYSLTKEE LEKYNISTDK TSQENFFKES	120
LIDNYGFANV SGLNLIFQHT KAIYDGVLLK VNRRNNNKILK KYKRKIEEGI EIDSPELEKA	180
IDESGHFVINP PGINKNIYCY QQVSPTIFNS FBTKIICPF NYKRNPNDDI QKGVIDRLAI	240
PFGEPEGYIPD HQRDVKVNHK KRIRKYYKNN ENKNKDAILA KINIGEDWVL FDLRGLLRNA	300
YWRKLIQKQG ITPQQLLDMF SGDPPVIDPIK NNITFIYKES IIPIHSEII KTKKSKELLE	360
KLTKDEQIALI VSIDLQQTNP VAARFSRLSS DLKPEHVSSS FLPDDELKNEI CRYREKS DLL	420
EIEIKNKAIA MLSQEQCDEI KLVNDISSEE LKNSVCKKYN IDNSKIPWDK MNGFTTFIAD	480
EFINNGGDKS LVYFTAKDKK SKKEKLVLKLS DKKIANSFKP KISKETREIL NKITWDEKIS	540
SNEYKKLSKR KLEFARRATN YLINOAKKAT RLNNVVVLVVE DLNSKFHGS GKREDGWDF	600
FIPKKENRWF IQALHKSLTD VSIHRGINVI EVRPERTSIT CPKCGCCDKE NRKGEDFKCI	660
KCDSVYHADL EVATFNIEKV AITGESMPKP DCERLGGEES IG	702
SEQ ID NO: 437	moltype = AA length = 661
FEATURE	Location/Qualifiers
source	1..661
	mol_type = protein
	organism = unidentified
SEQUENCE: 437	
VAPLDGKEVD EPYTLQPPAK CHILAVSRPI EEWPIARVTM AVQEHWYALP VHEVEKSRPE	60
TTEGSRSAWF KNSGVSNHGKV THAQLTNAIL KNAYNVYNGV IKKVENRANK KRDSLAAKNK	120
SRRRKGLPHF KADPPLEATD EQGYLLQPPS PNNSVYLVQQ HLRTPOIDLP SGYTGPVVDP	180
RSPIPSLIPI DRLAIPPGQP GYVPLHDREK LTSNKHRRMK LPKSLRAQGA LPVCFRVFDD	240
WAVVDGRGLI RHAQYRRLAP KNVSIAELLE LYTGDPVIDI KRNLMTRFA EAVVEVTARK	300
IVEKYHNKYI LKLTEPKGKP VREIGLVSID LNQVRLLALIA IYRVHQGTGES QLALSPCLH	360
EILPAKGLGD FDKYKSKFQNQ LTEEILTAAV QTTLTSAQOEE YQRVVEESSH EAKADLCLKY	420
SITPHELAWD KMTSSTQYIS RWLDRHGWNNA SDPTQITKGR KKVERLWSDS RWAQELKPKL	480
SNETRTRKLED AKHDLQRANP EWQLAKRKQ EYSRHLANTV LSMAREYTAC ETVVIAIENL	540
PMKGGFVDGNS GSRESGWDFN FTHKKENRWM IKDIHKALSD LAPNRGVHVL EVNPQYTSQT	600
CPECGHARDKA NRDPIQRERF CCTHCGAQRH ADLEVATHNI AMVATTGKSL TGKSLAPQRL	660
Q	661
SEQ ID NO: 438	moltype = AA length = 531
FEATURE	Location/Qualifiers
source	1..531
	mol_type = protein
	organism = unidentified
SEQUENCE: 438	
LEIPEGEPEGH VPWFQORMDIP EGQIGHVNKI QRPNFVHGKN SGKVFKSDKT GRVKRYHHSK	60
YKDATKPYKF LEESKKVSL DSILAIITIG DDWWVFDIRG LYRNVFYREL AQKGLTAVQL	120
LDLFTGDPVI DPKKGIITFS YKEGVVPVFS QKIVSRFKSR DTLEKLTSQG PVALLSVDLG	180
QNEPVAAARVC SLKNINDKIA LDNSCRIPFL DDYKKQIKDY RDSLDELEIK IRLEAINS	240
VNQVEIRDL DVFSADRAKA STVMDFDIDP NLISWDSMSD ARFSTQISDFL YLKNGGDESR	300
VYFEINNKRI KMTSSTQYIS VRPKLSDSTR KNLNDSIWLK KRTSEEYLKL SKRKELESRA	360
VVNYTIQSK LLSSGINDIVI ILEDLDVKKK FNFRGIRDIG WDNNFFSSRKI NRWFIPAFHK	420
SFSELSSNRG LCVIEVNPW TSATCPDCGF CSKENRDGIN FTCRKCGVSY HADIDVATLN	480
IARAVAVLGKP MSGPADRERL GGTKKPRVAR SRKDMKRDI SNGTVEVMVT A	531
SEQ ID NO: 439	moltype = AA length = 479
FEATURE	Location/Qualifiers
source	1..479
	mol_type = protein
	organism = unidentified
SEQUENCE: 439	
IPSFGYLDRL KIAKGQPGYI PEWQRETINP SKKVRRYWAT NHEKIRNAIP LVVFIGDDWV	60
IIDGRGLLRD ARRRKLADKN TTIEQLEMV SNDPVIDSTR GIATLSYVEG VVPVRSFIFI	120
GEKKGREYLE KSTQKESVT LSVDVGQINP VSCGVYKVSN GCSKIDFLDK FFLDKKHLD	180
IQKYRTLQDS LEASIVNEAL DEIDPSFKKE YQININSOTSN DVKKSLCTEY NIDPEAISWQ	240
DITAHSHTLIS DYLDIINNITN DVYRTVNKAQ YKTNDFGWYK KFSAKLSKEA REALNEKIWE	300
LKIASSKYKK LSVRKKEIAR TIANDCVKRA ETYGDNVVVA MESLTKNNKV MSGRGKRDPG	360
WHNLGQAKVE NRWFQIAISS AFEDKATHHG TPVLKVNPAY TSQTCPSCHG CSKDNRSSKD	420
RTIFVCKSCG EKFNADLDVA TYNIAHVAFS GKKLSPPSEK SSATKKPRSA RKSJKRSRK	479
SEQ ID NO: 440	moltype = AA length = 452
FEATURE	Location/Qualifiers
source	1..452
	mol_type = protein
	organism = unidentified

-continued

---

SEQUENCE: 440				
SPIEKLLNGL LVKITFGNDW IICDARGLLD NVQKGIIHKS YFTNKSSLVD LIDLFTCNPI	60			
VNYKNNVVTFCYKEGVVDVK SFTPIKSGPK TQEENLIKLLK YSRFQNEKDA CVLGVGVDVG	120			
VTPMPFAINGFKMPVDESSEW VMLNEPLFTI ETSQAFREEI MAYQQRTDEM NDQFNQQSID	180			
LLPPEYKVEF DNLPEDINEV AKYNLLHTLN IPNNFLWDKM SNTTQFISDY LIQIGRGTET	240			
EKTITTKKGK EKILTIRDVN WFNTFKPKIS EETGKARTEI KRDLQKNSDQ FQKLAKSREQ	300			
SCRTWVNNVT EEAQIKSGCP LIIFVIEALV KDNRNFSGKG HRAIGWHNFG KQKNERRWV	360			
QAIHKAFQEQ GVNHGPVIL CPPQTSQTC PKCNHVDRDN RSGEKFCLK YGWIGNADLD	420			
VGAYNIARVA ITGALKSKPL EQKKIKKAKN KT	452			
SEQ ID NO: 441	moltype = AA length = 425			
FEATURE Location/Qualifiers				
source 1..425				
	mol_type = protein			
	organism = unidentified			
SEQUENCE: 441				
LLDNVQKGII HKSYFTNKSS LVLDLIDLFTC NPIVNYKNNV VTFCYKEGVV DVKSFTPIKS	60			
GPKTQENLIK KLKYRSRFQNE KDACVLGVGV DVGVTNPFAI NGFKMPVDES SEWVMLNEPL	120			
FTIETSQAFR EEIMAYQORT DEMNDQFNQQ SIDLLPPEYK VEFDNLNPEDI NEVAKYNLLH	180			
TLNIPPNFLW DKMSNTTQF1 SDYLIQIGRQ TEETEKTTT KKGKEKLTIR DVNWFTFKP	240			
KISEETGKAR TEIKRDLQKN SDQFQKLAKS REQSCRTWVN NVTEEAKIKS GCPLIIFVIE	300			
ALVKDNRVFS GKGHRAGWH NFGKQKNERR WWVQAIHKAF QEQGVNHHGYP VILCPPQYTS	360			
QTCPKCNHVD RDNRSGEKFK CLKYWGIGNA DLDVGAYNIA RAVITGKALS KPLEQKKIKK	420			
AKNKT	425			
SEQ ID NO: 442	moltype = AA length = 735			
FEATURE Location/Qualifiers				
source 1..735				
	mol_type = protein			
	organism = unidentified			
SEQUENCE: 442				
MKPTVQSFL TPFGKLIRNH SRTAGLKLKN EGEEACKKVF RENEIPKDEC PNFQGGPAIA	60			
NIIAKSREFT EWEIYQSSLR IQEVIFTLPK DKLPEPILKE EWRAQWLSEH GLDTVPYKEA	120			
AGLNLLIINKA VNTYKGQVQK VDNKNKNNLA KINRKNEIAK LNGEQEISEF EIKAPDDKGY	180			
LLQKPSFLW DQKMSNTTQF1 SFYLIQIGRQ TEETEKTTK KKDWCVFDMDR GLLRTNHWWK	240			
GEPGYVPKWQ YTFLSKKENK RRKLSKRIKN VSPILGII CI KKDWCVFDMDR GLLRTNHWWK	300			
YHKPTDSIND LFDFYFTGDPV IDTKANVVRF RYKMENGIVN YKPVREKKKGK ELLENICDQN	360			
GSCKLATVDV GONNPAVIGL FELKKVNGEL TKTTLISRHT PIDFCNKITA YRERYDKLES	420			
SIKLDAAIKL TSEQKIEVDN YNNNPTQPNQ KQIVCSKLNI NPNDLWPDKM ISGTHFISEK	480			
AQVSNKSEIY FTSTDKGKTK DVMKSLQWPF QDYPKPLSKE VRDALSDIEW RLRRESLEFN	540			
KLSKSRQDQA RQLANWISSM CDVIGIENLV KKNNFFGGSG KREPQWDNFY KPKKENRWWI	600			
NAIHKALTEL SQNKGKRVIL LPAMRTSITC PKCKYCDSKN RNGEKFNCLK CGIELNADID	660			
VATENLATVA ITAQSMPKPT CERSGDAKKP VRARKAKAPE FHDKLAPSYT VVLREAVKRP	720			
AATKKAGQAK KKKEF	735			
SEQ ID NO: 443	moltype = AA length = 1120			
FEATURE Location/Qualifiers				
source 1..1120				
	mol_type = protein			
	organism = Listeria seeligeri			
SEQUENCE: 443				
MWISIKTLIH HLGVLFCDY MYNRREKKII EVKTMRITKV EVDRKKVLIS RDKNGGKLVY	60			
ENEMQDNTQ IMHHKKSSFY KSVVNKTICR PEQKQMKLKV HGLLQENSQE KIKVSDVTKL	120			
NISNFLNHRF KKSLYYFPEN SPDKEEYRI EINLSQLLED SLKKQQGTFI CWESFSKDM	180			
LYINWAENYI SSKTKLIKK5 IRNNRIQSTE SRSGQLMDRY MKDILNKNKP FDIOQSVEKY	240			
QLEKLTSALK ATFKEAKND KEINYKLTQ LQNHERQI ELKENSELNQ FNIEIRKHLE	300			
TYPIKKTNR KVGDIRNLGI GEIQKIVNHR LKNKIVQRLI QEGKLASYEI ESTVNSNLSQ	360			
KIKIEEAFAL KFINACLFAS NNLRNMVYPV CKKDILMIGE FKNSFKEIKH KKFIRQWSQF	420			
F5QEITVDDI ELASWGLRGA IAPIRNEI1H LKKHSWKKF NNPTFKVFKS KIINGKTKDV	480			
TSEFLYKETL PKDYFYSELD SVPELIINKM ESSKILDYYS SDQLNQVFTI PNFEPLSLTS	540			
AVPFAPSFKR VYLKGPDYQW ODEAQPDYNL KLNLYNEKAQ NSEAFQAQYS LFKMVYYQVF	600			
LPQFTTNNDL FKSSVDFILT LNKERKGYAK AFQDIRKMNK DEKPSEYMSY IQSQLMLYQK	660			
KQEKEKEKINH PEKFINQVFI KGFNSFIEKN RLTYICHPTK NTVPENDNIE IPFHTDMDDS	720			
NIAFWLMCKL LDAKQLSLEK NEMIKFSCSL QSTEEISTFT KAREVIGLAL LNGEKGNDW	780			
KELFDDKEAW KKNMSLYVSE EDGQTPVIRN SIDLVKKYGT ETILEKLFSS	840			
SDDYKVSAKD IAKLHEYDVT EKIAQQESLH KOWIEKPGLA RDSAWTKKYQ NVINDISNYQ	900			
WAKTKVELTQ VRHLHQLTID LLSRLAGYMS IADRDFQFSS NYILERENSE YRVTSWILLS	960			
ENKNKNKYND YELYNLKNAS IKVSSKNDPQ LKVDLKQLRL TLEYLELFDN RLKEKRNNIS	1020			
HFNLYLNGQLG NSILELFDDA RDVLSYDRKL KNAVSKSLKE ILSSHGMETV FKPLYQTNNHH	1080			
LKIDKLQPKK IHHLGEKSTV SSNQVSNEYC QLVRTLLTMK	1120			
SEQ ID NO: 444	moltype = AA length = 1159			
FEATURE Location/Qualifiers				
source 1..1159				
	mol_type = protein			
	organism = Leptotrichia buccalis			

-continued

---

SEQUENCE: 444

MKVTKVGGIS	HKKYTKSEGRL	VKSEEEENRT	DERLSALLNM	RLDMYIKNPS	STETKENQKR	60
IGKLKKFFSN	KMVKLDNTL	SLKNGKKENI	DREYSETDIL	ESDVRDKKNF	AVLKKIYLNE	120
NVNSEELEV	RNDIKKKLNK	INSLKYSFEK	NKANYQKINE	NNIEKVEGKS	KRNIIYDYYR	180
ESAKRDAYVS	NVKEAFDKLY	KEEDIAKLVL	EIENLTKELEK	YKIREFYHEI	IGRKNDKENF	240
AKIYIEEIQN	VNNMKELIEK	VPDMSELKKS	QVFYKYYLDK	EELNDKNIKY	AFCHPVEIEM	300
SQLLKNVYVK	RLSNISNDKI	KRIFYEQNLK	KLIENKLLNK	LDTYVRNCGK	YNYYLQDGEI	360
ATSDDFIARNR	QNEAFLRNII	GVSSVAYFSL	RNILETENEN	DITGRMRGKT	VKNNKGEEKY	420
VSGEVDKIYN	ENKKNEVKEN	LKMFSYDFN	MDNKNEIEDF	FANIDEAIISS	IRHGIVHFNL	480
ELEGKDIAF	KNIAPSEISE	KMFQNEINEK	KLKLKIFRQL	NSANAFRYL	KYKILNLYLK	540
TRPEFVNKNI	TIPVPSFTKLY	SRIDDLKNSL	GIYWKTPKTN	DDNKTKEIID	AQIYLLKNIY	600
YGFGLNYFMS	MNGNFEEISE	EIELLNKNDK	RNLKTGFYKL	QKFEDIQEKK	PKEYLANIQS	660
LYMINAGNQD	EEEKDTYIDF	IQKIFLKGFM	TYLANNGRSL	LIYIGSDEET	NTSLAEKKQE	720
FDKFLKQYEQ	NNNIKIPYEI	NEFLREIKLG	NILKYTERLN	MFYLLKLLN	HKELTNLKGS	780
LEKYQSONAKE	EAFSQDLELI	NLLNLDNNRVE	TEDFELEADE	IGKFLLDFNGN	KVKDNKELKK	840
FDTNKKIYFDG	ENIIKHKRHY	NIKKYGMNLN	LEKIADKAGY	KISIEELKKY	SNKNEIEKHN	900
HKMQENLHRK	YARPRKDEKF	TDEDYESYKQ	AIENIEEYTH	LKNKVEFNEL	NLLQGLLRI	960
LHRLVGYTSI	WERDLFRFLK	GEFPENQYIE	EIPNFENKNN	VKYKGQOIVE	KYIKPYKELH	1020
QNDEVKINKY	SSANIKVLKQ	EKKDLYIRNY	IAHFNYIPHA	EISLLEVLEN	LRKLLSYDRK	1080
LKNAVAMKSVV	DILKEYGFVA	TFKIGADKKI	GIQTESEKII	VHLKNLKKKK	LMTDRNSEEL	1140
CKLVKIMFEY	KMEEKKSEN					1159

SEQ ID NO: 445      moltype = AA length = 1389

FEATURE      Location/Qualifiers  
source      1..1389  
mol\_type = protein  
organism = Leptotrichia shahii

SEQUENCE: 445

MGNLFGHKRW	YEVRDKKDFK	IKRKVKVCRN	YDGNKYIILN	NENNNEKEKID	NNKFIRKYIN	60
YKKNDNILKE	FTRKPHAGNI	LFKLKGKEGI	IRIENNDDFL	ETEEVVLVIE	AYGKSEKLKA	120
LGITKKKIID	EAIRQGITKD	DKKIEIKRQE	NEEEIEIDIR	DEYTNTKLND	CSIILRIEN	180
DELETKKSIY	EIFKNINMSL	YKIEEKJEN	ETEKVFNERY	YEEHLEKLL	KDDKIDVILT	240
NFMEIREKIK	SNLEILGFVK	FYLNVGGDKK	KSKNKKMLVE	KILNINVDLT	VEDIADFVIK	300
ELEFWNITKR	IEKVKKVNNNE	FLEKRRNRTY	IKSYVLLDKH	EKFKIERENK	KDKIVKFVFE	360
NIKNNNSIKEK	TIKLAELFKI	DELICKLEKE	LKKGNCDTEI	PGIFPKKHVK	NFDSKKFSKK	420
SDEEKELYKI	IYRYLKGRI	KILVNEQKVR	LKKMEKIEIE	KILNESILSE	KILKRVKQYT	480
LEHIMYLGKL	RHNDIDMTTV	NTDDFSRLHA	KEELDLELIT	FFASTNMELN	KIFSRENINN	540
DENIDFFGGD	REKNVYLDKK	IILNSKIKIIR	DLDPIDKNM	ITNNFIRKFT	KIGTNERNR	600
LHAISKERDL	QGTQDDYNK	IIINIIQNLNQ	DEEVSKALN	DVVFQDKKN	ITKINDIKIS	660
EENNNDIYL	PSFSKVLPEI	LNLYRNPNPK	EPPDTIETK	IVLNALIYVN	KELYKKLILE	720
DDLEENESKN	IFLQELKKTL	GNIDEIDENI	IENYYKNAQI	SASKGNKAI	KKYQKVKIEC	780
YIGYLRKNYE	ELFDFDSDFKM	NIQEIKKQIK	DINDNKTYER	ITVKTSDKTI	VINDDFEYII	840
SIFALLNSNA	VINKIRNRFF	ATSWLNLNTSE	YONIIDLIL	IMQLNTRLNE	CITENWNLN	900
EEFIQKMKEI	EKFDDFKIQ	TKKEFVNYY	EDIKNNILTE	FKDDINGCDV	LEKKLEKIVI	960
FDDETKFEID	KKSNIQDQE	RKLSNINKKD	LKKKVDQYIK	DKDQEIKSKI	LCRIFNSDF	1020
LKKYKKEIDN	LIEDMESENE	NKFQEIYYPK	ERKNELYIYK	KNLFLNIGNP	NFDKIYGLIS	1080
NDIKMADAKF	LFNIDGKNIK	KNKISEIDAI	LKNLNDKLN	YSKEYKEKYI	KKLKBENDDF	1140
AKNIQNKNYK	SEPKDYNRVS	EYKKIRDLV	FNYLNKIESY	LIDINWKLAI	QMARFERDMH	1200
YIVNGLRELG	IIKLSGYNTG	ISRAYPKRNG	SDGFYTTTAY	YKFFDEESYK	KFEKICYGFG	1260
IDLSENSEIN	KPENESIRNY	ISHFYIVRNP	FADYSIAEQI	DRVSNLLSYS	TRYNNSTYAS	1320
VFEVFKKDVN	LDYDELKKKF	KLIGNNDILE	RLMKPKKVS	LELESYNSDY	IKNLIELLT	1380
KIENTNDTL						1389

SEQ ID NO: 446      moltype = AA length = 1285

FEATURE      Location/Qualifiers  
source      1..1285  
mol\_type = protein  
organism = Rhodobacter capsulatus

SEQUENCE: 446

MQIGKVQGRT	ISEFGDPAGG	LKRKISTDGK	NRKELPAHLS	SDPKALIGQW	ISGIDKIYRK	60
PDSRKSDGKA	IHSPTPSKMQ	FDARDDLGEA	FWKLVSEAGL	AQDSDYDQFK	RRLHPYGDKF	120
QPADSGAKL	FEADPPEPQA	FHGRWYGAMS	KRGNDAKELA	AALYEHHLVD	EKRIDGQPKR	180
NPKTDKFAFG	LVVARALGIE	SSVLPRGMAR	LARNWGEEEI	QTYFVVDVAA	SVKEVAKAAV	240
SAAAQAFDPPR	QVSGRSLSPI	VGFALAEHLE	RVTGSKRCSF	DPAAGPSVLA	LHDEVKKTYK	300
RLCARGKNA	RAPPADKTEL	LALMRHTHE	RVRNQVMVRMG	RVSEYRQQQA	GDLAQSHYWT	360
SAQOTEIKES	EIVFVRLWVG	FALAGRSMKA	WIDPMGKIVN	TEKNDRLTA	AVNIRQVISN	420
KEMVAEAMAR	RGIYFGETPE	LDRLGAEGNE	GTVFALLRLY	RGCRNQTFHL	GARAGFLKEI	480
RKELEKTRWG	KAKEAEBVVL	TDKTVAAIRA	IIDNDAKALG	ARLLADLSGA	FVAHYASKEH	540
FSTLYSEIVK	AVKDAPEVSS	GLPLRKLLLK	RADGVGRGVH	GLRDTRKHAF	ATKLPPPAP	600
RELDDPATKA	RYIALLRLYD	GPFRAYASGI	TGTALAGPAA	RAKEAATALA	QSVNVTKAYS	660
DVMEGRSSRL	RPNPNDGETL	EYPLSALTGET	ATEFRVQIYQ	ESDSERANKQ	AEFIENYRRD	720
MLAFMFEDYI	RAKGFDWILK	IEPGATAMTR	APVLPEPIDT	RGQYEHWQAA	LYLVMHFVPA	780
SDVSNLHQ	RKWEALQGKY	ELVQDGDATD	QADARREALD	LVKRFDRDVLV	LFLKTGEARF	840
EGRAAPFDLK	PFRALFANPA	TFDRLFMATP	TTARPAEEDP	EGDGASEPEL	RVARTLRLGR	900
QIARYNHMAV	LSDLFAKHKV	RDEEVARLAE	IEDETQEKSQ	IVAAQELRTD	LHDKVMKCHP	960
KTISPEERQS	YAAAIIKTEE	HRFLVGRVYL	GDHRLRHRLM	MDVIGRLIDY	AGAYERDTGT	1020

-continued

---

FLINASKQLG AGADWAVTIA	GAANTDARTQ	TRKDLAHFNV	LDRADGTPDL	TALVNRAREM	1080
MAYDRKRKNA VPRSILDMIA	RLGLUTLKWQM	KDHLLQDATI	TQAAIKHLKD	VRLTVGGPAA	1140
VTEARFSQDY LQMVAVFNG	SQVNPKPRRR	DDGDAWHKPP	KPATAQSQPD	QKPPNKAPSA	1200
GSRLPPPCVG EVYEGVVVKV	IDTSLGLFIA	VEGVAGNIGL	HISRLRRIRE	DAIIVGRRYR	1260
FRVEIYVPPK SNTSKLNAAD	LVRID				1285

---

SEQ ID NO: 447	moltype = AA length = 1175
FEATURE	Location/Qualifiers
source	1..1175
	mol_type = protein
	organism = Carnobacterium gallinarum

SEQUENCE: 447		
MRITKVKIKL DNKLKYQVTMQ	KEEKYGTLLK NEESRKSTAE ILRLKKASFN KSFHSKTINS	60
QEKENKNATK KNQGDYISQIF	EKLGVGDTNK NIRKPCKMSLT DLKDLPKKDL ALFIKRKFKN	120
DDIVEIKNL DLSLFNALQ	KVPGHEFTDE SWADFCQEMM PYREYKNF1 ERKIILLANS	180
IENONKGFSIN PETFSKRKRV	LHQWAIEVQE RGDFSLILDEK LSKLAEIYNF KKMCKRVDQE	240
LNDLEKSMKK GKNPEKEKEA	YKKQKQNFKIK TIWKDYPYKT HIGLIEKIKE NEELNQFNIE	300
IGKYFEHYFP IKKERCTEDE	PYYLNSETIA TTVNYQLKNA LISYLMQIGK YKQFGLENQV	360
LDSKKLQEIG EYQFQTKFM	DACVFATSSL KNIIEPMRSSG DILGKREFKE AIATSSFVN	420
HHFPPYFPPFE LKGMKDRESE	LIPFGEQTEA KQMQNIWLR GSVQQIRNEI FHSFDKNQKF	480
NLPQLDKSNSF EFDASENSTG	KSQSYIETDY KPLFEAEAKNQ LEQFFIERIK SSGALEYYPL	540
KSLLEKLFACK EMKFLSGSVP	VAFAPSYYKLL VKKGHSYQTA TEGTANYLGL SYYNRYELKE	600
ESFQAQYLYL KLIYQYVQF	NFSQGNSPMAF RETVKAILRN NKDEARKMK KNNKFLRKYA	660
FEOVREMEFK BTPDQYMSYL	QSEMREEKVR KAEKNDKGFE KNITMNFEKL LMQIFVKGFD	720
VFLTTFFAGKE LLLSSEEKVI	KETEISLSSK INEREKTLKA SIQVEHQLVA TNSAISYWLF	780
CKLLDSDRHLN ELRNEMIJKF	QSRIFKNHTQ HABELIQNLIP IVELTILSND YDEKNDSQNV	840
DVSAYFEDKS LYETAPYVQI	DDRTTVSFRP ILKLEKYHTK SLIEALLKDN PQFRVAATDI	900
QEWMMHKREEI GELVEKRKNL	HTEAWEGQQT LGAEKREYR DYCKKIDRFN WKANKVTLTY	960
LSQLHYLTD LLGRMVGFS	LFERDLVYFQ RSFSELGGET YHISDYKNLS GVRLRNAEVK	1020
PIKIKNIKVI DNEENPYKGN	EPEVKPFLDR LHAYLENVIG IKAHVHGKIRN QTAHLSVLQL	1080
ELSMIESMNN LRDLMLAYDRK	LKNAVTKSMI KILDKHGMIL KLKIDENHKN FEIESLIPKE	1140
IHLKDKAIK TNQVSEYYCQ	LVLALLTTNP GNQLN	1175

SEQ ID NO: 448	moltype = AA length = 1285
FEATURE	Location/Qualifiers
source	1..1285
	mol_type = protein
	organism = Herbinix hemicellulosilytica

SEQUENCE: 448		
MKLTRRRISG NSVDQKITAA	FYRDMMSQGLL YYDSEDNDCT DKVIESMDFE RSWRGRILKN	60
GEDDKNPFYM PVKGLVGSND	KIVCEPIDVD SDPDNLILN NKNLTGFGRN LKAPDSNDTL	120
ENLIRKIQAG IPEEEVLPK	KKIKEMIQKD IVNRKEQLLK SIKNNRIPFS LEGSKLVPST	180
KMKWKLFLK DVPNKTNEK	MLEYKWEYID YDKLKANITN RLDKTDKKAR SISRASVSEEL	240
REYHKNLRTN KVNFQYVQF	AAGLDNGGSA LIPKKEEFLP LFLKEVEQYF KKQFVPSKH	300
SNKSKDLSV DKYKNYCSY	VVKKEVNRSI INQLVAGLIIQ QGKLLYYFFF NDTWQEDFLN	360
SYGLSYIQLV EAFKKSVMTS	LSWGINRLTS FFIDDSNTVK FDDITTKAK EAIESNYFNK	420
LRTCSRMQDH FKEKLAFFY	VYVYKDKDPRP DDDIENLIVL VKNNAIESVST LRNRTFFHFK	480
SSLLELLKEL DDKNSGQNKI	DYSVAAEFIK RDIENLYDVF REQIRSLGIA EYYKADMISD	540
CFKTCGLEFA LYSPKNSLMP	AFKNVYKRG A NLNKAYIIRDK GPKETGDQGQ NSYKALEEYR	600
ELTWYIEVKN NDQSYNAYKN	LLQLIYYHAF LPEVRENEAL ITDFINRTKE WNRKETEERL	660
NTKNNKKHHN FDENDDTIVN	TYRYESIPDY QGESLDDYLYK VLQRKQMARA KEVNEKEEGN	720
NNYIQFIRDV VVWAFGAYLE	NKLKNYKNEI QPPLSKENIG LNDTLKELFP EEKVKSPFNI	780
KCRFSISTFI DNKGKSTDNT	SAEAVKTGK EDEKDKKNIK RKDLLCFYLF LRLLDENEIC	840
KLQHQFQIKY CSLKERRFP	R NRTKLEKETE LLAELLEELME LVRFTMPsip EISAKAESGY	900
DTMIKKYFKD FIEKKVFKP	KTSNLNYYHSD SKTPVTRKYM ALLMRMSPHL LYKDFIKGYY	960
LITKCLEY I KLDKQKHD	QNSLNLHEQ LERIKLKSEK QNGKDSLQYD KKDFYKVKEY	1020
VENLEQVARY KHLQHKINFE	SLYRIFRHV DIAARMVGYT QDWERDMHFL FKALVNGVL	1080
ERRFEAIFN NNDNNNDGRI	VKKIQQNLLNN KNRELVSMLC WNKKLNKNEF GAIIWKRNP	1140
AHLNHFTQTE QNSKSSLES	INSRLILLAY DRKRQNAVTK TINDLLLNDY HIRIKWEGRV	1200
DEGQIYFNIK EKEDIENEP	IHLKHLHKKD CYIYKNSYMF DKQKEWICNG IKEEVYDKSI	1260
LCIGNLKFKF DYEDKKNKSSA	NPKHT	1285

SEQ ID NO: 449	moltype = AA length = 1154
FEATURE	Location/Qualifiers
source	1..1154
	mol_type = protein
	organism = Paludibacter propionicigenes

SEQUENCE: 449		
MRVSKVKVVD GGKDKMVVLH	RKTTGAQLVY SGQPVSNETS NILPEKKRQS FDLSTLNKTI	60
IKFDTAKKQK LNVDQYKIVE	KIFKYPKQEL PKQIKAEEIL PFLNHKFQEP VKYWKNGKEE	120
SFNLTLLIVE AVQAQDKRKL	QPYYDWKWTY IQTKSDLLKK SIENNRLIDLT ENLSKRKCAL	180
LAWETEFTAS GSIDLTHYHK	VYMTDVLCM LQDVPLTDD KGKINTNAYH RGLKKALQNH	240
QPAIFGTRV PNEANRADNQ	LSIYHLEVVK YLEHYFPKLT SKRRNTADDI AHYLLAQTLK	300
TTIEKQLVNA IRANIIQGK	TNHHELKADT TSNDLIRIKT NEAFVNLNTG TCAFAANNIR	360
NMVDEQTN	ILGKGDFIKS LLKDNNTNSQL YSFFFGEGLS TNKAEKETQL WGIRGAVQQI	420
RNNVNHYKKD ALKTVPNISN	FENPITDPK QQTNYADTIY KARFINELEK IPEAFAQQLK	480

-continued

---

TGGAVSYTTI	ENLKSLLTTF	QFSLCRSTIP	FAPGFKKVFN	GGINYQNAKQ	DESFYELMLE	540
QYLRKFAE	ESYNYAFMLP	KLIYNNFLP	GFTDRKAFA	DVGVFQMQN	KKQAEKVNP	600
KKEAYAFAEV	RPMTAADSIA	DYMAYVQSEL	MQEQNKEEK	VAEETRINFE	KFVLQVFIKG	660
FDSFLRAKEF	DFVQMPQPQL	TATASNNQKA	DKLNQLEASI	TADCKLTPQY	AKADDATHIA	720
FYVFCKLLDA	AHLSNLRNEL	IKFRESVNEF	KFHHHLLEIE	ICLLSADVVP	TDYRDLYSSE	780
ADCLARLRPF	IEQGADITNW	SDLFVQSDKH	SPVIHANIEL	SVKYGTTKLL	EQIINKDTQF	840
KTTEANFTAW	NTAQKSIEQL	IKQREDHHEQ	WVKAKNADDK	EKQERKREKS	NFAQKFIEKH	900
GDDYLDICDY	INTYNWLDRN	MHFVHLNRLLH	GLTIELLGRM	AGFVALFDRD	FQFFDEQQIA	960
DEFKLHGFBV	LHSIDKKLNE	VPTKKIKEIY	DIRNKIIQIN	GNKINESVRA	NLIQFISSSKR	1020
NYYNNNAFLHV	SNDEIKEKQM	YDIRNHIAHF	NYLTKDAADF	SLIDLINELR	ELLHYDRKLK	1080
NAVSKAFIDL	FDKHMILKL	KLNADHKLKV	ESLEPKKIYH	LGSSAKDKPE	YQYCTNQVMM	1140
AYCNMCRSSL	EMKK					1154

---

SEQ ID NO: 450	moltype = AA length = 1182
FEATURE	Location/Qualifiers
source	1..1182
	mol_type = protein
	organism = Leptotrichia wadei

SEQUENCE: 450		
MYMKITKIDG VSHYKKQDKG	ILKKWKDLD ERKQREKIEA RYNQKIESKI YKEFFRLKNK	60
KRIEKEEDQN IKSPLYFFIKE	LYLNEKNEEW ELKNINLEI DDKERVIGY KFKEDVYFFK	120
EGYKEYYLRI LFNNNLIEKVQ	NENREKVRKN KEFLDLKBEIF KKYKRNKIDL LLKSIINNNKI	180
NLEYKKENVN EEIYGINPTN	DREMTFYELL KEIEKKDDEQ KSLIEEKLDDN FDITNLFLENI	240
EKIFNEETEI NNIKGKVLNE	LREYIKEEE NNNSDNKLQOI YNLELKKYIE NNFSYKKQKS	300
KSKNGKNDYL YLNFLKKIMF	IEEVDEKKEI NKEKFKNKN SNFKNLFVQH ILDGKLLYY	360
KENDEYIKNT COLETKDLEY	IKTKETLIRK MAVLVSAAN SYYNLFCRVS GDILGTEVVK	420
SKSTNVKVG SHIFKEKMLN	YFFSDTEIFDA NKIVEILESTI SYSIYNRNG VGHFNKLILG	480
KYKKKDINTN KRIEEDLNNN	EEIKGYFIKK RGEIERKVKE KPLSNNLQQYY YSKEKIENYF	540
EVYEFEILKR KIPFAPNFKR	IIKKGEDLFN NKMNKKYEYF KNFDKNSAEE KKEFLKTRNF	600
LILKELYNNNF YKEFLSKKKE	FEKIVLEVKE EKKSRGNINN KKSGVFSQSI DDYDTKINIS	660
DIYASIIHKKE MERVEKYNEE	KQKDTAKYIR DFVVEEFLTG FINYLEKDKR LHFLKEEFSI	720
LCNNNNNNVND PNININEEKI	KEFLKENDSK TLNLYLFFNM IDSKRISERF NELVKKQFT	780
KKRLDEEKEF LGIKIELEYET	LIEFVILTRE KLDTKSEEI DAWLVDKLYV KDSNEYKEYE	840
EILKLFVDEK ILSSKEAPYY	ATDNKTPILL SNEFKTRKYQ TOSFLSEIQS NYKYSKVEKE	900
NIEDYNNKSNIE KIQKKSNSIE	KLQDNLKVELH KKWEQNKITE KEIEKYNNNT RKINEYNYLK	960
NKEELQNVYL LHEMLSDLLA	RNVAFPNKWE RDPKFIVIAI QKFLRENDEKE KVNEFLNPPD	1020
NSKGKKVYFS VSKYKNTVEN	IDGIHKNFMN LIFLNNKFMN RKIDKMNCAI WVYFRNYIAH	1080
FLHLHTKNEK ISLISQMNL	IKLFSYDKKV QNHILKSTKT LLEKYNIQIN FEISNDKNEV	1140
SKYKIKNRLY SKKGKMLGKN	NKFEILENEF LENVKAMLEY SE	1182

SEQ ID NO: 451	moltype = AA length = 1224
FEATURE	Location/Qualifiers
source	1..1224
	mol_type = protein
	organism = Bergeyella zoohelcum

SEQUENCE: 451		
MENKTSGLNN IYYNPFKPQD	KSYFAGYFNA AMENTDSVFR ELGKRLKGKE YTSENFFDAI	60
FKENISLVEY ERYVKLLSDY	FPMARLDDKK EVPIKERKEN FKKNFKGIK AVRDLRNFT	120
HKEHGEVEIT DEIFGVLDDEM	LKSTVLTVKK KVVKTDKTKTKE ILKKSKTEKQL DILCQKKLEY	180
LRDTARKIEE KRRNQRERGE	KELVAPFKYS DKRDDLIAAI YNDAFDVYID KKKDSLKESS	240
KAKYNTKSDP QOEQGDLKIP	ISKNGVVPLL SLFLTKQBEIH AFKSKTIAFGK ATVIDEATVS	300
EATVSHGKNS ICFMATHEIF	SHLAYKLKLKV RTAENINGY EAENAEQLSV YAKETLMMQM	360
LDELSKVPDV VYQNLSEDVQ	TKTFIEDWNEY LKENNGDVGTE MEEEQVHHPV IRKRYEDKFN	420
YFAIRFLDEF AQFPTLRFQV	HLGNYLHDSR PKENLISDRR IKEKITVGR LSELEHKKAL	480
FIKNTETMED REHYWEIFPN	PNDYFPKPEI SVNDKDFPIA GSILDREKQP VAGKIGIKVK	540
LLNQQYVSEV DKAVALHQKLQ	QRKASKPSIQ NIIIEEVPIN ESNPKEAIVF GGQPTAYLSM	600
NDIHSILYEF FDKWEKKKE	LEKKGEKELR KEIGKELEKK IVGKIAQOIQ QIIDKDTNAK	660
ILLPYQDGNS TAIDKEKLIK	DLKQEQNLIQ KLKDEQTVERE KEYNDFIAYQ DKNREINKVR	720
DRNHKQYLIKD NLKRKYPEAP	ARKEVLYYRE KGKVAVWLAI DIKRFMPTDF KNEWKGEQHS	780
LLQKSLAYYE QCKEELKNLL	PEKVFQHLPF KLGGYFQQKY LYQFYTCDL KRLEYISGLV	840
QQAEENFKSEN KVFKKVNENE	FKFLKKQNYT HKELDARVQS ILGYPFLER GFMDEKPTII	900
KGKTFKGNEA LFADWFERRYK	EYQNFQTFYD TENYPLVLE KKQADRKRKT KIYQQKKNDV	960
FTLMAKHIF KSVFKQDSID	QFSLEDLYQS REERLGNQER ARQTGERNTN YIWNKTVDLK	1020
LCDGKITVEN VKLKNVGDFI	KYEYDQRVQA FLKYEEEN QAFLIKESKE EENYPYVER	1080
EIEQYEVKRV EELLKEVHLL	EEYILEKVKD KEILKKGDNQ NFKYIILNGL LKQLKNEDVE	1140
SYKVFNLNTE PEDVNIQNQLK	QEATDLEQKA FVLTYIRNKF AHNQLPKKEF WDYQCEKYKG	1200
IEKEKTYAAY FAEVFKEKE	ALIK	1224

SEQ ID NO: 452	moltype = AA length = 1126
FEATURE	Location/Qualifiers
source	1..1126
	mol_type = protein
	organism = Prevotella intermedia

SEQUENCE: 452		
MEDDKKTTDS IRYELDKHNF	WAAFLNLRH NVYITVNHN KILEEGEINR DGYETTLKNT	60
WNEIKDINKK DRSLKLIKH	FPPFLEATYR LNPTDTTQK EEKQAEQSL ESLRKSFFVF	120

-continued

IYKLRDLRNH YSHYKHSKSL ERPKFEEGLL EKMYNIFNAS IRLVKEDYQY NKDINPDED	180
KHLDRTSEEF NYFTKDNEG NITESGLLF VSLFLEKKDA IWMQQKLRGF KDNRENKKM	240
TNEVFCRSRM LLPKLRQST QTQDWILLDM LNELIRCPKS LYERLREEDR EKFRVPIEIA	300
DEDYDABQEPE FKNTLVRHQI RFPYFALRYF DYNEIFTMLR FQIDLGYHF SIYKKQIGDY	360
KESHHLTHKL YGFERIQEFT KQNRPDEWRK FVKTFSNFET SKEPYIPETT PHYHLENQKI	420
GTRFRNDNDK IWPSLKTNSK KNEKSKYKLD KSFQAEAFPLS VHLLPPMMFY YLLLKTENTD	480
NDNEIETKKK ENKNDKQEKH KIEEIIENKI TEIYALYDTF ANGEIKSIDE LEYCKGKDI	540
EIGHLPQOMI AILKDEHKVN ATAEAERKQEE MLVDVQKSLE SLDNQINEEI ENVERKNSSL	600
KSGKIASWLV NDMMRFQPVQ KDNEGKPLNN SKANSTEYQL LQRTLAFFGS EHERLAPYFK	660
QTKLIESNPW FPLKTDTEWE KCNNLISFYR SYLEAKKNFL ESLKPEDWEI NQYFLKLKEP	720
TKTPKTLVQG WNGNFGNLPRG IFTEPKWF MKHRENITVA ELKRVGLVAK VIPLFFSEYY	780
KDSVQPFNY HFNVGNINKP DEKNFLNCEE RRELLRKDKD EFKKMTDKEV EENPSYLEFK	840
SWNKFERELR LVRNQDIVTW LLCMELFNKK KIKELNVEKI YLKNINTNTT KKEKNTEKN	900
GEENKIKEKN NIILNRMIMPME LPIKVYGEN FSKNKKKKIR RNTFFTVYIE EKGTKLLKQG	960
NFKALERDRR LGGLLFSFVKT PSKAESKSNT ISKLRVEYEL GEYQKARIEI IKDMLALEKT	1020
LIDKYNSLDT DNFNKMMLTDW LELKGEPDKA SFONDV DLLI AVRNAFSHNQ YPMRNRIAF	1080
NIINPFSLSSA NTSEEKGLGI ANQLKDKTHK TIEKIIIEIEK PIETKE	1126

SEQ ID NO: 453 moltype = AA length = 1127  
 FEATURE Location/Qualifiers  
 source 1..1127  
 mol\_type = protein  
 organism = Prevotella buccae

SEQUENCE: 453  
 MQKQDKLFLVD RKKNAIFAFP KYITIMENKE KPEPIYYEELT DKHFWAAFLN LARHNVYTTI 60  
 NHINRRLEIA ELKDDGYMMG IKGSWNEQAK KLDDKVRLRD LIMKHFFLE AAAYEMTNK 120  
 SPNPKEQREK EQSEALSNN LKNVLFIFLE KLQVLRNYYH HYKYSEESPK PIFETSLLN 180  
 MYKVF DANVR LVKRDYMHHE NIDMQRDFTH LNRRKKQVGRH KNIIDSPNHF YHFADKEGNM 240  
 TIAGLFFFVS LFLLDKKDAIW MQKKLGKFD GRNLREQMTN EVFCRSRISL PKLKLENVQT 300  
 KDWQMQLDMN ELVRCPSLW ERLREKDRS FKVPFDIISD DYNAEEEPFK NTLVRHQDRF 360  
 PYFVLRYFLRQD NEIFEQLRFQ IDLGTYHFSR YNKRIGDEDE VRHLTHHLYG FARIQDFAPQ 420  
 NQPEEWKLV KDLDHPTSQ EPYISKATPH YHLENEKIGI KFCSAHNNLF PSLQTDKTCN 480  
 GRSKFNLGTO FTAEAFLSVH ELLPPMFYYL LLTKDYSRKE SADKVEGIIR KEISNIYAIY 540  
 DAFANNEINS IADLTRLQN TNILQGHLPK QMISILKGRQ KDMGKEAERK IGEMIDDTQR 600  
 RLDDLCKQTN QKIRIGKRNA GLLKSGKIAW WLVDNMMRFQ PVQKDQNNIP INNSKANSTE 660  
 YRMLQRALN FGSENFRFLKA YFNQMNVLGN DNPHPFLAET QWEHQTNILS FYRNYLEARK 720  
 KYLKGLKPQW WKQYQHFLIL KVQKTNRNTL VTGWKNSFNL PRGIFTQPIR EWFEKHNNSK 780  
 RIYDQILSFD RVGFVAKAIP LYFAEYKDQ VQPFYDYPN IGNRLPKKQ QFLDKKERVE 840  
 LWQKNKELFK NYPSEKKKTQ LAYLDLFLSWK KFERELRLKI NQDIVTWLMF KELFNMATVE 900  
 GLKIGEIHRL DIDTNTANEE SNNILNRMIP MKLKVKTYET DNKGNILKER PLATFYIET 960  
 ETKVLKQGNF KALVKDRLN GLFSFAETTD LNLEEHPIK LSVDELILKY QTTRISIFEM 1020  
 TGLEKLLID KYSTLPTDSF RNMLERWLQC KANRPELKMY VNSLIAVRNA FSHNQYPMYD 1080  
 ATLFAEVKKF TLFPSVDTKK IELNIAPQOLL EIVGKAIKEI EKSENKN 1127

SEQ ID NO: 454 moltype = AA length = 1135  
 FEATURE Location/Qualifiers  
 source 1..1135  
 mol\_type = protein  
 organism = Porphyromonas gingivalis

SEQUENCE: 454  
 MNTVPASENK QGSRTVEDDP QYFGGLYLNLA RENLIEVESR VRIKFGKKL NEESLKQSSL 60  
 CDHLLSVDRW TKVYGHSSRY LPFLHYFDPD SQIEKDHDSK TGVDPSAQR LIRELYSLLD 120  
 FLRNDFSHNR LDGTTFEHLE VSPDISSFIT GTYSLACGRA QSRFAVFFKP DDFVLAKNRK 180  
 EQLISVADGK ECLTVSGFAF FICLFLDREQ ASGMLSIRIG FKRTDENWAR AVHETFCDL 240  
 IRHPHDRLS SNTKEALLL MLNLENLRCPR ILYDMLPEEE RAQFLPALDE NSMNNLSENS 300  
 LDEESLRLWD GSSDWAEEALT KRIRHGDRF YLMLRFIEEM DLLKGIRFRV DLGEIELDSY 360  
 SKKVGGRNGEY DRTITDHALA FGKLSDFQNE EEVSRMSIGE ASYPVRFSLF APRYAIYDNK 420  
 IGYCHSTDV YPKSKTGEKR ALSNPQSMGF ISVHDLRKL LMELLCEGSF SRMQSDLRK 480  
 ANRILDTEA GKLQFSLALFP EMRHRFIPPCQ NPKSQDKRREK AETTLEKYKQ EIKGRKDKN 540  
 SQLLSAFDMD QRQLPSRLL EWMNRPASH SVKLRTYVKQ LNEDCRLRLR KFRKDGDGKA 600  
 RAIPLVGEMA TFLSDQDIVRM IISEETKKLI TSAYYNEMQR SLAQYAGEEN RRQFRAIVAE 660  
 LRLLDPSGGH PFLSATMETA HRYTEGFYKC YLEKKREWLA KIFYRPEQDE NTKRRISVFF 720  
 VPDGEAKRLL PLLIRRRMKE QNDLQDWIRN KQAHPIDLPS HLFDSKVMEL LKVKGDKKKW 780  
 NEAFKDWWS T KYPDGQPFY GLRRELNIHG KSVSYIPSDG KKfadcythl MEKTVRDKKR 840  
 ELRTAGKWPV PDLAADIKR FHRAVNREFE MLRLVQEDDR LMLMAINKMM TDREEDILPG 900  
 LKNIDSILDE ENQFSLAVHZ KVLEKEGEGG DNSLSLVPAT IEIKSKRKDW SKYIRYRDR 960  
 RVPGLMSHFP EHKTATLDEVK TLLGEYDRCR IKIFDWAFAL EGAIMSDRDL KPYLHESSR 1020  
 EGKSGEHTSL VKMLVEKKGC LTPDESQYLI LIRNKAHHQ FPCAAEMLI YRDVSAKVGS 1080  
 IEGSSAKDLP EGSSLVDSLW KKYEMIIRKI LPILDPEPNF FGKLLNNMSQ PINDL 1135

SEQ ID NO: 455 moltype = AA length = 1115  
 FEATURE Location/Qualifiers  
 source 1..1115  
 mol\_type = protein  
 organism = Bacteroides pyogenes

SEQUENCE: 455

-continued

---

MESIKNSQKS	TGKTLQKDPP	YEGLYLNMAL	LNRKVENVHII	RKWLGDVALL	PEKSGFHSSL	60
TTDNLSSAKW	TRFYYYRSRKF	LPFLEMFDSD	KKSYENRET	AECLDTIDRQ	KISSLLKEVY	120
GKLQDIRNAF	SHYHDDQSV	KHTALIISSE	MHRFIENAYS	FALQKTRARF	TGVFVETDFL	180
QAEEKGDNKK	FFAIGGNEGI	KLKDNLALFL	I CLFLDRREEA	FKFLSRATGF	KSTKEKGFLA	240
VRETFCALCC	RQPHERLLSV	NPREALLMDM	LNELNRCPDII	LFEMLDEKDQ	KSFLPLLGE	300
EQAHILENSL	NDELCEAIDD	PFEMIASLSK	RVRYKNRFPY	LMLRYIEEKI	LLPFIRFRID	360
LGCLELASYP	KKMGEENNYYE	RSVTDHAMAF	GRLTDFHNED	AVLQQITKGI	TDEVIRFSLYA	420
PRYAIYNNKI	GFVRTSGSDE	ISFPTLKKKG	GECHCVAATL	QNTKSGFVIS	IYDLRKILL	480
SFLDKDKAKN	IVSGLLEQCE	KHWKDLSENLL	FDAIRTELQK	EFPVPLIRYT	LPRSKGGKLV	540
SSKLADKQEY	YESEFERRKE	KLTTEILSEKD	FDLSPQIPRRM	IDEWLNVLPY	SREKKLKGYV	600
ETLKLDCREP	LRVFEKREKG	EHPPLPPRIGE	MATDLAKDII	RMVIDQGVKQ	RITSAYSEI	660
QRCLAOYAGD	DNRRHLDSII	RELRLKDTKN	GHPFLGKVLRI	PGLGHTEKLY	QRYFEEKREW	720
LEATFYPAAS	PKRVPVFVNP	PTGKQKELPL	IIRNLMKERP	EWRDWKQRKN	SHPIDLPSQL	780
FENEICRLLK	DKIGKEBPSGE	LKWNEMFKLY	WDKEFPNGM	RFYRCKRVE	VFDKVVEY	840
SEEGGNKYKY	REALIDEVVR	QKISSKEKS	KLQVEDLTL	VRRVFKRAIN	EKEYQLRLLC	900
EDDRLLFMVA	RDLYDWKEAQ	LDLKDIDNMII	GEPVSVSQVI	OLEGGQPDAV	IKAECKLKV	960
SKLMRMYCYDG	RVKGLMPYFA	NHEATQEQUE	MELRHEDHR	RRVFNWVFA	EKSVLKNEKL	1020
RRFYEEQSQQG	CCHRRCIDAL	RKASLVSEEEE	YEFLVHIRNK	SAHNQFPDLE	IGKLPPNVT	1080
GFCECIWSKY	KAIICRIIPF	IDPERRFFGK	LLEQK			1115

SEQ ID NO: 456	moltype = AA length = 668					
FEATURE	Location/Qualifiers					
source	1..668					
	mol_type = protein					
	organism = unidentified					
SEQUENCE: 456						
METEKKSIIFK	NKSSVEIVKK	DIFSQTPDNM	IRNYKITLKI	SEKNPVRVVA	EIEDELMNSTI	60
LKDGRSARR	EKSMTERKLI	EEKVAENYSL	LANCPMEEVD	SIKIYKIKRF	LTYRSNMLLY	120
FASINSFLCE	GIKGKDNETE	EIWHLKDNV	RKEVKVENFK	NKLIQSTENY	NSSLKNQIEE	180
KEKLLRKESK	KGAFYRTI	KLQKQERIKEI	SEKSLTEDCE	KIILKLYSEL	HPLMHYDYQY	240
FENLKLRENK	IFKSLPLV	RKVNLY	EDNDTLEFLVQ	KTKKAKTLYQ	300	
TYDALCQKQN	GPNKFINDFF	VSDGEENTV	KQIINEKPOS	EMEFLKRIS	ESEKKEKNEKL	360
KKFDMSKAHF	HNIINSEDTKE	AYFWDIHSSS	NYKTKYNERK	NLVNEYTELL	GSSKEKKLLR	420
EETTQINRKL	LKLKQEMEEI	TKKNSLFLRLE	YKMKIAFGFL	CECFDGNI	FKDEFDASNQ	480
EKIIQYHKNG	EKYLTYFLKE	EEKECFNKL	MOKIIQKTEE	EDWLLPETKN	NLFKFYLL	540
LLLPYELKGD	FLGFVKHHY	DIKVNDFMD	NQNNIQVSQT	VEKQEDFYH	KIRLFENKNT	600
KYEIVVKYSIV	PNEKLKQYFE	DLGIDIKYLT	GSVESGEKWL	GENLGIDIY	LTVEQKSEVS	660
EEKIKKFL						668

SEQ ID NO: 457	moltype = AA length = 796					
FEATURE	Location/Qualifiers					
source	1..796					
	mol_type = protein					
	organism = unidentified					
SEQUENCE: 457						
MEKDKKGEKI	DISQEMIEED	LRKILILFSR	LRHSMVHYD	EFYQALYSGK	DFVISDKNNL	60
ENRMISQLD	LNIFKELSKV	KLIKDKAISN	YLDKNTTIH	LGQDIAKIRL	LDIYRDIICGS	120
KNGFNKPINT	MITISGEEDR	EYKEKVIEHF	NKKMENLSTY	LEKLEQDNA	KRNNKRVYNL	180
LQKQKLIBEQQK	LKEWFCCPYV	YDIHSSKRYK	ELYIERKKLV	DRHSKLFEEG	LDEKNKKELT	240
KINDELSKLN	SEMKEMTKLN	SKYRLQYKLQ	LAFLFILEEF	DLNIDTFINN	FDKDKDLIIS	300
NFMKKRDIYL	NRVLDRGDNM	LKNKIIKEYKF	RDTEDIFCND	RDNNLVKLYI	LMYILLPVEI	360
RGDFLGFKVKK	NYYDMKHVDF	IDKKDKEDM	TFFHDLRLF	KNIRKLEITD	YSLSSGFLSK	420
EHKVDIEKKI	NDFINRNGAM	KLPDITIEE	FNKSLLIPIM	KNYQINFKLL	NDIEISALFK	480
IAKDRSITFK	QAIDEIKNED	IKKNSKKNDK	NNHKDKNINF	TQLMKRALHE	KIPYKAGMYQ	540
IRMNISHIDM	BOLYIDPLN	YMNMSKNNIT	ISBQIEKI	VCVTGGVTGK	ELNNNNIINDY	600
YMKKEKLFVN	LKLRLQKNDIV	SIESQEKNN	EEFVFKKYGL	DYDKGEINII	EVIQVNSLQ	660
EELRNKETS	KEKLNKNETL	FRDISLINGT	IRKNINFKIK	EMVLDIVRMD	EIRHINIHIY	720
YKGENYTRSN	IIKFKYAIDG	ENKKYYLQH	EINDINLEK	DKFVTLCNM	DKHPNKNQ	780
INLESNYIQN	VKFIIIP					796

SEQ ID NO: 458	moltype = AA length = 897
FEATURE	Location/Qualifiers
source	1..897
	mol_type = protein
	organism = unidentified
SEQUENCE: 458	

MENKGNNKKI	DFDENYNILV	AQIKEYFTKE	IENYNNRIDN	IIDKKELLKY	SEKKEESEKN	60
KKLEELNKLK	SQKLKILTDE	EIKADVKII	KIPSDLRHSL	MHYEYKYPEN	LFENKNEEL	120
AELLNLNLFK	NLTLLRQMKI	ENKTNYLEGR	EEFNIIIGKNI	KAKEVIGHYN	LLAEQKNGFN	180
NFINSFFVQD	GTCENLEFKKL	IDEHFVNAKK	RLERNIKKSK	KLEKELEKME	QHYQRLNCAY	240
VWDIHTSTTY	KKLYNKRKSL	IEEYKNQINE	IKDKEVITAI	NVELLRKKE	MEETIKSNSL	300
FRLKYKMQIA	YAFLEIEFGG	NIAKFKDEF	CSKMEEVQKY	LKKGVKYLKY	YKDKEAQKNY	360
EFFPFEEIFEN	KDTHNEEWLE	NTSENNLKF	YILTYLLPM	EFKGDFLGVV	KKHYDDIKNV	420
DFTDESEKEL	SQVQLDKMIG	DSFFFHKirLF	EKNTKRYEII	KYSILTSDEI	KRYFRLLELD	480
VPYFYEKGT	DEIGIFNKNI	ILTIFKYYQI	IFRLYNDLEI	HGLFNISSSL	DKILRDLKSY	540
GNKNINPREF	LYVIKQNNNS	STEEEYRKI	ENLEAKYRL	HLLTPEKEEI	KTKTKEELEK	600

-continued

---

LNEISNLRNG	ICHILNYKEII	EEILKTEISE	KNKEATLNEK	IRKVINFIKE	NELDKVELGF	660
NFINDFFMVK	BQFMFGQIQQ	VKEGNSDSIT	TERERKEEKN	KKLKETYELN	CDNLSEFYET	720
SNNLRERANS	SSLLEDSAFL	KIGLYVKVN	NKVNSKVKDE	EKRNIKRK	LLKDSSDIMG	780
MYKAEVVKKL	KEKLILIFKH	DEEKRIYVT	YDTSKAVPEN	ISKEILVCRN	NSKEEYFFED	840
NNKKYVTEYY	TLEITETNEL	KVIPAKKLEG	KEFKTEKNKE	NKLMNNHYC	FNVKIIY	897

SEQ ID NO:	459	moltype = AA	length = 919
FEATURE		Location/Qualifiers	
source		1..919	
		mol_type = protein	
		organism = unidentified	

SEQUENCE:	459					
MEEIKHHKNK	SSIIRVIVSN	YDMTGIKEIK	VLYQKQGGVD	TFNLKTIINL	ESGNLEIISC	60
KPKEREKRY	EFNCKTEINT	ISITPKDKVL	KBEIRKYSLE	LYFKNEKKDT	VVAKVTDLK	120
APDKIEGERN	HLRKLSSSTE	RKLLSKTLCK	NYSEISKTP	EEIDSISIYK	IKRFLNYSRN	180
FLIYFALIND	PLCAGVKEDD	INEWVLIQDK	EHTAFLENRI	EKITDYIFDK	LSKDIENKKN	240
QFEKRIKKYK	TSLEELKTET	LEKNKTFYID	SIKTKITNL	NKITELESLYN	SKESLKEDLI	300
KIISIFTNLR	HSLMHYDYKS	FENLFENIEN	EELKNLLDLN	LFKSIRMSDE	FKTKNRTNYL	360
DGTESFTIVK	HKQNLKLYT	YYNNLDCDKK	GFPNTFINSFF	VTDGIENTDF	KNLJILHFEK	420
EMBEYKKSIE	YKIKISNEK	NKSKEKELKE	KIDLLOSELII	NMREHKNLK	QIYFFDIHNS	480
IKYKELYSER	KNLIEQYNLQ	INGVKDVTAI	NHINTKLLSL	KNKMDKITQ	NSYRLKYKL	540
KIAYSFMLIE	FDGDVSFKPN	NFDPNTLEKR	VEYLDKKEBY	LNYTAPKNKF	NFAKLEELQ	600
KIQSTSEMGA	DYLNVPENN	LFKFYILTYI	MLPVEFKGDF	LGFKVNHYYN	IKNVDFMDES	660
LLDENEVDSEN	KLNEKIEENLK	DSSFFNPKIRL	FEKNIKKYEI	VKYSVSTQEN	MKEYPKQLNL	720
DIPYLDYKST	DEIGIFNKNM	ILPIFKYYQN	VFKLCNDIEI	HALLALANKK	QQNLEYAIYC	780
CSKKNSLNLYN	ELLKTPNRK	YQNLSFIRNK	IAHLNYKELF	SDLFNNELDL	NTKVRCLIEF	840
SQNNKFDQID	LGMNFINDYY	MKKTRFIFNQ	RRLRDLNVPS	KEKIIDGKRK	QONDNNELL	900
KKYGLSRTNI	KDIFNKA					919

SEQ ID NO:	460	moltype = AA	length = 1110
FEATURE		Location/Qualifiers	
source		1..1110	
		mol_type = protein	
		organism = unidentified	

SEQUENCE:	460					
MVKRYRKQAQ	LDTFIKKTEI	VNNNDIFIksi	IIEKAREKRY	SFLFDGEEKY	HFKNKSSVEI	60
VKNIDFSQTP	DNMIRNYKIT	LKISEKNPRV	VEAEIEDLMN	STILKDGRS	ARREKSMT	120
KLIEEKVAEN	YSLLANCPIE	EVDSIKIYKI	KRFLTYRSNN	LLYFASINSF	LCEGIKGKD	180
ETEEIWHLKD	NDVRKEKV	NFKNKLQIQT	ENYNSSLKNQ	IEEKEKLSSK	EFKKGAFYRT	240
IIKKLQQPERI	KELSEKS	DCEKIKLSS	ELRHPMLHYD	YQYFENLFEN	KENSELT	300
NLDIFKSLPL	VRKMKLNK	NYLEDNDTLF	VLQKTKKAKT	LYQIYDALCE	QKNGPNKFIN	360
DFPVSDGEEN	TVFKQIINEK	FQSEMEFLEK	RISESEKKNE	KLKKKLD	SMK AHFRNINSED	420
TKEAYFWDIH	SSRNYKTKY	ERKNLVNEYT	KLLGSSKEKK	LLREEITKIN	RQLLKLQEM	480
EEITKKNSLF	RLEYKMKIAF	GFLFCFEDGN	ISKFKDEFDA	SNQEKKIYQH	KNGEKYLTSF	540
LKEEEEKEFKN	LEKMQKIIQK	TEEEDWLPE	TKNNLFKFYL	LTYLLL	PYEL KGDFLGFVKK	600
HYDDIKVNDF	MDENQNNI	QSTVEKQEDY	FYHKIRLFEK	NTKKYEIVKY	SIVPNEKLQ	660
YFEDGLIDIK	YLTGVESGE	KWLGENLQID	IKYLTVEQKS	EVSEEKNKVV	SLKNNGMFNK	720
TILLFVFKYY	QIAFKLNDI	ELYSLEFLRE	KSEKPFFEV	EELKDKMIGK	QLNPGQLYV	780
VYEVVLVKNKD	LDKILSKKID	YRKDKSFSP	IAYLRFNLSH	LNYSKFLDNF	MKINTNKSDE	840
NKEVLIPS	IQKMIQFIEK	CNLQNQIDFD	FNFVNDFYMR	KEKMFIIQLK	QIFPDINST	900
KQKKSKEE	LRKRYHLINK	KNEQIKDEHE	AQSQLYEKIL	SLQKIFSCDK	NNFYRRLKEE	960
KLLFLEKQGK	KKISMKEIKD	KIASDSDL	GILKKEITRD	IKDKLTEKFR	YCEEKLLNIS	1020
FYNHQDKKKE	EGIRVFLIRD	KNSDNFKFES	ILDDGSNKF	ISKNGKEITI	QCCKVLET	1080
MIEKNTLKIS	SNGKIIISLIP	HYSYSIDV	KY			1110

SEQ ID NO:	461	moltype =	length =
SEQUENCE:	461		
000			

SEQ ID NO:	462	moltype =	length =
SEQUENCE:	462		
000			

SEQ ID NO:	463	moltype = RNA	length = 10
FEATURE		Location/Qualifiers	
source		1..10	
		mol_type = other RNA	
		organism = synthetic construct	

SEQUENCE:	463		
ttttttttt			10

SEQ ID NO:	464	moltype = DNA	length = 10
FEATURE		Location/Qualifiers	
source		1..10	
		mol_type = other DNA	
		organism = synthetic construct	

---

-continued

---

```

misc_feature      1..4
                  note = DNA
misc_feature      5..6
                  note = RNA
misc_feature      7..10
                  note = DNA
SEQUENCE: 464
ttttttttt          10

SEQ ID NO: 465      moltype = length =
SEQUENCE: 465
000

SEQ ID NO: 466      moltype = length =
SEQUENCE: 466
000

SEQ ID NO: 467      moltype = length =
SEQUENCE: 467
000

SEQ ID NO: 468      moltype = length =
SEQUENCE: 468
000

SEQ ID NO: 469      moltype = length =
SEQUENCE: 469
000

SEQ ID NO: 470      moltype = length =
SEQUENCE: 470
000

SEQ ID NO: 471      moltype = length =
SEQUENCE: 471
000

SEQ ID NO: 472      moltype = DNA length = 10
FEATURE
source
1..10
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 472
ttttttttt          10

SEQ ID NO: 473      moltype = DNA length = 12
FEATURE
source
1..12
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 473
ttttttttt tt          12

SEQ ID NO: 474      moltype = DNA length = 14
FEATURE
source
1..14
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 474
ttttttttt tttt          14

SEQ ID NO: 475      moltype = length =
SEQUENCE: 475
000

SEQ ID NO: 476      moltype = length =
SEQUENCE: 476
000

```

---

-continued

---

```

SEQ ID NO: 477      moltype = length =
SEQUENCE: 477
000

SEQ ID NO: 478      moltype = AA length = 18
FEATURE          Location/Qualifiers
source           1..18
                  mol_type = protein
SEQUENCE: 478      organism = unidentified
KRPAATKKAG QAKKKKEF

```

18

What is claimed is:

1. A method of quantitating a target nucleic acid in a test sample, the method comprising:
  - (a) contacting the test sample with the following components (i) through (iii), resulting in at least two nanovolumes of reaction mixture:
    - (i) a CRISPR/Cas effector protein,
    - (ii) a guide RNA, comprising a region that is capable of binding to the CRISPR/Cas effector protein, and a guide sequence that is capable of hybridizing with the target nucleic acid,
    - (iii) a reporter nucleic acid that is single stranded and does not hybridize with the guide sequence of the guide RNA;
  - (b) measuring detectable signals detected from the at least two nanovolumes and generated by cleavage of the reporter nucleic acid by the CRISPR/Cas effector protein, and
  - (c) quantitating the target nucleic acid in the test sample based on the measured signals from the at least two nanovolumes.
2. The method of claim 1, wherein the contacting step comprises sequentially adding each of the components (i) through (iii) to the test sample in the at least two nanovolumes.
3. The method of claim 1, wherein the contacting step comprises: (a) adding each of the components (i) through (iii) to the test sample to generate a master reaction mixture, wherein the master reaction mixture has a volume of more than 1 nL, and (b) distributing the master reaction mixture into the at least two nanovolumes.
4. The method of any one of claims 1-3, wherein each of the at least two nanovolumes comprises no more than 1 molecule of the target nucleic acid.
5. The method of any one of claims 1-4, wherein the CRISPR/Cas effector protein and the guide RNA are incubated with each other prior to step (a).
6. The method of any one of claims 1-4, wherein the CRISPR/Cas effector protein and the guide RNA are not incubated with each other prior to step (a).
7. The method of any one of claims 1-6, further comprising quantitating multiple target nucleic acids in the test sample, wherein the at least two nanovolumes of reaction mixture comprises one or more guide RNAs, wherein at least one of the one or more guide RNAs comprises a guide sequence that is capable of hybridizing with each of the multiple target nucleic acids.
8. The method of claim 7, wherein each of the at least two nanovolumes of reaction mixture comprises at least one of

the one or more guide RNAs comprising a guide sequence that is capable of hybridizing with each of the multiple target nucleic acids.

9. The method of claim 7, wherein each of the at least two nanovolumes of reaction mixture comprises at least one of the one or more guide RNAs comprising a guide sequence that is capable of hybridizing with no more than one of the multiple target nucleic acids.

10. The method of any one of claims 1-9, wherein the method does not comprise amplifying the target nucleic acid in the test sample.

11. The method of any one of claims 1-10, wherein the test sample comprises about 10,000 molecules to about 100,000 molecules of the target nucleic acid.

12. The method of claim 11, wherein the test sample comprises about 50,000 molecules of the target nucleic acid.

13. The method of any one of claims 1-12, wherein the contacting step results in a number of nanovolumes in the range of about 5000 nanovolumes to about 100,000 nanovolumes.

14. The method of any one of claims 1-13, wherein step (b) comprises measuring a binary signal from each of the at least two nanovolumes.

15. The method of any one of claims 1-14, comprising contacting the test sample with a precursor guide RNA array, wherein the CRISPR/Cas effector protein cleaves the precursor guide RNA array to produce the guide RNA.

16. The method of any one of claims 1-15, wherein the target nucleic acid is DNA or RNA.

17. The method of any one of claims 1-16, wherein the target nucleic acid is a viral nucleic acid or a bacterial nucleic acid.

18. The method of claim 17, wherein the target nucleic acid is a viral nucleic acid.

19. The method of claim 18, wherein the target nucleic acid is derived from a papovavirus, a human papillomavirus (HPV), a hepadnavirus, a Hepatitis B Virus (HBV), a herpesvirus, a varicella zoster virus (VZV), an Epstein Barr virus (EBV), a Kaposi's sarcoma-associated herpesvirus, an adenovirus, a poxvirus, a parvovirus, an influenza virus, a respiratory syncytial virus, an orthopoxvirus, or a coronaviruses.

20. The method of claim 19, wherein the target nucleic acid is derived from a severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2).

21. The method of any one of claims 1-20, wherein the target nucleic acid is derived from a human cell.

22. The method of any one of claims 1-21, wherein the target nucleic acid is a human fetal nucleic acid or a cancer cell nucleic acid.

- 23.** The method of any one of claims **1-22**, wherein the target nucleic acid is single stranded.
- 24.** The method of any one of claims **1-22**, wherein the target nucleic acid is double stranded.
- 25.** The method of any one of claims **1-24**, wherein the test sample comprises DNA or RNA from a cell lysate.
- 26.** The method of any one of claims **1-25**, wherein the test sample comprises cells.
- 27.** The method of any one of claims **1-26**, wherein the test sample is a blood, serum, plasma, urine, aspirate, fecal, wastewater, or biopsy sample.
- 28.** The method of any one of claims **1-27**, further comprising quantitating a positive control target nucleic acid in a positive control sample, the method comprising:
- (a) contacting the positive control sample with the following components (i) through (iii) resulting in at least two nanovolumes of reaction mixture:
    - (i) a CRISPR/Cas effector protein;
    - (ii) a positive control guide RNA, comprising a region that is capable of binding to the CRISPR/Cas effector protein, and a guide sequence that is capable of hybridizing with the positive control target nucleic acid; and
    - (iii) a reporter nucleic acid that is single stranded and does not hybridize with the guide sequence of the positive control guide RNA;
  - (b) measuring detectable signals detected from the at least two nanovolumes and produced by cleavage of the reporter nucleic acid by the CRISPR/Cas effector protein, and
  - (c) quantitating the amount of positive control target nucleic acid in the positive control sample based on the measured signals from the at least two nanovolumes.
- 29.** The method of claim **28**, wherein the at least two nanovolumes comprises more than one type of CRISPR/Cas effector protein, and one or more positive control guide RNAs capable of binding to each of the more than one type of CRISPR/Cas effector protein.
- 30.** The method of claim **29**, wherein the at least two nanovolumes comprises a Cas12 protein and a Cas13 protein, a positive control guide RNA comprising a region that is capable of binding to Cas12 protein and a positive control guide RNA comprising a region that is capable of binding to a Cas13 protein.
- 31.** The method of any one of claims **1-30**, further comprising quantitating the target nucleic acid in a positive control sample, the method comprising:
- (a) contacting the positive control sample with the following components (i) through (iii) resulting in at least two nanovolumes of reaction mixture:
    - (i) a CRISPR/Cas effector protein;
    - (ii) the guide RNA, comprising a region that is capable of binding to the CRISPR/Cas effector protein, and a guide sequence that is capable of hybridizing with the target nucleic acid; and
    - (iii) a reporter nucleic acid that is single stranded and does not hybridize with the guide sequence of the guide RNA;
  - (b) measuring detectable signals detected from the at least two nanovolumes produced by cleavage of the reporter nucleic acid by the CRISPR/Cas effector protein, and
  - (c) quantitating the amount of target nucleic acid in the positive control sample based on the measured signals from the at least two nanovolumes.
- 32.** The method of claim **31**, wherein the method comprises generating a standard curve for the target nucleic acid in the positive control sample, and obtaining an absolute quantitation of the target nucleic acid in the test sample based on the standard curve.
- 33.** The method of claim **31**, wherein the method comprises obtaining a relative quantitation of the target nucleic acid in the test sample based on the quantitation of the target nucleic acid in a positive control sample.
- 34.** The method of any one of claims **1-33**, wherein the detectable signal is detectable in less than 90 minutes.
- 35.** The method of any one of claims **1-34**, wherein the detectable signal is detectable in less than 30 minutes.
- 36.** The method of any one of claims **1-35**, wherein the CRISPR/Cas effector protein comprises an amino acid sequence with at least 90% identity to the amino acid sequence of any one of SEQ ID NOs: 1-72.
- 37.** The method of any one of claims **1-36**, wherein the CRISPR/Cas effector protein comprises the amino acid sequence of any one of SEQ ID NOs: 1-72.
- 38.** The method of any one of claims **1-37**, wherein target nucleic acid is an RNA.
- 39.** The method of claim **38**, wherein the CRISPR/Cas effector protein is an RNA-targeting CRISPR/Cas effector protein.
- 40.** The method of claim **39**, wherein the RNA-targeting CRISPR/Cas effector protein comprises an amino acid sequence with at least 90% identity to the amino acid sequence of SEQ ID NO: 21, 62, 43, 41, or 42.
- 41.** The method of claim **40**, wherein the RNA-targeting CRISPR/Cas effector protein comprises the amino acid sequence of SEQ ID NO: 21, 62, 43, 41, or 42.
- 42.** The method of any one of claims **1-37**, wherein target nucleic acid is a DNA.
- 43.** The method of claim **42**, wherein the CRISPR/Cas effector protein is a DNA-targeting CRISPR/Cas effector protein.
- 44.** The method of claim **43**, wherein the DNA-targeting CRISPR/Cas effector protein comprises an amino acid sequence with at least 90% identity to the amino acid sequence of SEQ ID NO: 3, 34, 57, 36, 65, 67, 68, 89, 90, 91, or 17.
- 45.** The method of claim **44**, wherein the DNA-targeting CRISPR/Cas effector protein comprises the amino acid sequence of SEQ ID NO: 3, 34, 57, 36, 65, 67, 68, 89, 90, 91, or 17.
- 46.** The method of any one of claims **1-45**, wherein the reaction mixture comprises a buffer, wherein the buffer comprises tricine, MgOAc, BSA, TCEP, imidazole, KCl, MgCl<sub>2</sub>, BSA, Igepal Ca-630, glycerol, HEPES, KOAc, Triton-X 100, Tris-HCl, (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, Tween-20, TMAO, or any combination thereof.
- 47.** The method of any one of claims **1-46**, wherein the reporter nuclear acid is a RNA.
- 48.** The method of any one of claims **1-46**, wherein the reporter nuclear acid is a DNA.
- 49.** The method of any one of claims **1-48**, wherein the reporter nucleic acid comprises a modified nucleobase, a modified sugar moiety, and/or a modified nucleic acid linkage.
- 50.** A method of assaying for a target nucleic acid in a sample, the method comprising:
  - a) amplifying the target nucleic acid using at least one amplification primer;

- b) contacting the sample to:
  - i. a reporter; and
  - ii. a composition comprising a programmable nuclease and a guide nucleic acid that hybridizes to the target nucleic acid or an amplified product thereof, wherein the programmable nuclease cleaves the reporter upon hybridization of the guide nucleic acid to the target nucleic acid or the amplification product thereof; and
- c) assaying for a change in a signal, wherein the change in the signal is produced by cleavage of the reporter; wherein the target nucleic acid is a gene of a monkeypox virus or a segment thereof; and optionally wherein the at least one amplification primer comprises a nucleotide sequence at least 85%, at least 87%, at least 90%, at least 92%, at least 94%, at least 95%, at least 96%, at least 98%, or 100% identical to any one of SEQ ID NOs: 92-235.

\* \* \* \*